Iowa State University

From the Selected Works of Chenxu Yu

2006

Identification of Pathogenic Bacteria in Mixed Cultures by FTIR Spectroscopy

Chenxu Yu

Joseph Irudayaraj, *Purdue University*

Available at: https://works.bepress.com/chenxu_yu/14/

IDENTIFICATION OF PATHOGENIC BACTERIA IN MIXED CULTURES BY FTIR SPECTROSCOPY

C. Yu, J. Irudayaraj

ABSTRACT. Routine identification of pathogenic microorganisms predominantly based on nutritional and biochemical tests is a time-consuming process. In this study, a method based on FTIR spectroscopy was developed to detect and identify the presence of five possible pathogenic bacteria in ten different microorganism mixtures with each cocktail containing up to three different species at a concentration of 10^9 CFU/mL. A mathematical approach based on factoring out the common features in the spectral fingerprints was developed and implemented in conjunction with discriminant analysis. FTIR spectra of the mixtures were directly obtained and analyzed using canonical variate analysis based on the discriminant model for the presence of particular species in the pathogenic bacteria mixtures. In nine out of ten mixtures, the predictions were 100% accurate; in one mixture, only one false negative was reported, with no false positives. Results suggest that FTIR spectroscopy combined with a suitable analytical procedure has excellent potential as a fast, powerful, and reliable alternative for identifying a specific foodborne pathogen in a complex and/or competing system.

Keywords. Bacteria identification, FT-IR spectroscopy.

Routine microbiological identification of food and clinical samples is largely based on nutritional and biochemical characteristics of microorganisms. Normally, more than one species of microorganisms could be present in the samples; hence, the first step in an identification process is to isolate, collect, and amplify single colonies, which normally takes 24 to 48 h. After this step, a single colony is cultured for another 16 to 24 h on solid or broth culture medium. A biomass of 10^6 to 10^9 cells is then used for biochemical assays. Consequently, it usually takes about 3 to 4 days to obtain the results through a labor-intensive and time-consuming process. Therefore, fast and reliable methods to detect pathogenic microorganisms at the earliest time are highly desirable.

Molecular biological techniques are now being evaluated and used as methods for the identification of microorganisms and the detection of specific antibiotic resistance genes (Tang et al., 1997; Nikkari and Relman 1999; Tenover et al., 1994; Clark et al., 1999). Although these techniques are potentially rapid, they are quite expensive and elaborate, and they require skilled personnel. False positive reactions due to DNA contamination and false negative reactions due to inhibitors introduced while preparing or collecting a sample may generate problematic results in DNA amplification-based molecular diagnostics (Vanechoutte and Eldere, 1997; Noordhoek et al., 1996; Ieven and Goossens, 1997;

Fredricks and Relman, 1998). As a first step, isolation of a single strain is still necessary, and precious time could be lost. Currently, molecular diagnostics are usually second lines of investigation and are seldom the sole basis for microbial identification (Maquelin et al., 2000).

An alternative approach in microbial characterization is the use of spectroscopic methods. Pyrolysis mass spectrometry has been evaluated as a method for bacterial characterization (Goodacre et al., 1996; Timmins et al., 1998; Barshick et al., 1999; Demirev et al., 1999); however, high instrumentation costs have prevented the widespread use of this method (Busse et al., 1996). The use of Fourier transform infrared (FTIR) spectroscopy for microbial identification and characterization has been gaining acceptance since the pioneering work of Naumann and co-workers (Naumann et al., 1991; Helm et al., 1991;). Timmins et al. (1998) used both pyrolysis-mass spectrometry and FTIR spectroscopy to differentiate *Candida* species and strains with much success. These phenetic classifications were found to be very similar to those obtained by genotypic studies, which examined the *HinfI* restriction enzyme digestion patterns of genomic DNA, and by use of the 27A *C. albicans*-specific probe. Discrimination between closely related isolates of *C. albicans*, *C. dubliniensis*, and *C. stellatoidea* was achieved in 10 s using FTIR spectroscopy. FTIR spectroscopy has been used to identify drug resistance in bacteria (Zeroual et al., 1995; Sockalingum et al., 1997; Bouhedja et al., 1997). Raman spectroscopy has also been explored in microbiology. Studies have been reported on Fourier transform (FT) Raman spectroscopy (Williams and Edwards, 1994; Edwards et al., 1995) and ultraviolet (UV) resonance Raman spectroscopy (Chadha et al., 1993; Manoharan et al., 1990; Ghiamati et al., 1992) to investigate suspensions and dried films of microorganisms as well as hydrated microbial smears taken from a solid culture medium. One major problem in applying these spectroscopic methods is to distinguish between the contributions of microorganism to the spectrum and the contribu-

Submitted for review in September 2005 as manuscript number BE 6064; approved for publication by the Biological Engineering Division of ASABE in August 2006.

