

# Information from Complexity: Challenges of TOF-SIMS Data Interpretation

Daniel J. Graham<sup>a</sup>, Matthew S. Wagner<sup>b</sup>, David G. Castner<sup>a,\*</sup>

<sup>a</sup>*National ESCA and Surface Analysis Center for Biomedical Problems  
Departments of Bioengineering and Chemical Engineering  
University of Washington, Seattle WA 98195-1750 USA*

<sup>b</sup>*Proctor & Gamble Company, 11810 East Miami River Road, Cincinnati, OH 45252 USA*

\* Corresponding author. (D G Castner).

---

## Abstract

Time-of-flight secondary ion mass spectrometry (TOF-SIMS) data are complex, even for the simplest systems. Yet it is within this complexity that information about sample composition, molecular orientation, surface order, chemical bonding, sample purity, etc. is contained. The challenge is how to easily extract this information from the spectra and images. Multivariate analysis (MVA) has shown promise in taming the complexity challenges presented by TOF-SIMS data while using all the information in the entire spectrum. The recent success of MVA methods such as principal component analysis (PCA) and partial least squares (PLS) in the spectroscopic and imaging analysis of organic and biological materials has led to a great increase in the interest of MVA processing of TOF-SIMS data. However, there is still a need to better understand what data to use to answer a given question, how to optimally process the data before applying MVA, and how to correctly interpret the MVA results. The challenges of TOF-SIMS data interpretation will only get more complex, especially for biological samples, further increasing the need for well-controlled MVA methodologies.

*Key words:* Multivariate Analysis, Principal Component Analysis, TOF-SIMS, Adsorbed Proteins, Self-assembled monolayers

---

## 1. Introduction

Time-of-flight secondary ion mass spectrometry (TOF-SIMS) data are complex. A typical TOF-SIMS spectrum may contain hundreds of peaks. The relative intensities of many of these peaks are interrelated since they come from the same surface species. Changes in the surface chemistry can affect the relative intensities of the peaks for a given sample system. This is true even of the simplest single component samples. Yet it is within this complexity that information about sample composition, molecular orientation, surface order, chemical bonding, and sample purity is contained. The challenge is how to easily extract this information from the spectra. This challenge is exacerbated by the potential size of TOF-SIMS data sets. Comparing the relative intensities of even a hundred peaks across a moderate group of samples can become a daunting task. This is especially true when one starts to consider not only how single peaks vary across the sample set, but how two peaks vary with respect to each other across the samples. This challenge is multiplied when considering similar analysis of TOF-SIMS images, in which one must track changes in hundreds of peak intensities across thousands of image pixels (65,636 pixels for a 256 x 256 image).

Fortunately multivariate analysis (MVA) has shown promise in taming the complexity presented by TOF-SIMS spectra. MVA routines have been used successfully with many types of data including electron spectroscopy for chemical analysis (ESCA)[1,2], infrared spectroscopy[3,4], scanning tunneling microscopy[5], Auger electron spectroscopy[6], atomic force microscopy[7], tandem mass spectrometry[8], pyrolysis mass spectrometry[9,10], and other mass spectrometry techniques[11]. TOF-SIMS data are well suited for this type of analytical methodology since the spectra generated are inherently multivariate. Multiple peaks are generated from the same surface

molecules, and their relative yields are often interrelated. MVA methods allow utilization of the entire spectrum to determine which peaks correlate with various surface treatments or chemistry changes.

MVA methods have shown success with many types of sample systems including polymers [12-19], self-assembled monolayers [20,21], proteins [22-32], residual extracellular matrix proteins[33], and TOF-SIMS images[34-36]. The promise shown by these studies has led to a great increase in the interest of MVA in TOF-SIMS data processing. New multivariate methods such as multivariate curve resolution (MCR)[37,38] are being explored and more groups are starting to apply MVA in their research.

The application of MVA methods has opened new doors for the exploration of surfaces by TOF-SIMS. Hopefully MVA methods will allow users to explore and understand the increasingly complex surfaces that are being analyzed by TOF-SIMS. Nevertheless, the ToF-SIMS community has only begun exploring the capabilities of what methods such as principal component analysis (PCA) and partial least squares (PLS) can do, and possibly more importantly, no guidelines have been established for data preprocessing before these methods are applied. To realize the full power of these methods we need to better understand what data to use to answer a given question, how to optimally process the data before applying MVA and how to correctly interpret the outcomes from the analysis. Also, good experimental design is essential to maximize the amount of information MVA methods can leverage out of a dataset. To fully address these issues time and effort need to be spent on understanding how MVA methods work and how the various data pretreatment methods affect the data outcome.

To this end, this paper will highlight some of the successes of PCA as applied to TOF-SIMS spectra and images. Opportunities and challenges will also be addressed in how MVA methods can be applied to maximize the information gained from TOF-SIMS data and what needs to be done to assure quality in both education about MVA methods and their application to TOF-SIMS spectra and images.

## 2. Experimental

### 2.1 Current practice

The application of MVA methods to TOF-SIMS data is fairly new. It has only been within the last few years that MVA has become more widespread among the TOF-SIMS community[39]. The success of MVA methods such as PCA has increased interest in applying MVA methods to TOF-SIMS data. Though there are many MVA methods, most of them work towards one central task of summarizing the variance patterns within a data set. The variance in the data describes the differences between samples and sample groups. In the case of TOF-SIMS data, these differences are due to changes in the relative intensities of peaks within the sample spectra. Since a typical TOF-SIMS spectrum can contain hundreds of peaks, it would be difficult to try and determine how each of the peaks changes individually. For example with only 200 peaks it would take 200 plots to look at how individual peaks vary across the samples set. If this were extended to plots of 1 peak against another it would take 19,900 plots to visualize how all the peaks varied with each other. This, of course, is not a feasible analysis scheme and has resulted in a data processing approach where only a few key TOF-SIMS peaks are selected for analysis. This means most the information present in the TOF-SIMS data is not used. MVA methods, such as PCA, circumvent this problem by efficiently and concisely summarizing all the information across entire set of variables and samples. Since PCA is the most common MVA method used for TOF-SIMS data analysis, this paper will concentrate on the use of PCA with TOF-SIMS data. For more detailed information about PCA see the descriptions given by Wold [40] and Jackson [41,42].

The input to PCA is a matrix where the rows are samples (i.e. spectra) and the columns are variables (i.e. peak intensities). The inputs for each cell in the data matrix are the peak areas for a

given peak from a given spectrum. Before PCA is applied to this data, the matrix is often pre-processed as discussed below.

PCA describes the variance within the input data matrix by determining the directions of greatest variation within the data. This can be thought of graphically as an axis rotation to capture the sequential directions of greatest spread within a data set (See Figure 1). Mathematically PCA consists of the singular value decomposition of the variance-covariance matrix, yielding the characteristic vectors (eigenvectors) and characteristic roots (eigenvalues) of the variance-covariance matrix. Thus, the new variables (PC1, PC2, etc.) formed by this transform are linear combinations of the original variables (TOF-SIMS peak intensities). Nonlinear PCA data processing schemes are also available, but typically are not needed for analysis of TOF-SIMS data.

The routines for determining a PCA solution are well established and many PCA software packages are available. Many of these programs are highly automated, so PCA can be easily run by just importing the data and executing a few commands. However, before applying PCA to a data set, it is necessary to properly pre-treat the data to assure that the variance patterns highlighted are truly related to the chemical differences between the samples and not to mathematical differences in peak intensities. Data pretreatment can include scaling, centering, and non-linear transformations. A discussion of the assumptions made with these types of data treatments and their affects for TOF-SIMS data has recently been given [39]. Reviews of normalization [43] and centering and scaling [44] for multivariate analysis have also been given. Since PCA finds the variations in the dataset, good experimental design coupled with the proper data preprocessing is essential. For example if the set of samples being investigated have several properties that vary (substrates, type of proteins, surface coverage, protein conformation, protein orientation, etc.) it will extremely challenging, if not impossible, for MVA methods to separate the effect of the different variables using a single dataset. Thus, if the objective of the analysis is to examine the effect of one particular variable, then it is best to design an experiment where the other variables are held constant and the data pre-processing and peak selection is done in a manner that will focus on the variable of interest.

