Analytical methods for heterogeneous bio-drugs

Tim Rudd, Marco Guerrini and *Edwin Yates*

Istituto di Ricerche Chimiche e Biochimiche *"G. Ronzoni"*

UNIVERSITY OF LIVERPOOL

Overview

- Chemometrics

- Glycosaminoglycans/heparin

1. Principal component analysis of heparin 1H NMR spectra

2. Two-dimensional correlation spectroscopy and two-dimensional correlation spectroscopic filtering of heparin 1H NMR spectra

Chemometrics

- Chemometrics is the science of extracting information from chemical systems by data-driven means - Wikipedia.

- **Extracting patterns from complicated data sets.**

- Adding value to data.

- **Transforming data in to information.**

- R - http://www.R-project.org/

Glycosaminoglycans

• Heparin / Heparan Sulfate

Involved in:

•Development

•Diseases - cancer/neurodegenerative

• Heparin is the largest consumed pharmaceutical product in terms of mass, 400 - 700 million pigs per year!

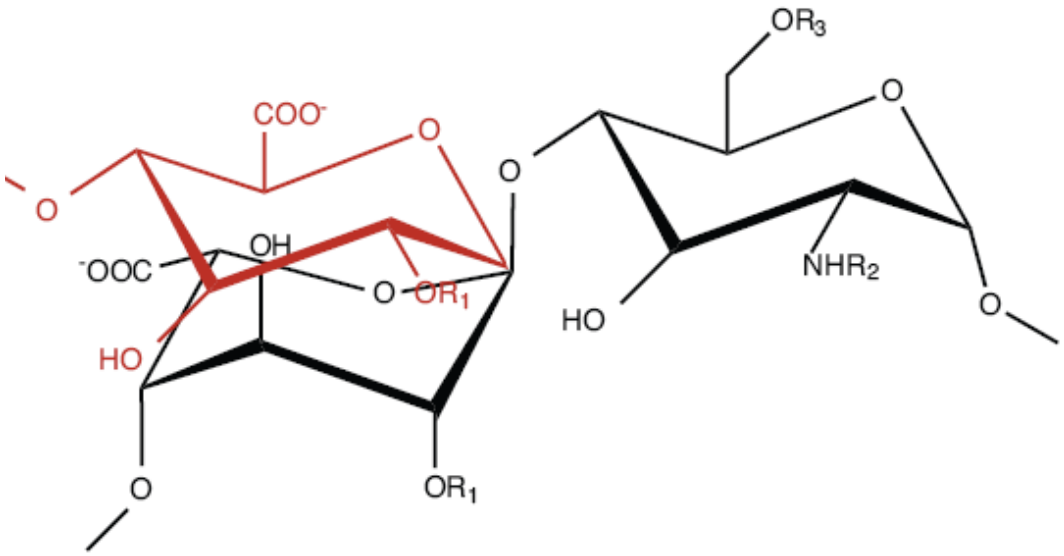
• The product has evolved, heparin has been depolymerised to low molecular weight heparins

• Its role is governed by protein-interactions; ~ 300 known at present - 1000s predicted.

- Conformation

- Sequence

-Cation



1. PCA of heparin 1H NMR spectra

A library of heparin sample should consist of:

- **Intrinsically diverse samples, covering the heterogeneity of heparin.**

- **It will be composed of what is considered to be "heparin" at that time and place. Samples should agree with the regulators requirements.**

- **As the library grows the definition of heparin will change.**

- **There will never be a "correct", "complete" and "final" library - it will always be in flux.**

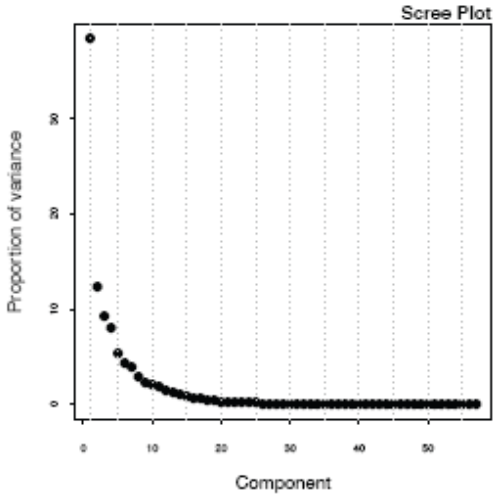
1. PCA of heparin 1H NMR spectra

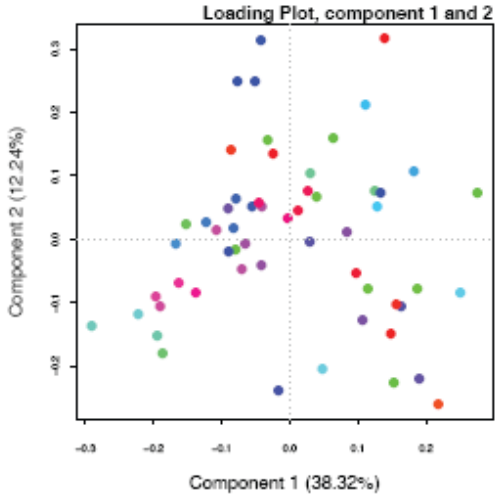
Principal component analysis

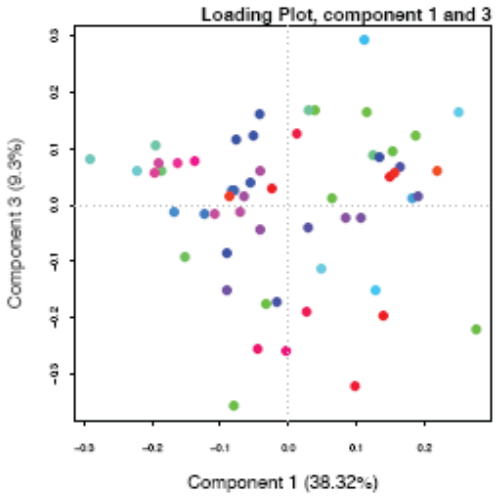
- Explorative multivariate technique - PCA can be used to explore the heparin library.
- Generates a number of independent features (components), that explains the variation within the dataset.
- PCA reduces a dataset to a lower number of components, greatly simplifying it.
- A linear combination of components can be used to reconstruct any member of the dataset used in the analysis.
- PCA can be used to find outliers within a dataset - heparin samples contaminated with OSCS?

