Journal of Fluorescence, Vol. 13, No. 6, November 2003 (©2003)

Rapid Identification of Bacterial Species by Fluorescence Spectroscopy and Classification Through Principal Components Analysis

Héctor Enrique Giana,[1,3] Landulfo Silveira Jr.,[2] Renato Amaro Zângaro,[2] and Marcos Tadeu T. Pacheco[2]

Received June 30, 2003; revised October 1, 2003; accepted October 2, 2003

This work presents the development of a method for rapid bacterial identification based on the autofluorescence spectrum. It was demonstrated differences in the autofluorescence spectrum in three bacterial species and the subsequent separation, through the Principal Components Analysis (PCA) technique, in groups with high likeness, that could identify the bacteria in less than 10 min. Fluorescence spectra of 60 samples of 3 different bacterial species (*Escherichia coli*, EC, *Enterococcus faecalis*, EF and *Staphylococcus aureus*, SA), previously identified by automated equipment Mini API, were collected in 10 excitation wavelengths from 330 to 510 nm. The PCA technique applied to the fluorescence spectra showed that bacteria species could be identified with sensitivity and specificity higher than 90% according to differences that occur within the spectra with excitation of 410 nm and 430 nm. This work presented a method of bacterial identification of three more frequent and more clinically significant species based on the autofluorescence spectra in the excitation wavelengths of 410 and 430 nm and the classification of the spectra in three groups using PCA. The results demonstrated that the bacterial identification is very efficient with such methodology. The proposed method is rapid, ease to perform and low cost compared to standard methods.

KEY WORDS: Bacterial identification; diagnosis; fluorescence spectroscopy; principal components analysis (PCA).

INTRODUCTION

In bacteriology, staining methods are used to identify generally two groups of bacteria, Gram positive and negative, without identify the species. Chromogenic media are capable to isolate and identify some of the microorganisms involved in human pathology, but it cannot identify all the possible species. Currently it is possible to identify around 20,000 different bacterial species with chemical methods. However, the great difficulty that still exists is the time of bacterial identification, which, for standard chemical methods using automated equipments, is between 18 and 24 hr. This long time could compromise the chemotherapic treatment, with the subsequent worsening of the patient's infection status due to eventual emergence of species resistant to the previously administered chemotherapic agents [1].

Some studies have proposed the use of fluorescence spectra for a rapid bacterial identification. Nelson [2] proposed that the use of multi-excitation fluorescence spectroscopy would be of great usefulness in the microbiology, allowing the selection of the best excitation wavelength and consequently the selective excitation of biological molecular groups, for best bacteria species identification.

[1] Laboratório Oswaldo Cruz, Rua Santa Clara, 393, V. Adyanna, CEP 12243-630, São José dos Campos, SP, Brazil.
[2] Instituto de Pesquisa e Desenvolvimento, Universidade do Vale do Paraíba, Av. Shishima Hifumi, 2911, Urbanova, CEP 12244-000, São José dos Campos, SP, Brazil.
[3] To whom correspondence should be addressed. E-mail: laboratorio@oswaldocruz.com

1053-0509/03/1100-0489/0

Roselle *et al*. [3], when investigating the differences in the spectra of viable bacteria, comparing culturable and unculturable bacteria (which do not grow in culture medium), described changes in the autofluorescence spectrum of *E. coli* after 210 days kept in saline solution. Ivnitski *et al*. [4] reviewed different techniques used for direct and indirect identification of bacteria, such as fluorescence and infrared spectroscopy, flow cytometry, chromatography and chemiluminescence, for construction of bacterial biosensors. Hill *et al*. [5] described a method for measuring the laser-induced fluorescence of biological micro-particles in aerosols constituted by organic and inorganic compounds, bacteria, smoke, allergens and others, evidencing the possible spectral differentiation of biological and non-biological aerosols generated in the laboratory atmosphere. Cheng *et al*. [6] described the implementation of an apparatus for the investigation of bio-aerosols using laser-induced fluorescence spectroscopy in the ultraviolet. It was verified that the fluorescence spectra of four bacteria, *E. coli*, *S. aureus*, *B. subtilis var niger* and *B. thuringiensis*, obtained with such device, was very similar. Spector *et al*. [7] described a method for bacterial identification in an animal model (chinchilla) of acute otitis, using fluorescence spectroscopy and fiber optic cable. *S. aureus*, *S. pneumoniae* or *H. influenzae* were inoculated in the medium ear of the animals. The type of bacteria responsible for the infection was identified with 100% of success.

