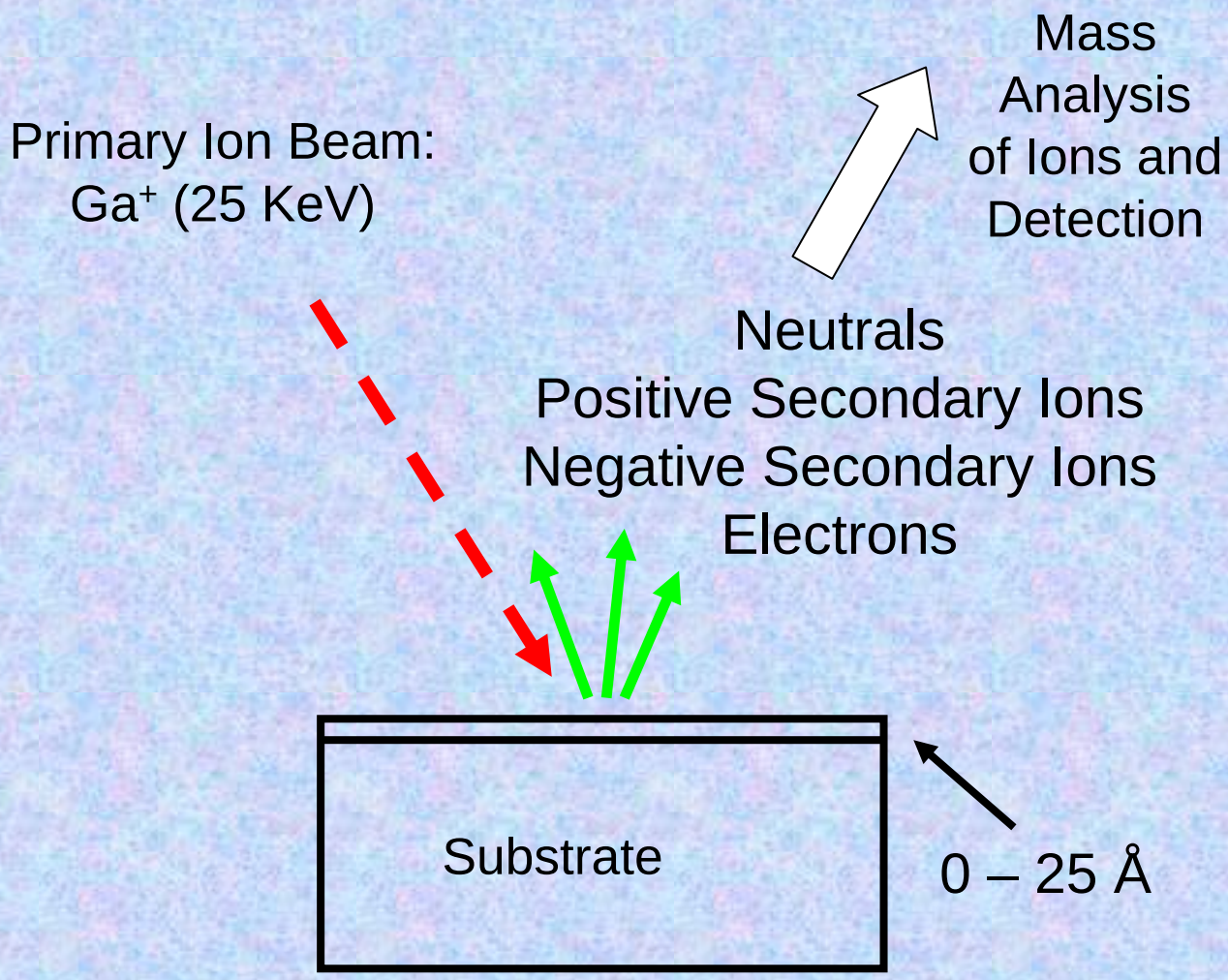
# Advanced Analytical Methods for Fuels Characterization

## A Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS)/ Chemometrics Analysis of Coal

Matthew R. Linford,[1,*] Lei Pei,[1] Larry L. Baxter[2]
[1]Department of Chemistry and Biochemistry
[2]Department of Chemical Engineering
Brigham Young University
Provo, UT 84602

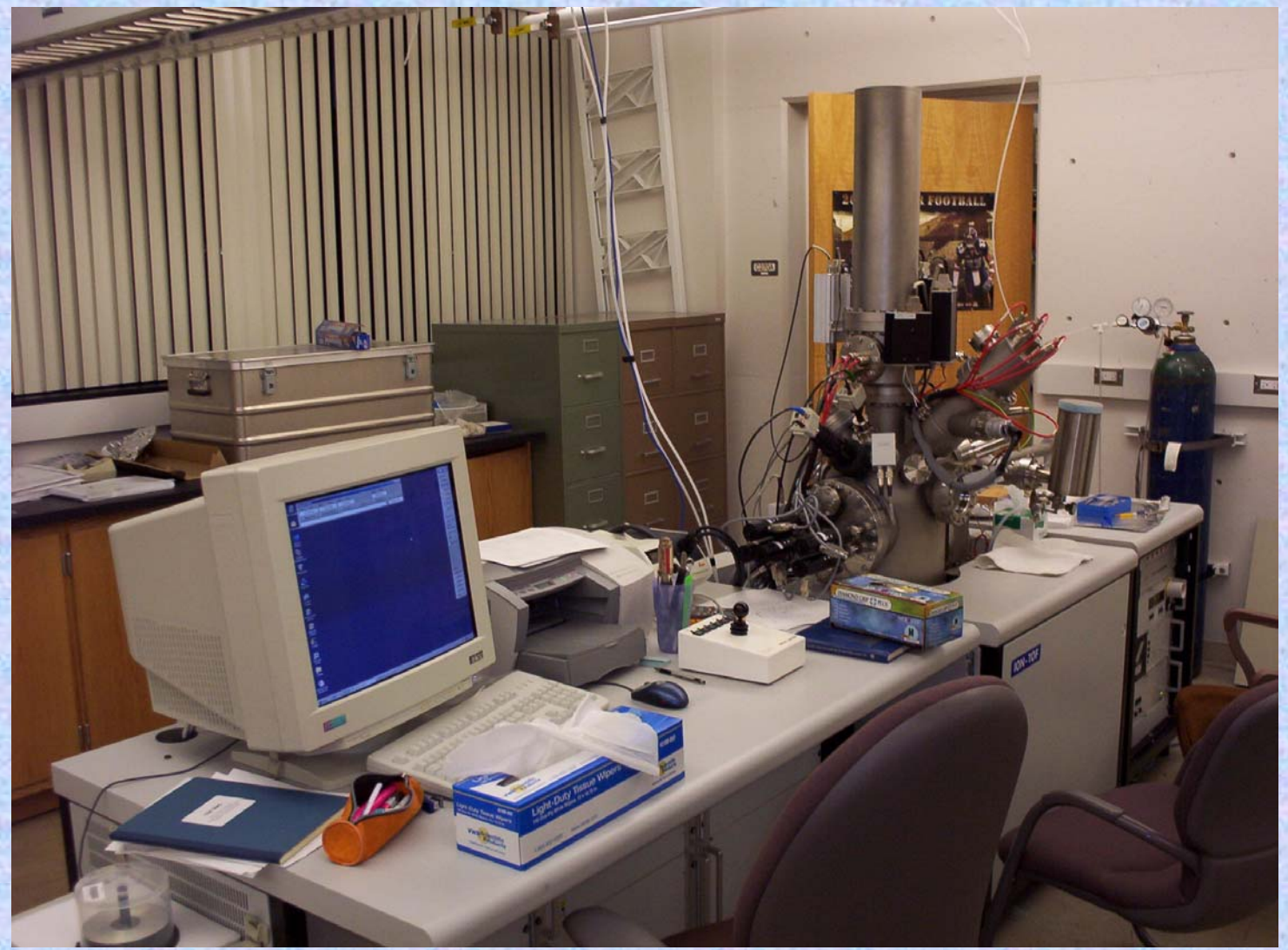
Presentation given at ACERC
February 28, 2007

