Anal Bioanal Chem (2013) 405:351–357
DOI 10.1007/s00216-012-6495-4

ORIGINAL PAPER

Identifying indoor environmental patterns from bioaerosol material using HPLC

Sarah J. R. Staton · Josemar A. Castillo · Thomas J. Taylor · Pierre Herckes · Mark A. Hayes

Received: 27 June 2012 / Revised: 16 September 2012 / Accepted: 10 October 2012 / Published online: 24 October 2012

Abstract A substantial portion of the atmospheric particle budget is of biological origin (human and animal dander, plant and insect debris, etc.). These bioaerosols can be considered information-rich packets of biochemical data specific to the organism of origin. In this study, bioaerosol samples from various indoor environments were analyzed to create identifiable patterns attributable to a source level of occupation. Air samples were collected from environments representative of human high-traffic- and low-traffic indoor spaces along with direct human skin sampling. In all settings, total suspended particulate matter was collected and the total aerosol protein concentration ranged from 0.03 to 1.2 μg/m^3. High performance liquid chromatography was chosen as a standard analysis technique for the examination of aqueous aerosol extracts to distinguish signatures of occupation compared to environmental background. The results of this study suggest that bioaerosol "fingerprinting" is possible with the two test environments being distinguishable at a 97 % confidence interval.

Keywords Bioaerosols · Environmental monitoring · Environmental pattern recognition · HPLC · Separation

S. J. R. Staton · J. A. Castillo · P. Herckes · M. A. Hayes (✉)
Department of Chemistry and Biochemistry,
Arizona State University,
Tempe, AZ 85281, USA
e-mail: MHayes@asu.edu

S. J. R. Staton
US Naval Research Laboratory,
Washington, DC 56944, USA

T. J. Taylor
Department of Mathematics and Statistical Sciences,
Arizona State University,
Tempe, AZ 85281, USA

Introduction

Bioaerosols are ubiquitous bits of biological material found suspended in the air [1, 2]. These materials are bacteria, viruses, and pollen as well as by-products of skin and hair of humans and animals along with insect and plant debris. The particles range in size from 0.3 to 100 μm [3] and contain a variety of different information-bearing biomolecules. For example, bioaerosols from vertebrates are largely composed of structural proteins like keratins [4–7]. These proteins are more stable and robust than other commonly targeted biomolecules (DNA for example), given their function as the barrier layer in the skin. Other potential molecules of interest found in shed skin and hair include truncated pieces of DNA, lipids, nucleic acids, amino acids, and glycosylated species [3]. Consequently, bioaerosols could be seen as biochemical "calling cards" of the organism of origin that could be utilized for identification. The shedding and aerosolization of material represents a fundamental biological process that analytical methods could exploit to expose the organism of origin and its potential whereabouts. The exploitation of bioaerosol material could expand current forensic detection tools allowing organisms to be tracked or found that currently would be undetectable or forensically "naked" [8].

Bioaerosols are omnipresent in both outdoor and indoor environments. In the atmosphere, proteins have recently been found to compose a much larger portion of the atmospheric particle budget than the trace amount previously assumed, with potentially up to 25 % of atmospheric particles having a biological origin [1, 9, 10]. Human bioaerosols rich in keratin have been reported in large quantities in indoor environments, such as classrooms, and have been linked to "sick building syndrome" [11, 12]. Aerosolized keratins are so prolific that they are also the most common contaminant in protein gels [13]. This is reasonable because the daily losses of the outer layer of human skin, called the

stratum corneum, are reported to range from 0.5 to 1.5 g per day in normal skin and up to 17 g per day in particular skin diseases [4, 14]. Over the course of a lifetime, this material can accumulate upwards of ~40 kg of skin debris for normal individuals.

By analyzing bioaerosol material, natural variations of protein expression in the skin and hair of different organisms can be exploited for identification. These protein patterns can be determined using a combination of different proteomic and separation techniques in conjunction with mathematical pattern recognition methods. Variations in protein expression of different types of skin and hair-related proteins have been studied [7, 15]. Limited research has also explored using these natural variations for pattern generation, including the use of bulk amino acid patterns from feathers to identify different bird species [16].

