Electronic Supplementary Material (ESI) for Analyst.

Rapid, accurate, and comparative differentiation of clinically and industrially relevant microorganisms *via* multiple vibrational spectroscopic fingerprinting

Howbeer Muhamadali[1], Abdu Subaihi[1], Mahsa Mohammadtaheri [2], Yun Xu[1], David I. Ellis[1], Rajesh Ramanathan[2], Vipul Bansal[2] and Royston Goodacre[1*]

[1]School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester, UK

[2]Ian Potter NanoBioSensing Facility, NanoBiotechnology Research Laboratory, School of Science, RMIT University, Melbourne, Australia

Despite the fact that various microorganisms (*e.g.* bacteria, fungi, viruses *etc.*) have been linked with infectious diseases, their crucial role towards sustaining life on Earth is undeniable. The huge biodiversity, combined with the wide range of biochemical capabilities of these organisms, have always been the driving force behind their large number of current, and, as of yet, undiscovered future applications. The presence of such diversity could be said to expedite the need for the development of rapid, accurate and sensitive techniques which allow for the detection, identification and classification of such organisms. In this study, we employed Fourier transform infrared (FT-IR), Raman, and surface enhanced Raman scattering (SERS) spectroscopies, as molecular fingerprinting techniques, combined with multivariate statistical analysis approaches for the classification of a range of industrial/ environmental or clinically relevant bacterial (*P. aeruginosa, P. putida, E. coli, E. faecium, S. lividans, B. subtilis, B. cereus*) and yeast (*S. cerevisiae*) cells. PC-DFA scores plots of the spectral data collected from all three techniques allowed for the clear differentiation of all the samples down to sub-species level. With the models generated based on the SERS spectral data displaying lower accuracy (74.9%) when compared to those obtained from conventional Raman (97.8%) and FT-IR (96.2%) analyses. In addition, our results suggest that while each of these spectroscopic approaches may favour different organisms (sample types), when combined, they could provide complementary and more in-depth knowledge (structural and/or metabolic state) of biological systems. To the best of our knowledge, this is the first time that such a comparative and combined spectroscopic study (using FT-IR, Raman and SERS) has been carried out on such a range microbial samples.

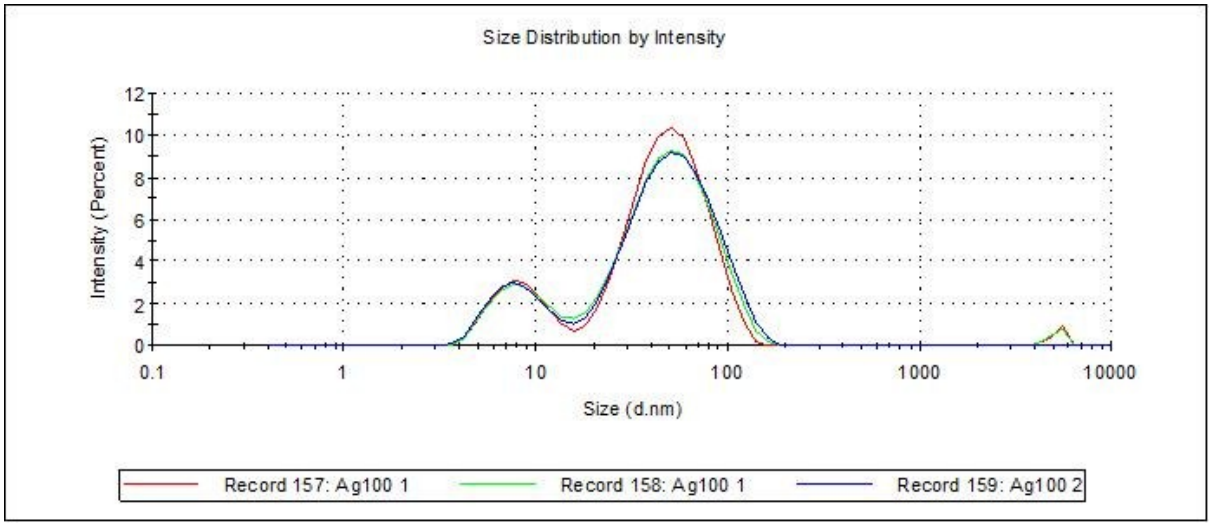

Figure S1. The DLS graph of the tyrosine capped AgNPs.

