CHAPTER 14

Spectral Fingerprinting of Pathogenic Bacteria by Fourier Transform Infrared Spectroscopy and Chemometric Models for Rapid Identification

Niladri Sekhar Chatterjee, Toms C. Joseph, Minimol V.A.

Email: niladri_icar@hotmail.com

Introduction

Identification and typing of pathogens employing rapid systems in the area of food safety, epidemiological surveillance and disease outbreak assumes more significance in the present scenario, as consumers are more concerned about the food that they consumed and the environment they live in (World Health Organization, 2008). The advanced microbiology techniques over conventional ones using biosensors and enzyme-based technologies ensures the rapid and accurate identification of pathogens within 6-9 h time (Franco-Duarte *et al*., 2019). Simultaneously, the DNA based molecular level identification system further eases the confirmation of specific bacteria or bacterial products (Adzitey *et al.,* 2013). DNA based techniques such as whole genome sequencing and next generation sequencing are widely used in routine diagnostics and surveillance studies where the most promising strain specific differences need to be studied (Besser *et al.,* 2018).

Similarly, the metabolomic fingerprinting can be used to monitor the conformation and composition of biochemical compounds in microbial cells and there by identification of microbes (Mester *et al*., 2018). Rapid analytical techniques for microbial identification include matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI TOF MS), near-infrared Fourier transform (NIR FT) Raman spectroscopy and Fourier transform infrared (FTIR) spectroscopy (Dinkelacker *et al.,* 2018). The accuracy of test results from these techniques are comparable and similar to most widely used DNA-based methods such as pulsed-field gel electrophoresis (PFGE), multi-locus sequence typing (MLST) and single nucleotide polymorphism (SNP) analysis of whole-genome sequencing (WGS) (Davis, 2012). The use of analytical techniques might provide useful information regarding the source or origin of contamination and their by necessary measures can be taken to reduce the contamination to safe level

Among the different analytical instruments, FTIR offers rapid discrimination, classification and identification of microbial cells by observing the vibration properties of chemical bonds present in the cell when excited by an infrared beam (Puzey et al., 2008). FTIR enables the identification of microbes upto subspecies level (Zarnowiec *et al.,* 2015) Use of FTIR spectroscopy has been successfully applied for the identification several pathogens including *Listeria, Bacillus, Staphylococcus, Clostridium*, *Escherichia coli* and *Lactobacillus* (Dawson and Upton,2014; Li *et al.,* 2018)

Methicillin resistant strains of *Staphylococcus aureus* have emerged as an important nosocomial infective agent and several clinical and foodborne outbreaks have been reported in recent years (Lakhundi, 2018). Similarly, certain strains of *Vibrio cholerae i.e*., *V. cholerae* O1, *V. cholerae* O139 and several serovariants of them are encountered in recent foodborne outbreaks (Finkelstein, 1996). The differentiation of these bacteria by conventional biochemical methods is difficult. However, this can be achieved by using FTIR which is rapid and effective.

There are very few studies as on date on the use of FTIR for the identification and differentiation of pathogenic bacteria of seafood origin. Amiali et al. (2007) reported the rapid identification of glycopeptide intermediate *S. aureus* (GISA) from glycopeptidesusceptible (GSSA) among methicillin resistant *S. aureus* (MRSA) by using FTIR. They concluded that FTIR can be used as an alternative to conventional susceptibility test for the identification of GISA and GSSA among MRSA isolates. Similarly, Li et al. (2018) established a subtyping method using FTIR with principal component analysis and hierarchical cluster anlasis for epidemiological surveillance of pathogenic *V. parahaemolytius.* Hence, the present study was carried out to analyze the spectral responses of biochemical compounds present in the cell wall by FTIR for identification of *Staphylococcus aureus* as methicillin sensitive and methicillin resistant strains as well as *Vibrio* isolates.

Bacterial ioslates

In this study, bacterial isolates recovered from seafood orgin were used. A total of sixty isolates each comprising of fifteen nos. of *S. aureus,* fifteen nos. of Methicillin resistant *S. aureus* (MRSA), fifteen nos. of Methicillin resistant coagulase negative *S. aureus* (MRCoNS), fifteen nos. of Coagulase negative *S. aureus* (CONS) isolated from retail fish markets and aquaculture farms of three district of kerala namely Ernakulam, Kottayam, and Alapuzha were selected for FTIR analyses. All *Staphylococcus* isolates were isolated during the period of 2012-2015 by Murugadas

et al. (2016). The identity of the isolates was further reconfirmed by following the procedure described in Bacteriological analytical manual for the isolation and identification of *Staphylococcus aureus* (Tallent et al., 2001). Briefly, the glycerol stock of each isolate was revived by inoculating a loopful of culture asceptically to trypticase soya broth and then streaked on to baired parker agar (Himedia, India) and incubated for 48 h at 37 ºC. Black colured colony from each BP agar plates were subjected to following biochemical tests such as Gram staining, catalase test and coagulase test. The methicillin susceptibility of the isolates was confirmed phenotypically by disc diffusion test (CLSI 2014). All the isolates from each category were maintained at -80º C in 30% glycerol until use.