The focus of the current study is to demonstrate a comprehensive analytical approach using complex bioaerosol composition for source identification, in contrast to single component analysis of bioaerosols (Fig. 1). This was achieved by collecting bulk complex environmental bioaerosol mixtures with human traffic as a variable building on current research that shows the importance of traffic levels on bioaerosol concentration [17]. These samples were then coarsely separated into reproducible patterns via high performance liquid chromatography (HPLC) and the resulting data were used to uniquely correlate to the conditions during sampling. Comparison of bioaerosols from indoor environments of varying human occupation levels (and to directly sampled freshly shed human epidermis) resulted in statistically significant and reproducible correlations.

Experimental

Total suspended particulate matter (TSP) samples were collected onto 47 mm Teflon filters with a 0.2-μm pore size (Pall Gellman, Port Washington, NY) fitted on an open face 47-mm filter holder (Advantec, Dublin, CA) using a linear piston pump (Medo® model VP0125, Hanover Park, IL). Samples were collected in indoor human high- and low-traffic environments as well as directly induced human epidermis bioaerosol samples. Sampling times varied from minutes (induced source samples) to 3 days (indoor samples). Filter assemblies faced upwards for both sample and blank collection [18]. Environmental blanks were taken from each environment tested.

The indoor air samples were collected in an occupied office/laboratory high-traffic area (at least three or more people on average over a 10-h work day with near constant habitation) and inside an unoccupied office/laboratory low-traffic area (one person or less for less than an hour of habitation on average over a 24-h period). Each sample type represents at least three separate sampling days. Source bioaerosol samples were collected by induced shedding. Specifically, the induced human sample was taken directly from individuals who had clean and dry skin. The human epidermis was collected using the side of a sterile glass to scrape loose skin into a Petri dish that was then placed near the sampling apparatus for sampling. The exact amount of skin collected was not quantified. After collection, the filter samples were stored frozen in the dark at ~20 °C until analysis. Prior to analysis, the filters were divided into four equal sections using a surgical blade, one filter section was used for bulk protein quantification, while the remaining three sections were used for HPLC characterization.

Bulk protein quantification was performed using the standard protocol [19] given in the NanoOrange Total Protein Quantitation Kit (Invitrogen®, Carlsbad, CA) and analyzed using a fluorometer (Shimadzu RF 551, Columbia, MD) at the excitation/emission wavelengths of 470/570 nm, respectively. The filter sections were extracted using ultrapure water (>18 MΩ, Millipore) under ultrasonication (Ultrasonic Power Corporation model 2000U 120- V, Freeport, IL) for 15 min. The resulting extracts were filtered through a 0.2-μm syringe filter, Puradisc 25 mm (Whatman, Kent,

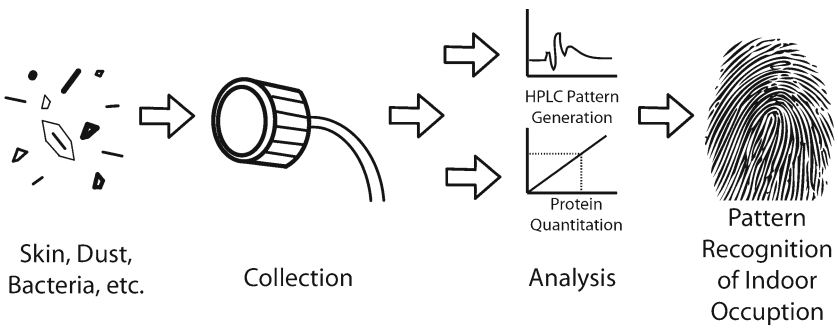

Fig. 1 Generalized schematic of human debris-based occupation detection. The various human occupation-related aerosolized material are collected via the filtration assembly along with other non-related aerosolized material. The collected material was then analyzed for total protein concentration as well as coarsely separated to generated unique pattern profiles. These profiles are information-rich enough to identify human occupation in an indoor space

U.K.). Reported (Table 1) protein concentrations have been corrected for environmental contamination through the subtraction of the environmental blank. The value of the environmental blanks was always less than 10 % of the sample concentration and generally much lower.

