Detection and Identification of Bacteria in a Juice Matrix with Fourier Transform–Near Infrared Spectroscopy and Multivariate Analysis

L. E. RODRIGUEZ-SAONA,1,2† F. M. KHAMBATY,2 F. S. FRY,2 J. DUBOIS,1 AND E. M. CALVEY2*

1Joint Institute for Food Safety and Applied Nutrition (JIFSAN)—University of Maryland, College Park, Maryland 20742; and 2Food and Drug Administration, Center for Food Safety and Applied Nutrition, 5100 Paint Branch Parkway, College Park, Maryland 20740, USA

MS 04-123: Received 30 March 2004/Accepted 9 June 2004

ABSTRACT

The use of Fourier transform–near infrared (FT-NIR) spectroscopy combined with multivariate pattern recognition techniques was evaluated to address the need for a fast and sensitive method for the detection of bacterial contamination in liquids. The complex cellular composition of bacteria produces FT-NIR vibrational transitions (overtone and combination bands), forming the basis for identification and subtyping. A database including strains of Escherichia coli, Pseudomonas aeruginosa, Bacillus subtilis, Bacillus cereus, and Bacillus thuringiensis was built, with special care taken to optimize sample preparation. The bacterial cells were treated with 70% (vol/vol) ethanol to enhance safe handling of pathogenic strains and then concentrated on an aluminum oxide membrane to obtain a thin bacterial film. This simple membrane filtration procedure generated reproducible FT-NIR spectra that allowed for the rapid discrimination among closely related strains. Principal component analysis and soft independent modeling of class analogy of transformed spectra in the region 5,100 to 4,400 cm⁻¹ were able to discriminate between bacterial species. Spectroscopic analysis of apple juices inoculated with different strains of E. coli at approximately 10⁵ CFU/ml showed that FT-NIR spectral features are consistent with bacterial contamination and soft independent modeling of class analogy correctly predicted the identity of the contaminant as strains of E. coli. FT-NIR in conjunction with multivariate techniques can be used for the rapid and accurate evaluation of potential bacterial contamination in liquids with minimal sample manipulation, and hence limited exposure of the laboratory worker to the agents.

Near infrared (NIR) spectroscopy allows fast, accurate, and nondestructive measurements of chemical components and can provide information about structural and physical properties of materials (15, 16). It is used extensively in the agricultural industry to assess product quality via measurements of protein and fat content, for example, and in the pharmaceutical, biomedical, textiles, polymer, and petrochemical fields (for reviews, see 4, 5, 25). This technique shows significant potential for monitoring food tampering, including assessing the presence of chemical and microbial threat agents, because of the short scanning time and chemical specificity.

Fourier transform–infrared (FT-IR) spectroscopic techniques (8–10, 18) have been used for the identification and classification of different bacterial taxa, occasionally down to the strain level. FT-IR spectroscopy provides fairly narrow bands arising from group vibrations with known assignment in most cases. As a result, the FT-IR spectra of bacteria have shown highly specific patterns that might be unique for individual strains and could be used as strain-specific fingerprints (1, 6–9, 12, 17, 18, 23, 24). NIR spectra are the result of relatively weak and broad overtone and combination bands of fundamental vibrational transitions associated mainly with C-H, N-H, and O-H functional groups (11). Consequently, the spectra of complex products are a result of overlapping contributions and often require multivariate mathematical treatment to extract the information. Although NIR spectra typically exhibit broad bands resulting from overtone vibrations or combinations, we previously demonstrated the use of the spectral range attributed to those bands in conjunction with multivariate analysis for the classification of different bacterial species and strains (20–22). In this work, we used information obtained with an interferometer by FT-NIR spectroscopy to benefit from improved spectral reproducibility and wave number precision compared with results from dispersion instruments. In addition, we developed a filtration procedure using aluminum oxide membranes to produce dry bacterial films that allowed for increased sensitivity and rapid discrimination among closely related strains.

