Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Association for Academic Surgery

Rapid Detection of *Clostridium difficile* Toxins in Serum by Raman Spectroscopy

Satya Kiran Koya, PhD,[a] *Michelle Brusatori, PhD,*[a] *Jonathan V. Martin, MD,*[b] *Sally Yurgelevic, MS,*[a] *Changhe Huang, PhD,*[a] *David M. Liberati, MS,*[b] *Gregory W. Auner, PhD,*[a] *and Lawrence N. Diebel, MD*[b,*]

[a] *Smart Sensors and Integrated Microsystems, Michael and Marian Ilitch Department of Surgery, School of Medicine, Wayne State University, Detroit, Michigan*

[b] *Michael and Marian Ilitch Department of Surgery, School of Medicine, Wayne State University, 6C University Health Center, Detroit, Michigan*

ARTICLE INFO

Article history:
Received 2 March 2018
Received in revised form
5 May 2018
Accepted 8 June 2018
Available online 9 July 2018

Keywords:
Clostridium *difficile* toxin A and B
Systemic CDI
Raman spectroscopy
Rapid detection
Diagnostic test

ABSTRACT

Background: *Clostridium difficile* infection (CDI) is due to the effects of toxins, toxin A and toxin B on the host. Severe CDI is associated with systemic signs of infection. Animal models of CDI demonstrate a strong correlation between systemic toxemia and the occurrence of severe disease. However, current technologies have low sensitivity to detect *C difficile* toxemia in human subjects. Raman spectroscopy (RS) is an upcoming technology that is used to detect bacteria and their toxins. We speculate that RS may be a sensitive method to detect clinically relevant concentrations of *C difficile* toxins in serum.

Materials and methods: Serum samples were spiked with varying concentrations of toxin A, toxin B, and both. RS was performed on an air-dried serum drop that was placed on a mirror-polished stainless steel slide. Raman spectra were obtained, background corrected, vector normalized, and analyzed by Partial Least Square Linear Discriminant Analysis and Support Vector Machine for Classification. Model accuracy was measured by cross-validation and bootstrap methods.

Results: Toxin-spiked sera of various concentrations (1 ng/mL, 1 pg/mL, and 0.1 pg/mL) were distinguished from control serum 100% with cross-validation error rate ranging from 0% to 18% and bootstrap error rate ranging from 0% to 12% for various concentrations. The sensitivity ranged from 87% to 100% and specificity ranged from 77% to 100% for various concentrations of toxin-spiked serum.

Conclusions: We conclude that RS may be a sensitive method to detect clinically relevant concentrations of *C difficile* toxins in serum and thus to help diagnose severe CDI in patients in real-time at the point of care.

* *Corresponding author*. Michael and Marian Ilitch Department of Surgery, 6C University Health Center, 4201 Saint Antoine, Detroit, MI 48201. Tel.: 313-577-5005; fax: 313-577-5310.
E-mail address: ldiebel@med.wayne.edu (L.N. Diebel).

0022-4804/$ – see front matter
https://doi.org/10.1016/j.jss.2018.06.031

Introduction

Clostridium difficile is the most common health care pathogen, with approximately a half million new cases with more than 29,000 deaths attributed to *Clostridium difficile* infection (CDI) each year.[1] Signs of severe CDI include hypotension and the need for pressors, profound leukocytosis, mental status changes, and worsening organ failure.[2] These systemic manifestations of CDI have been attributed to *Clostridium* pathogenic factors, principally toxin A (TcdA) and toxin B (TcdB), as well as the immunoinflammatory host response to infection.[3-5]

Animal studies suggest that *C difficile* systemic toxemia alone indicates severe disease.[6,7] Passive immunotherapy with intravenous administration of immunoglobulin resulted in remission of refractory *C difficile* colitis suggesting that toxemia may occur in fulminant CDI in humans.[8] This gives us the impetus to measure toxin levels in serum in severe CDI patients. Reports of toxemia in humans are rare and this is due to low levels of circulating toxins that are below the detection limit of assays.[8,9]

