

Defining and Using Microbial Spectral Databases

Jon G. Wilkes, Katherine L. Glover, Manuel Holcomb, Fatemeh Rafii, Xiaoxi Cao, and John B. Sutherland

National Center for Toxicological Research, Jefferson, Arkansas, USA

Susan A. McCarthy

Gulf Coast Seafood Laboratory, Food and Drug Administration, Dauphin Island, Alabama, USA

Simon Letarte and Michel J. Bertrand

University of Montreal, Montreal, Canada

This work shows how fingerprints of mass spectral patterns from microbial isolates are affected by variations in instrumental condition, by sample environment, and by sample handling factors. It describes a novel method by which pattern distortions can be mathematically corrected for variations in factors not amenable to experimental control. One uncontrollable variable is "between-batch" differences in culture media. Another, relevant for determination of noncultured extracts, is differences between the cells' environmental experience (e.g., starved environmental extracts versus cultured standards). The method suggests that, after a single growth cycle on a solid medium (perhaps, a selective one), pyrolysis MS spectra of microbial isolates can be algorithmically compensated and an unknown isolate identified using a spectral database defined by culture on a different (perhaps, nonselective) medium. This reduces identification time to as few as 24 h from sample collection. The concept also proposes a possible way to compensate certain noncultured, nonisolated samples (e.g., cells concentrated from urine or impacted from aerosol or semi-selectively extracted by immuno-affinity methods from heavily contaminated matrices) for identification within half an hour. Using the method, microbial mass spectra from different labs can be assembled into coherent databases similar to those routinely used to identify pure compounds. This type of data treatment is applicable for rapid detection in biowarfare and bioterror events as well as in forensic, research, and clinical laboratory contexts. (J Am Soc Mass Spectrom 2002, 13, 875–887) © 2002 American Society for Mass Spectrometry

Defining and Using Microbial Spectral Databases. The work described in this paper investigates a variety of issues revolving around a single question: The possibility of assembling and consulting databases of mass spectral patterns useful for rapid subspecies level identification of bacteria. The subject requires a rather diverse background presentation that may be summarized in seven phrases. Below we shall introduce:

1. Some of the advantages that would accrue from the development of workable databases;
2. an example from molecular biology that illustrates the degree of difficulty;
3. the potential of pyrolysis mass spectrometry (PyMS) methods to provide spectral patterns of sufficient

chemotaxonomic power to achieve rapid subspecies level identification;

4. the practical advantages of what we name coherent database methods compared to operational fingerprinting methods of using PyMS for chemotaxonomy;
5. the necessity for spectral drift compensation to assemble or consult a coherent database of microbial isolate patterns;
6. previously unrecognized effects limiting the validity of spectral drift correction; and finally the point of this paper, which is . . .
7. presentation of general protocols for addressing and correcting these effects so that coherent databases and rapid bacterial identification by MS methods can become a practical reality.

Published online May 20, 2002

Address reprint requests to Dr. J. G. Wilkes, Mass Spectrometry Branch, Chemistry Division, National Center for Toxicological Research, USFDA, 3900 NCTR Drive, Jefferson, AR 72079, USA. E-mail: jwilkes@nctr.fda.gov

Mass spectrometric, chromatographic, and many other bacterial identification methods share a requirement that bacteria be cultured and isolated, selectively

enriched or concentrated, purified and/or treated, or that their genetic material be extracted and amplified, before information can be gleaned for identification purposes. Such steps are time consuming and expensive. Much of the time/cost of sample handling and instrumental data acquisition can be reduced if only the one day's samples need to be handled. Development of a database containing representative spectra, chromatograms, gel patterns, etc. to compare with those acquired on a particular day would save an immense amount of time and labor. Any practical taxonomic method should be capable of generating a database of standard patterns. It should also be possible to evaluate each day's corresponding patterns, even when they are determined under experimental circumstances somewhat different from those used in assembling the database. We shall now illustrate some of the challenges and issues associated with rapid bacterial identification using as examples a group of techniques competitive with mass spectrometric approaches.

A variety of molecular genetic methods are currently available for rapid bacterial identification and taxonomy. They have become the methods of choice for many applications in diagnostic microbiology. However, Marshall et al. [1] showed that in at least one situation several of the genetic methods have limitations when applied to clinical or environmental samples isolated from disease outbreaks. The four methods they compared were enterobacterial repetitive intergenic consensus sequence-polymerase chain reaction (ERIC-PCR); flagellum locus restriction fragment length polymorphism (*Fla*-RFLP); pulsed field gel electrophoresis (PFGE); and ribotyping. For various reasons, the four methods classified the same isolates into differing numbers of class clusters. Marshall et al. [1] found that "...genetic changes that alter PFGE patterns can occur during the course of an outbreak." That is, mutations generated PFGE pattern variations of little or no therapeutic or diagnostic significance. Similarly "...rapid change in ERIC-PCR types could also obscure relationships between strains..." They found that *Fla*-PCR distinguished fewer classes than the other methods. We infer (though they did not specifically conclude this) that *Fla*-PCR did not provide sufficient taxonomic strength for the application. They concluded that "...[because of] the lower rate of change of ribotype patterns, [ribotyping] may be useful for evaluating the links between environmental strains and isolates from patients." However, ribotyping was very expensive and yielded identification only after several days' labor. Of the four methods that Marshall et al. [1] evaluated, only ribotyping might potentially produce stable patterns suitable for a database archive.

Like ribotyping, the instrumentation used in MS-based approaches is expensive. Therefore, the only way to justify MS for routine microbial taxonomy is through a combination of subspecies-level specificity (a requirement for excellent taxonomic power), rapidity of analysis (an important therapeutic and forensic advantage),

and amenability to automation (for reproducibility and economy). Use of mass spectrometric methods, such as PyMS, may be warranted if a coherent database can be assembled and consulted to yield subspecies-level identification.

PyMS has sufficient taxonomic power for distinguishing closely related, but nonidentical microbial isolates, e.g., antibiotic resistant and nonresistant strains of the same species. In microbial classification based on PyMS, the degree of specificity corresponds with therapeutically significant and/or forensically informative levels of diagnostic distinction. PyMS analysis differentiated similar isolates of the same bacterial species, i.e., isolates differing from each other based only on minor genetic mutations or serological variations: Wilkes et al. [2] used PyMS to distinguish *Vibrio parahaemolyticus* strains (the same species studied by Marshall et al. in 1999 [1]) from two outbreaks in which different *V. parahaemolyticus* serotypes were implicated. When isolates were cultured under the same conditions and analyzed on the same day using the same instrument, PyMS not only distinguished the two serotypes but also indicated small, significant differences between different strains of the same serotype implicated in a 1998 outbreak involving oysters from Galveston Bay, Texas. The latter observation paralleled Marshall et al.'s previously mentioned distinction of multiple strains implicated in the 1997 Pacific Northwest outbreak [1]. Goodacre and Berkeley [3] have reported the ability to distinguish small genotypic changes in *Escherichia coli* by PyMS.

Similarly, Sisson and coworkers [4–13] published multiple examples from clinical contexts in which PyMS was used for forensic purposes to elucidate vectors of nosocomial infections. In her work, samples from patients were isolated and then cultured under identical conditions. Pyrolysis mass spectra were acquired using an autosampler, and the database of spectra generated that day was used immediately to characterize patient isolates. Isolates could be grouped by pattern recognition conducted on this small spectral database. That is, spectral similarity was used to associate groups of patients who had acquired infections via the same vector after they were admitted to the hospital for other complaints. This work demonstrated the amenability to automation of MS-based taxonomic methods. However, Sisson's database was specific to her particular instrument as well as to the particular culture and handling conditions used on the day of analysis. Her methods could not be used for identification of samples generated by PyMS at different times or with different instruments.

