Electronic Supplementary Material (ESI) for RSC Advances.

SUPPORTING INFORMATION

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*

*** To whom correspondence should be addressed.**

Email: anja.boisen@nanotech.dtu.dk | mirko.palla@wyss.harvard.edu

CONTENTS

1. MINIMUM MAPPING AREA

1.1. Algorithm Description

The unpaired-sample Student's *t*-test (STT) was utilized to determine the minimum substrate area required to unambiguously distinguish between treatment and control samples. SERS intensity mappings were considered independent from each other for both treatment and control samples. Based on these measurements, two sets of square areas (corresponding to treatment and control) were sampled randomly at a time, starting from 1 × 1 to 35 × 35 pixels as square area. According to this sampling algorithm, we calculated the *p*-values of the STT[1,2] for unpaired samples evaluating different sizes of substrates. This sampling was repeated 1,000 times. Finally, the mean *p*-value of STT versus side length of square mapping area (in pixels) was plotted, with error bars displaying one standard deviation based on three independent mapping experiments (Fig. S1).

1.2 Experimental Procedures

The *p*-values from the STT for independent samples (or unpaired *t*-test) were calculated comparing a collection of SERS signals measured on two substrates treated with 1 nM TAMRA-labeled vasopressin (TVP) samples, each functionalized with vasopressin-specific aptamers (experiment) and with immunoglobulin E (IgE)-specific aptamers (control) respectively. Fig. S1 indicates that when using the STT test, in which TAMRA peak 1370 cm^{-1} is investigated, a total of 400 pixels – equivalent to a 20 μm × 20 μm SERS substrate area – are required to reach the *p*-value threshold ($p = 0.05$). This value can be considered as an inflection point, since as the mapping area increases from this point on, the *p*-value of STT converges to a fixed saturation level ($p = 0.01$).

2. STEM PLOT OF INTENSITY DISTRIBUTIONS

As an alternative representation emphasizing the discrete statistical nature of SERS intensity distributions, in an accumulative way, one can draw a vertical line at each SERS intensity value (*x*-axis) to mark off its occurrences collected during Raman mapping. In this way, one can easily determine the hotspot density of the heatmap identifying distribution trends and outliers. Similarly, as for the histogram representation (Fig. 5), we created stem plots of the top 5%, 10% and 20% (Fig. S2-4) of measured SERS signals for TAMRA peak 1370 cm^{-1} at varying TVP concentrations. We have observed that as the TVP concentration increases when keeping *X* constant, the number of higher SERS intensities increase as well, which is designated by the accumulation of intensity marks to the right along the *x*-axis (Fig. S4, X = 20%). This observation is consistent with the concentration dependence of SERS signals as previously shown using heatmaps (Fig. 1).