# Static ToF-SIMS: A Chemical Microscope




Note: 4 eV corresponds to ~ 92 kcal/mol

# BYU Time-of-Flight Secondary Ion Mass Spectrometer (ToF-SIMS)

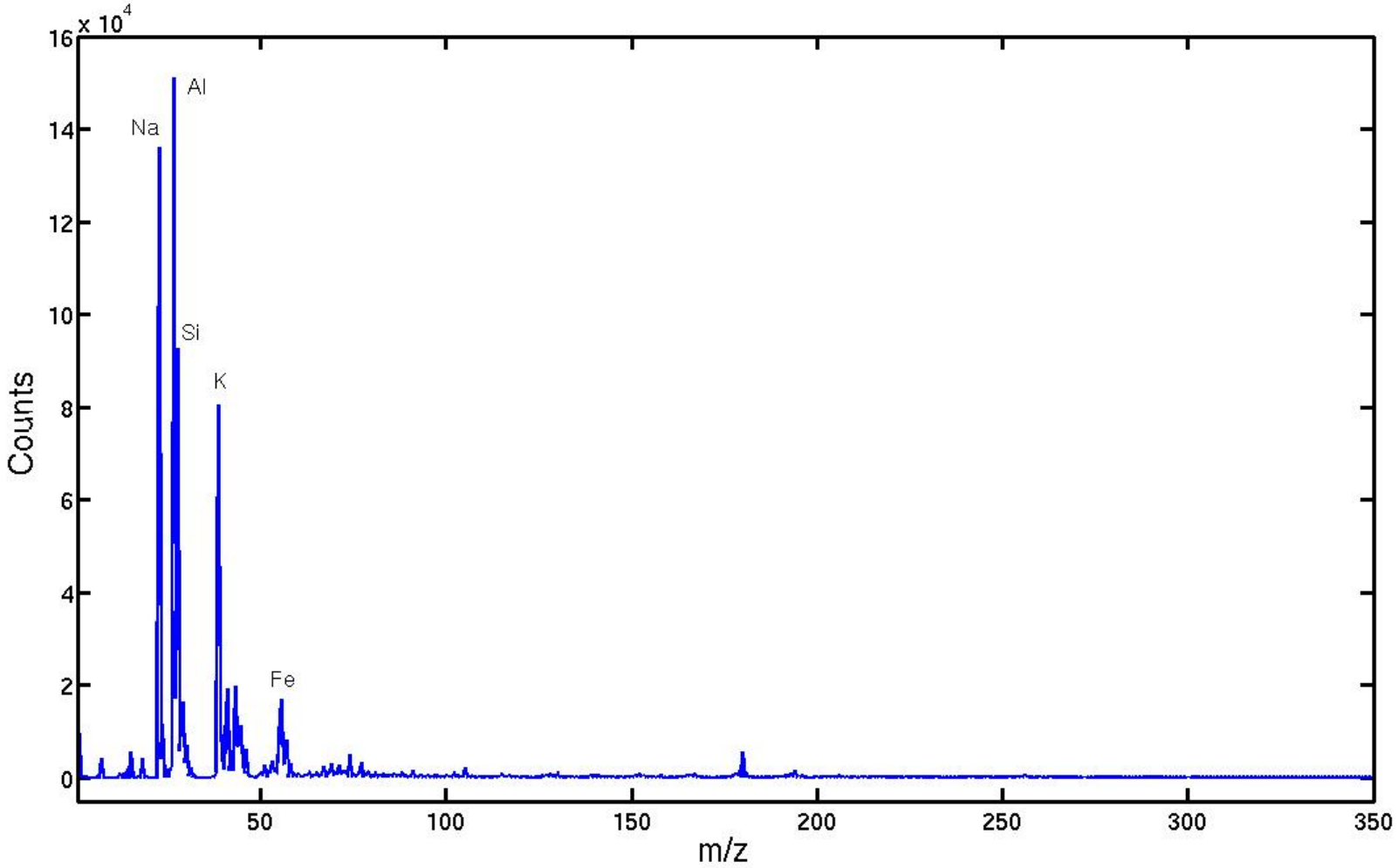


# ToF-SIMS Analysis of Coal

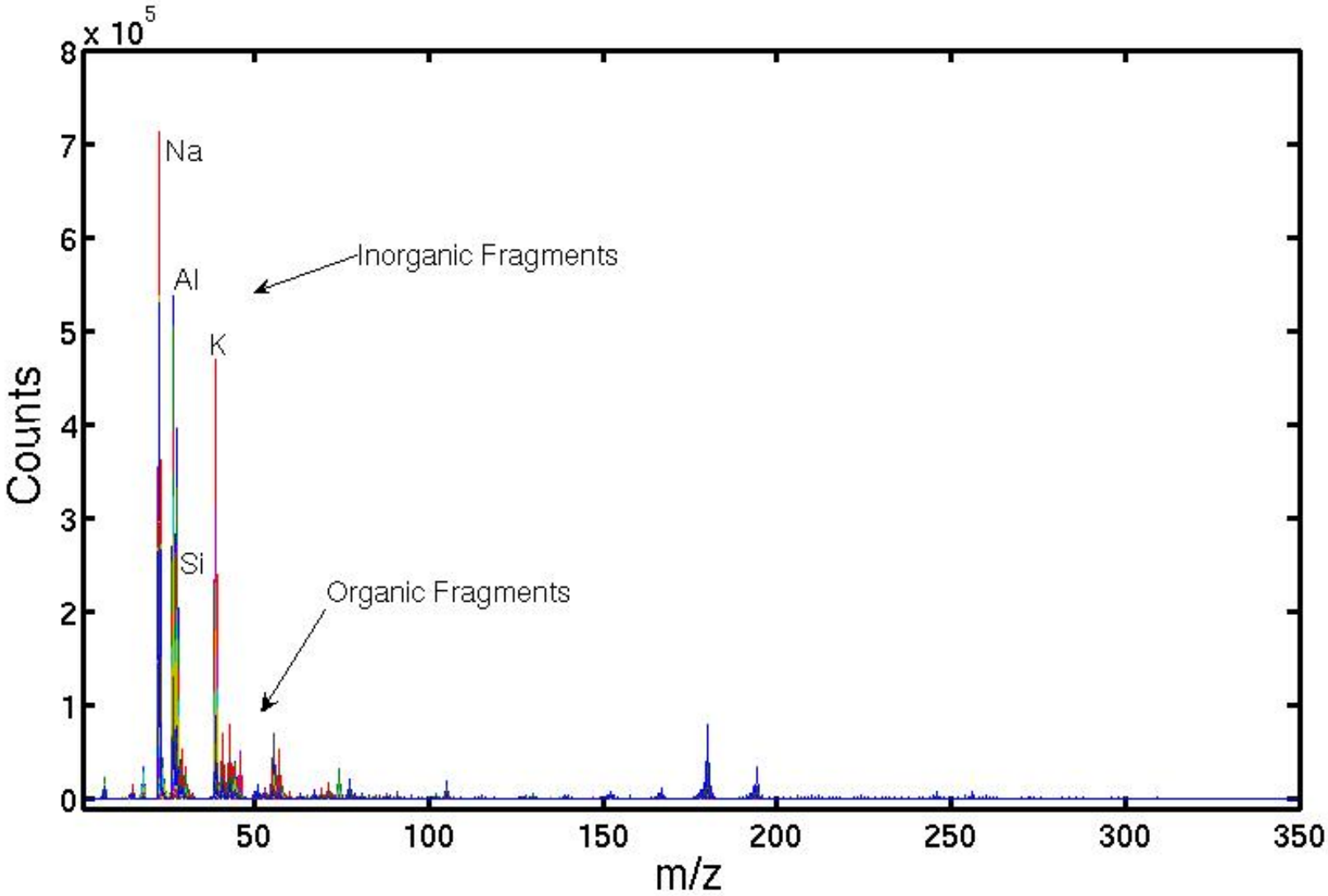
Sample Prep:

- Coal samples were ground with a mortar and pestle
- They were immediately mounted on double sticky tape
- 500μm x 500μm regions were then analyzed by ToF-SIMS
- *ca*. 100 spectra were taken of *ca*. 30 different coal samples

# Average ToF-SIMS Spectrum of Coal Samples

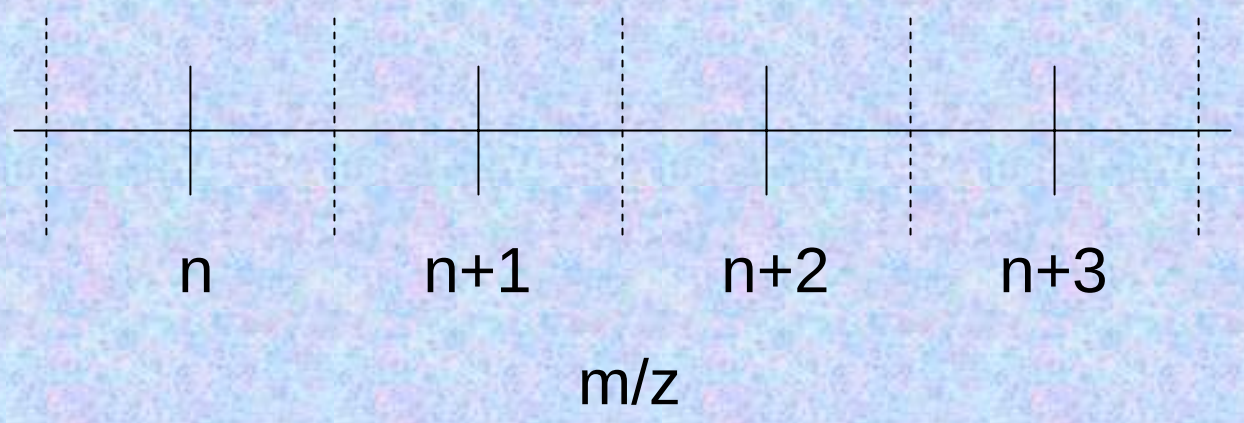


# All of the ToF-SIMS Spectra Superimposed

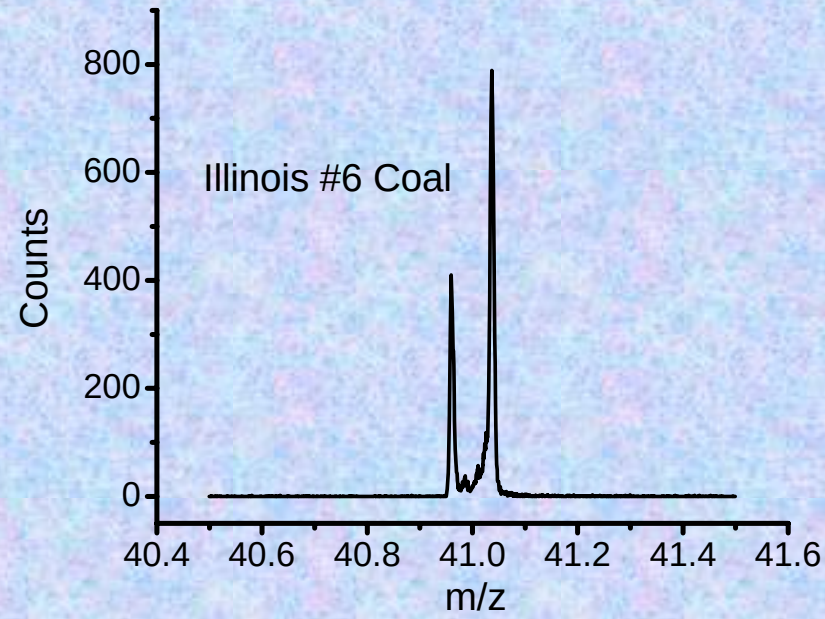


# Data Preprocessing

- To simplify spectral analyses of complicated spectra, mass spectral data are commonly binned, typically to unit mass resolution.
- For our spectra we binned 1099 regions.



# Binning to Half Unit Mass Resolution




Mass excess vs. mass deficit

| Isotope | mass | abundance |
|---|---|---|
| $^{1}H$ | 1.0079 | 99.99% |
| $^{4}He$ | 4.0026 | 100.00% |
| $^{7}Li$ | 7.0160 | 92.48% |
| $^{9}Be$ | 9.0123 | 100.00% |
| $^{11}B$ | 11.0093 | 81.02% |
| $^{12}C$ | 12.0000 | 98.89% |
| $^{14}N$ | 14.0031 | 99.64% |
| $^{16}O$ | 15.9949 | 95.77% |
| $^{19}F$ | 18.9984 | 100.00% |
| $^{20}Ne$ | 19.9924 | 90.92% |

# Data Preprocessing, Cont.

- All mass spectra were normalized with the 1-Norm. The 1-Norm is useful for correcting spectra that are identical, except to some constant factor.
- Other possible normalization schemes are the 2-Norm, which creates a vector of unit length, or the Inf-Norm, which gives the maximum peak a value of one.

$$x_{ij}^{1-Norm} = \frac{x_{ij}}{\sum_{i=1}^{I} x_{ij}}$$

$$x_{ij}^{2-Norm} = \frac{x_{ij}}{\sum_{i=1}^{I} x_{ij}^{2}}$$

$$x_{ij}^{Inf-Norm} = \frac{x_{ij}}{x_{ij,\max}}$$

# Normalization, Mean Centering

| | Var 1 | Var 2 | Var 3 | … | Var m |
|---|---|---|---|---|---|
| Spectrum 1 | $S_{11}$ | $S_{12}$ | $S_{13}$ | … | $S_{1m}$ |
| Spectrum 2 | $S_{21}$ | $S_{22}$ | $S_{23}$ | … | $S_{2m}$ |
| Spectrum 3 | $S_{31}$ | $S_{32}$ | $S_{33}$ | … | $S_{3m}$ |
| … | … | … | … | | |
| Spectrum n | $S_{n1}$ | $S_{n2}$ | $S_{n3}$ | … | $S_{nm}$ |

normalization -- a row operation

mean centering -- a column operation
(not essential for cluster analysis but critical for PCA)

$$^{cen}x_{ij} = x_{ij} - \bar{x}_j; \quad \bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$$

# Each Spectrum Can be Viewed as a Vector and/or a Point in a Hyperspace

Spectrum 1 = ($S_{11}$, $S_{12}$, $S_{13}$, …, $S_{1m}$)
Spectrum 2 = ($S_{21}$, $S_{22}$, $S_{23}$, …, $S_{2m}$)
Spectrum 3 = ($S_{31}$, $S_{32}$, $S_{33}$, …, $S_{3m}$)

