analytical chemistry

pubs.acs.org/ac Article

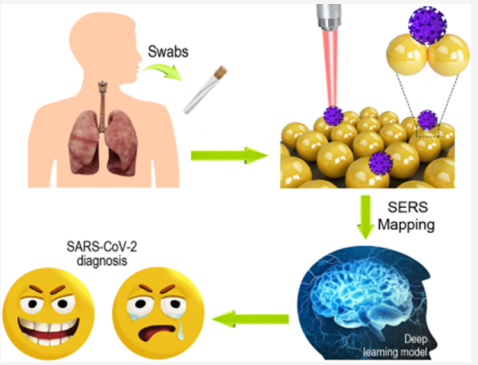
On-Site Detection of SARS-CoV-2 Antigen by Deep Learning-Based Surface-Enhanced Raman Spectroscopy and Its Biochemical Foundations

Jinglin Huang,[⊥] Jiaxing Wen,[⊥] Minjie Zhou, Shuang Ni, Wei Le, Guo Chen, Lai Wei, Yong Zeng, Daojian Qi, Ming Pan, Jianan Xu, Yan Wu, Zeyu Li, Yuliang Feng, Zongqing Zhao, Zhibing He, Bo Li, Songnan Zhao, Baohan Zhang, Peili Xue, Shusen He, Kun Fang, Yuanyu Zhao, and Kai Du*

Cite This: *Anal. Chem.* 2021, 93, 9174−9182 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A rapid, on-site, and accurate SARS-CoV-2 detection method is crucial for the prevention and control of the COVID-19 epidemic. However, such an ideal screening technology has not yet been developed for the diagnosis of SARS-CoV-2. Here, we have developed a deep learning-based surface-enhanced Raman spectroscopy technique for the sensitive, rapid, and on-site detection of the SARS-CoV-2 antigen in the throat swabs or sputum from 30 confirmed COVID-19 patients. A Raman database based on the spike protein of SARS-CoV-2 was established from experiments and theoretical calculations. The corresponding biochemical foundation for this method is also discussed. The deep learning model could predict the SARS-CoV-2 antigen with an identification accuracy of 87.7%. These results suggested that this method has great potential for the diagnosis, monitoring, and control of SARS-CoV-2 worldwide.

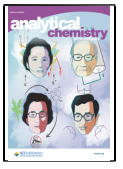


The outbreak of coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has become a serious problem worldwide.[1,2] As of January 27, 2021, approximately 100 million cases of the infection had been confirmed and over 20 million patients had died in over 224 countries and/or areas (https://www.who.int/emergencies/diseases/novel-coronavirus-2019). The human-to-human transmission, indefinite latency, and nonspecific symptoms of SARS-CoV-2 infection have caused a rapid spread of COVID-19.[3] The regular prevention and control measures for an epidemic could be the most effective measures to prevent the mass spread of pandemic situations before the availability of specific antiviral medicines or vaccines for COVID-19.[4,5] On-site, point-of-care, easy-to-use, portable, and accurate SARS-CoV-2 infection detection methods are of great importance for large-scale screening in the prevention and control of epidemics, especially at airports, railway stations, and other densely populated areas.

However, such ideal screening technology is still unavailable for the diagnosis of SARS-CoV-2. Reverse transcription-polymerase chain reaction (RT-PCR) for the detection of nucleic acids is the dominant method for the diagnosis of SARS-CoV-2.[6] The long detection time (approximately 4 h), laboratory requirements, and professional operation required for this method make it difficult to use for rapid and on-site detection.[7,8] Although using a reverse transcription loop-mediated isothermal amplification (RT-LAMP) assay can reduce the detection time to approximately 30 min, the complex operating procedure and high standard of biosafety requirements prevent the use of this method for the point-of-care detection of SARS-CoV-2.[9] The detection of specific antibodies, such as IgG and IgM, as well as antigens based on the enzyme-linked immunosorbent assay (ELISA) method could be a rapid, easy-to-use, and potentially point-of-care strategy for the diagnosis of SARS-CoV-2.[10] However, the limited sensitivity and the risk of missing detection during the window phase reduce the reliability of this strategy. In contrast, Raman spectroscopy can specifically identify target molecules with a short detection time and simple pretreatment requirements using a noncontact measurement procedure.[11] More importantly, it is easy to realize on-site and point-of-care detection using a portable Raman spectrometer.

Generally, the typical intensity of Raman spectroscopy is quite weak because of the small Raman scattering cross section for molecules, while advanced surface-enhanced Raman spectroscopy (SERS) can magnify the signal intensity by

Received: March 10, 2021
Accepted: June 14, 2021
Published: June 22, 2021



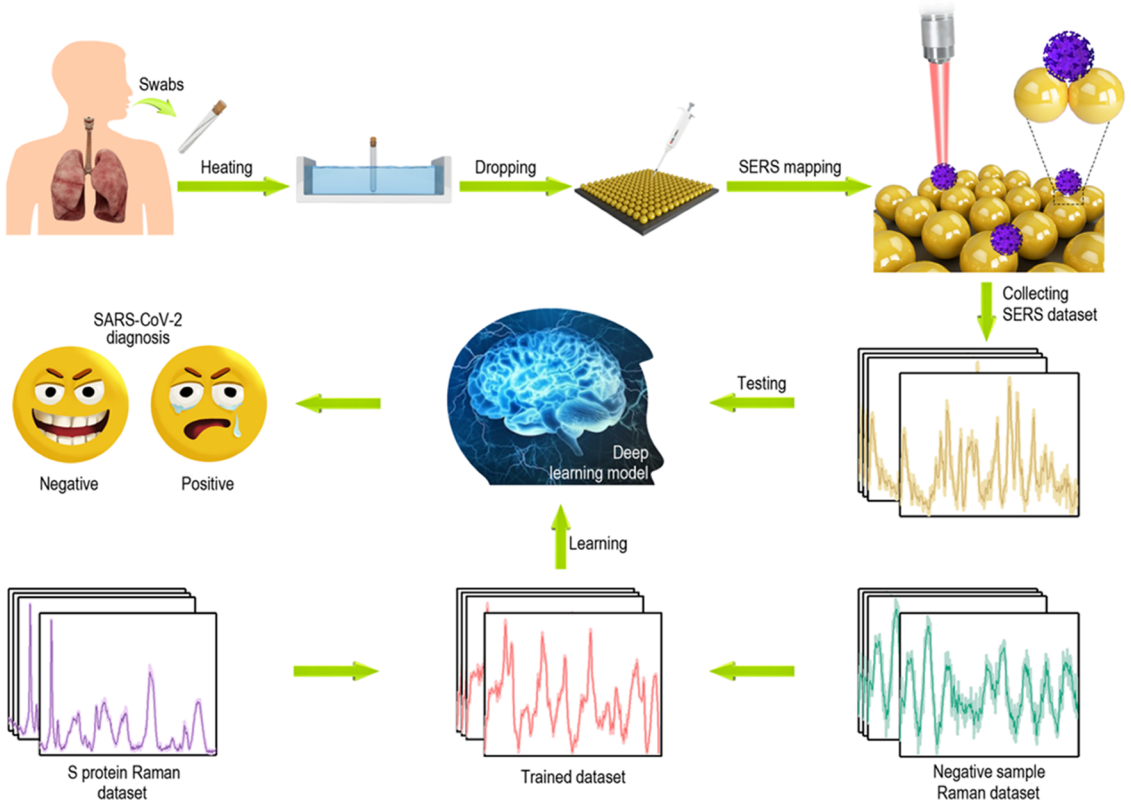


Figure 1. Schematic illustration of the detection process of the deep learning-based SERS technique.

over a million times and can even realize single-molecule detection.[12,13] The increased signal intensity of SERS mainly originates from electromagnetic (EM) field enhancement at the nanogaps between metal nanoparticles and/or nanostructures.[14] The use of SERS in biological fields has been widely reported.[15,16] Recently, SERS has been employed to detect the influenza virus and hepatitis B virus in human plasma.[17−19] However, the absence of clinical research makes it difficult for the practical applications of SERS in large-scale screening for SARS-CoV-2.

