Analyst

The multivariate detection limit for *Mycoplasma pneumoniae* as determined by Nanorod Array-Surface Enhanced Raman Spectroscopy and comparison with limit of detection by qPCR

Journal:	*Analyst*
Manuscript ID:	AN-ART-06-2014-001141.R1
Article Type:	Paper
Date Submitted by the Author:	30-Sep-2014
Complete List of Authors:	Henderson, Kelley; University of Georgia, Microbiology Sheppard, Edward; University of Georgia, Department of Microbiology Rivera, Omar; University of Georgia, Chemistry Choi, Joo-Young; University of Georgia, Chemistry Dluhy, Richard; University of Georgia, Department of Chemistry Thurman, Kathleen; Centers for Disease Control and Prevention, Pneumonia Response and Surveillance Laboratory Winchell, Jonas; Centers for Disease Control and Prevention, Pneumonia Response and Surveillance Laboratory Krause, Duncan; University of Georgia, Department of Microbiology

The multivariate detection limit for *Mycoplasma pneumoniae* as determined by Nanorod Array-Surface Enhanced Raman Spectroscopy and comparison with limit of detection by qPCR

Kelley C. Henderson[a], Edward S. Sheppard[a], Omar E. Rivera-Betancourt[b], Joo-Young Choi[b], Richard A. Dluhy[b], Kathleen A. Thurman[c], Jonas M. Winchell[c], Duncan C. Krause[a]

[a]*Department of Microbiology, University of Georgia, Athens, GA,*
[b]*Department of Chemistry, University of Georgia, Athens, GA*
[c]*Pneumonia Response and Surveillance Laboratory, Centers for Disease Control and Prevention, Atlanta, GA*

Abstract

Mycoplasma pneumoniae is a cell wall-less bacterial pathogen of the human respiratory tract that accounts for up to 20% of community-acquired pneumonia. At present, the standard for detection and genotyping is quantitative polymerase chain reaction (qPCR), which can exhibit excellent sensitivity but lacks standardization and has limited practicality for widespread, point-of-care use. We previously described a Ag nanorod array-surface enhanced Raman spectroscopy (NA-SERS) biosensing platform capable of detecting *M. pneumoniae* in simulated and true clinical throat swab samples with statistically significant specificity and sensitivity. We report here that differences in sample preparation influence the integrity of mycoplasma cells for NA-SERS analysis, which in turn impacts the resulting spectra. We have established a multivariate detection limit (MDL) using NA-SERS for *M. pneumoniae* intact-cell sample preparations. Using an adaptation of International Union of Pure and Applied Chemistry (IUPAC)-recommended methods for analyzing multivariate data sets, we found that qPCR had roughly 10× better detection limits than NA-SERS when expressed in CFU/ml and DNA concentration (fg). However,

the NA-SERS MDL for intact *M. pneumoniae* was 5.3 ± 1.0 genome equivalents (cells/µl). By comparison, qPCR of a parallel set of samples yielded a limit of detection of 2.5 ± 0.25 cells/µl. Therefore, for certain standard metrics NA-SERS provides a multivariate detection limit for *M. pneumoniae* that is essentially identical to that determined via qPCR.

1 Introduction

The cell wall-less prokaryote *Mycoplasma pneumoniae* is a major cause of respiratory disease in humans, accounting for 20% to 40% of all cases of community-acquired pneumonia (CAP), and the leading cause of CAP in older children and young adults.[1-5] The annual economic burden of CAP in adults alone exceeds $17 billion, and the incidence of infection in the very young and elderly is on the rise.[5, 6] Furthermore, extra-pulmonary sequelae occur in up to 25% of cases, and chronic *M. pneumoniae* infection can play a contributing role in the onset, exacerbation, and recurrence of asthma.[2]

M. pneumoniae infection is transmitted through aerosolized respiratory secretions and spreads efficiently but slowly within close living quarters, with incubation periods as long as three weeks.[7, 8] Symptoms tend to be nondescript, and the disease often has complex and variable presentations, making definitive diagnosis challenging.[3, 5, 9] As a result, diagnosis is often presumptive and relies heavily on the combination of physical findings and elimination of other possible causes.[1, 2, 8] Serologic testing has historically been considered the foundation for diagnosis of *M. pneumoniae* infection but has severe limitations in sensitivity and

specificity, a high tendency for false negatives, and often must be paired with another diagnostic method.[1-3, 8, 10] Of the currently existing methods, the most efficient means for detection is quantitative polymerase chain reaction (qPCR). At present, the only FDA-approved tests for the clinical detection of *M. pneumoniae* are the illumigene® automated detection system (Meridian Bioscience, Inc., Cincinatti, Ohio) and the FilmArray® Respiratory Panel (BioFire Diagnostics Inc., Salt Lake City, Utah). The illumigene® platform uses loop-mediated isothermal amplification and is capable of detecting *M. pneumoniae* in both throat and nasopharyngeal swab specimens with a high degree of sensitivity and specificity. The FilmArray® Respiratory Panel employs nested, multiplex qPCR with endpoint melt curve analysis on nasopharyngeal swabs to test for 21 different viral and bacterial respiratory pathogens, and is capable of detecting *M. pneumoniae* as low as 30 colony-forming units (CFU)/ml.[11] These methods can exhibit high sensitivity and allow for detection in the early stages of infection, but the expertise and complexity required and the lack of standardization between available tests and between labs limits the practicality of widespread use in hospitals and reference laboratories or point-of-care testing.[1-3, 8, 10] These limitations create a critical barrier to the accurate and timely diagnosis of *M. pneumoniae* infection, and a rapid, simple, diagnostic platform would greatly improve the control of *M. pneumoniae* disease.

