Development of a Rapid Method for Detection and Differentiation of *Escherichia coli* Serotypes and Strains Using Attenuated Total Reflectance Fourier Transform Infrared Microspectroscopy

**THESIS**

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Crystal N Cook

Graduate Program in Food Science and Technology

The Ohio State University

2016

Master's Examination Committee:

Dr. Luis Rodriguez-Saona, Advisor

Dr. Lynn Knipe

Dr. Ahmed Yousef
Copyrighted by
Crystal Cook
2016
Abstract

*Escherichia coli* is a common bacterium found in the intestinal tract of all mammals, environment, and foods. Pathogenic *E. coli* cause many outbreaks, hospitalizations, and deaths with Shiga toxin-producing *E. coli* (STEC) O157 being the leading source of foodborne outbreaks in the United States of America. Both STEC O157 and STEC non-O157 (e.g. O26, O45, O103, O111, O121, and O145) create millions of dollars of economic burden due to treatment costs and productivity losses.

Currently, detection and identification of pathogenic *E. coli* is dependent on time consuming, labor intensive procedures that often require a trained employee to perform. Today’s food industry needs a rapid, sensitive, and accurate method for detecting and identifying *E. coli* strains. Mid infrared spectroscopy is an easy-to-use, rapid detection technique that measures the molecular vibrations of atoms within a molecule producing unique spectral patterns or fingerprints. Furthermore, infrared microspectroscopy (IRMS) improves the sensitivity, reproducibility, differentiation, and speed capabilities of IR spectroscopy by coupling an infinity-corrected microscope to a high performance IR spectrometer. IRMS can provide capabilities for high-throughput screening of microorganisms and the ability to resolve spectral profiles within desired regions of the target. The objective of this study was to develop a rapid method for identification and differentiation of *E. coli* strains using IRMS and chemometrics.
The *E. coli* strains were generously provided by Dr. Joshua Daniels of the Department of Veterinary Clinical Sciences at The Ohio State University (Columbus, OH). *E. coli* strains (15) from 3 different serotypes were grown on tryptic soy agar (TSA) plates at 37°C for 24 hours. Bacteria cells were immersed into 100% ethanol (to enhance safe handling of the pathogenic bacteria), rinsed twice with water, transferred to a NEO-GRID membrane, vacuum dried, and analyzed directly by an IR microspectrometer. An attenuated total reflectance (ATR) germanium crystal was used to collect spectra in the 4000 cm\(^{-1}\) to 700 cm\(^{-1}\) range, and to increase the signal to noise ratio, 128 scans were co-added during collection of the spectra. Bacterial spectra were collected on 4 to 5 different days with 10-15 spectral measurements per day. Soft independent modeling of class analogy (SIMCA) (a supervised, classification technique) was employed to cluster *E. coli* strains and predict class membership for new additions. *E. coli* strains exhibited distinctive and reproducible infrared spectra in the fingerprint region (1500-900 cm\(^{-1}\)) of the IR spectrum. SIMCA permitted distinguishing *E. coli* strains through differences in bacterial cell envelope components. IRMS combined with chemometrics provide a simple and rapid phenotypic procedure for the reliable identification of *E. coli* strains and it could be potentially useful for an efficient and reliable monitoring of bacterial contamination in food with minimal sample manipulation.
For my family

Especially my parents: Randall and Natalie

Thank you for always loving me and believing in my dreams
Acknowledgments

Thank you Dr. Luis Rodriguez-Saona for your patience, guidance, and encouragement.

Thank you Dr. Ahmed Yousef for serving as my committee member, providing invaluable advice, and allowing me to use your laboratory and equipment.

Thank you Dr. Lynn Knipe for serving as my committee member and always providing your consistent smile and friendship.

Thank you Dr. Joshua Daniels for lending me your E. coli strains and advice.

Thank you Marcal Plans Pujolras, Mei-ling Shotts, Peren Aykas, Congcong Zhang, and Santosh Dhakal for helping me get through so much in the past two years.

Thank you Lisa Brannan and Kraig Barnhart for your love, advice, and encouragement.

A special thanks to my wonderful husband, Caleb Schwind, for your constant support, loyalty, and love.
Vita

September, 12 1991 .......................................................... Huntington, West Virginia
2010-2014 ................................................................. B.S. Microbiology, Auburn University
2014-2016 ...................... M.S. Food Science and Technology, The Ohio State University

Fields of Study

Major Field: Food Science and Technology
Table of Contents

Abstract ................................................................................................................................. ii
Acknowledgments .................................................................................................................. v
Vita ........................................................................................................................................ vi
Fields of Study ...................................................................................................................... vi
Table of Contents .............................................................................................................. vii
List of Tables ....................................................................................................................... x
List of Figures ..................................................................................................................... xi
Chapter 1: Literature Review ........................................................................................... 1
  1.1 The Principles of Infrared Spectroscopy .................................................................. 1
    1.1.1 Early History of Infrared Light ........................................................................ 1
    1.1.2 The Infrared Region of the Electromagnetic Spectrum .................................. 2
    1.1.3 Molecular Vibrations ....................................................................................... 3
    1.1.4 Infrared Spectrum ............................................................................................. 5
    1.1.5 Fourier Transform Infrared Spectroscopy ....................................................... 6
    1.1.6 Attenuated Total Reflectance .......................................................................... 8
    1.1.7 Infrared Microspectroscopy ............................................................................. 9
1.1.8 Soft Independent Modeling of Class Analogy (SIMCA) ........................................... 10

1.1.9 Bacterial Cell Surface .................................................................................................. 11

1.1.10 Band Assignments ...................................................................................................... 12

1.2. Applied Infrared Spectroscopy in Microbiology ................................................................. 13

1.2.1 Qualitative and Quantitative Analysis of the *Salmonella* Genus Using FTIR Spectroscopy .................................................................................................................. 13

1.2.2 Qualitative and Quantitative Analysis of the *Listeria* Genus Using FTIR Spectroscopy .................................................................................................................. 14

1.2.3 Qualitative and Quantitative Analysis of the *Escherichia coli* Species Using FTIR Spectroscopy .................................................................................................................. 16

1.3 Conclusions .................................................................................................................... 17

Chapter 2: Differentiating Between Different Serotypes and Strains of *Escherichia coli* Using Attenuated Total Reflectance Fourier Transform Infrared Microspectroscopy ..... 18

2.1 Abstract .......................................................................................................................... 18

2.2 Introduction .................................................................................................................... 19

2.3 Materials and Methods .................................................................................................. 22

2.3.1 Bacterial Cultures and Growth Conditions ................................................................. 22

2.3.2 Sample Preparation .................................................................................................... 23
2.3.3 Fourier-transform infrared spectroscopy ......................................................... 24