The particulate matter on the remaining three quarters of the filter sample was extracted with 3 mL of HPLC-grade methanol (Fisher Scientific, Pittsburgh, PA). This solution was then ultrasonicated for 15 min to lyse any cells. After sonication, the solution was filtered through a 0.2-μm syringe filter. The aerosol extract was then analyzed using a reverse phase Agilent 1100 HPLC system consisting of a 1100 binary pump, a 20-μL sample loop, and a 1100 variable wavelength detector (Agilent, Santa Clara, CA). The samples were separated with an Agilent Zorbax SB-C3 Analytical HPLC column 4.6× 150 mm with a Zorbax guard column (Agilent, Santa Clara, CA), using an isocratic mobile phase consisting of 90/10 methanol/water at a flow rate of 1 mL/min. Detection was performed through absorbance at 254 nm. Average run times were on the order of 2.6–3 min.

Along with visual inspection, recognition and distinction of data from HPLC chromatograms were approached with strategies established in the pattern recognition community [10, 20–23]. The statistical differences between the two different sample environments were evaluated using Euclidean distance, which here amounted to the root mean square difference between pairs of HPLC chromatogram data points. These statistical distances were then used to construct and train automated classifiers, as well as establishing their characteristics. The data were regarded as high-dimensional: each chromatogram is a list of thousands of numbers, each representing absorbance at a particular time as sampled at 200 Hz with each chromatogram representing a vector in thousands of dimensions. Automated classification was facilitated by dimension reduction methods. One of these methods is visual representation (e.g., graphing), which requires that the reduced dimension be less than or equal to three so that the data can be projected into three-dimensional or two-dimensional space. Optimal dimension reduction for data analysis, to any dimension, is the dimension reduction which best preserve distances between data points. Two different strategies have been employed in this regard [24]. The first strategy is principal component reduction, which projects onto the three orthogonal directions which capture the most variability of the data. The second is support vector machine (SVM), a non-statistical algorithm for finding the (hyper-) plane of maximal separation between two classes of multidimensional data. This is accomplished by finding the hyperplane that gives the greatest distance to the nearest points (support vectors) of one class one side and an equal distance to the nearest points of the other class on the opposite side. In a sense, it provides the best vantage point to detect the separation between two classes of data. The classifiers were constructed by first normalizing the data with respect to total absolute signal. A low concentration cutoff was applied removing samples with a largest feature size below 1 absorbance unit. A SVM was constructed from the two classes of data, low and high traffic, using the standard published methods, and the optimum separating hyperplane was obtained for these data. The chromatograms in each class at approximately equal (±50 %) total concentration are very close. Accordingly, the separation as measured by comparing standard deviation to difference of the means is very robust to the inclusion or exclusion of particular examples.

Results

Measurable quantities of protein were found in most samples (Table 1). In those environments where initial measurements showed that protein concentrations were below the detection limit, extended sampling times were used to boost to measurable levels. Protein concentrations are expressed as micrograms of protein per cubic meter relative to the BSA standard used in the NanoOrange kit. The largest amount of protein was collected through the induced sampling (0.07–1.1 $\mu g/m^3$), which could be expected given the assisted protein release near the filtered collection. The induced samples were directly collected skin debris resulting in more stable protein concentrations. Environmental

Table 1 Protein concentrations of various air samples

Sample type	Protein concentration range($\mu g/m^3$)	Reference
"High-traffic" indoor environment	0.05–1.2	This work
"Low-traffic" indoor environment	0.03–0.05	This work
Fresh human epidermis	0.07–1.1	This work
Indoor average over all particle sizes (<2.5–>10 μm)	0.6	Chen and Hildermann [17]
Average occupied classroom	86.8 (total dust)	Fox et al. [11]
Average unoccupied classroom	6.44 (total dust)	Fox et al. [11]

Concentration range of total protein found in high-traffic, low-traffic, and induced human epidermis samples. Concentrations are reported in microgram of protein material per cubic meter of air collected

samples showed a degree of variability in total protein concentration from 0.03 to 1.2 μg/m^3. This variation is likely the result of small differences in the exact amount of occupation, air handler activity, and the amount of shed material released by different individuals. The amounts of aerosolized protein recovered in these samples are nominally in the same range as other indoor protein studies previously reported [11, 17]. The bulk protein concentration does appear to display a large range when comparing the various studies, although the samples are not normalized with regard to air quantity measured, environmental conditions, and the level of occupation varies greatly.