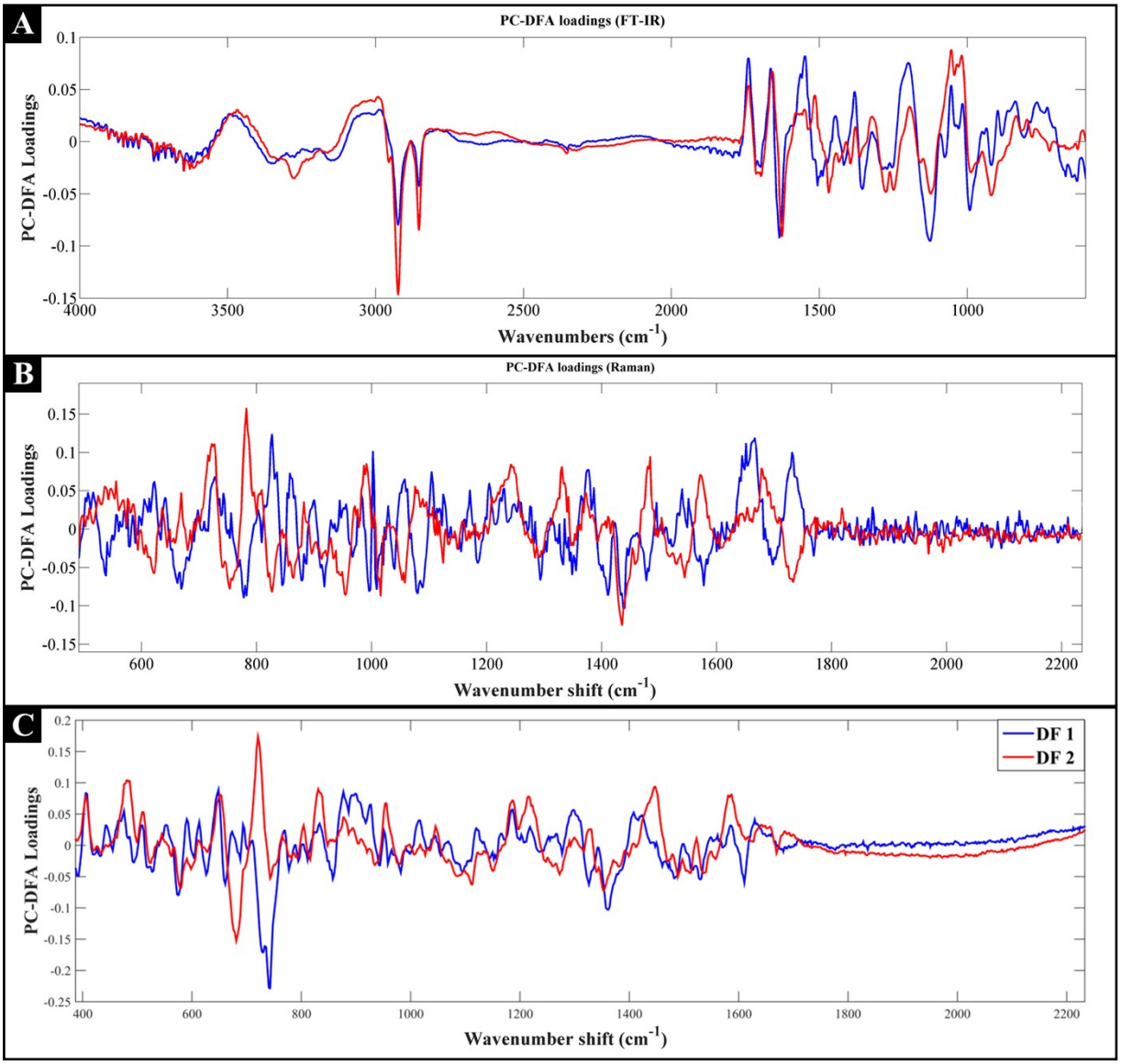

Figure S2. PC-DFA loadings plots of the FT-IR (A), Raman (B), and SERS (C) spectral data, demonstrating the most significant variables (peaks) that contribute to the clustering patterns.

