Inhibiting Analyte Theft in SERS Substrates: sub-Nanomolar Quantitative Drug Detection

Bart de Nijs[1], Cloudy Carnegie[1], István Szabó[2], David-Benjamin Grys[1], Rohit Chikkaraddy[1], Marlous Kamp[3], Steven J. Barrow[3†], Charlie A. Readman[3], Marie-Elena Kleemann[1], Oren A. Scherman[3], Edina Rosta[2], and Jeremy J. Baumberg[1*]

[1]NanoPhotonics Centre, Cavendish Laboratory, Department of Physics, JJ Thompson Avenue, University of Cambridge, Cambridge, CB3 0HE, United Kingdom

[2]Department of Chemistry, King's College London, 7 Trinity Street, London SE1 1DB, United Kingdom

[3]Melville Laboratory for Polymer Synthesis, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

ESI Section 1:

Following Le Ru *et al*, the SERS contributions as a function of analyte distribution can be extracted from numerical simulations assuming a uniform angular distribution of binding events around the spheres of a dimer, described below:

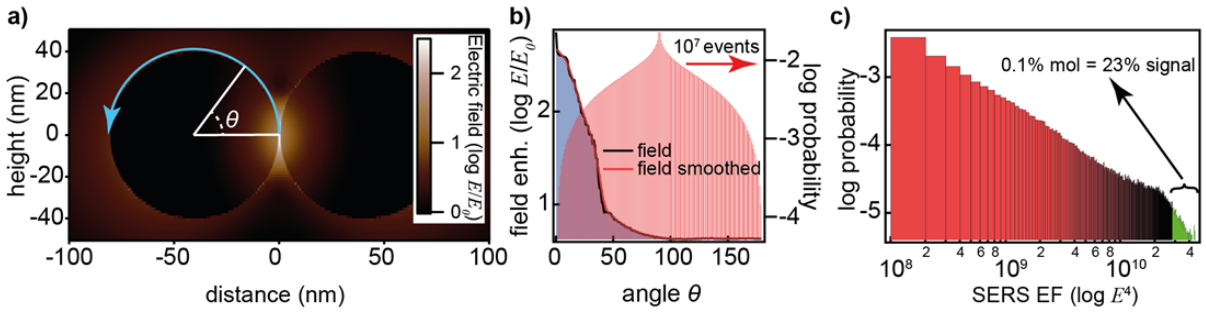

Figure S1: Estimated SERS contributions from a dimer of two 80 nm nanoparticles (NPs) based on numerical simulations. a) FDTD simulation of a dimer of two 80 nm nanoparticles using a 1 nm gap showing field enhancements up to $|\boldsymbol{E}/\boldsymbol{E_0}|$=500. **b**) Black trace is the extracted field enhancement around the NP in **a** as a function of θ, blue trace is the field-enhancement trace smoothed to mitigate the effects of finite resolution in the FDTD simulation. Distribution plot in red shows the angular probability of 10^7 random positions taken on the spheres with respect to the hot-spot. **c**) Probability distribution of different SERS enhancement factors ($\boldsymbol{E^4}$) based on the distribution presented in **b**, showing that based on the FDTD model and a uniform distribution of molecules, 0.1% of the analytes already contribute 23% of the total signal, which is in line with what has previously been shown by Le Ru *et al*.[1]

ESI Section 2:

Binding effects are expected to play a significant role when the ratio of the number of binding sites to the total number of analytes available in the system N_{ratio} starts to approach 1:

$$N_{\text{ratio}} \overset{\lim}{\longrightarrow} 1, \quad (1)$$

with N_{ratio} being defined as

$$N_{\text{ratio}} = \frac{N_{\text{bound}}}{N_{\text{analytes}}}, \quad (2)$$

and N_{analytes} and N_{bound} are calculated using the following expressions:

$$N_{\text{analytes}} = M_{\text{analytes}} \times V \times N_{\text{A}}, \quad (3)$$

$$N_{\text{bound}} = \frac{C_{\text{AuNPs}} \times V \times 4\pi r^2 \times \varphi}{K_{\text{occupancy}} \times A_{\text{binding site}}} \quad (4)$$

In eq. (3-4) V is the total volume, $\text{M}_{\text{analytes}}$ the concentration of analytes, N_{A} Avogadro's constant, C_{AuNPs} the concentration of nanoparticles, r the radius of the nanoparticles, φ the packing density of binding sites per area, $\text{K}_{\text{occupancy}}$ a factor to describe the number of sites that are occupied by an analyte, and $\text{A}_{\text{binding site}}$ the area taken up by a binding site.

