



**QUEEN'S
UNIVERSITY
BELFAST**

Quantitative Surface-enhanced Raman Spectroscopy of Single Bases in Oligodeoxynucleotides

Bell, S., & Dick, S. (2017). Quantitative Surface-enhanced Raman Spectroscopy of Single Bases in Oligodeoxynucleotides. *Faraday Discussions*. Advance online publication. <https://doi.org/10.1039/C7FD00134G>

Published in:
Faraday Discussions

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© The Royal Society of Chemistry 2017. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>



Quantitative Surface-enhanced Raman Spectroscopy of Single Bases in Oligodeoxynucleotides

S. Dick^a and S. E. J. Bell^a

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

To address the question of whether the SERS signals of ss-DNA are simply combinations of the signals from individual bases that comprise the sequence, SERS spectra of unmodified ss-DNA sequences were obtained using hydroxylamine-reduced Ag colloid aggregated with MgSO₄. Initially, synthetic oligodeoxynucleotides with systematic structural variations were used to investigate the effect of adding single nucleobases to the 3' terminus of 10-mer and 20-mer sequences. It was found that the resulting SERS difference spectra could be used to identify the added nucleobases since they closely matched reference spectra of the same nucleobase. Investigation of the variation in intensity of an adenine probe which was moved along a test sequence showed there was a small end effect where nucleobases near the 3' terminus gave slightly larger signals but the effect was minor (30%). More significantly, in a sample set comprising 25-mer sequences where A, T or G nucleobases were substituted either near the centre of the sequences or the 5' or 3' ends, the SERS difference spectra only matched the expected form in approximately half the cases tested. This variation appeared to be due to changes in secondary structure induced by the altering the sequences since uncoiling the sequences in a thermal pre-treatment step gave difference spectra which in all cases matched the expected form. Multivariate analysis of the set of substitution data showed that 99% of the variance could be accounted for in a model with just three factors whose loadings matched the spectra of the A, T, G nucleobases and which contained no positional information. This suggests that aside from the differences in secondary structure which can be eliminated by thermal pre-treatment, the SERS spectra of the 25-mers studied here are simply the sum of their component parts. Although this means that SERS provides very little information on the primary sequence it should be excellent for detection of post-transcription modifications to DNA which can occur at multiple positions along a given sequence.

Introduction

Surface enhanced Raman spectroscopy has several attributes that make it potentially valuable for RNA and DNA analysis, it is suitable for dilute aqueous samples, has good sensitivity (detection limits of 10⁻⁸ and 10⁻⁹ mol dm⁻³ have been reported for ss-DNA and ds-DNA, respectively)¹ and it can be carried out with microliter samples,² reducing the total sample to the picomole or femtomole range. Moreover, SERS does have chemical specificity which means that it can be used either to detect a chemically distinct label or RNA/DNA directly. Most current applications of SERS for DNA analysis involve development of assays in which the structure of the target sequence is detected through hybridisation with a complimentary strand and it is the hybridisation event which is detected, typically by detecting the SERS signal from a label on the complimentary strand.^{3,4} This

general approach is highly selective and sensitive and has been used to create clinically relevant assays.⁵ Conversely, direct label-free analysis of DNA or RNA, although it has the advantage of simplicity, is not expected to rival hybridisation-based techniques for sequence determination because of its low positional sensitivity. However, label-free methods do have the distinct advantage that they are intrinsically sensitive to chemical modifications of DNA, since chemical modification of nucleobases causes characteristic changes in their Raman spectra.⁶⁻¹⁴ For example, it has been shown that substitution of even a single base in 25-mer oligodeoxynucleotides (ODNs) can be detected using SERS difference spectroscopy.⁶ This sensitivity to changes in chemical structure means that SERS may be also used to detect post-transcription modifications of nucleobases, which are both important from a physiological/clinical viewpoint and difficult to detect through hybridisation assays. For example, Halas et al.⁷ have been able to detect methylation of adenine in 12-mer ODNs and Guerini et al.^{12,13} have also detected methylation of adenine and cytosine with similar sensitivity.

Although most studies have concentrated on well-defined synthetic ODNs where the objective is to explore the basic features of the approach, there are some studies which have gone directly to

School of Chemistry and Chemical Engineering, Queen's University, Belfast, BT9 5AG, U.K. E-mail: S.Bell@qub.ac.uk.

Electronic Supplementary Information (ESI) available: Illustration of spectral subtraction procedure, data illustrating reproducibility of spectra and spectral differences. See DOI: 10.1039/x0xx00000x

analysis of much more complex samples. For example, El-Sayed et al¹⁵ have observed large spectral changes when they used reactive oxygen species to create significant ds-DNA scission and oxidation in their samples. Driskel et al have studied mixtures of microRNAs¹⁶ while Chen et al¹⁷ have discriminated gastric cancer from normal serum based on the SERS spectra of extracted RNA.

In this work we examine some of the factors which determine the limits of label-free SERS of RNA and DNA, looking in particular at the question of whether the SERS spectra of DNA are simply the sum of their parts i.e. whether the total signal is simply the population weighted sum of the spectra of the individual components. In addition, we seek to establish whether there is any positional component within the SERS signals that might either be used to add to the information extracted from the SERS spectra or which might confound attempts to find a simple link between changes in spectra and changes in structure. For example, it has been shown that under some conditions the signals from nucleobases sited away from the ends of ODNs are too weak to be detected.¹⁸ If we are to use SERS to monitor nucleobase modification it is important to know which regions of the strands can be probed. Similarly, it is essential that we understand the extent of modification which can be detected by SERS and the factors which set this limit. Although these studies look forward to future studies using modified nucleobases, here we have used synthetic ODNs based on unmodified bases since these allow us to explore the main issues without introducing additional complexity by incorporating modified bases into the test set. Similarly, although there are now several methods to aggregate colloids which are appropriate for studies of ODNs¹³ we have used our original method based on silver colloid aggregated with MgSO₄¹⁹ in part because it allows us to compare with previous studies and also because it complements work on the literature which use other approaches, in particular Gerrini et al who use positively-charged spermine-modified colloids¹¹⁻¹⁴ rather than the negatively-charged colloids used here.

Experimental

Silver nitrate (99.9999%), hydroxylamine hydrochloride (99%), sodium hydroxide (≥97%) and magnesium sulfate were all purchased from Sigma-Aldrich. Aqueous solutions were prepared using distilled deionized (DDI, resistivity 18.2 MΩ cm) water from a Barnstead NANOpure Diamond System. The hydroxylamine hydrochloride silver colloids were prepared according to the Leopold and Lendl method and had a $\lambda_{\text{max}} = 420 \text{ nm}$.²⁰

All oligodeoxynucleotide sequences were purchased from Eurogentec Ltd. (Belgium) and were purified using the Selective Precipitation Optimised Process (SePOP) by the vendor. For SERS measurements, 50 μL of colloidal solution was mixed with 50 μL of analyte and then aggregated by 25 μL 0.1 M magnesium sulfate. Spectra were recorded using an Avalon Instruments RamanStation R1 which is fitted with a 785 nm diode laser and an echelle spectrograph. In this system the laser power at the sample was 160

mW and samples were held in 96 well polypropylene multiwell plates.