2.3.4 Multivariate Analysis .................................................................................. 26

2.4 Results and Discussion ................................................................................. 27

2.5 Conclusion .................................................................................................... 36

References ......................................................................................................... 38
List of Tables

Table 1. Shows the bond type and possible compound that is associated with bands at certain wavenumbers, which are seen when analyzing *E. coli* with FTIR spectroscopy. 33

Table 2. The interclass distances between *E. coli* serotypes and strains as determined by the SIMCA models of ATR FTIR spectra. ......................................................... 34

Table 3. Misclassification matrix, which is based on the SIMCA model of the spectra of *E. coli* serotypes and strains obtained using ATR FTIR spectroscopy, indicates the number of samples that were misclassified. ................................................................. 35
List of Figures

Figure 1. Representation of the infrared region and its proper place in association with the other regions within the electromagnetic spectrum. ................................................................. 3

Figure 2. Visual representation of how the difference in energy (ΔE) is dependent on the type of bond. .......................................................................................................................... 5

Figure 3. A typical absorbance spectrum obtained from E. coli O101:H9 bacteria using ATR FTIR spectroscopy. .................................................................................................................. 6

Figure 4. Schematic diagram of how a Michelson interferometer splits up and recombinates an IR light beam.................................................................................................................. 8

Figure 5. The benchtop Excalibur Series 3100 FTIR spectrometer used to collect spectra. ........................................................................................................................................ 25

Figure 6. The 600 UMA FTIR microscope with a slide-on ATR accessory attachment and a germanium objective, which were all used in spectra collection........................................... 26

Figure 7. ATR FTIR, bacterial spectra of E. coli O8:H9 (green), E. coli O101:H9 (blue), and E. coli O156:H8 (orange) serotypes with a range of 4000 to 700 cm\(^{-1}\). ......................... 28

Figure 8. SIMCA models developed from ATR FTIR spectra of (A) various strains from E. coli serotypes O8:H9, O101:H9, and O156:H8, (B) different strains of the E. coli O8:H9 serotype, (C) certain strains of the E. coli O101:H9 serotype, and (D) distinct strains of the E. coli O156:H8 serotype. .............................................................................. 30
Figure 9. Graphs displaying the discriminating power of each band for differentiating (A) various strains from E. coli serotypes O8:H9, O101:H9, and O156:H8 and (B) the different strains of E. coli O8:H9 (orange), E. coli O101:H9 (blue), and E. coli O156:H8 (green).
Chapter 1: Literature Review

1.1 The Principles of Infrared Spectroscopy

1.1.1 Early History of Infrared Light

Sir Frederick William Herschel (1738-1822) was a German astronomer that is accredited with being the first to discover the infrared (IR) region of the electromagnetic spectrum in 1800 (Mantsch 2001). Sir Herschel used a prism to disperse white light, creating a spectrum, and then detected a higher temperature just outside the red light region (Wehling 2010). The increased temperature provided evidence of energy, later identified as IR radiation, outside the visible light spectrum (Wehling 2010). In 1905, an American physicist named Dr. William Coblentz (1873-1962) published a book that provided hundreds of spectra for various substances and indicated that functional groups absorb specific and characteristic IR wavelengths (Schrader 1995). Dr. William Coblentz’s research on IR spectroscopy was largely ignored, since he was a physicist rather than a chemist (Schrader 1995).

World War II (1939-1945) initiated major interest within the United States of America and the United Kingdom in the applications of IR spectroscopy due to its ability to determine the origin of the gasoline used by German air force (Mantsch 2001).
However, publications in the field of applied IR spectroscopy were strongly restricted by wartime regulations, which inhibited the usage of IR spectroscopy in areas outside wartime applications (Mantsch 2001). The diverse capabilities of IR spectroscopy become broadly accepted within the scientific community in 1945 due to the publication of what is now recognized as the molecular spectroscopy bible “Infrared and Raman Spectra of Polyatomic Molecules” written by Dr. Gerhard Herzberg (1904-1999) (Mantsch 2001). Afterwards, IR spectroscopy became a widely accepted research tool for identifying the functional groups of organic compounds (Wehling 2010). IR spectroscopy has since made its way into various fields of research, but it was initially used in the food industry by the 1970s, where near IR spectroscopy was used to quantify moisture, fat, and protein content in foods, especially cereal grains (Wehling 2010).

1.1.2 The Infrared Region of the Electromagnetic Spectrum

The electromagnetic spectrum (Figure 1) is a range of all types of electromagnetic radiation (energy) (Alvarez-Ordóñez and Prieto 2012). Each portion of the electromagnetic spectrum is defined by its wavelengths or wavenumbers (Alvarez-Ordóñez and Prieto 2012).

\[
\text{wavenumber} = \frac{1}{\text{wavelength}}
\]

The IR region consists of light with wavenumbers from 5 to 12,500 cm\(^{-1}\) or wavelengths from \(10^{-1}\) to \(10^{-4}\) cm, and it, therefore, can be found between the visible and
microwave regions of the electromagnetic spectrum (Alvarez-Ordóñez and Prieto 2012).
The IR region can be broken down into three regions, including far IR (5 - 400 cm\(^{-1}\)), mid IR (MIR) (400 to 4000 cm\(^{-1}\)) and near IR (NIR) (4000 to 12,500 cm\(^{-1}\)) (Alvarez-Ordóñez and Prieto 2012).

![Figure 1. Representation of the infrared region and its proper place in association with the other regions within the electromagnetic spectrum.](image)

1.1.3 Molecular Vibrations

Light has wave-particle duality, since it consists of both particle-like and wave-like properties (Klein 2012). When representing light as a particle, light is made up of photons or discrete packets of energy (Klein 2012). The energy of each photon can be calculated using Planck’s equation (Klein 2012).

\[
\text{Energy} = \text{speed of light} \times \text{Planck’s constant}/\text{wavelength}
\]

Molecules can store energy via the molecular orbital that their electrons occupy and rotational and vibrational motions of their bonds (Klein 2012). However, the
molecule’s bonds can only rotate and vibrate at certain energy levels (Klein 2012). The nature of the bond (e.g. functional group and double, triple, cis, and trans bonds) determines the difference in energy (ΔE) between these energy levels (Figure 2) (Klein 2012). If a photon of light from a specific wavelength contains ΔE amount of energy, then the molecule’s bond will absorb the photon (Klein 2012). This absorption of energy causes vibrational excitation, where the photon is stored by vibrational and rotational motion of the bond and then released back into the environment, often as heat (Klein 2012). Vibrational motion consists of stretching and bending types of motion (Alvarez-Ordóñez and Prieto 2012). Whether the type of vibrational motion is bending or stretching depends on whether the bond’s angle or length is changing (Alvarez-Ordóñez and Prieto 2012). Stretching vibrational motions include asymmetric and symmetric stretching, and bending vibrational motions include scissoring, wagging, twisting, and rocking (Alvarez-Ordóñez and Prieto 2012).
1.1.4 Infrared Spectrum

IR spectra are formed based on the energy absorbed and released by bonds that have experienced vibrational excitation (Klein 2012). As mentioned, the ΔE is dependent on the type of bond, so each type of bond will absorb a characteristic frequency (Klein 2012). When we create an IR spectrum, we hit a substance with all the frequencies of IR light (Klein 2012). The frequencies that are absorbed by the sample are determined by measuring the percent transmittance, from which an absorption spectrum is created (Figure 3) (Klein 2012). Absorption spectra are usually reported in wavenumbers, the relationship of which can be explained by a simple equation (Alvarez-Ordóñez and Prieto 2012).
wavenumber = frequency/the speed of light

The nature of the bond is then determined based on the frequencies that the absorption spectrum indicates were absorbed (Klein 2012). Using IR spectra we can identify unknown substances, differentiate between different samples, and quantify various components of the sample, such as fat and protein content (Wehling 2010).

