Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Rapid detection of taxonomically important fatty acid methyl ester and steroid biomarkers using in situ thermal hydrolysis/methylation mass spectrometry (THM-MS): implications for bioaerosol detection

Angelo J. Madonna a, Kent J. Voorhees a,*, Ted L. Hadfield b

a Department of Chemistry, Colorado School of Mines, Golden, CO 80401, USA
b Armed Forces Institute of Pathology, Washington DC 20306-6000, USA

Received 26 September 2000; accepted 21 February 2001

Abstract

Implications for the rapid interrogation of biological materials collected from the atmosphere using a simple, one step, sample preparation technique was explored. For this purpose, various samples of whole bacteria, fungi, pollen, media contaminated with viruses, and proteins were treated with an aliquot of methanolic tetramethylammonium hydroxide prior to thermal introduction into the ion source of a triple quadrupole mass spectrometer. Molecular and fragment ions, consistent with fatty acid methyl esters (FAMEs) and steroids (non-methylated and methylated), generated during electron ionization (70 eV) of the volatile hydrolysates were subsequently detected. The varying distributions and relative intensities of these ions were used to discriminate between the different biological samples. More specifically, it was found that polyunsaturated FAMEs and steroids could be used to differentiate eukaryotic cells from prokaryotic cells since the latter do not generally synthesize either of these lipid membrane constituents. Further discrimination of the different eukaryotic samples was made based on the detection of ergosterol for fungi, cholesterol for the viral media, and C18:3Me for pollen. Multivariate statistical analysis was employed to evaluate and compare the large set of mass spectra generated during the study and to build a trained model for predicting the class membership of test samples entered as unknowns. Of 132 different samples subjected to the model as unknowns, 131 were correctly classified into their proper biological categories. Moreover, 29 out of 30 bacteria test samples representing five species

* Corresponding author.
E-mail address: kvorhee@mines.edu (K.J. Voorhees).
of pathogenic bacteria were correctly classified at the species level. © 2001 Elsevier Science B.V. All rights reserved.

**Keywords**: Thermal hydrolysis/methylation; Pyrolysis; FAMES; Steroids; Bioaerosols; Bacteria; Fungi; Pollen; Viruses; Proteins; Tetramethylammonium hydroxide

---

**1. Introduction**

Airborne microorganisms are ubiquitous in nature and exposure to either pathogenic or opportunistic species is known to cause a variety of adverse health effects [1–3]. In the interest of public health, rapid and accurate methods for differentiating and identifying aerosolized bacteria, fungi, and viruses are needed. Early recognition of airborne pathogens would be extremely important in preventing outbreaks of infectious diseases. Microbial identification systems designed for this purpose could effectively be used in industrial/agricultural settings, health facilities, and locations where large groups of people assemble for work, leisure, or travel. Mounting concerns regarding the deliberate use of biological substances as weapons by state-sponsored organizations and terrorist groups [4,5] also places a high priority on fast, reliable techniques for detecting airborne pathogenic microorganisms.

Atmospheric sampling of microorganisms is generally achieved by impaction onto a solid growth media or an adhesive surface, filtration through a filter membrane, or impingement into a liquid media [6–8]. The ensuing analysis of the collected bioaerosol has traditionally relied on culture-based techniques or direct microscopic examination. These standard microbiological techniques require a considerable amount of skill to perform, are time-consuming, and cannot be automated or developed as online, real-time monitoring procedures [9]. Furthermore, culture methods are limited to detecting only those organisms that can be cultivated on the specific media employed for the analysis. Therefore, nonculturable and nonviable airborne pathogens escape detection and give false-negative results for the assay.

Biochemical analysis using modern analytical instrumentation is an attractive alternative to conventional microbiological techniques. In particular, methodologies utilizing mass spectrometry display great promise. Mass spectrometry is well established as a rapid microbial characterization/identification method [10–12]. This technique can be used to search for a large number of diverse biological and chemical agents in a matter of minutes including nonculturable and nonviable organisms.

Procedures using mass spectrometry attempt to differentiate and identify biological specimens based on the detection of specific chemical markers. Using this chemotaxonomic approach, organisms are classified according to intrinsic differences in their individual chemical constituents. A chemical and/or physical prepro-
cessing step is usually used to release endogenous biochemical compounds producing a unique set of chemical signals of taxonomic significance. Subsequent mass analysis of these liberated chemical markers results in a distinguishing mass spectrum that can be used to ‘fingerprint’ and identify microorganisms.

Cellular lipids have been used extensively as signatory biomarkers for the chemical identification of microorganisms [13–15]. Lipid biomarkers, and in particular, the relative concentration and distribution of the fatty acid moieties, have long been recognized for their chemotaxonomic importance [16,17]. Classification, based on varying distributions of fatty acids, has become a routine method in microbiology laboratories. A typical method involves saponification of whole microbial cells releasing glycerol bound fatty acids followed by methylation and subsequent extraction of the derivatized fatty acid methyl esters (FAMEs) into a nonpolar organic solvent. Separation and analysis of the FAME extract is then completed by gas chromatography (GC). Exemplifying this technique is the well known commercialized Microbial Identification System (MIDI) [18]. Here, GC analysis and data interpretation are completely automated and a ranked list of probable matches based on the comparison of GC-FAME extract profiles to a database of known organisms is made available to the user.

The entire GC-based FAME procedure is extensive, requiring ~1.5 h per sample, in which 20–30 min are necessary for the chromatographic analysis and another 50–60 min are needed for the preceding sample preparation. Efforts to reduce the sample preparation time were successfully accomplished by Dworzanski [19,20] and Holzer [21]. Both research groups utilized an on-line derivatization procedure in which whole bacterial cells, in the presence of strongly basic methylating reagents (e.g. tetramethylammonium hydroxide or trimethylanilinium hydroxide), were pyrolyzed in the injector block of the GC. Thermal hydrolysis and subsequent methylation of the cellular lipids was observed to occur nearly instantaneously during each heating cycle. The resulting chromatograms of the volatile methyl derivatives showed a high degree of similarity to FAME profiles obtained from extracts prepared by conventional wet-chemistry methods. Aside from coating a pyrolysis filament with a mixture of sample and derivatizing reagent, no further sample manipulation was necessary, hence reducing the chance of losing or contaminating the sample. The thermal hydrolysis/methylation (THM) technique required only 5–10 µg of sample, whereas the wet chemistry procedure required 10–15 mg of sample. Therefore, the THM process is better suited for situations involving low levels of analyte. A thorough review of on-line derivatization procedures using quaternary ammonium salts has been published [22].

