

# Big-Data

*et traitement de données en MS*  
*le project MeasureHD*

*Marc-André Delsuc*  
*Prospectom - 21 novembre 2014*

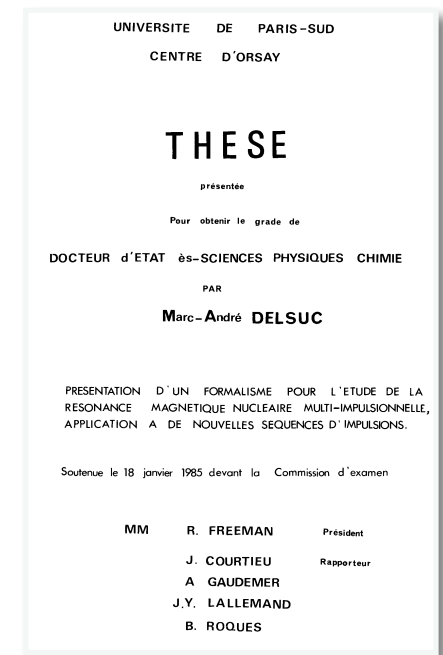


# présentation

- Spectroscopiste RMN
  - thèse sur les principes de Mécanique Quantique de la RMN 2D
  - 2D - 3D - mesure de Diffusion
  - (*en amont de la*) Biologie Structurale
- Labos
  - Gif-sur-Yvette / Montpellier / Strasbourg
- IGBMC
  - Strasbourg / Illkirch
  - 750 personnes / 52 équipes
- traitement
  - développement de plusieurs logiciels d'analyse de données  
Gifa / NPK / NMRNoteBook / NPKV2 / SpiKe
- Sujets d'études actuels:
  - Intrinsically Disordered Proteins / interactions complexes
    - Androgen Receptor / fibres amyloïdes
  - MS 2D-FTICR  
protéomique "top-down", analyse de PTM "exotiques"  
DDA, DIA

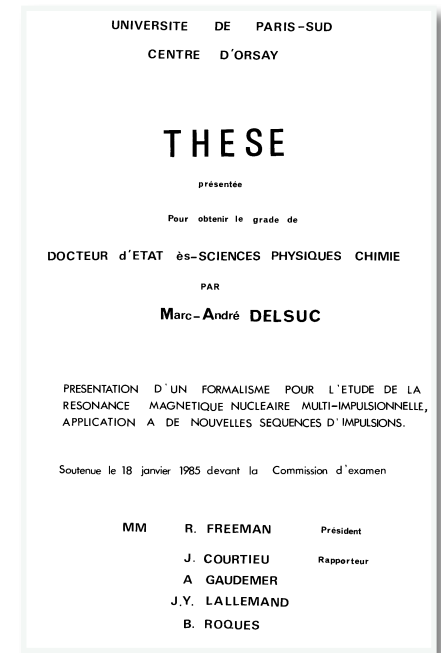
# présentation

- Spectroscopiste RMN
  - thèse sur les principes de Mécanique Quantique de la RMN 2D
  - 2D - 3D - mesure de Diffusion
  - (*en amont de la*) Biologie Structurale
- Labos
  - Gif-sur-Yvette / Montpellier / Strasbourg
- IGBMC
  - Strasbourg / Illkirch
  - 750 personnes / 52 équipes
- traitement
  - développement de plusieurs logiciels d'analyse de données  
Gifa / NPK / NMRNoteBook / NPKV2 / SpiKe
- Sujets d'études actuels:
  - Intrinsically Disordered Proteins / interactions complexes
    - Androgen Receptor / fibres amyloïdes
  - MS 2D-FTICR  
protéomique "top-down", analyse de PTM "exotiques"  
DDA, DIA



# présentation

- Spectroscopiste RMN
  - thèse sur les principes de Mécanique Quantique de la RMN 2D
  - 2D - 3D - mesure de Diffusion
  - (*en amont de la*) Biologie Structurale
- Labos
  - Gif-sur-Yvette / Montpellier / Strasbourg
- IGBMC
  - Strasbourg / Illkirch
  - 750 personnes / 52 équipes
- traitement
  - développement de plusieurs logiciels d'analyse de données  
Gifa / NPK / NMRNoteBook / NPKV2 / SpiKe
- Sujets d'études actuels:
  - Intrinsically Disordered Proteins / interactions complexes
    - Androgen Receptor / fibres amyloïdes
  - MS 2D-FTICR
    - protéomique "top-down", analyse de PTM "exotiques"
    - DDA, DIA



# Plan

- Mastodons
  - BigData
- Calcul et BigData en (FT)-MS
  - acquisition
    - optimisation - accélération
  - traitement de données
    - analyse “classique”
    - débruitage - approches low-rank - Compressed sensing
- Spectrométrie de masse
  - 2D FT-ICR MS
- Réseau

# Mastodons

- CNRS - Mission interdisciplinarité
  - Idée
    - ▶ utiliser les BigData pour structurer des projets “inter-disciplinaires”
  - inter-disciplinaires
    - ▶ généralement - informaticiens + autre communauté scientifique



**Mission interdisciplinarité**  
Centre national de la recherche scientifique



# Mastodons

- CNRS - Mission interdisciplinarité
  - Idée
    - ▶ utiliser les BigData pour structurer des projets “inter-disciplinaires”
  - inter-disciplinaires
    - ▶ généralement - informaticiens + autre communauté scientifique
- qu'est-ce que le Big-Data ?
  - taille des données
  - approches particulières
    - algorithmes linéaire / quasi-linéaire
      - ▶ Machine Learning
  - techniques informatiques particulières
    - ▶ requêtes sur des serveurs distants
    - ▶ calcul “dans les nuages”
  - ... ?



**Mission interdisciplinarité**  
Centre national de la recherche scientifique



# Mastodons

- CNRS - Mission interdisciplinarité

- Idée

- ▶ utiliser les BigData pour structurer des projets “inter-disciplinaires”

- inter-disciplinaires

- ▶ généralement - informaticiens + autre communauté scientifique

- qu'est-ce que le Big-Data ?

- taille des données

- approches particulières

- algorithmes linéaire / quasi-linéaire

- ▶ Machine Learning

- techniques informatiques particulières

- ▶ requêtes sur des serveurs distants

- ▶ calcul “dans les nuages”

- ... ?

- Pour moi - *et pour d'autres* -

- quand la taille des données exclu les approches classiques



**Mission interdisciplinarité**  
Centre national de la recherche scientifique





# Plusieurs types de *Big Data*

- Grandes bases de données
  - Google (NSA)
  - données : textuelles, irrégulières, nombreuses erreurs
  - analyse :
    - ▶ rechercher compter trier
    - ▶ recherche de corrélation, d'out-liers
    - ▶ Machine Learning
  - stratégies
    - ▶ systèmes distribués - map-reduce
    - ▶ cloud - hadoop
- Grandes mesures physiques
  - ▶ LHC : 40.000 Go / jour = 40 To / jour
  - ▶ Petasky : 3 Milliard de pixels / 17 secondes
  - données : numériques, régulières, peu d'erreur
  - besoins : recherche de motif / mesure
  - stratégies
    - ▶ adapter la puissance aux besoins
    - ▶ compresser / calcul en flux / calcul centralisé

# Plusieurs types de *Big Data*

- Grandes bases de données (Google)

- PDB ~100.000 structures
- Uniprot ~1M protéines
- Genome ~1000 organismes



- Grandes mesures physiques (LHC)

- BioMagResBank
  - ▶ données de RMN
- PRIDE/EBI
  - ▶ Proteomics : 38.000 assays
- COSMIC séquençage de tumeurs
  - ▶ ~1M échantillons ~2M mutations



### Statistics

Genes	28977
Samples	1058292
Coding Mutations	2710449
Papers	20247
Unique Variants	2139424

Fusions	10567
Genomic Rearrangements	61232
Whole Genomes	15047
Copy Number	702652
Gene Expression	118886698



Science

Daily News

Enter Search Term



ADVANCED

SUBSCRIBE

[News Home](#)[Hot Topics](#)[Categories](#)[From the Magazine](#)[ScienceInsider](#)[ScienceShots](#)[Sifter](#)[About Us](#)

NEWS

SCIENCE JOURNALS

CAREERS

MULTIMEDIA

COLLECTIONS

[News](#) > [Europe](#) > [E.U. Medicines Agency hands victory to advocates of trial data sharing](#)

## SCIENCEINSIDER

# E.U. Medicines Agency hands victory to advocates of trial data sharing

[Tweet](#) 484[g+](#) 3[Email Tania](#)[Follow @taniarabes](#)By [Tania Rabesandratana](#) | 3 October 2014 12:00 pm | [Comments](#)

After an 18-month saga, the European Medicines Agency (EMA) has approved the details of a new system allowing researchers to scrutinize unpublished data from clinical trials. Scientists and transparency campaigners have praised the decision as a big step forward but have voiced lingering concerns, in particular about information that could be redacted before clinical study reports are shared.

EMA is the "first entity in the world" to introduce such rules, Executive Director Guido Rasi pointed out at an event in the European Parliament earlier this week, adding that the agency is "setting new standards for transparency." The plan, which was approved yesterday at a managing board meeting in London, "represents a real shift in favour of ensuring research data is shared routinely and re-used effectively in the public interest," agrees Carl Heneghan, director of the Centre for Evidence Based Medicine at the University of Oxford

# Rencontre d'un troisième type

- Nombreuses mesures délocalisées
  - Exemples:
    - ▶ les caméras de surveillance
    - ▶ Les albums photo dans les familles françaises
- Caractéristiques:
  - stockage distribué
    - il n'y a pas de "*base de données*"
  - Analyse au vol, impossibilité de garder à long terme toutes les données
    - ▶ Il faut traiter tout, tout de suite
  - Traitement des données en local
  - difficulté de traitement optimal des données
  - Nécessité de traitements automatiques
    - ▶ robustes
    - ▶ autonomes
- Nécessité de développements méthodologiques

# Rencontre d'un troisième type

- En Biologie aussi
  - démocratisation des analyses biophysiques systématiques
    - ▶ Protéomiques / Métabolomique
    - ▶ MS, RMN,
    - ▶ flux d'échantillons et de données
  - médecine personnalisée
- Caractéristiques:
  - Analyse au vol, impossibilité de garder à long terme toutes les données
  - Difficulté de traitement optimal des données
  - Nécessité de traitements automatiques
    - ▶ robustes
    - ▶ “unattended”

# Les big data - Solutions de traitement

- Analyse
  - données : numériques, régulières, très peu d'erreur
  - besoins : recherche de motif / mesure
  - stratégies : compresser / calcul en flux / calcul décentralisé
- Utilisation du cloud
  - limité par les taux de transferts
  - confidentialité ?
- Développement de nouveaux algorithmes
  - permettre des analyses sans compromis sur la qualité de l'analyse et du traitement de données
  - adapter les protocoles de mesure et de traitement pour optimiser la chaîne de mesure
  - permettre le déploiement de cette approche chez les utilisateurs pour un traitement en temps réel
  - application à plusieurs secteurs industriels

**From:** Clinica Chimica Acta journals@mail.elsevier.com  
**Subject:** Rejoignez nous pour un webinar gratuit sur les applications et tendances de la spectrométrie de masse en laboratoire d'analyses médicales  
**Date:** 19 Nov 2014 16:54  
**To:** madelsuc@unistra.fr

# ELSEVIER

## Webinar

### Webinar à venir: Spectrométrie de masse en laboratoire d'analyses médicales: applications et tendances

**Date:** lundi 15 décembre 2014  
**Heure:** 15h00 (Paris/Bruxelles/Genève)

Joignez-vous à notre webinar gratuit. [Inscrivez-vous dès maintenant](#)

Cher Marc-André Delsuc,

[Rejoignez nous pour un webinar gratuit sur les applications et tendances de la spectrométrie de masse en laboratoire d'analyses médicales.](#)

Les laboratoires médicaux ont de plus en plus souvent à répondre à des demandes d'analyses spécifiques et de haute sensibilité, dans des délais courts. Dans le domaine du suivi thérapeutique pharmacologique par exemple, de nouvelles techniques d'analyse doivent être mises au point et validées chaque année pour de nouveaux médicaments, souvent pour ne devoir analyser qu'un nombre réduit de prélèvements. En toxicologie médicale ou médico-légale, l'amélioration des techniques de screening contribue à l'obtention de résultats plus pertinents et significatifs. Dans le domaine des biomarqueurs en protéomique et métabolomique, la standardisation des techniques devrait contribuer à la dissémination des innovations diagnostiques. Ce webinar couvrira ces différents aspects.

En association avec



# En protéomique par MS :

- Flux de données croissant

- appareils de plus en plus performants et automatisation des mesures

- LC-MS haute résolution tournant 24/24h

- ▶ Orbitrap : mesure à ~10 MHz, conversions sur 32 bits

format brut

40 Mb/sec    144 Gb/heure    3.5 Tb/jour

données analysées (analyse de Fourier, débruitage, peak-picking )

1 Mb/sec    4 Gb/heure    10 Gb/jour

- ▶ Remarque

ThermoFisher ne vous “*donne*” pas les formats bruts de données

- Limites

- exploitation en mode local impossible car saturation des capacités standards de traitement
- traitements classiques lourds et peu compatibles avec le volume de données



# En protéomique par MS :

- Flux de données croissant

- appareils de plus en plus performants et automatisation des mesures

- LC-MS haute résolution tournant 24/24h

- ▶ Orbitrap : mesure à ~10 MHz, conversions sur 32 bits

format brut

40 Mb/sec    144 Gb/heure    3.5 Tb/jour

données analysées (analyse de Fourier, débruitage, peak-picking )

1 Mb/sec    4 Gb/heure    10 Gb/jour

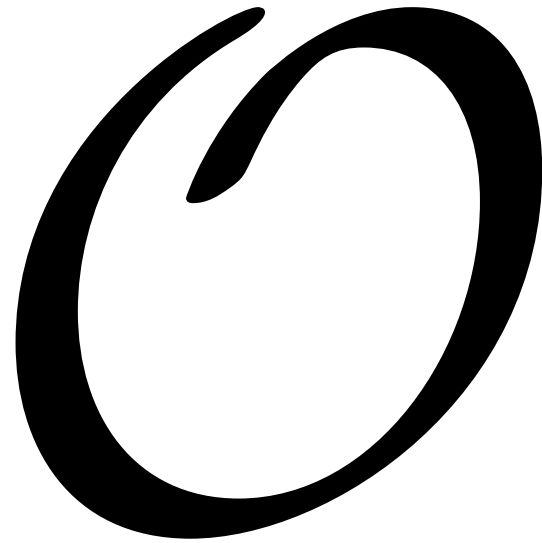
- ▶ Remarque

ThermoFisher ne vous “*donne*” pas les formats bruts de données

A red rectangular box with the word "BigData" written in white, sans-serif font. The box has a slight shadow and is positioned on the right side of the slide.

- Limites

- exploitation en mode local impossible car saturation des capacités standards de traitement
- traitements classiques lourds et peu compatibles avec le volume de données



# Spectrométrie de Masse

# Spectrométrie de Masse

- système physique

# Spectrométrie de Masse

- système physique
- signal électronique mesuré

# Spectrométrie de Masse

- système physique
- signal électronique mesuré
- données brutes

# Spectrométrie de Masse

- système physique
- signal électronique mesuré
- données brutes
- spectral processing

# Spectrométrie de Masse

- système physique
- signal électronique mesuré
- données brutes
- spectral processing
- pic / massif isotopique / quantification



# Spectrométrie de Masse

- système physique
- signal électronique mesuré
- données brutes
- spectral processing
- pic / massif isotopique / quantification
- PSM

# Spectrométrie de Masse

- système physique
- signal électronique mesuré
- données brutes
- spectral processing
- pic / massif isotopique / quantification
- PSM
- protéines identifiées / quantification

# Spectrométrie de Masse

- système physique
- signal électronique mesuré
- données brutes
- spectral processing
- pic / massif isotopique / quantification
- PSM
- protéines identifiées / quantification

signal  
TOF - Orbi

liste de pics

liste de PSM

liste de protéine

volcano plot

# Spectrométrie de Masse

- système physique
- signal électronique mesuré
- données brutes
- spectral processing
- pic / massif isotopique / quantification
- PSM
- protéines identifiées / quantification

ions

signal  
TOF - Orbi

liste de pics

liste de PSM

liste de protéine

volcano plot

# Spectrométrie de Masse

MS

MS/MS

- système physique
- signal électronique mesuré
- données brutes
- spectral processing
- pic / massif isotopique / quantification
- PSM
- protéines identifiées / quantification

ions

signal  
TOF - Orbi

liste de pics

liste de PSM

liste de protéine

volcano plot

# Spectrométrie de Masse

MS

MS/MS

- système physique
- signal électronique mesuré
- données brutes
- spectral processing
- pic / massif isotopique / quantification
- PSM
- protéines identifiées / quantification

ions

signal  
TOF - Orbi

liste de pics

liste de PSM

liste de protéine

volcano plot

# Spectrométrie de Masse

- système physique
- signal électronique mesuré
- données brutes
- spectral processing
- pic / massif isotopique / quantification
- PSM
- protéines identifiées / quantification

MS

MS/MS

ions

signal  
TOF - Orbi

liste de pics

liste de PSM

liste de protéine

volcano plot

# Orbitrap



**Alexander  
Makarov**



# Orbitrap



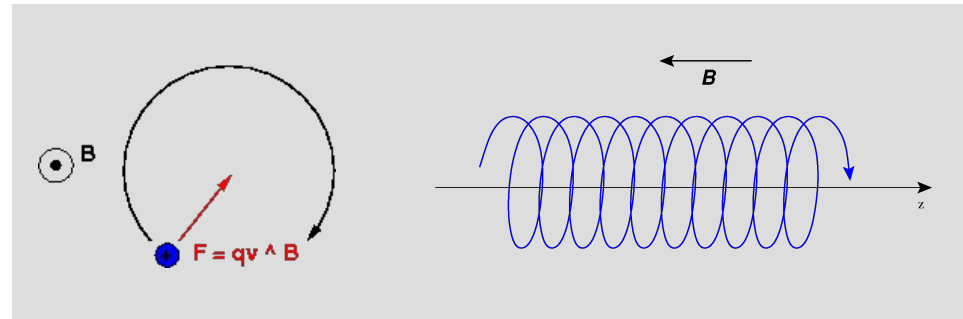
**2 generations of orbitrap  
cell**



**Alexander  
Makarov**

# FT-ICR

- A Magnet

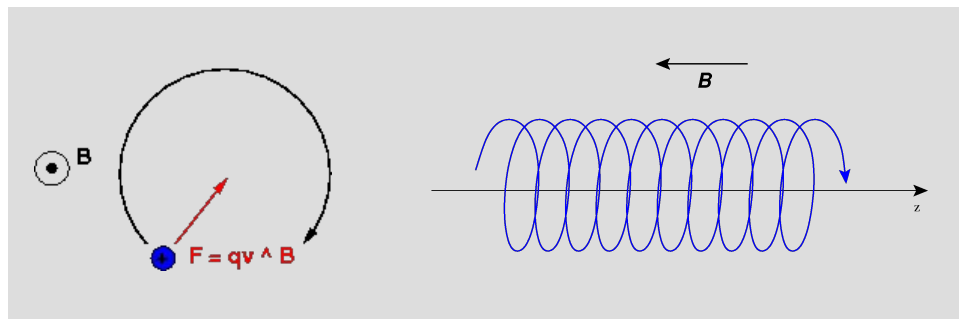


## Magnetic Resonance



# FT-ICR

- A Magnet



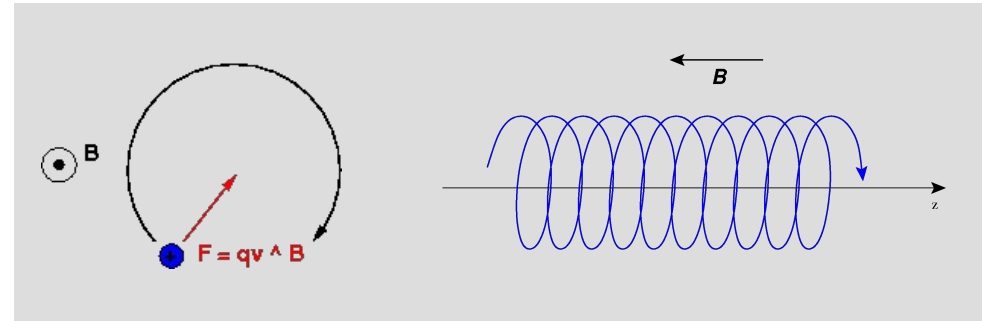
## Magnetic Resonance



# FT-MS

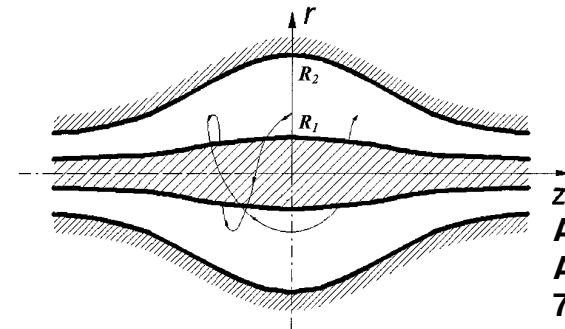
- ICR

- ions fly circularly in a homogeneous magnetic field
- cyclotron movement



- Orbitrap

- ions fly circularly in a (*carefully designed*) inhomogeneous electric field
- electrostatic trap



Alexander Makarov  
Anal. Chem. 2000,  
72, 1156-1162

Figure 1. Equipotentials of the quadro-logarithmic field and an example of a stable ion trajectory

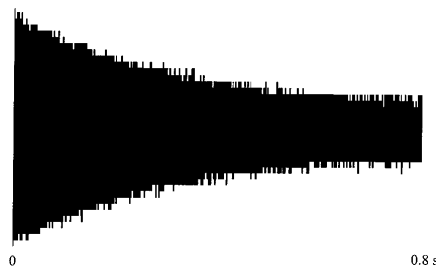


Figure 6. Typical transient for  $^{56}\text{Fe}^+$ . Hundreds of thousands of oscillations merge into the black envelope.

Fourier Transform



$R = 200\,000$  (*Orbi*)  
 $R = 2\,000\,000$  (*FTICR*)

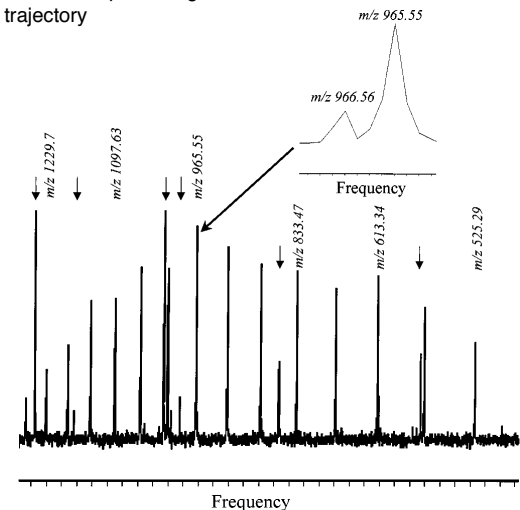


Figure 10. Wide mass range spectrum of PEG-1000 in the frequency domain (sum of 30 laser shots). Noise peaks of electronics are marked by arrows while all other peaks represent oligomers of PEG-1000. The most intense mass peak of the distribution (at mass 966.55 Da) is shown in detail in the inset.