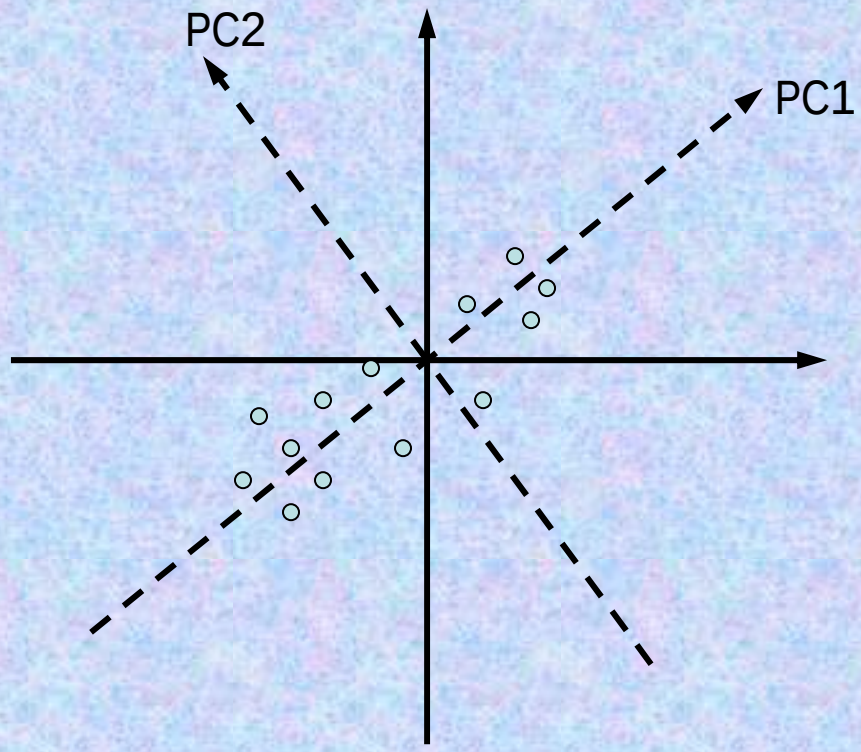
…

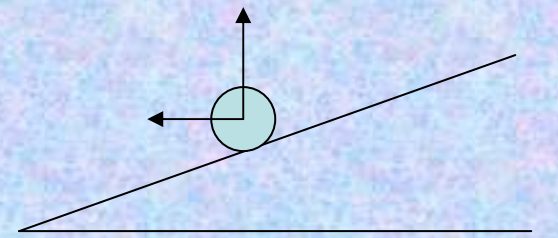
Spectrum n = ($S_{n1}$, $S_{n2}$, $S_{n3}$, …, $S_{nm}$)

# Principal Components Analysis (PCA) of Mass Spec Data

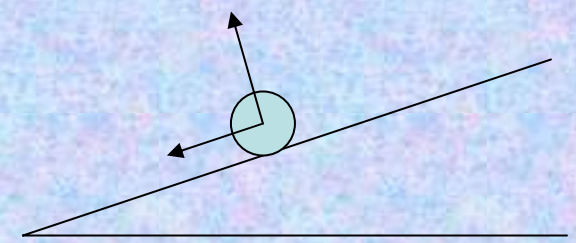
- Each peak or spectral region is considered to be a unique variable. Complete spectra are represented as vectors and/or single points in a hyperspace.
- The data are binned, normalized, and mean centered.
- Each spectrum is then “plotted” as a single point in a hyperspace of these variables
- The coordinate system is rotated to capture the maximum variation in the data along the axes.
- This analysis is called Principal Components Analysis (PCA).
- The projections of the original data points (spectra) on the new axes (principal components) are called scores.
- The contributions of the original axes (variables) to the new axes (principal components) are called loadings.

# A Simple Way to Explain PCA



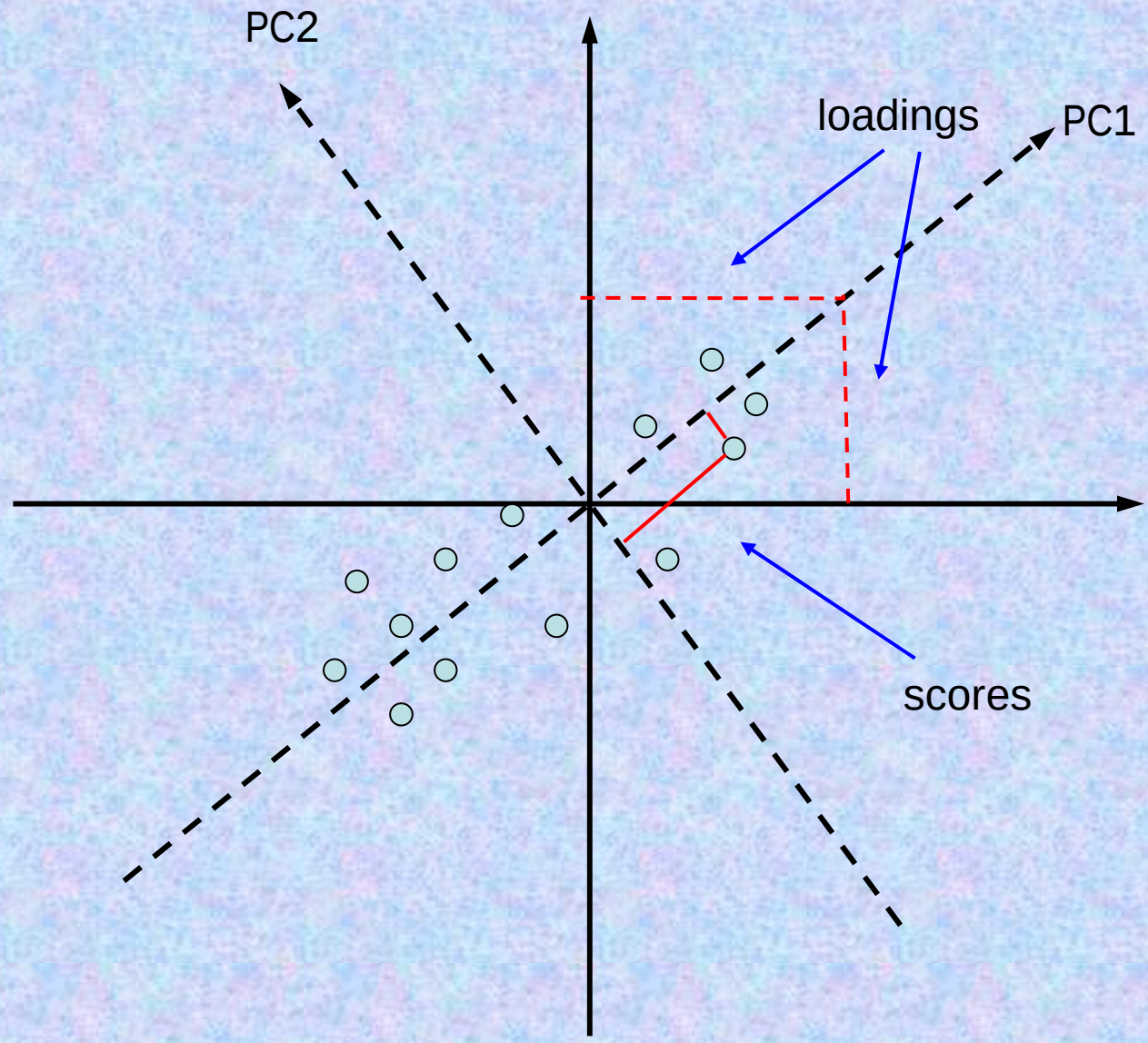
a



b

A la Bonnie Tyler.

# Scores and Loadings



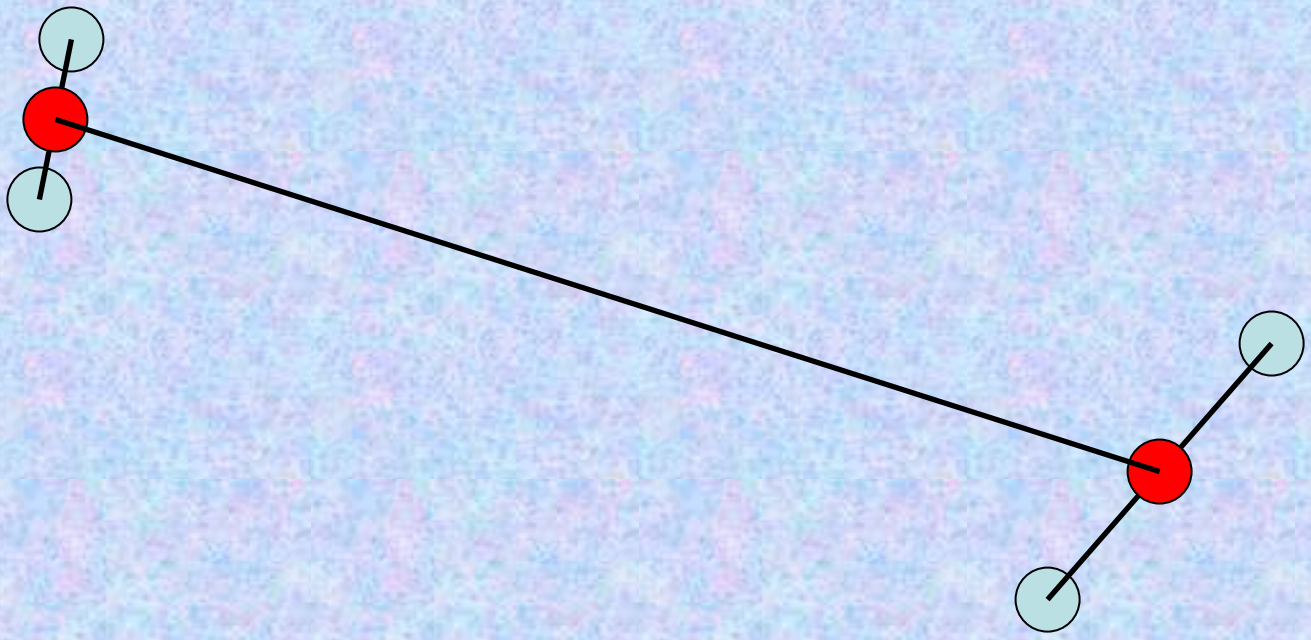
# Cluster Analysis

- Calculate the distance between points in a hyperspace.
- The simplest measure of distance between two points $x_i$ and $x_j$ in a hyperspace is the Euclidean distance:

$$d_{ij} = \left[ \left( \vec{x}_i - \vec{x}_j \right)^T \left( \vec{x}_i - \vec{x}_j \right) \right]^{1/2}$$

- Create a plot (a dendrogram) that reflects the distances between the points.
- This is an established method that has been used for many years, especially in the biological sciences as a classification tool.