The authors are **Chenxu Yu**, Post-Doctoral Researcher, and **Joseph Irudayaraj**, Associate Professor, Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana. **Corresponding author:** Joseph Irudayaraj, Department of Agricultural and Biological Engineering, 225 S. University St., Purdue University, West Lafayette, IN 47907; phone: 765-494-0388; fax: 765-496-1115; e-mail: josephi@purdue.edu.

tions of the background and environment (i.e., noise). Maquelin et al. (2000) developed a vector algebra based procedure to extract characteristic information from the Raman spectra of microorganisms by removing the contribution of the background medium and water.

Almost all of the published reports on applying spectroscopic techniques to identify microorganisms have focused on identification using single strains. In clinical diagnosis, the samples collected from a patient could potentially contain more than one microbial species. Hence, a more realistic problem would be to identify a particular pathogen from a mixture. However, this is extremely difficult since the spectrum of a microorganism could be due to the superposition of contributions from the biomolecules in a cell. Since most of these biomolecules are common to the microbial cells, the differences observed in the spectra of different species are usually quite small. To identify a particular microbial species in a microbial cocktail, it is therefore necessary to identify the key features that are characteristic of the species in question and to establish a database. Once this is achieved, the spectrum of an unknown mixture can be fingerprinted and analyzed for a match from the species of interest. Based on the analysis, the question of whether or not a particular species exists in the mixture could be addressed.

The main objective of this study was to develop a spectroscopic procedure for rapid identification of pathogenic microorganisms in a microbial mixture for clinical diagnosis of infectious diseases or to address food safety concerns. Specific objectives were to: (1) develop a mathematical procedure to obtain fingerprints of specific pathogenic microbes, (2) apply statistical tools to isolate and identify contributions from each ingredient of a microbial mixture, and (3) validate the method by identifying pathogens in a mixture. As of today, no method has been reported using FTIR spectroscopy to detect the presence or absence of bacteria in a cocktail of microorganisms. To demonstrate the validity of the methodology, model bacteria cocktail systems (in which the type and concentration of different microbes were controlled) were studied. Application of the method to real-world samples is to be addressed in future studies.

MATERIALS AND METHODS

SAMPLE PREPARATIONS

The five pathogenic bacteria analyzed (*Enterococcus faecium*, *Salmonella enteritidis*, *Bacillus cereus*, *Yersinia enterocolitica*, and *Shigella boydii*) were obtained from the Gastroenteric Disease Center (GDC) at the Pennsylvania State University (University Park, Pa.). Each species was

cultured in a 100 mL Luria-Bertani broth medium (5 g yeast extract, 8 g tryptone, and 5 g NaCl in 500 mL distilled water) at 35°C and shaken at 100 rpm for 24 h. The final concentrations of the bacterial cultures all reached 10⁹ CFU/mL after 24 h, as determined by plate counting. Ten possible combinations (table 1) of mixtures of three bacterial species (5 mL broth of each species) were made for analysis. In this study, the focus was to identify the presence or absence of a species in a bacterial cocktail. To ensure the acquisition of strong spectral signals, high concentrations of bacteria (10⁹ CFU/mL) were used. Improvement of the methodology for samples with lower concentrations will be addressed in the future.

Ten subsamples, each 8 mL, were drawn from each bacterial culture for replication of spectra acquisition and mixture preparation. The cultures were homogenized through stirring before subsamples were drawn. We fully understand that this cannot be counted as true replication; however, the main purpose of the work is to develop and test the mathematical procedure; hence, this practice of replication was considered to be sufficient.

FTIR MEASUREMENTS

Suspensions of each of the five bacterial species in the culture broth as well as mixtures were subjected to FTIR measurements using a Biorad Excalibur FTS 3000 spectrometer. The system consisted of a ZnSe ATR crystal, an air-cooled ceramic IR source, a deuterated triglycine sulfate detector (DTGS) at room temperature, and a KBr beam splitter. An 8 µL sample of the suspension at a concentration of 10⁹ CFU/mL was carefully pipetted onto the ZnSe ATR crystal at room temperature, and the spectrum was acquired after 2 min. FTIR signals of the suspensions on the ATR crystal were collected in the spectral region between 600 and 4000 cm⁻¹ at a resolution of 2 cm⁻¹. Plain broth was measured first as the background and subtracted from all of the sample spectra. An average of 256 scans was used as the spectrum of each sample. Ten spectra were taken for each bacterial species and each mixture sample. Only the fingerprint regions from 800 to 1800 cm⁻¹ and 2700 to 3100 cm⁻¹ (Maquelin et al., 2002) were used for microorganism differentiation.