# Improving the detection

- Sensitivity of the measure is governed by Signal/Noise ratio

⇒ increase signal

⇒ reduce noise

- Noise sources

- “standard”

▶ coming from the electronic on the apparatus

⇒ acquire more scan = takes time

- scintillation/jitter noise

▶ comes from the sample or the measure process

⇒ no counter action during acquisition

▶ preponderant in 2D (*t1-noise* in NMR)

$$S/N = \sqrt{N_{scan}}$$

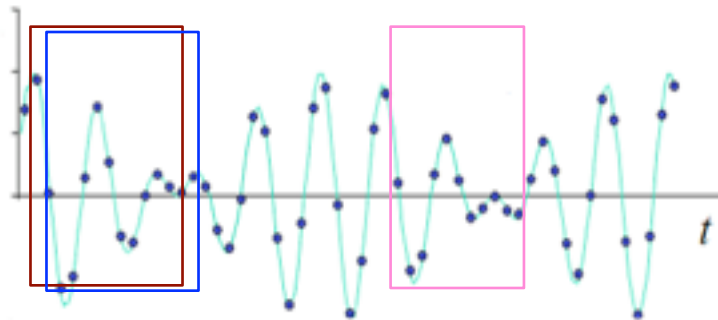
$$S/N \sim \text{invariant}$$

- Impact

- better detection of weaker compounds
- better coverage in bottom-up proteomics
- better detection of PTM
- faster acquisition

# Statistical treatment

Signal time-series :  $x = (x_1, x_2, x_3, \dots, x_L)$       P frequencies



Uniform sampling

Hankel matrix

$$H = \begin{bmatrix} x_1 & x_2 & \dots & x_N \\ x_2 & x_3 & \dots & x_{N+1} \\ x_3 & x_4 & \dots & x_{N+2} \\ x_4 & x_5 & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ x_M & x_{M+1} & \dots & x_{M+N-1} \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ M \end{matrix}$$

N

$$\begin{aligned} &M \times N \\ &L = M + N + 1 \\ &M < N \end{aligned}$$

**Hankel matrix:** Same terms on antidiagonals

# Cadzow procedure

- The idea is to decompose  $H$

- ▶ using **S**ingular **V**alue **D**ecomposition **SVD**

$$H = U\Sigma V^*$$

- ▶ singular values

$$\sigma_1 > \sigma_2 > \sigma_3 > \dots$$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \sigma_N \\ \dots & & & \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

- we keep only the  $k$  largest singular values

- ▶ and reconstruct a **denoised** signal from the **rank-reduced**  $H$  matrix

- ▶ projection of  $H$  on a subspace

$$\tilde{H} = U\Sigma_k V^*$$

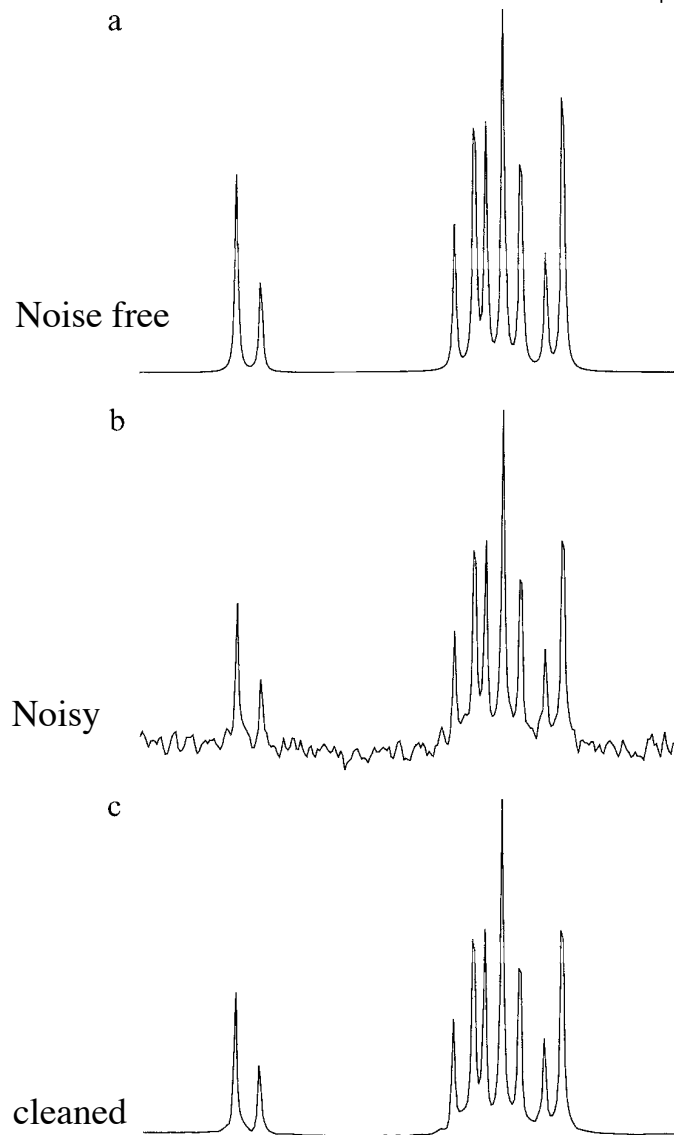
$$\tilde{H} = U_k U_k^* H$$



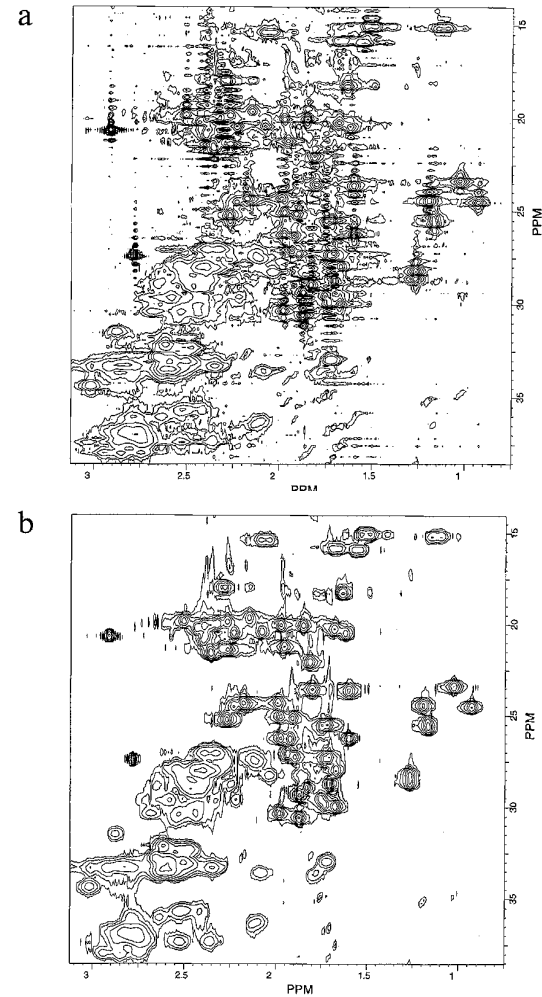
- ▶ then averaging on  $H$  antidiagonals

$$\tilde{x}_p = \langle \tilde{H}_{ij} \rangle_{i+j-1=p}$$

# Applied to NMR



Simulated



<sup>13</sup>C HMQC of pike parvalbumin

Brissac, C., Malliavin, T. E. & Delsuc, M.-A.  
*J Biomol NMR* **6**, 361–363 (1995).



# Approximate by random sampling

combine several new mathematical ideas

- use new developments linking between algebra and statistics
  - ▶ Johnson Linderstrauss Lemma (1984)
  - ▶ Compress Sensing approaches (Candès 2006, Donoho-Tanner 2007)
- Apply matrix approximation rather than complete matrices
  - ▶ Tygert, Martinsson ( 2007 )
- $\Rightarrow$  Estimate values rather than determining them
  - SVD can then be replaced by QR decomposition (faster)
  - precision and efficiency grows as the square root of the size  
hence efficient for Big Data
- uncoiled random QR denoising : urQRd
  - noise reduction from random sampling !

# New algorithm

- Build  $H$  :  $M \times N$
- Build a random matrix
  - $K$  is ~ number of signals
  - $K \ll M < N$
- Sample  $H$  with it
  - $Y$  smaller than  $H$
- Find main axes of  $Y$ 
  - QR decomposition MUCH faster than SVD
- make a rank reduction of  $H$  using  $Q$
- Reconstruction, as with Cadzow
  
- reminiscent with SVD truncature
  - much faster

$$\Omega : \mathcal{N}(0, 1) \ N \times K$$

$$H_{ij} = x_{i+j-1}$$

$$Y = H\Omega$$

$$Y = QR$$

$$\tilde{H} = QQ^* H$$

$$\tilde{x}_p = \langle \tilde{H}_{ij} \rangle_{i+j-1=p}$$

# expressing antidiagonal averaging as Fast Hankel matrix products

$$\tilde{H} = QQ^* H$$

$$\tilde{x}_p = \langle \tilde{H}_{ij} \rangle_{i+j-1=p}$$

$$U = Q^* H$$

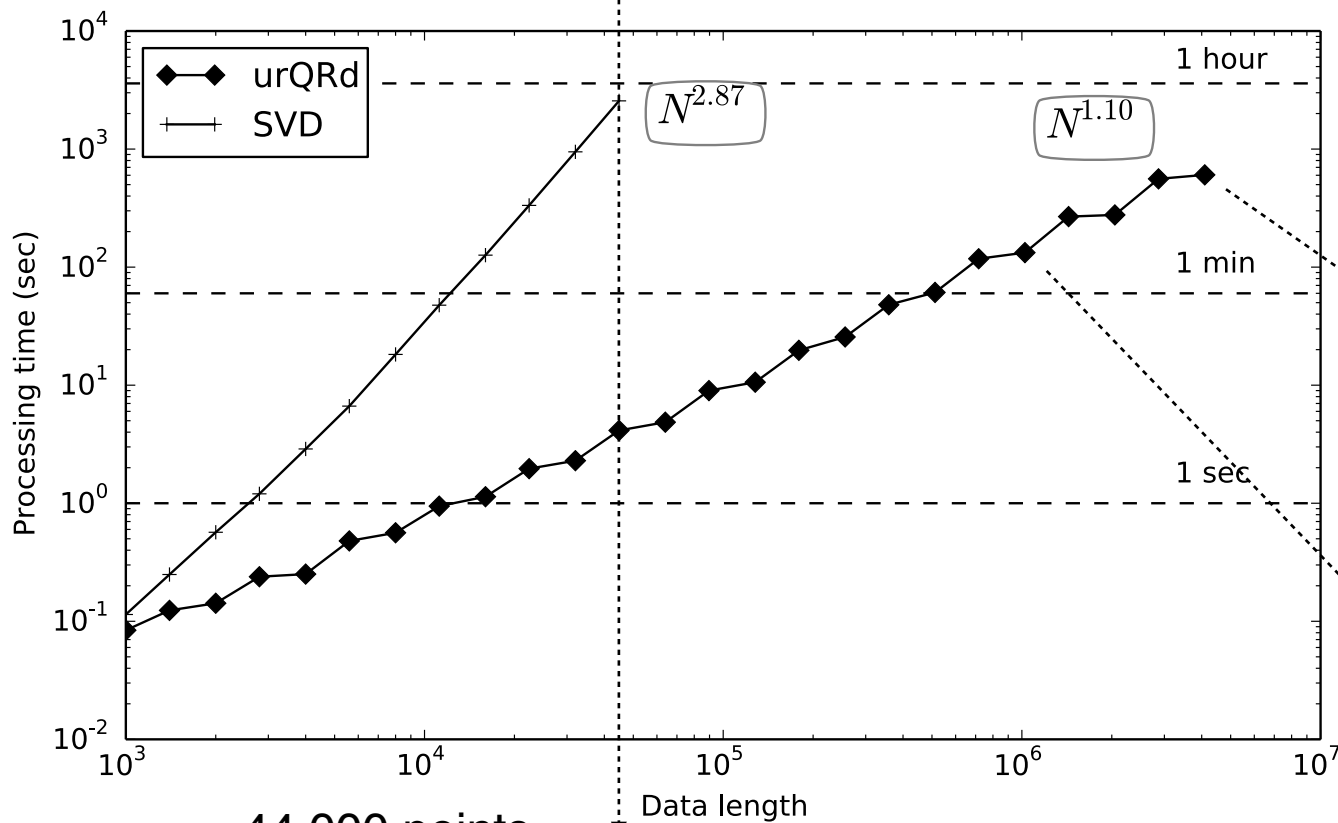
$$H_{i,j} \approx \tilde{H}_{i,j} = \sum_{k=1}^K Q_{i,k} U_{k,j}$$

$$\begin{aligned} \sum_{j=j_1}^{j_m} \tilde{H}_{i-j+1,j} &= \sum_{k=1}^K \sum_{j=j_1}^{j_m} Q_{i-j+1,k} U_{k,j} \\ &= \sum_{k=1}^K \sum_{j=j_1}^{j_m} Q_{i,j}^{(k)} U_j^{(k)} \\ &= \sum_{k=1}^K (Q^{(k)} \cdot U^{(k)})_i \end{aligned}$$

Here,  $Q^{(k)}$  is the  $L \times N$  Toeplitz matrix formed from the  $L + N - 1$  long vector  $[0, \dots, 0, Q_{k,1}, \dots, Q_{k,M}, 0, \dots, 0]$  with  $(N-1)$  zeros added on each extremities,  $U^{(k)}$  is the  $N \times 1$  vector whose entries are  $U_j^{(k)} = U_{k,j}$  and  $(Q^{(k)} \cdot U^{(k)})$  denotes the matrix-vector product, which is computed again using a fast algorithm.

sum of  $K$  Fast Matrix products

# much Faster - much Lighter



processing times  
 SVD in  $MN^2$   
 urQRd in  $KL\log(L)$   
 $K \ll M < N < L$

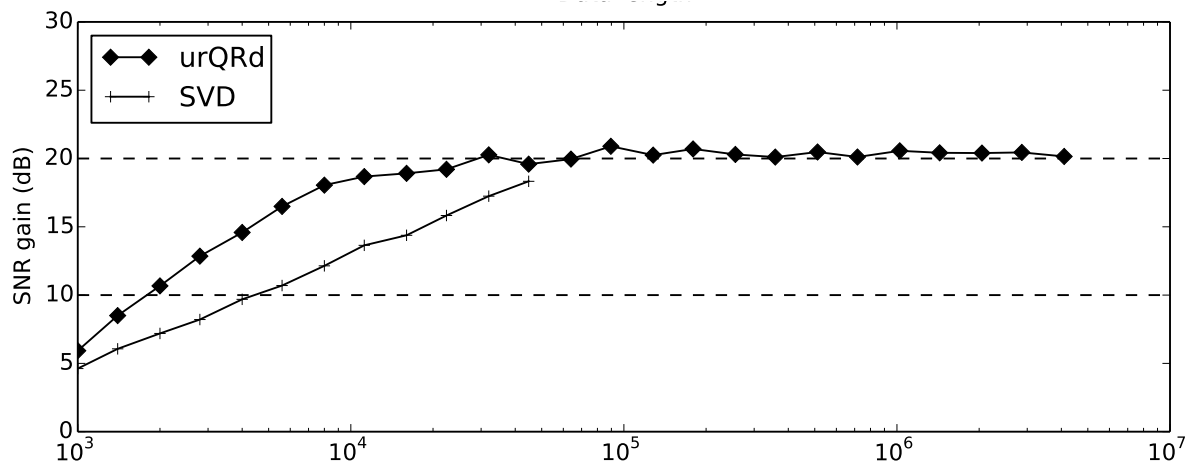
4.096.000 points  
 urQRd : 10 min

1.024.000 points  
 urQRd : 132 sec

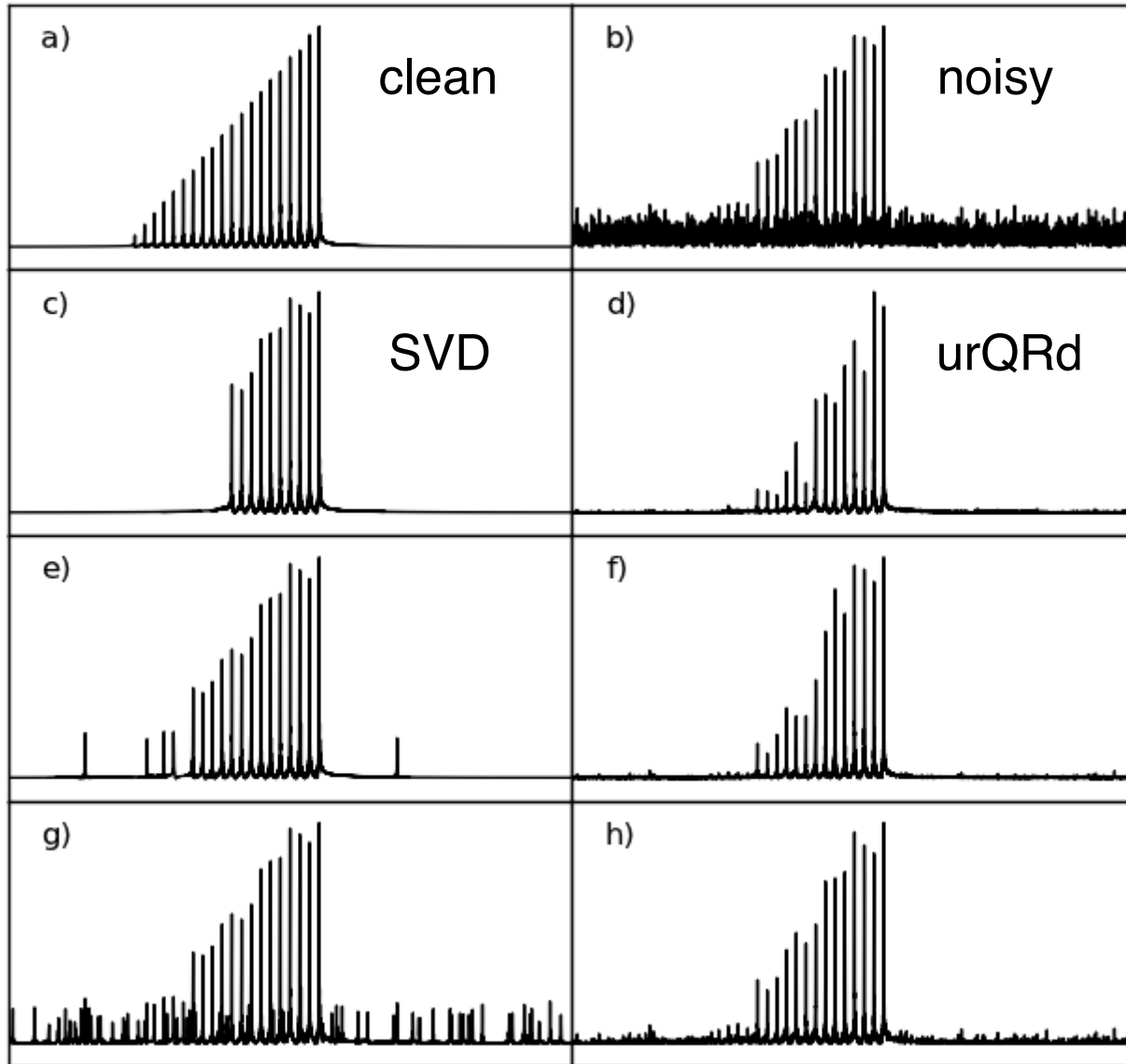
44.000 points  
 SVD : 42 min  
 urQRd : 4.1 sec

urQRd is very fast

better noise rejection



# Example of urQRd on synthetic data



*K* : estimate of number of signals

K=10

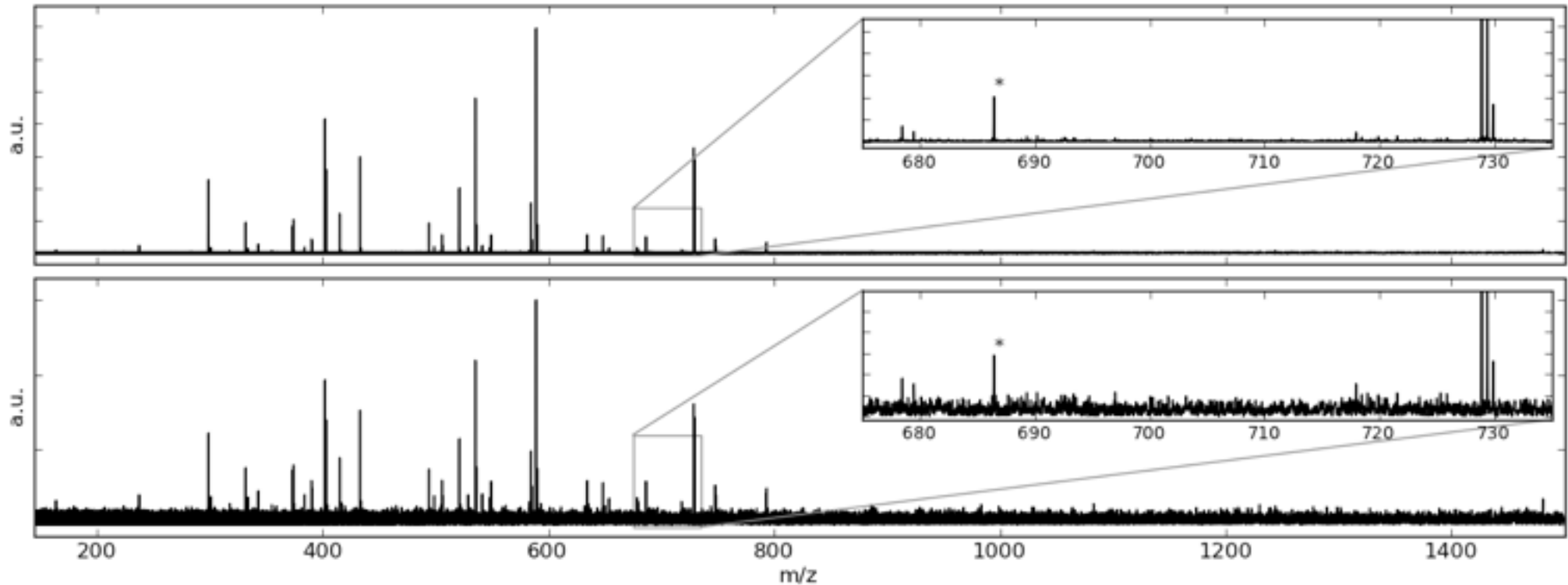
K=20

K=80

urQRd is very robust

# Applied to 1D FT-ICR data-set

Tryptic digest of CytC  
512k transient  
1 second acquisition



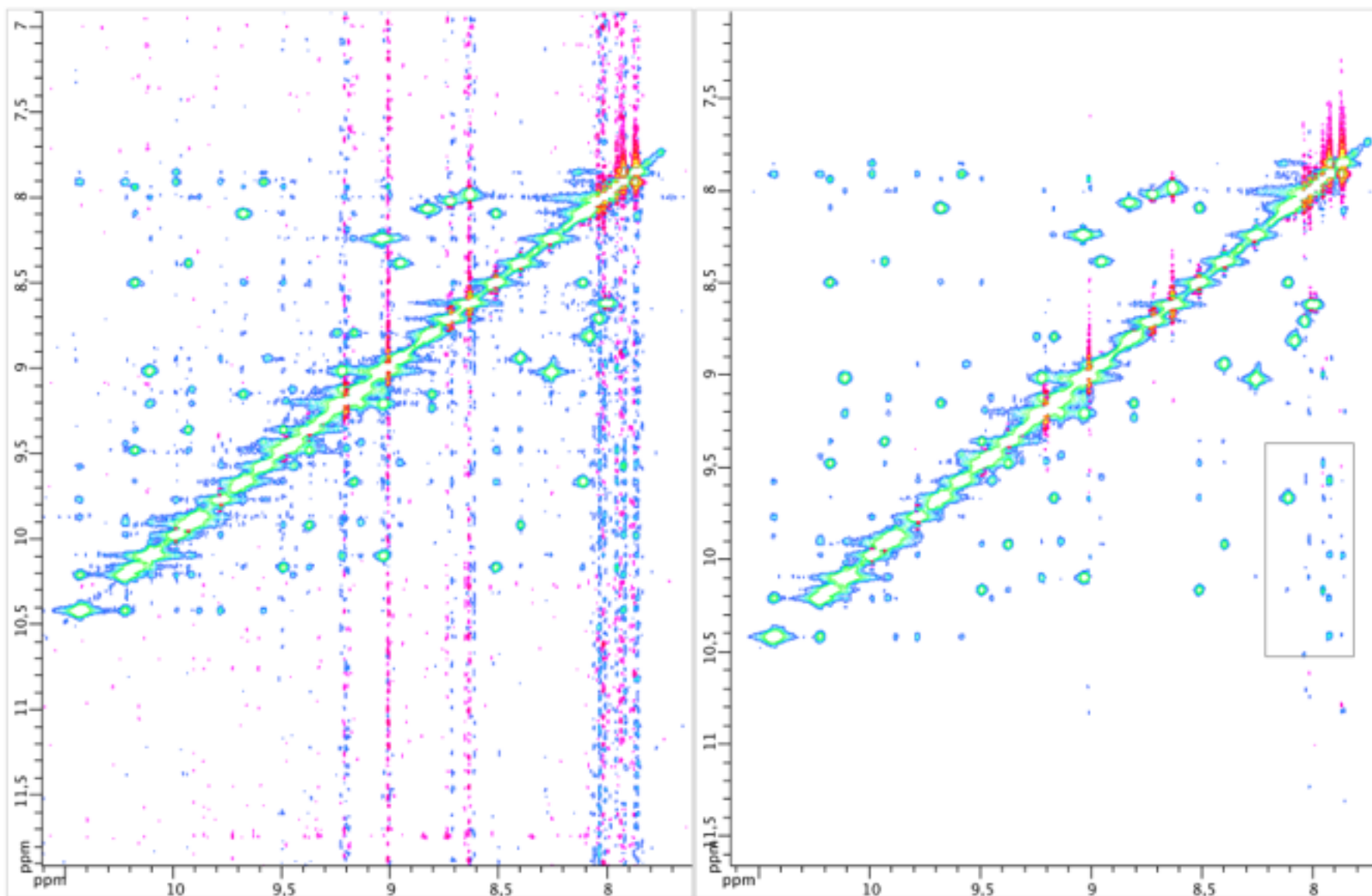
**Fig. 4.** Processing of a single-scan FT-ICR mass spectrum of a trypsin digest of Cytochrome C. *Bottom* original spectrum, SNR measured on the  $m/z$  728.8388 peak is 24.0 dB. *Top* same spectrum after urQRd processing ( $K = 1000$ ), SNR measured on the  $m/z$  728.8388 peak is 40.7 dB. *inset* the  $m/z$  728.8388 peak corresponds to the TGQAPGFSTDANK<sup>2+</sup> ion,  $m/z$  678.3821 to YIPGTK+ and  $m/z$  717.9012 to GEREDLIAYLKK<sup>2+</sup>. The peak labeled with a star at  $m/z=686.390$ , lacking isotopic structure, is likely to be an experimental artifact. The processed interferogram is 512k points, processed here with  $K = 1000$ .

