Anal. Chem. **2004,** *76,* 585−591

Characterization of Microorganisms Using UV Resonance Raman Spectroscopy and Chemometrics

E. Consuelo López-Díez* and Royston Goodacre

Department of Chemistry, UMIST, P.O. Box 88, Sackville Street, Manchester, M60 1QD, U.K.

The past decade has seen an increased interest in the application of several physicochemical analytical techniques for the rapid detection and identification of microorganisms. We report the development of UV resonance Raman (UVRR) spectroscopy for the reproducible acquisition of information rich Raman fingerprints from endospore-forming bacteria belonging to the genera *Bacillus* and *Brevibacillus*. UVRR was conducted at 244 nm, and spectra were collected in typically 60 s. Cluster analyses of these spectra showed that UVRR spectroscopy could be used to discriminate between these microorganisms to species level, and the clustering pattern from this phenotypic classification was highly congruent with phylogenetic trees constructed from 16S rDNA sequence analysis. Therefore, we conclude that UVRR spectroscopy when coupled with chemometrics constitutes a powerful approach to the characterization and speciation of microorganisms.

Accurate and rapid bacterial identification has always been necessary in many areas, and in particular, it is crucial within clinical diagnostics and for food quality controls. Over the past few years, physicochemical methods for whole-organism fingerprinting have attracted significant attention for the characterization of microorganisms. Current techniques include mass spectrometry (MS) methods based on pyrolysis-MS, electrospray-ionization (ESI-MS) and matrix-assisted laser desorption ionization (MALDI-MS), FT-IR spectroscopy, and Raman microspectroscopy (with excitation in the visible and near-infrared).[1−4] These techniques generate high dimensional data and are typically analyzed by multivariate analysis methods, such as principal components (PCA), discriminant function (DFA), and hierarchical cluster (HCA) analyses. However, there does not appear to be a consensus as to which analytical technique, or indeed chemometric processing, is the most suitable for the identification of microorganisms. Thus, techniques are still needed that combine accuracy with rapid data acquisition as well as those that require little sample preparation.

In the present study, UV resonance Raman (UVRR) spectroscopy (with excitation wavelength of 244 nm) in combination with chemometrics will be investigated as an alternative rapid method for bacterial detection. In UVRR, resonance Raman occurs because the laser wavelength used to excite the Raman spectrum lies under an intense electronic absorption band of a chromophore; under this condition, enhancement of Raman scattering may occur so that some of the band intensities are increased by a factor of 10^3−$10^{5.5}$.[5−7] UV-absorbing cell components, such as the nucleic acids and the aromatic amino acids, may be therefore selectively excited to give rise to characteristic resonance Raman bands. This selectivity may be important not only for identifying vibrations of these chromophores, but also for the sensitivity of UVRR to detect genetic and phenotypic differences between microorganisms. Beyond the maximizing of the Raman signal another advantage of UVRR spectroscopy is that measurements below 260 nm are not plagued by fluorescence interference, and so the obtained bacterial spectra are free of this low-frequency background when excited with deep UV light in the 190−250 nm region.[8] This effect not only increases the sensitivity of the spectra but also may reduce the preprocessing stage previous to any further analysis (such as any multivariate analysis method). These characteristics of UVRR have attracted interest in its potential biological applications and, in particular, to the development of a reliable rapid technique for the discrimination of microorganisms.

An excellent first step in this direction has been undertaken by Nelson, Sperry, and colleagues who first reported UVRR spectra using pulsed 242-nm excitation on bacteria and showed that it was possible to obtain reproducible spectra.[8] Since then, several studies by the same group and others using excitation light ranging from 222 to 257 nm have explored the use of UVRR to characterize spectral features that could aid in the identification of microorganisms. For instance, it has been shown that UVRR spectra can yield valuable information regarding the bacterial Gram type[9] and that the intensities of certain resonance Raman

* Corresponding author. Phone: +44 (0) 161 200 4414. Fax: +44 (0) 161 200 4519. E-mail:C.Lopez-Diez@umist.ac.uk.

(1) Claydon, M. A.; Davey, S. N.; EdwardsJones, V.; Gordon, D. B. *Nat. Biotechnol.* **1996**, *14*, 1584−1586.
(2) Goodacre, R.; Timmins, E. M.; Burton, R.; Kaderbhai, N.; Woodward, A. M.; Kell, D. B.; Rooney, P. J. *Microbiology-UK* **1998**, *144*, 1157−1170.
(3) Maquelin, K.; Choo-Smith, L. P.; van Vreeswijk, T.; Endtz, H. P.; Smith, B.; Bennett, R.; Bruining, H. A.; Puppels, G. J. *Anal. Chem.* **2000**, *72*, 12−19.
(4) Vaidyanathan, S.; Rowland, J. J.; Kell, D. B.; Goodacre, R. *Anal. Chem.* **2001**, *73*, 4134−4144.
(5) Carey, P. R. *Biochemical applications of Raman and Resonance Raman Spectorscopies*; Academic Press: New York, 1982.
(6) Asher, S. A. *Anal. Chem.* **1993**, *65*, 59−66.
(7) Asher, S. A. *Anal. Chem.* **1993**, *65*, 201−209.
(8) Dalterio, R. A.; Nelson, W. H.; Britt, D.; Sperry, J. F. *Appl. Spectrosc.* **1987**, *41*, 417−422.
(9) Manoharan, R.; Ghiamati, E.; Dalterio, R. A.; Britton, K. A.; Nelson, W. H.; Sperry, J. F. *J. Microbiol. Methods* **1990**, *11*, 1−15.