MATHEMATICAL PROCEDURE TO EXTRACT FINGERPRINT FEATURES OF BACTERIAL SPECTRUM

We hypothesize that the FTIR spectrum of each bacterial species is a superposition of contributions from many biomolecules that have absorbance in the mid-IR range. This will cause variations among different replications (different samples coming from the same batch of growing bacteria) in

Table 1. Compositions of the mixture samples.

Mixture	Species 1	Species 2	Species 3
m1	<i>Enterococcus faecium</i>	<i>Salmonella enteritidis</i>	<i>Bacillus cereus</i>
m2	<i>Enterococcus faecium</i>	<i>Salmonella enteritidis</i>	<i>Yersinia enterocolitis</i>
m3	<i>Enterococcus faecium</i>	<i>Salmonella enteritidis</i>	<i>Shigella boydii</i>
m4	<i>Enterococcus faecium</i>	<i>Bacillus cereus</i>	<i>Yersinia enterocolitis</i>
m5	<i>Enterococcus faecium</i>	<i>Bacillus cereus</i>	<i>Shigella boydii</i>
m6	<i>Enterococcus faecium</i>	<i>Yersinia enterocolitis</i>	<i>Shigella boydii</i>
m7	<i>Salmonella enteritidis</i>	<i>Bacillus cereus</i>	<i>Yersinia enterocolitis</i>
m8	<i>Salmonella enteritidis</i>	<i>Bacillus cereus</i>	<i>Shigella boydii</i>
m9	<i>Salmonella enteritidis</i>	<i>Yersinia enterocolitis</i>	<i>Shigella boydii</i>
m10	<i>Bacillus cereus</i>	<i>Yersinia enterocolitis</i>	<i>Shigella boydii</i>

the analysis, which are not just due to measurement errors. Cells grown under the same conditions do not necessarily have to be identical, and their IR spectral characteristics may differ. Spectra measured for microbial samples taken from the same batch of growing bacteria often differ from each other on a small but non-negligible scale. When the species in question is present in a microbial mixture, the problem becomes even more complicated because the spectra of different species share many common features; hence, the difference in spectral fingerprints between different types of bacteria could be small. In order to differentiate and identify microorganisms based on their IR spectra, it is critical to find the features in the spectra that can best characterize one particular species and be able to isolate these features from other variations or perturbation within the same species and among different species.

In order to extract these features, we developed a mathematical procedure that is inspired by the Hybrid Linear Analysis (HLA) multivariate calibration algorithm developed by Berger et al. (1998). In HLA, a calibration vector can be created to calculate concentrations of a component present in any unknown mixture, given that an accurate spectrum of this component and the concentrations of this component in a set of calibration mixture samples are known. The key to this procedure was to create the net analyte signal of the spectrum of the component in question, which is orthogonal to the spectral contributions from all other constituents of the mixture. However, an accurate spectrum of a particular species is very difficult to obtain, and replicate measurements of the same species often yield different results. Another critical factor is that the concentration of any particular species in the mixture samples is not known. The focus of the proposed effort is to address the issue of identifying a target organism in a mixture by utilizing the HLA concept to extract specific features from the spectrum of a species in conjunction with discriminant analysis.

One basic assumption used in our procedure is that the average over a number of replications of spectral measurements of one species will retain the key information in the spectra that best characterizes the species in question. In this work, each experiment was replicated ten times, and the average was used as a reference spectrum. Each individual spectrum from the ten replications was compared and normalized against this reference to identify the common features.

Let the spectra of mixture m1 be represented in matrix form as M . Each column of M is a vector that represents one spectrum of one sample of m1 measured during experiments; in this study, M has ten columns to account for the ten FTIR measurements from each sample. A three-step procedure was adopted to examine the presence of an unknown bacterial species (b) in this mixture. Note that b represents a single spectrum of b ; the entire set from the ten replications is denoted B , and the mean spectrum of b was calculated as b_a .

Step 2 involves removal of features common to b_a and M from M , as proposed by Maquelin et al. (2000), and is given by:

$$S = M - \text{proj}_{b_a} M \quad (1)$$

Equation 1 implies that the projection of m1 on b_a is subtracted from M . For each column m_i in M , the projection of m_i on b_a results in:

$$\text{proj}_{b_a} m_i = (b_a \cdot m_i / \|b_a\|^2) b_a \quad (2)$$

Step 3 is based on a concept similar to HLA, where b_a serves as a reference or base spectrum of one component in HLA analysis. After step 2, S is orthogonal to b_a . However, S is not orthogonal to every b due to differences between b_a and b . Now, following the same operations adopted in HLA, we can create a vector that contains only the characteristics in b and that is orthogonal to other spectra in M .