SVD ( ~45 days 512Go memory)  
urQRd 25 min 4Go memory

<http://urqrd.igbmc.fr>

Chiron, L., van Agthoven, M. A., Kieffer, B., Rolando, C. & Delsuc, M.-A.  
*Proc Natl Acad Sci USA* **111**, 1385–1390 (2014).

Very efficient



Reduction of t1-noise on a 2D NOESY spectrum ~2minutes

# Compression

- techniques
  - sans perte de données : zip
  - avec perte de données : jpeg mp4
- mesure de la compressibilité
  - quantité d'information
  - parcimonie
- mesure de la quantité d'information
  - par la compressibilité
  - entropie d'information
- théorème de Shannon
  - relie la quantité d'information maximale et la taille du canal d'information  
rapport signal/bruit taille de l'alphabet ...
- théorème de Candès
  - relie le nombre de signaux non nuls et le nombre de mesure à réaliser



# Compressed Sensing

- Dispersion des informations dans les données
  - transformée idéale
  - incertitude d'Eisenberg
  - parcimonie
- Exemples
  - les pommes d'Hadamard
  - carte des étoiles du ciel
- Compression
  - compression de la mesure et non pas des données
- Théorème de Candès
  - compression sans perte

E.Candès T.Tao 2006

Robust Uncertainty Principles:  
Exact Signal Reconstruction from Highly Incomplete  
Frequency Information

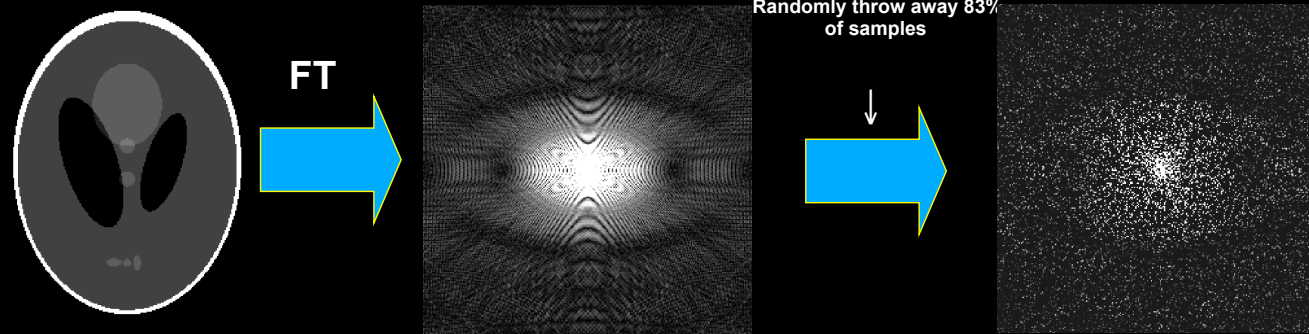
Emmanuel Candes<sup>†</sup>, Justin Romberg<sup>†</sup>, and Terence Tao<sup>#</sup>

<sup>†</sup> Applied and Computational Mathematics, Caltech, Pasadena, CA 91125

<sup>#</sup> Department of Mathematics, University of California, Los Angeles, CA 90095

E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. on Information Theory, 52 (2006), pp. 489–509.

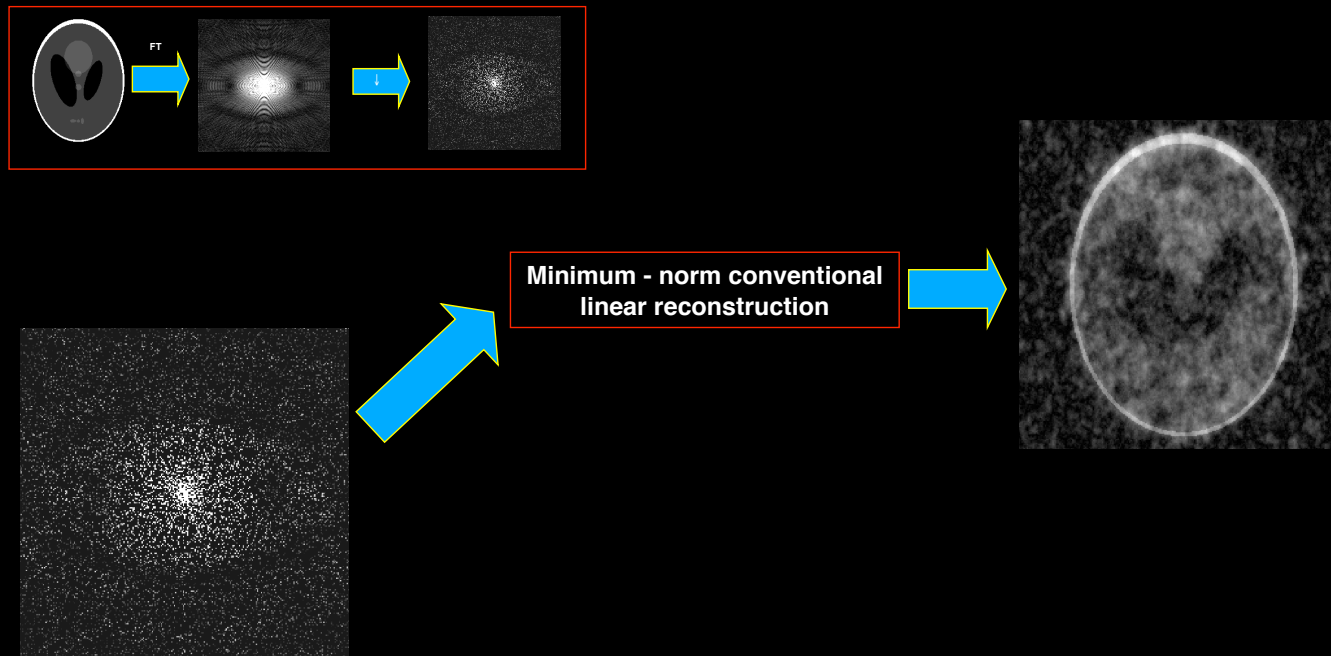
## A Surprising Experiment\*



\* E.J. Candes, J. Romberg and T. Tao.

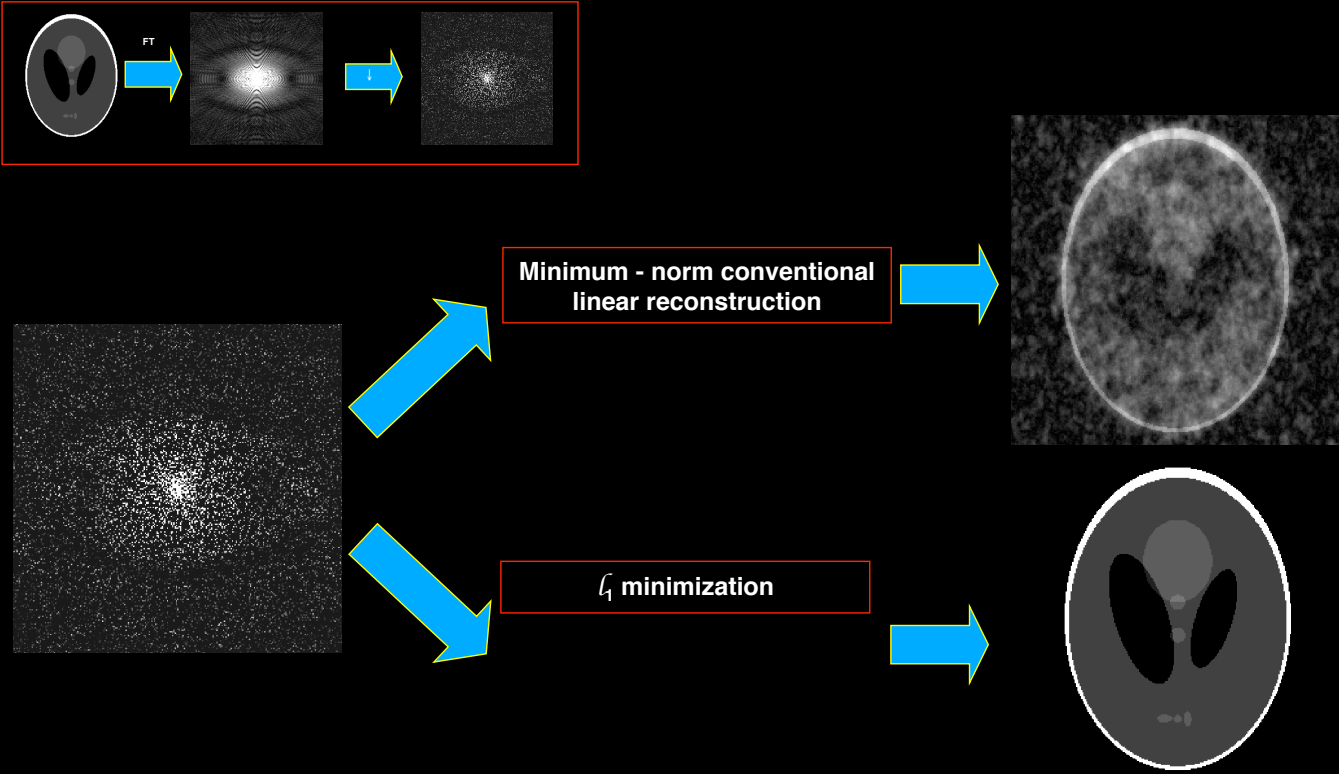
Jean-Luc Starck, Florent Sureau  
J. Bobin, N. Barbey, A. Woiselle

## A Surprising Result\*



\* E.J. Candes, J. Romberg and T. Tao.

# A Surprising Result\*



E.J. Candes

# Compressed Sensing

- régulation : données simples/creuses - *sparsity* -

- $K$  signaux -  $N$  mesures -  $P$  points dans le spectres
- hypothèse de peu de signaux :  $K \ll N \ll P$

- Le problème devient une simple optimisation convexe

$$\min(\|s\|_1) \quad \text{avec} \quad \|y - Ts\|_2 < \epsilon$$

- norme  $l_1$  ou  $l_0$  mais pas  $l_2$

$$\|s\|_1 = \sum |s_i| \quad \|s\|_p = \left( \sum s^p \right)^{\frac{1}{p}}$$

- Dans certaines conditions

## *Restricted Isometry Property*

- R.I.P.
- dispersion (FT par exemple)
- pseudo-inversible
- linéaire

# going further : Recital algorithm

## ▶ Resolution **E**nhan**C**ement by an **I**terative **A**lgorithm

### • Hypothesis

- ▶ The signal is the pulse response of the MS spectrum
- ▶ The MS spectrum is composed of isolated peaks

### • Idea

- ▶ Find the minimal number of signals that describes the data
- ▶ Using Inverse reconstruction

### • Method

#### • Minimize 2 targets

number of lines in spectrum

distance between experimental and reconstructed data (least-square)

#### • use noise estimate as stopping criterium

#### • adapted from the Fast Iterative Soft-Thresholding Algorithm

fast and robust

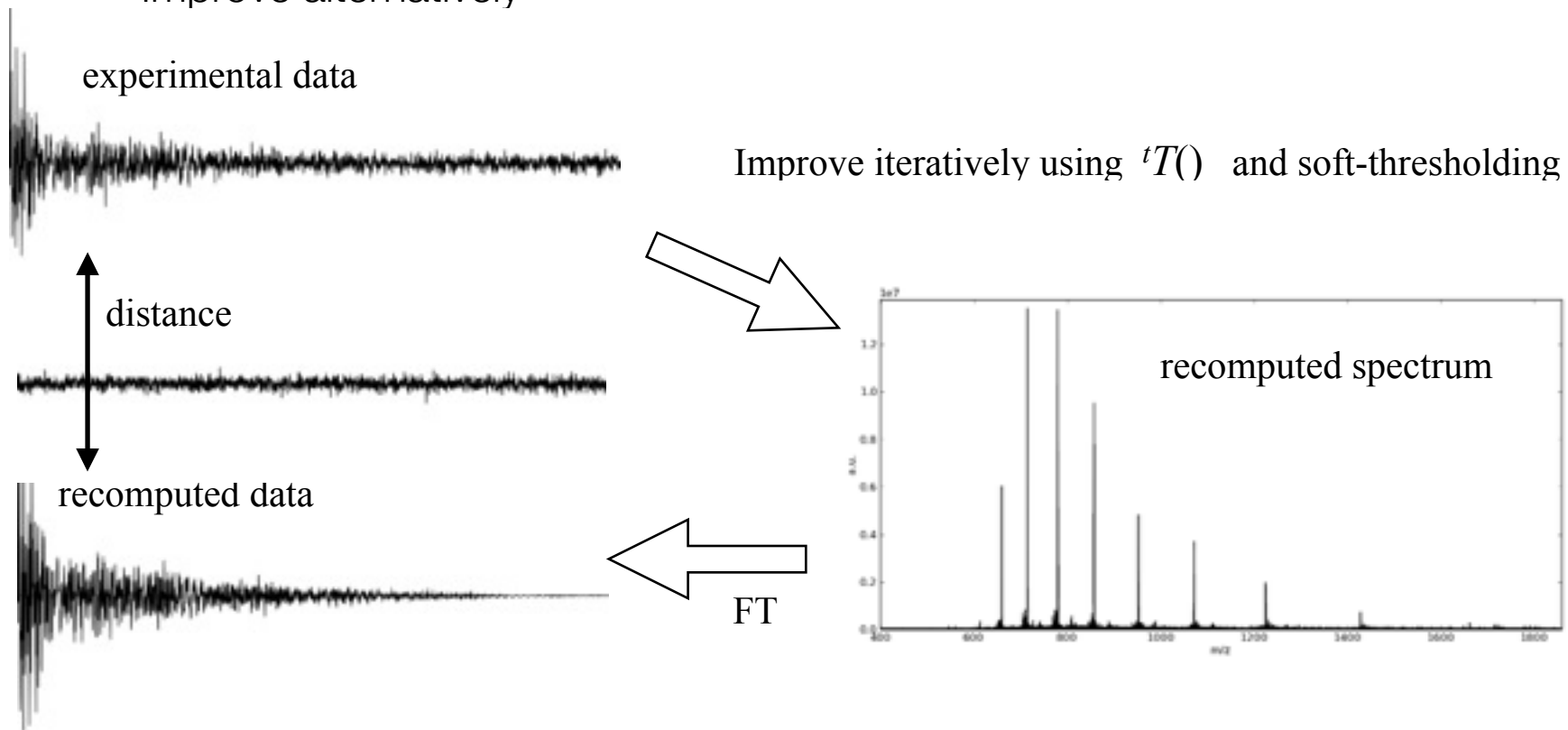
# Inverse reconstruction

- Inverse approach

- Split algorithm

$$\|s\|_{\ell_1} \quad \text{and} \quad \|y - Ts\|_{\ell_2}$$

- ▶ Improve alternatively



- using Fast Iterative Soft-Thresholding Algorithm (FISTA)



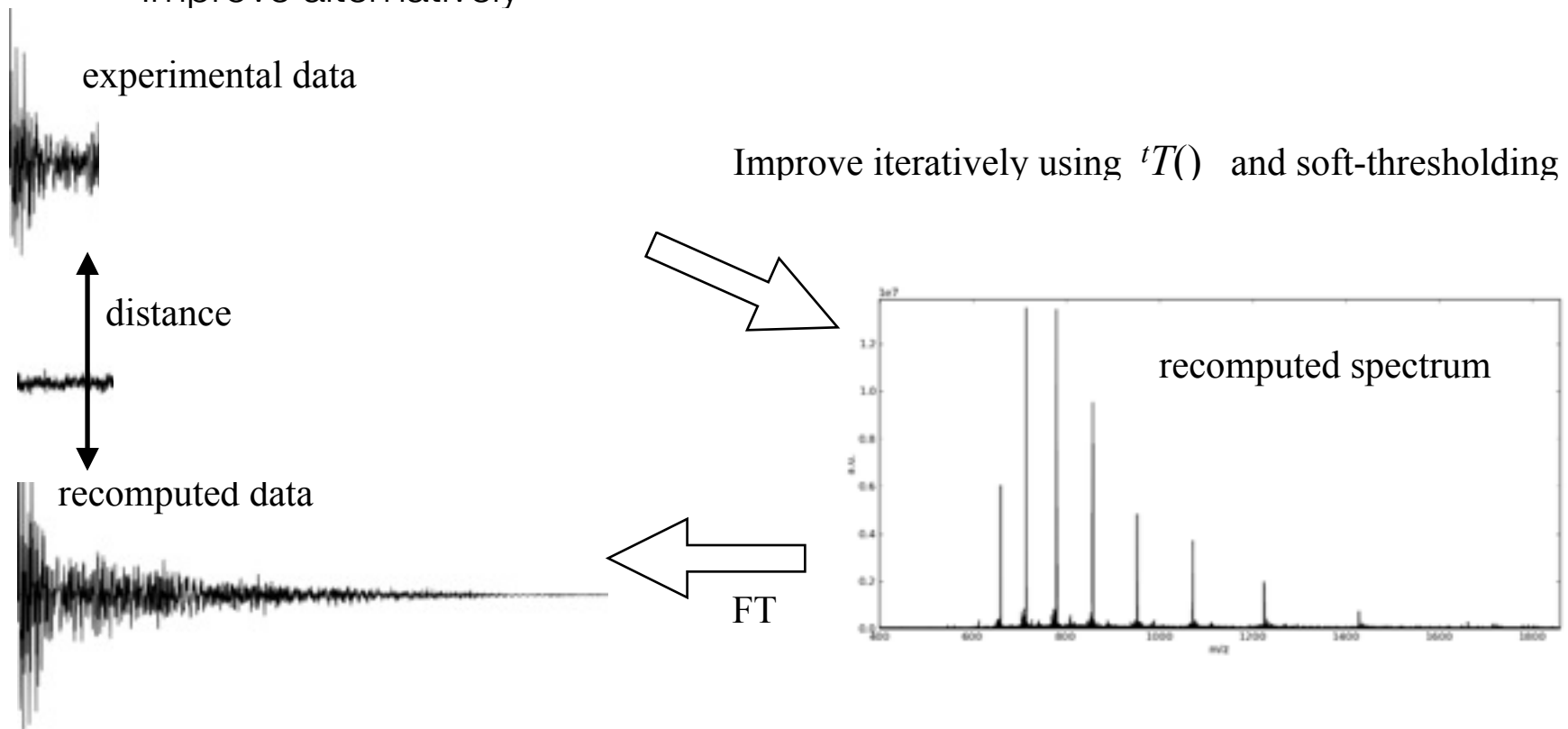
# Inverse reconstruction

- Inverse approach

- Split algorithm

▶ Improve alternatively

$$\|s\|_{\ell_1} \quad \text{and} \quad \|y - Ts\|_{\ell_2}$$



- using Fast Iterative Soft-Thresholding Algorithm (FISTA)

# What is this $\ell_1 - \ell_2$ stuff ?

- Generalization of the Euclidean norm :

Definition of the norm  $\ell_n$ . noted  $\|x\|_n$

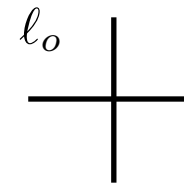
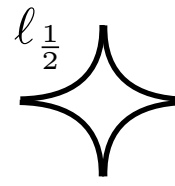
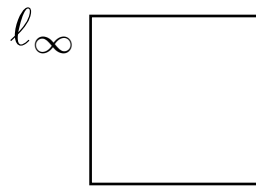
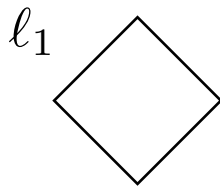
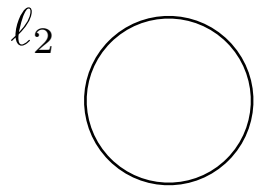
$$\|x\|_n = [\sum_i x_i^n]^{1/n}$$

So Euclidean norm is  $\ell_2$

- we can also define the special  $\ell_0$  norm :

$$\|x\|_0 = \lim_{p \rightarrow 0} \|x\|_p$$

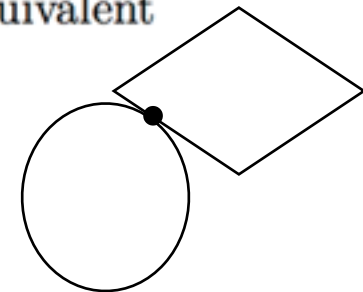
- what is a sphere in norm  $\ell_n$  ?



- $\ell_0$  counts the number of non-null values, so minimizing  $\|s\|_0 \equiv$  minimizing the number of lines.
- spheres in norm  $\ell_n$  with  $n \geq 1$  are convex  $\Rightarrow$  easy minimisation  
 $\ell_0$  is much more tricky !

By chance Candès et al<sup>1</sup>, in a seminal paper showed that  $\ell_0$  and  $\ell_1$  are equivalent on that matter. So the problem boils down to

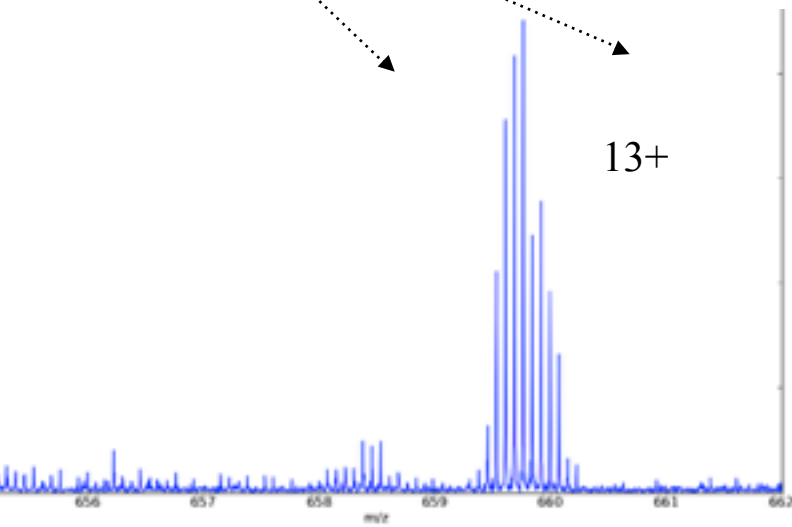
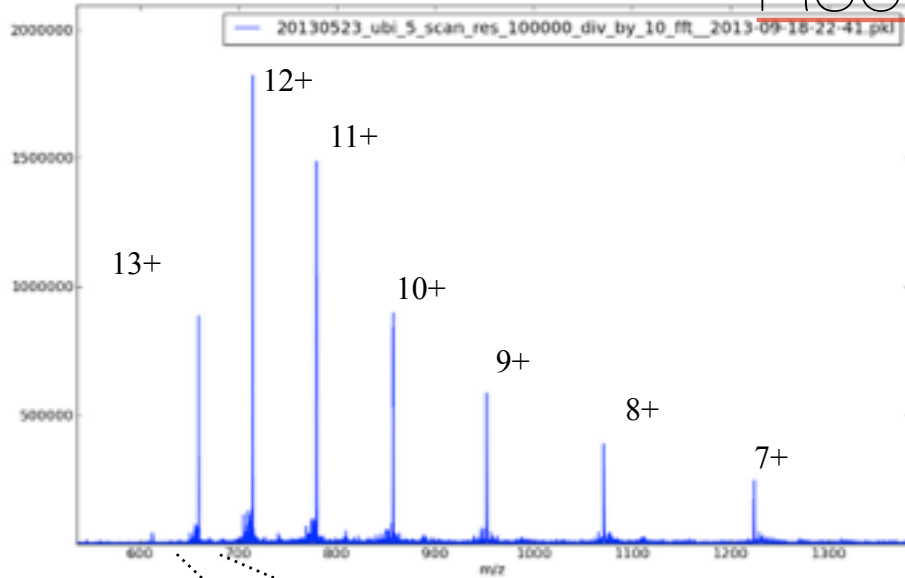
$$\min(\|s\|_1) \quad \text{with} \quad \|y - Ts\|_2 < \text{noise}$$



<sup>1</sup>Candès, E., Romberg, J. & Tao, T. "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information". *Inf Theo, IEEE Trans.* **52**, 489–509 (2006)

