Journal of proteome research

pubs.acs.org/jpr

Article

MALDI(+) FT-ICR Mass Spectrometry (MS) Combined with Machine Learning toward Saliva-Based Diagnostic Screening for COVID-19

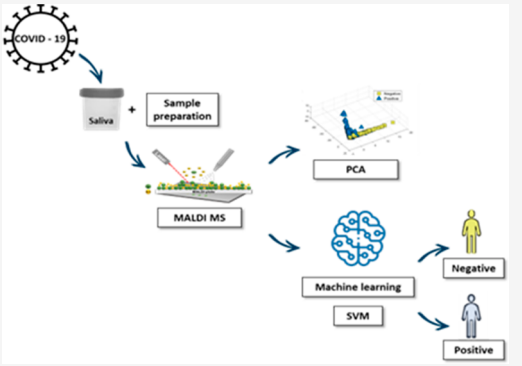
Camila M. de Almeida, Larissa C. Motta, Gabriely S. Folli, Wena D. Marcarini, Camila A. Costa, Ana C. S. Vilela, Valério G. Barauna, Francis L. Martin, Maneesh N. Singh, Luciene C. G. Campos, Nádia L. Costa, Paula F. Vassallo, Andrea R. Chaves, Denise C. Endringer, José G. Mill, Paulo R. Filgueiras, and Wanderson Romão*

Cite This: *J. Proteome Res.* 2022, 21, 1868−1875

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Rapid identification of existing respiratory viruses in biological samples is of utmost importance in strategies to combat pandemics. Inputting MALDI FT-ICR MS (matrix-assisted laser desorption/ionization Fourier-transform ion cyclotron resonance mass spectrometry) data output into machine learning algorithms could hold promise in classifying positive samples for SARS-CoV-2. This study aimed to develop a fast and effective methodology to perform saliva-based screening of patients with suspected COVID-19, using the MALDI FT-ICR MS technique with a support vector machine (SVM). In the method optimization, the best sample preparation was obtained with the digestion of saliva in 10 μL of trypsin for 2 h and the MALDI analysis, which presented a satisfactory resolution for the analysis with 1 M. SVM models were created with data from the analysis of 97 samples that were designated as SARS-CoV-2 positives versus 52 negatives, confirmed by RT-PCR tests. SVM1 and SVM2 models showed the best results. The calibration group obtained 100% accuracy, and the test group 95.6% (SVM1) and 86.7% (SVM2). SVM1 selected 780 variables and has a false negative rate (FNR) of 0%, while SVM2 selected only two variables with a FNR of 3%. The proposed methodology suggests a promising tool to aid screening for COVID-19.



KEYWORDS: *COVID-19, MALDI FT-ICR MS, machine learning, saliva, SARS-CoV-2, screening*

INTRODUCTION

Coronavirus disease is a viral infection caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Its spread may be due to airborne transport of aerosol particles from normal breathing or larger airborne suspended droplets from coughing and sneezing.[1] COVID-19 was declared a pandemic on March 11th, 2020 by the World Health Organization, with >420 000 000 people infected and 5 800 000 deaths worldwide as of March 2022.[2] Defining methods for early and rapid detection of COVID-19 is crucial to provide timely treatment, tracking individuals who have had contact with infected patients, and interrupting transmission routes.[3]

Reverse transcription−polymerase chain reaction (RT-PCR) in swabs collected from the nasopharyngeal mucosa is routinely used and considered the gold standard in detecting SARS-CoV-2 infection.[4] Smear collection requires qualified health professionals, and there is great concern regarding handling, transporting, and storing samples. In addition to the nasopharyngeal swab causing discomfort to patients, it can also cause bleeding. Therefore, there are several contraindications to this procedure, such as clinical conditions, as a consequence of coagulopathy or anticoagulant therapy and significant deviation of the nasal septum.[5] Additionally, the RT-PCR test has limitations associated with expenses and time that need to be overcome.[6,7]

To improve issues associated with transportation of highly contaminated materials, availability of equipment, and the need for faster testing, the use of rapid antigen tests (RATs) for COVID-19 was seen as a solution to quickly detect and track communities suspected of contracting COVID-19. A rapid antigen test is a rapid diagnostic test that detects the presence of the virus directly at the time of analysis.[8] Despite meeting the needs of a rapid response in a mass-screening program, we still need tests with greater sensitivity and accuracy as the frontline testing for COVID-19 diagnosis.[9] The diagnosis for

Received: March 12, 2022
Published: July 26, 2022

COVID-19 still has limitations for general use, especially in developing countries, due to the gold technique used and the need for inputs to apply numerous tests in the population. With high sensitivity and specificity, an alternative to these methods would be applying diagnostic tools based on mass spectrometry (MS), using the MALDI (matrix-assisted laser desorption/ionization) MS technique to identify disease-specific proteins. Recently, MS techniques have been explored to detect COVID-19 on the basis of the identification of salivary biomarkers.[10−12]

Saliva is a secretion produced by the salivary glands. It has several functions, such as protecting and cleaning the oral cavity and aiding digestion. In addition, due to the ease and safety of collecting saliva samples via noninvasive techniques, this biofluid may be an alternative source to screen disease biomarkers.[13,14] Thus, COVID-19 could be detected through salivary diagnostic tests, presenting advantages for healthcare professionals and patients.[15−17] The large amount of data obtained in MS spectra requires the combination of the output of this technique with robust statistical strategies to identify diagnostic patterns.[18] The current trend in associating algorithms labeled as artificial intelligence and "omics" techniques has yielded platforms involving machine learning (ML) for MS data analysis, aiming to identify disease biomarkers, including COVID-19 severity assessment.[19]

Recently, Nachtigall et al. (2020) reported the development of a MALDI MS method to diagnose SARS-CoV-2 infection. In this study, nasal swab samples were analyzed directly by MALDI TOF MS, and a spectral pattern was used to classify infected and noninfected patients, an interesting application of mass spectrometry to detect viral infections.[20] Another study, carried out by Chivte et al. (2021), used saliva samples and established five viral protein signals as potential biomarkers for COVID-19. MALDI TOF analysis showed ≥90% agreement with RT-qPCR results, concluding that MALDI-TOF can be used to develop assays for COVID-19 in a relatively quick and inexpensive way.[11]

Given the above, the development of a methodology using high-resolution MALDI FT-ICR MS combined with machine learning algorithms can provide a quick and highly accurate response to classify positive samples for diseases, favoring rapid and accurate diagnoses that complement genomic information. This may also allow us to increase our current knowledge of COVID-19.