A coarse level of separation was achieved using HPLC generating reproducible patterns from the material collected on the filters (Fig. 2). This material includes both the soluble organic and inorganic material collected that absorbs at the detection wavelength. The individual components of the sample were not determined for this study nor completely resolved due to the low concentration of subspecies. It is noted that higher resolving power strategies (other HPLC, GC, and MS) were attempted, but detection was hampered by low species concentration—hence the strategy of overlapping peaks to boost signal strength coupled with pattern recognition [25, 26]. The averaged HPLC trace displays three separate collections from the high-traffic indoor space and seven collections of the low-traffic indoor space. The standard deviation within each sample type between replicates was on average less than 1 absorbance unit in the raw chromatogram although this does scale with the normalization, and the average standard deviation over the information-rich regions are much smaller than the separation between the two data sets. The traces have been offset to aid in visualizing the details in each sample trace. Even with visual inspection of the chromatograms, differences are apparent at times ~1.68 and ~1.85 min (Fig. 3). To better distinguish these variations

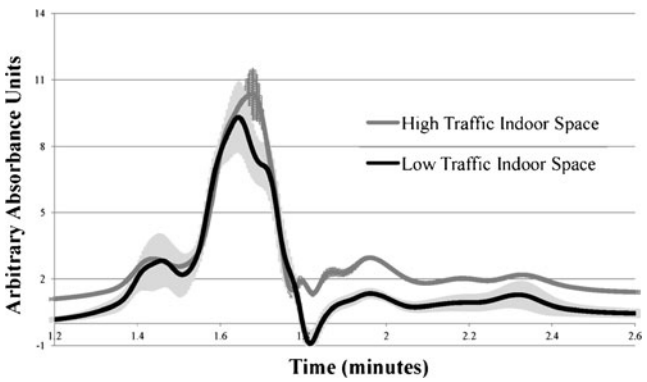

Fig. 2 Comparison of HPLC chromatograms of two environmental samples. The average chromatograms of three samples of high-traffic indoor space and seven sample of low-traffic indoor space along with standard deviation envelopes. The two samples types are offset to aid in visualizing variations in the HPLC traces that are plotted in arbitrary absorbance units verses time in minutes

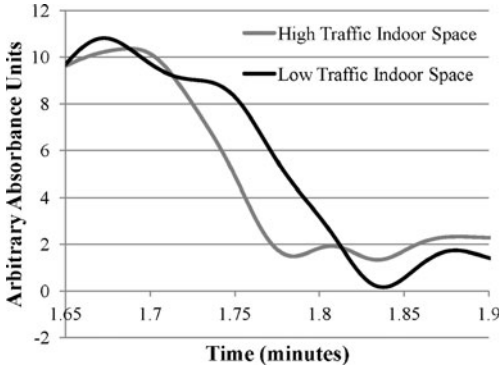

Fig. 3 Comparison of the identifying portions from each sample type enlarged. Closer view of the regions between 1.65 to 1.9 min of the low- and high-traffic chromatograms from Fig. 2 illuminating with greater detail the areas of greatest variation

and their magnitude, the chromatograms found in Fig. 2, without the offset, were subtracted from each other (Fig. 4). The resulting peaks, both positive and negative, more clearly illuminate the variations between the two model environments that can be potentially exploited for identification and classification. However, the pure visualization of the variations does not indicate the quantifiable statistical significance, if any. Pattern recognition tools were employed to quantify these distinctions.

Pattern analysis was performed to quantify the statistical difference of the two chromatogram types. The data from the high-traffic and low-traffic environments consisted of HPLC chromatograms truncated to a maximum common length of 821 segments, with each segment representing 0.3 s. The chromatograms were truncated to normalize the total number of time segments at the sampling rate of the HPLC, 0.3 s, while also preserving the data containing area of the chromatograms. The data from the segmented traces were treated as vectors in an 821 dimensional vector space. These vectors were normalized to minimize the effects of total concentration and to focus the analysis on the relative