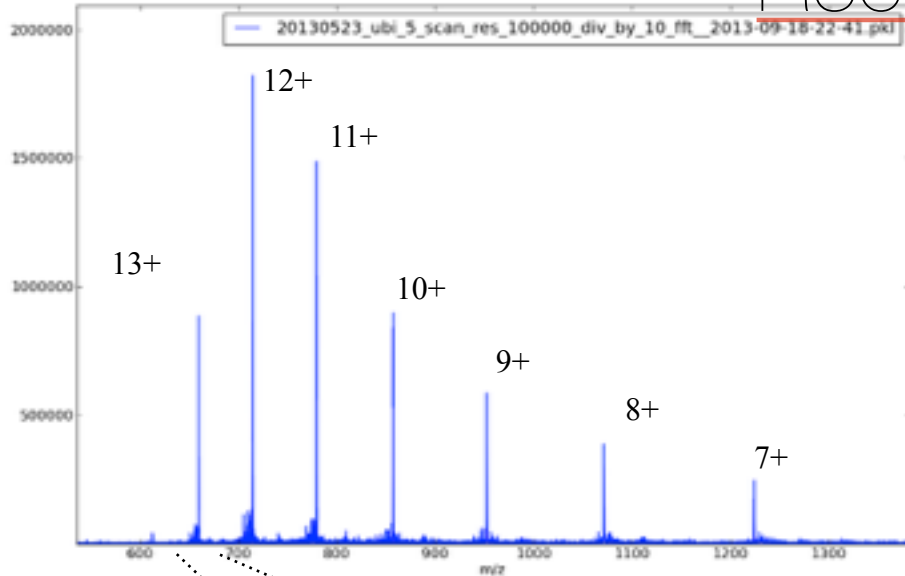
# Recital algorithm

## Orbitrap Ubiquitin spectrum



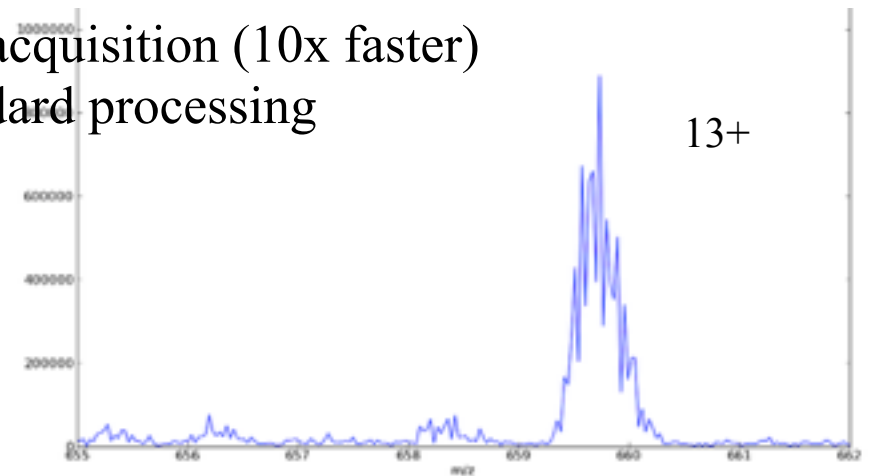
HiRes acquisition  
standard processing

# Recital algorithm

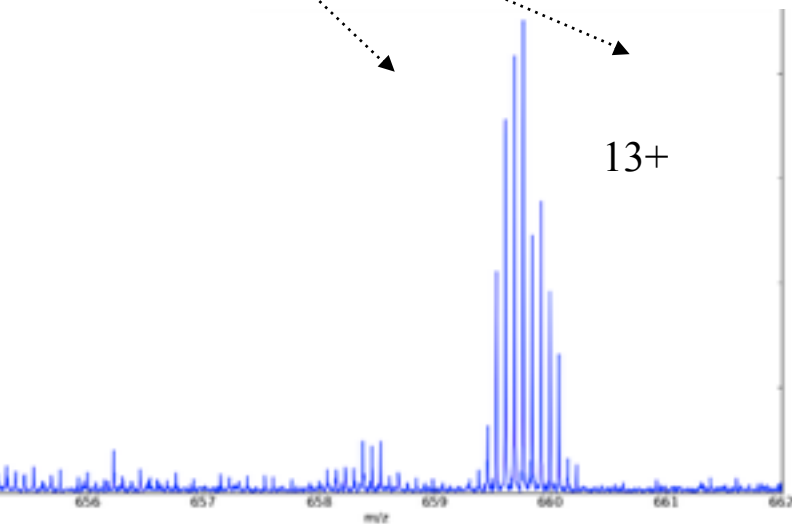
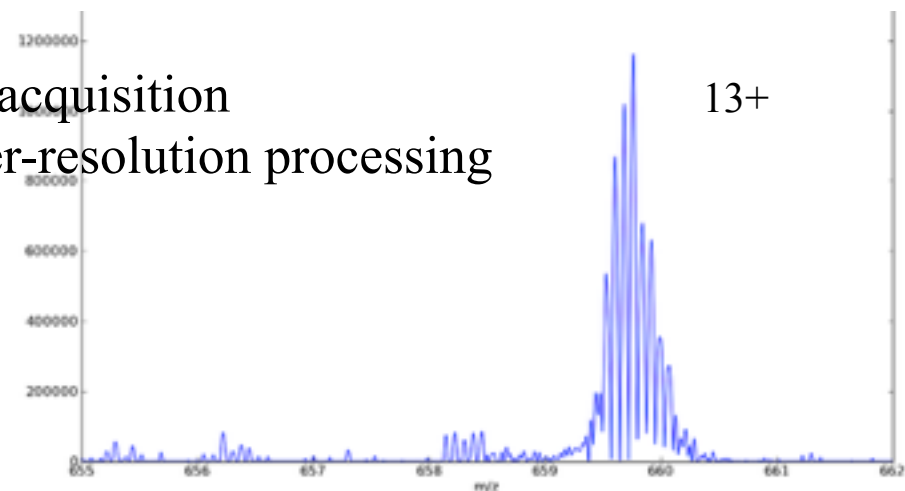


Orbitrap Ubiquitin spectrum

fast acquisition (10x faster)  
standard processing



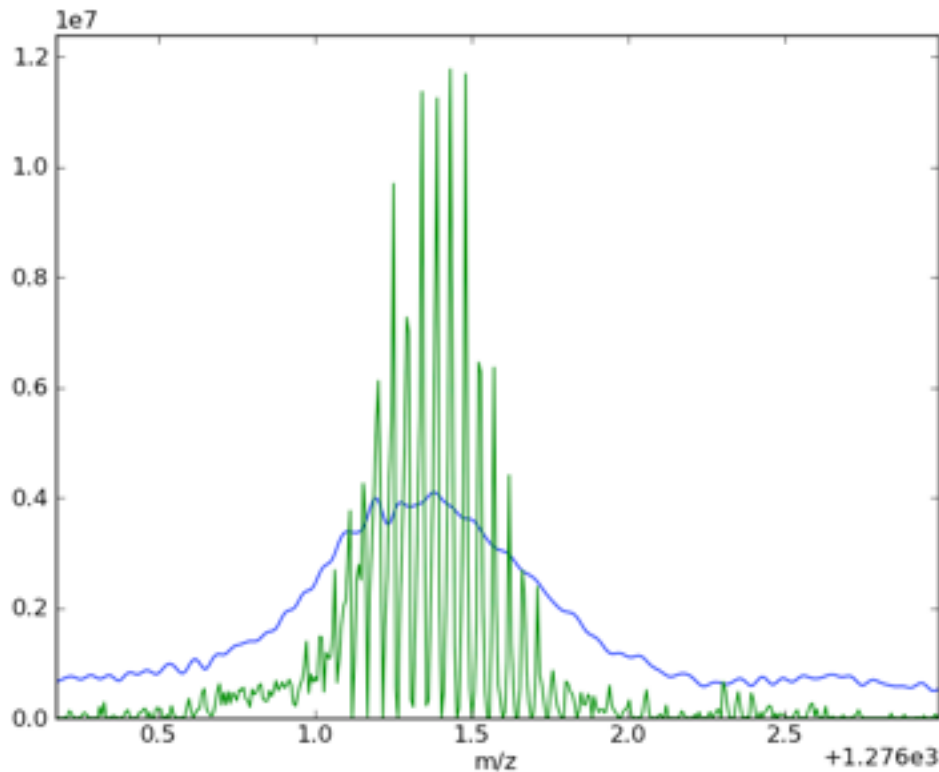
fast acquisition  
super-resolution processing



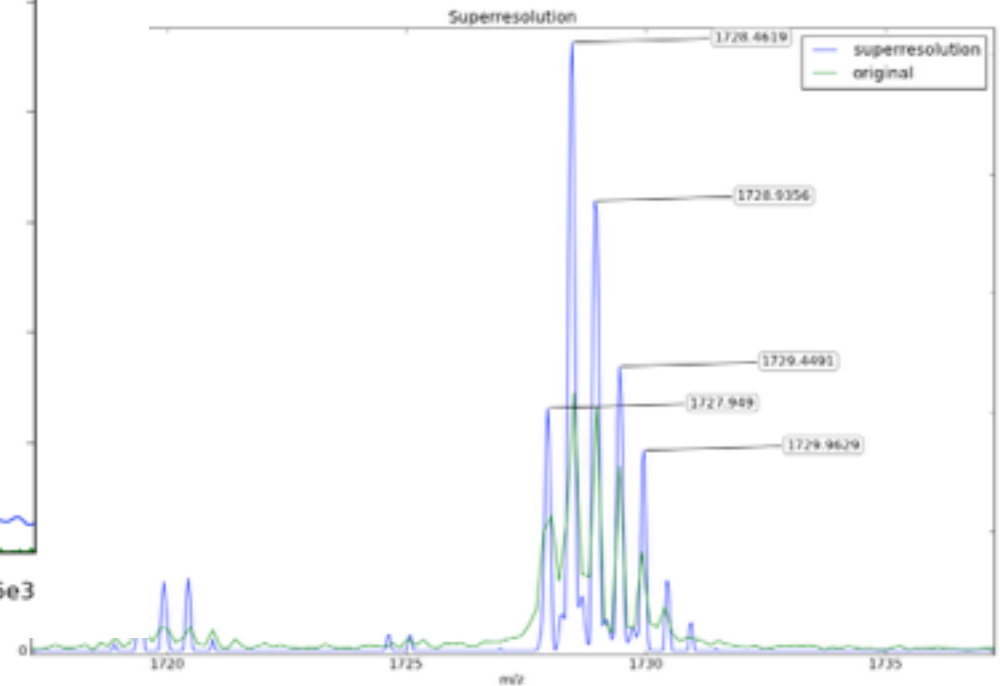
HiRes acquisition  
standard processing

# other example

- Standard FT vs Recital processing



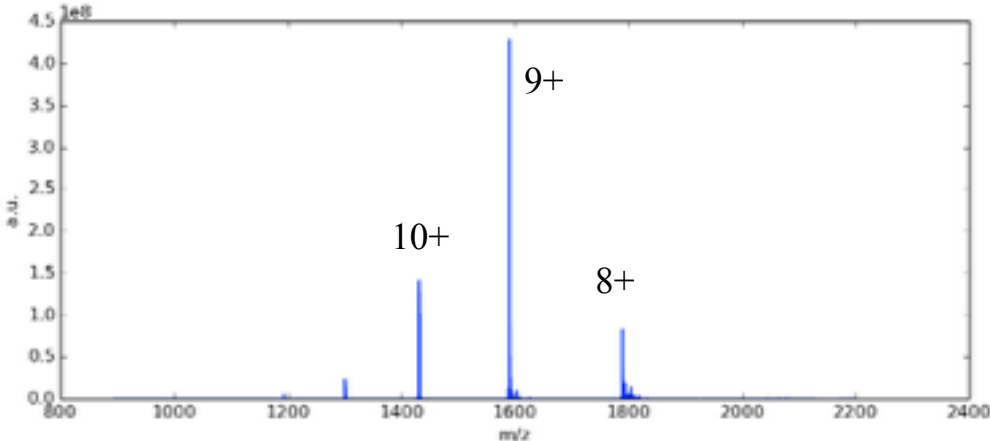
Fast Spectrum of Apolipoproteine A1  
FT-ICR data  
(30 kD  $z=24^+$   $m/z$  1271)



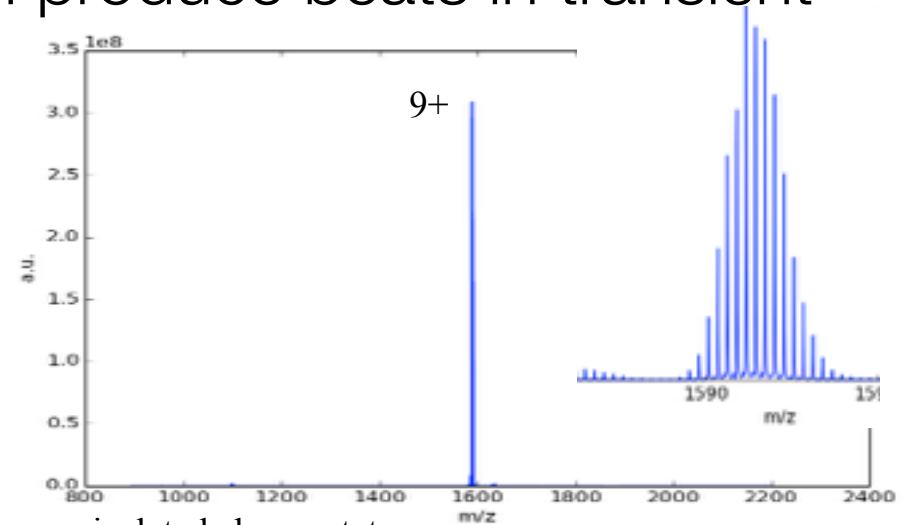
Fast Spectrum of Angiotensin  
FT-ICR data  
( $m/z$  1727)

# isotopic patterns

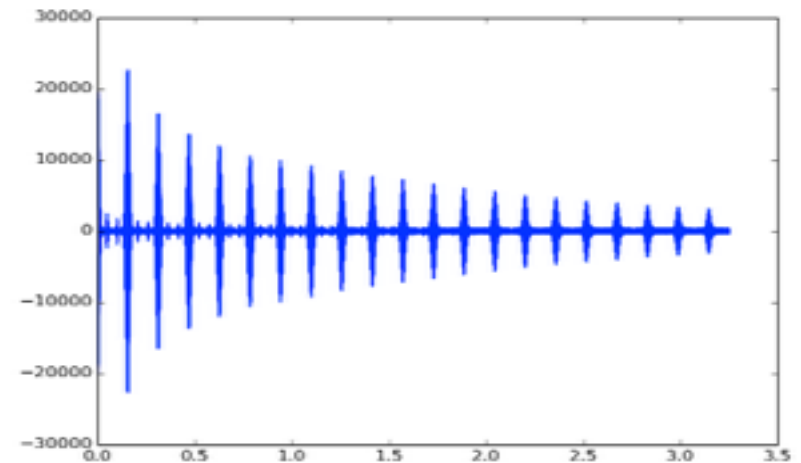
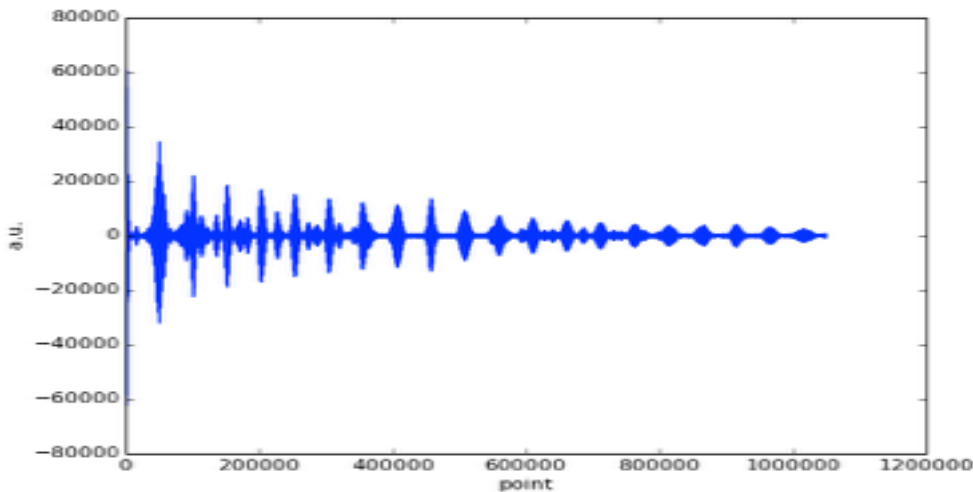
- isotopic patterns in high resolution produce beats in transient



Lysozyme - FT-ICR spectrum at 9.4T



isolated charge state



in FT-ICR beat period  $t_B$  is

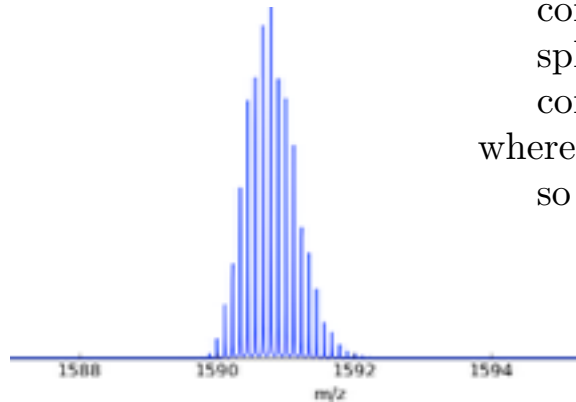
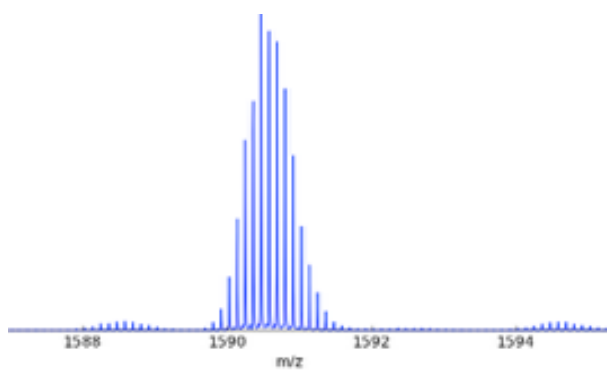
$$t_B \approx \frac{m^2}{zB_0}$$

# simulating the beats

Experimental data

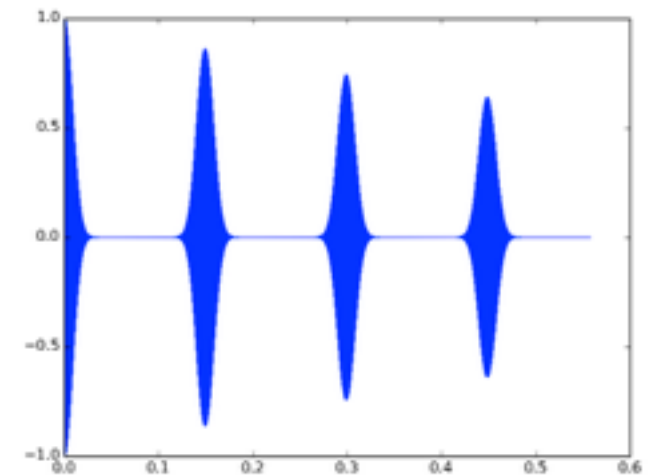
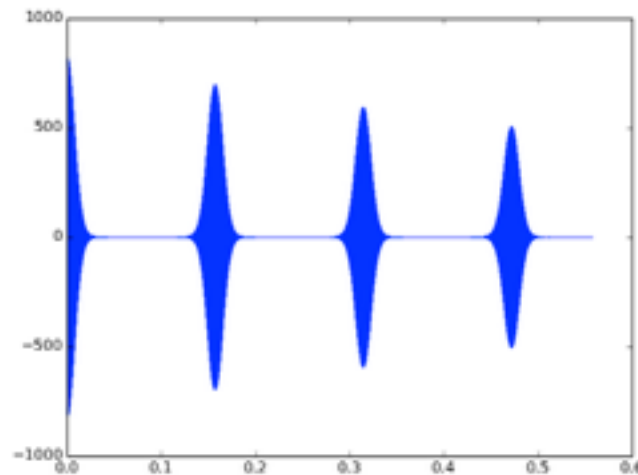
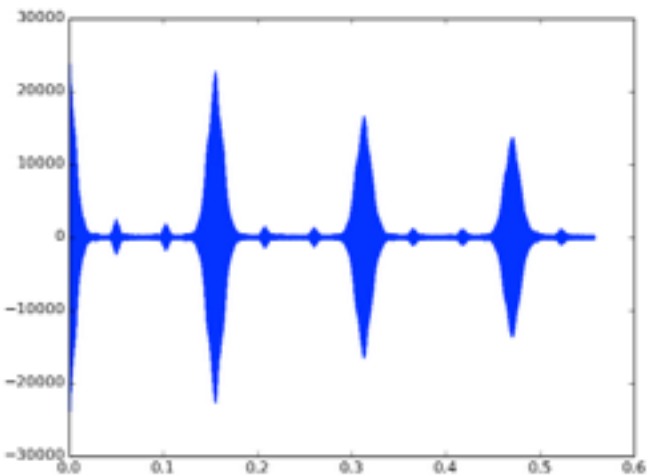
simulated data  
from primary sequence

simple mathematical model

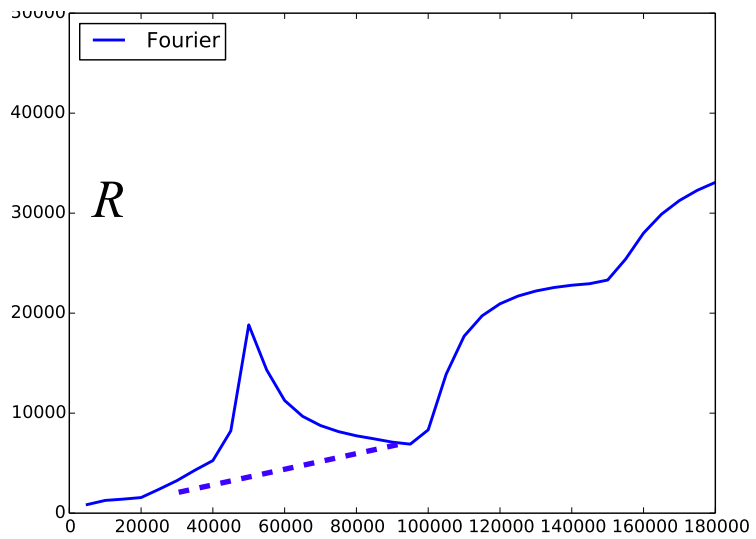
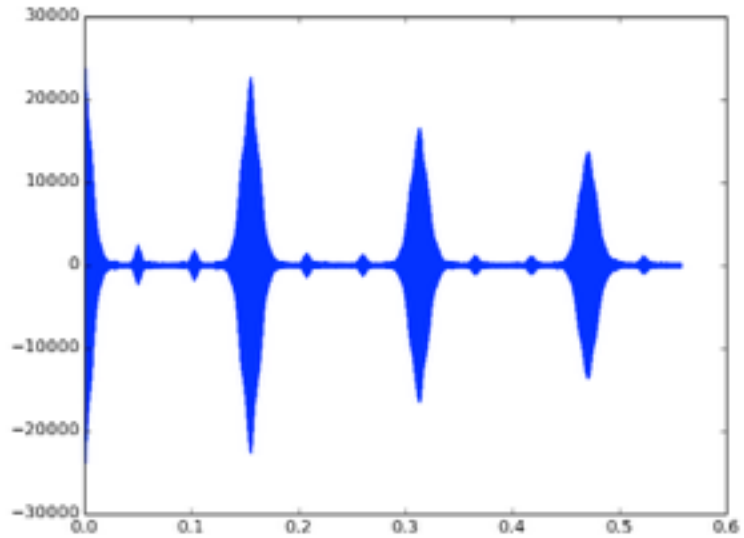


convolution product of all isotopic distributions  
splitting in spectrum  $\equiv$  cosine in transient  
convolution theorem  $\Rightarrow \cos^n$  itemizen transient  
where  $n \approx$  number of lines in isotopic pattern  
so  $s(t) \approx \cos^n(f_{\Delta z}t) \cos(\omega_o t)$

$$\cos^{30}(f_{\Delta z}t)$$



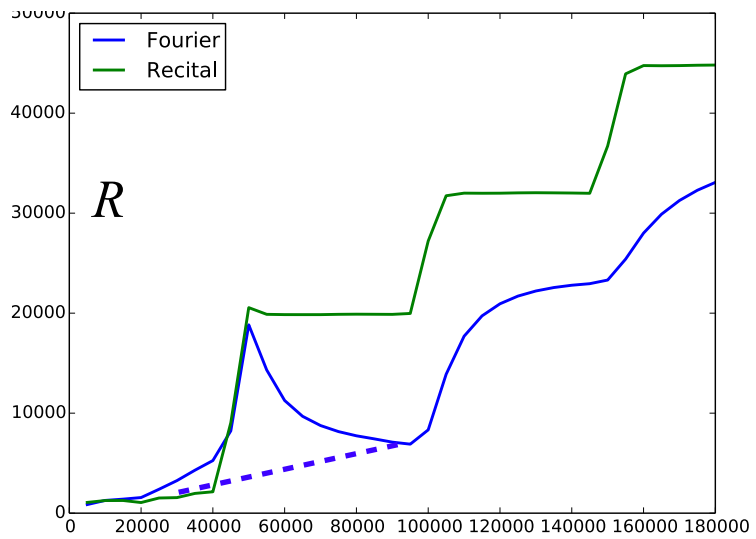
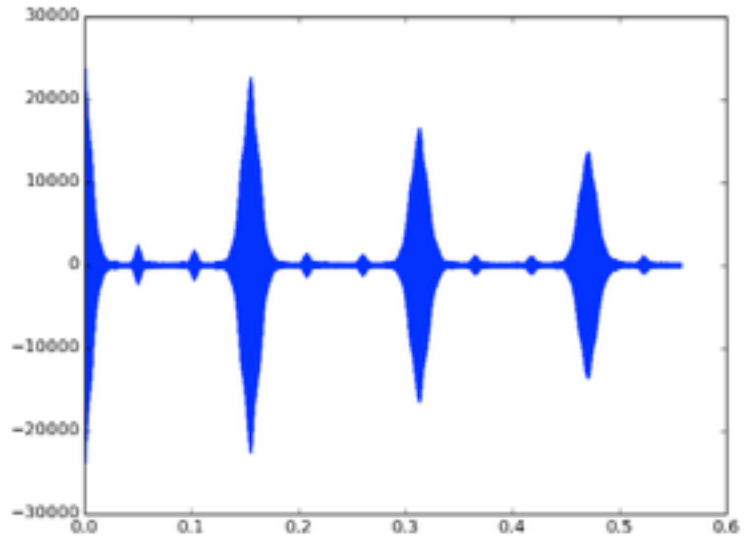
surprise !