The complex and heterogeneous chemical compositions of real clinical specimens, such as throat swabs, will significantly influence the SERS measurement of SARS-CoV-2. Actually, the identification of biomacromolecules by SERS remains a challenge because of the poor compatibility of biomolecules with SERS hot spots, the complexity of physiological environments, the differences between individuals, the random orientation of biomolecules, and the probable interactions of biomolecules with SERS substrates.[20] A labeled SERS strategy could be an effective way to overcome the complexity of direct SERS measurements, while it shares the drawbacks of typical antibody detection methods.[21]

Here, we demonstrated a method for the on-site detection of the SARS-CoV-2 antigen from human throat swabs and sputum within 20 min using a deep learning-based SERS technique. The corresponding biochemical foundations are discussed to indicate the reliability of this method and give an insight into the spectral characteristics of SARS-CoV-2. Figure 1 shows the detection process of our method. A gold nanoparticle (AuNP) array was used as the SERS substrate. According to the electromagnetic enhancement mechanism of SERS, the enhanced local electric field (hot spot) locates on the nanogaps of neighboring AuNPs, which have a feature size of 2−10 nm.[22] Considering the molecular architecture of the SARS-CoV-2 virus, the spike (S) protein, which has a nail-like shape with a width of 7 nm and a length of 23 nm, has the maximum probability to fall into the SERS hot spots.[23] Therefore, the S protein was chosen as the detection target in this work. To realize the rapid recognition of clinical specimens, a supervised deep-learning algorithm was established. A spectra database of pure S protein and negative clinical specimens was used to train the deep learning models. The supervised model could successfully identify the SARS-CoV-2 antigen in clinical specimens. Our method opens a door for the rapid on-site diagnosis of SARS-CoV-2 and other highly pathogenic viruses.

EXPERIMENTAL METHODS

Purification and Preconcentration of the Protein. The spike proteins used in this work were purchased from Sino Biological Inc. (Beijing, China). These proteins were buffered by phosphate-buffered saline (PBS) or 20 mM PB, 300 mM NaCl, and 10% glycerol with a concentration of about 1 mg/mL. An ultrafiltration centrifugation process was applied to remove the buffer and make a preconcentration for the purchased proteins. The ultrafiltration tubes were purchased from Merck Millipore with a nominal molecular weight limit (NMWL) of 10 and 30 kDa. The initial 50 μL protein solutions were centrifuged 5 times to give a final volume of ∼15 μL. The centrifugal process was carried out using a refrigerated centrifuge with a centrifugal force of 14 000g at 4 °C for 25 min each time.

Substrate Preparation. The normal Raman spectra of the protein were collected on Ag-coated Si substrates. The Ag coatings were deposited by the direct current magnetron sputtering technique. The Si slices were successively cleaned

https://doi.org/10.1021/acs.analchem.1c01061
Anal. Chem. 2021, 93, 9174−9182

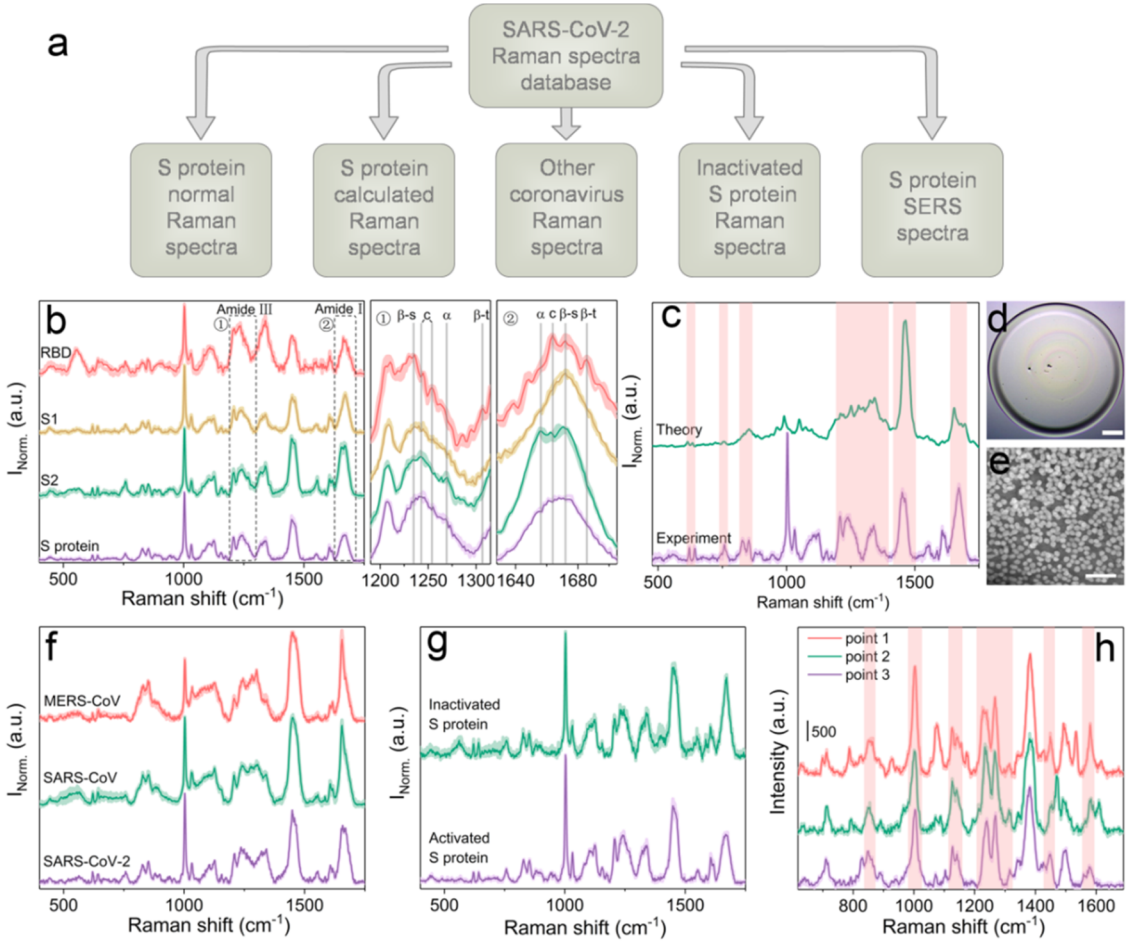


Figure 2. Raman spectra database of the S protein of SARS-CoV-2. (a) The architecture of the spectra database. (b) Normal Raman spectra of the S protein, the S1 subunit, the S2 subunit, and the RBD of SARS-CoV-2. (c) Comparisons of the theoretical and experimental Raman spectra of the S1 subunit. (d) White-light photograph of dried S protein on a Ag-coated Si slice. The scale bar is 300 μm. (e) The SEM image of the AuNP array with a scale bar of 200 nm. (f) Comparisons of the normal Raman spectra of the S proteins of SARS-CoV-2, SARS-CoV, and Middle East respiratory syndrome coronavirus (MERS-CoV) expressed in baculovirus-infected insect cells. (g) Comparisons of the normal Raman spectra of the S protein of SARS-CoV-2 before and after heat treatment. (h) SERS spectra of the S protein of SARS-CoV-2.

