RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Mathematical Model for Biomolecular Quantification Using Large-Area Surface-Enhanced Raman Spectroscopy Mapping

Mirkó Palla,[┼,†,‖,*] *Filippo G. Bosco,*[‡,‖] *Jaeyoung Yang,*[†,‖] *Tomas Rindzevicius,*[‡] *Tommy S. Alstrom,*[§] *Michael S. Schmidt,*[‡] *, Qiao Lin,*[†] *Jingyue Ju,*[⊥] *Anja Boisen*[‡,*]

[┼]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, United States

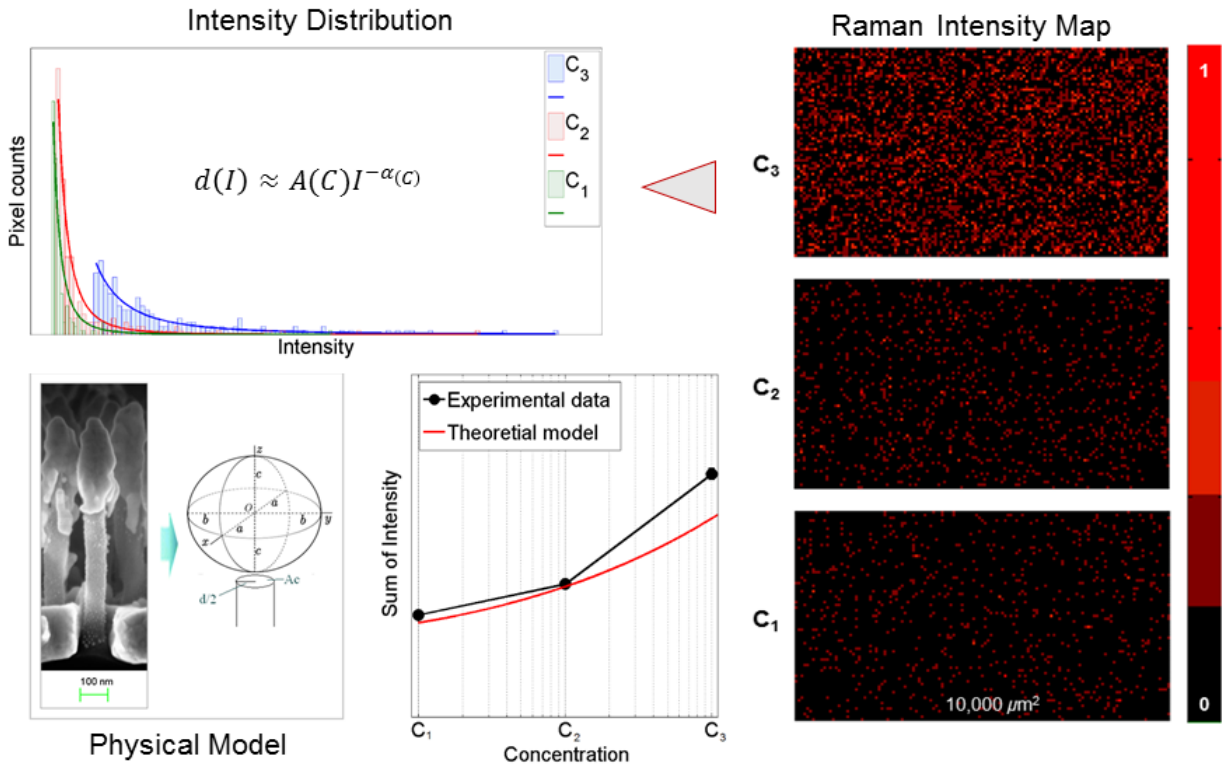
[†]Department of Mechanical Engineering, Columbia University, New York, New York 10027, United States

[⊥]Department of Chemical Engineering, Columbia University, New York, New York 10027, United States

[‡]Department of Micro- and Nanotechnology, Technical University of Denmark, Lyngby 2800, Denmark

[§]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby 2800, Denmark

Table of Contents Graphics



Based on single molecule surface-enhanced Raman spectroscopy (SERS) intensity distribution theory, a mathematical model is developed for highly sensitive biomolecular quantification using Raman mapping on SERS substrates with planar geometries.

ABSTRACT

Surface-enhanced Raman spectroscopy (SERS) based on nanostructured platforms is a promising technique for quantitative and highly sensitive detection of biomolecules in the field of analytical biochemistry. Here, we report a mathematical model to predict experimental SERS signal (or hotspot) intensity distributions of target molecules on receptor-functionalized nanopillar substrates for biomolecular quantification. We demonstrate that by utilizing only a small set of empirically determined parameters, our general theoretical framework agrees with the experimental data particularly well in the picomolar concentration regimes. This developed model may be generally used for biomolecular quantification using Raman mapping on SERS substrates with planar geometries, in which the hotspots are approximated as electromagnetic enhancement fields generated by closely spaced dimers. Lastly, we also show that the detection limit of a specific target molecule, TAMRA-labeled vasopressin, approaches the single molecule level, thus opening up an exciting new chapter in the field of SERS quantification.

KEYWORDS: surface-enhanced Raman spectroscopy, theoretical modeling, statistical quantification, signal intensity distribution, Raman mapping, biosensing

INTRODUCTION

The high sensitivity of surface-enhanced Raman scattering (SERS) has played an vital role in revitalizing the interest in SERS,[1,2] and in particular, applying this technique for low-abundance molecular quantification in various bioassays. A variety of SERS-active tags have been developed to label target molecules,[3,4,5] however, they have mainly been used for qualitative analysis. Quantitative analysis based on SERS labeling can be an important new analytical chemistry technique,[6,7,8] but requires the development of an effective quantification methodology by collectively measuring an ensemble of Raman signal intensities. Unfortunately, reliable biomolecular quantification using SERS has been hindered by poor reproducibility and uniformity of substrates, which frequently leads to large signal variations in the collected samples. This is a critical problem in the statistical interpretation of SERS experiments, which has been theoretically investigated in the literature,[9–11] but has not yet been rigorously addressed in connection with the applied sensor field.

During a typical measurement, Raman signals are amplified by large electromagnetic enhancements, so-called hotspots, and are collected from a SERS substrate. These highly localized hotspots are believed to be formed between metallic junctions of noble nanostructures or nanoparticles.[12,13] However, the acquisition of uniform SERS signals over a large area is particularly challenging as the signals significantly vary between hotspots with slightly different junction dimensions. Theoretically, an ideal approach to design such a sensing unit would be to precisely position one Raman active molecule (with surface functionalization strategies such as DNA origami or click chemistry) in the hotspot and eventually create an array-like configuration containing many of these units. Unfortunately, such precision for SERS substrate design and molecular positioning has not yet been achieved.

