MALDI-ToF Protein Profiling as Potential Rapid Diagnostic Platform for COVID-19

Prajkta Chivte[a#], Zane LaCasse[a#], Venkata Devesh R. Seethi[b], Pratool Bharti[b], Elizabeth R. Gaillard[a*]

Departments of Chemistry and Biochemistry[a] and Computer Science[b],

Northern Illinois University, DeKalb, IL 60115

#co-first authors

***corresponding author**

Abstract:

More than a year after the COVID-19 pandemic has been declared, the need still exists for accurate, rapid, inexpensive and non-invasive diagnostic methods that yield high specificity towards the current and newly emerging SARS-CoV-2 strains. Several studies have since established saliva as a more amenable specimen type in early detection and viral load quantitation as compared to the nasopharyngeal swabs. Considering the limitations and high demand for COVID-19 testing, we have employed MALDI-ToF mass spectrometry for the analysis of 60 gargle samples from human donors and compared the spectra with their respective RT-PCR results. Several standards including isolated human serum immunoglobulins and controls such as pre-COVID-19 saliva and heat inactivated SARS-CoV-2 virus were simultaneously analyzed to provide a relative view of the saliva and viral proteome as they would appear in this works methodology. Five potential biomarker peaks were established that demonstrated high consistency with PCR positive samples. Overall, the agreement of these results with RT-PCR testing was no less than 90% for the studied cohort, which consisted of young and largely asymptomatic student athletes. Further investigation of the potential biomarker peaks is necessary, however, from a clinical standpoint, these results are promising for a rapid and inexpensive COVID-19 assay.

Abbreviations

ACE2, angiotensin-converting enzyme 2; AUC, area under the curve; COVID-19, coronavirus disease 2019; DTT, dithiothreitol; E Protein, envelope protein; EUA, emergency use authorization; FDA, food and drug administration; IgA, immunoglobulin A; IgG, immunoglobulin G; IgM, immunoglobulin M; LCMS, liquid chromatography mass spectrometry; M Protein, membrane protein; MALDI-ToF MS, matrix-assisted laser desorption/ionization-time of flight mass spectrometry; N Protein, nucleocapsid protein; NP, nasopharyngeal; RBD, receptor binding domain; RNA, ribonucleic acid; RT-PCR, reverse transcriptase polymerase chain reaction; S

Protein, spike protein; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; VEP, viral envelope protein; WHO, world health organization

Keywords: COVID-19 testing, MALDI-ToF, asymptomatic, saliva, immunoglobulins

1. Introduction

Coronavirus disease 2019 (COVID-19), a highly transmissible disease caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus, was discovered in the Hubei Province of China in December 2019 and was soon declared a pandemic by the World Health Organization (WHO) in March 2020. Globally, there have been more than 123 million confirmed cases of COVID-19, including over 2.7 million deaths, as of March 2021 [1]. Clinical manifestations of COVID-19 predominantly include fever, dry cough, muscle pain, anosmia and fatigue [2]. Infections may also result in a number of medical complications, which worsen in patients of advanced age or with co-morbidities. Despite these symptoms and severe complications, evidence suggests that the most infections of SARS-CoV-2 are asymptomatic—such individuals can still transmit the virus as they possess the same infectivity as symptomatic patients [3].

Facing the ongoing pandemic caused by SARS-CoV-2, early diagnosis of COVID is of great importance for controlling the outbreak in communities and hospital. Thus, it has been crucial to design rapid, sensitive and accurate diagnostic tests to address this massive public health crisis. Broadly, there have emerged two major categories of laboratory diagnostic tests used for SARS-CoV-2: i) Detection of either the viral RNA genome or viral proteins and ii) measuring proteins associated with the host immune response to the virus [4].

The first type of test reports whether the virus is present at an abundance above the chosen threshold to minimize false positives. SARS-CoV-2 is an enveloped virus that encodes at least 29 proteins in its RNA genome, four of which are structural proteins: spike (S), membrane (M), envelope (E) and nucleocapsid (N) proteins. The S protein is responsible for binding to the cellular surface receptor angiotensin-converting enzyme 2 (ACE2) through the receptor binding domain (RBD), an essential step for membrane fusion. Activation of the S protein requires cleavage between the S1 and S2 subunits by a furin-like protease and subsequent conformational change [5]. The second type of test detects the presence of antibodies (mostly IgM and IgG) against SARS-CoV-2 within an individual who has been exposed to the virus. Studies have demonstrated that the strong antibody responses against the S and the N proteins are of high diagnostic utility [6].

The current gold standard test for COVID-19 is of the first type: molecular testing for the presence of SARS-CoV-2 RNA accomplished by reverse transcription polymerase chain reaction (RT-PCR). The sensitivity of RT-PCR is reported to be approximately 80% in practice while its specificity is near 95%. Occurrence of false negatives and/or false positives is commonly due to clinical laboratory errors involving steps such as sample storage, sample preparation, machine and/or operator error, and reporting errors. These test are usually carried out using specimens collected via a nasopharyngeal or oropharyngeal swab [4], [7], [8].

SARS-CoV-2 and the biomarkers of COVID-19 can be found in multiple other specimen types, however. These include tracheal aspirate, sputum, saliva, whole blood, plasma and serum [6]. Of particular interest in finding alternatives to nasopharyngeal 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 [9][10]. In fact, approval for using saliva as a primary test material for SARS-CoV-2 was first given to Rutgers' RUCDR Infinite Biologics and

collaborators under emergency use authorization (EUA) by the Food and Drug Administration (FDA) [9]. Similar saliva based tests were authorized for Yale School of Public Health and implemented by University of Illinois Urbana Champaign [11],[12].

The reliability of diagnostics involving salivary testing has been shown in the detection of SARS-CoV, Zika and Ebola [13]. 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 saliva is reported to be maximum within the first week of symptom onset, providing an avenue for early diagnosis of COVID-19 [14].

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. Molecular diagnosis of COVID-19 primarily depends on detection of viral RNA. However, RNA is extremely sensitive to degradation by ribonucleases and its extraction process is time-consuming, expensive and demands trained personnel as improper storage and extraction can be a major factor in contributing to false-negative results of COVID-19 [8]. Additionally, the RNA encoding the spike protein of SARS-CoV-2 has been undergoing mutations naturally since the time of its emergence [15]. This might prove to be a major challenge for tests that operate by detection of viral RNA.

On the other hand, proteins are more stable molecules that are present in higher amounts in the virus. Proteins cannot be directly amplified like nucleic acids which considerably reduces the chances of producing false positive results due to non-targeted amplification. Furthermore, several different proteins can be detected per 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 [8],[16]. Ultimately, proteomics combined with informatics can derive actionable information from large data sets and thus help in establishing biomarkers for this disease [17].