Cell cytotoxicity assay (CCAT) is the gold standard test for laboratory diagnosis of toxemia in serum, due its high sensitivity and specificity. However, this test is not performed by many laboratories due to slow turnaround time (24-72 h), requirement of cell culture facility, and technical complexity.[10] This method was improved drastically by ultrasensitive rapid immunocytotoxicity assay (ICT) for detection of TcdA at doses as low as 1-10 pg/mL. This was achieved by enhancing toxicity of TcdA on Fc gamma receptor I-expressing cells by adding anti-TcdA monoclonal antibody.[11] Turnaround time for ICT was brought down to approximately 3 h by using cryopreserved cells and measuring electrical impedance as a measure of cytotoxicity using real-time cell electronic sensing.[11] However, this test is not sensitive to TcdB at low concentrations (~10 pg/mL limit) and presence of toxin-specific neutralizing antibodies in sera further increased level of detection for both toxins.[9] A simple, rapid, and relatively cheap diagnostic assay for detection of TcdA and TcdB with high sensitivity and specificity is still lacking and is desirable.[9,12]

We propose a rapid, ultrasensitive Raman spectroscopy (RS)-based diagnostic modality for detection of TcdA and TcdB in serum. RS has been used as a diagnostic modality for infectious disease and toxin detection.[13,14] RS is a noncontact, nondestructive, reagent-less optical technique, which provides a unique spectroscopic fingerprint of a pathogen or toxin that is being detected. RS is very sensitive to slight changes in concentrations of toxins even at very low concentrations making it an ideal platform for detection of *C difficile* toxins in serum.

The RS-based diagnostic modality is cost-effective, rapid (<30 min turnaround time), and ultrasensitive (as low as 0.1 pg/mL) and may be employed as a point-of-care testing modality for continuous monitoring of patients and critically ill patients. This test will need verification in clinical trials and comparison with other modalities such as CCAT and ICT.

Materials and methods

Ethics statement

This study was approved by Institutional Review Boards of Wayne State University and Detroit Medical Center. Serum that was used in this study was obtained from healthy volunteer who provided written informed consent.

Toxin-spiked serum preparation

Recombinant TcdA and TcdB were purchased from R&D Systems (Cat# 8619GT020) and diluted to 2 μg/mL stock solutions in phosphate-buffered saline. Blood was collected using venipuncture into BD vacutainer redtop tubes from a healthy volunteer who provided a written consent. Blood was allowed to clot by leaving it undisturbed at room temperature. The clot was removed by centrifuging at 2000 × *g* for 10 min and was aliquoted into 0.5 mL tubes and stored at −80°C until RS. Toxins (TcdA alone, TcdB alone, or both TcdA and TcdB) were spiked into serum at the levels of 1 ng/mL, 1 pg/mL, and 0.1 pg/mL. Serum was used as negative control. Because of the toxicity of TcdA and TcdB, the preparation, dilutions, and RS were done in biosafety level II lab. Toxin-spiked serum samples were stored at −80°C until RS measurements were made.

Raman spectroscopy

RS probes molecular vibrations or rotations associated with chemical bonds in a sample to obtain information on molecular structure, composition, and intermolecular interactions. With this technique, a sample is illuminated with monochromatic light of a certain wavelength, typically from a laser. Although light can interact with the sample via the process of absorption, reflection, or scattering, it is the scattering event that is exploited in RS. Light can be scattered from a sample at the frequency of the incident light, termed Rayleigh (or elastic) scattering.[15] This occurs without a net energy transfer between the incident light and the sample. Light can also be scattered at higher (anti-Stokes shift) or lower (Stokes shift) frequencies than the incident light via an inelastic process and involves a net energy transfer between the incident photons and sample. An inelastic process, termed the Raman Effect, occurs in approximately 1 in 10^7 photon interactions with matter and depends on a change in the polarizability of a molecule as it vibrates or rotates.[16-18] By monitoring the intensity profile of inelastically scattered light as a function of frequency, the unique spectroscopic fingerprint of a sample (pathogen or toxin) is obtained. By convention, the frequency of scattered light is converted to Raman shifts, which is the difference in frequency between the incident and scattered light (usually in units of wave number). The wave numbers at which Raman bands occur is characteristic of vibrational modes of specific bond types in a molecule, with the intensity directly proportional to the concentration of species that give rise to the bands.[19] Since each sample has a unique composition, the spectroscopic profile arising from Raman-active