Evolution of resolution  
with acquisition time

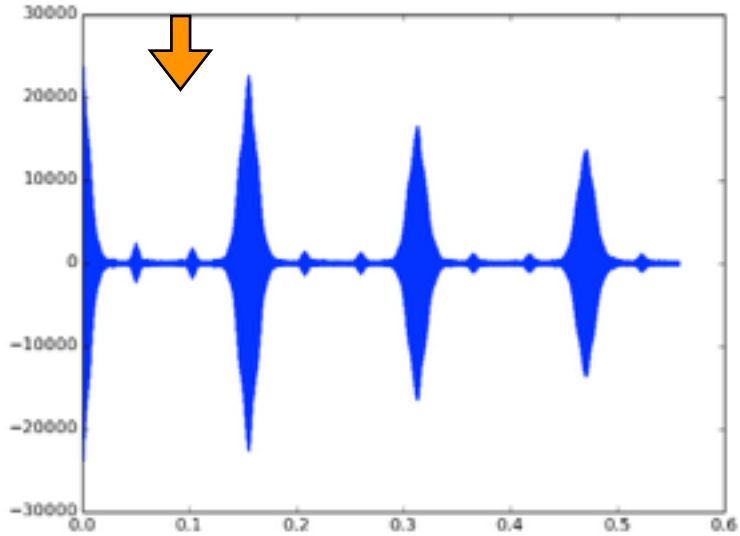


surprise !

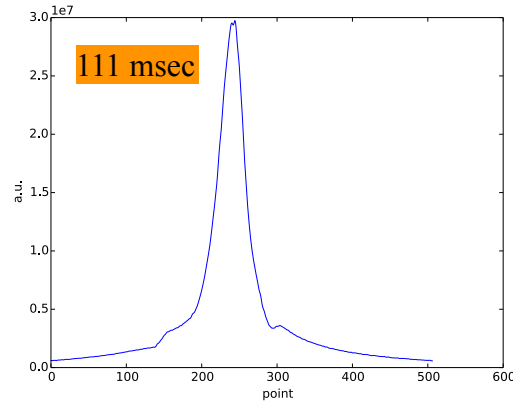


Evolution of resolution  
with acquisition time

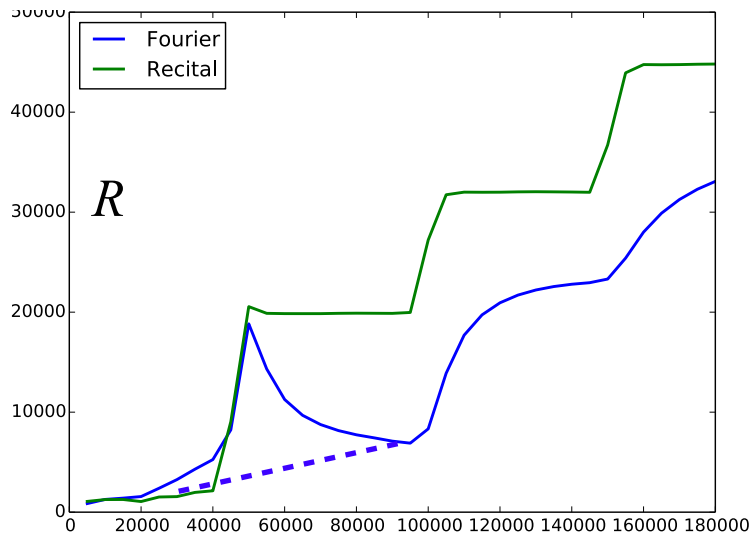
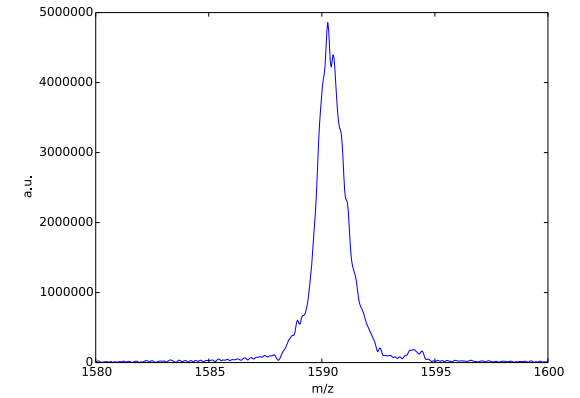
surprise !



Fourier



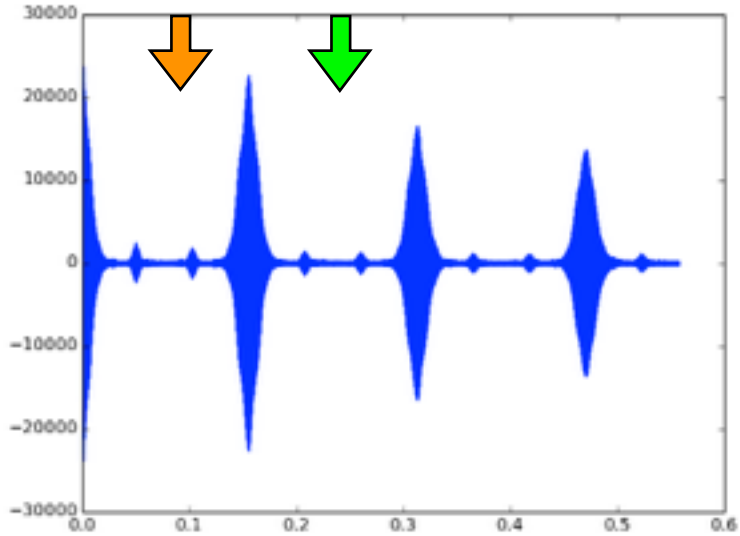
Recital



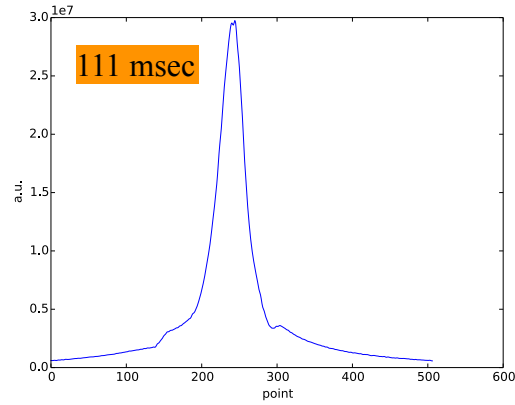
35k : 111msec  
Fourier  $R=4.300$   
Recital  $R=1.700$

Evolution of resolution  
with acquisition time

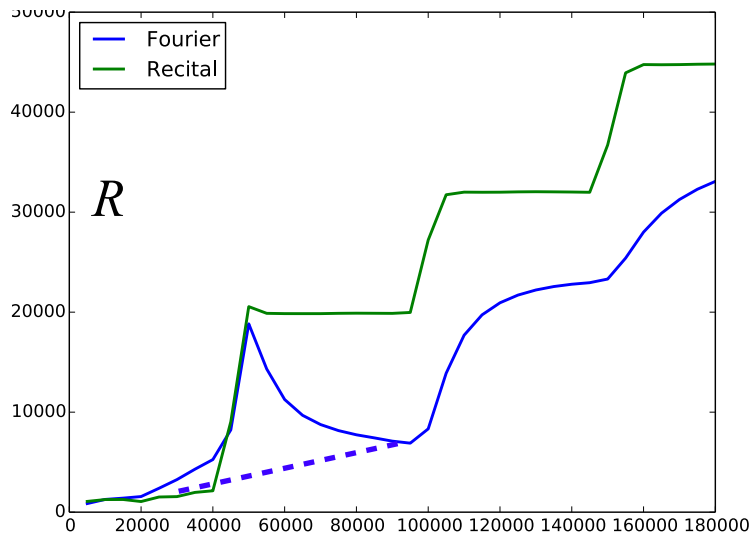
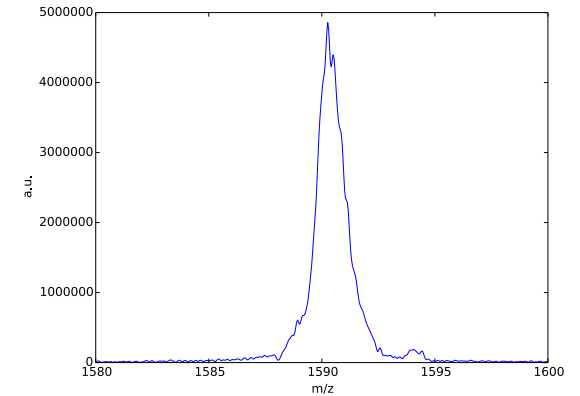
surprise !



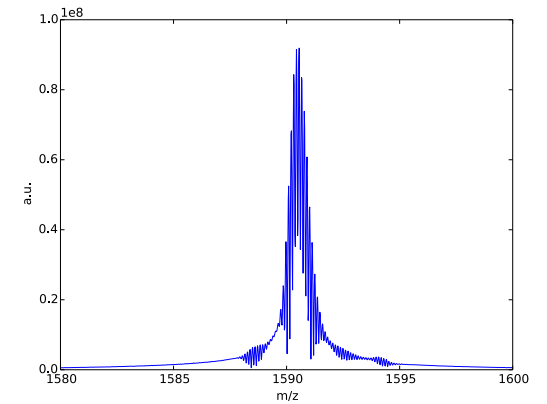
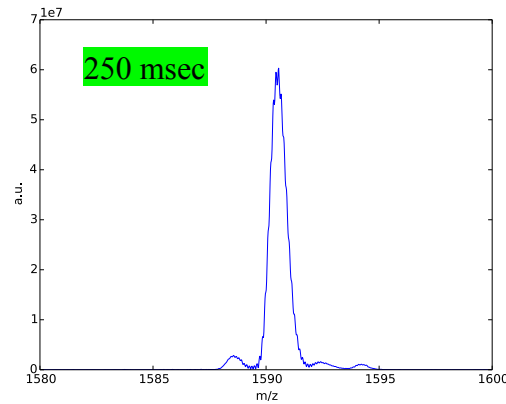
Fourier



Recital



35k : 111msec  
Fourier R=4.300  
Recital R=1.700

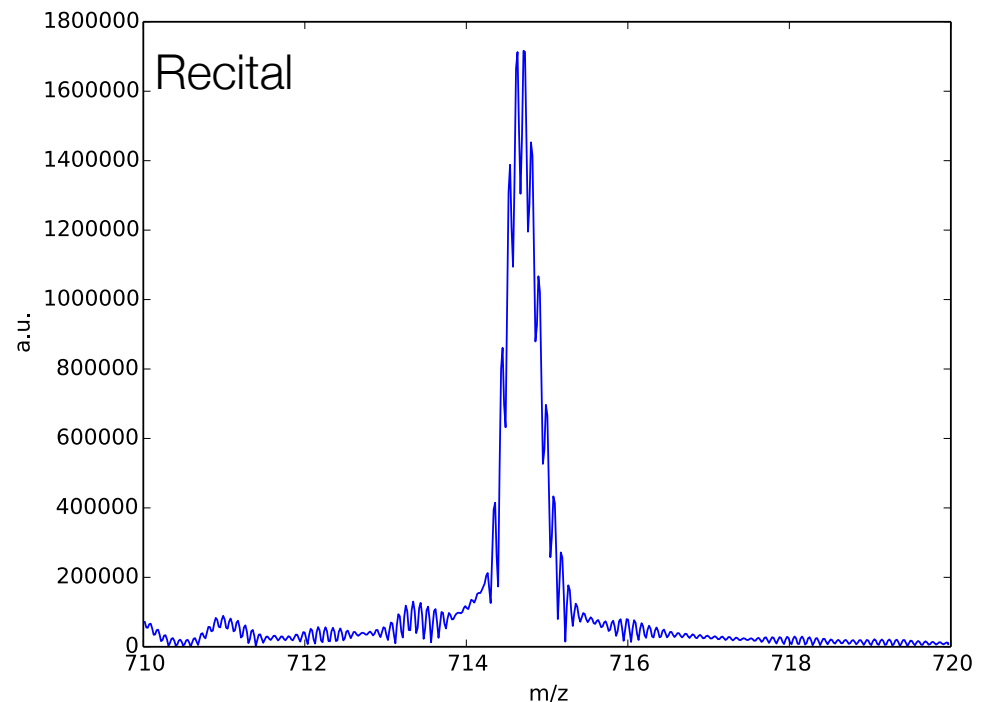
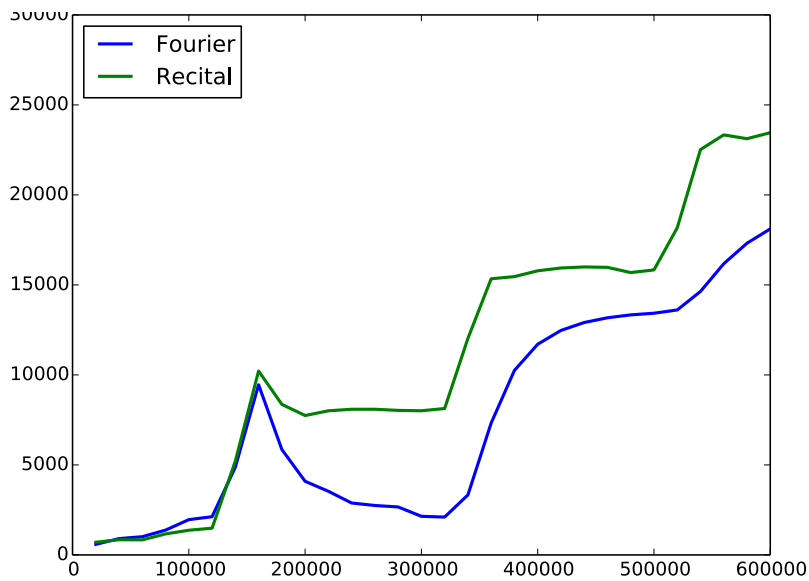
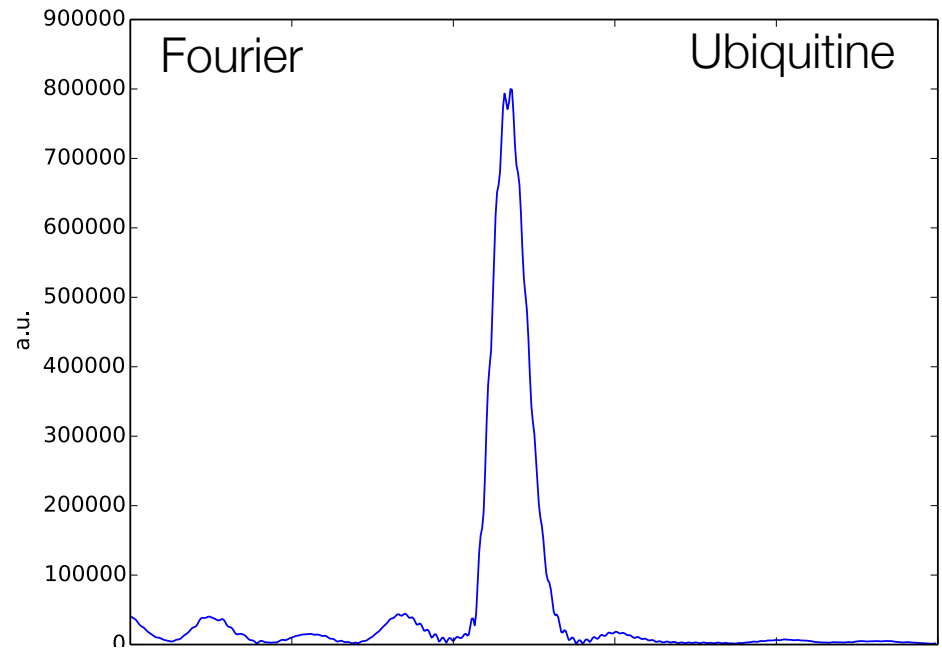
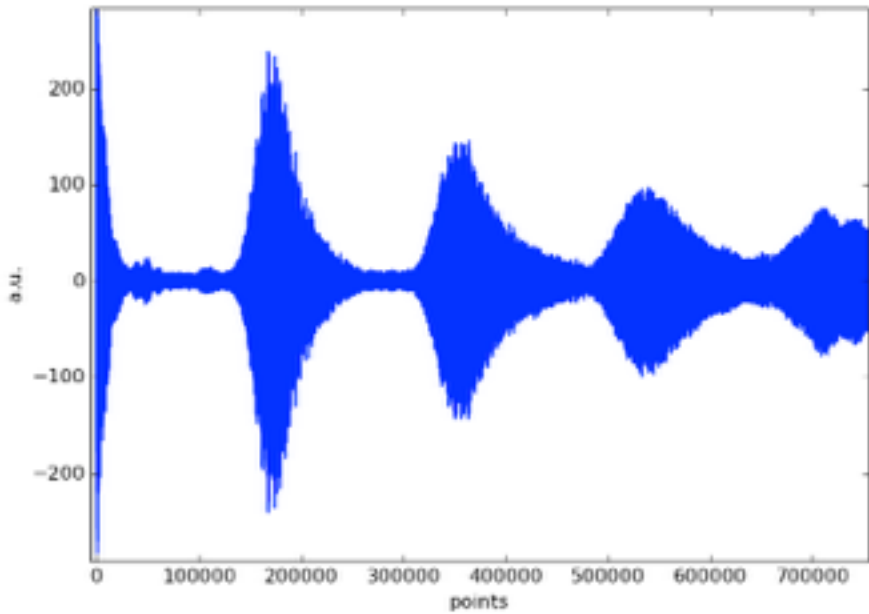


80k : 250msec  
Fourier R= 7.400  
Recital R=19.500

Evolution of resolution  
with acquisition time

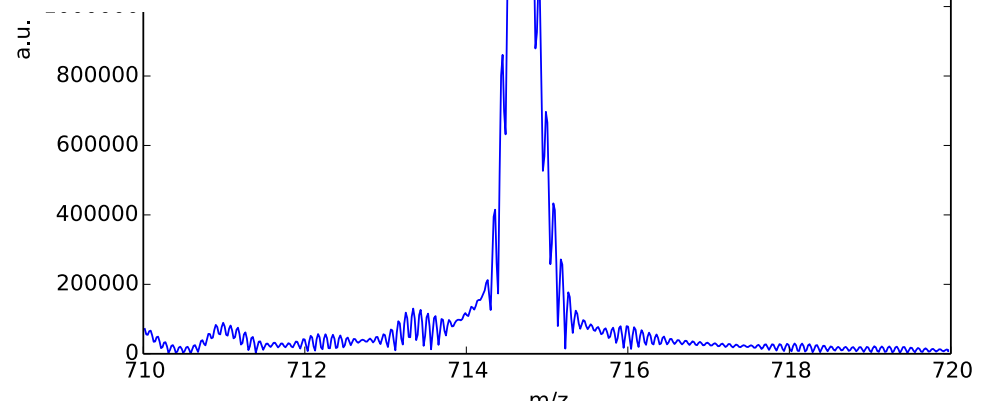
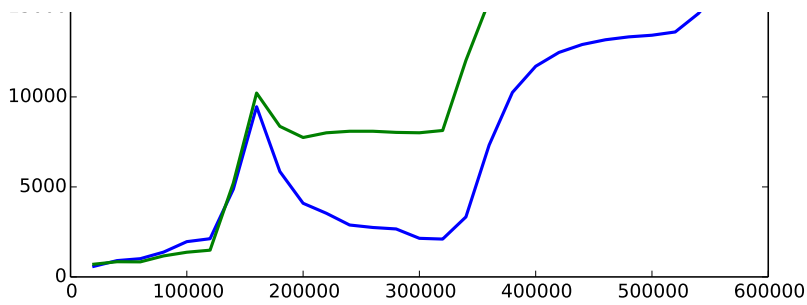
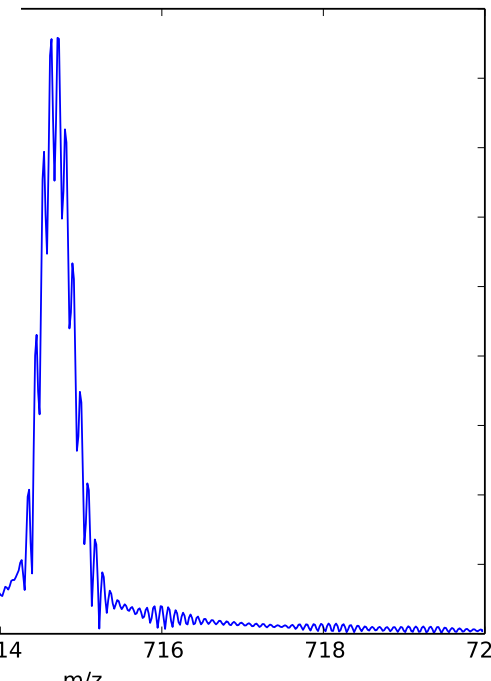
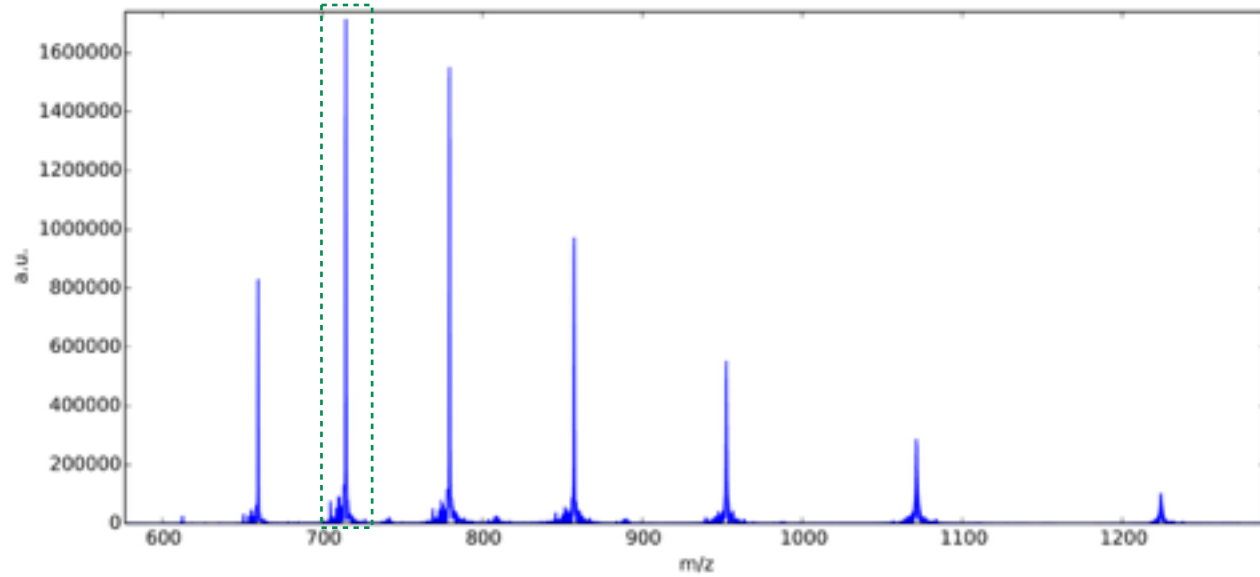
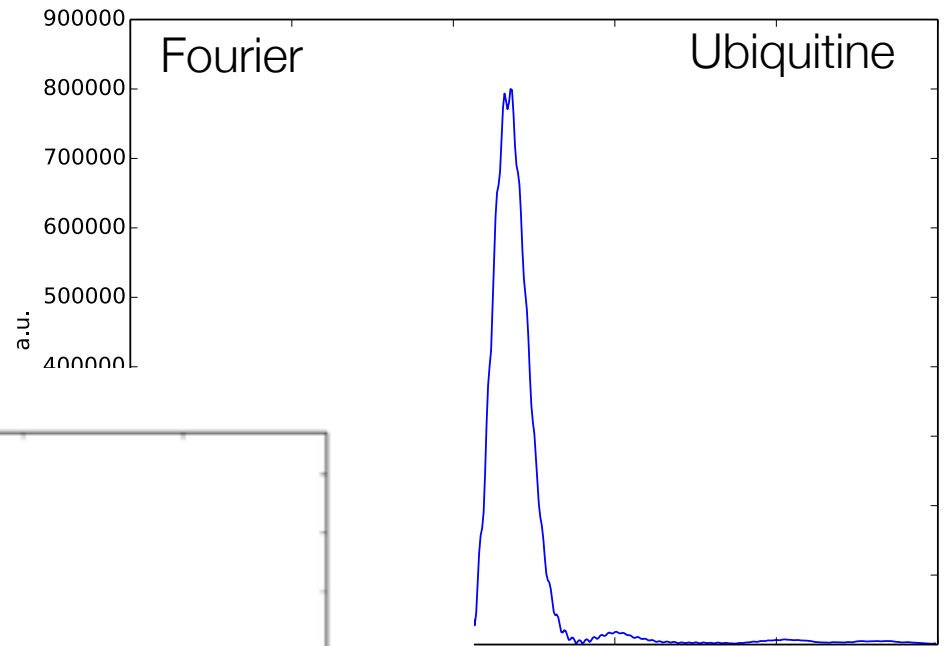
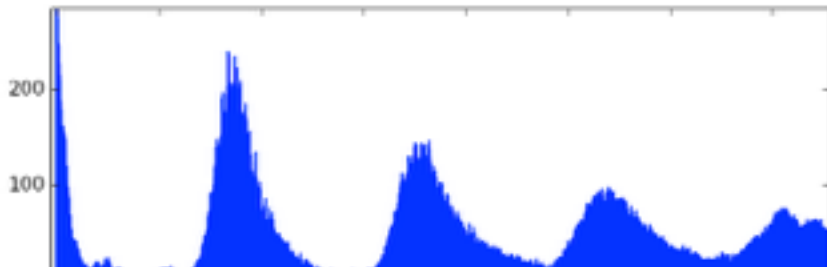
# the same in Orbitrap