3. FIGURES AND TABLES

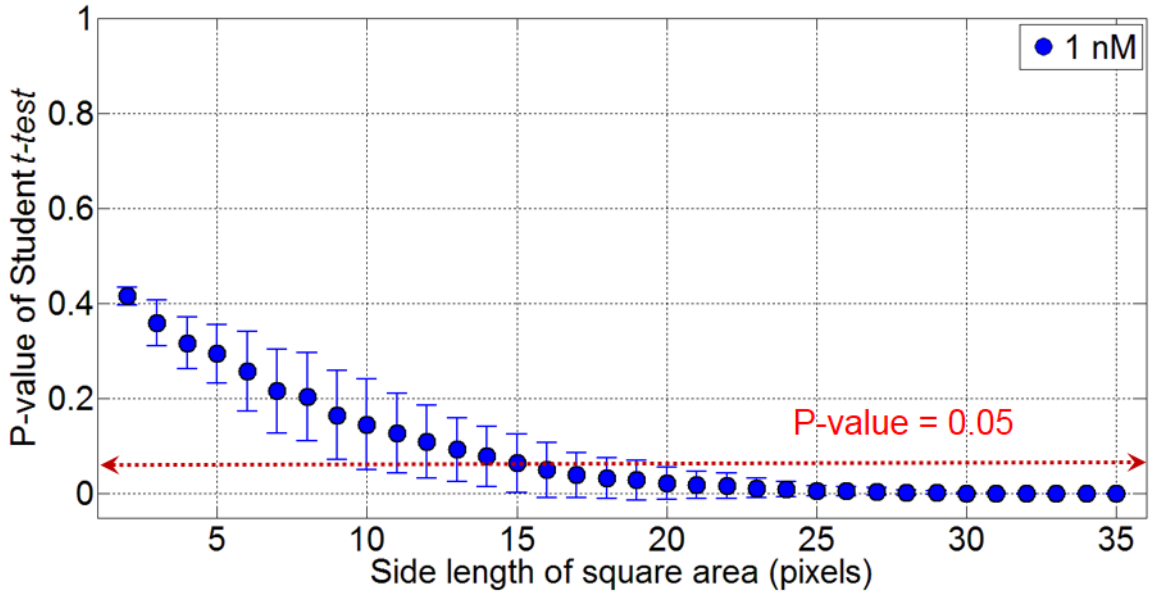

Fig. S1 Minimum mapping area determination using the unpaired Student's *t*-test (STT). The *p*-values of the STT test as a function of square area side length (pixel = 1 μm) of SERS mapping areas collected from heatmaps comparing vasopressin-specific aptamer (treatment) and IgE-specific aptamer (control) functionalized substrates at 1 nM TAMRA-labeled vasopressin concentration. Sampling of about 400 μm^2 area is sufficient enough to discriminate between treatment and control samples with high statistical significance (p = 0.05, red horizontal line). The error bars represent 2 standard deviations of the average *p*-value of 1,000 random samplings for three different SERS substrates.

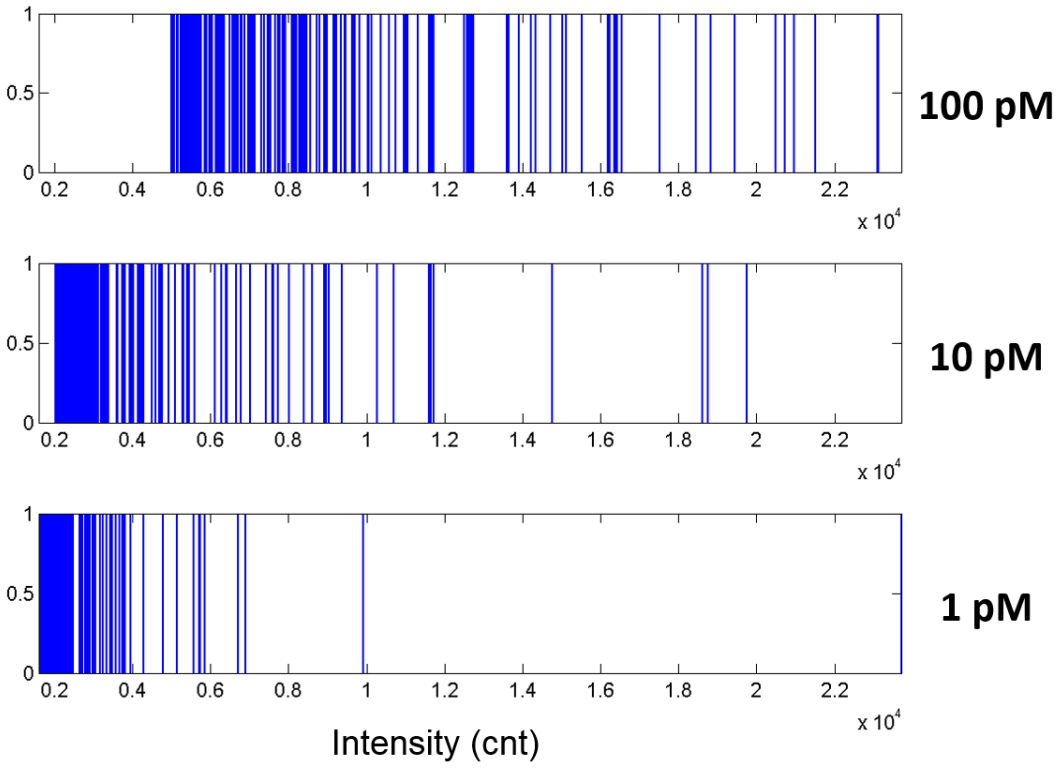

Fig. S2 Stem plots of SERS intensity distribution for the top 5% of hotspots of TAMRA peak 1370 cm^{-1} at 100 pM, 10 pM and 1 pM TVP concentrations.