Data normalization before MVA varies among researchers. One scaling method often used for TOF-SIMS data is normalization of the data by multiplication of the individual peak intensities by a scalar value. Normalization can account for variations in the data acquisition that may not relate to actual surface chemical differences, such as variations in primary ion current or detector efficiency. Common normalization methods seen in the literature for preprocessing TOF-SIMS data for MVA include normalization to the total intensity of a given spectra [20,21,45], to the most intense peak in a given spectrum [16,29,46-48], and to the sum of all selected peaks from the spectrum [12,14,15,17-19,22-28,30-32,49,50].

It is also common to mean-center data before applying MVA methods. Mean-centering is done by subtracting the mean value for a variable from that variable in each spectrum. This results in a data set where each variable varies across a common mean of zero. This data transformation makes it so that differences seen are due to sample variance instead of sample means.

Other data scaling methods include mean-scaling (scaling each variable to the mean value for that variable across the data set), variance scaling (scaling each variable to unit variance) and non-linear transformations (such as log transformations). Each data preprocessing method carries with it its own set of assumptions [39], and it is up to the researcher to understand if these assumptions are valid. The data processing methods used with each set of data described in this document are given with the dataset.

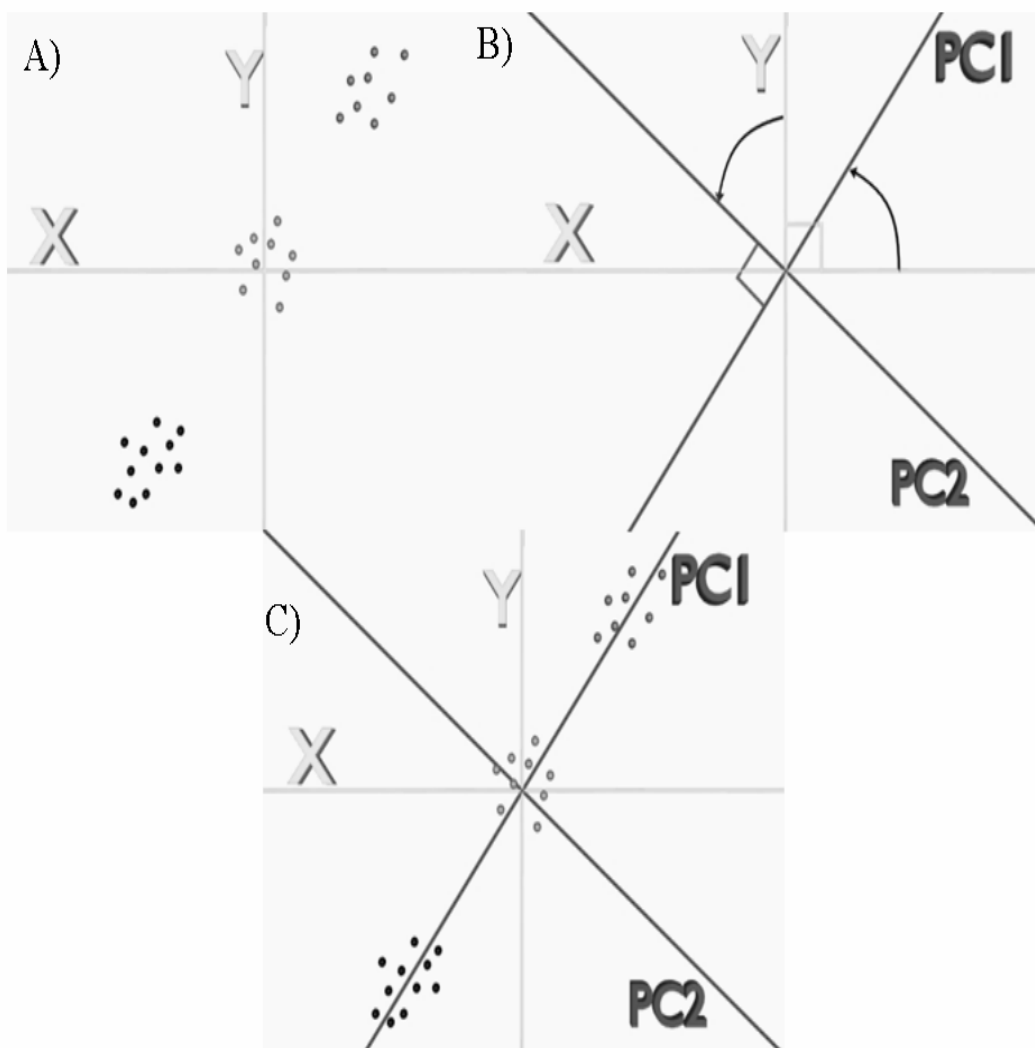


Fig. 1. Graphical representation of Principal Component Analysis.

PCA can be visualized as an axis rotation. The PC axes are rotated (B) to capture the greatest directions of variation in the original data (A). An overlay of the PC axes and the original axes is shown in C.

## 2.2 Future needs

The use of multivariate analysis will likely increase as will the complexity of the surfaces being probed by TOF-SIMS. The successful application of these methods to complex materials such as biological tissue will rely on our understanding of how these analytical methods work and how to properly execute them. For this we will need to understand how the various methods of data preprocessing affect the outcome of the data interpretation. Clear illustrations of the effects of normalization, scaling and data transforms will enable basic education of the TOF-SIMS community and will allow better application of multivariate analysis methods. This type of effort would be facilitated by the creation of standard data sets that can be used to explore data pretreatments and MVA. Papers should be written to report on how these data pretreatments affect the outcome of the MVA methods used. This type of information could help establish guidelines for applying MVA to TOF-SIMS data. Keenan and Kotula have shown good examples of this with their exploration of data scaling that takes into account the Poisson noise distribution of TOF-SIMS imaging data[51,52]. Additional work in this area is needed. It is likely that different data sets will require different methods of preprocessing. This may depend on the type of data (images versus

spectra) and the type of samples (SAMs, tissues, cells, etc.). Through exploring data pretreatments on various types of data, it is hoped that useful guidelines can be established for multivariate processing of TOF-SIMS data.

Along with the known, common data pretreatment methods, new methods of data pretreatment should be explored. Insight from the application of multivariate methods to other spectroscopic techniques should be investigated. A clear understanding of data preprocessing should be established to make sure that MVA is not a 'black box', but a tool that is both understood and utilized to its fullest extent. To date most of the multivariate analysis of TOF-SIMS data has analyzed positive and negative spectra separately. There is a need to develop normalization procedures that allow positive and negative secondary ion data to be combined into one dataset. Likewise, development of normalization procedures for combining TOF-SIMS data with other surface analysis data (x-ray photoelectron spectroscopy (XPS) data, contact angle data, etc.) into one dataset for MVA also needs to be addressed.

Coupled with this is the need for more work on standard spectral libraries for key components of the systems being investigated, such as cells, tissues, and diagnostic arrays. The effects on the fragmentation pattern of materials by the various cluster and atomic ion beams now being used also need to be understood. These developments coupled with enhanced data analysis routines should allow more efficient and insightful interpretation of TOF-SIMS data.

