The Risks of Promiscuous MVA of ToF-SIMS Data: The MVA Quickie

NESAC/BIO

Daniel J. Graham

University of Washington
Department of Bioengineering

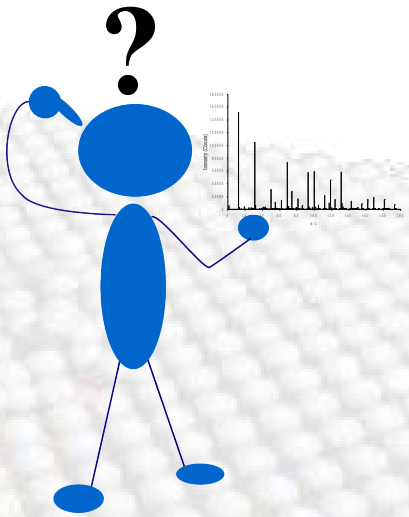
UNIVERSITY OF WASHINGTON

NESAC/BIO

SIMS Europe Sept 19-21, 2010
Münster Germany

ToF-SIMS Is Complicated

- Spectra contain hundreds of peaks
- Peak intensities can be interrelated
- Matrix effects can cause non-linear changes in peak intensities
  - Due to sample composition
  - Due to the presence of oxides
  - Due to the presence of salts
- Peak intensities may or may not correlate with surface composition
- Peak intensities may vary due to differential sputtering
- Often heavy fragmentation and lack of molecular ion signals (assuming you know what signal to expect)



ToF-SIMS Is Complicated

- **No one surface analysis method can provide a complete surface characterization alone**
- **ToF-SIMS in particular benefits from data from other methods**
- **The more complicated the surface chemistry, the more important this becomes**

AFM

SFG

XPS

Now it's starting to make sense

IR

NEXAFS

SIMS

Proper ToF-SIMS Analysis Requires:

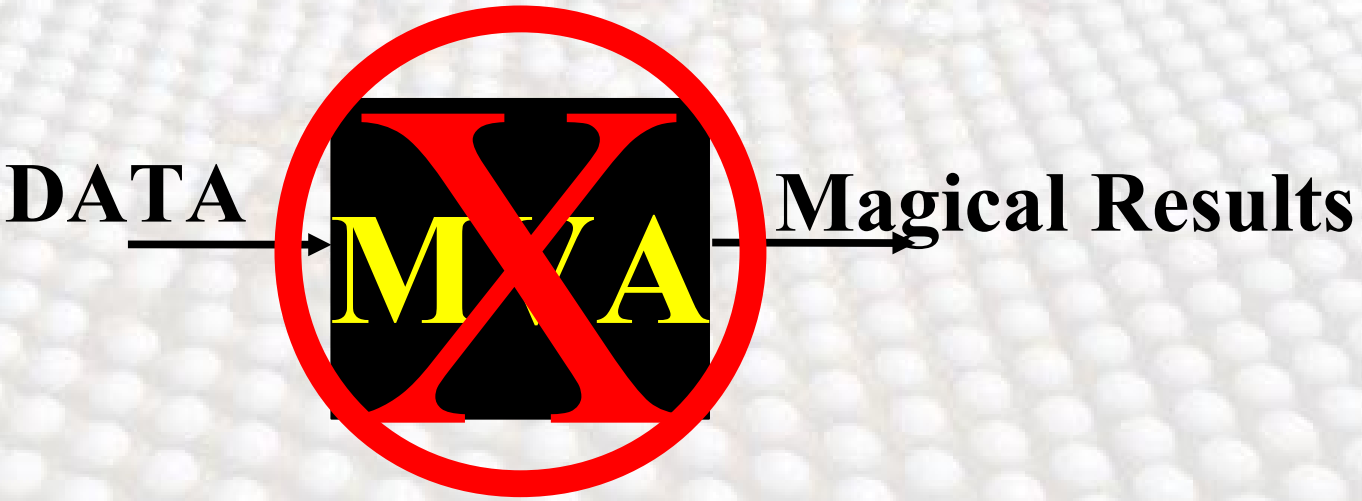
- Good experimental plans (controls)
- Proper sample preparation
- Careful data collection
- Consistent data calibration
- Sound understanding of the fundamentals of mass spectral analysis
- Knowledge of how to properly use the available tools to help with the analysis

MVA "To The Rescue..."

- MVA is becoming increasingly popular for ToF-SIMS spectra and images
- MVA can aid in reducing the complexity of a data set with regards to the magnitude of the data one needs to focus on (what is changing, where is it different, what peaks are changing)
- MVA cannot:
  - Remove all effects of contaminants or matrix effects
  - Fix a poorly designed experiment
  - Interpret your data
  - Find things that are not there

Avoid the MVA Quickie

- MVA should not be treated like a black box
- MVA should not be something you do as a last minute decision, it should be part of the experimental design from the beginning
- If someone says, “Lets see what we get when we use SIMS and MVA on my samples”...RUN...Or at least stop to think if it really makes sense

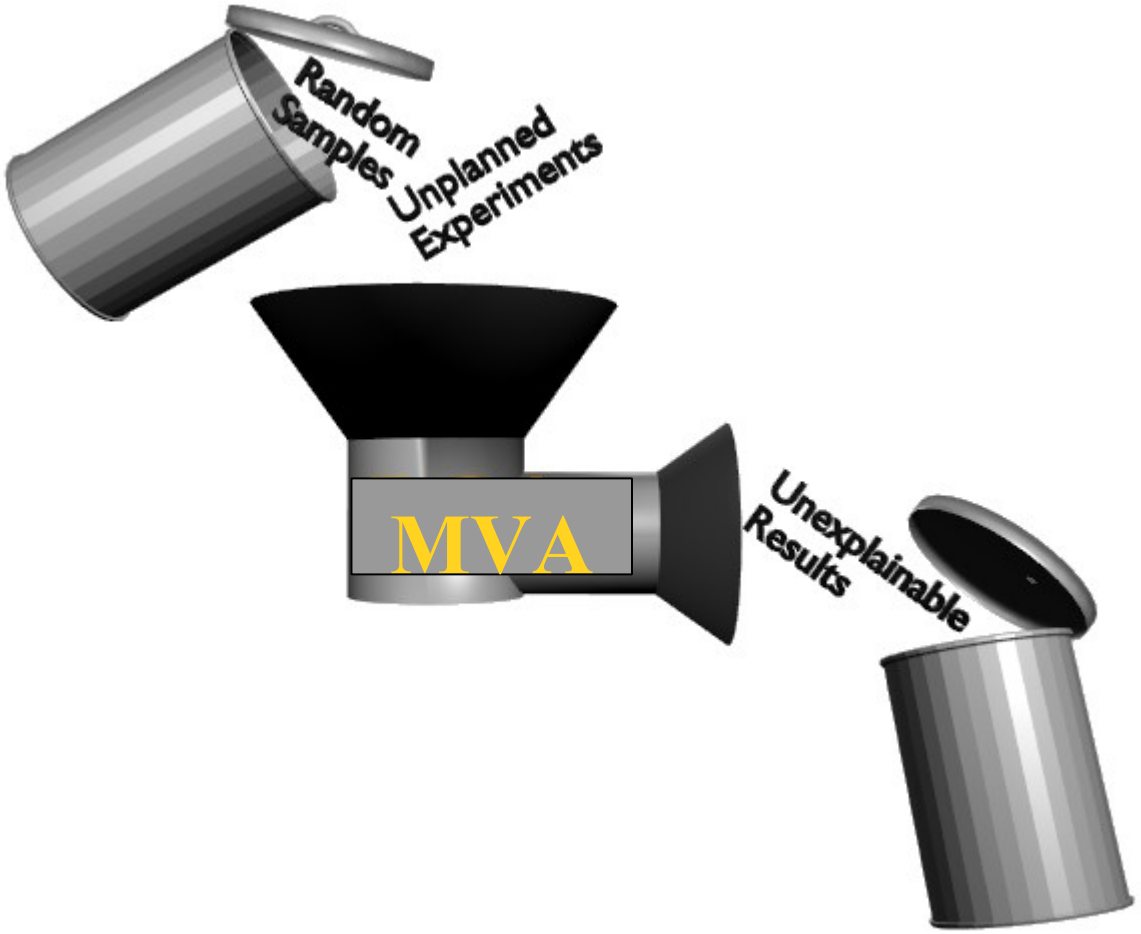


Plan

- What is the question you want to answer?
- What samples do you need to answer that question?
- How many samples/ replicates do you need?

Remember MVA will find the main differences between any samples

If you input garbage in



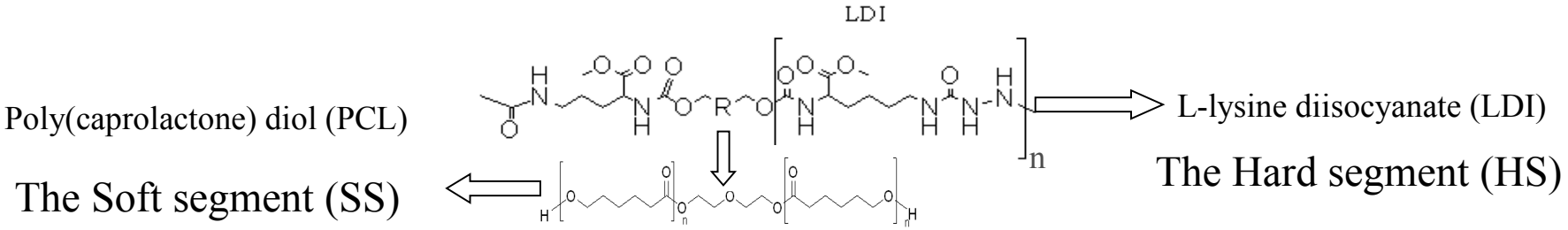
You will get garbage out!!!

When Should MVA be Used?