functional groups of nucleic acids, proteins, lipids, and carbohydrates allows for evaluation, characterization, and discrimination between sample types. We employed this technique to detect TcdA, TcdB, and TcdA and B levels in serum *in vitro*.

Acquisition of Raman spectra

A total of 320 (53 for control and ~30 for each test group) Raman spectra were obtained for serum (control) and toxin-spiked serum of specific concentrations (1 ng/mL, 1 pg/mL, and 0.1 pg/mL) for TcdA, TcdB, and TcdA and B. A drop (15 μL) of serum or toxin-spiked serum was placed on a sterile mirror-polished stainless steel substrate (alloy 304, Stainless Supply, Monroe, NC) and was air-dried by gently blowing sterile air for 5 min. Raman spectra of air-dried serum samples were recorded with an inVia Raman microscope (Renishaw, Gloucestershire, UK) equipped with a 785 nm excitation laser, 1200 L/mm grating, 576 × 400 pixel thermoelectric-cooled charge coupled device, and WiRE 3.4 software (Renishaw Inc., Hoffman Estates, IL). The laser light was focused onto the sample through a 50X objective (Nikon LU Plan Fluor EPI). Spectra were acquired using 100% laser power (corresponding to 37 mW at the sample) over a spectral range of 100-3200 cm^{-1} with 5 accumulations at an integration time of 20 s.

Preprocessing of Raman spectra

After acquisition, raw spectra were preprocessed by applying an in-house developed LabVIEW procedure for background subtraction and normalization. IIR zero phase filter was used to smooth the raw spectrum. For background subtraction, we used a LabVIEW background removal program from Donald J. Roth, which uses a low pass filter for background subtraction to obtain the best result.[20,21] The cosmic spikes were removed by cutting the peak with large second derivative value and the spectrum was normalized by unit vector normalization.

Multivariate analysis

Each Raman spectrum contained 3100 spectral features or Raman shifts that correspond to a certain biological molecule in the sample. Using multivariate analysis methods, we set to identify biologically relevant spectral features (biomarker signature of the toxin) and their relative changes between control and toxin-spiked serum to infer a biological relationship. This relationship is used to detect TcdA, TcdB, or both in the serum.

Most popular multivariate methods are Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) for classification. LDA was designed to maximize between-groups variance relative to within-groups variance.[22] This method uses Mahalanobis distance to calculate the probability of a new object belonging to a certain class (control or toxin-spiked serum). This method cannot be applied to data sets where the number of variables is higher than the number of observations. Such data sets are typical for Raman Spectral data (3100 variables and <500 observations). Therefore, a data dimension reduction method such as Principle Component Analysis (PCA) or Partial Least Squares (PLS) projection to latent structures is used before LDA.[22-24]

PCA is an unsupervised multivariate technique that describes major trends in the data based on the variation.[22] It reduces the number of variables (Raman shifts) to a new set of principle components that are linear combinations of the original ones. PCA does not consider sample class membership or groups, which is a disadvantage in some cases. PLS method solves this problem by considering sample groups by linking data with class membership and aims to maximize the covariance between independent variables (sample readings) and corresponding dependent variables or classes.[23] However, PLS aggressively separates experimental groups and for this reason it is essential to use PCA as a practical indicator of PLS model reliability.[25]