10.1021/ac035110d CCC: $27.50
Published on Web 12/23/2003

Table 1. List of the 27 *Bacillus* Strains Used in This Work

sample no.	species	strain no.	other strain nos.	identifier on plots
1	*B. sphaericus*	7134^T	DSM 28^T, G 1013	sph7134T
2		B0219	S2	sphB0219
3		B0408	F 2610/77	sphB0408
4		B0769[a]	NRS 592	sphB0769
5		B1147	1691	sphB1147
6	*B. subtilis*	B0014^T	DSM 10^T, ATCC 6051^T	subB0014
7		B0098	G 1007, Král, *B. niger*	subB0098
8		B0099	G 1208, *B. niger*	subB0099
9		B0410	F 2666/77	subB0410
10		B0501	M (Malaysia) 79A	subB0501
11		B1382[a]	B. globigii	subB1382
12	*Br. laterosporus*	B0043	8/1/15/1 colonial variant	latB0043
13		B0115	G 308, NTCT 7579	latB0115
14		B0262[a]	B0040 contaminant	latB0262
15		B0616	NRS 682, *B. orpheus*	latB0616
16	*B. cereus*	B0550	NRS 721, *B. albolactis*	latB0550
17		B0702	V 26	latB0702
18		B0712	NRS 201, *B. siamensis*	latB0712
19		B0851[a]	Isolate 1	latB0815
20	*B. amyloliquefaciens*	B0175	Campbell P, ATCC 23844	amyB0175
21		B0177^T	Campbell F, ATCC 23350^T	amyB0177
22		B0251[a]	duplicate of B0172	amyB0251
23		B0620	Harrman 1A35	amyB0620
24	*B. megaterium*	B0010^T	DSM32^T, ATCC 14581^T	megB0010
25		B0056	UB 8/1/5/1	megB0056
26		B0076[a]	G 659, NCTC 5637	megB0076
27		B0621	Hartman NRRL B-348	megB0621

[a] Strains used as a test set for the cluster analysis. T = type strain.

bands can be successfully related to the mole percent G + C content.[10] More recent studies have also shown that bacteria belonging to different genera are characterized by significantly different spectral intensity values,[11] and it has also been shown that the UVRR spectrum of *Escherichia coli* in log phase could be estimated on the basis of the average cellular molecular content (DNA, RNA, and protein contents).[12]

Together, the above studies show that UVRR is extremely sensitive to changes in the amount of nucleic acid in bacteria and show a clear relationship between the spectral features and the biochemical nature of the microorganisms. However, none of these studies have assessed whether the UVRR spectra are sufficiently discriminatory to allow for the robust characterization of bacteria by analyzing the spectra by cluster analysis, a process which is necessary to evaluate whether the discrimination of microorganisms based on their UVRR "fingerprints" is related to the known taxonomy of these bacteria.

Thus, the aim of the present study was to combine UVRR spectroscopy and chemometrics for the reliable discrimination of bacteria. As a model, a subset of endospore-forming bacteria were analyzed by UVRR. *Bacillus* and *Brevibacillus* represent two genera of Gram-positive bacteria which are ubiquitous in nature (soil, water, and airborne dust) and are of increasing relevance in the food processing industry as well as in the preparation of sterile products.[13] Furthermore, members of the endospore-forming bacteria are of extreme importance as potential biological warfare agents.[14] Therefore, the accurate discrimination of these bacteria is of crucial importance for health and safety issues.

MATERIALS AND METHODS

Microorganisms. A collection of 27 strains known to belong to the *Bacillus* and *Brevibacillus* genera were used in this study (see Table 1). Bacterial cells were harvested from axenic cultivations of the 27 strains on LabM blood agar base plates (without blood) and incubated at 37 °C for only 12 h to avoid sporulation. Each strain was cultured on four individual Petri dishes to give four biological replicates of each strain. The biomass was carefully removed from the Petri dishes by preparing cell suspensions with 400 μL of sterile distilled water. An aliquot (30 μL) of the suspension was then transferred to a CaF_2 substrate and dried in an oven at moderate temperature (50 °C) to form a uniform semitransparent film suitable for UVRR measurements. UVRR spectra were collected immediately after sample preparation so as to avoid sample degradation in the desiccated state.

UV Resonance Raman Spectroscopy. Raman spectra were excited with a deep-UV 244-nm laser with the power at the sampling point, typically at ~0.5 mW; backscattering radiation was collected by a Renishaw System 1000 (Renishaw plc, New Mills, Wotton-under-Edge, Gloucestershire, GL12 8JR, U.K.).[15,16] Bacterial samples dried onto CaF_2 substrates were placed on a spinning

(10) Nelson, W. H.; Manoharan, R.; Sperry, J. F. *Appl. Spectrosc. Rev.* **1992**, *27*, 67−124.

(11) Wu, Q.; Nelson, W. H.; Elliot, S.; Sperry, J. F.; Feld, M.; Dasari, R.; Manoharan, R. *Anal. Chem.* **2000**, *72*, 2981−2986.

