SECONDARY ION MASS SPECTROMETRY

DATA ANALYSIS IN 1D, 2D AND 3D

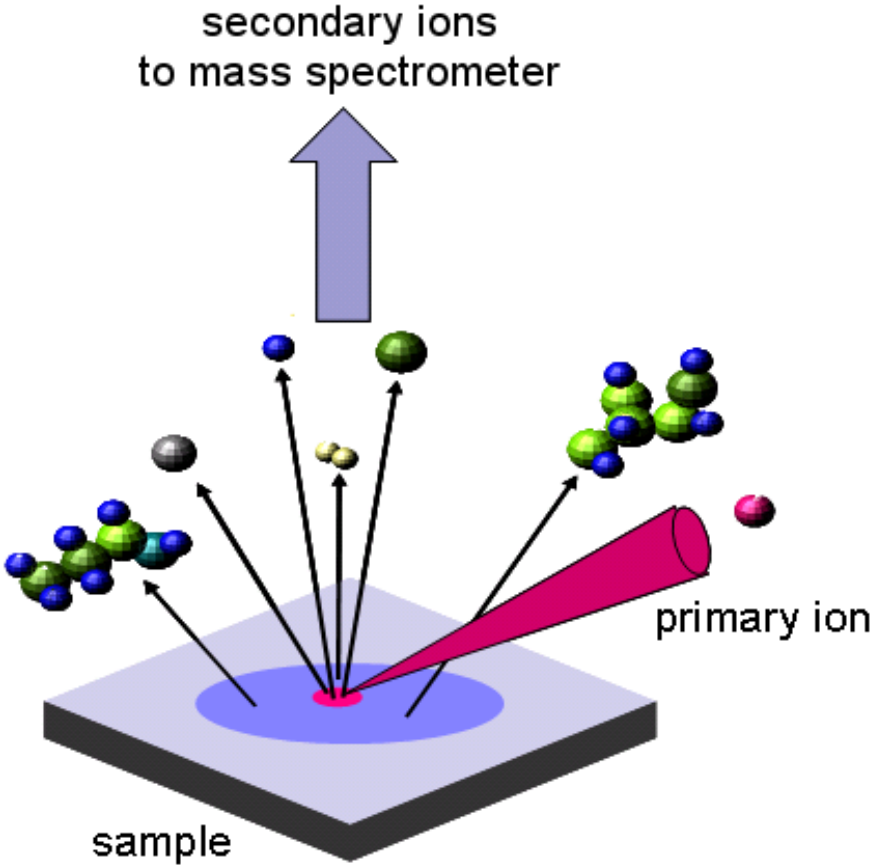
ALEX HENDERSON

Surface Analysis Research Centre
University of Manchester, UK

SIMS technique

- Fire energetic ions at a solid – primary ions
- Solid emits (sputters)
  positive ions
  negative ions
  electrons and neutrals
- These secondary ions produce a mass spectrum

Secondary **I**on
Mass **S**pectrometry



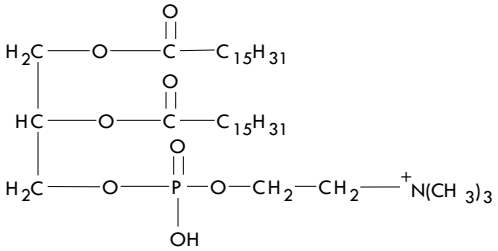
Comparison with other MS

- No (chromatographic) separation step so spectra are overlapping
- Limited amount of sample available
- Relatively low throughput so no large databases of spectra
- Highly surface sensitive so detects contamination easily (pro and con)
- Need to transport analyte to gas phase; proteins decompose, polymers generally OK
- MS/MS being developed (here)
- Negative ion spectra available

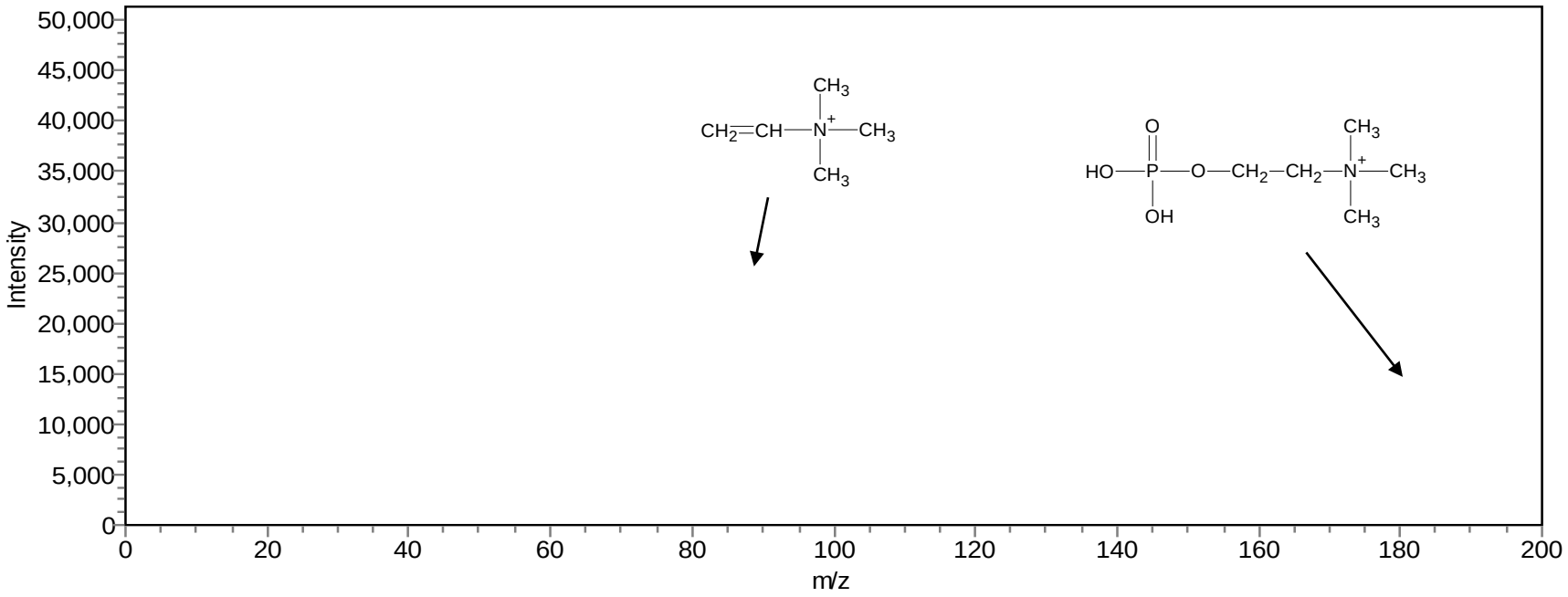
1D analysis

Collections of spectra

Typical SIMS spectrum



1,2-Dipalmitoyl-sn-glycero-3-phosphatidylcholine (DPPC) (positive ion)

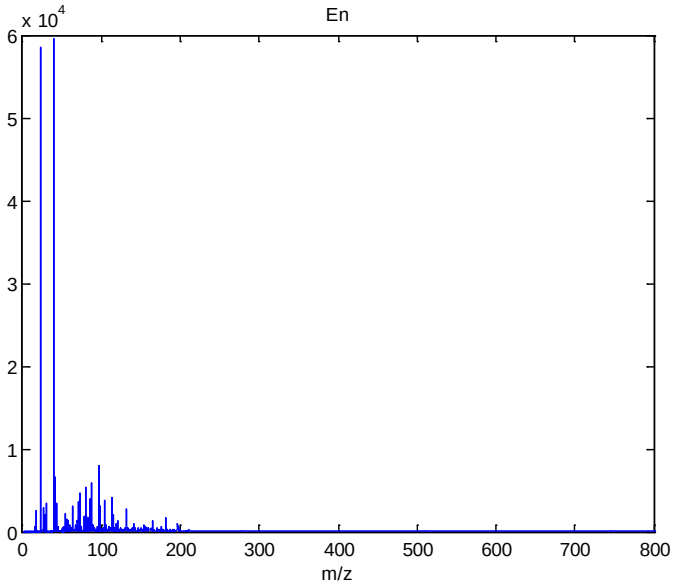


The Static SIMS Library, SurfaceSpectra Ltd

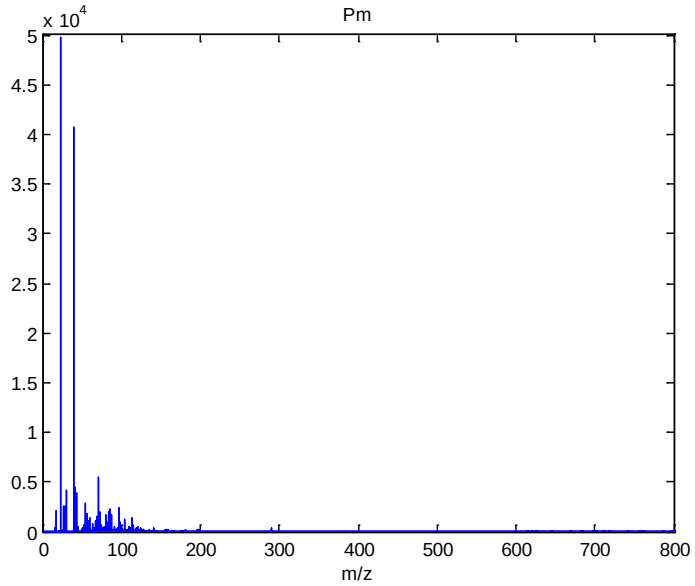
- 100 – 250k data points
- 1 – 3000 u mass range

