Diagnosing COVID-19 in human sera with detected immunoglobulins IgM and IgG by means of Raman spectroscopy

Ana Cristina Castro Goulart1 | Renato Amaro Zângaro1,2 | Henrique Cunha Carvalho2 | Landulfo Silveira Jr1,2

1Biomedical Engineering Program, Universidade Anhembi Morumbi – UAM, São Paulo, State of São Paulo, Brazil
2Laboratory of Vibrational Spectroscopy, Center for Innovation, Technology and Education – CITÉ, São José dos Campos, State of São Paulo, Brazil

Correspondence
Landulfo Silveira Jr., Biomedical Engineering Program, Universidade Anhembi Morumbi – UAM, Rua Casa do Ator, 275, São Paulo, State of São Paulo 04546-001, Brazil.
Email: landulfo.silveira@gmail.com; lsjunior@anhembi.br

Funding information
National Council for Scientific and Technological Development, Grant/Award Number: 306344/2017-3; São Paulo Research Foundation, Grant/Award Number: 2009/01788-5

Abstract
The severe COVID-19 pandemic requires the development of novel, rapid, accurate, and label-free techniques that facilitate the detection and discrimination of SARS-CoV-2 infected subjects. Raman spectroscopy has been used to diagnose COVID-19 in serum samples of suspected patients without clinical symptoms of COVID-19 but presented positive immunoglobulins M and G (IgM and IgG) assays versus Control (negative IgM and IgG). A dispersive Raman spectrometer (830 nm, 350 mW) was employed, and triplicate spectra were obtained. A total of 278 spectra were used from 94 serum samples (54 Control and 40 COVID-19). The main spectral differences between the positive IgM and IgG versus Control, evaluated by principal component analysis (PCA), were features assigned to proteins including albumin (lower in the group COVID-19 and in the group IgM/IgG and IgG positive) and features assigned to lipids, phospholipids, and carotenoids (higher the group COVID-19). Features referred to nucleic acids, tryptophan, and immunoglobulins were also seen (higher the group COVID-19). A discriminant model based on partial least squares regression (PLS-DA) found sensitivity of 84.0%, specificity of 95.0%, and accuracy of 90.3% for discriminating positive Ig groups versus Control. When considering individual Ig group versus Control, it was found sensitivity of 77.3%, specificity of 97.5%, and accuracy of 88.8%. The higher classification error was found for the IgM group (no success classification). Raman spectroscopy may become a technique of choice for rapid serological evaluation aiming COVID-19 diagnosis, mainly detecting the presence of IgM/IgG and IgG after COVID-19 infection.

KEYWORDS
COVID-19, diagnosis, immunoglobulins, Raman spectroscopy, serum

1 | INTRODUCTION

Since the beginning of the COVID-19 (COronaVIrus Disease 19) pandemic in early 2020, the World Health Organization has been advocating the massive testing of the population for controlling the infection by the SARS-CoV-2 virus (Severe Acute Respiratory Syndrome–CoronaVirus-2, the “new coronavirus”).[1] Massive
testing is extremely important to detect positive cases among the general population in order to promote adequate isolation to avoid disease spread (vertical isolation), thus minimizing the need for lockdowns and maintaining the economy losses to an acceptable level.\cite{2}

Therefore, a rapid, accurate, label-free, and cost-effective technique for massive testing is desirable worldwide.

Currently, the gold standard technique for the detection of the COVID-19 in suspected cases is based on the RT-PCR (Reverse Transcription–Polymerase Chain Reaction) test\cite{3} which identifies the virus genetic material (RNA, Ribonucleic Acid) in nasopharyngeal and oropharyngeal samples collected using a swab.\cite{4,5} Despite the widespread use, the RT-PCR technique has some disadvantages such as the need for specific physical space, equipments, and trained personnel to perform the assay, higher costs compared to serological tests for antibody detection (Immunoglobulins, Ig), discomfort that may cause sampling error, and they can present a false negative (Immunoglobulins, Ig), discomfort that may cause sampling error, and they can present a false negative result.\cite{5,6}

The detection of immunoglobulins M and G (IgM and IgG) is also employed for rapid diagnose of COVID-19 and population screening. These antibodies are molecules produced by lymphoid cells in response to the presence of antigens as a result of infectious process in order to recognize and assist in the destruction of the virus particles such as the SARS-CoV-2. The IgM is the first antibody produced in response to the viral infectious process and represents around 10% of the body’s immunoglobulins; since it can be detected after the fourth day of infection, peaking on the twentieth day, the presence of IgM suggests acute phase of the disease with possible presence of the virus in the body. The IgG represents 70–75% of the total immunoglobulins in the blood; it appears after the seventh day and reaches its maximum level around the twenty-fifth day, persisting in the serum for months to help preventing further infections by the same agent. Therefore, IgG is commonly used to indicate past infection.\cite{7,8} Despite researches showed that IgM may persist for months,\cite{8,9} there are still no studies claiming about permanent immunity\cite{10} mostly because of the novelty of the COVID-19. In addition, a positive correlation can be found between the antibody levels and disease severity.\cite{8,11}