In step 3, the results from step 2 are projected to obtain an orthogonal system. Spectra S from step 2 contain spectral features in M that are contributions from constituents (i.e., microorganism that make up mixture m1) in m1 other than b_a . Given that S contains enough entries to account for all of the constituents in m1, the resulting system S has the potential to describe most of the features of these constituents. Now a matrix of orthogonal basis spectra (V) comprised of the principal components of S is generated, so that all of the spectra contained in S can be modeled by a reduced set of components depicted in V . A correcting vector (c) can now be created as:

$$c = bV^T V \quad (3)$$

This vector represents the features in each b vector that are orthogonal to b_a and could be modeled by other components in M . These features are considered as non-common features to species b due to individual differences but are common features between b and other bacteria in the mixture. Therefore, in order to extract features in b that best characterize species b , vector c must be subtracted from b . In the next step, c was subtracted from each b to result in a residual spectrum (r) as given by:

$$r = b - c \quad (4)$$

Equation 4 represents the residual r , a portion of b that cannot be modeled by the spectra contained in V , i.e., it is orthogonal to V .

As a final step, r was normalized to obtain a characteristic spectrum of b with respect to the spectra of the mixture m1, and with all variations due to replications discounted:

$$b_f = r / (r \cdot b) \quad (5)$$

This procedure was performed for every b to obtain a set of spectra (BF) that was used in discriminant analysis for classification and differentiation. In an ideal case, where a specific microbial species always produces a consistent IR spectrum b_a , all of the individual spectra b_f in the matrix BF will be the same. However, since microbes are living cells and differ from each other in many ways, the best we can obtain for BF is a closely clustered group from a discriminant analysis procedure.

The matrix BF obtained in this manner represents the characteristic features in the spectrum of b with respect to mixture m1 and the common features from different measurements on the same batch of microorganism; hence, it has the potential to discount sample-specific variations. For a different mixture, the fingerprint for b needs to be generated independently.

The next step is to create a test data set to examine whether a specific pathogen in a fingerprint set BF is contained in mixture m1. First we project m1 onto the average of $b(b_a)$ to generate a matrix M_p . Each column of this matrix is the projection of one spectrum of m1 onto b_a , to give a vector m_p as given by:

$$m_p = (b_a \cdot m_i / \|b_a\|^2) b_a \quad (6)$$

Then each vector m_p is normalized against b to produce a test dataset M_p :

$$m_b = m_p / (m_p \cdot b) \quad (7)$$

If b is a constituent of m_1 , then M_b and BF should contain enough common information so that a discriminant analysis would show that M_b and B belong to the same group. On the other hand, if b is not an ingredient of m_1 , then it is quite likely that M_b and B will be different enough and separate into different groups during discriminant analysis. Thus, whether or not a particular species is present in a mixture can be determined.

The algorithm conceptualized using the projection and orthogonalization of the spectral data matrix was implemented in MATLAB (The Mathworks, Inc., Natick, Mass.).

DEVELOPMENT OF DISCRIMINANT MODEL AND TESTING OF SAMPLES

The data set BF created in the previous steps for a particular mixture was used to construct the discriminant model. First, PCA was conducted to reduce the dimension of the set BF to 10 (i.e., the first ten PCs were kept); then canonical variate analysis (CVA) was conducted on the reduced data set in which weighted linear combinations were structured to maximize the difference among the group means, relative to their variances. PAC and CVA were conducted using Win-DAS (1998 version, Wiley and Sons, Chichester, U.K.). The 1st and 2nd variates were plotted and analyzed by cluster analysis; each bacterium was represented by one cluster with an elliptic boundary indicating the 95% confidence interval. Win-DAS then created a discriminant model based on the CV scores; an unknown spectrum was tested against this model by calculating its 1st and 2nd CV scores and plotting it along the CV plane. If the data point fell inside the elliptic boundary, it was considered to be a member of the cluster. To further confirm the conclusions drawn by this visual investigation, a linear discriminant analysis (LDA) model based on Mahalanobis distance was constructed, and the unknown spectrum was tested using calculated Mahalanobis distances to each cluster; hence, its identity was evaluated.

