

DeltaPCA: A Statistically Robust Method for Detecting Protein Analyte Binding to Aptamer-Functionalised Nanoparticles using Surface-Enhanced Raman Spectroscopy

Fiona M. Given,^{*,†} Tamsyn Stanborough,^{*,†} Mark R. Waterland,^{*,‡} and
Deborah L. Crittenden^{*,†}

*[†]Biomolecular Interaction Centre and School of Physical and Chemical Sciences,
University of Canterbury, Christchurch, New Zealand*

*[‡]MacDiarmid Institute for Advanced Materials and Nanotechnology, Massey University,
Palmerston North, New Zealand*

E-mail: fiona.given@canterbury.ac.nz; tamsyn.stanborough@canterbury.ac.nz;

M.Waterland@massey.ac.nz; deborah.crittenden@canterbury.ac.nz

Abstract

In this work, we introduce a novel joint experimental design and computational analysis procedure to reliably and reproducibly quantify protein analyte binding to DNA aptamer-functionalised silver nanoparticles using slippery surface-enhanced Raman spectroscopy. We employ an indirect detection approach, based upon monitoring spectral changes in the covalent bond-stretching region as intermolecular bonds are formed between the surface-immobilized probe biomolecule and its target analyte. Sample variability is minimized by preparing aptamer-only and aptamer-plus-analyte samples under the same conditions, and then analysing difference spectra. To account for technical variability, multiple spectra are recorded from the same sample. Our new DeltaPCA analysis procedure takes into account technical variability within each spectral data set while also extracting statistically robust difference spectra between data sets. Proof of principle experiments using thiolated aptamers to detect CoV-SARS-2 spike protein reveal that analyte binding is mediated through the formation of $\text{N-H}\cdots\text{X}$ and $\text{C-H}\cdots\text{X}$ hydrogen bonds between the aptamer (H-bond donor) and protein (H-bond acceptor). Our computational analysis code can be freely downloaded from <https://github.com/dlc62/DeltaPCA>.

Keywords: Principal Component Analysis, Aptamer, Differential Analysis, Technical Variability, Sample Variability

Introduction

Surface-enhanced Raman spectroscopy is a promising technique for direct detection of environmental contaminants and/or biomolecular disease markers at extremely low concentrations.^{1–5} Advances in surface and substrate engineering have enabled selected analytes to be detected down to femtomolar or even attomolar concentrations, approaching the single-molecule detection limit.^{6–8}

However, the exquisite sensitivity and high spatial resolution of surface-engineered SERS often comes at the cost of reproducibility; substantial fluctuations in a SERS spectrum can be observed as molecules drift into and out of electric-field hot spots, and SERS spectra are very sensitive to differences in sample preparation, optical alignment and variations in laser power.^{9–12} In this work, we present a combined experimental design, data acquisition and data analysis strategy to overcome these limitations.

Our approach is based upon the SLIPSERS method,⁶ in which a microdroplets of citrate-capped silver nanoparticles - with and without the analyte of interest - are deposited on an omniphobic substrate and the solvent is evaporated off. Isotropic contraction of the near-spherical droplet leads to the formation of compact nanoparticle:analyte aggregates in which the local concentration of the analyte is enhanced in a reproducible and concentration-dependent manner. The main advantage of SLIPSERS over other surface-enhanced Raman methods is that it dramatically improves sensitivity and decreases detection limits, without requiring sophisticated substrate engineering.^{4,6}

However, questions of analyte specificity remain, particularly in the context of analysing complex chemical and biochemical mixtures. This can be overcome by nanoparticle surface functionalisation which selectively bind of the analyte of interest and enhance its local surface concentration. For example, SERS immunoassays take advantage of selective interactions between biological antigens and their associated antibodies.^{5,13} Upon binding of the target biomolecule to the surface-immobilized probe, spectral changes are observed in the so-called “fingerprint region” ($500 - 2000 \text{ cm}^{-1}$) that most clearly reflects differences in Raman spectra between biomolecules.¹⁴ Unfortunately, with experimental complexity – the presence of two large and complex biomolecules – comes spectral complexity, and

these assays have not proven reproducible enough for commercial development.^{5,13}

A key insight of the present work is that sample variability can be eliminated, or at least controlled, using an indirect detection approach; monitoring changes in the Raman spectrum of a surface-immobilized probe molecule upon analyte binding. Technical variability, on the other hand, can be quantified by taking multiple repeated measurements on the same sample. By combining an indirect detection experimental strategy with differential component analysis of the resultant spectra, sample and technical variability can be accounted for on the same footing, to ensure that observed spectral differences due to analyte binding are statistically significant and reproducible, *i.e.* are not technical artefacts.