Rapid immunochromatographic (ICG) tests are able to detect IgG and IgM antibodies specific to the SARS-CoV-2 virus using recombinant antigen for antigen/antibody conjugation in blood or serum. These tests can be used for mass testing, screening, and field research with minimal training for users and with results released in a short period of time, approximately 10 min.\cite{17} Commercial ICG tests from leading brands are adequate for a methodology of rapid identification of individual SARS-CoV-2 infection even though asymptomatic, with sensitivity of 65–99% and specificity of 92–100% for IgM and IgG, respectively.\cite{12} Since these tests present variable quality, studies indicate some results do not correlate with RT-PCR in the acute phase of the infection, especially when tested between 8 and 11 days after the onset of symptoms\cite{10,13,15}; nevertheless, they are widely used in order to identify late phase IgG antibodies in individual exposed to SARS-CoV-2. False-negative results with RT-PCR may occur due to viral evolution.\cite{16} Although the COVID-19 pandemic has been underway for a year, many studies are still needed to understand the immune response against SARS-CoV-2, and new methodologies must therefore emerge in order to obtain reliable diagnostic results.\cite{5}

Vibrational spectroscopic techniques such as Raman spectroscopy offers several advantages over molecular biochemical methods for the analysis of biological samples such as serum, including rapidness in obtaining the diagnostic, no need for sample preparation or no use of reagents (label-free), and assessment of biochemical information in a drop of sample. Raman spectroscopy is a nondestructive analytical technique that provides information about the molecular composition of the materials studied requiring a minimum amount of sample, permitting diagnosis in a few minutes using proper statistical and computational tools.\cite{17}

Recent studies showed that Raman spectroscopy could be used to quantify biomarkers of kidney disease such as urea and creatinine in serum,\cite{18} to quantify glucose and lipid components of serum,\cite{19} and to quantify prostate-specific antigen values in serum of prostate cancer patients\cite{20} aiming diagnosis. Very recently, two reviews thoroughly explore the potential of Raman spectroscopy for diagnosis of breast and oral cancers,\cite{21,22} and a research paper proposed the detection of infectious diseases such as hepatitis B in plasma.\cite{23}

Recently, two reviews thoroughly explore the potential of Raman spectroscopy and others optical technologies in detecting viruses and prospect some applications for the detection of infectious diseases like COVID-19.\cite{24,25} Up to date, six research papers have been published as a goal to clinically detect SARS-CoV-2 or diagnose COVID-19 using vibrational spectroscopic techniques such as Attenuated Total Reflection–Fourier Transform Infrared Spectroscopy (ATR-FTIR),\cite{26,27} Surface Enhanced Infrared Spectroscopy (SERS),\cite{28,29} and dispersive, near-infrared Raman spectroscopy.\cite{30,31} The main goal of these studies was the discrimination between healthy and infected subjects aiming diagnosis.

Studies indicate that the combined detection of IgG and IgM can be of great importance in improving the
detection of positive cases of the disease\cite{29,32} and the study of serum based on Raman spectroscopy may reveal the presence of increased amounts of immunoglobulins in the serum due to viral infection. Therefore, this study aims the diagnosis of COVID-19 by means of Raman spectroscopy and the multivariate technique Principal Component Analysis (PCA) and Partial Least Squares (PLS) for discriminant analysis of Raman features in serum samples of individuals with serological evaluation where immunoglobulins IgM and IgG were detected. We also present an exploratory analysis based on PCA as a tentative description of the possible biochemical differences of the serum positive for COVID-19 in different stages of the infection (as seen by the type of Ig detected—IgM and IgG) with regard the presence of spectral features assigned to proteins, amino acids, nucleic acids, lipids, and carotenoids.

2 | MATERIALS AND METHODS

2.1 | Serum samples

The study was approved by the Ethics and Research Committee of Universidade Anhembi Morumbi (UAM), protocol No. 26691419.6.0000.5492. Blood samples were collected in the Diagnostic Medicine Laboratory of CIPAX, São José dos Campos, SP, by venipuncture using a closed system and tubes with a separating gel (Vacutainer\textsuperscript{®}, BD Diagnostics, Franklin Lakes, NJ, USA). A total of 94 samples were collected from asymptomatic patients who were performing other laboratory tests. An amount of 3 mL of blood was extracted, and the samples were then centrifuged at 3,000 rpm for 10 min to obtain the serum. Then, the serum samples were sent to the sectors responsible for carrying out the immunochromatographic (ICG) test.