MATERIALS AND METHODS

Saliva Samples

The study included 149 saliva samples (97 RT-PCR positives for SARS-CoV-2 and 52 negatives for SARS-CoV-2). Among them, 86 samples (34 with positive RT-PCR for SARS-CoV-2 and 52 negatives) were obtained by the research group of the Center for Health Sciences at UFES (CAAE ethics committee: 30993920.1.0000.5071) and the other 63 positive samples were provided by the research group of the Faculty of Dentistry of the Federal University of Goiás (ethics committee CAAE: 38088920.9.0000.5083, 38088920.9.3002.5082, 38088920.9.3001.5078). All participants were >18 years old and were tested for SARS-CoV-2 infection using a qRT-PCR test.

Saliva collection took place at the hospitals of the respective cities, Vila Velha and São Mateus, in patients who had at least one characteristic symptom of COVID-19 and spontaneously attended the hospitals to undergo the RT-PCR test for COVID-19. Immediately after oropharyngeal swab collection, each patient received a sterile tube to facilitate self-collection of saliva (around 1 mL) under direct supervision of a nurse. Table S1 describes participant demographic profiles with gender and symptoms.

Sample Preparation

In the optimization of sample preparation, protein in saliva samples from voluntary donors without symptoms were determined by the dye-binding Bradford assay.[21] An aliquot of saliva (500 μL) was centrifuged at 1903.67*g* for 15 min at 4 °C. The supernatant aliquot (10 μL) was mixed with 190 μL Bradford reagent in a well plate. Protein absorbance was measured at 595 nm against a blank reagent. Standard curves were prepared using bovine serum albumin (BSA).

To the 50 μL-supernatant saliva sample 10 μL AmBic (50 mM; ammonium bicarbonate, *Sigma-Aldrich Chemicals*, USA) and 25 μL RapiGest SF (0.2%; *Waters*, Milford, MA, USA) were added, and the tubes were heated to at 37 °C for 1 h. Then, these tubes were submitted to disulfide reduction with 2.5 μL DTT at a concentration of 100 mM in AmBic (dithiothreitol in 50 mM AmBic) for 40 min at 37 °C and 2.5 μL IAA at a concentration of 300 mM in AmBic (Iodoacetamide, *Sigma-Aldrich Chemicals*, USA). Then, they were placed in the dark for 30 min at room temperature. Protein digestion was performed by adding 5 or 10 μL 0.05 μg/μL Trypsin (*Promega*, Madison, WI, USA) and incubating the samples at 37 °C for 2 h or overnight.

Digestion was terminated with 10 μL of TFA 5% (trifluoroacetic acid, *Sigma-Aldrich Chemicals*, USA), followed by heating the sample at 37 °C for 90 min. The samples were centrifuged and the pellets were discarded. The samples were purified with a C18 spin column (Pierce C18 Spin Column, Thermo Scientific, San Jose, CA, USA). These columns were activated by 2 × 200 μL of 50% ACN and equilibrated with 2 × 200 μL of 0.5% TFA. The tryptic peptides were adsorbed to the media using two repeat cycles of 60 μL (consisting of 15 μL 2% TFA in 20% ACN plus 45 μL sample) sample buffer loading, and the column was washed using 2 × 200 μL of 0.5% TFA. Finally, the peptides were eluted in 2 × 20 μL of 70% ACN.

After optimization, samples with suspected COVID-19 were digested and analyzed by MALDI MS.

MALDI MS Analysis

MALDI MS analyses were performed with an FT-ICR mass spectrometer (model 9.4 T Solarix, Bruker Daltonics, Bremen, Germany) equipped with a Smart beam II Nd: YAG laser (355 nm) and MALDI source. Two μL of each sample were spotted in triplicate on the MALDI target plate (Bruker Daltonics), followed by the addition of 1 μL of the matrix solution composed of 5 mM CHCA in methanol (α-ciano-hydroxy-cinnamic acid, *Sigma-Aldrich Chemicals*, USA) at 50% in acetonitrile:0.1% trifluoroacetic acid and air-drying.[22]

The mass spectra were acquired in positive ionization mode in the mass range *m*/*z* 200−4000. The main experimental parameters used were as follows: laser frequency, 200 Hz; plate offset, −100.0 V; deflector plate, −210.0 V; the number of laser shots, 200; laser power, 35%; laser focus, "small"; 16 scans, and the random walk. The time spent on the analysis and the numbers of signals were evaluated in the MALDI FT-ICR analysis, using different resolving powers (512k, 1M, 2M, 4M, and 8M).

https://doi.org/10.1021/acs.jproteome.2c00148
J. Proteome Res. 2022, 21, 1868−1875

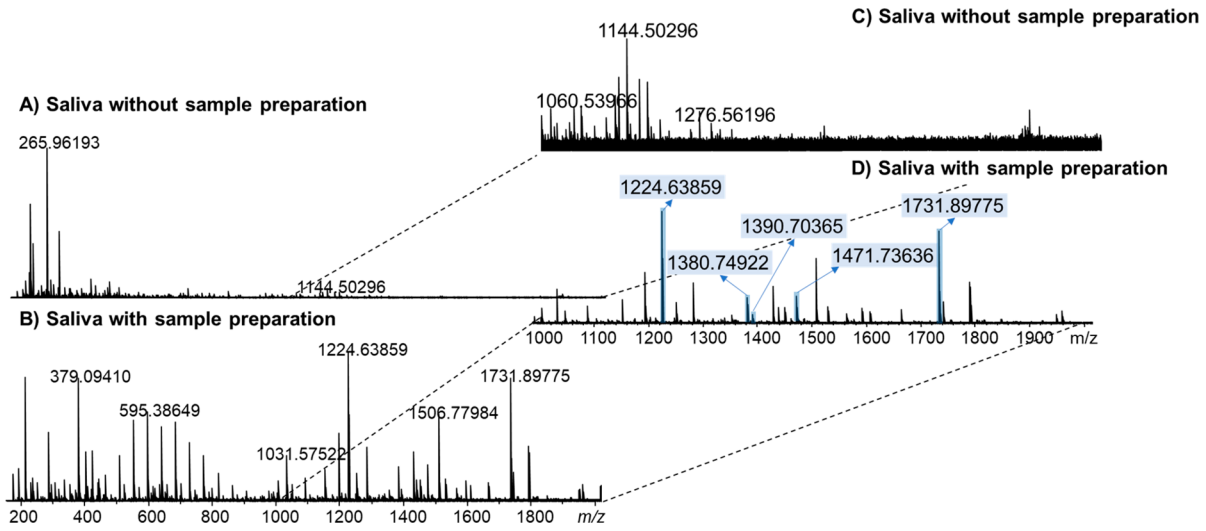


Figure 1. MALDI(+) mass spectra of saliva samples with and without sample preparation. (A,B) Mass range m/z 200−2000 (A) before and (B) after tryptic digestion and C18 spin column processing. (C,D) Mass range m/z 1000−2000 (C) before and (D) after tryptic digestion and C18 spin column processing.