As an illustrative example, we consider SARS-CoV-2 spike protein binding to thiolated aptamers immobilized on silver nanoparticles. Full experimental details are available in the literature,¹⁵ and the salient points summarised briefly below.

Methods

Experimental

Citrate-capped silver nanoparticles were synthesised following the approach of Kitaev *et al.*,^{16,17} using a borohydride reducing agent and bromide ions for faceting and shape control. 100 μ M stock solutions of 5'-thiolated DNA aptamer:¹⁸ and a sequence-scrambled analogue (Table 1) were prepared in 10 mM TCEP (pH 7) and incubated for 2 hours at room temperature, and then diluted to final concentrations of 1 nM, 500 pM and 1 pM in 50 mM Tris/HCl buffer (pH 8), with and without equimolar quantities of the associated Spike protein analyte.¹⁵ After 15 minutes, 10 μ L of these solutions were mixed with 50 μ L of colloidal Ag nanoparticle solution and mixed thoroughly. A blank containing 10 μ L of Tris buffer was also prepared. Each 60 μ L mixture was pipetted onto a polyfluoropolyether oil coated Teflon filters mounted on a glass slide, prepared according to the method of Yang *et al.*⁶ Following evaporation of the excess solvent at 65 °C, Raman spectra of the resultant aggregates were recorded under ambient conditions using

a custom-built Raman microscope.¹⁵ The detector exposure time was 1 second and 60 exposures (data frames) were captured and stored separately prior to data analysis. No background removal was applied during data collection.

Table 1: Aptamer sequences used in this work

Label	Sequence
A	5'-dithiol-CAGCACCGACCTTGTGCTTTGGGAGTGCTGGTCCAAGGGCGTTAATGGACA-3'
SA	5'-dithiol-AACGCGGAGCCATTGGTAAGGTGCGTCCGTCCTCAGTATCTAAGCTGTGGG-3'

Data Analysis

For each set of 60 data frames, uncentred principal component analysis is applied to obtain standardized scores and loadings, via singular values decomposition:

$$\mathbf{X} = \mathbf{USV}^T \quad (1)$$

$$\mathbf{Z} = \mathbf{U}\sqrt{n} \quad (2)$$

$$\mathbf{W} = \frac{\mathbf{SV}}{\sqrt{n}} \quad (3)$$

where \mathbf{X} is an $n \times p$ data array where n = number of data frames and p = number of intensity values recorded per frame, columns of \mathbf{US} and \mathbf{V} represent principal component scores and axes, respectively, and columns of \mathbf{Z} and \mathbf{W} correspond to standardized scores and associated loadings. The first column of \mathbf{W} , w_1 , is the principal loading vector that captures the similarity across each data set, and can be loosely interpreted as its average or central spectrum. The remaining columns account for variability between frames.

Baseline corrections are performed on each principal loading vector using Korepanov's *derpsalsa* algorithm,¹⁹ and subtracted from each data frame. The baseline-corrected central spectrum is also computed and stored.

To analyse variability between data sets (*e.g.* for samples with and without protein analyte), the baseline-corrected data sets are concatenated, and the central spectrum for the reference sample subtracted from each frame. The resultant set of 120 data frames

is then subjected to PCA analysis as described above. In this case, the first 60 frames contain only noise (by construction), while the last 60 frames contain noise plus signal from the difference in experimental conditions. The principal loading vector captures this signal, and can be interpreted as the difference spectrum.