The collected samples were subjected to ICG assay tests using the OnSite\textsuperscript{™} COVID-19 IgM/IgG rapid test kit (CTK Biotech Inc., Poway, CA, USA). This test presents sensitivity of 78.0% and specificity of 99.4% for IgM, sensitivity of 96.8%, and specificity of 100% for IgG.\cite{33} Among the samples with negative and positive results, 54 presented negative results for COVID-19 and 40 presented positive results, being six positive for antibodies IgM, 13 positive for both IgM and IgG, and 21 positive for IgG. All positive results were repeated and confirmed. After the conclusion of the analytical tests, the serum samples were placed in thermal boxes, with temperatures between 2°C and 8°C, and transported to the laboratory, being subjected to Raman spectroscopy analysis. Table S1 presents the demographic and clinical information associated to the COVID-19 positive subjects.

2.2 | Raman spectroscopy

At the time of spectroscopic analysis, 80 μL of serum was pipetted into the orifice of an aluminum sample holder (lab-made device with 12 wells, 5 mm diameter each well, and about 80 mm length, 30 mm width, and 10 mm height). The readings of the Raman spectra were performed in triplicate with 30 s collection time (3 s scanning and 10 accumulations) using a dispersive Raman spectrometer (Dimension P1 model, Lambda Solutions Inc., MA, USA) with 830 nm excitation and 350 mW laser power. A total of 278 spectra were acquired, being 159 for negative and 119 for positive COVID-19 samples (18 positive for antibodies IgM, 39 positive for both IgM and IgG, and 62 positive for IgG). The spectrometer has an estimated spectral resolution of 4 cm\textsuperscript{-1} in the spectral range between 400 and 1,800 cm\textsuperscript{-1}.

The data stored in the computer were subjected to preprocessing in order to subtract the baseline due to fluorescence background and scattering, and for this, a seventh-order polynomial was fitted to the entire spectrum range and subtracted. The cosmic rays were removed manually and the spectra normalized by the area under the curve (1-norm normalization). Mean spectra of each group were then obtained for comparison purposes.

2.3 | PCA and PLS for features extraction (exploratory analysis) and discrimination

The Raman spectra were submitted to PCA in order to unveil the spectral differences between the two groups related to the changes in the biochemical constitution of the serum due to COVID-19 (exploratory analysis). Also, the PCA variables, markedly the PCA loadings, the PCs, were used in a Discriminant Analysis (DA) model. The DA model was also implemented via PLS regression using the whole spectral information instead of the selected PCs. Multivariate models such as PCA and PLS have been used for classification of Raman spectra of sera in normal and anemias\cite{34} and for quantification of blood analytes in human serum for diagnosis.\cite{19,20} A Kolmogorov-Smirnov normality test was applied to check the normality of the PCs. Student’s t test (with Welch correction whenever needed) or Mann–Whitney U test were applied to the PCs in order to identify significant differences between Control and COVID-19 groups; ANOVA or Kruskal-Wallis (nonparametric ANOVA) tests were applied to the PCs in order to identify significant differences between Control and IgM+,
IgM+/IgG+, and IgG+ groups. The *P* value was considered significant between the groups when *P* < 0.05.

Principal component's features extraction was performed in Matlab® 7.4.0 (R2007a) (Mathworks Inc., Natick, MA, USA) through the function “princomp.m”. PCA variables Scores and Loadings (PCs) were extracted using the dataset, considering each spectrum a sample; the Scores resemble Raman spectra and represent the axis of spectral variance (in order of appearance), and the PCs represent the intensities of each Score in the original spectrum. Discriminant analysis models using both PCA and PLS (PCA-DA and PLS-DA) were performed using the routine Chemoface (www.ufla.br/chemoface/)[35] using the “leave-one-out” cross validation. Discrimination was tested without treatment and with first-order derivative filter.

### 3 | RESULTS AND DISCUSSION

#### 3.1 | Raman spectra of negative Control and positive immunoglobulins IgM and IgG

The human serum is composed mainly by proteins (albumin: 70% by weight; immunoglobulins IgG: 14%, IgA: 3%, IgM: 2%; transferring: 6%; fibrinogen: 2%); lipids (phospholipids; glycerolipids), and cholesterol [36]; therefore, the study of serum based on Raman spectroscopy may reveal the presence of increased amounts of immunoglobulins and changes in other metabolites in sera due to SARS-CoV-2 infection.

Figure 1 presents the mean Raman spectra of the samples in the Control and COVID-19 groups. The spectra show Raman features assigned to the serum constituents, mainly albumin. It was seen small differences between both groups; therefore, a more detailed exploratory analysis by PCA was performed to identify particular spectral changes associated to differences in the molecular constitution of the serum of the COVID-19 subjects.