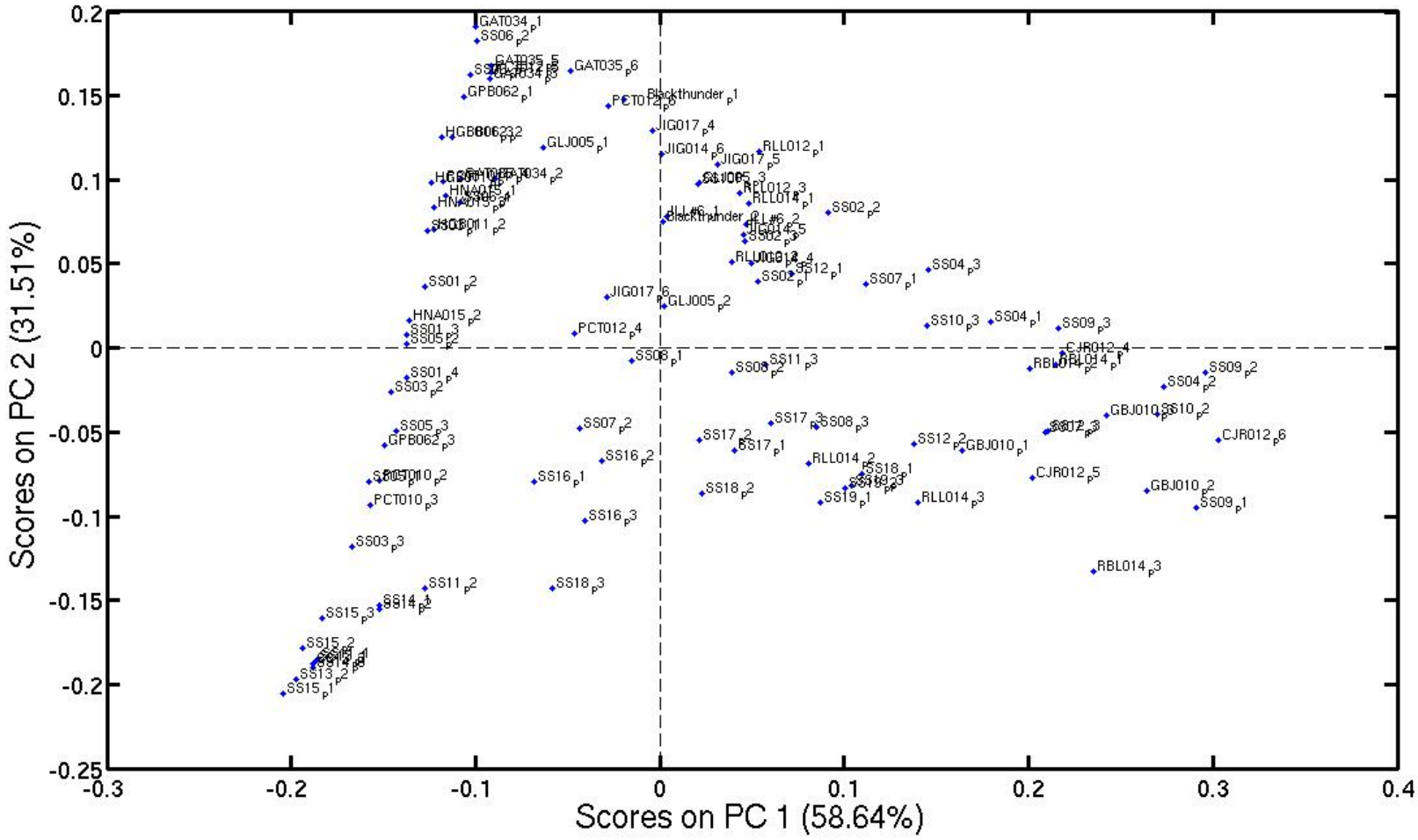
# Use the K-Means Method of Clustering



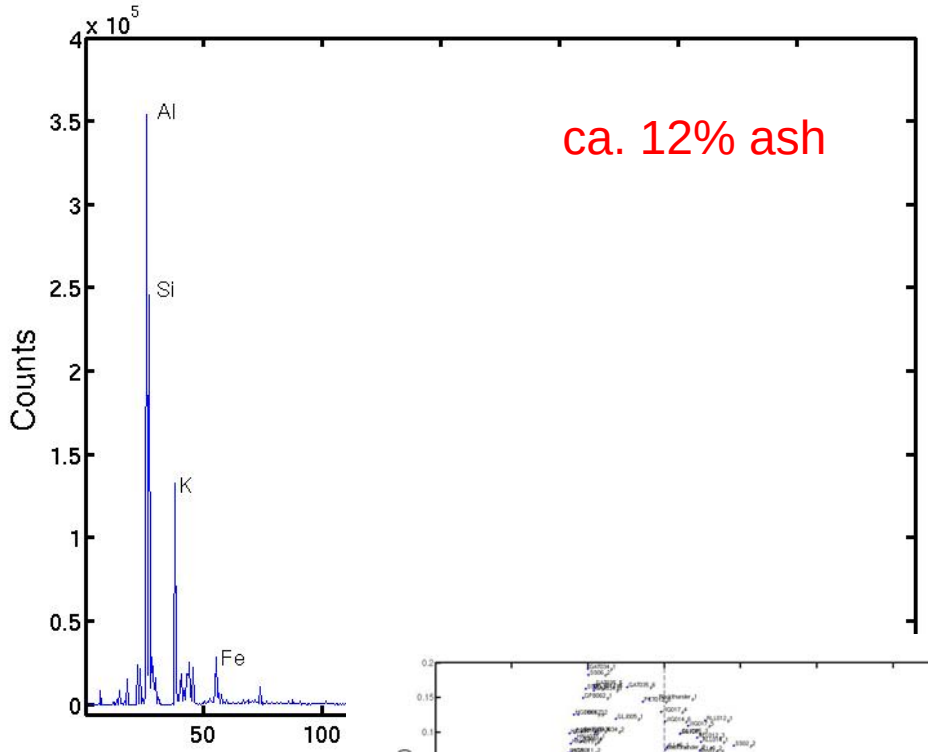
points in a hyperspace

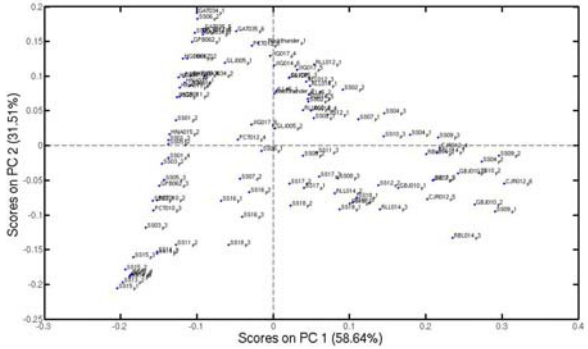
# Scores Plot of PC1 vs. PC2

(PC1 and PC2 account for just over 90% of the variance in the data.)