These algorithms have been implemented in a python script that is freely available from <https://github.com/dlc62/DeltaPCA>. Our code allows multiple pairwise comparisons to be performed and plotted simultaneously, or multiple samples to be compared against a common reference. It also allows different spectral regions to be analysed independently, which may be useful if there are strong background effects in some spectral regions but not others. The λ value that controls the stringency of the *derpsalsa* baseline algorithm can be specified via the command line.

Results and Discussion

The first step in our analysis process is to compute a central spectrum for each set of 60 technical replicates using principal component analysis, perform baselining on this reference spectrum, and subtract this common baseline from all 60 data frames. The results of this procedure are illustrated in Figure 1. The grey lines represent the baselined raw data while the coloured lines represent the primary principal component loading vectors. Visual inspection reveals that the aptamer-only spectra are quite similar to one another but qualitatively different from the aptamer + protein spectra, which display increased intensities in the fingerprint region ($500 - 2000 \text{ cm}^{-1}$) and a change in both intensity and spectral shape in the covalent bond stretching region ($2500 - 4000 \text{ cm}^{-1}$).

The next phase of our analysis procedure involves computing difference spectra by concatenating the baselined aptamer-only and aptamer-plus-protein spectral data sets, subtracting the aptamer-only reference spectrum, and then performing principal component analysis on the resultant data set and extracting the primary principal component loading vectors. The results shown in Figure 2(a) confirm a general increase in spectral intensity across most of the fingerprint region ($900 - 2000 \text{ cm}^{-1}$) but not in a

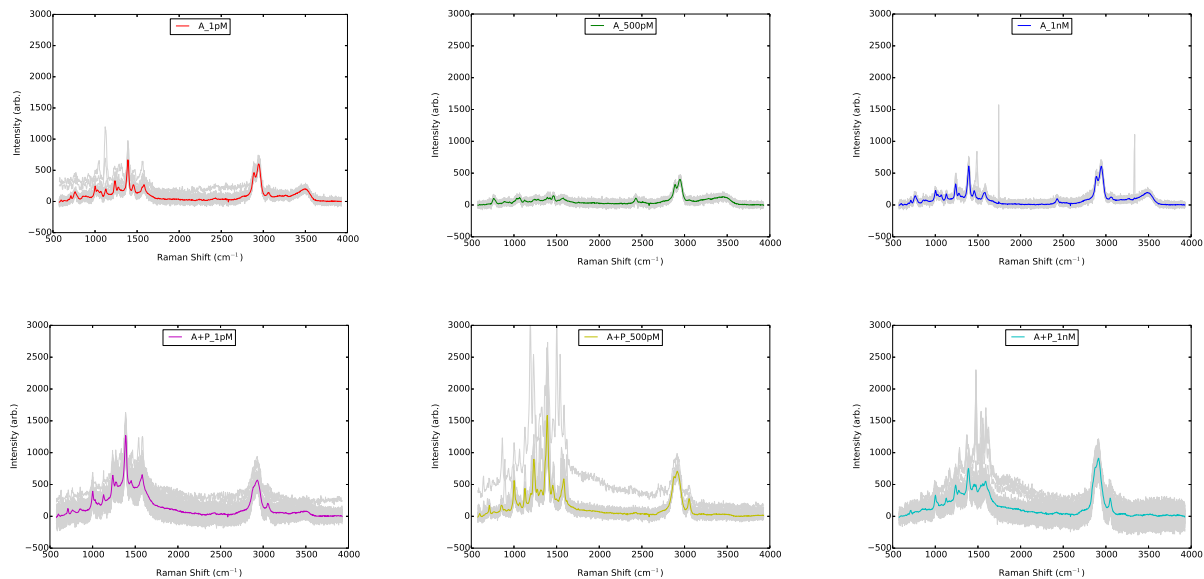


Figure 1: Primary principal component loadings (coloured lines) superposed on raw spectral data (grey lines) from which they are derived for aptamer (A) and aptamer plus protein (A+P) samples at a series of different concentrations (1 pM, 500 pM, 1 nM).