Developing a diagnostic test that overcomes the mentioned limitations would assist in fulfilling the worldwide demand of reliable, rapid and accurate testing. Iles *et al.* have published a preliminary report on utilizing water gargle samples to monitor COVID-19 associated proteins using MALDI-ToF mass spectrometry [18]. The current study proceeds in a similar direction that utilizes saliva for testing of COVID-19 through a simple gargle procedure followed by MALDI-ToF mass spectrometric analysis. However, in this study, the mass spectral analysis is directly compared to the RT-PCR status of the donor. The method described in this article is an amalgamation of both categories of COVID-19 testing: detection of viral proteins and the antibodies against them. For the 60 processed saliva samples that have been analyzed, the area under the curve (AUC) of viral protein and human immunoglobulin peaks have been used to suggest the presence of and response to SARS-CoV-2. The values for AUC of the biomarkers and viral proteins were compared to RT-PCR results acquired from nasopharyngeal swabs sampled in parallel.

2. Materials and Methods

2.1 Ethical and biosafety statement:

This work was approved by the Institutional Review Board and Institutional Biosafety Committee of Northern Illinois University (NIU) on August 12, 2020. 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-PCR results. Sample handling and processing followed all biosafety level guidelines.

2.2 Sample collection:

Samples were collected from a drive-thru gateway testing program for student athletes organized by NIU Athletics at the Yordon Center in DeKalb, Illinois between the months of August and September of 2020. Student athletes received a nasopharyngeal (NP) swab sample collection that was subjected to PCR testing through the DeKalb County Health Department. Students who consented to participate in the research study were asked to provide a water gargle sample at the same time as the NP swab sample collection. Subjects were asked to gargle 10 mL of bottled spring water for 30 seconds which was then deposited in a 50 mL conical centrifuge tube. Samples were stored at -20 °C until processing and analysis. A total of 550 gargle samples were collected to ensure a substantial pool of PCR-positive samples were included. For the current analysis, 60 (30 PCR positives + 30 PCR negatives) of the above total samples were processed and analyzed.

2.3 Preparation of samples and controls:

Gargle samples

Gargle samples were thawed and transferred to a 30 mL disposable polypropylene beaker (Fisher Scientific, Waltham, MA). Approximately 5 mL of each sample was filtered through a 0.45 µm polyethersulfone membrane filter (Fischer Scientific, Waltham, MA) 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 (Sigma-Aldrich, St. Louis, MO) to each tube. Samples were centrifuged in a Beckman Coulter Avanti J-E series centrifuge with a JA-20 rotor at 16,000 x g for 30 minutes at 4°C. The supernatant was disposed, the rim of the tube patted dry, and the pellet was resuspended using 100 µL of 1 M DTT (Sigma-Aldrich, St. Louis, MO) in LBSD-X buffer (MAPSciences, Bedford, UK), a proprietary membrane dissolution and viral envelope protein solubilization buffer. This buffer (reconstitution buffer) was prepared fresh daily by adding appropriate amounts of DTT solution before each analysis. 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 complete reconstitution, the samples were gently vortexed and incubated at room temperature for 15 minutes [18].

Standards and controls

Human serum antibodies IgA (I4036), IgG (I4506) and IgM(I8260) (Sigma-Aldrich, St. Louis, MO) were reduced with 1M DTT for 10 min to get a final concentration of 3 pmol of each antibody on the plate. Preparation of positive and negative controls for MALDI-ToF analysis followed closely to that of gargle samples. Human salivary α-amylase (A1031, Sigma-Aldrich, St. Louis, MO) was prepared at 100 pmol/µl in LC-MS grade water, 1 µl of which was spotted.The negative control, consisting of pooled human saliva (pre-COVID-19) collected before November 2019 (Lee Biosolutions, Maryland Heights, MO), was prepared by spiking 500 µL of the thawed stock to 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 (heat inactivated SARS-CoV-2) was obtained from BEI Resources (NR-52286). Briefly, this standard 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 x g for 15 minutes at 4 °C. The pellet was completely dissolved in 25 µL of the reconstitution buffer.

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 LCMS-grade water (OmniSolv, Sigma-Aldrich, St. Louis, MO) to acetonitrile (Oakwood Chemical, Estill, SC) solution containing 0.1% trifluoroacetic acid (Sigma-Aldrich, St. Louis, MO). First, 1 µL of sinapinic acid matrix was spotted in three wells of a 384 well stainless steel MALDI-MS sample plate (Shimadzu, Kyoto, Japan). After this air-dried, 1 µL of sample, control, or standard was spotted in each well immediately followed by 1 µL of sinapinic acid matrix. The matrix was made fresh every 7 days and stored at 4°C between analyses.

2.4 Data Acquisition:

Spectra acquisition was performed with a Shimadzu AXIMA Performance MALDI-ToF mass spectrometer (Shimadzu Kratos Analytical, Manchester, UK) equipped with a nitrogen laser set at 337.1 nm with a pulse width of 3 ns and maximum repetition rate of 60 Hz.

Instrument parameters

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.

Instrument calibration

The instrument was calibrated daily using the $(M+H)^+$ and $(M+2H)^{2+}$ peaks of ProteoMass Apomyoglobin MALDI-MS Standard (Sigma-Aldrich, St. Louis, MS), prepared at 100 pmol/µL in LCMS-grade water 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.

2.5 Data Analysis:

Shimadzu Biotech Launchpad was used to export a text file for each gargle sample that was 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 sub-ranges (shown in Table 1) were identified where m/z values were presumably indicative of host immune proteins or viral proteins. The intermittent values in each of the seven sub-ranges indicated the presence of similar protein masses, hence ion counts in each sub-range were coalesced together through integration of points to calculate the area under curve (AUC) which produced seven features for each data sample. AUC was computed by leveraging Composite Simpson's Rule for each spectral range [19]. Simpson's rule is a numerical integration method that divides the integral range into shorter sub-intervals (indicated by three consecutive data points) and uses a quadratic polynomial on each sub-interval to approximate the curve. If we define the width between each point as (h), the start and end of each range are x_0 and x_n respectively, where n is the total number of points, the width of each integral (h) can be calculated as shown in equation (1). One limitation of this approach is the requirement that h be uniform between each consecutive point. In the mass spectrometric data, protein peaks had two floating points precision which were unevenly spaced with different widths between two consecutive points. To make the widths uniform, each protein mass was rounded to the nearest integer and the mean of ion counts was taken at each integer value. This made the widths uniform with $h = 1$. If the points on the curve $\mathrm{x}_0, \mathrm{x}_1, \mathrm{x}_2, .. \mathrm{x}_\mathrm{n}$ have corresponding values $f(\mathrm{x}_0), \mathrm{f}(\mathrm{x}_1), f(x_2), \ldots f(x_n)$ respectively, then the area for the first sub-interval A_1 can be calculated using equation (2). Summing the areas for all sub-intervals i.e., $A_1, A_2, A_3, \ldots A_{\frac{n}{2}-1}$ gives the composite function to calculate AUC as shown in the equation (3). Since each curve was not smooth, lesser error can be achieved with Simpsons rule for integration as compared to Trapezoidal rule, which approximates the integral by forming trapezoids with straight lines in each sub-interval [20].