Multivariate data analysis was carried out using the PLSplus IQ (ThermoGalactic) module in GRAMS/AI v7.02 software. The only pre-processing function that was used was the Savitzky-Golay first derivative (2nd order, 11 points), which suppresses changes in the background without greatly distorting the signal.

Results and Discussion

The SERS spectra of nucleobases in oligodeoxynucleotides (ODNs) are known to be different from those of the free bases.^{21,22} In addition, the spectra reported in the literature show some quite striking differences from each other, partly through inter-laboratory variation but also because the detailed preparation method of the enhancing material does affect the spectra.¹³ Therefore the spectra of three of the four relevant polynucleotides obtained under our experimental conditions are shown for reference in Figure 1. The Poly G spectrum could not be recorded due to the tendency of 4 or more adjacent G nucleotides in a sequence to form guanine tetraplexes, which show poor adsorption to the metal colloid, so here the G Dimer was used instead. Also included in the Figure is the SERS spectrum of a typical synthetic ss-DNA sample, this was included to illustrate the common observation that the signals from adenine are significantly stronger than those of the other bases in the sequence. In this case the adenine bands dominate the spectrum, despite the fact that there are only 4 adenines in this 25 base ODN. However, bands due to the other bases can still be discerned, albeit with lower intensity than the strong adenine peaks.

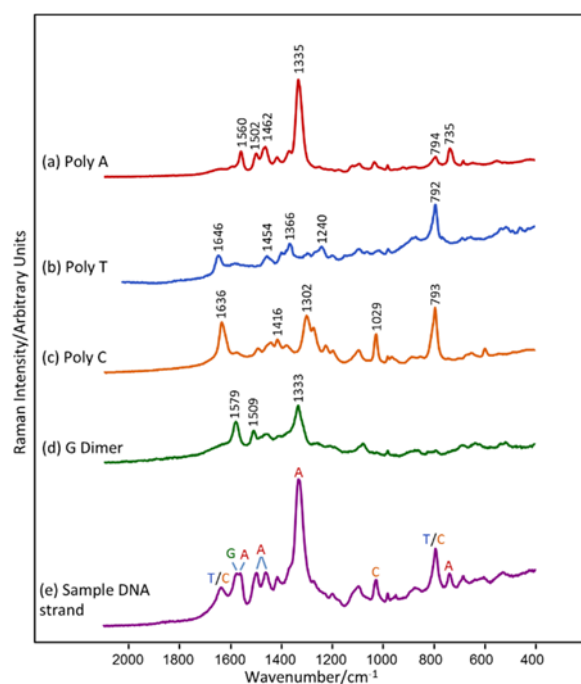


Figure 1. Reference SERS spectra of the 4 main DNA nucleotides, (a) polyadenosine, Poly A, (b) polythymidine, Poly T (c) polycytosine, Poly C and (d) 2'-OME-guanosine dinucleotide, G Dimer. These are compared with (e) the SERS spectrum of a ss-DNA sequence, CTT-TTT-CCT-GCA-TCC-TGT-CTG-GAA-G.

Single Base Extensions to ODN Sequences.

The first series of experiments were carried out to determine if SERS could be used to characterise the addition of nucleotides to ODN sequences, this is reminiscent of DNA sequencing where ODN sequences with single base extensions are measured.²³ The ODN sequences, listed in Table 1, allowed the addition of each of the 4 nucleobases in turn to ODNs to be investigated. Starting sequences with both 10 and 20 bases were used.

ODN code	Sequence (5' → 3')
OA-10-G	GCA-TAC-CGT-G
OA-11-A	GCA-TAC-CGT-GA
OA-12-T	GCA-TAC-CGT-GAT
OA-13-C	GCA-TAC-CGT-GAT-C
OA-14-G	GCA-TAC-CGT-GAT-CG
OA-20-G	GCA-TCC-TGT-CTG-GAA-TAC-CG
OA-21-A	GCA-TCC-TGT-CTG-GAA-TAC-CGA
OA-22-T	GCA-TCC-TGT-CTG-GAA-TAC-CGA-T
OA-23-C	GCA-TCC-TGT-CTG-GAA-TAC-CGA-TC
OA-24-G	GCA-TCC-TGT-CTG-GAA-TAC-CGA-TCG

Table 1. The ODNs used to test detection of additional nucleobases. The codes give the number of bases in the sequence and the base at the 3' terminus.

The SERS spectra of the 10-14 mer series are shown below in Figure 2, the longer series gave similar results. Data were obtained at both 10^{-4} mol dm⁻³ and 10^{-5} mol dm⁻³ but the increased concentration did not significantly improve the signal/noise ratio so 10^{-5} mol dm⁻³ samples were used for subsequent experiments. Since each sample differed from the preceding one by just a single base, the overall change in composition was small, which meant that the effect of adding the bases was not obvious in the raw spectra. To highlight the changes in the spectra the data were processed by subtracting the spectrum of the previous sample in the sequence. Since the aim was to show only the additional component given by each added nucleobase, the intensity of the subtracted spectrum was scaled by a factor that was just sufficient to give a resultant trace with no negative bands. The process is illustrated in Figure S1. The resulting difference spectra are shown below in Figure 3, where they are compared to the appropriate reference spectra for each base addition. In this Figure the area around 1335 cm⁻¹ has been blanked because there is interference from poorly subtracted adenine bands in this region. This is difficult to avoid since the high intensity of the adenine signal at this point means that even a small error in the adenine subtraction gives a noticeable signal which, although it does not interfere with the other bands in the spectrum, is distracting and therefore best removed.

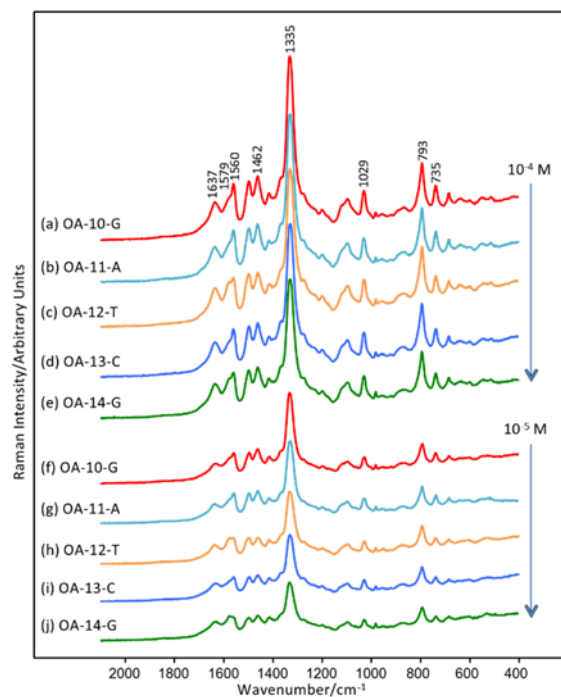


Figure 2. Raw SERS spectra of ODN sequences showing the effect on the spectra of adding nucleobases at the 3' end. Spectra were recorded using ODN concentrations 10^{-4} M and 10^{-5} M.

