Nanostructured Au-Based Surface-Enhanced Raman Scattering Substrates and Multivariate Regression for pH Sensing

Seju Kang,[a,b] *Wonil Nam*,[c] *Wei Zhou*,[c] *Inyoung Kim*,[d] and *Peter J. Vikesland*[a,b*]

[a]Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061;

[b]Virginia Tech Institute of Critical Technology and Applied Science (ICTAS) Sustainable Nanotechnology Center (VTSuN), Blacksburg24061, Virginia 24061;

[c]Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia 24061;

[d]Department of Statistics, Virginia Tech, Blacksburg, Virginia 24061;

***** *Corresponding author.* E-mail address: pvikes@vt.edu (P. Vikesland).

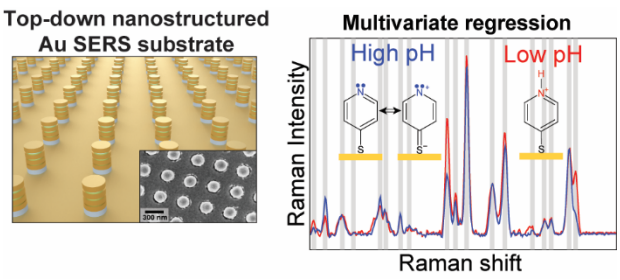
Postal address: 420 Durham, Virginia Tech, Blacksburg, VA 24061, USA

Keywords: *surface-enhanced Raman spectroscopy (SERS), pH, Top-down nanofabrication, Machine learning, multivariate regression*

Abstract

Compatibility in a range of media is vitally important for surface-enhanced Raman scattering (SERS) enabled pH detection. We report universal pH detection in a range of media using top-down nanostructured gold SERS substrates and multivariate regression. SERS substrates with vertically stacked multiple nanogap hotspots functionalized with the sensing molecule 4-mercaptopyridine (4-Mpy) exhibited high spatial uniformity. Standard ratiometric pH detection enabled development of a Boltzmann equation-based calibration curve for phosphate buffered saline. This calibration curve, however, could not be used to predict pH in other media such as carbonate buffer, apple juice, milk, and wastewater. To address SERS interferences that occur in these different media compositions, multivariate regression was successfully applied to pH prediction for all five media. A total of 19 spectral features in the 4-Mpy SERS spectra were extracted and used for model development. A nonparametric Gaussian process regression (GPR) model with 5/2 Matérn kernel function exhibited the greatest pH prediction accuracy with a root mean square error (RMSE) of 0.8139 among other multivariate regression models. This model was generalizable and capable of determining pH within media that had not been used for model training.

Graphical abstract



Introduction

Sensitive and accurate pH measurements in aqueous environments are important to a variety of science and engineering fields. Solution pH governs not only physical/chemical dynamics, but also affects biological activity. For example, colloidal particles in suspension show pH-dependent surface charges that affect measured interfacial potentials, and the enzymatic activity of nitrifying bacteria in biological wastewater treatment is known to be pH sensitive.[1] Similarly, the pH of clinical specimens (e.g., blood, serum, sputum) has physiological implications. Intracellular pH is an essential parameter for cell and organ function[2,3] and extracellular pH can be used to differentiate cancer and normal cells.[4] The consistent monitoring of the pH of commercial juice products can provide alarms about possible microbiological infection.[5,6]

Surface-enhanced Raman spectroscopy (SERS) has been recognized as a promising analytical technique for pH detection.[7–14] SERS has been proven to have high sensitivity for analyte detection and provides the benefits of rapidity, non-destructivity, and low-cost.[15–18] Furthermore, the capacity to access microenvironments and the inherent small sample volumes involved make SERS-based pH detection highly appealing. A number of SERS pH reporters containing a thiol group for surface anchoring and pH sensitive functional groups (e.g., amine, carboxyl, pyridyl) have been used to measure pH: 2-aminothiophenol (2-ATP),[19] 4-mercaptobenzoic acid (4-MBA),[3,20–25] and 4-mercaptopyridine (4-Mpy).[26–31] The protonation/deprotonation of the pH sensitive functional groups within these pH reporters result in measurable changes in the SERS spectra.

Despite a number of successes of the SERS pH sensing approach,[32–35] there remain challenges that must be addressed for field application of SERS pH nanoprobes. One major challenge arises from the use of bottom-up synthesized nanoparticles. Gold nanoparticles (AuNPs)

are commonly used for SERS due to their facile synthesis, stability, and biocompatibility.[36,37] pH reporters can be readily attached to the AuNP surface through strong covalent Au–S bonds.[38] SERS signals obtained by this method reflect the density of SERS hotspots generated by nanoparticle aggregation. The aggregation of pH reporter functionalized AuNPs is generally uncontrollable and the resulting heterogeneous spatial distribution can lead to poor uniformity. Numerous efforts have been made to resolve this issue. For instance, blocking agents such as bovine serum albumin (BSA) have been used to form a protective layer and prevent undesired AuNP aggregation.[2] In addition, co-solvent (water and ethanol) controlled aggregation has shown promise as a means to reproducibly control SERS hotspot generation.[23] Nonetheless, the intrinsic heterogeneity of nanoparticle aggregation may still impact reliability.

Another challenge affecting the use of SERS pH nanoprobes is the need to conduct ratiometric analyses in different media. To the best of our knowledge, all of the prior studies on the use of SERS pH nanoprobes have relied upon ratiometric analyses for pH determination. By taking two different SERS peaks whose intensities vary depending on the protonation/deprotonation of functional groups, it is possible to develop correlations between peak ratio and pH. However, the chemical/biological composition of the media can interfere with the SERS spectra and distort the correlations.[39] For example, it has been shown that halide ions make the pyridinium thiolate more likely to protonate in an acidic solution and alter the SERS spectrum of 4-Mpy.[40–44] Furthermore, it has been determined that SERS pH nanoprobes can be susceptible to BSA interference under physiological concentrations,[25] and different cation compositions (e.g., K^+, Na^+, Ca^{2+}, Mg^{2+}) in the cell culture medium can interfere with the SERS spectra.[45] Given the possible effects of the media on the SERS spectrum, the continued application of ratiometric analyses may not be the best option for robust SERS pH measurements.

To solve the challenges faced by previous SERS pH studies, we introduce 1) top-down nanostructured Au SERS substrates and 2) multivariate regression with the aim of accurate pH sensing in a variety of media. Lithography-based top-down nanofabrication enables the sophisticated design of SERS substrates. Top-down nanostructured SERS substrates have numerous advantages over bottom-up synthesized nanoparticles in that they inherently possess greater uniformity and reusability.[46] We furthermore hypothesized that the simultaneous selection of all pH sensitive SERS peaks would better reflect chemical/biological differences between media and enable more accurate and matrix agnostic pH detection. Prior studies have demonstrated the potential for SERS to measure dynamic changes within intracellular environments using multiple vibrational bands,[47] and recently, machine learning algorithms that are capable of learning high-dimensional variables have gained attention as a more robust means for SERS analysis than ratiometric approaches.[48–50] Multivariate regression based on Gaussian process (GP) is a supervised machine learning approach that can define a statistical model to study the relationship between several correlated predictor variables and a dependent variable.[51] Since GP is a family of functions, the GP regression (GPR) is also considered a nonparametric regression that is quite flexible for the building of such relationships without strong parametric assumptions. By considering changes in several spectral features in the SERS spectrum of a pH reporter in response to pH changes, a multivariate regression model based on GP is expected to accurately predict pH in the given media irrespective of potential spectral interferences.

Figure 1 illustrates the workflow for universal pH sensing using a top-down nanostructured SERS substrate and multivariate regression. We functionalized the SERS substrates with 4-Mpy as a pH reporter and incubated them in five different media: phosphate-buffered saline (PBS), carbonate buffer, apple juice, milk, and wastewater. These matrices range

from quite simple, such as PBS and carbonate buffer that only contain inorganic salts, to quite complex (e.g., milk and wastewater). PBS is the most commonly used buffer for biological research and has the lowest concentration of salts that can maintain stable pH. Carbonate buffer simulates many systems for which pH is controlled by the equilibrium between the CO_2 in the air and the relative concentrations of carbonate and bicarbonate.[52] The food industry uses pH as an indicator of product quality control and for this reason, apple juice and milk were chosen as representatives of complex liquid food matrices. Wastewater, in contrast, is a highly heterogeneous matrix whose pH varies widely due to local differences in water quality and sewage inputs. In this study, SERS spectra of 4-Mpy were collected over a range of pH values in each of these representative media. We then developed a multivariate regression model to accurately estimate pH given our SERS dataset and compared this accuracy with that obtained via ratiometric analysis.