$$h = \frac{x_n - x_0}{n} \tag{1}$$

$$A_1 = \frac{\mathrm{h}}{3}[\,\mathrm{f}(\mathrm{x}_0) + 4 * \mathrm{f}(\mathrm{x}_1) + \mathrm{f}(\mathrm{x}_2)] \tag{2}$$

$$AUC = \frac{h}{3}[f(x_0) + 2\sum_{j=1}^{\frac{n}{2}-1} f(x_{2j}) + 4\sum_{j=1}^{\frac{n}{2}-1} f(x_{2j-1}) + f(x_n)] \quad (3)$$

Index	Range
1	11,140 – 11,160
2	23,550 – 23,800
3	27,900 – 29,400
4	55,500 – 59,000
5	66,400 – 68,100
6	78,600 – 80,500
7	111,500 – 115,500

Table 1: *Potential protein biomarker ranges identified by MALDI-ToF from gargle samples.*

3. Results and Discussion

Recently, studies utilizing MALDI-MS as a potential method for the detection of SARS-CoV-2 have been reported [21],[22]. However, these studies involved nasopharyngeal swab sampling, were limited into the mass-to-charge range in collected mass spectra (< m/z =20,000) and lacked association of peaks to any molecular identity. In another study, MALDI-MS was employed to detect both viral proteins and immunoglobulins from samples of gargled water [18]. This method introduced a diagnostic method for COVID-19 that provides advantages over other tests including reduced costs, non-invasive sampling, rapid analysis time, and the ability to observe both viral and host proteins. This work, however, was limited in that no direct comparison was made between the MALDI-ToF diagnostic method of gargle samples and the results of a standard, accepted test.

Herein, we have elaborated on the work of Iles *et al.* [18] by conducting a comparison study based on the results of 60 nasopharyngeal RT-PCR specimens (the gold-standard test) and the MALDI-ToF spectral profiles of water gargle samples. By comparing the mass spectra of known PCR-positive and PCR-negative samples, clear distinctions can be observed in the range of 20,000 m/z to 200,000 m/z, a range not analyzed by previous COVID-19 MS studies that compare PCR results. Furthermore, to correlate the RT-PCR outcomes to the MALDI-ToF profiles, the area under the curve (AUC) of putative immunoglobulins and viral protein peaks was used.

3.1. MALDI-ToF profiling for PCR negative gargle sample:

In order to compare a true COVID-19 negative sample with PCR-negative samples, the spectrum for pre-COVID-19 saliva (see Materials and Methods) was acquired and considered as a negative control. Seen in Figure 1, the profile for a representative PCR-negative sample closely resembles that of the negative control. Of particular note are the peaks consistent between the two profiles, for salivary proteins would be present in both diseased and healthy individuals. The identity of these proteins is suggested later (Section 3.3).

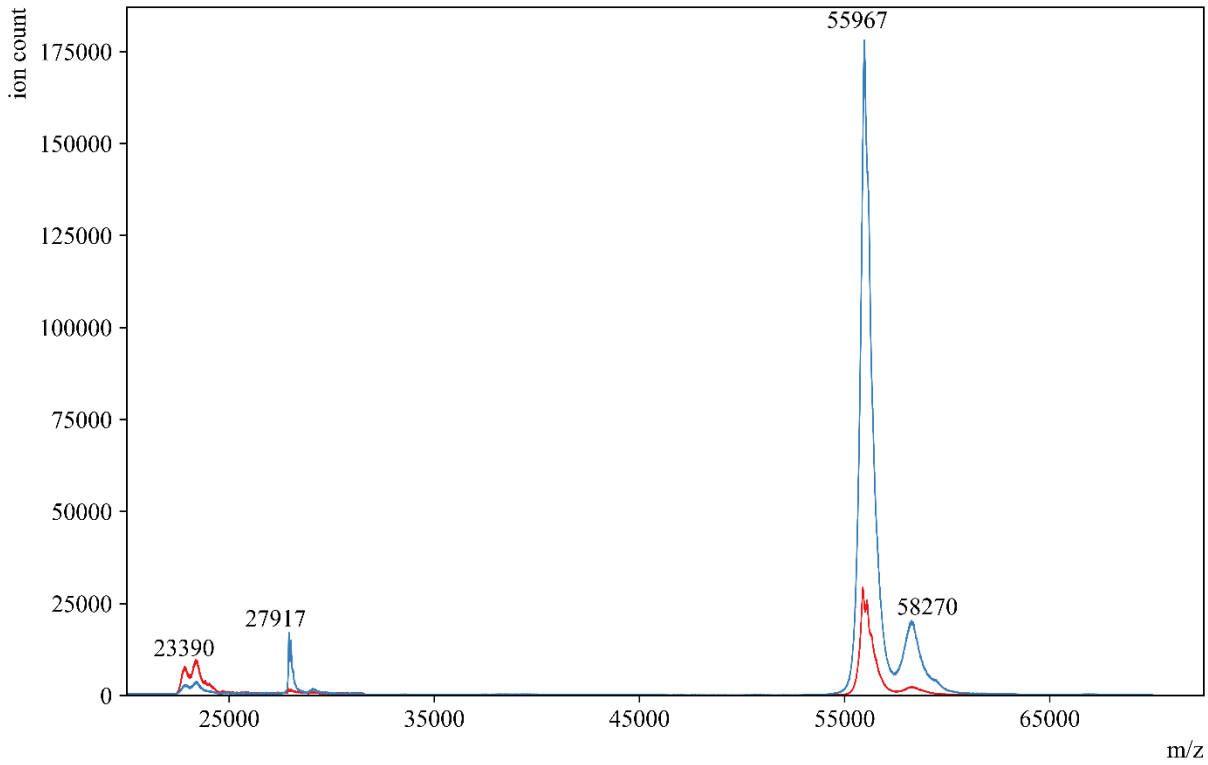

Figure 1 *displays the MALDI-ToF mass spectra of a gargle sample from an example donor who tested PCR negative (— red) along with the negative control (—blue). Both profiles closely overlapped each other, and no other signal was detected after 65000 m/z.*

3.2. Potential biomarkers for COVID-19 by MALDI-ToF:

Distinct differences were observed between PCR-negative and PCR-positive samples with respect to intensities and presence or absence of peaks (Figure 2). Overall, PCR-positive samples had higher intensities and numerous additional peaks in a given spectrum. The peaks around 23,000 m/z, 28,000 m/z and 56,000 m/z that are present in both PCR-negative and PCR-positive profiles generally had higher intensities in the PCR-positive cases (Figure 2A). The additional peaks unique to PCR-positive profiles were present between the ranges of 33,000 m/z to 51,000 m/z and 65,000 m/z to 120,000 m/z (Figure 2B and 2C).