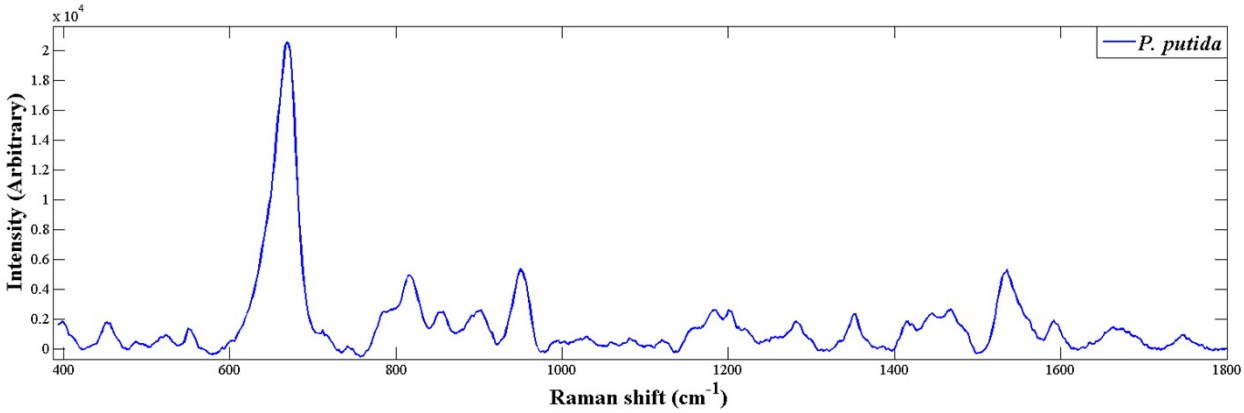

Figure S3. SERS spectrum of P. putida, obtained using hydroxylamine-reduced Ag NPs.

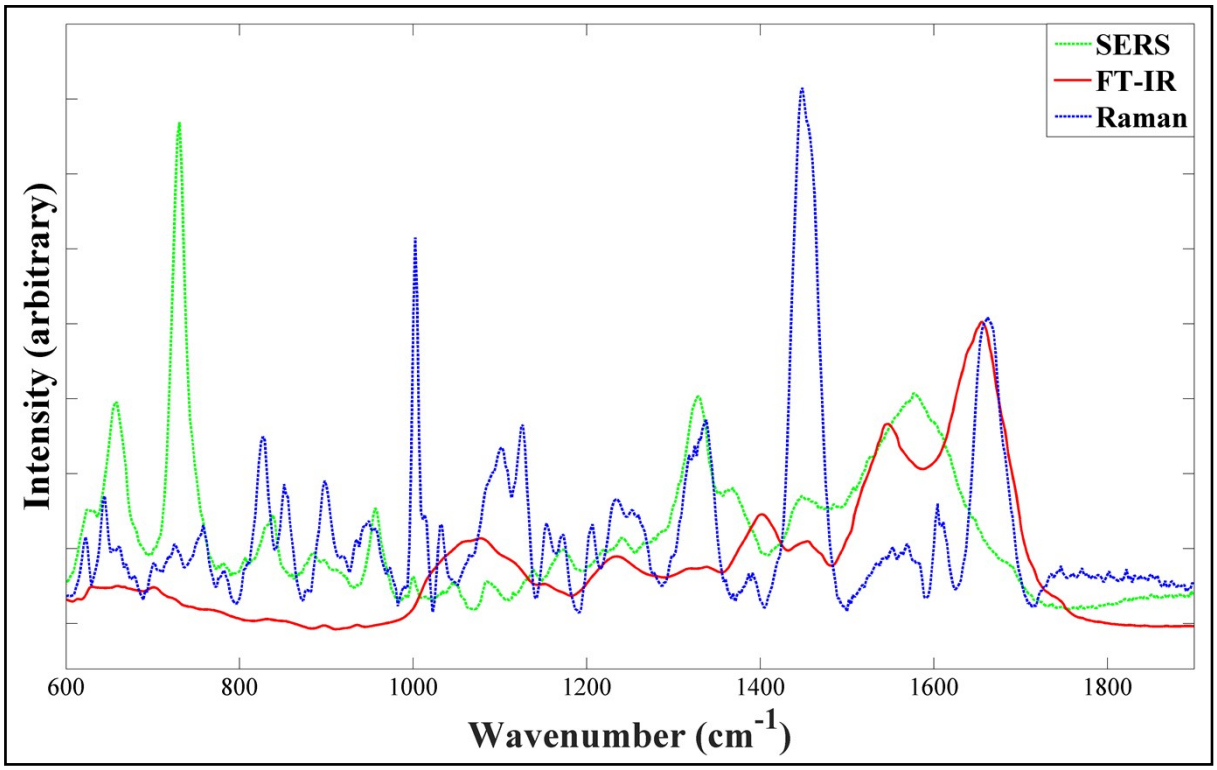

Figure S4. Comparison of FT-IR (red), Raman (blue), and SERS (green) averaged spectra of *B. subtilis*.

