SCIENTIFIC PAPERS
OF THE UNIVERSITY OF PARDUBICE
Series A
Faculty of Chemical Technology
10 (2004)

GC/MS AND FAB/MS ANALYSIS OF LIPID MOLECULES FOR TAXONOMIC CLASSIFICATION OF MICROORGANISMS

Petra MIKETOVÁ[a1], Ludmila KHAILOVÁ[b], David JELINEK[b], Karl H. SCHRAM[b], Ida M. (Ki) MOORE[a] and Jarmila VYTŘASOVÁ[c]
[a]College of Nursing, [b]College of Pharmacy, University of Arizona, AZ-85721, Tucson,
[c]Department of Biological and Biochemical Sciences, The University of Pardubice, CZ–532 10 Pardubice

Received September 30, 2004

The goal of this project was to identify the presence of microorganisms in the environment by using their gas chromatography/mass spectral (GC/MS) characteristics as biomarkers. A combination of GC/MS results and pattern recognition statistical evaluation were used to provide a mathematical model with defined parameters for different microorganisms, which allows the organisms to be classified as bacterial, fungal or viral in nature. Results from fast atom bombardment mass spectral (FAB/MS) analysis showed differences in the complex lipid profiles for most of the analyzed microorganisms. A general screening method that can provide early detection of microorganisms would permit initiation of most therapeutics. Various drugs are used to treat bacterial, fungal

[1] To whom correspondence should be addressed.

or viral diseases and, when treatment is started early, there is a dramatic reduction in mortality. A methodology for detection and classification of microorganisms is described in this paper.

Introduction

Early and accurate assessment of biological agents has gained momentum in light of recent events. The emerging concerns regarding the use of biological agents (bacteria, fungi or viruses) have raised issues associated with early detection and correct identification of these microorganisms. Such efforts could be organized into two major areas as illustrated in Figure 1.

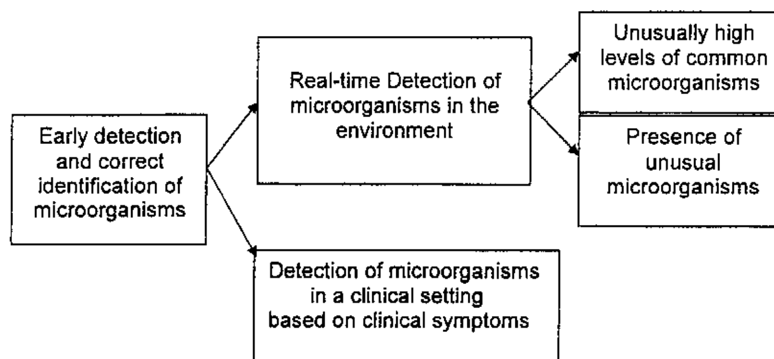


Fig. 1 Two major areas of interest for an early and correct identification of the microorganisms

There is an urgent need for technology which would allow detection of microorganisms in real time using universal instrumentation which is applicable to wide range of microorganisms. Such identification would provide information about the type of the microorganism and would be complemented by standard cultivation for confirmation of the microorganism (requires up to 48 hours). The early type-specific microbial identification would be used for administration of prophylactic antibiotic/antifungal/antiviral agents during the first 48 hours post exposure. With some microorganisms having a very narrow therapeutic window, such early treatment would significantly lower the morbidity and mortality of affected human.

A short overview of some limitations of currently used methods for the detection of the presence of microorganisms is given below.

A. *Staining for Gram-positive and Gram-negative bacteria, Culturing and Microscopic Evaluation:* A significant concentration of pathogen is required

for microscopic examination; thus use for early detection of dispersed material is limited. Cross-contamination with other bacteria in the environment can affect the cultivation process. In addition, culturing provides no information concerning the level of bacteria. Finally, due to slow growth of some microorganisms, the identification might take days, prolonging the period of empirical and possibly inadequate therapy.

B. *Molecular methods: Polymerase chain reaction (PCR)* based methods require identification of specific gene targets. Once developed, these methods can be very sensitive and relatively fast. However, most PCR assays applied in clinical and environmental diagnostics still suffer from contamination problems.

C. *Biosensors:* Different microorganism-specific biosensors are needed to cover wide variety of potential microorganisms.

D. *Fluorescence particle sizing:* This methodology is based on the ability to count and size airborne respirable particles. Fluorescence emission from specific biomolecules is utilized to distinguish different microorganisms. This technique is not suitable for localized microbial release and cannot be used in high environmental background conditions.

Mass spectrometry is proposed as a universal, ultra-sensitive bioanalytical method, which is characterized by low detection limits, high specificity and very short analysis time (5 – 10 minutes). The greatest advantage in the application of mass spectrometry to microbial identification is the use of a single, universal instrument for detection of all microorganisms. The infectious agent does not need to be cultured in order to make a specific identification. Also the presence of the agent is established using multiple components, which are characteristic for a particular microorganism, therefore, dramatically reducing false positive/negative results. The biomarkers are detected directly, i.e., antibodies and antigens are not necessary, and the biomarkers are indicative of the presence of living microorganism. Detection of the microbial agent does not depend on the transient existence of biomarker, i.e., as long as the agent itself is present, the characteristic signature will be recognized.