More recently, the in-situ THM procedure has been utilized for the analysis of whole cell bacteria in which the volatilized hydrolysates were introduced directly to the mass spectrometer eliminating the chromatographic step [23–25]. In these studies, sample preparation and mass analysis were accomplished in less then five minutes producing spectral profiles featuring taxonomically significant biochemical signals. For the present study, use of this rapid technique was extended to the analysis of intact fungi, pollen, viral infected cell culture media, and proteins. These biological samples were chosen to emulate constituents normally found dispersed in
the atmosphere. The aim of this current report was to identify characteristic biomarkers that could facilitate the differentiation of these samples from each other and from whole cell bacteria. For comparison, five known human pathogenic bacteria were included in the analysis. Positive identification of pathogenic bacteria collected during atmospheric sampling must include an initial step that is capable of differentiating the prokaryotic organism from a multitude of other possible airborne constituents. The simplicity and speed of THM-MS makes it an attractive approach to solving this problem.

2. Experimental section

2.1. Instrumental

All measurements were performed using an Extrel ELQ400 triple quadrupole mass spectrometer modified with a Curie-point pyrolysis inlet. The basic instrumental design and layout has previously been described [26]. Mass spectra were collected in the positive electron ionization (EI) mode at 70 eV electron energy over the scan range \( m/z \) 180–450. This spectral range was selected in order to emphasize FAME and steroid molecular ions. Mass spectra were collected and stored by averaging twenty scans from the pyrogram. Curie-point filaments used for sample introduction were heated to 358°C (rise time 100 ms) and held for 10 s. The transfer line interface (5-cm-long glass tube between the pyrolysis chamber and ion source) and the ion source were both maintained at 140°C.

GC-MS data were acquired on the Extrel ELQ 400 system described above using similar operating parameters. The interface between the GC and mass spectrometer was maintained at 225°C. Separations were performed using a J&W Scientific (Folsom, CA) DB-5 column (30 m × 0.320 mm × 0.25 μm). The GC oven was held at 150°C for 3 min and then ramped 5°C min \(^{-1}\) to 250°C with a final hold of 2 min. FAME mixtures were injected as 2 μl aliquots at a temperature of 225°C with a split ratio of 40:1. Lipid extractions and subsequent saponification and methylation were preformed according to the procedure outlined in the MIDI manual [18].

3. Reagent and biological samples

Tetramethylammonium hydroxide (TMAH) (Aldrich), prepared as a 0.1 M solution in HPLC grade methanol, was used as the derivatizing reagent in all in-situ THM reactions. Biological samples used in this study are listed in Table 1. Five replicate mass spectra were generated for each different sample by randomly alternating the order in which replicates of different samples were run. Bacteria samples were obtained from a continuous cell-line maintained at the Armed Forces Institute of Pathology (AFIP). Cells were cultured using standard procedures and harvested by washing with distilled water. The cells were
lyophilized and transferred to our laboratory where they were stored at $-20^\circ$C until use. To ensure safe handling, each bacterial sample was gamma killed in a cobalt radiation source prior to shipment. Fungi samples purchased from the American Type Culture Collection (ATCC) were grown on potato dextrose agar for 8 days at room temperature. Mycelia were carefully removed from the agar surface, washed with distilled water, freeze dried and stored at $-20^\circ$C. Lyophilized samples of cell culture media containing viruses were also received from the AFIP. A detailed description of how the media were prepared was previously published [27]. Pollen samples were hand collected from various locations in the Rocky Mountain region. Mortar and pestle were used to crush the dried pollen into a fine powder. A total of 10 mg of dried samples of bacteria, fungi, viral media, and pollen were resuspended in 1 ml of deionized water prior to application to the Curie-point filaments. The fungi suspensions were sonicated for 15 s in order to break apart the mycelia and produce more homogenous solutions. Standard protein samples (Sigma) were prepared as 1

Table 1

| Sample                  | Source       | PCA code |
|------------------------|--------------|----------|
| Bacteria               |              |          |
| *Bacillus anthracis*   | AFIP         | X        |
| *Yersinia pestis*      | AFIP         | Y        |
| *Vibrio cholera*       | AFIP         | C        |
| *Brucella melitensis*  | AFIP         | B        |
| *Francisella tularensis* | AFIP     | F        |
| Fungi                  |              |          |
| *Fusarium oxysporum*   | ATCC #48112  | 1        |
| *Penicillium camembertii* | ATCC #4845 | 2        |
| *Mucor rouxii*         | ATCC #24905  | 3        |
| *Aspergillus versicolor* | ATCC #9577  | 4        |
| *Rhizopus homothallicus* | ATCC #42221 | 5        |
| *Alternaria alternata* | ATCC #6663   | 6        |
| *Cladosporium herbarum* | ATCC #28987 | 7        |
| Viral media            |              |          |
| *Venezuelan equine encephalitis* | AFIP | v        |
| *Feline enteric coronavirus* | AFIP | v        |
| *Vero cells (noninfected)* | AFIP | v        |
| Pollen                 |              |        |
| *Sunflower*            | Private collection | #    |
| *Soybean*              | Private collection |        |
| *Mixed Colorado*       | Private collection | *    |
| *Cottonwood*           | Private collection | =    |
| Proteins               |              |          |
| *Albumin, bovine serum* | Sigma #A-8531 | t       |
| *Alcohol dehydrogenase* | Sigma #A-8656 | t       |
| *Carbonic anhydrase*   | Sigma #C-7025  | t        |
mg ml$^{-1}$ suspensions. Aliquots of 10 µl of each sample suspension were applied to Curie-point filaments, air-dried, and overlaid with 10 µl of the methanolic derivatizing reagent. The filaments were then introduced via direct inlet probe into the mass spectrometer.