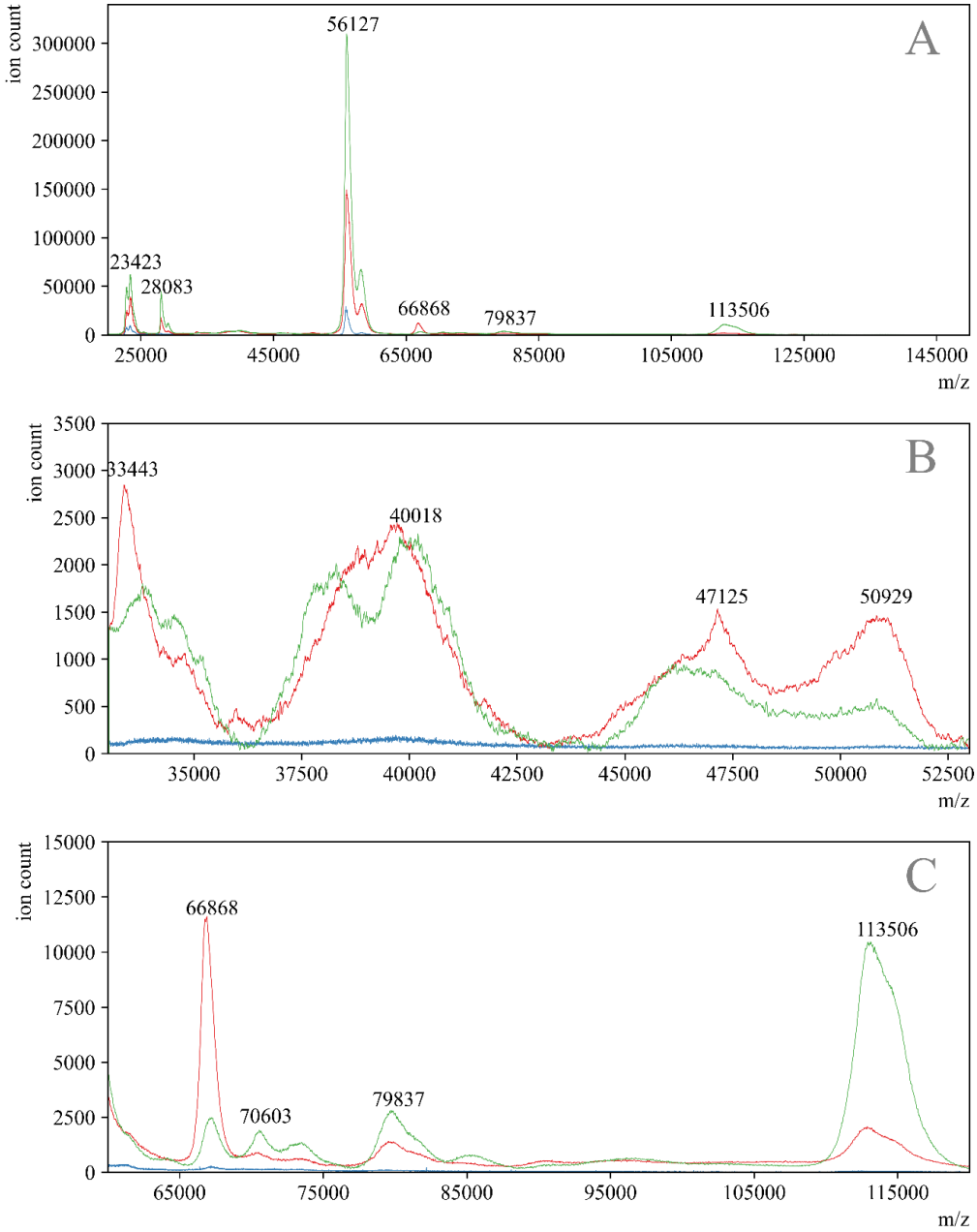

Figure 2 *displays the MALDI-ToF mass spectra of a gargle sample from a PCR-negative donor (—blue) and an overlay of two PCR-positive donors (— red, — green). Panel* ***A*** *is the full-range mass spectra of the samples. Panels* ***B*** *and* ***C*** *are specific ranges where differences in mass spectra are prominently observable between PCR-positive and PCR-negative samples.*

It is important to note that although the peaks found within these ranges reasonably separated the PCR-positive spectra from the PCR-negative spectra, there is variation in peak intensities found within the PCR-positive profiles. We suspect that this may be attributed to the time course of infection and its corresponding immune response: for the time from exposure to point of sample collection was not recorded, nor is such data easily collected.