#### 3.2 | Exploratory analysis by PCA

The exploratory analysis by means of the PCA features showed the differences in the biochemical composition of the groups Control versus COVID-19 evidenced by Raman peaks assigned to proteins, amino acids, nucleic acids, and lipids. Figure 2 shows the first six Scores, and Figure 3 shows the first six Loadings (PCs) as well as the statistical significance of the comparisons (non-parametrical tests Mann-Whitney and Kruskal-Wallis); they represent together 98.9% of the spectral variation. Score 1 represents alone 98.3% of the spectral variation, indicating low variability of the serum composition. The presence of several spectral features in the remnant five scores (0.6%) reveals low concentration in serum but complex nature of the organism's response following the infection with SARS-CoV-2. Figure S1 shows the scatterplot of the loadings seen in Figure 3, and Figure S2 shows the binary scatterplot of the loadings PC1 X PC2 and PC2 X PC4 seen in Figure 3.

Score 1 (Figure 2) showed spectral characteristics related to the serum; PC1 (Figure 3) showed a significant difference between groups, being of lower intensity for both COVID-19 IgM+ and IgG+ versus Control. This suggests a decrease in circulating albumin for the groups with COVID-19, which may be due to the inflammatory process. Studies have shown a decrease in albumin and other serum proteins due to the acute phase of infectious processes caused by increased capillary permeability (exudative phase, where more albumin enters the extravascular space), decreased synthesis in response to osmotic colloidal pressure, and decreased hepatic synthesis in response to inflammatory cytokines, especially interleukin.[39]

Score 2 (Figure 2) showed positive spectral characteristics that can be attributed to nucleic acids and negative characteristics related to proteins; PC2 (Figure 3) showed...
a significant difference between the groups, being of greater intensity (positive loading) for COVID-19 versus Control and for IgM+/IgG+ versus Control, IgM+ and IgG+. This may suggest an increase in the genetic material (RNA) of viruses for groups with COVID-19 (increased viral load), especially when the infection is in the transition phase (IgM+/IgG+); the negative characteristics of the proteins suggest a decrease in proteins due to the exudative phase with lower albumin values, as seen in PC1, and a general decrease in protein production that can be explained by the presence of the Ndp1 protein encoded in the presence of SARS-CoV-2, which binds to the 40 s ribosomal subunit, preventing its binding with messenger RNA and affecting normal protein production. This may explain the decreased protein characteristics in the IgM+/IgG+ group in Score 2. It would be expected that the low circulating protein would also occur in the IgM+ group due to the probable viral presence; however, as the selected patients were asymptomatic, they supposedly would have a lower viral load collected.

Score 3 (Figure 2) showed positive and negative spectral features that can be assigned to proteins: immunoglobulins, amino acids (glycine, tyrosine, and tryptophan), and nucleic acids (from RNA virus); PC3 (Figure 3) showed no significant difference between the groups; although not significant, the IgM+ group showed higher intensity (positive Loading) compared to the other groups. This suggests that virus genetic material and Ig are present in the acute phase of the infection (IgM+) for some patients of the groups with COVID-19 and may be detectable; in addition, the amino acids glycine and tyrosine may be in lower intensity in the IgM+ group than in the other groups. The glycine may be reduced in inflammatory processes, and the Ig contains tryptophan residues, an essential amino acid capable of producing bioactive compounds providing an anti-inflammatory response, where it promotes the attraction of macrophages in addition to promoting neurological and immune responses, among other mechanisms, being able to regress the disease. A deregulation in the nitrogen metabolism due to viral infection may occur, with consequent alteration in the production of amino acids. The Raman spectrum of a typical IgG has numerous features that can be easily assigned to both secondary and tertiary structure conformation, being the disulfide conformations (regions of ~500–550 cm⁻¹), tyrosine (region between 800–1,000 cm⁻¹, 1,200–1,400 cm⁻¹, and 1,600–1,700 cm⁻¹), and tryptophan the most relevant features for tertiary structure. Also, in patients with dengue virus, it was found increase in the Raman peaks associated to immunoglobulins, adenosine diphosphate, and hemoglobin.

**FIGURE 2** Plot of the first six principal components Scores. Biochemical assignments for the Raman features seen in Scores are discussed in the text.
Score 4 (Figure 2) showed negative spectral features related to phospholipids as well as carotenoid features; PC4 (Figure 3) showed significant difference between the groups, being higher intensity (negative Loading) for both COVID-19 and IgM+/IgG+ versus Control; a not significant higher intensity (negative Loading) was observed for the group IgM+. Other mechanisms associated to lipid features may be the increase of the arachidonic acid, platelet-activating factor, C-reactive protein,[45] associated to the inflammatory response for the groups with COVID-19, particularly when the infection is in the transition stage (IgM+/IgG+). The presence of carotenoid peaks at the intermediate stage of the COVID-19 infection (group IgM+/IgG+), negative Loading) in the present study using ICG assay is contradictory to the nonsignificant features seen in Score 3 in the serum of RT-PCR positive for COVID-19 in Goulart et al.,[31] suggesting that presence of carotenoids may be an important discrimination feature of COVID-19. The study of Khan et al.[47] showed that patients with dengue virus presented a decrease in peaks referred to carotenoids compared to non-infected ones.