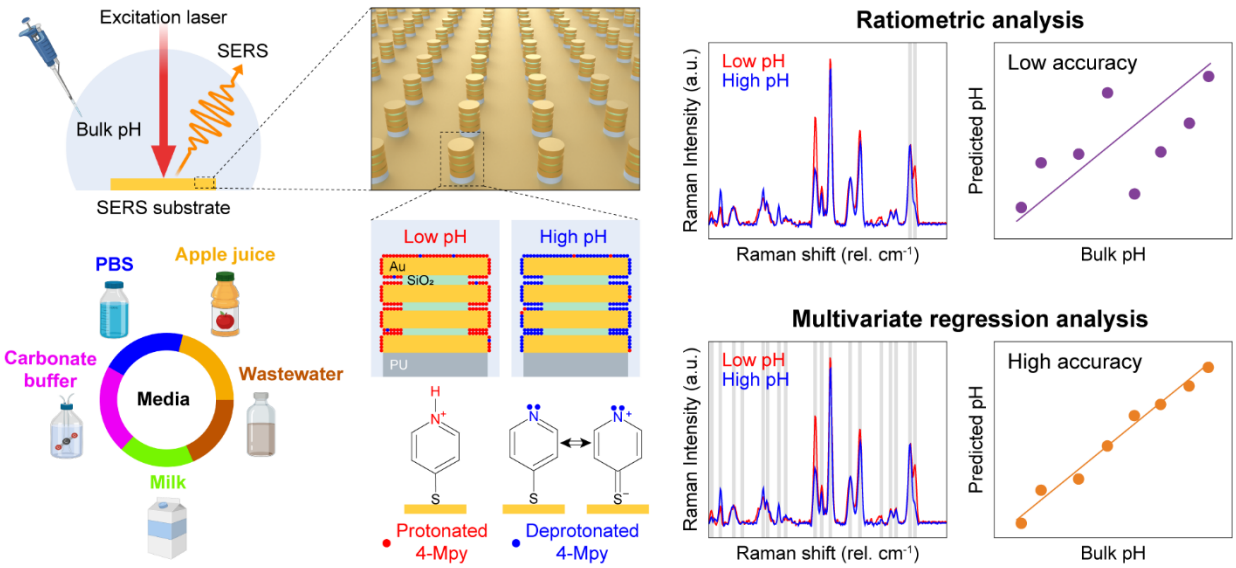


Figure 1 Schematic illustration of universal pH sensing enabled by the top-down nanostructured SERS substrates with different pH values (from 2 to 11) and five media: PBS, wastewater, carbonate buffer, apple juice, and milk. PU, polyurethane. The SERS spectrum of 4-Mpy deposited onto the SERS substrates was collected after incubation with different pH and media. Following SERS spectra collection, the accuracy of the calibrated pH prediction model from the multivariate regression algorithm was compared to ratiometric analysis.

Methods

Reagents and solution

Thiolated polyethylene glycol with a molecular weight of 1000 (1k-PEG-thiol), 4-Mpy, hydrochloric acid (HCl), sodium hydroxide (NaOH), $NaHCO_3$, and Na_2CO_3 were purchased from Sigma-Aldrich (St. Louis, MO). PBS at pH 7.4 was purchased from Invitrogen (Carlsbad, CA) and a wastewater sample was collected from the influent to the wastewater treatment plant in the Hampton Roads Sanitation District (HRSD; Virginia Beach, VA). Apple juice (100% Martinelli's Gold Medal apple juice, Watsonville, CA) and milk (2% reduced-fat milk, Simple Truth Organic®™, San Diego, CA) were purchased from the local grocery store. The pH values of these media were adjusted using HCl and NaOH except for carbonate buffer solution for which the pH was adjusted by mixing different volume ratios of two 0.1 M sodium bicarbonate (Na_2CO_3) and disodium carbonate ($NaHCO_3$) solutions. The bulk pH of the solution was measured using an Orion Versa Star pH meter (Thermo Fisher Scientific, Waltham, MA). All solutions were stored at 4 °C until use.

Fabrication of top-down nanostructured SERS substrates

Detailed SERS substrate fabrication processes are described elsewhere.[46,48,53] Briefly, a composite polydimethylsiloxane (PDMS) stamp (diameter = 120 nm, periodicity = 400 nm, height = 150 nm) was prepared by soft lithography.[54] We used UV-curable polyurethane (PU; NOA83H, Norland Product Inc., USA) to fabricate periodic nanopillar arrays on a flexible and optically transparent polyester (PET) film. After 10 min UV curing, an additional overnight heat-curing was performed in a convection oven at 80 °C. We then deposited alternating layers of Au (30 nm) and SiO_2 (6 nm, 8 nm, and 12 nm from the bottom) by electron-beam deposition (PVD250, Kurt J. Lesker Company, USA). One nm of Cr on PU nanopillar arrays and 1 nm of Ti at every metal-

dielectric interface were deposited as adhesion layers. We used 10:1 buffered oxide etchant (BOE) solution (Transene Inc., USA) to partially etch SiO_2 layers for 20 s to open dielectric nanogaps, thereby activating SERS hotspots.

Collection of SERS spectra of 4-Mpy in response to pH

To functionalize 4-Mpy on the top-down nanostructured SERS substrate, the substrate was cut into pieces of 5 mm × 5 mm. One piece was attached to the small petri dish in the diameter of 47 mm using a UV curable PU optical adhesive (NOA 61, Norland Product Inc., USA) with 5 s of UV curing. Then, the piece was immersed in 5 mL of 10 μM 4-Mpy solution for one hour. To stabilize the substrate, after 4-Mpy functionalization, the suspension was replaced by 5 mL of 1k-PEG-thiol (0.5% weight ratio) and further incubated for another hour. PEG coating can enhance substrate biocompatibility[55,56] and inhibit metal dissolution.[57] Then, the suspension was decanted and the substrate was washed several times by deionized water. The final functionalized SERS substrate was stored in deionized water at room temperature and covered with aluminum foil until use. For pH sensing, the supernatant was decanted, and the substrate was dried under a gentle flow of N_2 gas. An aliquot of 100 μL of pH solution was pipetted onto the SERS substrate and the SERS spectrum was immediately collected. The pHs of PBS, apple juice, milk, and wastewater solutions were adjusted in the range of pH 2-11 at one pH unit increments by adding HCl and NaOH. The pH range of carbonate buffer was limited between 8.6 to 10.2. The carbonate buffer pH was adjusted by mixing different ratios of Na_2CO_3 and $NaHCO_3$ to minimize pH change with time due to CO_2 transfer from the air. For milk, due to the high turbidity of the milk solids, the pipetted volume was reduced to 10 μL.

Instrumentation and data processing

The SERS spectra of 4-Mpy with different pH values on the SERS substrate were collected

using a confocal Raman spectrometer (Alpha500R, WITec, Germany) with a 785 nm diode laser (Toptica Photonics, Germany) and 10× objective lens. The laser power was set to 10 mW, and a 300 grooves/mm grating was used. Single point SERS spectra were measured using 10 s integration time per point with 6× accumulation. Single acquisitions from three different regions of the substrate were collected as triplicates. Before measurement, instrumental calibration was verified by the silicon peak at 520 cm^{-1}. The collected SERS spectra were processed using built-in software (Project Five v. 5.0, WITec, Germany). Each SERS spectrum was first processed by graph smoothing and cosmic ray removal (CRR). Then, the baseline was subtracted using the shape function.

Ratiometric analysis and multivariate regression

For ratiometric analysis, among the many spectral features, the intensities at 1576 and 1612 cm^{-1} (I_{1576} and I_{1612}) were selected.[26,28] The peak ratio (I_{1576}/I_{1612}) for each medium was plotted against the bulk pH measured by a pH meter. The best-fit calibration for the PBS solution was obtained using the Boltzmann equation and its compatibility to other media was investigated.

For multivariate regression, all spectral features in the SERS spectrum were extracted after normalization by the peak at 77 cm^{-1}. This pseudo-peak generated by a long-pass filter originates from plasmon-enhanced electronic Raman scattering (ERS). We have recently reported that ERS-based SERS calibration can more accurately determine concentrations at plasmonic hotspots with reduced spatial and temporal variations, thus enabling more rigorous quantitative analysis.[58–60] Based on this benefit of ERS normalization, it is expected that the multivariate regression model can be developed across a variety of nanostructured SERS substrates with spatially varying SERS hotspots. Several distinct SERS peaks between 350 and 1800 cm^{-1} were selected using automated peak labeling within the WITec Control Five (v. 5.0) software with the minimum relative height

(i.e., height to the average of all peaks) set to 0.1. Using this function, 19 spectral features were identified: 371, 390, 421, 484, 553, 680, 708, 780, 821, 1003, 1040, 1095, 1208, 1274, 1412, 1462, 1500, 1576, and 1612 cm^{-1}. We collected a total of 144 (48 pH values × 3 replicates per pH) SERS spectra for different pHs and media. Each SERS spectrum can be labeled as a corresponding bulk pH measured by a pH meter (numeric). The SERS dataset thus consists of 144 pHs × 19 spectral features. The multivariate regression models were trained using the regression learner application in Mathworks MATLAB/SIMULINK (ver. R2020a). It offers a variety of regression models (e.g., GPR, support vector machine (SVM), linear regression, regression trees, the ensemble of trees) with sub-specialized functions. To validate the model, 12-fold cross-validation was conducted. Simply put, the 144 pH labeled dataset was divided into 12 subsets of equal size. Iteratively, 11 of the subsets were used for training, while one was used for testing. For each test, the root mean square error (RMSE) of the test points from the prediction model was calculated. We evaluated the model by the average of RMSEs from 12 test results.

Results and Discussion

Nanostructured SERS substrate characterization

Figures 2A and **B** show photographic and top-view scanning electron microscope (SEM) images of our large-area (≈16 cm^2) top-down nanostructured SERS substrate. Our recent studies indicate that vertical stacking of multiple metal-insulator-metal (MIM) layers on periodic nanopillar arrays create vertically-oriented (out-of-plane) nanogaps that provide uniform 3D plasmonic hotspots with a SERS enhancement factor (EF) $\approx 5\times10^7$.[46,48,53] More detailed characterization of our SERS substrate is provided in our prior publications.[46,48,53] Top-down molding-based soft nanolithography offers fabrication scalability and enables cost-effective, mass production of reusable nanostructured SERS substrates.