SVM is a popular supervised method, which plots each data item as a point in n-dimensional space (where *n* is number of features you have) with the value of each feature being the value of a coordinate. Then, it optimally separates the data by finding a hyperplane that differentiates the two classes. Next, it projects the new sample into the same space and classifies it based on which side of the margin it falls.[26] Linear SVM allows us to extract Variable ImPortance (VIP) score for each variable revealing which variable that contributed most to the separation of test and control samples in the model. A VIP score is calculated from the absolute value of the t-statistic for each model parameter in linear SVM. Both PLS-LDA and SVM models' performances were validated for accuracy using cross-validation and bootstrap methods.[27,28] Multivariate analysis procedures were performed using the R statistical programming language and Tanagra open source project software.[29,30]

Results

We demonstrated the feasibility of RS to detect *C difficile* toxins TcdA, TcdB, and TcdA and B in serum at various concentrations (1 ng/mL, 1 pg/mL, and 0.1 pg/mL). Figure 1 shows mean Raman spectra for control serum, TcdA, TcdB, and TcdA and B spiked serum at all concentrations.

Raman spectra from each toxin-spiked serum and control serum were examined individually (e.g., TcdA 0.1 pg/mL serum *versus* control serum) using multivariate analysis to identify relevant biomarker signature for the toxin (Raman shifts). This biomarker signature would allow us to distinguish toxin-spiked serum from the control serum in the future analysis. Multivariate analysis revealed several spectral regions with spectroscopic differences enough to distinguish toxin-spiked serum from control serum and they are shown in (Fig. 2).

Before performing PLS-LDA, we identified variance between control and each toxin-spiked serum of various concentrations using PCA (data not shown). PCA is used as a practical indicator for PLS model reliability. PLS-LDA model successfully separated each toxin-spiked serum from its control serum 100% at every concentration tested. PLS-LDA model accuracy was measured using cross-validation and bootstrap methods and the error rates, sensitivity, and specificity were given in Supplemental Table 1. Using more than one multivariate model for classification helps us to avoid