- **MVA should be used to help answer questions**
  - **Are surfaces A and B different?**
  - **How does treatment X change the surface chemistry?**
  - **How is fragmentation pattern affected by ____?**
  - **Can TOF-SIMS data distinguish Protein A from Protein B?**
  - **What regions in my image are different and why?**
- **The question should be part of the experimental design and not an afterthought**

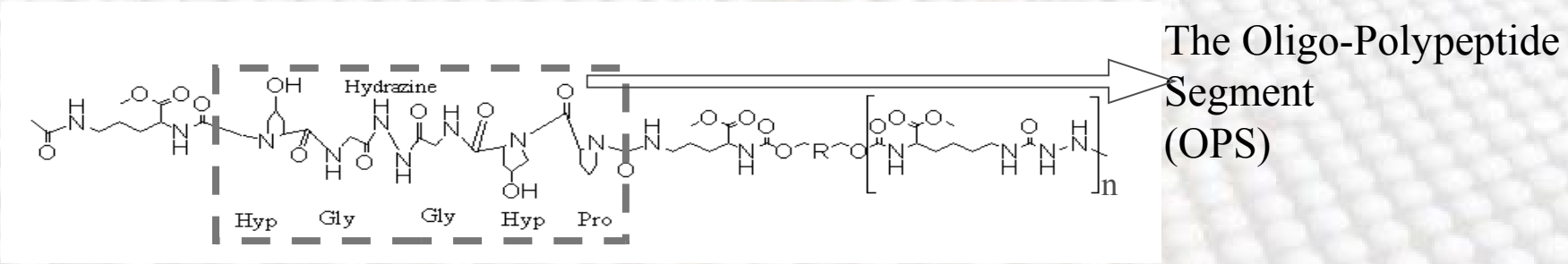
Structures of the polyurethane units

Control sample



PUU 817, LDI/PCL/Hyd =8/1/7

Structure of the Poly(peptide-urethaneurea)

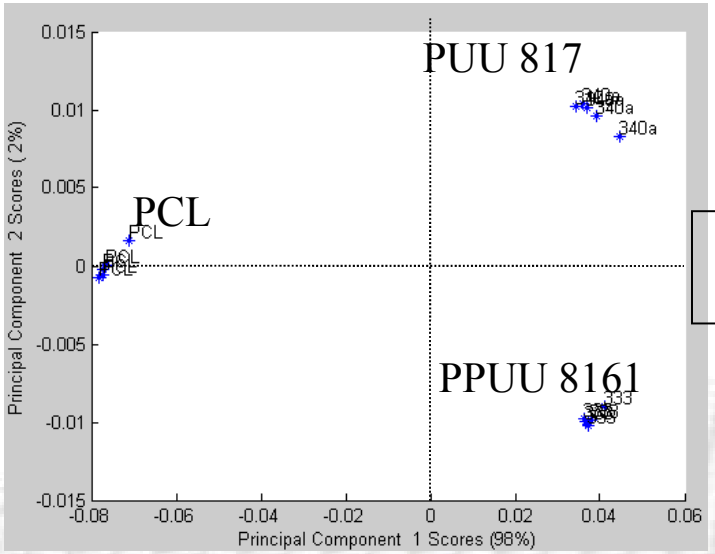


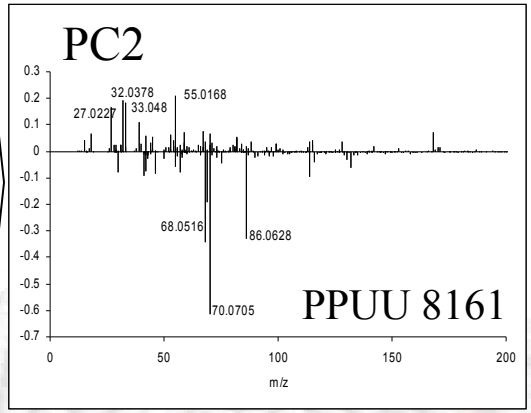
PPUU 8161, LDI/PCL/Hyd/(Gly/Hyp/Pro/Hyd) =8/1/6/1

*Slide and data from Gilad Zorn Ph.D., NESAC/BIO University of Washington

Characteristic peaks

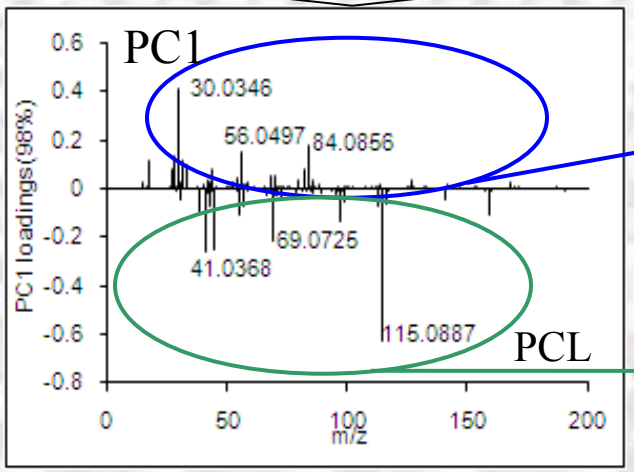
Principal Component Analysis (PCA) of ToF-SIMS data





OPS peaks:

C_4H_6N	68.052	Proline
C_4H_8N	70.071	Proline
C_4H_8NO	86.063	Hydroxy-proline



LDI-Hyd (HS) *

NH_4	18.036
CH_2N	28.019
CH_4N	30.035
N_2H_4	32.038
N_2H_5	33.048
$C_5H_{10}N$	84.086

PCL (SS) **

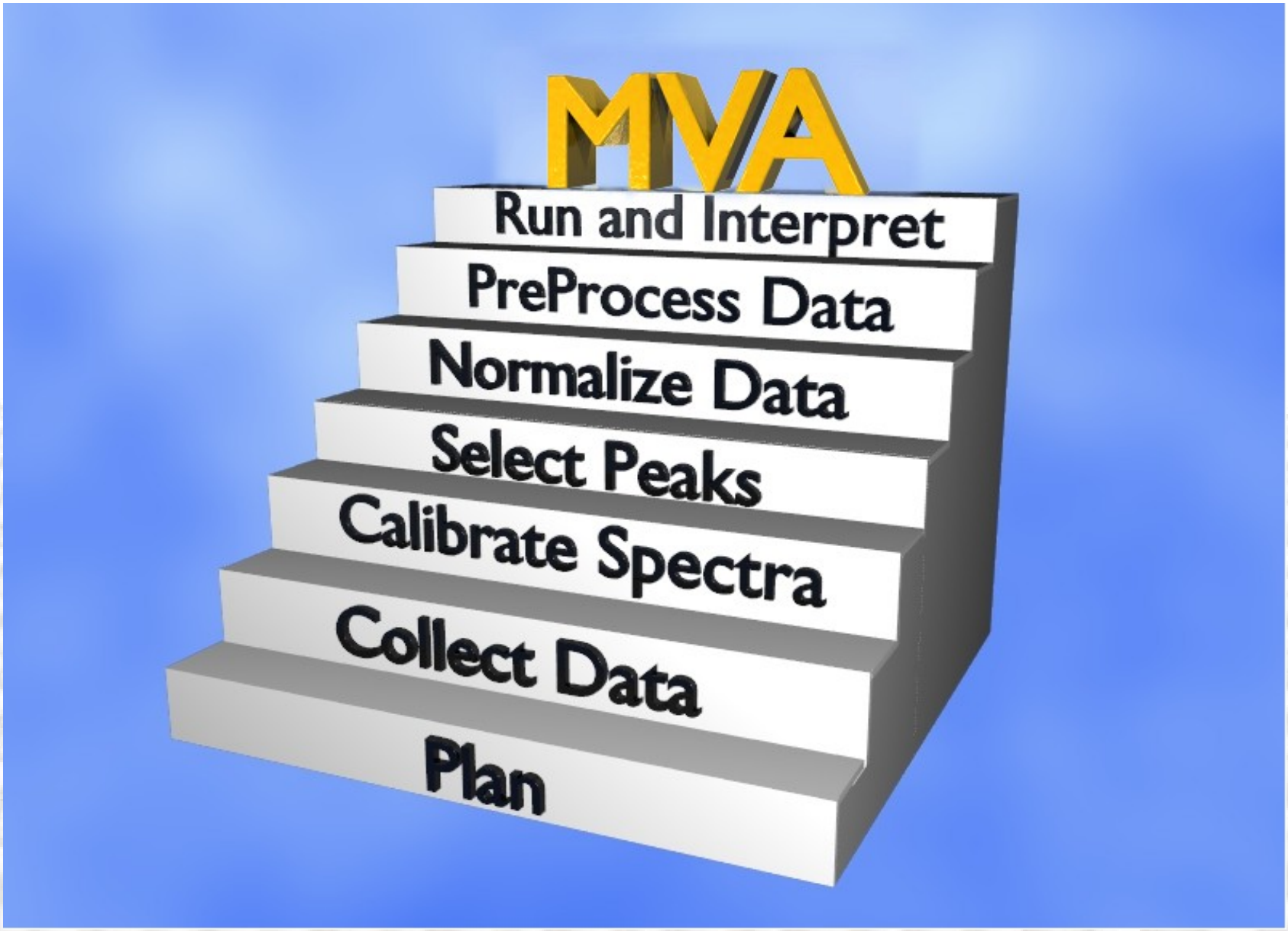
C_3H_3	39.023
C_3H_5	41.037
C_3H_3O	55.017
C_5H_9	69.073
C_6H_9O	97.068
$C_6H_{11}O_2$	115.089

* Comparable to ToF-SIMS studies of poly(L-lysine)

* * Similar to published ToF-SIMS data of PCL.