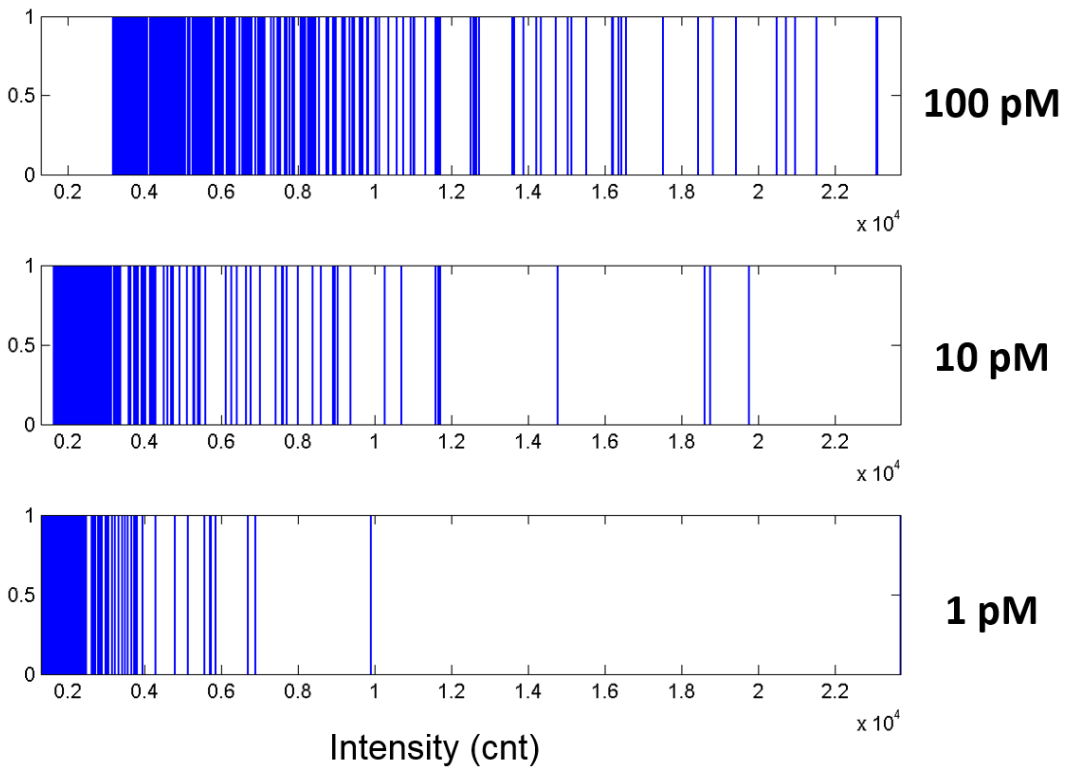

Fig. S3 Stem plots of SERS intensity distribution for the top 10% of hotspots of TAMRA peak 1370 cm^{-1} at 100 pM, 10 pM and 1 pM TVP concentrations.