Four species of *Vibrio* isolates comprising 60 nos each comprising of fifteen no's *V. cholerae,* fifteen no's *V. parahaemolyticus,* fifteen no's *V. mimicus, and* fifteen no's *V. vulnificus* were selected to anayse the discriminatory power of FTIR to differentiate the isolates to each category. All the *Vibrio* isolates were isolated from seafood samples from landning centres and retail markets of Cochin as a part of an ongoing project for the screening of seafood samples for the presence of emerging and reemerging foodborne pathogens by Microbiology, Fermentaion and Biotecnology Division, ICAR-CIFT (Unpublished data). Briefly, 190 seafood samples were screened for the presence of *Vibrio* species during the period of June 2018 to December 2019 and all the seafood samples were subjected to selective enrichment in alkaline peptone water (APW) and plating on selective agar (Thiosufate citrate bile salt sucrose agar (TCBS)) for the isolation of *Vibrio* species. Sucrose fermenting and non fermenting colonies were taken randomly from each plate, purified and subjected to a series of biochemical test for identification (Kaysner et al., 2004). Further all biochemically confirmed isolates from each category were further confirmed molecularly by polymerase chain reaction targeting their respective species specific prmers (Bej et al, 1999; Kumar and Lalitha, 2013; Guardiola- Avila et al., 2016; Canigral et al., 2010). All the confirmed isolates from each category were maintained at -80º C in 30% glycerol with 3% salt until use.

Measurement of FTIR spectral responses

The sample preparation was carried out as per the procedure of Grunert et al. (2013). The bacterial isolates were grown overnight at 37° C in trypticase soy broth (TSB). one ml of culture was taken and centrifuged at 5000 rpm for 5 min (Centrifuge 5430R, Eppendorf, India). The supernatant

was discarded and the bacterial pelled was washed with sretile water. The optical density (OD_{620}) of the bacterial suspension was adjusted to 0.5. Then, 30 µl of the suspension was placed on the ZnSe sample plate and dried at 37° C for 20-30 minutes. The transparent film obtained was taken for infrared measurement in Nicolet iS10 FTIR spectrometer, secondary Nicolet iZ10 module in 2010, loaded with Omni software for the data processing and anlysis (Thermo fisher scientific India PVT, LTD Mumbai). The individual spectra of each isolate were recorded in Absorbance/Transmission mode in the spectral range of 4000-400cm^{-1}. The spectral aqusition was carried out by averaging 50 scans for each sample and three replicate responses were recorded for each sample.

Principal of molecular spectroscopy

FTIR, FTNIR, Raman, and Hyperspectral imaging (HSI) techniques utilize the interaction of light with molecules in gas, liquid or solid matter to reveal crucial compositional details. When IR radiation is directed to a sample, some of it is absorbed, reflected back, or transmitted. The plot of %Absorbance/%Transmittance/%Reflectance against wavelength or wavenumber is the IR spectra which provides a molecular fingerprint of the sample. For FTIR spectra, the wavelength falls in the mid infrared region, typically 2500 to 25000 nm, whereas, the NIR spectral region is typically 800 to 2500 nm (Osborne, Fearn, & Hindle, 1993, Sun, 2009)

When IR light is pointed on a mater, the molecules within undergoes vibrational excitation. Infrared light is absorbed when its frequency matches with the frequency of energy transitions caused by the vibrational excitations. Molecular vibrations that cause a change in dipole moment of the molecules can only result in absorption of IR light. Hence, bonds with zero dipole moment show very weak absorbance. The IR light in mid-infrared region can cause fundamental transitions corresponding to stretching and bending vibrations, hence reveale crucial information on the functional group composition of the molecule. No two molecules have the same FTIR fingerprint and the unique fingerprints have important applications in food authenticity determination (Sun, 2009). Traditionally, to obtain a good quality FTIR spectrum, sample preparation involving KBr palletization and Nujol mall is required. Spectra can be recorded by transmitting mid IR radiation through a sample. However, the thickness of a liquid or solid sample has a direct bearing on the intensity of the spectral features. This limits the use of FTIR spectroscopy in food analysis. Advent of novel sampling accessories such as "Attenuated Total Reflectance (ATR)", and "Diffuse Reflectance" has largely solved this problem of direct sample analysis.