Vibrational spectroscopy has an inherent biochemical specificity that led to its consideration as a next-generation platform for the rapid detection, characterization, and identification of infectious agents.[12-15] Raman spectroscopy in particular has several advantages for application to biological samples, including

narrow bandwidths, good spatial resolution, and the ability to analyze aqueous samples due to the absence of interference by water molecules.[12, 13, 16] Additionally, Raman spectra provide detailed structural information on the chemical composition of a sample and can serve as a characteristic molecular fingerprint for pathogen identification.[15, 16] Despite these advantages, standard Raman spectra are inherently limited by low scattering cross-sections, which translate to weak signals for detection, and initially made the application of traditional Raman spectroscopy for biosensing applications impractical and inefficient.[7, 13, 16] However, in the late 1970s it was discovered that adsorption of molecules onto nanoscopically roughened metallic surfaces results in significant enhancements in Raman signal and spectral intensity.[15-17] The enhancement is attributed to the increased electromagnetic field experienced by molecules in close proximity to the metallic surface, with typical signal enhancements of 10^4 to 10^{14} with respect to normal Raman intensities.[12, 13] Most importantly, for biomedical applications, surface-enhanced Raman spectroscopy (SERS) retains the advantages of standard Raman spectroscopy, in addition to markedly improving sensitivity and allowing for considerable success in whole organism molecular fingerprinting.[12, 16, 18, 19] However, inconsistency and lack of reproducibility in the preparation of SERS-active substrates has hindered the widespread use of SERS for biosensing applications.[12, 13, 16]

Highly ordered silver nanorod array (NA) substrates fabricated using oblique angle deposition (OAD) yield consistent SERS enhancement factors of around 10^8, with less than 15% variation between substrate batches.[13] In addition, the usefulness of OAD-prepared substrates can be improved further when patterned

into a multiwell format with polydimethylsiloxane (PDMS).[12] The highly reproducible detection capabilities of NA-SERS substrates have been demonstrated for multiple infectious agents, including respiratory syncytial virus, rotavirus, influenza, HIV, adenovirus, SARS, and *M. pneumoniae*.[7, 14, 19-21]

Hennigan et al. previously described an NA-SERS-based assay capable of detecting *M. pneumoniae* in both simulated and true clinical throat swab samples, with statistically significant sensitivity and specificity.[7] Their initial evaluation of the NA-SERS biosensing platform capabilities indicate the potential for application as a next-generation diagnostic tool for the clinical detection of *M. pneumoniae*, but a more comprehensive analysis is needed prior to proceeding with clinical validation.[7] In addition, the initial study analyzed samples prepared in water, and we hypothesize that as a result the content of the analyte on the substrate consisted predominately of lysed cells, cytoplasmic content, and membrane debris. In the present study we further explored the impact of differences in sample preparation, defined the lower multivariate detection limit for *M. pneumoniae* intact-cell preparations by NA-SERS, and evaluated in parallel the limit of detection by qPCR, in order to continue the development of NA-SERS as a next-generation platform for the detection of *M. pneumoniae* in clinical samples.

2 Materials and Methods

Preparation of *M. pneumoniae* samples for SERS analysis

Wild type *M. pneumoniae* strain M129 was used in this study. Mycoplasma samples were cultured in SP4 medium[3, 22] in tissue culture flasks with a 1µl/ml inoculation,

incubated at 37°C, and harvested at log phase when the phenol red indicator turned an orange color upon reaching a pH of ~6.5. At time of harvest, spent growth medium was decanted and cells were scraped into 0.1× volume of SP4. Cells were then syringe-passaged 10× with a 25 gauge needle and aliquots made for determination of protein content, plating on PPLO agar[23] for colony-forming unit (CFU) determination, DNA extraction for qPCR analysis, and SERS analysis.

We used two protocols for preparation of *M. pneumoniae* samples for NA-SERS analysis. Initially we followed the protocol described previously.[7] Briefly, the spent SP4 medium was decanted and cells collected by scraping into 0.1× volume sterile deionized (DI) water and centrifuged (20,000×g for 25 min at 4°C). Mycoplasmas were then washed 3× in DI water, suspended in a final volume of 500 µl DI water, syringe-passaged 10× with a 25-gauge needle to disperse clumps, fixed with the addition of 500 µl of 8% formaldehyde in DI water, and stored at 4°C until time of SERS analysis. We anticipated that this protocol would yield significant lysis of the mycoplasma cells and therefore we also prepared samples by adding to a 500-µl aliquot of mycoplasma in SP4, 500 µl of 8% formaldehyde in SP4 (pH 7.0-7.5) and stored at 4°C until SERS analysis. Three independent M129 cultures were prepared for intact-cell SERS analysis. Growth medium control samples were prepared in parallel for the intact-cell sample preparation method. Briefly, uninoculated SP4 medium was incubated in the same volume as was used for *M. pneumoniae* cell growth. The SP4 medium-only negative control samples were treated identically as *M. pneumoniae* positive samples at time of harvest, washing, and fixation, as described above. At time of SERS analysis, mycoplasma and growth medium control

samples were serially diluted in DI water in ten-fold or hundred-fold increments to encompass and extend below the clinically relevant range of *M. pneumoniae* concentrations in order to determine the endpoint of the NA-SERS detection capabilities.

Preparation of *M. pneumoniae* samples for protein, DNA, and qPCR analysis

Aliquots designated for protein content and DNA extraction were prepared by centrifugation at 4°C and 20,000×g for 25 min. The supernatants were removed and the samples washed 2× in sterile PBS, pH 7.2. After the second wash the samples were suspended in 1 ml sterile PBS and analyzed for protein content via the colorimetric Bicinchoninic acid (BCA) assay,[24] or DNA extraction by the QIAamp DNA Blood Minikit (Qiagen, Valencia, CA) using the blood and body fluids protocol, including RNase A treatment. 200 µl of sample were used for DNA extraction, with a final elution volume of 200 µl for use to quantitate DNA content and in qPCR analyses. Quantitation of genomic DNA concentration was performed using a NanoDrop instrument (Model ND-1000, Thermo Scientific, Wilmington, DE) and analyzed by NanoDrop software V3.5.2. Genome equivalents of *M. pneumoniae* were calculated from DNA concentration obtained from this analysis and using the previously determined weight of the *M. pneumoniae* genome, 5.3×10^7 Daltons.[25]

Parallel analyses of the endpoint of detection by qPCR were done on three independent *M. pneumoniae* cultures using the CARDS toxin gene target[8] and assay cycling parameters developed by the U.S. Centers for Disease Control and Prevention (CDC).[8] DNA was extracted from the three independent cultures and serial dilutions of extracted DNA were made in nuclease free water prior to qPCR

analysis using an ABI 7500 real-time PCR system (Applied Biosystems, Foster City, CA) and SDS v1.4 software platform (Applied Biosystems, Foster City, CA) for analysis of fluorescence amplification. Briefly, qPCR mastermix reactions contained 12.5 µl 2× PerfeCTa® qPCR FastMix (Quanta Biosciences, Gaithersburg, MD, USA), forward and reverse primers (1µmol/L each), labeled probe (200 nmol/L), 5 µl of total nucleic acid extract, and nuclease free water to a final reaction volume of 25 µl.[26] Cycling conditions were as follows: 1 cycle of 95°C for 5 min followed by 45 cycles of 95°C for 15 sec and 60°C for 1 min. Upon completion of the cycling, positive amplification of a sample was defined as a sigmoidal fluorescence increase above the cycle threshold (Ct) limit assigned to the raw fluorescence data by the user[27]. Consistent with other qPCR platforms employed for clinical detection of *M. pneumoniae*[8], the limit for detection by qPCR was defined for each culture using Ct values from the fluorescence amplification analysis and defined as the lowest concentration for which positive amplification occurred in at least one of three replicates tested per individual dilution.

