technical note

Disciminating Mycobacterial Strains by LC and Multivariate Analysis

L. Scott Ramos, Infometrix, Inc., Suite 250,
10634 East Riverside Drive, Bothell, WA 98011

Abstract

Traditional methods of mycobacterial identification relied on a battery of biochemical tests. Most of these tests were low resolution—yes or no, for example—and could be subjective. Chromatographic procedures have replaced these tests in many laboratories because of ease of implementation and greater objectivity in the interpretation of the results.

Liquid chromatography (LC) of bacterial extracts has been demonstrated as a useful technique to separate the high molecular weight mycolic acids that are species specific. Multivariate analysis of the LC profiles can help the analyst distinguish among the different bacterial species. Tuberculosis is caused by one species of Mycobacteria—there are other species in this genus which are also human pathogens—thus reliable identification of species is critical.

Method

Mycobacterial extracts were analyzed by reversed phase LC using a C_{18} column, 7.5cm long by 4.6 mm diameter. A 10-minute step gradient from 98% methanol in methylene chloride was run to a final solvent composition of 35% methanol - 65% methylene chloride.

Retention time variability is known to influence the outcome of multivariate analysis, therefore, the chromatograms were aligned using LineUp™ (Infometrix, Inc.) prior to further analysis. Chromatographic profiles were assembled into a database in the KnowItAll® Informatics System, then principal component analysis (PCA) was run with the AnalyzeIt™ MVP package. Profiles were normalized by area % and mean-centering was used in preprocessing. Because the solvent region is composed of peaks not representative of the mycolic acid contribution, it was excluded from the computations.

Discussion

In Figure 1, the great variability among the different mycobacteria species can be seen. This species specific composition allows species to be distinguished based on their LC profiles.

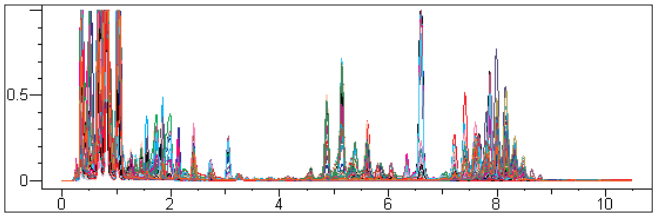

Figure 1. *Overlay of LC profiles from ten mycobacteria species.*

Subsequent PCA analysis was focused on the region from 3.5 to 9 minutes. The PCA scores are shown in Figure 2 where clear distinctions among most of the species are evident.

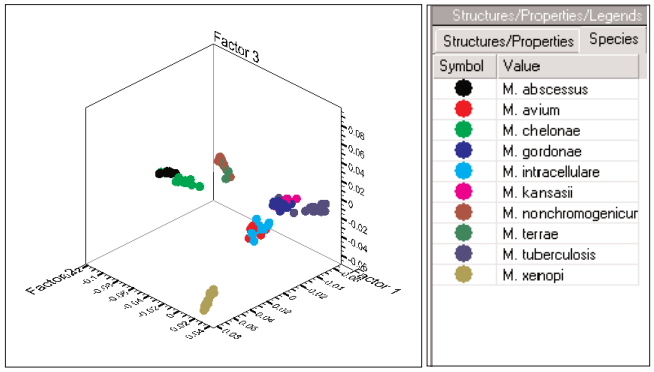

Figure 2. *PCA scores of PC profiles of several strains of ten mycobacteria species.*

Adjacent to the scores plot is the legend for the plot symbols, which were colored by their species designation. *M. tuberculosis (TB)* is of course the most important among these organisms, so it is imperative that it be separable from other species. From the plot, it appears that there may be some overlap among TB and two other species with similar profiles: *M. gordonae* and *M. kansasii*. In fact, the overlap among these three species is not so large.

BIO-RAD InfoMetrix

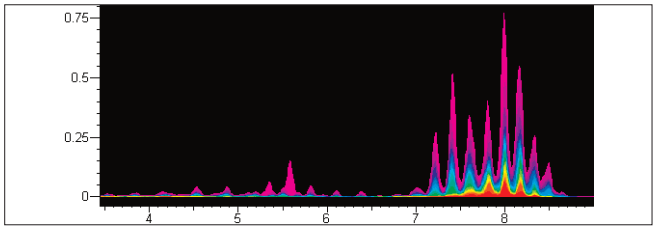

***Figure 3.** Overlay Density heatmap of profiles of three mycobacteria species.*

The red region of the overlap density (OD) heatmap indicates those portions of the chromatograms that are of greatest similarity. Although the overall similarity in the region of 7 to 8.5 minutes is significant, this plot indicates that it should be possible to discriminate among the species: the region of the heatmap in purple (unique features) covers a larger area of the common peaks than the region in red (common features).