particularly reproducible or concentration-dependent manner. The clearest - and most reproducible - determinants of analyte binding show up as *depletions* of the free C-H stretching band at 2956 cm^{-1} , the free N-H stretching band centred around 3480 cm^{-1} , and a nucleobase ring-deformation band at 763 cm^{-1} . Localised depletion of another C-H stretching band at 2889 cm^{-1} is also evident from these difference spectra. The strongest concentration-dependent response is observed in the appearance of new hydrogen-bonded C-H stretching bands at 2870 and 2912 cm^{-1} , giving rise to the calibration curves illustrated in Figure 2(b). A new hydrogen-bonded N-H stretching band also appears at 3050 cm^{-1} . These spectral changes indicate concentration-dependent binding of spike protein to surface-bound aptamer via $\text{N-H}\cdots\text{X}$ and $\text{C-H}\cdots\text{X}$ hydrogen bonding between exposed nucleobases and the target analyte.

To check that the spectral shifts we observe are due to specific binding, we performed the same experimental and analysis procedure using a scrambled aptamer, with and without the spike protein analyte. Results for the 1 pM sample are illustrated in Figure 3. It is immediately clear that the scrambled aptamer behaves very differently to the original. Most significantly, there is no longer evidence for the formation of specific intermolecular interactions between the aptamer and analyte. The characteristic spectral

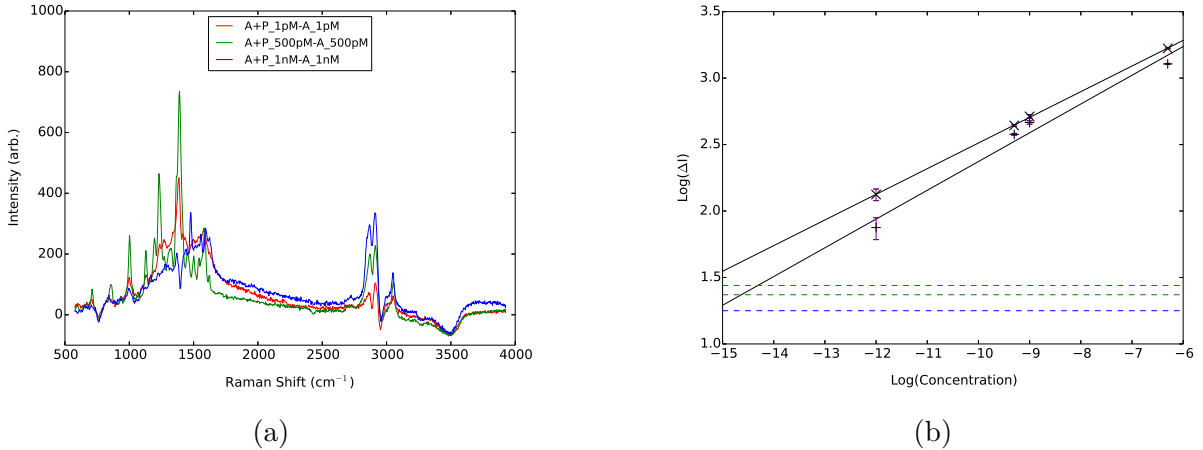


Figure 2: (a) Difference spectra computed as the primary principal component loadings that capture the majority of the variability between aptamer-only (A) and aptamer plus protein (A+P) SLIPSERS spectra at a range of different concentrations (1 pM, 500 pM, 1 nM) and (b) Log-log calibration curves^{6,20} obtained from differences in SERS intensities at 2872 cm^{-1} (x) and 2912 cm^{-1} (+) between silver nanoparticles treated with aptamer alone (A) at a given concentration vs those treated with aptamer plus protein (A+P) at the same concentration. Error bars show the standard error of the mean for the difference in intensity due to technical variability. Detection limits are obtained by analysis of baseline scatter in the 3650 - 4000 cm^{-1} region. Dotted horizontal lines indicate 99%, 99.9% and 99.99% one-sided confidence intervals.

shifts - and particularly covalent stretching regions in which the intensity is *depleted* upon analyte binding - are absent. This is consistent with the fact that the scrambled aptamer does not display a free N-H stretching band within its Raman spectrum, so it cannot form hydrogen bonds to the analyte. There is also no significant enhancement in the fingerprint region, which rules out the possibility of substantial non-specific binding.