Below we use the term coherent to describe a database that, in contrast to Sisson's, can be effectively consulted using spectra acquired in a variety of contexts. Previously, PyMS chemotaxonomy was limited to a less ambitious method named operational fingerprinting by Meuzelaar et al. in 1982 [14]. Sisson's work, even though it used a large capacity autosampler and in-

volved hundreds of samples, still exemplified operational fingerprinting.

To be justified by a cost/benefit criterion, routine use of mass spectrometry for rapid chemotaxonomy requires consultation of a database so coherent that it gives a reliable, subspecies-level identification of unknown isolates. To create such a database one must standardize instrumental tune and acquisition parameters, to maximize spectral reproducibility and instrumental sensitivity. One must also standardize sampling protocols as much as possible. If remaining instrumental or sample factors cannot be sufficiently controlled, one must develop a reliable way to compensate for and correct the resulting pattern distortions. A final, practical constraint is that all of this control and compensation be accomplished by automated processes operated by laboratory technicians rather than research scientists.

In 1996, Goodacre and Kell [15] published several algorithms for correcting PyMS spectra of bacteria for pattern distortions, spectral "drift", arising from instrumental variations. Their algorithms could be used effectively to compensate for instrumental drift as observed using one "known" isolate to track spectral variations of another of the same strain. They did not demonstrate that they could use spectral variations in a known isolate to track variations in bacteria of a different strain. Thus, they showed no practical way to drift-compensate spectra from isolates of unknown identity.

We modified one of the Goodacre and Kell [15] algorithms and used it to drift-compensate microbial PyMS spectra acquired at one time in an attempt to match others obtained two years earlier on the same instrument under nominally identical experimental conditions. We thought of the earlier spectra as a small database or library. Our algorithm appropriately corrected *E. coli* spectra using another *E. coli* spectrum acquired the same day in comparison to a typical *E. coli* spectrum in the two year old "library" (data not shown). However, the same algorithm failed when the *E. coli* reference spectrum was used to compensate spectra of other species (data not shown). We discovered that, in this case, a major contributor to spectral variation over the two years was the growth of bacteria on tryptic soy agar (TSA) from different suppliers. If the inability to compensate for differences in culture media was a general phenomenon and if analogous effects were also obtained for other uncontrollable experimental deviations, it might not be possible consistently to drift-compensate unknown isolates.

In this study, we explored the relationship between bacterial culture conditions and the ability to perform effective algorithmic compensation on the resulting spectra. Results suggested how the problem of tracking spectral variations can be solved to enable assembly and consultation of coherent microbial spectral databases. We believe the approach is suited not only for PyMS methods but also for other chemotaxonomically useful analytical techniques.

The results also suggested that the protocol may be used to correct spectra for some degree of variation in other environmental factors, e.g., duration or temperature of culture incubation. The results suggested a second protocol by which it might be possible to correlate, via a similar algorithmic compensation, spectra from cultured isolates to spectra from environmental extracts, thus avoiding analysis delays in emergency situations due to the requirement of culturing authentic standards for comparison. Finally, in support of the second protocol, they suggest a way to define an algorithmic transformation process ahead of the time when there is urgent need to identify a sample. This can allow for almost immediate identification of unknown microbial samples in certain situations. The ultimate goal of clinical diagnosticians is the identification of freshly isolated microorganisms within minutes of their arrival in the laboratory. [16]

Experimental

Experimental Overview

The experimental work comprised four sets. The first was intended to evaluate the kinds of bacterial PyMS spectral variability that might occur during a single analytical session, variability that could degrade patterns relative to standards in a database. In the second experimental set, spectral distortions for a test bacterium were produced in analytical sessions conducted several days apart. This was accomplished by intentional MS tune variations and the efficacy of a correction algorithm was evaluated by the resulting before- and after-correction locations on a display called a canonical variate (CV) score plot. The third experimental set introduced both instrumental tuning and sample culture medium as intentional sources of spectral variability and again used CV score plots to define the situations in which algorithmic spectral correction could successfully compensate for the variations. The results were eventually used to write protocols for defining and consulting microbial spectral databases. The fourth set was designed to further test the concepts on which the database building protocols for spectral compensation were premised. Some experimental details pertained to all or several of the experimental sets. These were described in more detail in four additional experimental subsections covering: (1) General bacterial culture conditions and sampling methods; (2) general PyMS method parameters; (3) the spectral compensation/correction algorithm; and (4) pattern recognition methods and the definition and uses of CV score plots for interpretation of results.

Set Number 1 Experimental Design

To model instrumental sources of pattern variability, we repeatedly acquired pyrolysis mass spectra for a concentrated isolate of suspended *E. coli* strain 1090.

The bacterial cells were grown on TSA from Difco (Becton, Dickinson & Co., Franklin Lakes, NJ). Enough bacterial cells were dispersed in 250 μL of 75% ethanol (HPLC Analyzed, J. T. Baker, Philipsburg, NJ) that the resulting suspension was opaque. Two μL of this concentrated suspension were dispensed from a syringe onto the wire loop tip of a direct exposure probe. After the droplet had dried on the wire, the probe was inserted into a Finnigan model 4500 mass spectrometer and heated as described later for sample pyrolysis. Pyrolysis 70 eV EI mass spectra were acquired and averaged (details found below). Early session spectra were compared to later examples considering particularly the changes in pattern distortion associated with continuous operation.

This vintage mass spectrometer was quite insensitive compared to today's commercially available systems. The F4500 ion source and optics were fairly contaminated from earlier work. The large sample loads used in these experiments along with the preexisting optics contamination were intended to model a worst-case scenario and to produce maximum pattern distortion in the course of a single analytical session.

Set Number 2 Experimental Design

E. coli 1090 was cultured on a variety of media: TSA from Difco, TSA from Remel (Remel, Ltd., Lenexa, KS), or sheep blood agar (5% sheep blood added to Remel TSA). A "library" spectrum for *E. coli* 1090 on Difco TSA was defined using PyMS acquired on January 14, 2000, referred to as Day 1. Typically, two analyses were done for isolates grown on each medium. PyMS acquisition conditions are described below. On the same day, replicate spectra for *Aeromonas hydrophila*, *Staphylococcus aureus*, and *Pseudomonas mendocina* strains were also acquired and added to a small database for use with experimental sets two and three. A few days later, the mass spectrometer was retuned to slightly different relative intensity specifications for the perfluorotributylamine (PFTBA) calibration/tuning compound. This retuning was performed to simulate instrumental relative intensity drift, not confusions in mass assignment. Newly cultured *E. coli* 1090 isolates, on each of the three media, were then reanalyzed using the same pyrolysis conditions. A spectrum of *E. coli* 1090 on one or the other of the three media types was selected as representative and, according to a multiplicative algorithm (details below), was used as a tracking reference in an attempt to adjust the day's other spectra to resemble the corresponding ones in the Day 1 Difco TSA-based "spectral library". The efficacy of these algorithmic manipulations was interpreted using CV score plots (explained below).

Set Number 3 Experimental Design

E. coli 1090, *A. hydrophila*, *P. mendocina*, and *S. aureus* were grown on Difco TSA, Remel TSA, or sheep blood

agar. Five days after Day 1 (and after the instrumental retuning), bacterial growth from the Difco and Remel TSA and blood agar plates for each isolate was pyrolyzed. Averaged spectra were obtained as in the other experiments. One of the Difco TSA *E. coli* Day 5 spectra was selected and via the multiplicative algorithm (details below) was used as a tracking reference in an attempt to adjust the other Day 5 spectra to resemble the appropriate one in the Day 1 spectral library. The effectiveness and consequences of these algorithmic manipulations were judged using CV score plots.