Testing data were created for unknown samples by calculating the M_b set following the method reported in the previous section. Bacteria are complex living systems in terms of their chemical composition; therefore, although a 100% match of enough replicated measurements over one unknown sample is the ideal criterion for judging the identity of the unknown sample, in reality a 100% match is not possible for most cases. In this study, a match rate of 80% or higher was considered to be an adequate indicator to positively associate the unknown samples to a certain bacterial identity.

RESULTS AND DISCUSSION

The means of the ten mid-IR spectra of the five pathogens studied are shown in figure 1a, with tentative band assignments of major chemical groups corresponding to the functional groups that comprise the microorganisms. To differentiate between these species, principal component analysis (PCA) followed by canonical variate analysis (CVA)

Figure 1a. Mean spectra for the five pathogenic bacteria in the fingerprint regions (800-1800 cm^{-1} and 2700-3100 cm^{-1}) with tentative band assignments.

Figure 1b. Mean residual spectra for the five pathogenic bacteria in the fingerprint regions (800-1800 cm^{-1} and 2700-3100 cm^{-1}).

Figure 1c. Typical spectra (mixture m_1) in the fingerprint regions (800-1800 cm^{-1} and 2700-3100 cm^{-1}).

Figure 2. Differentiation of the five pathogenic bacteria based on canonical variate using: (a) the original spectra (B), and (b) the manipulated spectra (BF).

on the ten samples showed that these five species could be well differentiated (fig. 2a).

Application of chemometrics to differentiate the species is quite straightforward, as seen from figure 2a as well as from similar work on other bacteria reported in the literature (Helm et al., 1991; Ieven and Goossens, 1997; Manoharan et al., 1990; Maquelin et al., 2000; Maquelin et al., 2002; Naumann et al., 1991). The unique contribution of our study is to identify the presence (or absence) of a particular species within a mixture of bacteria. To achieve this goal, first we need to extract those features (b_f) in the spectra that can best characterize the species. Figure 1b shows the mean residual spectra of the five pathogenic bacteria. Comparison of the original spectra shows that the differences at several peaks corresponding to the different functional groups [$\nu(\text{C-O})$, $\nu(\text{PO}_2^-)$, $\delta(\text{C-H})$, and amide II] in these species were amplified. Figure 2b shows the differentiation achieved by using the residual spectra (b_f). The data representing different species formed closer clusters, and the distances between these groups were also greater, clearly demonstrating that the residual spectra had a better differentiating capability.

In the next step, a test set (M_b) was created from the mixture spectra to test the presence or absence of each possible constituent in each mixture. Figure 1c shows typical spectra obtained for mixture m1 (table 1 gives the microorganisms present in the different mixtures). Based on these spectra and the spectra of the five species (fig. 1a), it was impossible to determine the presence or absence of a particular species in the mixture visually. In figure 3, the residual spectra of *Enterococcus faecium* and *Shigella boydii* calculated against m1 are shown and compared to the mean spectrum for these two species. It can be seen that the residual spectra show the same characteristics as their respective mean spectra. Furthermore, in figure 3, the test spectra (M_b) calculated for m1 against these two species are also shown. Mixture m1 contains *Enterococcus faecium* but not *Shigella boydii*. From visual examination of the test spectra (M_b) and the residual spectra (BF), the test set for *Enterococcus faecium* shows close similarity to its residual spectra (fig. 3), while the test set for *Shigella boydii* shows a marked difference in its residual spectra (fig. 4). Hence, we conclude that the presence of *Enterococcus faecium* in m1 is highly likely and that *Shigella boydii* is not. This conclusion was

Figure 3. Comparison between the residual spectra for *Enterococcus faecalis* calculated for m1 and testing spectra for *Enterococcus faecalis* of m1: (a) residual spectra, and (b) testing spectra).

strongly supported by chemometric analysis of the mixture containing organisms 1, 2, and 3 (*Enterococcus faecium*, *Bacillus cereus*, and *Salmonella enteritidis*) shown in figure 5.

Figures 5 through 7 show the typical analysis results for three mixtures. The analysis of m1 depicted in figure 5 clearly demonstrates that three species were present in the mixture and the other two species tested for were not present, to give a prediction accuracy of 100%. Five mixtures (m1, m3, m4, m6, and m9) show similar results. In figure 6 (mixture m10), two species can be positively identified as present, and two species can be identified as absent. The test

data set for the fifth species still showed quite significant differences compared to the absent species, which resulted in a good prediction. Similar results were obtained for mixtures m2, m5, m8, and m10. In figure 7 (mixture m7), a false negative was predicted. An ingredient of the mixture (species 2) was predicted to be not present. In all of the mixtures analyzed, no false positive was predicted, and no undetermined case was predicted. The predictions for all ten mixtures are given in table 2. Analysis of nine of the ten mixtures resulted in a 100% accurate prediction of the composition of the mixtures; in one case (m7), a false negative was predicted, giving a prediction accuracy of 80%

Figure 4. Comparison between the residual spectra for *Shigella boydii* calculated for m1 and testing spectra for *Shigella boydii* of m1: (a) residual spectra, and (b) testing spectra).