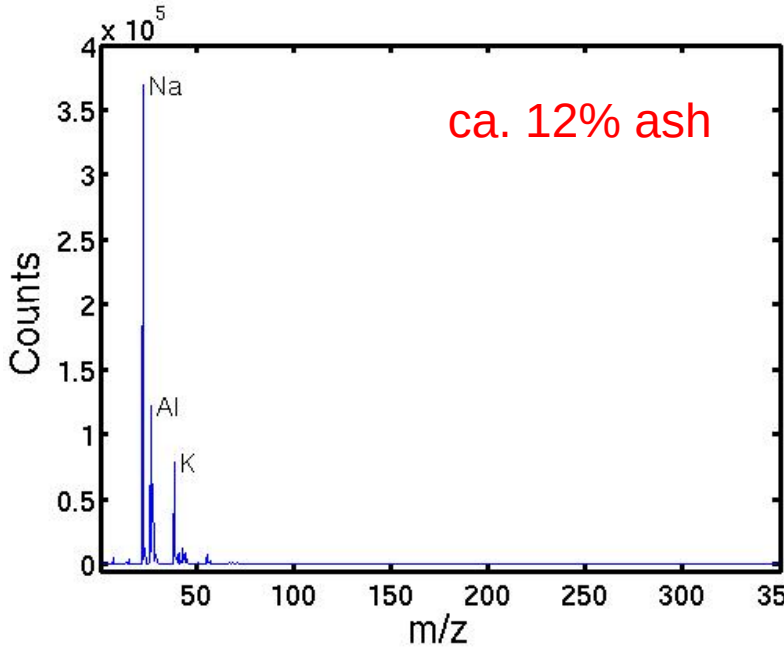
# Five Average Spectra at Each Vertex



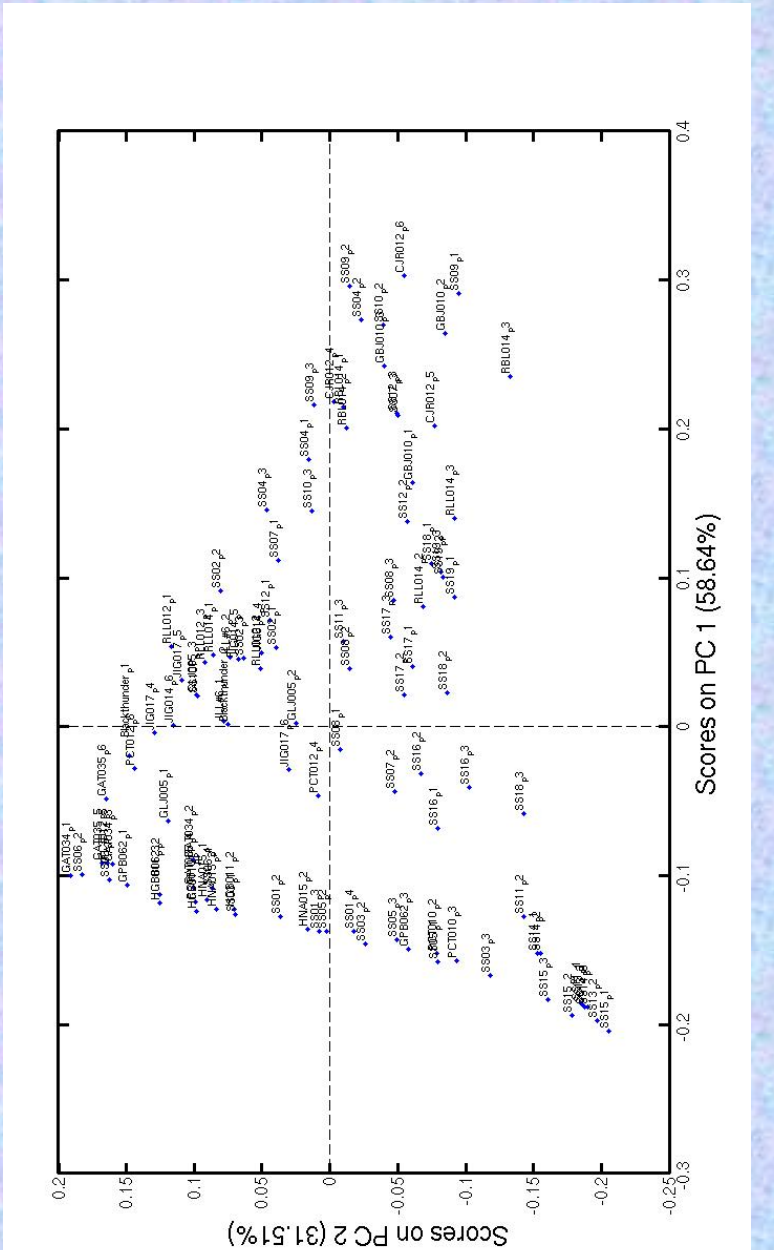



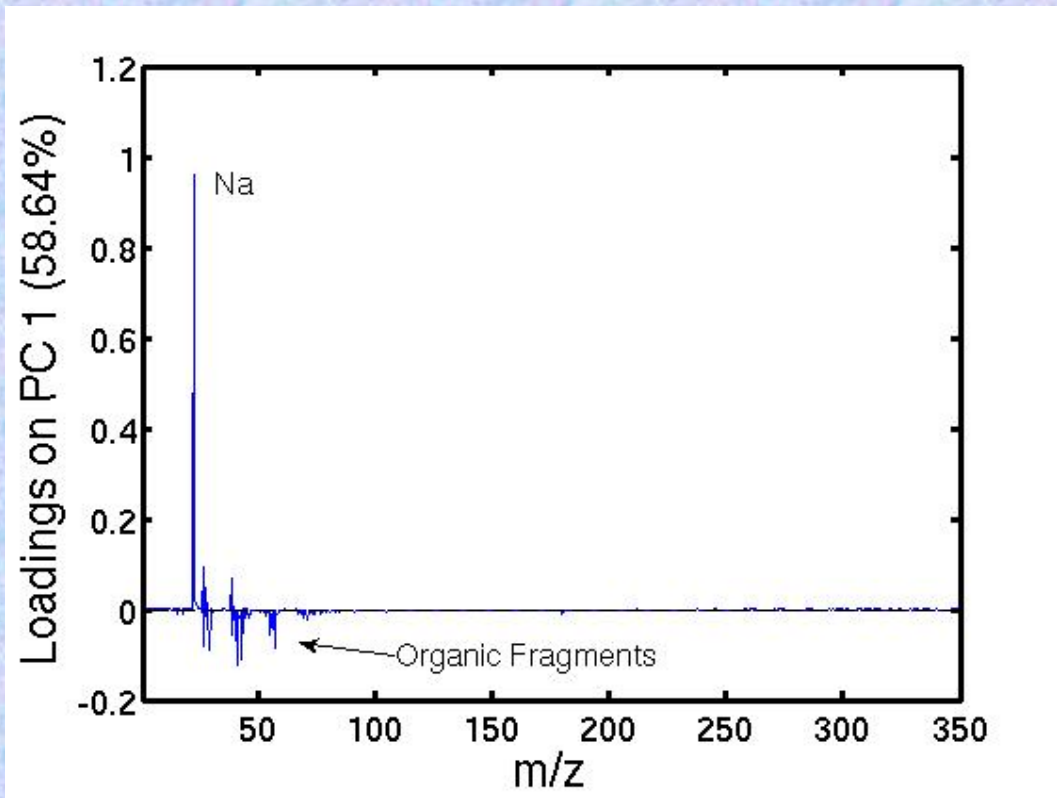




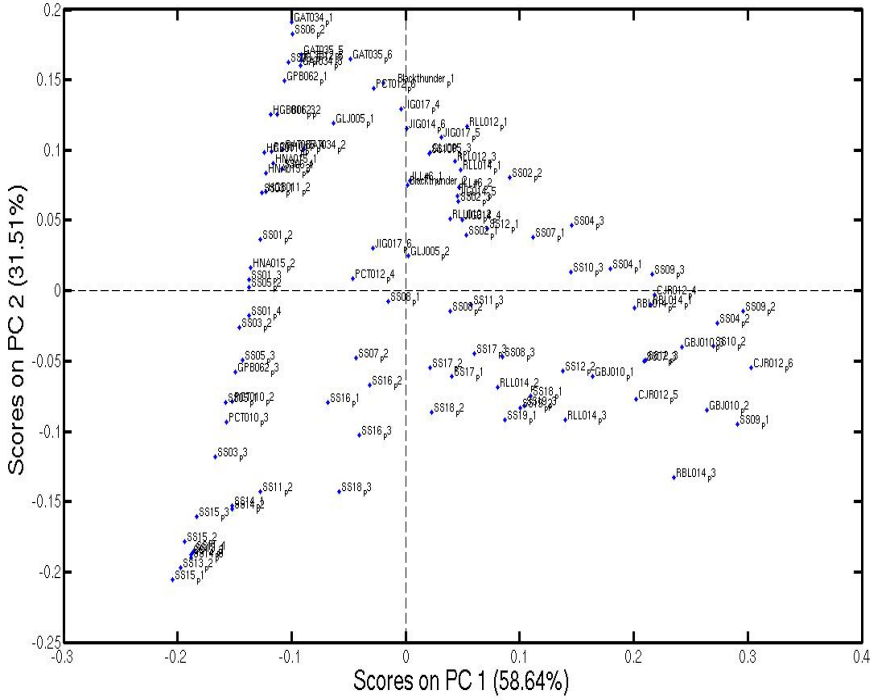


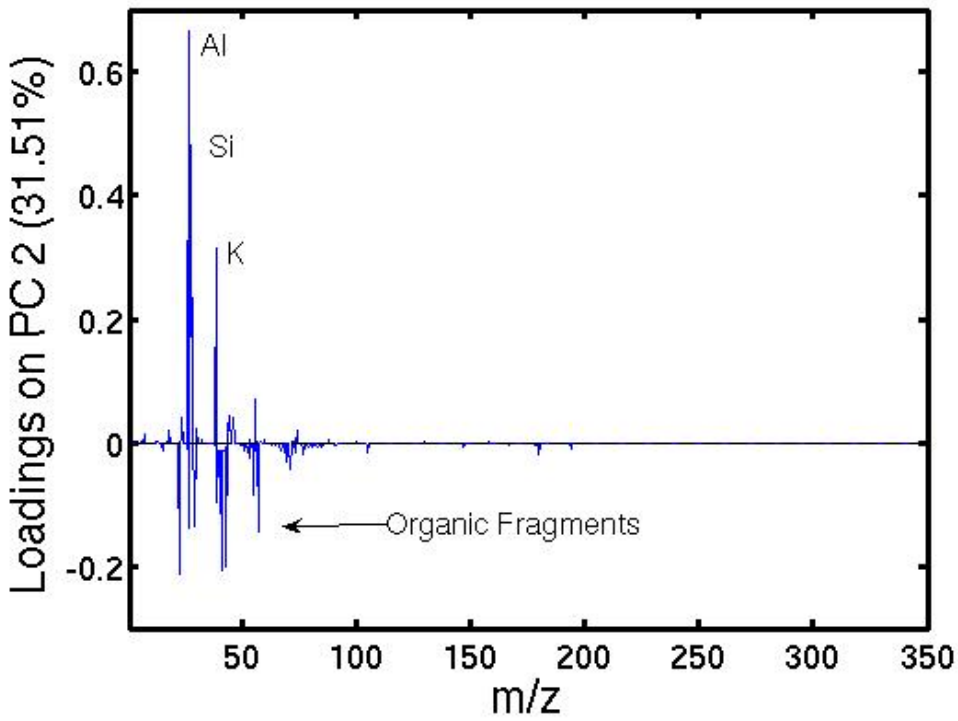
# Loadings on PC1



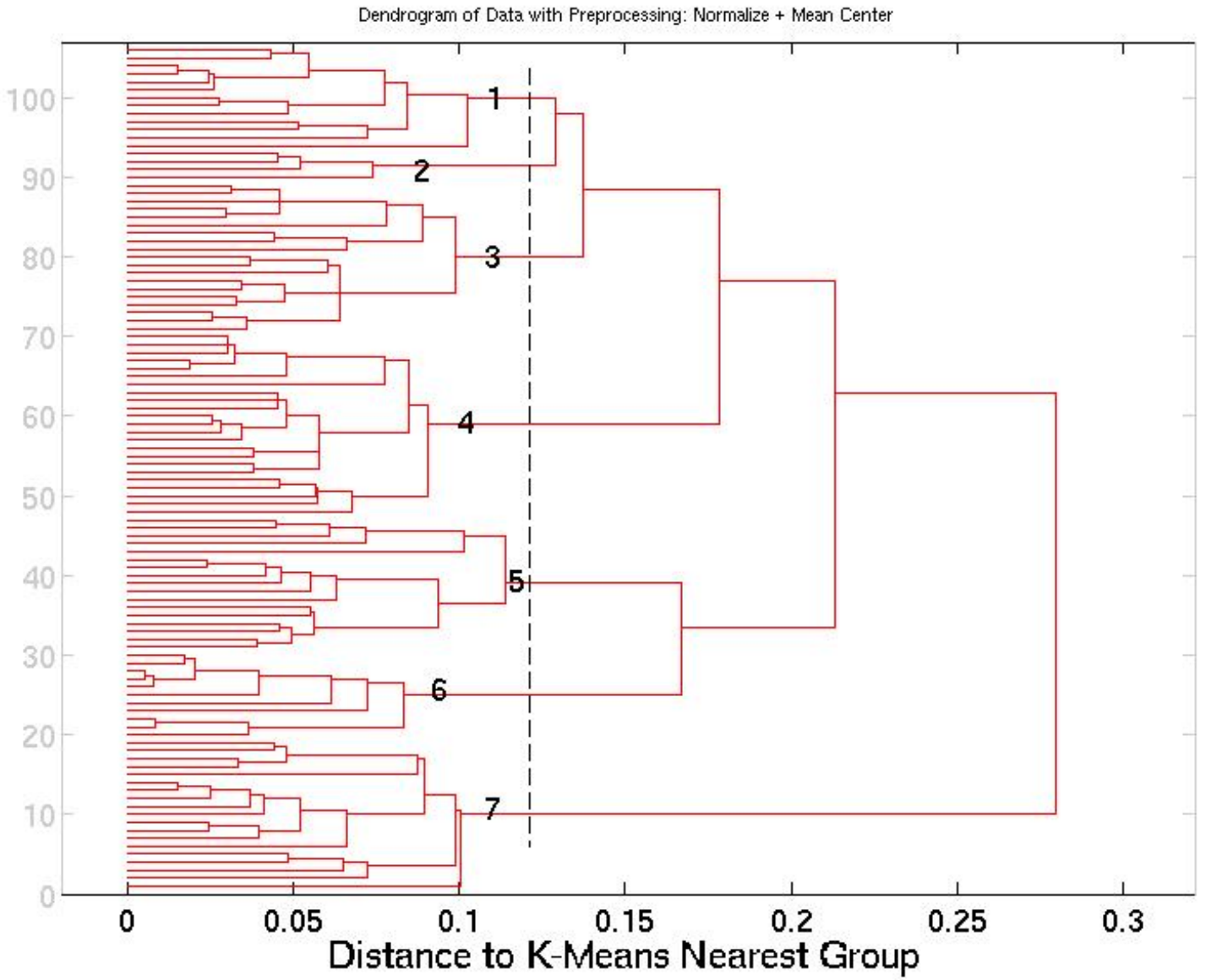



# Loadings on PC2

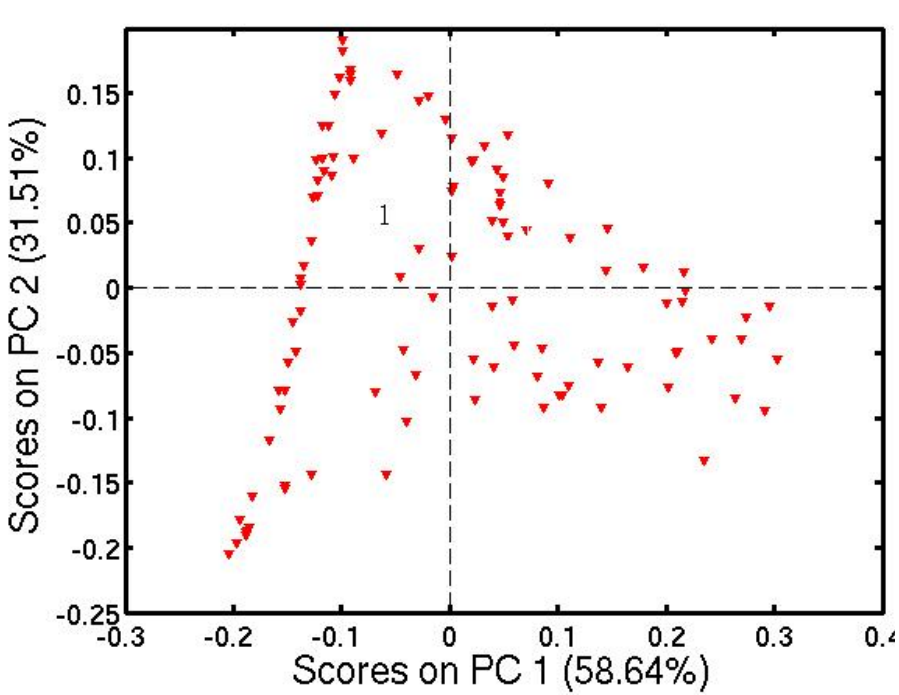