Set Number 4 Experimental Design

Five strains of *E. coli*, [*E. coli* HB101, *E. coli* JM109 (a derivative of Hoffman Berling strain 1100); *E. coli* 25922 (ATCC); *E. coli* PDTG112; and *E. coli* 98222 (ATCC)], a test set that demonstrated a large degree of genetic variation within the species, were each cultured on Remel TSA, sheep blood agar, and MacConkey agar (Difco). PyMS Spectra from these experiments were not adjusted algorithmically. Pairs of replicate spectra were grouped (total number of experiments numbering 30 = 5 strains \times 3 media \times 2 replicates each) were treated by pattern recognition as if each pair was from a different strain. That is, bacteria grown on different media were not assigned the same group symbol, so the pattern recognition program treated the 5 strains \times 3 media as if they were from 15 different strains. Spatial relationships among isolates of the same strain grown on different media as they appeared on CV score plots were used to determine which strains were metabolically similar.

General Culture Conditions and Sampling Methods

The cultures in all experimental sets were grown on solid media for 24 h at 37 °C. Cells were harvested by loop in sufficient numbers to provide a turbid suspension in 250 μL of 75% ethanol and 25% distilled water. In Experimental Set 1, the sample suspension was intentionally made highly concentrated to aggravate MS optics contamination and associated space charge effects. Suspensions were sonicated (Branson Model B-3, 15W) for about 2 min each to disperse cells.

General PyMS Method Parameters

The following procedure was used for all experimental sets. If necessary, each bacterial sample was resuspended before sampling by sonication or vortexing (Super Mixer, Cat. No. 422, from Cole-Parmer, Chicago, IL). A 2- μL amount was taken up with a GC sample injection syringe and dispensed on the rhenium wire loop of a direct exposure probe (DEP, Finnigan Corporation, Sunnyvale, CA). The droplet was air dried for a few minutes and the probe was inserted through the

vacuum lock into the ion source of the F4500 quadrupole MS (Finnigan Corporation).

Full scan acquisition was initiated at 1 scan/s over m/z 40 to 540 with 70 eV EI ionization. The DEP was heated at 5 mA/s from ambient to a maximum of 500 mA. The spectra across the evaporation peak were averaged and the average spectrum was stored as the result of the PyMS experiment.

Algorithm Used for Spectral Correction/Compensation

A bacterial strain (e.g., *E. coli* 1090) whose spectrum already existed in the database was chosen as a reference for the purpose of tracking spectral drift. Ratios of the tracking reference's relative ion intensities were used to define a correction factor matrix for the date of analysis. At a particular m/z value, i , the correction factor, F_i , was defined as $F_i = I_{i,\text{database}}/I_{i,\text{date}}$ where $I_{i,\text{database}}$ is the relative intensity at that ion as observed in the reference spectrum from the database and $I_{i,\text{date}}$ is the corresponding value for that same strain's spectrum as observed on the date of analysis. The column of F_i values, the Correction Factor Matrix, was used to multiply the Matrix of Raw Ion Intensities of an unknown isolate's spectrum. The unknown's spectrum was, of course, obtained the same day and under the same culture, handling, and PyMS instrumental acquisition parameters as that date's tracking reference. The product of this process for each unknown was a Matrix of Compensated Ion Intensities, the unknown's drift-compensated spectrum. Compensated and uncompensated spectra were combined into sets for comparison via pattern recognition as explained below.

Pattern Recognition Method and Use of Score Plots for Evaluation of Results

In Experimental Sets 2 through 4, multivariate pattern recognition—principal component (PC) and canonical variate (CV) analyses—and their associated score plot displays were produced using RESolve version 1.2 (Colorado School of Mines, Golden, CO). The spectra for each set of experiments were imported from the mass spectrometer and collected together (as *.res files) for comparison by the multivariate statistical methods available in the RESolve program.

Data pretreatment for a set of spectra included identification of replicates by assigning to each spectrum an alphanumeric symbol defining the group of replicate isolates to which it belonged. All spectra were also normalized to total ion intensity. This normalization step is required so that variations in ion intensity arising from differences in the amount of sample pyrolyzed (irrelevant to pattern definition) would not confound the pattern recognition.

For each comparison set (*.res file) the first 30 PCs were calculated. A smaller number of the PCs was used

for calculation of CVs. The optimal number of PCs in each experiment was determined by leave-one-out (LOO) cross-validation. In LOO cross-validation, each sample spectrum is temporarily removed from the comparison set and the CV model is recalculated and used to identify the removed sample. When this process has been repeated for every sample, the program reports the percentage of correct answers obtained. In this work, the number of PCs used for each model was that which gave the highest LOO cross-validation for the CV model. The number of CV vectors calculated in each pattern recognition experiment was $n - 1$, where $n =$ the number of groups defined by symbol assignments in the data-preprocessing step. The CVs produced using the optimal number of PCs were used to generate score plots described below.

Score plots are two-dimensional plots of the coordinates (the scores) associated with each experiment for the two selected vector dimensions displayed on the plot. Another way to describe them is as projections onto the plane of the two selected vectors (axes) for each experiment's coordinates. Similar microbial samples have similar scores in each dimension, so that alphanumeric symbols representing several samples show up on the score plot in proximity to each other whenever those samples have similar spectra. In this work, use of score plots in vector composite dimensions, especially CV plots, allowed for a qualitative estimate of overall spectral similarity that maximized the amount of variation represented in the fewest number of dimensions. This was possible because each CV included contributions from all of the factors (i.e., m/z ion intensities) in the original multi-dimensional data. Thus, CV scores of samples were similar only when almost all of their relative ion intensities were similar. Any significant relative intensity variation, even in a few ions, would change the CV score and, on the 2-D score plot, displace an experiment's alphanumeric symbol location relative to that of the original spectrum.

Spatial relationships on CV score plots permitted visualization of the degree and nature of changes resulting from algorithmic compensation of raw spectra. This capability was used for Experimental Sets 2, 3, and 4. In Sets 2 and 3, two clusters of alphanumeric symbols (for replicate analyses) before and after their algorithmic transformation could be connected using an arrow to characterize the change. For several before- and after-transformation cluster pairs, the direction and length of the arrows implied a variety of important conclusions that are reported below. In Experimental Set 4, arrows superimposed on the score plots had a different significance. They did not represent the results of algorithmic transformation, but joined symbols for the same isolates grown on different media, with the particular media difference in each case shown by the appearance of the arrow (solid or dotted).

The compensation algorithm was implemented using in-house code written in the PERL programming language. The algorithm included a capability to track

“within-session” spectral drift by a linear interpolation of correction factors for a “beginning-of-session” experiment and an “end-of-session” experiment for the same tracking reference sample. Experiments reported here did not require the “within-session” compensation feature. In these cases, the same reference spectrum was defined for the session-beginning and session-ending fields so that the time interpolation capability of the program was not actually used.

Results and Discussion

Set Number 1

Over the course of a single day, as the mass spectrometer ion source and downstream optics became increasingly contaminated with the *E. coli* pyrolysates, total ion intensities for replicate analyses decreased. Total ion pyrogram signal intensity decreased from a maximum of about 1,460,000 counts on the trace to a corresponding value of just over 60,000 counts. More important for pattern definition purposes, the *relative* ion intensities varied. Figure 1a shows a representative *E. coli* mass spectrum from early in the day and Figure 1b, near the end of the day. The way in which these changes progressed during the day can be illustrated by comparing two arbitrarily selected but representative examples; relative intensity of an ion at m/z 56 remained essentially unchanged while that of a higher mass ion at m/z 299 decreased by a factor of almost 10. (Figures 2 and 3) The lines drawn on the two graphs are the best linear fit to the data and indicate that a straight line is a reasonably good approximation for a particular ion's spectral drift. Relative intensities of ions plotted with the same scale expansion in Figure 1a and b diverge between the two graphs as a function of the m/z ratio. So, both plots show a base peak (100% relative intensity) for m/z 55. The ions at m/z 70 differ in relative intensity a small amount, whereas those at m/z 135 differ by a factor of 2.5 and those at m/z 299 by more than an order of magnitude.