![Figure 3. A typical absorbance spectrum obtained from E. coli O101:H9 bacteria using ATR FTIR spectroscopy.](image)

1.1.5 Fourier Transform Infrared Spectroscopy

Before Albert A. Michelson developed the two-beam interferometer (Figure 4) in 1891 (Griffiths and de Haseth 2007), researchers would spend years determining the spectrum of a sample with a simple molecular structure due to the slow and labor intensive process that went into making a spectrum with a dispersive instrument (Alvarez-Ordóñez and Prieto 2012). A dispersive spectrometer uses a monochromator, which disperses the light source into individual frequencies, and then passes each
frequency through the sample with a detector measuring the absorbance of each frequency (Wehling 2010). FTIR spectrometers use the Michelson interferometer to collect a sample’s spectrum in a matter of minutes by collecting and analyzing all the frequencies at once instead of the successive way the dispersive spectrometer collects spectrum (Alvarez-Ordóñez and Prieto 2012). To this day the Michelson interferometer is still the most commonly used model for an IR spectrometer (Griffiths and de Haseth 2007). It splits the IR light beam, creates an optical path difference, and then recombines the beam using mirrors, which creates inference (Griffiths and de Haseth 2007). The detector will then create a pattern of energy intensity (interferogram) based on the intensity of the IR radiation as a function of the optical path difference (Griffiths and de Haseth 2007). A FTIR spectrometer also uses FT (mathematical treatment) to transform the collected data into a generic IR spectrum (Griffiths and de Haseth 2007).
1.1.6 Attenuated Total Reflectance

Attenuated total reflectance (ATR) is a technique used to collect data from samples that are too thick to collect IR spectra via transmission measurements, (e.g. pastes and highly viscous liquids) by measuring the total amount of IR radiation reflected from the sample’s surface (Wehling 2010). An IR light beam is shown through an internal reflection element (IR transmitting crystal) made from material with a high refractive index (e.g. diamond, ZnSe, germanium, etc.) that is touching the surface of the sample (Wehling 2010). The high refractive index causes the light beam to bounce repeatedly off the sample’s surface (Wehling 2010). With each bounce the IR light beam
penetrates a short distance into the sample, and it is then reflected back into the IR transmitting crystal (Wehling 2010). The radiation, at the wavelengths of which the sample absorbs the radiation, decreases in intensity, creating a spectrum similar to a spectrum obtained via transmission measurements (Wehling 2010). ATR is the best MIR spectroscopy technique for lessening the effects of water without involving the complete dehydration of the sample, subtraction of the water’s signal, or usage of a thin, hydrated sample (Wang and Mizaikoff 2008).

1.1.7 Infrared Microspectroscopy

An IR microscope is an instrument that combines the abilities of an IR spectrometer and a microscope to allow the analysis of samples with a minimum diameter of 10 μm (Griffiths and de Haseth 2007). The IR microscope uses a Cassegrain lens to condense the light beam, allowing analysis of a small sample area (Griffiths and de Haseth 2007). It also uses a mercury cadmium telluride (MCT) detector, which is a photon detector, to allow rapid scanning of the sample (Alvarez-Ordóñez and Prieto 2012). The conductivity of the detector varies depending on the amount of radiation hitting the detector, allowing the detector to respond quickly to smaller alterations in radiation intensity (Alvarez-Ordóñez and Prieto 2012). However, cryogenic cooling using liquid nitrogen is needed to cut out thermal noise (Wehling 2010).
1.1.8 Soft Independent Modeling of Class Analogy (SIMCA)

SIMCA is a supervised, classification technique that combines principal components analysis (PCA) and class information with the goal of developing classification rules for each class. By using training data to create a confidence interval around each class, which allows the probable class membership of new data to be determined (Ballabio and Todeschini 2009). Supervised classification of data involves a machine learning task, which is based on pattern recognition, where the supervised learning algorithm can classify new samples by analyzing the training data (a set of examples that are manually classified) and produce an inferred function (Wang and Mizaikoff 2008). PCA is an unsupervised pattern recognition technique that does not use predefined classes, but rather aims to describe the overall variation of the data by determining the similarities of the data (Ballabio and Todeschini 2009). PCA projects the data into a condensed hypervolume, which is based on the principal components (PCs) or linearly uncorrelated variables. PCs are organized based on how much variance they cause within the data, where the first PC has the largest variance and so on (Ballabio and Todeschini 2009). SIMCA uses a different PC model for each class (Wang and Mizaikoff 2008). A new sample of an unknown class is then compared to each class’s PC model, and the residual distances between the unknown sample and each PC model are calculated (Wang and Mizaikoff 2008). If the unknown sample has a residual distance below the statistical limit for that class, it is predicted that the sample belongs in that
class (Wang and Mizaikoff 2008). As a result, an unknown sample can be indicated as belonging to multiple classes, if it has residual distances from the classes that are below the statistical limit for those classes, and not belonging to any class, if all the residual distances exceed the upper limit for all the classes (Wang and Mizaikoff 2008). The SIMCA algorithm can also easily handle data collinearities (Wang and Mizaikoff 2008).

1.1.9 Bacterial Cell Surface

The bacterial cell surface is an important factor that allows FTIR spectroscopy to differentiate and identify different bacteria (Burgula and others 2007). It consists of surface proteins, lipopolysaccharides (LPS), capsular polysaccharides, capsules, flagella, peptidoglycan (PG), fimbriae, outer membrane (OM) proteins, lipoteichoic acids, glycolates, lectins, and specific layers (S-layers) (Burgula and others 2007).