Instead, the main difference between aptamer-only and aptamer-plus-protein spectra is the accentuation of two free C-H stretching bands at 2892 and 2946 cm^{-1} . There are two possible explanations for this observation: experimental variability, or the formation of a weak and distant association complex that influences the dynamics of the aptamer and resolves the C-H environments but does not otherwise affect its SERS spectrum. Either way, however, it is clear that the strong and reproducible spectral changes observed using the original aptamer are unique and specific to that aptamer.

This example illustrates a number of key advantages of our approach:

1. Spectral shifts due to analyte binding are easy to observe, particularly in regions where there is concurrent depletion of some bands and appearance of others

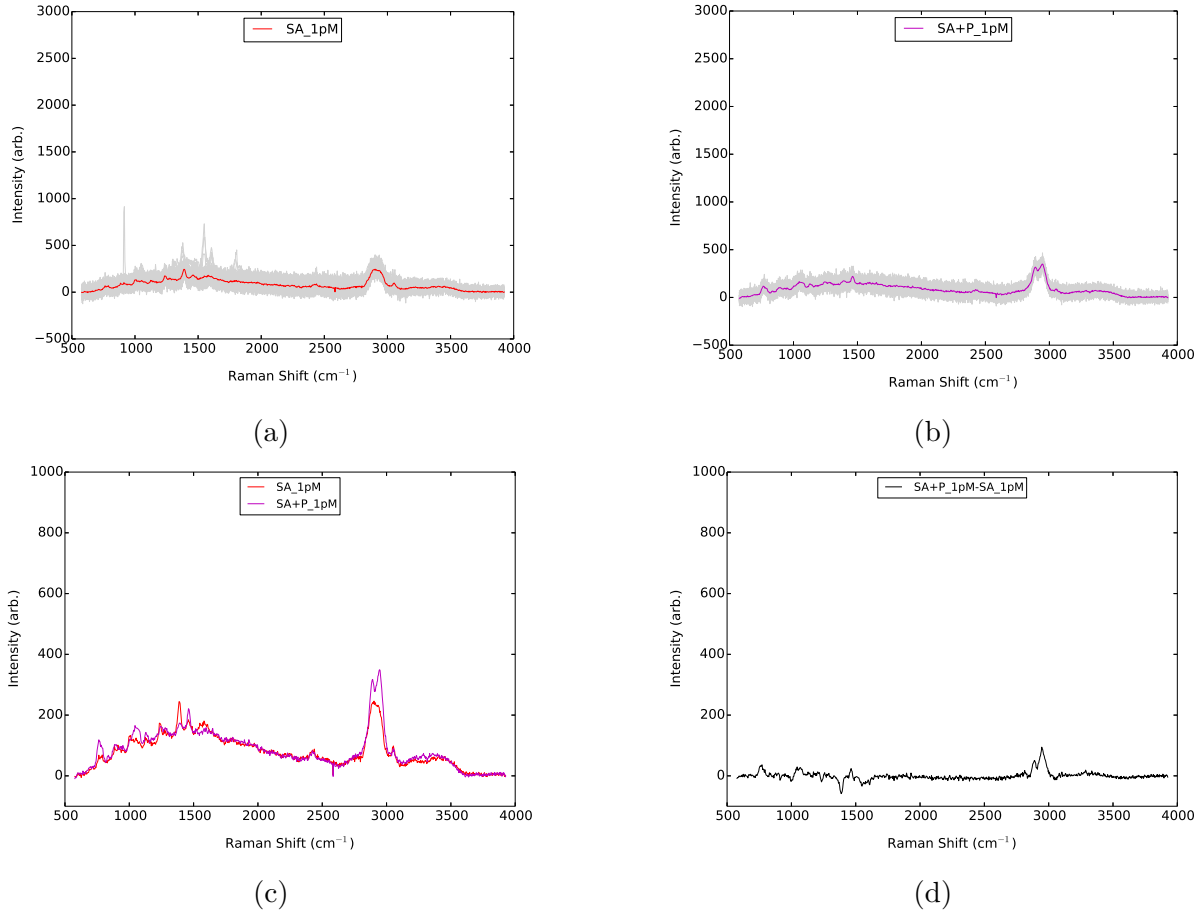


Figure 3: Primary principal component loadings (coloured lines) superposed on raw spectral data (grey lines) from which they are derived for (a) scrambled aptamer (SA) and (b) scrambled aptamer plus protein (SA+P) samples, along with (c) the superimposed principal component loadings from each sample and (d) the secondary principal component loading that accounts for the majority of the variability between the samples.

2. Spectral difference profiles are reproducible in shape, making them more statistically robust than peak-picking or monitoring changes in intensity at a single wavelength
3. Spectral changes that arise from specific interactions between the surface-immobilized probe and analyte can be readily separated from those due to non-specific adsorption
4. Observed spectral changes have clear, physically-interpretable meaning that yield molecular level information about bonding between the probe and analyte
5. Quantitative calibration curves can be obtained by monitoring changes in intensity

at specific wavelengths once robust spectral shift patterns have been identified.