Typically, analyte molecules are adsorbed onto the surface of a nanostructured SERS substrate at random locations by either drop coating or incubation methods using liquid samples containing the analyte. These methods for SERS measurements present several shortcomings. The surface coverage of the analyte molecules is not uniform and difficult to estimate due to the stochastic nature of the adsorption process; thus large spatial variations are expected in measurements over the SERS substrate area. At ultra-low concentrations, often used in single molecule (SM) studies, the probability for a molecule to be located exactly in a hotspot is extremely low, which leads to unreliable statistics due to the large portion of unoccupied hotspots, *i.e.*, low number of detectable SERS signals. On the other hand, for each type of nanostructured substrate and analyte to be detected, there is a concentration regime for which the surface coverage of adsorbed molecules is optimal, such that an exponentially increasing relationship can be observed in the SERS signal intensity (emanating from the Raman active molecules) as a function of analyte concentrations.[14]

Here, we developed a statistical method based on large-scale Raman mapping, which improves statistical reliability and reproducibility of collected SERS signals, by studying the intensity distribution of an ensemble of SERS signals and its statistical implications. This analytical model has the potential to be used as a quantitative tool in SERS sensing applications. To demonstrate the significance of our approach, the following issues will be examined in this paper: (1) development of a theoretical model to fit and (2) quantification of experimental results based on SM principles, and (3) dependence of SERS responses on the number of probed molecules per unit area at picomolar levels.

In brief, the approach developed here takes advantage of the characteristic power law distribution of SERS hotspots collected from an ensemble of measurements over a large area.

Thus, it provides not only a general theoretical framework to describe statistics of SERS signals requiring a minimum set of parameters, but also a novel method for low abundance biomolecular quantification.

THEORETICAL BASIS

SERS mapping experiments were carried out on gold (Au) decorated "leaning" nanopillar periodic arrays,[15,16] in which each nanopillar was geometrically approximated as a spheroid attached to a column, where the spheroid represents the Au coated head of the nanopillar, while the column represents the silicon shaft which leans upon drying. Hotspots are assumed to be formed at the junctions of two metallic particles, that produce a detectable SERS signal.[12,13,17] For our particular nanopillar substrate, close (< 10 nm) proximity of spherical tips leads to the formation of hotspots, which gives rise to large SERS enhancements when analyte is trapped between them. To study the correlation between the intensity distribution and the statistics of SERS signals, we have developed an analytical model based on two closely spaced spherical metallic particles,[18] which represent our leaning nanopillars. The single hotspot model addresses the distribution of SERS intensities on the metal surface only, assuming that all molecules are adsorbed on the surface in a monolayer. It ignores the contribution of multiple layers of molecules, and neglects the chemical and resonance component of SERS enhancement.

SERS enhancements on most substrates are highly non-uniform on the molecular scale. Points of large electromagnetic enhancement or hotspots are highly localized and are so sparsely distributed that they can be often found within tens of nanometers of points with zero enhancements. For this reason, we defined the SM enhancement factor (SMEF) as the SERS enhancement induced by a given molecule at a specific point on the substrate. The SMEF is

dependent upon the Raman tensor of the probe molecule, which describes its orientation on the substrate and with respect to the local electromagnetic field at that point. It is also dependent on the SERS substrate orientation with respect to incident laser polarization[19] and direction. We assumed these parameters to be constant, since the substrate orientation was kept orthogonal to the incident laser beam, *i.e.*, it was fixed, and 0° polarization was used in all experiments. Therefore, the SMEF is mathematically defined as:

$$SMEF = \frac{I_{SERS}^{SM}}{\langle I_{RS}^{SM} \rangle}, \quad (1)$$

where I_{SERS}^{SM} is the SERS intensity of the SM under consideration, while $\langle I_{RS}^{SM} \rangle$ the average Raman intensity of the same probe molecule, where the average is taken over all random orientations of the molecule in space. For our purposes, I_{SERS}^{SM} can be experimentally defined as the hotspot intensity originating from the fraction of probed TAMRA (5-carboxytetramethyl-rhodamine)-labeled vasopressin (TVP) molecules in the sampling spot (~ 1 μm^2) that give rise to detectable SERS signal due to their plasmonic interaction with the leaning nanopillars; whereas $\langle I_{RS}^{SM} \rangle$ as the average intensity measured on the SERS substrate. Then, by rearranging Equation (1), we obtain the SERS electromagnetic enhancement factor:

$$F = SMEF = \frac{I_{SERS}^{SM}}{\langle I_{RS}^{SM} \rangle} = \frac{I}{\langle I_{RS}^{SM} \rangle} \xrightarrow{implies} I = F\langle I_{RS}^{SM} \rangle. \quad (2)$$

Thus, we showed that F is proportional to the measured intensity I by a constant factor $\langle I_{RS}^{SM} \rangle$. In general, F can be calculated using the Raman tensor of the probe molecule, its adsorption and scattering geometry, as well as its vibrational mode energy.[18] For simplicity, F can be expressed in a more general form such that:

$$F \approx \left| \frac{E_{local}}{E_0} \right|^4 \quad (3)$$

where E_{local} is the local electromagnetic field; and E_0 is the incident field magnitude. Based on Equation (3) and a theoretical framework describing enhancement factor distribution around a single SERS hotspot,[18] the probability density function (PDF) *p(F)* that a randomly positioned probe molecule experiences a given enhancement *F* was derived as:

$$p(F) \approx AF^{-(1+k)} \tag{4}$$

which is a long-tailed distribution similar to the Pareto distribution. Since *F* is proportional to the measured intensity *I* for the same event (Equation (2)), an analytical expression analogous to Equation (4) can be derived to express the intensity distribution *d(I)* associated with the single hotspot model corresponding to a truncated Pareto distribution (TPD). It is truncated, since the distribution does not converge to $I \rightarrow \infty$, but has a maximum value at I_{max}, which is the largest experimentally obtained intensity value in the sample. Therefore, we can approximate the intensity distribution by a TPD with only four parameters: *k*, *A*, I_{max} and I_{min}, which provides an accurate description for the large intensity regime:

$$d(I) \approx AI^{-(1+k)} \tag{5}$$

where *k* is a decay constant that describes the rate of change of enhancement as we move away from the hotspot, *A* is a scaling factor that indicates the probability of finding a molecule when moving away from the hotspot (within 20 nm),[20] and I_{max}, I_{min} are the maximum, minimum intensity values of the hotspots respectively. The intensity distribution function, *d(I)* has the property that $d(I) \rightarrow 0$ as $I \rightarrow \infty$, which gives rise to the power rate convergence in the long-tail of the TPD.

While the intensity distribution model developed in the previous section only described a single hotspot, it is relevant to hotspot mapping experiments of leaning nanopillars, in particular for the picomolar concentration regime and < 1 μm^2 signal sampling areas. With increasing number of hotspots, the PDF of SERS intensity (Equation (5)) tends to a Gaussian distribution

(central limit theorem), whose upper tail - used in data processing - can be approximated fairly well with a TPD.