(12) Wu, Q.; Hamilton, T.; Nelson, W. H.; Elliott, S.; Sperry, J. F.; Wu, M. *Anal. Chem.* **2001**, *73*, 3432−3440.

(13) Doyle, M. P.; Beuchat, L. R.; Montville, T. J. *Food Microbiology: Fundamentals and Frontiers*; American Society of Microbiology Press: Washington, DC, 1997.

(14) Dando, M. *Biological Warfare in the 21st Century*; Brassey's Ltd.: London, 1994.

(15) Williams, K. P. J.; Pitt, G. D.; Batchelder, D. N.; Kip, B. J. *Appl. Spectrosc.* **1994**, *48*, 232−235.

plate mounted on the microscope stand. Rotation of the sample (at ~30 rpm) under the laser beam prevented the effect of sample degradation by UV light, and allowed reproducible spectra to be obtained. Spectra were collected for 1 min. Initially, spectra were acquired from 500 to 2300 wavenumbers (cm^{-1}). Visual inspection of these spectra showed that the most information-rich area was between 1100 and 1800 cm^{-1}, and this was used for data analysis. Spectral resolution was ~10 cm^{-1}. Calibration was periodically checked by recording the position of known Raman lines of a diamond crystal (1332 cm^{-1}), and the wavenumber accuracy was estimated to be ±4 cm^{-1}. All data were exported from the GRAMs WiRE software used to control the spectrometer into Matlab (The Mathworks, Inc. Natick, MA) for data analysis.

Data Analysis. Spectral preprocessing included a denoising routine that smoothed the spectra using a Savitzki−Golay (polynomial) smoothing filter with a 5-point (~20 cm^{-1})-wide window and a second-order polynomial. To account for cell density variation in the sample film, the data were then normalized so that the Σ photons from 1100 to 1800 cm^{-1} equalled 1.

To assess the ability of UVRR data to discriminate between the different bacilli, the UVRR spectra were analyzed by multivariate statistical methods. This process involved three different steps: (1) Principal components analysis (PCA)[17,18] was employed to reduce the dimensionality of the UVRR data while preserving most of the variance, and PCA was performed using the NIPALS algorithm.[19] (2) Discriminant factor analysis (DFA) was then used to discriminate between groups on the basis on the retained principal components (PCs) and the a priori knowledge of which spectra were biological replicates (i.e., the four spectra from the four cultures for each *Bacillus* strain). The number of principal components (PCs) used by the DFA was optimized by cross validation, which was performed by creating a model on a training data set and then projecting an independent test set (a strain from each of the six species was picked randomly) into the model as described elsewhere.[20] Fifteen PCs were used that accounted for 93−94% of the total explained variance. (3) Finally, the Euclidean distance between a priori groups centers in DFA space (the first 3 DFs) were used to construct a similarity measure with the Gower similarity coefficient, S_G, and these distance measures were then processed by an agglomerative clustering to construct a dendrogram. PCA, DFA, and hierarchical cluster analysis (HCA) were all performed in Matlab version 6 on an IBM-compatible PC running under Windows XP.

The above approach has been previously applied with success for the analysis of PyMS, FT-IR, and ESI-MS spectra,[4,21,22] and therefore, such methodology was also adopted in this study. Nevertheless the combination of PCA, DFA, and HCA does not readily yield information in terms of which spectral features are used for the discriminatory purposes.

By contrast, it has been shown that evolutionary computational-based methods can be used[23−25] to discover the relationship between an input vector (i.e., the spectra) and a particular property (i.e. the species analyzed) while also identifying the input features of a given spectrum that are relevant for such a discrimination. Consequently, further data analysis was performed by genetic programming (GP).[22,23,26] GP is part of the "so-called" evolutionary computational methods, which are based on the concepts of Darwinian selection to generate and to optimize a desired computational function or mathematical expression to produce explanatory rules. GP was performed using the genomic computing software Gmax-bio (Aber Genomic Computing, Aberystwyth, U.K.). Details on the Gmax-bio software can be found elsewhere.[24,27] The recommended default parameter settings for population size (1000) mutation and recombination rates were used throughout. The operators that were used were +, −, ÷, ×, 0.1, 1, 3, 5, rand, $\log_{10}$, 10^x, tanh. The fitness (F) calculation is determined by the FITNESS setting in the software so $F = 1/(0.01 + S/B)$, where the values of S is a statistic derived from the model which ranges between 0 and infinity and B is a normalizing quantity. The value of B is chosen such that a perfect model yields an F value of 100 and $F = 1$ describes a model that performs no better than random chance.

The phylogenetic tree of the type strains of the *Bacillus* species studied here was constructed using their 16S rDNA sequences. Sequences were downloaded from the Ribosomal Database Project (http://rdp.cme.msu.edu/html/) and were analyzed using TREECON. TREECON is a software package for the construction of phylogenetic trees on the basis of evolutionary distances inferred from nucleic and amino acid sequences.[28] The evolutionary distance was computed by the Juke−Cantor algorithm, and the tree topology was inferred using neighbor-joining for all pairs of organisms (or sequences).