Although principal component analysis has been widely used in analysing Raman spectra, it is usually applied to “blindly” categorise samples,^{21–23} although the importance of separating out technical and sample variability has recently been recognised.^{24,25} However, to the best of our knowledge, this work represents the first example of using differential principal component analysis in combination with a “probe recognition” experimental strategy to separate out technical and sample variability by design. We also note that our combined experimental and theoretical approach is fundamentally different to most Raman studies that monitor and analyse spectral changes in the fingerprint region, generally arising from direct adsorption of the analyte.¹² However, even in cases where substrate-specific probes have been used to attract and/or attach analytes more specifically, analysis has remained restricted to the fingerprint region and typically focusses on quite narrow spectral bands (peak-picking).¹³ Our results go some way to explaining why this approach may not be particularly robust.

Finally, we note that our differential PCA analysis approach is not solely restricted to the current use case. It could also be applied to quantify spectral changes due to direct analyte adsorption against an appropriate negative (no analyte/carrier medium only) control, and to establish whether observed changes are statistically significant or within the bounds of technical variability. Concurrent analysis of samples at different concentrations against a common reference spectrum could be performed to reveal the common spectral fingerprint whose intensity varies as a function of concentration. We look forward to seeing our analysis code applied widely to facilitate statistically robust generation and analysis of surface-enhanced Raman spectra.

Author Information

Corresponding author: Deborah Crittenden

Email: deborah.crittenden@canterbury.ac.nz

Conflict of Interest

The authors have no conflicts of interest to declare.