4. Multivariate analysis

Multivariate analysis was performed with the Resolve Software Package developed at the Colorado School of Mines [28]. Mass spectra were collected as a set of raw intensities and normalized to constant length to correct for variations in total ion current. The following equation was used:

$$x_{i,j}^* = \frac{x_{i,j}}{1000/\sqrt{\sum_{j=1}^{n} x_{i,j}^2}}$$

where $x_{i,j}^*$ is the normalized intensity of mass spectrum $i$ and mass $j$, 1000 is a constant scaling factor and $n$ is the number of masses in the spectrum. Mass peaks corresponding to $m/z$ 207, 267, 281, 327, 341, and 355 were omitted (silicone contamination) from the calculations. The multidimensional data set was mean centered and reduced by principal component analysis (PCA) to a smaller collection of factors still containing a significant proportion (~95%) of the total variance. Subsequent examination of this reduced set of factors or principal components (PCs) was conducted using linear discriminant analysis (LDA). In this way, mass spectra from samples listed in Table 1 (five replicates of each run in random order) were used to train a model for classifying unknown samples. Test sets were produced from the same samples examined in triplicate and submitted, as unknowns, to the LDA model for classification. Samples for the training set, first test set, and second test set, were analyzed on July 21, August 3, and August 19, 1999, respectively. Cross-validation was used as an assurance against over fitting the trained models with measurement noise [29].

5. Results and discussions

5.1. THM-MS of fungi

The THM mass spectra shown in Fig. 1 represent five different genera of fungi frequently encountered in the atmosphere [2]. For each fungal species studied, a distinctive series of peaks ($m/z$ 410, 396, 378, 363, 337, and 253) were detected and identified as products of electron ionization of ergosterol. The mass peak observed at $m/z$ 396 is the molecular ion of the free steroid while the mass peak observed at $m/z$ 410 corresponds to the methylated derivative. The incomplete derivatization of ergosterol is attributed to the fact that steroids are weakly acidic and as such do not favorably methylate to form ethers during the THM process [22]. Chemically, this process involves the liberation of bound ergosterol and fatty acids by saponification.
Fig. 1. Representative THM mass spectra from whole cell fungi. (a) *A. alternata*; (b) *A. versicolor*; (c) *F. oxysporum*; (d) *M. rouxi*; (e) *C. herbarum*. 
and transesterification (competing mechanisms) following the application of an aliquot of a methanolic solution of tetramethylammonium hydroxide to the intact fungal cells. The released fatty acids (as well as any naturally free fatty acids), being more acidic, are readily deprotonated in this highly basic medium and form salts with the quartenary ammonium cations. These salts thermally degrade (nucleophilic attack of the carboxylate anion upon the quartenary N-methylammonium cation) to FAMEs and tertiary amines when heated to a sufficiently high temperature.

In general, steroid biomarkers allow quick differentiation of eukaryotic (higher order organisms including mammalian, fungal, and vegetative) from prokaryotic (bacteria) cells, since the latter do not carry the biosynthetic machinery required to make steroids. Moreover, ergosterol is produced almost exclusively by fungi making it an excellent chemotaxonomic marker for differentiating fungi from other eukaryotic organisms.

Besides the group of ergosterol molecular and fragment ions, a series of electron ionization ions from FAMEs were also detected for the whole cell fungi. The predominant fatty acids synthesized by fungi are straight-chained and 16 or 18 carbons in length [30]. For each fungus analyzed, molecular ions for the methyl esters of hexadecanoic acid (C16:0), octadecanoic acid (C18:0), octadecenoic acid (C18:1), and octadecadienoic acid (C18:2) were detected (Fig. 1). Unlike the situation with ergosterol, near complete thermal methylation was observed for the cellular fatty acids. A compiled list of FAMEs and their observed electron ionization molecular and fragment ions are listed in Table 2. The common fragment ions detected included the series of carbomethoxy ions (185, 199, 213, 227, etc.) formed during electron ionization of saturated FAMEs. Also originating from

| Unsaturated FAME mass peaks detected by THM-MS |
| Monounsaturated FAME mass peaks detected by THM-MS |
| FAME | M^+ | [M-31]^+ | FAME | M^+ | [M-32]^+ | [M-74]^+ |
| C26:0Me | 410 | 379 | C24:1Me | 380 | 348 | 306 |
| C25:0Me | 396 | 365 | C18:1Me | 296 | 264 | 222 |
| C24:0Me | 382 | 351 | C16:1Me | 268 | 236 | 194 |
| C23:0Me | 368 | 337 | | | | |
| C22:0Me | 354 | 323 | Cyclopropane | | | |
| C21:0Me | 340 | 309 | FAME | M^+ | [M-32]^+ | [M-74]^+ |
| C20:0Me | 326 | 295 | cyC19:1Me | 310 | 278 | 236 |
| C19:0Me | 312 | 281 | cyC17:1Me | 282 | 250 | 208 |
| C18:0Me | 298 | 267 | | | | |
| C17:0Me | 284 | 253 | Polynsaturated | | | |
| C16:0Me | 270 | 239 | FAME | M^+ | | |
| C15:0Me | 256 | 225 | C18:2Me | 294 | | |
| C14:0Me | 242 | 211 | C18:3Me | 292 | | |

Table 2

Carbomethoxy ions
m/z 185, 199, 213, 227, 241, 255, ...
saturated FAMEs, but lower in abundance, is the series of fragment ions detected 31 Da. ([M-31]+) below the molecular weight of the precursor fatty acid. Mechanistically, these ions are created by homolytic cleavage of the methoxy group resulting in the formation of a stable acylium ion. For unsaturated FAMEs, the two major electron ionization fragment ions were observed at 32 Da. ([M-32]+) and 74 Da. ([M-74]+) below the molecular mass of the analyte. Generally, the M-32 ion is the base peak when electron ionization at 70 eV is used to analyze FAMEs containing double bonds.