In this case, the mass discrimination effect is a consequence of accumulating space charge due to sample ions consistently produced in large numbers, introduced into and transmitted through a contaminated ion source and ion optics. Without compensation, any such relative intensity discrimination is not acceptable for pattern recognition, especially when the higher mass ions (decreasing over time in relative intensity) also contain important taxonomic information, as they do in microbial PyMS experiments. Any MS instrument used for routine chemotaxonomy should be able to operate all day without significant pattern changes. One way to reduce the deleterious effect is for the instrument to be so inherently sensitive that it can obtain useful patterns from dilute samples. These would not significantly contaminate the ion source and cause or exacerbate space-charging.

Set Number 2

Figure 4 is a plot comparing the CV1 and CV2 scores (coordinates). The plot indicates locations of the experiments in the plane of CV1 versus CV2. Letters are for *E. coli* 1090 spectra of four different types. “L” locates the representation of a spectrum defined as an *E. coli* 1090 “library” standard grown on Difco brand TSA and acquired on Day 1 of these trials. The symbol “E” represents a variety of raw spectra for the same organism cultured on the same medium. These were acquired five days subsequent to the initial trials, after retuning the mass spectrometer to slightly different parameters (to simulate instrumental variation) and reculturing the isolate from the same stock. The solid arrow connects one of the Day 5 raw spectra to the library spectrum symbol. This relationship was used to define the correction factor matrix as discussed above. Symbols “e” and “2” locate corrected spectra obtained using two minor variant algorithms of the same general type. The variants proved equivalent as shown by superposition of the “e” and “2” symbols on the plot. The dotted arrow joins the center of the cluster of “E”s to the center of the cluster of “2”s and “e”s, and so represents graphically the effect of a correction based on the relationship shown by the solid arrow. Clearly, the algorithm shifted the Day 5 experiments much closer to the Day 1 library, indicating that the algorithmic correction worked properly. For Experimental Sets 2 and 3, we follow the convention of using a solid arrow to show the relationship used to define the algorithm and a dotted arrow to indicate the transformed results.

Apparently, the nutrients available from the TSA brands (Difco and Remel) were different enough that the locations of their symbols for *E. coli* 1090 varied on the CV score plots. When one of the Difco spectra taken six days after initial trials was used to define the compensation matrix relative to the Day 1 library spectrum, the algorithm transformed the Remel raw spectra in parallel so that the corrected spectra missed the Difco library spectrum by the same spatial distance and direction as between the Day 6 raw spectra (data not shown). However, when one of the Remel TSA spectra was used as reference, the other Remel spectra were transformed into the vicinity of the Difco library (Figure 5). Even spectra of *E. coli* grown on TSA agar differing substantially from that used for the library standard sample could be corrected to approximate the library spectrum.

The same results were obtained when transforming spectra of *E. coli* 1090 cultured on sheep blood agar. If the reference standard came from *E. coli* 1090 cultured on sheep blood agar, then all other *E. coli* 1090 sheep blood spectra could be transformed to the location of the *E. coli* 1090 Difco TSA library standard (Figure 6).

This result has potential significance as a way to shorten analysis times by one or more growth cycles. It indicates the possibility of acquiring raw spectra from isolates grown on a selective medium and using their

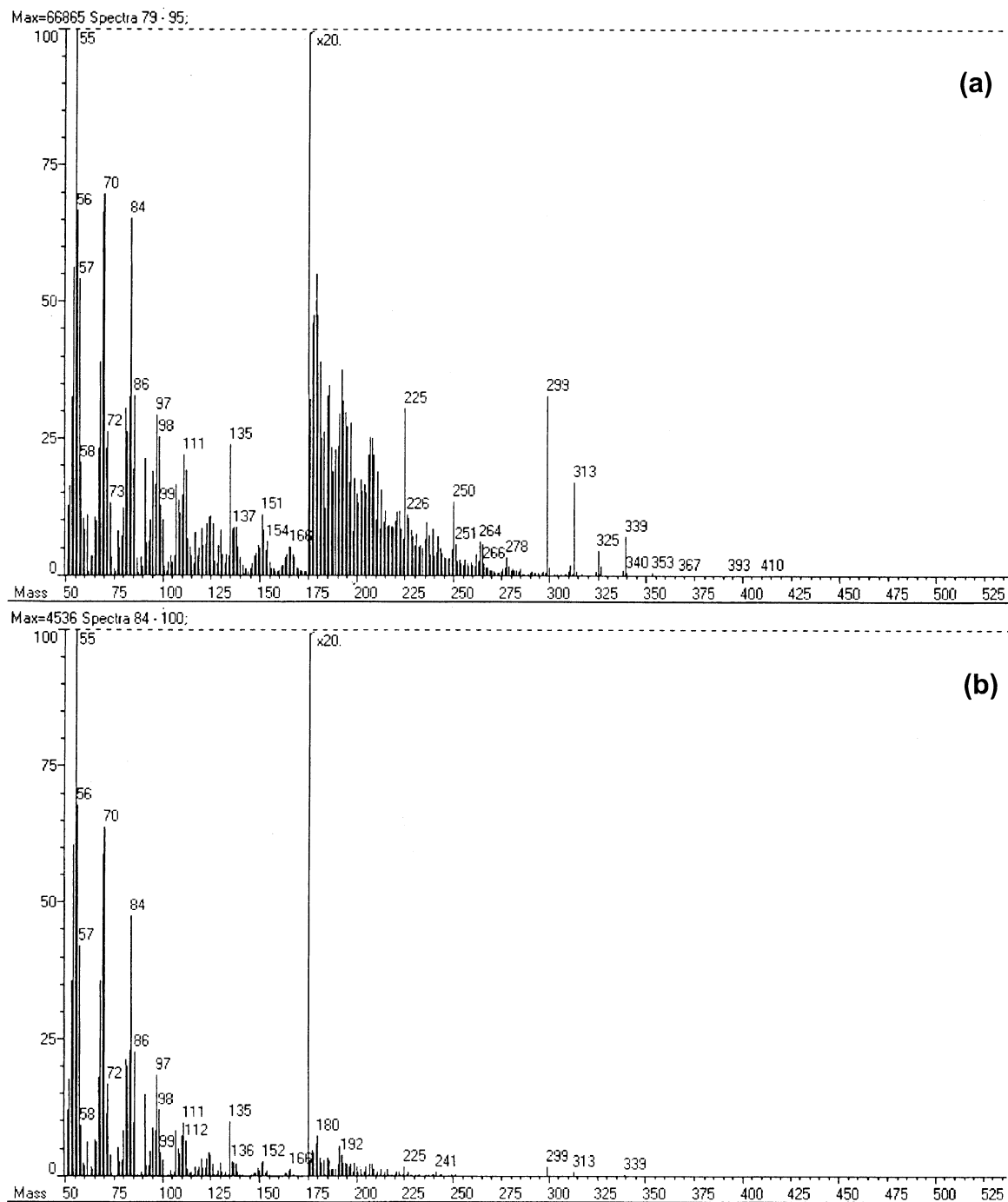


Figure 1. Typical average 70 eV EI spectra from (a) early in the day and (b) the end of the day for *E. coli* pyrolyzed repeatedly into a contaminated MS ion source.

appropriately transformed spectra to consult and compare the data with spectra of isolates in a library that was assembled using bacterial growth from a nonselective culture medium. It implies the possibility of consulting a library when the MS tuning for that day's analysis is not equivalent to that used for the library. (This can happen because of a less than fastidious initial instrument setup or because of purity differences in the PFTBA tune compound that leads to variant ion characteristic ratios.) The result also implies the possibility

that spectra obtained from clinical samples prior to culturing (extracted from urine [17]) can be correlated to library (database) spectra. However, in the latter case, these results are only useful for identification of unknowns if appropriate transform vectors can be defined when there is no way without a culture step to define a reference, drift-tracking standard.