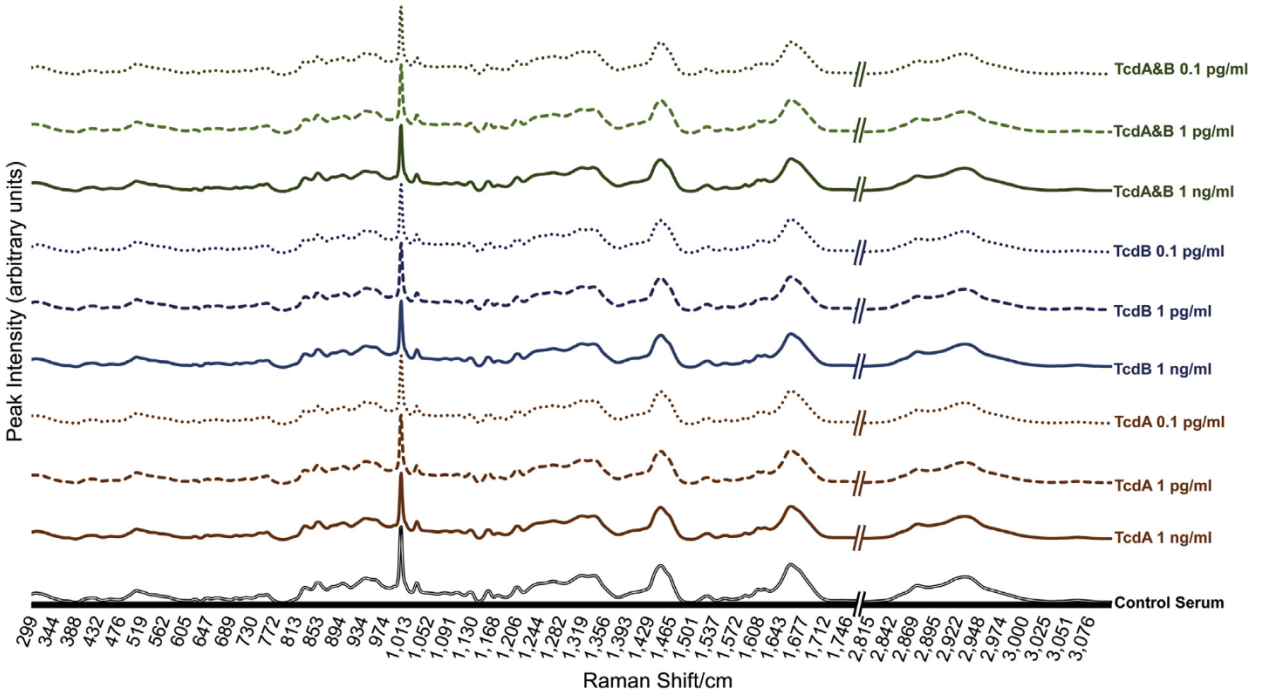

Fig. 1 – Mean Raman spectra. Mean Raman spectra for control serum (black double line), toxin-spiked serum for TcdA (brown), TcdB (blue), and TcdA and B (green) for 1 ng/ml (solid line), 1 pg/ml (dashed line) and 0.1 pg/ml (dotted line) concentrations for Raman shift 300-3100 cm^{-1} (excluding 1701-2799 cm^{-1}) were plotted on x-axis and their intensity in arbitrary units on y-axis. (Color version of figure is available online.)

statistically unreliable conclusions. For this reason, we used SVM for classification. SVM is a superior method compared to PLS-LDA for classification and linear SVM allows us to extract VIP scores for each variable toward the classification. SVM for classification successfully separated each toxin-spiked serum from its control serum 100% at every concentration tested.

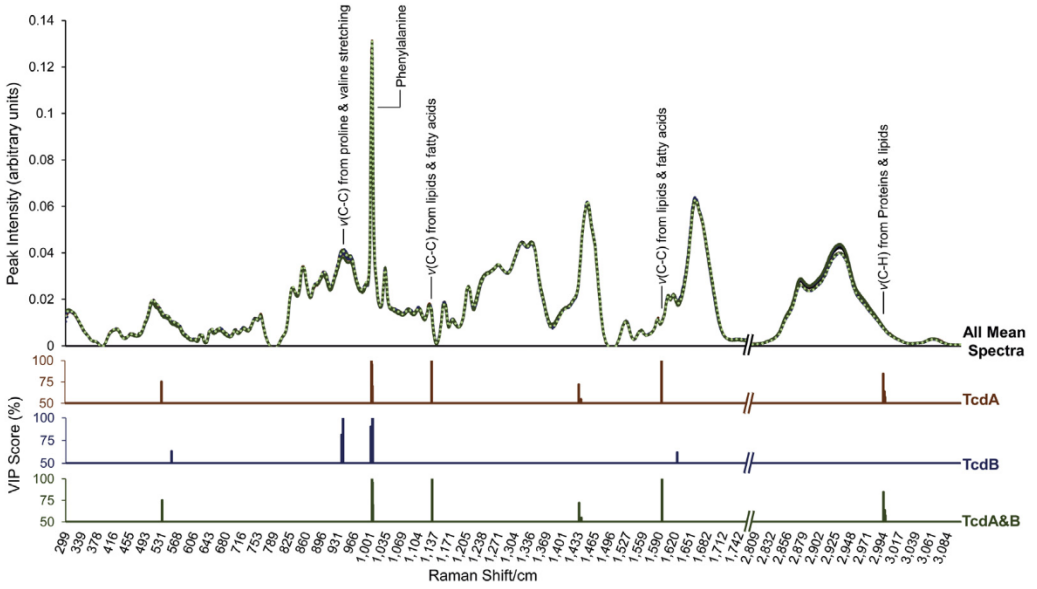

Fig. 2 – Variable Importance: All mean Raman spectra from Fig. 1 were plotted to scale for Raman shift 300-3100 cm^{-1} (excluding 1701-2799 cm^{-1}). VIP scores above 50% from SVM were plotted for TcdA (brown), TcdB (blue), and TcdA and B (green) below all mean spectra. The VIP score of a variable is calculated from the absolute value of the t-statistic for each model parameter used in SVM. Biological peak assignments for important Raman peaks identified through VIP scores were given. (Color version of figure is available online.)