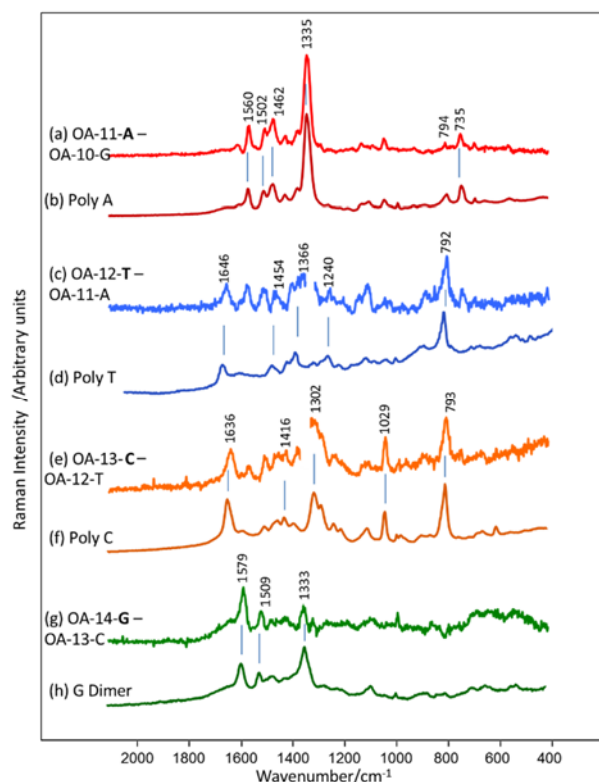


Figure 3. Comparison of the SERS difference spectra of ODNs which differ by a single bases with the appropriate reference spectra. (a), (c), (e) and (g) are experimental difference spectra corresponding to extension of initial ODN by A, T, C and G, respectively. (b), (d), (f), (h) are reference spectra of the same nucleobases.

With this scaled spectral subtraction, the difference spectra show only the additional intensity created by extension of the ODN and it is clear that these signals match those of their corresponding polynucleotides well. The addition of A gives a difference spectrum with a strong additional new band at 1335 cm^{-1} along with several other bands characteristic of adenine. However, after the addition of T the difference spectrum shown in (c) displays peaks at 1646 cm^{-1} , 1454 cm^{-1} and 792 cm^{-1} , matching closely to the spectra of (d) Poly T. The same is the case for (e) the addition of C where the band at 1302 cm^{-1} can now be seen clearly when the adenine band at 1335 cm^{-1} is removed. This corresponds well to (f) Poly C, along with the presence of characteristic bands at 1636 cm^{-1} , 1029 cm^{-1} and 793 cm^{-1} . Similar comparisons can be seen after the addition of G. Overall, these data appear to show that the signals due to the constituent bases in the spectra of these ODNs are simply additive; the addition of one base to a sequence can be detected and identified through the additional contribution to the overall signal that it produces. This effect was also observed for the 20- 24-mer samples whose raw spectra were shown in Figure 3.

This process mimics the Sanger sequencing method, where DNA sequences of discrete sizes are separated (e.g. by gel electrophoresis) and then read off in base sequence. Of course Sanger sequencing usually deals with DNA fragments with 300-1000 bases and so SERS will not be able to compete with normal sequencing techniques but the significance of the results above is that they support the idea that the SERS signals of DNA are simply a combination of the intrinsic signals of the individual nucleotides present. Interestingly, the Sanger technique produces poor quality sequencing in the first 15- 40 bases of the sequence due to primer binding, which is the sequence length where SERS works best.

End Effects

The data above show only the effect of extension at the 3' terminus of the sequences so it is important to test if the same approach might work when the bases probed are away from the 3' terminus. Intuitively it might be expected that end groups would be better at accessing a SERS enhancing surface than the nucleobases in the middle of a sequence. Moreover, it has been shown using "molecular rulers" (thiol terminated ODN sequences composed of adenine tracts in Poly C), that SERS enhancement of the nucleotide reduces the further it moved from the end of the chain (to a maximum of 8 nucleotides from the terminal position where the nucleotide was no longer enhanced at all).¹⁸ To test the possibility that there are end effects under the current experimental conditions, ODNs based on Poly C but with a single adenine at different positions along the sequence (see Table 2) were studied. An adenine probe was used because of its strong scattering which makes detecting its signal among those of the much larger number of cytosine bases easier than would be the case for other bases. In the resulting spectra (Figure 4(i)) the adenine bands at 1335 cm^{-1} and 735 cm^{-1} , although weaker than the bands of the much larger number of cytosine bases, are indeed clearly visible, indicating that the A nucleotide can be observed even when it is in the middle of the sequence. Figure 4(ii),

Code	Sequence (5' → 3')
OR-10-A1	CCC-CCC-CCC-A
OR-10-A2	CCC-CCC-CCA-C
OR-10-A3	CCC-CCC-CAC-C
OR-10-A4	CCC-CCC-ACC-C
OR-10-A5	CCC-CCA-CCC-C
OR-10-A6	CCC-CAC-CCC-C

Table 2. The ODNs used to test position sensitivity of a single adenine base in Poly C 10-mer ODNs.

a plot of the absolute intensities of the adenine signals against position in the sequence, shows that there is a small end effect in that the absolute intensity of the adenine signal is reduced by approximately 30% in moving 2 bases in from 3' end of the sequence but that it then remains constant. This result