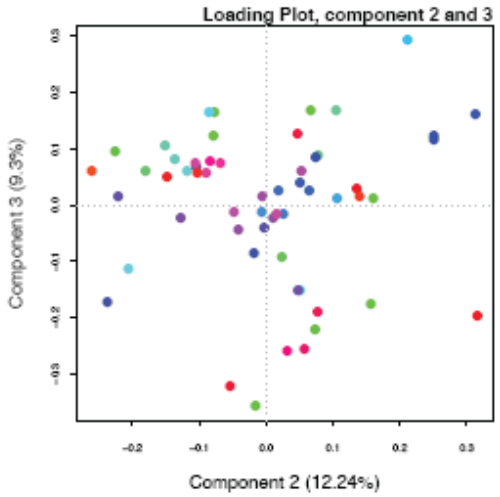
An example - A library of bona fide heparin samples.

1. PCA of heparin 1H NMR spectra









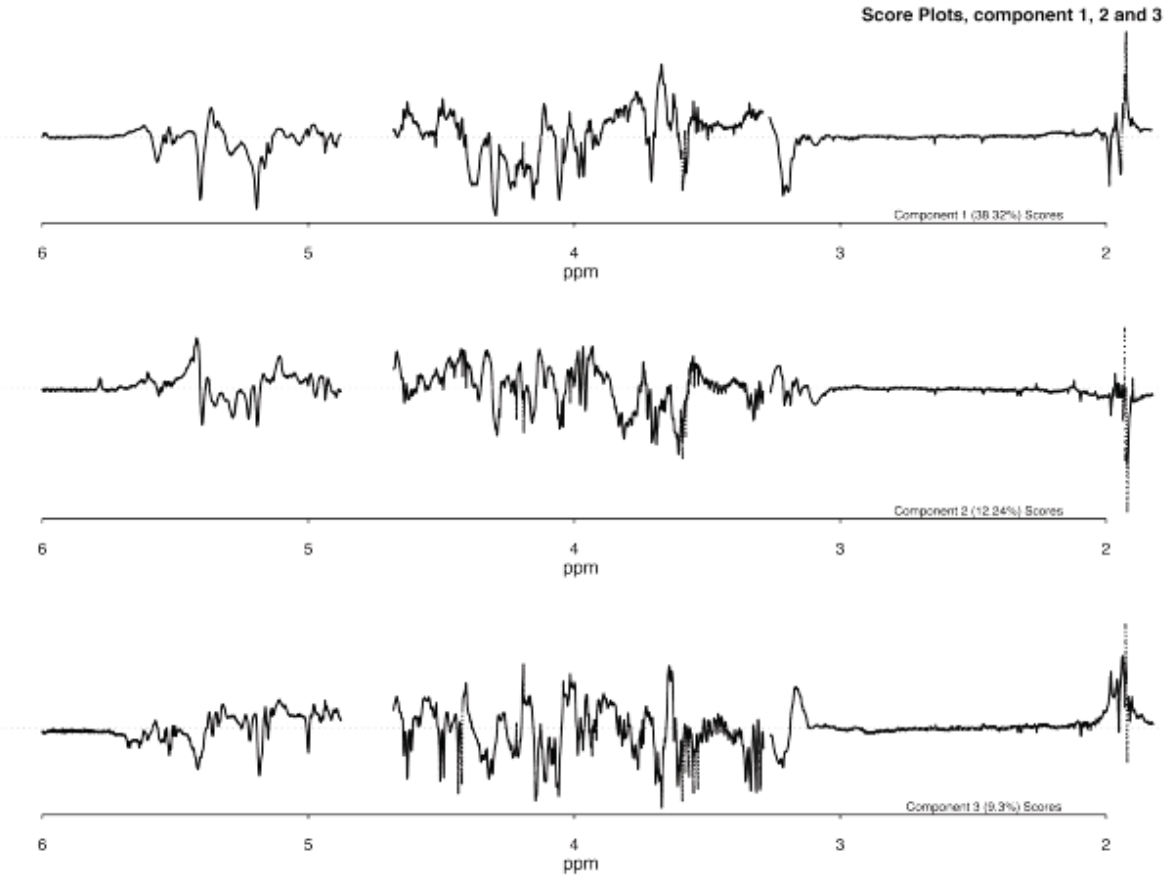
- Library of 57 *bona fide* heparin samples, which comply with EDQM regulations.

- 4 components explain over 70% of the variation within the dataset.

R:- rNMR package used to import Bruker NMR spectra directly. Data is normalised for area, mean centered and Pareto scaled. PCA is performed using the princomp function.