### 3. Results and Discussion

#### 3.1 Current progress

PCA has been successfully applied to many types of surfaces including self-assembled monolayers (SAMs), proteins, and images. Several examples from these systems will be presented below to highlight the strengths of PCA as applied to TOF-SIMS data.

##### 3.1.1 PCA of TOF-SIMS Spectra from Self-Assembled Monolayers

SAMs are ideal systems for studying the effects of surface chemical or structural changes on the fragmentation pattern generated by TOF-SIMS. The well-defined structure of SAMs allows direct control of the surface chemistry, which enables more in-depth interpretation of trends observed by TOF-SIMS. In addition, SAMs are excellent model systems for developing correlations between surface properties and material performance.

Previous work in our labs has shown that TOF-SIMS can detect changes in the structure of SAMs assembled for various assembly times[21]. In this work, PCA was used to analyze a series of SAMs of dodecanethiol assembled from a dilute ethanol solution. For this analysis all the peaks in the spectrum greater than 3 times the background level for a given region were included in the peak set. Since this included most of the peaks in a given spectrum, the data were normalized to the total intensity of the given spectrum. Before PCA, the data were mean centered. The positive and negative ion spectra were analyzed separately.

Differences in the spectra were seen for both high molecular weight molecular ion clusters and low molecular weight hydrocarbons. The samples from different assembly times could be separated from both the positive and negative ion spectra. In particular it was noted that the relative intensity of short hydrocarbons ( $C^-$ ,  $CH^-$ ,  $CH_2^-$ ,  $C_2^-$ ,  $C_2H^-$ ,  $C_2H_3^-$ ,  $C_3^-$ ,  $C_3H^-$ ,  $C_3H_2^-$ ) increased with increasing assembly time from a dilute thiol solution, while the relative intensity of longer hydrocarbons ( $C_5$  to  $C_7$ ) decreased.

One of the driving forces for SAM assembly is the attractive van der Waals forces between the methylene groups in the thiol chains. As the monolayer assembles, the attraction between the methylene groups stabilizes the layer and drives the assembly towards a crystalline structure. This attractive force decreases the probability of emitting larger hydrocarbon fragments from the SAM

layer as they are more tightly held in the crystalline lattice. Based on the trends seen in the PCA it was hypothesized that the short hydrocarbon fragments originated from the upper few carbons of the thiols in the monolayer since these atoms would have more rotational freedom and would be more susceptible to fragmentation and emission. This type of fragmentation mechanism is supported by molecular dynamic simulations of SAM bombardment.[53]

PCA of SAMs has also been shown to be able to distinguish SAMs with different end groups and quickly identify peaks that are associated with the different end group chemistries.[39] PLS of TOF-SIMS data from SAMs has been used to build a predictive model to determine the composition of mixed monolayers based on the SIMS fragmentation pattern.[39]

Here we show how PCA can also be used to track changes in the composition of mixed monolayers. Though PCA is not as robust as PLS for modeling, it has been shown to be useful for developing quantitative models from TOF-SIMS data.[13,49] Figure 2 shows the PCA scores and loadings from a set of TOF-SIMS spectra taken from mixed monolayers of  $\text{HS}(\text{CH}_2)_{15}\text{CH}_3$  and  $\text{HS}(\text{CH}_2)_{16}\text{OH}$ . Two to six spectra were acquired across two samples from each mixture composition (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 % OH thiol). All major organic peaks were selected with the exception of known inorganic contaminants (Cl, Na, etc.) and isotopes, which were excluded to simplify the peak set. Before PCA the data were normalized to the total intensity of the spectra,  $\log_{10}$  transformed and then mean centered. This data pretreatment accentuated the differences in the high mass molecular ions that tend to carry less weight in PCA models when linear data pretreatments are used.

As seen in Figure 2, PCA can clearly separate most of the concentrations in the mixture based on the TOF-SIMS peaks selected. The PC1 loadings plot (Figure 2) shows that the separation seen between the samples is due to the specific molecular ion peaks for the respective thiols. As the percent OH thiol increases, the relative intensity of the OH thiol molecular ions would be expected to increase. PCA clearly captures this trend. To illustrate this, Figures 3 and 4 show the relative intensity of the  $\text{Au}_2[\text{M-H}]$  molecular ion cluster for the  $\text{CH}_3$  and OH terminated thiols respectively. As would be expected the relative intensity of the  $\text{Au}_2[\text{M-H}]$  peak from  $\text{HS}(\text{CH}_2)_{15}\text{CH}_3$  decreases with increasing percent OH thiol, while the relative intensity of the peak from the  $\text{HS}(\text{CH}_2)_{16}\text{OH}$  increases. It is important to note that while Figures 3 and 4 illustrate the intensity trends of the two thiols, PCA is able to capture the trends in these plus all other peaks in the data set. Figure 5 shows a plot of the average PC1 score for each sample type versus their composition, as determined by ESCA. The PC1 scores are linearly correlated with the ESCA composition. Thus PCA is not only able to distinguish between the different sample chemistries, but can provide information regarding the relative composition of the surfaces.