Spectra of bacteria

***Enterococcus* spp.**



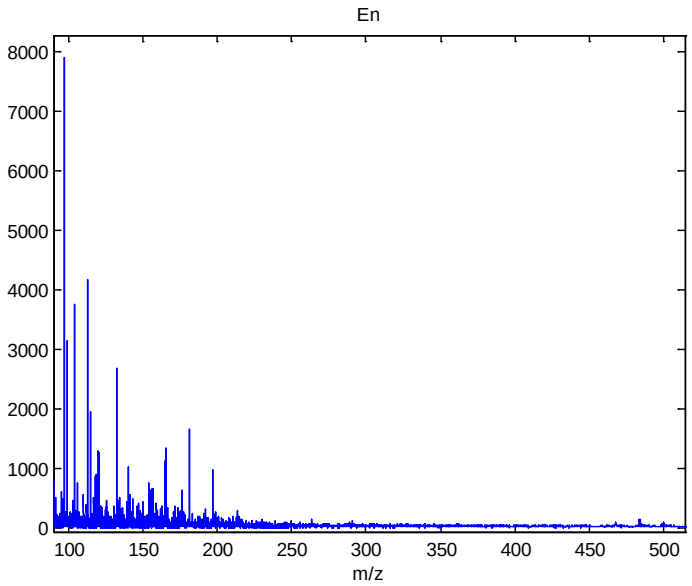
Proteus mirabilis



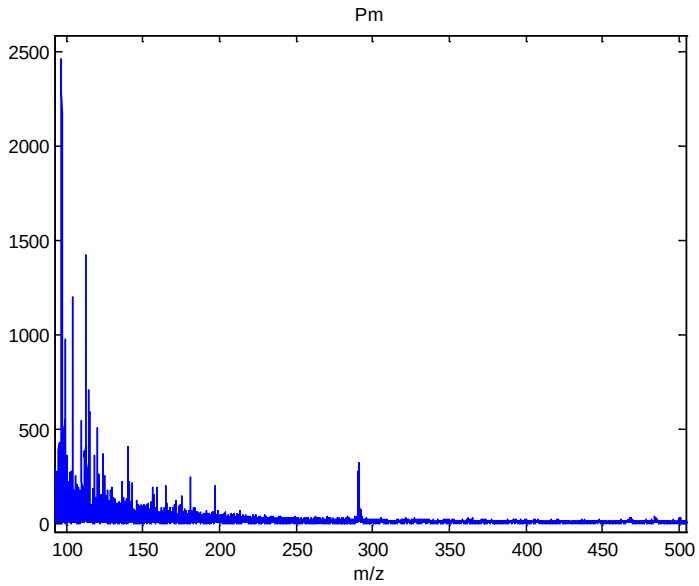
Applied Surface Science **252** (2006) 6719-6722

Spectra of bacteria

***Enterococcus* spp.**



Proteus mirabilis



Applied Surface Science **252** (2006) 6719-6722

PCA of bacteria

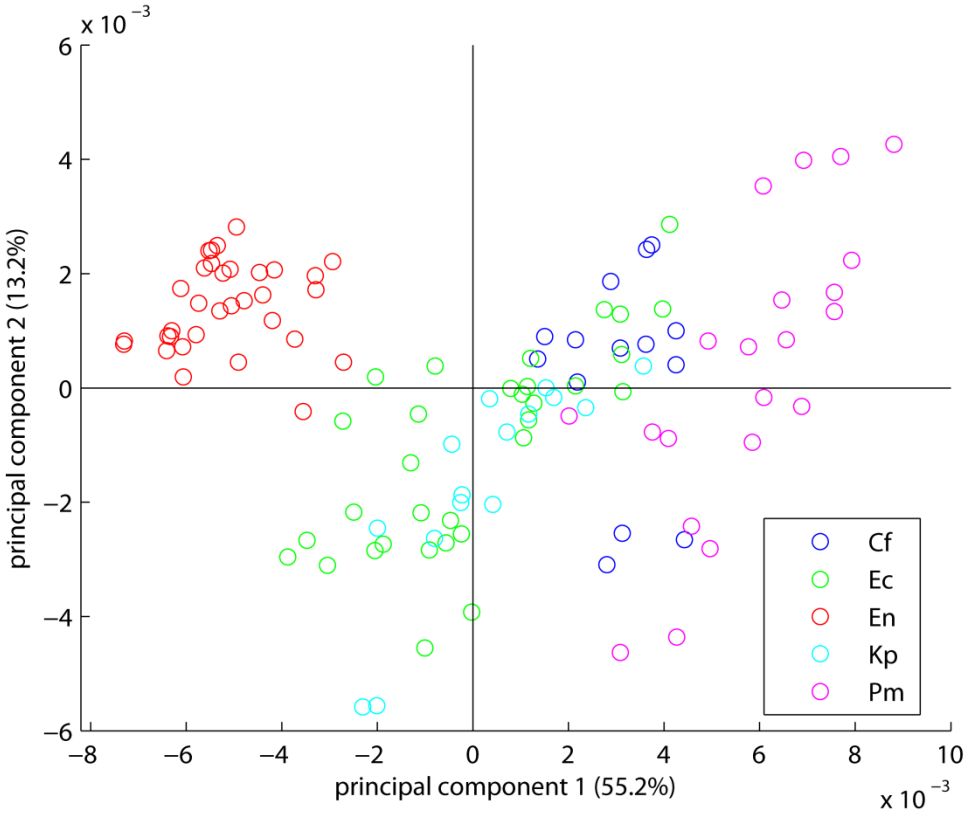
5 species of bacteria

No *a priori* knowledge used in PCA

Positive ion spectra 1-800 u, rebinned to 1 u steps

Square root of intensity

Sum-normalised

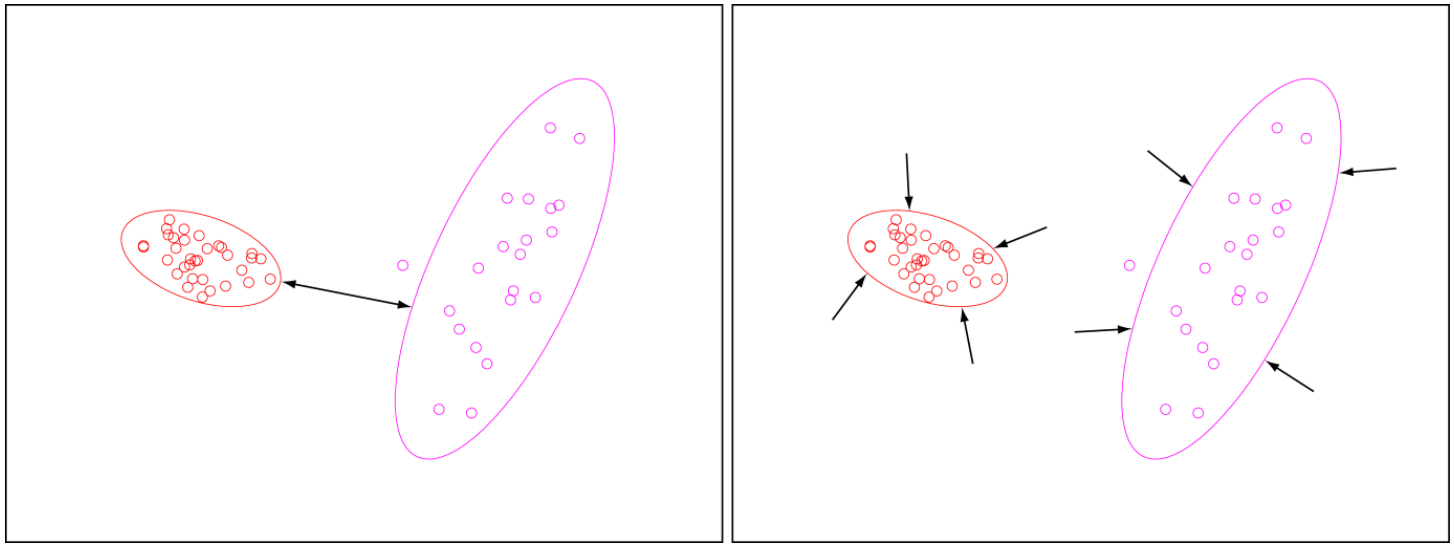


Applied Surface Science **252** (2006) 6719-6722

Canonical Variates Analysis (Discriminant Function Analysis)

- Principal components are pure (orthogonal) aspects of the data
- Canonical variates are combinations of PCs that best describe an *a priori* class structure
- Based on Fisher's Linear Discriminant, subtly different from Linear Discriminant Analysis (LDA): no assumption of normally distributed classes
- CVA is also referred to as DFA

Canonical Variates Analysis (Discriminant Function Analysis)



- Maximise ratio of between-group variance to within-group variance

PC-CVA bacteria classification

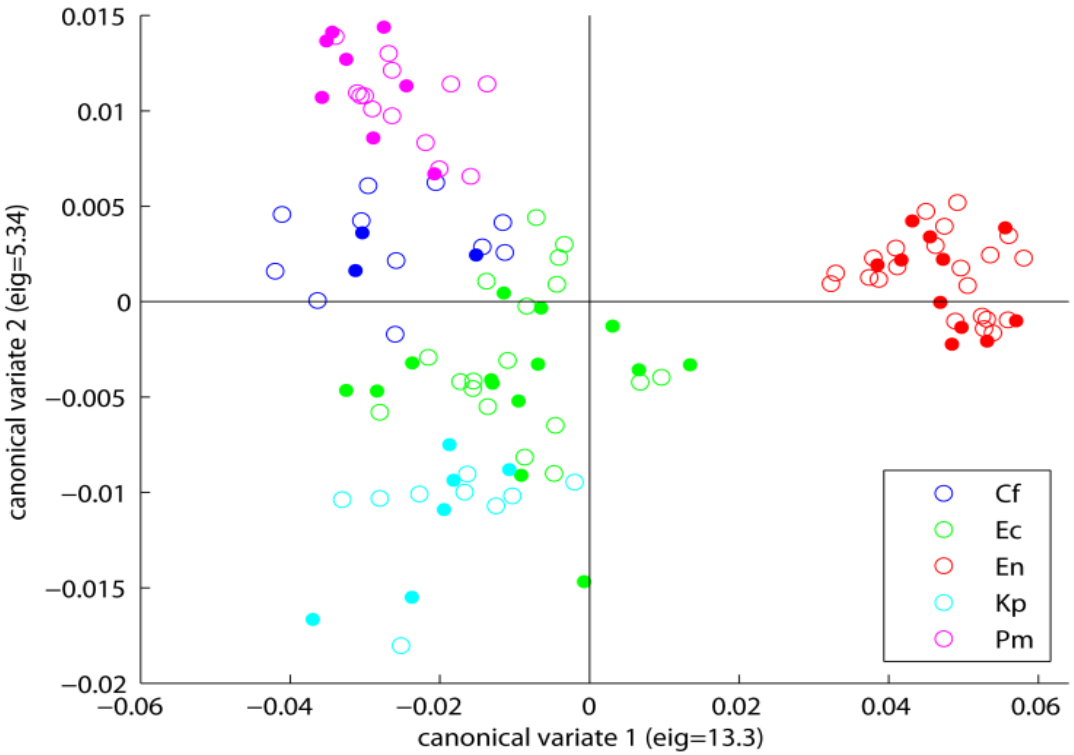
5 species of bacteria

Class structure used

9 PCs selected using PRESS test

Cross-validation indicates percent correctly classified:

Cf 75% CC
Ec 92% CC
En 100% CC
Kp 25% CC
Pm 50% CC



Applied Surface Science **252** (2006) 6869-6874

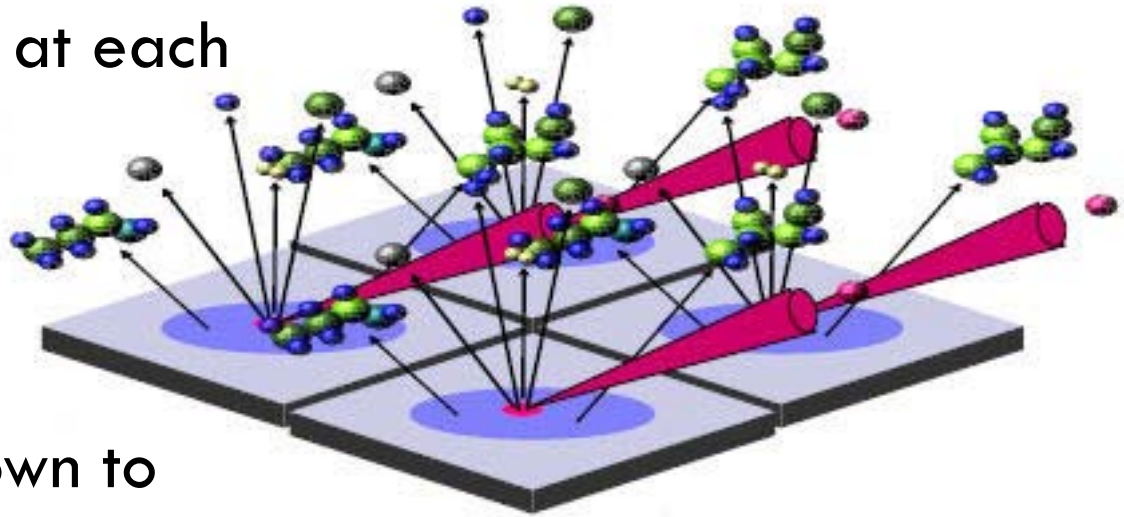
2D analysis

Spatially organised collections of spectra

Hyperspectral images

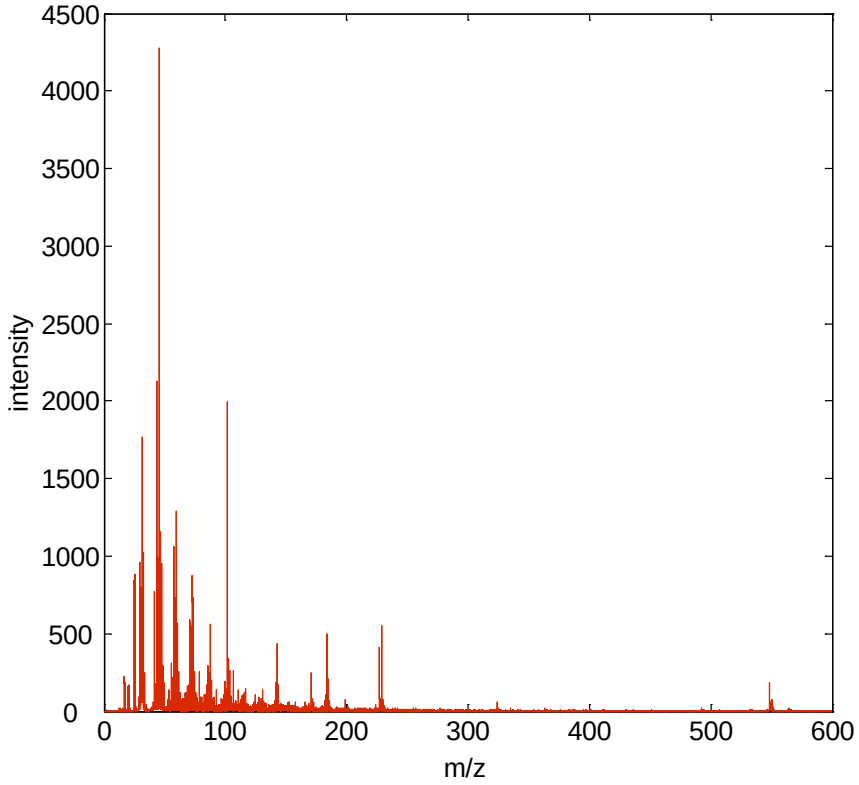
Imaging SIMS

- Focussed ion beam stepped from point to point
- 2D data generated
- Full mass spectrum at each point/pixel

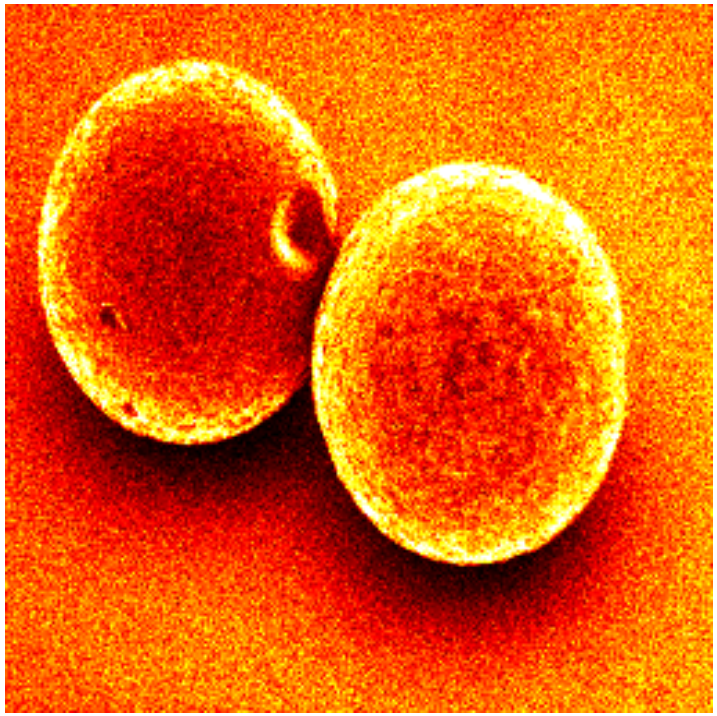


- Beam diameter down to ~1-2 µm
  (50 nm possible, but can damage sample)

SIMS image - ‘chemical photograph’



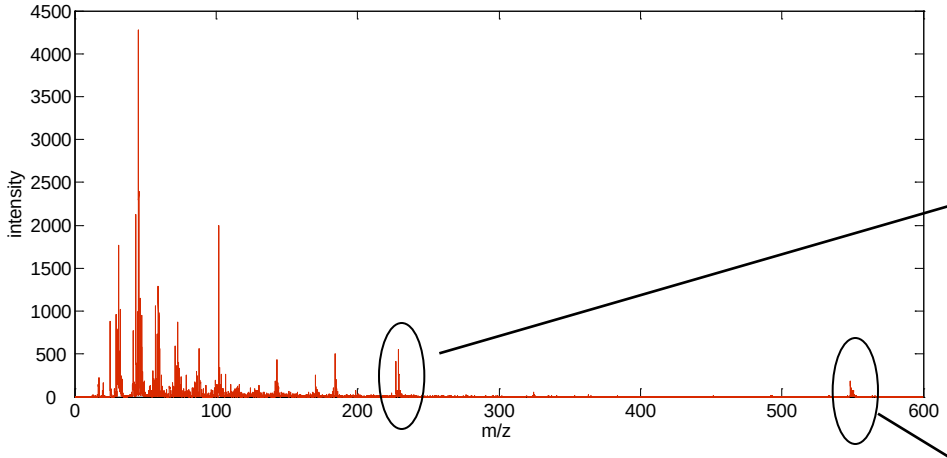
Total ion spectrum

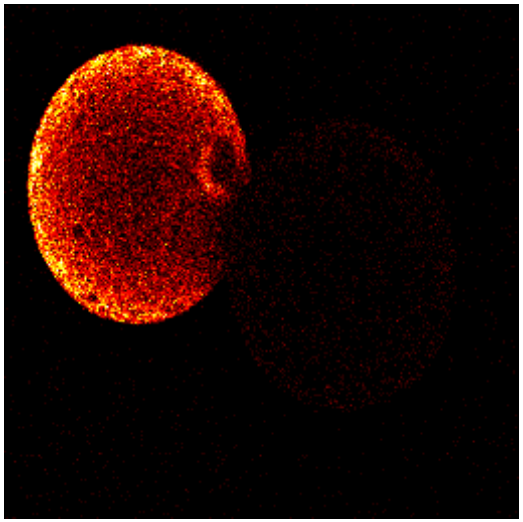


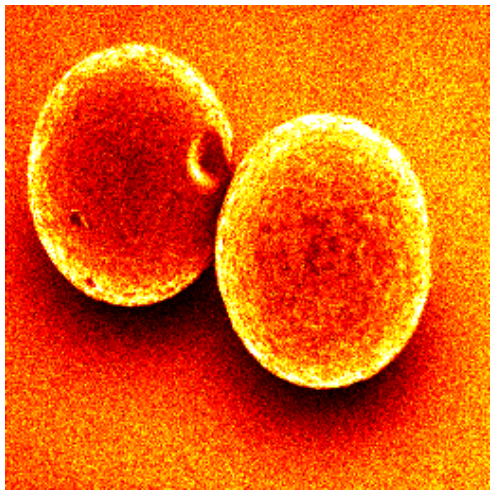
Total ion image
Bead diameter 50 µm

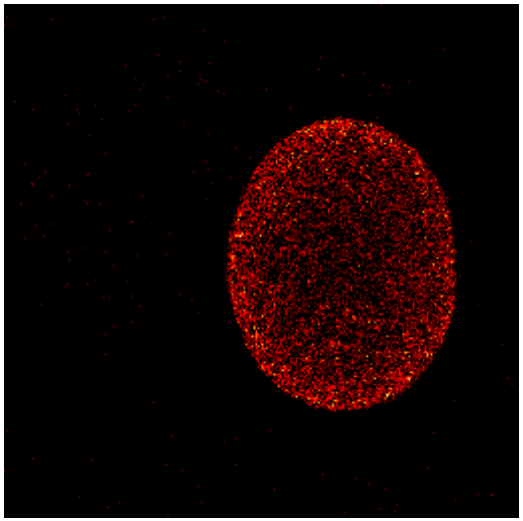
Data courtesy of Penn State University, USA