Very recently vibrational spectroscopy (Raman and infrared) have been used for rapid detection and identification of bacteria directly in culture plate [8], bacterial species in liquid suspension [9] and bacterial contamination in liquids [10].

The objective of this study is to develop an algorithm for rapid bacterial identification using the autofluorescence spectra applied to three different bacterial species, previously identified using an equipment calibrated with ATCC (American Type Culture Collection) standards. The algorithm is based on the determination of the typical autofluorescence pattern of each species, in a set of excitation wavelengths (between 330 and 510 nm). An identification model based on the spectral information obtained via Principal Components Analysis (PCA) method is proposed for fast diagnosis in these three types of human bacteria. The model is then tested prospectively in a new set of spectra from the same species, in order to verify the reliability and the robustness of the model.

MATERIALS AND METHODS

For the accomplishment of the study the bacterial species were chosen following the criteria:

(a) obtaining of the statistical frequency of most common bacterial infections in human biological fluids (furnished by Laboratório Oswaldo Cruz, São José dos Campos, SP, Brazil);
(b) identification and choice of the most frequent and more clinical importance species.

Based on the frequency of 2140 bacterial cultures, it was chosen the species of *Escherichia coli* (EC), with 35.7%, *Enterococcus faecalis* (EF), with 9.6% and *Staphylococcus aureus* (SA), with 3.0% of incidence. The last one (SA), although of small incidence, was chosen in function of the more clinical importance.

Bacteria were collected in different dates and from several human biological materials, such as urine, feces and secretions, and the bacterial species were obtained from cultures in standard (i.e., Blood Agar) and special (i.e., CPS ID2, BioMérieux) culture media, according to microbiology standardized techniques [11].

The bacterial species were partially identified according to the growth in special culture medium (CPS ID2, BioMérieux), which allows to separate the *Enterobacter*, as *E. coli* species, from the Gram positive bacteria, as *E. faecalis* and *S. aureus*, the last one having a special culture medium (Mac Conkey Agar, for gram negative, and CLED Agar and Blood Agar, for gram positive bacteria) and biochemical proofs for its previous identification (coagulase) [12]. Serial dilutions of each bacterial species were prepared following McFarland's scale from cultured agar plates using an automatic apparatus to standardize its concentration (Densimat, BioMérieux). In this work the scale 1 was used, representing 300,000 bacteria/mL.

Once accomplished the presumptive identification, appropriate biochemical culture media were used to identify with accuracy the bacteria species and type, using commercial strips containing these biochemical media (API strips, BioMérieux Inc, Missouri, USA), in which, according to changes in the color of the pH indicator or bacterial metabolites reaction with chemicals included in the strips, permits the identification [12].

After 18 to 24 hr in incubator at 37°C, the strips containing the culture were analyzed at the mini API®(BioMérieux Inc, Missouri, USA), being accepted the samples with identification accuracy higher than 99%, according to printed label furnished by the equipment. A total of 60 bacterial samples were collected and identified, being 20 of each species.