Scanning Electron Microscopy (SEM) characterization of *M. pneumoniae* samples

SEM images of the bacteria were obtained using a Zeiss 1450EP (Carl Zeiss MicroImaging, Inc., Thornwood, NY). The samples were fixed as previously described,[28] with modifications. As a control, cells grown on glass coverslips were fixed in 2% glutaraldehyde in sodium cacodylate buffer for one hr. Briefly, lysed- and intact-cell samples were dried onto glass coverslips, fixed with glutaraldehyde,

washed twice in sodium cacodylate buffer for five min each wash, post-fixed in 1% OsO_4 in sodium cacodylate buffer for one hr, washed once with sodium cacodylate buffer for ten min, and rinsed twice with water for five min. The SEM coverslips were then treated with an ethanol dehydration series sequentially (five min each step) with 25, 50, 75, 85, 95 and three 100% washes, critical point dried, and sputter coated with 20-nm diameter gold.

NA-SERS measurements and chemometric analysis

Silver nanorod array substrates were prepared for reproducible enhancement of the Raman signal using OAD.[13, 21, 29, 30] Briefly, an electron beam evaporation system was used to deposit three sequential layers onto 1×3'' glass microscope slides as follows: a 20-nm Ti film, a 500-nm Ag film, and an obliquely angled (86° with respect to the surface normal) as specified for optimum signal production.[30] Prior to their use, the nanorod substrates were cleaned for five minutes in an Ar^+ plasma using a plasma cleaner (Model PDC-32G, Harrick Plasma, Ithaca, NY) to remove any surface contamination.[31] The 1×3'' NA substrates were then patterned into 40 3mm diameter PDMS-formed wells. Raman spectra were acquired using a Renishaw inVia Reflex multi-wavelength confocal imaging microscope (Hoffman Estates, IL). A Leicha apochromatic 5× objective (NA 0.12) illuminated a 1265 μm^2 area on the substrate, which allows spatial averaging and minimization of the effect of potential random hot spots. A 785-nm near-infrared diode laser (Renishaw) operating at 10% power capacity (28 mW) provided the incoming radiation, and spectra were collected in 10-sec acquisitions.

A dilution series from each of the three *M. pneumoniae* NA-SERS cultures fixed in SP4 and their respective growth medium controls were analyzed on a single substrate. Each individual test dilution was analyzed in duplicate wells, and two wells were left blank on each substrate to obtain a background SERS reading on the naked nanorod substrate only. All samples were applied to the nanorod substrates in a volume of 1 µl per individual well. Samples were dried onto the nanorods overnight and spectra collected from five random locations within each sample spot for analysis. Ten spectra were collected per dilution (five spectra per well per µl of sample) for both experimental and control samples, with n=200 spectra per substrate. Three separate substrates were analyzed, resulting in a total of n=600 spectra. Raman spectra between 400-1800 cm^{-1} were acquired using Renishaw's WiRE 3.4 software. Instrument settings were optimized to maximize signal and minimize saturation or sample degradation arising from laser stimulation.

Raman spectra were first averaged using GRAMS32/A1 spectral software package (Galactic Industries, Nashua, NH) in order to assess signal-to-noise quality, and baseline-corrected using a concave rubber band algorithm which performed ten iterations on 64 points to aid in preliminary evaluation of the spectra and peak assignment (OPUS, Bruker Optics, Inc., Billerica, MA). Chemometric analysis was carried out with MATLAB version 7.10.0 (The Mathworks, Inc., Natick, MA) using PLS-Toolbox version 7.5.1 (Eigenvector Research Inc., Wenatchee, WA). Raw spectra were pre-processed using the first derivative of each spectrum and a fifteen-point, 2^{nd}-order polynomial Savitsky-Golay algorithm. Each dataset was then vector-normalized and mean-centered. Due to the inherently complex nature of the

spectral data, multivariate statistical analysis of the datasets was performed using principal component analysis (PCA), hierarchical cluster analysis (HCA), and partial least squares-discriminatory analysis (PLS-DA), using the PLS Toolbox software. The calculated principal components were used as inputs to the HCA algorithm, which used the K-nearest neighbor and Mahalanobis distance to evaluate minimum variances within clusters. Additionally, a method for estimating a multivariate limit of detection was used based on an extension of the IUPAC recommendations for univariate methods.

3 Results and discussion

SERS sample preparation and its effect on SERS spectra of *M. pneumoniae*

Previous studies[7] indicated a sub-CFU lower endpoint for detection by NA-SERS. In the initial development of the NA-SERS assay, mycoplasma samples were prepared in DI water rather than salt-based buffer in order to avoid potential damage to the Ag nanorods. As such, we hypothesized that the majority of cells in our sample were lysed, and consequently cytoplasmic contents and cell membrane debris encompassed the bulk of our analyte on the substrate, accounting for the sub-CFU detection limits observed. To investigate this point we compared the SERS sample preparation method used previously with a modified protocol expected to yield intact mycoplasma cells, visualizing each sample by SEM (Figs. 1a and 1c). As expected, we observed predominately intact cells with the characteristic flask shape of *M. pneumoniae*[32, 33] when samples were fixed in solution prior to dilution in DI H_20 (Fig. 1a), and an abundance of membrane vesicles characteristic of cell lysis

were present when samples were washed with DI H_20 prior to fixation (Fig. 1c). For comparison we also examined *M. pneumoniae* cells grown on coverslips and fixed in place. Those cells exhibited the expected elongated morphology of *M. pneumoniae* attached to an inert surface[33] (Supplementary Fig. 1).

Mycoplasmas are phylogenetically unique bacteria in that they lack a cell wall and are instead bound by only a cell membrane; this membrane has numerous surface-exposed membrane proteins and glycolipids.[25, 34] As such, the SERS spectra of intact-cell preparations should predominately originate from membrane lipids, glycolipids, and exposed regions of surface proteins accessible for interaction with the Ag nanorods. In contrast, SERS spectra from lysed-cell samples should also contain bands from a multitude of internal cellular components and membrane debris.