We functionalized the SERS substrates with 4-Mpy as a pH reporter. 4-Mpy consists of a pyridine ring with a thiol group that attaches to the Au surface via the Au−S bond while exhibiting a large Raman cross-section due to the pyridine ring (**Figure 1**). Changes in pH result in protonation/deprotonation of heterocyclic nitrogen that can be detected by SERS. **Figure 2C** illustrates the SERS spectra for 4-Mpy in PBS at pH 7. This result shows the successful functionalization of 4-Mpy on the SERS substrates. To evaluate the uniformity of 4-Mpy functionalization across a substrate, we collected SERS spectra from a 10 μm × 10 μm area with 10 × 10 (X × Y) points and 1 s integration time per point. Using the characteristic peak at 1096 cm^{-1} we then examined the spatial distribution of its intensity (I_{1096}) across the scan area. As shown in **Figure 2D**, I_{1096} was evenly distributed with a relative standard deviation (RSD) of 7.2%. The low RSD reflects the high uniformity of the functionalized SERS substrates.[61] Given the homogeneity of the SERS substrates, we then elected to collect SERS spectra from three different regions of a given substrate using a single point collection of 10 s integration time and six accumulations throughout this study. In this manner, we were able to rapidly collect high signal-to-noise ratio (SNR) 4-Mpy SERS spectra. A high SNR is critical to the development of a multivariate regression model since it is needed to facilitate monitoring of small peak responses to changes in pH.

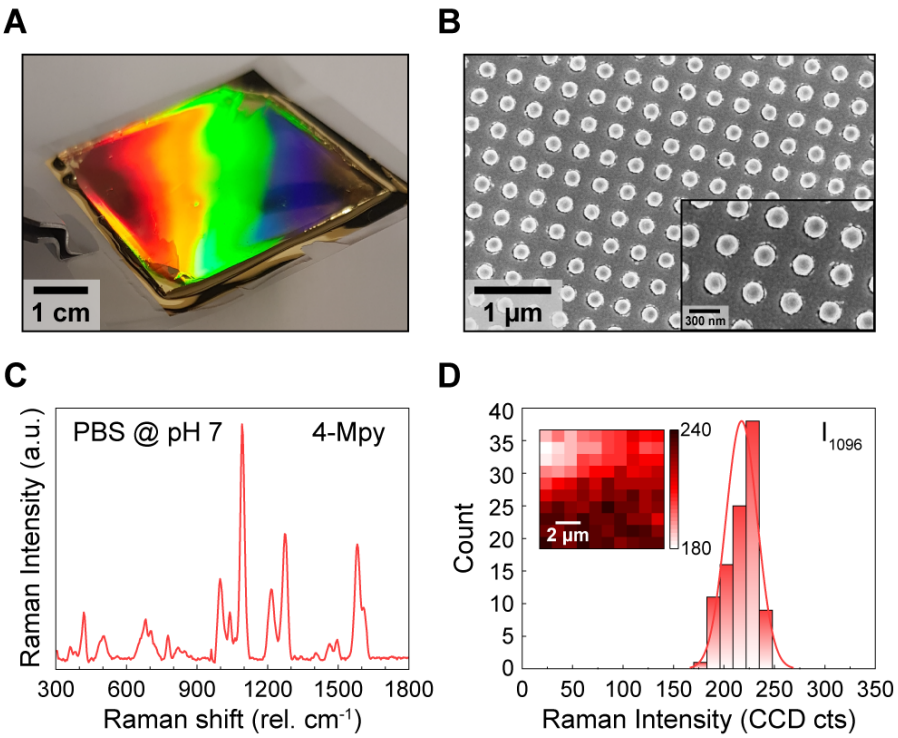


Figure 2 (A,B) Bright field and SEM images of the top-down nanostructured SERS substrates (C) The SERS spectrum of 4-Mpy functionalized SERS substrates in PBS at pH 7 (D) Histogram of the Raman intensity at 1096 cm^{-1} for 4-Mpy functionalized SERS substrates across a 10 µm × 10 µm scanning area with 10 × 10 (X × Y) points. The inset shows the spatial distribution of the Raman intensity at 1096 cm^{-1} across the scanning area.

SERS spectra of 4-Mpy with different pH and media

SERS spectra of 4-Mpy were collected as a function of pH in a number of different matrices. The collected SERS spectra are visually presented hereafter by normalization using the intensity of the 77 cm^{-1} pseudo-peak as an internal standard. **Figure 3A** shows the SERS spectra of 4-Mpy on the SERS substrates in PBS at pH 2 and 11. While their intensities vary, several distinct peaks with positions between 350 and 1800 cm^{-1} were observed in the SERS spectra at both pH values. The 4-Mpy structure contains a number of chemical bonds that produce SERS peaks that reflect different C—C, C—N, C—S, and C—H bonds, orientations, and vibrational modes (**Table S1**).

The SERS spectrum at pH 11 had remarkably higher signal intensities at 1000, 1095, 1274, and 1612 cm^{-1} than the SERS spectrum at pH 2. This difference reflects the different ionization states of 4-Mpy on the surface. Theoretically, there are two ways that 4-Mpy attaches to the Au surface (**Figure 1**). 4-Mpy can associate with Au either through the formation of a covalent Au−S bond or non-covalently through the lone pair of electrons on sulfur.[27] Under low pH conditions, 4-Mpy primarily has the aromatic ring with the protonated thiol group. Under this condition, both C=N and C=C have large Raman cross-sections. Under high pH conditions, two resonance ionization states of 4-Mpy (i.e., the thiol-thione tautomer) are common, leading to a decrease in the aromaticity of the pyridine ring and a relatively lower fraction of C=N.[27] The reduction in aromatic pyridine ring formation led to a decrease in the intensity at 1000 cm^{-1}. The different ratios of C=N to C=C bonds are reflected by whether the spectrum has two distinct peaks at 1576 and 1612 cm^{-1} or a single strong peak at 1576 cm^{-1}. Also, the C=S bond in the thione structure exhibits strong SERS intensity at 1095 cm^{-1} corresponding to the co-called X-sensitive band.[27]

We compared the SERS spectra of 4-Mpy in different media at pH 9 (**Figure 3B**). The consistency and clarity of the SERS spectra of 4-Mpy across this range of media illustrate the robustness of the SERS substrates. However, while the SERS profiles are fairly similar there remain distinct differences in peak intensities. Such differences reflect interferences in the SERS spectra that result from chemical/biological constituents within the various media. For example, prior studies have reported the effect of halide ions on the SERS spectrum of the pyridine ring.[28,40–44] At low pH, protonated pyridine (i.e., pyridinium) transforms into pyridine in the vicinity of halide ions. In addition, some proteins can form a steric protection layer with Au−S bonds that decrease the intensity of the ring breathing mode at 1000 cm^{-1}.[62,63] The matrices used in this study were purposely chosen as they are pertinent to a variety of fields: biotechnology (PBS), medical

science (carbonate buffer), food industry (apple juice and milk), and environmental science (wastewater). Each media has distinct levels of chemical/biological constituents such as ionic components and organic/inorganic matter. Such differences are expected to impact the SERS profiles of 4-Mpy at the same pH.

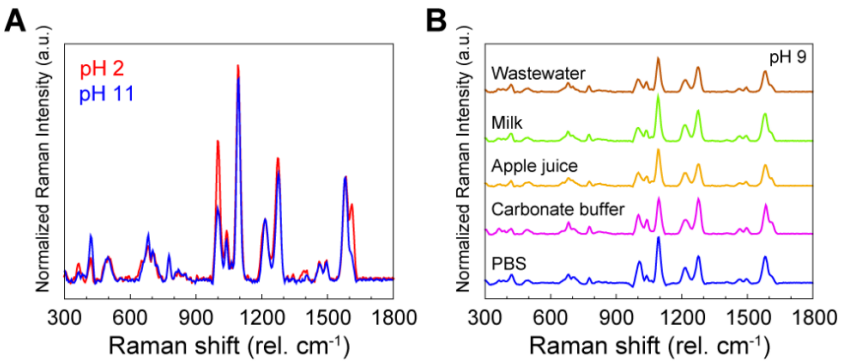


Figure 3 (A) Comparison of SERS spectra of 4-Mpy on the SERS substrates in PBS for pH 2 and 11. (B) Vertically stacked SERS spectra of 4-Mpy on the SERS substrates in five different media at pH 9

Ratiometric analysis to predict pH

Similar to prior work with SERS pH nanoprobes, we first applied ratiometric analysis to develop a pH calibration curve. Given the various pH-dependent ionization states of 4-Mpy different intensity ratios can be used to develop a ratiometric analysis. Previously, the peaks with the highest SNR have been commonly used for ratiometric analysis: 1000 vs. 1095 cm^{-1},[63] 1208 vs. 1274 cm^{-1},[2] or 1576 vs. 1612 cm^{-1}.[26,28] **Figure 4A** shows vertically stacked SERS spectra of 4-Mpy in PBS solution at pH 2-11. The inset spectra indicate the variation of two characteristic peaks (red-colored columns, 1576 and 1612 cm^{-1}) that showed the most sensitive spectral change in response to pH and represent the best-fit calibration curve amongst the three previously reported intensity ratios (**Figure S1**). When the pH increased, I_{1612} decreased, while I_{1576} was essentially constant since 4-Mpy was deprotonated. Given this dataset of 4-Mpy SERS spectra at different pH, we then plotted the peak ratio (I_{1576}/I_{1612}) for five media against the pH values measured by a

pH meter (**Figure S2**). In each case, a positive correlation between the peak ratio and pH was observed. Each calibration curve was readily fit by a Boltzmann expression with $R^2 > 0.93$ except for carbonate buffer. The small standard deviations for the majority of the peak ratios (~10%) for triplicate measurements illustrate the excellent homogeneity and uniformity of the SERS substrates. Unfortunately, however, the calibration curve for each plot differed considerably from one another. Such a result implies that interferents impact SERS spectra of 4-Mpy within the various media, and thus the calibration curves shift.