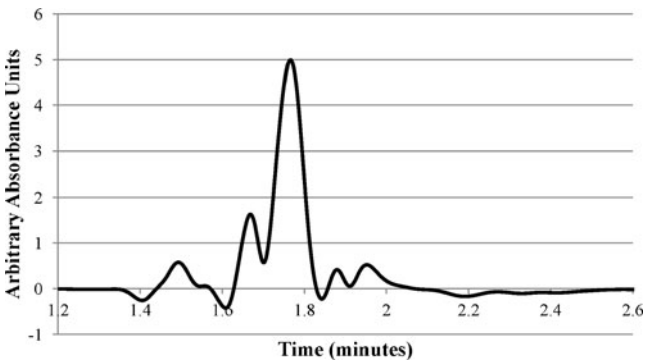

Fig. 4 Subtraction of the low- and high-traffic chromatograms. The high-traffic chromatogram was subtracted from the low-traffic chromatogram to aid the visualization of the differences existing in the two sample classes

concentration of individual components. The low concentration samples were removed due to their variation in peak width and resolution differences. These disparities greatly complicate the pattern recognition process leading to the concentration match cutoff for samples whose maximum concentration feature was less than 1 absorbance unit. The use of the concentration cutoff illustrates the lower limit of normalization as well as the difficulties of using chromatographic methods in pattern recognition regimes where size, shape, and resolution can all be affected by variables including concentration.

The principal components [23] for the ensemble of chromatograms from both environments were determined and ordered by singular difference values from the largest to the smallest. From this ordered list, the three directions with the largest magnitudes were selected for statistical comparison. This treatment, while increasing the ability to visualize the distinctions between traces, actually decreased the amount of difference between the traces of samples by only evaluating the three largest statistical distances along the chromatograms. Therefore, distinguishable differences derived from this analysis are at least as valid as had the analysis included all of the 821 dimensions. Statistical distances within a single sample type are a measure of the variance within the sample type while the distance between the sample types indicate the degree of variation between sample types. The use of the three largest distances captured 94 % of the variance in the data, while the addition of the subsequent components would not significantly increase the variance. For example, the addition of the fourth largest component would add only 4 % additional variance and the fifth largest component would add an additional 1.1 %.

From this analysis, it was determined that the two sample classes are distinguishable numerically. However, in classifying high-dimensional data, "viewpoint" plays an important role in establishing classifiability of distinct classes of data [20]. One common procedure of automated classification, the support vector machine [21, 22], as mentioned earlier, calculates a "maximum margin separating hyperplane." Applied here, the two classes fall on a hyperplane within the 821 dimensions such that the minimum perpendicular distance of the data points of both classes to the hyperplane is maximized. The maximum margin separating hyperplane was computed for the high-traffic- and the low-traffic HPLC traces. The one-dimensional displacements perpendicular to the hyperplane indicates not only the amount of separation between the two classes and the hyperplane but also describe the variance among the class samples as they vary in distance to the plane. The two data classes are more than 3 standard deviations from the mean of the other class along the direction of the normal to the separating hyperplane, suggesting that the high-traffic and low-traffic may be very well distinguishable by HPLC. Given the sample resolution of 3 standard deviations, this equates to a 97 % confidence interval that these sample types are distinguishable.

Discussion

Environmental samples are complicated and have wide ranging concentrations. It was the goal of this work to see, given all of the complicating factors, if differentiable patterns could arise from these samples using only a basic separation scheme, e.g., HPLC. It was found that both of the two indoor sample types and the induced human epidermis samples produced measureable quantities of protein within 3 days or approximately 30 m^3 of air, with most samples needing less than a day or 10 m^3 of air sampling. Even shorter sampling time could be achieved by using higher volume air pumps or more sensitive detection systems. Protein, as expected, was found to be ubiquitous in both environments, including the indoor environment with minimal human activity and no plants or animals. The exact composition and source of the proteinaceous material was not determined in this study as its focus was in the pattern recognition of the sample as a whole rather than its individual components. The range of protein concentration appears to largely depend on activity level within the indoor space, which stands to reason if human or human-related activity is the main source of the protein material. Variations in protein concentration maybe partially attributed to changes in air handler activity, environmental weathering from bacterial degradation and photolytic processes, as well as a number of other processes that have not been cataloged within the test environment. Other differences in the protein concentrations arose from the diffuse environmental sampling used for the high- and low-traffic samples in comparison to the consistently large quantities of protein observed in the actively induced samples. This increasing trend in concentration is consistent with direct transmission being the most efficient transfer method of protein containing fragments from the skin, hair, bacteria, mold, etc. to the sampling equipment.

