EDGEWOOD

CHEMICAL BIOLOGICAL CENTER

U.S. ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND

ECBC-TR-533

PERTURBATION BY UV LIGHT FOR RAPID CLASSIFICATION OF BIOLOGICAL PARTICLES BY FLUORESCENCE

Burt V. Bronk

AFRL-APG SITE

Jozsef Czege

UNIFORMED SERVICES UNIVERSITY OF HEALTH SCIENCES
Bethesda, MD 20814

Zhao Z. Li

SCIENCE AND TECHNOLOGY CORPORATION
Edgewood, MD 21040

Karl S. Booksh
Jeff Cramer

ARIZONA STATE UNIVERSITY
Tempe, AZ 85287

January 2007

Approved for public release; distribution is unlimited.

ABERDEEN PROVING GROUND, MD 21010-5424

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
XX-01-2007	Final	May 2005 - Oct 2006

4. TITLE AND SUBTITLE
Perturbation by UV Light for Rapid Classification of Biological Particles by Fluorescence

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Bronk, Burt V. (AFRL); Czege, Jozsef (USU); Li, Zhao Z. (STC); Booksh, Karl S.; and Cramer, Jeff (ASU)

5d. PROJECT NUMBER
622622

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
AFRL/HEPC, APG, MD 21010-5424
USU, Bethesda, MD 20814
STC, Edgewood, MD 21040
ASU, Tempe, AZ 85287

8. PERFORMING ORGANIZATION REPORT NUMBER
ECBC-TR-533

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Some time ago investigations were initiated into the fluorescence of calcium dipicolinic acid (CaDPA) to improve detection of bacterial endospores. Although the native chemical, CaDPA, fluoresced very weakly, the application of UV irradiation to CaDPA or DPA caused substantial blue-violet emission when excitation is applied later. Further investigation demonstrated a similar phenomenon for dry and wet endospores. The luminescence excitation-emission (Ex-Em or EEM) pattern of vegetative bacteria of various *Bacilli* as well as vegetative cells of other unrelated bacteria also changes markedly after UV irradiation. We found that the Ex-Em patterns for unirradiated bacteria taken together with the patterns after UV exposure provide a way to rapidly and inexpensively distinguish several different classes of biological particles and distinguish each of these from common background interferrants. In this report, the two dimensional experimental Ex-Em patterns using Parallel Factor Analysis and other modern pattern recognition techniques are analyzed. This analysis showed that gram positive bacteria and spores can be distinguished from gram negative bacteria, that vegetative bacteria can be distinguished from spores, and that all of these are distinguishable from certain common backgrounds using this potentially automatable technique.

15. SUBJECT TERMS
Discrimination of microorganisms from background after UV exposure
Gram negative from gram positive by fluorescence
Classification of microorganisms by fluorescence
Fluorescence of microorganisms
Microorganism fluorescence after UV
Excitation-emission patterns of microorganisms

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Sandra J. Johnson
U	U	U	UL	28	19b. TELEPHONE NUMBER *(include area code)* (410) 436-2914

Blank

PREFACE

The work described in this report was authorized under Project No. 622622 The work was started in May 2005 and completed in October 2006.

The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for public release. Registered users should request additional copies from the Defense Technical Information Center; unregistered users should direct such requests to the National Technical Information Service.

Blank

CONTENTS

FIGURES

TABLES

Blank

PERTURBATION BY UV LIGHT FOR RAPID CLASSIFICATION OF BIOLOGICAL PARTICLES BY FLUORESCENCE

1. INTRODUCTION

Many government and private institutions have an interest in developing instrumentation for rapidly assessing ambient air or water for pathogenic microorganisms. Since all microorganisms seem to exhibit fluorescence, this phenomenon was expected to be useful as a mode of detection. (In the present context, we will use the word "fluorescence" to encompass all luminescence where a longer wavelength of light is emitted due to electronic excitation by a more energetic shorter wavelength.) Indeed, the fluorescence of microorganisms following excitation by UV wavelengths has proven useful in distinguishing biological from non-biological particles in aerosols.[1,2.,3]

Almost all strains of species belonging to the bacterial genera *Bacillus* and *Clostridium* produce endospores when these bacteria run out of nutrient. Endospores are a particularly hardy life form, which have great resistance to damage by various environmental hazards such as sunlight and various chemicals. Therefore detection of endospores in aerosols has been a major concern to people monitoring the environment for dangerous particles. Other bacteria form different kinds of spores which are rather less resistant to damage. These are currently of lesser interest. We will therefore use the word spore and endospore interchangeably in the present discussion.

Some time ago we initiated studies to investigate fluorescence from the chemical dipicolinic acid (DPA) in various forms.[4-5] This chemical in the form of calcium dipicolinate (CaDPA) is the organic chemical usually predominant (~10% of the spore's drt weight) in endospores but rarely found elsewhere. Thus a characteristic fluorescence from CaDPA would indicate the presence of spores. It was found that fluorescence was hardly detectable when the chemical had been protected from light; however, fluorescence from DPA in various forms became strong in the violet-blue region after UV radiation.[4-5] This was followed by an investigation to see whether the effect of UV irradiation on fluorescence could also be observed in living spores.[6] The effect was indeed present, and distinguishable from the fluorescence resulting from the effect of UV on vegetative bacteria which do not contain DPA.[6] Later investigations showed that the enhanced fluorescence of the chemicals DPA and CaDPA could be observed in the dry state as well as in the wet state, and in dry or wet spores.[7-9] The suggestion was made that UV and other perturbations[5-6] could be used as a basis for rapid classification of bacteria found in the environment.

Recently investigations were undertaken to investigate further how the CaDPA or DPA contained in endospores affected the fluorescence of those spores. Excitation-emission (Ex-Em) graphs were obtained for the isolated chemical in both dry form and in solution.[7] These showed where one might expect to see emission from the chemical as a spore component. These were followed by studies of the Ex-Em graph of *Bacillus subtilis* spores of two types: The fluorescence Ex-Em (or EEM) graphs from spores of a normal, wild-type strain in which CaDPA was present, (DPA^+ PS 832) were compared with those produced from a mutant strain derived

from the (FB108, DPA$^-$, i.e., DPA-less) in which there is much less DPA (by a factor of 10 to 20).[10] The DPA-less spores showed much less fluorescence in the region influenced by DPA.

We show in Figure 1, an example of the comparison of the Ex-Em graphs for different *Bacillus subtilis* spores with the normal amount of CaDPA present (~10 % of dry weight) with graphs for very similar but modified spores with very little DPA present (less than 1 % dry weight). The situation illustrated is complicated. Both the two upper centers of luminescence increase in intensity after UV (spots at excitation ~350 nm and at ~370 nm). The spot occurring at the well known location for tryptophan fluorescence (excitation ~280 nm) diminishes for both DPA- and DPA+ spores by about the same percent (data not shown), but the scales in the figure were adjusted so that the tryptophan fluorescence appears at roughly the same brightness for all four graphs. The conclusion is that the two upper longer wave length fluorescence centers become brighter after fluorescence in both cases, but are much brighter when CaDPA is present. Since CaDPA is present in large proportions for almost all unmodified endospores, the situation should be similar for all spores of Bacillus or Clostridium species. As a matter of fact, the results from other endospores resemble those for the DPA+ spores shown here.