Figure 4B shows a comprehensive plot relating the peak ratios and the pH values measured by a pH meter for all media. As a standard, the best-fit curve and the corresponding 95% confidence region for the PBS data points were embedded to investigate how well a specific media's data points overlap with them. The best-fit curve for PBS using the Boltzmann equation had a high correlation ($R^2 = 0.97$, $I_{1576}/I_{1612} = \mathbf{4.95} - \frac{3.63}{1+\exp\left(\frac{pH-8.78}{1.98}\right)}$). However, the other media's measured points differed significantly from the PBS-based calibration curve. This result was especially true under highly acidic and basic conditions (e.g., pH 2-4 and 9-11) where the majority of the data points for the other four media deviated from the calibration curve and fell outside the 95% confidence interval. As discussed earlier, these differences reflect spectral inferences due to the chemical/biological components of the other media. In addition to the PBS-based calibration curve, a comprehensive calibration curve was obtained by fitting a Boltzmann expression to the results collected in all media to explore the applicability of ratiometric analysis for universal pH detection (**Figure S3**). The calibration curve exhibited a relatively poor fit to data points from all media with an R^2 of only 0.76. We, therefore, concluded that while ratiometric analysis may be appropriate for a single highly-controlled media its applicability in other media is questionable. A

more robust approach is required to achieve media agnostic comprehensive SERS-enabled pH detection.

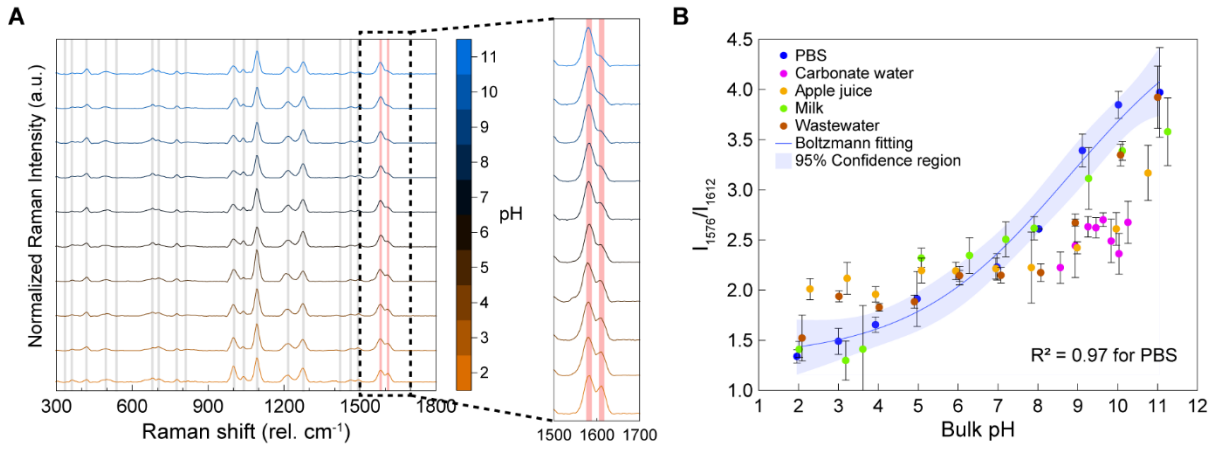


Figure 4 (A) Vertically stacked SERS spectra of 4-Mpy in PBS solution for bulk pH 2-11 measured by a pH meter. Gray regions indicate the 19 selected spectral features used for the development of the multivariate regression model. The zoomed-in SERS spectra show the two characteristic peaks at 1576 and 1612 cm^{-1} used for ratiometric analysis. (B) The comprehensive plot of I_{1576}/I_{1612} vs. bulk pH in different media. The blue line and region indicate the best-fit curve (Boltzmann equation) and 95% confidence region for the plot from the PBS solution.

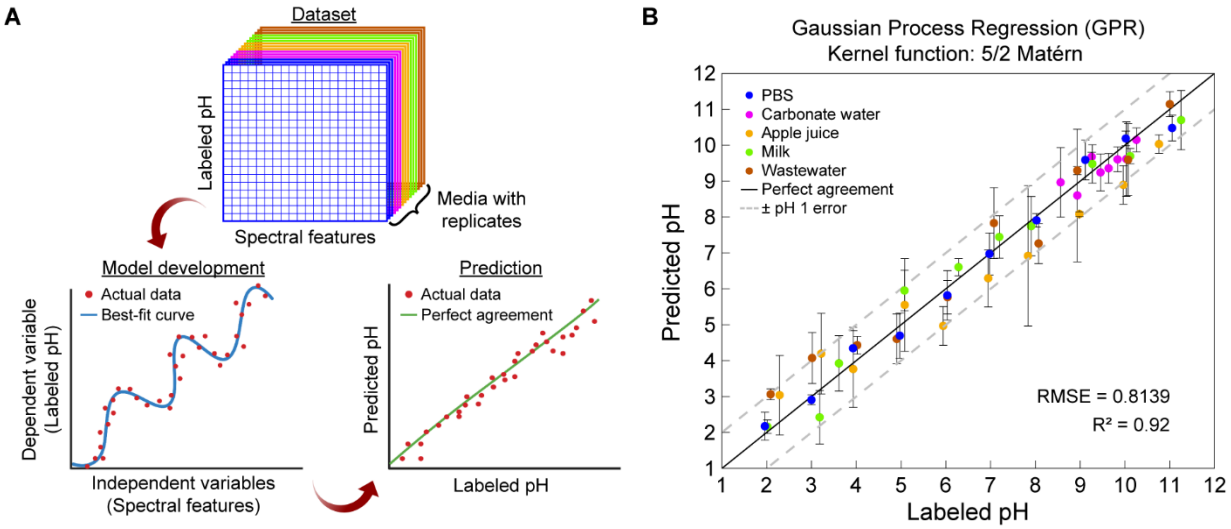
Multivariate nonparametric regression to predict pH

We used multivariate nonparametric-regression to build a pH prediction model that can be universally applied across highly variable media. To date, we found only a single literature report that used multivariate spectral analysis for SERS pH sensing.[64] In that effort, they analyzed spectral features using principal component analysis (PCA) and achieved significant improvement in pH prediction relative to a ratiometric approach. However, the capability of PCA is strictly limited to pH discrimination. In other words, because there is no defined calibration curve, pH cannot be quantitatively predicted with PCA. Hence, we developed a multivariate nonparametric regression model that can be used to directly predict pH by producing a predictive formula (i.e., the mathematical calibration equation) without imposing strong modeling assumptions. The use of a robust spectral analysis is expected to avoid false pH predictions that arise from disruptive

spectral signals that may arise due to variations between media. The gray-colored regions within **Figure 4A** indicate the 19 spectral features selected for the development of the multivariate regression model. **Figure 5A** summarizes the scheme used to obtain the spectral features and optimize the multivariate regression model. We tested 10 different pH values from 2 to 11 for PBS, apple juice, milk, and wastewater at 1-unit pH increments. In contrast to the other media, the pH of carbonate buffer was adjusted by mixing different ratios of $NaHCO_3$ and Na_2CO_3 to achieve 8 different pH values at ~0.2-unit increments (**Table S2**). A total of 144 (48 pH values × 3 replicates per pH) SERS spectra were numerically labeled based upon the corresponding pH meter determined bulk pH. Accordingly, a dataset consisting of 144 labeled pH values (numeric dependent variable) × 19 spectral features (independent variables) was collected. The multivariate regression model for pH detection was determined based upon the best-fit calibration curve where the difference between labeled and predicted pH values is minimized. The accuracy of the pH prediction model was estimated by plotting predicted pH values against those determined by a pH meter. Here, a GPR was applied to build a pH regression model based on the combined SERS dataset. GPR is the multivariate nonparametric regression that has the advantage of providing uncertainty bounds on the predictions while retaining suitability to small datasets. The Gaussian process is stochastic (i.e., a collection of random variables) with a multivariate Gaussian distribution and is a family of functions. Hence, GPR is considered a nonparametric regression. The process is governed by different kernel functions that measure the similarity between training inputs ($\boldsymbol{x_i}$) and the predicting input ($\boldsymbol{x'}$) ($|\boldsymbol{x_i} - \boldsymbol{x'}|$). The governing equations for the different GPR kernel functions used in this study are provided in **Table S3**.

Figure 5B illustrates pH prediction based on the GPR model with the 5/2 Matérn kernel function that is commonly used due to its' flexibility. The GPR model was trained by the dataset

consisting of the SERS spectra in all media and showed high agreement between the prediction and labeled pH (i.e., bulk pH measured by a pH meter). All of the data points from the different media fell within the regime of pH prediction with ± pH 1-unit deviations (gray dashed lines) with a RMSE of 0.8219 and an R^2 of 0.92. As well as improving the accuracy of prediction, the model showed a higher sensitivity for pH prediction than did the ratiometric analysis. As mentioned earlier, the ratiometric analysis in the case of carbonate buffer, apple juice, milk, and wastewater samples showed limited sensitivity for pH prediction in the range of pH 2-8. On the contrary, the data points for all samples were well linearized from pH 2 to 11 by the GPR model. The multivariate regression model showed higher superiority for pH prediction than the ratiometric analysis since it addresses the issue of background interferents within complex media simultaneous to model development. As discussed earlier, the ratiometric analysis failed to address the contribution of complex media for pH detection, showing a poor-Boltzmann fit across all media (R^2 = 0.76). Meanwhile, the PBS-based multivariate regression model also showed the poor predictive capacity for other media similar to the ratiometric analysis (**Figure S4**), the model based on all five media showed significant improvement in pH detection. We further compared the models developed in different numbers of media; one (PBS), two (PBS, carbonated buffer), three (PBS, carbonated buffer, apple juice), four (PBS, carbonated buffer, apple juice, milk) (**Figure S5**). All models exhibited high accuracies with RMSE of < 0.8834 and R^2 > 0.90.