RESULTS AND DISCUSSION

Bacterial UVRR Spectra. Typical spectra of the six species of *Bacilli* studied in this work are shown in Figure 1. Nucleic acids and aromatic amino acids strongly absorb light in the UV range, and therefore, excitation of bacteria at wavelengths between 222 and 257 nm has been shown to result in spectra dominated by the Raman bands of these chromophores.[8−10,29] In this study, spectra of the nucleic acid bases adenine (A), guanine (G), thymine (T), cytosine (C), and uracil (U), and the aromatic amino acids tyrosine (Tyr), tryptophan (Trp), and phenylalanine (Phe) were collected (data not shown) and compared with those of the bacteria studied. The main Raman bands excited at 244 nm that were observed together with the proposed peak components are listed in Table 2, and these band assignments are in accordance with previously published results.[12] It is notable that at this

(16) Williams, K. P. J.; Pitt, G. D.; Smith, B. J. E.; Whitley, A.; Batchelder, D. N.; Hayward, I. P. *J. Raman Spectrosc.* **1994**, *25*, 131−138.
(17) Jolliffe, I. T. *Principal component analysis*; Sringer-Verlag: New York, 1986.
(18) Causton, D. R. *A Biologist's advanced mathematics*; Allen and Unwind: London, 1987.
(19) Wold, H. *Multivariate analysis*; Academic Press: New York, 1966.
(20) Radovic, B. S.; Goodacre, R.; Anklam, E. *J. Anal. Appl. Pyrol.* **2001**, *60*, 79−87.
(21) Timmins, E. M.; Howell, S. A.; Alsberg, B. K.; Noble, W. C.; Goodacre, R. *J. Clin. Microbiol.* **1998**, *36*, 367−374.
(22) Goodacre, R.; Shann, B.; Gilbert, R. J.; Timmins, E. M.; McGovern, A. C.; Alsberg, B. K.; Kell, D. B.; Logan, N. A. *Anal. Chem.* **2000**, *72*, 119−127.
(23) Gilbert, R. J.; Goodacre, R.; Woodward, A. M.; Kell, D. B. *Anal. Chem.* **1997**, *69*, 4381−4389.
(24) Ellis, D. I.; Broadhurst, D.; Kell, D. B.; Rowland, J. J.; Goodacre, R. *Appl. Environ. Microbiol.* **2002**, *68*, 2822−2828.
(25) De Oliveira, P. P. B.; Vogler, O.; Matta, C. E. *Inverse Probl. Eng.* **2003**, *11*, 309−328.
(26) Koza, J. R. *Genetic programming: on the programming of computers by means of natural selection*; MIT press: Cambridge, Mass, 1992.
(27) Kell, D. B.; Darby, R. M.; Draper, J. *Plant Physiol.* **2001**, *126*, 943−951.
(28) Van de Peer, Y.; De Wachter, R. *Comput. Appl. Biosci.* **1994**, *10*, 569−570.
(29) Manoharan, R.; Ghiamati, E.; Chadha, S.; Nelson, W. H.; Sperry, J. F. *Appl. Spectrosc.* **1993**, *47*, 2145−2150.

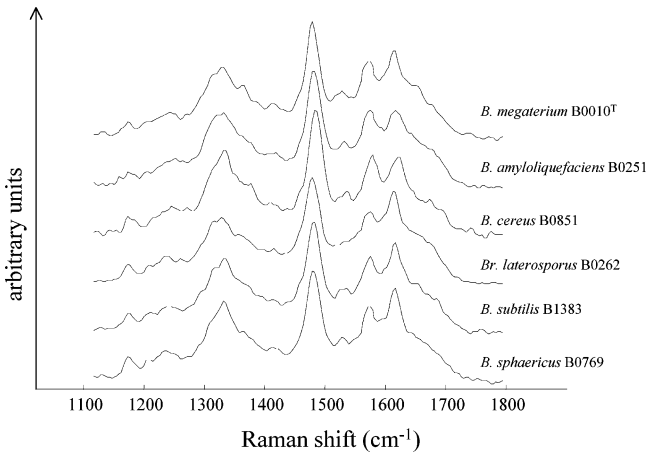

Figure 1. Typical processed UVRR spectra from *Br. laterosporus* and the five *Bacillus* species analyzed. The band assignments are shown in Table 2.

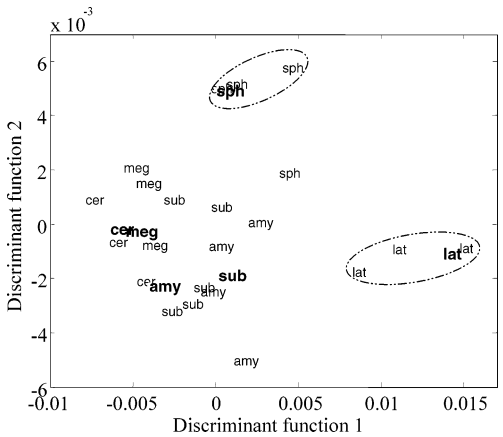

Figure 2. PC-DFA plot showing the relationships between the whole collection of endospore-forming bacteria analyzed in this study (see Table 1 for legends used). Fifteen PCs that contained 93% of the explained variance were used to construct the ordination plot. The group center of each strain is plotted. Bold typeface represent the test set spectra that were projected into the PC-DFA space resulting from calibration with the training set (nonbold typeface), see text for full details. Ellipses are drawn as a visual guide.

wavelength, the major contribution to the bacterial spectra is due to nucleic acids and in particular to the purine bases, which have two rings in their structure, and therefore, Raman scattering from these chromophores is enhanced further. In addition, it was found that the peaks at 1172 and 1607 cm^{-1} result solely from the vibration of the aromatic amino acids Tyr, and Trp and Tyr, respectively; while no major band could be assigned to Phe. Note that bacterial spectra were collected by rotating the sample (see material and methods), and therefore, the Raman spectra depicted in Figure 1 are likely to represent the spatial averaging of the cells rather than the spectra of a single cell.