With a AuNP concentration of $C_{\text{AuNPs}} = 2.6 \cdot 10^{13}$ particles per litre, a radius of $r = 40\ \text{nm}$, a packing fraction of $\varphi = 0.5$ and an occupancy of $K_{\text{occupancy}} = 0.50$, N_{ratio} reaches 1 at roughly 1 µM. This means for any lower concentrations the analyte 'theft' of the substrate has to be taken into account.

ESI Section 3:

DFT calculations were performed to predict the binding energetics for each THC@CB[n], n = 5, 6, 7, and 8 complex. The Gibbs free energy of binding is calculated as:

$$\Delta G_{\text{bind}}^{\text{RRHO/QH}}(\text{l}) = \Delta G_{0}^{\text{RRHO/QH}}(\text{l}) + \delta E^{\text{CP}}(\text{g}), \quad (5)$$

and the enthalpy gain is analogously obtained as:

$$\Delta H_{\text{bind}}^{\text{RRHO}}(\text{l}) = \Delta H_{0}^{\text{RRHO}}(\text{l}) + \delta E^{\text{CP}}(\text{g}), \quad (6)$$

where $\boldsymbol{\delta E^{\text{CP}}(\mathbf{g})}$ denotes the counterpoise correction, and $\boldsymbol{\Delta G_{0}^{\text{RRHO/QH}}(\mathbf{l})}$ and $\boldsymbol{\Delta H_{0}^{\text{RRHO}}(\mathbf{l})}$ are the binding Gibbs free energy and enthalpy, respectively prior to the correction for the basis superposition error (BSSE). The results for each of the complexes are given in Table S1, with (**g**) and (**l**) denoting gas phase (vacuum) and liquid with implicit water molecules as the simulated environment respectively.

Table S1. Calculated binding Gibbs free energies and enthalpies (in kcal·mol^{-1}) at the B3LYP/6-31G*+GD3BJ level of theory in SMD implicit water for each of the THC@CB[n] complexes, using the rigid-rotor harmonic oscillator approximation (RRHO), as well as the quasi-harmonic approximation (QH) following Grimme[2] where the translational entropy was also corrected.

Complex	$\Delta E(\text{g})$[b]	$\delta E^{\text{CP}}(\text{g})$	$\Delta H_{\text{bind}}^{\text{RRHO}}(\text{l})$	$\Delta G_{\text{bind}}^{\text{RRHO}}(\text{l})$	$\Delta G_{\text{bind}}^{\text{QH}}(\text{l})$[c]
THC@CB[5]	−4.4	11.6	22.7	37.2	33.4
THC@CB[6]	−40.0	16.1	−17.4	−6.1	−10.6
THC@CB[7] (i)[a]	−42.7	17.9	−28.1	−12.7	−18.7
THC@CB[7] (ii)[a]	−40.6	15.4	−30.9	−16.2	−21.9
THC@CB[8]	−41.3	16.2	−22.8	−18.9	−20.7

[a]The two complexes slightly differ in the THC conformation, but have virtually the same depth of penetration into the CB[7] cavity.
[b]Gas phase potential energy in kcal·mol^{-1} at the B3LYP/6-31G*+GD3BJ level of theory.
[c]The frequency cutoff applied in Grimme's quasi-harmonic approximation was set to 100 cm^{-1}.

ESI section 4: Principal component analysis

Principal component analysis (PCA) is a powerful tool in characterising spectral changes in Raman spectroscopy as it takes into account the entire spectrum. It is therefore able to pick up on changes in

spectra which are otherwise barely distinguishable when examining individual peaks. We use a version of PCA on a dilution series of analytes, to extract the LOD. For a typical experiment here, a stock analyte solution is made in an appropriate carrier solvent (water for MV^{2+}, methanol for cannabinoids), and the stock solution is sequentially volumetrically diluted resulting in an analyte concentration series. For each analyte concentration, 5 new aggregate solutions were prepared by mixing 5 µL of 1 mM CB[n] (in this case n=5) stock solution with 220 µL of AuNPs solution and allowed to self-assemble for 10 minutes, after which 20 µL of analyte solution was added (*ie.* 1 unique sample realisation). Between each block of 5 sample realisations two control measurements with only carrier solvent were measured to test for consistency. At the end of each series five samples were measured without analyte or carrier solvent added (Figure S2a).

