Sensors & Diagnostics

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: A. B. Walter, L. Whitehead, A. Taylor and A. Locke, *Sens. Diagn.*, 2025, DOI: 10.1039/D5SD00092K.

View Article Online
DOI: 10.1039/D5SD00092K

High-wavenumber Raman spectroscopy for the detection of *Mycobacterium tuberculosis* in saliva

Alec B. Walter[1,2], Luke Whitehead[1,2], Amelia Taylor[1,2], Andrea K. Locke[1,2,3,*]
[1]Vanderbilt Biophotonics Center, Vanderbilt University, Nashville, TN, USA
[2]Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA
[3]Department of Chemistry, Vanderbilt University, Nashville, TN, USA

Abstract

Despite the increased availability of low-cost and effective treatments for tuberculosis (TB), ~1 million people continue to die from TB-related symptoms annually. One major challenge limiting the effectiveness of TB treatment is delays in diagnosis, largely due to current detection methods requiring either weeks of culture time or complex processing steps that cannot be performed at the point-of-care. Thus, there is a need for alternative methods that are easier to use yet still effective in providing an accurate TB diagnosis. This work investigates the feasibility of using high-wavenumber Raman spectroscopy to detect the presence of the causative agent of TB*, Mycobacterium tuberculosis,* in human saliva. To accomplish this, raw saliva was collected from healthy participants and inoculated with a fixed, physiologically relevant, concentration of bacteria (10^6 CFU/mL) and concentrated into a pellet. The samples were measured using Raman spectroscopy and analyzed with a spectral unmixing approach to determine the relative biochemical composition. The presence of *M. tuberculosis* resulted in a significant increase in the lipid signal of saliva pellets containing the spiked bacteria, with a median percent increase of 423.6% as compared to the control samples. Control experiments using *Streptococcus mutans*, a common oral bacterium, only resulted in a slight increase of 9.8%. Additionally, using linear regression analysis, a predictive relationship was found between the Raman lipid fractions of the raw saliva and the control saliva pellets. Using the 95% prediction interval of this relationship as a classification threshold, the presence of *M. tuberculosis* was accurately determined for all samples with an overall training accuracy of 98.5% and a cross-validation accuracy of 100%. These results showcase the potential of high-wavenumber Raman spectroscopy as a reagent-free method of detecting *M. tuberculosis* in saliva samples.

View Article Online
DOI: 10.1039/D5SD00092K

*Andrea K. Locke, E-mail: andrea.locke@vanderbilt.edu

Introduction

Tuberculosis (TB) remains one of the top infectious diseases, resulting in a high global burden. Despite the availability of low-cost and effective TB treatment, annually, approximately 10 million people acquire TB, and over 1 million people die from disease-related symptoms.[1,2] During the COVID-19 pandemic, there was a rise in TB cases and deaths as well as drug-resistant cases, according to the World Health Organization.[3] This may have been due to the disruption in services and screening due to the COVID-19 focus.[4]The pathogen driving TB infections is *Mycobacterium tuberculosis* (*M. tuberculosis*), a bacterial pathogen that is transmitted through the air. One of the main challenges with TB treatment is the late-stage diagnosis due to self-reported symptoms not occurring until the late onset of the disease.[5] Another challenge is the issue of latent TB (LTBI). The TB pathogen is transmitted airborne and can live from three months to two years in its host without active signs of the disease. About 5-10% of LTBI cases will go on to develop active TB which can then result in the spreading of the disease to others.[6] Typically, the first signs are noted in the lungs, and then evidence of the disease can be found in other parts of the body, detectable in bodily fluids such as urine, blood, or saliva.[7–10]

The gold-standard tools for detecting bacterial infections are not always translatable to the point of care for certain pathogenic species. These techniques rely on culture-based methods or molecular sensing for high accuracy.[11] However, culture-based methods can take days to grow for an adequate pathogen count required for sensing, which limits applicability for rapid detection in the field. In the case of *M. tuberculosis*, it takes an estimated four to eight weeks for an adequate colony count.[12] Thus, alternative approaches for detection are required, such as deoxyribonucleic acid (DNA) amplification via polymerase chain reaction (PCR) or immunoassays. While commercially-available, urine-based, lipoarabinomannan (LAM) antigen tests are an easy-to-use, rapid diagnostic, they suffer from poor sensitivity (<50%), especially for individuals who are not HIV positive or those with high white blood cell counts.[13] Overall, immunoassays such as the antigen-based test either lack sensitivity or specificity (<70%), while the skin test has

View Article Online
DOI: 10.1039/D5SD00092K

a high rate of false positives. PCR-based techniques such as the Xpert tests are complex, requiring multiple reagents, buffers, and standards for accurate diagnosis, resulting in a high financial cost. Thus, there is still a need for cost-effective, easy-to-use, culture-free platforms that deliver robust and accurate pathogen detection at the point of care.

Another challenge is the type of sample collected to confirm the presence of the *M. tuberculosis*. The most common practice is the assessment of sputum samples from suspected patients. However, sputum samples, acquired in a cough-like manner, can be difficult to produce by most patients and present additional safety concerns for healthcare workers. Moreover, children and those living with HIV are often unable to provide this type of sample.[14,15] This leads to a need for alternative sample types that are less invasive but still effective in allowing for a proper TB diagnosis. In recent years, tongue swabs and saliva samples have been examined as alternative samples for the detection.[16] Oral swabs analysis (i.e., tongue swabs) and saliva testing, is known to be faster, safer, and easier to collect compared to sputum. However, these samples have also been shown to carry a significantly smaller bacterial load, averaging around 1% the amount found in sputum samples [17–21]. Despite this, researchers have shown sensitivity between 65-90% and specificity between 95-100% when analyzing tongue swabs versus sputum using the Xpert PCR analysis system with the sputum reference standard.[22] Other researchers have examined saliva samples as another alternative approach to sputum collection since it was widely used for rapid detection of COVID-19 pathogen. In general, the analysis of saliva has proven to be comparable in sensitivities in the low 90% for HIV-negative cases.[23] One drawback is both methods utilized a form of PCR to assess the presence of the *M. tuberculosis*. Access to Xpert PCR systems is still limited due to cost, and the turn-around time for sample analysis is much greater than 20 minutes, which is not ideal based on the World Health Organization criteria. Therefore, the call for new diagnostic tools for detecting *M. tuberculosis* in a variety of sample types was a challenge put forth in 2018.[24] In addition, the need for rapid, culture-free technologies could potentially help the early detection of the vast majority of TB cases that typically go undetected for long periods of time. Herein, we assessed the design a culture-free platform using high-wavenumber Raman spectroscopy for rapid detection of the pathogen using its biochemical "fingerprint."

View Article Online
DOI: 10.1039/D5SD00092K

Raman spectroscopy (RS) is an analytical technique that provides a unique spectral fingerprint representative of the biochemical components of a specimen, based on light interaction with the chemical bonds within a sample. This interaction results in inelastic scattering when the frequency of the input light is close to the chemical bonds' vibrational frequencies.[25] For biological applications, RS has been well-documented for the analysis of biomolecules such as lipid membranes, proteins, amino acids, carbohydrates, nucleic acids, and glucose, amongst others.[26] It has a wide range of applications, from *in vivo* detection of cancerous tissues and inflammatory sites to *ex vivo* application of bacterial differentiation, including drug-resistant strains.[27–30] Overall, Raman spectroscopy is a label-free, non-destructive method that can be used *in vivo* or *ex vivo* for rapid assessment of collected tissue specimens or biofluids.[31,32]

While most biological Raman studies have utilized the fingerprint region (400-1800 cm^{-1}) of the spectrum due to its multiple discrete peaks, there has been an increased interest in using the high-wavenumber region (2800-3800 cm^{-1}) instead. This is because the high-wavenumber region has been shown to provide significantly stronger Raman intensities, as compared to the autofluorescence background, without the need to include external agents, such as with surface-enhanced Raman spectroscopy.[32] This results in an improvement in the accuracy of spectral processing and a decrease in the required measurement times when specifically designed spectrometers are used. Additionally, despite only having two major Raman bands that are representative of Oxygen-Hydrogen bonds and Carbon-Hydrogen ($-CH_x$) bonds, the high-wavenumber region has been shown to provide comparable diagnostic accuracies as the fingerprint region, indicating that these two bands are similarly rich in biochemical information.[33–35] Moreover, the high-wavenumber Raman peak at 2850 cm^{-1} is characteristic of lipids and not shared with the other classes of biomolecules.[26,36] This is in contrast with the fingerprint region which has greater number of lipid-associated peaks that commonly overlap with those of proteins and carbohydrates.[26,36,37] These overlaps can confound the detection of lipids in complex media, especially when they are at relatively lower concentrations. As *Mycobacterium* species are characterized by their cell wall-associated lipids, known as mycolic acids, the ability to easily measure the presence of lipids with high-wavenumber Raman