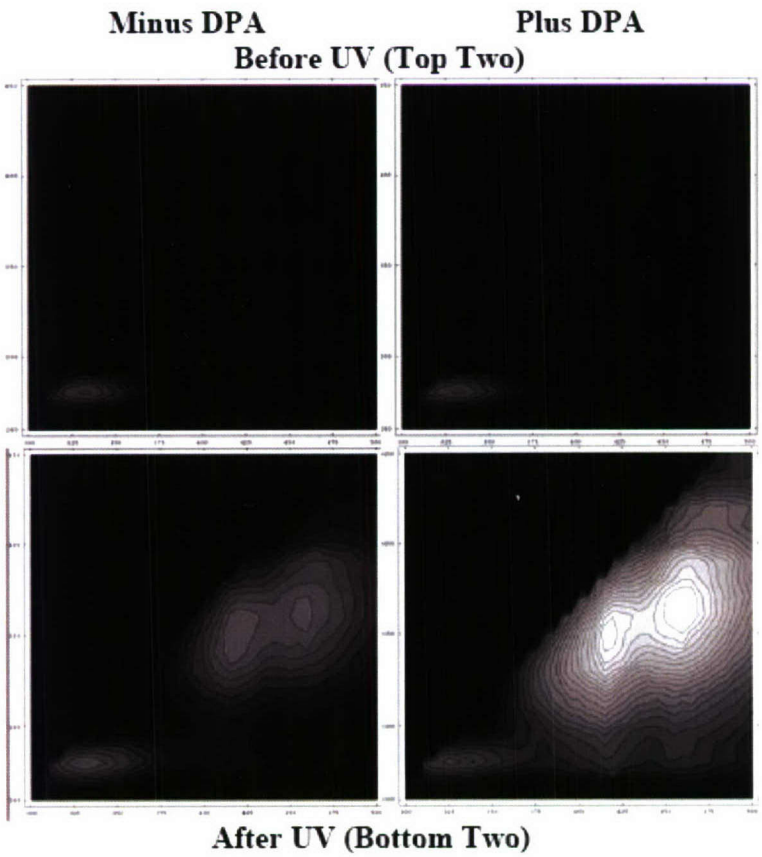

Figure 1. Fluorescence of two isogenic *Bacillus subtilis* spore samples (DPA- and DPA+) before and after UV exposure. The two graphs on the left are DPA-, while the two on the right are DPA+. The top graphs have not had UV applied, while the bottom two have been subjected to 3.1 J/cm^2 of UVC (254 nm) during a 60 min. irradiation.

2. EXPERIMENTATION

In the course of this investigation we have looked at a variety of microorganisms prepared grown in a number of different media, and with other variations in preparation. Table 1 lists the organisms we have used in this project so far, and Table 2 lists the growth media in most of the cases.

Table 1. Microorganisms, Gram Classification, and Interferrants for Experiments Reported.

Gram Positive (GP) Bacteria	**DataSet No.**
Species	
Vegetative Prep.	
Staphylococcus epidermidis	1, 2, 3, 4, 5,45,46
Enterococcus durans	50, 58
Bacillus atrophaeus (BG vegetative)	42
Endospores	
Bacillus atrophaeus(Field BG)	9,18
Bacillus atrophaeus (Fluidized BG)	28,41
Bacillus subtilis (PS832)	13
Bacillus subtilis (FP122 plus DPA)	22,24,26,34
Bacillus subtilis (FP122 minus DPA)	23,25,27,33,44
Bacillus thuringiensis (kurstoki,clean)	15,17,19
Bacillus thuringiensis (kurstoki, dirty)	29
Bacillus thuringiensis (israeliensis)	47,51
Bacillus cereus (T)	37
Clostridium perfringens	59
Gram Negative (GN) Bacteria	
Species (all vegetative)	
Escherichia coli (B/r)	6,7,8,10,11,12,14,38,39
Escherichia coli (K12)	49,52,55,56,57
Pantoea agglomerans (formerly Erwinia h)	16,20,21,32
Interferrants Studied	
Diesel Oil; Household Dust (Abingdon, MD); Household Dust (Highland, MD); Outdoor and Indoor Dust (Tempe, Arizona);Lycopodium spores (nonbacterial); Brain Heart Infusion medium; Luria Broth (fresh); Luria Broth (depleted)	

The growth media along with a key to the experiments where they were used is presented in Table 2.

Table 2. Media Used in Experiments.

Growth Medium	Recipe (per liter distilled H_20)	DataSet No.
Luria broth (LB)	Tryptone--- 10.0 g; Yeast Extract--- 5. g; NaCl---10.0 g	6- (short growth-log phase,6 only)7,8, 12,14,38,49,52,55,56,57,
Tripticase Soy Broth (TSB)	Trypticase Soy Broth (Bacto cat# 211825),used 30g/liter with no other additive	16,20,21,32
Brain Heart Infusion broth (BHI)	Brain Heart Infusion Broth (Difco 237500)--- 25.0 g; Nutrient Broth (BD234000) --- 5.4 g Yeast extract--- 2.5 g	1,2,5,47,50,58
M1 minimal medius	M1 medium NH4Cl---2.0 g; Na2HPO4--- 6.0 g; KH2PO4--- 3.0 g; NaCl--- 3.0 g Autoclave and add following 2 chemicals separately for final concentration per liter: MgSO4.7H2O--- 0.25g glucose--- 2.0 g bring to pH 7.0 before autoclave	3&46 (add .25 gm Yeast Extract these 2 exp. only); 4 (add 39 μmoles tryp exp. 4 only),11,39
Danish Prep--Dugway	Prepared under contract for Dugway with following recipe: Marcor Inc peptone HCT (a hydrolyzed protein digest from pork) 6.0g; Amberex 1003 (yeast extract—Sensient Technologies) 3.0g; Antifoam, Pluronic ~0.3 gm $MgSO_4 .7H_2O$—0.31g; $MnSO_4 .1H_2O$—0.08g; $CaCl_2.2\ H_2O$—0.16g; $K_2\ HPO_4$--0.16g; Dextrose (autoclave separately)—6.0 g pH adjusted 6.8 –7.2 w NaOH or Sulfuric acid before autoclave. Grown in fermenter with aeration.	9,10,15,17,18,19,28,29, 41,42,

Table 2. Media Used in Experiments (Continued).

Growth Medium	Recipe (per liter distilled H_20)	DataSet No.
Leighton-Doy Also called 2XSG in P.Setlow papers	Difco Nutrient Broth 16 g 1 M $MgSO_4$ 2 ml 2 M KCl 13 ml 1 M $MnCl_2$ 100 ul 0.36 M $FeSO_4$ 3 µl H_2O 970 ml For plates add 15 g/L agar. Autoclave, then add sterile 50 x $Ca(NO_3)_2$.Glucose------20 ml 50 x $Ca(NO_3)_2$.Glucose $Ca(NO_3)_2.4H_2O$ 1.18 g Glucose 5 g H_2O to 100 ml **For DPA plus plates, add** 200 µg/ml filter sterilized DPA before pouring plates (1.2mM in plate)	13,22,23,24,25,26,27, 33,34, 44,
DSM medium—from A. Driks	Difco Nutrient Broth 8 g 1.2% MgSO4 10 ml 10% KCl 10 ml 1N NaOH 0.5 ml Cool, then add Autoclaved Seperately components 1ml each. then add autoclaved supplements individually just before use: 1 M Ca(NO3)2; 0.01 M MnCl2; 1 mM FeSO4	47
SNB (supplemented nutrient broth)	Difco Nutrient Broth 8 g SNB salts 8 ml Bactoagar (Difco) 15 g Add 980 ml H2O autoclave, cool, Add sterile Ca-glu soln 20 ml **Ca-glu solution:** 0.5 M CaCl2 10 ml Glucose 5 g H2O to 100 ml **SNB salts** 0.5 M FeSO4 0.28 ml 1 M MnCl2 2 ml KCL 100 g MgSO4.7H2O 25 g H2O to 800 ml	37

So far, variations in the growth medium as well as the final washes and optical density before the luminescence experiment have not affected the classification results in the cases where we have varied the preparation for a single strain of bacteria.