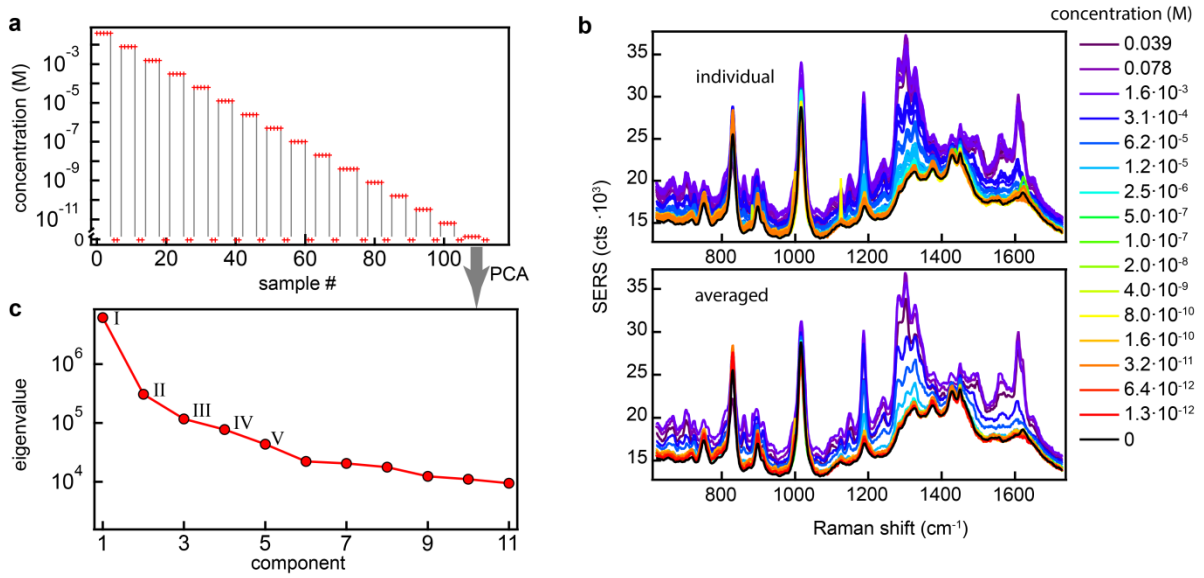

Figure S2: Experimental data acquired for principal component analysis (PCA). a) Analyte concentration series for measuring SERS spectra with PCA analysis. b) (top) Individual SERS spectra for analyte (**2**), colour denotes concentrations. (bottom) SERS spectra at same concentration averaged together for better visualisation showing clear trends with analyte concentration. c) Eigenvalues obtained from PCA analysis on the dataset for each of the components generated.

In all experiments a clear drop in analyte peak intensities is observed with the reduction of analyte concentration. Raw spectra show characteristic peaks around 1300 cm^{-1}, that drop with the reduction in analyte concentration (Figure S2b,top), visualized more clearly by averaging the 5 individual measurements for each analyte concentration (Figure S2b, bottom). The Igor implementation of principal component analysis by WaveMetrics was used to process the set of raw spectra into a combination of eigenvectors (components) and eigenvalues (Figure S2c). The typical decrease in eigenvalues with increase in component number is observed. Plotting the PCA scores for each of the analytes *vs* their concentration (here labelled by sample number corresponding to Figure S2a), clear trends emerge (Figure S3a). For comp I, II, IV an undulating pattern is observed with sample number matching the artificially-generated periodic change in analyte concentration (Figure S2a). This makes evident that comp I, II, IV are related to the probed analyte. For comp III we see a sudden rise in score for the last 5 sample numbers where no carrier solvent methanol was added. Finally, the small comp V scores show both matching and non-matching behaviour (partly linked to comp IV). As the eigenvalues decrease with greater component number, their contributions become less relevant and more complicated. Comparing the loading plots (Figure S3b, red), a CB[5]:AuNP spectrum is identified

as comp I but with an additional peak at 1000 cm^{-1} and peaks around 1300 cm^{-1}. These extra peaks are a result of comp II and III contributing to the loading plot of comp I. This mixing of the states is inherent to PCA analysis and requires post processing, or 'rotation', to separate the components by enforcing a known condition which is the spectrum without any analyte.