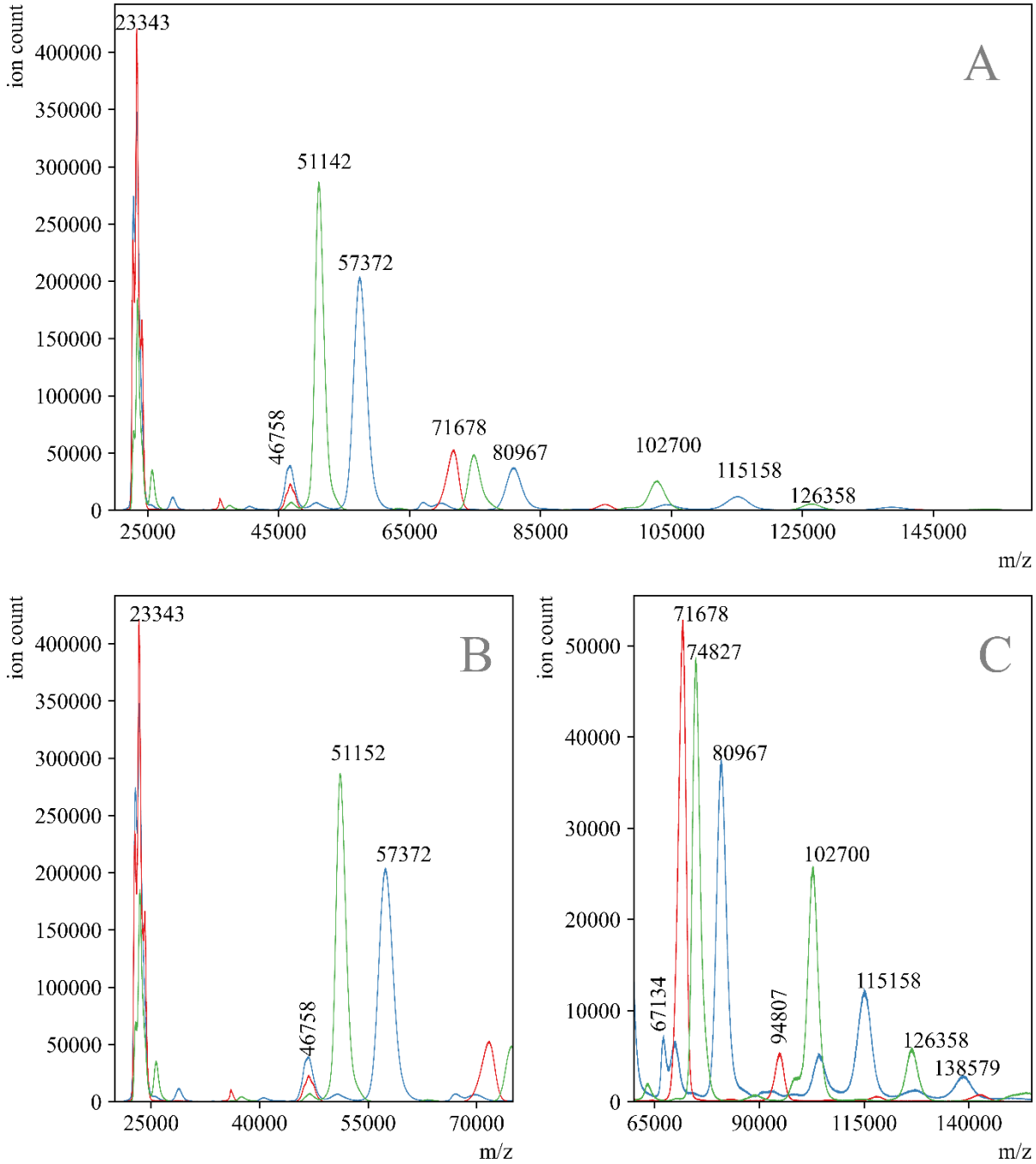

Figure 3 *displays the MALDI mass spectra of human serum derived IgA (—blue), IgG(— green) and IgM (— red) isolated from human serum reduced with 1M DTT for 10min. Panel* ***A*** *shows the entire mass range collected and Panels* ***B*** *and* ***C*** *show zoomed-in ranges.*

3.3. Comparison and identification of peaks in gargle samples:

As mentioned previously, the mass spectra of gargle specimens would include the host salivary proteins (as well as the viral proteins if PCR-positive). Hence, it was crucial to identify these peaks to confirm the presence of SARS-CoV-2 in PCR-positive sample as well as to gain further insights into the other proteins residing in the saliva of donors.

Figure 1 displays the mass spectra of samples of pre-COVID-19 saliva and PCR negative saliva. The most prominent peak in these spectra occurs near 56,000 m/z and two other features near 23,300 m/z and 28,000 m/z are also observed. The peaks near 56,000 m/z and 23,000 m/z have been reported for a gargle sample [18] and assigned as IgA heavy chain and Ig light chains, respectively.

To this end, three human serum immunoglobulins (IgA, IgG and IgM) under reducing conditions were analyzed using equivalent parameters as gargle samples (Figure 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 24,000 m/z. Various other peaks were observed throughout the overlapped spectra which were considered combinations of heavy and light chains, dimers of heavy chains, and multiply-charged ions.

Basal levels of secretory IgA are always present in a healthy saliva specimen [14]. It follows that, within the negative control profile and PCR-negative samples (Figure 1), the two major peaks observed correspond to the light chain (23,390 m/z) and the heavy chain (55,967 m/z) of IgA. Moreover, the intensities of these peaks are comparatively larger in PCR-positive samples. We suspect that this may be attributable to a heightened immune response upon antigen invasion. Along the same lines, another feature to note is the heavy chain peak of IgM at 71,678 m/z in Figure 3A and the existence of a peak at 70,603 m/z seen in Figure 2C (green, PCR-positive spectrum). This could indicate early infection of COVID-19, as it has been reported that viral-specific IgM antibodies are the first to be produced followed by IgA and IgG [23].

Additionally, human α-amylase is found in two forms in saliva with molecular weights of ca. 56 kDa (unglycosylated) and 62 kDa (glycosylated) [24],[25]. 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 [26],[27]. 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 S1) therefore α-amylase likely contributes to the peak found in this region of gargle sample spectra.

There were observable peaks within the range of 65,000 m/z to 120,000 m/z, particularly in the PCR-positive spectra and more prominently in the ranges of 66,400-68,100 m/z and 78,600-80,500 m/z (Figure 2C). Comparing the UniProt database (UniProtKB P0DTC2) and the work of Iles *et al*., the peak at 79837 m/z could be a signal for the S1 fragment of the SARS-CoV-2 spike protein. This same peak was observed in the positive control spectrum of SARS-CoV-2 (Figure 4). Furthermore, the peaks described by Iles *et al.*[18] as viral envelope proteins (VEPs) were also visible within PCR-positive profiles (Figure 2B).

A signal was consistently observed in the range of 66,000 – 68,000 m/z for PCR positive samples. By comparing numerous PCR positive spectra, it appears that there are two distinct species with closely overlapping m/z envelopes. This peak may be a signal from 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 [28]. Alternatively, this peak may arise from a fragment of IgA (Figure 3C). This signal cannot be unequivocally identified at this time. An additional caution is that, in the mass spectra of control samples that are isolated from cell culture, 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 4).