Bacteria can be divided into Gram negative and Gram positive categories based on the structure of their cell wall (Burgula and others 2007). The cell wall of Gram negative bacteria, like E. coli, have lipoproteins covalently bound to a thin PG layer, while Gram positive cell walls have teichoic or teichuronic acids covalently bound to a thick PG layer (Burgula and others 2007). Gram positive bacteria cells have a rigid PG layer to provide strong support to the cell wall during times of osmotic stress (Burgula and others 2007). Though Gram negative bacteria cells lack this kind of protection from osmotic stress, they have an OM, which consists of an inner layer with phospholipids and
an outer layer with LPS (Burgula and others 2007). The OM prevents many harmful substances, such as antibiotics, bile salts, digestive enzymes, and hydrophobic drugs from entering the cell membrane (Burgula and others 2007). LPS is a surface molecule that plays roles in cell transport and adhesion (Walker and others 2004). LPS includes the O-antigen, core domain, which consists of inner and outer oligosaccharides, and lipid A (Burgula and others 2007). The O-antigen is a heteropolysaccharide that has antigenic properties (Burgula and others 2007). The lipid A is named so due to its function as a lipid anchor for the LPS into the OM (Burgula and others 2007).

1.1.10 Band Assignments

From 3000-2800 cm⁻¹ are bands indicating the presence of fatty acids (Burgula and others 2007). The presence of proteins and peptides is indicated by the amide A band at 3200 cm⁻¹, amide I band at 1695-1675 cm⁻¹, the amide II band at 1550-1520 cm⁻¹, and the amide III band at 1310-1240 cm⁻¹ (Burgula and others 2007). A band at 1468 cm⁻¹ indicates the CH₂ bending of methylene (Burgula and others 2007). From 1250-1220 cm⁻¹ are bands indicating the PO₂ stretching of phosphodiesters (Burgula and others 2007). Bands from 1200-900 cm⁻¹ indicates the presence of polysaccharides (Burgula and others 2007). The fingerprint region consists of the wavenumbers from 900-700 cm⁻¹ (Burgula and others 2007). The fingerprint regions is named so because the weak bands found in
this region have been shown to be unique to the type of bacteria being analyzed (Burgula and others 2007).

1.2. Applied Infrared Spectroscopy in Microbiology

1.2.1 Qualitative and Quantitative Analysis of the Salmonella Genus Using FTIR Spectroscopy

FTIR spectroscopy was used to differentiate between nonpathogenic and pathogenic Salmonella enterica serovars from different serogroups (Preisner and others 2012). ATR FTIR spectroscopy has been used to differentiate between viable and heat killed S. Typhimurium and S. Enteritidis using a zinc selenide crystal (Sundaram and others 2012a). ATR FTIR spectroscopy has been used to identify and differentiate between Salmonella serovars based on DNA characteristics using a diamond crystal (Sundaram and others 2012b). Rapid isolation and detection of Salmonella serovars has been performed with hydrophobic grid membranes (HGM) FTIR and S. enterica strains has been performed with ATR FTIR using a zinc selenide crystal (Männig 2007). Gas phase FTIR spectroscopy has been used to identify packaged beef that is contaminated with S. typhimurium by analyzing the headspace volatiles in a FTIR gas cell (Amamcharla and others 2010). Alvarez-Ordóñez and others (2010) used S. enterica serovars Typhimurium and Enteritidis to show how spectra can be influenced by growth
temperature, sodium chloride concentration in the growth medium, acidification of the
growth medium, and aerobic versus anaerobic growth conditions. IMS FTIR
spectroscopy has been used to detect, quantify, and differentiate between viable and heat-
treated *S. enterica* serovars on chicken breasts (Davis and others 2010a). De Lamo-
Castellvi and others (2010) determined that IMS FTIR could detect *Salmonella* serovars,
but the anti-*Salmonella* magnetic beads used created an interference signal that prevented
the differentiation between the *Salmonella* serovars. FTIR spectroscopy was used to
differentiate between five closely related phage types within the *S. enterica* serovar
Enteritidis (Preisner and others 2010). Zoumpopoulou and others (2010) used FTIR to
observe changes in major cellular constituents of *S. enterica* serovar Typhimirium and
investigate the anti-*Salmonella* properties exhibited by certain *Lactobacillus* strains.

1.2.2 Qualitative and Quantitative Analysis of the *Listeria* Genus Using FTIR
Spectroscopy

Romanolo and others (2015) used FTIR spectroscopy to differentiate between
*Listeria* species and *L. monocytogenes* serotypes. FTIR spectroscopy has also been used
to differentiate between pathogenic and nonpathogenic *Listeria monocytogenes* serotypes
(Rebuffo-Scheer and others 2007), strains of *L. monocytogenes*, and strains of *Listeria
innocua* (Nyarko and Donnelly 2015). Davis and Mauer (2011) subtyped the *L.*
monocytogenes species at the haplotype level using FTIR spectroscopy. Al-Qadiri and others (2008) was able to use FTIR spectroscopy to predict the extent of sublethal, heat injury in S. enterica serovar Typhimurium ATCC 14028 and L. monocytogenes ATCC 19113. The mode of action for polyhexamethylene biguanide (PHMB) (antibacterial agent) has been investigated using FTIR spectroscopy, the spectra of which support the hypothesis that PHMB disturbs the cytoplasmic membrane of L. innocua LRGIA 01 by interacting with the first layer of the membrane lipid bilayer (Chadeau and others 2012). FTIR spectroscopy can be used to detect organic acid stress for the L. innocua ATCC 33090 strain when it is treated with various sublethal and lethal levels of acetic acid (Wu and others 2013). Nyarko and others (2014) used FTIR to classify epidemic clones of L. monocytogenes and to further determine whether the strains were intact or heat-killed. In addition, FTIR was used to determine whether L. monocytogenes R2-764 cells injured by heat or acid return to their original physiological integrity after repairing themselves (Nyarko and Donnelly 2015). Since FTIR spectra showed repaired bacteria cells clustering closer to the intact cells, it was concluded that L. monocytogenes R2-764 cells do return to their original physiological integrity after recovering from heat or acid injury (Nyarko and Donnelly 2015).
1.2.3 Qualitative and Quantitative Analysis of the *Escherichia coli* Species Using FTIR Spectroscopy

FTIR spectroscopy can differentiate between *E. coli* O157:H7 and *Shigella sonnei* (Yang and others 2014). Dawson and others (2014) discriminated between 95 different uropathogenic *E. coli* (UPEC) strains from 6 different multilocus sequence types via FTIR spectroscopy. It has also been used to analyze the DNA structure, from which the screening characteristics and structural radiotolerance of nonpathogenic *E. coli* strains can be determined (Muntean and others 2014). FTIR spectroscopy was used to identify and differentiate between papG⁺ and papG⁻ strains of *E. coli* (Lechowicz and others 2013). Multilocus variable number tandem repeat analysis (MLVA) types of *E. coli* O157:H7 have been differentiated via FTIR spectroscopy (Davis and others 2012). *E. coli* clones from B2 (n=39) and D (n=13) phylogenetic groups have been differentiated and identified using ATR FTIR spectroscopy (Sousa and others 2013). Scholz and others (2011) used FTIR spectroscopy to estimate the concentrations of relevant bioprocess monitoring variables, like plasmid, biomass, and carbon sources (glucose, glycerol, acetate) in recombinant *E. coli* cultures. Wang and others (2010) quantified the amount of *E. coli* K12 that had been internalized into baby spinach samples via ATR FTIR spectroscopy. *E. coli* O157:H7 has been shown to be detectable on ground beef at a minimal level of $10^7$ CFU/g using FTIR spectroscopy with filtration and immunomagnetic separation (Davis and others 2010b).
1.3 Conclusions