The objectives of this study were twofold: first, to develop a small database and reproducibly identify and classify bacterial strains through the use of FT-NIR spectroscopy and multivariate statistical methods, and second, to determine whether the database could be used to identify bacteria in a spiked fruit juice sample. The second objective was meant to investigate the effect of a food matrix on the predictability of a model developed without any contribution from a food product. We previously described a calibration model developed independently of the juice matrix for predicting sugar ratios in fruit juices (19). This work

* Author for correspondence. Tel: 301-436-1981; Fax: 301-436-2641; E-mail: elizabeth.calvey@cfsan.fda.gov.
† Present address: Food Science and Technology Department, The Ohio State University, 110 Parker Food Science and Technology Building, 2015 Fyffe Road, Columbus, OH 43210, USA.
assesses the applicability of FT-NIR to solve food microbiology problems.

**MATERIALS AND METHODS**

**Strains and growth conditions.** The following bacteria were examined: three *Escherichia coli* (HB101, ATCC 43888, and a nonpathogenic wild-type food isolate #1254), *Pseudomonas aeruginosa*, *Bacillus cereus*, *Bacillus thuringiensis*, and a *Bacillus subtilis*. All strains were either ATCC strains or part of the U.S. Food and Drug Administration’s microbial collection.

All strains were cultivated aerobically on brain heart infusion agar at 37°C for 16 h. From each plate, an isolated single colony was carefully collected with a sterile plastic loop and suspended in 25 ml of brain heart infusion broth in a screw-capped Erlenmeyer flask (250 ml). The bacteria were grown at an incubation temperature of 37°C with shaking at 110 rpm. The bacterial strains were grown for up to 33 h, and aliquots were taken at different time points to establish the growth phase providing the most reproducible NIR spectra. The optical density (absorbance at 610 nm) was measured to monitor the growth phase of the cultures.

**Sample preparation.** An aliquot (~1 ml) of the cell suspension was transferred to a microcentrifuge tube. From each bacterial suspension, two tubes were prepared to evaluate same day reproducibility. The reproducibility of cultures grown on different days was also examined. The tubes were centrifuged at 10,000 rpm for 2 min, the supernatant was removed, and the resultant wet pellet was weighed and resuspended in 0.5 ml of sterile saline (0.9% NaCl) solution. For experiments evaluating the effects of the saline solution, no further transfers were made. Aliquots containing 2, 1, or 0.5 mg of bacterial suspension were transferred to microcentrifuge tubes, and ethanol was added to obtain a final 70% (vol/vol) alcohol level. The suspension was filtered through an aluminum oxide Anodisc membrane (Whatman International, Maidstone, UK) with a 0.2-µm pore size to harvest the bacterial cells and eliminate the solution. The bacterial film formed was approximately 8 mm in diameter to match the optical window of the FT-NIR instrument used. The membrane provided a smooth and flat surface for the retention of the bacterial cells, allowing the formation of a bacterial film, and did not contribute significantly to the absorbance spectra. To avoid substantial changes in the spectra because of the ambient relative humidity, the membranes were placed in desiccators containing silica gel or other hygroscopic material and kept under vacuum for a minimum of 5 min before analysis.

**FT-NIR measurements.** All FT-NIR spectra were recorded with a Perkin Elmer (Shelton, Conn.) Spectrum 100 spectrometer operating at 8 cm⁻¹ resolution. Measurements were made with the use of a diffuse reflectance accessory incorporating an integrating sphere and a lead sulfide (PbS) detector. The Anodisc membrane, supporting the dry bacterial film, was placed on the reflectance accessory for direct measurement by reflection with a Teflon-based material (Spectralon, Perkin Elmer) to direct the beam back to the detector. The FT-NIR spectra were recorded from 10,000 to 4,000 cm⁻¹ at intervals of 4 cm⁻¹. Interferograms (128) were co-added, followed by strong Beer-Norton apodization. The absorbance spectrum was computed as a ratio of the single-beam spectrum of the sample to that of the supporting Anodisc membrane. Before calibration, the FT-NIR reflectance data were mean centered and baseline corrected.