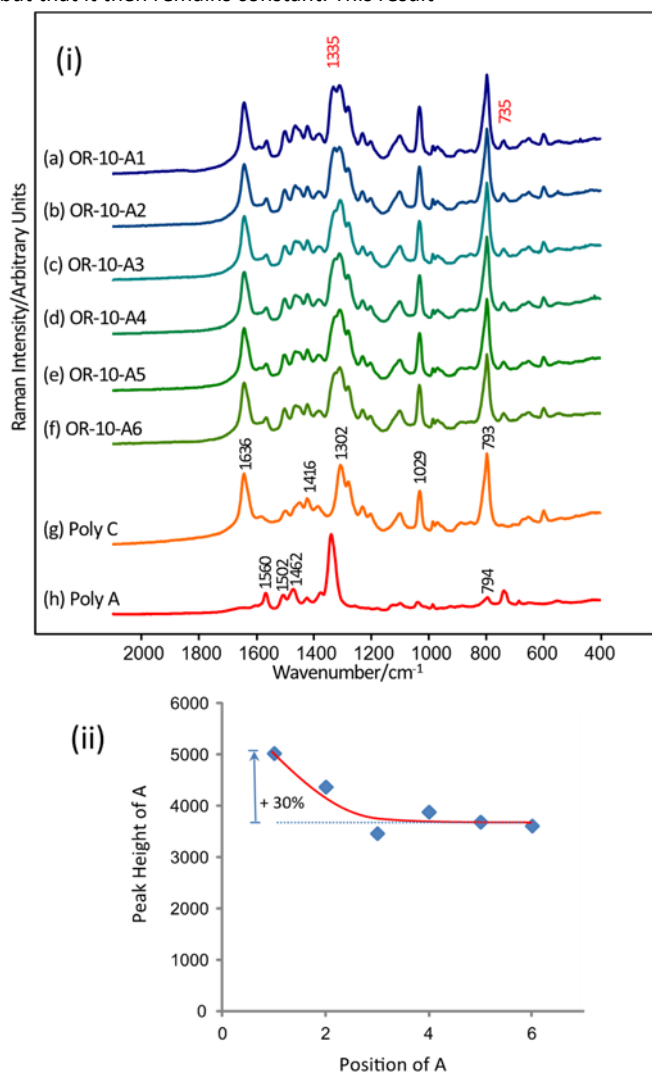


Figure 4. (i) SERS spectra of (a)-(f) ODNs where the position of a single adenine in a Poly C sequence is systematically moved along the sequence. The SERS spectra of (g) Poly C and (h) Poly A are included for comparison. (ii) Plot of the absolute intensity of the adenine band at 1335 cm^{-1} against the position of the base in the ODN.

contrasts with a previous study which reported a significant change between spectra of samples where the nucleotides were positioned at the end versus the middle of the ODN sequences.¹⁸ However, in that work the strands were thiolated and therefore the ODNs would interact with silver nanoparticles via the sulfur of the thiol group, causing the strands to be upright, pointing out from metal surface. Under such conditions it would be expected that moving away from the end also means moving away from the enhancing surface. However, in the current experiments the ODNs were not thiolated, so the strands would be expected to lie flat on the surface to maximise the non-specific binding interaction of the ODN with the metal surface. This should minimise end effects but, as shown above, it does not appear to completely eliminate them, presumably because even in a flat orientation the ends have greater conformational freedom than the middle of the sequence and so may be able to access slightly more favourable orientations.

Single Base Substitutions in ODN Sequences

The next step up in complexity was to detecting substitutions of a single nucleotide at different positions in 25-mer ODNs. Previous work has shown that a single base substitution at the 3' end of 2 different 25-mer sequences gives characteristic changes which allow the bases involved to be identified,⁶ similarly a 5' substitution in a 10-mer ODN has been detected¹⁰ and PLS has been used to discriminate between several different 20-mers with single base substitutions in the first few bases at each end.¹⁴

As a first step, here G to A substitution at the 3' end of the sequence: CTT-TTT-CCT-GCA-TCC-TGT-CTG-GAA-X was studied. The raw spectra and resulting difference spectrum are shown below in Figure 5.

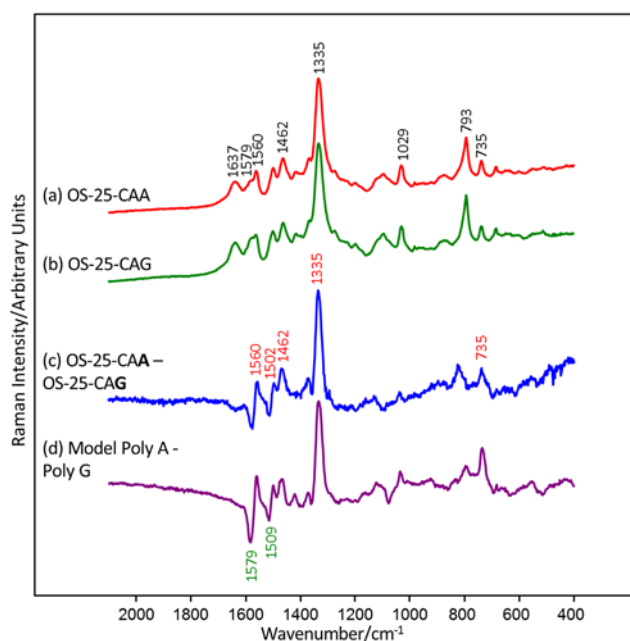


Figure 5. Raw SERS spectra of (a) CTT-TTT-CCT-GCA-TCC-TGT-CTG-GAA-A (OS-25-CAA) and (b) CTT-TTT-CCT-GCA-TCC-TGT-CTG-GAA-G (OS-25-CAG), (c) is the 1:1 spectral subtraction of the two while (d) is the model spectrum created by subtracting reference spectra of Poly A – G dimer.

As was the case with single base additions, the differences between the raw OS-25-CAA and OS-25-CAG spectra and were too small to detect by visual inspection. The 1:1 subtraction of the two raw spectra is shown in (c) where the spectrum shows two sets of bands; a positive set with peaks at 1560 cm⁻¹, 1335 cm⁻¹ and 735 cm⁻¹, which are features from the added A and a second set of negative bands with features at 1579 cm⁻¹ and 1509 cm⁻¹, which are due to loss of the G. A better way to interpret the difference spectra is to generate model spectra since these give not only the expected band positions for differences but also the relative band intensities. Figure 5(d) shows the model difference spectrum calculated as a 1:1 subtraction of Poly A – G dimer. There is a very close fit not only in band positions but also in intensities between the experimental and model difference spectra.

In these spectra 1:1 subtraction was used, rather than the scaled subtractions used for the base extension studies because here the total number of bases remained the same, so the difference signal should have both negative and positive contributions from the leaving and incoming bases. In contrast, the extension experiments involved addition of an extra base, so the required signal was that which arose only from the new component while the background sequence was removed by scaled subtraction. Data illustrating the reproducibility of the spectra and difference spectra recorded under our experimental conditions are shown in Figures S2, S3 and S4.

Further testing was carried out using the set of ODNs shown in Table 3, which allowed for systematic investigation of the effects of substitution at both ends and the centre of the sequences.

Code	Sequence (5' → 3')
OS-25-CAA	CTT-TTT-CCT-GCA-TCC-TGT-CTG-GAA- <u>A</u>
OS-25-CAC	CTT-TTT-CCT-GCA-TCC-TGT-CTG-GAA- <u>C</u>
OS-25-GAG	<u>G</u> TT-TTT-CCT-GCA-TCC-TGT-CTG-GAA- <u>G</u>
OS-25-CAG	<u>C</u> TT-TTT-CCT-GCA-TCC-TGT-CTG-GAA- <u>G</u>
OS-25-CGA	<u>C</u> TT-TTT-CCT-GC <u>G</u> -TCC-TGT-CTG-GAA- <u>A</u>
OS-25-CCA	<u>C</u> TT-TTT-CCT-GC <u>C</u> -TCC-TGT-CTG-GAA- <u>A</u>
OS-25-AAG	<u>A</u> TT-TTT-CCT-GCA-TCC-TGT-CTG-GAA- <u>G</u>
OS-25-AAA	<u>A</u> TT-TTT-CCT-GCA-TCC-TGT-CTG-GAA- <u>A</u>
OS-25-GAA	<u>G</u> TT-TTT-CCT-GCA-TCC-TGT-CTG-GAA- <u>A</u>
OS-25-AGA	<u>A</u> TT-TTT-CCT-GC <u>G</u> -TCC-TGT-CTG-GAA- <u>A</u>
OS-25-GGA	<u>G</u> TT-TTT-CCT-GC <u>G</u> -TCC-TGT-CTG-GAA- <u>A</u>

Table 3. The ODNs used in the analysis of nucleotide substitutions. In the reference code the number refers to the length of the sequence while the 3 letter sequence refers to the identity of the bases at the 5', middle and 3' positions of the sequence, respectively.