FTIR spectroscopy is a rapid, easy-to-use, accurate method for differentiating and identifying bacteria in the food industry. When combined with the various techniques and statistical analysis methods available, the advantages and possibilities that can be attained with the use of FTIR spectroscopy increase dramatically. FTIR spectroscopy has made great strides within the last ten years, and, as a result, interest in FTIR spectroscopy has grown. FTIR spectroscopy would be a good alternative to genotype-based techniques.
Chapter 2: Differentiating Between Different Serotypes and Strains of *Escherichia coli* Using Attenuated Total Reflectance Fourier Transform Infrared Microspectroscopy

2.1 Abstract

*E. coli* is a common bacterium present in the intestinal tract of all mammals, environment, and foods. Pathogenic *E. coli* cause millions of dollars of economic burden and many outbreaks, hospitalizations, and deaths in the United States. The food industry needs a sensitive, accurate method for rapid detection and identification of *E. coli* strains. Mid infrared spectroscopy is a simple, rapid method for detection and identification that measures the molecular vibrations and rotations within a molecule to produce fingerprints (unique, spectral patterns) of the bacterial strains. IRMS combines an infinity-corrected microscope and a high performance IR spectrometer to further enhances the rapidity, discrimination, reproducibility, and sensitivity abilities of IR spectroscopy. IRMS also offers high throughput, sample screening and the ability to resolve spectral profiles within desired regions of the target. The experiment’s objective was to use IRMS and SIMCA to develop a rapid method for identification and differentiation of *E. coli* strains. The Department of Veterinary Clinical Sciences (Dr. Daniels) at The Ohio State University (Columbus, OH) kindly provided the *E. coli* strains, which were grown on
TSA plates at 37°C for 24 hours. Ethanol-killed bacterial cells were washed twice, transferred to a NEO-GRID membrane, vacuum dried, and directly analyzed with an IRMS using a germanium crystal. Spectra (4000-700 cm$^{-1}$) were collected for 4-5 days with 10-15 spectral measurements per day. The resulting spectra were then analyzed with SIMCA (a supervised, classification method) to cluster strains together and determine the probable class of new data. The fingerprint region (1500-900 cm$^{-1}$) provided unique spectra for discriminating and identifying *E. coli* strains. SIMCA clustered bacteria from 3 different serotypes (O8:H9, O101:H9, and O156:H8) and further differentiated 5 different strains within each serotype. Differences were associated to bacterial cell envelope components. IRMS coupled with chemometrics is an easy-to-use, rapid method for reliable identification of *E. coli* strains and would be useful for monitoring bacterial contamination in food with minimal sample preparation.

2.2 Introduction

The Foodborne Diseases Active Surveillance Network (FoodNet) published that 19,542 infections, 4,445 hospitalizations, and 71 deaths in 2014 were caused by mostly preventable, foodborne illnesses, which demonstrates that foodborne illnesses are a major burden to the health of the American population (Crim and others 2015). Foodnet has also conservatively estimated that foodborne illnesses cause 15.5 billion dollars per year
of economic burden, which is based on the cost of treatment and loss of wages due to illness and death (Hoffmann and others 2015). Shiga toxin-producing *E. coli* (STEC) O157 is a leading cause of foodborne illness outbreaks in the United States of America (Hoffmann and others 2015). In 2014, STEC accounted for 35% of the hospitalizations and 0.7% of the deaths, while STEC non-O157 accounted for 15% of the hospitalizations and 0.0% of the deaths in the United States (Crim and others 2015). Foodnet estimates that STEC O157 creates an economic burden of over $271 million, while STEC non-O157 probably creates an economic burden of over $27 million (Hoffmann and others 2015).

The official, detection methods for *E. coli* mandated by the United States Food and Drug Administration (FDA) are time consuming, labor intensive, and often require training to perform correctly, but the FDA deems them essential since identification of *E. coli* strains is needed to determine how dangerous the contamination will be and the source of contamination (Feng and others 2011). Tissue culture, serologic, and genotype-based techniques are all used to detect the presence of the pathogenic *E. coli* responsible for causing foodborne illnesses (Feng and others 2011). Tissue culture identification techniques are effective, easy-to-use, very sensitive, and provides a basis for supplemental techniques, like serologic and antigenic assays (Houpikian and Raoult 2002). However, tissue culture techniques can be time consuming as they can take several days, a large quantity of sample is often needed, and temperature, duration, and
atmosphere of incubation must be closely monitored (Houpikian and Raoult 2002).

Serological methods for identifying pathogenic *E. coli* strains are rapid (<1 days), convenient, highly specific, and effective at detecting very low antigen concentrations, but they can be expensive and takes trained personnel to get accurate results (Paton and Paton 2003). Genotype identification techniques are sensitive, specific, rapid (~4 hours) and require only a small sample size (Paton and Paton 2003). On the other hand, genotype-based methods can be easily contaminated, are expensive, and require trained personnel to attain accurate results (Paton and Paton 2003).

**FTIR spectroscopy** is a rapid, accurate technique that requires minimal sample preparation and a small sample size of less than a few milligrams (Alvarez-Ordóñez and Prieto 2012). Mid infrared spectroscopy provides unique spectral patterns by measuring the molecular vibrations of a compound (Klein 2012). Though instruments can be expensive, the cost of regular use of an infrared spectrometer would be less expensive compared to other identification methods (Alvarez-Ordóñez and Prieto 2012). Since infrared spectrometers are easy-to-use, employees would require little training (Alvarez-Ordóñez and Prieto 2012). **IRMS** further enhances the throughput, reproducibility, speed, and sensitivity of infrared spectroscopy by combining it with a microscope (Griffiths and de Haseth 2007). IRMS also allows the analysis of a smaller sample area (Griffiths and de Haseth 2007). **ATR** is a technique that enabled the collection of thick samples by touching the sample with highly refractive material to increase the intensity of the signal.
It’s also the best way to minimize water’s effects on a spectrum when using MIR spectroscopy (Wang and Mizaikoff 2008).

The objective of this study was to develop a rapid method using ATR FTIR microspectroscopy in combination with SIMCA for differentiation and identification of E. coli serotypes and strains.