Detection of C18:2, which is produced by most eukaryotic cells, is a valuable signatory marker since prokaryotic organisms usually do not produce polyunsaturated fatty acids. Therefore, the detection of FAMEs with two or more double bonds could be used to corroborate the presence of eukaryotic organisms. In addition, if mass peaks from ergosterol are also detected, then it may be inferred that the sample contains fungal cells. Unfortunately, fungi do not produce a large variety of fatty acids, as such; there is not much variation between different FAME profiles [31]. This limits the ability of using THM-MS to identify whole fungal cells at the genus or species level. However, the THM mass spectra exhibited by fungi do differ significantly from profiles generated by bacteria, pollen, viral infected media, and proteins.

5.2. THM-MS of viruses

Viruses are parasitic in nature. Because of this host/guest relationship, interest was focused, not on a purified viral preparation, but rather, on media recovered (supernatant) from cell cultures used to propagate the viruses. Two such media were examined by THM-MS and the mass spectra are shown in Fig. 2(a–b). These two media were inoculated separately, one with feline enteric coronavirus (FECV), and the other with Venezuela Equine Encephalitis (VEE). For comparison, a non-infected medium (same medium used to propagate VEE) was also analyzed and is shown in Fig. 2(c). From an earlier study, it was found that viruses do not contribute significantly to the overall mass spectrum from cell culture media [27]. The relative contribution of the viruses to the total mass of the cell culture medium is only about a tenth of a percent and accordingly does not generate an intense set of ion signals (relative to the cell culture media). Hence, the two mass spectra (Fig. 7(b–c)) of media recovered from Vero cell cultures (one with the virus the other without) are essentially indistinguishable. It was further discovered that the major source of fatty acids and steroids detected from the media supernatant originated from blood serum. Serum is commonly added to cell cultures as a rich supplement of proteins and lipids.

The FECV sample shown in Fig. 2(a) was taken from a feline kidney cell culture supplemented with 10% (v/v) horse serum while the VEE sample (Fig. 2(b)) was recovered from a Vero cell culture supplement with 5% (v/v) fetal bovine serum. While both spectra show peaks consistent with the methyl esters of C16:0 (M+, m/z 270), C18:0 (M+, m/z 298), and C18:1 (M+, m/z 296 and [M-32]+, m/z 264), only the FECV spectrum contains a peak indicative of methylated C18:2 (M+, m/z 294).
Fig. 2. THM mass spectra of supernatants recovered from cell culture media. (a) Feline kidney cell culture infected with FECV. (b) Vero cell culture infected with VEE. (c) Non-infected Vero cell culture.

Comparable results were obtained for the THM mass spectra of just the horse and fetal bovine sera (data not shown) demonstrating their overall strong contributions to the culture media.

The series of peaks in the upper scan range ($m/z$ 386, 368, and 353) are known electron ionization products of cholesterol; the most abundant steroid found in
mammalian blood [32]. Again, no ions were detected indicating the formation of the methyl ether derivative of cholesterol. The detection of cholesterol indicates the presence of mammalian type cellular systems. This biomarker has also been detected from a purified mouse hepatitis virus [23] further substantiating its general use as a mammalian/viral signatory biomarker. For enveloped viruses, cholesterol is incorporated within the lipid bilayer that covers the viral protein coat. Furthermore, the virus acquires both steroids and lipids from the host and therefore the composition of these biomarkers will resemble the media that was used in propagation. It can further be postulated that if a selective trapping method is employed to capture airborne viruses prior to THM-MS analysis then mass spectra similar to those shown in Fig. 4 will be generated and directly attributed to the viruses.

5.3. THM-MS of pollen

Pollen grains are normal constituents found in relatively high concentrations in the atmospheric load of particulant matter. Spectra from four different pollens, indigenous to the Rocky Mountain area, are shown in Fig. 3. Similar to other eukaryotic organisms, the fatty acids synthesized by pollen are also predominately 16 and 18 carbons in length with varying degrees of unsaturation. The methyl ester of hexadecanoic acid (M⁺, m/z 270) is the base peak in each of the pollen mass spectra. Furthermore, each different pollen investigated exhibited a unique distribution of saturated (C18:0Me, M⁺ m/z 298) monenoic (C18:1ME, M⁺ m/z 296), dienoic (C18:2, M⁺ m/z 294), and trienoic (C18:3Me, M⁺ m/z 292) methylated fatty acids of 18 carbons atoms in length. Similar FAME profiles of pollen from maize [33], tea [34], and rapeseed [35] have been reported. The distinct FAME profiles not only permit a quick way to differentiate the different types of pollen shown in Fig. 3, but also can be used to separate this taxonomic group of airborne constituents from bacteria and other potential air contaminates.

Another noteworthy feature in each of the pollen mass spectra shown in Fig. 3 was the emergence of mass peaks of low relative abundance at m/z 414, 382, and 354, respectively. A lipid extraction was prepared off-line for sunflower pollen and analyzed using GC/MS. The elution peak corresponding to m/z 414 produced a mass spectrum that best matched sistosterol (comparison with unmethylated steroid in NIST database), a familiar pollen steroid [36]. Meanwhile, m/z 382 and m/z 354, were found in trace amounts and gave corresponding retention times and comparable mass spectra to C24:0Me and C22:0Me, respectively. The prominent peak at m/z 316 in the mass spectrum of sunflower (Fig. 3(d)) was not found in the lipid extraction and will require additional examination for positive identification.