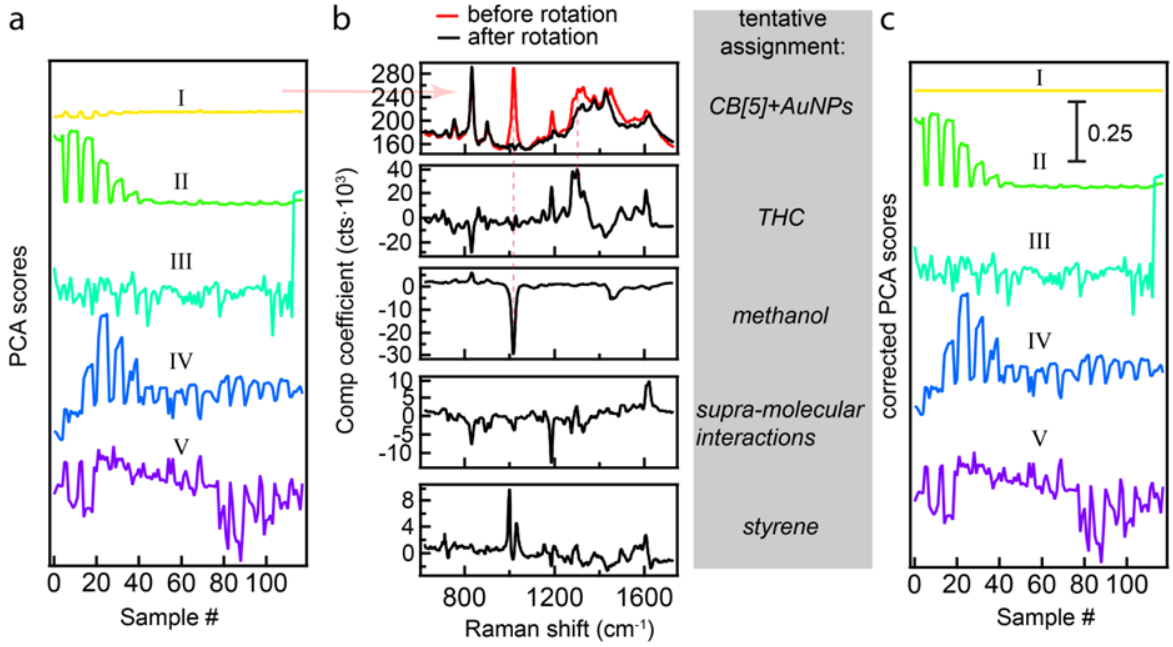

Figure S3: PCA results and rotation for analyte (2). a) PCA scores plotted versus sample number. b) Corresponding loading plots for component with comp I shown before and after rotation. c) PCA scores for comp I to V after rotation plotted vs sample number.

To remove the contribution of comp II to the loading plot of comp I when no analyte is present, we use:

$$E_{\mathrm{I}}' = E_{\mathrm{I}} - E_{\mathrm{II}} \times \frac{C_{\mathrm{II}}(0\%)}{C_{\mathrm{I}}(0\%)}$$

which gives the black comp I curve in Figure S3b (top), and where E_{I} and E_{II} are the loading plots of comp I and comp II respectively, and $C_{\mathrm{I}}(0\%)$ and $C_{\mathrm{II}}(0\%)$ the component scores for the sample with 0% analyte. The rotation performed for the comp scores follows:

$$c_{\mathrm{II}}'(m) = c_{\mathrm{II}}(m) - c_{\mathrm{I}}(m) \times \frac{C_{\mathrm{II}}(0\%)}{C_{\mathrm{I}}(0\%)}$$

Hence for comp II specifically, the score corresponding to analyte concentration m=0% is forced to be 0. The resulting transformed scores are shown in Figure S3c (black), with comp I nearly constant and comp II following the analyte concentration closely. Close examination of the other loading plots now reveals for comp III a resemblance to methanol.[3] Comp IV shows a relation to the analyte concentration and shows a peak in the region of 1600 cm^{-1}, this we tentatively attribute to hydrogen bonding or other forms of supra-molecular interaction, as discussed in [3]. Comp V shows a double peak around 1000 cm^{-1} which is very similar to styrene, a polymer commonly used in laboratory consumable such as pipette tips, multi-well plates and cuvettes, and likely a source of contamination though of minor significance compared to the signal from the analytes (Figure S2c). To normalise the PCA scores and extract a spectral response, the scores were multiplied by the loading plot normalised for power (P) and spectral integration time (T), using:

$$value_c = score_c \cdot \sum_{m=\lambda} \frac{|\text{comp II}_\lambda|}{P \cdot T}$$

ESI section 5: comp II and analyte identification

The obtained comp II from each analyte concentration series can be used as a fingerprinting technique to identify the types of component present in mixtures. Using a simple Pearson correlation between the raw unprocessed spectra and the comp II for each analyte vs concentration shows the analyte can be clearly distinguished above concentrations higher than 10^{-8} M (Figure S4).