Score 5 (Figure 2) showed positive spectral features related to methylene/methyl and nucleic acids and negative spectral features related to tryptophan; PC5 (Figure 3) showed no significant difference between the groups; although not significant, the IgM+ group showed higher intensity (positive Loading) of nucleic acids for the IgG+ group and higher intensity (negative Loading) of tryptophan compared to the other groups, corroborating the findings of not significant higher intensity of tryptophan (i.e., presence of Ig) in the group IgM+ seen in Score 3. The methylene/methyl features (not significantly higher for the IgG+) may be associated to the CH modes
of the carbohydrates in the IgG. The presence of nucleic acid features in the IgG+ group may suggest the occurrence high viral load even in the late stage of the infection of asymptomatics.\cite{48} Increased levels of glucose were found in individuals with a positive result for COVID-19 due to stress hyperglycemia, even in patients without diagnosed diabetes.\cite{45,49}

Score 6 (Figure 2) showed positive spectral features which may be assigned to tyrosine and negative features related to lipids; PC6 (Figure 3) showed significant difference between the groups, being the negative Loading suggesting significantly higher intensity of lipids and significantly lower intensity of tyrosine for both COVID-19 and COVID-19 IgM+/IgG+ versus Control, suggesting a decrease in amino acids, particularly tyrosine, due to virus and increase in blood circulating lipids (lipid metabolites associated to inflammatory process due to infections) for the groups with COVID-19, and especially when the infection is in the intermediate stage (IgM+/IgG+).

This exploratory analysis evidenced that the COVID-19 changed the Raman spectral profile of the serum and the biochemical changes depend on the status of the infection. Mahmood et al.\cite{50} found that protein peaks, including the ones assigned to IgG, were higher in sera samples positive for dengue. The same study showed decrease in peaks associated to tyrosine and increase of peaks associated to lipids (phospholipids) for dengue, thus evidencing changes in body metabolism due to virus infection. Goulart et al.\cite{31} showed that sera of positive COVID-19 subjects diagnosed by RT-PCR presented an increase in lipids, nitrogen compounds such as amines, amides and urea, presence of nucleic acid, and decrease of proteins and amino acids. Interesting to observe that the group IgM+/IgG+ was the one with the highest significant difference compared to the other groups (PC2 and PC4) and principally to Control (PC6). Therefore, lower levels of proteins and amino acids and higher levels of nucleic acids, tryptophan, lipids (including phospholipids), and carotenoids, seen in the Raman features of Scores 2, 4, and 6, are well aligned with changes in the metabolism of positive patients and the presence of viral load\cite{40,42–45,48} already confirmed by our previous study.\cite{31} Thus, it is evident that Raman spectral features seen by PCA showed changes in the biochemistry of the serum and the presence of genetic material and antibodies in subjects infected with SARS-CoV-2.

The nonessential amino acid glycine appeared to be in low concentration in some IgM+ and IgM+/IgG+ patients despite not significant difference (positive PC3, negative Score 3). This amino acid is responsible for providing therapeutic, anti-inflammatory effects, regulation of the immune response.\cite{42} In addition, glycine has been used as a therapeutic supplement for patients with COVID 19 against the tissue damage and the cytokine storm, since it holds a cytoprotective and anti-inflammatory role.\cite{42,51}

This study showed nonsignificant increase in Trp features for the IgM+ and IgM+/IgG+ groups (positive PC3, positive Score 3; negative PC5, negative Score 5). The spectral features of tryptophan and phenylalanine, two nonessential aromatic amino acids, were found to be in high amount in the Raman spectra of saliva from positive COVID-19 subjects\cite{28} in the same two peaks were also identified as characteristic signals from viruses of the coronavirus family, probably being involved in the viral protein structure or with the interactions with physiologically expressed molecules.\cite{52} Recent studies conducted on SARS-CoV-2 and on other types of coronaviruses demonstrated the important presence of aromatic amino acids, including tryptophan, in the virus spike glycoproteins, identifying the so-called tryptophan-rich regions involved in the interaction between the virus and the receptor angiotensin converting enzyme type 2 (ACE2).\cite{28,53,54}

The increased bands of phospholipids for IgM+/IgG+ group (Score 4) may be due to the high amount of sphingolipids and lyssolecithin, which were observed in a plasma metabolomic and lipidomic study, with a 3.5- to 5.5-fold increase for COVID-19 positive patients.\cite{55} Zhang et al.\cite{26} showed a strong near-infrared absorption band centered at 1,078 cm\(^{-1}\) in COVID-19 patients that may reflect the abundance of sphingolipids in serum. Immunoglobulin secondary structure (\(\beta\)-sheet), with peaks centered at 1,650 and 1,670 cm\(^{-1}\) (Score 3), may provide hints toward the proportion of albumin and immunoglobulins in human sera, respectively; despite not significant, there was a visual negative correlation of the proportion of immunoglobulin versus albumin and PC3 intensity following the infection status.