View Article Online
DOI: 10.1039/D5SD00092K

spectroscopy is expected to further improve the capability of detecting *M. tuberculosis* in a biologically complex envirnonment.[38–41]

Herein, we demonstrate the feasibility of utilizing high-wavenumber RS (HWRS) for the detection of *M. tuberculosis* strain H37Ra spiked in human saliva samples. We show that the HW spectra of the bacteria generate a strong lipid-based spectrum, which we believe is associated with its mycolic acid components. We also show that using a non-negative least-squares spectral unmixing approach with known endmember spectra, we can determine the fraction of lipids present within the pelleted cells of the saliva and use this fraction to detect the presence of the biochemical fingerprint of the bacteria. This work demonstrates the potential for HWRS to be used as a label-free method for detecting TB, requiring no reagents and minimal sample preparation for translation to point-of-care applications.

Experimental

Participant Recruitment and Saliva Collection

This study was approved by the Vanderbilt University Medical Center Institutional Review Board (IRB #191004). For this work, 9 participants with no known oral diseases were recruited after obtaining written informed consent. Saliva collection was performed by asking participants to slowly spit into a sterile collection tube (5016.02, Salimetrics) without stimulating the production of saliva or rinsing out the mouth. In total, 12 saliva samples were collected and utilized in this work. Half of these samples were collected from the participants in a non-fasting state, usually after midday. The other half of the samples were collected with the participants in a fasting state. This was accomplished by providing the participants with sterile collection tubes and asking them to collect a sample themselves the following morning before eating or performing any oral hygiene. Once received, the samples were aliquoted and stored at -80 °C until used.

Bacterial Species and Culturing

The H37Ra strain of *Mycobacterium tuberculosis* (ATCC #25177) and the UA159 strain of *Staphylococcus mutans* (*S. mutans*; ATCC #700610), a bacterial species commonly found in the oral cavity, were used in this work.[42] Liquid cultures of *M. tuberculosis* were made by diluting 0.5 mL of frozen stock into 9.5 mL of Middlebrook 7H9 broth (M0178,

View Article Online
DOI: 10.1039/D5SD00092K

Millipore Sigma) supplemented with 10% albumin-dextrose-catalase growth supplement (BD 211887, Fisher Scientific), 0.2% glycerol (G7893, Sigma- Aldrich), and 0.1% Tween-80 (P4780, Sigma-Aldrich). Cultures were grown for 6 days on a rolling incubator at 37 °C and 5% CO_2. *S. mutans* was cultured from frozen stock on brain heart infusion (BD 237500, Fisher Scientific) agar plates at 37 °C and 5% CO_2 for 24 hours. Liquid cultures were made by collecting and inoculating single colonies into 3 mL of brain heart infusion broth and incubating for an additional 24 hours. All liquid cultures were arrested at 4 °C for at least 24 hours prior to use.

High-wavenumber Raman Characterization

Before being used together, high-wavenumber Raman spectra of the fasting saliva, non-fasting saliva, and bacterial species were determined. Whole saliva samples were prepared without any additional processing by depositing three 1 µL droplets onto an aluminum foil-coated (Reynolds Wrap, Reynolds) microscope slide and allowing them to dry for at least 5 minutes. This process was repeated for each of the collected saliva samples. To prepare the bacterial cultures for measurement, they were first washed to remove the liquid media. 1 mL culture samples were washed twice with sterile phosphate-buffered saline, centrifuging at 3300g for 8 minutes each time. To maximize the signal collected from the bacteria, the bacterial pellet was isolated by centrifuging a third time and removing the supernatant. Similar to the saliva, samples were prepared by depositing three 1 µL droplets of the bacterial pellet onto an aluminum foil-coated microscope slide. This process was repeated three time for both species, using a different culture each time to account for any biological variability.

High-wavenumber Raman measurements of the dried saliva and bacteria samples were acquired using a Renishaw inVia Qontor Raman microscope with a 785 nm excitation laser. The wavenumber axis for the system was calibrated before each set of measurements using an internal silicon standard. Sample spectra were acquired using a 50x/0.5NA objective which supplied ~90 mW of laser power at the sample. Each measurement was obtained using a total integration time of 50 seconds (5 accumulations with 10 seconds of exposure) to compensate for the low quantum efficiency of the detector across the measurement range. Three spectra were acquired along the edge of each dried droplet where the material was thickest, resulting in 9 total spectra per sample.

View Article Online
DOI: 10.1039/D5SD00092K

Spectral Preprocessing

All acquired high-wavenumber Raman spectra were preprocessed prior to analysis. Collected spectra were first calibrated against the spectral response of the system using a 785 nm luminescence standard reference material (SRM 2242, NIST), before being truncated from 2765 to 3205 cm^{-1}. The autofluorescence background of the measurements were removed using a 1st order polynomial estimate obtained through the modified polynomial method.[43] Background estimation was performed while excluding the region from 2800 and 3100 cm^{-1} to prevent the strong -CH_x band from influencing the shape of the estimated background. Noise reduction was accomplished by smoothing the resulting Raman spectra using a 2nd order Savitsky-Golay filter with a window size of 11, corresponding to a spectral width of approximately 8.25 cm^{-1}, matching established smoothing parameters for fingerprint Raman spectra.[44] Lastly, to isolate the -CH_x band for further analysis, the resulting spectra were binned using a 1 cm^{-1} bin-width and truncated between 2800 and 3100 cm^{-1}.

Saliva Sample Preparation

Prior to use, arrested bacterial cultures were removed from storage and washed twice using sterile deionized water containing 0.001% Tween-80. Optical density readings at 600 nm were made on the washed cultures to quantify the bacterial concentrations. The concentrations of the measured cultures were subsequently adjusted to be $2x10^7$ CFU/mL. This process was kept consistent for both *M. tuberculosis* and *S. mutans*. After preparing the bacterial culture, a saliva sample was removed from -80 °C storage and allowed to thaw at room temperature. Once completely liquid, the saliva was vortexed for 2-3 seconds to ensure homogenization before aliquoting 95 uL into two sterile microcentrifuge tubes. Using these two aliquots, three different measurement samples were made for each saliva sample.

In one of the tubes, 5 µL of sterile deionized water containing 0.001% Tween-80 was added to serve as a negative control. In the other, 5 µL of the washed $2x10^7$ CFU/mL *M. tuberculosis* culture was spiked into the saliva, resulting in final sample concentration of 10^6 CFU/mL. This represents the upper range reported for the *M. tuberculosis* load in saliva, tongue-swab, and other non-sputum oral samples during an active infection.[17–21] As such, it serves as an ideal concentration for this feasibility study by falling within

View Article Online
DOI: 10.1039/D5SD00092K

physiological parameters while providing a sensitivity buffer for the new and unoptimized diagnostic approach. For the first measurement sample, designated as the raw saliva sample, a 1 µL droplet of the spiked saliva was dropped onto a microscope slide coated with non-stick aluminum foil (Reynolds Wrap Nonstick, Reynolds) and allowed to dry. Non-stick aluminum foil was utilized as its modestly hydrophobic nature serves to reduce the size of the dried droplet, thus increasing the effective concentration, while contributing minimal background to the Raman measurement. After plating the raw saliva sample, both saliva tubes were concentrated by centrifuging at 3,300g for 5 minutes, removing 95% of the total volume from the supernatant, and resuspending the cellular pellet. Similar to the raw saliva sample, a 1 µL droplet from both the bacteria-containing pellet sample and the control pellet sample was placed on the nonstick aluminum foil and allowed to dry. In total, this process was repeated for 4 of the fasting saliva samples and 4 of the non-fasting saliva samples. The remaining 4 sets of saliva were used to create a set of control samples where everything was kept the same except that the *M. tuberculosis* was replaced with the same concentration of *S. mutans*.