2.3 Materials and Methods

2.3.1 Bacterial Cultures and Growth Conditions

Dr. Joshua Daniels of the Department of Veterinary Clinical Sciences at The Ohio State University (Columbus, OH) generously provided fifteen strains from three different serotypes of E. coli, including O8:H9 (strains 88.1590, 5.2674, 10.0579, and 11.0709), O101:H9 (strains 75.0385, 82.1356, 87.1833, 84.0980, and 98.1212), and O156:H8 (strains 96.0943, 78.0150, 97.1583, 12.2722, and 99.0184). These strains were originally obtained through Pennsylvania State University’s E. coli Reference Center, which performed the O and H typing. The O antigens were determined by using antisera via a procedure developed by Orskov and others (1977), while the H antigens were determined via Polymerase Chain Reaction-Restriction Fragment Length Polymorphism analysis of the fliC gene with a procedure developed by Machado and others (2000). Strains of each serotype were isolates from different cow breeds from different years to decrease the
change of having clones. Stock cultures are streak plated onto tryptic soy agar (TSA) plates (Difco, Becton Dickinson and Company, Franklin Lakes, NJ) and incubated at 37°C for 24 hours.

2.3.2 Sample Preparation

Sample preparation was based on a method developed by Mannig and others (2007) with modifications. Bacterial matter was transferred via a calibrated bacterial loop (10 μL, Thermo Fisher Scientific, Waltham, MA) from the TSA plates. The bacteria were killed for safety purposes by suspending them in 200 μL of ethanol (Thermo Fisher Scientific, Waltham, MA) in 2 mL centrifuge tubes (Thermo Fisher Scientific, Waltham, MA). The mixture was then vortexed for approximately 10 seconds, and samples were then centrifuged at 10,000 rpm for 5 minutes in a microcentrifuge (Model 5415 R, Eppendorf, Hamburg, Germany). After removal of the supernatant, the bacterial pellets went through a washing step twice, where they were resuspended in 200 μL of HPLC grade water (Thermo Fisher Scientific, Waltham, MA), vortexed for approximately 10 seconds, and centrifuged at 10,000 rpm for 5 minutes. The supernatant was again removed, and the pellet was transferred to hydrophobic grid membranes (hydrophobic-grid membrane filter, 0.45 μm porous, Neogen Corporation, Lansing, MI) within an NEO-GRID filtration system (NEO-GRID test system, Neogen Corporation, Lansing, MI) to concentrate the sample, allow high throughput by providing an easy platform array
for the collection of spectra, and prevent the samples from slipping after the drying process and spilling into each other. The drying process involved using a vacuum until the samples were flat (approximately 15 minutes) and using the desiccator until the samples were dry in 5 minute increments (approximately 5-10 minutes). The samples were analyzed immediately after preparation using ATR-FTIR.

2.3.3 Fourier-transform infrared spectroscopy

A benchtop Excalibur Series 3100 FTIR spectrometer (Agilent Technologies, Santa Clara, CA) (Figure 5) combined with a 600 UMA FTIR microscope (Agilent Technologies, Santa Clara, CA) (Figure 6) were used to collect spectra via ATR. The benchtop FTIR spectrometer and IR microscope were equipped with a Michelson interferometer, KBr beamsplitter, MCT detector, motorized x-y stage, a widefield 10X objective with an 18 mm diameter, and a 15X Cassegrain objective, to which a slide-on ATR accessory is attached. The IR microscope had a slide-on ATR germanium objective (refractive index 4, numerical aperture 2.4) (Agilent Technologies, Santa Clara, CA), which was cleaned with 70% ethanol and then acetone (Thermo Fisher Scientific, Waltham, MA) before and after analyzing each strain. Resolutions Pro Software (Version 4.1.0.101, Agilent Technologies, Santa Clara, CA) was used to calibrate the instruments each day and observe spectra during collection. Before a strain was analyzed, a
background spectrum was collected for later subtraction by taking a measurement of the air using the clean germanium objective. FTIR spectra, which included wavenumbers from 4000 to 700 cm\(^{-1}\), were obtained with a resolution of 4 cm\(^{-1}\) by co-adding 128 scans to increase the signal to noise ratio. For each strain, 18 spectra were collected per day for 5 days to include the variability that comes with growing bacteria in the training set.

Figure 5. The benchtop Excalibur Series 3100 FTIR spectrometer used to collect spectra.
2.3.4 Multivariate Analysis

Pirouette® software (Version 4.5 rev 1, Infometrix, Bothell, WA) was used to develop multivariate training models via SIMCA. SIMCA, a supervised pattern recognition discriminant analysis, was used to determine whether ATR FTIR spectroscopy can be used to differentiate between the different serotypes and strains of *E. coli*. The FTIR spectra were then transferred as *.spc* files into the Pirouette® software.
FTIR spectra were vector normalized, transformed to their second derivative (gap 25 points), and smoothed (gap 25 points). To identify the outliers within the data, Mahalanobis distance and sample residuals were used. If the interclass distance between 2 classes was 3 or above, the classes were considered statistically different. The data was projected onto the PC axes, creating a score plot to make sample clustering, including patterns, groupings, and outliers, easier to identify. Class residuals are used to assess the class distances. The discriminating power can be calculated by comparing the average residual variance of each class to all classes and residual variance of all classes to themselves. The discriminating power is then used to assess which variables or wavenumbers have a predominant effect on sample classification. It also helps to group the data together in a cluster and minimize the differences between the samples within a cluster, while maximizing the differences of samples from other clusters. At this point, class borders are determined based on 95% probability clouds (confidence interval), creating a hypervolume, which allows the identification of outliers that fall outside the confidence interval of a class and classification of new samples when they fall within the boundaries of a class’s confidence interval.

2.4 Results and Discussion

As can be seen in Figure 7, bacterial spectra provide a lot of information about the molecular structure of the bacterial cell components. The first large band at around 1650
cm\(^{-1}\) has been identified as amide I of a \(\alpha\)-helical structure, while the second large band at about 1520 cm\(^{-1}\) is known to be amide II (Burgula and others 2007). The small band at 1468 cm\(^{-1}\) indicates CH\(_2\) bending of methylene (Burgula and others 2007). The stretching of C=O in a COO\(^-\) group is signified by the band at 1390 cm\(^{-1}\) (Männig 2007). The band at around 1250 cm\(^{-1}\) was created by the overlapping bands of amide III and PO\(_2\) stretching of phosphodiesters (Burgula and others 2007). Any bands within the 1200 to 900 cm\(^{-1}\) range is due to C-O-C and C-O bonds in different polysaccharides (Burgula and others 2007).

**Figure 7.** ATR FTIR, bacterial spectra of *E. coli* O8:H9 (green), *E. coli* O101:H9 (blue), and *E. coli* O156:H8 (orange) serotypes with a range of 4000 to 700 cm\(^{-1}\).