The working principal of the ATR accessory is presented in the schematic Figure 1. When infrared light is pointed on an optically dense diamond crystal at a certain angle, the internal reflectance caused by the high refractive index creates an evanescent wave at the surface of the crystal and interacts with the sample placed in contact with the crystal. The sample absorbs energy from the evanescent wave thereby altering or attenuating it. The attenuated evanescent wave meets back with the IR beam, which then exits the opposite end of the crystal and reaches the detector. The detector system then generates the IR spectrum. For this technique to be successful, the sample should be in good contact with the crystal surface, since the evanescent waves do not extend a few microns beyond the crystal surface (Cocciardi, 2003).

The working principle of a diffuse reflectance accessory is presented in the schematic Figure 2. When light is illuminated on an uneven surface, the unevenness causes the light to reflect at all angles. This phenomenon is called diffuse reflectance. Fuller and Griffiths in 1978, first described the diffuse reflectance accessory for recording FTIR spectra in this particular mode (Fuller & Griffiths, 1978). Ever since, the accessory has been modified and adopted in commercial instruments. During measurements, the IR light passing through a lens system illuminates the sample at normal incidence. The diffused light is collected back through the lens system along the same axis as the incident IR beam. The plot of %Reflectance against the wavelength generates the diffuse reflectance spectra. This technique of FTIR analysis has recently found application in food analysis (Yang et al., 2021). High-throughput transmission (HTT) is another sampling technique where a transmission cell is used for holding liquid samples. This technique is particularly suitable for authenticity testing of oils and fats (Lerma-Garcia et al., 2010).

Chemometric data analysis tools and algorithms

An untargeted analysis of food even with a miniature spectrometer generates a large amount of data. Moreover, visual discrimination between the spectral fingerprints of an authentic vis-a-vis adulterated food sample is not straight forward. Such spectral dataset contains large number of variables depending upon the resolution of the spectrometer used. Meaningful interpretation of such food fingerprints need assistance from chemometrics. Broadly, chemometrics can be defined as the tool box of various statistical algorithms, machine learning, and cloud computing for extracting chemically relevant information from food. During the last decade, the chemometrics

aided detection of food fraud, adulteration, determination of quality, freshness, and traceability have seen rapid growth in innovations (Medina et al., 2019). Figure 5 presents a summary of various types of chemometrics techniques used in spectroscopy data analysis for food authentication.

The digital signals generated from a spectroscopy sensor may contain several unwanted information arising from electrical interferences, surrounding effects, and baseline shifts. To correct these aberrations in signal, several pre-processing techniques are recommended including peak alignment (Jellema, 2009; Savorani et al., 2010). Thankfully, spectroscopy signals are less prone to signal drift or horizontal shifts as compared to chromatographic signals. For chemometric analysis, the digital signals are arranged into a data matrix where the variables (usually wavelength or wavenumber) are assembled in a column and the corresponding signal values (usually, absorbance, % reflectance, transmittance etc.) are presented in a row. The pre-processing algorithms apply independently on each signal of the data matrix.

Smoothing, Baseline correction, Normalization, Signal correction, and Derivatization are the major pre-processing treatment, applied singly or in a chain to a data matrix. The major smoothing algorithms include “Savitzky-Golay”, “Exponentially Weighted Moving Average” (EWMA), “Wavelet Denoise Spectral” (WDS), “Moving window”, and “Asymmetric Least Square Smooting” (AsLS). Among the smoothing techniques, Savitzky-Golay is most commonly used where noise is removed by applying a moving polynomial to the data matrix (Rahman et al., 2019). WDS technique is particularly used for removing high frequency coefficients from the data set.

For baseline correction, “Row-Center”, “Offset correction”, “Linear correction”, and AsLS correction techniques are used. Row-Center technique that shifts each spectrum by subtracting the row mean from each row value is often successful for building chemometric models using spectroscopy data. The “Standard Normal Variate (SNV)” filter is commonly used for spectra normalization that subtracts the mean from each signal value and then divides by the standard deviation of the signal values (Mishra et al., 2021). Variation in spectroscopy signals due to different path-length and scatter effect is corrected using the “Multiplicative Signal Correction (MSC) technique. The MSC technique normalizes each spectrum by regressing against the average spectrum from a selected set of spectra (Mishra et al., 2021). Further, to reduce scatter

effect first and second order derivatization of the spectra is performed. For new application development, initially different pre-processing techniques are evaluated either singly or in combination to arrive at most optimum latent data set for model building (McVey et al., 2021).