Discrimination at Species Level. Simple visual inspection of the UVRR spectra in Figure 1 (and indeed, of all the spectra acquired) shows they are *qualitatively* very similar. However, on closer inspection subtle *quantitative* differences can be seen, and it was hoped that these would contain enough information to allow for the characterization of the bacteria analyzed. It is obvious that such simple visual inspection (that is to say, "stare and compare" analysis) of these spectra will not allow one to discover the relationship between the bacteria based on their UVRR "fingerprints", and alternative strategies need to be adopted. Therefore, multivariate statistical methods were used, and these included PCA, DFA, and HCA, as detailed above.

In the initial stage of the data analysis, all of the spectra (27 × 4) were analyzed. The resultant PC-DFA ordination plot and dendrogram are shown in Figures 2 and 3, respectively. It can be seen from both figures that three main groups are recovered: (1) All the *Brevibacillus laterosporus* strains are recovered in one cluster very separately from all the *Bacillus* species. Given that these two major groups reflect the two different genera, this was perhaps not surprising. (2) All the *Bacillus sphaericus* were recovered together. (3) The rest of the bacilli comprising *Bacillus amyloliquefaciens*, *Bacillus cereus*, *Bacillus megaterium*, and *Bacillus subtilis* were recovered in a heterogeneous group. Note that on closer inspection and as indicated in Figure 3 that within this group, the *B. cereus* and *B. megaterium* strains appear separated from those of the *B. amyloliquefaciens* and *B. subtilis* species.

In the construction of Figures 2 and 3, to test the reproducibility of these cluster analysis approaches, one strain from each of the six species was selected randomly and used to cross-validate the PC-DFA and HCA, while the remainder (21 × 4 spectra) were used to construct the models. After construction of the PC-DFA plot and dendrogram, the test set spectra were projected into this space. That the test set spectra in the ordination plot (Figure 2) are recovered very closely to the same species shows that the analysis is wholly accurate.

These results from the initial cluster analysis are highly encouraging because they show that for *Br. laterosporus* and *B. sphaericus*, different strains from these species are recovered together. This suggests that the UVRR spectra do contain enough information to allow for the differentiation between these bacteria. However, the adopted strategy (that is, PCA followed by DFA and HCA) does not readily give details on whether a particular spectral region or the whole spectra is used for discrimination of *B. sphaericus* or *Br. laterosporus*. As mentioned above, recent studies have demonstrated that GP can be used to find the relevant spectral features for a given discriminatory purpose (i.e., distinguish between fresh and spoiled chicken[24]). Therefore, in the present study, a number of GPs were evolved to determine whether discrimination of these two species could be carried out on the basis of a given region in the spectrum. Thus, using GP, two different models (a model for each species), were produced to discriminate the *B. sphaericus* and *Br. laterosporus* strains from

Table 2. Prominent Raman Bands for *Bacillus* Species Excited at 244 nm, Contributions from the Amino Acids and Nucleic Acid Vibrations

	Raman bands (cm^{-1})								
	1172	1247	1324	1359	1475	1524	1567	1607	1638
peak components	Tyr	G + A + U	A + G + Tyr	T + A	G + A	C	G + A	Tyr + Trp	T

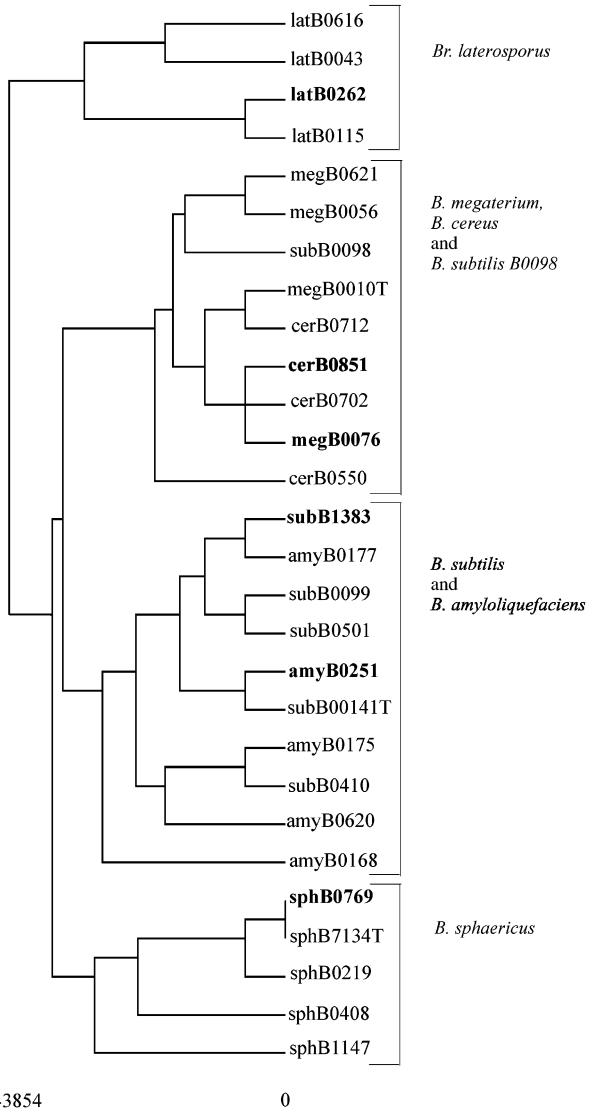

Figure 3. Dendrogram resulting from HCA showing the relationship among all 27 bacteria analyzed by UVRR. Test set spectra are shown in bold; the training set is in normal font.