The SERS spectra of the two sample preparation types (Figs. 1b and 1d) exhibited both similarities and differences. Qualitatively, the key peaks found within the intact-cell spectra (Fig. 1b) consisted of a broad peak at 895 cm^{-1}, a sharper peak at 1051 cm^{-1}, and three more broad peaks at 1402, 1613, and 1645 cm^{-1}. For the lysed-cell spectra (Fig. 1d) the peaks were more numerous, sharper, and of an overall greater intensity, with the strongest bands falling at 607, 767, 932, 959, 1051, 1137, 1402, 1613, and 1645 cm^{-1}. Several peaks were present in both intact- and lysed-cell samples, including those at approximately 465, 1051, 1284, 1402, 1613, and 1645 cm^{-1}, though the intensity of the bands was different between the two sample types at all peaks other than 465 cm^{-1}.

Vibrational mode assignments for the major Raman shift peaks observed in Fig. 1 are given in Table 1. The region between 550-1000 cm^{-1} contained the majority of the spectral variation between the two sample types. Bands present in both the intact- and lysed-cell samples were more frequently associated with bond vibrations present in amino acids and lipids, whereas the lysed-cell spectra contained additional peaks that commonly correspond with nucleotide, amino acid, and lipid/carbohydrate bond vibrations.[35-41] The spectral differences seen in Fig. 1b and Fig. 1d are likely explained by the differences in the two sample preparation types. The sharper band profile seen in Fig. 1d may also be due to the small vesicle size in lysed-cell preparations, which allows greater surface contact with the Ag nanorod array, with correspondingly greater signal enhancement.

NA-SERS multivariate detection limit (MDL) for intact-cell *M. pneumoniae* preparations

Because clinical samples are likely to have predominantly intact mycoplasmas present, we have assessed the limit of detection of intact-cell *M. pneumoniae* preparations by NA-SERS. Due to sample complexity and heterogeneity, we have employed a whole-spectrum approach to analyze the *M. pneumoniae* SERS spectra, rather than discrete band analysis. Multivariate analysis based on the intrinsic SERS spectrum of the analyte is possible since vibrational spectroscopy is sensitive to the same chemical and structural variations in pathogenic organisms that govern their infectivity and serotype.[42] Thus, vibrational spectra have the ability to differentiate microorganisms based on their inherent biochemical

differences, a technique known as whole organism spectral fingerprinting.[43, 44] The unique biochemical specificity inherent to vibrational spectroscopy has led to its evaluation as a clinical method for detection, identification and classification of pathogenic organisms with species and strain specificity. [45-47]

Unlike the case for univariate calibration, there is no generally accepted methodology for determining the limit of detection in the multivariate case.[48] However, several groups have published protocols for estimating a detection limit for multivariate data based on an extension of International Union of Pure and Applied Chemistry (IUPAC) recommendations for univariate calibration.[49, 50] We have adapted one of these methods[50] to calculate a multivariate detection limit (MDL) for serial dilutions of *M. pneumoniae* as analyzed by NA-SERS. This approach relies on the spectral residuals and a calculated regression vector from a multivariate regression model while taking account of type I (false positive) and type II (false negative) errors in the following manner.

$$\hat{c}_{MDL} = \left(t_{\alpha,v} + t_{\beta,v}\right) \cdot \sqrt{\frac{\sum_{i=1}^{I}\left(\hat{s}_i - s_i\right)^2}{I-2}} \cdot \beta$$

In this expression, $\hat{c}_{MDL}$ is the estimated MDL, $t_{\alpha,v}$ and $t_{\beta,v}$ are coefficients of a Student's t distribution with v degrees of freedom taking into account the probabilities of both type I (α) and type II (β) errors, s_i is the spectral response, $\hat{s}_i$ is the spectral response predicted from the multivariate model, I is the number of samples used in the calibration, and β is the multivariate regression vector. The

values of the $t_{\alpha,v}$ and $t_{\beta,v}$ coefficients were chosen to correspond to IUPAC recommended criteria in which the probabilities of either a type I or type II error are approximately 7%.[50]

Three separate *M. pneumoniae* culture preparations were prepared for the three independent data sets used in this SERS MDL analysis. The details of these cell cultures are presented in Table 2. Three separate dilution series with concentration ranges from 10^8 to 10^{-4} CFU/ml were prepared from these cell cultures and analyzed on three independent Ag nanorod substrates. Baseline-corrected and normalized SERS spectra of the 10^3 CFU/ml dilutions from each of the three independent NA-SERS substrates are given in Supplementary Information Fig. 2 to show spectral reproducibility and consistency among the three datasets.

Due to the propensity for mycoplasma cells to clump, a confounding factor in using CFU values to define endpoints for detection is the potential discrepancy between CFU value and actual cell number, which can differ by as much as three logs.[51] Furthermore, clumping and small cell size prevents quantifying cell number by direct microscopic count.[25, 32] To account for this potential issue, analyses to determine total protein and genomic DNA concentration and calculate genomic equivalents were included to supplement the CFU values for each culture and better define the content of the samples at each detection endpoint. Sample content for all three cultures fell within comparable ranges (Table 2). The molecular content of our samples is consistent with published values for bacterial cells. For example, Zubkov, *et al*. reported an average of 60-330 fg total protein per bacterial cell.[52] *M. pneumoniae* is much smaller than model bacteria, roughly 5% by volume the size of

E. coli, corresponding to 3-16 fg of protein per *M. pneumoniae* cell based on the Zubkov, *et al*. study, and in good agreement with our results of 5.6 fg protein per *M. pneumoniae* cell (Table 2). As expected, the greatest variation observed between cultures was for CFU values, whereas the remaining measures were more consistent among independently prepared samples. As such, for the purposes of describing the dilutions within the multivariate models and comparing MDL, genomic equivalents in cells/ml will be used for consistency and ease of reference.

The multivariate regression vector β was calculated from an optimized partial least squares (PLS) calibration model. This PLS calibration model was constructed using NA-SERS spectra obtained from the *M. pneumoniae* serial dilutions in the range of ~1 to ~10^4 cells/ml, a concentration range that encompasses clinically relevant concentrations of *M. pneumoniae* in respiratory secretions. In development of this multivariate regression model, 2/3 of the spectra were assigned to the calibration set, while 1/3 was assigned to the validation set. Cross validation was performed by leaving out a random selection of 1/3 of the spectra, followed by optimization. This procedure was repeated for 200 iterations; after which the optimum number of latent variables was calculated for data sets (a), (b), and (c) in Table 3 as 2, 3, and 3, respectively.