73 msec



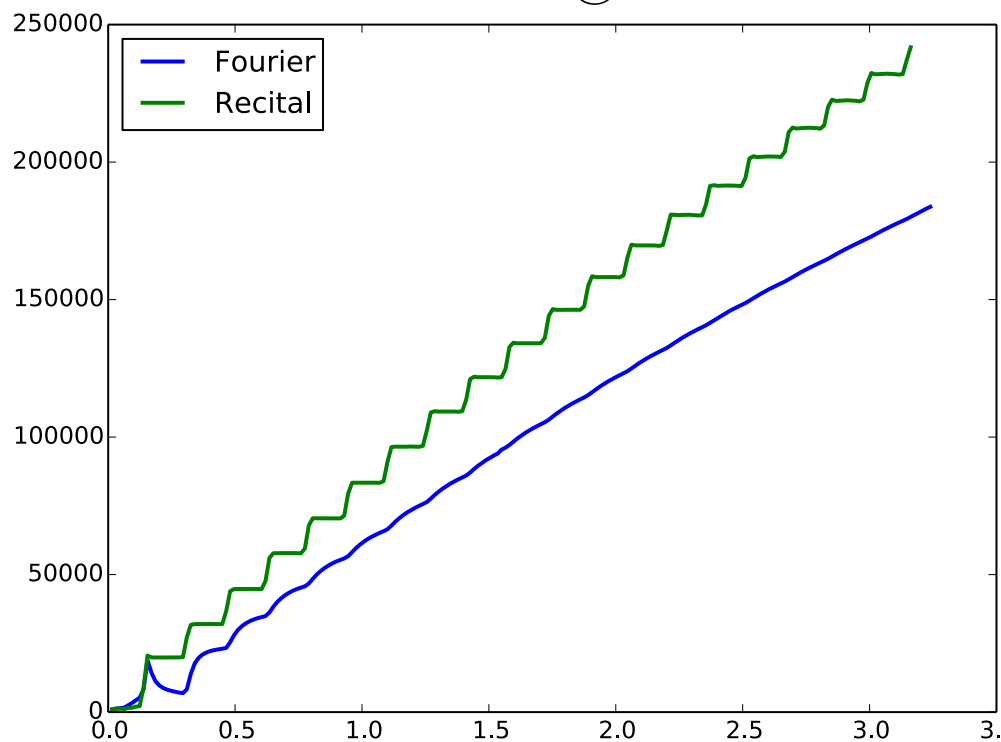
# the same in Orbitrap

73 msec



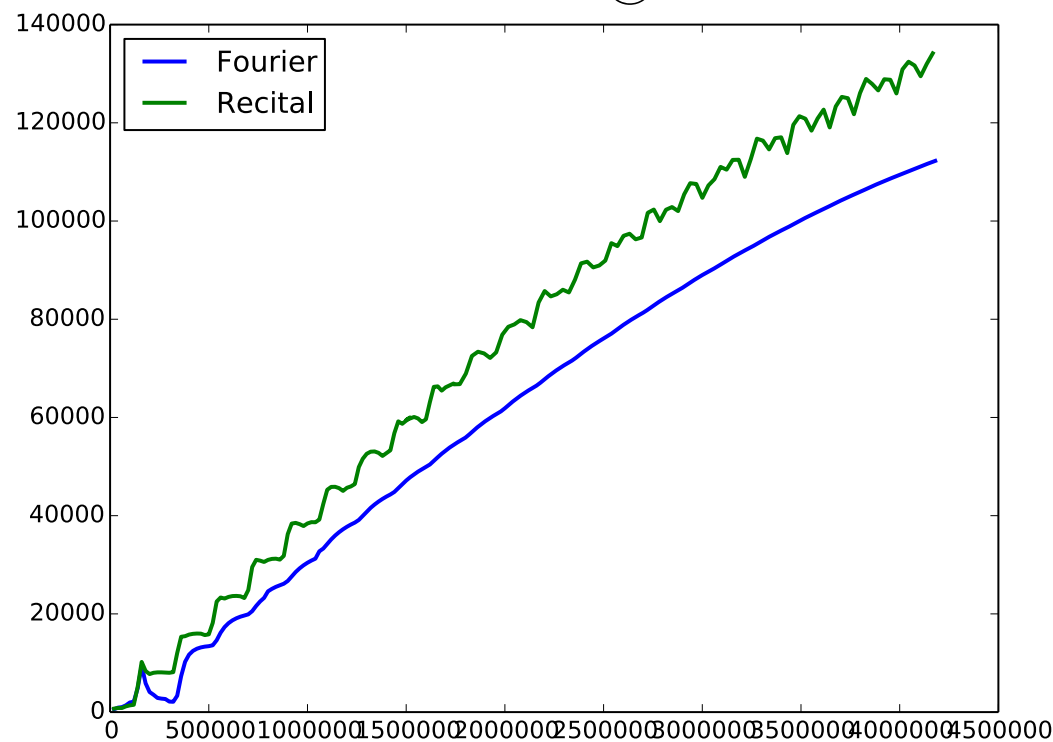
# what about full resolution ?

R = 240.000  
@ m/z = 1590



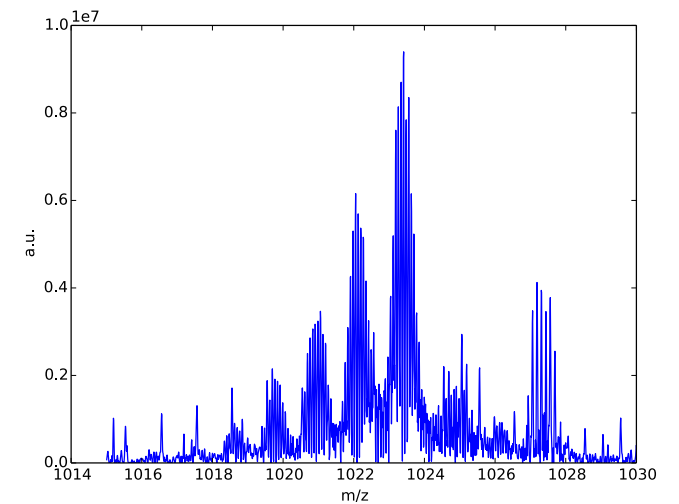
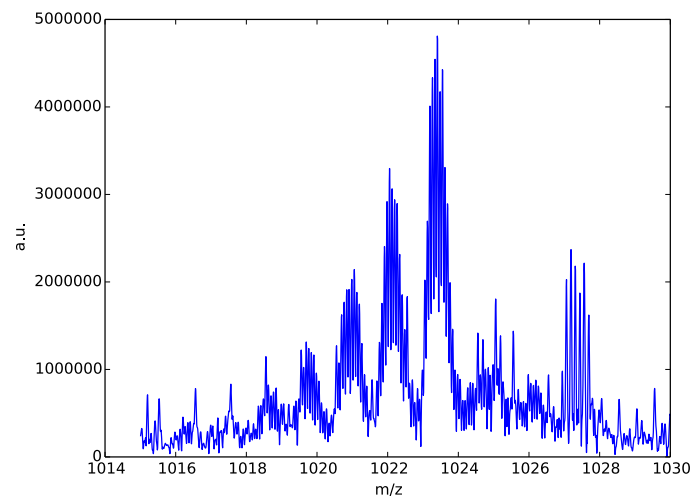
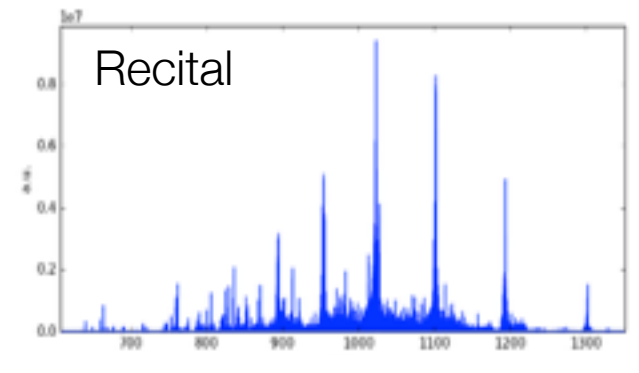
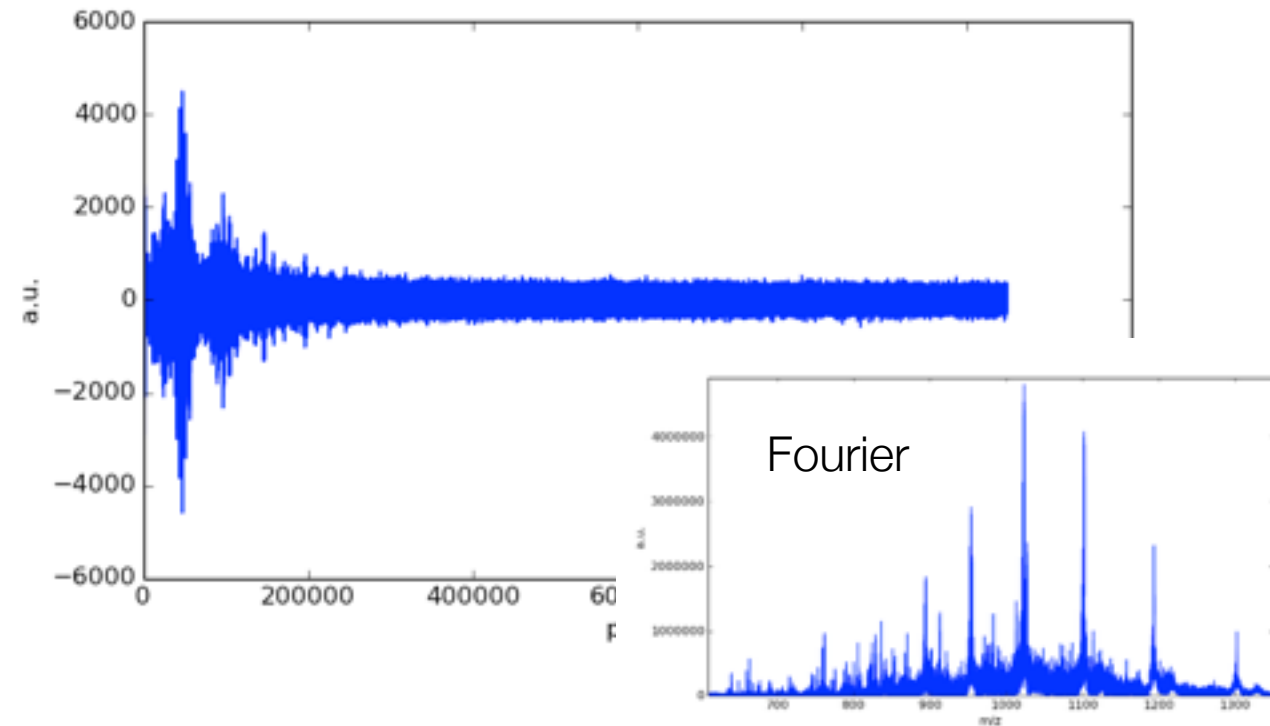
30% faster 2.2 sec vs 3.3 sec  
+30% resolution R=240k vs 180k

R = 133.000  
@ m/z = 715

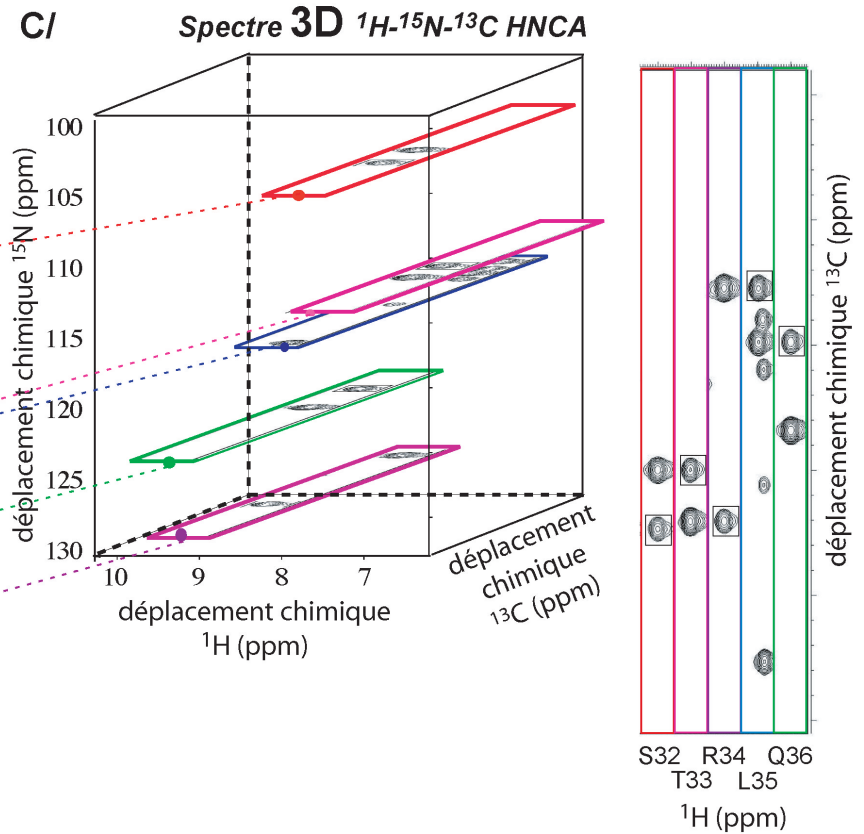
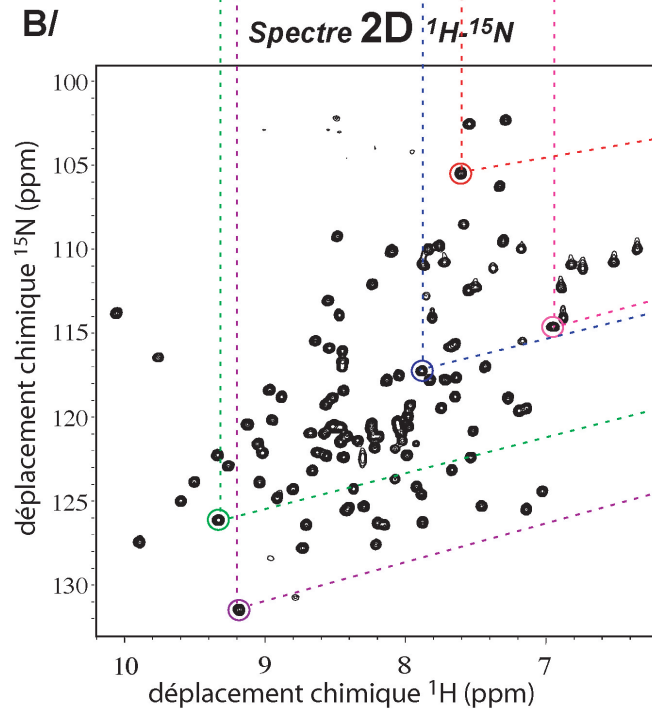
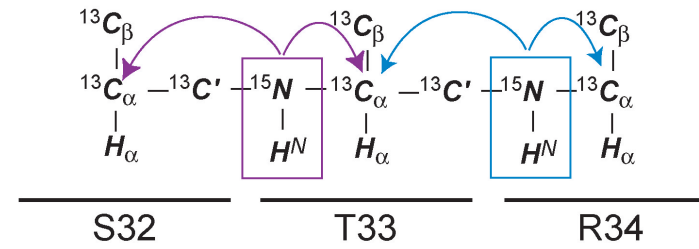
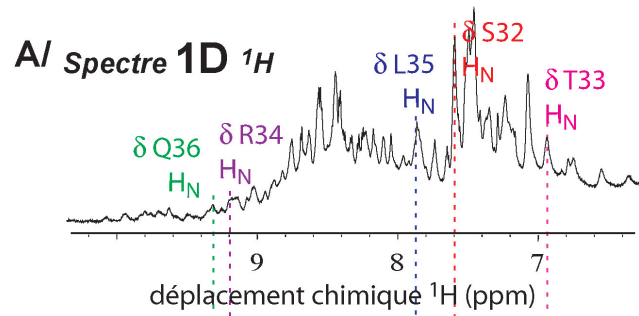


25% faster 1.6 sec vs 1.17 sec  
+18% resolution R=133k vs 112k

# Denaturated Lysozyme is another story



# Spéctroscopie n-Dimensionnelle



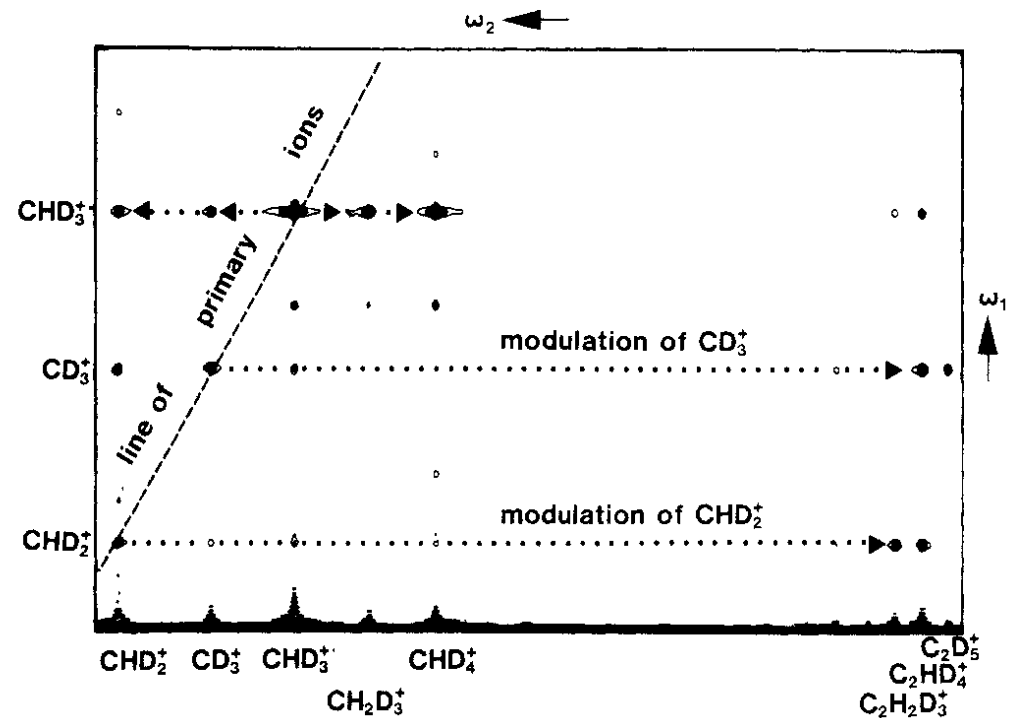
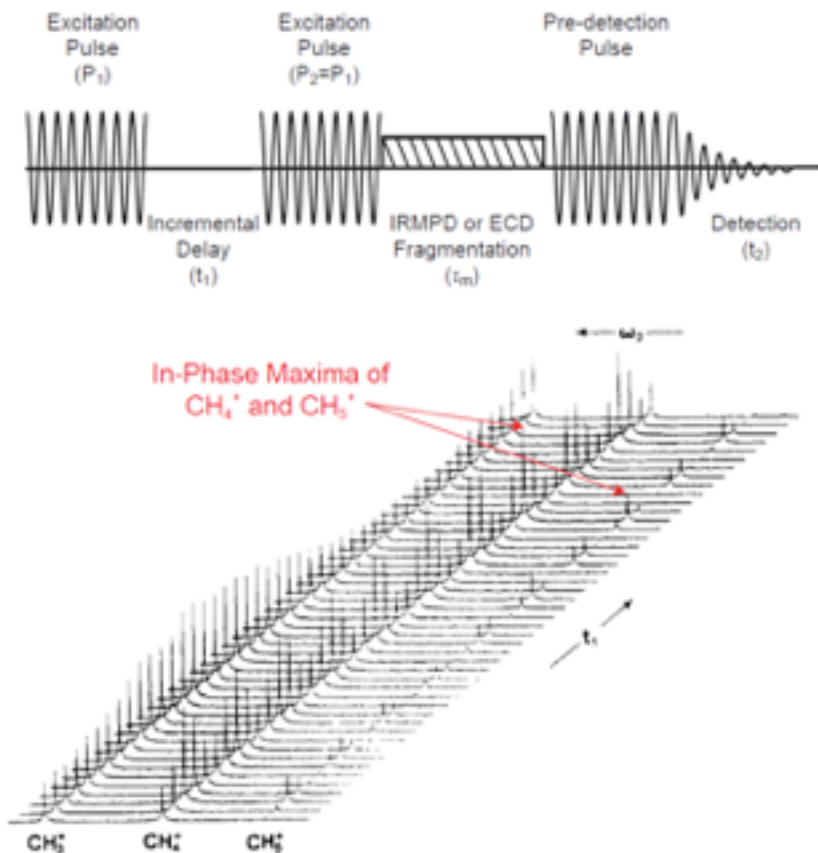