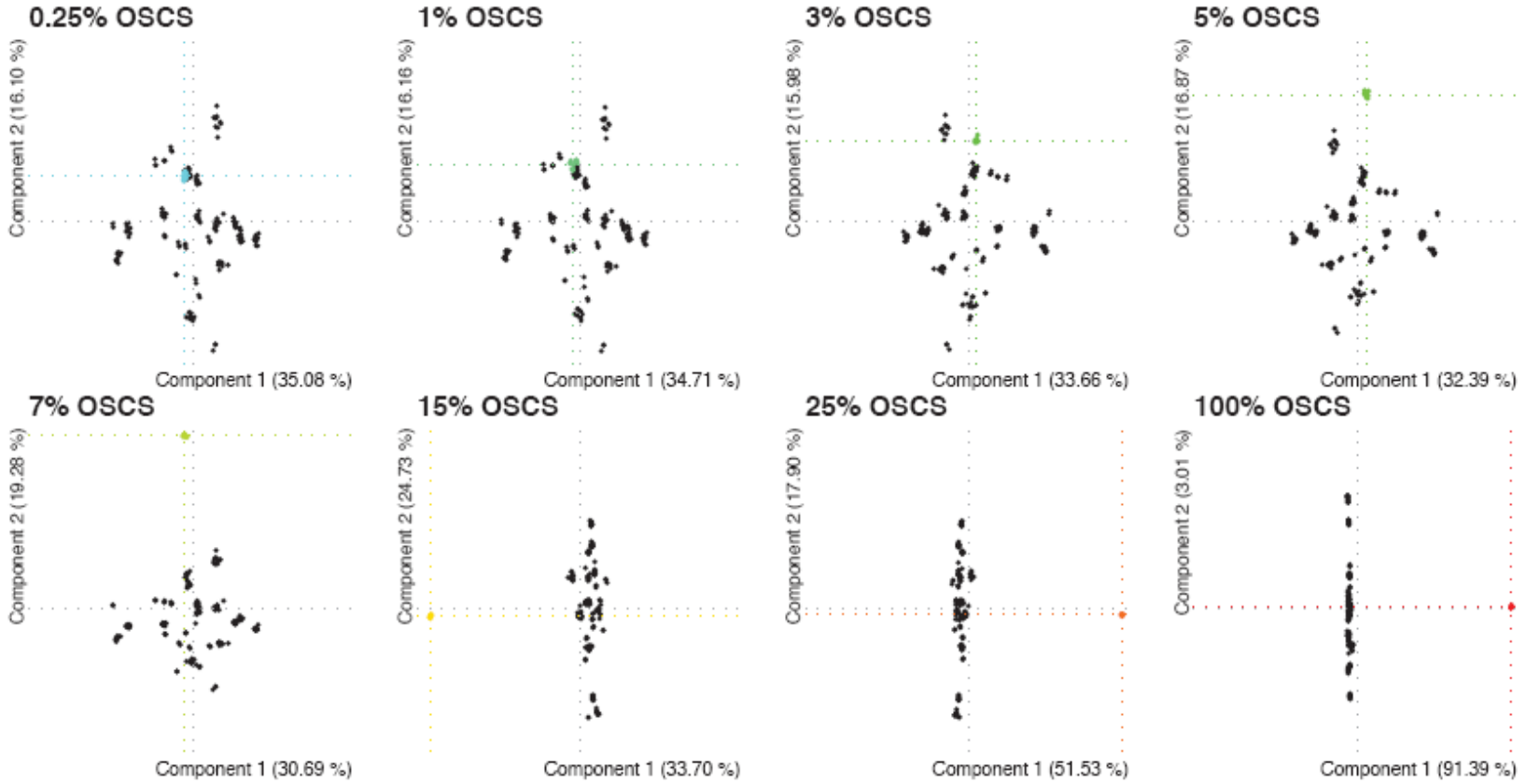
1. PCA of heparin 1H NMR spectra

- Scores - spectral features extracted by PCA.



1. PCA of heparin 1H NMR spectra

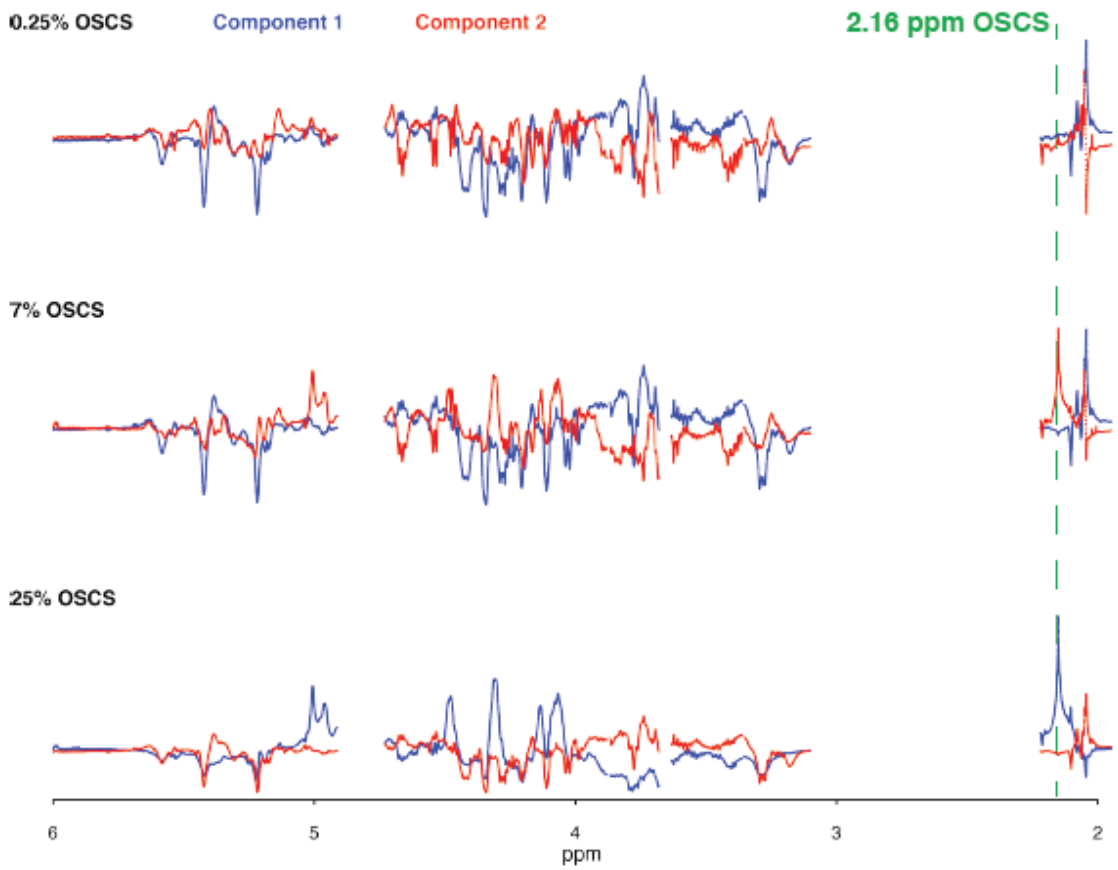
- Heparin contaminated with OSCS.



- Heparin contaminated with 7% OSCS can be clearly differentiated .