Raman Mapping

High-wavenumber Raman spectra of the raw, control pellet, and bacteria-containing pellet samples were acquired using a Renishaw inVia Qontor Raman microscope. As the localization of the bacteria within the dried droplets is unknown, measurements were taken across all relative radial locations throughout the droplets. This was accomplished by collecting a Raman map consisting of a series of measurements around the edge of a circle inscribed between the center and the edge of the droplet (see Figure 2). Measurements were taken 25 µm apart using LiveTrack, the automatic focus tracking feature built into WiRE 5.6 (Renishaw), to account for variable sample height across the droplet. Measurements at each location were acquired in a similar manner as was previously described. A 50x/0.5NA objective was used to supply ~90 mW of 785 nm laser power to the sample, and high-wavenumber Raman measurements were acquired using a total of 3 accumulations with 10 seconds of exposure each. Each spectrum was preprocessed using the same methodology as described above except that the spectra were smooth using an increased Savitzky-Golay windows size of 21, corresponding to an approximate spectral width of 15.75 cm^{-1}. While this is wider than what is traditionally

View Article Online
DOI: 10.1039/D5SD00092K

used, the spectral features of the high-wavenumber region are, on average, broader than the fingerprint region. Therefore, it was found that this combination of order and window size minimized the increased noise present in the mapping measurements without significantly altering the shape or intensity of the -CH_x band. However, some of the measurements in the map were found to have too low of a Raman signal to be usable, even after smoothing, due to a lack of sample material or inadequate autofocusing at different locations across the region of interest. To account for such measurements, the signal-to-noise ratio (SNR) of each spectrum was determined before any smoothing was applied. SNR was approximated as the average Raman value between 2910 and 2950 cm^{-1}, the major protein-dominated peak of the -CH_x band, divided by the standard deviation of the values found with the lower signal region between 3000 and 3100 cm^{-1}. Spectra that were found to have an SNR less than 8 were excluded from any subsequent analysis.

Spectral Unmixing Analysis

To approximate the composition of the dried saliva samples, a total of 8 pure components were selected across the 4 major biochemical categories of proteins, nucleic acids, carbohydrates, and lipids.[35] Salivary proteins were represented by amylase (A3176, Sigma-Aldrich) and γ-globulin (G5009, Sigma-Aldrich), two of the most abundant proteins found in saliva, while bovine serum albumin (A2153, Sigma-Aldrich) was used to approximate the remaining population of general proteins.[45] DNA (D1501, Sigma-Aldrich) was used to represent any nucleic acids found in saliva, while carbohydrates were represented by glycogen (G0855, Sigma-Aldrich). As the focus of this analysis, three different lipid components were used to account for any potential sources of variability. Oleic acid (364525, Sigma-Aldrich) was used as the representative free fatty acid, the most common type of salivary lipid, while phosphatidylethanolamine (PE; P1223, Sigma-Aldrich) was chosen to represent the less common polar salivary lipids.[46] Finally, to account for the unique *Mycobacterium* lipid inclusion, the most abundant form of mycolic acid produced by *M. tuberculosis*, α-mycolic acid (791280; Avanti Research), was also included.[39] High-wavenumber Raman spectra of the pure components were acquired and processed using the methodology described above. Three separate spectra were taken

View Article Online
DOI: 10.1039/D5SD00092K

for each component, with the average of the normalized spectra being used as the endmembers in the subsequent spectral unmixing analysis.

Spectral unmixing analysis was performed on each viable spectrum obtained from the Raman maps using a non-negative least-squares approach.[35] Prior to analysis, both the target and component spectra were normalized such that their respective area-under-the-curves (AUC) were equal to 1. This helps to directly approximate the fractional contributions each component has to the measured Raman spectra. All eight components described above were used as the endmembers to spectrally unmix each high-wavenumber Raman measurement of the different Raman maps. The overall fitting error for each spectrum was determined by summing the absolute residuals and multiplying by 100. Due to the AUC normalization, this metric approximates the absolute percentage fitting error. After unmixing, the individual component contributions were combined to obtain an estimate of the Raman fractions originating from the broader biochemical classes.

Statistical Analysis and Detection Model

The Raman lipid fractions, obtained from the spectral unmixing analysis, of the paired control and bacteria-spiked pellet samples were compared against each other using a Wilcox rank sum test. The change in the relative lipid content caused by the centrifugation and concentration of the saliva was assessed using linear regression analysis with individual samples represented by the average Raman lipid fraction across their viable map measurements. The linear regression compared the values of the raw saliva samples to that of their respective control pellet sample, fixing the intercept at 0 under the assumption that having no detectable lipid signal in the raw saliva would result in the same for the concentrated sample.

In addition to the linear fit, the 95% prediction interval (α=0.05) of the regression was determined and compared to the results of the bacteria-spiked samples. To evaluate if this relationship could be used to detect the presence of *M. tuberculosis*, the upper bound of the prediction interval was set as the classification threshold, with values exceeding the interval being categorized as containing *M. tuberculosis*. The performance of this classification model was assessed using a leave-one-sample-out cross-validation. For each fold of the cross-validation, both the control pellet and the bacteria-spiked pellet

View Article Online
DOI: 10.1039/D5SD00092K

from the test sample were classified. To further assess the model's performance, an analysis of the training accuracy was conducted in parallel with the cross-validation. This was accomplished by classifying each of the bacteria-spiked samples that were paired with the training data, for each fold of the cross-validation. While not a direct metric of performance, this analysis provides insight into the robustness of the model by determining if any significant changes are occurring in the construction of the model due to withholding specific training data.

Results and Discussion

High-wavenumber Characterization

To assess the feasibility of using high-wavenumber Raman spectroscopy to detect the presence of *M. tuberculosis* in saliva, an initial characterization of saliva collected before and after a minimum 6-hour overnight fast was performed. As seen in Figure 1, overnight fasting had minimal effects on the high-wavenumber Raman spectra, with both fasting and non-fasting samples remaining primarily protein dominant with the commonly associated peak at 2935 cm^{-1}.[26,47] However, it was observed that fasting served to reduce the spectral differences between the saliva samples, particularly around the 2850 and 2975 cm^{-1} shoulders, which are indicative of lipid and protein/nucleic acid content, respectively.[26,36,37]This variability could become impactful when trying to detect *M. tuberculosis* as the primary feature that distinguishes the pathogen from saliva is a strong 2850 cm^{-1} lipid peak (Figure 1). The strength of this peak, which also distinguishes it from other oral bacteria like *S. mutans*, is likely due to the abundance of mycolic acids, which are long-chain fatty acids that *Mycobacterium* species incorporate into their cell walls.[39] The presence of mycolic acids has been shown to be indicative of *M. tuberculosis* infection, and improved detection methods are being investigated.[38,40,41] With the high-wavenumber lipid peak having minimal overlap with other biological components, and the observed low lipid content of human saliva, these results support the potential for rapid and nondestructive detection of *M. tuberculosis,* or its shed mycolic acids, in saliva samples using HWRS.

Saliva Raman Maps

View Article Online
DOI: 10.1039/D5SD00092K

To assess the use of HWRS for *M. tuberculosis* detection in saliva, the pathogen was spiked at physiologically relevant concentrations and Raman maps were performed across the dried droplets to mitigate any variation in biochemical signal due to heterogeneous drying. A representation of the processing steps for the Raman spectral maps is show in Figure 2. With each map taking the form of a circle inscribed between the center and edge of the dried droplet, sampling was achieved across all relative radial distances. Due to differences in spot sizes between the dried saliva samples, the total number of measurements taken varied between samples, with an average number of 67 per sample. Out of these, a number of Raman measurements were unusable due to a poor signal-to-noise ratio. This was caused by both bare spots, where minimal solid materials were deposited, and thicker regions where too much material was deposited and the autofocus could not work against the dominance of the reflective substrate. Using a threshold value of 8 for an acceptable SNR, an average of 70% of points in each map were deemed viable and used for further analysis. Overall, the remaining Raman spectra had an average SNR of 13.3, with a standard deviation of 2.64.