We have shown that the relationship between an *E. coli* spectrum acquired under variant experimental conditions compared to one under standard conditions can

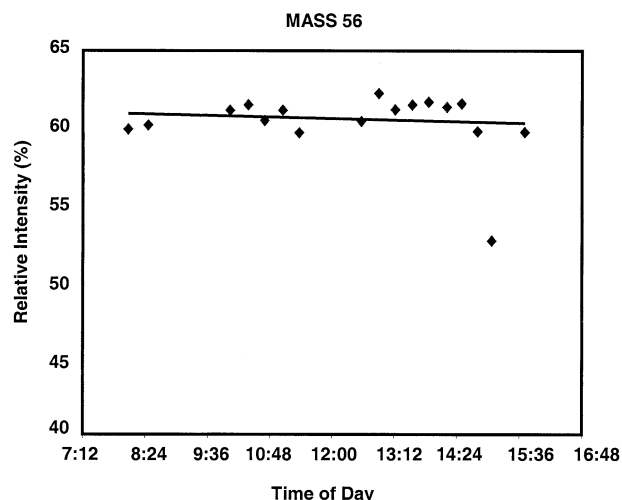


Figure 2. A typical low mass PyMS relative ion intensity, here m/z 56, for replicate measurements, as a function of time of day.

be used to standardize another *E. coli* spectrum obtained under the same variant conditions. Bluntly, this fact doesn't serve a purpose if one doesn't know in advance that the isolate to be transformed is *E. coli*. But, unless the sample is unknown, there is no need for the transformation.

Set Number 3

Figure 7 demonstrates the successful transformation of *A. hydrophila* raw spectra "A" acquired on Day 27 to a new location "a" near an *A. hydrophila* Day 1 library spectrum "C". The *E. coli* 1090 Day 27 spectrum "D" was compared to the Day 1 *E. coli* 1090 library spectrum "L" to define this transformation. In this case, all isolates were grown on Difco TSA. Experimental Set Two results demonstrated successful transformations involving different experimental conditions when the tracking reference was of the same strain as the other

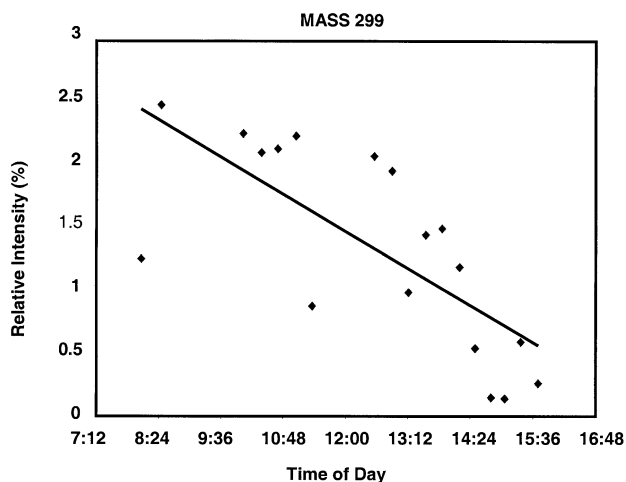


Figure 3. A typical PyMS relative ion intensity, here m/z 299, for replicate measurements, as a function of time of day. For microbial pyrolysis with EI ionization, m/z 299 is a high mass ion.

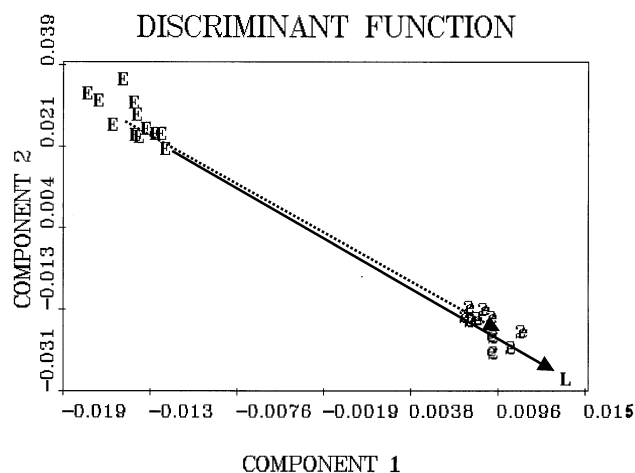


Figure 4. CV1 versus CV2 score plot showing successful compensation of simulated instrumental drift (see experimental section for explanation; the two CV vectors are named Components in this plot). All isolates are *E. coli* cultured on tryptic soy agar (TSA) manufactured by Difco. L = library spectrum generated on the Julian calendar date 01400 (January 14, 2000, referred to in text as initial trial date or Day 1); E = raw spectra generated on 01800 (Day 5); e = 01800 spectra corrected via algorithm 1; 2 = 01800 spectra corrected via algorithm 2. Results show that the two algorithms gave equivalent transformations.

spectra transformed. Figure 7 shows that an appropriate tracking reference can be of a genus different from that of the unknown isolate.

Plots for analogous transformations involving an *E. coli* reference correction of *P. mendocina* or *S. aureus* raw spectra failed (data not shown). In each case, a transform vector defined by the corresponding *E. coli* 1090 relationship had the proper direction but the wrong

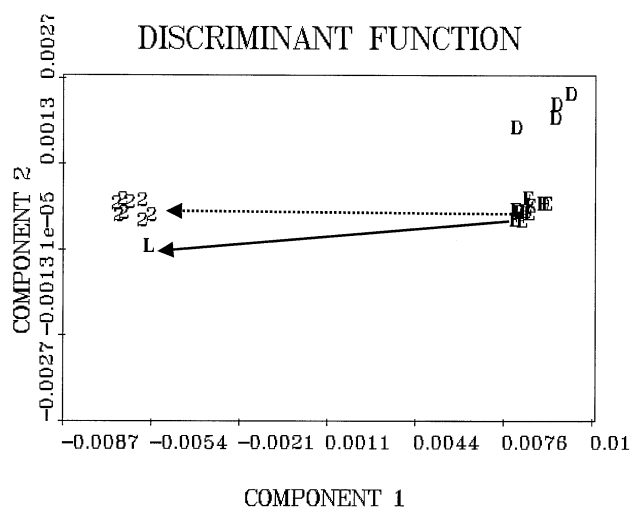


Figure 5. CV1 versus CV2 score plot showing successful compensation for *E. coli* spectral drift induced by use of different brands of TSA culture media (see experimental section). L = Difco, 01400, library, (Day 1); D = Difco, 01900 (Day 6), raw; E = *E. coli*, TSA, Remel, 01900 (Day 6), raw; 2 = *E. coli*, TSA, Remel, 01900 corrected, using one of the Day 6 Remel spectra as reference to Day 1 Difco library.

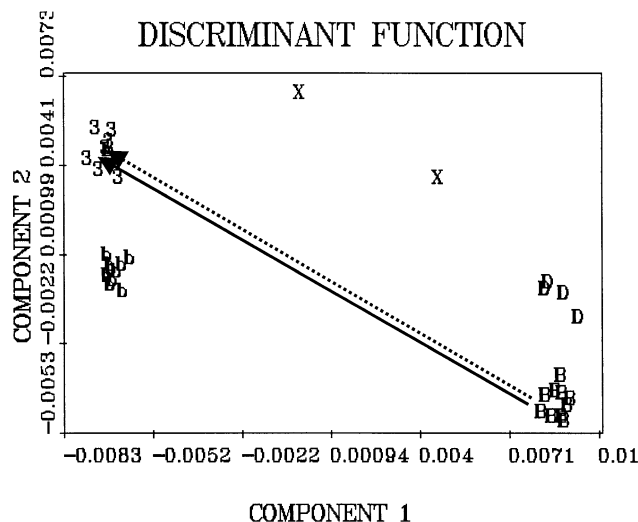


Figure 6. CV1 versus CV2 score plot showing successful compensation for *E. coli* spectral drift induced by use of either TSA or sheep blood agar as culture media (see experimental section). L = TSA, Difco, 01400 (Day 1) library; D = TSA, Difco, 01900 (Day 6); X = blood agar, 01400 (Day 1) raw; 3 = blood agar, 01900 (Day 6) corrected, using B for reference; B = blood agar, 01900 raw; b = blood agar, 01900 corrected, using D for reference.

length to correct the *P. mendocina* or *S. aureus* raw spectra.