Table 1. Peptides Identified in Saliva Samples after Digestion of Proteins and SPE

m/z (M + H$^+$)	TIC (total ion current)	peptide sequence	parent protein	ref
1224.63859	4.5×10^7	GPPQQGGHPRPP	Acidic proline-rich phosphoprotein or PRH1Protein (P02810)	30
1380.74922	1.0×10^7	GPPQQGGHPRPPR		30
1390.70365	3.6×10^6	GGRPQGPPQGQSPQ		22,29
1471.73636	1.1×10^7	GRPQGPPQQGGHQQ		30
1731.89775	3.7×10^7	GPPQQGGHPPPPQGRPQ		22

Chemometrics

The data set obtained with the DataAnalysis software (Bruker Daltonics, USA) was exported in ASC format and imported into the MATLAB 9.0 software (R2013a). Pretreatments such as mean centering, normalization, autoscaling, and combinations were performed.

Principal Component Analysis (PCA)

PCA pattern recognition model was developed to identify possible anomalous samples, visualize similarities and possible natural groupings between samples, and analyze the behavior and dispersion of spectral variables. This modeling is given by decomposing the study X matrix into two matrices: scores containing sample information and loadings containing variable information.

Support Vector Machine (SVM)

SVM classification models are designed to identify relevant ion disease-specific biomarkers and classify SARS-CoV-2 positive and negative samples. For each set of techniques and properties to be estimated, the sample set was previously divided into calibration (70%) and prediction (30%) by the Kennard Stone method.[23] The cost (C), margin (ε), and gamma (γ) parameters of the kernel function were optimized.

RESULTS AND DISCUSSION

The MALDI FT-ICR MS technique is a widely used tool for studying biological compounds due to its high sensitivity and mass accuracy. Its combination with machine learning can favor molecular diagnostics by using inexpensive, safe, easy, and noninvasive collection samples, such as saliva.[24]

Saliva is a bodily fluid produced by the salivary glands, composed of various substances and a rich reservoir of proteins and peptides; in fact, it brings together >3652 proteins and 12 562 peptides. In general, saliva contains molecules of high molecular weight, such as mucins, lysozyme, immunoglobulins, proline-rich protein (PRP), and statherins.[13,25] In this sense, protein digestion is an essential step for MALDI FT-ICR MS analysis since the worked m/z range (200−4000) does not identify high molecular weight proteins.

Protein identification is based on m/z measurements for a set of peptides and digestion products of the original protein.[26] However, peptide analysis can present several problems that affect the detector response due to unwanted interference from the matrix or from reagents and other additives used to facilitate protein digestion.[27] Thus, an alternative to eliminate these interfering signals is using C18 spin columns after tryptic digestion, which removes interfering contaminants and releases peptides in MS-compatible solutions, resulting in greater sensitivity and a high-quality spectrum.[28]

After the saliva digestion protocol, it was possible to observe the spectral differences between an unprepared saliva sample (Figure 1A) and another sample after the protein digestion step (Figure 1B). This allows to make a comparison of the ions detected by the MALDI FT-ICR MS analysis, giving indications of some peptides already identified in saliva studies in the literature (Table 1).

The digestion protocol was adapted. The trypsin volume (0.05 μg/μL) and the digestion time were evaluated to optimize and speed up the sample preparation process. The volumes used were 5 and 10 μL of trypsin and the digestion times of 2 and 16 h ("overnight"), totaling 4 variations in the digestion process of the saliva sample.

The analysis results by MALDI(+)FT-ICR MS after digestion showed no significant spectral differences between

https://doi.org/10.1021/acs.jproteome.2c00148
J. Proteome Res. 2022, 21, 1868−1875

Table 2. Number of Signals Detected and Time of Acquisition of the MALDI FT-ICR Mass Spectra in the Optimization of the Parameters

data acquisition size	number of signals with signal/noise > 4%	number of signals with relative intensity > % noise	time of acquisition (min)	total time acquisition (h)	total ion current
8 M	43 228	169	1.89	14.1	1.1×10^8
4 M	22 699	234	1.06	7.9	2.1×10^8
2 M	11 729	223	0.69	5.2	1.9×10^8
1 M	6 493	221	0.51	3.8	1.2×10^8
512 k	3 345	169	0.36	2.7	3.9×10^7

the four variations in the digestion protocol (Figure S1) in the m/z range from 1000 to 2200. When analyzing the full scan spectra of the samples that used 10 μL of trypsin for the digestion, these presented a higher ratio of signals detected with a signal-to-noise ratio (S/N) > 4, compared to the samples that used 5 μL of trypsin for the digestion. The samples that had overnight digestion also exhibited more signals, but the 2 h time was satisfactory and sped up the diagnostic process by at least 14 h (Table S2). Additionally, the sample with 10 μL of trypsin and 2 h of digestion showed a higher TIC for all peptide sequences (Table S3). All four samples after digestion showed a different spectrum from the one of the saliva sample without digestion (Figure S1a), in which it was possible to identify the peptide sequence, represented by the ions of m/z 1224; 1380; 1390; 1471; and 1731. They are phosphoprotein derivatives, rich in salivary acid proline (1/2), P02810.[22,30]

Human glandular salivary secretions contain several phosphoproteins rich in acidic proline (PRP). These proteins have essential biological functions that provide a protective environment for teeth and appear to have other activities associated with the modulation of bacterial adhesion to oral surfaces.[31] In optimizing the parameters of MALDI analysis, the resolving power and the acquisition time of the samples were evaluated. The size of the generated data set is related to the number of signals in the mass spectrum, which increases according to the resolving power but decreases the acquisition speed. Thus, it is necessary to use an optimal resolving power that does not affect the amount of information acquired faster in the FT-ICR MS spectrum.[32]

In the analysis of the raw spectra, we observed in Table 2 that the analysis time for an acquisition at the highest resolution (8 M) was 1.89 min. It generated a greater number of signals with S/N > 4 (43 228), but the total time for the 447 analyses was approximately 14 h. Noise filtering separates component signals from the background originating from the chemical matrix or instrumental interference. Thus, the spectra were corrected by eliminating signals with relative intensity below the noise.[33]

After this correction, the number of signals decreased significantly, making the difference between the data set size minimal and the time difference significant. Therefore, the resolution power chosen to perform the analysis was 1 M, which generated 221 signals with relative intensity above the noise and an analysis time of 3.8 h, only 1.1 h more than the fastest analysis of the resolution of 512 k, in which it presented only 169 signals above the noise.