Coarse level separation of the aerosols was achieved using HPLC and created reproducible patterns of the particular sample types over several duplicate samples from the same environments. Separation of the extracted TSP samples included all of the soluble components smaller than 0.2 μm, both organic and inorganic. Although the resulting pattern generated was reproducible, the individual component peaks were not resolved or identified given the use of UV–Vis detection at a generic protein absorption wavelength. The overall peak pattern for the two sample types display visible differences in the chromatograms (Figs. 2, 3, and 4). Within the difference of the two chromatograms

(Fig. 4), the variations between the two sample types become more clearly distinguishable. These variations allow for a gross qualitative identification using only HPLC separation. Given the coarse nature of the separation including the short retention times, generic C3 column used, and the nonspecific UV–Vis detection mode, these results show early promise, with the possibility that with more advanced separation and/or detection methods further more detailed information could be gleaned for identification purposes.

Pattern recognition methodologies were employed to characterize the distinguishability of the high-dimensional data of the HPLC traces from the different environments. This characterization quantified the statistical distance between the two environments tested. The first method calculated the root mean square difference between the various data points along the chromatograms. This analysis helps to determine areas of the chromatograms that contain the most diagnostic information for classification. These areas in the chromatograms may be indicative of various organic or inorganic components associated with human habitation. One likely suspect is human debris. Human debris known to be a large contributor to indoor protein contamination and is stable enough to survive past the time of direct human activity [11]. The stability and ubiquity of human epidermis makes it a likely protein source in both high- and low-traffic environments; however, the material maybe in various states of decomposition.

In order to provide a numerical understanding of the degree of variation between the two environments, support vector machine was also performed. Through support vector machine analysis, the statistical distances of the chromatograms were determined to demonstrate the separation of the high-traffic- and low-traffic data from a constructed plane [20, 22]. The data's perpendicular displacement from either side of this plane offers a measure of separability of the two data sets. The standard deviations of the displacements of the low-traffic data point clusters are small enough that they can be identified as statistically distinct with 97 % confidence. This strongly supports the idea that distinguishing these environments via bioaerosol samples is feasible. Even using basic methods of collection (impact filtration) and coarse separation (HPLC), distinguishing features are observable at levels comparable to more developed and accepted technologies, e.g., DNA sampling of aerosolized bacteria and viruses.

An alternative method of evaluating the value of bioaerosols for identification purposes involves the production of synthetic samples that are well characterized and artificially simple. Although this approach is valid and has strengths, directly evaluating the challenges faced by using real environmental samples has insured that the results generated directly reflect the viability of the method. More testing is warranted, including synthetic standards; however, by tackling the environmental samples first, this method has provided a proof-of-concept. A large untapped wealth of bioanalytical possibilities are suggested to be locked in bioaerosols and their potential applications for identification. Even at the current level of this technology, it demonstrates the ability to sense the presence of different level of habitation through air collection. One of the most powerful attributes of this type of investigation is its remote and noninvasive nature. Without having to come into direct contact with the organism of interest, the method exploits a fundamental biological process of most species, the release of bioaerosol material, and represents a potentially powerful forensic tool. Expanding this method by using higher resolution separations and detection of specific motifs to determine the species of origin could find several future applications. Those potential applications could include the remote detection of criminals and their movements as well as the impact on the environment of different policies. However, even in its current state, the use of noninvasive air sampling of bioaerosols have opened the doors to a new and unique means to detect the habitation of an environment without direct contact, which could prove valuable to several fields like forensic science.

The results of this preliminary study suggest that even with all the complexity of minimally controlled real environmental samples and standard protein analysis strategies (HPLC with UV detection) together with standard pattern, recognition algorithms show statistically significant features that can differentiate and group the types of environmental bioaerosols collected. When these samples are concentration matched and normalized, high-traffic indoor samples can be differentiated from low-traffic indoor samples with a 97 % confidence interval. The results of this work strongly suggest that with an adequate amount of sample and resolution in separation, various types and subtypes of samples can generate distinct patterns supporting the idea of bioaerosol "fingerprinting." The ability to determine the level of habitation remotely via air sampling could have several applications from defense to environmental settings. With further development, this technique could hold promise of identifying different humans and potentially animals remotely from environmental air samples.

Acknowledgments Special thanks for the assistance of Dr. Jim Anderson for providing the air samplers. This research was financially supported in part through the Department of State Fulbright Fellowship and the Department of Education Foreign Language and Area Studies Fellowship (FLAS) as well as NIH grants 1R03AI094193-01, 1R03AI099740-01, and R21EB010191-01A1.