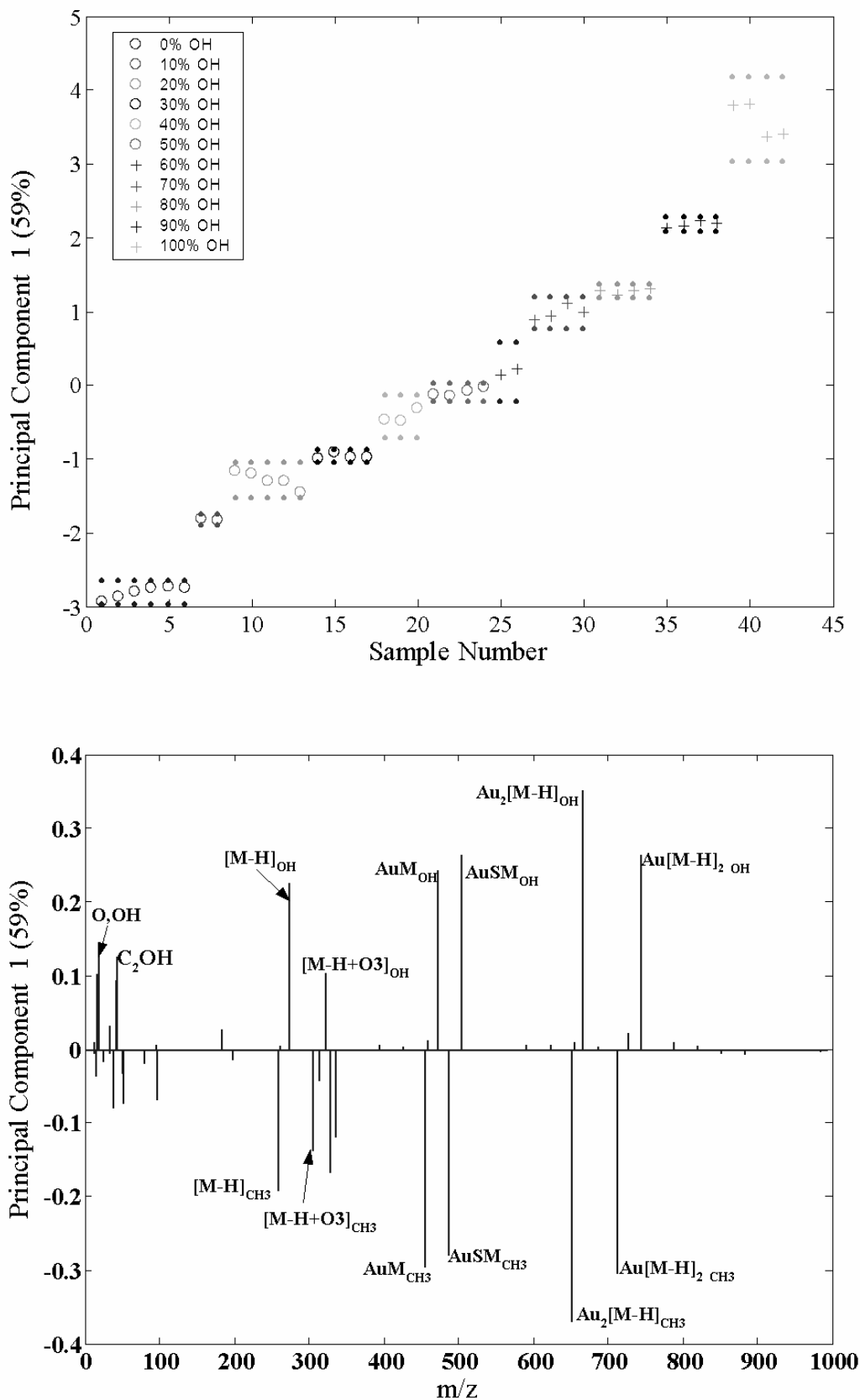


Fig. 2. PCA scores and loadings plot from mixed monolayer TOF-SIMS data.

The PC1 scores (59%) in the upper plot show clear separation of most compositions for the mixed SAMs (the dashed lines represent the 95% confidence limits). The PC1 loadings in the lower plot show the main peaks responsible for the separation seen in the scores plot. As seen from the loadings the main peaks responsible for the sample differences are the molecular ion peaks for the respective thiols and peaks indicative of the thiol head groups such as O, OH and  $C_2OH$ .

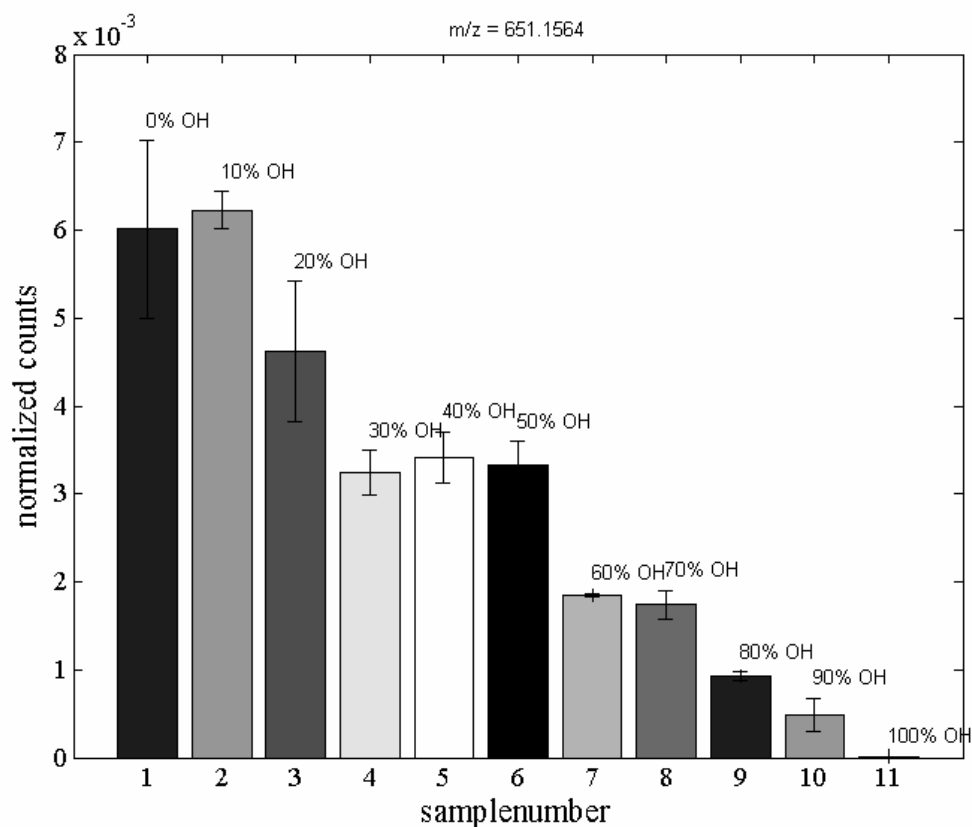


Fig. 3. Normalized raw data for the Au<sub>2</sub>[M-H] peak from HS(CH<sub>2</sub>)<sub>15</sub>CH<sub>3</sub>. The bar heights show the average value (n=2 to 6). The error bars show 1 standard deviation for each sample set.

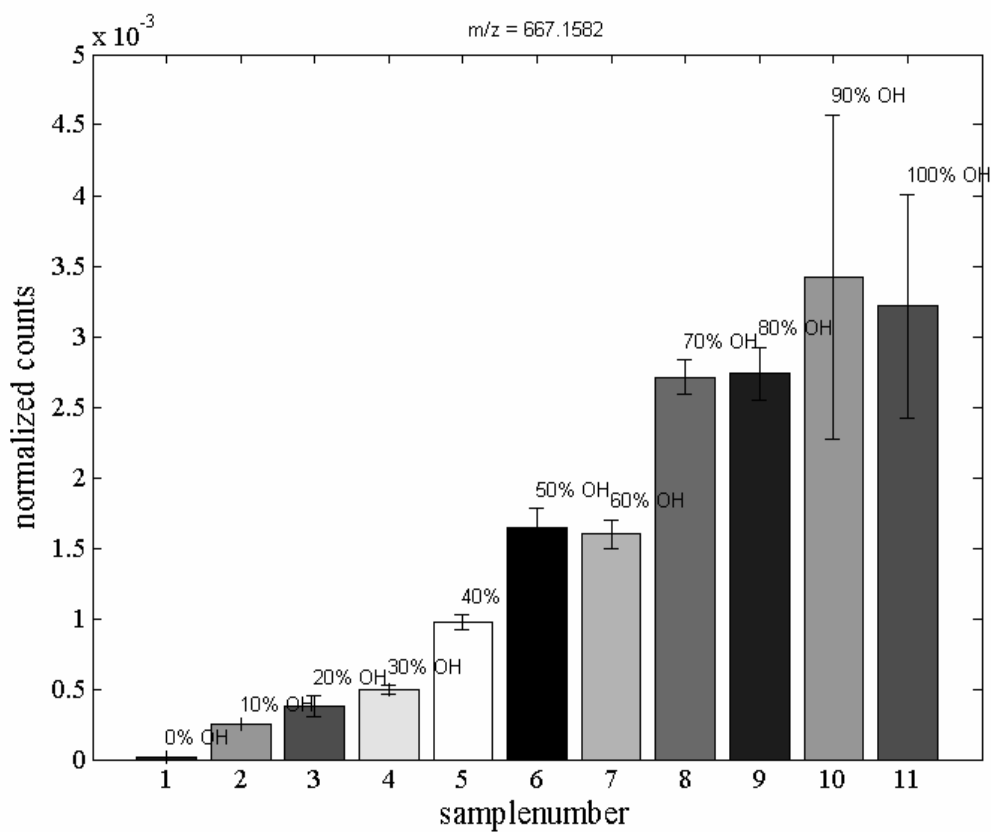


Fig. 4. Normalized raw data for the Au<sub>2</sub>[M-H] peak from HS(CH<sub>2</sub>)<sub>16</sub>OH. The bar heights show the average value (n=2 to 6). The error bars show 1 standard deviation for each sample set.