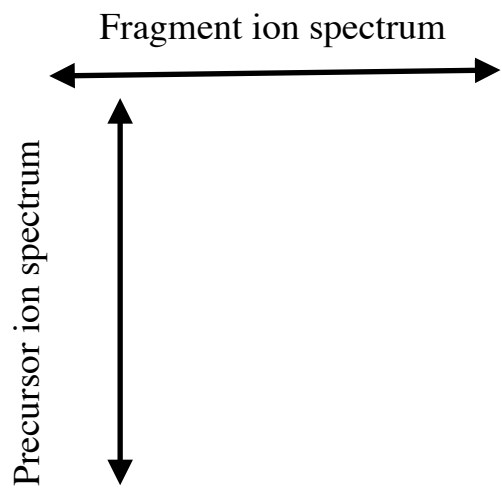
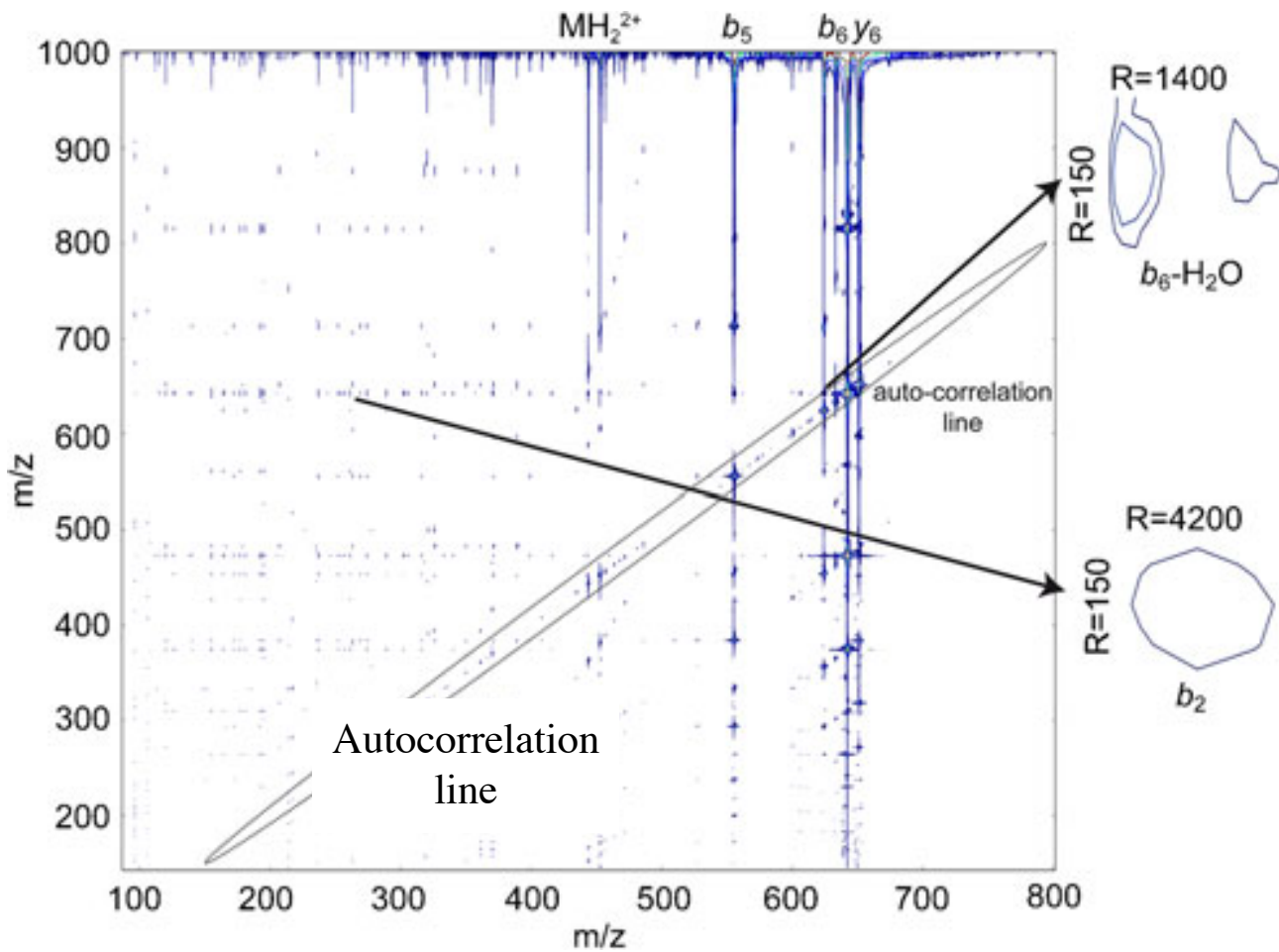
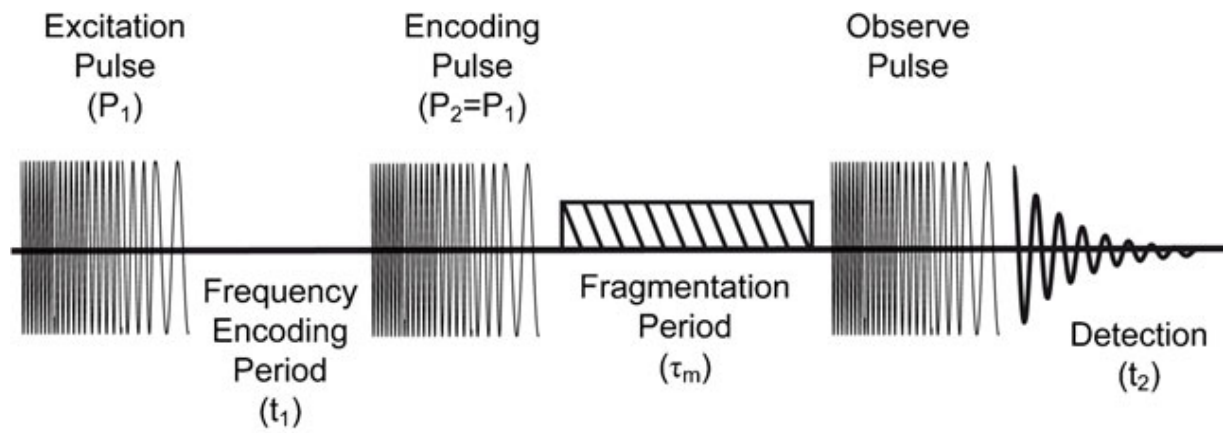
# FTICR-2D

- principle of 2D FTICR proposed in 1987-88  
nearly as old as 2D NMR

P Pfändler and G Bodenhausen and J Rapin and R  
Houriet and T Gäumann  
*Chem Phys Let* (1987) vol. 138 (2) 195-200

P Pfaendler, G Bodenhausen, J Rapin, M Walser, T Gäumann  
*J Am Chem Soc* (1988) vol. 110 (17) 5625-5628





Two-dimensional FT-ICR MS spectrum of bradykinin using IRMPD fragmentation and Cadzow denoising (30 lines). Inserts enlargements of the  $b_6 \rightarrow b_6\text{-H}_2\text{O}$  and the  $b_6 \rightarrow b_2$  peaks.

van Agthoven, M. A., Delsuc, M.-A., Bodenhausen, G. & Rolando, C. *Anal Bioanal Chem* **405**, 51–61 (2013).

# Is 2D FTMS useful ?

- ultimate DIA

- Acquisition Times

- ▶ assume MS-MS : 200 scans / 2k increment 2D FTICR

peaks to fragment	MS-MS	2D FTICR
5	1 min	30 min
50	10 min	30 min
500	2 hours	30 min
5000	20 hours	30 min

- 2D FTICR

- Redundancy

- all fragments are measured only once - in one 2D peak.

- Sensitivity

- Fourier Gain on all peaks : as if each fragments has been measured 30 min (x 0.5)

- Resolution

- ▶ MS1 : R ~200      F1 in 2D : R ~1000
    - ▶ 2D spectroscopic space    MUCH LARGER

# Resolution in 2D

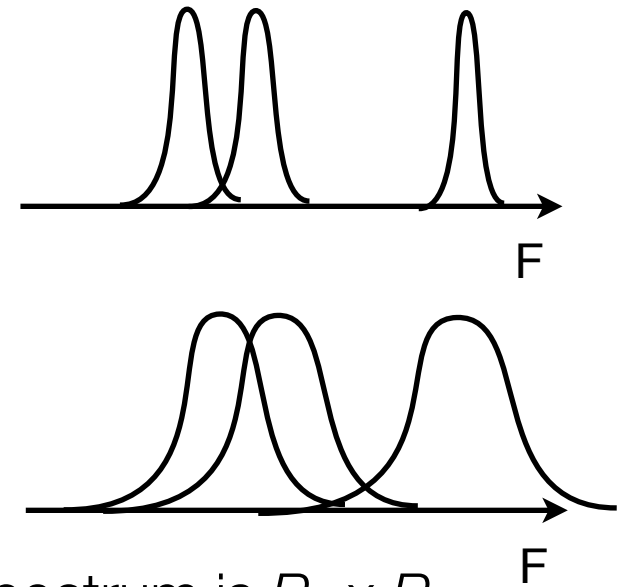
- Resolution

- $R$  is related to the *maximum* number of signals that can be packed in the spectrum

$$R = \frac{m}{\Delta m}$$

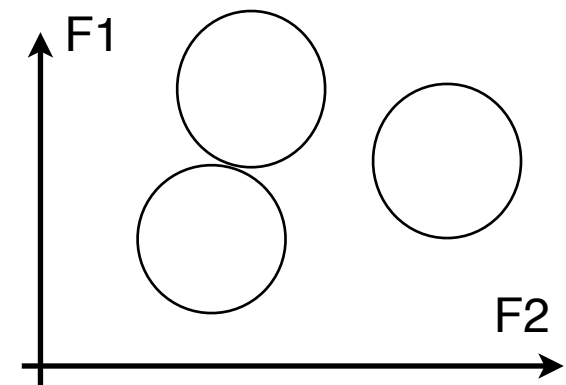
- In 2D  $R$  is asymmetric

- $R_1 \ll R_2 < R_{\text{opt}}$
- eg :  $R_1 = 1.000$   $R_2 = 20.000$



- But

- maximum number of signals to pack in the 2D spectrum is  $R_1 \times R_2$
- eg :  $R_{2D} = 20.000.000$



# Different optimisations

- Fragmentation

- initially by CID (1987)

- reduces resolution

- IRMPD

- van Agthoven, M. A., Delsuc, M.-A. & Rolando, C. Two-dimensional FT-ICR/MS with IRMPD as fragmentation mode. *International Journal of Mass Spectrometry* **306**, 196–203 (2011).

- ECD

- van Agthoven, M. A., Chiron, L., Coutouly, M.-A., Delsuc, M.-A. & Rolando, C. Two-Dimensional ECD FT-ICR Mass Spectrometry of Peptides and Glycopeptides. *Anal Chem* **84**, 5589–5595 (2012).

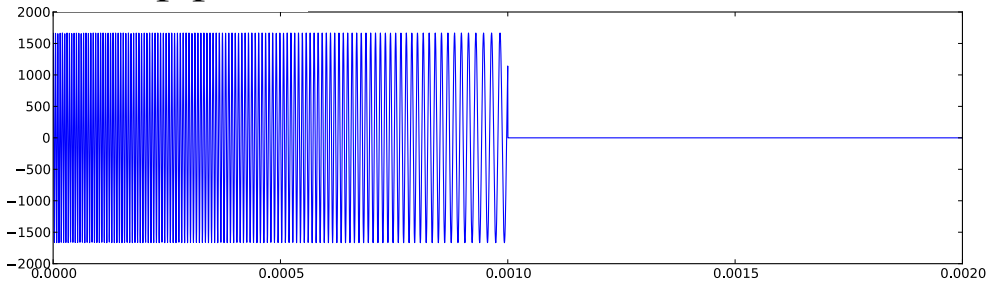
- “Pulse sequences”

- optimisation du signal

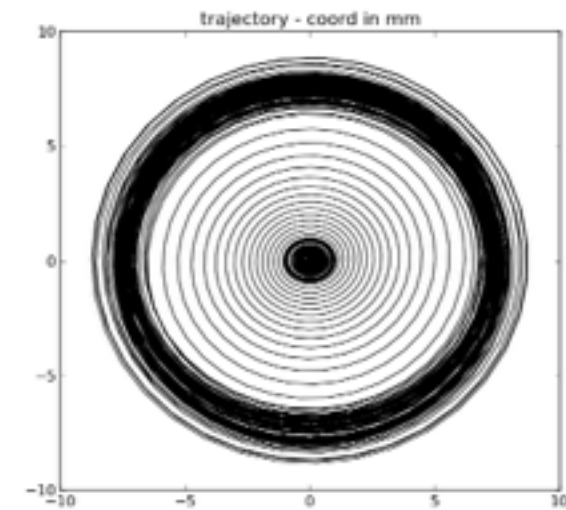
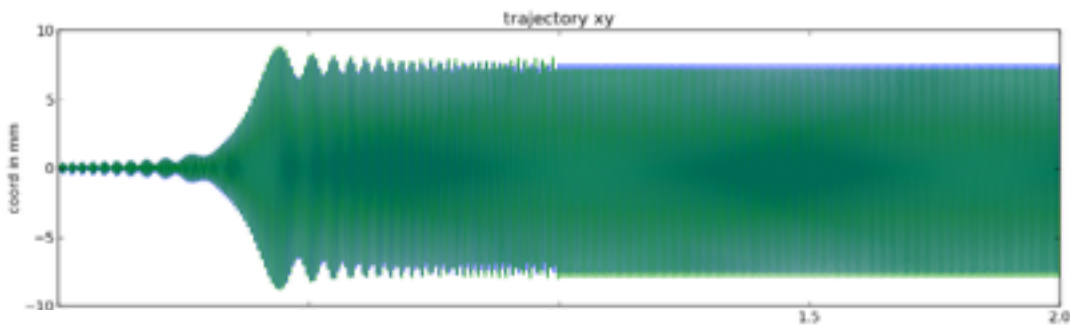
- van Agthoven *et al.* Optimization of the discrete pulse sequence for two-dimensional FT-ICR mass spectrometry using infrared multiphoton dissociation. *International Journal of Mass Spectrometry* (2014).

# FT-ICR simulator

Chirp pulse

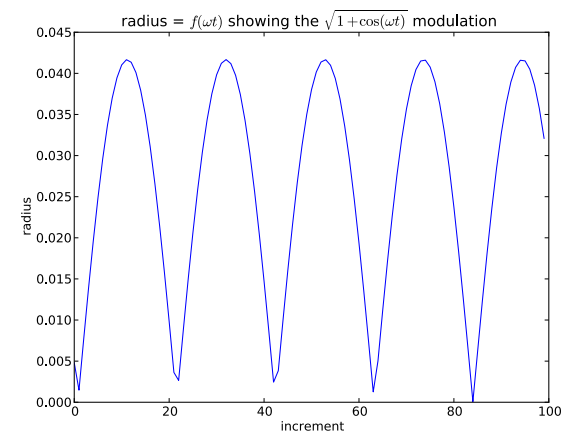
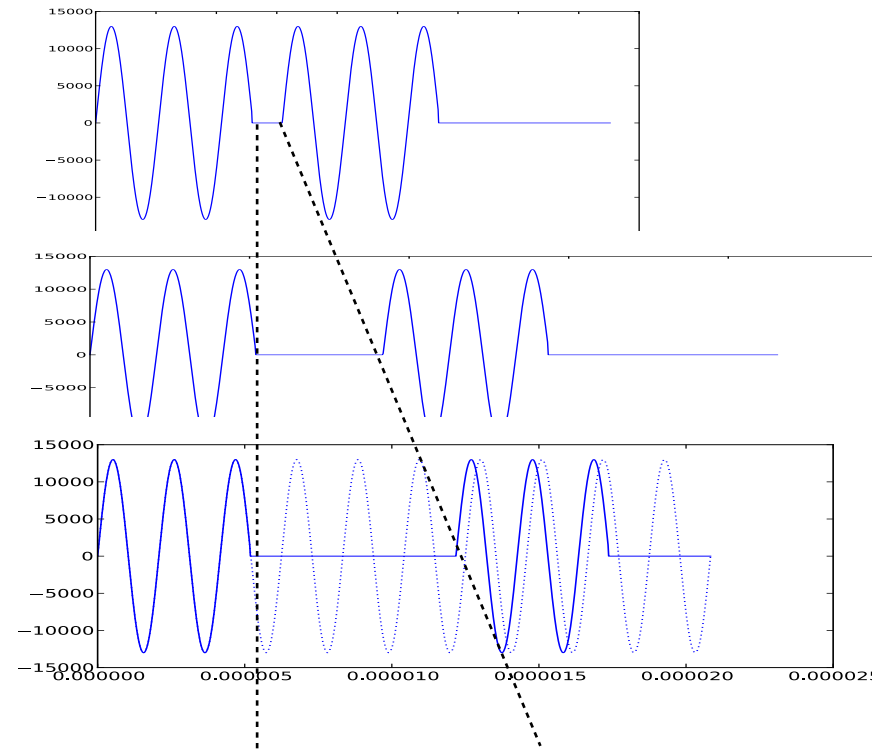


ion trajectory

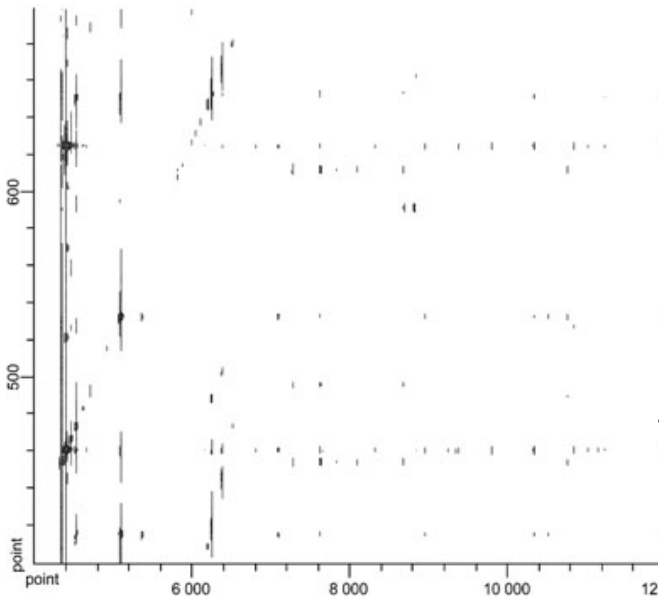


Swept pulse  
 Frequency : 200.000-50.000 kHz  
 sweep width : 150.000 kHz    sweep steps : 1000  
 duration : 1.000 msec  
 Epp : 1666.67 V/m  
 approx excitation radius : 11.82 mm

resonant frequency 144151.41 Hz  
 final radius : 7.56 mm

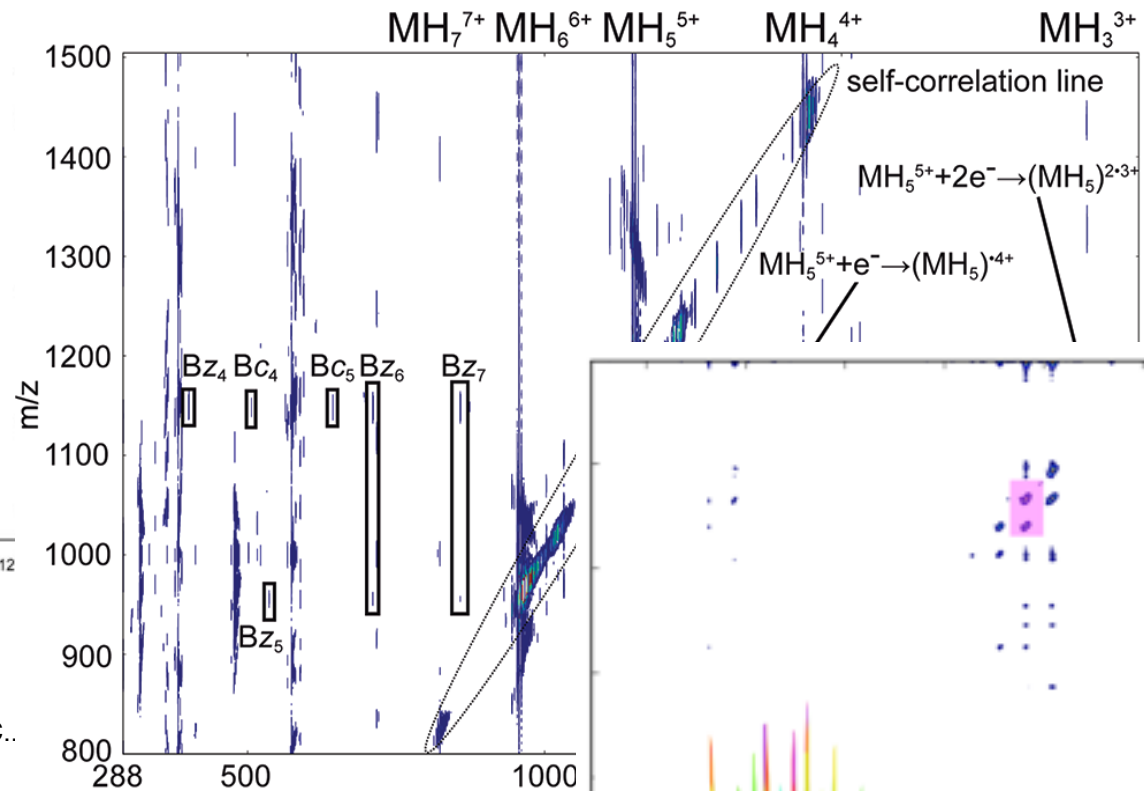


# Improvements over time



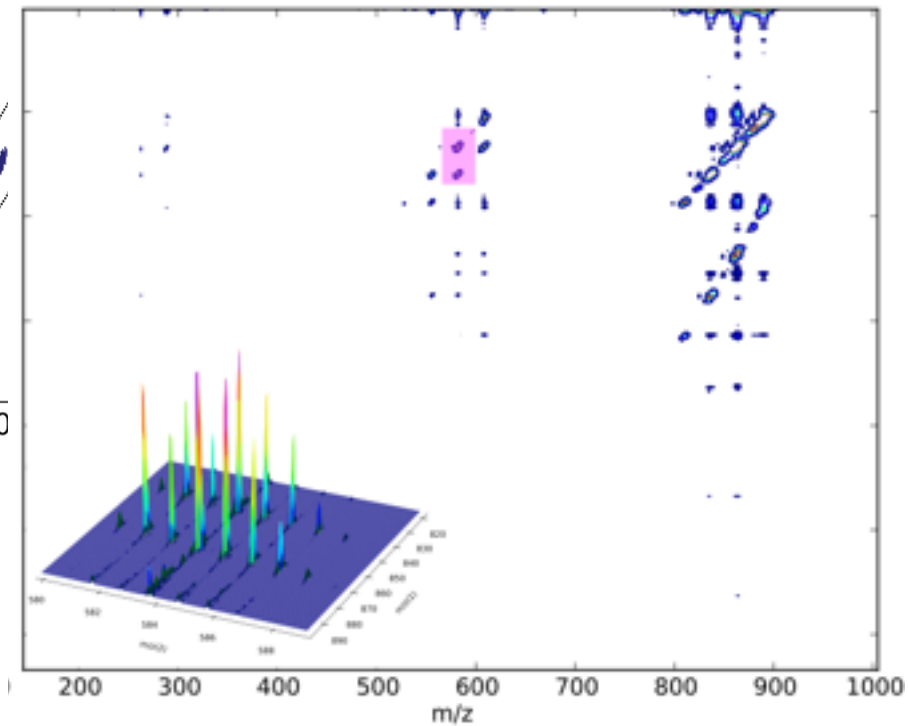
2011 : bradykinin - IRMPD

van Agthoven, M. A. V., Delsuc, M.-A. & Rolando, C..  
*J Mass Spect.* **306**, 196–203 (2011).



2012 : BPTI - ECD

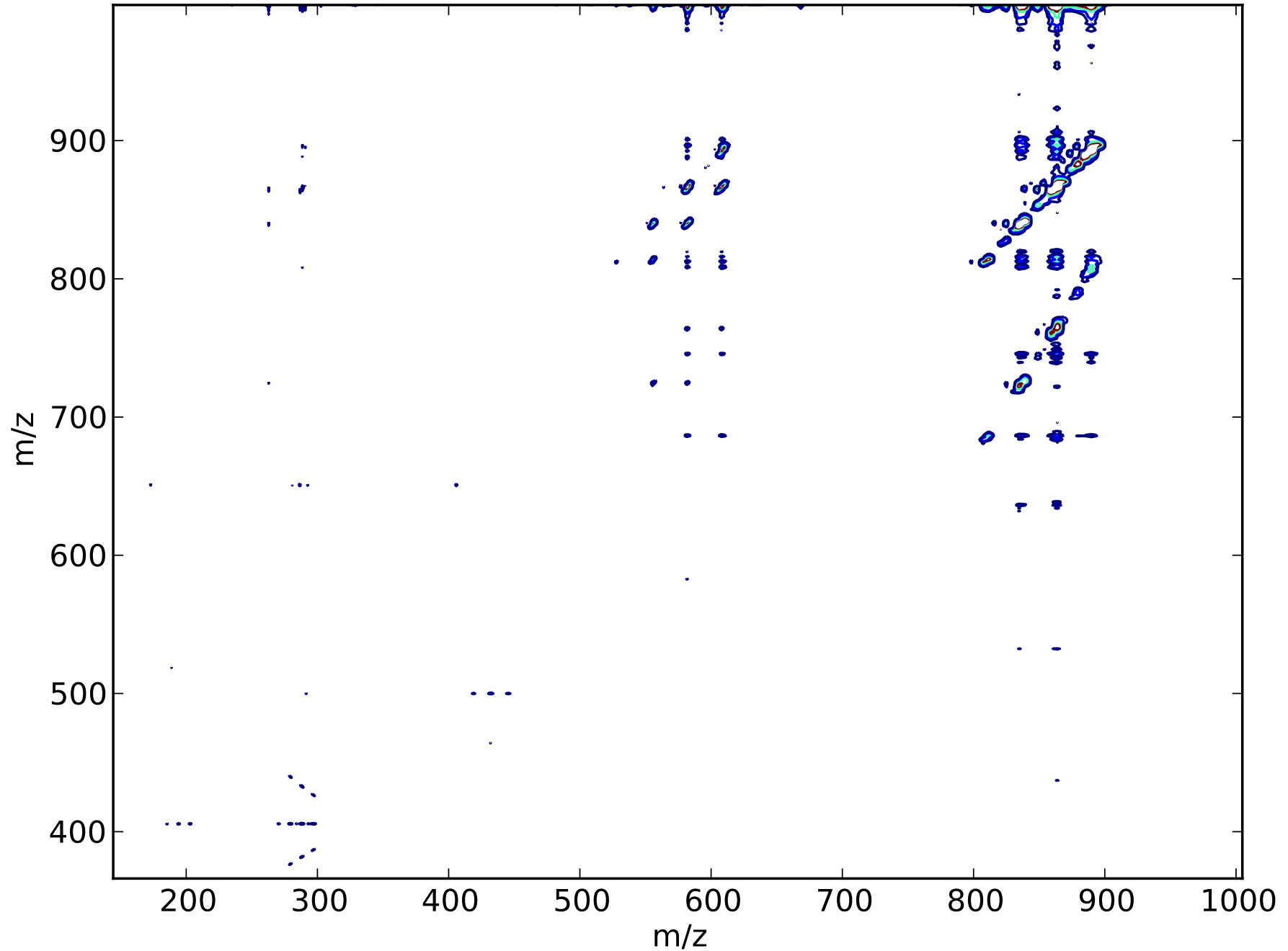
van Agthoven, M. A., Chiron, L., Coutouly, M.-A.,  
Delsuc, M.-A. & Rolando, C.  
*Anal Chem* **84**, 5589–5595 (2012).