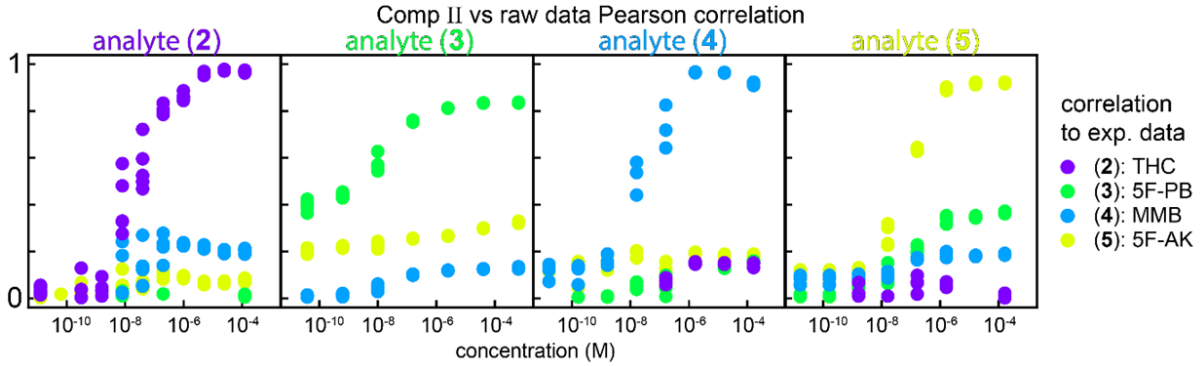

Figure S4: Pearson correlations between raw unprocessed spectra and different comp II spectra. The plots show each components is clearly distinguishable from its sister compounds once the analyte concentration reaches 10^{-8} M or higher.

In addition to fingerprinting correlation, a comparison can be made between the comp II spectra and their respective bulk powder Raman spectra (Figure S5), offset for clarity.

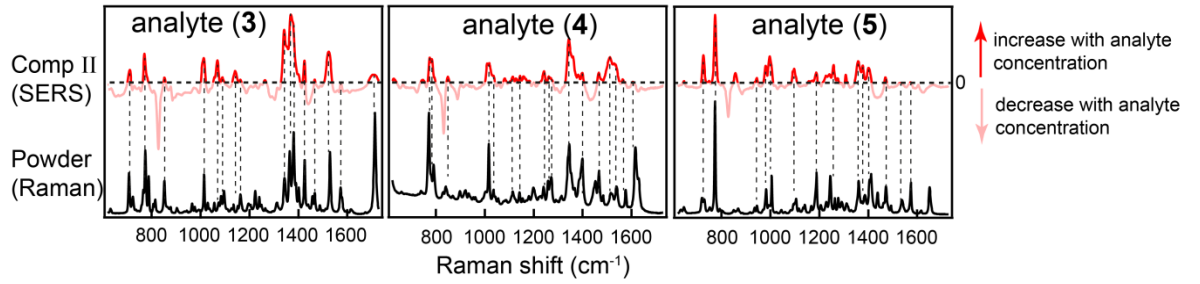

Figure S5: Comparisons between comp II (SERS) and bulk powder Raman spectra for each of the synthetic cannabinoids measured. Distinct peaks from the powder spectra can be clearly recognised in the SERS spectra, while some smaller peaks have shifted or change intensity ratios. A notable difference is the increased width of the SERS lines with respect to the Raman lines resulting in peaks with narrow separations showing up as shoulders in the SERS spectra.

For the majority of the peaks from each compound a good correlation can be made between the PCA comp II (SERS) and the powder Raman spectra. There are however, a couple of notable differences with slight peaks shifts (expected in SERS) and ratio changes in peak intensities. In particular in the region 1100 cm^{-1} and 1250 cm^{-1} of analyte (**4**) the peak ratios for the smaller peaks differ significantly between the comp II and the Raman.