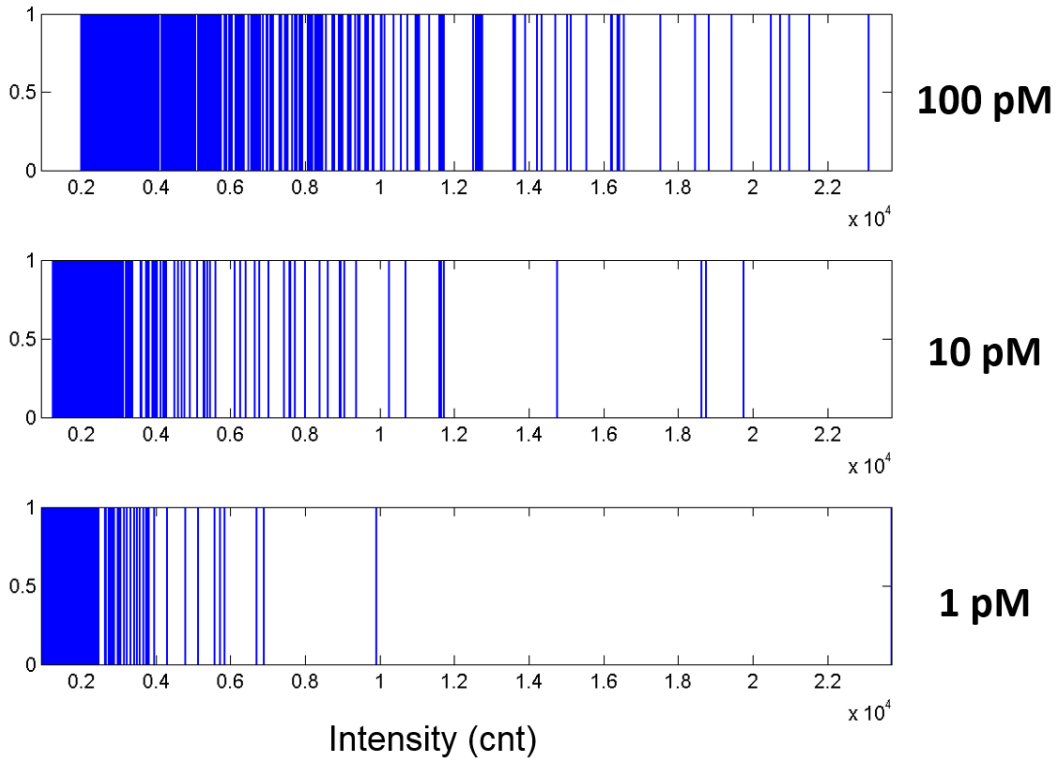

Fig. S4 Stem plots of SERS intensity distribution for the top 20% of hotspots of TAMRA peak 1370 cm^{-1} at 100 pM, 10 pM and 1 pM TVP concentrations.

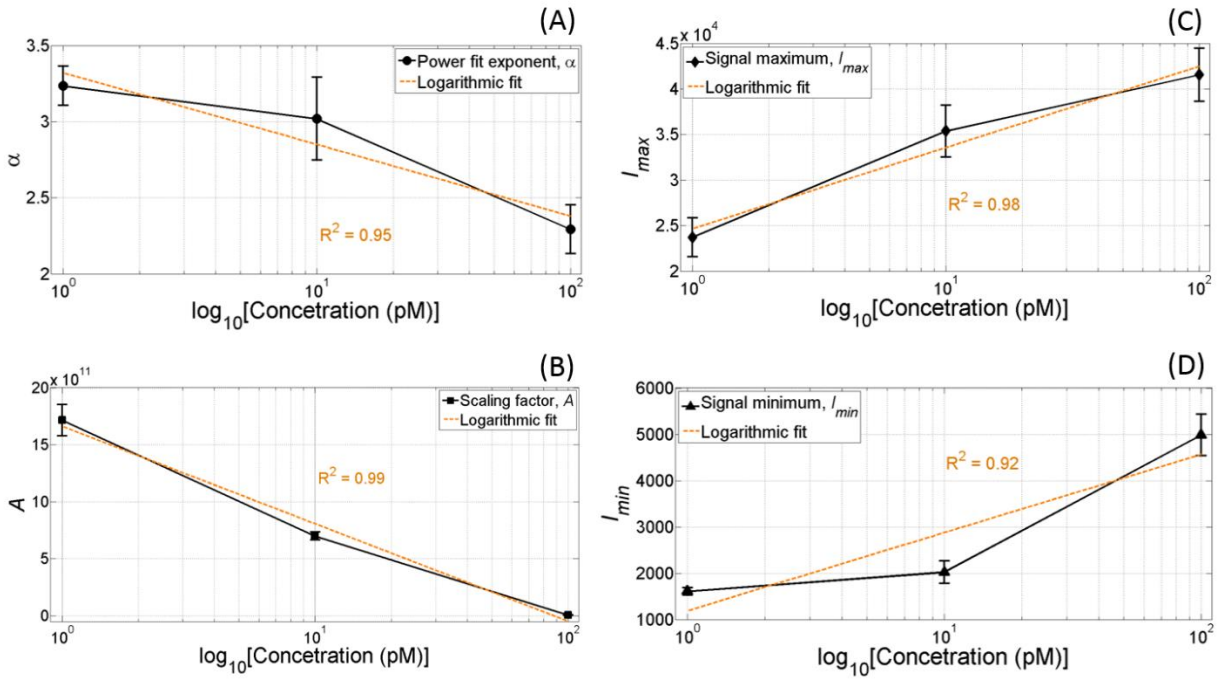

Fig. S5 SERS intensity distribution fitting parameters as a function of TAMRA-labeled vasopressin (TVP) concentration. (A) Power fit exponent (α) and (B) scaling factor (A) of the SERS intensity distribution as a function of analyte concentration for peak 1370 cm^{-1} of TVP. Similar curves are generated for parameters: (C) I_{max} and (D) I_{min} as well. Plots in semi-log scale. The error bars represent standard deviations on three different samples for the same TVP concentration. All fit with a logarithmic function in the general form: $y(x) = a\ln(x) + b$ (dotted orange lines).

REFERECNES

1 D. G. Altman, *Practical Statistics for Medical Research (Chapman & Hall/CRC Texts in Statistical Science)*, Chapman & Hall, 1st ed., 1990.

2 P. Armitage, G. Berry and J. N. S. Matthews, *Statistical Methods in Medical Research*, Wiley, 2001.