In these ODN sequences only G, A and C substitutions were included since these allowed a complete set of substitutions to be carried out while still keeping the number of difference spectra to a reasonable size. The effect of G, C and A nucleotides at the 3' terminus have been published but here the same substitutions were also carried out in the middle of the sequence and at the 5' terminus. For example, C

to A nucleotide substitution at the 3' terminus had already been shown to give spectra that matched the expected reference spectrum.⁶ Here the same bases were substituted but, in addition to 3' substitution, a pair of samples where a C lying in the centre of the sequence was substituted by A were studied, as were a pair where the substitution was at the 5' end. Figure 6 shows the resulting difference spectra and it is clear that substitution at either the 3' or 5' ends or in the centre of the sequence all yield very similar difference spectra, which matches the expected form. This is consistent with previous studies where changes in order but not composition gave only very small changes in spectra, as reported by Guerrini et al for two 20-mer ODNs on spermine-aggregated colloid.¹⁴ On negatively charged colloid no difference was detected between different 12-mers which carried three A and nine C bases in blocks of three i.e. Poly A₃C₉, Poly C₃A₃C₆ and Poly C₃AC₃AC₃A.¹ Indeed, it was only with dsDNA samples designed to maximise the size of the perturbation, Poly[A]₁₂[C]₁₂ and Poly[AC]₁₂ (which compares a strand with an alternating structure to one that has two distinct ends) that a difference just larger than the experimental error could be detected.¹

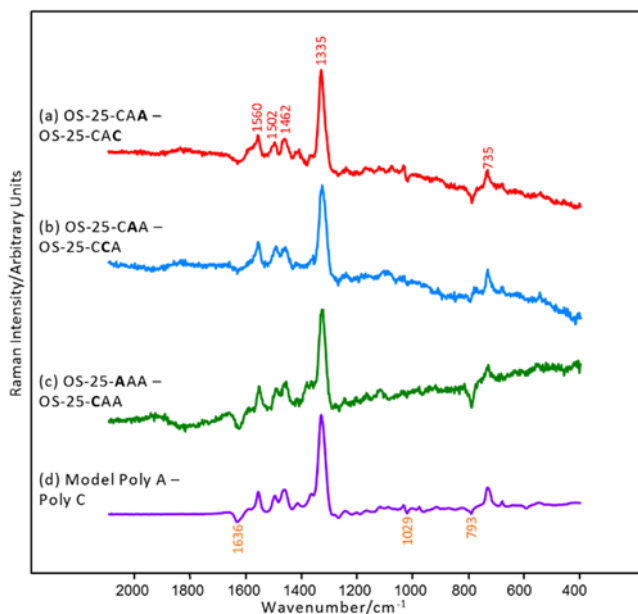


Figure 6. SERS difference spectra resulting from 1:1 spectral subtraction of the spectra of (a) OS-25-CAA – OS-25-CAC (C to A substitution, 3' end), (b) OS-25-CAA – OS-25-CCA (C to A substitution, middle), (c) OS-25-AAA – OS-25-CAA (C to A substitution, 5' end) compared with (d) the expected model difference spectrum created by subtraction of Poly A – Poly C reference spectra.

Surprisingly, in light of these observations, analysis of the entire sample set showed that many substitutions gave difference spectra which clearly did not match the expected model differences. For example, Figure 7 shows data for G to C substitution at the 3' end and centre of a pair of ODNs. The 3' difference spectrum is as expected and reflects the same spectral pattern observed in the model, with positive bands at 1636 cm⁻¹, 1029 cm⁻¹ and 793 cm⁻¹, corresponding to the addition of the C nucleotide and negative bands

at 1579 cm⁻¹ and 1509 cm⁻¹, due to removal of the G nucleotide. However, substitution in the middle of the chain yields an overall negative signal with a band pattern similar to that of Poly A, which is very different from the expected form. This was not a random fluctuation or due to a poor batch of ODN. When the experiment was repeated using a different batch of ODN and different colloid the same result was obtained. Similarly, since the difference procedure gave spectra which were in excellent agreement with expectations for some of the substitutions, the basic method did not seem to be at fault, which suggests that the anomalous behaviour arose from the properties of the ODNs themselves.

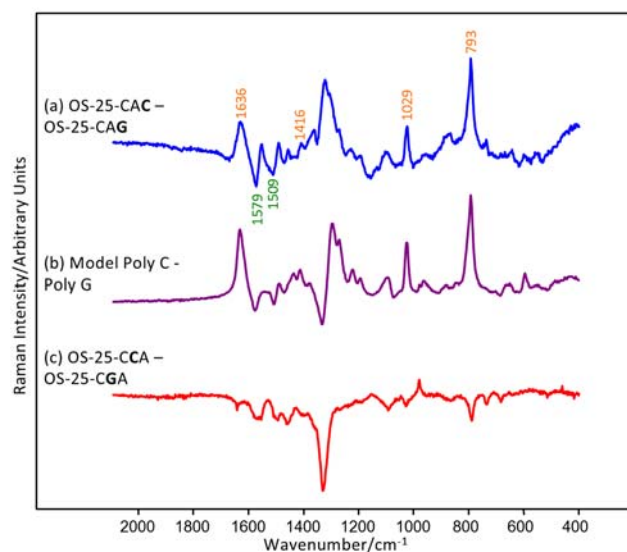


Figure 7. SERS difference spectrum (a) resulting from 1:1 subtraction of the spectra of OS-25-CAC – OS-25-CAG (G to C substitution, 3' end) which matches that of (b) the expected model difference spectrum created by subtraction of Poly C – Poly G. (c) SERS difference spectrum for OS-25-CCA – OS-25-CGA (G to C substitution, middle) which clearly does not have the expected form.