The fluorescence spectrum of each bacterial suspension, placed in quartz cuvette, was obtained using the SPEX Fluoromax 2 spectrofluorimeter (Jobin Yvon Inc, NJ, USA). The fluorimeter was configured to collect the Excitation-Emission Matrix (EEM) automatically, with

10 excitation wavelengths ranging from 330 to 510 nm, in 20 nm steps, and the emission spectra of each sample in the range of 350 to 700 nm, in 2 nm steps. The integration time for each spectrum was 3 min and the total time for each set of spectra was 30 min. The choice of the excitation and emission wavelength range was based on the automatic laser spectrofluorimenter described by Zângaro *et al*. [13]. The background spectrum (from saline) was taken but not subtracted from spectra. It was found that Raman band of water (most intense peak that appear in all bacteria spectra) gives an intensity reference for normalization.

After saved and stored in ASCII format, the autofluorescence emission spectra were normalized by the most intense peak and separated in three groups according to the bacterial species (EC, EF and SA). For developing the identification model, normalized data were again separated in training and prospective groups, each one composed by 30 spectra, 10 of each species. To the matrix formed by 30 spectra of EC, EF and SA from training group, it was applied PCA algorithm. As the result, the principal components scores were used to correlate the most relevant spectral features described by the first principal components and the bacterial species.

Basically PCA technique concentrates the maximum variance of the data in a reduced number of independent variables, called principal components (PCs) [14]. Normally the largest variation in the spectral data is due to differences in the constitution of the samples, that means a few variables. A significant portion of the variance is represented by the first PCs, in other words, higher differentiation among the spectra, and the last PCs carry uncorrelated information or noise. In order to achieve the best spectral differentiation among the three bacterial species which will allow the correct identification, it was used the first four PCs, that represent about 90% of the variance of the spectra.

The PCA was applied to the fluorescence data in the following ways: (a) using all the 10 excitation wavelengths; (b) using each excitation wavelength individually; (c) using couples of excitation wavelengths. The obtained PCA scores were plotted in pairs, from the 1st to the 4th, in order to verify which PC is able to promote better clustering in the data, thus separating the bacterial species.

The evaluation of the PCA model was performed in the prospective group. For so, the fluorescence spectrum of each new bacterium species was fit to the first four PCs calculated in the training group using a least squares minimization method and the fitting coefficients (the scores) were plotted using the same scale [15]. PCA, least square minimization and data handling (normalization) routines were made using the software Matlab (The Mathworks, MA, USA) and plots were made using software Excel 97 (Microsoft Corp., SP, Brazil).

Table I. Sensitivity and Specificity Values of the Bacterial Autofluorescence Spectra Identification Model Using PCA

	Sensitivity (%)	Specificity (%)
E. coli	100	100
S. aureus	90	100
E. faecalis	100	95

Sensitivity and specificity indices can be used as a evaluation criterion of a particular identification or diagnostic method. Sensitivity is defined as the percentage of samples with a specific disease who have a positive test result. Specificity is defined as the percentage of samples without the disease who have a negative result. The sensitivity and specificity indices of PCA model were calculated for each one of the bacteria types and are shown in Table I.

RESULTS

Figure 1 shows the autofluorescence spectra of *E. coli* (Fig. 1a), *E. faecalis* (Fig. 1b) and *S. aureus* (Fig. 1c) bacteria under 410 nm excitation wavelength (training samples) without background subtraction. It can be observed that the fluorescence profile does not present great variations among samples from the same group and that there is a perceptible difference in the fluorescence emission among the three species, mainly the EC, which presents strong emission in the range 500 to 550 nm, and EF, which presents fluorescence relatively intense in the same range. The peak that appears at 470 nm in all samples is due to the Raman scattering of the water used in the dilutions.

In order to verify which excitation wavelength could best differentiate the emission spectra from the three bacterial species in the training group, the PCA was applied to the data in the following order: all 10 wavelengths, individual wavelengths and pairs of wavelengths.

By using the first four PC scores and the 10 excitation wavelengths, best results were obtained by the scores of the couple PC2 and PC3. Figure 2 shows the plot of the PC2 and PC3 using all 10 excitation wavelengths. Figure 3 shows the plot of the PC2 and PC3 using the excitation wavelengths of 410 and 430 nm. The result using individual wavelengths was not plot due to the poor cluster formation, consequently causing a great number of misclassifications.