Chemical profiles of microorganisms have been used for taxonomical purposes by many researchers. Gas chromatography/mass spectrometry is usually the method of choice allowing the separation of extracted components and providing a characteristic mass spectrum used for definitive identification [1]. Besides electron ionization, chemical ionization has been used as a valuable tool for biomarker ions in the fatty acid class [2]. Chemical ionization provides reduced fragmentation, relative to EI, while enhancing the intensity of the molecular ion of fatty acids. As a result, chemical ionization was shown to be more effective than electron ionization in bacterial profiling. Fast atom bombardment mass spectrometry has been utilized for direct determination of phospholipid structures in microorganisms for the taxonomic purposes [3]. Tandem mass spectrometry

Table I Microbial cultures used in the study

Name of the Microorganism	ATCC Number	Classification	Medium	Abbreviation (PCA symbol)
Seratia marcescens	264	G– bacterium	Difco 0001 Difco 0003	SMR (m)
Enterobacter aerogenes	29008	G– bacterium	Difco 0001 Difco 0003	EAE (a)
Bacillus cereus	6464	G+ bacterium	Difco 0001 Difco 0003	BCE (c)
Bacillus subtilis	82	G+ bacterium	Difco 0001 Difco 0003	BSU (s)
Bacillus circulans	61	G+ bacterium	Difco 234000 Difco 213000	BCI (r)
Bacillus pumilus	72	G+ bacterium	Difco 234000 Difco 213000	BPU (p)
Beet curly top geminivirus clone PBCT051	45037	Virus	Bacteria-free phage	037 (1)
Tobacco etch potyvirus clone PTL-0059IN	45035	Virus	*E. coli* HB101	035 (2)
Tomato apical stunt viroid clone PAS64-H4 in	45053	Virus	*E. coli* K-12 strain JM83	053 (3)
Aspergillus niger	9642-U	Fungus	#336	ANG (N)
Aspergillus clavatus	9192	Fungus	#325	ACL (C)
Aspergillus nidulans	10074	Fungus	#336	AND (D)
Aspergillus fumigatus	96918	Fungus	#325	AFU (F)
Candida albicans	14053	Fungus	#200	CAB (A)
Fusarium oxysporum	48112	Fungus	#338	FOX (O)

Difco 0001: #3 Nutrient Aga
Difco 0003: Nutrient Broth
Difco 234000: #3 Broth:Nutrient Broth
Difco 213000: #3 Agar:Nutrient Agar

#336: Potato Dextrose Agar
#325: Malt Extract Agar
#200: YM Agar
#338: Potato Sucrose Agar

(MS/MS) measurements allowed structural characterization of characteristic phospholipids used for classification of microorganisms. Electrospray mass spectrometry (ESI/MS) in conjunction with MS/MS measurements, provided distinctive patterns for microbial classification [4]. The latest development in the matrix assisted desorption-time of flight mass spectrometry (MALDI-TOF/MS)

technique is its use for profiling large biomolecules, which are used in the construction of a classification algorithm [5,6].

The presence of a microbe can be determined by using the microbial fingerprint approach. Such a fingerprint consists of a set of biomarkers, chemical entities, which are characteristic of the presence of specific microorganisms. Specific biomarker molecules are derived from the differences in the composition of the cellular membranes in bacteria, fungi and viruses. The compound classes to be examined as a source of the biomarkers are lipids, in particular fatty acids.

Lipids constitute a suitable class of molecules with chemotaxonomic utility since, being major constituents of cell walls, they are directly accessible in intact microorganisms by mass spectrometry. In a method developed by Fenselau [7], the microorganisms are screened based on the predominant phospho- or other polar lipids detected in their mass spectra. However, due to only a single biomarker, this method has a large number of false positive results.

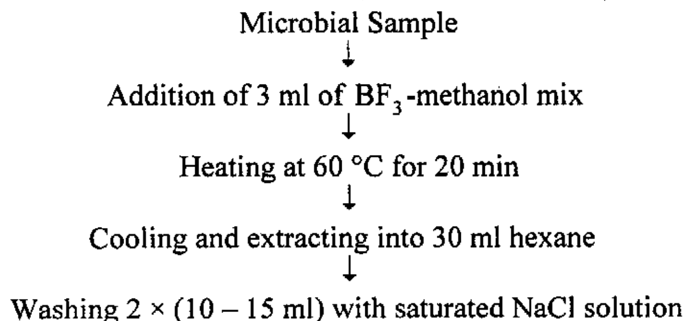
Methyl ester derivatives of fatty acids were used for taxonomical classification of 15 different bacteria after being pyrolyzed and profiled using GC/MS instrument [8].

In this project, we set out to test the hypothesis that we can use lipid-derived biomarkers to distinguish the microbial origin as bacterial, fungal or viral. GC/MS profiles of different microorganisms were evaluated using pattern recognition software [9] and compared with the classification potential of complex lipid profiles analyzed by FAB/MS from the same samples.