The MDL by NA-SERS as defined by CFU, protein content, and genome equivalents are shown in Table 3. We determined MDL mean values of 20.3 ± 17.5 CFU/ml, 29.8 ± 8.8 fg protein, and 5,312 ± 1,038 cells per ml for the three data sets. Since 1 µl of the *M. pneumoniae* suspension was applied to the NA-SERS substrate, these data correspond to 5.3 ± 1.0 cells and 29.8 ± 8.8 fg protein per microliter

volume applied. While the standard deviation was higher for some of these metrics than for others, it is important to keep in mind that these values are representative of the very endpoint of the dilution series and range, which is where the greatest amount of variation is to be expected.

Limit of detection by qPCR analysis

At present, the most reliable and rapid test for detecting *M. pneumoniae* in a clinical sample is real-time PCR.[2] We compared the detection capabilities of the NA-SERS assay with a highly sensitive mycoplasma assay developed and employed by the CDC for outbreak detection. A singleplex version of the assay was used for this study, and qPCR experiments were conducted in the Pneumonia Response and Surveillance Laboratory at the CDC in Atlanta, Georgia.[8] As with the NA-SERS data discussed above, three separate M. pneumoniae cell culture preparations were used in preparing samples for qPCR analysis; information on these cell cultures are provided in Table 4.

Similarly with the NA-SERS MDL experiments, dilution series were generated for qPCR analysis from three independent cultures, in which the concentration varied from 10^7 to 10^0 cells/ml. All samples were tested in triplicate, and positive vs. negative amplification of each sample was compared to crossing threshold (Ct) values of positive and negative template controls. Samples amplifying above the Ct value with the *M. pneumoniae* template control were considered positive and those failing to amplify were considered negative.[8] All Ct value data are given in

Supplementary Tables 1-3, while the limits of detection for the individual datasets as determined by qPCR are summarized in Table 5.

The mean values for the lower limit for detection by qPCR was 2.45 ± 0.39 CFU/ml, 44.7 ± 5.0 fg of genomic DNA, and 2,533 ± 251 cells/ml, corresponding to 223.5 ± 5 fg of genomic DNA or 12.67 ± 1.25 cells per 5 µl of sample examined by qPCR. These findings are consistent with those established by the CDC of approximately 1-5 CFU/ml and 50 fg of DNA.[8]

A comparison of the mean LOD's from the qPCR assay (Table 5) with the MDL's from the NA-SERS assay (Table 3) showed that the qPCR LOD's calculated from Ct values were consistently lower, by approximately a factor of 10×, than the MDL's calculated from the SERS data. However, several aspects of these calculations suggest an equivalence of qPCR and SERS methods. First, the NA-SERS analysis used a 1 µl volume of sample whereas qPCR analysis required a 5 µl sample. This is reflected in the genome equivalent limits for each technique, where qPCR exhibited a virtually identical limit of detection (within a factor of 2) when compared with NA-SERS (2.5 ± 0.25 vs. 5.3 ± 1.0 cells/µl, respectively). Also, a key consideration in comparing the two technologies arises from the fact that they detect fundamentally different molecular properties. NA-SERS detects any cell component of *M. pneumoniae* that interacts with the nanorods upon adsorption to the substrate, whereas qPCR amplifies only *M. pneumoniae* DNA.

Sensitivity and specificity calculations by NA-SERS using partial least squares-discriminatory analysis (PLS-DA)

The overall goal of this project is the development of NA-SERS as a platform for clinical determination of *M. pneumoniae* infections. Differentiation of the SERS spectra for identification of *M. pneumoniae* is an important component of clinical applications. To that end, we have analyzed the SERS spectra to determine the sensitivity and specificity of detection using partial least squares discriminant analysis (PLS-DA). PLS-DA is a full-spectrum, multivariate, supervised method whereby prior knowledge of classes is used to yield more robust discrimination by minimizing variation within classes while emphasizing latent variables arising from spectral differences between classes.[53, 54] When using PLS-DA, it is important to include an appropriate negative control to avoid over- or under-fitting the statistical models. For this purpose a mycoplasma-free growth medium control was processed in parallel, in accordance with the intact-cell sample preparation, and serially diluted to match the corresponding *M. pneumoniae* dilution series. This allowed us to build PLS-DA models for each dilution that included both growth medium and substrate negative controls to ensure that any differences in growth media and nanorod background signal within the substrate did not affect the ability of the model to discriminate between the presence or absence of *M. pneumoniae.* The ability to distinguish presence or absence at 90% accuracy has clinical relevance and is consistent with the performance capabilities of existing platforms for *M. pneumoniae* detection. [2, 8] An example of the PLS-DA modeling system used herein is shown in Supplementary Information, Fig. 3.

For each individual dilution for all three dilution series, PLS-DA models were generated to discriminate between three classes: (1) a positive control *M.*

pneumoniae dilution (10^3 CFU/ml) and each individual *M. pneumoniae* test sample; (2) the growth medium control; and (3) the substrate background. PLS-DA models for all individual dilutions contained a total of n=30-to-40 pre-processed NA-SERS spectra (10 spectra per class for substrate background and growth medium control samples, 20 spectra for *M. pneumoniae* control and test sample dilution class) and were cross-validated using a Venetian blinds algorithm with five to six data splits. The clinically relevant concentration of *M. pneumoniae* in respiratory secretions is ~10^3 – 10^5 organisms/ml.[3] Within this range, the sensitivity and specificities calculated from the SERS spectra of intact *M. pneumoniae* dilutions by PLS-DA after cross-validation were between 90-100%. Furthermore PLS-DA was able to classify with ≥ 90% cross-validated sensitivity and specificity for *M. pneumoniae* dilutions spanning 10^2 to 10^8 cells/µl, qualitatively detecting *M. pneumoniae* down to an average of 0.66 ± 0.1 cells/µl. Thus, qualitative detection of *M. pneumoniae* by PLS-DA modeling was more sensitive than detection by qPCR. Full PLS-DA modeling statistics for all intact-cell dilution ranges can be found in Supplementary Information Tables 4-6.