One possible source of this effect might be that the secondary structure of the ODNs is being changed by the differences in the primary sequence which alter self-complementarity in the sequences. This could lead to changes in the interaction between the ODNs and the enhancing surface. Previously attempts to use sample heating to remove the effect of secondary structure were unsuccessful but the temperatures used were quite high (95°C)²⁴ and so here a series of more gentle heating conditions were investigated. Samples of the same ODNs shown in Figure 7(c), OS-25-CAG and OS-25-GAG, were pre-treated before analysis by heating at a series of temperatures and heating times. Temperatures in the range 55 – 70°C with heating times varying from 10 – 30 minutes followed by a 1 minute ice quenching step were tested. For samples at the extremes of the temperature and time ranges the difference spectra did not match the expected reference data but heating to 60°C for 15 minutes yielded spectra which were much closer to the expected form. Figure 8 shows the change in the difference spectra induced by sample pre-treatment for the test system.

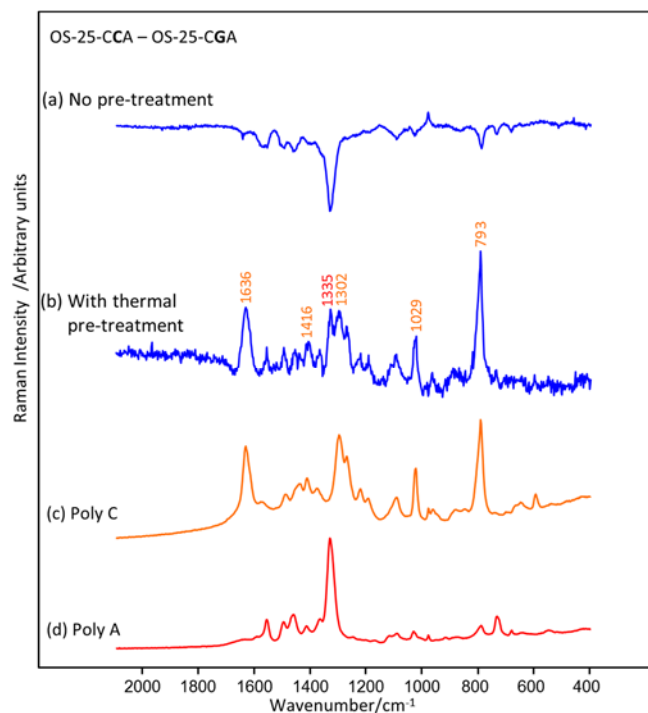


Figure 8. SERS difference spectra resulting from 1:1 subtraction of the spectra of OS-25-CCA and OS-25-CGA (G to C substitution, middle), (a) without the thermal pre-treatment and (b) with the thermal pre-treatment, compared with (c) Poly C and (d) Poly A. Without pre-treatment the difference spectra show strong negative adenine bands, after treatment they show the expected growth of cytosine bands with a small artefact at 1335 cm^{-1} due to unsubtracted adenine.

Further tests showed that almost 50% of the pairs of samples tested produced spectral subtractions which differed significantly from the expected spectra created from subtractions of the corresponding reference polynucleotides (see Table 4). Most of the variation produced similar spectra to the negative spectrum seen in Figure 7. The thermal pre-treatment was then applied throughout the sample set for consistency. After thermal pre-treatment all the substitutions gave the expected patterns, so the SERS difference spectra allowed the identification of all substituting nucleotides at each of the three positions tested along the ODN sequence. This is certainly strong evidence that it is the secondary structure of the ODNs which gives rise to the anomalous behaviour of some of the ODNs, since the thermal pre-treatment is not sufficiently harsh to cause any changes larger than uncoiling the strands. Although we have no direct evidence of the mechanism of how the secondary structure can give rise to the anomalous subtraction data it is obvious that the spectral difference procedure requires that the only change in the samples is the substitution of the base. If the substitution leads to a difference in secondary structure that may also change the way in which the ODN interacts with the enhancing surface, bringing different regions of the sequence in closer proximity and therefore altering the enhancement pattern.

	CAA	CAC	GAG	CAG	CGA	CCA	AAG	AAA	GAA	AGA	GGA
CAA											
CAC	Yes										
GAG											
CAG	Yes	Yes	No								
CGA	Yes										
CCA	Yes				No						
AAG			No	No							
AAA	Yes						Yes				
GAA	No		Yes					No			
AGA					No			Yes			
GGA					No				No	Yes	

Table 4. A summary of the substitutions which did or did not give the expected difference spectra without any thermal pre-treatment of the samples. After thermal pre-treatment all substitutions gave the expected patterns.

The observation that secondary structure affects the substitution data highlights the fact that although ODNs behave in many ways as a simple combination of their constituent nucleosides, the particular nature of ss-DNA does mean that in some instances the ODNs are not simply the sum of their parts. It is interesting to compare this result to work on spermine-aggregated colloid where the interaction between the ODN and the positively charged surface is through interactions with the phosphates on the backbone rather than direct interaction of the nucleobases with the surface. Under those conditions a series of 20-mer ODNs with single base differences showed rational changes in their spectra but the same analysis did not allow substitution to be detected in 35-mers.¹⁴ It is not possible at this point to be certain if this was due to secondary structure effects in the particular 35-mers studied but it does seem possible. In addition, the strength of the interactions between the ODNs and negatively or positively charged surfaces are likely to be quite different so their tendency to disrupt the hydrogen bonding interactions observed in solution on binding may also be different. In future it may be possible to exploit this by tuning the interaction to either retain or reduce the effect of secondary structure on binding.

Multi Variate Data Analysis – Partial Least Squares

Overall, the results above show that thermally pre-treating the ODN sequences gives SERS spectra which show rational changes on substitution of a single base. The observation that the difference

spectra are very similar regardless of the position in the ODN where the nucleotide is substituted suggests that the SERS difference method will be able to identify the type but not the position of the substitution. However, the effects of substitution position is expected to be subtle and therefore PLS was used to investigate the changes in the sample set, since it is a much more rigorous tool than simple visual inspection for quantitative analysis involving simultaneous changes in several variables. The rationale was that if it was possible to build a PLS model which gave the correct predictions for the number and type of substitutions which were carried out, irrespective of the positions, it would suggest that there is no systematic difference in signals created by substitutions at the ends or middle of the ODNs.

The PLS-1 model was constructed from a sample set in which, because substitutions were carried out at just 3 positions, the samples could be described using only the identity of the base at each position, so the sequence CTT-TTT-CCT-GCA-TCC-TGT-CTG-GAA-A was coded as CAA. However, in the model the only input information was the total number of each base at the substitution sites, so both CAA and ACA would be coded 1C: 2A: 0G and the input data therefore did not carry any information of the position of the bases. This input was then used to build a PLS model, it was found that just 3 principal components explained 99% of the variance in the 11 spectra in the sample set. This model could then be used to predict the total number of each of the bases within each sample from their SERS spectra, so for example, the CAA sample which was 1C: 2A: 0G might be predicted as 0.95C: 1.93A: 0.21G.