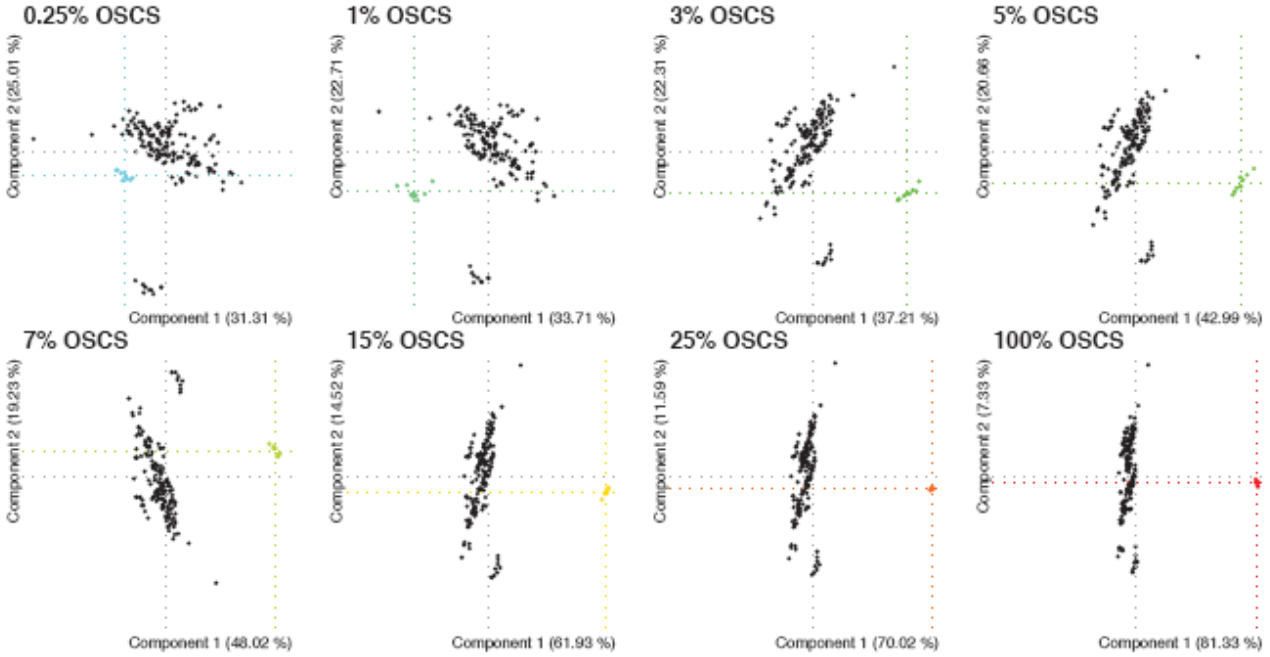
1. PCA of heparin 1H NMR spectra

- N-acetyl signal of OSCS at 2.16 ppm is the primary diagnostic signal.



1. PCA of heparin 1H NMR spectra

- **OSCS has a *free signal at 2.16 ppm. By selecting these data,* sensitivity can be improved - heparin contaminated with 1% OSCS is differentiable.**



Rudd TR, *et al.* (2011) **Construction and use of a library of bona fide heparins employing ^{1}H NMR and multivariate analysis**, Analyst, 136, 1380.

1. PCA of heparin 1H NMR spectra

Conclusions

- **“Heparin” has no conventional definition, but a statistical one.**
- **The pragmatic definition of heparin takes the form of a library of 1H NMR spectra.**
- **PCA can be used to explore the heparin library/definition.**
- **A library of heparin 1H NMR spectra can be used to differentiate a heparin sample contaminated with 7% OSCS - full spectrum.**
- **Selecting the spectral region can increase the limit of detection for OSCS, with samples containing 1% OSCS differentiable - N-acetyl region, OSCS signal at 2.16 ppm.**

2. 2D-COS/2D-COS-f

Two-dimensional correlation spectroscopy (2D-COS):

Orthodox 2D-COS is the correlation of a set of linearly perturbed spectra, the sample can be mechanically or chemically perturbed.

Noda *et al.* 2000 Appl. Spectrosc.

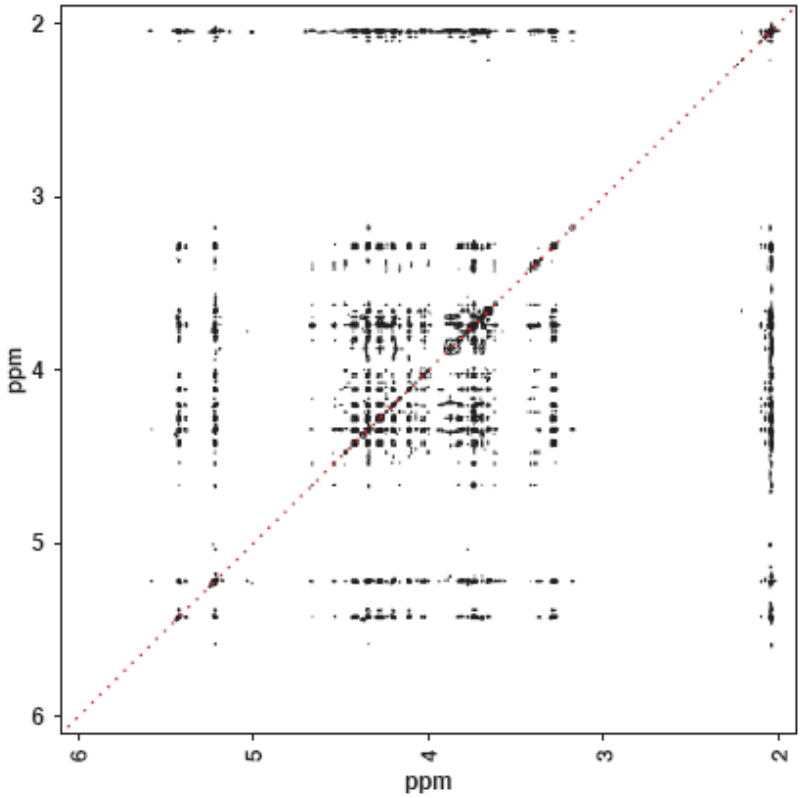
$$\text{2D-COS spectrum} = \text{Covariance matrix} = 1/(n-1)XX^{T}$$

Linear algebra formalisation Šašic *et al. 2000 J. Phys. Chem. B*

A variation of this idea has been used for Metabolomics, instead of using a covariance matrix a correlation matrix is used, (STOCSY - Cloreac *et al.* 2005 Anal. Chem.).