Experimental

Sample description. The following microbial cultures were purchased from ATCC (Manassas, VA, USA):

Sample Preparation and GC/MS Analysis. Dry microbial material was weighted to 0.1 mg and dissolved in 3 ml of boronitrofluoride (BF_3)-methanol mix (14 % BF_3: 86 % methanol, Pierce, Rockford, IL, USA). The sample preparation followed the scheme below:



↓

Collecting hexane layer and drying it under N_2

↓

Adding 25 μl hexane and injecting 1 μl into the GC

The resulting fatty acid methyl esters (FAMEs) were analyzed by GC/MS (HP5890/HP5970, Agilent Technologies, Foster City, CA, USA) using fused silica capillary column SPB-5 (30 m × 0.25 mm × 0.25 μm, Supelco, Bellefonte, PA, USA). 1 μl of each sample was injected by using an autosampler (HP 6890 series). The initial temperature on the column was 70 °C and was held constant for 2 minutes and a temperature gradient of 10 °C min^{-1} was finished at 280 °C (held for 7min). The detector temperature was 290 °C and the injector temperature was 250 °C. Ultra-pure helium was used as a carrier gas. Samples were analyzed using an HP5970 mass selective detector at an ionizing voltage of 70 eV. All measurements were conducted in triplicate in order to ensure reproducibility and to gain variability for the statistical evaluation.

Sample preparation and FAB/MS Analysis. All microbial samples were weighed out in the amount of 0.1 mg and dissolved in 1 ml phosphate buffer. The lipids were extracted with methanol:chloroform (1:2), and the organic phase containing lipids was dried under nitrogen. The dried extract was dissolved in thioglycerol and subjected to the FAB/MS analysis. The analyses were performed on a JEOL HX110A instrument (Peabody, MA, USA).

Data Analysis. Pattern recognition analyses were performed using the RESOLVE software developed by Dr. Voorhees at Colorado School of Mines, Golden, Colorado [9]. The intensities of individual GC peaks (total ion current) were used as independent variables. All data were preprocessed by normalizing to the total ion intensity and mean centered to allow direct relationship between the principal components and the variance in the data set. Both supervised and unsupervised pattern recognition analyses were performed. In unsupervised analysis, the class assignment of the samples is unknown and the natural clustering of the data points is examined. In principal component analysis (PCA), the data set is mathematically evaluated in order to express the variation among a large number of variables using a small number of factors [10]. The factors are then reduced to two-dimensional plots and are used to show the distribution of data. Linear discriminant analysis (LDA) is a supervised pattern recognition technique where the class assignment is considered during the course of analysis. The aim of analysis is to develop a model by defining common features of a particular class in the data set and use the model to predict the class assignment for future measurements [10]. The results of PCA and LDA were visualized as score plots (factor-factor) along with their corresponding loading plots.

Results and Discussion

Six bacterial specimens (three Gram-positive and three Gram-negative), three viral specimens and six fungal specimens were weighed into two separate vials in the amount of 0.1mg. One portion of the samples was prepared for the GC/MS analysis following the scheme outlined in the experimental section. The other portion of the specimens underwent lipid extraction and the crude lipid extract was dissolved in the thioglycerol and analyzed on FAB/MS.

Results from GC/MS analysis. Due to the complexity of the obtained data, a pattern recognition approach was used in order to evaluate the differences between individual microbial specimens. The intensities of fatty acid peaks were used as independent variables during the PCA. The identities of fatty acids were confirmed during the electron ionization mass spectral analysis. The fatty acids found in the microbial samples are listed in Table II.

During the pattern recognition analysis, two different categorization methods were applied. First, all microbial specimens obtained from the ATCC were categorized based on their origin: b indicating bacteria, v indicating viruses and F indicating fungi. The PCA results are shown in Fig. 2. Unfortunately, unsupervised analysis did not provide tight and well-separated clusters based on the microbial origin. Most bacteria are found in the middle section of the PCA plot, while fungal and viral samples are in the left portion of the PCA plot. However, clusters are partially merging and coinciding, hence, we have to conclude that there is only a limited natural clustering observed in this data set. The loading components of this separation are shown in the bottom portion of Figure 2. The total variance for component 1 is 38.073 %; for component 2 the total variance is 64.883 %. The number of the peak in the loading is coincidental with the number assigned to individual fatty acid in Table II. Specimens, whose position is indicated by positive values on the factor plots, are characterized by peaks in the upper portion of the loading plot. Inversely, negative values on the PCA plot are associated with the bottom portion of the loading plot.

An additional approach for categorization was also used while comparing the same set of specimens. Each microorganism was assigned individually with one letter/number as indicated in Table I. The assignment of different categories does not affect the clustering process during the PCA. All analytical results were collected in triplicate and no outliers were removed. All data were mean centered and normalized to total intensity prior to PCA. The PCA plot shows the same clusters with individual microorganisms (each analyzed in triplicate) as in the categorization approach based on the origin. However, this plot provides additional information about the proximity of the three analyses of individual microorganisms; all triplicates are closely clustered together. The loadings are identical to those used for the first categorization approach.