Additionally, the detected SERS signal results from the summation of many SM signals located within the probed surface area (~ 1 μm^2), since the observed signal intensity can be expressed as the sum of signals from independent hotspots by the superposition principle. Since the Au nanopillar leaning direction is random, each dimer will be randomly oriented with respect to the incident field polarization. Only a dimer with (i) a smallest gap and (ii) parallel to the incident field polarization displays highest electromagnetic field enhancement, and (iii) a dimer should also contain a number of target molecules within the hotspot area (within 10-20 nm, picomolar concentration regime) which then produces a detectable SERS intensity. We therefore believe that recorded SERS intensities for each measurement point are dominated by a single hotspot.

Specifically, when the nanopillars are functionalized with vasopressin-specific aptamers, upon liquid evaporation, the pillars lean on each other forming highly enhancing hotspots, modeled as dimers of metallic spheres. When the TVP is introduced, it binds to the aptamers randomly distributed on the pillar heads. When vasopressin is trapped by an aptamer close enough to a hotspot, the TAMRA SERS signal is highly enhanced. Because of the random nature of Au-thiol bonding between aptamers and Au-coated pillar heads and the TVP binding to aptamers, statistically only a few TAMRA molecules may be positioned close to the hotspot at a time in the scattering volume. However, the majority of SERS signal will be derived from these particular molecules due to the predominant enhancement effects.

From the above formulated statistical arguments, in the theoretical domain, this particular problem deals with statistically equivalent observations of SERS signals collected on identical

dimers of metallic spheres. In the case of vasopressin detection, using the leaning nanopillar approach, we experimentally measure the SERS signal from TAMRA molecules deposited on the Au-coated pillar heads. Here, in the experimental domain, statistics are determined by collecting a large ensemble of random measurements from the SERS substrate and compiling them into histograms. These two approaches are analogous to each other and are the basis for correlating SERS signal intensity to analyte concentration in this analytical quantification method.

RESULTS AND DISCUSSION

To experimentally determine the SERS intensity distribution, we chose the TAMRA peak at 1370 cm^{-1} as a diagnostic peak. For the quantification of TVP, we collected about 10,000 Raman spectra for the various analyte concentrations in the picomolar regime using point-by-point mapping (see Materials and Methods for acquisition algorithm). A typical SERS intensity distribution map for the diagnostic peak can be represented as a two-dimensional (2D) matrix (or heatmap) rendering intensity values at each 1 μm^2 mapping coordinate (Fig. 1). As can be seen in Fig. 1, the higher the TVP concentration, the more and brighter the pixels are seen in the map, implying that the probability of obtaining SERS hotspots increases with the total number of analytes adsorbed on the nanopillar surface.

To represent the SERS intensity distribution in another way, one can generate an intensity histogram rendering a large number of SERS events of randomly distributed molecules, in which the smallest and largest value on the *x* axis corresponds to the minimum and maximum of the top *X*% of all SERS signals measured during the Raman mappings. At 100 pM TVP, for the diagnostic peak 1370 cm^{-1}, such histograms were constructed for the top 5%, 10%, 20% and 100% of measured SERS signals (Fig. 2A-D), and was fitted with a Pareto function in the general form:

$y(x) = ax^n$. It was observed that there is a functional relationship between the coefficient of determination value (R^2) for the Pareto fit and the number of hotspots. This is demonstrated in Fig. 3 for the full range of hotspot percentages, which plateaus approximately at the top 20% cutoff value ($R^2 \approx 1.00$, $X = 20\%$). At this inflection point (Fig. 2C), the collective SERS intensities reflect a highly skewed or long-tailed TPD distribution similar to that described by Equation (5). Then, the R^2 values moderately decrease until the full range of SERS intensities are utilized ($X = 100\%$) (Fig. 3). This observation is consistent with the general Pareto principle and experimental SM SERS intensity distributions studies,[21,22] thus this threshold was chosen in further analysis for the development of our quantification model.

The constructed SERS intensity histograms, containing the top 20% of hotspots, were fitted with a power function for the diagnostic TAMRA peak 1370 cm^{-1} at the three different concentration levels in the picomolar regime (Fig. 4). The high coefficient of determination (R^2) values (> 0.9) demonstrate a strong power relationship between TVP concentration and SERS intensity integrals (Table 1). Since the experimentally obtained SERS intensity distribution fits well to the predicted long-tailed TPD distribution, we believe that the analytical model – based on single hotspot theory developed in Equation (3-5) – is suitable to describe the SERS enhancement mechanism of leaning nanopillars for TVP detection. By arbitrarily defining $\alpha = 1 + k$, we can now rewrite Equation (5) for a given TVP concentration (C) as:

$$d(I) \approx A(C) I^{-\alpha(C)}, \quad (6)$$

where the equation indicates that parameters α and A can possible depend on the analyte concentration. Even though A is assumed to be an intrinsic characteristic of the SERS substrate, the leaning nanopillars could potentially lump together in various three-dimensional

configurations under different concentrations, thus changing the probability for a Raman tag (TAMRA) to be located close to a hotspot.

The SERS substrate is considered as a collection of distinct hotspots. The distribution of corresponding SERS intensities – to a first approximation – can be obtained by summing their respective intensities. Now, based on the experimentally obtained fitting parameters at the three picomolar concentration levels (Table 2), it is reasonable to assume, that α and A may be (arbitrarily) represented as logarithmic functions of the analyte concentration with general forms:

$$\alpha(C) = b\ln(C) + d, \tag{7}$$

$$A(C) = g\ln(C) + h, \tag{8}$$

where b, d, g and h are constants whose values are determined by logarithmic fits (Fig. S5A-B in ESI). Then we can express the SERS intensity distribution analytically as a function of concentration by substituting Equation (7-8) into (6) as:

$$d(I) \approx (g\ln(C) + h)I^{b\ln(C)+d}. \tag{9}$$

For a given analyte concentration, the integral of the intensity distribution between the local intensity minimum (I_{min}) and maximum (I_{max}) values give the sum of all intensities (I_{sum}) in a particular mapping experiment. More precisely, I_{min} and I_{max} are defined as the lowest (minima) and highest (maxima) experimentally determined hotspot intensity values in each experiment. Now, this integral can be expressed as a function of analyte concentration (C) such that:

$$I_{sum} = \int_{I_{min}}^{I_{max}} d(I)\,dI = \int_{I_{min}}^{I_{max}} A(C)I^{-\alpha(C)}\,dI = A(C)\int_{I_{min}}^{I_{max}} I^{-\alpha(C)}\,dI, \tag{10}$$

or

$$I_{sum} = A(C)\left[\frac{1}{1-\alpha(c)} I^{1-\alpha(C)}\right]_{I_{min}}^{I_{max}} = \left[\frac{A(C)}{1-\alpha(C)}\right]\left[{I_{max}}^{1-\alpha(C)} - {I_{min}}^{1-\alpha(C)}\right], \quad (11)$$