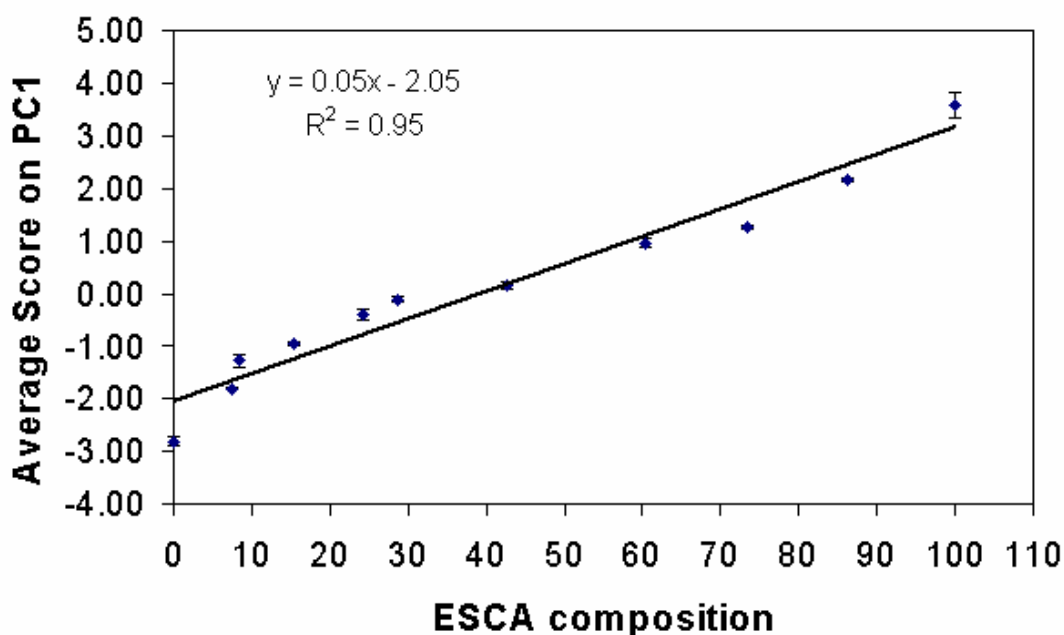


Fig. 5. PC1 scores vs ESCA composition for the mixed CH<sub>3</sub>/OH SAMs. The data are shown with increasing percent OH thiol, as determined by ESCA, plotted on the x-axis versus the average PC1 score (diamonds) from the TOF-SIMS data on the y-axis. The error bars show the standard deviation of the PC1 score measurement.

### 3.1.2 PCA of TOF-SIMS Spectra from Proteins

The TOF-SIMS spectra of proteins are complex not only because of the number of peaks, but because under typical TOF-SIMS analysis conditions no large clusters are emitted from proteins. Nevertheless, the spectra do contain fragments of all the amino acids present within the protein and most of the amino acids produce fragments that are unique. The problem is that since all proteins are made up of the same 20 amino acids, they produce the same set of TOF-SIMS peaks. The only thing that changes from protein to protein is the relative intensity of the amino acid peaks. The challenge then is to determine how the intensity pattern of amino acid peaks is related to the surface structure of the adsorbed protein film. PCA is an ideal method for this type of problem since it can summarize differences in the data and tell what peaks are responsible for these differences. PCA of proteins has been shown to be able to separate spectra from different proteins[23,24,30,54]and to determine conformational and orientational changes within proteins on surfaces[31,32,55,56].

PCA of proteins is often done using just the peaks known to be related to amino acids. This is done because inclusion of substrate peaks in the data sets often results in the differences in the substrate peaks overwhelming the differences in the amino acid peaks. Since PCA captures the largest differences in the spectra, if all peaks are included in the data set, the differences between the proteins may not be clearly seen (i.e., the variations in surface coverage of the different proteins can be larger than the variation of the amino acid fragment intensities among the proteins). While this information may also be interesting, the goal of this analysis was understanding the effect of different proteins on the ToF-SIMS spectra. PCA has been shown to be able to separate spectra from different proteins on various substrates including mica, Teflon and silicon.[26,30,57]

Figure 6 shows the scores plot from PCA carried out on 16 different single protein films adsorbed onto mica[54]. The data set included only the amino acid peaks and the data were normalized to the sum of selected peaks and mean centered before analysis. Normalization to the sum of selected peaks was chosen to monitor the relative changes in the peak intensities of just the amino acid peaks. This study showed that not only could PCA separate out the spectra based on the type of protein, but that the PCA loadings (not shown) showed consistency with the total amino acid composition of the various proteins in the data set. For example, the peaks with high positive

loadings on PC1 corresponded with fragments from amino acids that are known to have higher relative abundances in proteins with positive scores on PC1. The same was found true of the negative loadings and scores. Thus PCA distinguished the various proteins based on their overall amino acid composition.

Figure 7 shows the scores and loadings plot from 3 different protein compositions (100% fibrinogen, 50% fibrinogen/50% albumin, and 100% albumin) adsorbed onto poly(DTB suberate). As seen in the figure, PCA clearly separates the pure fibrinogen and pure albumin surfaces. The 50/50 surfaces bridge between the two pure surfaces suggesting a successful mixture of the two proteins. The loadings plot shows that peaks with positive loadings, corresponding with samples with positive scores (100% albumin and some 50/50 spectra), include amino acids that have a higher relative abundance in albumin such as  $m/z = 44$  (Cys, Ala),  $m/z = 72$  (Val),  $m/z = 84$  (Lys),  $m/z = 120$  (Phe). It is noted that the peak at mass 107 can be attributed to both Tyr and the underlying polymer. Removal of this peak does not significantly affect the PCA results. Since albumin is a smaller protein than fibrinogen it is logical to expect signal from the polymer to be more prominent on the albumin surfaces compared to the fibrinogen surfaces. Peaks with negative loadings include peaks from amino acids known to have higher relative abundances in fibrinogen such as  $m/z = 59$  (Arg),  $m/z = 60$  (Ser), and  $m/z = 130$  (Trp).