Table 1 – C-SVC model performance.

Toxin	Conc.	% Error		% Sensitivity	% Specificity
		Cross validation	0.632 bootstrap		
TcdA	1 ng/mL	18.75	12.19	87.7	77.3
	1 pg/mL	11.25	8.41	90.6	85.2
	0.1 pg/mL	0	0	100	100
TcdB	1 ng/mL	3.75	2.32	100	90
	1 pg/mL	7.5	4.98	96	86.7
	0.1 pg/mL	8.75	6.15	91	91.7
TcdA and B	1 ng/mL	15	9.88	87	80.1
	1 pg/mL	2.5	2.03	93.4	100
	0.1 pg/mL	0	0.09	100	100

Error rates, sensitivity, and specificity from cross-validation and bootstrap methods were given in Table 1. SVM for classification performed better than PLS-LDA (Supplemental Table 1).

Using the biomarker signature that was identified using SVM, TcdB at 1 ng/mL concentration was clearly distinguished from the control serum with 100% sensitivity and 90% specificity. As the concentration of the TcdB decreased, the sensitivity and specificity slightly decreased. This trend was opposite for TcdA and TcdA and B. At 0.1 pg/mL concentration, TcdA and TcdA and B were distinguished from control serum with 87.7 % and 87% sensitivity and 77.3% and 80.1% specificity, respectively. As the concentration of TcdA and TcdA and B decreased, the sensitivity and specificity increased to 100% (Table 1). TcdA and B followed the trend of TcdA suggesting that TcdA signature is highly influencing the detection of TcdA and B in serum.

Next, using the VIP scores from SVM, we identified Raman shifts that contributed most to the classification model performance. VIP scores that contributed more than 50% for the distinguishing toxin-spiked serum from control serum were plotted in Figure 2.

Overall, the Raman biomarker signature we identified in this analysis, successfully and rapidly identified TcdA, TcdB, and both in serum at 0.1 pg/mL concentrations with high sensitivity and specificity.

Discussion

The Raman spectra are complex in nature and often contain broad Raman shifts due to the superposition of contributions from all the molecules present. In general, Raman spectral features over the range of 400 to 800 cm^{-1} are associated with nucleic acids and nucleotide conformation; 800 to 1200 cm^{-1} contain spectral contributions from nucleic acids, lipids (C-C, C-O), proteins (C-N, C-C), and C-O-C stretching of carbohydrates (saccharides); 1200 to 1400 cm^{-1} result from Amide III vibrations (C-N, N-H) with contributions from CH2/CH3 vibrations of proteins and lipids (phospholipids), as well as from nucleic acids; and 2800-3100 cm^{-1} corresponds to CH, CH2, and CH3 stretching vibration of lipids and proteins.[19]

Using VIP scores from SVM, significant spectral differences between control and toxin-spiked serum were identified (Fig. 2). Spectral regions between 1000 and 1008 cm^{-1}, which correspond to vibrations from phenylalanine were significantly different between control serum and all toxin-spiked serum (TcdA, TcdB, and TcdA and B). But spectral features 1132 cm^{-1} and 1595 cm^{-1} that correspond to carbon and carbon (C-C) vibrations from lipids and fatty acids and region between 2800-3000 cm^{-1} that correspond to carbon and hydrogen (C-H) vibrations from proteins and lipids were specific for TcdA. Similarly, spectral regions 928-940 cm^{-1} that correspond to vibrations from carbon-carbon (C-C) stretching probably from proline and valine were specific to TcdB. Spectral features of TcdA significantly contributed to TcdA and B model suggesting that TcdA signature is stronger compared to TcdB when both are present. Overall, these spectral differences suggest underlying differences in proteins and lipids in the serum from the toxin.