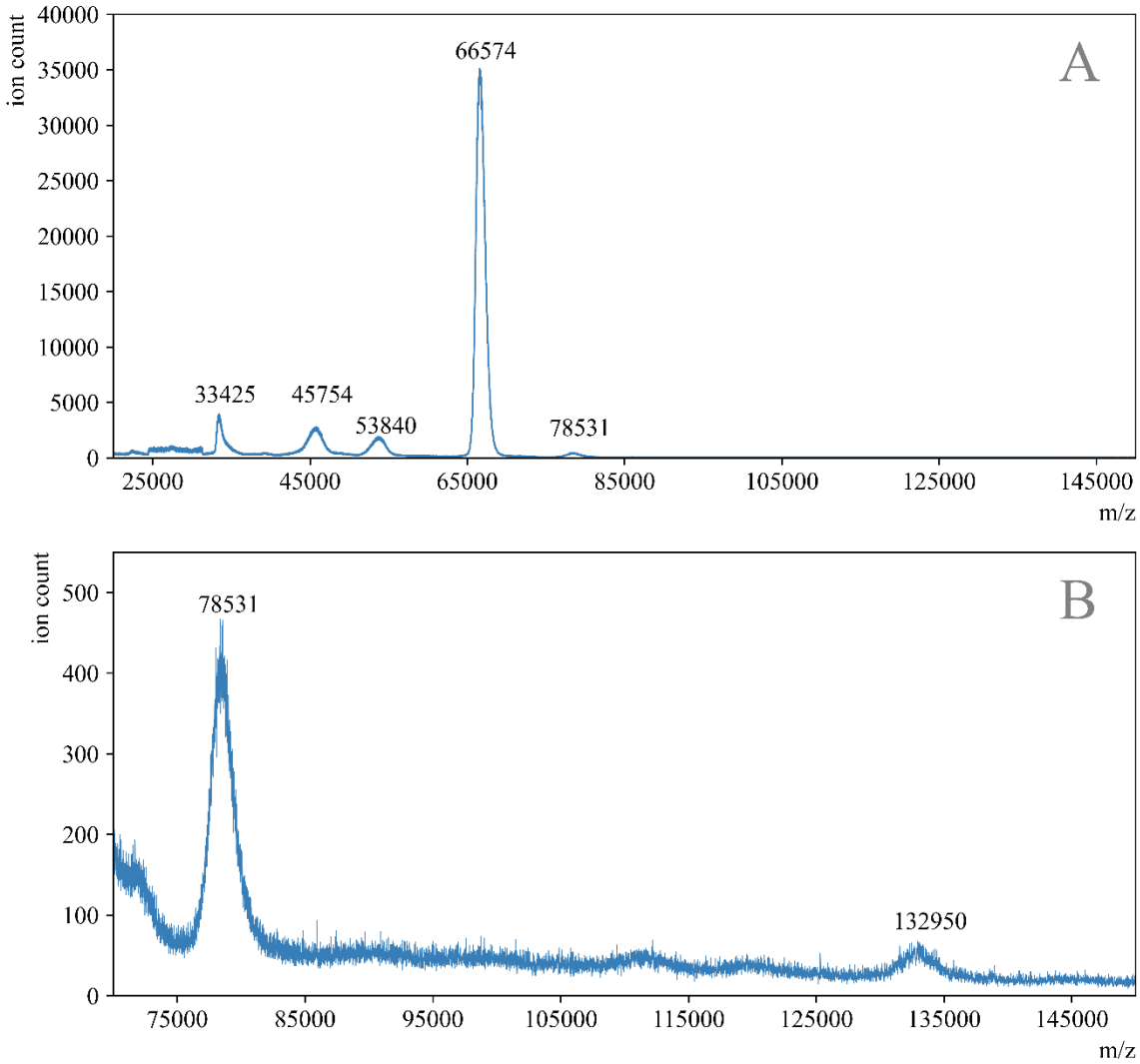

Figure 4 *displays the MALDI-ToF spectrum of heat inactivated SARS-CoV-2 utilized as the positive control. Panel* ***A*** *shows the entire mass range collected and Panel* ***B*** *shows the major peak signals beyond 70,000 m/z.*

3.4. MALDI-ToF criteria for diagnosis:

In order to compare the results of the nasopharyngeal RT-PCR samples and the protein profiles collected with gargle samples via MALDI-ToF, the AUC was calculated for each sample under the seven peak ranges of interest listed in Table 1. Included among the viral and immune proteins discussed is a peak located at 11,150 m/z used as a quality control feature, most likely to be cystatin A, a resident protein of saliva [16]. The presence of this peak was used to deem a sample as successfully gargled and appropriate to be included for analysis. Table S1 shows these seven features organized for each specimen, named as the reported RT-PCR result.

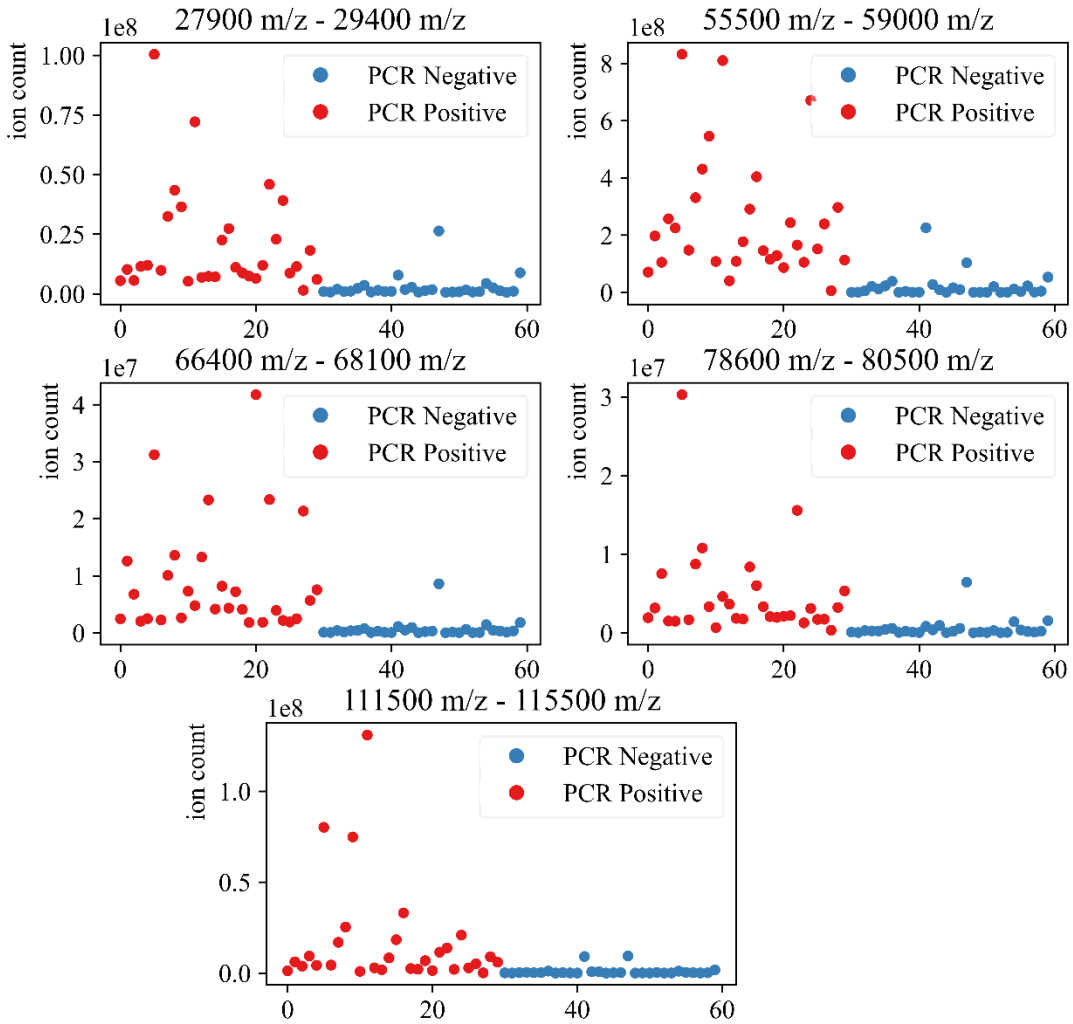

Figure 5 *displays AUC values for the 60 gargle samples for the five potential biomarker peak range. The sample data files were labeled 1-60 and the AUC for the biomarker range for each file is depicted. The PCR Positive files (red) were clustered and labelled as 1-30 while the PCR Negative files (blue) were clustered and labelled as 31-60. A marked difference for the AUC in the biomarker ranges of PCR Positive versus PCR Negative samples can be observed.*