394

395 **Figure 5** (A) The scheme of the dataset consisting of different labeled pH values (dependent
396 variable), spectral features (independent variables), and different media with replicates, the
397 development of a multivariate regression model, and pH prediction (B) pH prediction by the
398 optimized GPR model with 5/2 Matérn kernel for five media. Data points and error bars indicate
399 the average of predicted pH values from triplicate measurements and the standard deviations. The
400 black and gray dashed lines indicate the perfect agreement and its ± pH 1-unit differences.

401

402 To further validate the GPR model and investigate its field applicability, we predicted the
403 pH of a set of commercially available media that were not used for training under a wide range of
404 pH values: sports electrolyte drink (pH 2.11), white wine (pH 2.54), aloe drink (pH 3.42), green
405 tea (pH 6.02), pond water (pH 6.28), soy milk (pH 7.13), and cleaning solution (pH 10.20) (**Figure**
406 **6**). The pH values predicted by the GPR model were all within ± pH 1-unit of the labeled pH
407 values except for two cases. The pH for pond water predicted by the model was higher than the
408 labeled pH by 1.28 pH unit. It is reasonable that the large quantity of natural organic matter in the
409 pond water could interfere with the SERS spectrum.[65] The pH for cleaning solution predicted by
410 the model was lower than the labeled pH by 1.15 pH unit. The cleaning solution tested in this study
411 contained a surfactant for cleaning purposes that could also result in SERS interference.[66] The
412 effect of such potential interferents could be minimized by the removal of natural organic matter
413 and surfactant. Also, the model can be improved with better compatibility by feeding additional

SERS datasets from a wider range of media that include comparable amounts of natural organic matter and surfactant. We also adjusted the pH of the media under acidic, neutral, and basic conditions to investigate if these independent media also exhibited the linear pH dependence in the GPR model (**Figure S6**). The result showed that the GPR model can linearize the pH values of these media as well. Based on the high accuracy and compatibility of the multivariate regression model for pH prediction, we believe that multivariate regression can be applied in other parameter analysis in addition to pH. For example, hydrogen peroxide (H_2O_2) and hypochlorite (ClO^-) can be quantitatively detected using the SERS reporters, 3-mercaptophenylboronic acid (3-MPBA) and 4-mercaptophenol (4-MP).[67–72] Similar to pH detection, to the best of our knowledge, ratiometric analyses are typically applied for both targets and the compatibilities of these calibrations in other media have yet to be fully explored. It is expected that the application of multivariate regression to other targets can further expand the applicability of SERS-based detection in variable media.

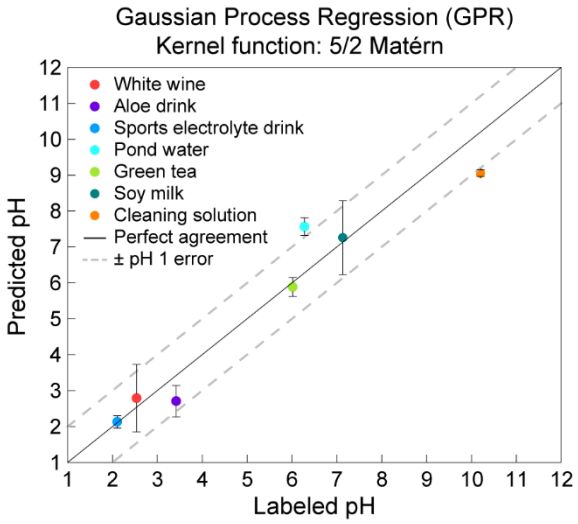


Figure 6 pH prediction of test media (sports electrolyte drink (pH 2.11), white wine (pH 2.54), aloe drink (pH 3.42), green tea (pH 6.02), pond water (pH 6.28), soy milk (pH 7.13), and an cleaning solution (pH 10.20)) by the optimized GPR model with 5/2 Matérn kernel. Data points and error bars indicate the average of predicted pHs from triplicate measurements and the standard deviations. The black and gray dashed lines indicate the perfect agreement and its ± pH 1-unit differences.

Comparison of multivariate regression models

We compared a number of multivariate regression models to determine the most powerful regression model for pH prediction by SERS spectra: GPR, SVM, linear regression, regression tree, and the ensemble of trees. The set of boxplots and scatter points in **Figure 7** shows the RMSE and R^2 values of the multivariate regression models for pH prediction. RMSE and R^2 are indicators of the model prediction performance. Each boxplot consists of 12 test results from the 12-fold cross-validation. Both GPR and SVM are kernel-based regressions. The GPR model uses the kernel to define the covariance of a prior distribution (i.e., the probability before the new data is collected) while SVM uses a kernel-based hyperplane that separates data points. We applied commonly used kernel functions for GPR (i.e., exponential, rational quadratic, 5/2 Matérn, and squared exponential). For SVM, parametric (linear, quadratic, cubic), and non-parametric (fine, medium, coarse Gaussian) kernel functions were applied. Linear regression is the simplest parametric regression that assumes a linear relationship between independent and dependent variables. A robust objective function (i.e., robust linear regression), can make the model less sensitive to outliers. The regression tree generates the tree from the root node with two-way branches to a leaf node. Several binary nodes with the branches from the root to the leaves determine the final response. Depending on the number of leaves, a fine tree (a large number of small leaves) and a medium/coarse tree (the fewer large leaves) can be applied. Finally, the ensemble of trees is the regression that combines multiple regression trees with least-squares boosting (boosted) or bootstrapping bagging (bagged) methods.

Among the regression models, all GPR, the SVM with some of the kernel functions (i.e., linear, quadratic, and medium Gaussian), and the linear/robust linear regression models showed the comparably lowest RMSE values based on the post-hoc Duncan's method ($\boldsymbol{\alpha} = 0.05$). The

458 RMSE values for the others followed the order of coarse Gaussian SVM ≤ boosted ensemble ≤
459 cubic SVM ≈ bagged ensemble ≤ medium tree ≤ fine tree ≤ coarse tree ≤ fine Gaussian SVM.
460 Generally, the regression tree and the ensemble models showed significantly larger RMSE values
461 than the GPR, SVM, and linear regression models.

462 To evaluate the stability of the regression models, the coefficient of variation (i.e.,
463 variabilities) of RMSEs from the 12 points were compared. The GPR models had small standard
464 deviations in the range of 25.4 to 28.2%. The linear and medium Gaussian SVM models showed
465 comparable variabilities of ~25.0%. Even though linear regression models showed great pH
466 predicition accuracy, they had relatively higher variabilities of 29.8 and 34.6%. The larger
467 variabilities imply a greater chance of overfitting. Besides, since there is no clearly defined
468 mathematical relationship between pH and the spectral features, nonparametric regression (GPR
469 or Gaussian SVM) would be preferred over parametric linear regression. Overall, among the
470 multivariate regression models, the GPR and medium Gaussian SVM models showed the best
471 performance for pH prediction with the lowest RMSEs and variabilities.

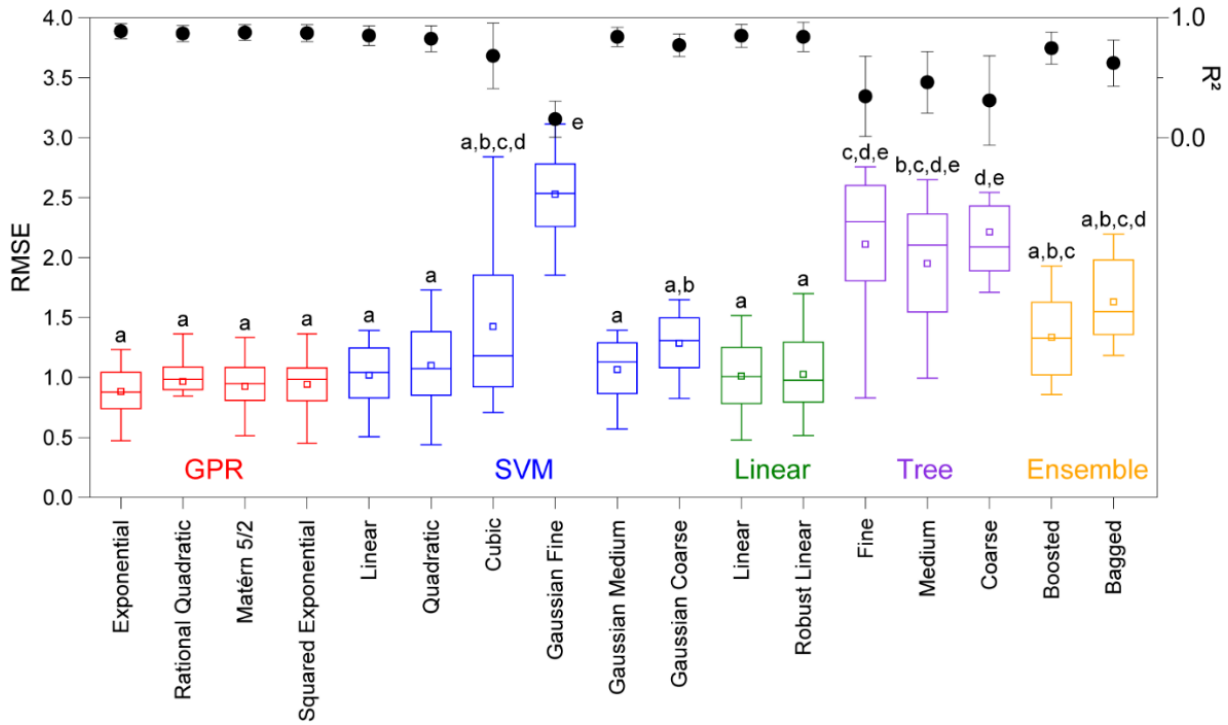


Figure 7 Comparison of the RMSEs of different multivariate regression models: (red) GPR with exponential, rational quadratic, 5/2 Matérn, squared exponential kernel functions, (blue) SVM models with linear, quadratic, cubic, fine/medium/coarse Gaussian kernel functions, (green) linear and robust linear regression, (purple) fine/medium/coarse regression tree, (orange) boosted and bagged ensemble tree. Each boxplot was made from 12 points from 12-fold cross-validation results, upper and lower quartiles, the outlier whisker, median, mean (square). Corresponding R^2 values for each model are shown in the upper graph. Error bar indicates the standard deviation of 12 points.