We subjected the bacteria to a final wash for all experiments (usually there were two washes). The wash, centrifugation, and final suspension were with filtered deionized H_2O or with 0.9% NaCl solution. These were tested periodically and did not have detectable fluorescence.

The spectra were taken from a spot of 0.2 ml of particle suspension dried onto a filter. The filter used had negligible fluorescence. The suspension from which the spot was made was adjusted to ~ 0.03 to 0.06 mg of spores or roughly 5 x 10^5 colony forming units (cfu) for BG spores several years old. The spot was roughly circular about 5 to 8 mm in diameter. For vegetative cells (freshly grown), OD_{600} was adjusted to the range 0.1 to 0.3 and about 2 to 4 x 10^7 cfu in the spot.

Spots were formed and dried before measuring fluorescence for the "Before UV" sample and the "After UV" sample. The UV irradiation was given in two ways. In the protocol used first (P1), Data sets 1—17, the spot itself was irradiated in the fluorometer at excitation 270 nm with 1 mm excitation slits for ~38 min. A rough measure of the effectiveness of this irradiation on BG spores was made by following disappearance of the tryptophan fluorescence. This showed this method roughly equivalent to be about 10% less effective than the same time of exposure with our UVC lamp. Irradiation protocol P1 has the advantage that the Ex-Em "After UV" graph is measured on exactly the same cells as are measured for the "Before UV" graph. A second protocol (P2), was used for samples with a Data set numbered greater than 17 with the exception of Ex-Em for some liquid exposures—Data sets 30,51,53,54—taken only Before and in a quartz cuvette. In protocol P2, ~3 x 10^4 J/ m^2 was given during a 60 min exposure by a lamp emitting UVC (predominantly 254 nm) light to the cells in a quartz cuvette. This dose is lethal to the bacteria. This had the advantage of a more accurate measure of the dose given, but a separate spot was formed and dried for Ex-Em measurements after the UV. This made the number of cells exposed for excitation before and after only equivalent to ~50%.

Fluorescence measurements for this report were made on spots dried from suspensions onto nonfluorescent filters as described above. The instrument used was a Spex Fluorolog-2 Spectrofluorometer equipped with double grating excitation and emission spectrometers. Two excitation and two emission slits were all opened to 1 mm, giving an excitation bandpass of 1.70 nm and an emission resolution of 3.40 nm. The data was taken at excitation intervals of 10 nm, between 260 and 450 nm, and emission intervals of 5 nm between 300 and 450 nm. The UV irradiation during measurement of Ex-Em graphs gave a measurable effect for the time spent with the 1 mm slit open for excitations of 290 nm and less. An illustration of the effect of irradiation during the scan is shown in Figure 2. One sees that fluorescence at excitations near 350 and 370 nm becomes much increased due to the result of the previous scan only. To minimize this effect, the fluorometer was programmed so that all the longer wavelength excitations where most of the important luminescence occurs were taken prior to exposure to the short wavelength excitations.

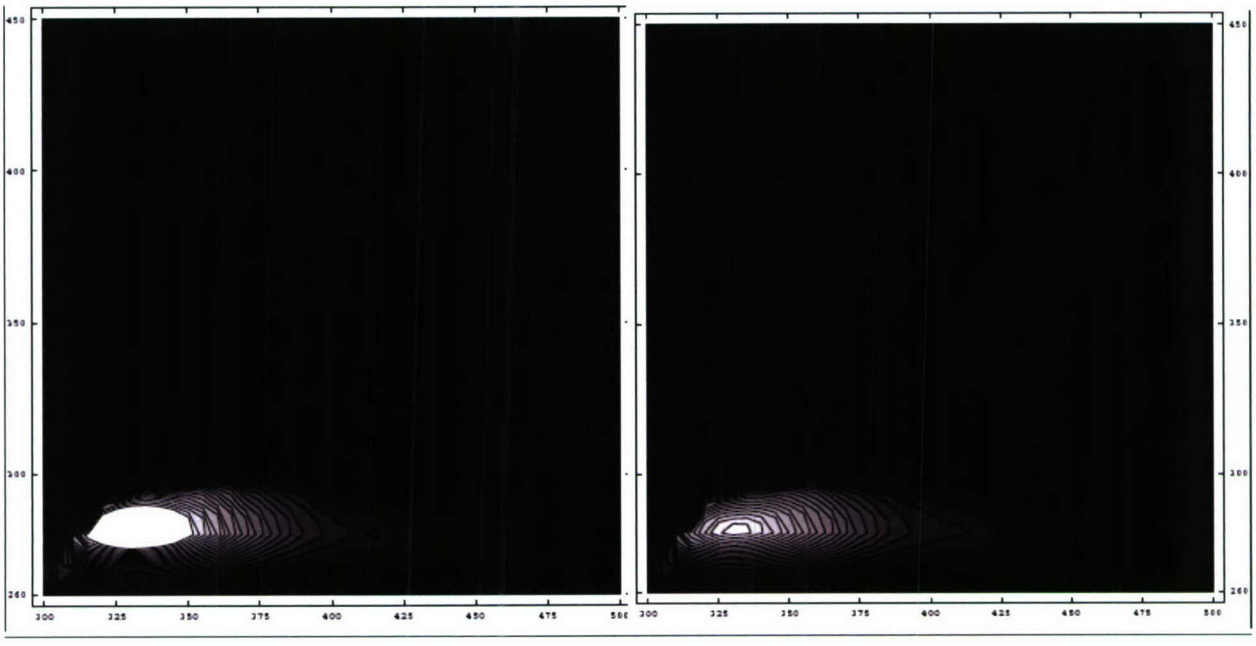

Figure 2. Ex-Em graphs for *Staphylococcus epidermidis* cells on a fluorescence free filter. Left graph is first scan with 1 mm excitation slit where exposure to wavelengths below 300 nm lasted ~3 min. Graph on Right shows the second scan taken with long wavelength excitations first. The only prior UV exposure was during the first scan. A comparison shows substantial changes due to UV exposure during the first scan. On all Ex-Em graphs the vertical axis gives excitation wavelength.

3. RESULTS

Some of the results using protocol P2 are shown for before and after UV in the graphs below. All the following graphs showing Ex-Em data have the same scale before and after UV. In Figure 3, the Ex-Em graphs are shown for an overnight growth of *Escherichia coli* B/r in rich medium.