By substituting Equation (7) and (8) into Equation (11), we finally gain an analytical expression for the total intensity as a function of the TVP concentration, *C*:

$$I_{sum} = \left[\frac{g\ln(C)+h}{1-b\ln(C)-d}\right]\left[{I_{max}}^{1-b\ln(C)-d} - {I_{min}}^{1-b\ln(C)-d}\right]. \quad (12)$$

It should be noted that in Equation (12) the upper and lower limits of integration need to be substituted with their function counterparts, since I_{min} and I_{max} may also be described as logarithmic functions of the analyte concentration (Fig. S5C-D in ESI). Here, we simplified our quantification model by redefining I_{min} and I_{max} as the lowest and highest experimentally determined hotspot intensity values in all mapping experiments, *i.e.*, global extremes (Table 2); but this observation may be investigated in a future work using a more rigorous framework.

Fig. 5 shows the total intensities obtained from experiments at various analyte concentrations (solid black line) along with their theoretical fits based on the analytical expression for the intensity integral (Equation (12)) – using the global intensity extremes (dotted red line) or the mean of local extremes (dashed red line) respectively. This result demonstrates that the long-tailed (Pareto-like) intensity distribution model developed for a single hotspot (Equation (6)) can predict experimental values for the TVP quantification using the leaning nanopillar-based SERS substrate (Table 3). The model fits the experimental data well in the tested picomolar concentration regimes and calls for further experiments to evaluate its feasibility in ultralow (femotmolar) analyte concentrations.

In order to minimize the variability in measurements for individual substrates with the same functionalization/treatment conditions, coefficient of variation (CV) analysis was used to

determine the optimal mapping area (see Materials and Methods). This statistical method determines the extent of variability in relation to the mean of the population, *i.e.*, it is a normalized measure of the dispersion in the intensity distribution obtained from the Raman mapping experiments. Nanopillar substrates functionalized with the vasopressin-specific aptamer and treated with 1 nM TVP were sampled using the CV algorithm, computing the mapping-to-mapping variability under the same substrate functionalization conditions. The 2D scatter plots in Fig. 6 indicate that the CV exponentially decays as the mapping area increases when the diagnostic TAMRA peak 1370 cm^{-1} is investigated. A total of 100 pixels – equivalent to a 10 μm × 10 μm SERS substrate area – are required to reach a variability threshold of 1% for both 1 pM and 100 pM TVP concentrations (Fig. 6A-B). Thus, the CV analysis verified that the measurements from the different substrates were not statistically different, using at least square mapping areas of ~100 μm^2 in the picomolar concentration regime, demonstrating the robustness of experimental repeatability of the mapping technique and providing an optimal area threshold to minimize variability between experiments.

The quantitative method denoted above provides information about the ensemble of SERS signals in a collective way, but does not detail the number of molecules per SERS enhancement event, which might be a critical parameter for SM studies. To address this issue, we approximated the leaning nanopillar heads as prolate spheroids and the pillar columns as cylinders based on high resolution SEM images with dimensions shown in Fig. 7. Semi-axes *a, b* (a = b in our case) and *c* are 75 nm, 75 nm and 120 nm respectively, while the column radius (*d*/2) was measured to be about 40 nm. Now, the surface area of the elongated spheroid (*S*) was calculated using the formula:

$$S = \left[\frac{a^{2p} + 2a^{p}c^{p}}{3}\right]^{\frac{1}{p}} \quad (13)$$

with $p \approx 1.6075$, which is the parameter utilized for nearly spherical ellipsoids. The total area of a single nanopillar (A_{pill}) was then calculated by subtracting the pillar cross-section (A_c) from the ellipsoid area:

$$A_{pill} = S - A_c = S - \frac{d^2}{4}\pi \quad (14)$$

Given the calculated nanopillar density, $D_{pill} = \frac{20}{\mu m^2}$,[15] we estimated the total number of nanopillars on the 5 × 5 mm chip (N_{pill}), which was used in all Raman mapping measurements, to be: $N_{pill} = A_{chip} D_{pill} = 5 \times 10^8$. By using these approximations, the total active gold area of the nanopillars (A_{tot}) – assuming that gold is only deposited on the spheroid and not on the pillar – over the entire chip was calculated to be: $A_{tot} = N_{pill} A_{pill} = 0.436\ cm^2$. Now, by considering the maximum packing density of thiol-DNA bonds (10^{12}/cm^{-2}),[23] the number of adsorbed aptamers (N_{apt}) on this surface was estimated to be: $N_{apt} = 4.36 \times 10^{11}$. Then, using the Langmuir Isotherm, we calculated the total number of vasopressin molecules bound by aptamers on the nanopillar surfaces: $N_{analyte} = N_{apt} \frac{C}{C+K}$, where c is the TVP concentration and κ is the dissociation constant (1.17 nM)[24] (Table 4). We then estimated that less than ~20 molecules per cluster would be probed by a single SERS measurement at 1 pM TVP concentration. This estimation suggests the possibility of SM detection, just one order of magnitude away, using this statistical SERS quantification method.

CONCLUSIONS

In this paper, we have developed a novel statistical method for quantifying trace amounts of biomolecules (TAMRA-labeled vasopressin) by SERS using a rigorous mathematical derivation. This method was built upon a theoretical framework dealing with the enhancement factor distribution around a single surface-enhanced Raman scattering hotspot and its relation to

SERS detection over a large substrate area. We demonstrated that the experimentally obtained SERS hotspot data fits the analytical predictions well in the picomolar concentration regimes studied here. We hypothesize that this quantification framework could be generalized for planar SERS substrates, in which the nanostructured SERS features can be approximated as a closely spaced electromagnetic dimer problem. We also showed that our approach is robust in terms of experimental repeatability. Finally, we demonstrated the potential for SM detection by estimating the number of analyte molecules probed in the Raman scattering volume during each laser excitation using the Langmuir Isotherm. This opens up an exciting opportunity for future work focusing on the optimization of SERS substrates, surface functionalization methods and data acquisition strategies for potential SM SERS studies.

MATERIALS AND METHODS

Experimental Protocol. The nanopillar SERS substrate was functionalized with vasopressin-specific aptamer sequences, then TVP samples were prepared as described in our previous work (see Materials and Methods).[25] TVP samples in the concentration range of 1 pM to 100 pM were used. Next, the aptamer functionalized substrates were incubated with the TVP samples for 1 hr at 37 °C in a thermal cycler (Eppendorf AG, Germany) to provide optimal conditions for the aptamer-vasopressin binding. Then, the substrates were transferred into vasopressin buffer solutions and kept at 37 °C for 15 min to remove unbound molecules. Lastly, the substrates were briefly rinsed once with deionized water to prevent salt aggregation on the dried substrates.