The SIMCA models displayed in Figure 8 A-D are two-dimensional representations of the three-dimensional score plots that were developed by projecting the
data onto PC axes. The first SIMCA model (Figure 8 A) shows how the *E. coli* strains of the different serotypes separated into different clusters, which demonstrates how ATR FTIR spectroscopy can be used to differentiate *E. coli* strains based on serotype. In Figure 8 B, the SIMCA model shows various strains of the *E. coli* O8:H9 serotype clustering into separate groups. Figure 8 C presents some strains of the *E. coli* O101:H9 serotype dispersing into individual clusters, but *E. coli* O101:H9 82.1356 and 98.1212 bunched together, which suggests they are chemically similar to each other. As Figure 8 D illustrates, the *E. coli* O156:H8 78.0150 and 12.2722 developed into nice clusters, but the other *E. coli* O156:H8 strains remained clumped together, indicating they were very chemically similar.
Figure 8. SIMCA models developed from ATR FTIR spectra of (A) various strains from *E. coli* serotypes O8:H9, O101:H9, and O156:H8, (B) different strains of the *E. coli* O8:H9 serotype, (C) certain strains of the *E. coli* O101:H9 serotype, and (D) distinct strains of the *E. coli* O156:H8 serotype.

As can be observed in Figure 9, molecules that absorbed IR light within the 920 cm\(^{-1}\) to 1360 cm\(^{-1}\) range were the most influential in discriminating between *E. coli* serotypes and strains. Polysaccharides, the presence of which are represented by bands...
from 1200 cm\(^{-1}\) to 900 cm\(^{-1}\) (Burgula and others 2007), must play a major role in differentiating between both \(E.\ coli\) serotypes and strains, since Figure 9 A and B shows compounds that absorb IR light in the polysaccharide range are high discriminating powers. As seen in Figure 9 A, bands at 1059 cm\(^{-1}\) and 1083 cm\(^{-1}\) are shown to have the highest discriminating power when differentiating between the \(E.\ coli\) O8:H9, O101:H9, and O156:H8 serotypes. As can be seen in Table 1, the band at 1059 cm\(^{-1}\) is associated with C-OH bending (Kiwi and Nadtochenko 2005) and is probably caused by the presence of carbohydrates in the cell wall (Davis and Mauer 2010). On the other hand, the band at 1083 cm\(^{-1}\) is associated with symmetric vibration of \(\text{PO}_2^-\) bonds in LPS (Kiwi and Nadtochenko 2005), DNA, RNA, and phospholipids (Davis and Mauer 2010). Figure 9 B also shows that polysaccharides play major roles in differentiating between \(E.\ coli\) strains of the O8:H9 (via bands at 1016 cm\(^{-1}\), 1076 cm\(^{-1}\), and 1155 cm\(^{-1}\)), O101:H9 (via bands at 981 cm\(^{-1}\) and 1301 cm\(^{-1}\)), and O156:H8 (via the band at 1030 cm\(^{-1}\)) serotypes. As Table 1 shows, the band at 981 cm\(^{-1}\) is associated with bending of a HRC=CH\(_2\) bond (Klein 2012), which may be associated with carbohydrates in cells walls (Davis and Mauer 2010), and the band at 1016 cm\(^{-1}\) is linked with C-OH bending possibly due to lipopolysaccharide (Kiwi and Nadtochenko 2005). Table 1 also reveals that the band at 1030 cm\(^{-1}\) has been related to the stretching of C-O-C bonds in carbohydrates (Davis and Mauer 2010) and peptidoglycan, while the band at 1076 cm\(^{-1}\) has been conjoined to the stretching of C-O bonds in the rings of sugars (Kiwi and Nadtochenko 2005). Also, the
band at 1155 cm\(^{-1}\) has been associated with C-O stretching (Kiwi and Nadtochenko 2005) of carbohydrates in the cell wall (Davis and Mauer 2010), while 1219 cm\(^{-1}\) has been related to asymmetric vibration of a PO\(_2^-\) bond (Kiwi and Nadtochenko 2005) in the phospholipid bilayer (Lu and others 2011). The band at 1301 cm\(^{-1}\) has been connected with wagging of CH\(_2\) bonds in lipopolysaccharides (Kiwi and Nadtochenko 2005).

Figure 9. Graphs displaying the discriminating power of each band for differentiating (A) various strains from *E. coli* serotypes O8:H9, O101:H9, and O156:H8 and (B) the different strains of *E. coli* O8:H9 (orange), *E. coli* O101:H9 (blue), and *E. coli* O156:H8 (green).
Table 1. Shows the bond type and possible compound that is associated with bands at certain wavenumbers, which are seen when analyzing *E. coli* with FTIR spectroscopy.

<table>
<thead>
<tr>
<th>Frequency (cm(^{-1}))</th>
<th>Bond</th>
<th>Possible Compound</th>
</tr>
</thead>
<tbody>
<tr>
<td>981</td>
<td>HRC=CH(_2) bending</td>
<td>Carbohydrates in cell wall</td>
</tr>
<tr>
<td>1016</td>
<td>C-OH bending</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>1030</td>
<td>C-O-C stretching</td>
<td>Carbohydrates; Peptidoglycan</td>
</tr>
<tr>
<td>1059</td>
<td>C-OH bending</td>
<td>Carbohydrates in cell wall</td>
</tr>
<tr>
<td>1076</td>
<td>C-O stretching</td>
<td>Sugar rings</td>
</tr>
<tr>
<td>1083</td>
<td>PO(_2^-) symmetric vibration</td>
<td>DNA, RNA, phospholipids</td>
</tr>
<tr>
<td>1155</td>
<td>C-O stretching</td>
<td>Carbohydrates in cell wall</td>
</tr>
<tr>
<td>1219</td>
<td>PO(_2^-) asymmetric vibration</td>
<td>Phospholipid bilayer</td>
</tr>
<tr>
<td>1301</td>
<td>CH(_2) wagging</td>
<td>Lipopolysaccharide</td>
</tr>
</tbody>
</table>

Table 2 quantifies the distance between the clusters observed in Figures 8 A-D by indicating the interclass distances calculated by the SIMCA model. Classes with an interclass distance of 3 or above are considered statistically different from each other. As can be viewed in Table 2, the interclass distances between the serotypes ranged from 2.81 to 5.41 and strains from 1.09 to 8.20 with the highest interclass distances between strains of *E. coli* O8:H9 and the lowest interclass distances between strains of *E. coli* O156:H8 due to how chemically similar they are.
Table 2. The interclass distances between *E. coli* serotypes and strains as determined by the SIMCA models of ATR FTIR spectra.