Figure 9 show histograms of representative data for 3 samples in which the actual and predicted number of each of the bases can be compared visually. Within each of these samples the prediction is

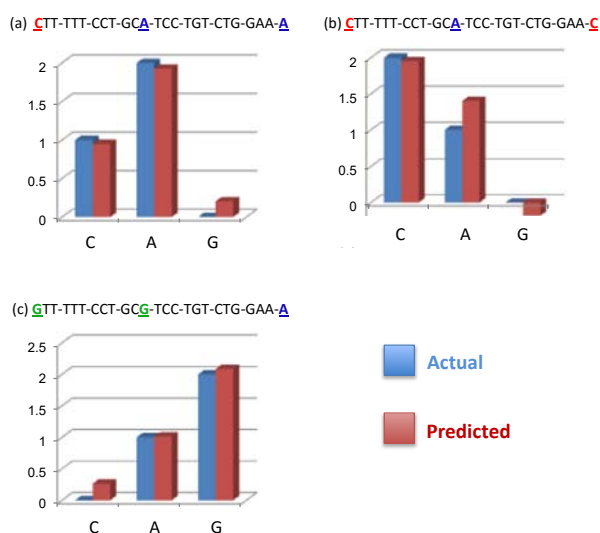


Figure 9. Actual and predicted (PLS-1) base composition for the three substitution positions in (a) OS-25-CAA, (b) OS-25-CAC and (c) OS-25-GGA.

good, particularly since the result must be an integer for each base, so that the outputs from the model for CAA should be rounded to the nearest integer and 0.95C: 1.93A: 0.21G therefore becomes 1C: 2A: 0G., the correct result. The greatest difference between actual versus predicted occurs in (b), where the sequence containing 2C: 1A: 0G nucleotides were predicted to have 1.95C: 1.40A: -0.18G, however, this would still be rounded to indicate the correct composition.

As an alternative to showing histograms for each of the samples, the predicted versus actual number of bases for all samples are plotted in Figure 10. This allows all the predicted values to be displayed in a compact form and any systematic errors to be uncovered. It is clear from the plots that the predictions are excellent, particularly because as mentioned above, predicted values should be rounded to the nearest whole number and if this is done all the predictions are correct across the 11 samples.

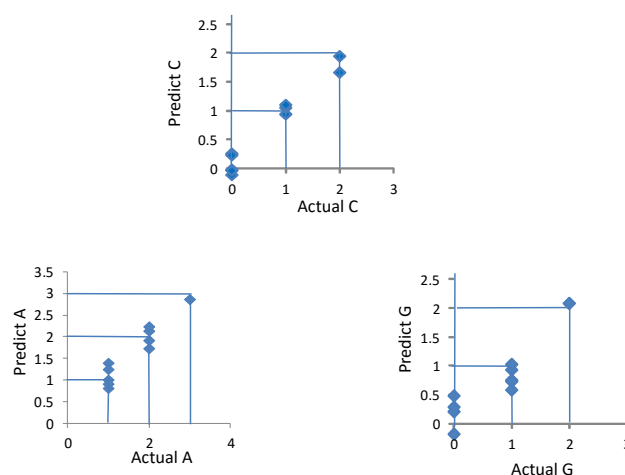


Figure 10 Plots of actual versus predicted (PLS-1 model) sample composition for (a) C, (b) A and (c) G across the entire sample set of 11 ODNs.

Finally, as with all PLS models, it was important to verify that the changes detected in the PLS model were not the result of over-fitting. Since there were just 11 spectra and 3 variables, this would have been a real possibility. However, the factor loadings which are shown in Figure 11., demonstrate that the model correctly identified the sources of variance within the data, since these plots match very closely the spectra of the individual nucleobases. For example, trace (a), the Factor 1 Loading, contains bands at 1560 cm^{-1} , 1335 cm^{-1} and 735 cm^{-1} which closely align to the spectrum of (b) Poly A. Similarly (c), Factor 2 Loading, has strong bands at 1636 cm^{-1} , 1302 cm^{-1} , 1029 cm^{-1} and 793 cm^{-1} , which match the bands of Poly C. Finally, trace (e), Factor 3 loading, contains bands at 1579 cm^{-1} and 1509 cm^{-1} , corresponding to (f), the G dimer. This close correspondence between the loading plots and the known spectra of the components is expected in this case since PLS works by explaining variance in terms of factors which are orthogonal to each other in vector space

while here the dataset is composed the spectra of samples where the variations between them were chemically “orthogonal”, in the sense that in each instance where one nucleobase was removed, a second nucleobase with a different spectrum was substituted in its place.

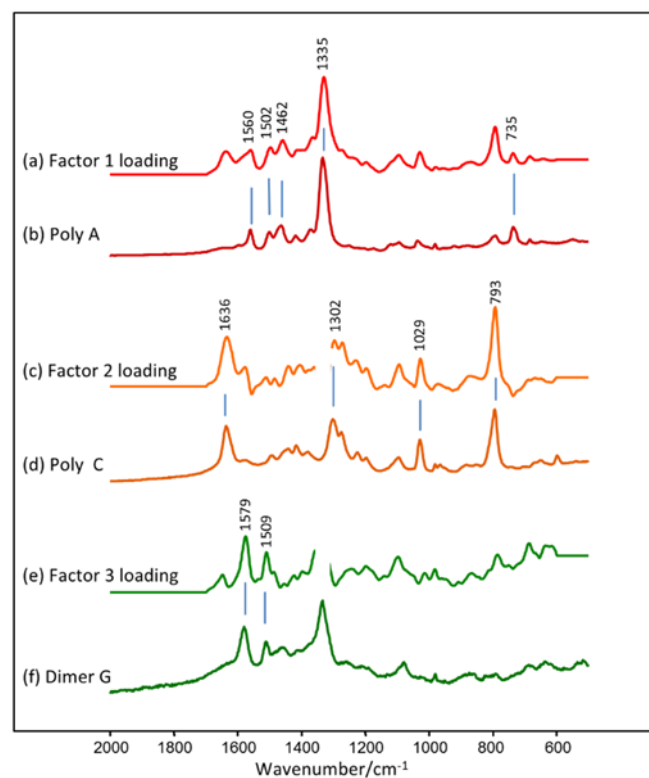


Figure 11. Loading plots for the first three factors in the PLS-1 model (which account for 99% of the variance) compared to reference spectra. (a) Factor 1 loading compared with (b) Poly A, (c) Factor 2 loading compared with (d) Poly C and (e) Factor 3 loading compared with (f) the 2'-OMe-guanosine dinucleotide.

Irrespective of the sample compositions, the close correspondence between loadings and component spectra is impressive because the loading plots arise from the vector algebra used in the minimisation and were generated entirely without any input of what the expected forms might be. The ordering of the factors suggests that the size of the SERS signal decreases from A > C > G, as the loading number indicates its contribution to the overall variance within the data set, Factor 1 having the largest contribution etc. This order is consistent with the data shown in Figure 1.