with acetone, ethanol, and deionized water (DI, 18 MΩ) before deposition. The deposition processes were carried out in pure Ar at a working pressure of 0.5 Pa (3.8 mTorr) and room temperature under a base pressure of ~1.2×10^{-4} Pa (9×10^{-7} Torr). The power of the Ag target was set constant to 100 W with a deposition time of 370 s. The thickness of the deposited Ag coating was approximately 300 nm. The aqueous suspensions of AuNPs with a diameter of ~40 nm were synthesized according to the Frens method.[24] The self-assembly of AuNPs was carried out at the oil/water interface. First, a specific concentration of AuNPs was injected into a bottle. Then, 1 mL of hexane was added above the AuNP solution. Third, 100 μL of ethanol was injected drop by drop. The AuNPs were observed to move to the hexane/water interface on the addition of ethanol and form a nanoparticle monolayer film. Finally, the film was transferred onto Si slices to form the AuNP array substrates.

Raman Measurements. The Raman spectra were obtained using a portable Raman spectrometer (Renishaw Virsa) with a 532 nm laser. A wavenumber calibration was performed using a Si sample before the measurement. A ×50 objective long focal lens was used to collect the Raman signals. The laser power was set as 15 mW with an integration time of 20 s. The spectra were background-corrected using the built-in arithmetic of the Renishaw WiRE 5.3 program. The spectral range was selected as 400−1750 cm^{-1}, which contains the majority of information on proteins. The spectra were normalized to 0−1 within this spectral range.

DFT Calculations. A calculation model for the S1 subunit was created based on the experiment structure (PDB code: 6VSB[25]). The S1 subunit was first divided into fragments with each fragment of ~200 atoms. After hydrogen saturation, the fragment structures were optimized using the Vienna ab initio simulation package (VASP)[26] with the nonhydrogen atoms fixed. All hydrogen atoms were relaxed until the force on each atom was less than 0.03 eV/Å. Geometry optimizations were performed with γ point only. The total energies were converged to 10^{-4} eV. The Perdew, Burke, and Ernzerhof (PBE) exchange−correlation functional[27] within the generalized gradient approximation and the projector augmented-wave pseudopotential[28] were adopted. Then, the frequencies and Raman activities of each fragment were calculated by the Gaussian 09 package at the hybrid B3LYP[29] functional level, with 6-31G(d) basis sets. The spectrum of the S1 subunit consisted of the frequencies and Raman activities of all of the

https://doi.org/10.1021/acs.analchem.1c01061
Anal. Chem. 2021, 93, 9174−9182

fragments and was smoothed using a Lorentzian convolution of 4 cm^{-1}.

SERS Characterization. A standard surface-enhanced Raman spectroscopy testing was performed using a portable Raman spectrometer (Renishaw Virsa) with an excitation laser wavelength of 785 nm to avoid the interference of fluorescence. A rapid Raman mapping mode was used to collect the SERS spectra with an integration time of only 0.2 s for a single spot. Two 50 × 50 discrete spot maps were taken for each sample with 2 μm spacing between spots to avoid overlap between the spectra. A spectral range of 618−1689 cm^{-1} was selected to perform the SERS measurements.

Dataset. It was hard to obtain the positive dataset because of the uncertainty in the virus distributions in the clinical specimens from COVID-19 cases. Therefore, the positive dataset was obtained from the negative spectra superimposed with the randomly modulated normal Raman spectra of the S protein. The random modulation factor for the intensity was uniformly set from 0.1 to 1. The spectral range was from 618 to 1689 cm^{-1} and consisted of 1015 one-dimensional float data points. The training dataset for our RNN model contained 93 750 negative spectra collected from 75 healthy controls and 93 750 positive spectra obtained by the method mentioned above. During the training process, 95% of the dataset was used as the training set and 5% as the test set. Only the training set was used to train the model.

Data Preprocessing. The preprocessing algorithm included cosmic muon noise removal, convolution filtering, wideband background correction, and 0−1 normalization. This method processed data in the following order: (1) cosmic muon noise removal, (2) 0−1 normalization, (3) convolution filtering, and (4) wideband background correction. In the cosmic muon noise removal algorithm, there was a moving window with a width of 43 points. The mean value (MEAN) and standard deviation (SD) of the first 10 points and the last 10 points in the window were calculated. If the value at the center point of the window was larger than (20 × SD + MEAN), the value at the center point of the window was replaced by the MEAN. The convolution filtering was implemented by a 5-point convolution window. The wideband background correction was implemented using a sensitive nonlinear peak clipping (SNIP) algorithm, which is an iterative background elimination method that has been widely used in the γ-ray spectrum processing.[30] The iteration time was set to 20 in this work.

RESULTS AND DISCUSSION

Raman Spectra Database of SARS-CoV-2. To precisely identify the SARS-CoV-2 antigen in clinical specimens by SERS, a normative Raman spectra database needs to be first established. Figure 2a illustrates the architecture of this spectra database, which has five components: (1) the normal Raman spectra of the S protein, including full-length S protein, S1 subunit, S2 subunit, and the receptor-binding domain (RBD) to provide reference spectra and an insight into the Raman spectra of the molecular structures of the S protein; (2) the theoretically calculated Raman spectra, which were used to further demonstrate the reliability of the experimental data; (3) the Raman spectra of other coronaviruses to confirm the specificity of the method for a particular virus; (4) the Raman spectra of inactivated S protein to evaluate the influence of the inactivation treatment on the Raman spectra; and (5) the SERS spectra of the S protein to establish a link between the normal Raman spectra and the SERS spectra.

The classical Raman spectroscopy technique requires relatively high concentrations of samples, while it is not easy to acquire such high concentrations of the recombinant S protein of SARS-CoV-2. The signal-to-noise ratio (SNR) for the Raman spectra of the S protein with a concentration of ∼1 mg/mL is too low for the signals to be recognized (Figure S1). A drop-coating deposition Raman (DCDR) spectroscopy technique was used to overcome this problem. Previous studies have demonstrated that DCDR spectroscopy can provide significant signal amplification by preconcentration of the analytes through a coffee-ring effect.[31,32] The DCDR measurements were performed by pipetting 1 μL of purified protein onto a Ag-coated Si slice and drying for ∼15 min under temperature and humidity-controlled conditions. Figure 2d shows the ring pattern of the dried S protein on the Si slice under white light.