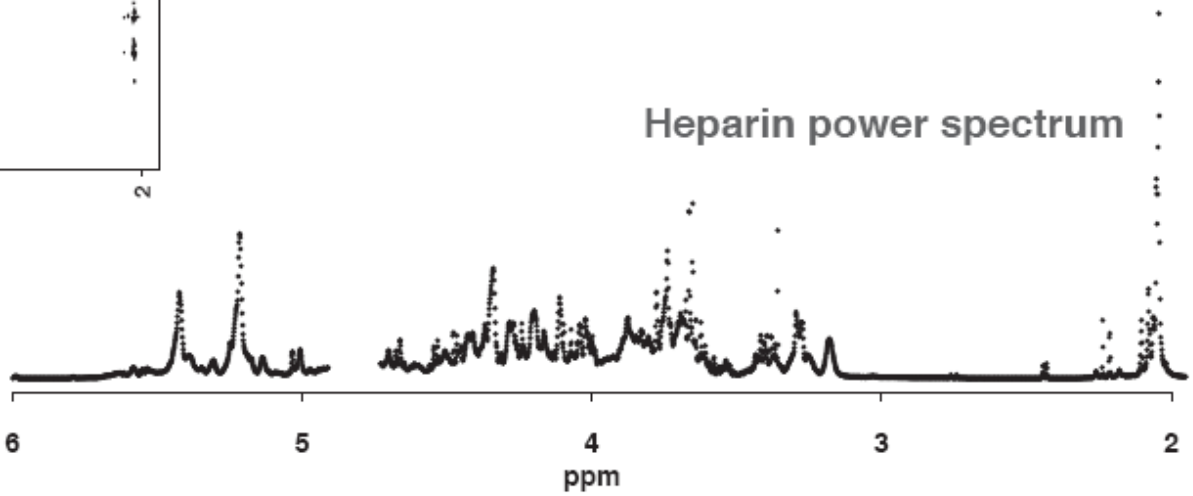
Pre-data processing is VERY important, the data is normally mean centred and then scaled (Pareto or auto scaling), data can also be normalised for area.

2. 2D-COS/2D-COS-f

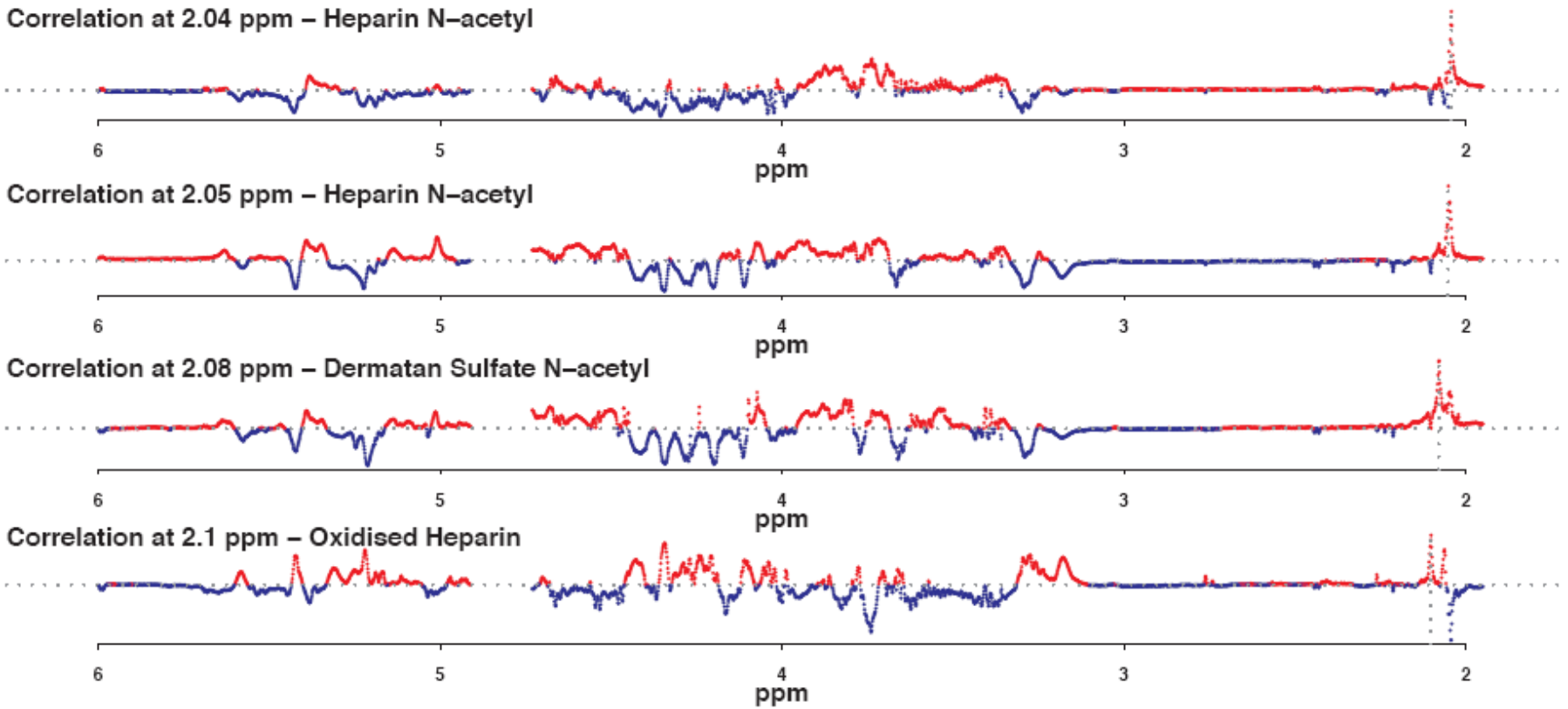


2D-COS of heparin library

- A psuedo-TOCSY spectrum is produced
- **Diagonal** - Variances of the covariance matrix - power spectrum.
- **Off-diagonal** - Covariances - cross peak correlations.