Besides signal pre-processing, removal of redundant or uninformative variables is also important for creation of rugged chemometric models. Inclusion of so many variables from a complex spectrum may result in overfitting of the model. Unfortunately, for a new application, it is not known which part of the spectra will be most useful. For variable selection, initially, a full model is created including all the available variables. The model is optimized for all pre-processing treatments and classification/regression algorithms. Then, the variables that fall above a defined threshold of variable importance in projection (VIP) are selected (Chatterjee et al., 2019, Gomes et al. 2022). Hence, a subset using the selected variables are created which is further used to develop the final chemometric model. Variable selection is more important for Multiple Linear Regression (MLR), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) methods which use a dataset of original registered variables and without any pre-processing treatments (Yun et al., 2013).

Chemometric methods are selected based on the objective of the study. The methods are broadly classified as unsupervised/exploratory methods and supervised methods. Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) are the unsupervised multivariate analysis tools. PCA is highly useful for initial data structure exploration. A PCA scores plot reveals important insights on the intrinsic variability of the data set and natural groupings of samples. This in turn helps in selecting the important variables. However, one should keep in mind that PCA is strictly a data exploration technique and not to be used as a classification tool (Jiménez-Carvelo et al., 2019). HCA analysis, presented as dendrograms, shows the similarity or relationship between cluster of samples. Each node in the dendrogram denotes a cluster (Drab & Daszykowski, 2014).

Purpose of the supervised methods can be discrimination/classification, quantification or both. The major statistical tools for discrimination or class analysis are “k nearest neighbour (kNN)”, “Soft Independent Modelling of Class Analogies (SIMCA)”, “Partial Least Square Discriminant Analysis (PLSDA)”, “Orthogonal Partial Least Square Discriminant Analysis (OPLSDA)”, “Linear Discriminant Analysis (LDA)”, and “Artificial Neural Network (ANN)”. Essentially,

different classification models define the class boundaries for sets of samples based on a training set of representative samples comprising different classes (Oliveri, 2017). In case of kNN and SIMCA, the classification is strictly qualitative and works best for binary classification. On the other hand, the PLS algorithm-based analysis performs a multivariate regression and assigns a numeric value to each sample before classifying them into different classes (Brereton & Lloyd, 2014). There are merits and demerits of each of these classification models and they face several issues for application in regulatory control of food authenticity (Rodionova et al., 2016). However, recently one class classifier, particularly one class SIMCA (ocSIMCA) has been particularly successful in providing reliable results (Rodionova et al., 2016; Horn et al., 2018).

Newer algorithms, such as “Support Vector Machine (SVM)”, “Classification and Regression Tree (CART)”, and “Random Forest (RF)” are also finding application in the area of food authentication. The advantage of these algorithms is that it can be used for both classification and quantification models. Despite their advantage over conventional classification methods and successful applications in the field of metabolomics and ecology, examples of application in food authentication are still limited (Cutler et al., 2007; Gromski et al., 2015). On the other hand, purely quantitative algorithms such as “Principal Component Regression (PCR)”, “Partial Least Square Regression (PLSR)”, and “Multiple Linear Regression (MLR)” are widely used for non-destructive determination of a quantitative parameter in food material. Particularly, determination of the content of an adulterant in food non-destructively or with minimum sample preparation is important application of these quantitative statistical methods (Ejeahalaka et al., 2020).

Conclusion

Using chemometrics and molecular spectroscopy it is possible to identify microbial pathovers in real time. However, these chemometrics models are built initially using authentic isolates and comprehensive spectral datbases needs to be built. Once the chemometric algorithms are trained, they will be able to identify unknown sample in real time. These rapid methods can be highly accurate and saves lot of time in identification.

Reference:

1. Adzitey, F., Huda, N. and Ali, G.R.R., 2013. Molecular techniques for detecting and typing of bacteria, advantages and application to foodborne pathogens isolated from ducks. *3 Biotech*, *3*(2), pp.97-107.