The profiles from these three species were transferred to KnowItAll's MineIt™ application to make a separate hit list, then this list of profiles were transferred back to AnalyzeIt MVP in order to run a separate PCA. The same preprocessing parameters were used but a more narrow time range was specified—6.5 to 9 minutes—because the diagnostic peaks were in this limited retention time region. The scores from this analysis are shown below.

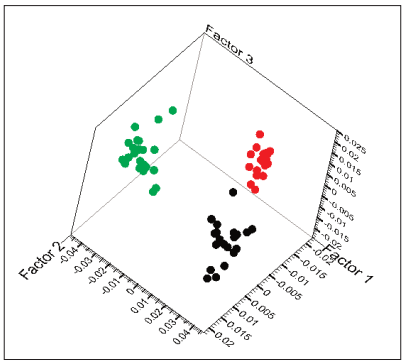

***Figure 4.** PCA scores of three mycobacteria species.*

From the analysis in Figure 4, it can be seen that *M. tuberculosis* samples are totally separated from those of similar profile, and the other two species are distinguishable as well. This separation is borne out by observation of the corresponding profiles, shown in the following figures. The color-by-species legend allows selecting just samples from a given species, and these can be shown as an overlap density heatmap or as the actual profiles.

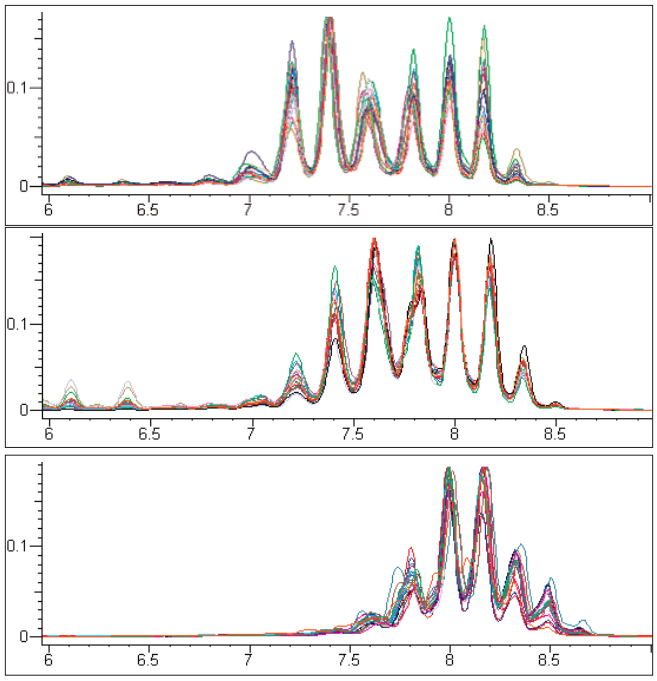

***Figure 5.** LC profiles of M. gordonae, M. kansasii and M. tuberculosis.*

Conclusions

This application note provides both a strategy and a procedure for extracting the information content from large, complex chromatography databases. If the data is aligned to remove retention time variation, principal component analysis is successful in organizing the data into natural groupings. Using an interactive graphical display of PCA scores provides a means to identify samples of interest, where the overlap density map can take over to highlight portions of the chromatogram that reflect similarities and differences.

References

Mycolic Acid Pattern Standards for HPLC Identification of Mycobacteria, U.S. Department of Health and Human Services, Public Health Service, 1999, http://www.cdc.gov/ncidod/dastlr/TB/TB_HPLC.htm.

L.S. Ramos (1994). Characterization of Mycobacteria Species by HPLC and Pattern Recognition. *J. Chromatog. Sci.*, 32:219-227.

BIO-RAD

Bio-Rad Laboratories, Inc.

Informatics Division
www.knowitall.com

U.S.	*Phone: +1 215 382 7800 • 1 888 5 BIO-RAD (888-524-6723) • E-mail: informatics.usa@bio-rad.com*
Europe	*Phone: +44 20 8328 2555 • E-mail: informatics.europe@bio-rad.com*
Japan	*Phone: +81 03 (5811) 6287 • E-mail: informatics.nbr@bio-rad.com*
Rest of World	*Phone: +1 215 382 7800 • E-mail: informatics.row@bio-rad.com*