A limitation of this study is that we used toxin-spiked serum for our analysis. In the future experiments, we will use serum from severe CDI patients and to account for variability in serum from person to person, we will use larger sample. Next, in case of TcdA and TcdA and B as the concentration of toxin decreased (1 ng-0.1 pg/mL), error rates decreased (18.75% to 0% and 15% to 0%, respectively) suggesting that it was easier to separate TcdA or TcdA and B spiked serum from control serum at lower concentrations. We expect it to be harder to separate toxin-spiked serum from control serum at the lower concentrations of the toxin. This was not the case for the TcdB and as the concentration decreased (1 ng-0.1 pg/mL), as expected model performance suffered (3.75%-8.75%). This unusual behavior might be contributed by TcdA as this is only observed in case of TcdA and TcdA and B spiked serum. Further studies are needed to determine if this might be due to presence of any neutralizing antibodies against TcdA in the serum. It is known that neutralizing antibodies in the sera make it difficult to detect TcdA and TcdB in cell-based toxicity assays. We did not account for that in this study and in our future experiments, we will determine the effect of neutralizing antibodies on the model performance.

Despite these differences, TcdA, TcdB, and TcdA and B were detected in serum with relatively high sensitivity and

specificity at lower concentrations (1 ng-0.1 pg/mL). Overall, error rates can be decreased by increasing sample number and employing better classification algorithms such as deep learning that take advantage of large sample data. The turnaround time for RS is less than 30 min. This can be further decreased considerably by decreasing wavelength range to be recorded during the RS to the important regions such as phenylalanine and CH vibrations. These results suggest that RS has great potential to detect *C difficile* toxins in serum at very low concentrations.

Further evaluation of this technique will be done with larger toxin-spiked serum sample and its control from severe CDI patients and results will be validated with CCAT and ICT.

Conclusion

RS gives a molecular fingerprint of the sample and it does not require any antibodies or labels. Currently, most diagnostic modalities are either not sensitive enough or have long turnaround time and require expertise to perform. RS is cost-effective and has 30-min turnaround time that can be further reduced on optimization to a few minutes. Because of these advantages, RS has great potential to be a point-of-care testing tool for rapid detection of *C difficile* toxins in serum for critically ill CDI patients and continuous-monitoring patients.

Acknowledgment

Author's contributions: SKK was involved in study design, protocol development, statistical and computational analysis, interpretation of the data, and manuscript preparation and revision. SY was involved in acquisition of Raman spectra from serum samples, study design, Raman protocol development, and manuscript approval. JVM and DML were involved in toxin-spiked serum preparation, study design, toxin protocol development, and manuscript approval. DML was involved in IRB application and approval. CH was involved in preprocessing of Raman spectra and manuscript approval. MB, GWA, and LND were involved in all aspects of this study, including study design, analysis, manuscript preparation, and final approval. GWA and LND are coprinciple investigators.

Disclosure

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors have no financial disclosures and report no conflicts of interest.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jss.2018.06.031.