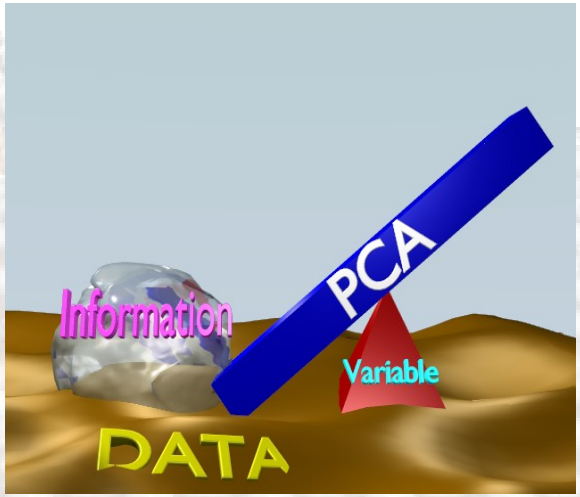
*Slide and data from Gilad Zorn Ph.D., NESAC/BIO University of Washington

Steps to MVA

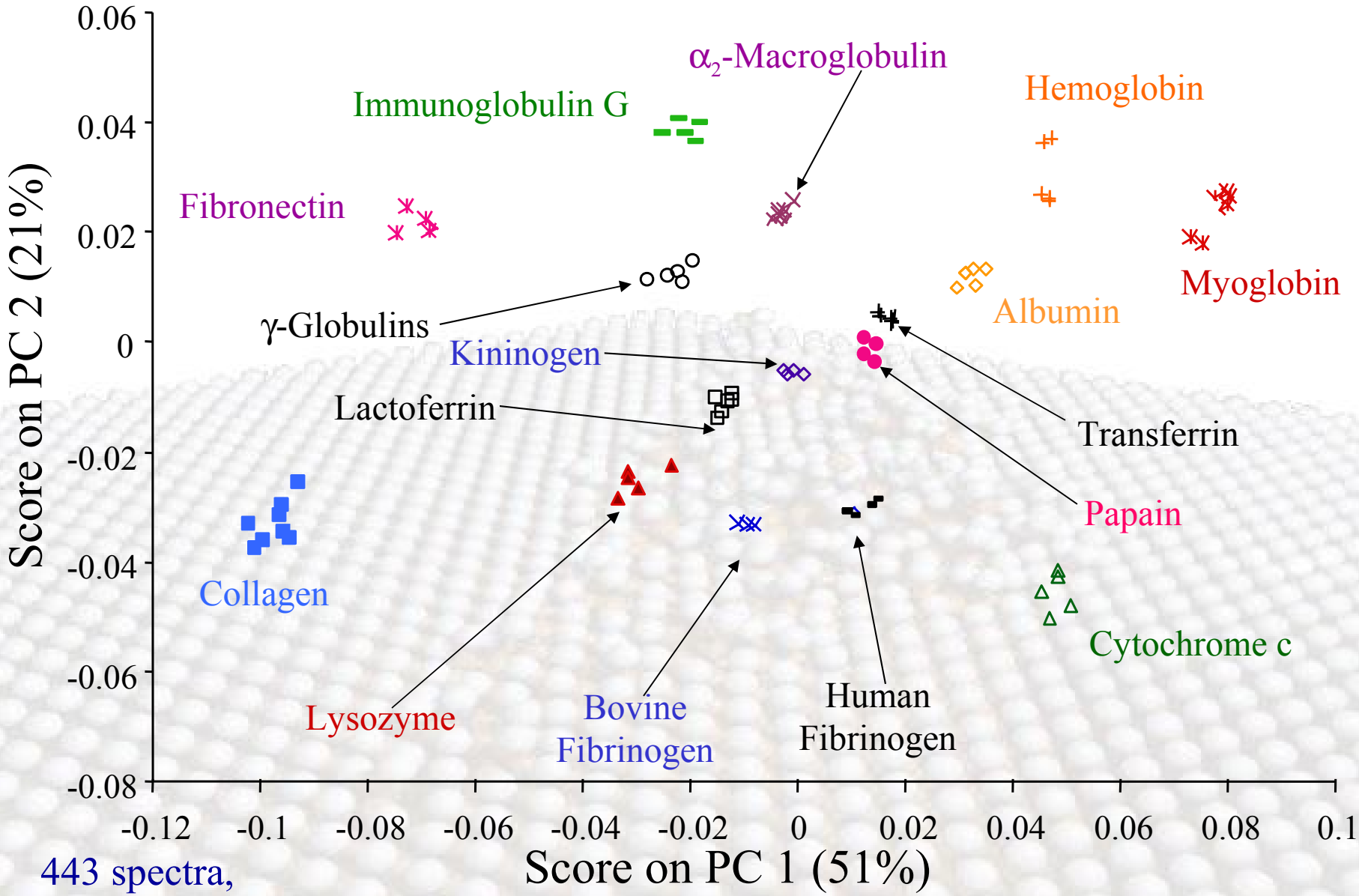

MVA
Run and Interpret
PreProcess Data
Normalize Data
Select Peaks
Calibrate Spectra
Collect Data
Plan

Experimental Design/Data Collection

- **Not all systems are well defined, but your experimental design can be:**
  - **Think about what you want to learn from SIMS**
  - **Simplify the number of variables you are dealing with per experiment**
  - **Plan appropriate controls**
  - **Run enough replicates to determine reproducibility**
    - **Homogeneous => 3 to 5 spots on 2 samples**
    - **Inhomogeneous => 5 to 7 spots on 3 to 5 samples**



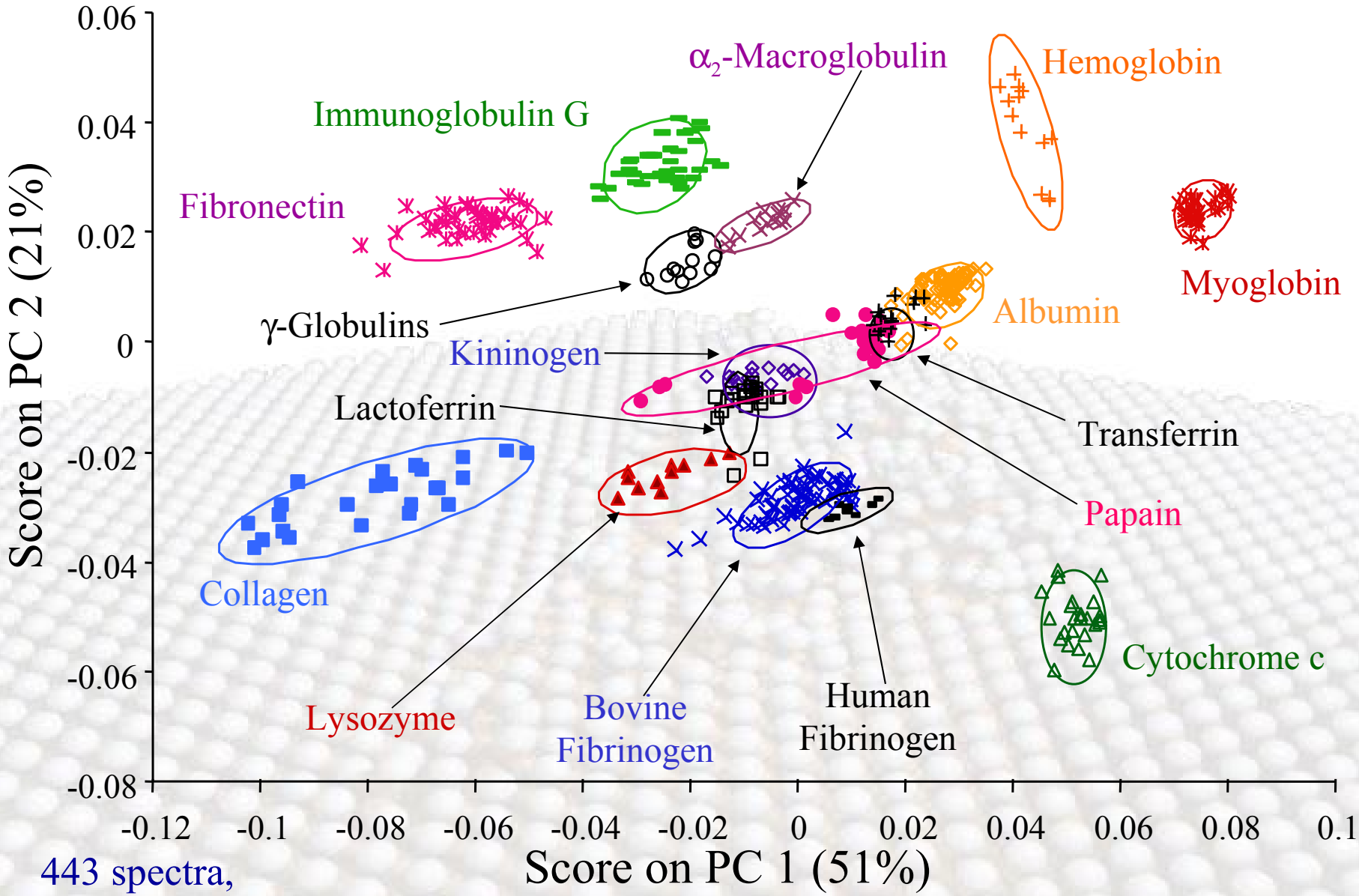
Proteins adsorbed onto Mica: PCA



443 spectra,
16 different proteins

Modified fromWagner & Castner, Langmuir **17** (2001) 4649.

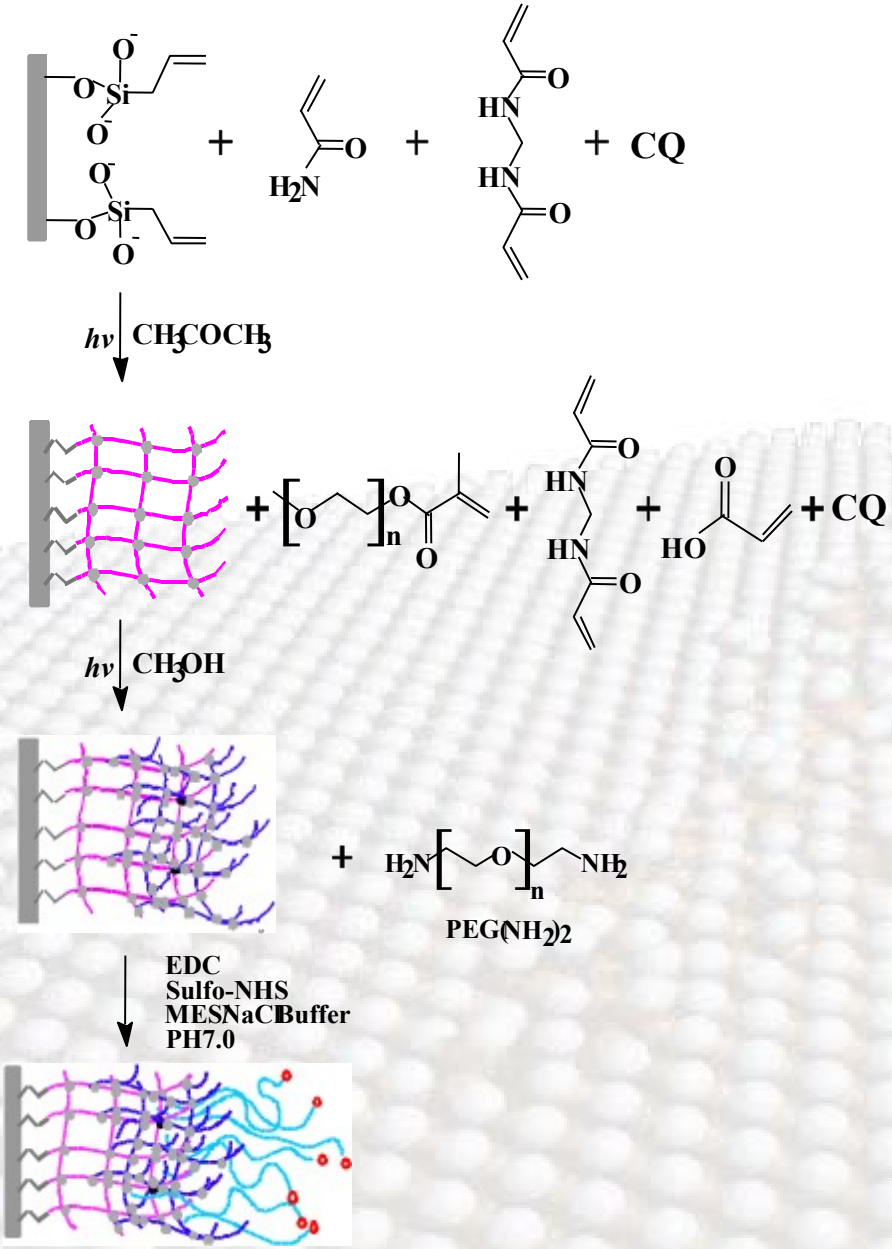
Proteins adsorbed onto Mica: PCA



443 spectra,
16 different proteins

Wagner & Castner, Langmuir **17** (2001) 4649.