References

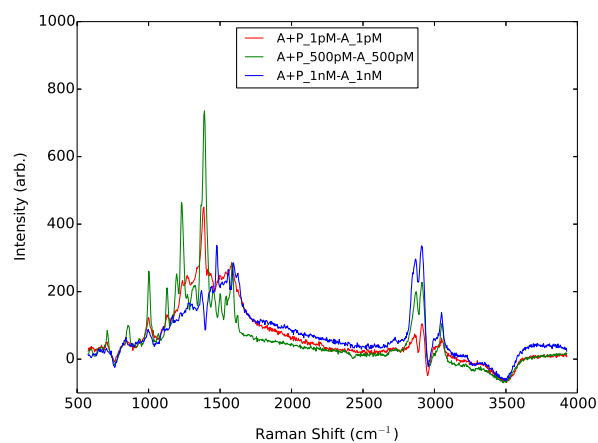
- (1) Kim, K.; Kashefi-Kheyraadi, L.; Joung, Y.; Kim, K.; Dang, H.; Chavan, S. G.; Lee, M.-H.; Choo, J. Recent advances in sensitive surface-enhanced Raman scattering-based lateral flow assay platforms for point-of-care diagnostics of infectious diseases. *Sens. Act. B* **2020**, 129214.
- (2) Saviñon-Flores, F.; Méndez, E.; López-Castaños, M.; Carabarin-Lima, A.; López-Castaños, K. A.; González-Fuentes, M. A.; Méndez-Albores, A. A Review on SERS-Based Detection of Human Virus Infections: Influenza and Coronavirus. *Biosens.* **2021**, *11*, 66.
- (3) Yadav, S.; Sadique, M. A.; Ranjan, P.; Kumar, N.; Singhal, A.; Srivastava, A. K.; Khan, R. SERS Based Lateral Flow Immunoassay for Point-of-Care Detection of SARS-CoV-2 in Clinical Samples. *ACS Appl. Bio Mater.* **2021**,
- (4) Mehta, M.; Waterland, M. Ultrasensitive surface-enhanced Raman scattering detection of biological pollutants by controlled evaporation on omniphobic substrates. *Heliyon* **2020**, *6*, e04317.
- (5) Porter, M. D.; Lipert, R. J.; Siperko, L. M.; Wang, G.; Narayanan, R. SERS as a bioassay platform: fundamentals, design, and applications. *Chem. Soc. Rev.* **2008**, *37*, 1001–1011.
- (6) Yang, S.; Dai, X.; Stogin, B. B.; Wong, T.-S. Ultrasensitive surface-enhanced Raman scattering detection in common fluids. *Proc. Nat'l Acad. Sci.* **2016**, *113*, 268–273.
- (7) De Angelis, F.; Gentile, F.; Mecarini, F.; Das, G.; Moretti, M.; Candeloro, P.; Coluccio, M. L.; Cojoc, G.; Accardo, A.; Liberale, C.; Zaccaria, R. P.; Perozziello, G.; Tirinato, L.; Toma, A.; Cuda, G.; Cingolani, R.; Di Fabrizio, E. Breaking the diffusion limit with super-hydrophobic delivery of molecules to plasmonic nanofocusing SERS structures. *Nature Phot.* **2011**, *5*, 682–687.
- (8) Lee, H. K.; Lee, Y. H.; Koh, C. S. L.; Phan-Quang, G. C.; Han, X.; Lay, C. L.;

- Sim, H. Y. F.; Kao, Y.-C.; An, Q.; Ling, X. Y. Designing surface-enhanced Raman scattering (SERS) platforms beyond hotspot engineering: emerging opportunities in analyte manipulations and hybrid materials. *Chem. Soc. Rev.* **2019**, *48*, 731–756.
- (9) Ding, S.-Y.; Zhang, X.-M.; You, E.-M.; Ren, B.; Tian, Z.-Q. Surface-Enhanced Raman Spectroscopy: General Introduction. *Encyc. Anal. Chem.* **2006**, 1–42.
- (10) Deckert, V.; Deckert-Gaudig, T.; Diegel, M.; Götz, I.; Langelüddecke, L.; Schneidewind, H.; Sharma, G.; Singh, P.; Singh, P.; Trautmann, S.; Zeisberger, M.; Zhang, Z. Spatial resolution in Raman spectroscopy. *Faraday Discuss.* **2015**, *177*, 9–20.
- (11) Willets, K. A. Super-resolution imaging of SERS hot spots. *Chem. Soc. Rev.* **2014**, *43*, 3854–3864.
- (12) Fornasaro, S.; Alsamad, F.; Baia, M.; Batista de Carvalho, L. A.; Beleites, C.; Byrne, H. J.; Chiadò, A.; Chis, M.; Chisanga, M.; Daniel, A.; Dybas, J.; Eppe, G.; Falgayrac, G.; Faulds, K.; Gebavi, H.; Giorgis, F.; Goodacre, R.; Graham, D.; La Manna, P.; Laing, S.; Litti, L.; Lyng, F. M.; Malek, K.; Malherbe, C.; Marques, M. P. M.; Meneghetti, M.; Mitri, E.; Mohacek-Grosev, V.; Morasso, C.; Muhamadali, H.; Musto, P.; Novara, C.; Pannico, M.; Penel, G.; Piot, O.; Rindzevicius, T.; Rusu, E. A.; Schmidt, M. S.; Sergo, V.; Sockalingum, G. D.; Untereiner, V. R.; Vanna; Wiercigroch, E.; Bonifacio, A. Surface Enhanced Raman Spectroscopy for quantitative analysis: results of a large-scale European multi-instrument interlaboratory study. *Anal. Chem.* **2020**, *92*, 4053–4064.
- (13) Wang, Z.; Zong, S.; Wu, L.; Zhu, D.; Cui, Y. SERS-activated platforms for immunoassay: probes, encoding methods, and applications. *Chem. Rev.* **2017**, *117*, 7910–7963.
- (14) Kuhar, N.; Sil, S.; Verma, T.; Umapathy, S. Challenges in application of Raman spectroscopy to biology and materials. *RSC Adv.* **2018**, *8*, 25888–25908.