Table II List of identified fatty acids in the examined microorganisms with assigned ID numbers. The identity of the fatty acids was confirmed by mass spectrometry and NIST mass spectral database. The FAMEs are abbreviated in the following manner: the number following the letter C indicates the total number of carbons in the fatty acid chain. The presence of double bonds is indicated by a 2n symbol, and the number following indicates the number of the double bonds. The word "(Cyclo)" indicates the presence of a cyclopropyl group in the fatty acid chain, (diCOOH) indicates dicarboxylic acid and (oxo) indicates carbonyl group in fatty acid chain.

	Fatty acids	
ID #	Ret. Time	FAME
1	7.46	C8
2	8.36	C8
3	9.93	C9
4	10.83	C8(oxo)
5	11.35	C10
6	12.29	C9(oxo)
7	13.5	C12
8	13.79	C9(diCOOH)
9	14.34	C13
10	14.44	C13
11	14.79	C13
12	15.53	C14
13	15.86	C14(2n:1)
14	16.01	C14
15	16.72	C15
16	16.82	C15
17	17.13	C15
18	17.56	C16(2n:1)
19	17.67	C16(2n:1)
20	17.81	C16
21	17.97	C16(2n:1)
22	18.16	C16(2n:1)
23	18.35	C16
24	18.64	C17(2n:1)
25	18.79	C17(cyclo)
26	18.87	C17
27	18.97	C17
28	19.03	C17(cyclo)
29	19.13	C17(cyclo)
30	19.26	C17
31	19.54	C18(2n:1)
32	19.81	C18(2n:3)
33	19.9	C18
34	19.98	C18(2n:2)
35	20.02	C18(2n:1)
36	20.07	C18(2n:1)
37	20.28	C18
38	20.43	C18(2n:2)

Table II – Continued

	Fatty acids	
ID #	Ret. Time	FAME
39	20.82	C18(2n:2)
40	20.94	C19(2n:1)
41	20.94	C19(cyclo)
42	21.18	C19
43	21.53	C20(2n:4)
44	21.68	C20(2n:3)
45	21.84	C20(2n:1)
46	22.07	C20
47	22.93	C21
48	23.08	C22
49	24.76	C23
50	25.84	C24

Since strong natural clustering was not observed using the PCA approach, LDA was employed as an alternative strategy. During LDA, a mathematical model is developed with pre-assigned categories and this model is used for classification of unknown samples. Figure 3 shows results from LDA. As in PCA, the results are expressed as two dimensional plots of factors. Each plot is characterized by a set of two components (1 and 2), which show the distribution of data. Components 1 and 2 are determined by the canonical variate plots, a graphic representation of the elements responsible for the difference in the data. The LDA resulted in very well defined model of differences in the microorganisms based on their origin. There are three well defined clusters, b-bacteria, v-viruses, and F-fungi. All the clusters are tight and well differentiated, thus making this identification model a good one for classification purposes. At the bottom of the Fig. 3 are the canonical variates 1 and 2. The bacteria cluster is defined by the negative values of component 1 (the lower portion of the canonical variate 1 lists ID numbers of the characteristic fatty acids as listed in Table II) and positive values of component 2 (the upper portion of the canonical variate 2 lists ID numbers of the characteristic fatty acids as listed in Table II). The viral cluster is characterized by positive values of component 1 (the upper portion of the canonical variate 1 lists ID numbers of the characteristic fatty acids as listed in Table II) and negative values of component 2 (the lower portion of the canonical variate 2 lists ID numbers of the characteristic fatty acids). The fungal cluster is characterized by positive values of component 1 (the upper portion of the canonical variate 1 lists ID numbers of the characteristic fatty acids as listed in Table II) and positive values of component 2 (the upper portion of the canonical variate 2 lists ID numbers of the characteristic fatty acids as listed in Table II).

During the next LDA, the identification of Gram-positive and Gram-negative bacteria was examined. The results are shown in Fig. 4. There are four well-differentiated clusters: + signs indicate the Gram-positive bacteria, – signs