From these usable portions of the Raman maps, no particular region of the dried droplets was found to contain more of the *M. tuberculosis* induced lipid signal than the others. This indicates that the bacteria and their lipid products are accumulating throughout the entirety of the droplet. This is likely a consequence of the material properties of the substrate, including the hydrophobicity and the surface roughness, acting together to suppress the coffee ring effect.[48] Additionally, ellipsoidal particles, such as the rod-shaped *M. tuberculosis*, have also been shown to inhibit the formation of coffee rings, leading to a more uniform deposition during droplet drying.[49] As such, it can be concluded that no singular region of the droplet should be used alone. Instead, measurements approaching averages of the entire droplet should be utilized to account for any variances in the deposition location of the bacteria. For this work, all viable measurements within the map are used, with average results representing the sample as a whole. Future work using this approach will aim to leverage low magnification, large spot size measurements to sample most, if not all, of the droplet at once. This can be paired with optimization of both the physical properties of the substrate and the droplet

View Article Online
DOI: 10.1039/D5SD00092K

drying conditions to enable more direct control of the coffee ring effect, ultimately improving the consistency and concentration of the deposited particles.

Spectral Unmixing

After acquiring the Raman maps, spectral unmixing analysis was performed on the viable spectra, using 8 pure biochemical components, to determine the Raman lipid fractions of each droplet.[35] The -CH_x bands for the protein, nucleic acid, carbohydrate, and lipid components used in the spectral unmixing analysis can be found in Figure 3a. These four biochemical classes cover the majority of the spectral variation observed within the -CH_x band of biological samples. While a single representative biomolecule is expected to be sufficient to capture the nucleic acid and carbohydrate contribution of the saliva spectra, an increased number of protein and lipid components were used to account for the more nuanced spectral differences that can occur between samples. Since protein is the primary component of dried saliva, differences in the relative protein content can have significant effects on the overall spectral shape which, if unaccounted for, can severely diminish overall fit accuracy.[45] Additionally, as the primary goal of this analysis is to determine the lipid content of the sample, accounting for the spectral differences between the polar salivary lipids, non-polar salivary lipids, and mycolic acids expected in the samples is critical.[46]

The chosen pure components yielded accurate spectral fits for all samples when used as the endmembers of the spectral unmixing analysis, as evidenced by the representative result presented in Figure 3 that show the computed spectra to closely match the measured spectra. Despite variations in composition between saliva samples, the average relative fitting errors across all analyzed spectra were consistently close to 5% for the raw saliva (5.98%), control pellet (5.65%), and bacteria-spiked pellet (4.97%) samples. This represents an improvement to the fitting accuracy obtained from using the same approach for fingerprint spectra, likely due to the lower molecular specificity of the high-wavenumber region allowing the spectra of the selected representative biomolecules to better approximate the spectral shapes of the components within the saliva samples.[50–52] This has enabled the direct spectral unmixing approach to achieve reconstruction errors on par or better than those obtained when using automated endmember retrieval strategies on fingerprint spectra.[53,54] As expected, all samples were

View Article Online
DOI: 10.1039/D5SD00092K

found to be predominantly protein-based (61.3 - 92.2%), with varying levels of contribution attributed to nucleic acids, carbohydrates, and lipids. Across the 12 raw saliva samples, most were determined to have a minimal lipid fraction of 2% or less. As lipids have a higher Raman cross-section in the -CH_x region compared to the other classes of biomolecules, this represents an even smaller fraction of the overall mass.[36] However, a quarter of the samples, including both fasting and non-fasting samples, were found to have a much higher lipid fraction, ranging from 5 to 11%, than what was expected for normal saliva composition.[46] These increased values are likely due to either increased cellularity or remnant food debris within the samples, as no rinsing was performed prior to saliva collection. While this results in a greater variability in the measured samples, oral rinsing has been shown to affect the relative prevalence of bacteria within saliva sample, such procedures should likely be avoided for the diagnostic detection of *M. tuberculosis* until their effects are better understood.[55]

Despite these differences in the raw saliva composition, comparisons between the negative control and bacteria-spiked pellet samples (Figure 4) found that the inclusion of *M. tuberculosis* resulted in a significant increase in the Raman lipid fraction for both fasting and non-fasting saliva, with a median percent difference of 423.6%. Additionally, it was found that saliva spiked with the same concentration of *S. mutans* did not lead to a correspondingly significant increase in the Raman lipid fraction, with a median percent difference of only 9.8%. This indicates that the high lipid content of *M. tuberculosis*, thought to be due to their mycolic acids, induces a specific and detectable increase in the relative lipid content of concentrated saliva, when present at concentrations that have been reported during infection.[17–21] However, due to the variability observed in the amount of lipid present between samples, a single threshold value capable of determining *M. tuberculosis* presence could not be set. Moreover, as paired control saliva will not exist in a diagnostic setting to compare against, another approach that takes into account this inherent variability in saliva composition is required.

Linear Regression Model

As the variability in the lipid content was observed in both the concentrated and unconcentrated samples, it was hypothesized that a raw saliva measurement could be used to predict the expected lipid content of the saliva pellet. To understand how the

View Article Online
DOI: 10.1039/D5SD00092K

concentration process alters the observed lipid fraction, the relationship between the raw saliva and the control saliva pellet measurements was assessed using linear regression. For the 12 samples used in this work, a strong linear relationship (R^2=0.994) was found between the two sets of values with a fixed intercept through 0 and a slope of 1.647 (Figure 5a). Using this relationship, measurement of a raw saliva sample can be used to predict the lipid fraction of the corresponding pelleted sample. To account for processing, sampling, and other unknown sources of variability that may occur between samples, a range of acceptable values for this prediction was set using the 95% prediction interval of the linear fit. As seen in Figure 5b, when this range is compared to the results of the bacteria-spiked pellets, all samples containing *M. tuberculosis* were found to exceed the upper bounds of the prediction interval. In contrast, the samples containing *S. mutans* remained within the normally expected bounds. This suggests that the upper bounds of the prediction interval can serve as a detection threshold that scales with the initial composition of the saliva.

Using a leave-one-sample-out cross-validation approach, the accuracy of the proposed detection method was assessed. As can be seen from the resulting confusion matrix (Figure 5c), the accuracy in detecting the presence *M. tuberculosis* was maintained at 100% for both the negative controls and the bacteria-spiked samples. This showcases that the observed linear relationship is robust enough that it is maintained despite the removal of a sample. However, it was observed that not being trained on certain samples resulted in small, but noticeable, shifts in the linear fit and the bounds of the prediction interval. While these shifts did not result in classification errors for the test samples, an assessment of the model's training accuracy, made using the measurements of the bacteria-spiked samples paired to the training data (Figure 5d), found that the differences in the linear model caused some of the samples containing *S. mutans* to be inaccurately categorized as containing *M. tuberculosis*. This decrease in performance occurred when samples with higher lipid fractions were withheld from training, as their relative scarcity gave them more weight in the overall fit. As this relationship between raw and concentrated saliva becomes better characterized through an increased sample size, the potential impact that a single sample has on the model will be minimized. Additionally,

View Article Online
DOI: 10.1039/D5SD00092K

increasing the confidence level of the prediction interval would help prevent false positives at the risk of increased false negatives.

Despite this, however, the overall training accuracy of the model still displayed a strong performance with an average accuracy of 98.5%, positive predictive value of 97.8%, and sensitivity of 100% and the worst performing fold maintaining a training accuracy of 81.8%, a positive predictive value of 80%, and a sensitivity of 100%. Taken together, it can be concluded that a combination of spectral unmixing analysis and linear regression is capable of detecting the presence of *M. tuberculosis* in human saliva, showcasing the potential of diagnostic approaches leveraging high-wavenumber Raman spectroscopy.