Selected ion images







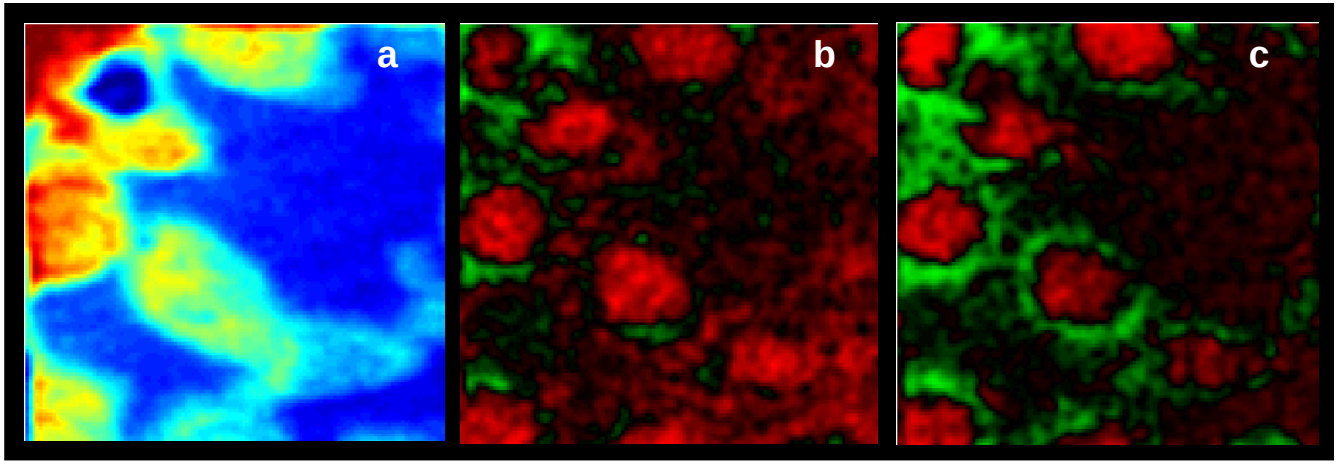


Data courtesy of Penn State University, USA

Maximum Autocorrelation Factor Analysis (MAF)

- Incorporates spatial information with spectral information
- Determine the correlation with the same image offset diagonally by 1 pixel
- Autocorrelation means scaling invariant
- Computation time ~4× PCA
- Possible issue when image feature size ≈ pixel size

MAF versus PCA of HeLa cells

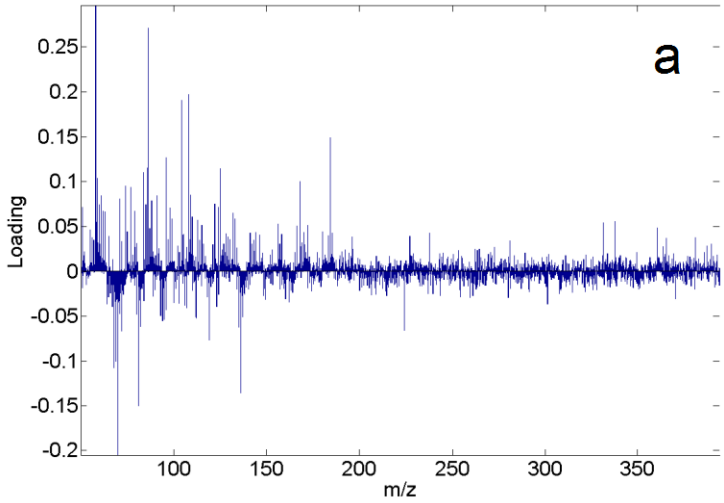


a) Total ion image, scaled and smoothed
b) Score on PC 6, scaled and smoothed
c) Score on MAF 4, scaled and smoothed

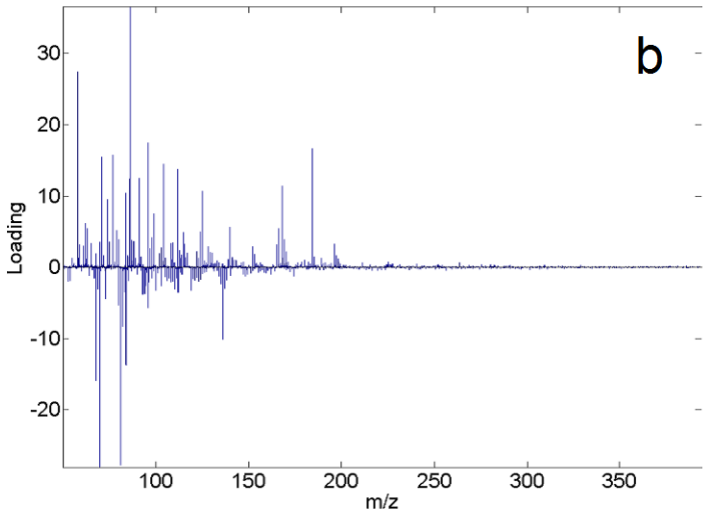
MAF captures the distinction between inner and outer cellular regions more clearly than PCA. Comparison with the total ion image shows the additional information available following multivariate analysis.

Surface and Interface Analysis **41** (2009) 666-674

MAF versus PCA of HeLa cells



Loading on PC 6



Loading on MAF 4

- MAF loadings more 'informative'

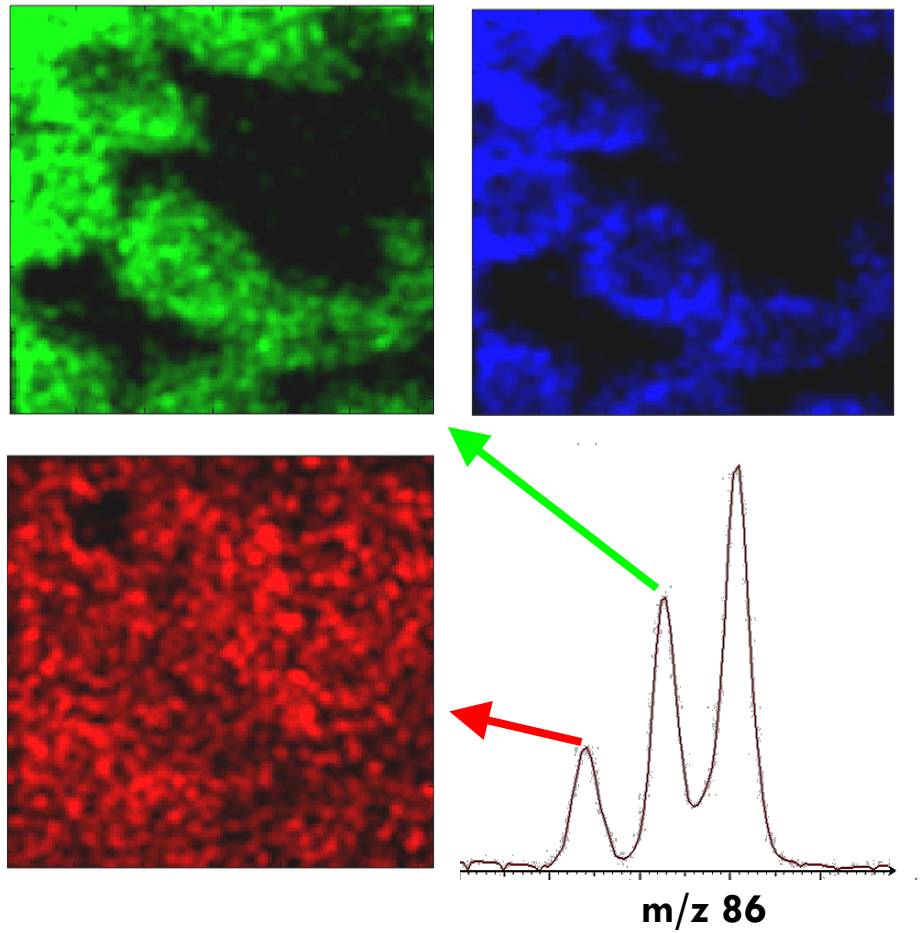
Surface and Interface Analysis **41** (2009) 666-674