This table (Table S1) was used to sort the samples by AUC values for a given feature from largest to smallest. Sorting in this way showed a reasonable separation between PCR-positive and PCR-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 with any diagnostic test, an output value threshold was necessary to establish in order to distinguish between positive/negative cases as well as to compare our test to the results of RT-PCR. This was done by examining how the PCR-positive and PCR-negative specimens were separated when sorted by AUC values for each feature (Figure 5). A threshold for the AUC values was determined that maintained as similar an outcome as possible to RT-PCR results. In this way, AUC values above the threshold value were deemed MALDI-ToF positive and AUC values below deemed negative. The cut-off values for a given feature chosen under this criterion are reported in Table 2, along with the percent agreement between RT-PCR results and our analysis.

For all five potential biomarkers mentioned in Table 2, we achieved 90% and higher agreement with the RT-PCR results. Although the identities of the proteins are yet to be confirmed, these m/z ranges certainly exhibit a relationship to the PCR results. In future, more rigorous analyses, more than one biomarker can be used in combination to improve the overall specificity of the test by expanding the discrimination criteria.

It is interesting to note that 89% of the PCR 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, however this cohort has been an elusive group to track, particularly early in the pandemic when only symptomatic persons were eligible for PCR testing. The PCR positive, asymptomatic group are reported to be as contagious as symptomatic persons [3] and have 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, recreational facilities etc. are predominant.

(Ig heavy chain/amylase)$^{+2}$ (27900-29400 m/z)	***Cut-off AUC : 5.2E+06***	
	PCR-Positive	**PCR-Negative**
MALDI-ToF Positive	96.67%	10.00%
MALDI-ToF Negative	3.33%	90.00%
Ig heavy chain/amylase (55500-59000 m/z)	***Cut-off AUC : 4.0E+07***	
	PCR-Positive	**PCR-Negative**
MALDI-ToF Positive	96.67%	10.00%
MALDI-ToF Negative	3.33%	90.00%
Spike protein S2/immune protein (66400-68100 m/z)	***Cut-off AUC : 1.5E+06***	
	PCR-Positive	**PCR-Negative**
MALDI-ToF Positive	100.00%	6.67%
MALDI-ToF Negative	0.00%	93.33%
Spike protein S1 (78600-80500 m/z)	***Cut-off AUC : 1.0E+06***	
	PCR-Positive	**PCR-Negative**
MALDI-ToF Positive	93.33%	10.00%
MALDI-ToF Negative	6.67%	90.00%
Asymptomatic Biomarker (111500-115500 m/z)	***Cut-off AUC : 1.2E+06***	
	PCR-Positive	**PCR-Negative**
MALDI-ToF Positive	93.33%	10.00%
MALDI-ToF Negative	6.67%	90.00%

Table 2: *Percent agreement between RT- PCR results and MALDI-ToF results as determined by threshold 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 PCR-positive from PCR-negative results.*

4. Conclusion

A rapid and inexpensive saliva/gargle test was employed for the diagnosis of COVID-19 using MALDI-ToF mass spectrometry. For the first time, direct comparison of RT-PCR results with MALDI-ToF analysis over a wide range of 2,000-200,000 m/z was carried out. Five potential biomarkers of COVID-19 were selected after MALDI-ToF-based analysis of 60 gargle samples (30 PCR-positive and 30 PCR-negative). It is crucial to further investigate these potential

biomarker peaks for identity in order to understand the correlation of these 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 the mentioned assay and comparing peak masses of these controls to those found in gargle samples. Further confirmatory analysis such as sequencing of proteins that occur in a healthy and SARS-CoV-2-infected saliva proteome is required. 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. While currently detection of the infection is based on AUC of the separate potential biomarkers, we aim to develop machine learning models for achieving unbiased, higher accuracies using multiple peaks in tandem as diagnosis criteria.

5. Acknowledgements

We thank Dr. Ray Iles, MAPSciences, for the kind gift of LBSD-X buffer and Northern Illinois University Athletics (Yordon Center) for assisting us with the sample collection. The positive control reagent (Figure 4) was deposited by the Centers for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate USA-WA1/2020, Heat Inactivated, NR-52286.

Competing Interests: P. Chivte, Z. LaCasse, V. Seethi, P. Bharti and E.R. Gaillard have no competing interests to declare.

IRB information: The protocol "Use of MALDI TOF mass spectrometry to analyze SARS CoV2 viral proteins" was approved by the Northern Illinois University Institutional Review Board on August 12, 2020.

References

[1] World Health Organization, "WHO Coronavirus Disease (COVID-19) Dashboard": https://covid19.who.int/

[2] Centers for Disease Control and Prevention, "Symptoms of Coronavirus": https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

[3] Z. Gao, Y. Xu, C. Sun, X. Wang, Y. Guo, and S. Qiu, K. Ma,"A systematic review of asymptomatic infections with COVID-19," *J. Microbiol. Immunol. Infect.*, 2020. https://doi.org/10.1016/j.jmii.2020.05.001

[4] E. C. Stites and C. B. Wilen, "The interpretation of SARS-CoV-2 diagnostic tests," *Med, Cell Press*, vol. 1, pp. 1–12, 2020. https://doi.org/10.1016/j.medj.2020.08.001

[5] H. Yao, Y.Song, Y.Chen, N. Wu, J. Xu, C. Sun, J. Zhang et al., "Molecular architecture of the SARS-CoV-2 virus," *Cell*, 183, pp. 1–9, 2020. https://doi.org/10.1016/j.cell.2020.09.018

[6] N. Ravi, D. L. Cortade, E. Ng, and S. X. Wang, "Diagnostics for SARS-CoV-2 detection: A comprehensive review of the FDA-EUA COVID-19 testing landscape," *Biosens. Bioelectron.*, vol. 165, 2020. https://doi.org/10.1016/j.bios.2020.112454

[7] J. Prasad and R. Yadav, Rajkumar, "Laboratory diagnosis of SARS-CoV-2 - A review of current methods," *J. Infect. Public Health*, vol. 13, pp. 901–905, 2020. https://doi.org/10.1016/j.jiph.2020.06.005