Figure 5. Discriminant testing results for microbial mixture m1 indicating organisms 1, 2, and 3 present and organisms 4 and 5 not present (100% accurate).

for that mixture. No false positive was observed for all ten mixtures.

To further confirm these results acquired through visual observations of the CV plot, an LDA model using Mahalanobis distance as the differentiating criteria was also implemented against the test set. For each entry in the test set, a Mahalanobis distance to each of the five clusters was calculated. An average of the Mahalanobis distances of the ten entries of the whole set was compared, and the identity of

the sample was determined. When a bacterium was present in the mixture, the average Mahalanobis distance between the test set and the target cluster (representing the bacterium tested for) was smaller by a magnitude of 1 to 4 than the distances between the test set and other clusters. On the other hand, if the bacterium was absent in the mixture, the Mahalanobis distances between test set and all clusters were at the same magnitude. The average Mahalanobis distances for mixtures m1, m7, and m10 are listed in tables 3 through

Figure 6. Discriminant testing results for microbial mixture m10 indicating organisms 3, 4, and 5 present, and organisms 1 and 2 not present (100% accurate).

Figure 7. Discriminant testing results for microbial mixture m7 indicating organisms 3 and 4 present, and organisms 1, 2, and 5 not present (80% accurate).

Table 2. Accuracy of prediction of composition of mixtures.

	Positive (present)	Negative (absent)	False Positive	False Negative	Prediction Accuracy
M1	3	2	0	0	100%
M2	3	2	0	0	100%
M3	3	2	0	0	100%
M4	3	2	0	0	100%
M5	3	2	0	0	100%
M6	3	2	0	0	100%
M7	2	3	0	1	80%
M8	3	2	0	0	100%
M9	3	2	0	0	100%
M10	3	2	0	0	100%

5, corresponding to figures 5 through 7, respectively. It is clear that visual observation of the CV plot provided

predictions consistent with the LDA predictions through Mahalanobis distance calculation.

These results demonstrate that the mathematical procedure developed worked extremely well. A database of single-species spectra can be established. If compositional information of a mixture sample is required, the spectra of the mixture can be measured and subjected to the same analysis, and the presence of the species in the database can be tested.

Only one incubation operation is needed to prepare the samples for investigation, and the whole process can be completed within 24 h from the collection of bacteria samples. This is a significant improvement compared to traditional approaches.

To make the most use of this method, accurate spectra for single bacterial species are required. In addition, more spectra will help to improve the accuracy and statistical

Table 3. Prediction of identity based on average Mahalanobis distances of entries in test set to each cluster for mixture m1 using LDA model.

Test Data Set	Average Distance from Group 1	Average Distance from Group 2	Average Distance from Group 3	Average Distance from Group 4	Average Distance from Group 5	Prediction
Test for 1	5.7645e-1	6.8807e1	3.4671e1	3.5883e1	8.3440e1	Present
Test for 2	6.6454e1	2.2346e-1	3.3812e1	3.1293e1	1.4865e1	Present
Test for 3	3.1968e1	3.6262e1	1.0476e-1	2.1677e1	4.6510e1	Present
Test for 4	4.9639e1	1.8595e1	1.5541e1	1.4421e1	3.3136e1	Absent
Test for 5	2.0217e1	4.8013e1	1.3877e1	1.4998e1	6.2555e1	Absent

Table 4. Prediction of identity based on average Mahalanobis distances of entries in test set to each cluster for mixture m10 using LDA model.

Test Data Set	Average Distance from Group 1	Average Distance from Group 2	Average Distance from Group 3	Average Distance from Group 4	Average Distance from Group 5	Prediction
Test for 1	5.2082e2	1.6486e2	1.9481e2	5.0555e2	1.2576e2	Absent
Test for 2	1.9382e2	2.4465e2	1.9481e2	1.8837e2	2.5258e2	Absent
Test for 3	3.8009e2	6.0161e2	4.2738e-2	1.5032e2	1.4690e2	Present
Test for 4	1.4255e2	6.5201e2	7.9728e1	1.3940e0	4.2935e2	Present ^[a]
Test for 5	8.5268e2	5.6732e2	1.4954e2	5.8499e2	2.4343e-2	Present

^[a] Test for group 4 yields a distance that is still significantly smaller than others; hence, a positive prediction can be made with confidence.