Assisted by the DCDR technique, the SNRs of the Raman spectra obtained for the S protein, S1 and S2 subunits, and the RBD of SARS-CoV-2 were considerably improved, as shown in Figure 2b. The spectra of these four proteins were similar to each other, although some differences were observed. To further understand these differences and provide an insight into the relationship of the Raman spectra with the protein structures, the Raman band assignment was performed based on previous studies (see Table S1).[15,33−37] All of the Raman bands could be attributed to standard shifts of the biochemical entities of protein, including the backbone conformation, the side-chain, and the glycans on the surface. The characteristic Raman bands of the backbone were associated with the vibrational modes of the CONH group and the C−C, C−N, and C−H bonds. The Raman bands of the side-chain mainly originated from the aromatic amino acid residues (e.g., the benzene rings of Phe and Tyr, and the indole ring of Trp) and sulfur-containing residues (e.g., C−S and S−S bonds). It has been reported that each monomer of the S protein of SARS-CoV-2 is highly glycosylated with 22 predicted N-linked glycosylation sites and 4 *O*-glycosylation sites.[38−40] Therefore, the glycans on the surface can also contribute to the Raman spectra of the S protein. The characteristic Raman bands of the side-chain and glycans were similar for the four studied proteins of SARS-CoV-2, while there were obvious differences in the Raman bands of the backbones, especially for the amide III (coupled C−N stretching and N−H bending vibrations, 1230−1300 cm^{-1}) and amide I (stretching vibration of C═O, 1600−1690 cm^{-1}) bands, as shown in Figure 2b and Table S1. The S1 subunit showed substantial Raman bands of β-sheets/coils but a few bands from α-helix. The S2 subunit exhibited obvious Raman bands of α-helical components and β-sheets/coils, while the RBD contained many Raman bands of β-sheets and coils/turns. Meanwhile, the Raman bands of the full-length S protein had the same characteristics of S1 and S2 subunits. According to previous structural studies of the purified S protein of SARS-CoV-2 by cryoelectron microscopy (cryo-EM), the S1 subunit is mainly formed by β-sheets, the S2 subunit is formed by central α-helices, twisted β-sheets, and loops, and the RBD is formed by β-sheets and coils/turns.[25,41,42] Therefore, the characteristic Raman bands of the S protein were consistent with the structures, which indicated that the Raman spectra can reflect the intrinsic biochemistry information of the S protein.

https://doi.org/10.1021/acs.analchem.1c01061
Anal. Chem. 2021, 93, 9174−9182

To further confirm the reliability of the experimental Raman spectra of the S protein of SARS-CoV-2, a theoretical calculation of the Raman spectra for the S1 subunit was carried out using density functional theory (DFT) calculations. According to the experimental structure of SARS-CoV-2 (PDB code: 6VSB), the S1 subunit contains 681 amino acid residues and nearly 4000 nonhydrogen atoms.[25] To deal with such a large structure, the S1 subunit was divided into several fragments with each fragment containing approximately 200 atoms. After the calculations of frequencies and Raman activities for each fragment, the Raman spectra of the S1 subunit consisted of the frequencies and Raman activities of all of the fragments. Figure 2c presents a comparison of the calculated and experimental Raman spectra of the S1 subunit of SARS-CoV-2. It should be noted that the Raman intensity of the simulated spectrum was not optimized. Therefore, the intensity of the simulated spectrum and experimental spectrum cannot be compared; however, the Raman shifts of the simulated spectrum and the experimental spectrum were very close (see the noted areas in Figure 2c). Usually, the Raman shift is the main feature to identify a molecule. Therefore, the above results indicated the high reliability of the experimental Raman spectra of the S protein.

A reliable diagnostic technique requires specific diagnostic criteria. To explore whether the Raman spectra of the S protein have a fingerprint character, the Raman spectra of the S protein for additional four types of β-genus human coronavirus, including SARS-CoV, Middle East respiratory syndrome coronavirus (MERS-CoV), and human HKU1 and OC43 coronaviruses (HCoV-HKU1 and HCoV-OC43), were obtained as shown in Figures 2f and S2a. It should be noted that the S proteins of these four coronaviruses were expressed by baculovirus-infected insect cells. Thus, the Raman spectra of the SARS-CoV-2 spike protein expressed by the same system are provided in Figure 2f for comparison. The Raman bands presented obvious differences between the five coronaviruses, especially in the amide I and amide III bands (Figure S2b). To make a relatively quantitative comparison of these differences, the Raman band assignment was performed as shown in Table S2. The amide I and amide III bands of SARS-CoV-2, HCoV-HKU1, and HCoV-OC43 showed evidence of more obvious β-sheets/coils (1665−1669 cm^{-1} for amide I, and 1233−1247 cm^{-1} for amide III) than α-helices (1651−1655 cm^{-1} for amide I, and 1264−1283 cm^{-1} for amide III); whereas the Raman bands of amide I/III for SARS-CoV and MERS-CoV showed distinct α-helical components (1655 cm^{-1} for amide I, and 1277−1283 cm^{-1} for amide III). These differences in the Raman bands might be caused by different conformations of their S proteins. Except for the backbone vibration, there were distinct differences in the Raman bands of the side chains. The spectra of the S proteins of HCoV-HKU1 and HCoV-OC43 showed more intense Raman bands than SARS-CoV-2, SARS-CoV, or MERS-CoV at a lower wavenumber region (400−800 cm^{-1}), which could be caused by differences in the amino acid sequences as the latter three coronaviruses share higher amino acid identities. In addition, the Raman bands for the S protein of SARS-CoV-2 expressed by different systems were almost identical (Figure 2b,f). The above results demonstrated that different viruses present distinct Raman bands, which implies that Raman spectra should be an effective technique to distinguish viruses.

An inactivation treatment is essential before the detection of clinical specimens to ensure the safety of the operators. Therefore, the effect of inactivation treatment on the Raman spectra of the S protein needs to be studied. Figure 2g shows the Raman spectra of the S protein after heat treatment for 30 min at 56 °C, which is a standard inactivation procedure for nucleic acid assays. The Raman bands showed negligible changes for the inactivated S protein compared with the activated protein, which indicated that the inactivation treatment did not appreciably influence the Raman spectra of the S protein. Actually, the Raman spectra mainly represent the vibration modes of the primary structures and a few secondary structures of proteins,[43] while the inactivation treatment mainly destroys the higher structures of proteins. Therefore, the Raman bands of the S protein remained largely unchanged after the heat treatment.