Synthesis of an IPN of P(AAm-*co*-EG/AA)

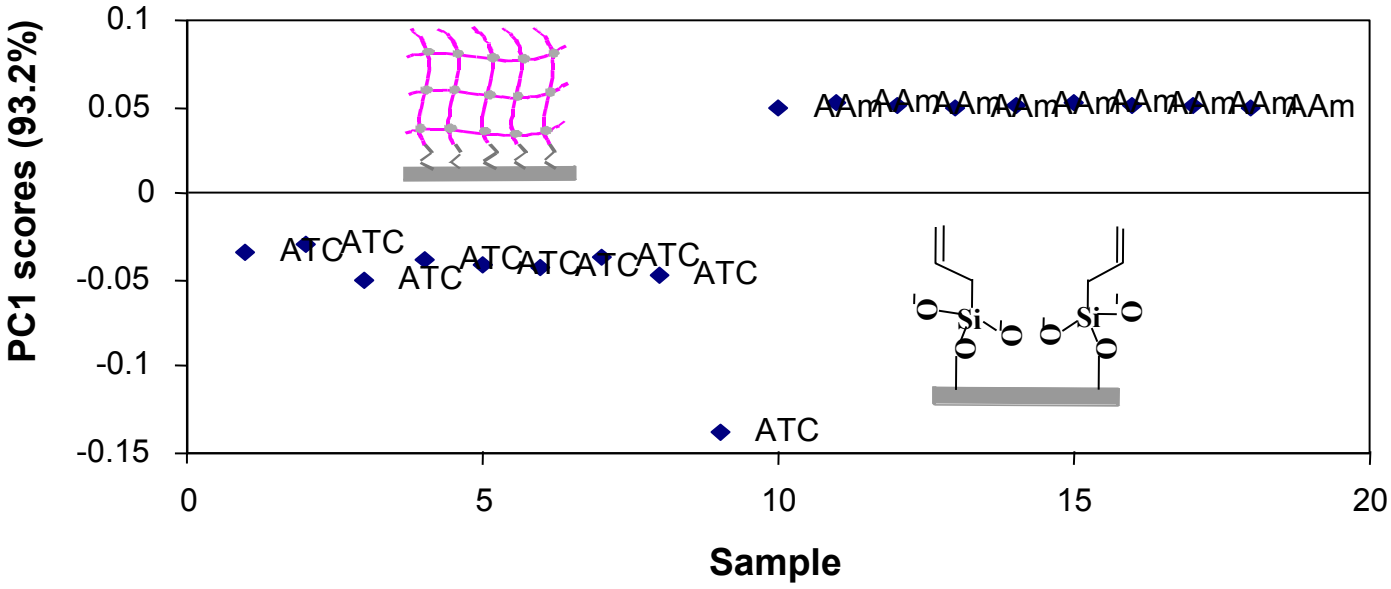


PEG spacer

IPN

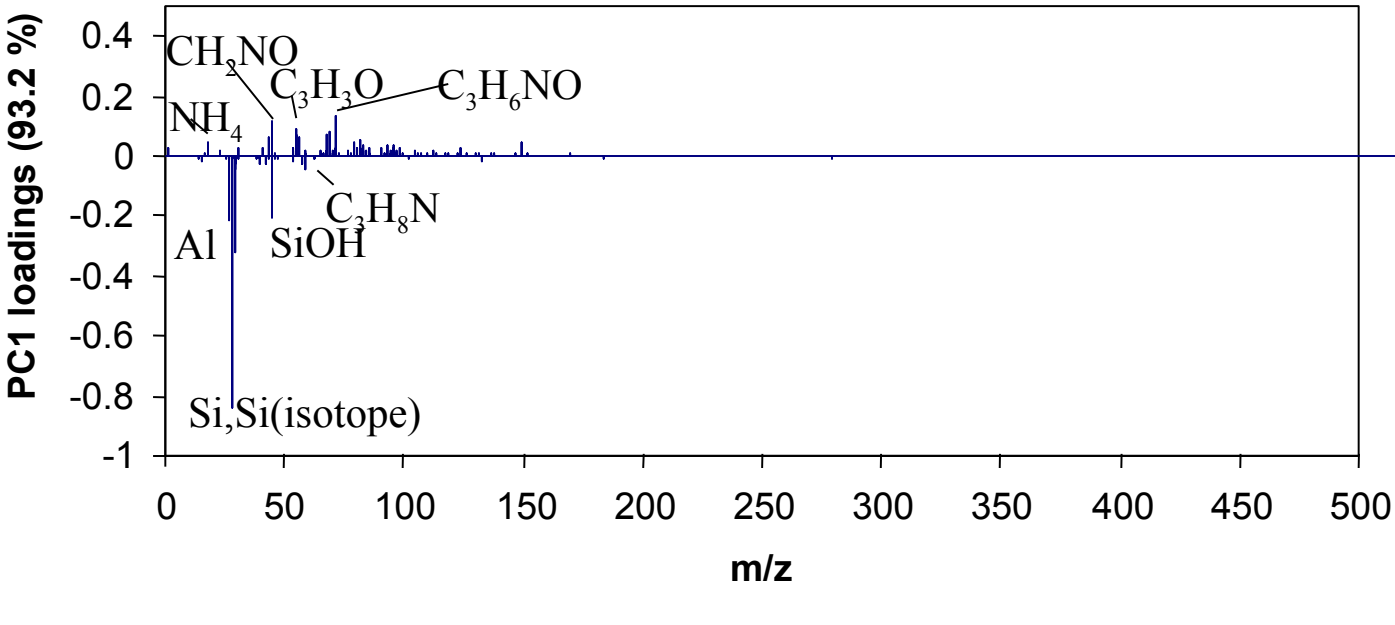
Substrate

PCA of ATC and AAm samples



-PCA clearly separates the two sample types

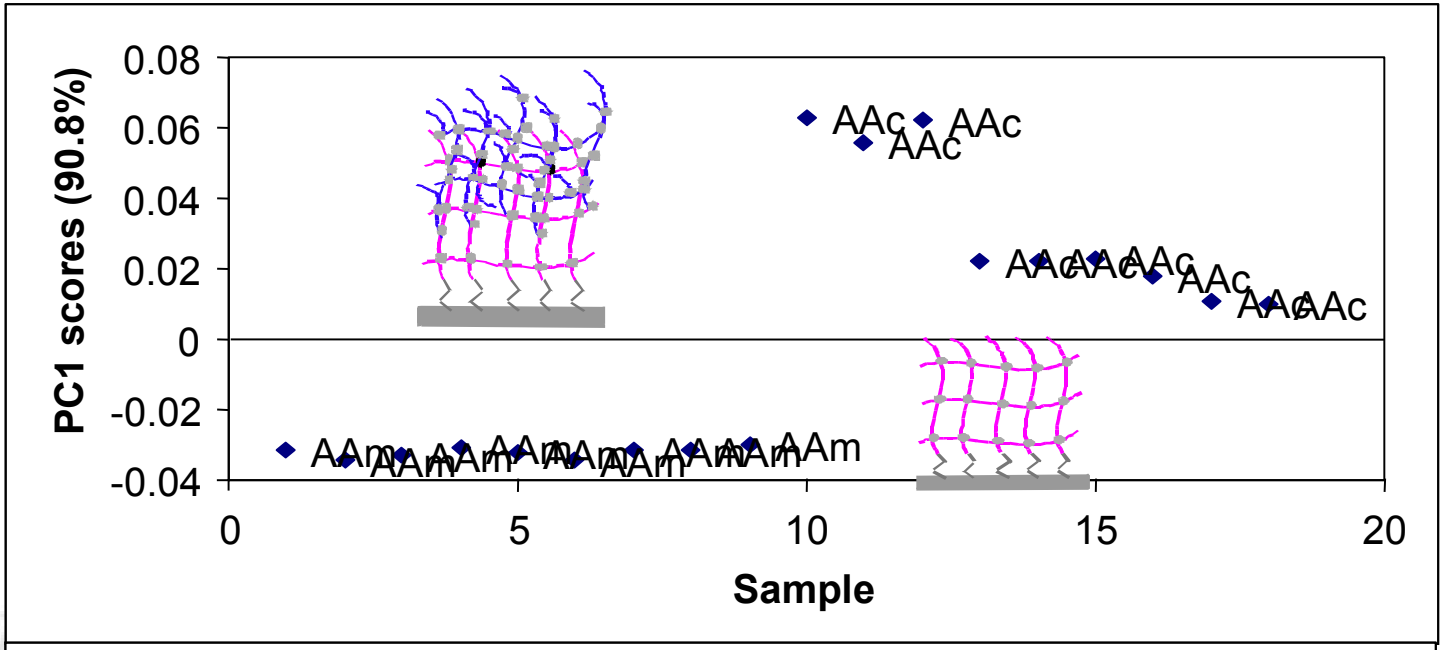
-low scatter in the data

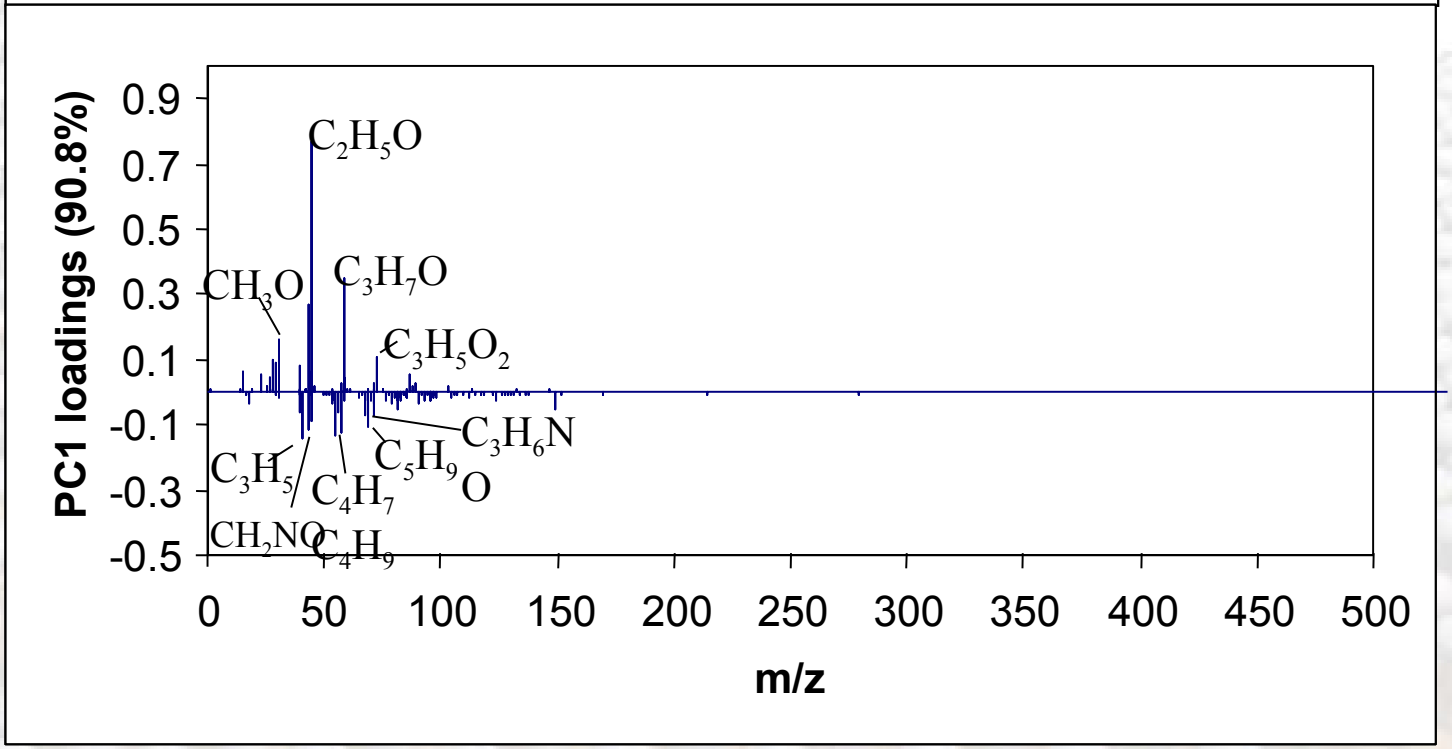


PCA of AAm and AAc samples

-PCA separates the two sample types