2. 2D-COS/2D-COS-f



Reveals library contents

2. 2D-COS/2D-COS-f

2D-COS-Filtering: **Effectively a 2D-COS difference spectrum between the reference heparin library and a test heparin sample.**

2D-COS-F spectrum = $1/(n_{lib+Test\ Sample}-1)(X_{lib+Test\ Sample}\ X_{lib+Test\ Sample}^{T}) - 1/(n_{lib}-1)(X_{lib}X_{lib}^{T})$

- **Auto-correlations between signals from the heparin samples in the reference library and heparin test sample.**
- **Subtraction removes correlated heparin signals, leaving 'alien' features.**

Library

- **The library is a reference definition for the conventionally undefinable heparin.**
- **The library encompasses the natural variation within pharmaceutical heparin.**

2. 2D-COS/2D-COS-f

2D-COS-Filtering Procedure:

1. Measure ^{1}H NMR spectra - the same conditions!

2. Import data in to R [no other programmes are available... ;-)]: http://www.R-project.org/.

3. Cut data - water signal, other signals can be removed - solvents.

4. Normalise for area.

5. Mean centre.

6. Pareto Scale, scale to the square root of the standard deviation. van der Berg *et al.* 2006 BMC Genomics - Good scaling reference.

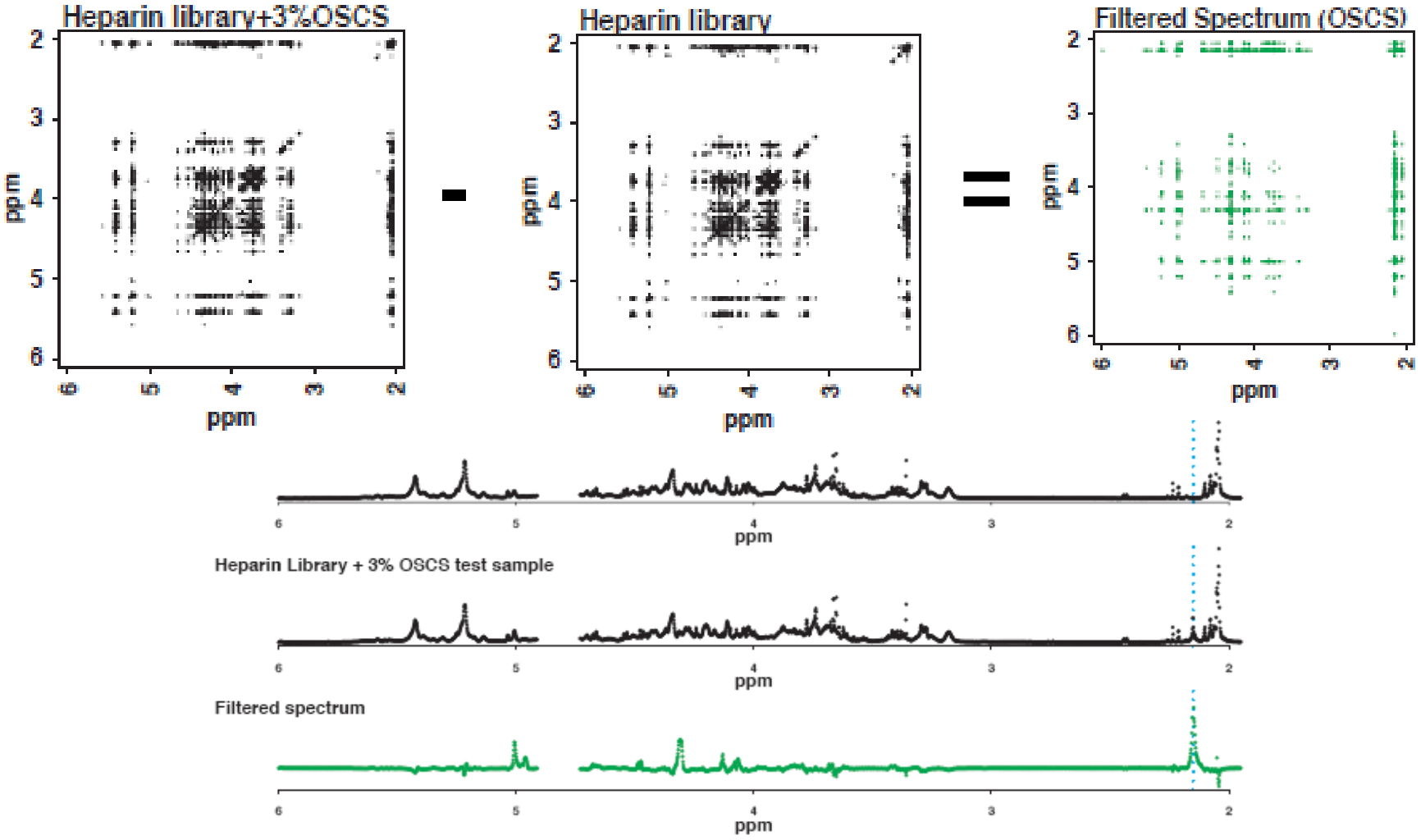
7. Determine the covariance matrix of the reference heparin library [A].