However, there are a number of limitations that will need to be overcome to better establish the feasibility of using high-wavenumber Raman spectroscopy as a diagnostic tool for TB infection. The first limitation of this study centers on the type of samples used in this work. While spiking bacteria into human saliva mimics the complex fluid matrix that would be found in a clinical sample, it does not account for any changes that can be expected to occur due to an active TB infection. The cellular and biochemical composition of both the saliva and the *M. tuberculosis* are known to change in response to the immune response. While this work did account for variability in saliva composition between patients and fasting states, future work using confirmed clinical samples will be required to fully evaluate the effectiveness of the proposed technique. Second, before clinical samples can be effectively tested, the overall sensitivity of this high-wavenumber Raman approach will need to be determined. While this work demonstrated the potential to achieve high detection accuracies, the use of a concentration at the top of the reported range is not enough to draw conclusions about clinical viability as concentrations of *M. tuberculosis* in saliva have been reported to range down to as low as between 10^2 and 10^3 CFU/mL.[17,18] As such, future work will be required to fully determine the potential limit of detection. Third, current clinical tools for TB detection, such as the Xpert MTB/RIF assay and loop-mediated isothermal amplification for TB (LAMP-TB), can reach limits of detections between 10 and 10^2 CFU/mL.[56,57] To realize these sensitivities using HWRS, future optimizations will be needed. These could include utilizing more effective methods of concentrating the saliva samples, improving the substrate's ability to better localize the

View Article Online
DOI: 10.1039/D5SD00092K

bacteria within the measurement samples, and using larger laser spot sizes, potentially using a cheaper portable Raman spectrometer, to reduce the number of measurements that need to be taken on each sample. Alternatively, single-cell localization and measurement techniques using higher magnification objectives could be integrated into this approach to further improve the sensitivity and enable the determination of antibiotic resistance status, at the cost of requiring more complex instrumentation than the current approach.[58–60] Despite these limitations, this work is the first to demonstrate the feasibility of using high-wavenumber Raman spectroscopy as a reagent-free and culture-free method of detecting *M. tuberculosis* directly in human saliva samples.

Conclusion

This study aimed to determine the feasibility of using high-wavenumber Raman spectroscopy to detect the presence of *M. tuberculosis* in human saliva. To achieve this, six fasting and six non-fasting saliva samples were collected from healthy human participants and used to create a series of measurement samples including raw saliva, negative control saliva pellets, and bacteria-spiked saliva pellets. Using spectral unmixing analysis, it was found that the presence of *M. tuberculosis* caused the bacteria-spiked saliva pellets to have a significantly greater Raman lipid fraction than the paired negative control. Additionally, substituting *M. tuberculosis* with the common oral bacterial species *S. mutans* eliminated this difference, indicating the specificity of this measurement. However, due to the inherent variability in the lipid content between saliva samples, a single threshold value to detect the presence of *M. tuberculosis* could not be found. Instead, a linear regression model was used to predict the Raman lipid fraction of a control pellet given the lipid fraction of its corresponding raw saliva. When compared to the bacteria-spiked samples, it was found that only samples containing *M. tuberculosis* had lipid fractions exceeding the 95% prediction interval. Using the upper bounds of this interval as a classification threshold, leave-one-sample-out cross-validation found that the model was able to accurately determine the presence of *M. tuberculosis* in all of the samples tested. Overall, this work showcases the potential of using high-wavenumber Raman spectroscopy as a point-of-care method of detecting *M. tuberculosis* in human saliva without the need for reagents or complex sample preparations.

View Article Online
DOI: 10.1039/D5SD00092K

Conflicts of Interest

The authors declare that they have no relevant financial interests regarding the materials discussed in this paper and no potential conflicts of interest to disclose.

Ethical Statement

Approval for the saliva sample collection was obtained from the Vanderbilt University Medical Center Institutional Review Board (protocol #191004), and written informed consent was obtained from each participant prior to collection. All associated protocols and methods were performed in accordance with approved guidelines and regulations, and in accordance with the Declaration of Helsinki and its amendments.

Data Availability

The data underlying this study are available in the published article. Additional data related to this research may be requested from the authors.

Acknowledgement

Financial support for this project was provided by the Bill and Melinda Gates Foundation (INV-050756).

References

1 World Health Organization, *Global tuberculosis report 2024*, World Health Organization, Geneva, 2024.

2 J. Chakaya, M. Khan, F. Ntoumi, E. Aklillu, R. Fatima, P. Mwaba, N. Kapata, S. Mfinanga, S. E. Hasnain, P. D. M. C. Katoto, A. N. H. Bulabula, N. A. Sam-Agudu, J. B. Nachega, S. Tiberi, T. D. McHugh, I. Abubakar and A. Zumla, *International Journal of Infectious Diseases*, 2021, **113**, S7–S12.

3 P. Comella-del-Barrio, M. L. De Souza-Galvão, C. Prat-Aymerich and J. Domínguez, *Arch Bronconeumol*, 2021, **57**, 5–6.

4 F. Di Gennaro, G. Gualano, L. Timelli, P. Vittozzi, V. Di Bari, R. Libertone, C. Cerva, L. Pinnarelli, C. Nisii, S. Ianniello, S. Mosti, N. Bevilacqua, F. Iacomi, A. Mondi, S. Topino, D. Goletti, F. Vaia, G. Ippolito, E. Girardi and F. Palmieri, *Antibiotics*, 2021, **10**, 272.

View Article Online
DOI: 10.1039/D5SD00092K

5 S. Bello, R. F. Afolabi, D. T. Ajayi, T. Sharma, D. O. Owoeye, O. Oduyoye and J. Jasanya, *BMC Public Health*, 2019, **19**, 820.

6 M. A. Behr, P. H. Edelstein and L. Ramakrishnan, *BMJ*, 2019, **367**, l5770.

7 P. Escalante and J. W. Wilson, *Am J Respir Crit Care Med*, 2021, **203**, 1460–1461.

8 D. Pandey and D. Ghosh, *J Proteomics*, 2024, **305**, 105245.

9 N. Khambati, L. Olbrich, J. Ellner, P. Salgame, R. Song and E. M. Bijker, *Front Pediatr*, 2021, **9**, 756043.

10 J. Yu, J. Yuan, Z. Liu, H. Ye, M. Lin, L. Ma, R. Liu, W. Ding, L. Li, T. Ma, S. Tang and Y. Pang, *Clin Proteomics*, 2024, **21**, 66.

11 P. Rajapaksha, A. Elbourne, S. Gangadoo, R. Brown, D. Cozzolino and J. Chapman, *Analyst*, 2019, **144**, 396–411.

12 S. Gupta and V. Kakkar, *Biosens Bioelectron*, 2018, **115**, 14–29.

13 P. K. Drain, X. Niu, A. E. Shapiro, Z. P. Magcaba, Z. Ngcobo, M. W. Ngwane, K. K. Thomas, R. R. Dalmat, J. F. Morton, E. Budiawan, A. Pinter, J. Cantera, C. Anderson, R. Buchmann, D. Wilson and B. Grant, *EBioMedicine*, 2024, **108**, 105353.

14 M. Armstrong-Hough, J. Ggita, P. Turimumahoro, A. J. Meyer, E. Ochom, D. Dowdy, A. Cattamanchi, A. Katamba and J. L. Davis, *The International Journal of Tuberculosis and Lung Disease*, 2018, **22**, 1152–1159.

15 U. Mathebula, C. Emerson, T. Agizew, S. Pals, R. Boyd, A. Mathoma, J. Basotli, G. Rankgoane-Pono, C. Serumola, A. Date, A. F. Auld and A. Finlay, *Public Health Action*, 2020, **10**, 11–16.

16 A. K. Luabeya, R. C. Wood, J. Shenje, E. Filander, C. Ontong, S. Mabwe, H. Africa, F. K. Nguyen, A. Olson, K. M. Weigel, L. Jones-Engel, M. Hatherill and G. A. Cangelosi, *J Clin Microbiol*, DOI:10.1128/JCM.01847-18.

17 H. Yeager, J. Lacy, L. R. Smith and C. A. LeMaistre, *Am Rev Respir Dis*, 1967, **95**, 998–1004.

View Article Online
DOI: 10.1039/D5SD00092K

18 S. H. J. van den Elsen, T. van der Laan, O. W. Akkerman, A. G. M. van der Zanden, J.-W. C. Alffenaar and D. van Soolingen, *J Clin Microbiol*, 2017, **55**, 3292–3293.

19 M. N. Karinja, T. M. Esterhuizen, S. O. Friedrich and A. H. Diacon, *J Clin Microbiol*, 2015, **53**, 1087–1091.

20 C. S. Ealand, A. Sewcharran, J. S. Peters, B. G. Gordhan, M. Kamariza, C. R. Bertozzi, Z. Waja, N. A. Martinson and B. D. Kana, *Front Cell Infect Microbiol*, 2023, **13**, 1186191.

21 A. Shaikh, K. Sriraman, S. Vaswani, V. Oswal and N. Mistry, *Int J Infect Dis*, 2019, **86**, 5–11.

22 A. Andama, G. R. Whitman, R. Crowder, T. F. Reza, D. Jaganath, J. Mulondo, T. K. Nalugwa, F. C. Semitala, W. Worodria, C. Cook, R. C. Wood, K. M. Weigel, A. M. Olson, J. Lohmiller Shaw, M. Kato-Maeda, C. M. Denkinger, P. Nahid, G. A. Cangelosi and A. Cattamanchi, *J Clin Microbiol*, DOI:10.1128/jcm.00421-22.