After parameter optimization, 149 samples (97 SARS-CoV-2 positives and 52 negatives) were analyzed by MALDI FT-ICR MS (Figure S2). An independent PCA was performed as an exploratory data analysis. The PCA model built with SARS-CoV-2 positive and negative saliva samples showed 96.88% of the total variance, and the PC1 × PC2 × PC3 scores and loadings plot are shown in Figure 2. From the scores, it is possible to notice a tendency of separation between the two main groups, with the influence of m/z 525.4 and 598.5 in positive samples and 1410.4 and 2242.3 in negative samples.

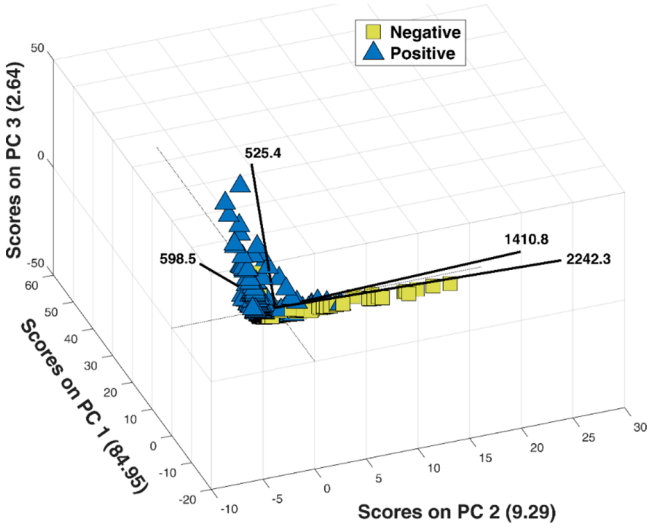


Figure 2. PCA scores and loadings plot of MALDI(+) spectra of saliva samples positive or negative for SARS-CoV-2.

Loadings that present vectors with an angle close to 90° to each other, that is, perpendicular vectors, indicate little or no intercorrelation. That is obeyed for the variables of m/z of 525.4 with 1410.8 and 2242.3, and m/z of 598.5 with 1410.8 and 2242.3. One could suggest that the variables of m/z 525.4 and 598.5 are independent of m/z 1410.8 and 2242.3. Also, they do not impact the variables of m/z 1410.8 and 2242.3. However, loadings that present vectors with angles close to zero among themselves, superimposed, present a strong intercorrelation. That explains the behavior of variables of m/z 525.4 with 598.5 in m/z, m/z 1410.8, and 2242.3, in which the variation of m/z 525.4 influences the m/z 598.5 and the variation of m/z 1410.8 influences the m/z 2242.3.

In the graph of the most important variables for the SVR nonlinear classification model (Figure 3), it is noted that the m/z of 525.4 with 598.5 showed higher intensity in the positive samples, while the m/z 1410.8 and 2242.3 showed higher intensity in negative samples. The dependence and independence of the respective variables are justified since the m/z vectors below 1000 influence positive samples and m/z above 1000 influence negative ones.

Subsequently, ML analysis was employed to identify the best-performing model for distinguishing positive and negative samples for COVID-19.

https://doi.org/10.1021/acs.jproteome.2c00148
J. Proteome Res. 2022, 21, 1868−1875

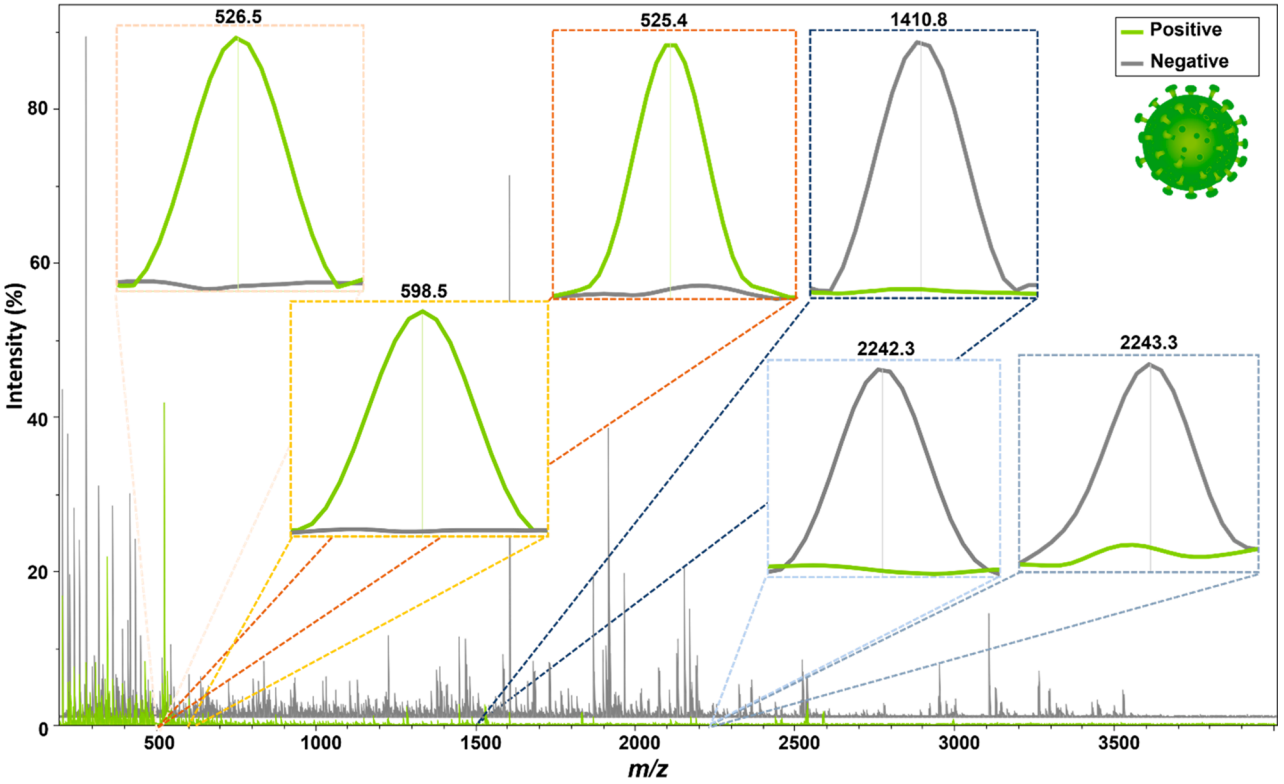


Figure 3. Characteristic signal in the individual MALDI(+) spectra among SARS-CoV-2 positive samples (green) versus negative samples (gray).