8. Determine the covariance matrix of the reference heparin library with the test sample added [B].

9. B-A - plot the data...

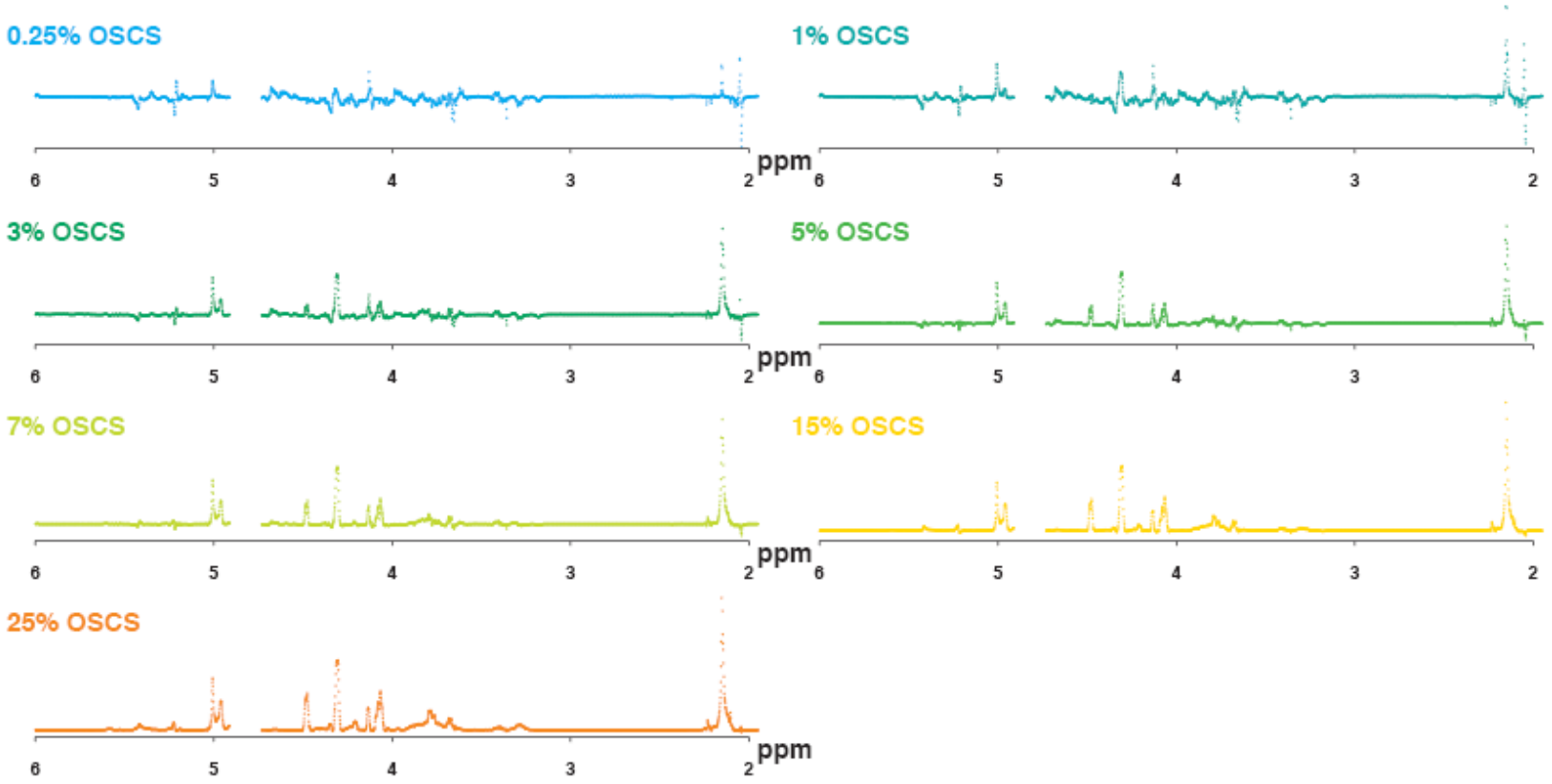
2. 2D-COS/2D-COS-f

OSCS contamination of heparin



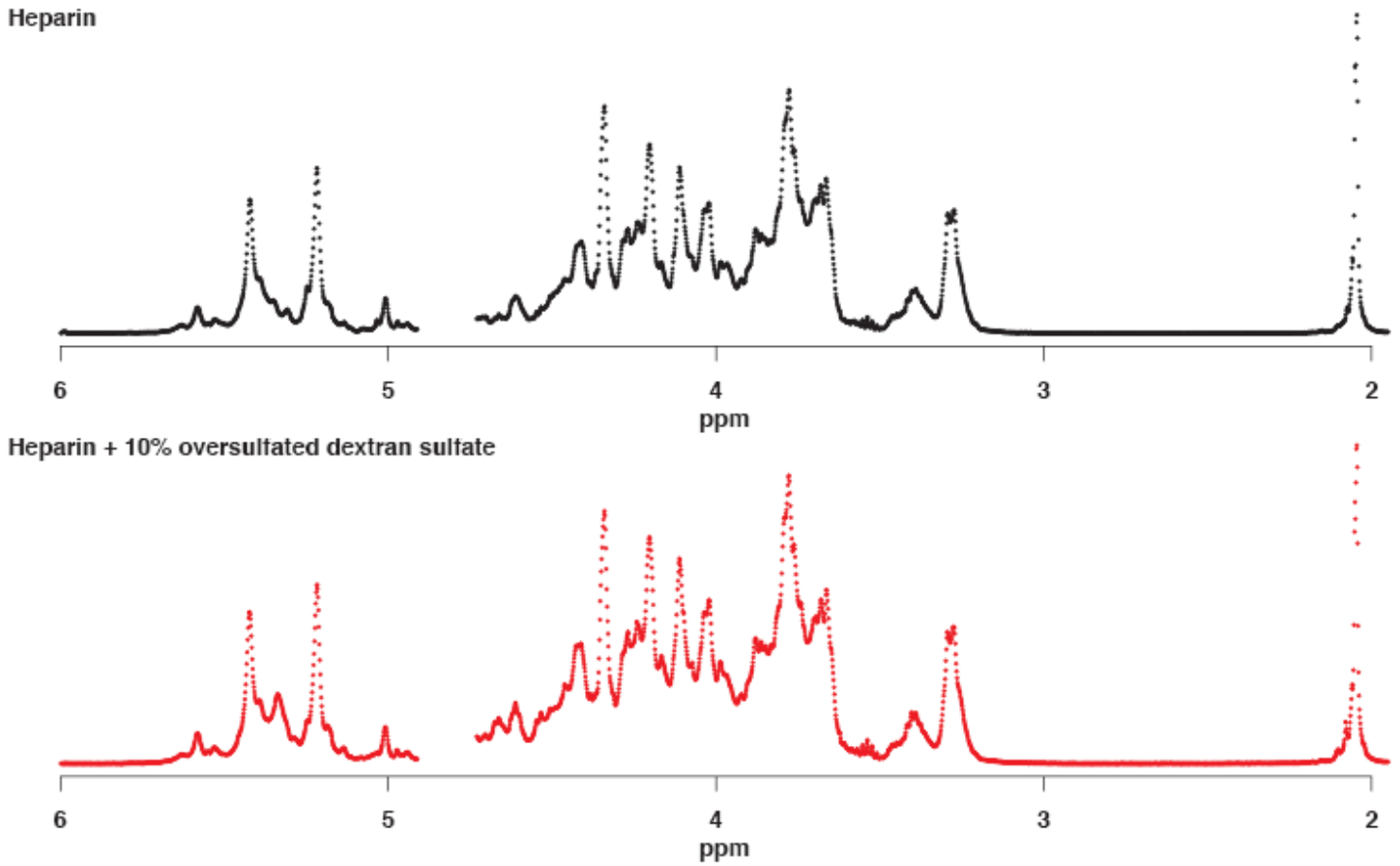
2. 2D-COS/2D-COS-f

OSCS contamination of heparin

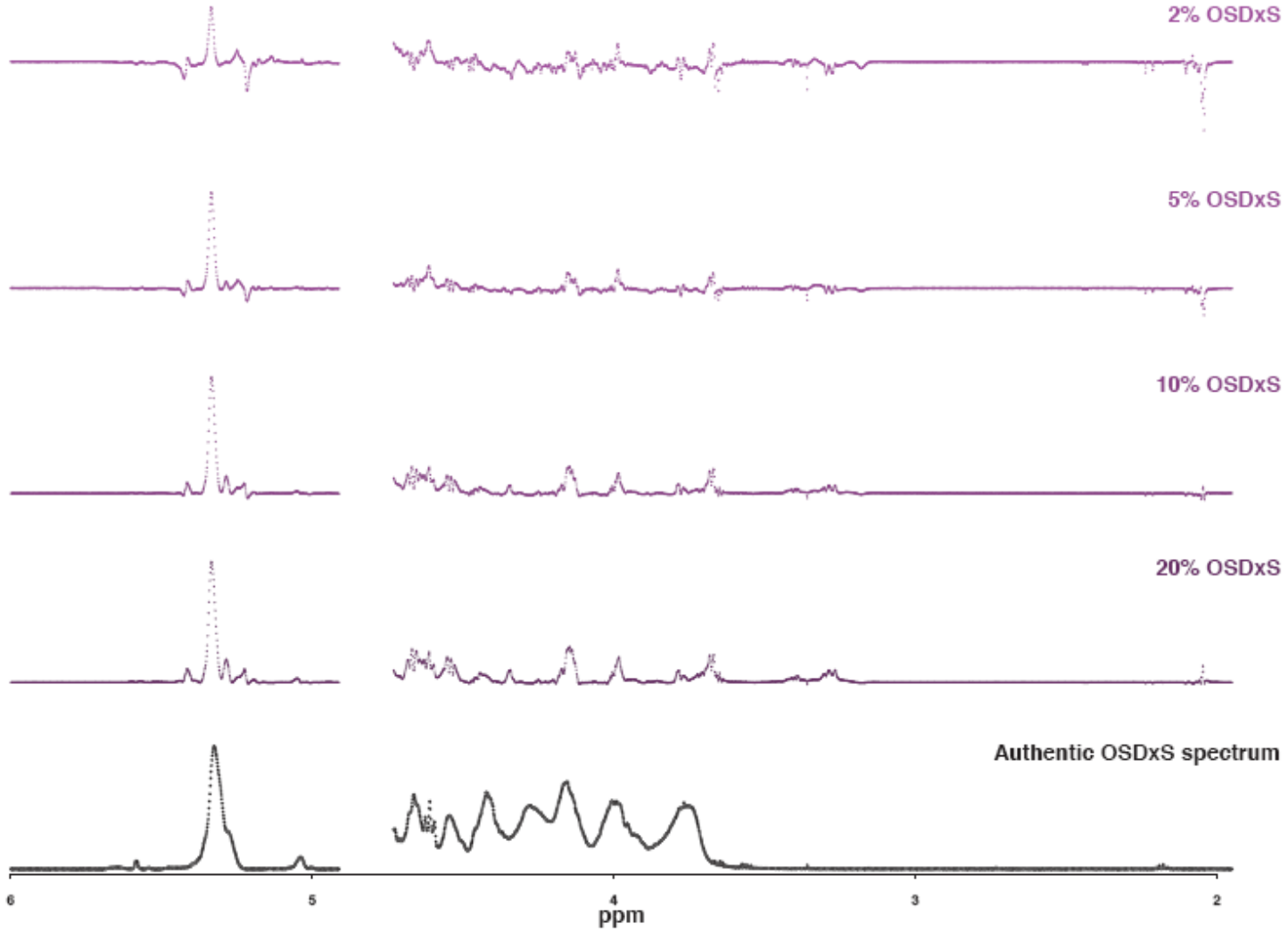


2. 2D-COS/2D-COS-f

- Oversulfated Dextran Sulfate - example of an economically relevant contaminant
- No N-acetyl signals - all other signals are within heparin signals



2. 2D-COS/2D-COS-f



2. 2D-COS/2D-COS-f

- 2D-COS filtering is a means of visualising signals which are under a complicated spectrum.

- **The technique relies on having a library of reference samples to filter out signals from the test sample.**

- In this circumstance the library is a definition of the polydisperse carbohydrate heparin.

- **2D-COS filtering can detect signals from compounds that do not have N-acetyl signals, which would be hidden under the mass of heparin signals.**

- 2D-COS filtering can be used to visualise OSCS signals from samples contaminated with as little as 0.25 % and 2% oversulfated dextran sulfate.

Final Remarks

- 2D-COS - a means of visualising signals under complicated spectra.
- Employs library approach to define heterogeneous product
- Sensitive detection of samples with separate NMR signals and those without
- Applicable to any heterogeneous compounds, natural products, industrial products etc

Acknowledgements

Tim Rudd - Ronzoni Institute

To whom any technical questions can be addressed:

trudd@liv.ac.uk. eayates@liv.ac.uk

Istituto di Ricerche Chimiche
e Biochimiche *"G. Ronzoni"*

UNIVERSITY OF LIVERPOOL

Acknowledgements

Ronzoni Institute, Italy:

Marco Guerrini
Davide Gaudesi
Stefano Elli
Guilherme Sassaki
Cesare Cosentino
Eleonora Macchi
Christina Gardina
Laura Muzi
Giangiacomo Torri

University of Keele, UK:

Mark Skidmore

Universidade Federal de São Paulo, Brazil:

Marcelo Lima

Istituto di Ricerche Chimiche e Biochimiche *"G. Ronzoni"*

UNIVERSITY OF LIVERPOOL

Any Questions?