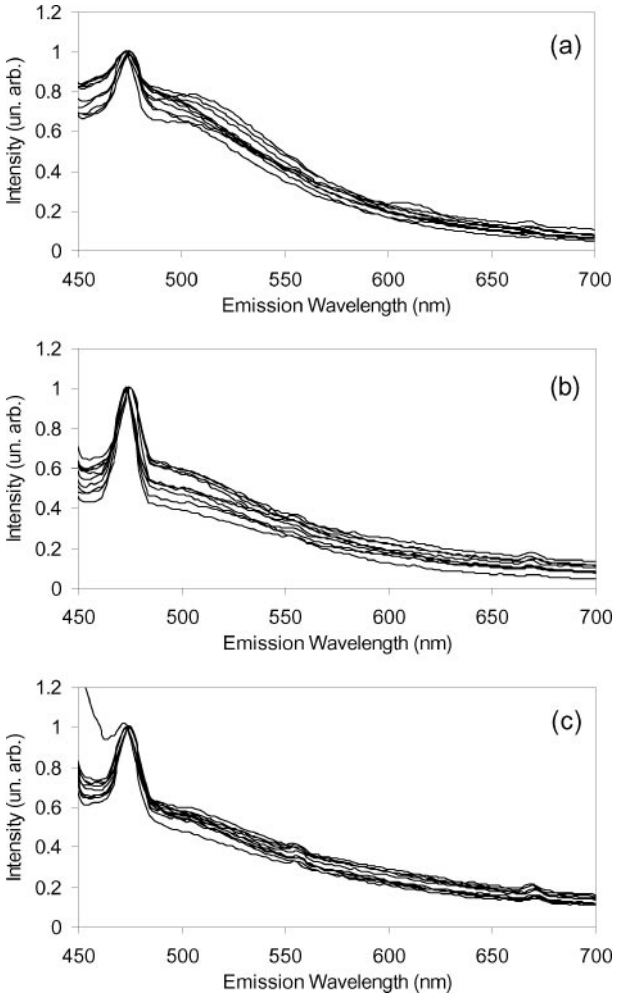

Fig. 1. Autofluorescence spectra of the 30 samples of bacterial species (10 of each species) with excitation wavelength at 410 nm: (a) *Escherichia coli*, (b) *Staphylococcus aureus*, and (c) *Enterococcus faecalis*.

Separation lines were empirically drawn, so that they could separate the clustered data in agreement with the bacterium species. By analyzing the Fig. 3, it can be observed that there are three separation zones, in which *E. coli* and *E. faecalis* totally separate, having however an overlap of 1 case (10%) of *S. aureus* in the *E. faecalis* group.

It was determined that the best performance in terms of separation of the three bacterial species can be obtained with a reduced number of spectra, decreasing the number of wavelengths from 10 to 2, and improving much the time for collection as for processing the spectra (6 min instead of 30). It was found (not plotted) that a single excitation wavelength does not provide a good group separation, mainly because of its low

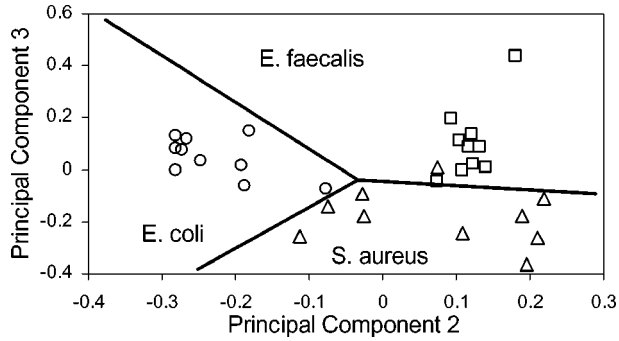

Fig. 2. Scores of the principal components 2 and 3 obtained from the 30 samples in the training group using all 10 excitation wavelengths. Lines provide separation of bacterial species in three groups.

fluorophore selectivity, besides the rapid data acquisition (3 min).