Table 5. Prediction of identity based on average Mahalanobis distances of entries in test set to each cluster for mixture m7 using LDA model.

Test Data Set	Average Distance from Group 1	Average Distance from Group 2	Average Distance from Group 3	Average Distance from Group 4	Average Distance from Group 5	Prediction
Test for 1	3.8036e2	1.1102e2	1.2724e2	3.3260e2	4.1989e2	Absent
Test for 2	4.2089e2	2.9653e1	6.1781e1	3.1606e2	6.3764e2	Absent ^[a]
Test for 3	7.8597e2	1.4559e2	1.5140e0	1.0986e2	9.5455e2	Present
Test for 4	1.3505e2	4.8096e2	1.0690e2	5.7355e-1	1.3409e2	Present
Test for 5	5.9939e2	2.1259e2	1.0817e2	2.0035e2	5.4236e2	Absent

^[a] Test for group 2 yields a distance that is not significantly smaller than the others; hence, a false negative prediction was made.

reliability of the prediction and decrease possible errors. Thus, more work needs to be done to establish the database of important pathogenic microorganisms.

CONCLUSION

A novel mathematical procedure was developed to analyze mid-infrared spectral data to determine the presence or absence of a specific microorganism in mixed cultures of microorganisms. Ten microbial mixtures, each consisting of at least three pathogenic bacteria, were prepared and analyzed by FTIR spectroscopy. Of the mixtures analyzed, the prediction was 100% accurate in nine mixtures and 80% accurate in the tenth mixture. Compared to traditional microbiological approaches, the method developed will allow foodborne pathogens (a mixture of microorganisms) to be identified and characterized within 1 h after an 18 to 24 h incubation time. With respect to other reported spectroscopic identification procedures, the method developed has the unique feature of working on mixture samples directly without going through the time-consuming separation process. Once the database for key pathogens is established, the procedure could be used for rapid diagnosis of infectious pathogens.