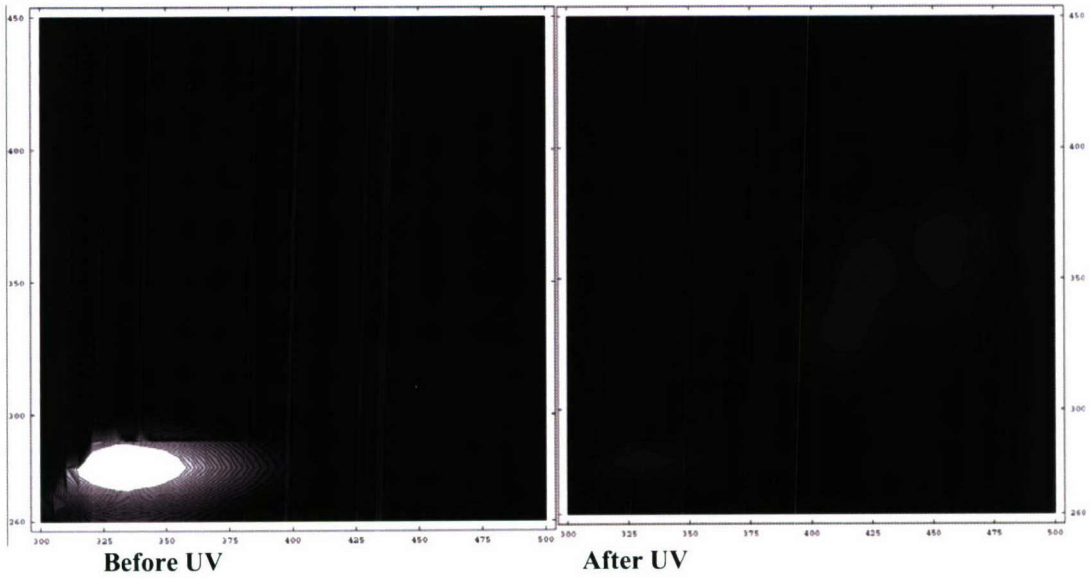

Figure 3. Ex-Em graphs for *E. coli* grown overnight in rich medium

The excitation axis is vertical and ranges from 260 nm to 450 nm. The horizontal emission axis ranges from 300 to 500 nm. The Ex-Em graph for *E. coli* seen in Figure 3 has a quite modest change in fluorescence for excitations near 350 and 360 nm in contrast to the change shown in Figure 1 for DPA+ *B. subtilis* spores. A similar change for *Bacillus spores* of a different species is shown in Figure 4.

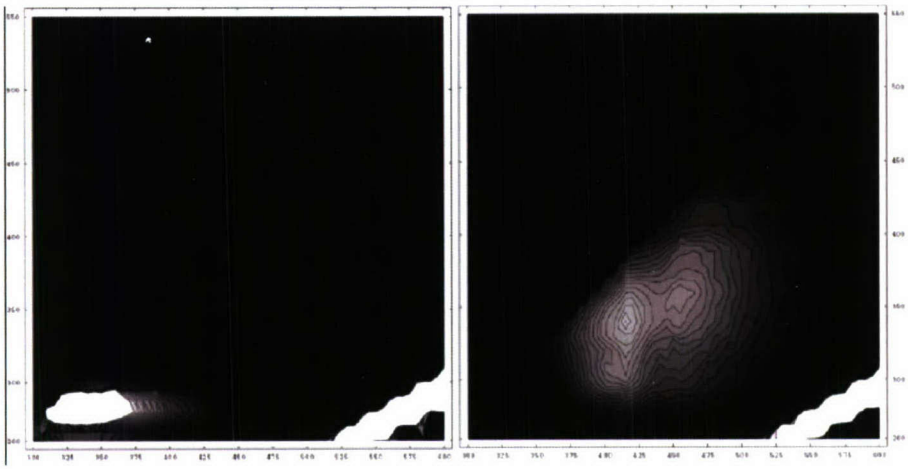

Figure 4. *Bacillus thuringiensis* spores. Left is "Before UV". Right is "After UV".

The change in contrast between the long wavelength emission and that for the tryptophan emission at 280 nm again appears much greater for the spore Ex-Em graph shown in Figure 4, than for the *E. coli* graph shown in Figure 3. We will show that an algorithm for automated recognition of these differences can be obtained with the use of pattern recognition techniques later in this section.

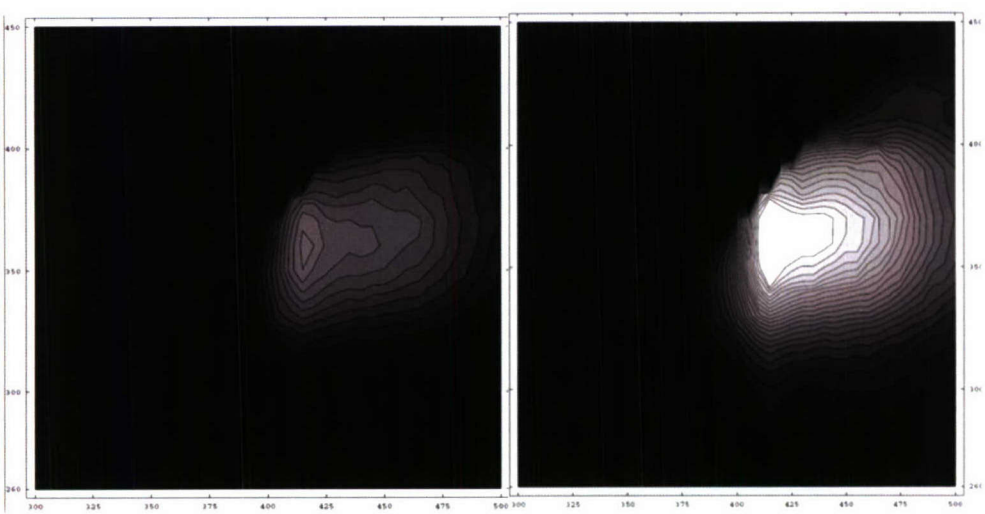

Figure 5. Ex-Em graphs for a sample of dust from a house in Highland, MD. Left is Before UV. Right is After UV.

The graphs shown in Figure 5, give one example of a possible background Ex-Em measurement. This graph (as with other background graphs examined) is quite different from those for endospores (Figures 1 and 4), for Gram negative (GN) bacteria (Figure 3), and for Gram positive (GP) vegetative bacteria (Figure 2); whereas, each of the graphs for bacteria appears similar to other graphs for the same class and dissimilar to graphs of other classes. We could take ratios of the emissions at various wavelengths and arrive at a fairly simple method of discriminating between the classes considered here (i.e., GN bacteria; GP bacterial spores; GP vegetative bacteria; various background materials likely to be found in aerosols). It is preferable to take a more general approach provided by modern pattern recognition techniques. In the following section, we explore this option.

4. PATTERN RECOGNITION APPLIED TO BACTERIAL LUMINESCENCE

4.1 Brief Non-Expert's Introduction.

We start with a PARAFAC type of analysis. Excitation-emission scans (Ex-Em or EEM) from a specific sample naturally form two dimensional matrices with zeros for emission at wavelength less than the excitation wavelength and zero entered for second order emission values. For each specific sample, we have M excitation values (rows index m) and N emission values (columns index n). The Ex-Em data thus form M by N rectangular matrices. Suppose we have P experiments. If we take "Before" UV exposure scans and "After" UV exposure scans as separate experiments, we then have 2 x P = P' of these matrices. We can arrange these P' matrices in a stack like a deck of cards.

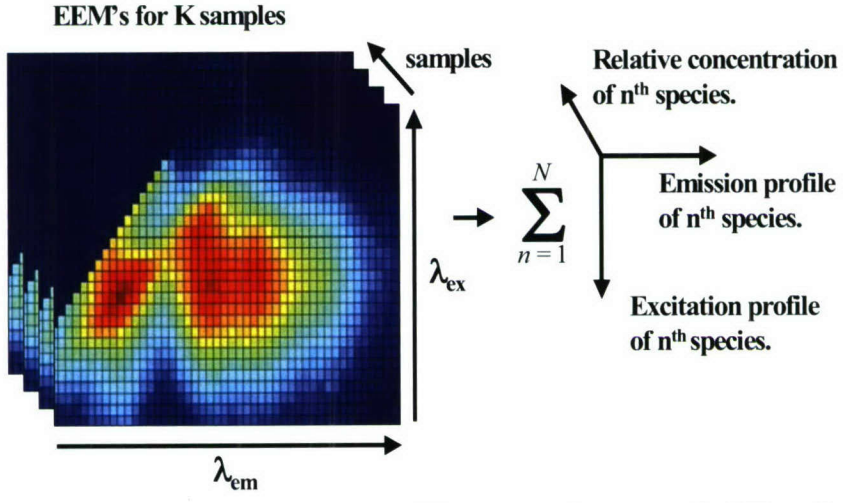

Figure 6. Notional sketch of Ex-Em matrices for different experiments stacked like a deck of cards.