Why would an algorithm that functioned properly at the computational level succeed for some isolates and fail for others? The spectral relationships among the four bacteria suggest a possible answer to the question (Figure 8). Experimental Set One involved culturing the four different bacteria on a variety of media. When CV scores for these experiments were plotted together in two dimensions, spectral variations arising from media variations were small compared to those due to the

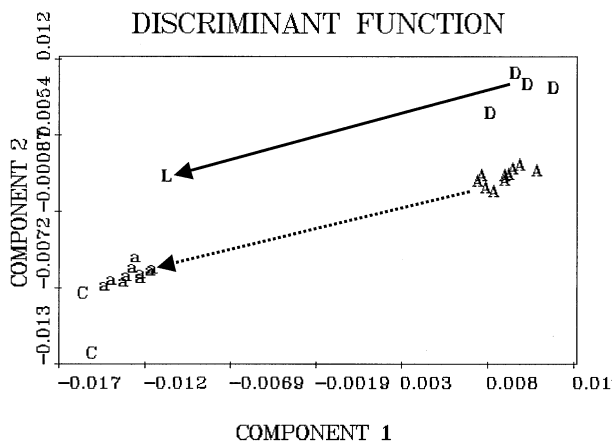


Figure 7. CV1 versus CV2 score plot showing successful transformation of *A. hydrophila* spectra to the *A. hydrophila* library plot location using corresponding relationships of *E. coli* as reference. All isolates were cultured on Difco TSA. L = *E. coli*, 01400 (Day 1) library; D = *E. coli*, 04000 (Day 27); A = *A. hydrophila*, 04000 (Day 27) raw; a = *A. hydrophila*, 04000, corrected; C = *A. hydrophila*, 01400 (Day 1) library.

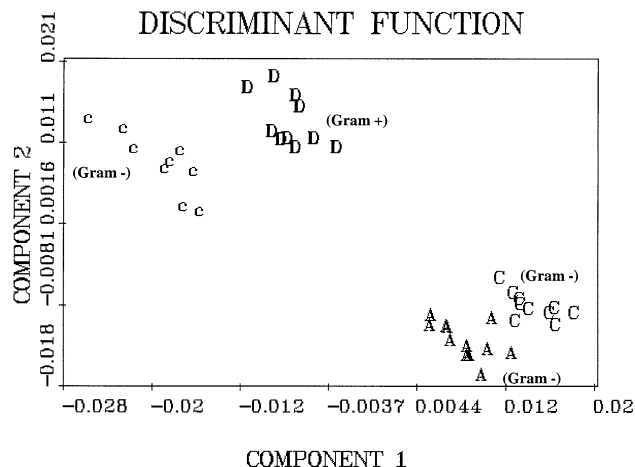


Figure 8. CV1 versus CV2 score plot indicating degree of similarity for spectra of four test bacteria cultured on Difco or Remel TSA or sheep agar, and analyzed by PyMS with EI ionization on Day 1. For this pattern analysis, samples were labeled by bacterial identity and were not differentiated with respect to culture media used. Using 11 PCs, 100% correct identifications were obtained by leave-one-out (LOO) cross-validation. A = *E. coli* 1090, all media; c = *P. mendocina*, all media; C = *A. hydrophila*, all media; D = *S. aureus*, all media. For inferences derived from this plot see discussion of results.

variations among the bacteria because of their different identities. Also, *E. coli* 1090 and *A. hydrophila* appeared adjacent to each other on the plot, whereas *P. mendocina* and *S. aureus* were shown at a substantial distance from the other two. Apparently, the *E. coli* 1090 could be used successfully as a reference to correct its adjacent “neighbor” *A. hydrophila* but not for correcting spectra of the other two whose symbols were more distant. We hypothesized that in an unknown identification situation, only “neighbors” appearing on a score plot could be used to transform each other’s spectra. A step-by-step rationale for this hypothesis follows:

1. Two neighbors are adjacent because they have similar spectra. (This argument is based on the mathematical definition of a CV score plot in relation to the raw data from which the PCs and CVs were calculated.)
2. Two spectra are similar because the bacterial biochemical profiles are similar. (This argument is based on the structure-spectra relationships inherent in organic mass spectrometry.)
3. The biochemical profiles are similar because the bacteria have similar metabolism. (That is, given a particular substrate and growth conditions, similar bacteria will metabolize the substrate to produce, within their cells, most of the same biochemical constituents and in similar proportions.)
4. If bacteria have similar metabolism, they should respond to environmental variations in a similar fashion. (So, working backward, their spectra should change similarly under environmental variation and this would be reflected by a parallel and congruent

shift in their corresponding symbol locations on a CV score plot derived from those spectra.)

- Therefore, either one can serve as an effective tracking reference for compensating its adjacent neighbor's spectral variations using this or another suitable algorithm.

The quality of result obtained by using Goodacre and Kell's [15] or our algorithms for unknowns does not depend on the integrity of the computation, which in both cases is excellent. The accuracy of the result depends critically on the choice of an appropriate reference to define the Compensation Factor Matrix. We saw that identification of an appropriate reference can be accomplished by examining neighbor relationships on a 2-D CV score plot. This can also be accomplished automatically and algebraically (entirely in silico) using Euclidean or Mahalanobis distance values to define neighbors. An in silico method could be implemented automatically and would obviate the need, as a prerequisite to implement the Correction Factor Matrix concept, that an expert select the most appropriate reference for each unknown isolate.

Effective employment of this realization for identifying each unknown depends on the availability of an adjacent neighbor for that unknown. This implies that, before the unknown's analysis begins, a set of potential references must be identified and cultured for analysis on the same day as the unknown. Then one or more potential references will be available to serve for any likely unknown.

Although the species of bacteria present in nature are numerous, only a few are generally regarded as pathogenic. One practical way to reduce the number of reference species used in the spectral transformation necessary for rapid, subspecies-level identification is to include the major anticipated pathogenic strains in the day's reference isolates. This will insure the availability of reference bacteria with similar metabolism for samples in which rapid analysis is critical. Further identification for less critical samples may be performed by growing standards similar to the still unidentified bacteria on the database medium and repeating the PyMS analysis for confirmation the next day.

Which and how many reference isolates must be cultured ahead of time for tracking purposes varies with particular applications. For example, identification of the causative agent for an outbreak due to contaminated oysters need not include a precultured reference suitable for correcting airborne *Legionella* or *Bacillus anthracis* spectra.

Set Number 4

How does one determine the number and identity of precultured references necessary for effective spectral compensation? Figure 9 shows how to use CV score plots to group bacteria into classes having similar metabolism. Five different *E. coli* strains were cultured

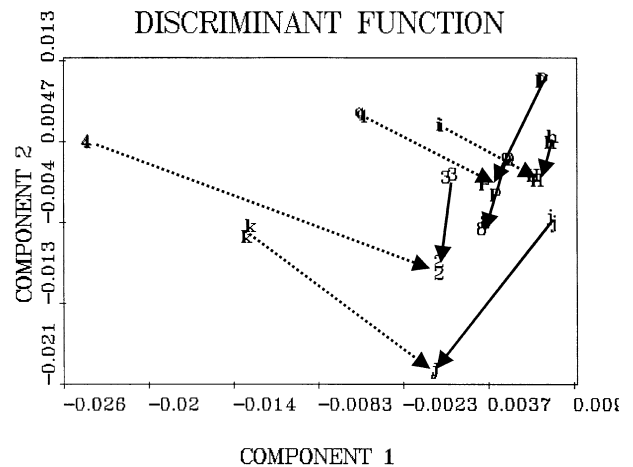


Figure 9. CV1 versus CV2 score plot for five *E. coli* strains cultured on three different types of agar. *E. coli* isolates used: Strain HB101, shown by symbols H, h, and i for Remel TSA, blood agar, and MacConkey agar, respectively. For the same sequence of agar, the others: Strain JM109: J, j, k; strain 25922: 2, 3, 4; strain PDTG112: P, p, q; and strain 98222: 8, 9, for TSA and blood agar, respectively (strain 98222 did not grow on MacConkey agar). For the same strains, solid arrows point from blood agar to TSA agar spectra. Similarly, dotted arrows point from MacConkey agar to TSA. (For significance of spatial relationships indicated by the arrows, see the text.)