5.4. THM-MS of proteins

The focus of this study was to identify key FAME and steroid biomarkers of chemotaxonomic significance. Proteins were included as a comparative measure of the isobaric effects made by a biological compound containing no lipids. The mass spectra from three protein standards, mass analyzed following thermal treatment
with TMAH, are shown in Fig. 4. Peaks were observed at virtually every mass signifying the complicated nature involved with the thermal cleavage of high molecular weight proteins. Thermal degradation coupled with electron ionization produces a multitude of possible fragmentation mechanisms resulting in a complex

Fig. 3. THM mass spectra of pollen. (a) Cottonwood; (b) mixed Colorado; (c) soybean; and (d) sunflower.
mixture of predominantly low molecular weight gaseous ions. The most abundant ions observed in the subsequent mass spectra are the smaller (generally below 150 Da), more volatile distillents. The ions observed in the mass spectra illustrated in Fig. 4 are most likely dipeptide remnants and diketopiperazines that were formed by a known intra-polypeptide cyclization mechanism [37].

Relatively high intensity mass peaks with nominal masses equivalent to FAME and steroid biomaker compounds were not observed in the mass spectra from the
protein standards. The experimental setup was designed to enrich the cellular lipid fraction and minimize contributions from proteins, carbohydrates, and DNA. Lipid products, from thermal degradation of biological samples, are more volatile than other constituent biochemicals, especially for the mass range above 180 Da. Consequentially, by maintaining the transfer line from the heating chamber at lower temperatures (140°C), we could effectively prevent (by condensation onto the cold walls) less volatile thermal products from reaching the ion source. In fact, for proportionally similar amounts of samples, the total ion current (TIC) from whole cells (containing lipids) was an order of magnitude or more greater than the TIC produced from pure proteins. Needless to say, there does exist isobaric contributions from cellular proteins, but the overall contribution is minimal. In view of their negated significance to this study, no further attempt was made to identify the protein mass peaks demonstrated in Fig. 4.

5.5. THM-MS of bacteria

The THM mass spectra generated for the five species of bacteria selected for this study are depicted in Fig. 5. These bacteria were of interest because of their potential use as biological weapons [4]. A brief discussion of the bacterial FAMEs detected is given here. The results were consistent with a more detail description reported by Basile [38].

*B. anthracis*, the only Gram-positive species examined (Fig. 5(a)), contains peaks at *m/z* 242, 256, 270, and 284. These peaks are the molecular ions for C14:0Me, C15:0Me, C16:0Me, C17:0Me, respectively [38]. Gram-positive bacteria characteristically contain significant amounts of odd-chain fatty acids (both saturated and branched congeners) [31]. The only unsaturated fatty acid reported for *B. anthracis* [39] was C16:1 which, following methylation, produces notable ions at *m/z* 268 (M+ ) and *m/z* 236 ([M-32]+, loss of methanol).

Unlike *B. anthracis*, Gram-negative bacteria generally do not synthesize a substantial amount of saturated fatty acids with an odd number of carbon atoms [31]. Instead, even-chain saturated and monounsaturated fatty acids characteristically predominate. This phenomenon is illustrated in Fig. 5 for *Y. pestis* (Fig. 5(b)) and *V. cholera* (Fig. 5(c)), respectively. The base peak observed in both spectra is from the molecular ion from C16:0Me. Monounsaturated, C18:1 (*m/z* 296) fatty acid, is also prevalent to Gram-negative bacteria and is noticeable in both spectra by the appearance of the electron impact fragment ions at *m/z* 264 ([M-32]+) and *m/z* 222 ([M-74]+). Similarly, C16:1 (*m/z* 268), detected by its fragment ions at *m/z* 236 ([M-32]+) and *m/z* 194 ([M-74]+), was also detected in both spectra. Cyclopropyl heptadecanoic acid (cyC17:0), known to be produced by *Y. pestis* (and not by *V. cholera*) [40], accounts for the greatest difference between the cellular fatty acid profiles of these two bacteria species. Chemical signals corresponding to the losses of 32 and 74 Da from the methyl ester of this cyclopropane fatty acid (*m/z* 282) were observed at *m/z* 250 and *m/z* 208, respectively. Each bacterium gave expected FAME profiles when compared to published data [40,41].
Fig. 5. THM mass spectra of pathogenic bacteria. (a) B. anthracis; (b) Y. pestis; (c) V. Cholera; (d) B. melintensis; and (e) F. tularensis.
The most abundant fatty acid detected for \textit{B. melintensis} was cyclopropyl nonadecanoic acid (cyC19:0). The molecular ion for cyC19:0Me (M$^+$, \(m/z\) 310) was observed (Fig. 5(d)) as were the intense fragment ions at \(m/z\) 278 ([M-32]$^+$) and \(m/z\) 236 ([M-74]$^+$). In addition, \textit{B. melintensis} produced a substantial amount of C18:1 evident by the appearance of the molecular ion for the methyl ester at \(m/z\) 296 and the M-32 and M-74 fragment ions at \(m/z\) 264 and 222, respectively. The methyl ester for C16:0 was also detected (M$^+$, \(m/z\) 270 and [M-31]$^+$, \(m/z\) 239) for this species.

The fifth bacterium analyzed during this study was \textit{F. tularensis}. This human pathogen exhibits a unique profile featuring an unusual amount of long-chain fatty acids consisting of 20 or more carbon atoms. C26:0, C25:0, C24:0, C24:1, C23:0, C22:0, and C20:0 are all biosynthesized by \textit{F. tularensis} \cite{42} and were detected (see Table 2 for assigned mass to charge ratios) in the THM mass spectrum shown in Fig. 5(e). Furthermore, a compliment of esterified fatty acids of shorter length were also detected including: C18:0, C18:1, C16:0, and C14:0. As with the other bacteria used in this study, the FAME profile exhibited by \textit{F. tularensis} correlated well with an earlier report \cite{42}.