23 P. Byanyima, S. Kaswabuli, E. Musisi, C. Nabakiibi, J. Zawedde, I. Sanyu, A. Sessolo, A. Andama, W. Worodria, L. Huang and J. L. Davis, *Microbiol Spectr*, DOI:10.1128/spectrum.00860-22.

24 A. S. Fauci and R. W. Eisinger, *Am J Trop Med Hyg*, 2018, **98**, 650–652.

25 E. Smith and G. Dent, *Modern Raman spectroscopy: a practical approach*, John Wiley & Sons, 2019.

26 A. C. S. Talari, Z. Movasaghi, S. Rehman and I. ur Rehman, *Appl Spectrosc Rev*, 2015, **50**, 46–111.

27 L. Rodriguez, Z. Zhang and D. Wang, *Analytical Science Advances*, 2023, **4**, 81–95.

28 M. Paraskevaidi, B. J. Matthew, B. J. Holly, B. J. Hugh, C. P. V. Thulya, C. Loren, C. StJohn, G. Peter, G. Callum, K. G. Sergei, K. Kamila, K. Maria, L. M. G. Kássio, M.-H. L. Pierre, P. Evangelos, P. Savithri, A. A. John, S. Alexandra, S. Marfran, S.-S. Josep, T. Gunjan, W. Michael and W. Bayden, *Appl Spectrosc Rev*, 2021, **56**, 804–868.

View Article Online
DOI: 10.1039/D5SD00092K

29 J. Haskell, T. Hubbard, C. Murray, B. Gardner, C. Ives, D. Ferguson and N. Stone, *Analyst*, 2023, **148**, 4373–4385.

30 A. K. Locke, F. R. Zaki, S. T. Fitzgerald, K. Sudhir, G. L. Monroy, H. Choi, J. Won, A. Mahadevan-Jansen and S. A. Boppart, *Front Cell Infect Microbiol*, 2022, **12**, 869761.

31 B. Kaewseekhao, N. Nuntawong, P. Eiamchai, S. Roytrakul, W. Reechaipichitkul and K. Faksri, *Tuberculosis*, 2020, **121**, 101916.

32 X.-S. Zheng, I. J. Jahn, K. Weber, D. Cialla-May and J. Popp, *Spectrochim Acta A Mol Biomol Spectrosc*, 2018, **197**, 56–77.

33 J. Mo, W. Zheng, J. J. H. Low, J. Ng, A. Ilancheran and Z. Huang, *Anal Chem*, 2009, **81**, 8908–8915.

34 K. Lin, W. Zheng, C. M. Lim and Z. Huang, *Theranostics*, 2017, **7**, 3517–3526.

35 E. J. Haugen, A. K. Locke, L. H. Dao, A. B. Walter, P. K. Rasiah, J. S. Baba, M. A. Buendia, H. Correa, G. Hiremath and A. Mahadevan-Jansen, *Sci Rep*, 2025, **15**, 22471.

36 K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor and M. Baranska, *Journal of Raman Spectroscopy*, 2015, **46**, 4–20.

37 E. Wiercigroch, E. Szafraniec, K. Czamara, M. Z. Pacia, K. Majzner, K. Kochan, A. Kaczor, M. Baranska and K. Malek, *Spectrochim Acta A Mol Biomol Spectrosc*, 2017, **185**, 317–335.

38 J. M. Viader-Salvadó, C. A. Molina-Torres and M. Guerrero-Olazarán, *J Microbiol Methods*, 2007, **70**, 479–483.

39 M. Jackson, *Cold Spring Harb Perspect Med*, 2014, **4**, a021105–a021105.

40 J. Perumal, U. Dinish, A. Bendt, A. Kazakeviciute, C. Y. Fu, I. L. H. Ong and M. Olivo, *Int J Nanomedicine*, 2018, **Volume 13**, 6029–6038.

41 S. Rafał, D. Magdalena, M. Karol, S. Bartłomiej and K. Konrad, *Sci Rep*, 2025, **15**, 13118.

42 R. J. Palmer Jr, *Periodontol 2000*, 2014, **64**, 20–39.

View Article Online
DOI: 10.1039/D5SD00092K

43 C. A. Lieber and A. Mahadevan-Jansen, *Appl Spectrosc*, 2003, **57**, 1363–1367.

44 M. Unal, R. Ahmed, A. Mahadevan-Jansen and J. S. Nyman, *Analyst*, 2021, **146**, 7464–7490.

45 J. A. Loo, W. Yan, P. Ramachandran and D. T. Wong, *J Dent Res*, 2010, **89**, 1016–1023.

46 J. Matczuk, M. Żendzian-Piotrowska, M. Maciejczyk and K. Kurek, *Advances in Clinical and Experimental Medicine*, 2017, **26**, 1021–1029.

47 N. K. Howell, G. Arteaga, S. Nakai and E. C. Y. Li-Chan, *J Agric Food Chem*, 1999, **47**, 924–933.

48 D. Mampallil and H. B. Eral, *Adv Colloid Interface Sci*, 2018, **252**, 38–54.

49 P. J. Yunker, T. Still, M. A. Lohr and A. G. Yodh, *Nature*, 2011, **476**, 308–311.

50 C. M. O'Brien, E. Vargis, A. Rudin, J. C. Slaughter, G. Thomas, J. M. Newton, J. Reese, K. A. Bennett and A. Mahadevan-Jansen, *Am J Obstet Gynecol*, 2018, **218**, 528.e1-528.e18.

51 E. E. Spurlin, K. Belcher, F. Shofu, J. A. E. Esteves, P. T. Jimenez and C. M. O'Brien, *npj Women's Health*, 2025, **3**, 32.

52 A. Daniel, A. Prakasarao and S. Ganesan, *Spectrochim Acta A Mol Biomol Spectrosc*, 2018, **190**, 409–416.

53 B. P. Yakimov, A. V Venets, J. Schleusener, V. V Fadeev, J. Lademann, E. A. Shirshin and M. E. Darvin, *Analyst*, 2021, **146**, 3185–3196.

54 D. Georgiev, Á. Fernández-Galiana, S. Vilms Pedersen, G. Papadopoulos, R. Xie, M. M. Stevens and M. Barahona, *Proc Natl Acad Sci U S A*, 2024, **121**, e2407439121.

55 E. Caselli, C. Fabbri, M. D'Accolti, I. Soffritti, C. Bassi, S. Mazzacane and M. Franchi, *BMC Microbiol*, 2020, **20**, 120.

56 M. Seki, C.-K. Kim, S. Hayakawa and S. Mitarai, *Eur J Clin Microbiol Infect Dis*, 2018, **37**, 1405–1410.

View Article Online
DOI: 10.1039/D5SD00092K

57 S. Chakravorty, A. M. Simmons, M. Rowneki, H. Parmar, Y. Cao, J. Ryan, P. P. Banada, S. Deshpande, S. Shenai, A. Gall, J. Glass, B. Krieswirth, S. G. Schumacher, P. Nabeta, N. Tukvadze, C. Rodrigues, A. Skrahina, E. Tagliani, D. M. Cirillo, A. Davidow, C. M. Denkinger, D. Persing, R. Kwiatkowski, M. Jones and D. Alland, *mBio*, DOI:10.1128/MBIO.00812-17/SUPPL_FILE/MBO004173453S1.DOCX.

58 H. Jayan, H. Pu and D.-W. Sun, *Crit Rev Food Sci Nutr*, 2022, **62**, 4294–4308.

59 S. Stöckel, S. Meisel, B. Lorenz, S. Kloß, S. Henk, S. Dees, E. Richter, S. Andres, M. Merker, I. Labugger, P. Rösch and J. Popp, *J Biophotonics*, 2017, **10**, 727–734.