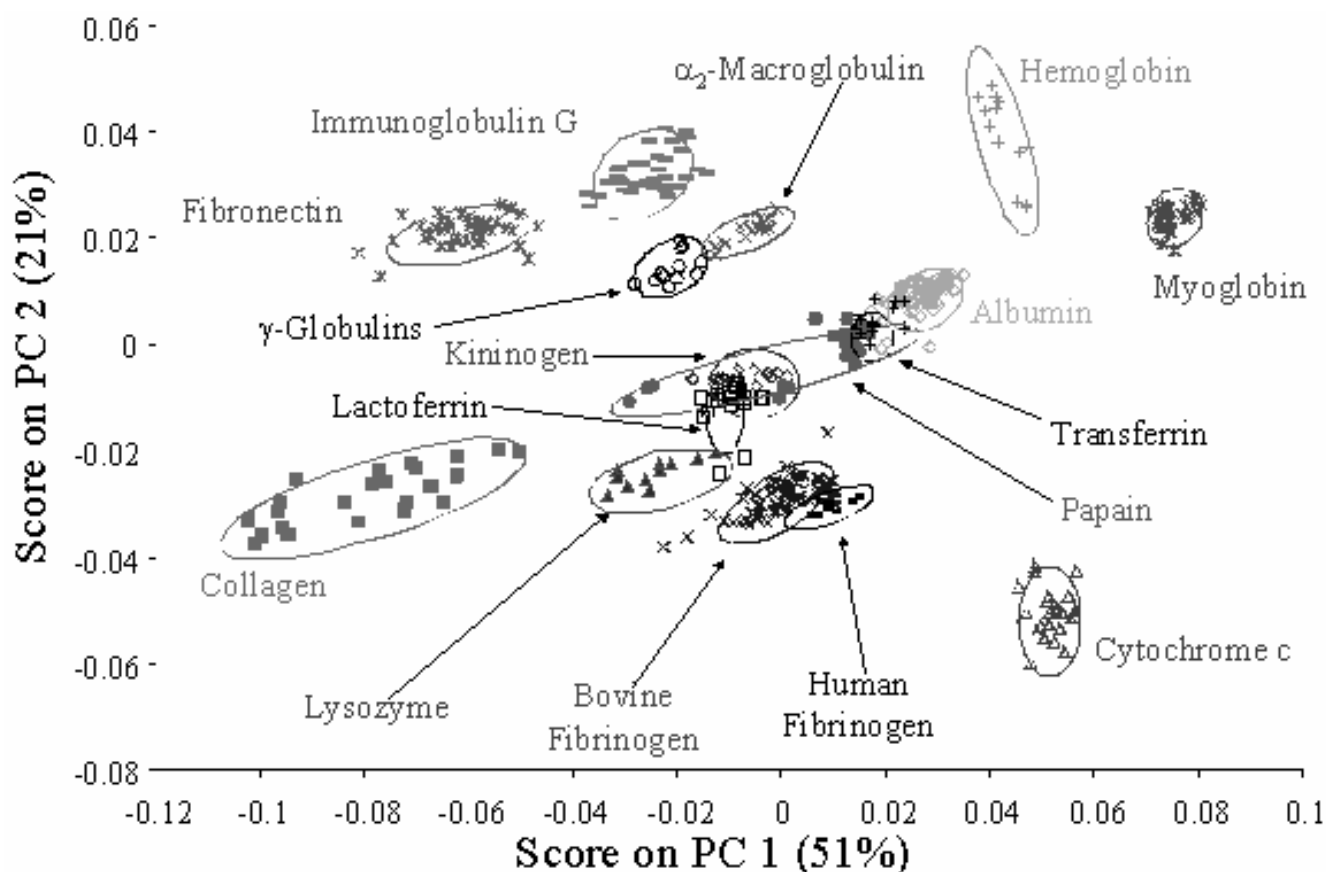


Fig. 6. PCA scores plot from single component protein films on mica. The PC1 (51%) and PC2 (21%) scores are able to separate most of the proteins based solely on the relative intensities of the amino acid fragments in the TOF-SIMS data. The circles shown are the 95% confidence limits.[54]

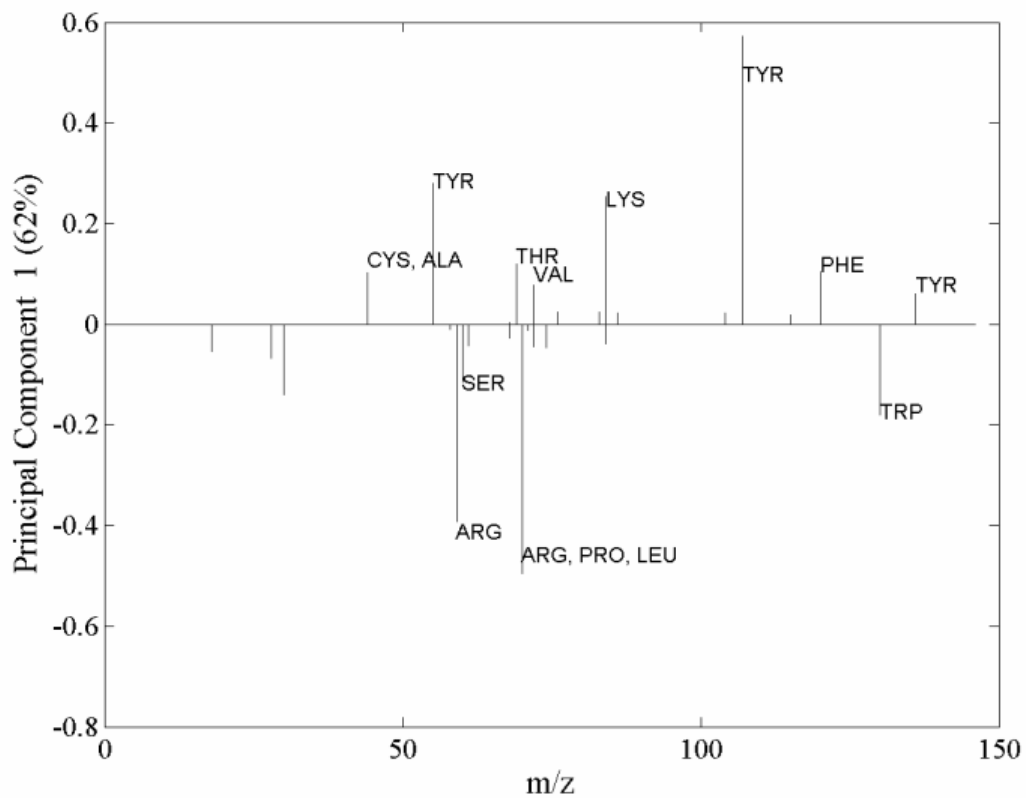
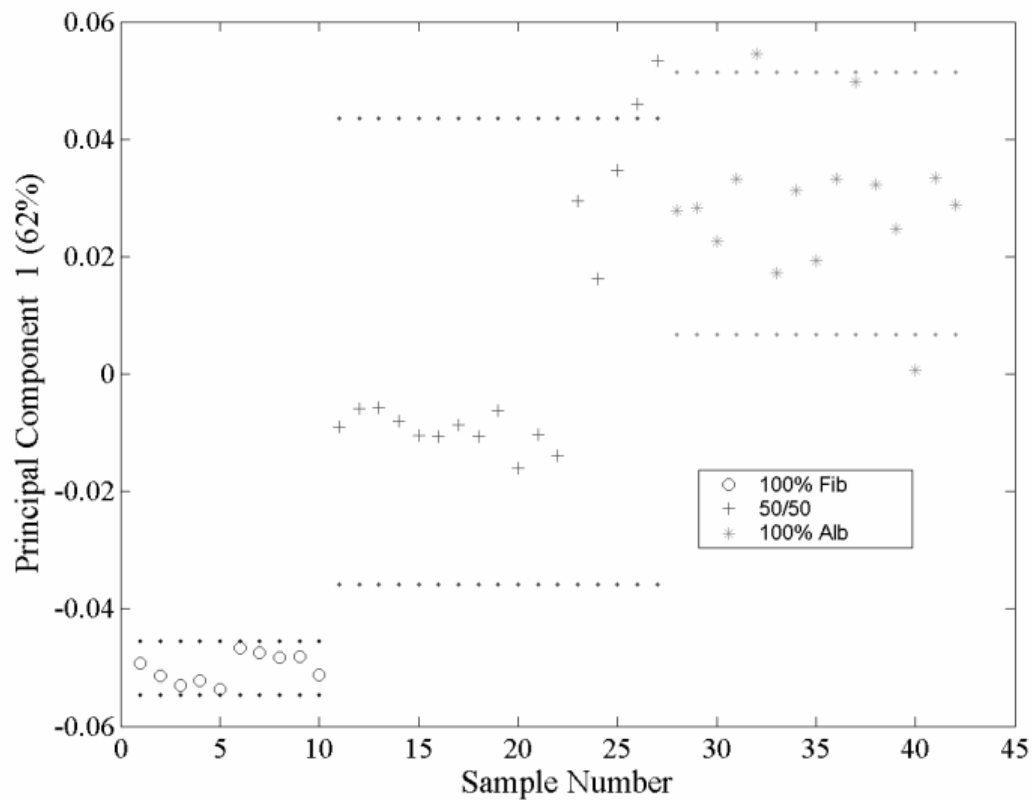


Fig. 7. PCA scores and loadings from protein adsorption onto poly(DTB suberate). PC1 (82%) is able to separate the surfaces based on the type of proteins adsorbed. The location of the 50/50 data suggests there is a mixture of both proteins on the surface. The lines represent the 95% confident limits.