### 3.3 Discrimination and classification between Control versus immunoglobulins IgM and IgG by PCA-DA and PLS-DA

The first six PCs were used for PCA-DA discrimination model with leave-one-out cross validation. PLS-DA discrimination model was also performed with the maximum of 10 latent variables; it was selected the number of latent variables that produced the best classification, and it was different depending on the type of grouping (Control vs. COVID-19 and Control vs. IgM+ vs. IgM+/IgG+ vs. IgG+). A first-order derivative filter (second-order polynomial, 12 data points) was applied to the Raman dataset since it returned better discrimination.
than no processing. Table 1 shows the confusion matrix for the discrimination of Raman spectra of sera using PLS-DA, and Table 2 presents the results of the sensitivity, specificity, and accuracy for the classification of Control versus positive COVID-19 sera and Control versus different antibodies (IgM, IgM/IgG, and IgG) sera using the PLS-DA discrimination. The results of the PCA-DA were worse compared to the PLS-DA (accuracy of 73.7% for Control vs. COVID-19 and accuracy of 65.8% for Control vs. IgM+/IgM+/IgG+/IgG+).

Vibrational spectroscopy techniques have been employed for COVID-19 diagnosis. ATR-FTIR in serum have been used by Zhang et al.\cite{26} to differentiate COVID-19 from normal controls and some common respiratory viral infections or inflammation with PLS-DA analysis, reaching sensitivity of 87% and specificity of 98% depending on the decision threshold of the Receiver Operating Characteristic (ROC) curve; authors found spectral features of antibodies and serum phospholipids. Another ATR-FTIR study was done by Barauna et al.,\cite{27} where Genetic Algorithm–Linear Discriminant Analysis (GA-LDA) detected SARS-CoV-2 particles in pharyngeal swabs with known PCR results. The authors found sensitivity of 95% and specificity of 89% using 61 negatives and 20 positives in the validation dataset. Raman spectroscopy has been used by other authors. Carlomagno et al.\cite{28} evaluated saliva of RT-PCR positive subjects by SERS to significantly discriminate patients with a current infection by COVID-19 from healthy subjects and/or subjects with a past infection (with or without symptoms), with accuracy ranging from 87.6% to 97.8% depending on the arrangement of the Control versus COVID-19 groups and the discrimination technique used. Liu et al.\cite{29} developed a SERS-based lateral flow immunoassay biosensor for the detection of anti-SARS-CoV-2 IgM/IgG in clinical samples, by immobilizing anti-human IgM and IgG and characterizing the immunoglobulins; authors reached 100% of specificity and accuracy. Yin et al.\cite{30} used dispersive Raman spectroscopy in sera of confirmed COVID-19 patients, suspected cases, and healthy individuals as a control groups to build a diagnostic algorithm based on machine learning method—Support Vector Machine (SVM) applied to the spectrum dataset. Goulart et al.\cite{31} also employed dispersive Raman spectroscopy associated to PCA and linear discriminant analysis to discriminate serum of RT-PCR positive patients from negative ones; authors found that selected principal component loadings could classify spectra of COVID-19 patients with sensitivity of 87% and specificity of 100%.

Previous studies aiming diagnosis of virus infection by SARS-CoV-2 using vibrational spectroscopy found spectral differences related to the biochemical compounds of infection (nucleic acids, immune system response and metabolic alterations) which are present in sera.\cite{26,31} The present study using the ICG test for IgM and IgG showed sensitivity of 84.0%, specificity of 95.0%, and accuracy of 90.3% for diagnosing COVID-19 versus Control in sera. When evaluating the COVID-19 groups

| TABLE 1 | Confusion matrix for the discrimination of sera within the groups Control versus COVID-19 and within the groups Control versus IgM and IgG (IgM+, IgM+/IgG+, and IgG+) |
|---------|--------------------------------------------------------------------------------------------------|
|         | Control | IgM+ | IgM+/IgG+ | IgG+ |
| Control (159) | 151   | 0    | 0         | 4   |
| COVID-19 (119) | 19    | 100  |           |     |
| Raman diagnosis using PLS-DA (6 latent variables) |  |  |  |  |
| Control | 155 | 0 | 0 | 4 |
| IgM+ (18) | 11 | 0 | 3 | 4 |
| IgM+/IgG+ (39) | 5 | 0 | 27 | 7 |
| IgG+ (62) | 11 | 0 | 6 | 45 |

Note: The number of latent variables used in each PLS-DA model is also mentioned.