High mass resolution required

Nominally all at m/z 86

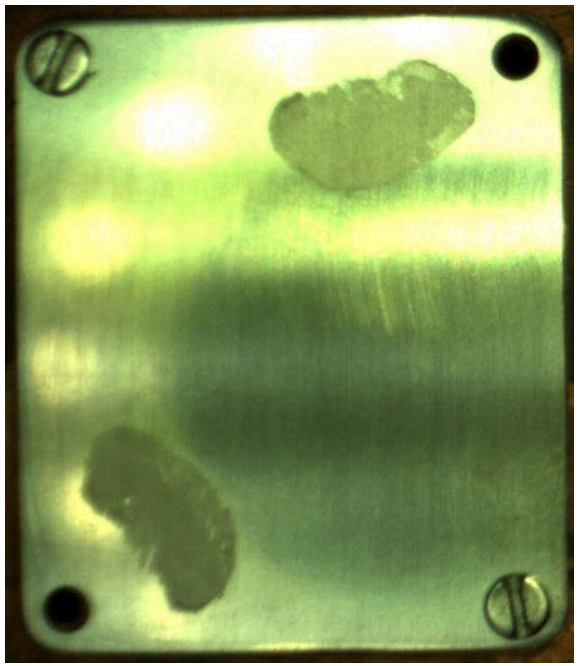
- Lipid (DPPC)
  m/z = 86.0969692
- Unknown
- Silicon substrate $[Si_3H_2]^+$
  m/z = 85.9464332

Separation = 0.15 u



Analytical chemistry **80** (2008) 9058-9064

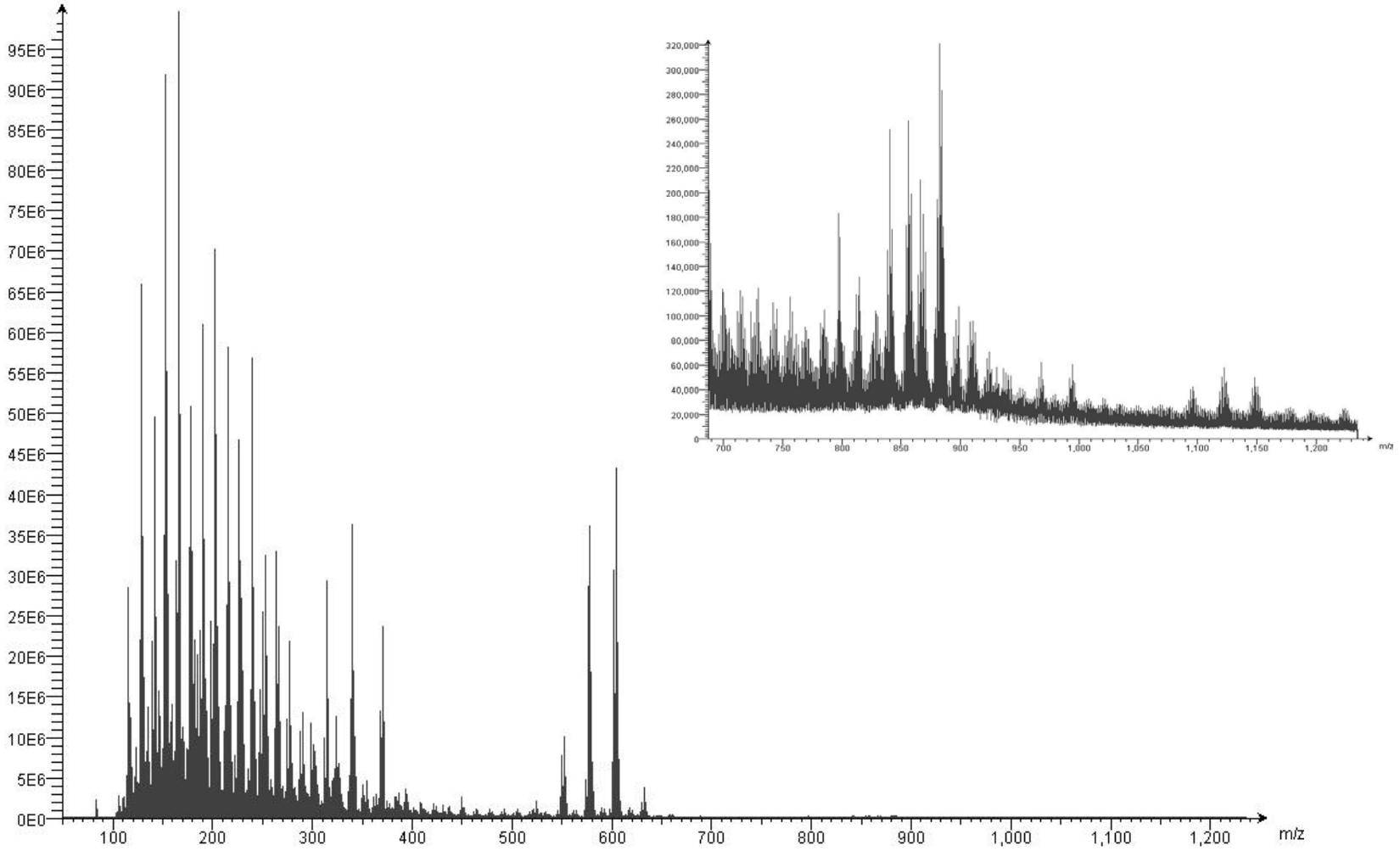
Rat kidney tissue



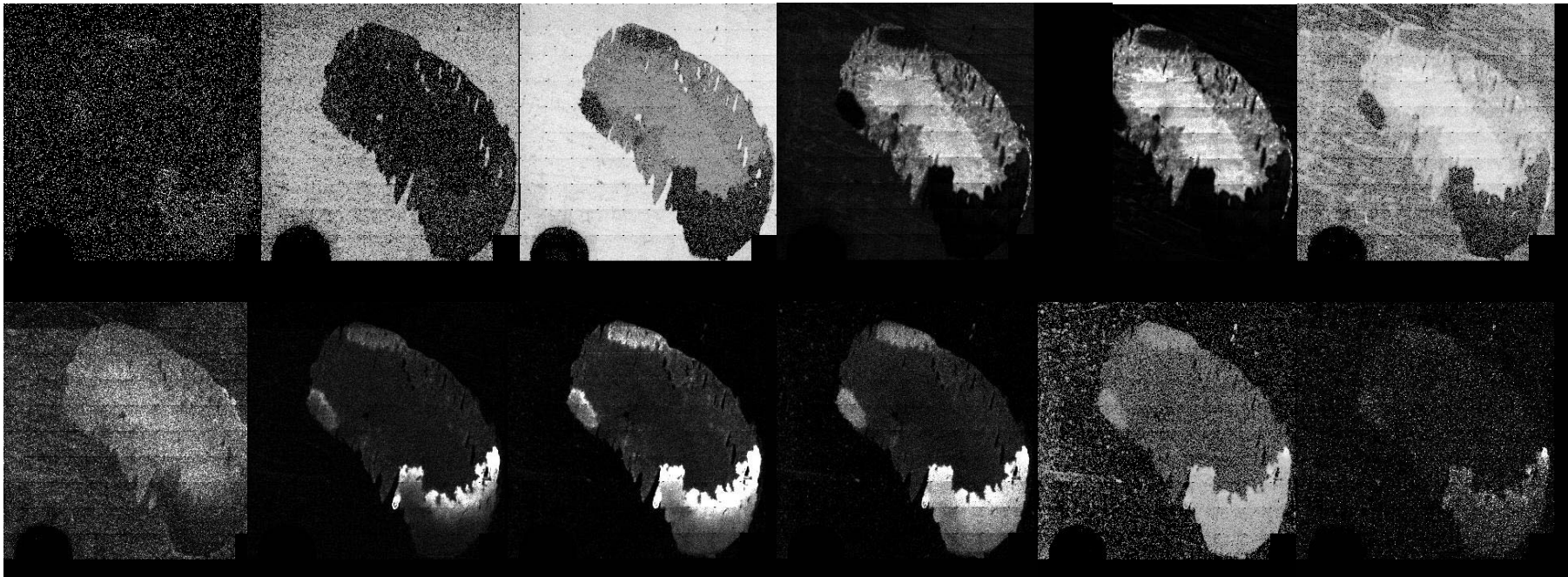
- Cryo-microtomed slices
- Sequential layers
  - Unwashed
  - Washed in methanol
- Data is
  - 10x10 tiles
  - 32x32 pixels per tile
  - 24000 channels per pixel
  - 2.4 billion datapoints
  - 19.2 GB (as doubles in MATLAB)

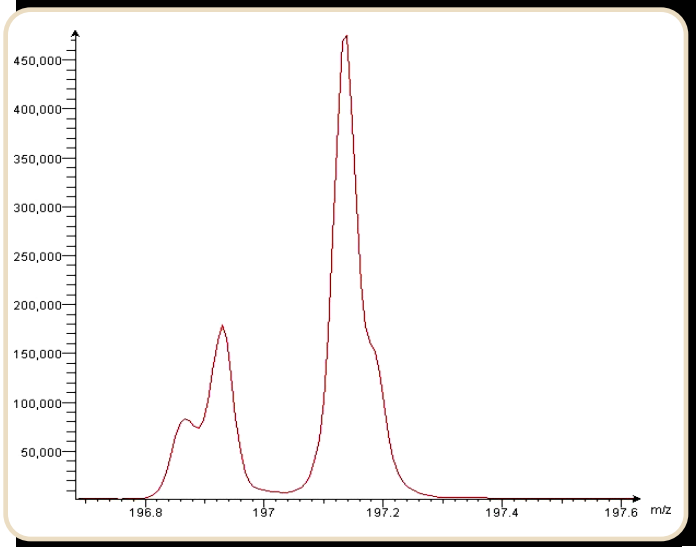
Almost all zeros, so use sparse matricies to handle the data