Table 3. Parameter Performance of SVM Models by MALDI Spectra with Fisher's Discriminant (FD)[a]

model	FD (%)	varsel	group	TP	TN	FP	FN	SENS (%)	SPEC (%)	ACC (%)	PPV (%)	PNV (%)	FPR (%)	FNR (%)	MCC (%)
1	30	780	cal	204	107	1	0	100.0	99.1	99.7	99.5	100.0	0.1	0.0	99.3
			test	87	42	6	0	100.0	87.5	95.6	93.6	100.0	12.5	0.0	90.5
2	85	2	cal	204	108	0	0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0
			test	84	33	15	3	96.6	68.8	86.7	84.9	91.7	31.3	3.5	70.7

[a]FD (Fisher's Discriminant); varsel (selected variables); TP (true positive); TN (true negative); FP (false positive); FN (false negative); SENS (sensitivity); ACC (accuracy); SPEC (specificity); PPV (prevalence positive value); PNV (prevalence negative value); FPR (false positive rate); FNR (false negative rate); MCC (Matthews correlation coefficient).

Despite the differences shown, no specific biomarkers for any positive or negative sample groups were found in the analyses performed by MALDI FT-ICR MS. The ML approach was then applied to analyze the data from the entire range of the selected m/z pattern, applying the FD (Fisher's Discriminant) for the variable selection. Two models showed better results than the built models, one with a FD of 30% and the other with 85%. When training the different models and selecting variables with a FD above 30%, Model 1 gave better performance, despite having a greater number of selected variables (780), with a sensitivity of 100%, specificity of 87.5%, and accuracy of 95.6%. However, model 2, which used 85% of the FD, selected only 2 variables, m/z 525.4 (FD = 0.70) and m/z 1410.8 (FD = 0.58) and presented a sensitivity of 96.6%, specificity of 68.7% and accuracy of 86.7%. The most important variables were the same and presented the same profile as observed in the PCA model. Herein, variables with higher m/z show a higher correlation with negative samples, while variables with lower m/z show a strong correlation with positive samples. This profile can also be seen in the raw spectra and after processing (Figure S2) either the positive or negative samples for SARS-CoV-2. There are more signals and a greater relative intensity at higher m/z for negative sample spectra. In comparison, positive samples have a higher intensity and lower signal numbers for m/z.

The models gave an excellent performance, and it is possible to discriminate the positive and negative classes since there is little spectral variance between samples of the same class. The results of model 1 (Table 3) show a FNR (false negative rate) of 0.0%, making the methodology used efficient in distinguishing positive samples for SARS-CoV-2 from negative ones in saliva in a noninvasive matrix, avoiding false-negative diagnoses.

After obtaining the supervised and unsupervised classification models, it was noted that signals of m/z below 1000 have a higher spectral intensity in positive samples and m/z above 1000 have a higher spectral intensity in negative samples.

Figure 3 represents the most important m/z for SVM models. A positive (green) and a negative (gray) sample were used. The higher intensities in m/z below 1000 are correlated with positive samples, and there is an intensity predominance for negative samples in m/z above 1000. The greater importance of variables was with lower m/z for positive samples (525.4 and 598.5) and higher m/z for negative samples (1410.8 and 2242.3). This can also be seen in the results of the PCA model, Figure 2.

https://doi.org/10.1021/acs.jproteome.2c00148
J. Proteome Res. 2022, 21, 1868−1875

On the basis of the above results, we compared other studies within the literature that also applied multivariate data analysis or ML to data obtained by the MALDI MS technique to study biological samples from patients suspected of being contaminated by SARS-CoV-2. Rocca et al. (2020) obtained preliminary results for detecting protein profiles directly from SARS-CoV-2 positive and negative nasopharyngeal swab samples, where accuracy, sensitivity, and specificity were 67.66%, 61.76%, and 71.72%, respectively.[34] In the study by Nachtigall et al. (2020), 362 nasal swab samples without prior preparation were analyzed using the MALDI−MS with ML. The model showed higher accuracy (93.9%) with 7% false positives and 5% false negatives.[21] Deulofeu et al. (2021) developed a methodology that combines MALDI-TOF MS with ML. The best model showed high performance in terms of accuracy, sensitivity, and specificity, reaching values greater than 90%.[19]

Unlike the studies mentioned above, the following investigations employed blood samples (plasma or serum). Lazari et al. (2021) obtained optimized results with an accuracy of 92%, sensitivity of 93%, and specificity of 92% in the analyzed data set.[35] Gomila et al. (2021) performed MALDI-TOF MS analysis of serum peptides in patients classified as mild, severe, critical, and healthy control. The ML approach was rated with 100% accuracy in clinically stable patients according to the severity and correctly predicted the outcome of unstable patients in all cases.[36] The study by Wan et al. (2021) developed a new approach to detect asymptomatic SARS-CoV-2 infection using metabolic patterns combined with ML. This approach provided 93.4% accuracy with only 5% false-negative and 7% false-positive rates.[37] Despite the excellent results found in the literature relating the MALDI MS technique with ML in biological samples from patients suspected of having COVID-19, the protocols for collecting these samples are still very invasive, uncomfortable, and have some risk of viral transmission for the operator who performs the procedure.

Thus, the use of saliva as a biological sample has already aroused interest in diagnostic studies for COVID-19, such as in the study by Chivte et al. (2021) and Costa et al. (2021), which identified potential biomarkers of infection by SARS-CoV-2.[11,24] The high accuracy of mass measurement of the FT-ICR MS reduces the probability of an incorrect assignment of a protein in the peptide mass profile. Additionally, the high resolution of the equipment provides a greater amount of variables that make the separation more accurate.[38] The absence of studies using MALDI FT-ICR MS in rapid diagnosis of biological samples makes this work of great value for developing new techniques with results concerning sensitivity, specificity, and accuracy superior to other techniques previously studied. However, the uniqueness of MS profiles among individuals requires the development of sophisticated biostatistical analysis, such as machine learning approaches, making our study promising in advancing research using the MALDI FT-ICR MS technique combined with ML.