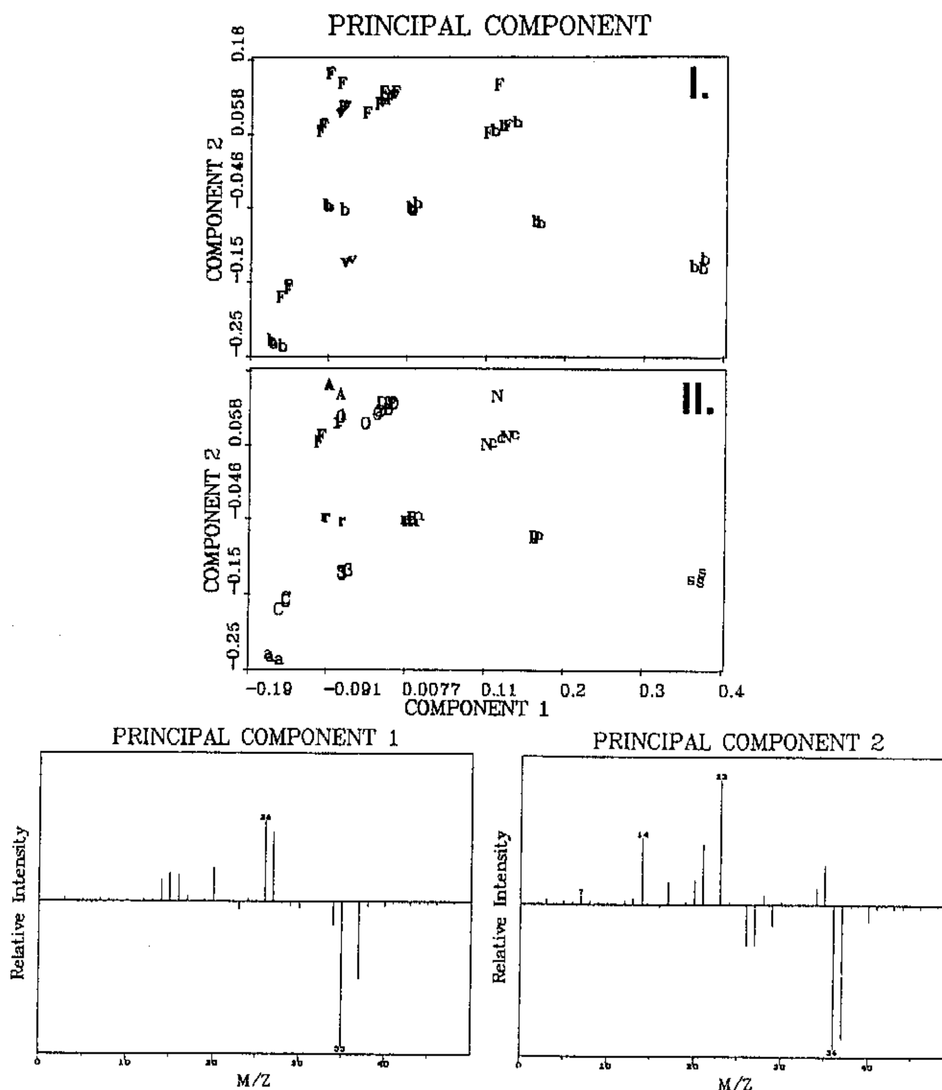


Fig. 2 PCA results from the origin-based classification of microorganisms based on the GC/MS results with three categories (I; bacteria [b], viruses [v], and fungi [F]) and individual microorganisms categories (II; abbreviations of individual microorganisms are listed in Table I). The loadings for both PCA are found in the lower portion of the figure

indicate the Gram-negative bacteria, v indicates viruses and F indicates fungi. The interpretation of the canonical variates is done the same way as in Fig. 3.

PCA was also employed for the differentiation of individual species of microorganisms. Figures 5 and 6 show the results of such analyses: the *Bacillus*

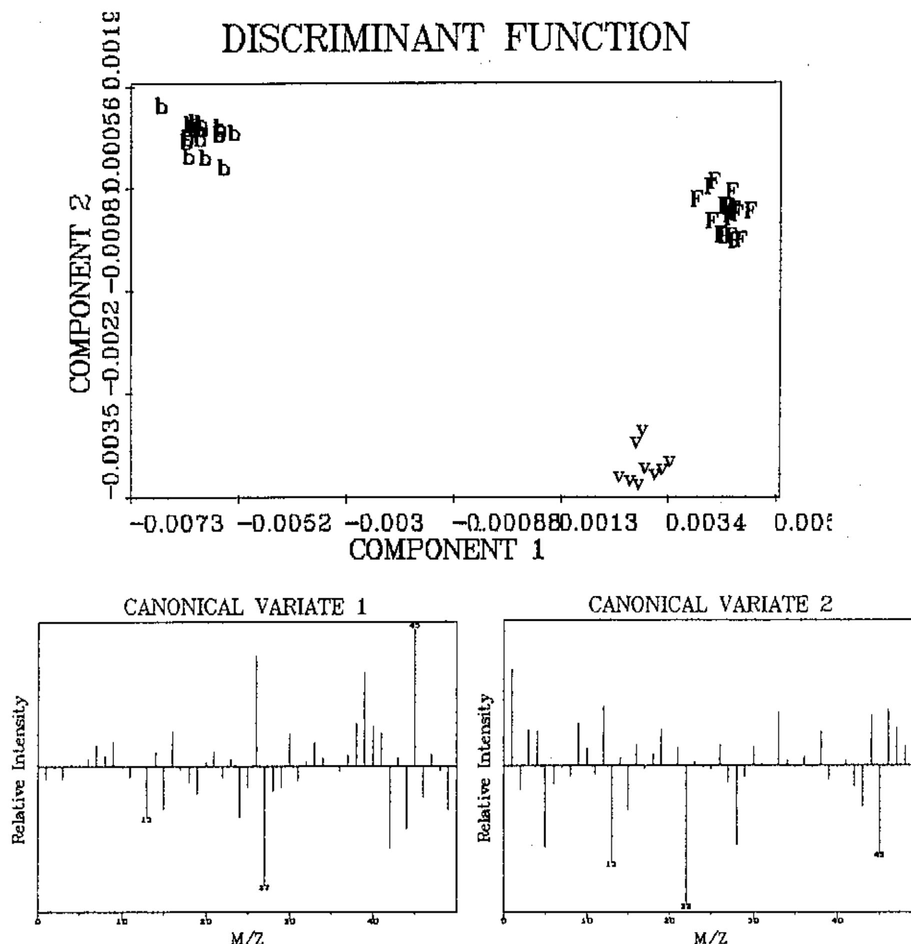


Fig. 3 LDA generated model for classification of microorganisms into three groups: bacteria (b), viruses (v), and fungi (F). Canonical variates 1 and 2 summarizing the differences responsible for the separation of the clusters are shown in the lower part of the figure