An aggregated spherical AuNP array was prepared to perform the SERS measurements (Figure 2e). The AuNP array was fabricated by self-assembly of 40 nm gold nanoparticles onto Si slices at an oil/water interface. A 300 nm silver film was deposited on the Si slices before the self-assembly of the AuNPs to shield the Raman signal of Si. The aggregated AuNPs can generate large EM fields in the nanogaps of adjacent AuNPs. The enhancement activity of the AuNP array was tested using a dye molecule, as shown in Figure S3. The characteristic peaks of rhodamine 6G (R6G) can be observed at a concentration of 1×10^{-10} M, which indicated that the limit of detection (LOD) for R6G on the AuNP array was about 50 ppt. The S protein aqueous solutions of SARS-CoV-2 at a concentration of ∼1 ppm were dropped on the SERS substrate and thoroughly dried to acquire the SERS spectra (Figure S4). The SERS spectra were collected with a 785 nm laser to restrain the fluorescence signals. The largest measurement range of our Raman spectrometer was ∼1070 cm^{-1} without revolving the grating. Therefore, to improve the detection efficiency, a spectral range of 618−1689 cm^{-1} was selected to perform the SERS measurements. Figure 2h shows the SERS signals of the S protein collected at three different points. The main Raman bands around 853, 1003, 1125−1175, 1240−1320, 1450, and 1586 cm^{-1} (highlighted in Figure 2h) of the S protein could be observed by comparison with the normal Raman spectra (Figure 2b), indicating the detection of SARS-CoV-2 by SERS. Besides, the signal consistency of these SERS spectra was not good, which should be caused by the random impurity peaks from the AuNP substrates.

SERS Measurements of Clinical Specimens. The SERS measurements of throat swabs from 102 healthy people were carried out at Sichuan Science City Hospital, Mianyang, China (see Table S3). Meanwhile, the SERS spectra of throat swabs or sputum from 30 confirmed COVID-19 patients were collected at the Sichuan Provincial Center for Disease Control and Prevention, Chengdu, China (see Table S4). This work was approved by the research board of the Ethics Committee of the Sichuan Science City Hospital. All patients provided informed consent in accordance with the Declaration of Helsinki. To guarantee the safety of operators, all specimens were inactivated at 56 °C for 30 min in a validated biological safety cabinet before the SERS measurement. A droplet of 1.5 μL of each specimen was dropped onto a SERS substrate and thoroughly dried to acquire the SERS spectra. After drying, the substrate was sealed into a homemade quartz box to avoid the possible production of aerosols during the laser irradiation (Figure S5). Considering the low concentration of virosomes in the clinical specimens, two Raman mappings (2 μm step

https://doi.org/10.1021/acs.analchem.1c01061
Anal. Chem. 2021, 93, 9174−9182

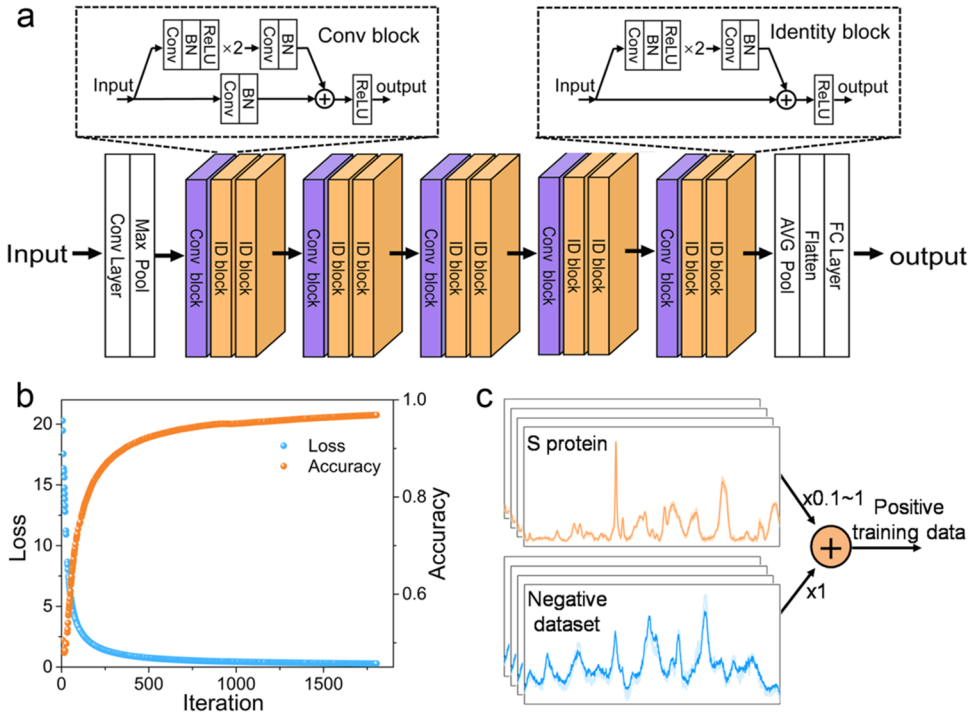


Figure 3. Deep learning-based SERS for the diagnosis of SARS-CoV-2. (a) The architecture of the RNN deep learning model. (b) Training loss and accuracy obtained with the dropout layer deactivated after training at each iteration. (c) Production of positive training data by the superposition of negative SERS spectra with normal Raman spectra of the S protein.

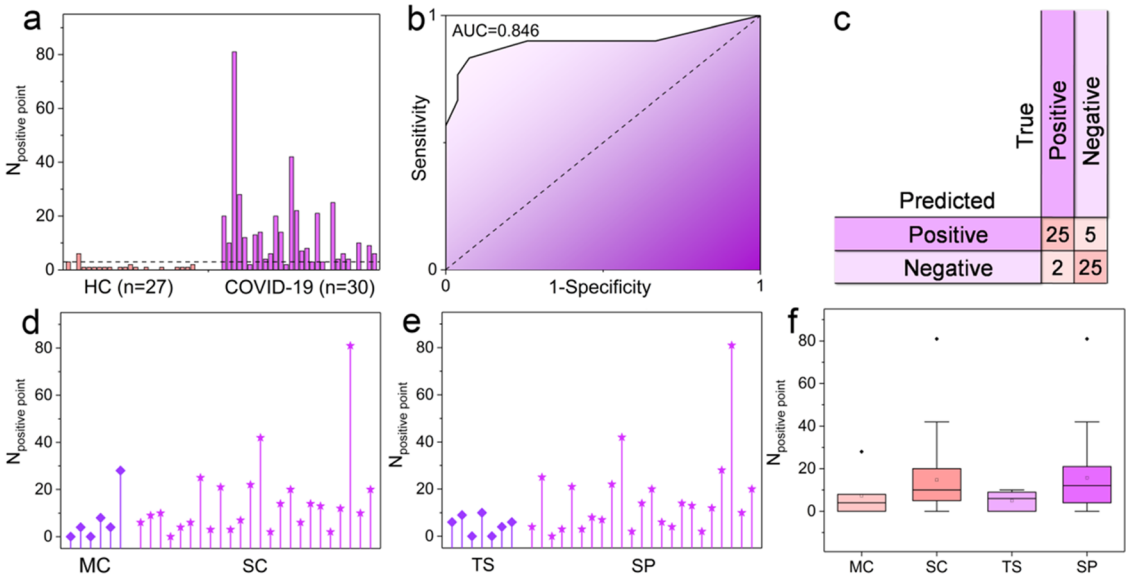


Figure 4. SARS-CoV-2 diagnosis results with deep learning-based SERS. (a) The bar graph shows the number of the predicted positive points of clinical specimens based on the output scores from the RNN model. HC represents healthy controls. (b) The ROC curve and the corresponding AUC value. (c) A binary classifier summarized the diagnostic results at the optimal cutoff value. (d) Drop lines of the predicted positive point number for mild cases (MC) versus severe cases (SC). (e) Drop lines of the predicted positive point number for throat swab cases (TS) versus sputum cases (SP). (f) The corresponding box plots of (d) and (e).

size, 50 μm × 50 μm) were obtained for each specimen to ensure the detection of virus (Figure S6a). Figure S6b,c presents the collected SERS spectra of clinical specimens with and without SARS-CoV-2. Although these SERS spectra exhibited distinct Raman peaks, because of the complexity and randomness of these spectra, which should be caused by impurities in the clinical specimens, it was difficult to directly distinguish the positive and negative specimens. Therefore, a deep learning method was implemented to analyze the heterogeneous signal patterns and to identify SARS-CoV-2 signals.