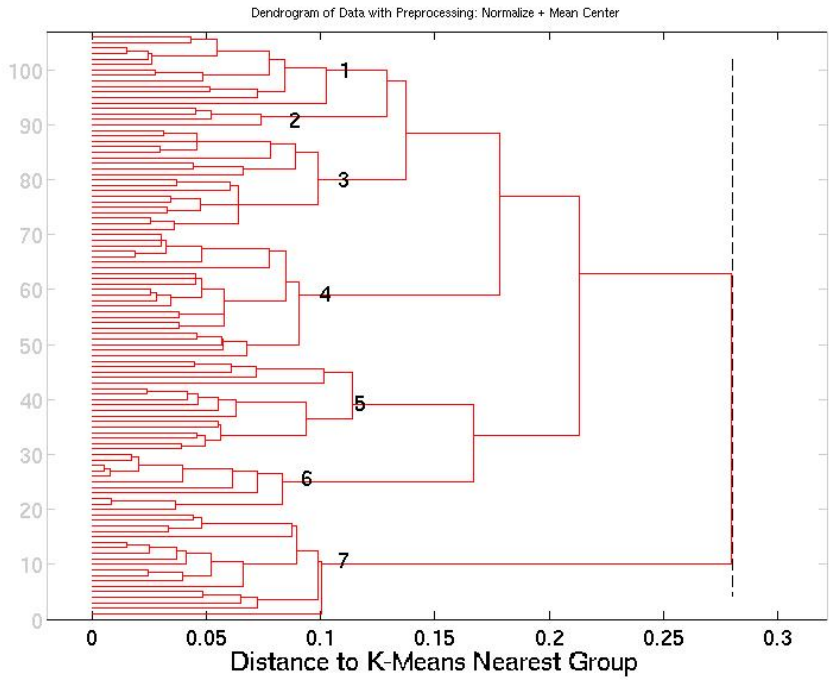



# Dendrogram from a Cluster Analysis

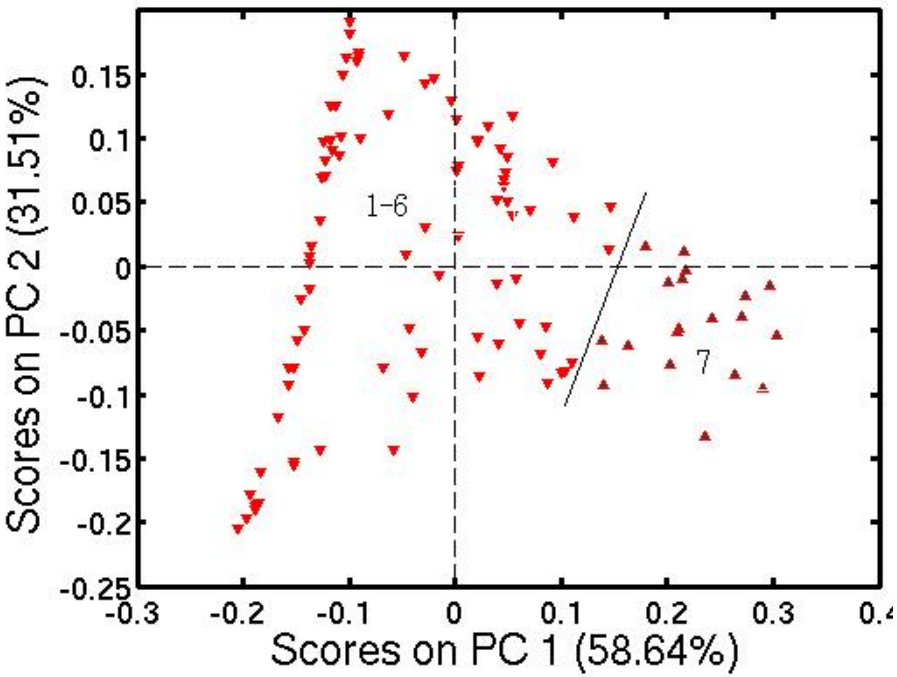


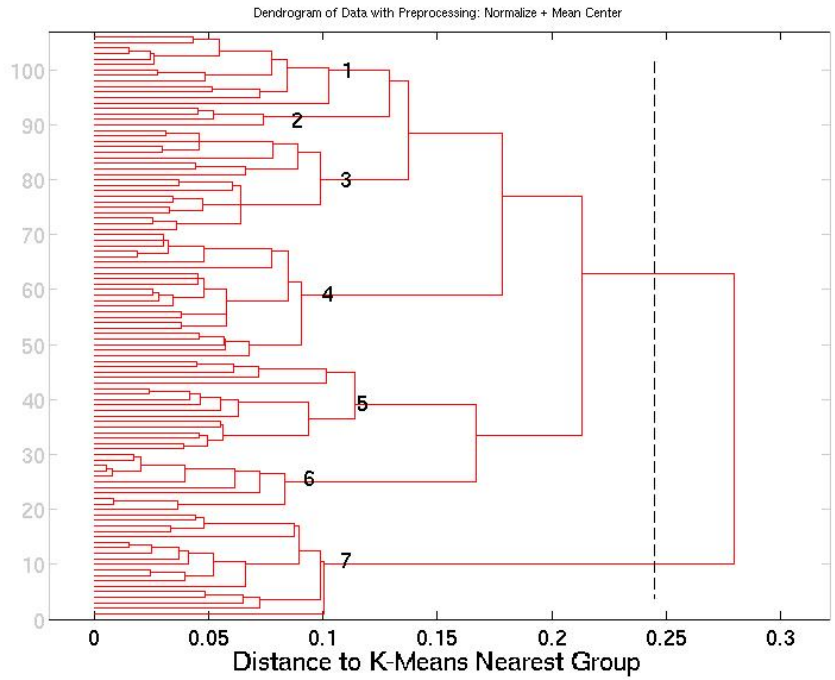
# Integrating the PCA and Cluster Analyses: One Cluster



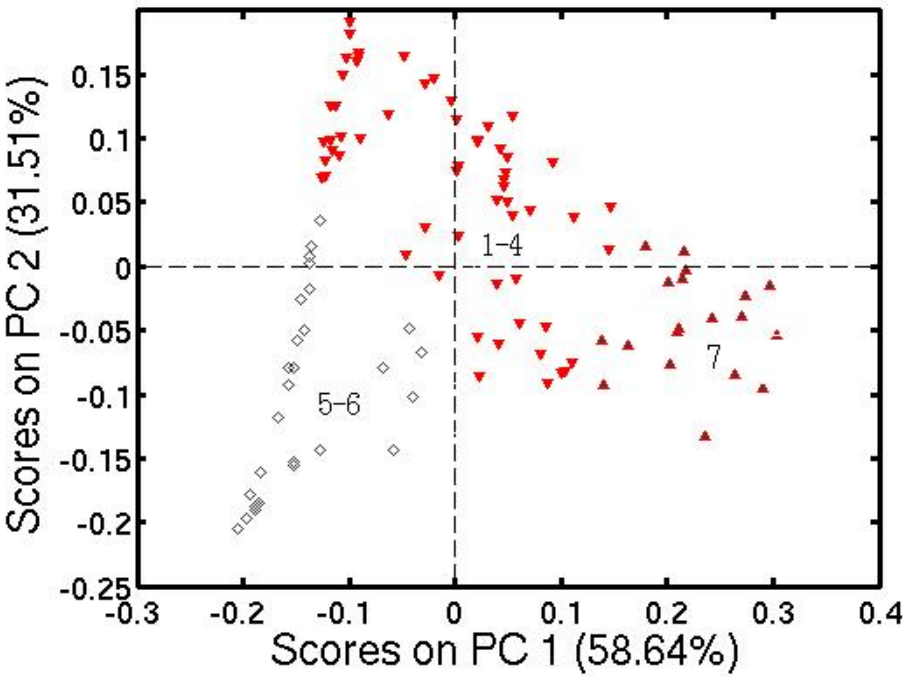



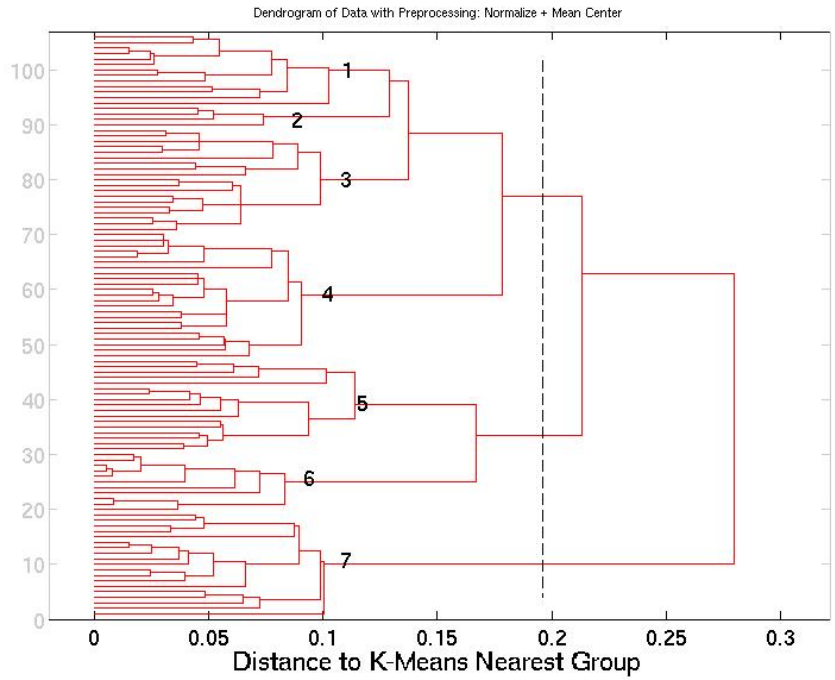
# Integrating the PCA and Cluster Analyses: Two Clusters