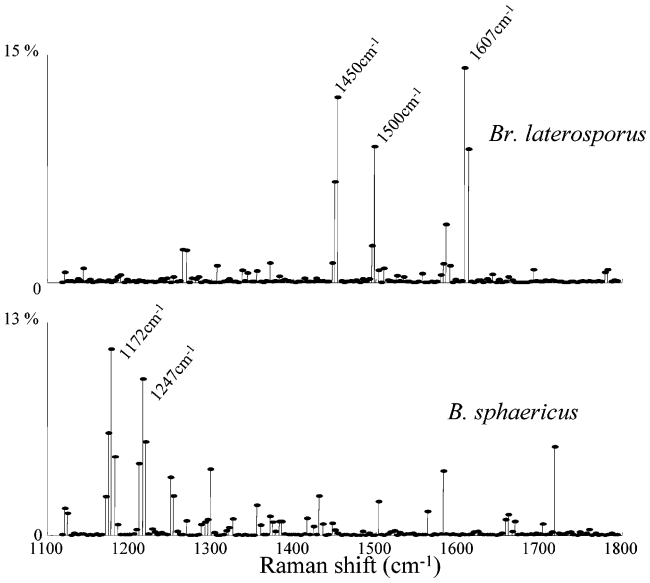

Figure 4. Frequency plot of the number of times an input variable (Raman shift) was used in 10 independent GPs, evolved to discriminate (a) *Br. laterosporus* and (b) *B. sphaericus* from the rest of the species studied.

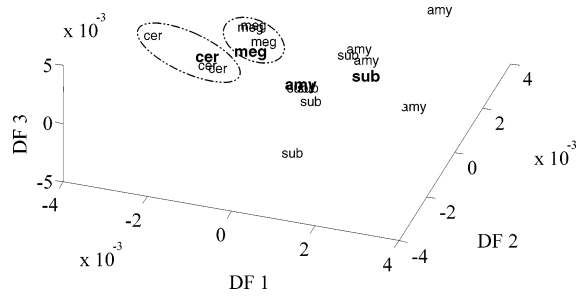

Figure 5. Pseudo 3-D PC-DFA plot showing the relationship between strains of *B. amyloliquefaciens*, *B. cereus*, *B. megaterium*, and *B. subtilis* (see Table 1 for legends used) where the group center of each strain is plotted. Fifteen PCs that contained 94% of the explained variance were used to construct the ordination plot. Bold typeface represent the test set spectra that are projected into the PC-DFA space resulting from calibration with the training set (nonbold typeface). See text for full details. Ellipses are drawn as a visual guide.

the rest of the bacterial collection. For each model, since initial populations were generated randomly, 10 separate GPs were evolved to ensure reproducibility. Subsequently, for each model, the number of times each input variable (i.e., Raman shift) was selected to build the 10 discriminatory models was calculated and plotted against the Raman shift (Figure 4).

The GP trees (or mathematical explanatory rules) generated in this way had an average fitness of 75 for both models (note the fitness is described by a nonlinear function) and produced the frequency plots shown in Figure 4. As can be seen, there are two relevant areas of the spectra for discriminating between *Br. laterosporus* and the rest of the collection, the first area being at 1607 cm^{-1} and the second at 1450 and 1500 cm^{-1} around the edges of the 1475 cm^{-1} band. The former is typically attributed to the in-plane stretching of the aromatic ring of the amino acids tyrosine and tryptophan, and the 1475 cm^{-1} band is tentatively assigned to ring vibrations of the nucleotides bases adenine and guanine. It is possible that the *Br. laterosporus* species has a different protein complement compared with the other bacilli, but there is no direct evidence of this. Nelson and others,[10] using UVRR with excitation at 242 nm, have shown that the intensity of the peak at 1475 cm^{-1} is related to the GC content of the bacteria and that the ratio of the intensities of the 1530 and 1475 cm^{-1} peaks shows a linear relationship with the bacterial GC content. An obvious question that arises is "Are we simply measuring a GC-AT bias in our analysis?" However, literature percent GC values for these organisms are all very similar,[30] with *Br. laterosporus* being 40−43% while the other *Bacillus* species are 32−48%. These values vary within a large range and, although this could still indicate that bacterial discrimination is based on the percent GC content, they do not contain enough information to conclude whether UVRR spectra simply measure the GC-AT bias or measure a combination of biochemical and genetic features of the microorganisms.

By contrast, the discrimination of *B. sphaericus* was predominantly based on the spectral region around the 1172-cm^{-1} band,

(30) Priest, F. G. In *The Aerobic Endospore-Forming Bacteria. Classification and Identification*; Berkeley, R. C. W., Goodfellow, M., Eds.; Academic Press: New York, 1981; pp 33−57.