Deep Learning Model to Identify the SARS-CoV-2 Antigen. The deep learning methods have been previously demonstrated to be effective in the analysis of complex Raman spectra.[44,45] Therefore, to predict COVID-19 infections using the standard Raman spectra of the S protein, a residual neural

https://doi.org/10.1021/acs.analchem.1c01061
Anal. Chem. 2021, 93, 9174−9182

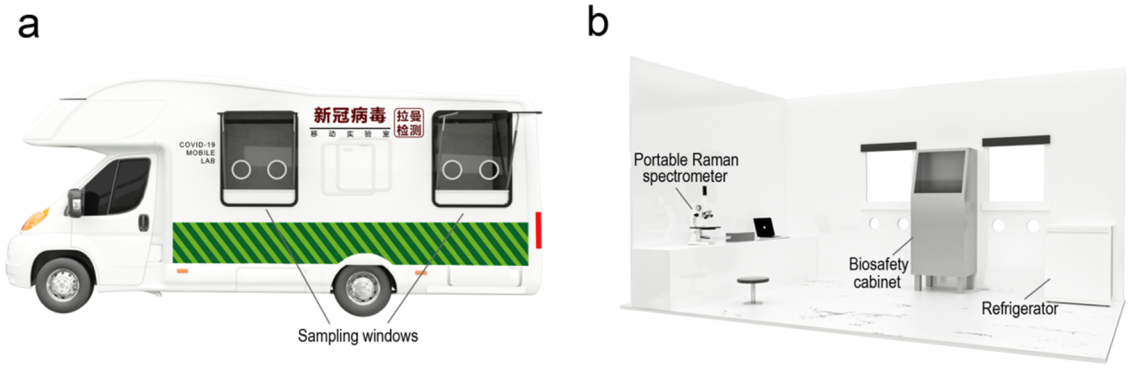


Figure 5. Schematic illustration of the SERS mobile detection platform for COVID-19. (a) Exterior view of the COVID-19 mobile lab. (b) Interior view of the COVID-19 mobile lab.

network (RNN)-based deep learning model was carried out, as shown in Figure 3a. The RNN depends on a shortcut connection by tagging previous input data to the output data to maintain previous gradient information and solve the vanishing gradient problem of deep architecture. In view of the computing power and efficiency of training, the framework of the RNN was organized with an initial convolution layer, followed by batch normalization (BN), a rectified linear (ReLU) transformation, and a max pooling layer (Max Pool), followed by five serial blocks. These blocks were one convolution block followed by two identity blocks with shortcuts. Finally, these blocks were followed by an average pooling layer (AVG Pool), a flatten layer, and a fully connected (FC) layer.

A well-prepared dataset is essential to acquire a functional RNN model. The Raman spectra collected from the throat swabs of healthy people were used as negative samples and labeled as 0, while the Raman spectra obtained from COVID-19 cases could not be simply labeled as positive (1) or negative (0) because of the uncertainty of the virus distributions. Therefore, the negative spectra superimposed with the intensity of the randomly modulated normal Raman spectra of the S protein were used as positive training data and labeled as 1 (see Figure 3c). In addition, a data preprocessing method was applied to the input data, including cosmic muon noise removal, convolution filtering, wideband background correction, and 0−1 normalization (Figure S7). Consequently, the RNN model was trained with the positive and negative data sets to identify SARS-CoV-2 in clinical specimens.

SARS-CoV-2 Diagnosis with Deep Learning-Based SERS. We further tested whether the RNN model could identify SARS-CoV-2 in clinical specimens. As described before, to ensure the detection of the virus, two Raman mappings were collected for each sample. Every mapping contained 625 spectra, thus there were 1250 spectra for each sample. To improve the diagnostic efficiency, a homemade computer program was used to perform the identification process (Figure S8). The number of the predicted positive points based on the output scores from the RNN model was used as the diagnostic scale as shown in Figure 4a. The mean number of positive points for COVID-19 patients was 13.2-fold higher than that for the healthy controls, and the difference was statistically significant with a p-value of 1.3×10^{-8} (Figure S9). The receiver operating characteristic (ROC) curve of our method is given in Figure 4b with an area under the curve (AUC) value of 0.846 for all specimens. The optimal cutoff value was determined from the ROC curve. Figure 4c summarizes the diagnostic results at this cutoff value. Consequently, our method achieved a sensitivity of 83.3% and a specificity of 92.5%, indicating a high identification accuracy of 87.7%. It should be noted that the case severity and sample source displayed some interesting relationships with the SERS diagnosis results, as shown in Figure 4d−f. Severe cases and sputum samples exhibited more positive points than mild cases and throat swabs, respectively. Meanwhile, the sputum samples contained 5 mild cases and 18 severe cases. The throat swabs were from 1 mild case and 6 severe cases. Therefore, the increased positive points could be caused by the higher viral loads in the severe cases and sputum samples.[46] Besides, false-negatives were observed from 2 mild cases and 3 severe cases. This result may be caused by the use of only a few positive samples, which may have influenced the training of the neural network and reduced the accuracy. More positive samples will be collected in our future works.

This study used a portable Raman spectrometer to implement the experiment, which indicates the possibility of the integration of Raman spectroscopy with a mobile detection platform. An example is given in Figure 5. This mobile platform could integrate the whole detection process, including sample collections, pretreatments, and spectra measurements. Considering the small size of the required equipment for our detection method, the length of the car can be controlled within 6−8 m, which provides high maneuverability to this mobile platform. More importantly, a rapid Raman mapping mode was used in this study. The two mappings for each sample were collected in less than 5 min. Furthermore, the Raman detection pretreatment was simple and could be accomplished in 10−15 min. Therefore, the detection time for a single specimen could be controlled within 20 min without considering the inactivation treatment. These results indicated that our method has the potential as an effective on-site SARS-CoV-2 antigen detection technique.

Our findings revealed that there are obvious differences in the Raman spectra of the S protein for different coronaviruses, even those that share high amino acid identities. Generally, structural studies of proteins are performed using high-resolution cryo-EM, X-ray crystallography, or mass spectrometry.[47−50] Our data suggest that Raman spectroscopy could be an effective and easy-to-use method for the identification and structural analysis of proteins. Much effort should be made to understand the relationships of the Raman shifts with the structures of proteins.

https://doi.org/10.1021/acs.analchem.1c01061
Anal. Chem. 2021, 93, 9174−9182

CONCLUSIONS

We have applied a deep learning-based SERS method to realize the on-site and rapid detection of the SARS-CoV-2 antigen. A Raman spectra database for SARS-CoV-2 was first established based on the S protein to provide standard references. An RNN model pretrained on our database was used to achieve the detection of clinical specimens with high identification accuracy. We believe that this method could be extended to the identification of other viruses and could be an effective method to complement conventional detection techniques. The establishment of a reliable spectra database or references, such as the one demonstrated here, could be the most important factor to achieve this goal.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.1c01061.