on three types of agar: tryptic soy, blood, and MacConkey. One strain did not grow in 24 h on the MacConkey agar. The fourteen classes of spectra appear on the plot with each sample type defined as if they were fourteen different bacterial strains. In this plot, a dotted arrow leads from each cluster of MacConkey spectra to the cluster of TSA spectra corresponding to the same strain. Similarly, the solid arrows join corresponding blood/TSA agar spectra. The blood/TSA solid arrows are roughly parallel to each other. The MacConkey/TSA dotted arrows are also roughly parallel to each other but their direction in the CV1 versus CV2 plane differs greatly from that of the blood/TSA solid arrows. These results demonstrate two approximately parallel sets of shifts in which the shift direction depends on the particular difference in culture substrate. This further exemplifies the previous observation that different strains of the same bacterial species respond in a similar fashion to a particular change in environmental conditions.

By interpreting spatial relationships in the plot of CV3 versus CV4 for the same experiment (data not shown) with those in Figure 9, we concluded that the three strains appearing in the upper right corner of the figure can be categorized in the same "metabolic similarity group" for spectral drift compensation purposes. The other two strains differ in some degree from each other and also from the group of three strains. This discussion illustrates one way to identify a diverse set of potential reference bacteria in an unknown identification situation.

An alternative strategy for defining reference strains

would reduce the number of such strains necessary on a particular day. Rather than define the Correction Factor Matrix for a particular unknown with respect to the most proximate strain, it could be defined using several neighbors, each perhaps more distant in multi-dimensional space than the single, adjacent reference isolate. In this approach, the contributions of each would be inversely distance weighted (i.e., proximity weighted) to emphasize the spectral shift relationships of the nearest neighbors relative to those of the more distant ones.

A proximity weighted multiple reference strategy could be easily automated in user-friendly software so that no expert decisions would be required. This would reduce the level of complexity for analysis and interpretation of results so that an integrated system could be developed for rapid, routine field analyses performed by a technician rather than by a research scientist.

General Protocol When a One-Day Culture Step is Used

We have used the concepts derived from the results shown above to develop two protocols for rapid identification of unknown bacterial samples. We present the first of these without further explanation:

1. Spectra are obtained for samples of unknown identity grown on a selective agar plate. The technician would isolate one colony. (Several other morphologically similar or dissimilar small colonies can be isolated from the same plate and identified separately.)
2. In the same session, spectra are obtained for several reference isolates having representative diversity and are cultured ahead of time on the same media.
3. All raw spectra are compared to each other using one or more 2-D CV score plots.
4. A neighbor "known" is selected as an unknown's tracking reference.
- 5a. Closest neighbor raw spectrum is used with its library spectrum to transform the unknown's raw spectrum to a corresponding database spectrum.
- 5b. Alternatively, the calculation for each unknown may use a Correction Matrix derived from a distance-weighted average of all or several known raw isolate spectra and their corresponding library (database) spectra.
6. Consult the database using each properly transformed spectrum to identify each unknown.

The six steps of this protocol can be executed in an automated fashion for reliable and rapid implementation using software, some of which awaits development.

The concepts underlying the results and their interpretation suggest a second protocol that could be im-

plemented in some situations for automated data acquisition, spectral averaging, spectral drift transformation, and sample identification in a few minutes after sampling rather than subsequent to a 24 h cell culture step.

Proposed Protocol Without a One-Day Culture Step

1. Frequently grow (on a nonselective medium such as TSA), sample, and analyze the same set of tracking reference strains one would use to transform spectra by the previous one-day protocol.
- 2a. Use these "day" spectra to transform in reverse the entire standard database, or a relevant subsection thereof, i.e., make a database-for-the-day. This transformation is possible with noniterative calculations and could be executed by a PC in a matter of seconds, even for a large database.
- 2b. At the same time, algorithmically modify a table of transform vectors from standard conditions to the variants appropriate for that day. The table would contain transform vectors for each tracking reference but with different vectors for each significantly different patient population or environmental extract condition as described by the adjunct protocol below.
3. Run any unknowns and compare their spectra to the newly compiled database-for-the-day.
4. Select from the day's table of transform vectors, the one relevant to a particular non-media-cultured unknown specimen. One such vector might be for an *E.coli* HB101 to transform that species and similar samples in urine [17] to their equivalents on TSA agar. Another example might be a transform vector for *Bacillus* isolates to transform from aerosol-collected [18], starved cells to TSA. A third vector might serve for *E. coli* 1090 similarity types to transform spectra from bacteria separated out of stool samples by immunoaffinity techniques [19, 20] to their corresponding spectra on TSA.
5. Use the appropriate transform vector to correct each raw unknown spectrum of that day to its corresponding pattern had it actually been grown on TSA database-for-the-day media and analyzed on the same instrument.
6. For subspecies-level I.D., compare each transformed unknown spectrum to the database-for-the-day using pattern recognition.

Proposed Adjunct Protocol for Generating Sample Transform Vectors

This process is illustrated using a transform for males suffering urinary tract infections due to *E. coli* HB101:

1. Collect the urine sample from the patient, centrifuge it, suspend the pellet in 70% ethanol, and streak this suspension on the available TSA.

2. The next day, harvest a colony that grew on the TSA plate, suspend the cells and analyze under identical conditions both these cultured cell suspensions and the noncultured ones from the urine extract.
3. Use the ratios of the ion relative intensities to define the transform vector associated with the particular bacterial neighbor set from the particular class of patient; e.g., the transform vector that changes the spectrum from an *E. coli* HB101 extract from a male's urine into a corresponding TSA-cultured spectrum. Note that this approach eliminates factors due to instrumental and sample handling variability.
4. Save the transform vector as the representative of this type of patient having a urinary infection caused by this *E. coli* HB101 or a metabolically similar bacterium.
5. Perform the same process on sample extracts until you have a transform vector for at least one bacterium in each metabolic similarity group that is a likely cause of male urinary infection.
6. Assemble an analogous set of transform vectors for females and for other patient subsets (e.g., male or female diabetics) for which a medical condition or other factors might contribute to a consistent pattern of spectral variability.
7. For a particular patient's sample, use the appropriate transform vector in Step 5 of the proposed protocol without a culture step.

For unknown samples whose growth conditions provide no internal reference available for comparison (e.g., extracts, concentrates), the challenge is to find a reliable way to track the inevitable spectral variations. The principles illustrated in this paper speak to this need. They indicate that a strategy that implements the second and the adjunct protocols may be able to produce same-day strain-level identification of noncultured samples. Figure 8 suggests that substantial changes in culture media do not distort spectra enough to prevent discerning score plot adjacent-neighbor relationships for purposes of defining a useful transform vector. (Using the raw spectral patterns, one can identify a likely subset to which the unknown sample belongs.) Similarly, variations in environmental conditions, under certain well-defined conditions, may not distort raw microbial PyMS spectra so greatly that it is impossible to select appropriate transform vectors for noncultured unknowns. By well-defined conditions we mean, for example, that the clinical analyst will know a sample's history (perhaps that it came from a male patient suffering diabetes) and thus that the difference in environmental experience is compensated by a vector that transforms extracts from the relevant source into their spectral equivalent under library medium culture conditions. The analyst's table of transform vectors will contain such a transform vector only when experimentation has demonstrated that this particular difference in environmental conditions can be reliably corrected using the designated transform vector.