-scatter seen in the AAc data suggesting less uniform chemistry spot to spot.

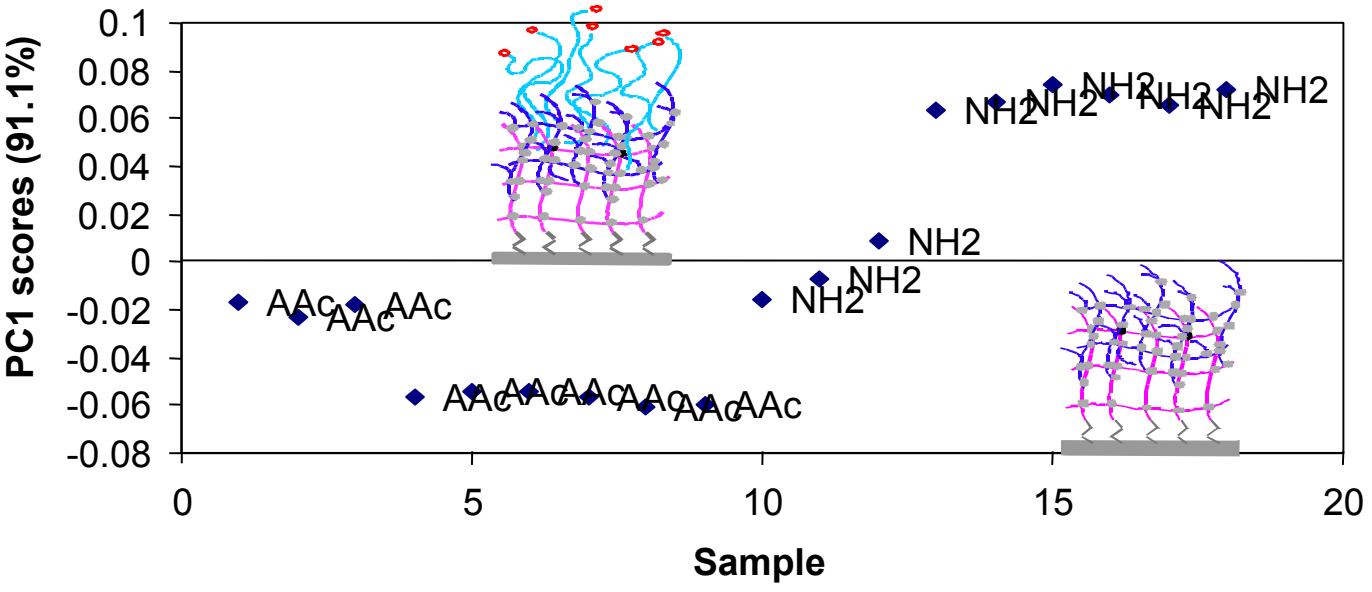


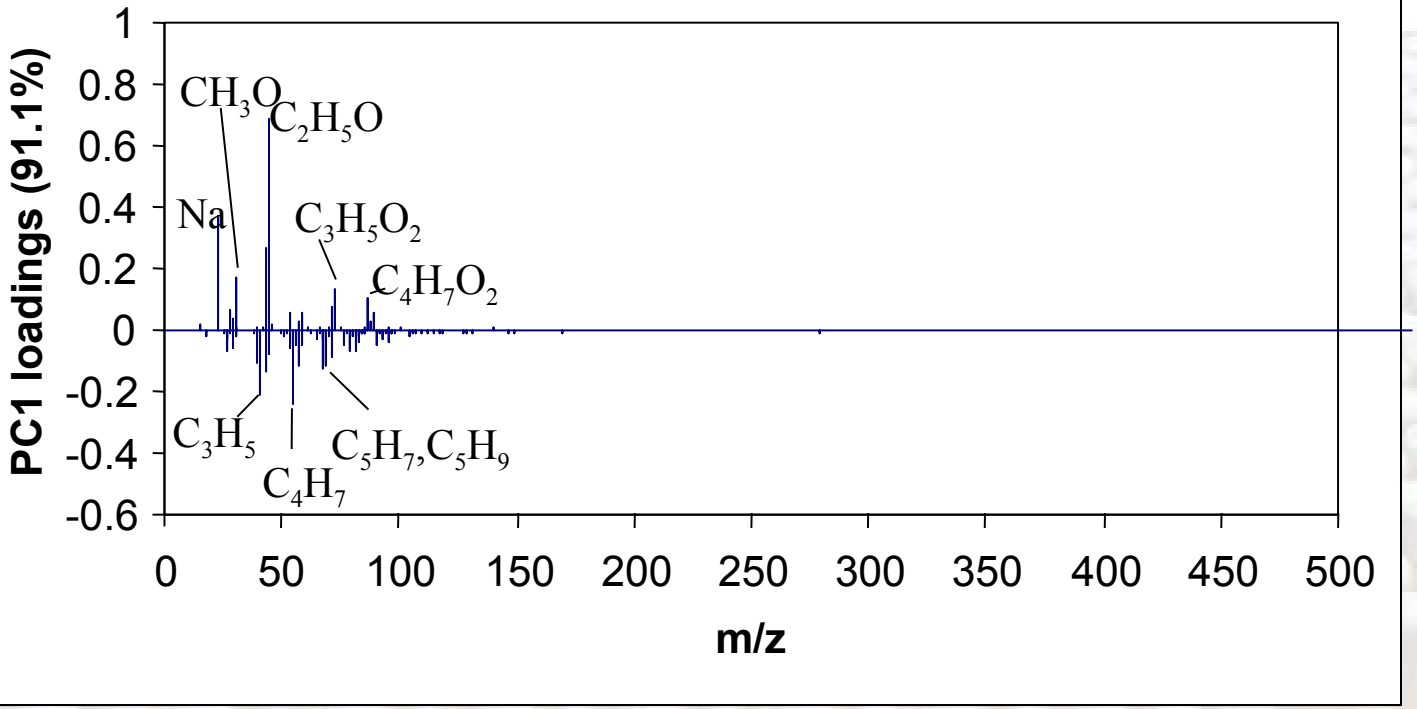


PCA of AAc and NH2 samples

-PCA separates the two sample types (some overlap with 1 NH2 sample)

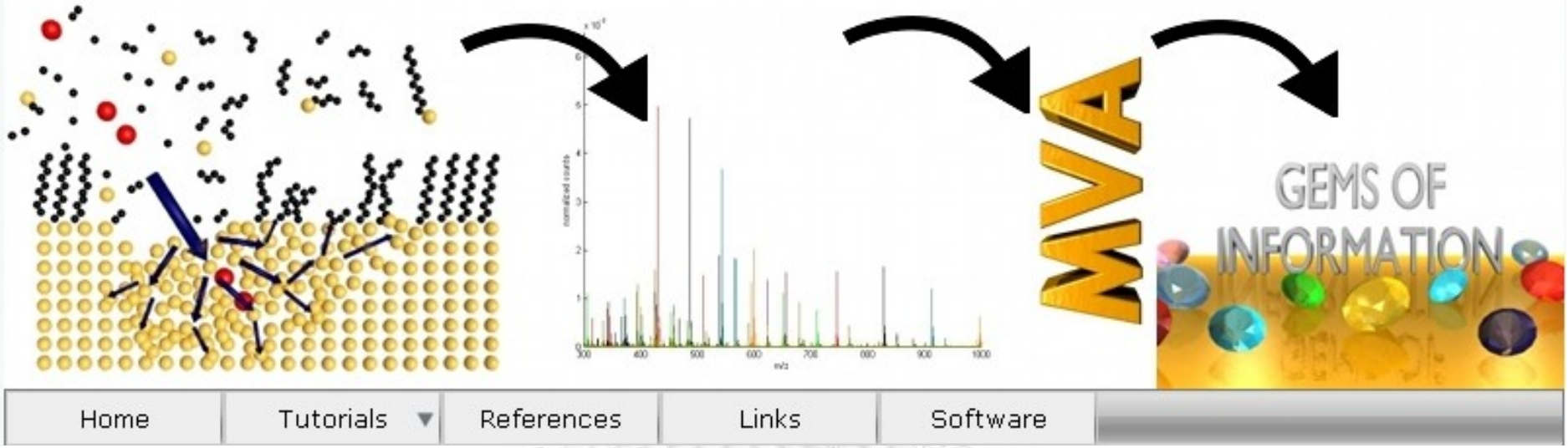
-1 sample from each set is seen to have noticeably different scores (reason unknown)