4 Conclusions

M. pneumoniae is a significant human respiratory tract pathogen in both incidence of infection and public health impact, but diagnostic strategies are complicated by the atypical and complex presentation of disease, non-descript symptoms, and the numerous challenges posed by direct culture. Serologic testing was historically the

gold standard for detection but suffered from severe limitations that made it both unreliable and impractical for widespread use. Advances in qPCR technologies have overcome many issues with sensitivity and reliability, but the cost of reagents and requirement for technical expertise is still high, limiting diagnosis by qPCR to advanced laboratory facilities and making it impractical for point-of-care use. Here we have shown that NA-SERS has a sensitivity that equals qPCR for *M. pneumoniae* detection, when expressed in units of genome equivalents (cells/μl). Additionally, our findings stress the significance of sample preparation when using NA-SERS technology. The question of whether cell lysis improves or hinders the detection capabilities of NA-SERS in the presence of a complex clinical background remains to be determined. An important potential advantage of NA-SERS technology is the existence of handheld Raman instruments that have the potential to be employed for point-of-care clinical detection.[55-57] In combination with the minimal sample preparation requirements and expedient detection, NA-SERS shows great promise for future application as a potential platform to apply for point-of-care *M. pneumoniae* diagnostics.

Acknowledgments

This work was supported by Public Health Service research grants AI096364 (DCK) and GM102546 (RAD) from the US National Institutes of Health.

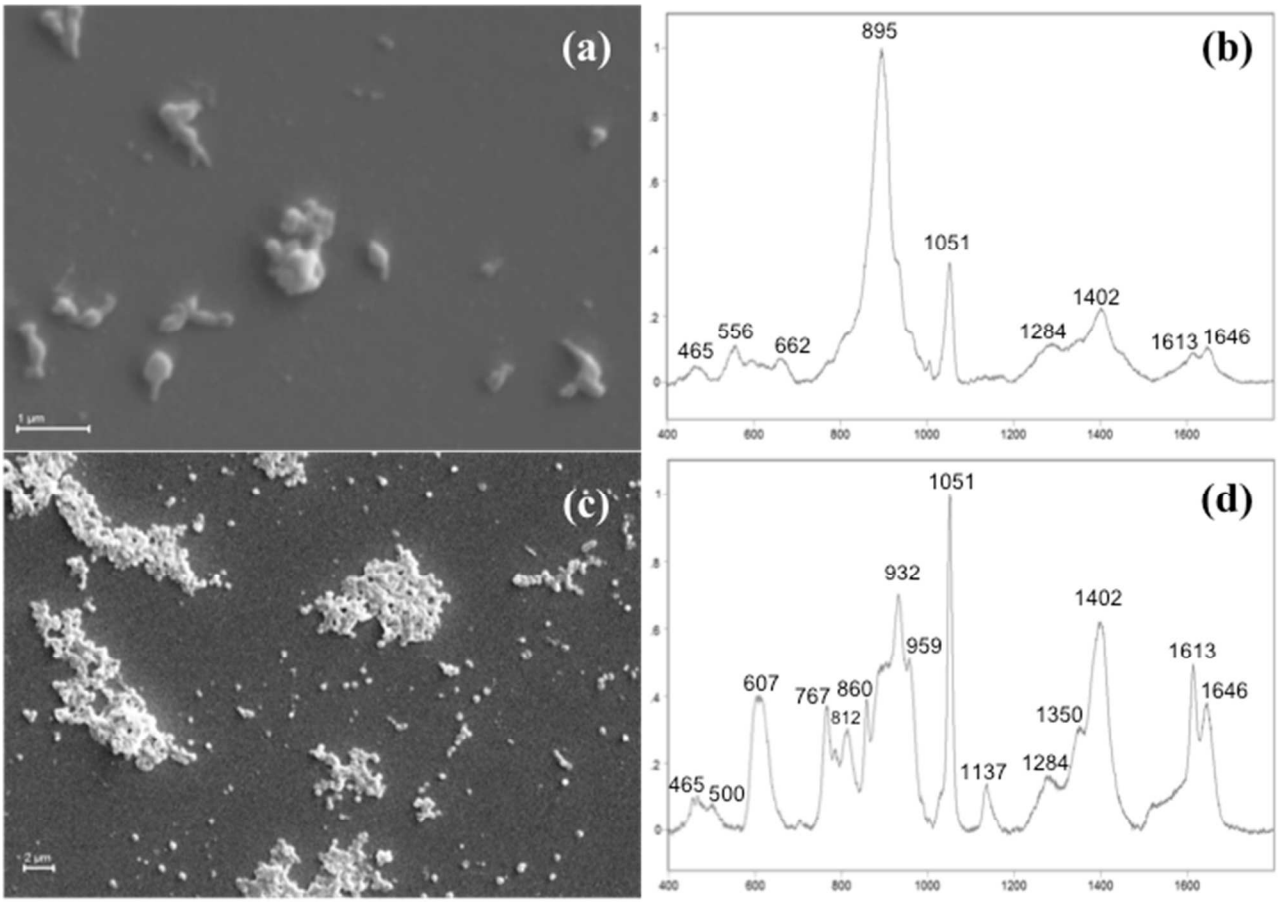

Fig. 1. (a) SEM image of intact *M. pneumoniae* cells fixed in suspension; (b) corresponding SERS spectrum of intact *M. pneumoniae* cells fixed in suspension; (c) SEM image of lysed-cell *M. pneumoniae* preparations; (d) corresponding SERS spectrum of lysed-cell *M. pneumoniae* preparation. For (b) and (d), spectra were averaged (n=10), baseline-corrected, and normalized; initial concentrations were $2x10^3$ CFU/ml (b) and $6.2x10^3$ CFU/ml (d), respectively.

Table 1. Representative Raman bands appearing in the NA-SERS spectra of intact- and lysed- cell *M. pneumoniae* samples. Peaks present in both sample types are shown in green; peaks present in lysed-cell only are shown in blue; peaks found in only the intact cells are shown in black.

Raman Shift (cm-1)	Vibrational mode assignment
1646	Amide I[58]
1613	Tyr[59]
1402	COH bend; $(CH_2)_n$ in-phase twist, COC str[59]
1350	Amide III[59], Trp[39]
1284	COH bend, Amide III,[35] CH in-plane (lipid)[35]
1137	C-N and C-C stretch[59], deoxyribose phosphate[35, 37]
1051	Gln, C-N stretch[41]
1005	Phenylalanine[59]
959	C-C stretch[39], PO_4[36]
932	Thr, Trp, Glu, Gln, Asp, Met, His C-COO stretch Tyr[41]
895	COC str[59]
860	C-C str, COC-1,4 glycosidic link[59]
812	Xylose,[38] O-P-O[35]
786	Cystosine, Uracil (stretch, ring)[59], O-P-O symmetric stretch[37]
767	Trp[41]; Glucose, Galactose[38]
662	Guanine,[59] C-S[39]
607	COO – wag[41]
556	Trp, C-SS-C[35, 39]
500	Deoxyribose phosphate[37]
465	Protein S-S stretching[60]

Table 2. Information on *M. pneumoniae* cell culture preparations used in NA-SERS sample datasets.