<table>
<thead>
<tr>
<th>A</th>
<th>Serotypes</th>
<th>O8:H9</th>
<th>O101:H9</th>
<th>O156:H8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>@5</td>
<td>@5</td>
<td>@5</td>
</tr>
<tr>
<td>O8:H9@5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O101:H9@5</td>
<td>2.81</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O156:H8@5</td>
<td>3.64</td>
<td>5.41</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>O8:H9 Strains</td>
<td>10.0579</td>
<td>11.0709</td>
<td>5.2674</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@5</td>
<td>@5</td>
<td>@5</td>
</tr>
<tr>
<td>10.0579@5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.0709@5</td>
<td>3.13</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.2674@5</td>
<td>4.31</td>
<td>6.96</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>88.1590@5</td>
<td>4.06</td>
<td>3.32</td>
<td>8.20</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>O101:H9 Strains</td>
<td>82.1356</td>
<td>75.0385</td>
<td>87.1833</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@5</td>
<td>@5</td>
<td>@5</td>
</tr>
<tr>
<td>82.1356@5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75.0385@5</td>
<td>2.23</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>87.1833@5</td>
<td>4.69</td>
<td>6.17</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>84.0980@5</td>
<td>4.58</td>
<td>8.15</td>
<td>3.39</td>
<td>0</td>
</tr>
<tr>
<td>98.1212@5</td>
<td>2.05</td>
<td>3.10</td>
<td>5.39</td>
<td>5.96</td>
</tr>
<tr>
<td>D</td>
<td>O156:H8 Strains</td>
<td>78.0150</td>
<td>12.2722</td>
<td>97.1583</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@6</td>
<td>@6</td>
<td>@6</td>
</tr>
<tr>
<td>78.0150@6</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.2722@6</td>
<td>5.45</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.1583@6</td>
<td>2.08</td>
<td>1.48</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>99.0184@6</td>
<td>2.03</td>
<td>2.95</td>
<td>1.25</td>
<td>0</td>
</tr>
<tr>
<td>96.0943@6</td>
<td>3.15</td>
<td>1.48</td>
<td>1.09</td>
<td>1.71</td>
</tr>
</tbody>
</table>

@# specifies the number of principal components that were needed to differentiate the serotypes and strains

Table 3 A shows that 99% of the 913 samples were correctly classified whether the SIMCA model was differentiating at the serotype level or the strain level. Strains of
$E. coli$ O8:H9 (Table 3 B) and $E. coli$ O101:H9 (Table 3 C) were not misclassified, but only 96% of the 296 samples of $E. coli$ O156:H8 were correctly classified.

Table 3. Misclassification matrix, which is based on the SIMCA model of the spectra of $E. coli$ serotypes and strains obtained using ATR FTIR spectroscopy, indicates the number of samples that were misclassified.

<table>
<thead>
<tr>
<th></th>
<th>Serotypes</th>
<th>O8:H9 @5</th>
<th>O101:H9 @5</th>
<th>O156:H8 @5</th>
<th>No Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>O8:H9@5</td>
<td>341</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>O101:H9@5</td>
<td>0</td>
<td>296</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>O156:H8@5</td>
<td>0</td>
<td>0</td>
<td>270</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>O8:H9 Strains</td>
<td>10.0579@5</td>
<td>11.0709@5</td>
<td>5.2674@5</td>
<td>88.1590@5</td>
</tr>
<tr>
<td></td>
<td>10.0579@5</td>
<td>65</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>11.0709@5</td>
<td>0</td>
<td>71</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5.2674@5</td>
<td>0</td>
<td>0</td>
<td>84</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>88.1590@5</td>
<td>0</td>
<td>0</td>
<td>76</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>O101:H9 Strains</td>
<td>82.1356@5</td>
<td>75.0385@5</td>
<td>87.1833@5</td>
<td>84.0980@5</td>
</tr>
<tr>
<td></td>
<td>82.1356@5</td>
<td>86</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>75.0385@5</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>87.1833@5</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>84.0980@5</td>
<td>0</td>
<td>0</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>98.1212@5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>D</td>
<td>O156:H8 Strains</td>
<td>78.0150@6</td>
<td>12.2722@6</td>
<td>97.1583@6</td>
<td>99.0184@6</td>
</tr>
<tr>
<td></td>
<td>78.0150@6</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>12.2722@6</td>
<td>0</td>
<td>49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>97.1583@6</td>
<td>0</td>
<td>0</td>
<td>67</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>99.0184@6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>96.0943@6</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

@# specifies the number of principal components that were needed to differentiate the serotypes and strains.
2.5 Conclusion

The developed method successfully uses ATR FTIR microspectroscopy combined with SIMCA to discriminate between different serotypes and strains of ethanol-treated *E. coli* on a membrane-based platform with minimal sample preparation. However, as can be seen by observing the score plot of the class projections and the table of the interclass distances, the technique was better at separating the serotypes into their respective classes than it was at grouping the strains. The misclassification table shows that the samples could be classified into their correct serotypes and the strains of the *E. coli* O8:H9 and *E. coli* O101:H9 serotypes could be identified as the correct strain. The strains of the *E. coli* O156:H8 serotype were more difficult to separate, which could possibly mean their much more genetically similar. The SIMCA model also showed that the differentiating bands between the O8:H9, O101:H9, and O156:H8 serotypes were centered at 1059 cm\(^{-1}\) and 1083 cm\(^{-1}\). These frequencies correspond to absorption bands of carbohydrates in the cell wall represented by C-OH bending and PO\(_2^-\) symmetric vibration. The C-OH bending in LPS (1016 cm\(^{-1}\)), C-O bonds of sugar rings (1076 cm\(^{-1}\)), and C-O stretching from carbohydrates in the cell wall (1155 cm\(^{-1}\)) were all indicated as having high discriminating power for differentiating between the strains of the *E. coli* O8:H9 serotype. The stretching of C-O-C bonds in carbohydrates and peptidoglycan (1030 cm\(^{-1}\)) gave the highest discriminating power when differentiating between the strains of the *E. coli* O156:H8. The bending of HRC=CH\(_2\) bonds in the carbohydrates of the cell wall (981
cm$^{-1}$), PO$_2^-$ bonds in the phospholipid bilayer (1219 cm$^{-1}$), and wagging of CH$_2$ bonds in the LPS (1301 cm$^{-1}$) have been the highest discriminating powers responsible for the clustering of strains of E. coli O101:H9 serotype. ATR FTIR microspectroscopy would be an ideal method for discriminating and identifying E. coli strains in the industry setting since it is accurate, reproducible, quick, and easy-to-use. Future research should focus on using FTIR spectroscopy to detect, differentiate, and identify E. coli on food samples and quantify the amount of E. coli.
References


diseases active surveillance network, 10 U.S. sites, 2006-2014. MMWR 64(18): 495-9.


Rebuffo-Scheer CA, Schmitt J, Scherer S. 2007. Differentiation of listeria monocytogenes serovars by using artificial neural network analysis of fourier-