The validity of the proposed identification method was tested using a new group of 30 samples, 10 of each type, previously identified by the automated method (mini API). The fluorescence spectra with excitation at 410 and 430 nm were collected and the new PC scores were determined and plotted. The result of the prospective analysis using 30 samples can be seen in the Fig. 4, whereas the separation lines and scales were kept the same. It was verified that 100% of the samples of *E. coli* and *E. faecalis* were grouped inside the correct region, and 10% (1 case) of the *S. aureus* was again located inside the *E. faecalis* group. The sensitivity and specificity for each bacterial species is shown in Table I. The algorithm provided higher values for both sensitivity and specificity. Highest values are obtained for *E. coli*. The prospective analysis results show that the profile of the bacterial

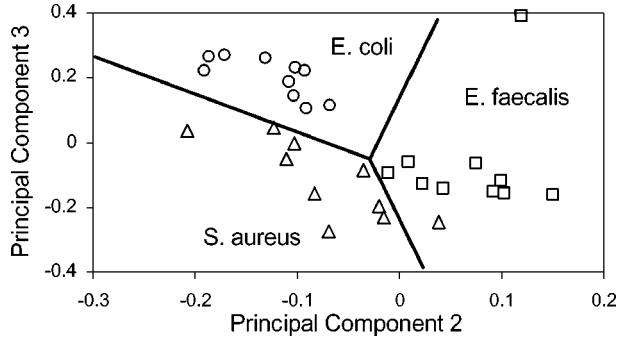

Fig. 3. Scores of the principal components 2 and 3 obtained from the 30 samples in the training group using excitation wavelengths of 410 and 430 nm.

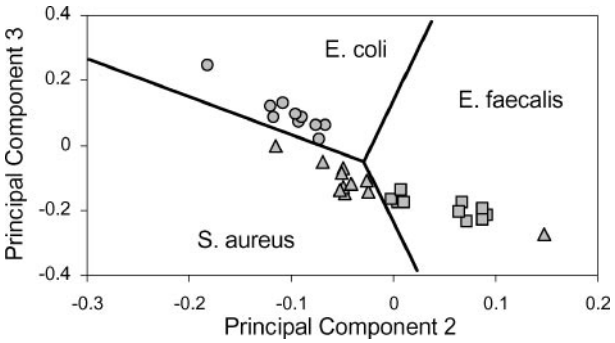

Fig. 4. Scores of the principal components 2 and 3 obtained from the 30 samples in the prospective group using excitation wavelengths of 410 and 430 nm, with separation lines kept the same as Fig. 3.

autofluorescence behaves in the same way, validating the employed method.

CONCLUSIONS

This work presents a method of bacterial identification of three most frequent and more clinically important species (*E. coli*, *E. faecalis* and *S. aureus*) based on the autofluorescence spectra collected from 60 samples in the excitation wavelengths of 410 and 430 nm, and the separation and classification of the spectra in three groups according to the differences observed using the scores calculated through PCA. The results demonstrated that the bacterial identification was very efficient, with cases of false result or wrong identification just for the bacterium *S. aureus*, with 10% of error.

The proposed method can bring a progress in the bacterial identification, with rapidness, high sensitivity and low costs, representing a promising tool for the clinical analysis laboratories, since the costs of preparing the sample can be reduced, as well as the time to obtain the identification being limited by the time for collecting and processing the fluorescence spectra. The time saving with such method can be on the order of 18 to 24 hr, eliminating the need of culture in specific biochemical media.

According to the statistics obtained by the Laboratório Oswaldo Cruz (São José dos Campos, SP, Brazil), almost half of the most common and clinically significant bacteria can be identified by this method. With a higher number of samples and the eight more frequent species of bacteria, which is under way, the algorithm would become an important resource for the triage of the material to be analyzed in clinical analysis laboratories.