This is shown in Figure 6. ParaFac analysis has typically been used by chemists to determine the amount of several known chemicals mixed in unknown proportions and with an unknown background. Consider the case for three chemicals. It is straight-forward to construct a three dimensional plot of excitation, emission, and concentration from laboratory measurements for the three chemicals. These graphs are then used to analyze a mixture for the concentration of the

three chemicals in initially unknown proportion and with one or more unknown contaminants for their relative concentration in the mixture. The effectiveness of this approach has been demonstrated in the lab of one of the authors in a case where UV photodegradation was applied to chemicals in a manner similar to the way microorganism exposure to UV was used in the present project. The approach was successfully applied to the photolysis Ex-Em spectra of pesticide[11] and polycyclic aromatic hydrocarbons.[12]

The present problem is related to the above, but differs in an important aspect from the above. We start with laboratory measurements of a number of known specific microorganisms. However, we may not know exactly the condition of the constituent chemicals giving rise to the luminescence. Gram positive spores or GP vegetative bacterial cells or GN bacterial cells have a fairly well-known chemical makeup. However, cells of a particular bacterial species have localized structure which affects how these chemicals respond to the excitation. Hence, any one chemical may fluoresce with observable differences in two different microbes, depending on its local environment. For such reasons, it is not possible at this point to assign all hot spots of luminescence to the spectra of single chemicals.

In the present case considering microorganisms, the vertical axis of Figure 6 gives Excitation (index i), the horizontal axis corresponds to Emission (index j), and the axis into the page is the index indicating the measurement (index k), with Before UV and After UV labeled as separate experiments. The initial stack appears smoother and is more compatible with a fit if we arrange the deck like a new deck of cards so that similar samples are grouped and normalized so absolute values of entries are similar. Since we are studying known preparations at this time, this may be accomplished from a-priori knowledge and inspection. Since fluorescence spectra are usually smooth, we could do a three dimensional smoothing of the data.

4.2 PARAFAC Analysis.

The PARAFAC approach applied to the present problem derives a fit of all the spectra in the stack with the two dimensional spectra of a small number of latent factors representing "surrogate chemicals" or "pseudo-chemicals". The spectrum of each of these "pseudo-chemicals" differs from card to card in the deck only as to concentration. The Ex-Em spectrum of a particular "pseudo-chemical" may correspond to that for an actual chemical constituent of cells of a particular species, but does not necessarily do so. One well known center of luminescence, the location of which corresponds almost directly with that of a known chemical, is the peak for the amino acid, tryptophan, which almost always appears in the Ex-Em spectrum for bacteria at excitation near 280 nm. The number of factors, N, was in the present case, selected from the set 3, 4, ...8. The actual number finally decided on was determined from the number giving the best results for the analysis using linear combinations of the Ex-Em graphs for the N factors.

The PARAFAC program then iteratively, starting from random entries, develops N three dimensional contour graphs for each of these factors which added together in appropriate proportions fit each graph contained in the stack of Figure 6. The fourth dimension, the concentration of each pseudo-chemical in a given experiment, corresponds to the relative

luminescence contributed by that chemical to a given matrix in the stack. The sum of these for all N factors is fit to the smoothed data stack or three dimensional matrix.

Next, we decide on the number of classes, C, for which we are testing the experimental data. (e.g., Gram positive vegetative bacteria, Gram negative bacteria; endospores;... etc.) and determine whether the best number of pseudo-chemicals for the whole set of experiments is able to characterize each of these classes. If not, we try again one more pseudo-chemical and check again.

The best fit for trials with different numbers of factors occurred for N = 5. Each of these five pseudo-chemicals may occur in different proportions for each k value in the stack shown in Figure 6. The contour plots in an Ex-Em diagram of contour plots for each of the five pseudo-chemicals are shown in Figure 7, where two of these are near a tryptophan location.

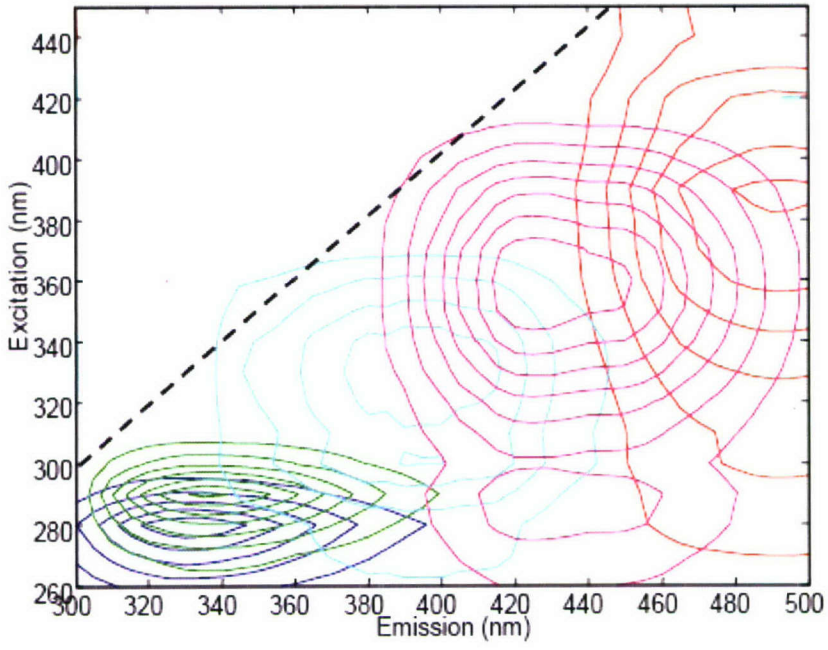

Figure 7. Ex-Em graph for the five pseudo-chemicals whose linear combinations give reasonable fit to matrix stack of Figure 6 for all the data.

The contours indicating concentration, are centered about peaks, which show large emission for each of the five pseudo-chemicals.

Now, we restate the approach to be used more precisely, assuming familiarity with the ideas above. We are using both Parallel Factor (PARAFAC) Analysis and Partial Least Squares – Discriminant Analysis (PLS-DA) to differentiate between the three classes of microbes: Gram positive vegetative bacteria, GN bacteria, and GP endospores and later discrimination from several common backgrounds. PARAFAC is used to extract spectral features common to most samples analyzed. The relative contributions of the extracted spectral features to each sample are used in the PLS-DA model to distinguish among the three classes. A

nested PLS-DA model is used. First, a model is built to distinguish Gram positive from Gram negative microbes. Then, a second model is applied to just the Gram positive microbes to distinguish between vegetative bacteria and endospores.

Ex-Em spectra, both before and after UV photo-conversion, were collected for 37 microbial samples: 6 GP vegetative bacteria, 14 GN bacteria, and 17 GP endospores. The spectra were collected at 20 excitation wavelengths from 260 nm to 450 nm at 10 nm resolution and 41 emission wavelengths from 300 nm to 500 nm at 5 nm resolution. The spectra were formed into a 20 x 41 x 74 three-dimensional data cube (Figure 6). The 'Before UV exposure' and 'After UV exposure' Ex-Em spectra or each sample are treated independently as unique objects in this **PARAFAC** analysis. However, the data could be formed into a 20 x 41 x 37 x 2 four-dimensional cube and equivalently analyzed by a 4-way PARAFAC model.