and *Aspergillus* genera were examined for their natural clustering capabilities. In Fig. 5, the *Bacillus* genus is represented by four different species: *B. cereus* (c), *B. subtilis* (s), *B. circulans* (r), and *B. pumilus* (p). The PCA results show four well-defined clusters for each individual species (component 1: 58.426 % of total variance; component 2: 83.463 % of total variance). The loadings for the PCA components are shown in the lower part of Fig. 5. Figure 6 shows the PCA results from the *Aspergillus* genus separation. There are four different *Aspergillus* spe-

cies: *A. niger* (N), *A. clavatus* (C), *A. nidulans* (N) and *A. fumigatus* (F). The best separation was achieved when using component 1 (63.377 % total variance) and component 2 (94.659 % total variance). Well-defined clusters for each species of *Aspergillus* genus are observed. The loadings for the PCA components are shown in the lower part of Fig. 6.

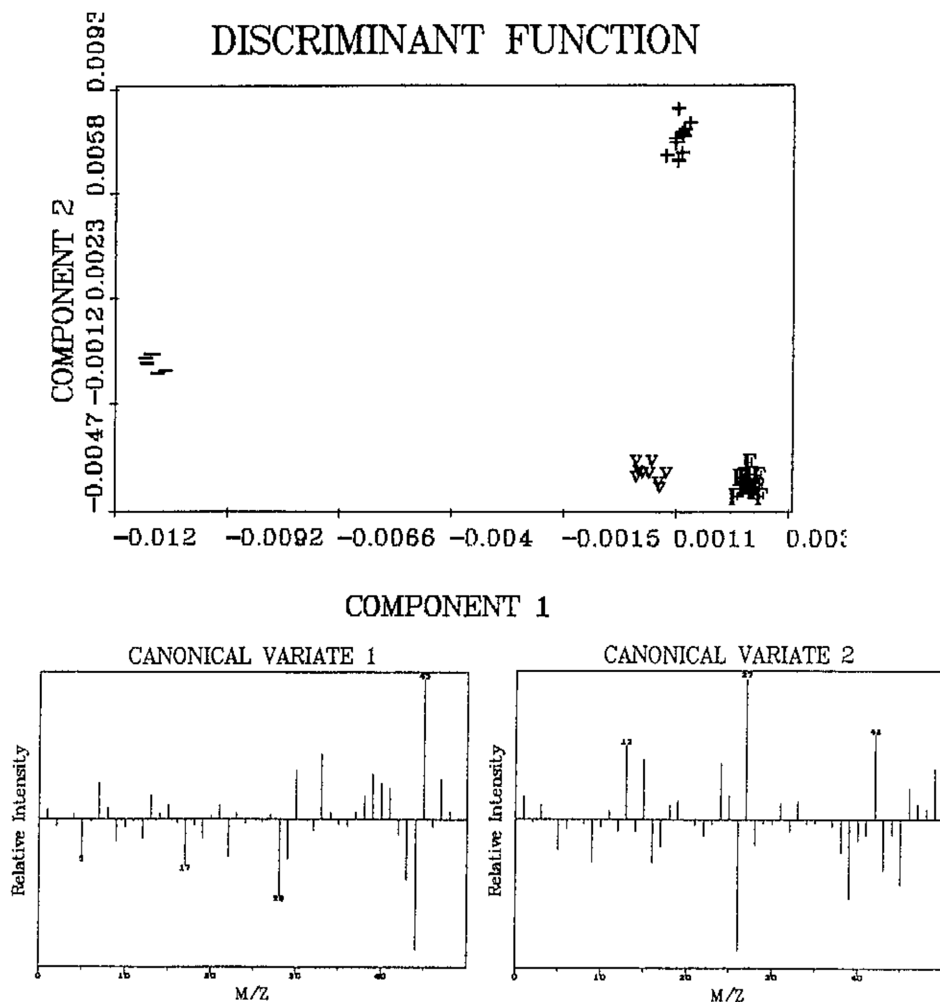


Fig. 4 LDA generated model for classification of microorganisms into four groups: Gram-positive bacteria (+), Gram-negative bacteria (–), viruses (v), and fungi-F. Canonical variates 1 and 2 summarizing the differences responsible for the separation of the clusters are shown in the lower part of the figure

Results from FAB/MS analysis. The FAB/MS lipid profiles were evaluated in the following manner: since we did not have triplicates of each measurement, the statistical analysis could not be employed. The mass region of complex lipids was examined, since the previously published literature suggested that these molecules might produce unique biomarkers. The mass region from 400 to 1000 Da was examined in detail. Peaks that were found in all specimens in the bacterial category were defined as biomarker peaks for bacterial presence; peaks found in