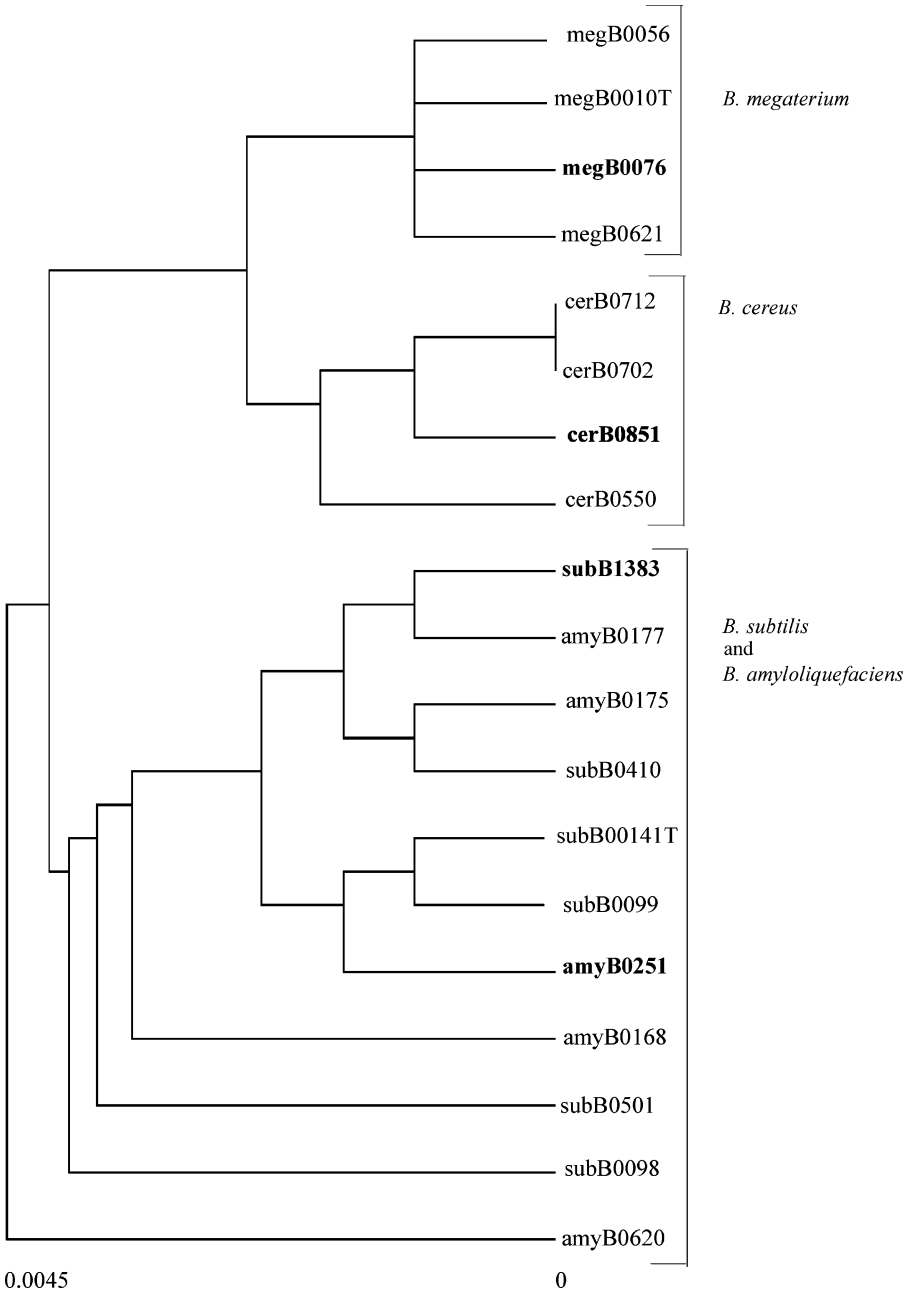

Figure 6. Dendrogram resulting from HCA showing the relationship between strains of *B. amyloliquefaciens*, *B. cereus*, *B. megaterium*, and *B. subtilis* analyzed by UVRR. Test set spectra are shown in bold; the training set is in normal font.

which has been attributed to the vibrational mode ν9a, in plane CH bend, of tyrosine[31] and, therefore, reflects to some extent the protein content of the bacterial cell. It is known that *B. sphaericus* strains are covered by a cell surface structure formed by a 2D crystalline array of protein or glycoprotein subunits (termed the S-layer).[32] It is, therefore, possible that the discrimination by the GP rules has been made on the basis of the elevated protein content of the S-layer coat.

From the previous analyses, it is clear that spectra of the species *B. sphaericus* and *Br. laterosporus* show characteristic features that differentiate them from those of the rest of the studied species (see Figure 1), and this is reflected in the clustering pattern (Figures 2 and 3) as well as in the frequency plots (Figure 4). It is possible that those differences might distort the PC scores in the direction of those groups and therefore might be masking the differences due to the other species.[4] Consequently, spectra from these two species were excluded from the data set, the remaining spectra were reanalyzed, and projection analysis was carried out as detailed above. By contrast to analyzing all the bacilli together, the resulting PC-DFA plot and dendrogram from the reduced data set (Figures 5 and 6, respectively) show that now the two species *B. cereus* and *B. megaterium* can be clearly separated into two different clusters. *B. subtilis* and *B. amyloliquefaciens* were clustered together in a mixed group and this was not resolved by analyzing the two species separately.