CONCLUSIONS

SVM is efficient for diagnosing COVID-19 with a FNR (false negative rate) of 0.0%. Thus, the methodology created, besides being noninvasive, avoids false-negative diagnoses. In addition, the SVM had a sensitivity of 100%, specificity of 87.5%, and accuracy of 95.6%. Signals with m/z below 1000 have higher intensity in positive samples, and signals with m/z above 1000 have higher intensity in negative samples for the PCA and SVM models. The rapid screening method for patients with suspected COVID-19 developed in this work allows for a simple, safe, and noninvasive collection. The MALDI FT-ICR MS technique possessed high sensitivity and specificity for COVID-19 detection combined with machine learning methods. In optimizing the methodology, using 10 μL of trypsin and 2 h of digestion showed a higher TIC for all peptide sequences, streamlining the sample preparation process. As for the MALDI FT-ICR MS analysis, 1 M resolving power showed an excellent ratio of detected signals (221 signals) with a relative intensity above the noise and an individual analysis time of half a minute.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00148.

Table with health data of participants (Table S1); MALDI(+) mass spectra of saliva samples with and without digestion protein (Figure S1); Table with the number of signals detected in the analysis of the saliva samples with and without sample preparation (Table S2); Table with peptide sequence identified (Table S3); and MALDI(+) mass spectra used in SVM models (Figure S2) (PDF)

AUTHOR INFORMATION

Corresponding Author

Wanderson Romão − *Chemistry Department, Federal University of Espírito Santo, Vitória, ES 29040-090, Brazil; Science Department, Federal Institute of Education, Science, and Technology of Espírito Santo, Vila Velha, ES 29106-010, Brazil;* orcid.org/0000-0002-2254-6683; Email: wandersonromao@gmail.com

Authors

Camila M. de Almeida − *Chemistry Department, Federal University of Espírito Santo, Vitória, ES 29040-090, Brazil*

Larissa C. Motta − *Chemistry Department, Federal University of Espírito Santo, Vitória, ES 29040-090, Brazil*

Gabriely S. Folli − *Chemistry Department, Federal University of Espírito Santo, Vitória, ES 29040-090, Brazil*

Wena D. Marcarini − *Department of Physiological Sciences, Federal University of Espírito Santo, Vitória, ES 29040-090, Brazil*

Camila A. Costa − *School of Dentistry, Department of Stomatology (Oral Pathology), Federal University of Goiás, Goiânia, GO 74000-000, Brazil*

Ana C. S. Vilela − *School of Dentistry, Department of Stomatology (Oral Pathology), Federal University of Goiás, Goiânia, GO 74000-000, Brazil*

Valério G. Barauna − *Department of Physiological Sciences, Federal University of Espírito Santo, Vitória, ES 29040-090, Brazil*

Francis L. Martin − *Biocel UK Ltd., Hull HU10 6TS, U.K.*

Maneesh N. Singh − *Biocel UK Ltd., Hull HU10 6TS, U.K.*

Luciene C. G. Campos − *Department of Biological Science, Santa Cruz State University, Ilhéus, BA 45662-900, Brazil*

Nádia L. Costa − *School of Dentistry, Department of Stomatology (Oral Pathology), Federal University of Goiás, Goiânia, GO 74000-000, Brazil*

https://doi.org/10.1021/acs.jproteome.2c00148
J. Proteome Res. 2022, 21, 1868−1875

Paula F. Vassallo − *Clinical Hospital, Federal University of Minas Gerais, Belo Horizonte, MG 31270-901, Brazil*
Andrea R. Chaves − *Chromatography and Mass Spectrometry Laboratory, Institute of Chemistry, Federal University of Goiás, Goiânia, GO 74690-900, Brazil;* orcid.org/0000-0002-1600-1660
Denise C. Endringer − *Pharmaceutical Science Graduate Program, Universidade Vila Velha, Vila Velha, ES 29106-010, Brazil*
José G. Mill − *Department of Physiological Sciences, Federal University of Espírito Santo, Vitória, ES 29040-090, Brazil*
Paulo R. Filgueiras − *Chemistry Department, Federal University of Espírito Santo, Vitória, ES 29040-090, Brazil*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.2c00148

Author Contributions

Investigation: W.R., V.G.B., L.C.G.C., and P.R.F. Funding acquisition: J.G.M., V.G.B., N.L.C., L.C.G.C., and W.R. Methodology and sample preparation: C.M.A., L.C.M., W.D.M., C.A.C., A.C.S.V., V.G.B., N.L.C., L.C.G.C., and J.G.M. MS method development and data acquisition: C.M.A., L.C.M., L.C.G.C., V.G.B., and W.R. Data analysis: G.S.F. and P.R.F. Writing—review and editing: C.M.A., L.C.M., G.S.F. Manuscript review: F.L.M., M.N.S., L.C.G.C., N.L.C., A.R.C., P.F.V., D.C.E, P.R.F., J.G.M., V.G.B., and W.R.

Funding

This work was supported by the CAPES (Finance Code 1), FAPES (492/2021, 165/2021, 06/2019, and 151/2020), CNPq (422555/2018-5, 308541/2018-9, 310057/2020-5, 401870/2020-0, and 313500/2021-5), FAPESB-PPSUS (0027/2021), and UESC (#073.11012.2020.0007594-29) for their financial support.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank CAPES, FAPES, CNPq, FAPESB-PPSUS, and UESC for the financial support. The authors would also like to thank Núcleo de Competências em Química do Petróleo and LabPetro for using their installations and all the hospitals (Vila Velha Hospital, Hospital Universitário Cassiano Antônio Moraes, Pronto Atendimento da Glória, and Hospital Roberto Arnizaut Silvares), employees, nurses, doctors, and patients that accepted to participate in this study, as well as the municipality of Vila Velha and Secretaria Municipal de Saúde. Finally, the authors would like to thank Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi), and Secretaria Estadual de Saúde do Espírito Santo (SESA) for clinical patient data and RTPCR results.

■ ABBREVIATIONS

MALDI FT-ICR MS, matrix-assisted laser desorption/ionization Fourier-transform ion cyclotron resonance mass spectrometry; SVM, support vector machine; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; RT-PCR, reverse transcription−polymerase chain reaction; RAT, rapid antigen tests; MS, mass spectrometry; ML, machine learning; BSA, bovine serum albumin; CHCA, α-ciano-hydroxy-cinnamic acid; AmBic, ammonium bicarbonate; DTT, dithiothreitol; IAA, iodoacetamide; TFA, trifluoroacetic acid; PCA, principal component analysis; PRP, proline-rich protein; FD, Fisher's Discriminant; varsel, selected variables; TP, true positive; TN, true negative; FP, false positive; FN, false negative; SENS, sensitivity; SPEC, specificity; PPV, prevalence positive value; PNV, prevalence negative value; FPR, false positive rate; FNR, false negative rate; MCC, Matthews correlation coefficient.