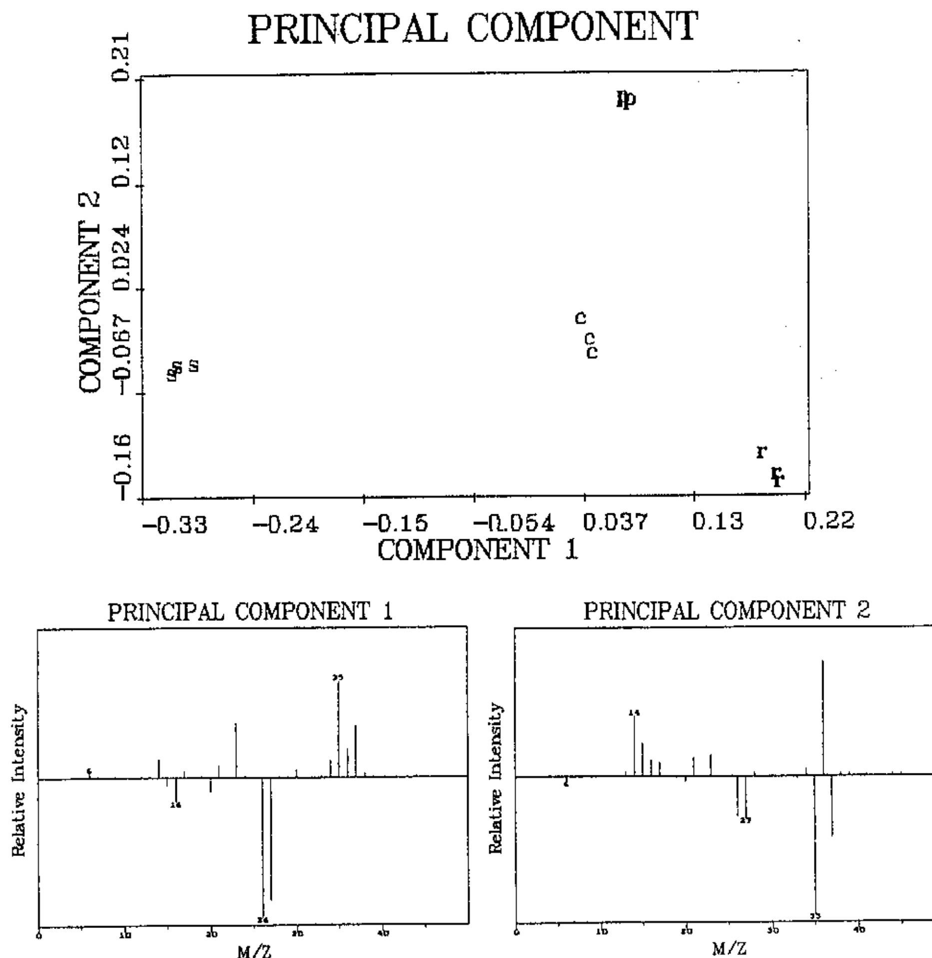


Fig. 5 PCA results from the species-based classification of *Bacillus* genus represented by four different species: *B. cereus* (c), *B. subtilis* (s), *B. circulans* (r), and *B. pumilus* (p). PCA results show four well-defined clusters for each individual species (component 1: 58.426 % of total variance; component 2: 83.463 % of total variance)

all viruses were defined as biomarkers for viral presence; and peaks found in all fungi were defined as biomarkers for fungal presence. The taxonomical classification can be accomplished by the presence/absence of the biomarker peaks. The results from this recognition algorithm is presented in Table III.

The advantage of this algorithm in comparison with previously published recognition technique [7] is the presence of multiple biomarkers, which reduce the number of false positive identifications.