### 3.1.3 PCA of TOF-SIMS Images

The complexities of analyzing TOF-SIMS data are even more challenging for TOF-SIMS imaging. Each pixel of a TOF-SIMS image contains an entire mass spectrum. This means that there are 65,536 spectra in a 256 x 256 image. This in itself is a challenge for data processing, but it is further compounded by the low signal intensities found across the image. Furthermore, to fully utilize TOF-SIMS imaging it is desirable to not only identify the components within an image, but to maintain high spatial resolution during image processing.

Wickes et. al. initiated work applying standard image filtering methodologies to TOF-SIMS images before applying PCA.[36] TOF-SIMS images were processed by down-binning the data (data compression), wavelet filtering, and boxcar filtering. Peaks were manually selected from the m/z 0 to 200 range. The data were normalized to the sum of the selected peaks for the given pixel. After filtering the data sets were autoscaled. The results from PCA were then compared. It was found that all the preprocessing methods compressed the information from the PCA into the first few PCs, whereas the main features in the unfiltered data were spread across many PCs. Figure 8 shows PCA score images from the unfiltered and down binned data sets of a sample with protein linked onto gold squares modified with NHS terminated SAMs. As seen in the figure, PC1 for both data sets captures the primary pattern feature of the protein islands, but the location of the secondary feature of background crystallites of PEG is affected by the image processing. For the unprocessed data the secondary feature appears in PC8 and is barely visible, while in the down binned data, the secondary feature shows up clearly in PC2. It was also noted by Wickes et. al. that in the unfiltered data set the secondary feature was often spread over several PCs while in the filtered data sets it was contained in one PC.[36] Thus, using image processing methods helped compress the pertinent information into the first few PCs that capture the highest percent variance in the data. This makes it easier to ensure all important features in the image are identified.

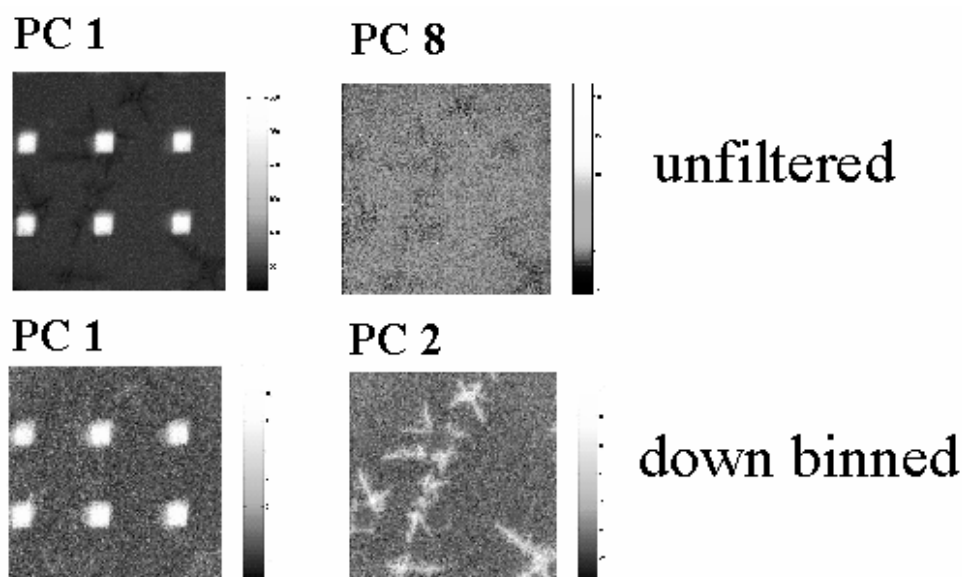


Fig. 8. PCA scores images from patterned protein on SAM modified gold islands.

PCA scores images are shown for the PCs that capture the primary feature (first column) and secondary feature (second column) of the patterned protein samples. The scores images from the unfiltered data are shown on top, while the scores images from the down binned data are shown on the bottom. The islands are 20  $\mu\text{m}$  x 20  $\mu\text{m}$  squares separated by 60  $\mu\text{m}$  spaces.[36]

### 3.2 Future needs

As TOF-SIMS analysis continues to evolve and explore new sample systems, the data analysis methodologies used will also need to evolve. The application of TOF-SIMS to cell and tissue systems is already creating new excitement in the field, but is also brings with it new

challenges. Since most biological components in cells and tissues are made of the same components, it becomes difficult to distinguish what fragment come from which cellular component. The ability to process the data from these systems will likely involve the use of MVA methods such as PCA. This is especially true of TOF-SIMS cell and tissue images that contain thousands of spectra and hundreds of peaks. Extracting useful, detailed information from these systems will present an exciting challenge for the TOF-SIMS analyst.

Part of the success of MVA methods is determined by the ability to identify the peaks that show high loadings in the PCA results. Though some of this is left to standard brute force peak identification methods, the establishment of peak databases and standard peak tables can greatly enhance the interpretation of PCA results. This has already begun with work done on characterizing the amino acids [58], and DNA building blocks[59]. These types of resources can facilitate spectral interpretation and enhance the ability to understand TOF-SIMS results. Yet these are just two of many components that make up a biological system. As TOF-SIMS becomes utilized more with cell and tissue analysis, new databases of other characteristic structures need to be developed. Standard peak libraries from homogeneous lipid films bilayers need to be created[60]. This should then be extended to mixed lipid layers. Lipid protein mixtures should also be explored to determine if there are proximity affects to the yield of characteristic lipid and protein fragments.

As MVA methods become more common among the TOF-SIMS community it is important that proper education is available on how to apply these methods. The process of normalization and scaling needs to be explored to determine the affects of various data pretreatments on the results. Through testing different methods of data pretreatment and understanding their affects on the MVA outcomes it may be possible to establish guidelines for the application of MVA methods to TOF-SIMS data. This may help facilitate understanding of TOF-SIMS spectra and how to interpret the trends and changes that occur across complex sample sets.

#### **4. Conclusions**

The challenges of TOF-SIMS data interpretation will only get more complex. This is already being demonstrated with the analysis of cells, tissues and the increasingly complex engineered surfaces used in research and development today. These developments will increase the need for well-controlled MVA methodologies. Guidelines need to be established in the proper application of MVA methods to TOF-SIMS data. Model systems need to be studied to generate databases of peaks related to the various surface components of cells and tissues. The fragmentation process of the various ion beams used today need to be more fully understood. Moving forward in these areas will aide in unlocking the information encoded in the complexities of TOF-SIMS spectra.

#### **5. Acknowledgments**

The authors gratefully acknowledge support from National Institutes of Health grants EB-002027 (National ESCA and Surface Analysis Center for Biomedical Problems) and EB-001046 (Resource for Integrated Technologies for Polymeric Biomaterials). The authors also acknowledge the stimulating discussion on MVA with many past and present members of the NESAC/BIO research group, especially Bonnie Tyler, Buddy Ratner, and Bronwyn Wickes. Caren Tidwell and Ryan Hartmaier are thanked for their technical expertise with the preparation and data acquisition for the mixed SAM and poly(DTB suberate) datasets.