Summation of all tissue pixels



Clear changes in chemical distribution over small mass range





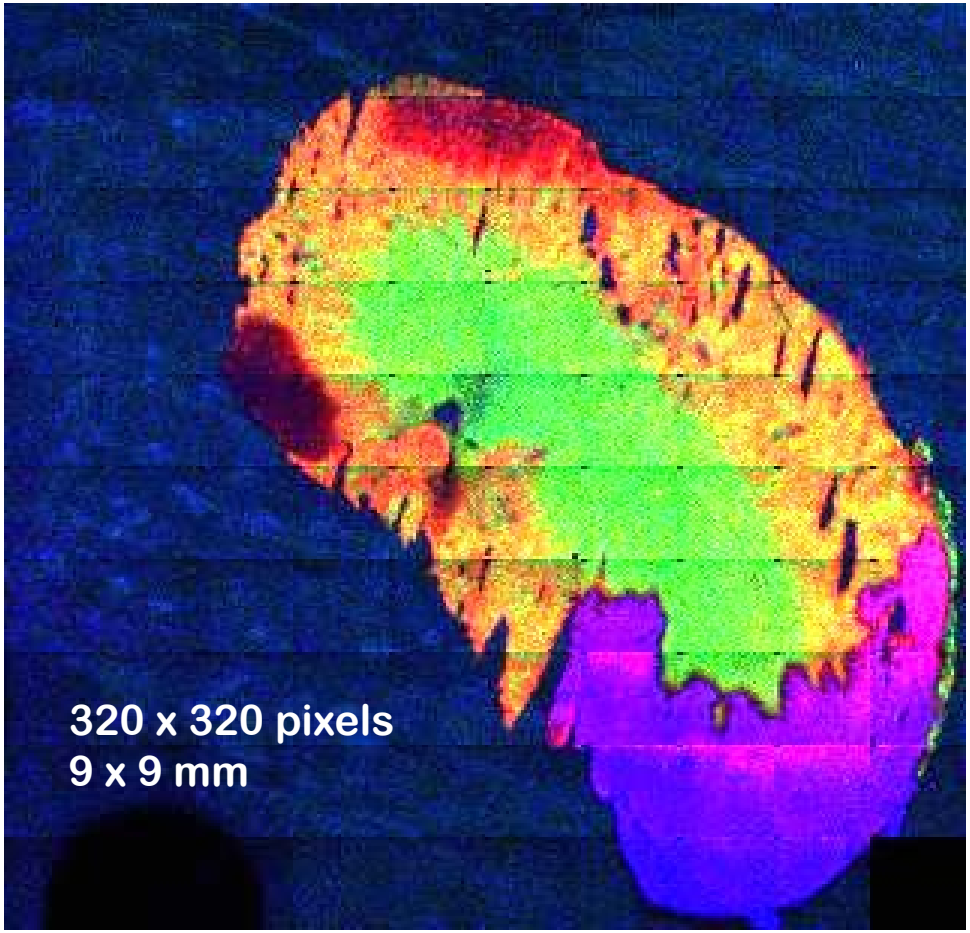
Individual images from nominal mass m/z 197

Images generated at 0.05 u increments
from 196.75 to 197.30

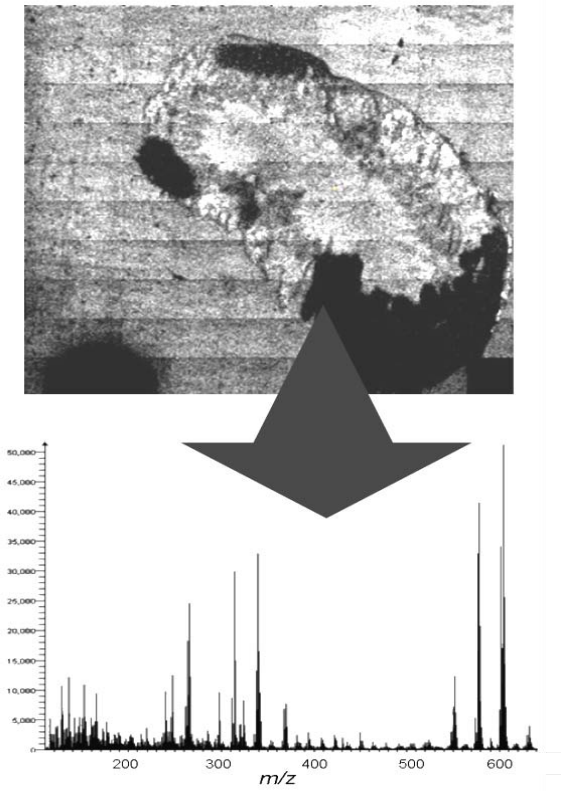
Images are 9 x 9 mm^2 comprising
10 x 10 tiles of 32 x 32 pixels

Analytical and Bioanalytical Chemistry **396** (2010) 85-104

Rat kidney section - overlay



Lipid (DPPC) **DAG** **Salt related**

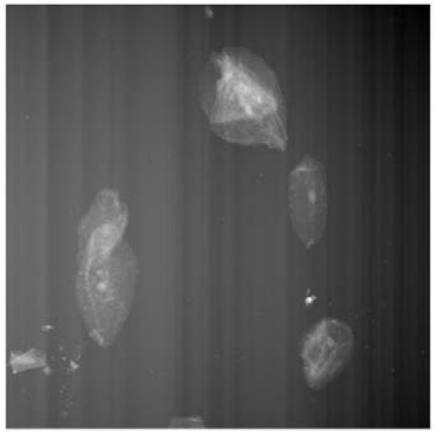


Full mass spectrum available at each pixel

Analytical and Bioanalytical Chemistry **396** (2010) 85-104

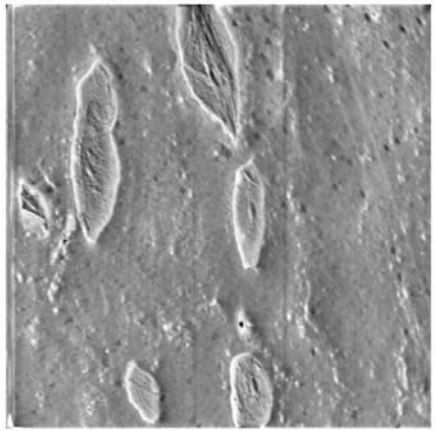
SIMS and AFM

AFM



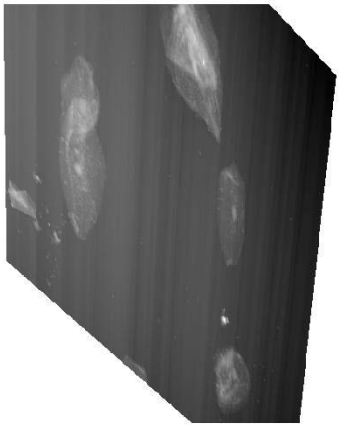
Identify points of reference between images

SIMS (total ion)

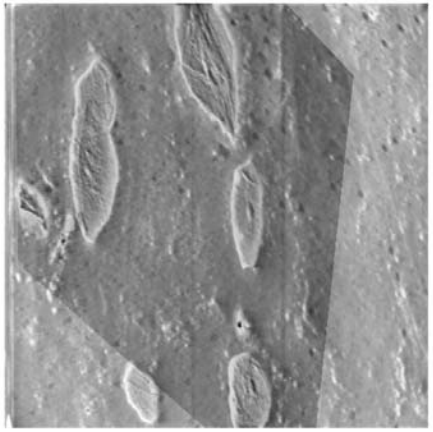


(Human cheek cells)

AFM following 'projective' transform and cropping in MATLAB

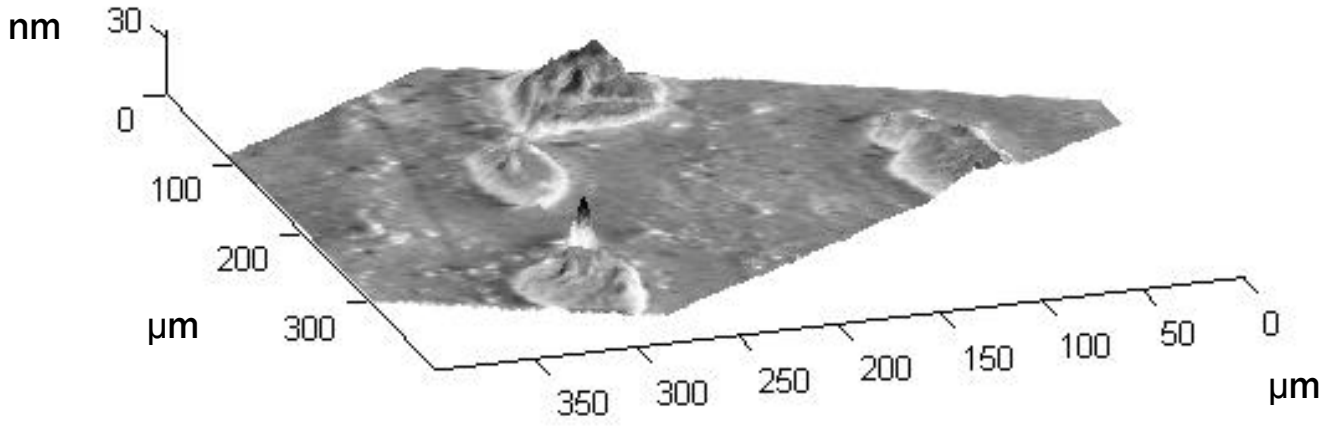


Register images and overlay



Applied Surface Science **255** (2008) 1264–1270

SIMS and AFM



- Use AFM *z*-height to offset SIMS data
- Use AFM *x*- and *y*-position to accurately scale SIMS *x*- and *y*-positions