Table S1. Prediction accuracy of the classification model, generated using PLS-DA and validated by bootstrapping method (1000 iterations), of the FT-IR spectral data. The numbers in the brackets represent the max and min of the prediction accuracy of the 1000 bootstrap iterations.

	B. cereus	*B. subtilis*	*E. coli*	*E. faecium*	*P. aeruginosa*	*P. putida*	*S. cerevisiae*	*S. lividans*
B. cereus	92.62% (100-51.6)	5.71%	0.14%	0.52%	0.00%	0.85%	0.00%	0.17%
B. subtilis	0.06%	99.24% (100-95.6)	0.06%	0.36%	0.00%	0.15%	0.00%	0.13%
E. coli	0.03%	0.07%	97.51% (100-84.4)	0.01%	0.37%	1.85%	0.00%	0.15%
E. faecium	0.00%	0.15%	0.03%	99.43% (100-96.9)	0.00%	0.37%	0.02%	0.00%
P. aeruginosa	0.00%	0.03%	0.68%	0.03%	96.82% (100-80)	2.27%	0.00%	0.18%
P. putida	0.00%	0.10%	1.61%	0.27%	2.95%	94.46% (100-60.1)	0.00%	0.61%
S. cerevisiae	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	99.90% (100-99.1)	0.08%
S. lividans	0.18%	1.91%	0.11%	0.03%	1.03%	0.54%	0.08%	96.11% (100-66.5)
							Average	96.19%

Table S2. Prediction accuracy of the classification model, generated using PLS-DA and validated by bootstrapping method (1000 iterations), of the Raman spectral data. The numbers in the brackets represent the max and min of the prediction accuracy of the 1000 bootstrap iterations.

	B. cereus	*B. subtilis*	*E. coli*	*E. faecium*	*P. aeruginosa*	*P. putida*	*S. lividans*
B. cereus	99.90% (100-97.1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%
B. subtilis	0.11%	99.73% (100-90.9)	0.01%	0.00%	0.00%	0.06%	0.10%
E. coli	0.00%	0.08%	99.70% (100-73.9)	0.10%	0.07%	0.05%	0.00%
E. faecium	0.00%	0.00%	0.00%	99.90% (100-90.4)	0.00%	0.10%	0.00%
P. aeruginosa	0.00%	0.00%	0.03%	0.10%	98.50% (100-67.6)	1.37%	0.00%
P. putida	0.00%	0.00%	0.10%	0.28%	4.34%	95.29% (100-60.4)	0.00%
S. lividans	0.00%	0.00%	0.00%	0.00%	0.30%	2.69%	97.01% (100-76.6)
						Average	97.79%

Table S3. Prediction accuracy of the classification model, generated using PLS-DA and validated by bootstrapping method (1000 iterations), of the SERS spectral data. The numbers in the brackets represent the max and min of the prediction accuracy of the 1000 bootstrap iterations.

	B. cereus	*B. subtilis*	*E. faecium*	*E. coli*	*P. aeruginosa*	*S. cerevisiae*	*S. lividans*
B. cereus	72.47% (100-42.3)	6.06%	4.47%	6.09%	6.37%	2.91%	1.62%
B. subtilis	5.51%	84.56% (100-57.8)	1.79%	3.65%	2.19%	2.25%	0.05%
E. faecium	3.77%	0.09%	85.67% (100-58.2)	2.87%	3.55%	0.30%	3.74%
E. coli	3.32%	1.12%	1.76%	71.49% (100-42.4)	14.62%	5.28%	2.41%
P. aeruginosa	4.84%	0.42%	1.36%	17.91%	67.43% (100-37.9)	5.84%	2.20%
S. cerevisiae	1.51%	0.01%	0.45%	8.70%	8.23%	69.43% (100-34.9)	11.68%
S. lividans	0.95%	0.00%	6.60%	3.99%	4.34%	2.48%	81.64% (100-51.6)
						Average	74.89%