SERS Hotspot Mapping. SERS measurements were performed under the 632.8 nm excitation line of a He−Ne laser on an inVia Raman microscope (Renishaw, UK) in a standard backscattering configuration. The Raman signal was collected through a confocal pinhole of 25 μm diameter by using 0.75 NA dry objective of 50× magnification (Leica Microsystems, Germany). A computer controlled XY translation stage was used to acquire a total of 10,087 SERS spectra within the scan area of 130 × 76 μm divided into an acquisition grid by 1 μm step size in both dimensions. All spectra in this work were obtained with an exposure time of 1 sec and at 0.3 mW laser power before the objective. The SERS intensity maps of the selected diagnostic TAMRA vibration mode (1370 cm^{-1}) were derived using the signal to baseline map generation algorithm of the WiRE 3.2 software (Renishaw, UK) for the spectral range of 1350-1390 cm^{-1}.

Repeatability of SERS Mapping Experiments. Coefficient of variation (CV) analysis was used to determine the minimum (or optimal) scanning area to minimize variation between Raman mapping experiments where SERS signals from TVP on the substrate were collected. For both lowest and highest concentrations, 1 pM and 100 pM respectively, varying areas within the SERS substrate were sampled randomly at a time, starting from 1 × 1 to 35 × 35 pixels as square area, for three independent mapping experiments respectively. SERS intensity integrals were calculated for each experiment, then the coefficient of variation (c_v), defined as:

$$c_v = \frac{\sigma}{\mu}$$

where σ is the standard deviation; and μ is the mean of the collected SERS intensities, was computed comparing mapping-to-mapping variability under the same substrate functionalization conditions. This sampling was repeated 1,000 times. Finally, the CV values versus side length of

square mapping area (in pixels) was plotted on a 2D scatter plot, in which, at each side length parameter tick mark, all (1,000) CV values were displayed.

FIGURES

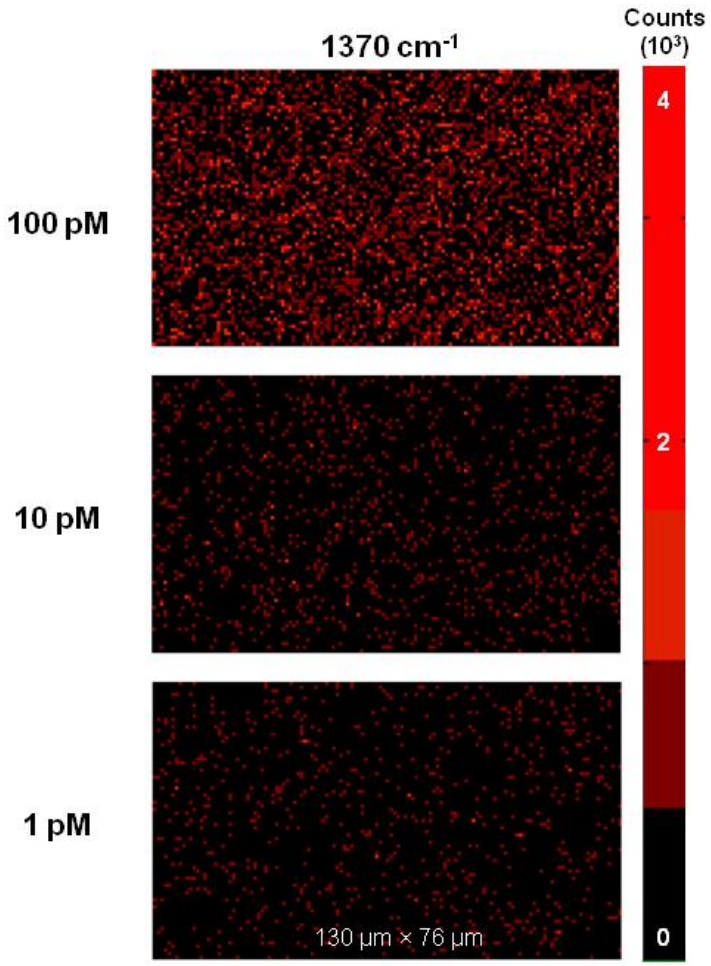


Fig. 1 Spectroscopic Raman mapping of SERS intensity distribution. Intensity map of TAMRA peak 1370 cm^{-1} in a single mapping experiment for 100 pM (top), 10 pM (middle) and 1 pM (bottom) TAMRA-labeled vasopressin quantification. A substrate area about 130 × 80 μm^2 is shown, along with an intensity scale bar with a maximum intensity value of 4.00 × 10^3 counts.

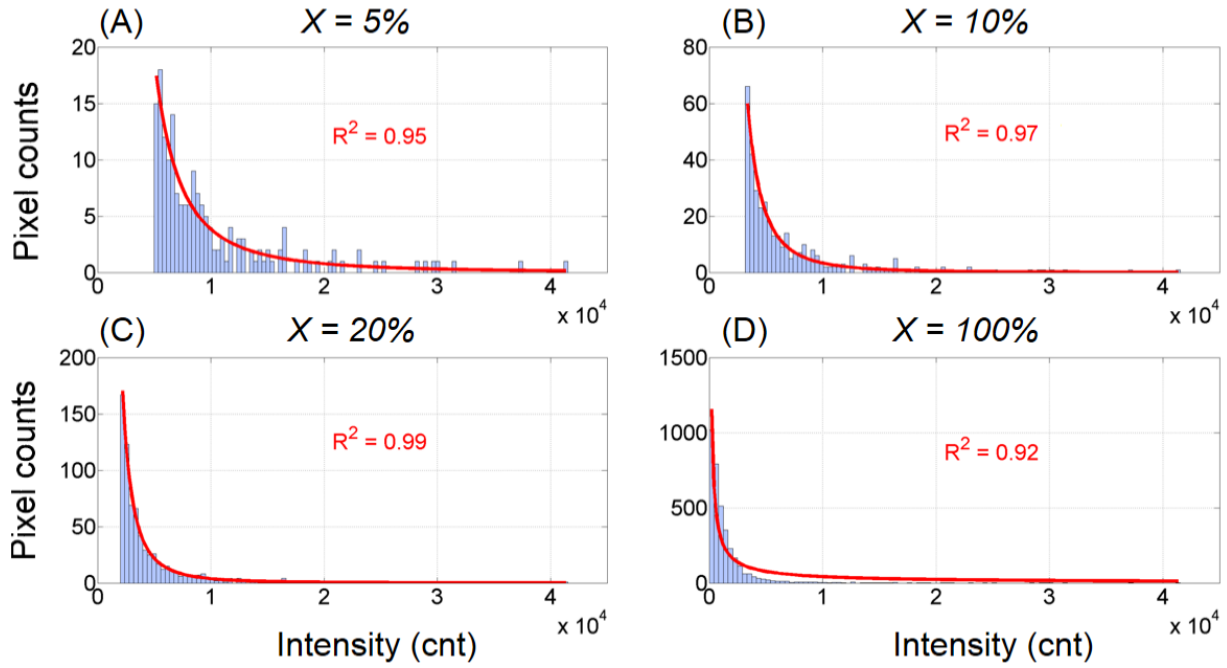


Fig. 2 Histogram of SERS intensity distribution for the top (A) 5%, (B) 10%, (C) 20% and (D) 100% of hotspots of peak 1370 cm^{-1} of TAMRA-labeled vasopressin at 100 pM concentration. Intensity histograms fit power fit functions (solid red line). R^2 values are also shown. Note, the different y-axis scales in each histogram.