Raman band assignments; baseline characteristics of clinical specimens; Raman spectra of low concentration S protein; Raman spectra of purified S proteins of different coronaviruses; SERS spectra of R6G on AuNP array; SERS measurement of S protein; photograph of the homemade quartz box; SERS measurement of clinical specimens; preprocessing of the SERS spectra; operation interface of the homemade software; and average positive point of healthy controls and confirmed patients (PDF)

AUTHOR INFORMATION

Corresponding Author

Kai Du − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China;* orcid.org/0000-0003-0317-990X; Email: dukai@caep.cn

Authors

Jinglin Huang − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China;* orcid.org/0000-0001-6316-5898

Jiaxing Wen − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China; Department of Engineering Physics, Tsinghua University, 100084 Beijing, China*

Minjie Zhou − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Shuang Ni − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Wei Le − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Guo Chen − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Lai Wei − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Yong Zeng − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Daojian Qi − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Ming Pan − *Sichuan Provincial Center for Disease Control and Prevention, 610041 Chengdu, China*

Jianan Xu − *Sichuan Provincial Center for Disease Control and Prevention, 610041 Chengdu, China*

Yan Wu − *Sichuan Science City Hospital, 621000 Mianyang, China*

Zeyu Li − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Yuliang Feng − *Sichuan Provincial Center for Disease Control and Prevention, 610041 Chengdu, China*

Zongqing Zhao − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Zhibing He − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Bo Li − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Songnan Zhao − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Baohan Zhang − *Laser Fusion Research Center, China Academy of Engineering Physics, 621900 Mianyang, China*

Peili Xue − *Sichuan Science City Hospital, 621000 Mianyang, China*

Shusen He − *Sichuan Provincial Center for Disease Control and Prevention, 610041 Chengdu, China*

Kun Fang − *Sichuan Science City Hospital, 621000 Mianyang, China*

Yuanyu Zhao − *Sichuan Science City Hospital, 621000 Mianyang, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.1c01061

Author Contributions

⊥J.H. and J.W. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Chen Wang, Lili Ren, Yulong Liu, Weiqian Zhao, Bing Zhao, Huiping Yang, and Juan Li, and the support from the Young Talents Foundation of Laser Fusion Research Center, CAEP (grant no. RCFCZ3-2019-9), the Sichuan Science and Technology Program (nos. 2020FYS0015 and 2020FYS0017), and the National Natural Science Foundation of China (no. 82041033).

REFERENCES

(1) Al-Mandhari, A. S.; Brennan, R. J.; Abubakar, A.; Hajjeh, R. *Lancet* **2020**, *396*, 1786−1788.

(2) Gurdasani, D.; Bear, L.; Bogaert, D.; Burgess, R. A.; Busse, R.; Cacciola, R.; Charpak, Y.; Colbourn, T.; Brury, J.; Friston, K.; Gallo, V.; Goldman, L. R.; Greenhalgh, T.; Hyde, Z.; Kuppalli, K.; Majumder, M. S.; Martin-Moreno, J. M.; McKee, M.; Michie, S.; Mossialos, E.; Nouri, A.; Pagel, C.; Pimenta, D.; Popescu, S.; Priesemann, V.; Rasmussen, A. L.; Reicher, S.; Ricciardi, W.; Rice, K.; Silver, J.; Smith, T. C.; Wenham, C.; West, R.; Yamey, G.; Yates, C.; Ziauddeen, H. *Lancet* **2020**, *396*, 1800−1801.

(3) Kam, K. Q.; Yung, C. F.; Cui, L.; Lin, R. T. P.; Mak, T. M.; Maiwald, M.; Li, J. H.; Chong, C. Y.; Nadua, k.; Tan, N. W. H.; Thoon, K. C. *Clin. Infect. Dis.* **2020**, *71*, 847−849.

(4) WHO Solidarity Trial Consortium. *N. Engl. J. Med.* **2021**, *384*, 497−511.

(5) Bloom, B. R.; Nowak, G. J.; Orenstein, W. *N. Engl. J. Med.* **2020**, *383*, 2202−2204.

(6) Corman, V. M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D. K. W.; Bleicker, T.; Brunink, S.; Schneider, J.; Schmidt, M. L.; Mulders, D. G. J. C.; Haagmans, B. L.; van der Veer, B.; van den Brink, S. vd.; Wijsman, L.; Goderski, G.; Romette, J. L.; Ellis, J.;