### 5.6. Multivariate analysis

Principal component analysis (PCA) was used to interrogate 110 THM mass spectra (five replicates of each of the 22 biological samples listed in Table 1). The outcome of PCA is the reduction of a multivariable problem into a smaller set of factors which collectively retain much of the total variance of the data. These factors can be assembled into \(X-Y\) coordinate systems, referred to as score plots. In this manner, a simple visual representation of a very complex set of mass spectra can be produced and used for data evaluation and interpretation. Each sample is assigned to either a letter, number, or symbol code which is then plotted on the score plot (see Table 1 for assigned codes). Where a specific code plots within the factor space depends on the distributions and normalized intensities of peaks detected during mass analysis. Proximity of one code relative to another is used to gauge how similar the mass spectra are between two independently analyzed samples. Hence, samples from the same taxonomoic class (bacteria with bacteria, fungi with fungi, etc.) are anticipated to produce similar mass spectra and then group together on the PCA score plot.

Using PCA, two samples from the original 110 mass spectra were judged to be outliers by visual examination of the score plots. These two samples, one \textit{B. anthracis} and the other FECV, did not plot within a reasonable distance of the other four replicates and were thus eliminated from the data set and any further calculations. A specific cause (instrumental or method error) of these two outliers was not identified. The remaining 108 mass spectra were used to construct the score plots shown in Fig. 6. Negative values on each axis scale result from mean centering the data prior to factor analysis. For the purpose of clarity, irregular shaped enclosures were included to highlight the five different classes of biological samples investigated. The boundaries of these enclosures have no mathematical relevance.
and were used only to emphasize clustering between members of the same defined classes. As expected, samples of the same category group together within the space defined by the principal components (PCs).

The upper plot in Fig. 6 was constructed from the first and second PCs, which, together, retained nearly half (49.4%) of the total variance. In this plot, tight clusters were formed by the viral and protein samples while the other three biological sample classes showed moderate intra-class dispersion. Samples of *B. melintensis*

![Graph showing THM-MS data](image)

Fig. 6. PC score plots showing the THM-MS data from samples listed in Table 1. (a) Components 1–2 representing 49.4% of the total data variance. (b) Components 2 and 4 representing 20.9% of the total data variance.
and *P. camembertii* exhibited the greatest variations from other members of the same class. For *B. melitensis*, deviation from the bacteria cluster is along the second factor axis while *P. camembertii* differed somewhat for the other fungi going across the first factor axis.

The second and fourth PCs were used to construct the lower plot shown in Fig. 6. Here, 20.9% of the variance is retained by these two factors. Again, it can be seen, along the second factor, that the five replicated *B. melitensis* samples diverge from the other bacteria. In addition, the bacteria spread apart according to genera along the fourth PC axis while the other four biological classes stayed grouped. This observation can be attributed to the greater diversity of fatty acids synthesized by prokaryotes opposed to the more conservative distribution of fatty acids made by eukaryotes.

Loading plots for the first four PCs are illustrated in Fig. 7. Drawing attention to Fig. 7(a), it can be seen that the four peaks, *m/z* 294, 292, 270, and 264, are the most intense and all point in the negative direction. Thus, along PC1, samples which generated mass spectra with these peaks in high abundance will plot on the negative side of the axis while those samples with these peaks in lesser abundance will plot to the positive side. The polyunsaturated FAMEs are responsible for the grouping of both the pollen and fungi samples on the negative half of PC1. Several bacterial samples were also slightly negative along PC1. Both *Y. pestis* and *V. cholera* exhibited an intense C16:0Me peak while *B. melitensis* produced *m/z* 264 ([M-32]⁺, C18:1Me) in high abundance. The positive loadings are mainly the mass peaks extending from *m/z* 186 to *m/z* 230. This region of the mass spectrum appears to be most characteristic of the protein standards which consequently had the most positive scores on PC1. Although the viral media samples showed both FAME and steroid peaks in their mass spectra, they also contained high abundances of the lower mass peaks common to the proteins. On account of this proteinaceous influence, the viral media samples also grouped to the positive half of PC1. Furthermore, the cholesterol molecular ion (*m/z* 386), common only to the viral media samples, contributes to the slightly positive loading value for PC1.

Examining the loading from PC2 (Fig. 7(b)), it can be seen that a large contribution (in the positive direction) is made by *m/z* 292, the C18:3Me biomarker characteristic of pollen. In fact, the influence of this biomarker on both PC1 and PC2 makes it possible to separate the pollen samples from the other classes. The fungi can be isolated due to the ergosterol peaks which cause this class to plot negative on the second factor axis. In addition, there is a large negative contribution to PC2 from *m/z* 264 and *m/z* 194, which together, causes the differentiation of *B. melitensis* from the other bacteria samples.

The loading plot for PC3 (Fig. 7(c)) displays many peaks ascribe to the FAMEs and steroids listed in Table 2. This observation provides further support to the notion that these lipid biomarkers can, and do play a large part in differentiating different classes of organisms analyzed by mass spectrometry. For example, from the loading plot for PC3, it can be seen that mass peaks originating from ergosterol impart a positive contribution to the variance while the contribution from the odd-chain FAMEs (C15:0, C17:0) is in the opposite direction. Thus, PC3 can effectively be used to differentiate fungi from Gram-positive bacteria.
Variation within the bacterial category occurs along PC4. Replicate samples of *B. anthracis* plotted on the positive side of this component due (Fig. 6(b)) to an abundance of the odd-chain methyl esters of C15:0Me and C17:0Me. *Y. pestis*, on the other hand, plotted to the negative side on account of the contributions made by C16:0Me and cyC17:0Me ([M-74]^+). *Brucella melitensis* also plotted to the negative side of PC4, here the determinates were peaks originating from C18:1Me ([M-74]^+) and C16:1 ([M-74]^+, m/z 194). Meanwhile, *Vibrio cholera* remained virtually neutral, plotting right on top of the PC4 axis. The fifth bacterium, *F.*
*tularensis*, plotted to the positive half of PC4 due to the effects from C24:0Me, C24:1Me ([M-32]+), C22:0Me, and C14:0Me to the variance in this direction. Because of the cholesterol molecular ion, the virus samples cluster on the lower half of PC4 while the protein samples fell almost exactly at the intersection formed by PC2 and PC4 (see Fig. 7(b)). The characteristic, low molecular weight, protein mass peaks basically make a negligible contribution to either PC4 or PC2, and thus this class of samples clustered about the origin of the score plot.