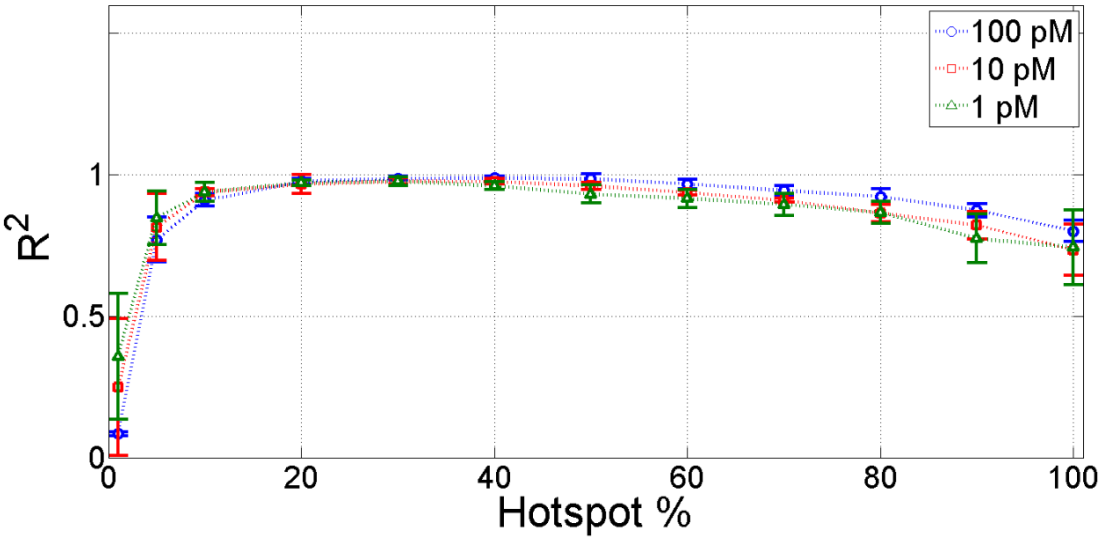


Fig. 3 Coefficient of determination (R^2) of power (Pareto) fit for the SERS intensity distribution as a function of percentage (%) of hotspots evaluated in the fitting of peak 1370 cm^{-1} of TAMRA-labeled vasopressin (TVP) at 100 pM (blue curve), 10 pM (red curve) and 1 pM (green curve) concentrations. Note, that the R^2 values follow a similar trend for all TVP concentrations and plateau at about 20%, which is predicted by the theoretical Pareto principle.[22] Error bars represent standard deviations for three independent mapping measurements. Fig. 2 represents four slices from this figure.

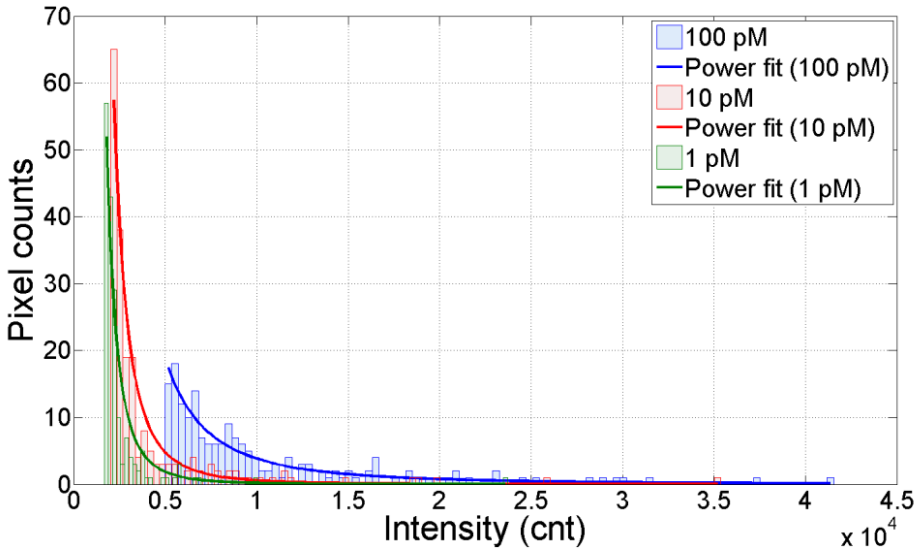


Fig. 4 Histogram representation of SERS intensity distribution for the top 20% of hotspots at different concentrations of TAMRA-labeled vasopressin (TVP). Histograms are displayed in overlaid configuration for peak 1370 cm^{-1} of TVP corresponding to various analyte concentrations. Blue, red and green distributions represent the 100 pM, 10 pM and 1 pM TVP concentrations. All fit with a power function in the general form: $y(x) = ax^n$ and the colors of the curve fits match those of the concentrations.