Applied Surface Science **255** (2008) 1264–1270

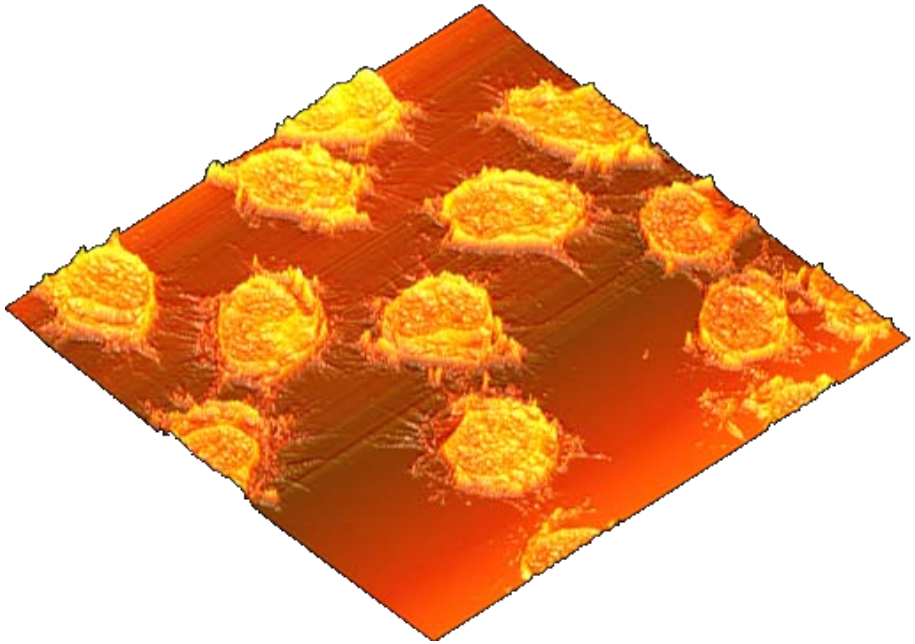
3D analysis

Three-dimensional spatially organised collections of spectra

‘Hyper-’hyperspectral images

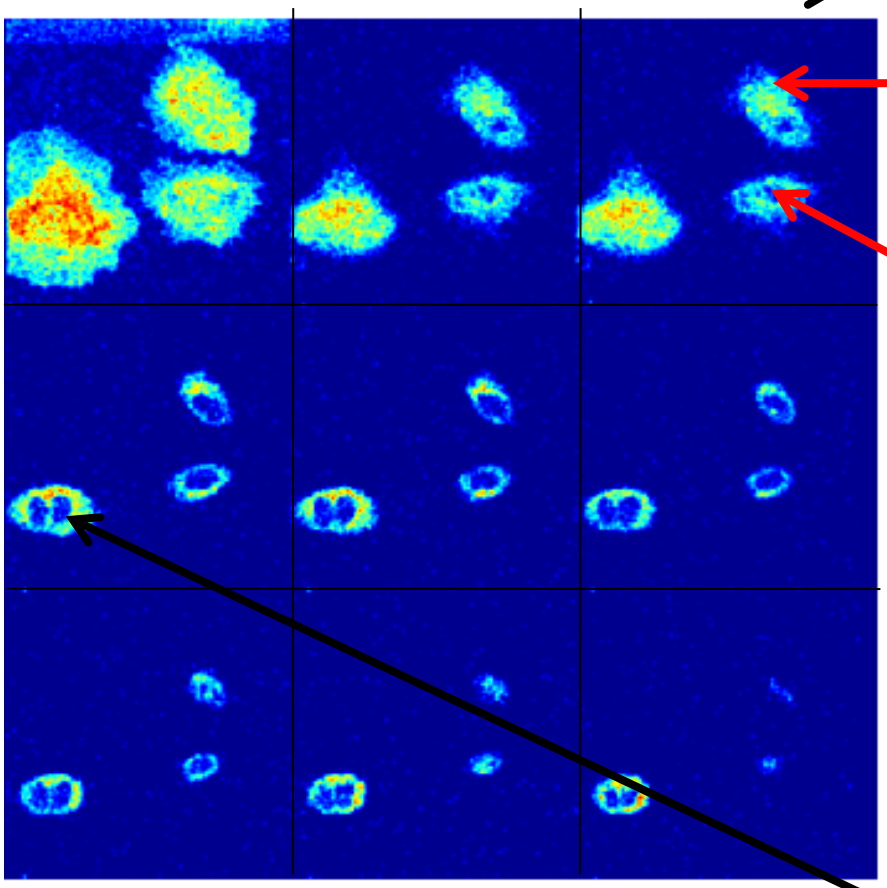
HeLa-M cells

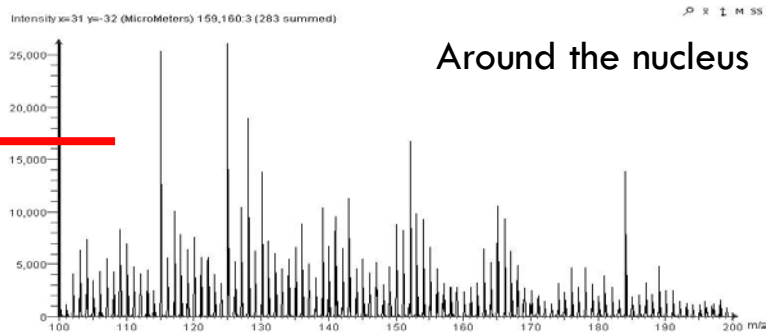
- Immortalised cervical cancer cell line commonly used as a model system for biology
- Cultured on poly-L-lysine coated substrates
- AFM after freeze drying:
  30 – 50 μm diameter
  0.6 – 1μm thick

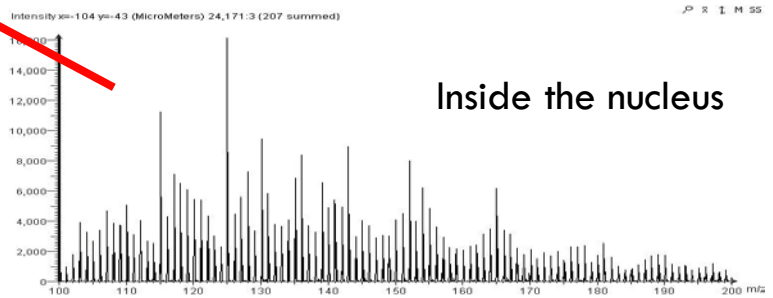


Formalin fixed, freeze-dried HeLa cells

Lipid (DPPC) increasing ion dose







- 1st layer: evidence of chemistry having spread outside cell area
- 3rd layer: spectra show chemistry still mixed up between nucleus and membrane
- 4th layer: one cell exhibits bimodal nucleus indicating the dividing stage of cell cycle – mitosis

Rapid Commun. Mass Spectrom. **25** (2011) 925-932

3D visualisation – issues

- Problem with point of view of 'observer'
- Mass analyser is normal to sample
- Different sample heights appear to be at same level

- Data size:
  - 256 x 256 pixels x 10 layers x 11400 channels
  - 7.5 billion datapoints = 58 GB
  - >99% of data is zero so only ~5 GB in RAM

3D visualisation – our approach

- PCA of whole dataset (sparse, MATLAB)
- Identify PC that separates substrate from organic overlayer
- Assume substrate is flat
- Slide columns of pixels to align with crossover in score of that PC
- Discard substrate-rich pixels
- Repeat PCA to identify peaks or components
- Render results (AVS Express)

Rapid Commun. Mass Spectrom. **25** (2011) 925-932

Two sample preparation methods

Aim is to maintain the biological integrity of cells in vacuum system

Formalin fixed, freeze-dried cells

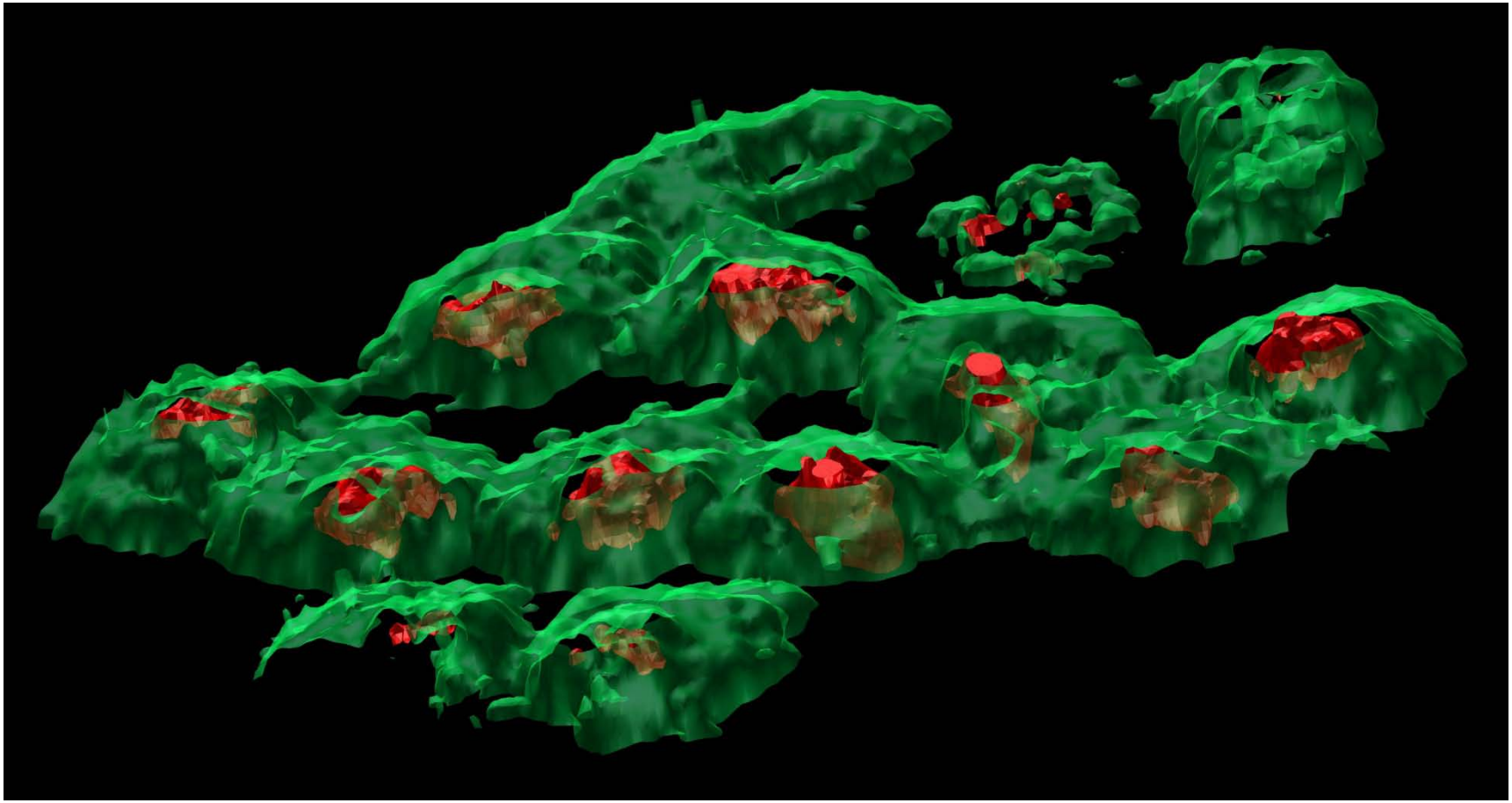
- Hela-M cells cultured on poly-L-lysine coated silicon wafer
- Formalin fixed (4%)
- Cells are washed with ammonium formate (0.15 M) prior to analysis to remove salts
- Freeze-dried in vacuum