[a,b,c,d,e] Data annotated with the same character are not significantly different based on the post-hoc Duncan's method ($\alpha = 0.05$)

Conclusions

In this study, top-down nanostructured SERS substrates were used for universal pH sensing. The SERS substrates functionalized with the pH reporter 4-Mpy showed high spatial uniformity of a SERS signal at 1096 cm^{-1} with a RSD of 7.2% across the scan area (10 μm × 10 μm, 10 × 10 points). We collected SERS spectra of 4-Mpy on the top-down nanostructured SERS substrates with different solutions and pH values. pH-dependent protonation of 4-Mpy was reflected by changes in the SERS spectra. We initially applied ratiometric analysis to estimate the pH of five

media (i.e., PBS, carbonate buffer, apple juice, milk, and wastewater). The peak ratio I_{1578}/I_{1612} was plotted against the bulk pH, and the clear positive correlation between the two was observed with a well-fitted Boltzmann equation. However, the PBS-based calibration curve showed poor compatibility with other media due to SERS interferences. To collectively reflect all potential effects of different media on the SERS spectra in response to pH, we developed a multivariate regression model that was trained with 19 spectral features. The GPR model with a 5/2 Matérn kernel function showed the highest accuracy pH prediction with an RMSE of 0.81. The low variability of 12 cross-validation test results and accurate pH prediction for other media that were not used for training indicate the generalizability of the approach.

Associated content

SERS peaks and corresponding assignments; carbonate buffer solution recipe by mixing Na_2CO_3 and $NaHCO_3$ solutions with different volume ratio; governing equations for GPR kernel functions; The set of the plots of three SERS peak intensity ratios (I_{1000}/I_{1095}, I_{1208}/I_{1274}, I_{1576}/I_{1612}) and bulk pH measured by a pH meter in PBS; The set of the plots of SERS peak intensity ratios (I_{1576}/I_{1612}) and bulk pH measured by a pH meter in five different media: PBS, wastewater, carbonate buffer, apple juice, and milk; The comprehensive plot of I_{1576}/I_{1612} vs. bulk pH in five media; pH prediction by the optimized PBS-based GPR model with 5/2 Matérn kernel for five media; pH prediction by the optimized one to four media based GPR model with 5/2 Matérn kernel; pH prediction of commercially available media that were not used for training by the GPR model with 5/2 Matérn kernel

Acknowledgments

This research was supported by National Science Foundation grants (CBET-1705653 and CBET-2029911). Some of the icons used were obtained from Biorender.com

References

(1) Prinčič, A.; Mahne, I.; Megušar, F.; Paul, E. A.; Tiedje, J. M. Effects of pH and Oxygen and Ammonium Concentrations on the Community Structure of Nitrifying Bacteria from Wastewater. *Appl. Environ. Microbiol.* **1998**, *64* (10), 3584–3590. https://doi.org/10.1128/aem.64.10.3584-3590.1998.

(2) Zheng, X. S.; Hu, P.; Cui, Y.; Zong, C.; Feng, J. M.; Wang, X.; Ren, B. BSA-Coated Nanoparticles for Improved SERS-Based Intracellular pH Sensing. *Anal. Chem.* **2014**, *86* (24), 12250–12257. https://doi.org/10.1021/ac503404u.

(3) Jaworska, A.; Jamieson, L. E.; Malek, K.; Campbell, C. J.; Choo, J.; Chlopicki, S.; Baranska, M. SERS-Based Monitoring of the Intracellular pH in Endothelial Cells: The Influence of the Extracellular Environment and Tumour Necrosis Factor-α. *Analyst* **2015**, *140* (7), 2321–2329. https://doi.org/10.1039/c4an01988a.

(4) Hao, G.; Xu, Z. P.; Li, L. Manipulating Extracellular Tumour pH: An Effective Target for Cancer Therapy. *RSC Adv.* **2018**, *8* (39), 22182–22192. https://doi.org/10.1039/c8ra02095g.

(5) Patil, S.; Valdramidis, V. P.; Cullen, P. J.; Frias, J.; Bourke, P. Inactivation of *Escherichia Coli* by Ozone Treatment of Apple Juice at Different pH Levels. *Food Microbiol.* **2010**, *27* (6), 835–840. https://doi.org/10.1016/j.fm.2010.05.002.

(6) Wang, H.; Hu, Z.; Long, F.; Guo, C.; Niu, C.; Yuan, Y.; Yue, T. Combined Effect of Sugar Content and pH on the Growth of a Wild Strain of Zygosaccharomyces Rouxii and Time for Spoilage in Concentrated Apple Juice. *Food Control* **2016**, *59*, 298–305. https://doi.org/10.1016/j.foodcont.2015.05.040.

(7) Wang, Y.; Yan, B.; Chen, L. SERS Tags: Novel Optical Nanoprobes for Bioanalysis. *Chem. Rev.* **2013**, *113* (3), 1391–1428. https://doi.org/10.1021/cr300120g.

(8) Zou, X.; Wang, Y.; Liu, W.; Chen, L. M -Cresol Purple Functionalized Surface Enhanced Raman Scattering Paper Chips for Highly Sensitive Detection of pH in the Neutral pH Range. *Analyst* **2017**, *142* (13), 2333–2337. https://doi.org/10.1039/c7an00653e.

(9) Zhang, Q.; Wen, H.; Watanabe, K.; Kotani, I.; Ricci, M.; Fortuni, B.; Dao, A. T. N.; Masuhara, A.; Hirai, K.; Kasai, H.; Inose, T.; Uji-I, H. Low-Cytotoxic Gold-Coated Silver Nanoflowers for Intracellular pH Sensing. *ACS Appl. Nano Mater.* **2020**, *3* (8), 7643–7650. https://doi.org/10.1021/acsanm.0c01278.

(10) Zhao, X.; Campbell, S.; Wallace, G. Q.; Claing, A.; Bazuin, C. G.; Masson, J. F. Branched Au Nanoparticles on Nanofibers for Surface-Enhanced Raman Scattering Sensing of Intracellular pH and Extracellular pH Gradients. *ACS Sensors* **2020**, *5* (7), 2155–2167.

https://doi.org/10.1021/acssensors.0c00784.

(11) Bai, L.; Wang, X.; Zhang, K.; Tan, X.; Zhang, Y.; Xie, W. Etchable SERS Nanosensor for Accurate pH and Hydrogen Peroxide Sensing in Living Cells. *Chem. Commun.* **2019**, *55* (86), 12996–12999. https://doi.org/10.1039/c9cc06485k.

(12) Bi, Y.; Di, H.; Zeng, E.; Li, Q.; Li, W.; Yang, J.; Liu, D. Reliable Quantification of pH Variation in Live Cells Using Prussian Blue-Caged Surface-Enhanced Raman Scattering Probes. *Anal. Chem.* **2020**, *92* (14), 9574–9582. https://doi.org/10.1021/acs.analchem.0c00714.

(13) Wang, J.; Geng, Y.; Shen, Y.; Shi, W.; Xu, W.; Xu, S. SERS-Active Fiber Tip for Intracellular and Extracellular pH Sensing in Living Single Cells. *Sensors Actuators, B Chem.* **2019**, *290* (December 2018), 527–534. https://doi.org/10.1016/j.snb.2019.03.149.

(14) Zhang, Z.; Bando, K.; Mochizuki, K.; Taguchi, A.; Fujita, K.; Kawata, S. Quantitative Evaluation of Surface-Enhanced Raman Scattering Nanoparticles for Intracellular pH Sensing at a Single Particle Level. *Anal. Chem.* **2019**, *91* (5), 3254–3262. https://doi.org/10.1021/acs.analchem.8b03276.

(15) Nam, W.; Song, J.; Safiabadi Tali, S. A.; Lezec, H. J.; Agrawal, A.; Zhou, W. Au/SiO2-Nanolaminated Plasmonic Nanoantennas as Refractive-Index-Insensitive and Transparent Surface-Enhanced Raman Spectroscopy Substrates. *ACS Appl. Nano Mater.* **2021**. https://doi.org/10.1021/acsanm.1c00389.

(16) Langer, J.; de Aberasturi, D. J.; Aizpurua, J.; Alvarez-Puebla, R. A.; Auguié, 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.; Javier García de Abajo, F.; Goodacre, R.; Graham, D.; Haes, A. J.; Haynes, C. L.; Huck, C.; Itoh, T.; Käll, M.; Kneipp, J.; Kotov, N. A.; Kuang, H.; Le Ru, E. C.; Lee, H. K.; Li, J. F.; Ling, X. Y.; Maier, S. A.; Mayerhöfer, T.; Moskovits, M.; Murakoshi, K.; Nam, J. M.; Nie, S.; Ozaki, Y.; Pastoriza-Santos, I.; Perez-Juste, J.; Popp, J.; Pucci, A.; Reich, S.; Ren, B.; Schatz, G. C.; Shegai, T.; Schlücker, S.; Tay, L. L.; George Thomas, K.; Tian, Z. Q.; van Duyne, R. P.; Vo-Dinh, T.; Wang, Y.; Willets, K. A.; Xu, C.; Xu, H.; Xu, Y.; Yamamoto, Y. S.; Zhao, B.; Liz-Marzán, L. M. Present and Future of Surface-Enhanced Raman Scattering. *ACS Nano* **2020**, *14* (1), 28–117. https://doi.org/10.1021/acsnano.9b04224.