To further accentuate similarities between members of the same classification, the data was subjected to LDA. This more rigorous numerical approach takes linear combinations of PCs and a priori knowledge of sample classification and reduces the ‘intra-class’ variance while increasing the ‘inter-class’ distances. In this manner, a model was fashioned using the 108 mass spectra discussed above as the training set. Pictured in Fig. 8 is the resulting LDA score plot determined using the first 13 PCs (93.2% of the total variance). Much tighter clustering between members of the same class and increased distances between different classes is readily apparent. The loadings from the first and second linear discriminant axis or canonical variates (CVs) are shown in the bottom half of Fig. 8.

Positioning of the pollen class of samples in the upper left quadrant of the LDA score plot is due mainly to the intense loading for *m/z* 292 which is negative for CV1 and positive for CV2. The fungi, meanwhile, fall into the left, lower quadrant. Ergosterol peaks, contributing in the negative direction of CV2, are most responsible for the isolation of this class. Mass spectra from the protein standards plotted furthest to the right on CV1 owing to the positive contributions from masses in the lower scan range. Like the protein standards, the viral media samples also plotted in the right, upper quadrant (although their determined coordinates placed them to the left and above the protein samples). The most notable influence on the viral samples came from the loading for *m/z* 386 which is in the positive direction on CV2. Finally, samples representative of the bacteria plotted close to the score plot origin because of their more heterogenous distribution of fatty acids. There is not a specific set of peaks that are inclusive to all bacteria. Therefore, intense loadings representative of the bacterial class as a whole were not found by the LDA algorithm.

In order to validate the reliability of the trained LDA model, two test sets were formed. Each of these sets were constructed by reanalyzing the same samples (ran in triplicate) used in the training set. Samples from both test sets were designated as unknowns and submitted to model for classification. All 66 samples of the first test set were correctly classified while one misclassification was made on the second test set. On the whole, 131 out of 132 unknown samples were accurately classified with the one misclassification being a protein sample classified as viral media. Moreover, the model was able to differentiate the bacteria samples from other biological materials commonly found suspended in the atmosphere.

Extending the data analysis, new training and test sets were constructed with only the bacteria samples. The LDA score plot for the bacteria training set is shown in Fig. 9 accompanied by the first two CV loadings. The peaks observed to make *B. melinensis* plot furthest to the left on CV1 were mainly those associated with C18:1Me (*m/z* 296, 264, and 222) and cyC19:1Me (*m/z* 278 and 194). *Francisella*
species were also negative (slightly) along CV1 caused mainly by C24:0Me, C24:1Me ([M-32]^+, m/z 348) and C22:0Me. For CV2, the positive contribution of these same peaks is much greater indicated by the intensity of their loadings.

Fig. 8. LDA score plot and loadings showing the THM-MS data from samples listed in Table 1. (a) LDA score plot constructed from CV1 and CV2. (b) Loading plot from CV1. (c) Loading plot from CV2.
Meanwhile, *B. anthracis* plotted to the extreme right of CV1 accounted for by the odd-chained FAMEs C17:0Me and C15:0Me. These same peaks also strongly influenced the positioning of the *anthrax* samples to the negative side of CV2. The other two genera, *Yersinia* and *Vibrio*, although well separated from the others, remained fairly close to each other. Slight separation was achieved by CV2 chiefly affected by \( m/z \) 250 and \( m/z \) 208 from \( \text{cyC}_{17}:1\text{Me} \) (*Yersinia*) and \( m/z \) 194 from \( \text{C}_{16}:1\text{Me} \) (*Vibrio*).
A classification model was built by preforming LDA on the first six PCs (97.4% of the total variance) determined from the training set of 25 bacteria observations. Of 30 bacteria tested as unknowns, 29 were correctly classified. The one misclassification was a *Y. pestis* sample, from the second test set, incorrectly identified as *B. melintensis*.

6. Conclusions

Representative samples of common airborne constituents were analyzed using mass spectrometry and evaluated for chemotaxonomic significance using multivariate statistical analysis. An online sample preparation technique, using a thermally activated hyrolyzing/derivatizing reagent, was utilized and manipulated to produce profiles from whole cell organisms enriched in FAME and steroid signals. The extent to which these different lipid biomarkers were methylated was observed to be a function of their relative acidities. The fatty acid carboxylate cation was easily formed in the strongly basic milieu created during sample preparation while the hydroxyl functionality of the steroids was much less prone to deprotonation in the same environment. Therefore, the fatty acids were observed to be methylated to a much greater extent than were the steroid biomarkers which, instead, were liberated and thermally desorbed mainly as the underivatized molecule.

The THM technique is fast and ameanable to field operation. Furthermore, the mass spectra generated could be used to discriminate between the five different categories of biological samples. Eukaryotic organisms were differentiated from prokaryotic microorganisms by the detection of polyunsaturated and steroid biomarkers. For the eukaryotic cells, C18:3Me was the most characteristic of the pollen samples while the steroids ergosterol and cholesterol were unique to samples of fungi and viral infected media, respectively. Suspensions of standard proteins were included in the analysis to evaluate differences between whole cell samples and non-lipid biological samples. The main criteria for differentiating the whole cell samples from the proteins was the detection of FAME and steroid biomarkers for the former and the lack of the same biomarkers for the latter. Finally, linear discriminant analysis was used to train models for predicting cell type and bacteria species. These models correctly predicted the cell type (or a protein) of samples from two test sets with a 99.2% accuracy. Moreover, the speciation of five different bacteria was correctly made at a level of 96.7%.