60 A. Zyubin, A. Lavrova, O. Manicheva, M. Dogonadze, V. Belik and I. Samusev, *Laser Phys Lett*, 2019, **16**, 085602.

View Article Online
DOI: 10.1039/D5SD00092K

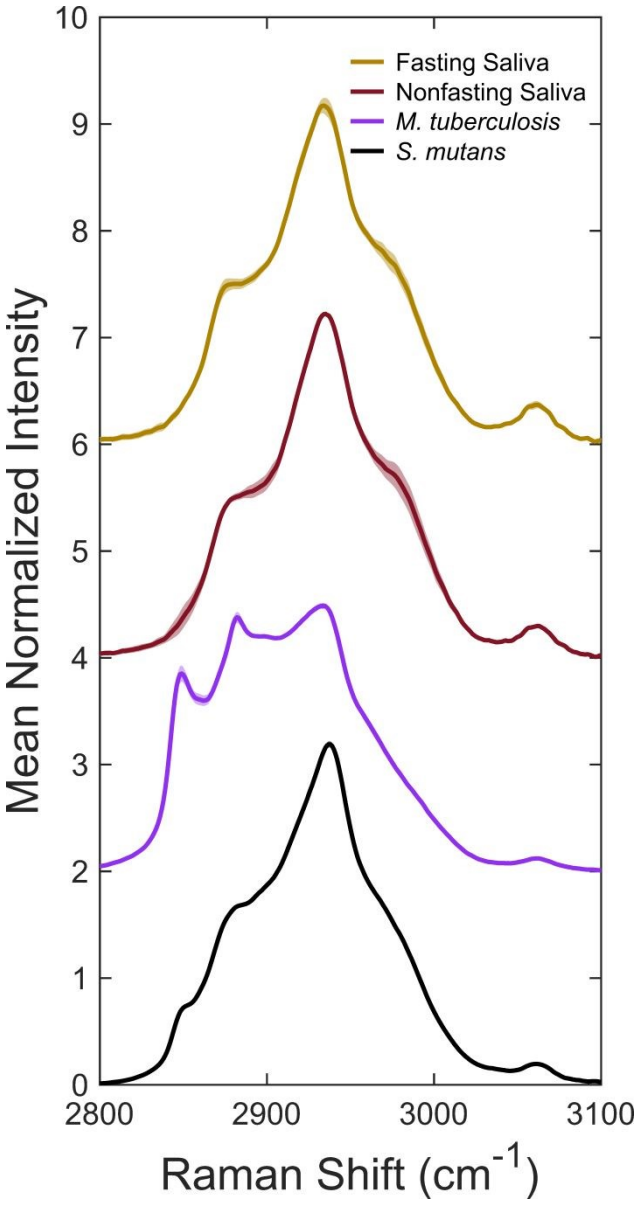

Figure 1. High-wavenumber Raman spectra for fasting and non-fasting human saliva compared to *M. tuberculosis* and *S. mutans*. Each spectra represents the mean (solid line) and standard deviation (shaded region) of the saliva samples (n=5) and bacterial cultures (n=3). Spectra are mean normalized and vertically offset for visualization.

View Article Online
DOI: 10.1039/D5SD00092K

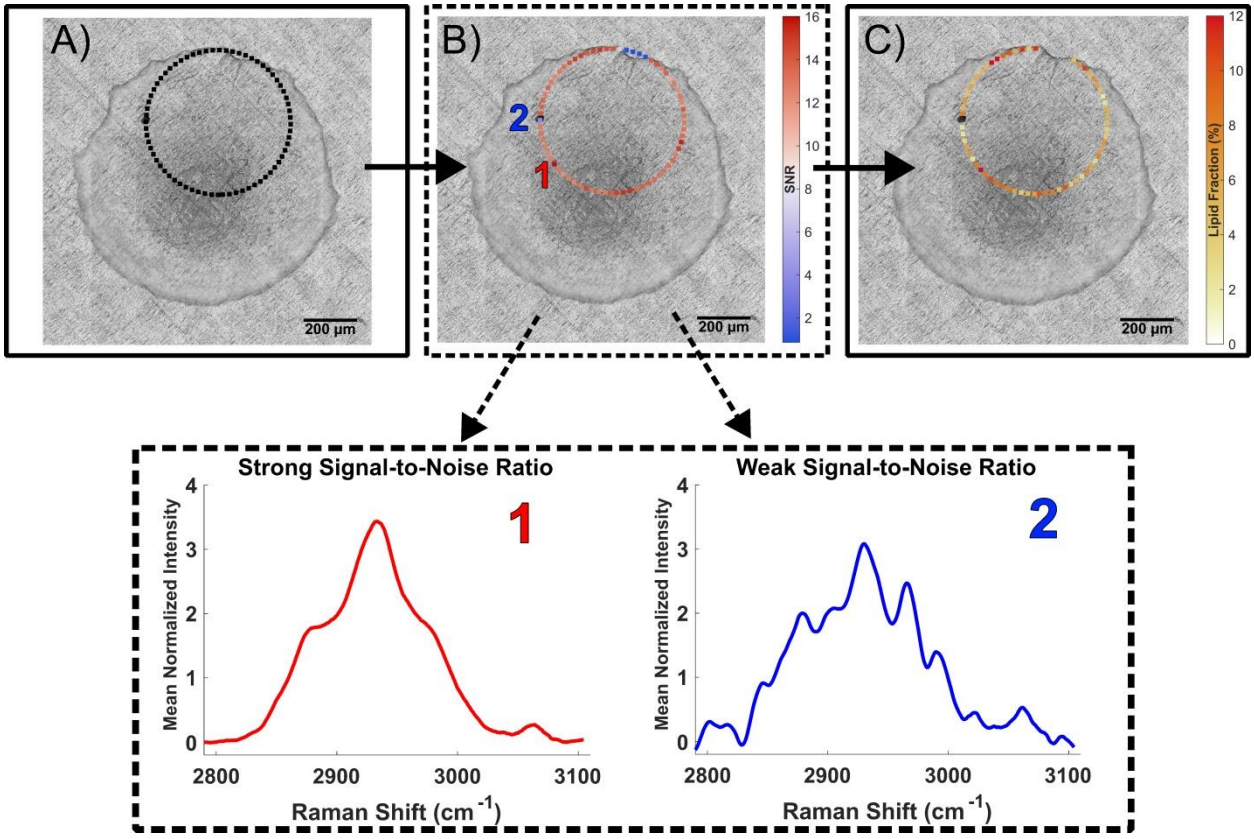

Figure 2. Raman map analysis of dried saliva droplets. (A) For each sample, a circle was inscribed between the center and the edge of the droplet, with measurements, denoted by black squares, made every 25 µm along the circumference, (B) In some locations, the signal-to-noise ratio (SNR) was too low to be usable. Spectra with an SNR above 8 (red) were kept for further analysis while those with SNR below 8 (blue) were discarded. (C) For each spectrum, spectral unmixing analysis was performed to determine the Raman lipid fraction.