ESI Section 6: Reproducibility and noise

The strength of this self-assembled SERS aggregation and measurement technique lies in its extremely high reproducibility. Small variations around 1000 cm^{-1} arise from trace amounts of styrene

contamination leached from the plastic well plates (Figure S6a). This allows for the background (CB[n]AuNP aggregates without analyte) to be reliably subtracted from the raw data to isolate spectral changes, see Figure 7 of the main text. The level of noise in our spectra was estimated by taking a background region of the spectrum without peaks (above 1600 cm^{-1}), fitting a sigmoidal curve, and taking the residuals as noise (Figure S6b).

To estimate the noise threshold, the standard deviation of the residuals was used giving 0.03 $cts \cdot mW^{-1} \cdot s^{-1}$.

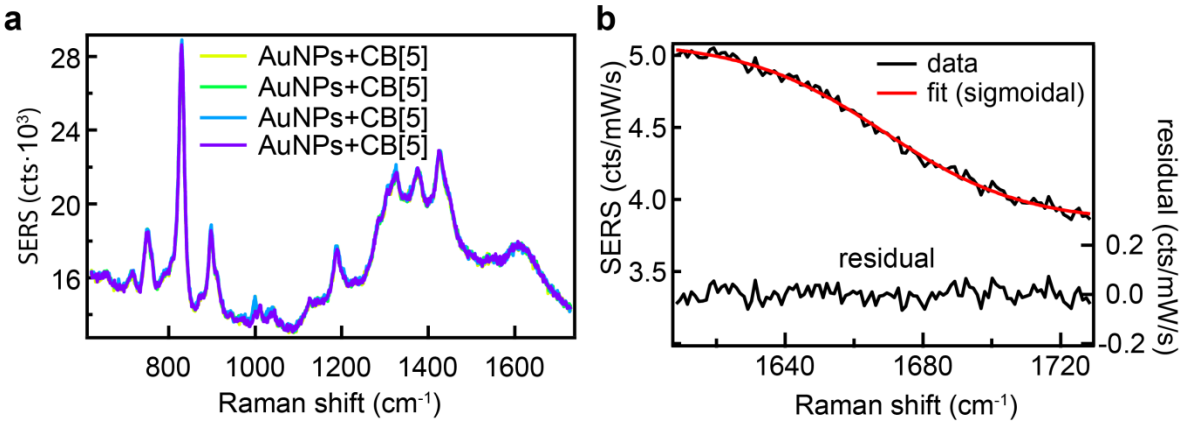

Figure S6: Experimental determination of the LOD. a) Four different background scans plotted on top of each show near identical spectra. Small variations around 1000 cm^{-1} arise from trace amounts of styrene contamination leached from the plastic well plates. b) Region of the spectrum without peaks fitted to a sigmoidal curve, with residuals giving the noise level.

Table S2. Hill-Langmuir fit parameters: K_D and N.

Compound	K_D	N
THC(2):CB[5]	$5.8 \cdot 10^{-7} \pm 0.77 \cdot 10^{-7}$	0.53±0.03
5F-PB-22(3)	$13.5 \cdot 10^{-7} \pm 3.95 \cdot 10^{-7}$	0.34±0.05
MMb-CHMICA(4)	$2.42 \cdot 10^{-7} \pm 1.96 \cdot 10^{-7}$	0.56±0.20
5F-AKB48(5)	$1.94 \cdot 10^{-7} \pm 0.27 \cdot 10^{-7}$	2.53±0.001

The Hill-Langmuir fits for each of the components returned a K_D value around 10^{-7}, and for 3 out of 4 a Hill coefficient between N=0.3 and 0.6. However, a clearly different Hill coefficient of N=2.53 is found for compound (**5**) suggesting a different binding behaviour from the other compounds. This shows that understanding the difference in binding behaviour can help bring down the limits of detection. Elucidating the origin of this change in N and how to modify this is subject to ongoing research.

REFERENCES

(1) Le Ru, E. C.; Etchegoin, P. G.; Meyer, M. Enhancement Factor Distribution around a Single Surface-Enhanced Raman Scattering Hot Spot and Its Relation to Single Molecule Detection. *J. Chem. Phys.* **2006**, *125* (20), 204701.

(2) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465.

(3) Nijs, B. de; Kamp, M.; Szabó, I.; J. Barrow, S.; Benz, F.; Wu, G.; Carnegie, C.; Chikkaraddy, R.; Wang, W.; M. Deacon, W.; et al. Smart Supramolecular Sensing with Cucurbit[N]Urils: Probing Hydrogen Bonding with SERS. *Faraday Discuss.* **2017**, *205* (0), 505–515.