Culture prep type	CFU/ml	Protein conc. (µg/ml)	Genome Equivalents (cells/ml)	DNA conc. (µg/ml)
Intact (a)	$8x10^7$	310	$7.3x10^{10}$	6.47
Intact (b)	$5x10^8$	250	$5.4x10^{10}$	4.77
Intact (c)	$2x10^8$	540	$7.1x10^{10}$	6.27
Mean ± Std. deviation	$2.6x10^8 \pm 2.2x10^8$	370 ± 153	$6.6x10^{10} \pm 1.0x10^{10}$	5.48 ± 0.93

Table 3. NA-SERS multivariate detection limit for *M. pneumoniae* based on the data presented in Table 2.

Intact-cell culture dataset	MDL by CFU/ml	MDL by protein concentration (fg/µl)	MDL by genome equivalents (cells/ml)	MDL by DNA concentration (fg)
(a)	5.4	26.6	5110	218.2
(b)	39.6	23.0	4390	215.6
(c)	16.0	39.8	6440	300.0
Mean ± Std. deviation	20.3 ± 17.5	29.8 ± 8.8	5313 ± 1040	244.6 ± 48.0

Table 4. Information on *M. pneumoniae* cell culture preparations used in qPCR sample datasets.

qPCR dataset	CFU/ml	Protein conc. (µg/ml)	Genome equivalents (cells/ml)	DNA content (µg/ml)
(a)	2.03×10^{8}	190	2.3×10^{11}	19.95
(b)	2.53×10^{8}	175	2.8×10^{11}	25.03
(c)	2.79×10^{8}	220	2.5×10^{11}	22.33
Mean ± Std. deviation	$2.45 \times 10^{8} \pm 3.86 \times 10^{7}$	195 ± 23	$2.5 \times 10^{11} \pm 2.5 \times 10^{10}$	22.45 ± 2.54

Table 5. Lower limit of detection of *M. pneumoniae* by qPCR analysis, based on initial culture data presented in Table 4.

qPCR dataset	LOD by CFU/ml	LOD by genome equivalents (cells/ml)	LOD by DNA concentration (fg)
(a)	2.03	2300	39.8
(b)	2.53	2800	49.8
(c)	2.79	2500	44.4
Mean ± Std. Deviation	2.45 ± 0.39	2533 ± 251	44.7 ± 5.00

References

1. K. Loens, H. Goossens and M. Ieven, *Eur J Clin Microbiol Infect Dis*, 2010, **29**, 1055-1069.
2. K. B. Waites, M. F. Balish and T. P. Atkinson, *Future Microbiol*, 2008, **3**, 635-648.
3. F. Daxboeck, R. Krause and C. Wenisch, *Clin Microbiol Infect*, 2003, **9**, 263-273.
4. C. M. Kung and H. L. Wang, *Jpn J Infect Dis*, 2007, **60**, 352-354.
5. K. P. Thibodeau and A. J. Viera, *Am Fam Physician*, 2004, **69**, 1699-1706.
6. V. Chalker, T. Stocki, D. Litt, A. Bermingham, J. Watson, D. Fleming and T. Harrison, *Euro Surveill*, 2012, **17**.
7. S. L. Hennigan, J. D. Driskell, R. A. Dluhy, Y. Zhao, R. A. Tripp, K. B. Waites and D. C. Krause, *PloS one*, 2010, **5**, e13633.
8. J. M. Winchell, K. A. Thurman, S. L. Mitchell, W. L. Thacker and B. S. Fields, *J Clin Microbiol*, 2008, **46**, 3116-3118.
9. R. Dumke, P. C. Luck, C. Noppen, C. Schaefer, H. von Baum, R. Marre and E. Jacobs, *J Clin Microbiol*, 2006, **44**, 2567-2570.
10. K. A. Thurman, N. D. Walter, S. B. Schwartz, S. L. Mitchell, M. T. Dillon, A. L. Baughman, M. Deutscher, J. P. Fulton, J. E. Tongren, L. A. Hicks and J. M. Winchell, *Clin Infect Dis*, 2009, **48**, 1244-1249.
11. K. Kanack, E. Amiott, F. Nolte, H. Saliminia, B. Rogers, M. Poritz, B. Lingenfelter and K. Ririe, Idaho Technology, Inc., 2011.
12. J. L. Abell, J. D. Driskell, R. A. Dluhy, R. A. Tripp and Y. P. Zhao, *Biosens Bioelectron*, 2009, **24**, 3663-3670.
13. J. D. Driskell, S. Shanmukh, Y. Liu, S. B. Chaney, X. J. Tang, Y. P. Zhao and R. A. Dluhy, *J Phys Chem C*, 2008, **112**, 895-901.
14. Y. J. Liu, Z. Y. Zhang, Q. Zhao, R. A. Dluhy and Y. P. Zhao, *J Phys Chem C*, 2009, **113**, 9664-9669.
15. D. F. Willemse-Erix, M. J. Scholtes-Timmerman, J. W. Jachtenberg, W. B. van Leeuwen, D. Horst-Kreft, T. C. Bakker Schut, R. H. Deurenberg, G. J. Puppels, A. van Belkum, M. C. Vos and K. Maquelin, *J Clin Microbiol*, 2009, **47**, 652-659.
16. M. Harz, P. Rösch and J. Popp, *Cytometry Part A*, 2009, **75A**, 104-113.
17. A. Otto, *Journal of Raman Spectrosc*, 2002, **33**, 593-598.
18. R. S. Golightly, W. E. Doering and M. J. Natan, *ACS Nano*, 2009, **3**, 2859-2869.
19. S. Shanmukh, L. Jones, J. D. Driskell, Y. Zhao, R. A. Dluhy and R. A. Tripp, *Nano Letters*, 2006, **6**, 2630-2636.
20. H. Chu, Y. J. Huang and Y. Zhao, *Appl Spectrosc*, 2008, **62**, 922-931.
21. J. D. Driskell, Y. Zhu, C. D. Kirkwood, Y. Zhao, R. A. Dluhy and R. A. Tripp, *PloS one*, 2010, **5**, e10222.
22. P. A. Granato, L. Poe and L. B. Weiner, *J Clin Microbiol*, 1983, **17**, 1077-1080.
23. R. P. Lipman, W. A. Clyde, Jr. and F. W. Denny, *J Bacteriol*, 1969, **100**, 1037-1043.