| TABLE 2 | Sensitivity, specificity, and percentage of correct classification (accuracy) for the classification of Control sera versus positive COVID-19 and its different antibodies (IgM, IgM/IgG, and IgG) sera using the PLS-DA discrimination |
|---------|--------------------------------------------------------------------------------------------------|
|         | Control x COVID-19 | Control x IgM+ x IgM+ /IgG+ x IgG+ |
| No. of latent variables | 6 | 8 |
| Sensitivity | 84.0% | 77.3%a |
| Specificity | 95.0% | 97.5%a |
| Accuracy | 90.3% | 81.7% and 88.8%a |

aConsidering immunoglobulins IgM+, IgM+/IgG+, and IgG+ as a single COVID-19 group.
individually, the accuracy reached 81.7%. The group IgM+ presented the higher number of classification error proportional to the number of sample spectra (Table 1). In fact, all sample spectra were misclassified (61% were classified as Control, and 39% were classified as IgM/IgG and IgG). This may be explained by the early phase of the infection, where small and not detectable biochemical changes may still “escape” from being detected in sera by the Raman analysis and different organism’s response to the infection in these asymptomatic patients (most of the PCs presented high standard error bars, not allowing enough significance in the statistical tests); also, the small number of samples in this group (6 serum samples and 18 sample spectra) may difficult the model to identify very tiny but consistent features associated to the early infection. Therefore, in such case, the Raman-based model failed to detect the changes in the serum associated to the presence of the early virus infection.

### 3.4 Impact of the findings in COVID-19 diagnosis and pandemic control

The RT-PCR test is the gold-standard for diagnosing the RNA virus, and the ICG assay, despite not being the gold standard, is commonly used for screening and testing the immune response due to its speed, ease of use, and low cost. RT-PCR and ICG tests have high sensitivity and sensitivity (for RT-PCR: 95% sensitivity and 100% specificity,[56] and for ICG: sensitivity of 78% for IgM and 96.8% for IgG, and specificity of 99.4% for IgM and 100% for IgG[33]). It is important to note the lower sensitivity for the ICG when detecting IgM.

The present study with Raman technique showed sensitivity and specificity numbers comparable to ICG test; for IgM detection, the sensitivity of the ICG assay is about 78%.[33] In the Raman discrimination model presented in this study, IgM+ could not be detected. On the other hand, RT-PCR is not a good option for detecting early infection, since the viral load can be undetectable and late. In fact, both RT-PCR and the ICG assay can have high failure rates to detect early infection. It was reported by Long et al.[57] that patients with COVID-19 and mild or asymptomatic symptoms IgG levels remained high while IgM gradually decreased when patients recovered from the infection. Also, a relevant disadvantage of methods based on the detection of antibodies in sera is the possibility of cross-reactions with other viruses, especially those from the same family which cause colds and other respiratory diseases. This is due to the identification between the sequences between the coronaviruses (which cause the common flu) and the SARS-CoV-2 leading to false positives and demonstrating an overestimated prevalence of infection. On the other hand, false negatives can occur due to viral load, inadequate material collection, and small amount of antibodies, among other issues.[58,59]

Although information regarding the cut-off point for the days after symptoms to perform the RT-PCR and immunological tests with high sensitivity and specificity are still under investigation, some studies indicate that the more at the onset of symptoms, the greater the chance of positive RT-PCR, and the more far from the onset of symptoms, a greater chance of positive IgM and IgG. Tests of molecular biology and immunology are essential at different moments of the infection, and simultaneous application of both tests has been demonstrated with greater diagnostic and prognostic proficiency.[60] Thus, the detection of antibodies, especially IgM, which are produced at the beginning of the infection, can be a tool combined with RT-PCR to improve sensitivity and diagnostic accuracy. Therefore, Raman spectroscopy may become a tool for rapid, reagent-free, and accurate detection of COVID-19 in intermediate infection and also infection in remission since it could detect antibodies (IgG and tryptophan), viral load (RNA virus), and metabolic changes associated to the infection (decrease in proteins and amino acids and increase in lipids).

Due to the biochemical nature of the Raman spectroscopy technique, future works could evaluate the efficacy of vaccines in the generation of immune response by evaluating serum from nonvaccinated and vaccinated subjects, the long-term effects of the infections and the detections of reinfections.

### 4 Conclusion

This study showed that the main differences in the Raman spectra of positive IgM and IgG versus Control were spectral features assigned to proteins including albumin (lower in the group COVID-19 and in the group IgM/IgG and IgG positive), and features assigned to lipids, phospholipids, and carotenoids (higher the group COVID-19 and in the group IgM/IgG positive). Features referred to nucleic acids, tryptophan, and immunoglobulins were also seen (higher the group COVID-19). The discriminant model based on PLS (PLS-DA) found sensitivity of 84.0%, specificity of 95.0%, and accuracy of 90.3% for discriminating positive Ig groups versus Control. When considering individual Ig group versus Control, it was found sensitivity of 77.3%, specificity of 97.5%, and accuracy of 88.8%. The higher classification error was found for the IgM group (no success classification), which could be due to low number of samples, the group...
study composed of asymptomatic patients, and the lower sensitivity for IgM in the ICG test. Raman spectroscopy may become a technique of choice for rapid serological evaluation aiming COVID-19 diagnosis, mainly detecting the presence of IgM/IgG and IgG after COVID-19 infection.