ACKNOWLEDGMENTS

We thank Prof. Dr. Paulo Ricardo Galhanone for providing the PCA routines, Prof. Dr. Marcelo Pelisson for the use of the spectrofluorimeter, Dominique Giraldi Barbosa and Simone Rodrigues from Laboratório Oswaldo Cruz for the preparation of bacterial samples.

REFERENCES

1. E. Jawetz, J. L. Melnick, and E. A. Adelberg (1995). *Microbiologia Médica*, Guanabara Koogan, Rio de Janeiro, Brazil, pp. 35–38.
2. W. H. Nelson (1985). *Instrumental Methods for Rapid Microbiological Analysis*, VCH, New York, p. 22.
3. D. C. Roselle, M. Seaver, and J. D. Eversole (1998). Changes in intrinsic fluorescence during the production of viable but nonculturable *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* **20**, 265–267.
4. D. Ivnitski, I. Abdel-Hamid, P. Atanasov, and E. Wilkins (1999). Biosensors for detection of pathogenic bacteria. *Biosens. Bioelectron.* **14**, 599–624.
5. S. C. Hill, R. G. Pinnick, S. Niles, Y. L. Pan, S. Holler, R. K. Chang, J. Bottiger, B. T. Chen, C. S. Orr, and G. Feather (1999). Real-time measurement of fluorescence spectra from single airborne biological particles. *Field Anal. Chem. Tech.* **3**, 221–239.
6. Y. S. Cheng, E. B. Barr, B. J. Fan, P. J. Hargis, D. J. Rader, T. J. O'Hern, J. R. Torczynski, G. C. Tisone, B. L. Preppernau, S. A. Young, and R. J. Radloff (1999). Detection of bioaerosols using multiwavelength UV fluorescence spectroscopy. *Aerosol Sci. Tech.* **31**, 409—421.
7. B. C. Spector, L. Reinisch, D. Smith, and J. A. Werhaven (2000). Noninvasive fluorescent identification of bacteria causing acute otitis media in a chinchilla model. *Laryngoscope* **110**, 1119–1123.
8. K. Maquelin, L. P. 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. *Anal. Chem.* **72**, 12–19.
9. T. Udelhoven, D. Naumann, and J. Schmitt (2000). Development of a hierarchical classification system with artificial neural networks and FT-IR spectra for the identification of bacteria. *Appl. Spectrosc.* **54**, 1471–1479.
10. L. E. Rodriguez-Saona, F. M. Khambaty, F. S. Fry, and E. M. Calvey (2001). Rapid detection and identification of bacterial strains by Fourier transform near-infrared spectroscopy. *J. Agric. Food Chem.* **49**, 574–579.
11. G. L. Daguet (1977). *Técnicas en Bacteriologia, Vol. 3,* JIMS, Barcelona, pp. 45–50.
12. C. H. P. M. Silva (1999). *Bacteriologia: Um Texto Ilustrado*, Eventos, São Paulo, Brazil, p. 125.
13. R. A. Zângaro, L. Silveira Jr., R. Manoharan, G. Zonios, I. Itzkan, R. R. Dasari, J. Vam-Dam, and M. S. Feld (1996). Rapid multiexitation fluorescence spectroscopy system for in vivo tissue diagnosis. *Appl. Opt.* **35**, 5211–5219.
14. D. Frenkel and J. Nadal (2000). Comparação de métodos de representação do segmento ST na detecção automática de isquemia de miocárdio. *Braz. J. Biom. Eng.* **16**, 153–162.
15. L. Silveira Jr., S. Sathaiah, R. A. Zângaro, M. T. T. Pacheco, M. C. Chavantes, and C. A. G. Pasqualucci (2002). Correlation between near-infrared Raman spectroscopy and the histopathological analysis of atherosclerosis in human coronary arteries. *Lasers Surg. Med.* **30**, 290–297.