REFERENCES

- Barshick, S. A., D. A. Wolf, and A. A. Vass. 1999. Differentiation of microorganisms based on Pyrolysis-ion trap mass spectrometry using chemical ionization *Analytical Chem.* 71(3): 633-641.
- Berger, A. J., T.-W. Koo, I. Itzkan, and M. S. Feid. 1998. An enhanced algorithm for linear multivariate calibration. *Analytical Chem.* 70(3): 623-627.
- Bouhedja, W., G. D. Sockalingum, P. Pina, P. Allouch, C. Bloy, R. Labia, J. M. Millot, and M. Manfait. 1997. ATR-FTIR spectroscopic investigation of *E. coli* transconjugants β -lactams-resistance phenotype. *FEBS Lett.* 412(1): 39-42.
- Busse, H. J., E. B. Denner, and W. Lubitz. 1996. Classification and identification of bacteria: Current approaches to an old problem. Overview of methods used in bacterial systematics. *J. Biotech.* 47(1): 3-38.
- Chadha, S., R. Manoharan, P. Moëne-Loccoz, W. H. Nelson, W. L. Peticolas, and J. F. Sperry. 1993. Comparison of the UV resonance Raman spectra of bacteria, bacterial cell walls, and ribosomes excited in the deep UV. *Applied Spectrosc.* 47(1): 38-43.
- Clark, N. C., O. Olsvik, J. M. Swenson, C. A. Spiegel, and F. C. Tenover. 1999. Detection of a streptomycin/spectinomycin adenyltransferase gene (*aadA*) in *Enterococcus faecalis*. *Antimicrobial Agents and Chemotherapy* 43(1): 157-160.
- Demirev, P. A., Y. P. Ho, V. Ryzhov, and C. Fenselau. 1999. Microorganism identification by mass spectrometry and protein database searches. *Analytical Chem.* 71(14): 2732-2738.
- Edwards, H. G. M., N. C. Russell, R. Weinstein, and D. D. Wynn-Williams. 1995. Fourier transform Raman spectroscopic study of fungi. *J. Raman Spectrosc.* 26(8-9): 911-916.
- Fredricks, D. N., and D. A. Relman. 1998. Improved amplification of microbial DNA from blood cultures by removal of the PCR inhibitor sodium polyanethanesulfonate. *J. Clin. Microbiol.* 36(10): 2810-2816.
- Ghiamati, E., R. Manoharan, W. H. Nelson, and J. F. Sperry. 1992. UV resonance Raman spectra of *Bacillus* spores. *Applied Spectrosc.* 46(2): 357-364.
- Goodacre, R., S. J. Hiom, S. L. Cheeseman, D. Murdoch, A. J. Weightman, and W. G. Wade. 1996. Identification and discrimination of oral asaccharolytic *Eubacterium* spp. by pyrolysis mass spectrometry and artificial neural networks. *Current Microbiol.* 32(2): 77-84.
- Helm, D., H. Labischinski, G. Schallehn, and D. Naumann. 1991. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *J. Gen. Microbiol.* 137(1): 69-79.
- Ieven, M., and H. Goossens. 1997. Relevance of nucleic acid amplification techniques for diagnosis of respiratory tract infections in the clinical laboratory. *Clin. Microbiol. Rev.* 10(2): 242-256.
- Manoharan, R., E. Ghiamati, R. A. Dalterio, K. A. Britton, W. H. Nelson, and J. F. Sperry. 1990. UV resonance Raman spectra of bacteria, bacterial spores, protoplasts, and calcium dipicolinate. *J. Microbiol. Methods* 11(1): 1-15.
- Maquelin, K., L. Choo-Smith, T. van Vreeswijk, H. P. Endtz, B. Smith, R. Bennett, H. A. Bruining, and G. J. Puppels. 2000. Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. *Analytical Chem.* 72(1): 12-19.
- Maquelin, K., C. Kirschner, L. Choo-Smith, N. van den Braak, H. P. Endtz, D. Naumann, and G. J. Puppels. 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. *J. Microbiol. Methods* 51(3): 255-271.
- Naumann, D., D. Helm, and H. Labischinski. 1991. Microbiological characterization by FT-IR spectroscopy. *Nature* 351(6321): 81-82.
- Nikkari, S., and D. A. Relman. 1999. Molecular approaches for identification of infectious agents in Wegener's granulomatosis and other vasculitides. *Current Opinion in Rheumatol.* 11(1): 11-16.
- Noordhoek, G. T., J. D. van Embden, and A. H. Kolk. 1996. Reliability of nucleic acid amplification for detection of *Mycobacterium tuberculosis*: An international collaborative quality control study among 30 laboratories. *J. Clin. Microbiol.* 34(10): 2522-2525.
- Sockalingum, G. D., W. Bouhedja, P. Pina, P. Allouch, C. Mandray, R. Labia, J. M. Millot, and M. Manfait. 1997. ATR-FTIR spectroscopic investigation of imipenem-susceptible and -resistant *Pseudomonas aeruginosa* isogenic strains. *Biochem. Biophys. Res. Comm.* 232(2): 240-246.
- Tang, Y. W., G. W. Procop, and D. H. Persing. 1997. Molecular diagnostics of infectious diseases. *Clin. Chem.* 43(11): 2021-2038.
- Tenover, F. C., M. B. Huang, J. K. Rasheed, and D. H. Persing. 1994. Development of PCR assays to detect ampicillin resistance genes in cerebrospinal fluid samples containing *Haemophilus influenzae*. *J. Clin. Microbiol.* 32(11): 2729-2737.

- Timmins, E. M., S. A. Howell, B. K. Alsberg, W. C. Noble, and R. Goodacre. 1998. Rapid differentiation of closely related *Candida* species and strains by pyrolysis-mass spectrometry and Fourier transform-infrared spectroscopy. *J. Clin. Microbiol.* 36(2): 367-374.
- Vanechoutte, M., and J. Van Eldere. 1997. The possibilities and limitations of nucleic acid amplification technology in diagnostic microbiology. *J. Med. Microbiol.* 46(3): 188-194.
- Williams, A. C., and H. G. M. Edwards. 1994. Fourier transform Raman spectroscopy of bacterial cell walls. *J. Raman Spectrosc.* 25(7-8): 673-677.
- Zeroual, W., M. Manfait, and C. Choisy. 1995. FT-IR spectroscopy study of perturbations induced by antibiotics on bacteria (*Escherichia coli*). *Pathol. Biol.* 43(4): 300-305.

NOMENCLATURE

- b = represents a bacterium
- b = vector that represents a spectrum measured for a bacterium

- b_a = vector that represents the mean spectrum of a bacterium
- b_f = vector that represents the normalized residual spectrum for a bacterium
- B = matrix that contains all of the spectra measured for a bacterium
- BF = matrix of all normalized residual spectra for a bacterium
- c = correcting vector
- M = matrix that contains all spectra measured for a bacterial mixture
- m_i = vector that represents a spectrum measured for a bacterial mixture
- M_p = matrix of test data set
- m_p = vector that represents a spectrum in the test data set
- r = residual spectrum for a bacterium
- S = matrix that contains components in M orthogonal to b_a
- V = PCA basis of S