■ REFERENCES

(1) Baloch, S.; Baloch, M. A.; Zheng, T.; Pei, X. The Coronavirus Disease 2019 (COVID-19) Pandemic. *Tohoku J. Exp. Med.* **2020**, *250* (4), 271−278.
(2) Jebril, N. World Health Organization Declared a Pandemic Public Health Menace: A Systematic Review of the Coronavirus Disease 2019 "COVID-19". *SSRN Electron. J.* **2020**, *24* (9), 2784−2795.
(3) Kucharski, A. J.; Russell, T. W.; Diamond, C.; Liu, Y.; Edmunds, J.; Funk, S.; Eggo, R. M.; Sun, F.; Jit, M.; Munday, J. D.; Davies, N.; Gimma, A.; van Zandvoort, K.; Gibbs, H.; Hellewell, J.; Jarvis, C. I.; Clifford, S.; Quilty, B. J.; Bosse, N. I.; Abbott, S.; Klepac, P.; Flasche, S. Early Dynamics of Transmission and Control of COVID-19: A Mathematical Modelling Study. *Lancet. Infect. Dis.* **2020**, *20* (5), 553−558.
(4) Corman, V. M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D. K. W.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M. L.; Mulders, D. G. J. C.; Haagmans, B. L.; Van Der Veer, B.; Van Den Brink, S.; Wijsman, L.; Goderski, G.; Romette, J. L.; Ellis, J.; Zambon, M.; Peiris, M.; Goossens, H.; Reusken, C.; Koopmans, M. P. G.; Drosten, C. Detection of 2019 Novel Coronavirus (2019-NCoV) by Real-Time RT-PCR. *Eurosurveillance* **2020**, *25* (3), 23−30.
(5) Lippi, G.; Simundic, A. M.; Plebani, M. Potential Preanalytical and Analytical Vulnerabilities in the Laboratory Diagnosis of Coronavirus Disease 2019 (COVID-19). *Clin. Chem. Lab. Med.* **2020**, *58* (7), 1070−1076.
(6) Kim, Y. G.; Yun, S. G.; Kim, M. Y.; Park, K.; Cho, C. H.; Yoon, S. Y.; Nam, M. H.; Lee, C. K.; Cho, Y. J.; Lim, C. S. Comparison between Saliva and Nasopharyngeal Swab Specimens for Detection of Respiratory Viruses by Multiplex Reverse Transcription-PCR. *J. Clin. Microbiol.* **2017**, *55* (1), 226−233.
(7) Sasaki, T.; Inoue, O.; Ogihara, S.; Kubokawa, K.; Oishi, S.; Shirai, T.; Iwabuchi, K.; Suzuki-Inoue, K. Detection of SARS-CoV-2 RNA Using RT-QPCR in Nasopharyngeal Swab, Saliva, Lingual, and Buccal Mucosal Swab. *Jpn. J. Infect. Dis.* **2022**, *75*, 102−104.
(8) Yamayoshi, S.; Sakai-Tagawa, Y.; Koga, M.; Akasaka, O.; Nakachi, I.; Koh, H.; Maeda, K.; Adachi, E.; Saito, M.; Nagai, H.; Ikeuchi, K.; Ogura, T.; Baba, R.; Fujita, K.; Fukui, T.; Ito, F.; Hattori, S. I.; Yamamoto, K.; Nakamoto, T.; Furusawa, Y.; Yasuhara, A.; Ujie, M.; Yamada, S.; Ito, M.; Mitsuya, H.; Omagari, N.; Yotsuyanagi, H.; Iwatsuki-Horimoto, K.; Imai, M.; Kawaoka, Y. Comparison of Rapid Antigen Tests for COVID-19. *Viruses* **2020**, *12* (12), 1420.
(9) Scohy, A.; Anantharajah, A.; Bodéus, M.; Kabamba-Mukadi, B.; Verroken, A.; Rodriguez-Villalobos, H. Low Performance of Rapid Antigen Detection Test as Frontline Testing for COVID-19 Diagnosis. *J. Clin. Virol.* **2020**, *129*, 104455.
(10) Ihling, C.; Tänzler, D.; Hagemann, S.; Kehlen, A.; Hüttelmaier, S.; Arlt, C.; Sinz, A. Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. *J. Proteome Res.* **2020**, *19* (11), 4389−4392.
(11) Chivte, P.; LaCasse, Z.; Seethi, V. D. R.; Bharti, P.; Bland, J.; Kadkol, S. S.; Gaillard, E. R. MALDI-ToF Protein Profiling as a Potential Rapid Diagnostic Platform for COVID-19. *J. Mass Spectrom. Adv. Clin. Lab* **2021**, *21*, 31−41.
(12) Frampas, C. F.; Longman, K.; Spick, M. P.; Lewis, H. M.; Costa, C. D. S.; Stewart, A.; Dunn-Walters, D.; Greener, D.; Evetts, G. E.; Skene, D.; Trivedi, D.; Pitt, A. R.; Hollywood, K.; Barran, P.; Bailey, M. J. Untargeted Saliva Metabolomics Reveals COVID-19 Severity. *medRxiv*, July 9, 2021, 21260080. DOI: 10.1101/2021.07.06.21260080.