References

1. Jaenicke R, Matthias-Maser S, Gruber (2007) Omnipresence of biological material in the atmosphere. Environ Chem 2:217–220

2. Matthias-Maser S, Peters K, Jaenicke R (1995) Seasonal variations of primary biological aerosol particles. J Aerosol Sci 26:S545–S546
3. Cox CS, Wathes CM (1995) Bioaerosols handbook. Lewis, Boca Raton
4. Klingman AM (1964) The epidermis. Academic, New York
5. Wood EJ, Bladon PT (1985) The human skin. Camelot, London
6. Chao Y-CE, Nylander-French LA (2004) Determination of keratin protein in a tape-stripped skin sample from jet fuel exposed skin. Ann Occup Hyg 48(1):65–73
7. Plowman JE (2007) The proteomics of keratin proteins. J Chromatogr B 849:181–189
8. Castillo JA, Staton SJR, Taylor TJ, Herkes P, Hayes MA (2011) Exploring the feasibility of bioaerosol analysis as a novel fingerprinting technique. Anal Bioanal Chem. doi:10.1007/s00216-012-5725-0
9. Jaenicke R (2005) Abundance of cellular material and proteins in the atmosphere. Science 308:73
10. Snyder AP, Maswadeh WM, Tripathi A, Eversole J, Ho J, Spence M (2004) Orthogonal analysis of mass and spectral based technologies for the field detection of bioaerosol. Anal Chim Acta 513:365–377
11. Fox K, Castanha E, Fox A, Feigley C, Salzberg D (2008) Human K10 epithelial keratin is the most abundant protein in airborne dust of both occupied and unoccupied school rooms. J Environ Monit 10:55–59
12. Luoma M, Batterman SA (2001) Characterization of particulate emissions from occupant activities in offices. Indoor Air 11:35–48
13. Arnott D, Henzel WJ, Stults JT (1998) Rapid identification of comigrating gel-isolated proteins by ion trap-mass spectrometry. Electrophoresis 19:968–980
14. Cox RN, Clark RP (1973) Natural-convection flow about the human body. Rev Gen Therm 12:11–19
15. Ku N-O, Liao J, Omary MB (1997) Apoptosis generates stable fragments of human type I keratins. J Biol Chem 272(52):33197–33203
16. Zhongwu S, Xueliang G, Hongfei Z, Qing W, Bing B (2008) Analysis on keratin amino acids in the feather of cranes and storks and the application in species identification. Sci Silvae Sin 44 (3):102–106
17. Chen Q, Hildemann LM (2009) The effects of human activities on exposure to particulate matter and bioaerosols in residential homes. Environ Sci Technol 43:4641–4646
18. Kitto A-MN, Colbeck I (1999) Filtration and denuder sampling techniques. In: Spurny KR (ed) Analytical chemistry of aerosols. Lewis, Boca Raton, pp 103–132
19. Menetrez MY, Foarde KK, Dean TR, Betancourt DA, Moore SA (2007) An evaluation of the protein mass of particulate matter. Atmos Environ 41:8264–8274
20. Kaski S (1998) Dimensionality reduction by random mapping. In: Proceedings of the International Joint Conference on Neural Networks, p 413–418
21. Arriaga RI, Vempala S (1999) An algorithmic theory of learning: robust concepts and random projection. Paper presented at the Proceedings of the 40th Annual Symposium on Foundations of Computer Science
22. Bingham E, Mannil H (2001) Random projection in dimensionality reduction: applications to image and text data. In: International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, p 245–250
23. Jolliffe IT (1986) Principal component analysis. Springer, New York
24. Abe S (2005) Support vector machines for pattern classification (advances in pattern recognition). Advances in pattern recognition. Springer, London
25. Davis JM, Giddings JC (1985) Statistical method for estimation of number of components from single complex chromatograms—theory, computer-based testing, and analysis of errors. Anal Chem 57(12):2168–2177
26. Whiting CE, Dua RA, Duffy CF, Arriaga EA (2008) Determining under- and oversampling of individual particle distributions in microfluidic electrophoresis with orthogonal laser-induced fluorescence detection. Electrophoresis 29(7):1431–1440