Acknowledgements

The authors acknowledge the DOE for their generous support of this study (Contract # DE-AC05-00OR22725).
References

[1] R. Burrell, Immunol. Allergy Clin. North Am. 14 (1994) 537.
[2] W.E. Horner, S.B. Lehrer, J.E. Salvaggio, Immunol. Allergy Clin. North Am. 14 (1994) 551.
[3] W.D. Griffiths, G.A.L. DeCosemo, J. Aerosol Sci. 25 (1994) 1425.
[4] R.M. Atlas, Crit. Rev. Microbiol. 24 (1998) 157.
[5] G.W. Christopher, T.J. Cieslak, J.A. Pavlin, E.M.J. Eitzen, J. Am. Med. Assoc. 278 (1997) 412.
[6] S. Terzieva, J. Donnelly, V. Ulevicius, S.A. Grinshpun, K. Willeke, G.N. Stelma, K.P. Brenner, Appl. Environ. Microbiol. 62 (1996) 2264.
[7] P.A. Jensen, B. Lighthart, A.J. Mohr, B.T. Shaffer, in: B. Lighthart, A.J. Mohr (Eds.), Atmospheric Microbial Aerosols Theory and Applications, Chapman and Hall, New York, 1994, p. 226.
[8] B. Crook, in: C.S. Cox, C.M. Wathes (Eds.), Bioaerosol Handbook, CRC Press, Boca Raton, FL, 1995, p. 247.
[9] K.R. Spurny, J. Aerosol Sci. 25 (1994) 1533.
[10] C. Fenselau, in: C. Fenselau (Ed.), Mass Spectrometry for the Characterization of Microorganisms, vol. 451, American Chemical Society, Washington, DC, Chapter 1, 1994.
[11] H.L.C. Meuzelaar, J. Haverkamp, F.D. Hileman, Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterial, vol. 3, Elsevier, Amsterdam, 1982.
[12] A. Fox, S.L. Morgan, L. Larsson, G. Odham, Analytical Microbiology Methods: Chromatography and Mass Spectrometry, Plenum Press, New York, 1990.
[13] T.G. Tornabene, Methods Microbiol. 18 (1985) 209.
[14] S.J. Macnaughton, T.L. Jenkins, S. Alugupalli, D.C. White, Am. Ind. Hyg. Assoc. J. 58 (1997) 270.
[15] G. Odham, A. Tunlid, G. Westerdahl, L. Larsson, J.B. Guckert, D.C. White, J. Microbiol. Methods 3 (1985) 331.
[16] C.W. Moss, J. Chromatogr. 203 (1981) 337.
[17] M.P. Lechevalier, Crit. Rev. Microbiol. 5 (1977) 109.
[18] Microbial ID, Microbial Identification System operating manual, Version 3.0, Newark, DE, 1993.
[19] J.P. Dworzanski, L. Berwald, H.L.C. Meuzelaar, Appl. Environ. Microbiol 56 (1990) 1717.
[20] J.P. Dworzanski, L. Berwald, W.H. McClennen, H.L.C. Meuzelaar, J. Anal. Appl. Pyrolysis 21 (1991) 221.
[21] G. Holzer, T.F. Bourne, W. Bertsch, J. Chromatogr. 468 (1989) 181.
[22] W.C. Kossa, J. MacGee, S. Ramachandran, A.J. Webber, J. Chromatogr. Sci. 17 (1979) 177.
[23] F. Basile, M.B. Beverly, C. Abbas-Hawks, C.D. Mowry, K.J. Voorhees, T.L. Hadfield, Anal. Chem. 70 (1998) 1555.
[24] A.D. Hendrick, C. Abbas-Hawks, F. Basile, K.J. Voorhees, T.L. Hadfield, Int. J. Mass Spectrom. Ion Phys. 190/191 (1999) 331.
[25] S.A. Barshick, D.A. Wolf, A.A. Vass, Anal. Chem. 71 (1999) 633.
[26] S. DeLuca, E.W. Sarver, P. Harrington, K.J. Voorhees, Anal. Chem. 62 (1990) 1465.
[27] A.J. Madonna, T.L. Hadfield, E.J. Hilyard, K.J. Voorhees, J. Am. Soc. Mass Spectrom. 10 (1999) 502.
[28] P.B. Harrington, K.J. Voorhees, Anal. Chem. 62 (1990) 729.
[29] K.E.K. Defernez, Trends Anal. Chem. 16 (1997) 216.
[30] J.D. Weete, Fungal Lipid Biochemistry; Distribution and Metabolism, Plenum Press, New York, 1974.
[31] H. Lechevalier, M.P. Lechevalier, in: R.C., S.G. Wilkinson (Eds.), Microbial Lipids, vol. 1, Academic Press, San Diego, 1988, p. 869.
[32] J.A. Erwin, Lipids and Biomembranes of Eukaryotic Microorganisms, Academic Press, New York, 1973.
[33] G. Bianchi, C. Murelli, E. Ottaviano, Phytochemistry 29 (1990) 739.
[34] J. Sekiya, K. Yamashita, S. Nakagawa, Y. Shibata, A. Hatanaka, Agric. Biol. Chem. 52 (1988) 243.
[35] D.E. Evans, N.E. Rothnie, M.V. Palmer, D.G. Burke, J.P. Sang, R.B. Knox, E.G. Williams, E.P. Hilliard, P.A. Salisbury, Phytochemistry 26 (1987) 1895.
[36] C. Hitchcock, Plant Lipid Biochemistry: the Biochemistry of Fatty Acids and Acyl Lipids with Particular Reference to Higher Plants and Algae, Academic Press, New York, 1971.
[37] A.D. Hendricker, K.J. Voorhees, J. Anal. Appl. Pyrolysis 48 (1998) 17.
[38] F. Basile, M.B. Beverly, T.L. Hadfield, K.J. Voorhees, Trends Anal. Chem. 17 (1998) 95.
[39] T. Kaneda, J. Bacteriol. 95 (1968) 2210.
[40] T.G. Tornabene, Biochim. Biophys. Acta 30 (1973) 173.
[41] M.A. Lambert, F.W. Hickman-Brenner, J.J. Farmer, C.W. Moss, Int. J. Syst. Bacteriol. 33 (1983) 777.
[42] P.D. Nichols, W.R. Mayberry, C.P. Antworth, D.C. White, J. Clin. Microbiol. 21 (1985) 738.