## 6. References

- [1] S. Oswald, S. Baunack, 25 (1997) 942-947.
- [2] D. Michaud, M. Baril, G. Perrault, 43 (1993) 729-735.
- [3] T.F. Kaltenbach, G.W. Small, 63 (1991) 936-944.
- [4] D.M. Haaland, H.D.T. Jones, E.V. Thomas, 51 (1997) 340-345.
- [5] A.M. Bouchard, G.C. Osbourn, B.S. Swartzentruber, 321 (1994) 276-286.
- [6] P. DeVolder, R. Hoogewijs, R.D. Gryse, L. Fiermans, J. Vennik, 64 (1993) 41-57.
- [7] H. Rothe, 264 (1995) 282-290.
- [8] S. Kornig, R. Hoogerbrugge, W.R.v. Witzenburg, P.G. Kistemaker, 89 (1989) 111-124.
- [9] D. Garozzo, G. Montaudo, 9 (1985) 1-17.
- [10] B.K. Alsberg, R. Goodacre, J.J. Rowland, D.B. Kell, 348 (1997) 389-407.
- [11] M. Preu, M. Petz, 840 (1999) 81-91.
- [12] G. Coullerez, S. Lundmark, M. Malkoch, H. Magnusson, E. Malmstrom, A. Hult, H.J. Mathieu, *Appl. Surf. Sci.* 203-204 (2003) 620-624.
- [13] X.V. Eynde, P. Bertrand, *Appl. Surf. Sci.* 141 (1999) 1-20.
- [14] G. Coullerez, S. Lundmark, E. Malmstrom, A. Hult, H.J. Mathieu, *SIA* 35 (2003) 693.
- [15] G. Coullerez, D. Leonard, S. Lundmark, H.J. Mathieu, *SIA* 29 (2000) 431.
- [16] R. Canteri, G. Speranza, M. Anderle, S. Turri, S. Radice, *SIA* 35 (2003) 318.
- [17] S.M. McArthur, M.S. Wagner, P.G. Hartley, K.M. McLean, H.J. Griesser, D.G. Castner, *SIA* 33 (2002) 924.
- [18] N. Medard, C. Poleunis, X. Vanden-Eynde, P. Bertrand, *SIA* 34 (2002) 565.
- [19] N. Medard, A. Benninghoven, D. Rading, A. Licciardello, A. Auditore, T.M. Duc, H. Montigaud, F. Vernerey, C. Poleunis, P. Bertrand, *Appl. Surf. Sci.* 203-204 (2003) 571.
- [20] D.J. Graham, D.D. Price, B.D. Ratner, *Langmuir* 18 (2002) 1518-1527.
- [21] D.J. Graham, B.D. Ratner, *Langmuir* 18 (2002) 5861-5868.
- [22] S. Ferrari, B.D. Ratner, *SIA* 29 (2000) 837-844.
- [23] J.-B. Lhoest, M.S. Wagner, C.D. Tidwell, D.G. Castner, *J. Biomed. Mater. Res.* 57 (2001) 432-440.
- [24] M.S. Wagner, D.G. Castner, *Appl. Surf. Sci.* 203-204 (2003) 698-703.
- [25] M.S. Wagner, T.A. Horbett, D.G. Castner, *Langmuir* 19 (2003) 1708-1715.
- [26] M.S. Wagner, M. Shen, T.A. Horbett, D.G. Castner, *Appl. Surf. Sci.* 203-204 (2003) 704-709.
- [27] M.S. Wagner, T.A. Horbett, D.G. Castner, *Biomaterials* 24 (2003) 1897-1908.
- [28] M.S. Wagner, M. Shen, T.A. Horbett, D.G. Castner, *J. Biomed. Mater. Res.* 64A (2003) 1-11.
- [29] M. Shen, M.S. Wagner, D.G. Castner, B.D. Ratner, T.A. Horbett, *Langmuir* 19 (2003) 1692-1699.
- [30] M.S. Wagner, B.J. Tyler, D.G. Castner, *Anal. Chem.* 74 (2002) 1824-1835.
- [31] N. Xia, D.G. Castner, *J. Biomed. Mater. Res.* 67A (2003) 179-190.
- [32] N. Xia, C.J. May, S.L. McArthur, D.G. Castner, *Langmuir* 18 (2002) 4090-4097.
- [33] H.E. Canavan, X. Cheng, D.J. Graham, B.D. Ratner, D.G. Castner, *Langmuir* 21 (2005) 1949-1955.
- [34] B. Tyler, *Appl. Surf. Sci.* 203-204 (2003) 825-831.
- [35] M.C. Biesinger, P.-Y. Paepegaey, N.S. McIntyre, R.R. Harbottle, N.O. Petersen, *Anal. Chem.* 74 (2002) 5711-5716.
- [36] B.T. Wickes, Y. Kim, D.G. Castner, *SIA* 35 (2003) 640-648.
- [37] N.B. Gallagher, J.M. Shaver, E.B. Martin, J. Morris, B.M. Wise, W. Windig, *Chem. and Intell. Lab. Sys.* 73 (2004).

- [38] J.A.T. Ohlhausen, M.R. Keenan, P.G. Kotula, D.E. Peebles, *Appl. Surf. Sci.* 231-232 (2004) 230-234.
- [39] M.S. Wagner, D.J. Graham, B.D. Ratner, D.G. Castner, *Surf. Sci.* 570 (2004) 78-97.
- [40] S. Wold, K. Esbensen, P. Geladi, *Chemom. and Intell. Lab. Sys.* 2 (1987) 37-52.
- [41] J.E. Jackson, *J. Qual. Technol.* 12 (1980) 201-213.
- [42] J.E. Jackson, *A Users's Guide to Principal Components*, New York, John Wiley & Sons, Inc., 1991.
- [43] S.N. Deming, J.A. Palasota, J.M. Nocerino, *J. Chemom.* 7 (1993) 393-426.
- [44] R. Bro, A.K. Smilde, *J. Chemom.* 17 (2003) 16-33.
- [45] H.B. Lu, C.T. Campbell, D.J. Graham, B.D. Ratner, *Anal. Chem.* 72 (2000) 2886-2894.
- [46] V.H. Perez-Luna, T.A. Horbett, B.D. Ratner, *J. Biomed. Mater. Res.* 28 (1994) 1111-1126.
- [47] A. Chilkoti, B.D. Ratner, D. Briggs, *Anal. Chem.* 65 (1993) 1736-1745.
- [48] A. Chilkoti, A.E. Schmierer, V.H. Perez-Luna, B.D. Ratner, *Anal. Chem.* 67 (1995) 2883-2891.
- [49] X.V. Eynde, P. Bertrand, *SIA* 25 (1997) 878-888.
- [50] S.L. McArthur, M.W. Halter, V. Vogel, D.G. Castner, *Langmuir* 19 (2003) 8316.
- [51] M.R. Keenan, P.G. Kotula, *Appl. Surf. Sci.* 231-232 (2004) 240-244.
- [52] M.R. Keenan, P.G. Kotula, *SIA* 36 (2004) 203-212.
- [53] K.S.S. Liu, C.W. Yong, B.J. Garrison, J.C. Vickerman, *J. Phys. Chem. B* 103 (1999) 3195-3205.
- [54] M.S. Wagner, D.G. Castner, *Langmuir* 17 (2001) 4649-4660.
- [55] M. Henry, C. Dupont-Gillain, P. Bertrand, *Langmuir* 19 (2003) 6271-6276.
- [56] H. Wang, D.G. Castner, B.D. Ratner, S. Jiang, *Langmuir* 20 (2004) 1877-1887.
- [57] O.D. Sanni, M.S. Wagner, D. Briggs, D.G. Castner, J.C. Vickerman, 33 (2002) 715-728.
- [58] D.S. Mantus, B.D. Ratner, B.A. Carlson, J.F. Moulder, *Anal. Chem.* 65 (1993) 1431-1438.
- [59] C.J. May, H.E. Canavan, D.G. Castner, *Anal. Chem.* 76 (2004) 1114-1122.
- [60] A.G. Sostarecz, D.M.C. Jr., C.M. McQuaw, S. Sun, A.G. Ewing, N. Winograd, *Langmuir* 20 (2004) 4926-4932.