View Article Online
DOI: 10.1039/D5SD00092K

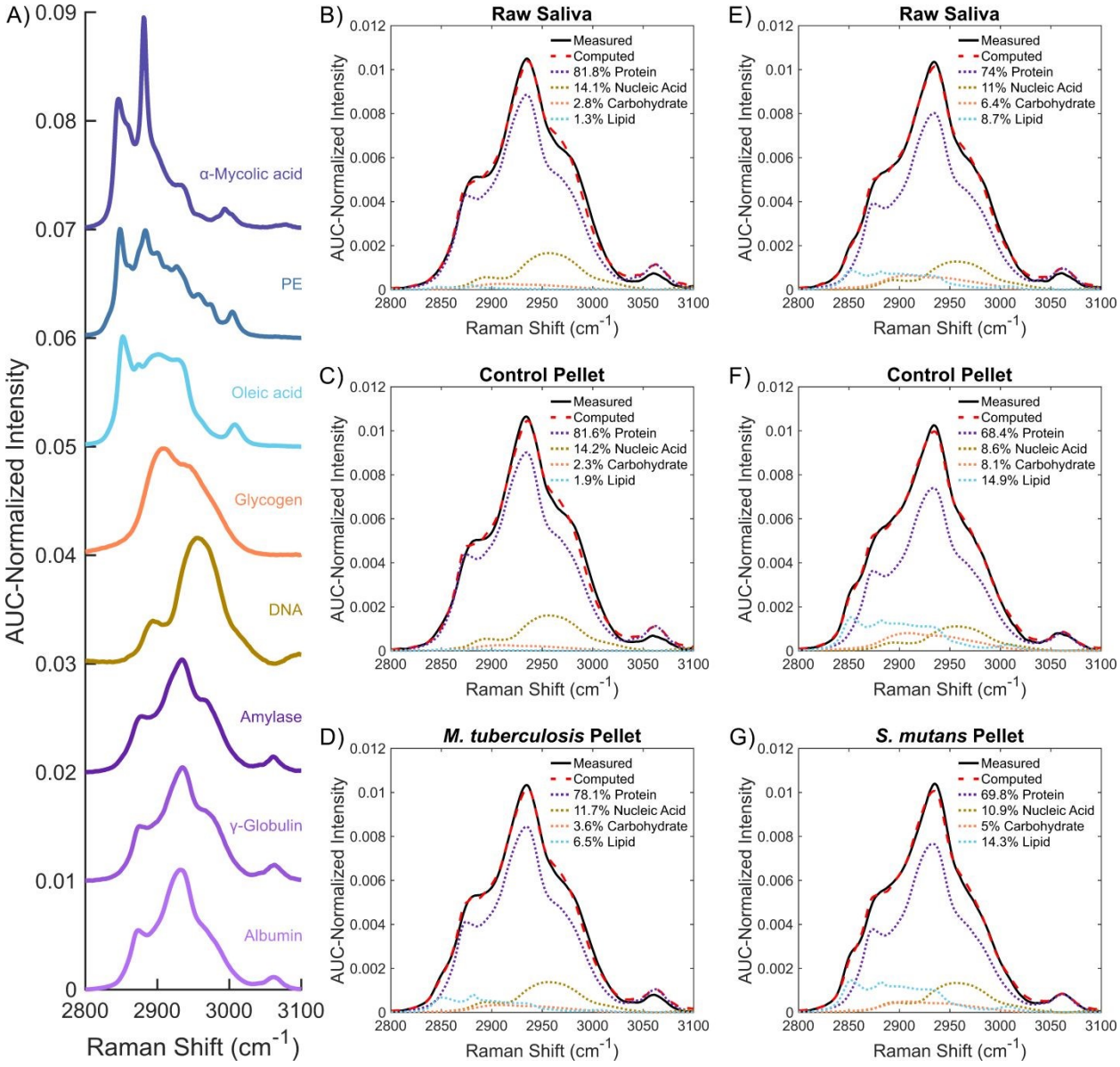

Figure 3. (A) High-wavenumber Raman spectra for the 8 components used in the spectral unmixing analysis. Components are categorized as either protein (purple), nucleic acid (gold), carbohydrate (orange), or lipid (blue). Representative spectral unmixing results for the raw saliva (B, E), control pellets (C, F), and bacteria-spiked pellets (D, G) for a set of samples spiked with *M. tuberculosis* (B-D) and *S. mutans* (E-G). Pure components (dotted lines) are grouped by biochemical category and are represented by the weighted sum of the individual components. The overall computed spectra (dashed line) represent the sum of each category multiplied. Percentage contributions reported in each legend are the average values across the Raman map measurement for each category. Individual spectra are normalized to their area-under-the-curve (AUC) for the unmixing analysis and the pure component spectra are offset for visualization.

View Article Online
DOI: 10.1039/D5SD00092K

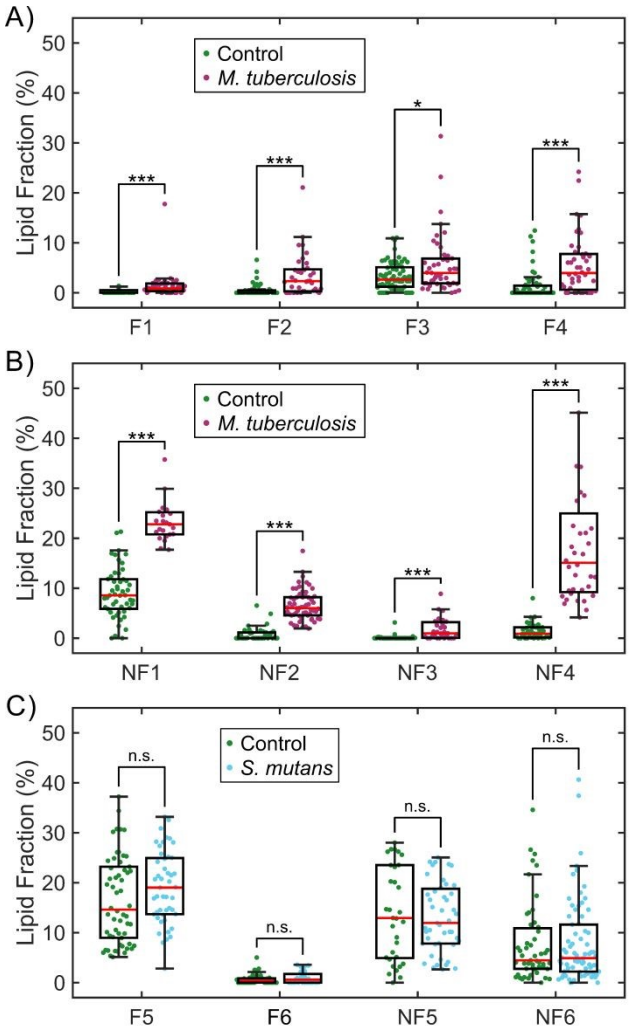

Figure 4. Comparison of the Raman mapping results for the negative control and bacteria-spiked saliva pellets. *M. tuberculosis* presence significantly increased the Raman lipid fraction of both (A) fasting (F) and (B) non-fasting (NF) saliva samples. (C) Inclusion of *S. mutans* did not significantly alter the relative lipid concentration. (n.s. $p > 0.05$, *$p < 0.05$, ***$p < 0.001$).

View Article Online
DOI: 10.1039/D5SD00092K

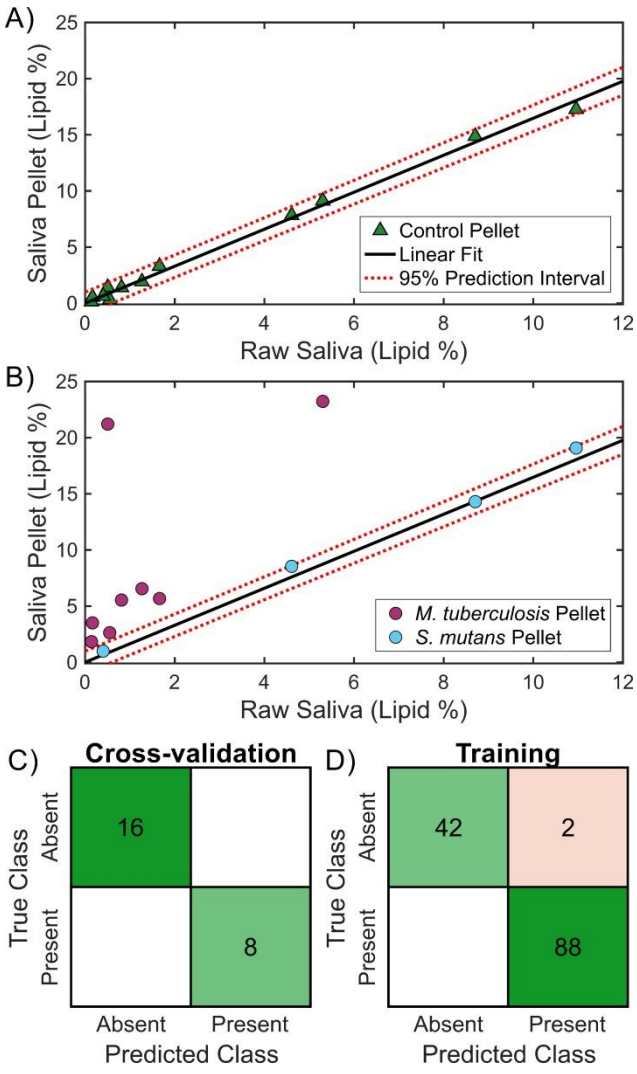

Figure 5. (A) Linear regression model used to predict the Raman lipid fraction of a pelleted saliva sample from a measurement of the raw saliva. (B) Comparing the bacteria-spiked pellets to the linear model found that samples containing *M. tuberculosis* exceeded the 95% prediction interval while those containing *S. mutans* remained within. A leave-one-sample-out cross-validation of the model found (C) high classification accuracies and (D) high training accuracies when using the upper prediction interval as a classification threshold.

View Article Online
DOI: 10.1039/D5SD00092K

Data Availability

The data underlying this study are available in the published article. Additional data related to this research may be requested from the authors.