Conclusions

- **PCA has great potential to aid in spectral interpretation and analysis**
  - **can aid in determining sample differences**
  - **requires well thought out experiments**
  - **cannot do analysis for you**
- **Plan your experiments with a central question and minimize the number of variables**
  - **This can greatly simplify the interpretation**
  - **Can maximize what you get out of your data**



- **mvsa.nb.uw.edu**
  - **Tutorials**
  - **References**
  - **Links**
  - **Software**
  - **Practice Data Sets**

Website is online now
If you have information, tutorials, links, software you would like to contribute please contact me at:
graham@nb.uw.edu

The Risks of Promiscuous MVA of ToF-SIMS Data: The MVA Quickie

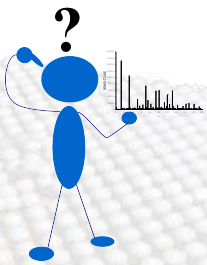
Daniel J. Graham
University of Washington
Department of Bioengineering
UNIVERSITY OF WASHINGTON

SIMS Europe Sept 19-21, 2010
Münster Germany

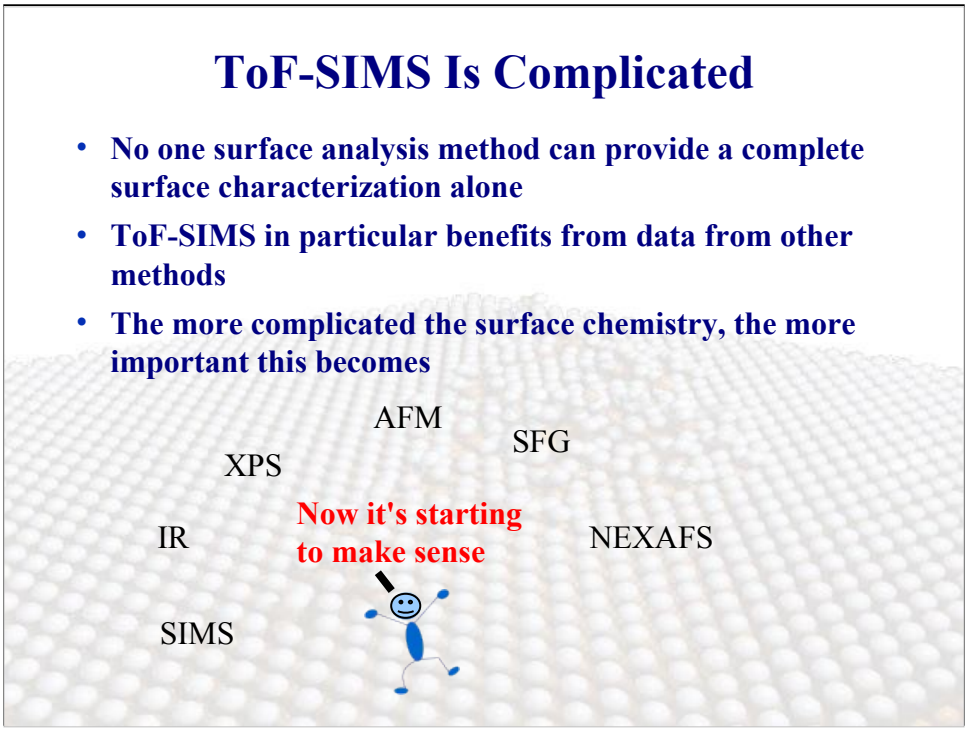
The goal of this presentation is to drive home the fact that MVA is a tool that can aid in the analysis of ToF-SIMS data, but it cannot do the analysis for you. Proper use of MVA requires good planning, sound experimental design and careful review and interpretation of the results.

ToF-SIMS Is Complicated

- **Spectra contain hundreds of peaks**
- **Peak intensities can be interrelated**
- **Matrix effects can cause non-linear changes in peak intensities**
  - **Due to sample composition**
  - **Due to the presence of oxides**
  - **Due to the presence of salts**
- **Peak intensities may or may not correlate with surface composition**
- **Peak intensities may vary due to differential sputtering**
- **Often heavy fragmentation and lack of molecular ion signals (assuming you know what signal to expect)**



There is no way around it. ToF-SIMS data is complicated. There can be multiple factors that can change the relative intensities of peaks within a spectrum or image. These changes can be due to instrumentation, sample preparation, sample composition and more.



As with many surface analytical methods, ToF-SIMS should not be used alone. Complimentary information from other methods can help elucidate and clarify the interpretation of the ToF-SIMS data.

Proper ToF-SIMS Analysis Requires:

- **Good experimental plans (controls)**
- **Proper sample preparation**
- **Careful data collection**
- **Consistent data calibration**
- **Sound understanding of the fundamentals of mass spectral analysis**
- **Knowledge of how to properly use the available tools to help with the analysis**

One should always plan carefully when doing any experiment. This is particularly important when using ToF-SIMS. Controlling extraneous variables and sources of potential variation within a data can be critical to the success of an experiment.

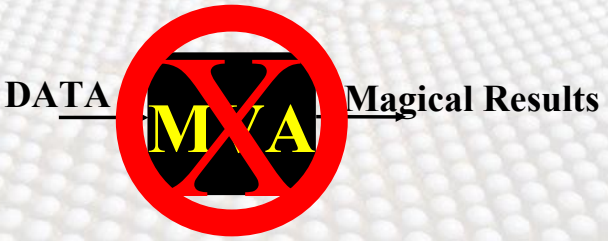
MVA "To The Rescue..."

- **MVA is becoming increasingly popular for ToF-SIMS spectra and images**
- **MVA can aid in reducing the complexity of a data set with regards to the magnitude of the data one needs to focus on (what is changing, where is it different, what peaks are changing)**
- **MVA cannot:**
  - **Remove all effects of contaminants or matrix effects**
  - **Fix a poorly designed experiment**
  - **Interpret your data**
  - **Find things that are not there**

Due to the complexities of ToF-SIMS data, researchers a turning to MVA to aid in sorting through their data. MVA can aid in reducing the complexity of a data set and help highlight what is changing, however MVA cannot interpret your data for you.

Avoid the MVA Quickie

- **MVA should not be treated like a black box**
- **MVA should not be something you do as a last minute decision, it should be part of the experimental design from the beginning**
- **If someone says, "Lets see what we get when we use SIMS and MVA on my samples"...RUN...Or at least stop to think if it really makes sense**



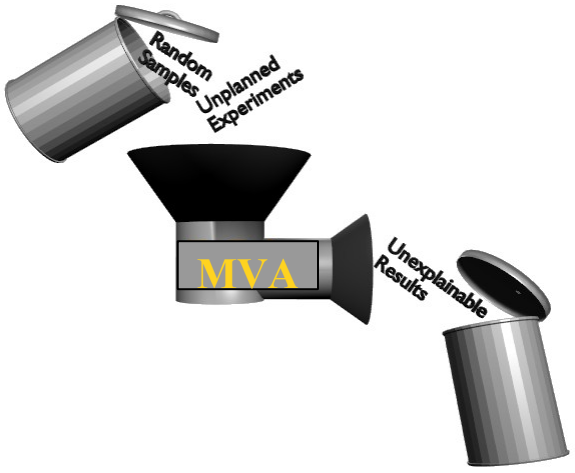
MVA should not be treated as a black box. It cannot help make sense of a poorly designed experiment.

Plan

- What is the question you want to answer?
- What samples do you need to answer that question?
- How many samples/ replicates do you need?

Remember MVA will find the main differences between any samples

If you input garbage in



You will get garbage out!!!

Garbage in = Garbage out

Plan before you start any experiment.

When Should MVA be Used?

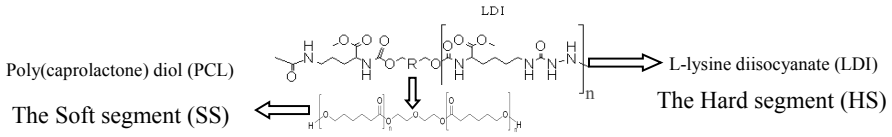
- **MVA should be used to help answer questions**
  - **Are surfaces A and B different?**
  - **How does treatment X change the surface chemistry?**
  - **How is fragmentation pattern affected by ____?**
  - **Can TOF-SIMS data distinguish Protein A from Protein B?**
  - **What regions in my image are different and why?**
- **The question should be part of the experimental design and not an afterthought**

MVA should be part of the experimental design, not an afterthought of "wow this is complex, maybe we should use MVA."

Design your experiments around a specific question/hypothesis to be answered. Think about how MVA can help and which method would be best suited to the experimental design.