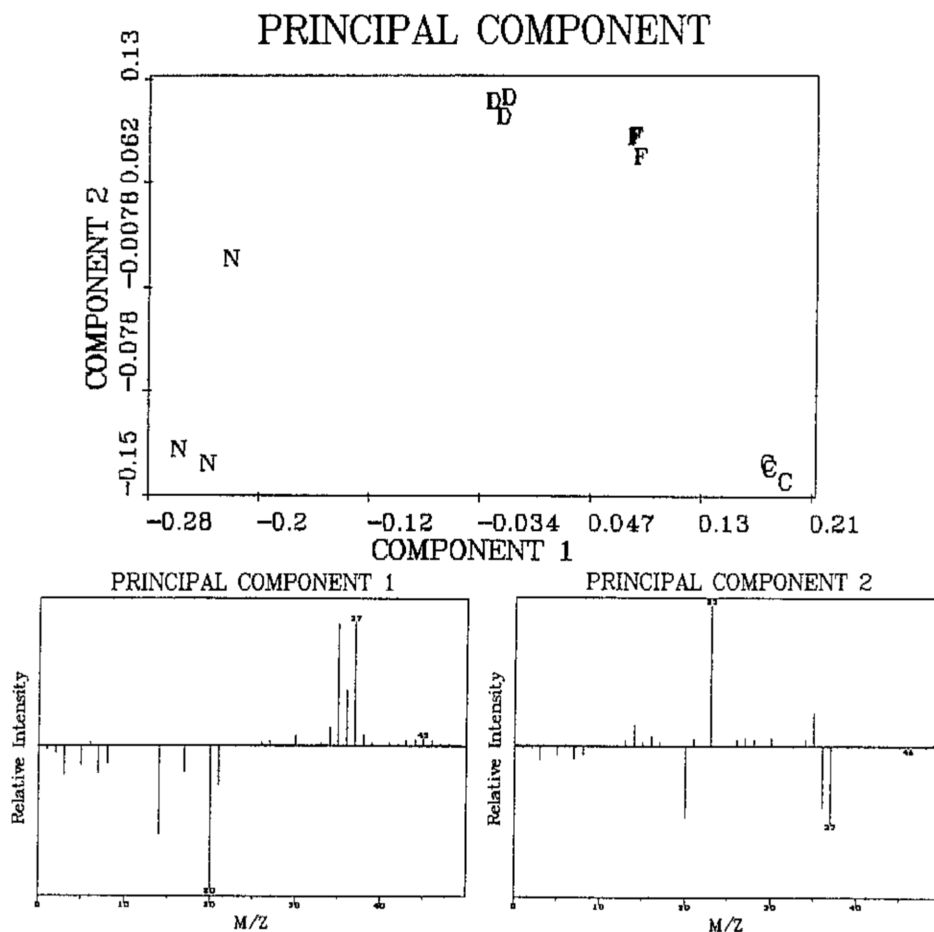


Fig. 6 PCA results from the *Aspergillus* genus separation represented by four different *Aspergillus* species: A. *niger* (N), A. *clavatus* (C), A. *nidulans* (N) and A. *fumigatus* (F). The best separation was achieved when using component 1 (63.377 % total variance) and component 2 (94.659 % total variance). Well-defined clusters for each species of *Aspergillus* genus are observed. The loadings for the PCA components are shown in the lower part of the figure

Table III Characteristic lipid derived FAB/MS peaks that are used as an algorithm for classification of microorganisms based on their origin. The classification is accomplished by observing the presence/absence of the biomarker peaks

Characteristic lipid derived FAB/MS peaks		
Bacteria	Viruses	Fungi
430	-	430
439	439	-
-	522	522
525	-	-
535	-	-
552	552	-
-	575	575
-	578	578
-	579	579
-	601	601
-	603	603
-	604	604
-	605	605
-	706	706
-	734	734
-	759	759
-	788	788

Conclusion

The results from this pilot study showed that lipid-derived biomarkers are suitable for distinguishing the microbial origin of microorganisms as bacterial, fungal or viral. GC/MS profiles of 15 different microorganisms were evaluated using pattern recognition software. There was a limited natural clustering observed in the PCA,; therefore, an alternative approach was used for classification. A taxonomical model was constructed using the results from LDA, which has a good working resolution for different microorganisms. Also, the model can be used for classification of Gram-positive and Gram-negative bacteria. PCA was also employed in classification of different species within the same genus for the *Bacillus* and *Aspergillus* genera. Natural clustering allowed separation of all four species in each genus. The utility of complex lipid profiles and their classification potential generated during the FAB/MS analyses was also examined. A set of biomarker peaks was defined for each of the categories. This methodology must be verified on a larger data set to ensure the reproducibility for all bacteria, fungi and viruses.

Acknowledgements

This work was partially supported by the University of Arizona Foundation and the Office of the vice President for Research and Graduate Studies at the University of Arizona (grant # 210109), Tucson, Arizona and by the Arizona Disease Control Research Commisssion grant # 6014. Support of the Grant Agency of the Czech Republic (project # 203/02/0023) is also appreciated.

References

[1] Larsson L., Saraf A.: Mol. Biotechnol. **7**, 279 (1997).
[2] Barshick S.A., Wolf D.A., Vass A.A.: Anal. Chem. **71**, 633 (1999).
[3] Cole M.J., Enke C.G.: Anal. Chem. **63**, 1032 (1991).
[4] Xiang F., Anderson G.A., Veenstra T.D., Lipton M.S., Smith R.D.: Anal. Chem. **72**, 2475 (2000).
[5] Vater J., Gao X., Hitzeroth G., Wilde C., Franke P.: Comb. Chem. High Throughput Screening. **6**, 557 (2003).
[6] Wahl K.L., Wunschel S.C., Jarman K.H., Valentine N.B., Petersen C.E., Kingsley M.T., Zartolas K.A., Saenz A.J.: Anal. Chem. **74**, 6191 (2002).
[7] Heller D.N., Murphy C.M., Cotter R.J., Platt J.A., Uy O.M., Fenselau C.: *Advances in Mass Spectrometry*, P. Heydon & Sons Ltd., London, 1989.
[8] Basile F., Beverly M.B., Abbas-Hawks C., Mowry C.D., Voorhees K.J., Handfieldl T.L.: Anal. Chem. **70**, 1555 (1998).
[9] Harrington P.B.: RESOLVE Tutorial, Colorado School of Mines, 1990.
[10] Beebe K.R., Pell R.J., Seasholtz M.B.: *Chemometrics: A Practical Guide*, John Wiley & Sons, Inc., New York, 1999.