**Multivariate analyses.** Principal component analysis (PCA) and soft independent modeling of class analogy (SIMCA) (26) were carried out with the use of Pirouette pattern recognition software (version 3.1 for Windows, Infometrix Inc., Woodinville, Wash.). The spectra were imported as JCAMP-DX files into the Pirouette multivariate analysis program. The FT-NIR spectra were area normalized and transformed to their second derivatives on the basis of a Savitzky-Golay polynomial filter (15-point window). PCA was performed with the NIPALS (nonlinear iterative partial least squares) algorithm. Outlier diagnostics were done by sample residual and Mahalanobis distance. The data comprised repeated observations on cells collected from a single log stationary-phase culture (two measurements) and from cultures grown on subsequent days (replicates). The scores plot (a projection of the original data onto the principal component axes) allowed the visualization of clustering among samples (sample patterns, groupings, or outliers).

The FT-NIR data are arranged as a matrix comprising samples (rows) and spectral frequencies (columns). In PCA, a new set of orthogonal axes (principal components or PCs) is computed such that the first PC is oriented in the direction of the maximum variation within the data, the second PC is oriented orthogonally to PC 1 in the direction of the maximum remaining variation in the data, and so on. The data are projected onto these new axes to observe any clustering of samples within the data. The data for all samples are used to infer differentiation among samples. Samples not used in creating the model can be projected onto the PC axes to determine similarity to those samples used for the model. In SIMCA, a PCA model is created for each distinct class (i.e., species or strain) of the known samples. The data from independent (unknown) samples (or from samples in other classes) can be projected onto the PCs for each class to observe clustering. Additionally, probabilistically determined boundaries based on Mahalanobis distances for each class can be computed, and membership in one (or none of the class models can be predicted.

**Spiked apple juices with bacterial cells.** A commercial apple juice (2 liters) was filtered under vacuum through a Nalgene (Rochester, N.Y.) cellulose nitrate filter (0.2-µm pore size) to sterilize the juice samples. Four juice samples (200 ml each) were inoculated with approximately 2-mg pellets of stationary-phase *E. coli* strains (HB101 and 1254, duplicates for each strain) suspended in 0.9% saline solution, resulting in approximately 10⁴ to 10⁵ cells per ml of juice. A control inoculated with saline solution only was included. Modifications of the method used to develop the calibration were required to concentrate the bacteria from the apple juice matrix. These modifications included filtration of the homogenized apple juice samples through a 45-mm-diameter Durapore (Millipore, Billerica, Mass.) polycarbonate membrane filter (0.22-µm pore size) under vacuum for the rapid and efficient collection of the bacterial cells from a large juice volume. The membrane was then placed in a screw-capped glass tube, and 2 ml of 70% (vol/vol) ethanol was added to isolate the cells from the membrane. The liquid and the suspended cells were transferred to a centrifuge tube. After centrifugation at 12,000 rpm for 2 min, the resulting pellet was resuspended in 50 µl of saline solution and applied to an Anodisc membrane for FT-NIR analysis. Duplicate samples were prepared.

Cell enumeration was performed by the standard plate count method. Briefly, an aliquot (0.1 ml) of the inoculated apple juice samples was diluted in 9.9 ml of brain heart infusion broth (10⁻² dilution). A serial dilution was prepared (10⁻⁻ dilution). Aliquots (0.1 ml) were plated (in duplicate) on brain heart infusion agar plates, and colonies were counted after incubating the plates at 35°C for 24 h.
RESULTS AND DISCUSSION

We previously described the use of a combination of FT-NIR and a filtration system to concentrate bacterial cells and eliminate the effect of the matrix (20–22) for the rapid and reproducible identification of bacteria. The FT-NIR spectra of bacterial films reveal important chemical information with little interference from water absorption peaks. Nevertheless, differences in the spectra between bacteria are subtle, and mathematical processing is needed to extract the relevant information for qualitative and quantitative analysis (3). The spectral range 5,200 to 4,000 cm\(^{-1}\) was found to be optimal for identification following computation of the second derivative, which allowed removal of baseline variations and resolution of overlapping peaks.