ACKNOWLEDGMENTS
L. Silveira Jr. thanks FAPESP (São Paulo Research Foundation) for granting the Raman spectrometer (Grant No. 2009/01788-5) and CNPq (National Council for Scientific and Technological Development) for the Productivity Fellowship (Process No. 306344/2017-3).

DATA AVAILABILITY STATEMENT
Data available on request from the authors

ORCID
Landulfo Silveira Jr https://orcid.org/0000-0002-6616-3334

REFERENCES
[1] Word Health Organization. Laboratory testing strategy recommendations for COVID-19: interim guidance. 2021. WHO Publications. Available at: https://www.who.int/publications/i/item/laboratory-testing-strategy-recommendations-for-covid-19-interim-guidance
[2] A. Z. Schuchmann, B. L. Schnorrenberger, M. E. Chiquetti, R. S. Gaiki, B. W. Raimann, M. A. Maeyama, J. R. Giles, A. E. Baxter, D. A. Oldridge, A. R. Weisman, C. A. G. Ittner, O. Kuthuru, J. Dougherty, K. Nzingha, N. Han, J. Kim, A. Pattekar, E. C. Goodwin, E. M. Anderson, M. E. Weirick, S. Gouma, C. P. Arevalo, M. J. Bolton, F. Chen, S. F. Lacey, H. Ramage, S. Cherry, S. E. Hensley, S. A. Apostolidis, A. C. Huang, L. A. Vella, U. Penn COVID Processing Unit, M. R. Betts, N. J. Meyer, E. J. Wherry, Science 2020, 369, eabc8511. https://doi.org/10.1126/science.abc8511
[3] M. P. Verotti, M. C. Ramos, C. M. P. Henriques, F. T. S. Elias, E. B. Camargo, Rep. Health Sci. 2020, 31, 217. https://doi.org/10.5172/ccc.v31iSuppl%20.1.S771
[4] V. M. C. H. Dias, M. Carneiro, L. Michelin, C. F. L. Vidal, L. A. T. da Costa, C. E. S. Ferreira, E. A. Rosseto-Welter, R. S. Lins, R. Kfouri, S. F. Costa, C. A. da Cunha, A. Chebabo, J. L. L. Rocha, L. C. Von Bahten, L. E. da Silva, R. V. Cohen, J. A. Moura-Neto, M. M. do Nascimento, A. F. Oliveira, H. S. C. Ribeiro, R. Ribeiro, C. M. D. M. Carrilho, J. Infect. Control. 2020, 9, 90.
[5] Z. Shen, Y. Xiao, L. Kang, W. Ma, L. Shi, Z. Zhang, Z. Zhou, J. Yang, J. Zhong, D. Yang, L. Guo, G. Zhang, H. Li, Y. Xu, M. Chen, Z. Gao, J. Wang, L. Ren, M. Li, Clin. Infect. Dis. 2020, 71, 713. https://doi.org/10.1093/cid/ciaa203
[6] E. Smith, G. Dent, Modern Raman Spectroscopy: A Practical Approach, John Wiley & Sons, Chichester 2013.
[7] M. L. Almeida, C. J. Saatkamp, A. B. Fernandes, A. L. B. Pinheiro, L. Silveira, Lasers Med. Sci. 2016, 31, 1415. https://doi.org/10.1007/s10103-016-2003-y
[8] S. Giansante, H. E. Giana, A. B. Fernandes, L. Silveira, Lasers Med. Sci. 2021. https://doi.org/10.1007/s10103-021-03247-8
[9] N. A. Correia, L. A. T. Batista, R. J. M. Nascimento, M. C. T. Cangussú, P. J. L. Crueger, L. G. P. Soares, L. Silveira, A. L. B. Pinheiro, J. Photochem. Photobiol. B. 2020, 204, 111801. https://doi.org/10.1016/j.jphotobiol.2020.111801
[10] Q. Zhan, Y. Li, Y. Yuan, J. Liu, Y. Li, J. Raman Spectrosc. 2020, 51, 2377. https://doi.org/10.1002/jrs.5940
[11] S. Sabtu, S. F. A. Sani, D. A. Bradley, L. M. Looi, Z. Osman, J. Raman Spectrosc. 2020, 51, 380. https://doi.org/10.1002/jrs.5774
[12] M. Saleem, S. Ali, M. B. Khan, A. Amin, M. Bilal, H. Nawaz, M. Hassan, J. Raman Spectrosc. 2020, 51, 1067. https://doi.org/10.1002/jrs.5896
[13] J. Lukose, S. Chidangil, S. D. George, Biosens. Bioelectron. 2021, 178, 113004. https://doi.org/10.1016/j.bios.2021.113004
SUPPORTING INFORMATION
Additional supporting information may be found in the online version of the article at the publisher’s website.

How to cite this article: A. C. C. Goulart, R. A. Zângaro, H. C. Carvalho, L. Silveira Jr, J Raman Spectrosc 2021, 52(12), 2671. https://doi.org/10.1002/jrs.6235