The PARAFAC model assumes that there is a finite set of N fluorophores (or pseudo-chemicals), that contribute to the Ex-Em spectra of all 74 samples. Each of these N fluorophores will have the same excitation profile and emission profile in each sample; the only change will be the relative concentration of the N fluorophores throughout the 74 samples. The outer product of the n^{th} resolved excitation profile and n^{th} resolved emission profile presents the extracted Ex-Em spectra of a given fluorophore that contributes to the overall Ex-Em spectra. Figure 7 shows 5 resolved Ex-Em spectra that were extracted from the 74 samples by PARAFAC analysis. The PARAFAC model provides N sets of three vectors: an excitation spectrum, an emission spectrum, and a 74 element long vector containing the relative contribution of the n^{th} fluorophore to each of the 74 samples. Thus, if the correct value of N is chosen, the data is reproduced by the sum of the outer products of these N triads.

To fit the PARAFAC model to the collected data, a weighted PARAFAC algorithm was used. The weighted algorithm assigns weights of zero to Ex-Em wavelengths containing Rayleigh scattering and Ex-Em wavelengths where the emission energy is less than or equal to half the excitation energy. All other Ex-Em wavelengths are assigned a weight of 1. Based on this algorithm, PARAFAC models using from N = 1 to N = 8 factors are constructed. Based on fit of the PARAFAC models to the data, the model with N=5 was found to be best. The resolved Ex-Em spectra from the 5 factors of this model are shown in Figure 7. The relative contributions of these 5 Ex-Em spectra extracted with the PARAFAC model were used for the PLS-DA model below.

We recognize that there are likely to be more than 5 different fluorophores present in the microbes. However, at the level of sensitivity accepted for the present experiments, the additional fluorophores would give rise to patterns indistinguishable from experimental noise. At the same time, a single fluorophore may occur in several different environments within a class of microbes, and its Ex-Em spectrum could appear to be two different pseudochemicals. While the 5 resolved profiles may not all represent identifiable chemicals, they do provide a solid description of the Ex-Em spectra from which to determine class differences. One further note on the experimental data. These data have not yet been corrected for the spectrum of the xenon lamp in the fluorometer used. Since there will be a one to one correspondence of the corrected spectra with the uncorrected spectra, this is not expected to affect the separation of classes, but the Ex-Em graphs for the data and the pseudo-chemicals will change their appearance.

4.3 Partial Least Squares-Discriminant Analysis (PLS-DA).

The PLS-DA is analogous to Partial Least Squares Regression (PLSR). Where PLSR is the inverse least squares formulation of multiple linear regression (MLR), PLS-DA is the inverse least squares formulation of Linear Discriminant Analysis (LDA). The PLS-DA has the same error reduction and variable selection advantages over LDA as PLSR has over MLR.

4.3.1 Model 1: Differentiating GN and GP Samples.

In PLS-DA, samples within the target class are assigned a value of 1, and samples external to the target class are assigned a value of 0. A PLS model is built to predict the assigned value for each sample. There are two parameters that must be optimized for PLS-DA. The optimal number of factors in the model is found by cross validation to best predict the 'score' values of 0 or 1, which were assigned to the samples. A cut-off value is found by Bayesian statistics applied to the distribution of 'score' values such that a sample achieving above the cut-off has >50% chance of truly being included in the target class.

To use PLS-DA, 12 new variables were created from the five factors extracted with the PARAFAC model. To have an accurate value for the UV dose (see Section 2), we chose the method, which required separate spots to be measured for the Before and After spectrum for most of the samples. This allowed the possibility for substantial variation of the number of bacteria in the excitation light between Before and After measurements for one preparation. Thus, instead of comparing absolute concentrations for the linear combination of pseudo chemicals, which fit a given experiment, we took ratios of the concentrations for each of the other four factors to the "tryptophan" factor for the before and after sample separately. The tryptophan factor was designated as that with its excitation peak closest to 280 nm. A 37 x 12 matrix is formed and each of the 37 samples is associated with a class value of 0 or 1 for PLS regression.

Each of the 37 bacteria is associated with two Ex-Em spectra: 'Before UV exposure' and 'After UV exposure'. Thus, for the 5 pseudo-fluorophores extracted by PARAFAC, there are 4 Ratios (i.e., concentrations relative to the tryptophan concentration) associated with the 'Before' spectrum and 4 Ratios with theAfter. This yields 8 new variables; 4 from the 'Before' spectra and 4 from the 'After' spectra for each sample. The remaining 4 new variables are constructed by calculating the ratio of the normalized factors between the 'Before' and 'After' spectra (i.e., ratio of ratios). We call the above 12 ratios the Ratio Variables (RV). We then determine the subspace of the RV space in which the most variability or best separation is exhibited between the Gram positive (GP) and Gram negative (GN) sets. In doing this, we construct orthogonal m dimensional subspaces, first of dimension 1; then 2; then 3;.....12 in which the two data sets, GP and GN show the most separation from each other. Each axis in the m dimensional subspace is formed of a linear combination of the 12 RV or latent variables, and is orthogonal to the preceding m-1 dimensional subspace. This continues from m = 1 through m = 12.

In Figure 8, we use PLS-DA to determine the total variance of the GP samples from their predicted value.

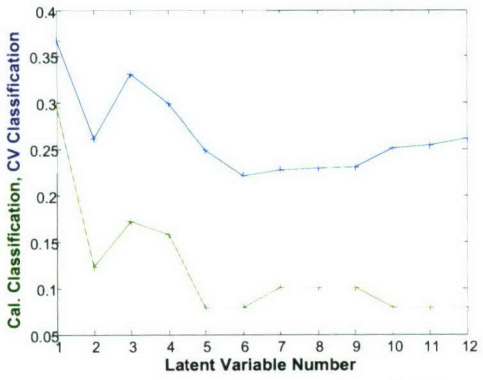

Figure 8. Variance of separation of GP from GN microorganisms.

The PLS-DA was then used to differentiate the GP microbes as a single class from the GN microbes. The result from Figure 8 was that a 6 latent variable (i.e., m=6) PLS-DA model was found to be optimal (least variability from expected value) based on the root mean squared error (RMSE) of classification for the PLS-DA model (Figure 8, green, lower curve) and RMSE from leave-3-out-crossvalidation (Figure 8, blue, upper curve). The meaning of the crossvalidation test is that the program leaves out 3 randomly chosen data sets (12 times) while evaluating the error. Leaving out a small number of data sets lets us determine if one or several of these sets has had too much effect on the final result. It also provides a better feel for how a model will perform on future samples. The 6 latent variable model (RV) corresponds to a minimum in both of these two curves. With this model, 95% of the variance in the X-block (measured variable block) and 57% of the variance in the Y-block (predictor variable block) is captured.

The 6-factor PLS-DA model achieved a 96% classification rate for GP microbes (22 of 23) and a 93% classification rate for GN microbes (13 of 14). Figure 9 presents the predicted scores for the GP and GN microbes for the 6 latent variable PLS-DA model.

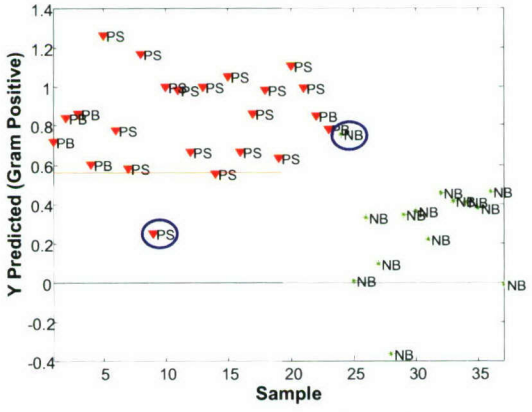

Figure 9. GP (values above red line) vs GN assignment. Samples were renumbered (horizontal axis) and are different from tables.