[8] W. Feng, A.M. Newbigging, C. Le, B. Pang, H. Peng, Y. Cao, J.Wu et al., "Molecular Diagnosis of COVID-19: Challenges and Research Needs," *Anal. Chem.*, vol. 92, pp. 10196–10209, 2020. https://dx.doi.org/10.1021/acs.analchem.0c02060

[9] A. Aita, D. Basso, A.M. Cattelan, P. Fioretto, F. Navaglia, F. Barbaro, A. Stoppa et al., "SARS-CoV-2 identification and IgA antibodies in saliva : One sample two tests approach for diagnosis," *Clin. Chim. Acta*, vol. 510, August, pp. 717–722, 2020. https://doi.org/10.1016/j.cca.2020.09.018

[10] N. Huang, P. Pérez, T. Kato, Y. Mikami, K. Okuda, R. C. Gilmore, C. D. Conde *et al.*, "SARS-CoV-2 infection of the oral cavity and saliva," *Nat. Med.*, 2021. https://doi.org/10.1038/s41591-021-01296-8

[11] C. B. F. Vogels, A.E. Watkins, C.A. Harden, D.E. Brackney, J. Shafer, J. Wang, C. Caraballo et al., "SalivaDirect : A simplified and flexible platform to enhance SARS-CoV-2 testing capacity ," *Cell Press*, Med 2, pp. 1–18, 2020. https://doi.org/10.1016/j.medj.2020.12.010

[12] D. R. E. Ranoa, R.L. Holland, F.G. Alnaji, K.J. Green, L. Wang, C.B. Brooke, M.D. Burke et al., "Saliva-based molecular testing for SARS-CoV-2 that bypasses RNA extraction," 2020. https://doi.org/10.1101/2020.06.18.159434

[13] L. Azzi, G. Carcano, F. Gianfagna, P. Grossi, D.D. Gasperina, A. Genoni, M. Fasano et al., "Saliva is a reliable tool to detect SARS-CoV-2," *J. Infect.*, vol. 81, pp. e45–e50, 2020. https://doi.org/10.1016/j.jinf.2020.04.005

[14] D. Sapkota, T.M. Soland, H.K. Galtung, L.P. Sand, S. Giannecchini, K.K.W. To, M.C. Mendes-Correa et al., "COVID-19 salivary signature : diagnostic and research opportunities," *J. Clin. Pathol.*, 0, pp. 1–6, 2020. http://dx.doi.org/10.1136/jclinpath-2020-206834

[15] Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao et al., "The Impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity," *Cell*, vol. 182, pp. 1284–1294, 2020. https://doi.org/10.1016/j.cell.2020.07.012

[16] C. Ihling, D. Tanzler, S. Hagemann, A. Kehlen, S. Huttelmaier, C. Arlt, A. Sinz, "Mass spectrometric identification of SARS-CoV-2 proteins from gargle solution samples of COVID-19 patients," *J. Proteome Res.*, 2020. https://dx.doi.org/10.1021/acs.jproteome.0c00280

[17] A. D. Whetton, G. W. Preston, S. Abubeker, and N. Geifman, "Proteomics and informatics for understanding phases and identifying biomarkers in COVID-19 disease," *J. Proteome Res.*, 2020. https://dx.doi.org/10.1021/acs.jproteome.0c00326

[18] R. K. Iles, R. Zmuidinaite, J.K. Iles, G. Carnell, A. Sampson, J.L. Heeney, “Development of a clinical MALDI-ToF mass spectrometry sssay for SARS-CoV-2 : Rational design and multi-disciplinary team work,” *Diagnostics*, vol. 10, no. 746, pp. 1–16, 2020. https://doi.org/10.3390/diagnostics10100746

[19] M. Jorge, N.M. Garrido, A. J. Queimada, I. G. Economou, E.A. Macedo “Effect of the integration method on the accuracy and computational efficiency of free energy calculations using thermodynamic integration,” *J. Chem. Theory Comput*., vol. 6, pp. 1018–1027, 2010. https://doi.org/10.1021/ct900661c

[20] R. J. Tallarida, R. B. Murray "Area under a curve: trapezoidal and Simpson’s rules," Manual of Pharmacologic Calculations, pp. 77–81, 1987. https://doi.org/10.1007/978-1-4612-4974-0_26

[21] F. M. Nachtigall, A. Pereira, O. S. Trofymchuk, and L. S. Santos, “Detection of SARS-CoV-2 in nasal swabs using MALDI-MS,” *Nat. Biotechnol*., 2020. https://doi.org/10.1038/s41587-020-0644-7

[22] M. F. Rocca, J.C. Zintgraff, M. E. Dattero, L.S. Santos, M. Ledesma, C. Vay, M. Prieto et al., “A combined approach of MALDI-TOF mass spectrometry and multivariate analysis as a potential tool for the detection of SARS-CoV-2 virus in nasopharyngeal swabs,” *J. Virol. Methods*, vol. 286, 2020. https://doi.org/10.1016/j.jviromet.2020.113991

[23] A. K. Azkur, M. Akdis, D. Azkur, M. Sokolowska, W. van de Veen, M-C. Bruggen, L. O’Mahony et al., “Immune response to SARS-CoV-2 and mechanisms of immunopathological changes in COVID-19,” *Allergy*, vol. 75, pp. 1564–1581, 2020. 10.1111/all.14364

[24] S. Z. Fisher, L. Govindasamy, C. Tu, M. Agbandje-McKenna, D.N. Silverman, H.J. Rajaniemi, R. McKenna, “Structure of human salivary α-amylase crystallized in a C-centered monoclinic space group,” *Protein Struct. Commun*., vol. F62, pp. 88–93, 2006. 10.1107/S1744309105042491

[25] R.A. Bank, E.H. Hettema, F. Arwert, A.V.N. Amerongen, J.C. Pronk, “Electrophoretic characterization of posttranslational modifications of human parotid salivary a-amylase,” Electrophoresis, vol. 12, pp. 74–79, 1991.

[26] L. Kandra, Gyonyi, J. Remenyik, C. Ragunath, N. Ramasubbu.,“Transglycosylations catalysed by Y151M mutant of human salivary α -amylase (HSA),” *Biologia (Bratisl).*, vol. 60, pp.57–64, 2005.

[27] C. Hirtz, F. Chevalier, D. Centeno, V. Rofidal, J-C. Egea, M. Rossignol et al., “MS characterization of multiple forms of alpha-amylase in human saliva,” *Proteomics*, vol. 5, pp. 4597–4607, 2005. 10.1002/pmic.200401316

[28] N. G. Herrera, N.C. Morano, A. Celikgil, G.I.Georgiev, R.J, Malonis, J.H. Lee, K.Tong et al., “Characterization of the SARS-CoV-2 S protein: Biophysical, biochemical, structural, and

antigenic analysis," *ACS Omega*, vol. 6, pp. 85–102, 2021. https://dx.doi.org/10.1021/acsomega.0c03512