24. P. K. Smith, R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson and D. C. Klenk, *Anal Biochem*, 1987, **150**, 76-85.
25. R. Himmelreich, H. Hilbert, H. Plagens, E. Pirkl, B. C. Li and R. Herrmann, *Nucleic Acids Res.*, 1996, **24**, 4420-4449.
26. M. H. Diaz and J. M. Winchell, *Diagn Microbiol Infect Dis*, 2012, **73**, 278-280.
27. J. Logan, K. Edwards and N. A. Saunders, *Real-time PCR: Current Technology and Applications*, Caister Academic Press, Norfolk, UK, 2009.
28. C. E. Romero-Arroyo, J. Jordan, S. J. Peacock, M. J. Willby, M. A. Farmer and D. C. Krause, *J Bacteriol*, 1999, **181**, 1079-1087.
29. C. L. Leverette, S. A. Jacobs, S. Shanmukh, S. B. Chaney and R. A. Dluhy, *Appl Spectrosc*, 2006, **60**, 906-913.
30. Y. P. Zhao, S. B. Chaney, S. Shanmukh and R. A. Dluhy, *J Phys Chem B*, 2006, **110**, 3153-3157.
31. P. Negri, N. E. Marotta, L. A. Bottomly and R. A. Dluhy, *Appl Spectrosc*, 2011, **65**, 66-74.
32. K. B. Waites and D. F. Talkington, *Clin Microbiol Rev*, 2004, **17**, 697-+.
33. J. M. Hatchel and M. F. Balish, *Microbiol*, 2007, **154**, 286-295.
34. J. R. Maniloff, R. N. McElhaney, L. R. Finch and J. B. Baseman, *Mycoplasmas: molecular biology and pathogenesis*, American Society of Microbiology, Washington, D.C., 1992.
35. M. Culha, A. Adigüzel, M. M. Yazici, M. Kahraman, F. Sahin and M. Güllüce, *Appl Spectrosc*, 2008, **62**, 1226-1232.
36. R. M. Jarvis, E. W. Blanch, A. P. Golovanov, J. Screen and R. Goodacre, *Analyst*, 2007, **132**, 1053-1060.
37. J. Y. Ling, Q. Z. Yang, S. S. Luo, Y. Li and C. K. Zhang, *Chin Chem Lett*, 2005, **16**, 71-74.
38. M. F. Mrozek and M. J. Weaver, *Anal Chem*, 2002, **74**, 4069-4075.
39. E. Podstawka, Y. Ozaki and L. M. Proniewicz, *Appl Spectrosc*, 2004, **58**, 1147-1156.
40. S. Stewart and P. M. Fredericks, *Spectrochim Acta Part A*, 1998, **55**, 1615-1640.
41. S. Stewart and P. M. Fredericks, *Spectrochim Acta Part A*, 1998, **55**, 1641-1660.
42. K. Maquelin, L.-P. Choo-Smith, C. Kirschner, N. A. Ngo-Thi, D. Naumann and G. J. Puppels, in *Handbook of Vibrational Spectroscopy*, eds. J. M. Chalmers and P. R. Griffiths, John Wiley & Sons, Ltd., Chichester, 2002, vol. 5, pp. 3308-3334.
43. D. Naumann, in *Biomedical Vibrational Spectroscopy*, eds. P. Lasch and J. Kneipp, J. Wiley & Sons, Hoboken, 2008, pp. 1-8.
44. M. Harz, P. Roesch, K. D. Peschke, O. Ronneberger, H. Burkhardt and J. Popp, *Analyst*, 2005, **130**, 1543-1550.
45. K. Maquelin, C. Kirschner, L. P. Choo-Smith, N. A. Ngo-Thi, T. van Vreeswijk, M. Stammler, H. P. Endtz, H. A. Bruining, D. Naumann and G. J. Puppels, *J. Clin. Microbiol.*, 2003, **41**, 324-329.
46. D. Graham and R. Goodacre, *Chem Soc Rev*, 2008, **37**, 883-884.

47. R. Jarvis, S. Clarke and R. Goodacre, in *Surface-Enhanced Raman Scattering: Physics and Applications*, 2006, vol. 103, pp. 397-408.
48. A. C. Oliveri, N. K. M. Faber, J. Ferre, R. Boque, J. H. Kalivas and H. Mark, *Pure Appl Chem*, 2006, **78**, 633-661.
49. R. Boqué, M. S. Larrechi and F. X. Rius, *Chemom Intell Lab Syst*, 1999, **45**, 397-408.
50. M. Ostra, C. Ubide, M. Vidal and J. Zuriarrain, *Analyst*, 2008, **133**, 532-539.
51. P. F. Smith, *The Biology of Mycoplasmas*, Academic Press, Inc., New York, New York, 1971.
52. M. V. Zubkov, B. M. Fuchs, H. Eilers, P. H. Burkill and R. Amann, *Appl Environ Microbiol*, 1999, **65**, 3251-3257.
53. M. Barker and W. Rayens, *J Chemom*, 2003, **17**, 166-173.
54. G. Musumarra, V. Barresi, D. F. Condorelli, C. G. Fortuna and S. Scire, *J Chemom*, 2004, **18**, 125-132.
55. B. M. Cullum, J. Mobley, Z. Chi, D. L. Stokes, G. H. Miller and T. Vo-Dinh, *Rev Sci Instrum*, 2000, **71**, 1602-1608.
56. D. S. Moore and R. J. Scharff, *Anal Bioanal Chem*, 2009, **393**, 1571-1578.
57. J. Jehlička, A. Culka, P. Vandenabeele and H. G. M. Edwards, *Spectrochim Acta Part A*, 2011, **80**, 36-40.
58. X. Yang, C. Gu, F. Qian, Y. Li and J. Z. Zhang, *Anal Chem*, 2011, **83**, 5888-5894.
59. K. Maquelin, C. Kirschner, L. P. Choo-Smith, N. van den Braak, H. P. Endtz, D. Naumann and G. J. Puppels, *J Microbiol Methods*, 2002, **51**, 255-271.
60. T. A. Alexander, *Anal Chem*, 2008, **80**, 2817-2825.