The sample preparation step involved the suspension of aliquots (~1 mg) of bacteria (grown to stationary phase) in sterile saline (0.9% NaCl) or ethanol (final 70% [vol/vol] concentration) to inactivate the microorganisms, thereby addressing safety concerns raised when evaluating pathogenic or potentially pathogenic strains. Alcohol treatment obviously affects the spectral profile of the bacteria by dissolving some lipids and possibly changing the structure of proteins; as a result, the discrimination of the different strains by PCA also shows an effect from the ethanol treatment (Fig. 1). As expected, the largest changes in the spectra were noticed in the ranges of 4,700 to 4,330 cm\(^{-1}\) and 4,676 to 4,611 cm\(^{-1}\), affecting bands associated with amide \(1\), whereas the band at 4,676 cm\(^{-1}\) matched the frequency of HC=O stretching modes, carbonyl overtone (4,920 cm\(^{-1}\) C=O), and amino acid combination bands (4,688 and 4,468 cm\(^{-1}\)), whereas the band at 4,676 cm\(^{-1}\) matched the frequency of HC=CH bond vibrations (4). The spectral range between 4,400 and 4,000 cm\(^{-1}\) is associated with C-H overtones and combination bands of lipids and carbohydrates; the increased variability in this spectral range could be indicative of changes in the bacte-
FT-NIR measurements of the bacterial films on membranes (Fig. 3) collected from apple juices (200 ml) spiked with 2 mg of bacterial mass showed that the filtration process was able to recover the bacteria because the spectra exhibited band features similar to those obtained for the *E. coli* bacteria. The control samples showed FT-NIR absorption signals with peaks at the 4,400 to 4,100 cm\(^{-1}\) range related to aliphatic C-H groups. The bacterial levels in the 200-ml apple juice samples ranged between \(10^4\) and \(10^5\) CFU/ml. Normal aerobic microflora in fresh apple cider or juice from tree-harvested, unsorted fruit ranges from 89 to \(2.3 \times 10^4\) CFU/ml depending on apple variety (13). Therefore, the model developed for this analysis would require that 1 to 2 liters of juice be filtered if the bacterial load was \(10^3\) CFU/ml.

The SIMCA model was used to classify the samples. The test samples spiked with *E. coli* HB101 were predicted as *E. coli* HB101 and *E. coli* ATCC43888, whereas the samples spiked with *E. coli* 1254 were predicted as *E. coli* HB101 and *E. coli* 1254 (Fig. 4). These results demonstrate that simple models developed with FT-NIR spectra and multivariate analysis (SIMCA) can correctly predict the contamination of juices with *E. coli* species but might not be able to determine the strain. Nevertheless, the FT-NIR method described here predicted the presence of *E. coli* in spiked juice samples. The overall sensitivity of the method depends on the inherent absorbance bands of the chemical moieties being measured, the optics and detector of the FT-NIR spectrophotometer, and the degree of difference among the species and strains involved. Having now demonstrated the ability to detect bacteria in a food matrix with the use of a classification model developed independently of this matrix, future work is needed to expand the database of spectra to include various food bacterial contaminants and natural flora. With the proper classification models, FT-NIR offers an additional quality control tool for the rapid and
safe detection of bacterial contamination for the juice industry.

The use of FT-NIR spectral information and a multivariate classification technique shows great potential for the identification and subtyping of different bacterial species. When building a classification model with cells harvested in the stationary growth phase and treated with ethanol to increase the safety of the process, we were able to use a simple PCA or SIMCA to reproducibly group the samples into different clusters. Analysis of bacteria extracted from apple juice samples inoculated with E. coli and filtered through an aluminum oxide membrane demonstrates the ability of FT-NIR spectroscopy to predict bacterial contamination on the basis of a SIMCA multivariate model. This methodology can be applied for the monitoring of bacterial contamination in liquid systems and could become a powerful tool for monitoring the safety of our food supply.

ACKNOWLEDGMENTS

This project was funded by the Technical Support Working Group through the Department of the Army (BFRC no. DAAD05-98-R0548). We express our appreciation to the Technical Support Working Group for sponsoring this research effort.

REFERENCES