- Amiali, N.M., Mulvey, M.R., Berger-Bächi, B., Sedman, J., Simor, A.E. and Ismail, A.A., 2007. Evaluation of Fourier transform infrared spectroscopy for the rapid identification of glycopeptide-intermediate Staphylococcus aureus. *Journal of antimicrobial chemotherapy*, *61*(1), pp.95-102.
- Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L. and Trees, E., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical microbiology and infection*, *24*(4), pp.335-341.
- Dinkelacker, A.G., Vogt, S., Oberhettinger, P., Mauder, N., Rau, J., Kostrzewa, M., Rossen, J.W., Autenrieth, I.B., Peter, S. and Liese, J., 2018. Typing and species identification of clinical Klebsiella isolates by Fourier transform infrared spectroscopy and matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Journal of clinical microbiology*, *56*(11), pp.e00843-18.
- Franco-Duarte, R., Černáková, L., Kadam, S., Kaushik, K.S., Salehi, B., Bevilacqua, A., Corbo, M.R., Antolak, H., Dybka-Stępień, K., Leszczewicz, M. and Relison Tintino, S., 2019. Advances in Chemical and Biological Methods to Identify Microorganisms—From Past to Present. *Microorganisms*, *7*(5), p.130.
- Grunert, T., Wenning, M., Barbagelata, M.S., Fricker, M., Sordelli, D.O., Buzzola, F.R. and Ehling-Schulz, M., 2013. Rapid and reliable identification of Staphylococcus aureus capsular serotypes by means of artificial neural network-assisted Fourier transform infrared spectroscopy. *Journal of clinical microbiology*, *51*(7), pp.2261-2266.
- Kumar, R. and Lalitha, K.V., 2013. Prevalence and molecular characterization of Vibrio cholerae O1, non-O1 and non-O139 in tropical seafood in Cochin, India. *Foodborne pathogens and disease*, *10*(3), pp.278-283.
- Lakhundi, S. and Zhang, K., 2018. Methicillin-resistant Staphylococcus aureus: molecular characterization, evolution, and epidemiology. *Clinical microbiology reviews*, *31*(4), pp.e00020-18.
- Li, Z., Chen, S., Xu, C., Ju, L. and Li, F., 2018. Rapid subtyping of pathogenic and nonpathogenic Vibrio parahaemolyticus by fourier transform infrared spectroscopy with chemometric analysis. *Journal of microbiological methods*, *155*, pp.70-77.

6. Mester, P., Robben, C., Witte, A.K., Kalb, R., Ehling-Schulz, M., Rossmanith, P. and Grunert, T., 2018. FTIR Spectroscopy Suggests a Revised Mode of Action for the Cationic Side-Chain Effect of Ionic Liquids. *ACS combinatorial science*, *21*(2), pp.90-97
7. Puzey, K.A., Gardner, P.J., Petrova, V.K., Donnelly, C.W. and Petrucci, G.A., 2008, April. Automated species and strain identification of bacteria in complex matrices using FTIR spectroscopy. In *Chemical, biological, radiological, nuclear, and explosives (CBRNE) sensing IX* (Vol. 6954, p. 695412). International Society for Optics and Photonics.
8. World Health Organization, 2008. *Foodborne disease outbreaks: guidelines for investigation and control*. World Health Organization.
9. Zarnowiec, P., Lechowicz, L., Czerwonka, G. and Kaca, W., 2015. Fourier transform infrared spectroscopy (FTIR) as a tool for the identification and differentiation of pathogenic bacteria. *Current medicinal chemistry*, *22*(14), pp.1710-1718.
10. Cañigral, Irene, Yolanda Moreno, Jose L. Alonso, Ana González, and María A. Ferrús. "Detection of Vibrio vulnificus in seafood, seawater and wastewater samples from a Mediterranean coastal area." *Microbiological research* 165, no. 8 (2010): 657-664.
11. Guardiola-Avila, I., Acedo-Felix, E., Sifuentes-Romero, I., Yepiz-Plascencia, G., Gomez-Gil, B. and Noriega-Orozco, L., 2016. Molecular and genomic characterization of Vibrio mimicus isolated from a frozen shrimp processing facility in Mexico. *PloS one*, *11*(1).
12. Kaysner, C.A., DePaola, A.J. and US Food and Drug Administration, 2004. bacteriological analytical manual; methods for specific pathogens; Chapter 9 Vibrio.
13. Bej, A.K., Patterson, D.P., Brasher, C.W., Vickery, M.C., Jones, D.D. and Kaysner, C.A., 1999. Detection of total and hemolysin-producing Vibrio parahaemolyticus in shellfish using multiplex PCR amplification of tl, tdh and trh. *Journal of microbiological methods*, *36*(3), pp.215-225.