In an attempt to find out which regions of the spectra are relevant for the discrimination of the *B. cereus* and *B. megaterium*

(31) Asher, S. A.; Ludwig, M.; Johnson, C. R. *J. Am. Chem. Soc.* **1986**, *108*, 3186−3197.
(32) Sara, M.; Sleytr, U. B. *Prog. Biophys. Mol. Biol.* **1996**, *65*, 83−111.

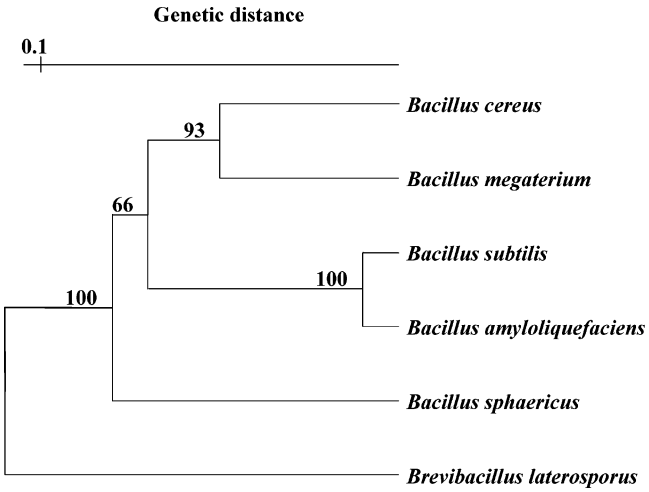

Figure 7. Phylogenetic tree of the type strains of the *Bacillus* species studied. The tree was constructed using the sequences of the 16S rDNA gene.

strains from those of *B. subtilis* and *B. amyloliquefaciens*, GP rules were again evolved. Unfortunately, the generated GP models were characterized by a poor fitness, below 10, and thus, the frequency plots would not be considered to contain reliable information. This indicates that GP failed to distinguish those species from the rest, which probably reflects the great (bio)chemical similarity between those and rest of the collection.

The UVRR analyses performed in this study resulted in a clustering pattern in which (1) the *Br. laterosporus* strains were the most different bacteria; (2) the *B. sphaericus* strains were the next most different group of organisms; (3) *B. megaterium* and *B. cereus* appeared similar, but could be separated into two groups reflecting the two different species; and (4) *B. amyloliquefaciens* and *B. subtilis* clustered together, but that these two species were unresolved. The *Bacillus* species characterized in this study have been previously subjected to a series of miniaturized biochemical tests (API tests).[33] These phenotypic tests separated these species as *B. cereus* belonging to group I; *Br. laterosporus* to group II; *B. sphaericus* to group III; and *B. subtilis* and *B. amyloliquefaciens*, together with *B. megaterium*, fell in group IV, in which the latter appears significantly separated from the rest. Phylogenetic analysis on the sequences of the 16S rDNA gene (the tree is shown in Figure 7) indicates that *Br. laterosporus* is very distinct from the *Bacillus* spp.; that *B. sphaericus* is the most dissimilar *Bacillus* spp. and is clearly separated from the rest; and that *B. subtilis* and *B. amyloliquefaciens* comprise one group, whereas the *B. cereus* and *B. megaterium*, although easily separable, come off the tree at a similar point. It is clear that there is excellent correlation between the species level discrimination observed with UVRR and that based on the phylogenetic analysis of the 16S rDNA sequences. Furthermore, the correlation with the biochemical data is also very encouraging.

The agreement between UVRR and 16S rDNA is perhaps not surprising, since the Raman spectra, excited at 244 nm, are dominated by the nucleic acid vibrations,[8,9,11,12,29] although the genotype is obviously reflected in the organism's phenotype, since its phenotype = its genotype + its environment. Thus, while the relative peak heights and positions may partly reflect the average nucleic acid content for the different species, the UVRR spectra also contain a wealth of additional information regarding the biochemical composition of the bacteria, as clearly demonstrated by the GP models that selected protein signals as well as nucleic acid ones as being important for differentiating *Br. laterosporus* and *B. sphaericus*.

CONCLUSION

The work presented here demonstrates that UV resonance Raman, together with appropriate chemometrics, can be used as a tool for discriminating between very closely related endospore-forming bacteria. In accordance with previously published results, it has been shown that UVRR bacterial spectra excited at 244 nm are dominated by nucleic acid and the aromatic amino acids. In addition, the UVRR spectra required little spectra preprocessing, and cluster analyses produced robust and accurate models which were wholly congruent with the known taxonomy of these organisms, as judged by biochemical tests and 16S rDNA sequences. Furthermore, evolutionary computational methods indicated that for *Br. laterosporus* and *B. sphaericus*, it was possible to select a spectral region in the UVRR spectra that may allow for rapid detection of these species. In conclusion, we believe that UVRR when combined with cluster analysis presents itself as a powerful method for the whole organism fingerprinting of bacteria.

ACKNOWLEDGMENT

We thank U.K. EPRSC for funding. We are very grateful to Dr. Niall Logan for providing us with the bacterial strains and Professor Douglas B. Kell for useful discussions.

Received for review September 22, 2003. Accepted November 10, 2003.

AC035110D

(33) Logan, N. A.; Berkeley, R. C. W. *J. Gen. Microbiol.* **1984**, *130*, 1871−1882.