https://doi.org/10.1021/acs.analchem.1c01061
Anal. Chem. 2021, 93, 9174−9182

Zambon, M.; Peiris, M.; Goossens, H.; Reusken, C.; Koopmans, M. P. G.; Drosten, C. *Eurosurveillance* **2020**, *25*, No. 2000045.
(7) Wong, M. L.; Medrano, J. F. *BioTechniques* **2005**, *39*, 75−85.
(8) Huggett, J.; Dheda, K.; Bustin, S.; Zumla, A. *Genes Immun.* **2005**, *6*, 279−284.
(9) Yan, C.; Cui, J.; Huang, L.; Du, B.; Chen, L.; Xue, G.; Li, S.; Zhang, W.; Zhao, L.; Sun, Y.; Yao, H.; Li, N.; Zhao, H.; Feng, Y.; Liu, S.; Zhang, Q.; Liu, D.; Yuan, J. *Clin. Microbiol. Infect.* **2020**, *26*, 773−779.
(10) Lin, Q.; Wen, D.; Wu, J.; Liu, L.; Wu, W.; Fang, X.; Kong, J. *Anal. Chem.* **2020**, *92*, 9454−9458.
(11) Cao, Y. C.; Jin, R.; Mirkin, C. A. *Science* **2002**, *297*, 1536−1540.
(12) Langer, J.; Aberasturi, D. J. d.; Aizpurua, J.; Alvarez-Puebla, R. A.; Auguie, B.; Baumberg, J. J.; Bazan, G. C.; Bell, S. E. J.; Boisen, A.; Brolo, A. G.; Choo, J.; Cialla-May, D.; Deckert, V.; Fabris, L.; Faulds, K.; Abajo, F. J. G. d.; Goodacre, R.; Graham, D.; Haes, A. J.; Haynes, C. L.; Huck, C.; Itoh, T.; Kall, M.; Kneipp, J.; Kotov, N. A.; Kuang, H.; Ru, E. C. L.; Lee, H. K.; Li, J.-F.; Ling, X. Y.; Maier, S. A.; Mayerhöfer, T.; Moskovits, M.; et al. *ACS Nano* **2020**, *14*, 28−117.
(13) Ding, S.-Y.; Yi, J.; Li, J.-F.; Ren, B.; Wu, D.-Y.; Panneerselvam, R.; Tian, Z.-Q. *Nat. Rev. Mater.* **2016**, *1*, No. 16021.
(14) Ding, S.-Y.; You, E.-M.; Tian, Z.-Q.; Moskovits, M. *Chem. Soc. Rev.* **2017**, *46*, 4042−4076.
(15) Hernández-Arteag, A. C.; Delgado-Nieblas, F. C.; Ojeda-Galvan, H. J.; Velazquez-Salazar, J. J.; Vinogradova, E.; Jose-Yacaman, M.; Guirado-Lopez, R. A.; Navarro-Contreras, H. R. *J. Phys. Chem. C* **2017**, *121*, 21045−21056.
(16) Luo, S.-C.; Sivashanmugan, K.; Liao, J.-D.; Yao, C.-K.; Peng, H.-C. *Biosens. Bioelectron.* **2014**, *61*, 232−240.
(17) Lim, J.-y.; Nam, J.-s.; Shin, H.; Park, J.; Song, H.-i.; Kang, M.; Lim, K.-i.; Choi, Y. *Anal. Chem.* **2019**, *91*, 5677−5684.
(18) Wang, Y.; Ruan, Q.; Lei, Z.-C.; Lin, S.-C.; Zhu, Z.; Zhou, L.; Yang, C. *Anal. Chem.* **2018**, *90*, 5224−5231.
(19) Kamińska, A.; Witkowska, E.; Winkler, K.; Dzięcielewski, I.; Weyher, J. L.; Waluk, J. *Biosens. Bioelectron.* **2015**, *66*, 461−467.
(20) Zong, C.; Xu, M.; Xu, L.-J.; Wei, T.; Ma, X.; Zheng, X.-S.; Hu, R.; Ren, B. *Chem. Rev.* **2018**, *118*, 4946−4980.
(21) Liu, H.; Dai, E.; Xiao, R.; Zhou, Z.; Zhang, M.; Bai, Z.; Shao, Y.; Qi, K.; Tu, J.; Wang, C.; Wang, S. *Sens. Actuators, B* **2021**, *329*, No. 129196.
(22) Masango, S. S.; Hackler, R. A.; Large, N.; Henry, A.-I.; McAnally, M. O.; Schatz, G. C.; Stair, P. C.; Van Duyne, R. P. *Nano Lett.* **2016**, *16*, 4251−4259.
(23) Liu, C.; Yang, Y.; Gao, Y.; Shen, C.; Ju, B.; Liu, C.; Tang, X.; Wei, J.; Ma, X.; Liu, W.; Xu, S.; Liu, Y.; Yuan, J.; Wu, J.; Liu, Z.; Zhang, Z.; Wang, P.; Liu, L. *bioRxiv* **2020**, DOI: 10.1101/2020.03.02.972927.
(24) Frens, G. *Nat. Phys. Sci.* **1973**, *241*, 20−22.
(25) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. *Science* **2020**, *367*, 1260−1263.
(26) Kresse, G.; Furthmuller, J. *Phys. Rev. B* **1996**, *54*, 11169−11186.
(27) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(28) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758−1775.
(29) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem. A* **1994**, *98*, 11623−11627.
(30) Morháč, M.; Kliman, J.; Matousek, V.; Veselsky, M.; Turzo, I. *Nucl. Instrum. Methods Phys. Res., Sect. A* **1997**, *401*, 113−132.
(31) Barman, I.; Dingari, N. C.; Kang, J. W.; Horowitz, G. L.; Dasari, R. R.; Feld, M. S. *Anal. Chem.* **2012**, *84*, 2474−2482.
(32) Kočišová, E.; Prochazka, M. *J. Raman Spectrosc.* **2011**, *42*, 1606−1610.
(33) Wiercigroch, E.; Szafraniec, E.; Czamara, K.; Pacia, M. Z.; Majzner, K.; Kochan, K.; Kaczor, A.; Baranska, M.; Malek, K. *Spectrochim. Acta, Part A* **2017**, *185*, 317−335.
(34) Rygula, A.; Majzner, K.; Marzec, K. M.; Kaczor, A.; Pilarczyk, M.; Baranska, M. *J. Raman Spectrosc.* **2013**, *44*, 1061−1076.
(35) De Gelder, J.; Gussem, K. D.; Vandenabeele, P.; Moens, L. *J. Raman Spectrosc.* **2007**, *38*, 1133−1147.
(36) Arboleda, P. H.; Loppnow, G. R. *Anal. Chem.* **2000**, *72*, 2093−2098.
(37) Zhu, G.; Zhu, X.; Fan, Q.; Wan, X. *Spectrochim. Acta, Part A* **2011**, *78*, 1187−1195.
(38) Woo, H.; Park, S.-J.; Choi, Y. K.; Park, T.; Tanveer, M.; Cao, Y.; Kern, N. R.; Lee, J.; Yeom, M. S.; Croll, T. I.; Seok, C.; Im, W. *J. Phys. Chem. B* **2020**, *124*, 7128−7137.
(39) Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P. *Glycobiology* **2020**, *30*, 981−988.
(40) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. *Science* **2020**, *369*, 330−333.
(41) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. *Nature* **2020**, *581*, 215−220.
(42) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. *Cell* **2020**, *181*, 281−292.
(43) Tuma, R. *J. Raman Spectrosc.* **2005**, *36*, 307−319.
(44) Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y.-g.; Choi, B. H.; Kang, K.-W.; Jeong, H.; Park, Y.; Hong, S.; Kim, H. K.; Choi, Y. *ACS Nano* **2020**, *14*, 5435−5444.
(45) Ho, C.-S.; Jean, N.; Hogan, C. A.; Blackmon, L.; Jeffrey, S. S.; Holodniy, M.; Banaei, N.; Saleh, A. A. E.; Ermon, S.; Dionne, J. *Nat. Commun.* **2019**, *10*, No. 4927.
(46) Pan, Y.; Zhang, D.; Yang, P.; Poon, L. L. M.; Wang, Q. *Lancet Infect. Dis.* **2020**, *20*, 411−412.
(47) Wu, N. C.; Yuan, M.; Bangaru, S.; Huang, D. L.; Zhu, X. Y.; Lee, C. C. D.; Turner, H. L.; Peng, L. H.; Yang, L. L.; Burton, D. R.; Nemazee, D.; Ward, A. B.; Wilson, I. A. *PLoS Pathog.* **2020**, *16*, No. e1009089.
(48) Yuan, M.; Wu, N. C.; Zhu, X. Y.; Lee, C. C. D.; So, R. T. Y.; Lv, H. B.; Mok, C. K. P.; Wilson, I. A. *Science* **2020**, *368*, 630.
(49) Ye, Q. Z.; West, A. M. V.; Silletti, S.; Corbett, K. D. *Protein Sci.* **2020**, *29*, 1890−1901.
(50) Hurlburt, N. K.; Seydoux, E.; Wan, Y. H.; Edara, V. V.; Stuart, A. B.; Feng, J. L.; Suthar, M. S.; McGuire, A. T.; Stamatatos, L.; Pancera, M. *Nat. Commun.* **2020**, *11*, No. 5413.

https://doi.org/10.1021/acs.analchem.1c01061
Anal. Chem. 2021, 93, 9174−9182