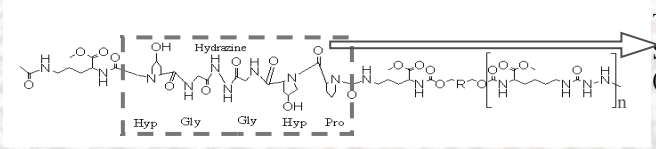
Structures of the polyurethane units

Control sample



PUU 817, LDI/PCL/Hyd =8/1/7

Structure of the Poly(peptide-urethaneurea)



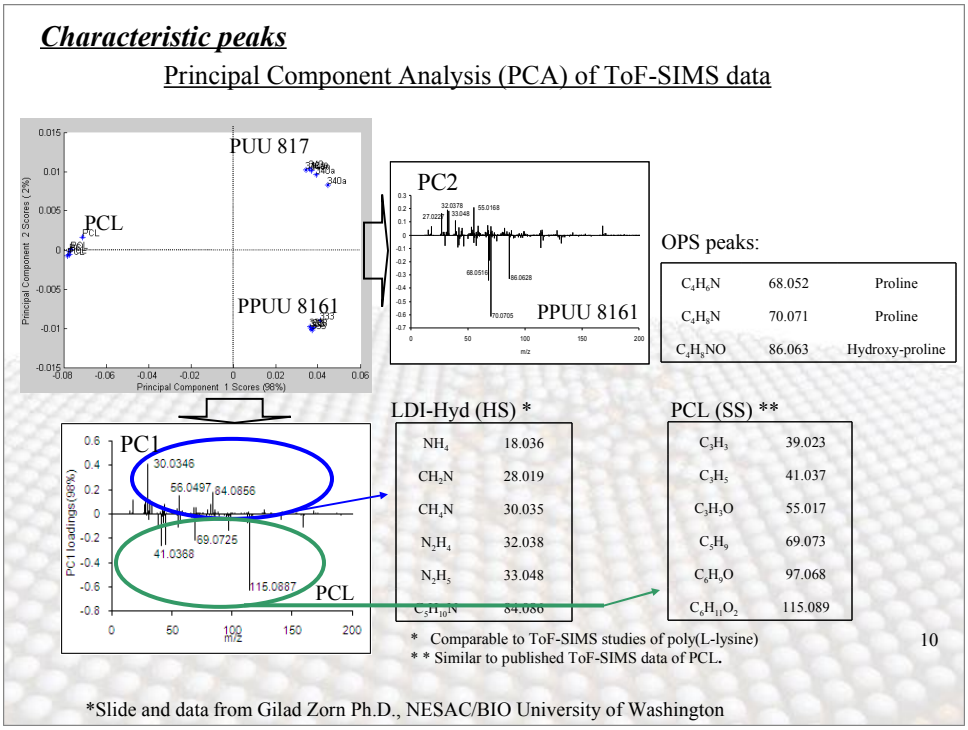
PPUU 8161, LDI/PCL/Hyd/(Gly/Hyp/Pro/Hyd) =8/1/6/1

*Slide and data from Gilad Zorn Ph.D., NESAC/BIO University of Washington

9

In this example from Zorn .et .al. the authors compared a set of polyurethane samples. Both polymers had a PCL softsegment. One of the polymers included a polypeptide segment within the hard segment.

The authors wanted to verify the presence of the peptide within the polymer and determine which peaks were characteristic of each material. For this they used PCA.

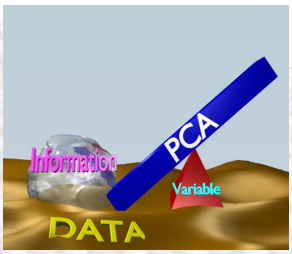


It was found that the PCA PC1 scores clearly separated the PCL soft segment from the two different hard segment materials. PC2 separated the two different polymers. It was seen that peaks characteristic of the peptide segment were found to have high loadings corresponding with the peptide containing polymer (negative scores and loadings on PC2).

This graphic illustrates the "Steps" required to properly carrying out MVA of ToF-SIMS data. This presentation will not cover all of them. These steps are covered in more detail in the "PCA Step by Step" tutorial on the NESAC/BIO MVSA website (http://mvsa.nb.uw.edu).

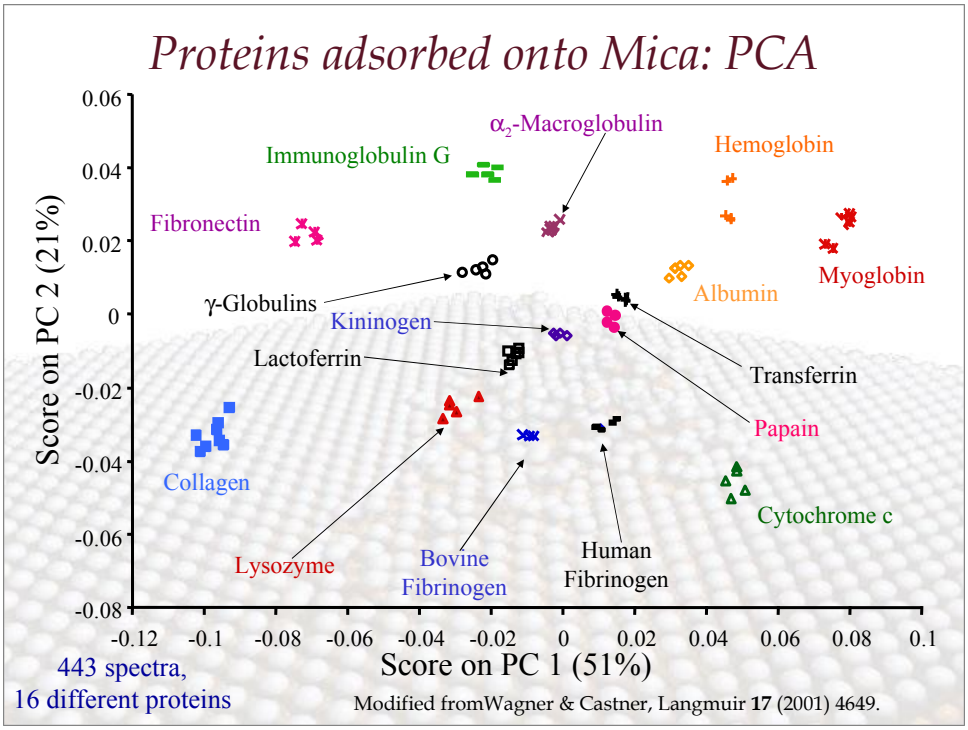
Experimental Design/Data Collection

- **Not all systems are well defined, but your experimental design can be:**
  - **Think about what you want to learn from SIMS**
  - **Simplify the number of variables you are dealing with per experiment**
  - **Plan appropriate controls**
  - **Run enough replicates to determine reproducibility**
    - **Homogeneous => 3 to 5 spots on 2 samples**
    - **Inhomogeneous => 5 to 7 spots on 3 to 5 samples**



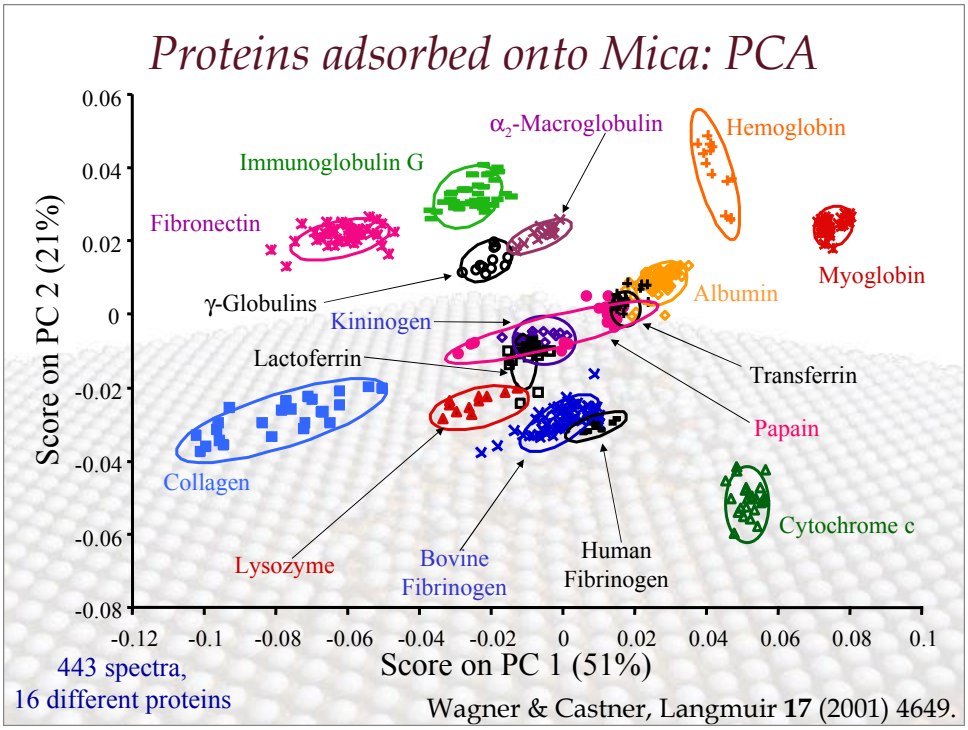
It is important to think about what one wants to learn from ToF-SIMS data and also to plan experiments where a minimal number of variables change within a sample set. By reducing the number of variables that are changing, one can more easily relate changes seen in the spectra to the variable of interest. Ideally only one variable should change within a given set of samples.

It is also important to include the proper number of replicates in order to get statistical significance to the MVA results. This slide provides a general guideline, but ultimately the number of data points to collect will depend on the sample type. In general, more homogenous samples require fewer data points, and less homogeneous samples require more data points.

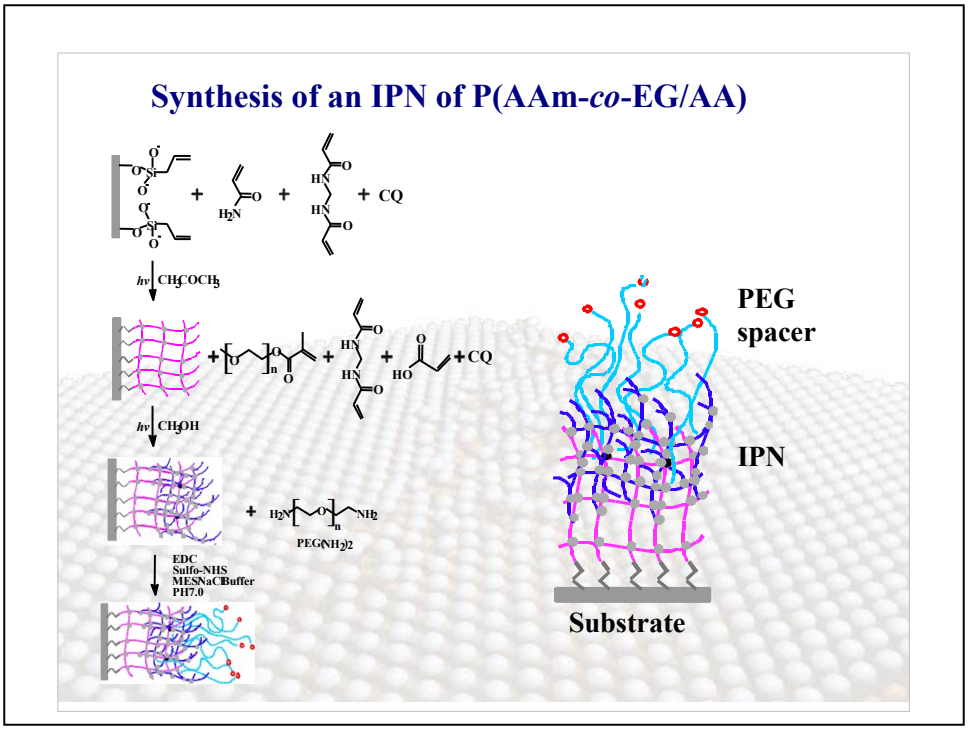


This slide was adapted from the work of Matt Wagner and Dave Castner. In this slide I have deleted most of the data points from the protein data set generated by Matt. With this set of data points one could look at the data and conclude that all of the proteins are clearly separated and that the scatter in the data is minimal.