(17) Halvorson, R. A.; Vikesland, P. J. Surface-Enhanced Raman Spectroscopy (SERS) for Environmental Analyses. *Environ. Sci. Technol.* **2010**, *44* (20), 7749–7755. https://doi.org/10.1021/es101228z.

(18) Golightly, R. S.; Doering, W. E.; Natan, M. J. Surface-Enhanced Raman Spectroscopy and Homeland Security: A Perfect Match? *ACS Nano* **2009**, *3* (10), 2859–2869. https://doi.org/10.1021/nn9013593.

(19) Wang, Z.; Bonoiu, A.; Samoc, M.; Cui, Y.; Prasad, P. N. Biological pH Sensing Based on Surface Enhanced Raman Scattering through a 2-Aminothiophenol-Silver Probe. *Biosens. Bioelectron.* **2008**, *23* (6), 886–891. https://doi.org/10.1016/j.bios.2007.09.017.

(20) Capocefalo, A.; Mammucari, D.; Brasili, F.; Fasolato, C.; Bordi, F.; Postorino, P.; Domenici, F. Exploring the Potentiality of a SERS-Active pH Nano-Biosensor. *Front. Chem.* **2019**, *7* (June), 1–11. https://doi.org/10.3389/fchem.2019.00413.

(21) Jamieson, L. E.; Jaworska, A.; Jiang, J.; Baranska, M.; Harrison, D. J.; Campbell, C. J. Simultaneous Intracellular Redox Potential and pH Measurements in Live Cells Using SERS Nanosensors. *Analyst* **2015**, *140* (7), 2330–2335. https://doi.org/10.1039/c4an02365j.

(22) Gühlke, M.; Heiner, Z.; Kneipp, J. Combined Near-Infrared Excited SEHRS and SERS Spectra of pH Sensors Using Silver Nanostructures. *Phys. Chem. Chem. Phys.* **2015**, *17* (39), 26093–26100. https://doi.org/10.1039/c5cp03844h.

(23) Wei, H.; Willner, M. R.; Marr, L. C.; Vikesland, P. J. Highly Stable SERS pH Nanoprobes Produced by Co-Solvent Controlled AuNP Aggregation. *Analyst* **2016**, *141* (17), 5159–5169. https://doi.org/10.1039/c6an00650g.

(24) Pallaoro, A.; Braun, G. B.; Reich, N. O.; Moskovits, M. Mapping Local pH in Live Cells Using Encapsulated Fluorescent SERS Nanotags. *Small* **2010**, *6* (5), 618–622. https://doi.org/10.1002/smll.200901893.

(25) Wang, F.; Widejko, R. G.; Yang, Z.; Nguyen, K. T.; Chen, H.; Fernando, L. P.; Christensen, K. A.; Anker, J. N. Surface-Enhanced Raman Scattering Detection of pH with Silica-Encapsulated 4-Mercaptobenzoic Acid-Functionalized Silver Nanoparticles. *Anal. Chem.* **2012**, *84* (18), 8013–8019. https://doi.org/10.1021/ac3018179.

(26) Xu, M.; Ma, X.; Wei, T.; Lu, Z. X.; Ren, B. In Situ Imaging of Live-Cell Extracellular pH during Cell Apoptosis with Surface-Enhanced Raman Spectroscopy. *Anal. Chem.* **2018**, *90* (23), 13922–13928. https://doi.org/10.1021/acs.analchem.8b03193.

(27) Yu, H. Z.; Xia, N.; Liu, Z. F. SERS Titration of 4-Mercaptopyridine Self-Assembled Monolayers at Aqueous Buffer/Gold Interfaces. *Anal. Chem.* **1999**, *71* (7), 1354–1358. https://doi.org/10.1021/ac981131+.

(28) Guo, H.; Huang, Q.; Leng, W.; Zhan, Y.; Behkam, B.; Willner, M. R.; Wei, H.; Marr, L. C.; Vikesland, P. J. Bromide Ion-Functionalized Nanoprobes for Sensitive and Reliable pH Measurement by Surface-Enhanced Raman Spectroscopy. *Analyst* **2019**, *144* (24), 7326–7335. https://doi.org/10.1039/c9an01699f.

(29) Jensen, R. A.; Sherin, J.; Emory, S. R. Single Nanoparticle Based Optical pH Probe. *Appl. Spectrosc.* **2007**, *61* (8), 832–838. https://doi.org/10.1366/000370207781540105.

(30) Shen, Y.; Liang, L.; Zhang, S.; Huang, D.; Zhang, J.; Xu, S.; Liang, C.; Xu, W. Organelle-Targeting Surface-Enhanced Raman Scattering (SERS) Nanosensors for Subcellular pH Sensing. *Nanoscale* **2018**, *10* (4), 1622–1630. https://doi.org/10.1039/c7nr08636a.

(31) Zhu, G.; Cheng, L.; Liu, G.; Zhu, L. Synthesis of Gold Nanoparticle Stabilized on Silicon Nanocrystal Containing Polymer Microspheres as Effective Surface-Enhanced Raman Scattering (SERS) Substrates. *Nanomaterials* **2020**, *10* (8), 1–9. https://doi.org/10.3390/nano10081501.

(32) Wei, H.; Vejerano, E. P.; Leng, W.; Huang, Q.; Willner, M. R.; Marr, L. C.; Vikesland, P. J. Aerosol Microdroplets Exhibit a Stable pH Gradient. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (28), 7272–7277. https://doi.org/10.1073/pnas.1720488115.

(33) Zhang, Y.; de Aberasturi, D. J.; Henriksen-Lacey, M.; Langer, J.; Liz-Marzan, L. M. Live-Cell Surface-Enhanced Raman Spectroscopy Imaging of Intracellular pH: From Two

Dimensions to Three Dimensions. *ACS Sensors* **2020**, *5* (10), 3194–3206. https://doi.org/10.1021/acssensors.0c01487.

(34) Huang, Q.; Wei, H.; Marr, L. C.; Vikesland, P. J. Direct Quantification of the Effect of Ammonium on Aerosol Droplet pH. *Environ. Sci. Technol.* **2020**. https://doi.org/10.1021/acs.est.0c07394.

(35) Talley, C. E.; Jusinski, L.; Hollars, C. W.; Lane, S. M.; Huser, T. Intracellular pH Sensors Based on Surface-Enhanced Raman Scattering. *Anal. Chem.* **2005**, *76* (23), 7064–7068. https://doi.org/10.1021/ac049093j.

(36) Boisselier, E.; Astruc, D. Gold Nanoparticles in Nanomedicine: Preparations, Imaging, Diagnostics, Therapies and Toxicity. *Chem. Soc. Rev.* **2009**, *38* (6), 1759–1782. https://doi.org/10.1039/b806051g.

(37) Lu, L.; Ai, K.; Ozaki, Y. Environmentally Friendly Synthesis of Highly Monodisperse Biocompatible Gold Nanoparticles with Urchin-like Shape. *Langmuir* **2008**, *24* (3), 1058–1063. https://doi.org/10.1021/la702886q.

(38) Kokkin, D. L.; Zhang, R.; Steimle, T. C.; Wyse, I. A.; Pearlman, B. W.; Varberg, T. D. Au-S Bonding Revealed from the Characterization of Diatomic Gold Sulfide, AuS. *J. Phys. Chem. A* **2015**, *119* (48), 11659–11667. https://doi.org/10.1021/acs.jpca.5b08781.

(39) Wei, H.; Willner, M. R.; Marr, L. C.; Vikesland, P. J. Highly Stable SERS pH Nanoprobes Produced by Co-Solvent Controlled AuNP Aggregation. *Analyst* **2016**, *141* (17), 5159–5169. https://doi.org/10.1039/c6an00650g.

(40) Cai, W. Bin; She, C. X.; Ren, B.; Yao, J. L.; Tian, Z. W.; Tian, Z. Q. Surface Raman Spectroscopic Investigation of Pyridine Adsorption at Platinum Electrodes - Effects of Potential and Electrolyte. *J. Chem. Soc. - Faraday Trans.* **1998**, *94* (20), 3127–3133. https://doi.org/10.1039/a803240h.

(41) Jung, H. S.; Kim, K.; Kim, M. S. Raman Spectroscopic Investigation of the Adsorption of 4-Mercaptopyridine on a Silver-Sol Surface. *J. Mol. Struct.* **1997**, *407* (2–3), 139–147. https://doi.org/10.1016/S0022-2860(97)00006-9.

(42) Sun, S. C.; Bernard, I.; Birke, R. L.; Lombardi, J. R. The Effect of pH, Chloride Ion and Background Electrolyte Concentration on the SERS of Acidified Pyridine Solutions. *J. Electroanal. Chem.* **1985**, *196* (2), 359–374. https://doi.org/10.1016/0022-0728(85)80033-4.

(43) Chang, H.; Hwang, K. C. The Behavior of Pyridine, Pyridinium Ion, and Pyridinium Halide on a Ag Electrode and Their SERS Spectra. *J. Am. Chem. Soc.* **1984**, *106* (22), 6586–6592. https://doi.org/10.1021/ja00334a023.

(44) Saito, H. Behaviour of the 1025 cm^{-1} Band of SERS Spectra Due to Pyridine on a Silver Electrode in Aqueous Halide Media. *J. Raman Spectrosc.* **1993**, *24* (4), 191–197. https://doi.org/10.1002/jrs.1250240403.

(45) Sun, F.; Zhang, P.; Bai, T.; David Galvan, D.; Hung, H. C.; Zhou, N.; Jiang, S.; Yu, Q. Functionalized Plasmonic Nanostructure Arrays for Direct and Accurate Mapping Extracellular pH of Living Cells in Complex Media Using SERS. *Biosens. Bioelectron.* **2015**, *73*, 202–207. https://doi.org/10.1016/j.bios.2015.05.060.