This and other significant issues need to be explored to determine whether such a strategy would yield consistent, rapid, subspecies-level identification in a clinical or other public health setting. PyMS approaches depend on the reproducibility of relative intensities rather than the presence of unique ion biomarkers, so they are applicable only to patterns obtained from isolates or nearly pure (one-strain) extracts. The rapid I.D. protocol using PyMS is limited to cases in which a single strain has multiplied in a normally sterile fluid (e.g., urine) or a selective isolation method was used to extract a single cell type from a matrix contaminated by multiple species (e.g., immunoaffinity extraction of a target pathogen followed by PyMS, algorithmic compensation, and pattern recognition to confirm that any cells extracted are not cross-reactive strains). These steps are necessary because PyMS and most other bacterial chemotaxonomic methods require pattern determination based on pure isolates rather than mixtures [21].

If the spectral compensation method described in the proposed protocol without a culture step cannot be made to work, there is probably no conceptually distinct alternative method for an ultra-rapid assay (real time or a few minutes) based on the phenotypic characteristics of bacterial specimens. The causes of spectral pattern shift are phenomena that cannot be ignored if one requires rapid identification with subspecies-level precision and if there are more than a handful of possible identities for an unknown sample.

Conclusion

This work demonstrates how to track and compensate for changes in environmental factors. Thereby, it establishes the theoretical basis for general protocols useful in assembling and consulting mass spectrometric and other analytical spectrometric, spectroscopic, or chromatographic databases. These results support the feasibility of using PyMS with automated spectral compensation and pattern recognition for routine, rapid bacterial chemotaxonomy.

Since the relevant phenomena affecting classification of microbial samples reside in the biochemical composition of each microbial specimen and not in peculiarities of PyMS, it also follows that the principles related to drift compensation apply to chemotaxonomy based on any other mass spectral, chromatographic, or spectroscopic assay method.

Acknowledgments

The authors gratefully acknowledge support of the U.S. Food and Drug Administration (FDA) for this research. However, the opinions expressed herein are solely those of the authors and do not reflect official FDA policy.

References

- Marshall, S.; Clark, C. G.; Wang, G.; Mulvey, M.; Kelly, M. T.; Johnson, W. M. Comparison of Molecular Methods for Typing *Vibrio parahaemolyticus*. *J. Clin. Microbiol.* **1999**, *37*(8), 2473–2478.
- Wilkes, J. G.; Holland, R.; Holcomb, M.; Lay, J. O., Jr.; McCarthy, S. Comparison of Py-MS, MALDI-TOF/MS, and Molecular Methods for Rapid Classification of *Vibrio parahaemolyticus* Outbreak Strains. *Proceedings of the 48th ASMS Conference on Mass Spectrometry and Allied Topics* Long Beach, CA. **2000**, 650–651.
- Goodacre, R.; Berkeley, R. C. W. Detection of Small Genotypic Changes in *Escherichia coli* by Pyrolysis Mass Spectrometry. *FEMS Microbiol. Lett.* **1990**, *59*(1/2), 133–137.
- Sisson, P. R.; Kramer, J. M.; Brett, M. M.; Freeman, R.; Gilbert, R. J.; Lightfoot, N. F. Application of Pyrolysis Mass Spectrometry to the Investigation of Outbreaks of Food Poisoning and Nongastrointestinal Infection Associated with *Bacillus* species and *Clostridium perfringens*. *Int. J. Food Microbiol.* **1992**, *17*(1), 57–66.
- Cartmill, T. D.; Orr, K.; Freeman, R.; Sisson, P. R.; Lightfoot, N. F. Nosocomial Infection with *Clostridium difficile* Investigated by Pyrolysis Mass Spectrometry. *J. Med. Microbiol.* **1992**, *37*(5), 352–356.
- Low, J. C.; Chalmers, R. M.; Donachie, W.; Freeman, R.; McLaughlin, J.; Sisson, P. R. Pyrolysis Mass Spectrometry of *Listeria monocytogenes* Isolates from Sheep. *Res. Vet. Sci.* **1992**, *53*(1), 64–67.
- Sisson, P. R.; Freeman, R.; Magee, J. G.; Lightfoot, N. F. Rapid Differentiation of *Mycobacterium xenopi* from Mycobacteria of the *Mycobacterium avium-intracellulare* Complex by Pyrolysis Mass Spectrometry. *J. Clin. Pathol.* **1992**, *45*(4), 355–357.
- Gould, F. K.; Freeman, R.; Sisson, P. R.; Cookson, B. D.; Lightfoot, N. F. Inter-Strain Comparison by Pyrolysis Mass Spectrometry in the Investigation of *Staphylococcus aureus* nosocomial infection. *J. Hospital Infect.* **1991**, *19*(1), 41–48.
- Sisson, P. R.; Freeman, R.; Magee, J. G.; Lightfoot, N. F. Differentiation Between Mycobacteria of the *Mycobacterium tuberculosis* Complex by Pyrolysis Mass Spectrometry. *Tubercle* **1991**, *72*(3), 206–209.
- Orr, K.; Gould, F. K.; Sisson, P. R.; Lightfoot, N. F.; Freeman, R.; Burdess, D. Rapid Inter-Strain Comparison by Pyrolysis Mass Spectrometry in Nosocomial Infection with *Xanthomas maltophilia*. *J. Hospital Infect.* **1991**, *17*(3), 187–195.
- Sisson, P. R.; Freeman, R.; Gould, F. K.; Lightfoot, N. F. Strain Differentiation of Nosocomial Isolates of *Pseudomonas aeruginosa* by Pyrolysis Mass Spectrometry. *J. Hospital Infect.* **1991**, *19*(2), 137–140.
- Freeman, R.; Gould, F. K.; Wilkinson, R.; Ward, A. C.; Lightfoot, N. F.; Sisson, P. R. Rapid Inter-Strain Comparison by Pyrolysis Mass Spectrometry of Coagulase-Negative Staphylococci from Persistent CAPD Peritonitis. *Epidemiol. Infect.* **1991**, *106*(2), 239–246.
- Sisson, P. R.; Freeman, R.; Lightfoot, N. F.; Richardson, I. R. Incrimination of an Environmental Source of a Case of Legionnaires' Disease by Pyrolysis Mass Spectrometry. *Epidemiol. Infect.* **1991**, *107*(1), 127–132.
- Meuzelaar, H. L. C.; Haverkamp, J.; Hileman, F. D. *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials*. Elsevier: Amsterdam, 1982.
- Goodacre, R.; Kell, D. Correction of Mass Spectral Drift Using Artificial Neural Networks. *Anal. Chem.* **1996**, *68*, 271–280.
- Berkeley, R. C. W.; Goodacre, R.; Helyer, R.; Kelly, T. Pyrolysis-MS in the Identification of Micro-Organisms. *Lab. Prac.* **1990**, *39*, 81–83.
- Letarte, S.; Wilkes, J. G.; Bertrand, M. J. Detection and Identification of Bacteria in Urine Samples by Py-MAB-MS. *Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics*; Chicago, **May 2001**, in press.
- Basile, F.; Beverly, M. B.; Voorhees, K. J.; Hadfield, T. L. Pathogenic Bacteria—Their Detection and Differentiation by Rapid Lipid Profiling with Pyrolysis Mass Spectrometry. *TRAC-Trends Anal. Chem.* **1998**, *17*(2), 95–109.
- Rees, J.; Madonna, A.; Ferrer, I.; Voorhees, K. J.; Furlong, E.; Basile, F. Capture and Detection of Bacteria Using Immobilized Antibodies and MALDI-TOF Mass Spectrometry. *Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics*; Chicago, **May 2001**, in press.
- Madonna, A. J.; Basile, F.; Furlong, E.; Voorhees, K. J. Detection of Bacteria from Biological Mixtures Using Immunomagnetic Separation Combined with Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15*(13), 1068–1074.
- Goodacre, R. Characterisation and Quantification of Microbial Systems Using Pyrolysis Mass Spectrometry: Introducing Neural Networks to Analytical Pyrolysis. *Microbiol. Europe* **1994**, *2*(2), 16–22.