The GP samples are represented by red triangles and labeled either 'PB' for GP vegetative bacteria or 'PS' for GP spores. The Gram negative bacteria are labeled NB. A cut-off value of 0.58 was determined to differentiate between the two classes. Samples with a score greater than 0.58 would be determined to be GP. Samples with a score less than 0.58 would be classified as other than GP. In reality, such samples could be either GN or just random background sample fluorescence. However, because no environmental background spectra were included in this preliminary analysis, we are realistically performing a binary classification between GP and GN microbes. The sample numbers in Figure 9 and the other Figures below are from the same data, but with different numbering from Tables 1 and 2.

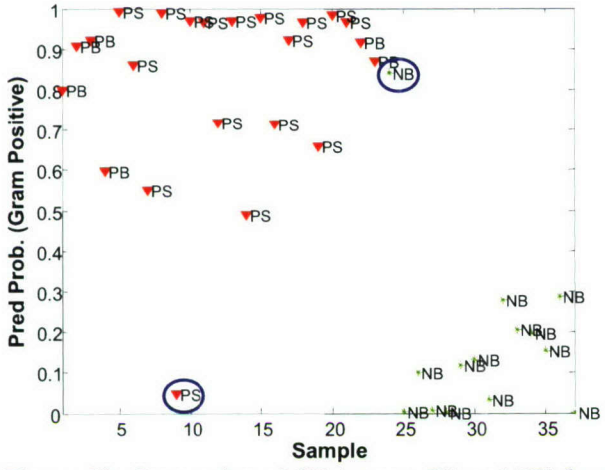

Figure 10. Separation of GP (upper, PS and PB) from GN bacteria (NB).

Bayesian statistics can be used to convert the predicted scores on the Y-axis to probabilities of each sample belonging to the GP class (Figure 10). Samples with a score greater than 1 are capped at a 100% probability of being GP. Samples with a score less than 0 are assigned a 0% probability of being GP. Only 2 samples were misclassified (blue circles). One GP sample was classified as being GN, and one GN sample was misclassified as being GP. The reasons for these misclassification are under investigation. We note that since many different culture and preparation conditions were used. this is a source of variability in the analysis.

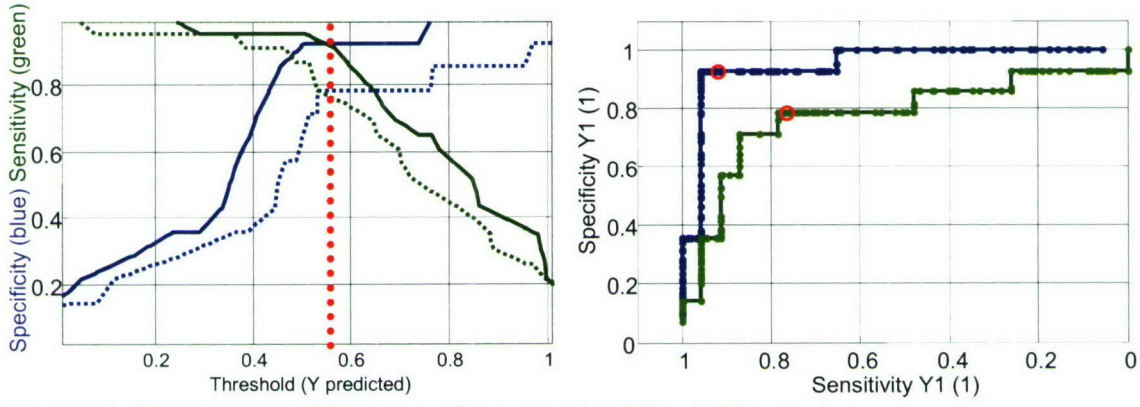

Figure 11. Threshold and ROC curves for separating GP and GN samples.

The performance of the classification model can be seen in the threshold (Figure 11, left) and ROC curves (Figure 11, right). The threshold graph shows the effect on sensitivity (green, descendng to right--the lower the sensitivity, the more GPs are missed) and specificity (blue, ascending to right—the higher the specificity, the less GNs are included as GPs) of the model of choosing the cut-off value (threshold). These figures of merit are presented for the model applied to all the data (solid lines) and estimated values from leave-3-out cross-validation (dashed lines). The cross validation figures of merit are believed to more accurately predict future performance of the model than are the figures of merit from self-fit. The Y-axis presents the sensitivity of the model as the fraction of GP samples correctly classified and the specificity of the model as the fraction of GN samples classified as not being GP. The vertical dotted red line is the cut-off chosen for the model.

Increasing the cut-off threshold increases specificity the at the expense of decreasing the sensitivity. Similarly, decreasing the cut-off threshold increases the sensitivity at the expense of the specificity. The trade-off between the sensitivity and specificity at different thresholds is seen in the ROC curve (Figure 11, right), which is obtained simply as a parametric evaluation of sensitivity and specificity for each threshold value. The blue line is the ROC curve for the fit of the model to the 37 training samples. The green line is the ROC curve based on cross validation. The red circles are the locations along the ROC curves of the threshold value shown in the previous plots. Taken together, the threshold and ROC curves indicate that although ~95% specificity and sensitivity were observed based on fit of the model to the training set, 80% sensitivity and specificity are predicted for future samples being applied to this model. However, the data used here are preliminary and were collected under a variety of culturing and processing conditions.

4.3.2 Model 2: Differentiating GP Spores from GP Bacteria.

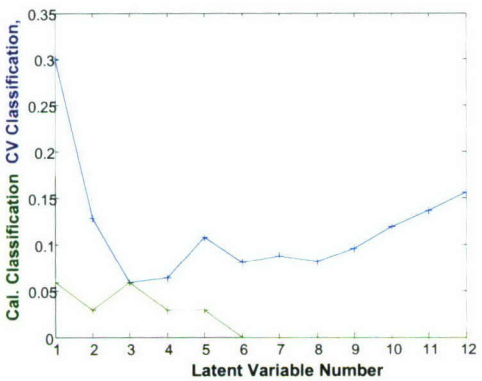

Figure 12. Performance for separation of GP spores from GP vegetative bacteria.

A second PLS-DA model was constructed to differentiate among the two classes of GP microbes. The 6 GP vegetative bacteria and 17 GP spores were used as a training set. The RMSE of calibration (Figure 12, green, lower graph) and RMSE from leave-2-out cross validation (Fig. 12 upper graph) indicate that a 4-factor PLS-DA model be used. While the cross validation performance of the two models was very similar, the 4- factor model was chosen over the 3-factor model based on performance in fit to the training set. However, it is recognized that a 3-factor model may, in fact, prove slightly more robust with future analyses.

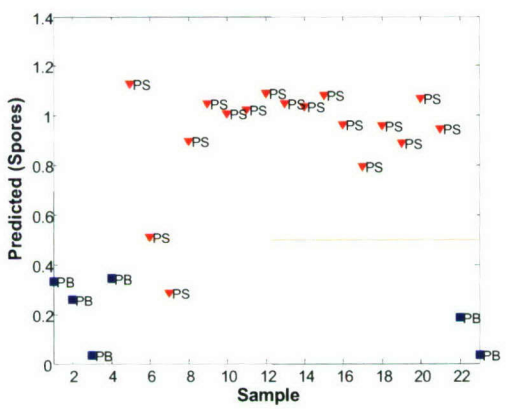

Figure 13. Separation of GP spores (upper, PS, red) from GP vegetative bacteria (PB, lower, blue).