REFERENCES

1. Lessa FC, Winston LG, McDonald LC. Emerging infections program CdST burden of Clostridium difficile infection in the United States. *N Engl J Med*. 2015;372:2369–2370.
2. Korman TM. Diagnosis and management of Clostridium difficile infection. *Semin Respir Crit Care Med*. 2015;36:31–43.
3. Kelly CP, Kyne L. The host immune response to Clostridium difficile. *J Med Microbiol*. 2011;60:1070–1079.
4. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, et al. The role of toxin A and toxin B in Clostridium difficile infection. *Nature*. 2010;467:711–713.
5. Solomon K. The host immune response to Clostridium difficile infection. *Ther Adv Infect Dis*. 2013;1:19–35.
6. Carter GP, Chakravorty A, Pham Nguyen TA, Mileto S, Schreiber F, et al. Defining the roles of TcdA and TcdB in localized gastrointestinal disease, systemic organ damage, and the host response during Clostridium difficile infections. *MBio*. 2015;6:e00551.
7. Steele J, Chen K, Sun X, Zhang Y, Wang H, et al. Systemic dissemination of Clostridium difficile toxins A and B is associated with severe, fatal disease in animal models. *J Infect Dis*. 2012;205:384–391.
8. Salcedo J, Keates S, Pothoulakis C, Warny M, Castagliuolo I, et al. Intravenous immunoglobulin therapy for severe Clostridium difficile colitis. *Gut*. 1997;41:366–370.
9. Yu H, Chen K, Wu J, Yang Z, Shi L, et al. Identification of toxemia in patients with Clostridium difficile infection. *PLoS One*. 2015;10:e0124235.
10. Chang TW, Lauermann M, Bartlett JG. Cytotoxicity assay in antibiotic-associated colitis. *J Infect Dis*. 1979;140:765–770.
11. He X, Wang J, Steele J, Sun X, Nie W, et al. An ultrasensitive rapid immunocytotoxicity assay for detecting Clostridium difficile toxins. *J Microbiol Methods*. 2009;78:97–100.
12. Pollock NR. Ultrasensitive detection and quantification of toxins for optimized diagnosis of Clostridium difficile infection. *J Clin Microbiol*. 2016;54:259–264.
13. Grow AE, Wood LL, Claycomb JL, Thompson PA. New biochip technology for label-free detection of pathogens and their toxins. *J Microbiol Methods*. 2003;53:221–233.
14. Pahlow S, Meisel S, Cialla-May D, Weber K, Rosch P, et al. Isolation and identification of bacteria by means of Raman spectroscopy. *Adv Drug Deliv Rev*. 2015;89:105–120.
15. Strutt JW. XV. On the light from the sky, its polarization and colour. *Lond Edinb Dublin Phil Mag*. 1871;41:107–120.
16. Chase B. A new-generation of Raman instrumentation. *Appl Spectrosc*. 1994;48:A14–A19.
17. Raman CV. A change of wave-length in light scattering. *Nature*. 1928;121:619.
18. Raman CV, Krishnan KS. A new type of secondary radiation. *Nature*. 1928;121:501–502.
19. Talari ACS, Movasaghi Z, Rehman S, Rehman IU. Raman spectroscopy of biological tissues. *Appl Spectrosc Rev*. 2015;50:46–111.
20. Roth D. *Automated background removal from spectra forums.ni.com: LabVIEW formus*. 2008. Available at: https://forums.ni.com/t5/LabVIEW/Automated-background-removal-from-spectra/td-p/733511.
21. Schulze G, Jirasek A, Yu MM, Lim A, Turner RF, et al. Investigation of selected baseline removal techniques as candidates for automated implementation. *Appl Spectrosc*. 2005;59:545–574.
22. Wold S, Esbensen K. Geladi P Principal component analysis. *Chemometrics Intell Lab Syst*. 1987;2:37–52.
23. Helland I. *Partial least squares regression*. Wiley StatsRef: Statistics Reference Online; 2014.

24. Gromski PS, Xu Y, Correa E, Ellis DI, Turner ML, et al. A comparative investigation of modern feature selection and classification approaches for the analysis of mass spectrometry data. *Anal Chim Acta*. 2014;829:1–8.
25. Worley B. Powers R PCA as a practical indicator of OPLS-DA model reliability. *Curr Metabolomics*. 2016;4:97–103.
26. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press; 2000:189.
27. Westerhuis JA, Hoefsloot HCJ, Smit S, et al. Assessment of PLSDA cross validation. *Metabolomics*. 2008;4:81–89.
28. Jain AK, Dubes RC, Chen CC. Bootstrap techniques for error estimation. *IEEE Trans pattern Anal Mach Intell*. 1987;5:628–633.
29. R-Core-Team R. *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018.
30. Rakotomalala R. TANAGRA: un logiciel gratuit pour l'enseignement et la recherche. *EGC*. 2005:697–702.