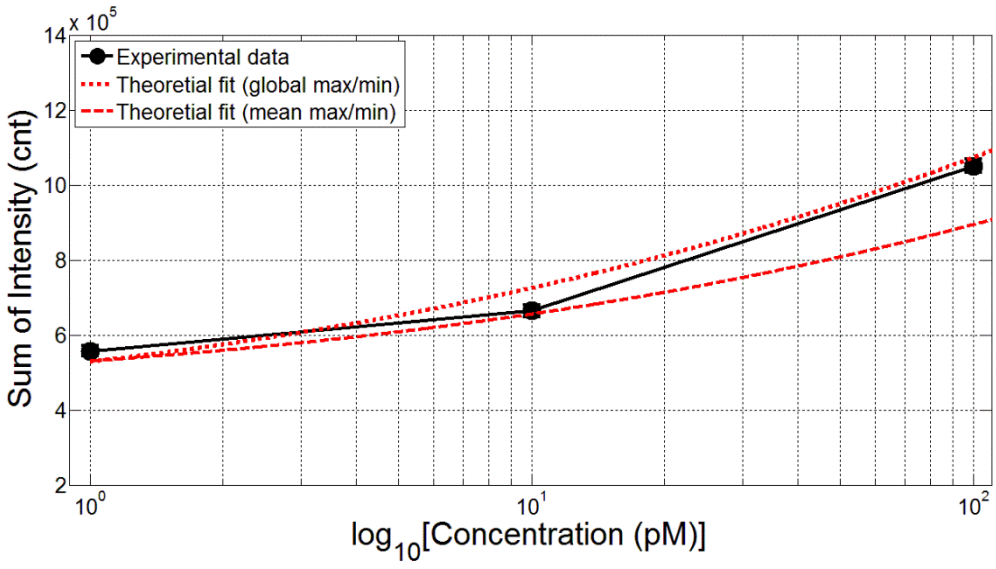


Fig. 5 Comparison of intensity integrals of experimental data and theoretical fit for the diagnostic TAMRA peak 1370 cm^{-1} as a function of TAMRA-labeled vasopressin (TVP) concentration. The experimental curve (solid black line) is predicted by theoretical fits relatively well in the picomolar TVP concentration regimes. Both theoretical models using the global intensity maxima/minima values (dotted red line) and the mean of local extremes (dashed red line) as limits of integration in the analytical expression for the intensity integral defined by Equation (12) are shown.

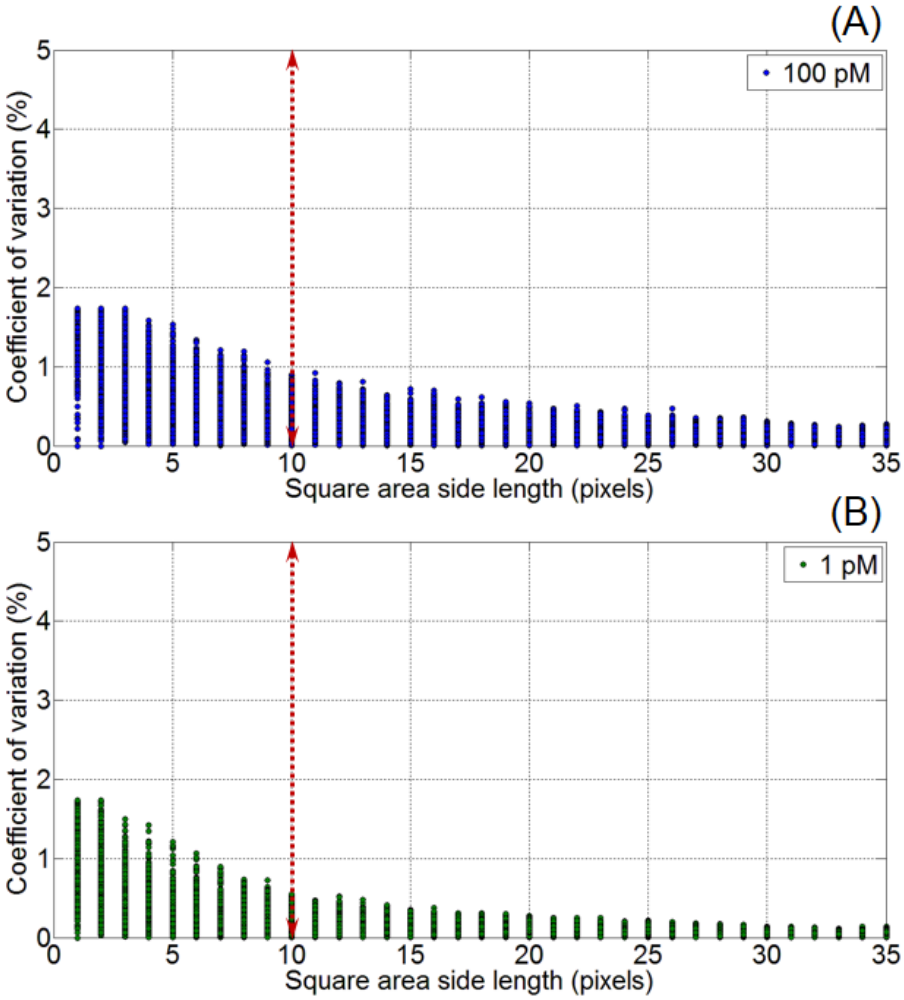


Fig. 6 Determination of optimal Raman mapping area to minimize variation between measurements. Coefficient of variability values *versus* side length of square mapping area (in pixels) exponentially decays for both (A) 100 pM (blue) and (B) 1 pM (green) TAMRA-labeled vasopressin concentrations. Optimal mapping area is estimated to be 100 to minimize experimental variability (vertical red lines) below the 1% CV threshold. All sampled CV values are displayed.

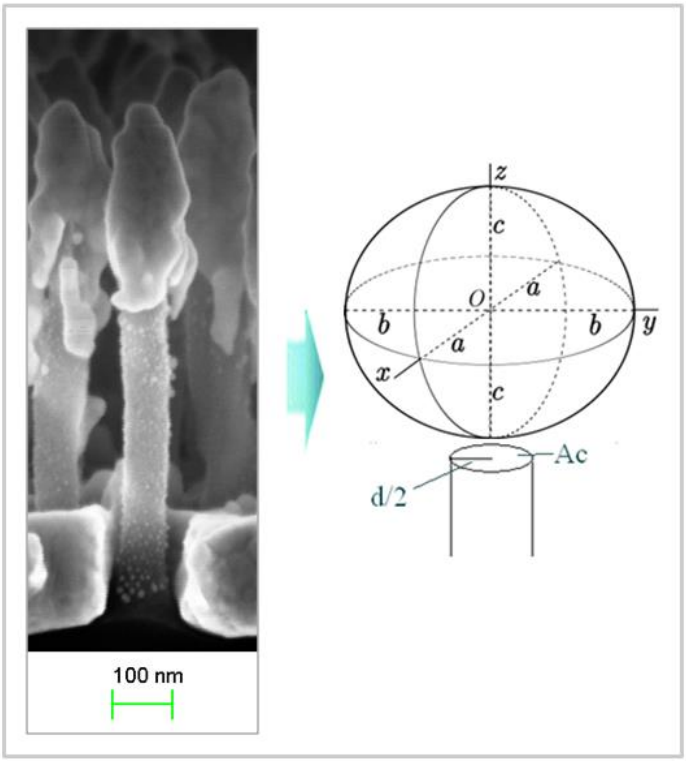


Fig. 7 Geometrical approximation of a leaning nanopillar. The pillar head can be modeled as a prolate spheroid (a=b<c) with semi-axes *a*, *b* (*a*=*b* in our case) and *c*; while the pillar shaft can be approximated as a column with a circular cross-section characterized by diameter, *d*. The cross-sectional area of the pillar shaft is denoted by A_c.

TABLES

Table 1 Coefficient of determination of power fit for the SERS intensity distribution of peak 1370 cm^{-1} of TAMRA-labeled vasopressin (TVP). Each entry is the average of three independent measurements.