- (15) Stanborough, T.; Given, F. M.; Koch, B.; Sheen, C. R.; Stowers-Hull, A. B.; Waterland, M. R.; Crittenden, D. L. Optical Detection of CoV-SARS-2 Viral Proteins to Sub-Picomolar Concentrations. *ACS Omega* **2021**, *6*, 6404–6413.
- (16) Frank, A. J.; Cathcart, N.; Maly, K. E.; Kitaev, V. Synthesis of Silver Nanoprisms with Variable Size and Investigation of their Optical Properties: A First-year Undergraduate Experiment Exploring Plasmonic Nanoparticles. *J. Chem. Ed.* **2010**, *87*, 1098–1101.
- (17) Cathcart, N.; Frank, A. J.; Kitaev, V. Silver Nanoparticles with Planar Twinned Defects: Effect of Halides for Precise Tuning of Plasmon Resonance Maxima from 400 to > 900 nm. *Chem. Comm.* **2009**, *46*, 7170–7172.
- (18) Song, Y.; Song, J.; Wei, X.; Huang, M.; Sun, M.; Zhu, L.; Lin, B.; Shen, H.; Zhu, Z.; Yang, C. Discovery of Aptamers Targeting Receptor-Binding Domain of the SARS-CoV-2 Spike Glycoprotein. *Anal. Chem.* **2020**, *92*, 9895–9900.
- (19) Korepanov, V. I. Asymmetric least-squares baseline algorithm with peak screening for automatic processing of the Raman spectra. *J. Raman Spectrosc.* **2020**, *51*, 2061–2065.
- (20) Sackmann, M.; Materny, A. Surface enhanced Raman scattering (SERS)—a quantitative analytical tool? *J. Raman Spectrosc.* **2006**, *37*, 305–310.
- (21) Khan, S.; Ullah, R.; Javaid, S.; Shahzad, S.; Ali, H.; Bilal, M.; Saleem, M.; Ahmed, M. Raman spectroscopy combined with principal component analysis for screening nasopharyngeal cancer in human blood sera. *Appl. Spectrosc.* **2017**, *71*, 2497–2503.
- (22) He, X.; Liu, Y.; Huang, S.; Liu, Y.; Pu, X.; Xu, T. Raman spectroscopy coupled with principal component analysis to quantitatively analyze four crystallographic phases of explosive CL-20. *RSC Adv.* **2018**, *8*, 23348–23352.

- (23) Delfino, I.; Ricciardi, V.; Manti, L.; Lasalvia, M.; Lepore, M. Multivariate analysis of difference raman spectra of the irradiated nucleus and cytoplasm region of SH-SY5Y human neuroblastoma cells. *Sensors* **2019**, *19*, 3971.
- (24) Van de Sompel, D.; Garai, E.; Zavaleta, C.; Gambhir, S. S. A hybrid least squares and principal component analysis algorithm for Raman spectroscopy. *PLoS One* **2012**, *7*, e38850.
- (25) Guo, S.; Rösch, P.; Popp, J.; Bocklitz, T. Modified PCA and PLS: Towards a better classification in Raman spectroscopy-based biological applications. *J. Chemomet.* **2020**, *34*, e3202.

ToC Graphic



Difference Raman spectra showing concentration-dependent SARS-CoV-2 spike protein analyte binding to aptamer-functionalised silver nanoparticles