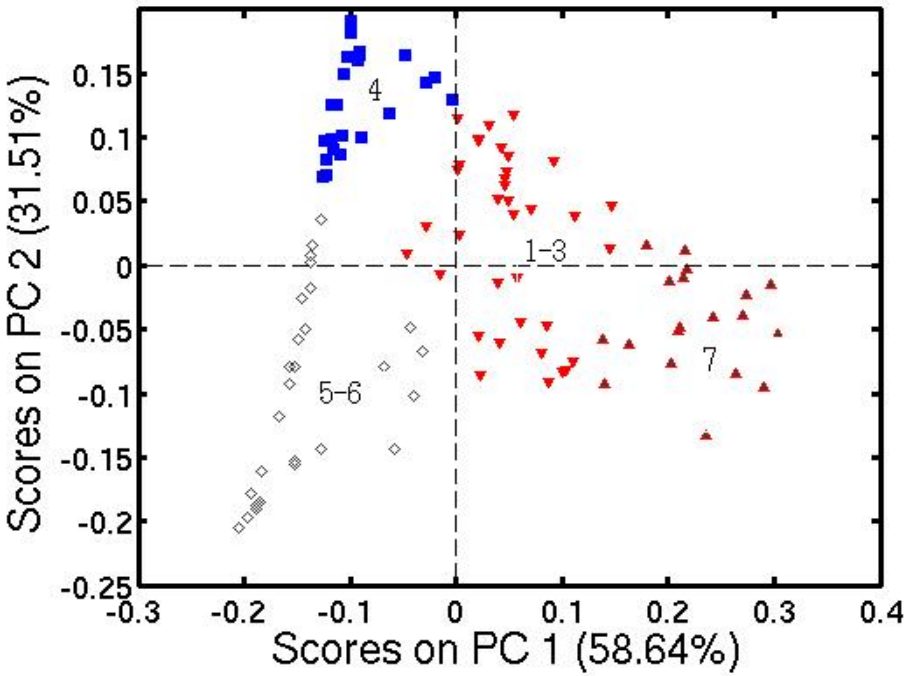



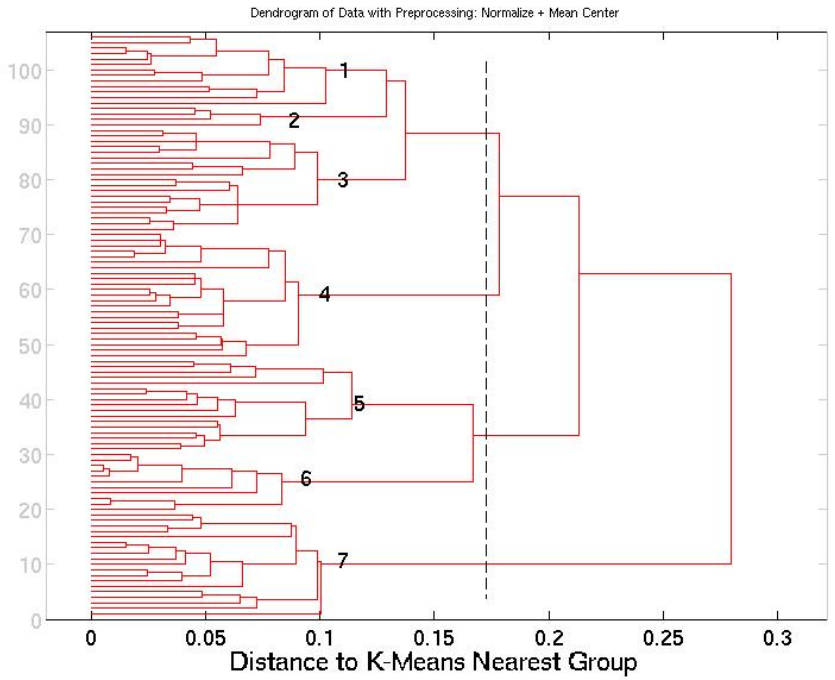
# Integrating the PCA and Cluster Analyses: Three Clusters



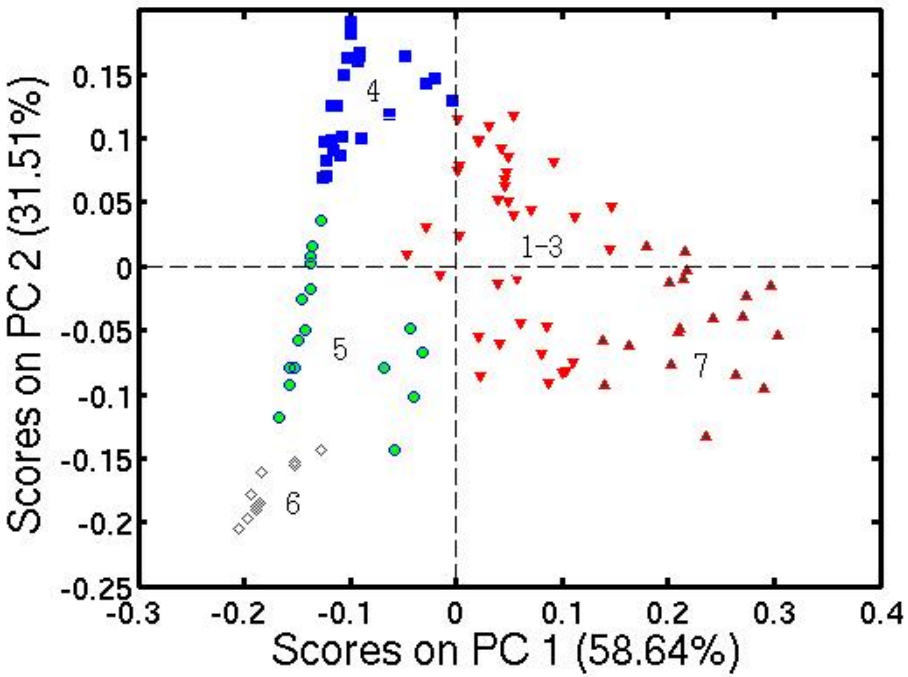



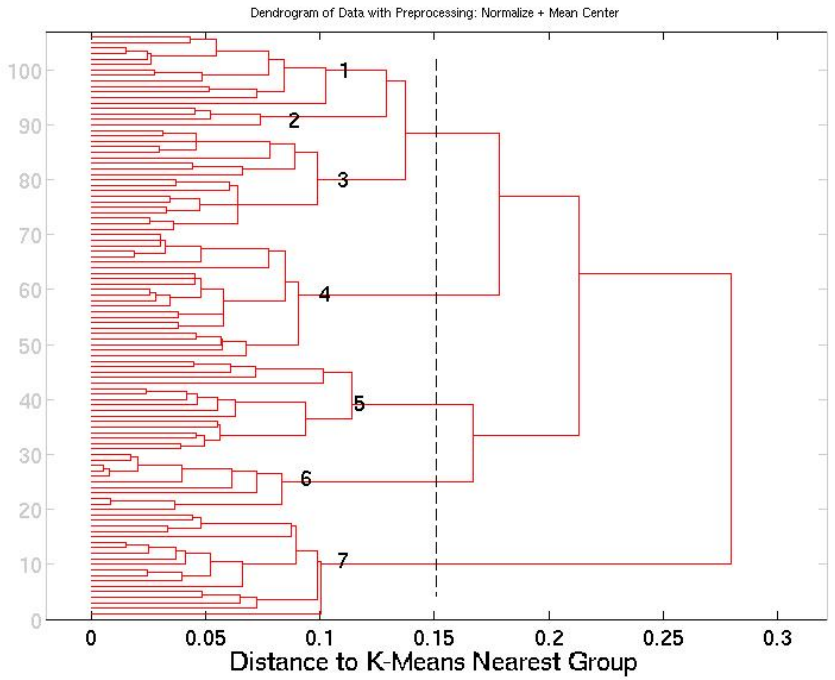
# Integrating the PCA and Cluster Analyses: Four Clusters



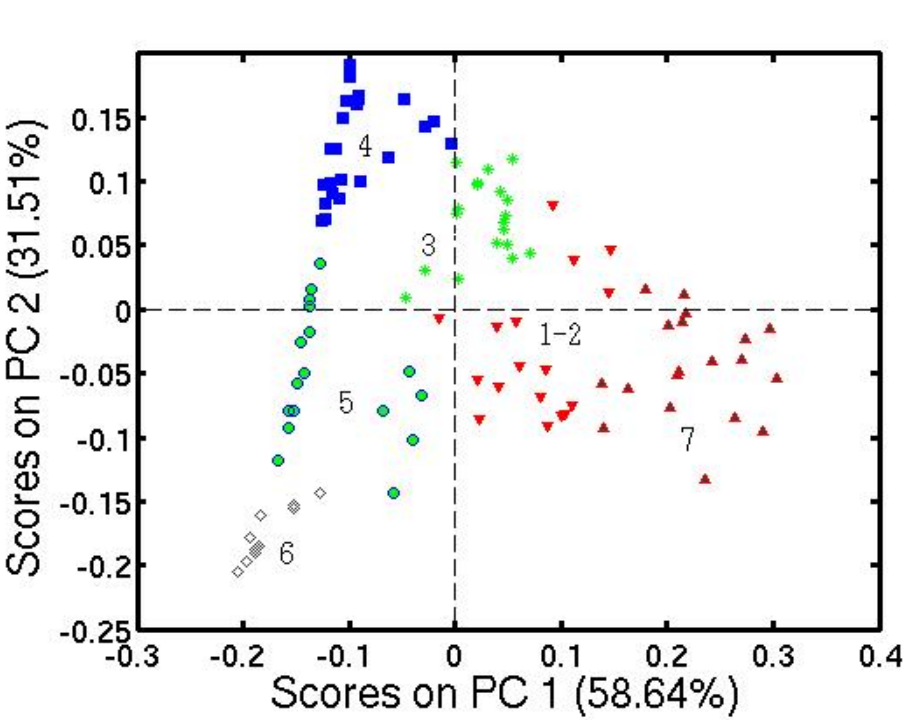



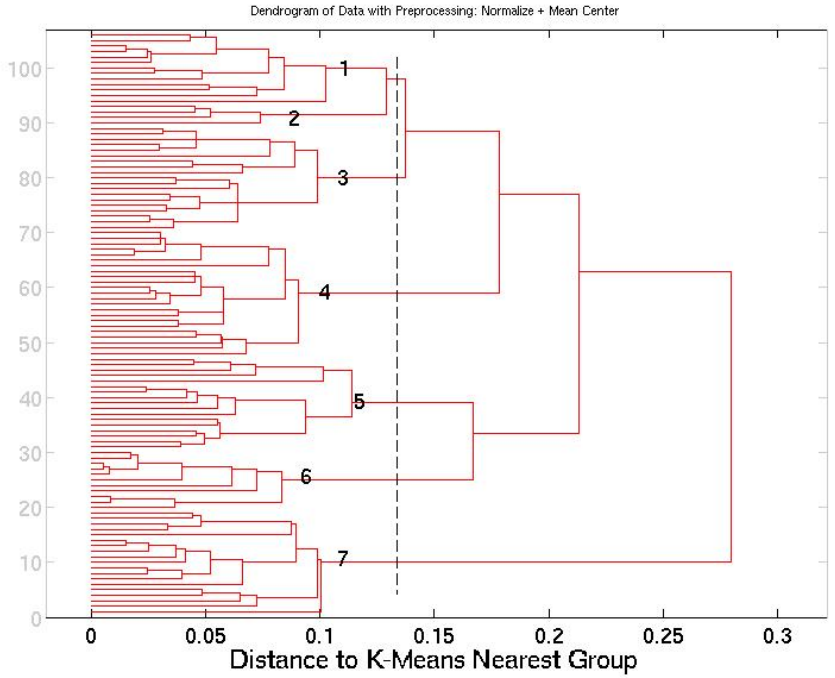
# Integrating the PCA and Cluster Analyses: Five Clusters