Frozen hydrated cells

- Hela-M cells cultured on one plate of hinged two-plate steel substrate coated in poly-L-lysine
- Cells are washed with ammonium formate (0.15M) to remove salts
- Cells trapped between plates and rapidly frozen in liquid propane
- Sandwich fractured in vacuum at 160 K to reveal cells

Rapid Commun. Mass Spectrom. **25** (2011) 925-932

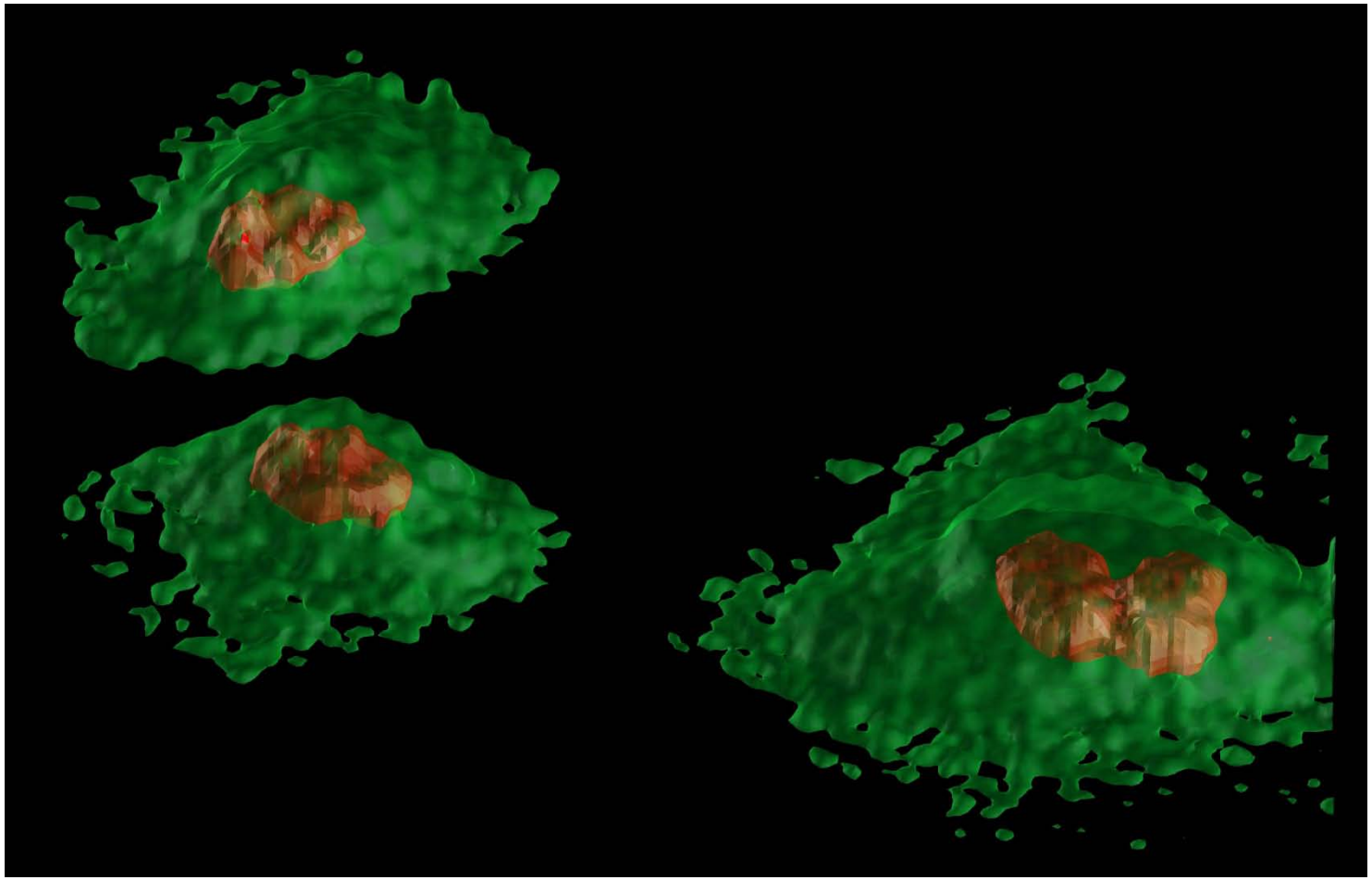
Frozen hydrated HeLa cells



Lipid (DPPC headgroup) @ m/z 184.1 **Nucleic acid (adenine) @ m/z 136.1**

Rapid Commun. Mass Spectrom. **25** (2011) 925-932

Formalin fixed, freeze-dried HeLa cells



Lipid (DPPC) @ m/z 184.1

Principal component indicating guanine & tryptophan

Rapid Commun. Mass Spectrom. **25** (2011) 925-932

Movies available online

256 x 256 pix
x 10 layers
x 11400 chans

7.5 billion pts
= 58 GB

>99% of data is zero so only ~5 GB in RAM

- Frozen hydrated HeLa cells
  - http://www.youtube.com/watch?v=sPuMYUQPvHQ
  - Visualisation of the membrane (green, m/z 184.1, phosphocholine) and nucleus (red, m/z 136.1, adenine)
- Formalin fixed, freeze-dried HeLa cells
  - http://www.youtube.com/watch?v=2phMkYo7yhg
  - Visualisation of the membrane (green, m/z 184.1, phosphocholine) and nucleus (purple, negative scores on principal component 5)
  - Possible mitosis visible at lower right

Summary

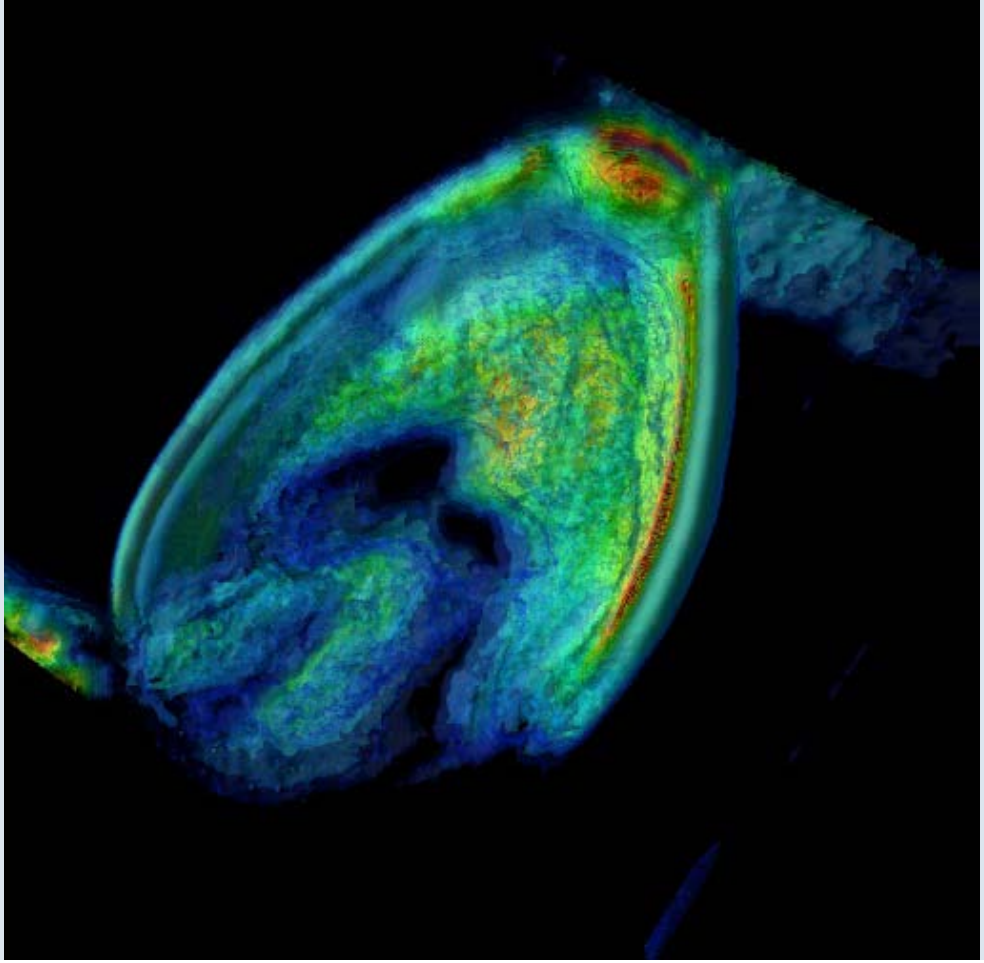
- What SIMS lacks in ultimate species identification it makes up for in spatial location
- New ionisation sources produce data nearer to ‘traditional’ MS, opening up resources
- Tandem MS approaches are breaking through

Acknowledgements

- SARC
  - John Vickerman
  - Nick Lockyer
  - John Fletcher
  - Sadia Rabbani (now Sheraz)
  - ... and the rest!
- Validation algorithms
  - Roy Goodacre
  - Yun Xu
  - Elon Correa
- Visualisation group in Research Computer Services

EPSRC
Engineering and Physical Sciences Research Council
BBSRC
bioscience for the future

www.sarc.manchester.ac.uk



www.sarc.manchester.ac.uk

alex.henderson@manchester.ac.uk