TVP Concentration (pM)	Coefficient of Determination (R^2)
10^0	0.9270
10^1	0.9562
10^2	0.9003

Table 2 Power fit exponent (α), scaling factor (A), signal maximum (I_{max}) and signal minimum (I_{min}) of the SERS intensity distribution of various analyte concentrations in the picomolar regime for peak 1370 cm^{-1} of TAMRA-labeled vasopressin (TVP).

TVP Concentration (pM)	**10^2**	**10^1**	**10^0**
A	5.7244e+09	6.9896e+11	1.7156e+12
α	2.2933	3.0187	3.2350
I_{max}	4.1550e+04	3.5370e+04	2.3710e+04
I_{min}	4.9875e+03	2.0253e+03	1.6108e+03

Table 3 Comparison of experimentally and theoretically obtained intensity integral values for peak 1370 cm^{-1} of TAMRA-labeled vasopressin (TVP). Both theoretical models, using the global and mean intensity extremes as limits of integration in Equation (12), are shown.

Theoretical Limits of Integration	TVP Concentration (pM)	Experimental Mean Value (cnt)	Theoretical Prediction (cnt)	Percentage Error (%)
Global maxima/minima	10^0	5.5073E+05	5.2994E+05	-3.78 ± 6.04
	10^1	6.6482E+05	7.2652E+05	+9.28 ± 8.70
	10^2	1.0502E+06	1.0749E+06	+2.35 ± 0.83
Mean of local maxima/minima	10^0	5.5073E+05	5.2994E+05	-3.78 ± 6.04
	10^1	6.6482E+05	6.5573E+05	-1.37 ± 8.70
	10^2	1.0502E+06	8.9447E+05	-14.83 ± 0.83

Table 4 Number of TAMRA-labeled vasopressin (TVP) molecules on the SERS biosensor. Summary of approximating calculations estimating the total number of TVP molecules adsorbed on the SERS substrate surface and the total number of TVP molecules probed in the Raman detection volume.

TVP Concentration (pM)	Total number of absorbed molecules (#)	Number of molecules per cluster ($1/\mu m^2$)
100	3.96×10^{10}	1585
10	4.32×10^{9}	172
1	4.35×10^{8}	17

ASSOCIATED CONTENT

Supplementary Information. Statistical algorithm and experimental procedures for determining the minimum mapping area; alternative representation of the SERS intensity distributions; figures supplementing the statistical tests; and additional references.

AUTHOR INFORMATION

Corresponding Author

*E-mail: anja.boisen@nanotech.dtu.dk (A.B).

*E-mail: mirko.palla@wyss.harvard.edu (M.P).

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‖These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We acknowledge the support from the project MUSE (Innovation Fund Denmark) and the project NAPLAS (Sarpere Aude, The Danish Council for Independent Research).

ABBREVIATIONS

SERS, surface-enhanced Raman spectroscopy; SM, single molecule; EF, enhancement factor; TVP, TAMRA (5-carboxytetramethyl-rhodamine)-labeled vasopressin; PDF, probability density function; TPD, truncated Pareto distribution; CV, coefficient of variation.

REFERENCES

1 S. Nie, *Science,* 1997, **275**, 1102–1106.

2 K. Kneipp, Y. Wang, H. Kneipp, L. Perelman, I. Itzkan, R. Dasari and M. Feld, *Phys. Rev. Lett.*, 1997, **78**, 1667–1670.

3 E. S. Allgeyer, A. Pongan, M. Browne and M. D. Mason, *Nano Lett.*, 2009, **9**, 3816–3819.

4 Y. C. Cao, R. Jin and C. A. Mirkin, *Science*, 2002, **297**, 1536–1540.

5 Y. C. Cao, R. Jin, J. M. Nam, C. S. Thaxton and C. A. Mirkin, *J. Am. Chem. Soc.*, 2003, **125**, 14676–14677.

6 L. Wang, M. B. O'Donoghue and W. Tan, *Nanomedicine (Lond).*, 2006, **1**, 413–426.

7 R. Wilson, A. R. Cossins and D. G. Spiller, *Angew. Chem. Int. Ed. Engl.*, 2006, **45**, 6104–6117.

8 W. E. Doering, M. E. Piotti, M. J. Natan and R. G. Freeman, *Adv. Mater.*, 2007, **19**, 3100–3108.

9 P. G. Etchegoin and E. C. Le Ru, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6079–6089.

10 P. G. Etchegoin, M. Meyer and E. C. Le Ru, *Phys. Chem. Chem. Phys.*, 2007, **9**, 3006–3010.

11 P. G. Etchegoin, M. Meyer, E. Blackie and E. C. Le Ru, *Anal. Chem.*, 2007, **79**, 8411–8415.

12 A. M. Michaels, Jiang and L. Brus, *J. Phys. Chem. B*, 2000, **104**, 11965–11971.

13 Xu, Aizpurua, Kall and Apell, *Phys. Rev. E. Stat. Phys. Plasmas. Fluids. Relat. Interdiscip. Topics*, 2000, **62**, 4318–4324.

14 A. Kim, S. J. Barcelo, R. S. Williams and Z. Li, *Anal. Chem.*, 2012, **84**, 9303–9309.

15 M. S. Schmidt, J. Hübner and A. Boisen, *Adv. Mater.*, 2012, **24**, OP11–OP18.

16 K. Wu, T. Rindzevicius, M. S. Schmidt, K. B. Mogensen, A. Hakonen and A. Boisen, *J. Phys. Chem. C*, 2015, **119**, 2053–2062.

17 M. Futamata, *Faraday Discuss.*, 2006, **132**, 45–94.

18 E. C. Le Ru, P. G. Etchegoin and M. Meyer, *J. Chem. Phys.*, 2006, **125**, 204701.

19 P. G. Etchegoin, C. Galloway and E. C. Le Ru, *Phys. Chem. Chem. Phys.*, 2006, **8**, 2624–2628.

20 K. Wu, T. Rindzevicius, M. S. Schmidt, K. B. Mogensen, S. Xiao and A. Boisen, *Opt. Express*, 2015, **23**, 12965–12978.

21 Y. Fang, N.-H. Seong and D. D. Dlott, *Science*, 2008, **321** , 388–392.

22 M. E. J. Newman, *Contemp. Phys.*, 2005, **46**, 323–351.

23 A. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez and P. G. Schultz, *Nature*, 1996, **382**, 609–611.

24 W. G. Purschke, D. Eulberg, K. Buchner, S. Vonhoff and S. Klussmann, *Proc. Natl. Acad. Sci. United States Am.* , 2006, **103** , 5173–5178.

25 J. Yang, M. Palla, F. G. Bosco, T. Rindzevicius, T. S. Alstrøm, M. S. Schmidt, A. Boisen, J. Ju and Q. Lin, *ACS Nano*, 2013, **7**, 5350–5359.