Figure 13 presents the scores for classification between GP spores and bacteria using 4 latent variables. Figure 14 presents the same data converted to probability of classification as a GP spore. No GP bacteria are classified as a GP spore (100%, 6 of 6) and only 1 GP spore is misclassified (94% correct, 16 of 17). The probability of inclusion of each sample as GP shows that besides the misclassified sample, only one other sample has a probability of classification between 5% and 95%. Most samples are very unambiguously and correctly classified as either spores or vegetative bacteria.

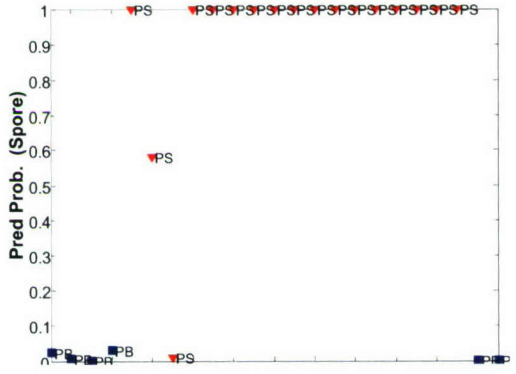

Figure 14. Probability of correct classification for GP spores vs GP vegetative bacteria.

The threshold and ROC curves, shown in Figure 15, predict better performance for PLS-DA differentiating between GP vegetative bacteria and GP endospores, than for differentiating between GP and GN microbes. Cross validation predicts 100% specificity and almost 90% sensitivity.

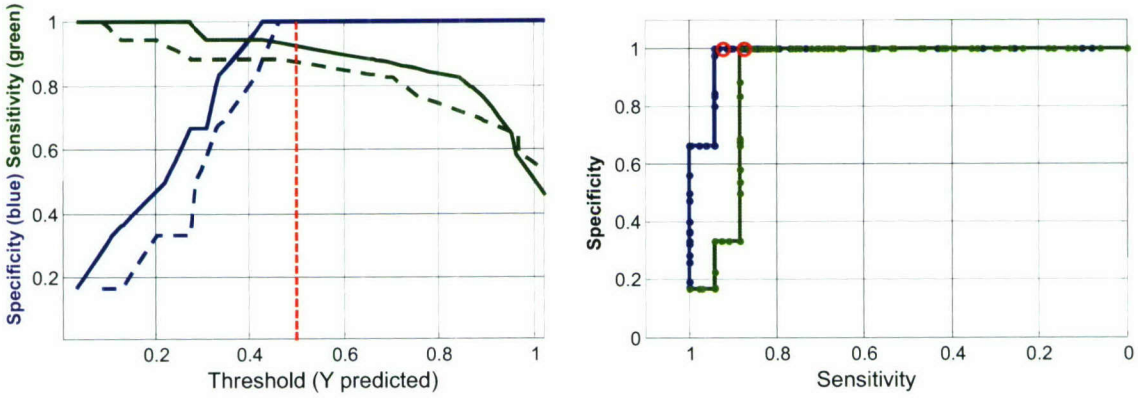

Figure 15. Threshold and ROC curves for differentiating GP spores from GP vegetative bacteria.

5. CONCLUSIONS

The above analysis suggests that based on only one perturbation, i.e., exposure to a single UV dose, we can achieve good separation among several classes of bacteria. There are other additional perturbative physical treatments, which could be inexpensively incorporated into field instruments in a way in which further discrimination of microbial classes could be rapidly and automatically achieved.[13] It is not unreasonable to expect that incorporation of additional perturbations would allow separation of unknown biological particles into additional well-defined classifications. The analysis is potentially fast and direct.[13]

LITERATURE CITED

1. Reyes, F.L., Jeys, T.H., Newbury, N R., Primmerman, C.A., Rowe, G.S., Sanchez, A. "Bio-aerosol Fluorescence Sensor." *Field Analyt. Chem. Technol.* Vol. **3,** (4-5), pp 240-248 (1999).

2. Faris, G.W., Copeland, R.A., Mortelmans, K., Bronk, B.V. "Spectrally-Resolved Absolute Fluorescence Cross Sections for Bacillus Spores." *Appl. Optics* Vol. **36,** pp 958-967 (1997).

3. Ho, J., Spence, M., Hairston, P. "Measurement of Biological Aerosol with a Fluorescent Aerodynamic Particle Sizer (FLAPS): Correlation of Optical Data with Biological Data." *Aerobiologia* Vol. **15**, pp 1573-3025 (1999).

4. Nudelman, R., Feay, N., Hirsch, M., Efrima, S., Bronk B. "Fluorescence of Dipicolinic Acid as a Possible Component of the Observed UV Emission Spectra of Bacterial Spores." In *Proceedings of SPIE*, Vol. **3533**, pp 190-195, Air Monitoring and Detection of Chemical and Biological Agents. J. Leonelli and M.L. Althouse, Eds. (1999).

5. Nudelman, R., Bronk, B.V., Efrima, S. "Fluorescence Emission Derived from Dipicolinic Acid, Its Sodium, and Its Calcium Salts." *Appl. Spectrosc.* Vol. **54**, pp 445-449 (2000).

6. Bronk, B., Nudelman, R., Shoaibi, A., Akinyemi, A. "Physical Perturbation for Fluorescent Classification of Microorganism Particles." *In Proceedings of SPIE,* Vol. **4036**, pp 169-180, Chemical and Biological Sensing, Chair: Patrick J. Gardner (April 2000).

7. Sarasanandarajah, S., Kunnil, J., Bronk, B.V., Reinisch, L. "Two Dimensional Multiwavelength Fluorescence Spectra of Dipicolinic Acid and Calcium Dipicolinate." *Appl. Optics* Vol. **44**, pp 1182-1187 (2005).

8. Bronk, B.V., Reinisch, L., Setlow, P. "The Role of DPA in the Fluorescence of Bacillus Spore." In *6th Joint Conference on Standoff Detection for Chemical and Biological Defense.* CBIAC Report CB-193503, APG, MD (October 2004).

9. Sarasanandarajah, S., Kunni, J., Chacko, E., Bronk, B.V., Reinisch, L. "Reversible Changes in Fluorescence of Bacterial Endospores found in Aerosols due to Hydration/Drying." *J. Aerosol. Sci.* Vol. **36**, pp 689-699 (2005).

10. Paidhungat, M., Setlow, B., Driks, A., Setlow, P. "Characterization of Spores of *B. subtilis* which Lack DPA." *J. Bact.* Vol. **182**, pp 5505-5512 (2000).

11. Nahorniak, M.L., Cooper, G.A., Kim, Y.C., Booksh, K.S. "Three- and Four-Way Parallel Factor (PARAFAC) Analysis of Photochemically Induced Excitation-Emission Kinetic Fluorescence Spectra." *Analyst*, Vol. **130**, pp 85-93 (2005).

12. Kim, Y.C., Jordan, J.A., Nahorniak, M.L., Booksh, K.S. "Photocatalytic Degradation-Excitation-Emission Matrix Fluorescence for Increasing the Selectivity of Polycyclic Aromatic Hydrocarbon Analyses." *Anal. Chem.* Vol. **77**, pp 7679-7686 (2005).

13. Nahorniak, M.L., Booksh, K.S. "Optimizing the Implementation of the PARAFAC Method for Near-Real Time Calibration of Excitation-Emission Fluorescence Analysis." .J. *Chemometric.* Vol. **17,** pp 603-608 (2003).