(13) Spielmann, N.; Wong, D. Saliva: Diagnostics and Therapeutic Perspectives. *Oral Dis.* **2011**, *17* (4), 345−354.
(14) Kawas, S. A.; Rahim, Z. H.A.; Ferguson, D. B. Potential Uses of Human Salivary Protein and Peptide Analysis in the Diagnosis of Disease. *Arch. Oral Biol.* **2012**, *57* (1), 1−9.
(15) Baghizadeh Fini, M. Oral Saliva and COVID-19. *Oral Oncol.* **2020**, *108*, 104821.
(16) Nascimento, M. H. C.; Marcarini, W. D.; Folli, G. S.; da Silva Filho, W. G.; Barbosa, L. L.; Paulo, E. H. de; Vassallo, P. F.; Mill, J. G.; Barauna, V. G.; Martin, F. L.; de Castro, E. V. R.; Romão, W.; Filgueiras, P. R. Noninvasive Diagnostic for COVID-19 from Saliva Biofluid via FTIR Spectroscopy and Multivariate Analysis. *Anal. Chem.* **2022**, *94* (5), 2425−2433.
(17) Barauna, V. G.; Singh, M. N.; Barbosa, L. L.; Marcarini, W. D.; Vassallo, P. F.; Mill, J. G.; Ribeiro-Rodrigues, R.; Campos, L. C. G.; Warnke, P. H.; Martin, F. L. Ultrarapid On-Site Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High Sensitivity and Specificity. *Anal. Chem.* **2021**, *93* (5), 2950−2958.
(18) Deulofeu, M.; García-Cuesta, E.; Peña-Méndez, E. M.; Conde, J. E.; Jiménez-Romero, O.; Verdú, E.; Serrando, M. T.; Salvadó, V.; Boadas-Vaello, P. Detection of SARS-CoV-2 Infection in Human Nasopharyngeal Samples by Combining MALDI-TOF MS and Artificial Intelligence. *Front. Med.* **2021**, *8*, 398−410.
(19) Tran, N. K.; Howard, T.; Walsh, R.; Pepper, J.; Loegering, J.; Phinney, B.; Salemi, M. R.; Rashidi, H. H. Novel Application of Automated Machine Learning with MALDI-TOF-MS for Rapid High-Throughput Screening of COVID-19: A Proof of Concept. *Sci. Rep.* **2021**, *11* (1), 8219.
(20) Nachtigall, F. M.; Pereira, A.; Trofymchuk, O. S.; Santos, L. S. Detection of SARS-CoV-2 in Nasal Swabs Using MALDI-MS. *Nat. Biotechnol.* **2020**, *38* (10), 1168−1173.
(21) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72* (1−2), 248−254.
(22) Huang, C.-M.; Zhu, W. Profiling Human Saliva Endogenous Peptidome via a High Throughput MALDI-TOF-TOF Mass Spectrometry. *Comb. Chem. High Throughput Screen.* **2009**, *12* (5), 521−531.
(23) Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11* (1), 137−148.
(24) Costa, M. M.; Benoit, N.; Saby, F.; Pradines, B.; Granjeaud, S.; Almeras, L. Optimization and Standardization of Human Saliva Collection for MALDI-TOF MS. *Diagnostics* **2021**, *11* (8), 1304−1319.
(25) Lorenzo-Pouso, A. I.; Pérez-Sayáns, M.; Bravo, S. B.; López-Jornet, P.; García-Vence, M.; Alonso-Sampedro, M.; Carballo, J.; García-García, A.; Zalewska, A. Protein-Based Salivary Profiles as Novel Biomarkers for Oral Diseases. *Dis. Markers* **2018**, *2018*, 1.
(26) Bogdanov, B.; Smith, R. D. Proteomics by FTICR Mass Spectrometry: Top down and Bottom Up. *Mass Spectrom. Rev.* **2005**, *24* (2), 168−200.
(27) Cantú, M. D.; Carrilho, E.; Wulff, N. A.; Palma, M. S. Seqüenciamento de Peptídeos Usando Espectrometria de Massas: Um Guia Prático. *Quim. Nova* **2008**, *31* (3), 669−675.
(28) Ishikawa, S.; Ishizawa, K.; Tanaka, A.; Kimura, H.; Kitabatake, K.; Sugano, A.; Edamatsu, K.; Ueda, S.; Iino, M. Identification of Salivary Proteomic Biomarkers for Oral Cancer Screening. *In Vivo* **2021**, *35* (1), 541−547.
(29) Hardt, M.; Witkowska, H. E.; Webb, S.; Thomas, L. R.; Dixon, S. E.; Hall, S. C.; Fisher, S. J. Assessing the Effects of Diurnal Variation on the Composition of Human Parotid Saliva: Quantitative Analysis of Native Peptides Using ITRAQ Reagents. *Anal. Chem.* **2005**, *77* (15), 4947−4954.
(30) Lucchi, G.; Chambon, C.; Truntzer, C.; Pecqueur, D.; Ducoroy, P.; Schwartz, C.; Nicklaus, S.; Morzel, M. Mass-Spectrometry Based Characterisation of Infant Whole Saliva Peptidome. *Int. J. Pept. Res. Ther.* **2009**, *15* (3), 177−185.
(31) Hay, D. I.; Bennick, A.; Schlesinger, D. H.; Minaguchi, K.; Madapallimattam, G.; Schluckebier, S. K. The Primary Structures of Six Human Salivary Acidic Proline-Rich Proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-F). *Biochem. J.* **1988**, *255* (1), 15−21.
(32) Ferey, J.; Marguet, F.; Laquerrière, A.; Marret, S.; Schmitz-Afonso, I.; Bekri, S.; Afonso, C.; Tebani, A. A New Optimization Strategy for MALDI FTICR MS Tissue Analysis for Untargeted Metabolomics Using Experimental Design and Data Modeling. *Anal. Bioanal. Chem.* **2019**, *411* (17), 3891−3903.
(33) Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y. Chemometric Methods in Data Processing of Mass Spectrometry-Based Metabolomics: A Review. *Anal. Chim. Acta* **2016**, *914*, 17−34.
(34) Rocca, M. F.; Zintgraff, J. C.; Dattero, M. E.; Santos, L. S.; Ledesma, M.; Vay, C.; Prieto, M.; Benedetti, E.; Avaro, M.; Russo, M.; Nachtigall, F. M.; Baumeister, E. 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* **2020**, *286*, 113991.
(35) Lazari, L. C.; de Rose Ghilardi, F.; Rosa-Fernandes, L.; Assis, D. M.; Nicolau, J. C.; Santiago, V. F.; Dalçóquio, T. F.; Angeli, C. B.; Bertolin, A. J.; Marinho, C. R. F.; Wrenger, C.; Durigon, E. L.; Siciliano, R. F.; Palmisano, G. Prognostic Accuracy of MALDI-TOF Mass Spectrometric Analysis of Plasma in COVID-19. *Life Sci. Alliance* **2021**, *4* (8), 1−12.
(36) Gomila, R. M.; Martorell, G.; Fraile-Ribot, P. A.; Doménech-Sánchez, A.; Albertí, M.; Oliver, A.; García-Gasalla, M.; Albertí, S. Use of Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry Analysis of Serum Peptidome to Classify and Predict Coronavirus Disease 2019 Severity. *Open Forum Infect. Dis.* **2021**, *8* (6), 1−8.
(37) Wan, Q.; Chen, M.; Zhang, Z.; Yuan, Y.; Wang, H.; Hao, Y.; Nie, W.; Wu, L.; Chen, S. Machine Learning of Serum Metabolic Patterns Encodes Asymptomatic SARS-CoV-2 Infection. *Front. Chem.* **2021**, *9*, 746134.
(38) Yang, H. J.; Park, K. H.; Shin, S.; Lee, J. H.; Park, S.; Kim, H. S.; Kim, J. Characterization of Heme Ions Using MALDI-TOF MS and MALDI FT-ICR MS. *Int. J. Mass Spectrom.* **2013**, *343*−*344*, 37−44.

https://doi.org/10.1021/acs.jproteome.2c00148
J. Proteome Res. 2022, 21, 1868−1875