The broader conclusion from this analysis is that the PLS-1 model did not reveal any underlying pattern in the data which was needed to explain the variance in the spectra, other than the number of each of the nucleobases present. Specifically this shows that, since there is no requirement to add positional information to the input to obtain a good model, the spectral data do not have a strong positional component. This generalises the observation that the same

substitution creates the same spectral change, irrespective of the position of the substituted nucleobase, which was illustrated for a single example of C to A substitution in Figure 6. Of course it is worth stressing that this is only the case when the samples are deliberately uncoiled and that if sample pre-treatment is not used there may be large differences between the spectra of samples with the same overall base composition but different primary structure (as shown in Figure 7). It could be argued that these differences do themselves give a method for distinguishing between samples with different primary sequences. However, the differences which are observed do not appear to be systematic, which means that interpreting the observed spectra at any level other than identifying that something has changed will be difficult. In principle, it might be possible to build libraries of sequences and use spectral matching to identify specific sequences but this would certainly be challenging, particularly for mixed samples. Alternatively, for diagnostic applications using complex biological samples if a particular spectral profile indicates a disease state then this can be detected and interpreted without any detailed knowledge of the constituents that lead to that characteristic profile.

Conclusions

In conclusion, this work has shown that high quality, reproducible SERS signals of unmodified ss-DNA can be obtained using simple standard Ag colloids. In general the spectral changes induced by adding or substituting a single nucleobase are too small to detect by eye but the reproducibility and signal-to-noise levels are sufficiently high to allow meaningful difference spectra to be obtained. These difference spectra can be used, for example, to detect the addition of a single nucleobase to the 3' terminus of 10-mer and 20-mer sequences. The central question in this work was whether the SERS signal of ss-DNA is simply a combination of the individual bases that comprise the sequence, a related question is whether the spectra show any positional information i.e. if sequences with the same total numbers of each base will give the same spectra, irrespective of the ordering along the chain? Here we did find a small end effect where nucleobases near the 3' terminus gave slightly larger signals but the effect is minor (30%) and very much smaller than the orders of magnitude intensity change observed with thiol-tethered “molecular rulers”, due to difference in orientation with the enhancing surface between these two sample types. One sequence-dependent complication did arise when the SERS difference spectra were found to be sensitive to secondary structure, which prevented reliable detection of nucleobase substitutions. However, thermal pre-treatment removed this effect and allowed the identity of the incoming and outgoing nucleobases, but not their position, to be detected reliably. This was confirmed by multivariate data analysis which showed that a simple model with three factors and no positional information could account for 99% of the variance in the data.

These results show that SERS can be used to detect even relatively small changes in the composition of ss-DNA at the single base level.

The fact that the spectra are not position dependent means that cannot be used to obtain sequence information directly but this is something which is better addressed using hybridisation methods in any case. Conversely, the lack of sensitivity to position means that it is well suited to detecting post-transcription modifications to DNA which can occur at multiple positions along a given sequence.

Acknowledgements

SD was supported by a Department for Employment and Learning studentship, N.I.

References

- [1] E. Papadopoulou, S. E. J. Bell, *Chem. Eur. J.* **2012**, *18*, 5394-5400.
- [2] M. Cheung, W.W.Y. Lee, D.P. Cowcher, R. Goodacre, S. E. J. Bell, *Chem. Commun.* **2016**, *52*, 9925-9928.
- [3] K. Gracie, D. Lindsay, D. Graham, K. Faulds, *Anal. Methods* **2016**, *7*, 1269-1272.
- [4] K. Gracie, M. Morres, W.E. Smith, K. Harding, M. Girolami, D. Graham, K. Faulds, *Anal. Chem.* **2016**, *88*, 1147-1153.
- [5] S. Mabbott, D. Thompson, N. M.S. Sirimuthu, G. McNay, K. Faulds, D. Graham, *Faraday Discuss.* **2016**, *187*, 461-472.
- [6] E. Papadopoulou, S. E. J. Bell, *Angew. Chem. Int. Ed.* **2011**, *50*, 9058-9061.
- [7] A. Barhoumi, N. J. Halas, *J. Phys. Chem. Lett.* **2011**, *2*, 3118-3123.
- [8] A. J. Hobro, S. Abdalib, E. W. Blanch, *J. Raman Spectrosc.* **2012**, *43*, 187-195.
- [9] L. E. Camafeita, S. Sanchez-Cortes, J. V. Garcia-Ramos, *J. Raman Spectrosc.* **1995**, *26*, 149-154.
- [10] L. J. Xu, Z. C. Lei, J. X. Li, C. Zong, C. J. Yang, B. Ren, *J. Am. Chem. Soc.* **2015**, *137*, 5149-5154.
- [11] J. Morla-Folch, H. N. Xie, P. Gisbert-Quilis, S. Gómez-de Pedro, N. Pazos-Perez, R. A. Alvarez-Puebla, L. Guerrini *Angew. Chem. Int. Ed.* **2015**, *54*, 13650-13654.
- [12] L. Guerrini, Z. Krpetic, D. van Lierop, R. A. Alvarez-Puebla, D. Graham, *Angew. Chem. Int. Ed.* **2015**, *54*, 1144-1148.
- [13] A. Tores-Nunez, K. Faulds, D. Graham, R. A. Alvarez-Puebla, L. Guerrini, *Analyst*, **2016**, *141*, 5170-5180.
- [14] J. Morla-Folch, P. Gisbert-Quilis, M. Masetti, E. Garcia-Rico, R. A. Alvarez-Puebla, L. Guerrini *Angew. Chem. Int. Ed.* **2017**, *129*, 2421- 2425.
- [15] S. R. Panikkanvalappil, M. A. Mahmoud, M. A. Mackey, M. A. El-Sayed, *ACS Nano* **2013**, *7*, 7524-7533.
- [16] J. D. Driskell, O. M. Primera-Pedrozo, R. A. Dluhy, Y. Zhao, R. A. Tripp, *Appl. Spect.* **2009**, *63*, 1107-1114.
- [17] Y. Chen, G. Chen, X. Zheng, C. He, S. Feng, Y. Chen, X. Lin, R. Chen, H. Zeng, *Med. Phys.* **2012**, *39*, 5664- 5666.
- [18] N. E. Marotta, K. R. Beavers, L. A. Bottomley, *Anal. Chem.* **2013**, *85*, 1440-1446.
- [19] S. E. J. Bell, N. M. S. Sirimuthu, *J. Am. Chem. Soc.* **2006**, *128*, 15580-15581.
- [20] N. Leopold, M. Haberkorn, T. Laurell, J. Nilsson, J. R. Baena, J. Frank, B. Lendl, *Anal. Chem.* **2003**, *75*, 2166-2171.
- [21] E. Papadopoulou, S. E. J. Bell, *J. Phys. Chem. C* **2010**, *114*, 22644-22651.
- [22] E. Papadopoulou, S. E. J. Bell, *J. Phys. Chem. C* **2011**, *115*, 14228-14235.
- [23] A. Munshi, in *DNA Sequencing – Methods and Applications*, InTech, **2012**.
- [24] A. Barhoumi, D. Zhang, F. Tam, N. J. Halas, *J. Am. Chem. Soc.* **2008**, *130*, 5523-5529.