(46) Song, J.; Nam, W.; Zhou, W. Scalable High-Performance Nanolaminated SERS Substrates Based on Multistack Vertically Oriented Plasmonic Nanogaps. *Adv. Mater. Technol.* **2019**, *4* (5), 1–7. https://doi.org/10.1002/admt.201800689.

(47) Willets, K. A. Surface-Enhanced Raman Scattering (SERS) for Probing Internal Cellular Structure and Dynamics. *Anal. Bioanal. Chem.* **2009**, *394* (1), 85–94. https://doi.org/10.1007/s00216-009-2682-3.

(48) Nam, W.; Ren, X.; Tali, S. A. S.; Ghassemi, P.; Kim, I.; Agah, M.; Zhou, W. Refractive-Index-Insensitive Nanolaminated SERS Substrates for Label-Free Raman Profiling and Classification of Living Cancer Cells. *Nano Lett.* **2019**, *19* (10), 7273–7281. https://doi.org/10.1021/acs.nanolett.9b02864.

(49) Lussier, F.; Missirlis, D.; Spatz, J. P.; Masson, J. F. Machine-Learning-Driven Surface-Enhanced Raman Scattering Optophysiology Reveals Multiplexed Metabolite Gradients Near Cells. *ACS Nano* **2019**. https://doi.org/10.1021/acsnano.8b07024.

(50) Banaei, N.; Moshfegh, J.; Mohseni-Kabir, A.; Houghton, J. M.; Sun, Y.; Kim, B. Machine Learning Algorithms Enhance the Specificity of Cancer Biomarker Detection Using SERS-Based Immunoassays in Microfluidic Chips. *RSC Adv.* **2019**, *9* (4), 1859–1868. https://doi.org/10.1039/C8RA08930B.

(51) Alexopoulos EC. Introduction to Multivariate Regression Analysis. *Hippokratia* **2005**, *14* (1), 23–28.

(52) Lundegard, P. D.; Land, L. S. Carbonate Equilibria and pH Buffering by Organic Acids - Response to Changes in PCO2. *Chem. Geol.* **1989**, *74* (3–4), 277–287. https://doi.org/10.1016/0009-2541(89)90038-7.

(53) Ren, X.; Nam, W.; Ghassemi, P.; Strobl, J. S.; Kim, I.; Zhou, W.; Agah, M. Scalable Nanolaminated SERS Multiwell Cell Culture Assay. *Microsystems Nanoeng.* **2020**, *6* (1). https://doi.org/10.1038/s41378-020-0145-3.

(54) Qin, D.; Xia, Y.; Whitesides, G. M. Soft Lithography for Micro- and Nanoscale Patterning. *Nat. Protoc.* **2010**, *5* (3), 491–502. https://doi.org/10.1038/nprot.2009.234.

(55) Niidome, T.; Yamagata, M.; Okamoto, Y.; Akiyama, Y.; Takahashi, H.; Kawano, T.; Katayama, Y.; Niidome, Y. PEG-Modified Gold Nanorods with a Stealth Character for in Vivo Applications. *J. Control. Release* **2006**, *114* (3), 343–347. https://doi.org/10.1016/j.jconrel.2006.06.017.

(56) Fernández-López, C.; Mateo-Mateo, C.; Álvarez-Puebla, R. A.; Pérez-Juste, J.; Pastoriza-Santos, I.; Liz-Marzán, L. M. Highly Controlled Silica Coating of PEG-Capped Metal Nanoparticles and Preparation of SERS-Encoded Particles. *Langmuir* **2009**, *25* (24), 13894–13899. https://doi.org/10.1021/la9016454.

(57) Liu, C.; Leng, W.; Vikesland, P. J. Controlled Evaluation of the Impacts of Surface Coatings on Silver Nanoparticle Dissolution Rates. *Environ. Sci. Technol.* **2018**, *52* (5), 2726–2734. https://doi.org/10.1021/acs.est.7b05622.

(58) Wei, H.; Leng, W.; Song, J.; Willner, M. R.; Marr, L. C.; Zhou, W.; Vikesland, P. J. Improved Quantitative SERS Enabled by Surface Plasmon Enhanced Elastic Light Scattering. *Anal. Chem.* **2018**, *90* (5), 3227–3237.

https://doi.org/10.1021/acs.analchem.7b04667.

(59) Nam, W.; Ren, X.; Kim, I.; Strobl, J.; Agah, M.; Zhou, W. Plasmonically Calibrated Label-Free Surface-Enhanced Raman Spectroscopy for Improved Multivariate Analysis of Living Cells in Cancer Subtyping and Drug Testing. *Anal. Chem.* **2021**. https://doi.org/10.1021/acs.analchem.0c05206.

(60) Nam, W.; Zhao, Y.; Song, J.; Ali Safiabadi Tali, S.; Kang, S.; Zhu, W.; Lezec, H. J.; Agrawal, A.; Vikesland, P. J.; Zhou, W. Plasmonic Electronic Raman Scattering as Internal Standard for Spatial and Temporal Calibration in Quantitative Surface-Enhanced Raman Spectroscopy. *J. Phys. Chem. Lett.* **2020**, *11* (22), 9543–9551. https://doi.org/10.1021/acs.jpclett.0c03056.

(61) Lee, T.; Wi, J. S.; Oh, A.; Na, H. K.; Lee, J.; Lee, K.; Lee, T. G.; Haam, S. Highly Robust, Uniform and Ultra-Sensitive Surface-Enhanced Raman Scattering Substrates for MicroRNA Detection Fabricated by Using Silver Nanostructures Grown in Gold Nanobowls. *Nanoscale* **2018**, *10* (8), 3680–3687. https://doi.org/10.1039/c7nr08066b.

(62) Xie, J.; Zhang, Q.; Lee, J. Y.; Wang, D. I. C. The Synthesis of SERS-Active Gold Nanoflower Tags for in Vivo Applications. *ACS Nano* **2008**, *2* (12), 2473–2480. https://doi.org/10.1021/nn800442q.

(63) Bi, L.; Wang, Y.; Yang, Y.; Li, Y.; Mo, S.; Zheng, Q.; Chen, L. Highly Sensitive and Reproducible SERS Sensor for Biological pH Detection Based on a Uniform Gold Nanorod Array Platform. *ACS Appl. Mater. Interfaces* **2018**, *10* (18), 15381–15387. https://doi.org/10.1021/acsami.7b19347.

(64) Williams, A.; Flynn, K. J.; Xia, Z.; Dunstan, P. R. Multivariate Spectral Analysis of pH SERS Probes for Improved Sensing Capabilities. *J. Raman Spectrosc.* **2016**, *47* (7), 819–827. https://doi.org/10.1002/jrs.4910.

(65) Guo, H.; Hamlet, L. C.; He, L.; Xing, B. A Field-Deployable Surface-Enhanced Raman Scattering (SERS) Method for Sensitive Analysis of Silver Nanoparticles in Environmental Waters. *Sci. Total Environ.* **2019**, *653*, 1034–1041. https://doi.org/10.1016/j.scitotenv.2018.10.435.

(66) Chen, J.; Qin, G.; Shen, W.; Li, Y.; Das, B. Fabrication of Long-Range Ordered, Broccoli-like SERS Arrays and Application in Detecting Endocrine Disrupting Chemicals. *J. Mater. Chem. C* **2015**, *3* (6), 1309–1318. https://doi.org/10.1039/c4tc02224f.

(67) Zhang, K.; Wang, Y.; Wu, M.; Liu, Y.; Shi, D.; Liu, B. On-Demand Quantitative SERS Bioassays Facilitated by Surface-Tethered Ratiometric Probes. *Chem. Sci.* **2018**, *9* (42), 8089–8093. https://doi.org/10.1039/c8sc03263g.

(68) Peng, R.; Si, Y.; Deng, T.; Zheng, J.; Li, J.; Yang, R.; Tan, W. A Novel SERS Nanoprobe for the Ratiometric Imaging of Hydrogen Peroxide in Living Cells. *Chem. Commun.* **2016**, *52* (55), 8553–8556. https://doi.org/10.1039/c6cc03412h.

(69) Wang, W.; Zhang, L.; Li, L.; Tian, Y. A Single Nanoprobe for Ratiometric Imaging and Biosensing of Hypochlorite and Glutathione in Live Cells Using Surface-Enhanced Raman Scattering. *Anal. Chem.* **2016**, *88* (19), 9518–9523. https://doi.org/10.1021/acs.analchem.6b02081.

(70) Ryoo, D.; Xu, X.; Li, Y.; Tang, J. A.; Zhang, J.; Van Zijl, P. C. M.; Liu, G. Detection and Quantification of Hydrogen Peroxide in Aqueous Solutions Using Chemical Exchange Saturation Transfer. *Anal. Chem.* **2017**, *89* (14), 7758–7764. https://doi.org/10.1021/acs.analchem.7b01763.

(71) Gu, X.; Wang, H.; Schultz, Z. D.; Camden, J. P. Sensing Glucose in Urine and Serum and Hydrogen Peroxide in Living Cells by Use of a Novel Boronate Nanoprobe Based on Surface-Enhanced Raman Spectroscopy. *Anal. Chem.* **2016**, *88* (14), 7191–7197. https://doi.org/10.1021/acs.analchem.6b01378.

(72) Dong, W.; Ren, Y.; Bai, Z.; Yang, Y.; Chen, Q. Fabrication of Hexahedral Au-Pd/Graphene Nanocomposites Biosensor and Its Application in Cancer Cell H 2 O 2 Detection. *Bioelectrochemistry* **2019**, *128*, 274–282. https://doi.org/10.1016/j.bioelechem.2019.04.018.