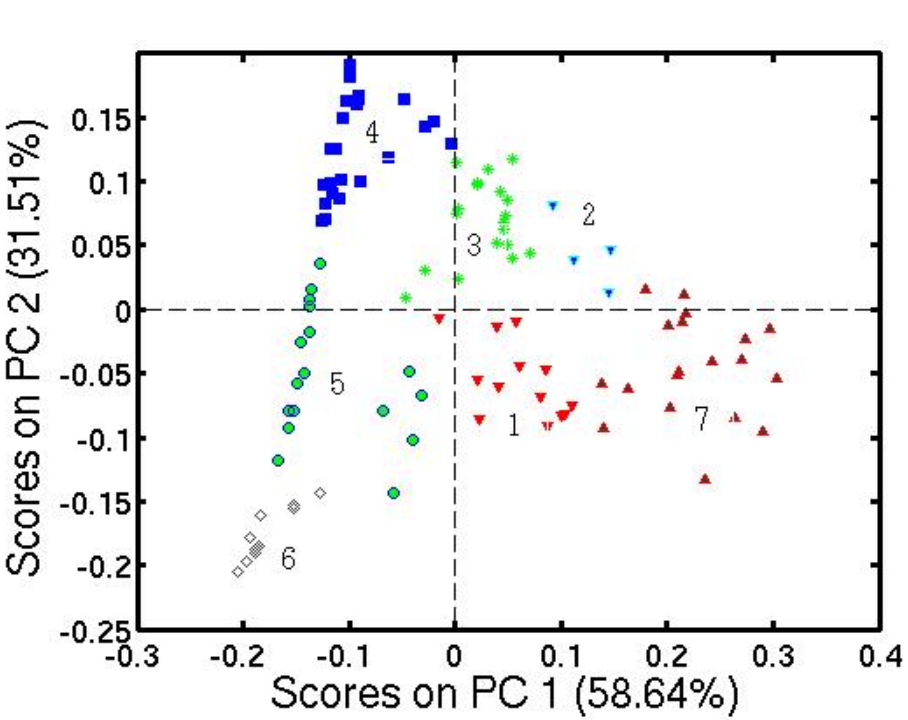



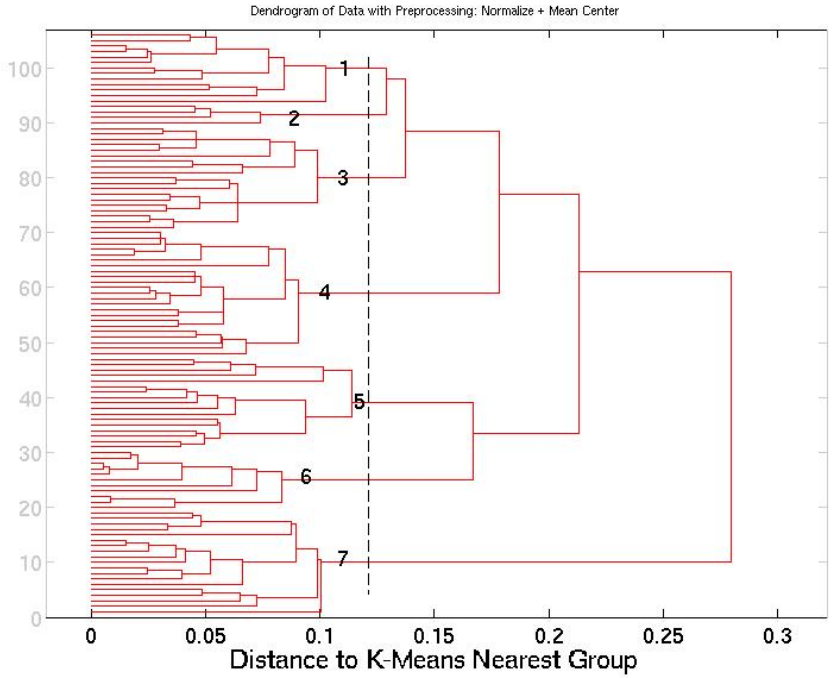
# Integrating the PCA and Cluster Analyses: Six Clusters



# Integrating the PCA and Cluster Analyses: Seven Clusters

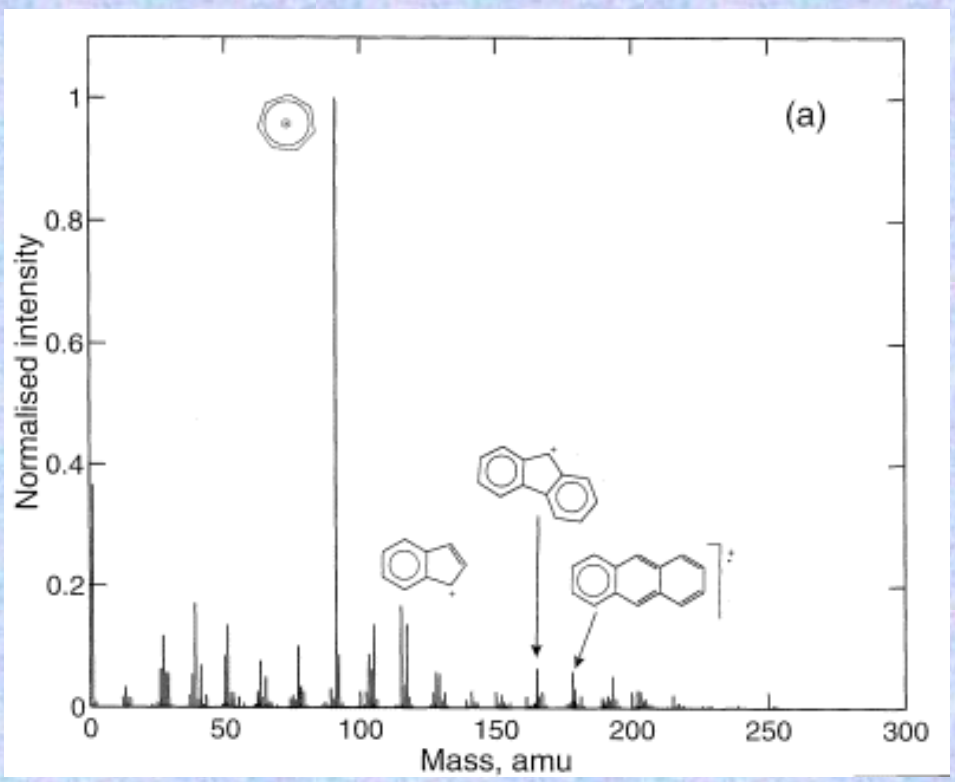


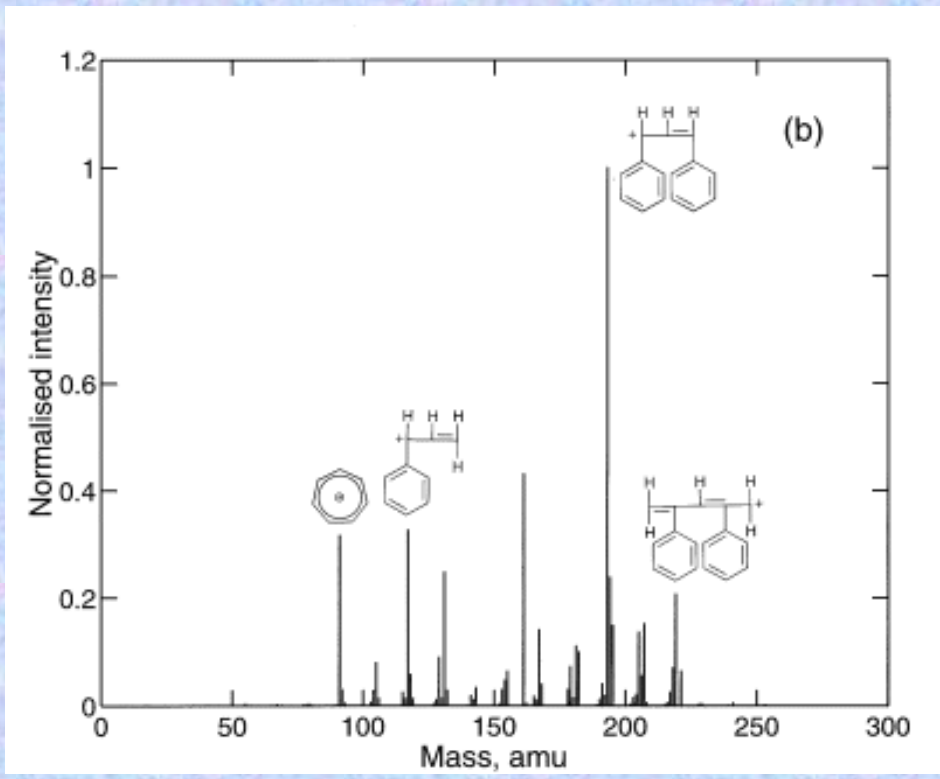



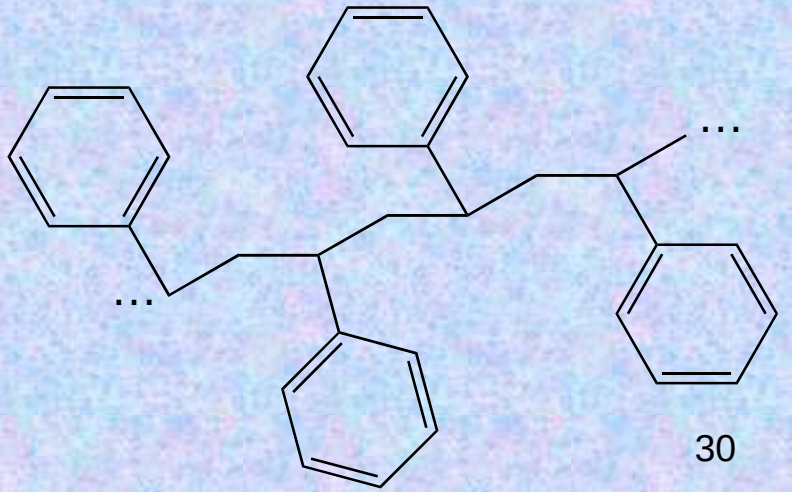
# Future Work: New Sources

- New sources, such as gold and bismuth clusters, or $C_{60}$ ions can produce significantly more high mass, organic ions than monatomic sources, such as $Ga^+$.
- High mass ions are highly characteristic of the sample they come from.
- These ion sources would allow some speciation of the coal.
- They should also increase the number of organic, compared to inorganic, ions that are produced by ToF-SIMS.

# Future Work: G-SIMS



Example: G-SIMS of polystyrene
Problem: We need another source

# Conclusions

- Good spectra can be obtained when coal is analyzed by ToF-SIMS.
- The spectra contain both an organic and an inorganic component.
- Principal components analysis and cluster analysis group the spectra in a similar way.
- There seems to be some correlation between the ToF-SIMS spectra and the physical/chemical properties of the coal samples.
- Other sources would allow us to obtain more high-mass ions, which are highly characteristic of the materials they come from.
- Other sources would also allow us to do G-SIMS.

# Thank You