However if you add back in all the data points, the story changes....

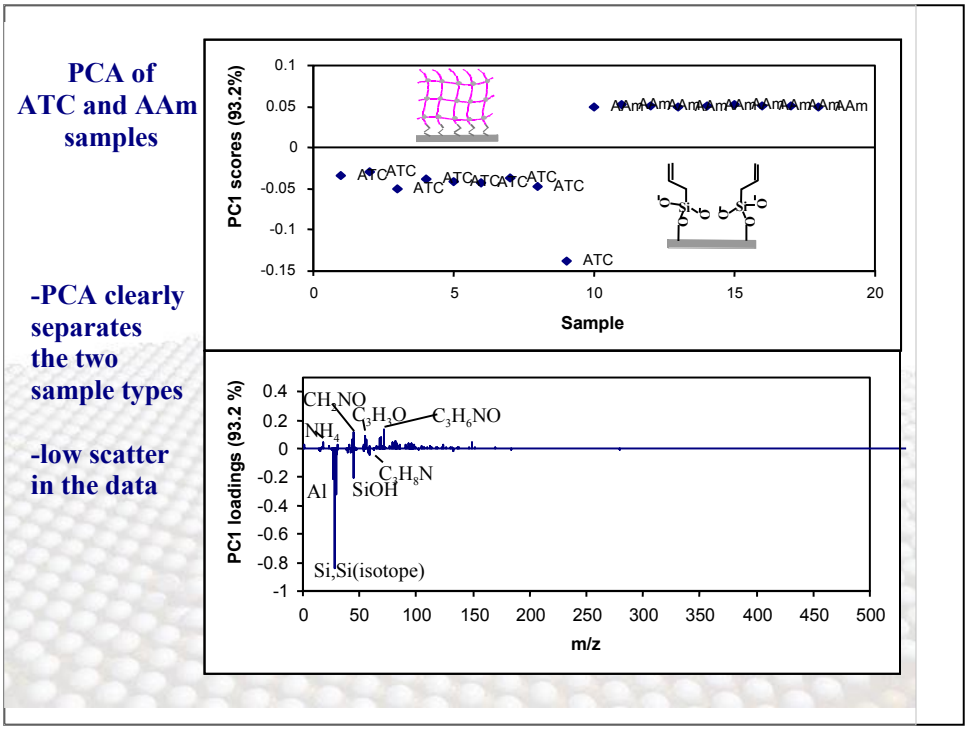


This slide shows the Wagner data with all of the data points. Most all of the proteins are still clearly separated from each other, however there is significant scatter in the data of some protein. Collecting a proper number of data points is critical in understanding the true variance within a data set.

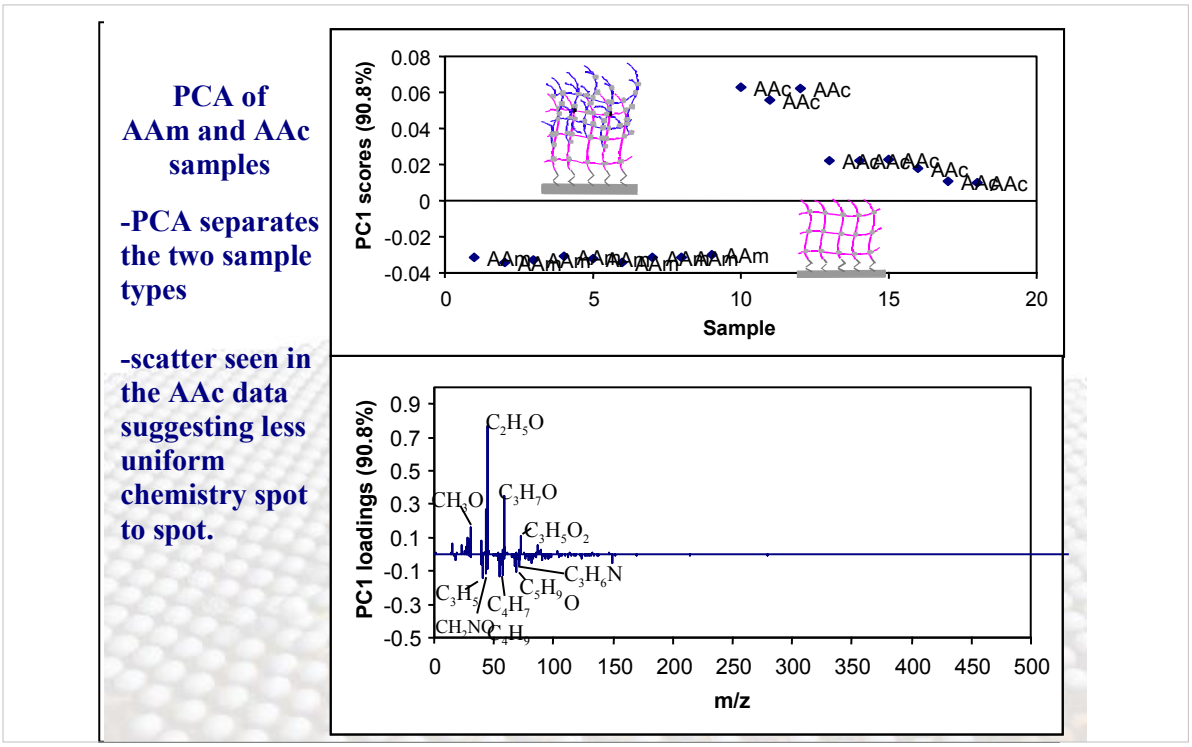


This example is taken from data collected from samples prepared from Kevin Healy's group at Berkely. The sample consists of an interpenetrating polymer network (IPN) created by sequentially attaching various compounds to the substrate.

Due to the similarity in the chemistry of the various compounds when looking at all of the data together, much of the data overlaps in the PCA scores. However, when the samples are compared step by step throughout the IPN creation process one can separate out the various chemistries and find peaks that are characteristic of each compound.



Here the base silanized substrate is compared to the surface after addition of the AAm polymer. PCA separates the two sample surfaces and highlights the characteristic peaks for each of the compounds on the surface.

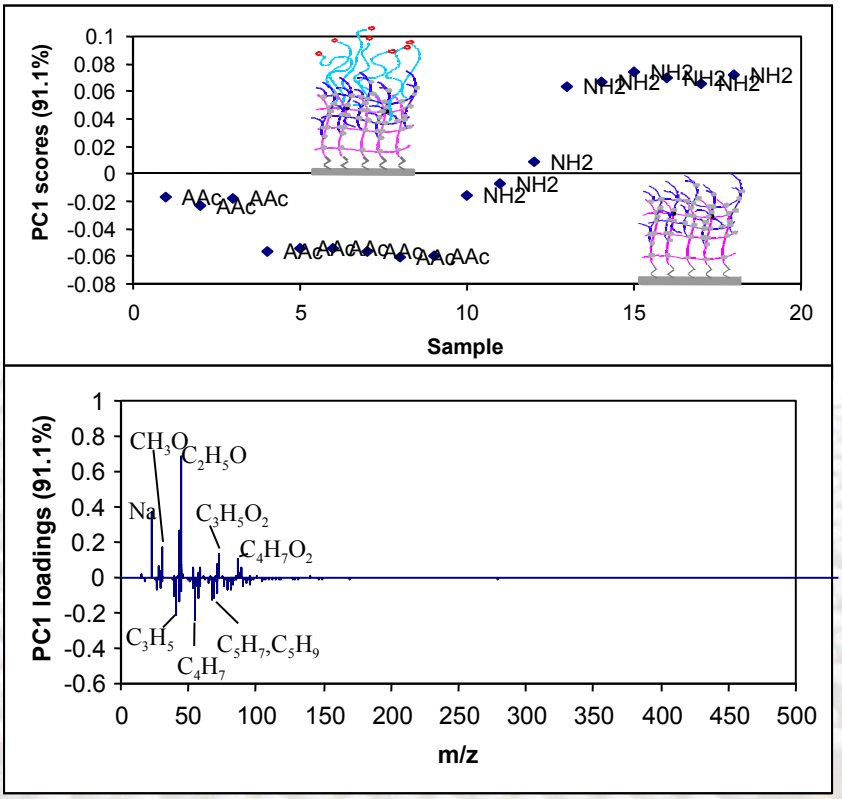


Here the AAm polymer is compared to the AAc polymer. The two samples types are separated on the scores plot, though it is noted that there is more scatter within the AAc samples.

PCA of AAc and NH2 samples

-PCA separates the two sample types (some overlap with 1 NH2 sample)

-1 sample from each set is seen to have noticeably different scores (reason unknown)



Comparing the AAc samples to the NH2 samples shows some overlap on PC1 and scatter in both samples types. However, overall the two samples types are separated on PC1. Since both polymers contain similar chemical structures, the loadings appear to be showing that the NH2 samples have a higher relative intensity of the oxygen containing fragments than the AAc samples. This is verified in the raw data (not shown).

Conclusions

- **PCA has great potential to aid in spectral interpretation and analysis**
  - **can aid in determining sample differences**
  - **requires well thought out experiments**
  - **cannot do analysis for you**
- **Plan your experiments with a central question and minimize the number of variables**
  - **This can greatly simplify the interpretation**
  - **Can maximize what you get out of your data**

MVA encompasses a powerful set of methods that can be valuable for the ToF-SIMS analyst. However, MVA is a tool and not a replacement of common sense and good experimental planning.

The ToF-SIMS analyst still needs to know "standard" data analysis methodologies and also needs to understand why they are using a given MVA method and the assumptions that go along with each method.

The NESAC/BIO MVSA website is a community resource providing information about the application of MVA to surface analytical data.

The website provides:

Tutorials, references, links and software to aid in the application of MVA to ToF-SIMS and other surface analytical data.

If you have materials that could be useful for the community, please consider contributing them to the website.