2014 : triglycerid - high resolution

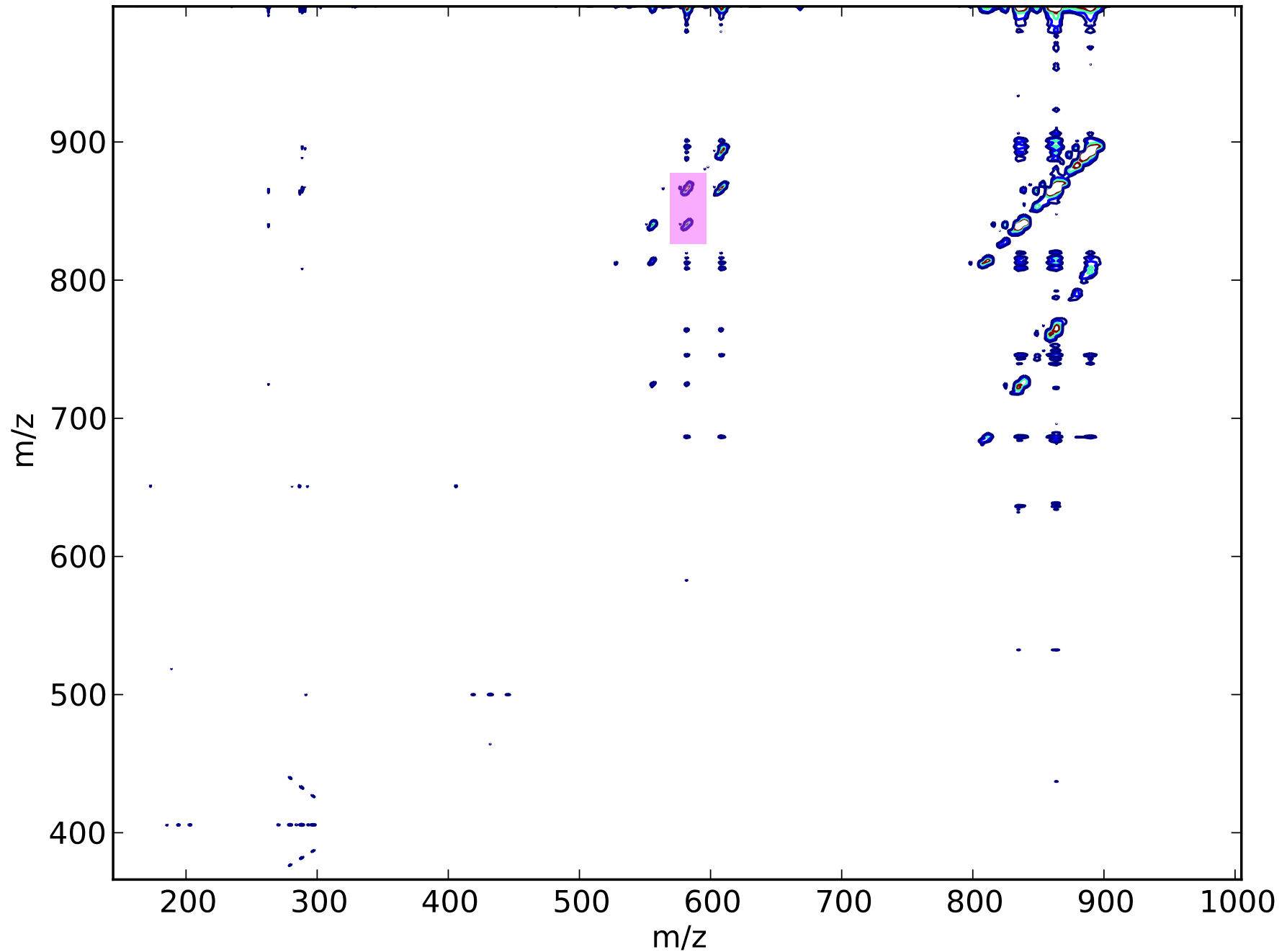
Chiron, L., van Agthoven, M. A., Kieffer, B., Rolando, C. & Delsuc, M.-A.  
*Proc Natl Acad Sci USA* **111**, 1385–1390 (2014).

# Tri Acyl Glycerol (TAG)

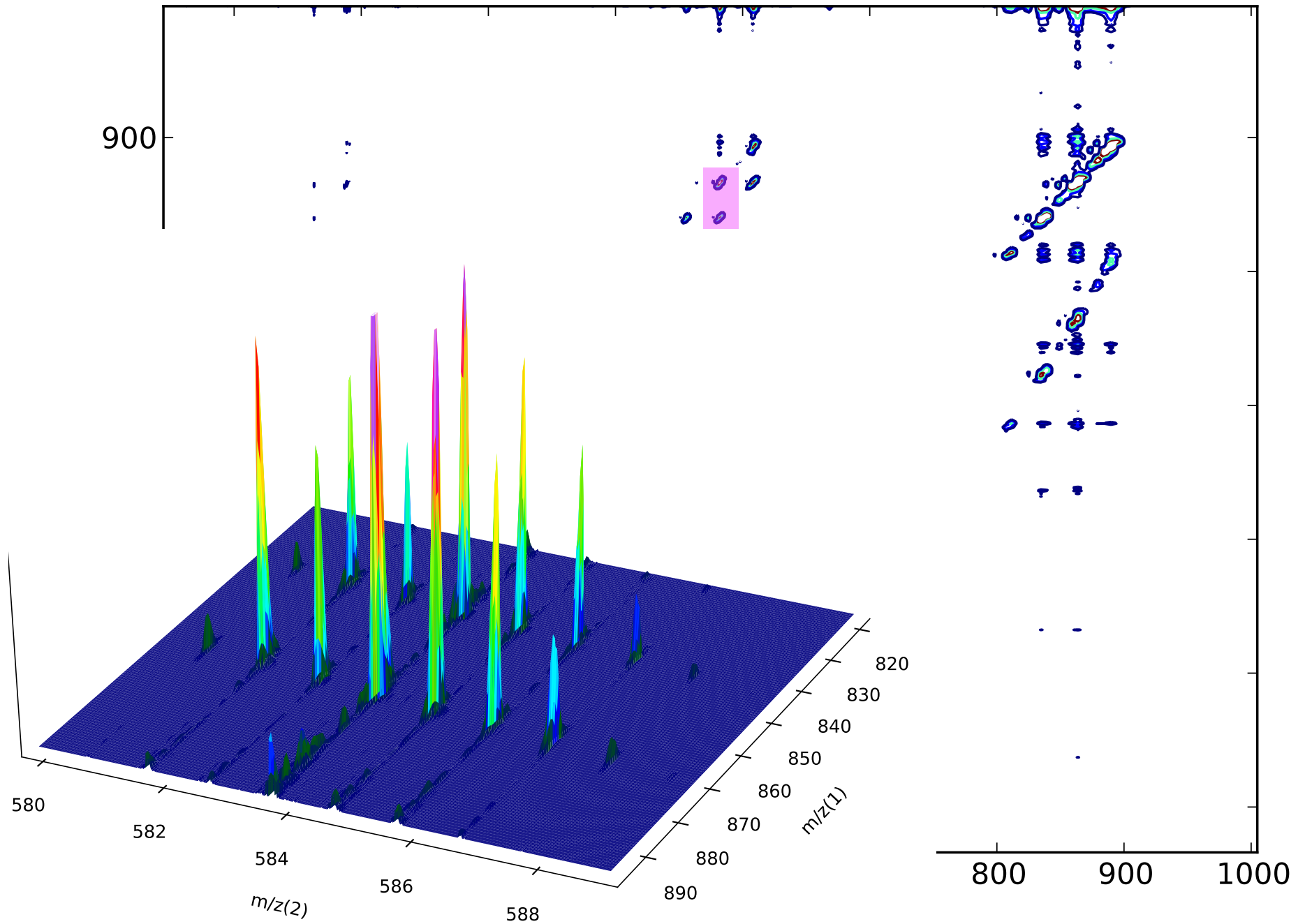




# Tri Acyl Glycerol (TAG)



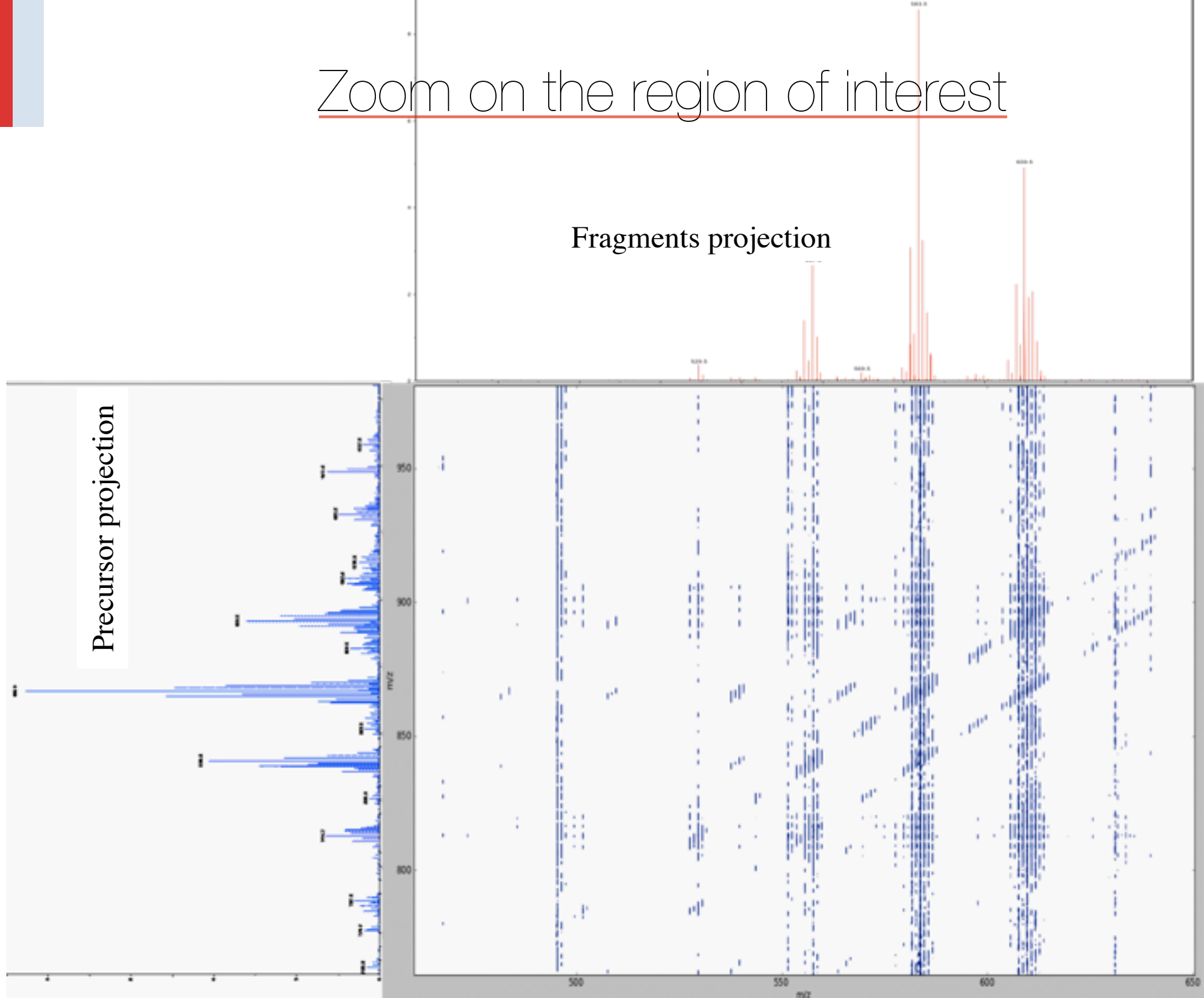
# Tri Acyl Glycerol (TAG)



# Zoom on the region of interest

Fragments projection

Precursor projection

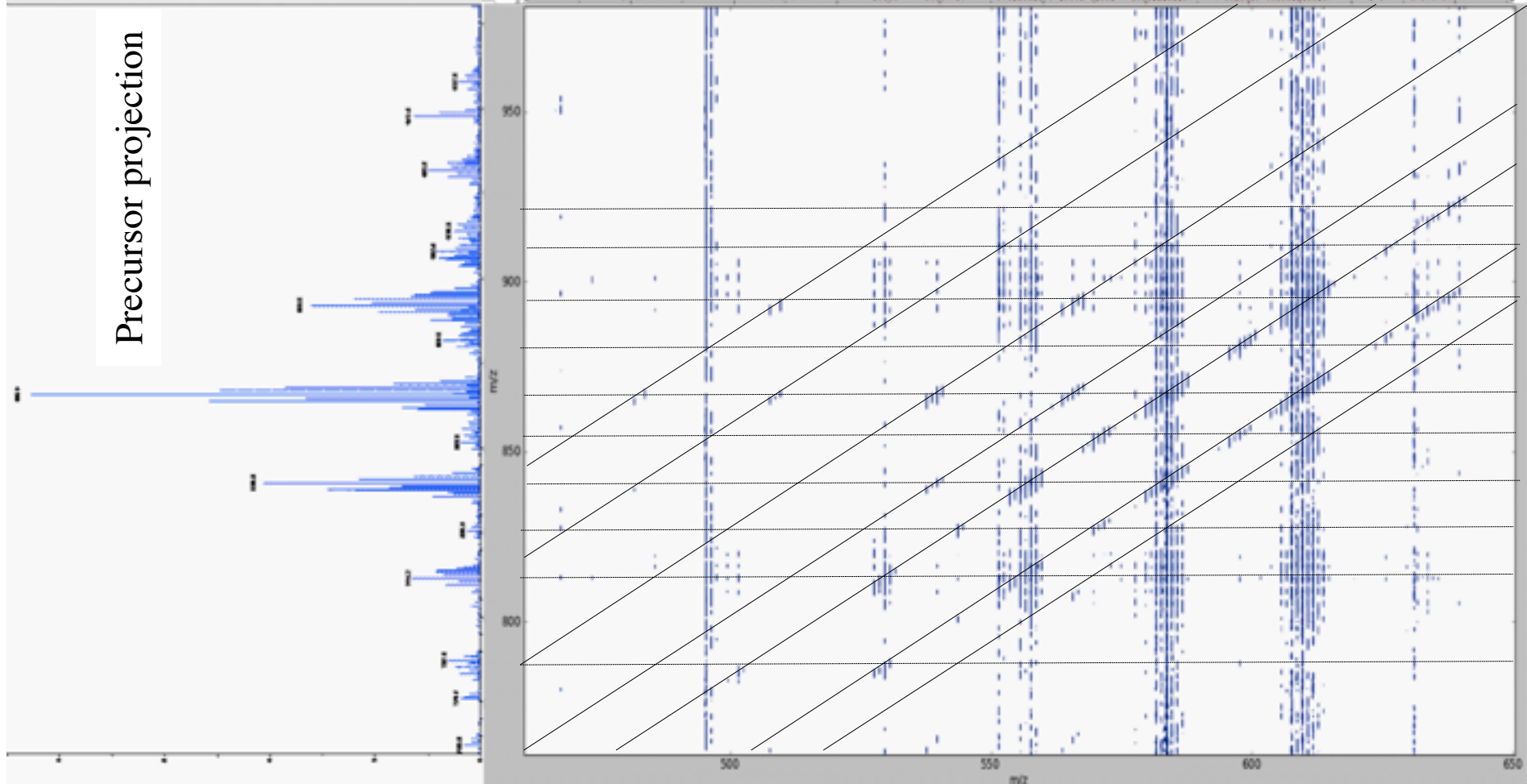


# Zoom on the region of interest

Analysis of fatty acid  
unsaturation ( $\Delta m = 2!$ )

Fragments projection

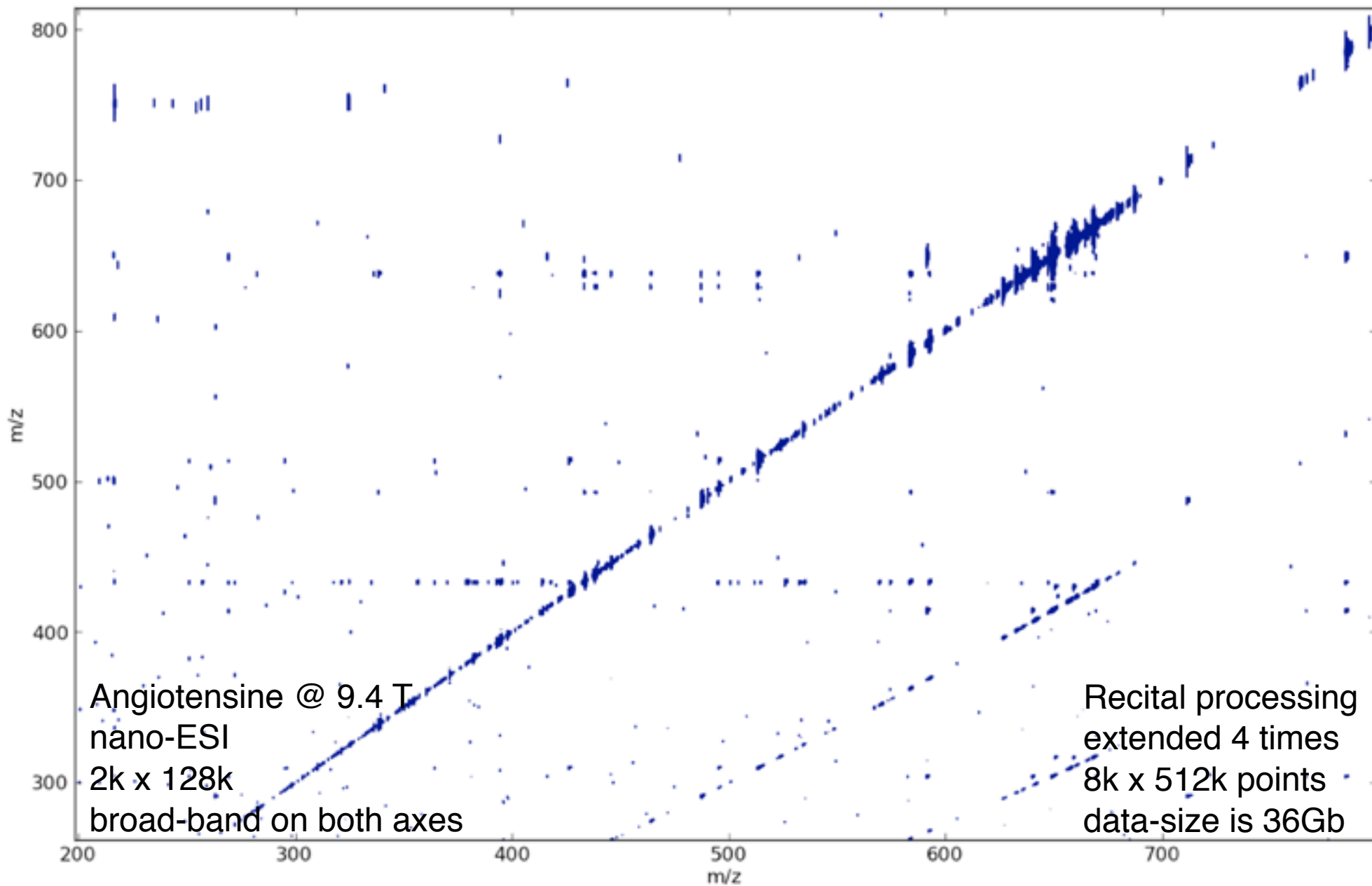
Precursor projection

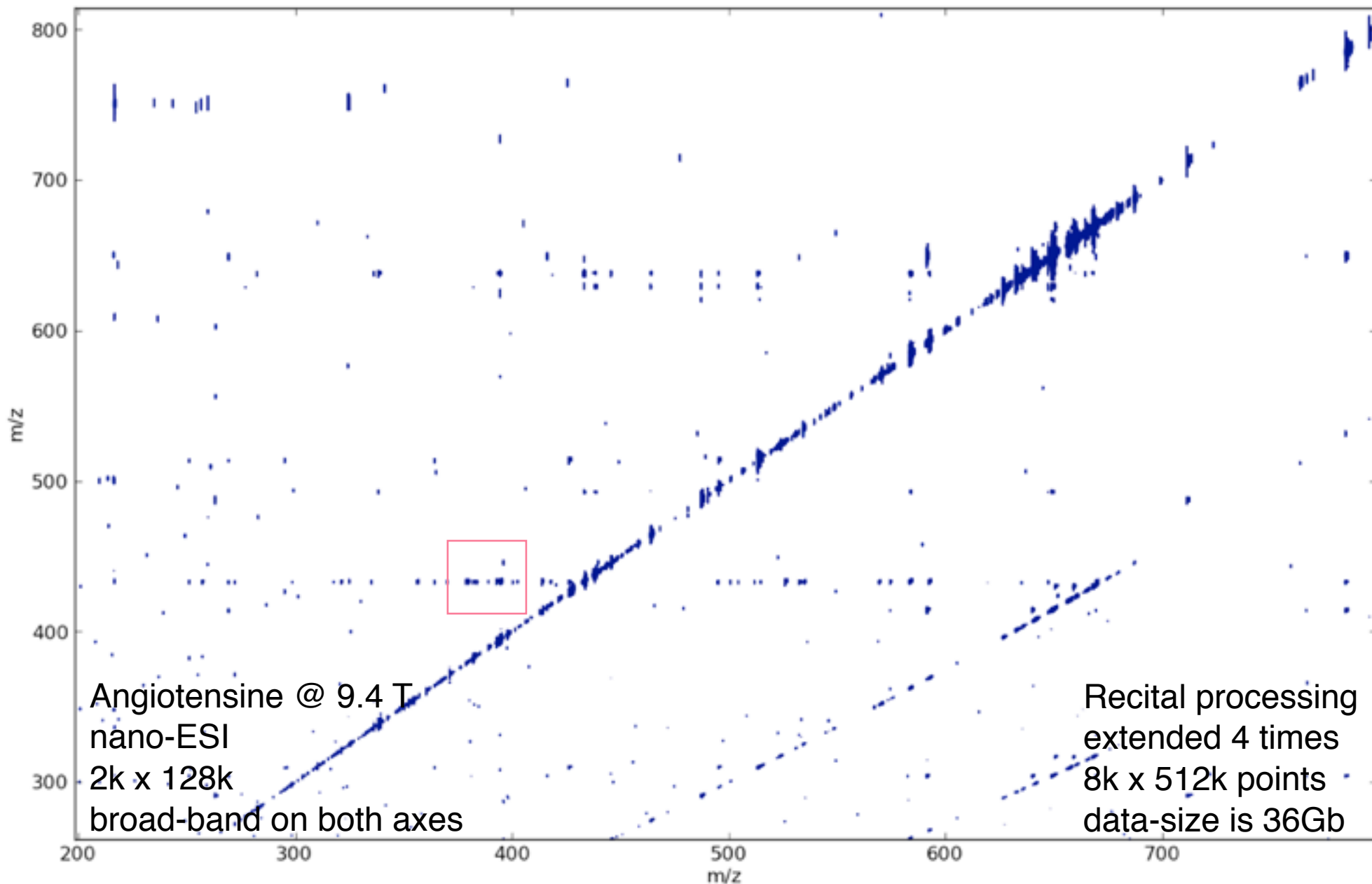


# Practically

- SIZE is a problem
  - ▶ “small” : 2k x 128k ⇒ file size is already 2 Gb on disk
  - ▶ “ideal” : 16k x 512k ⇒ 64 Gb !
- to be processed at once
  - ▶ rely on fast algorithms : FFT
  - ▶ rely on robust and efficient processing language : python
  - ▶ rely on sophisticate file management : NASA HDF5 file format
  - ▶ rely on dedicated display : hierarchical “google-maps-like” display engine
- be performed “at home”
  - ▶ able to run on desktop
- Resolution is a goal
  - Goes with size
    - ▶ Gabor-Heisenberg relation
  - longer  $t_1$  ⇒ longer acquisition

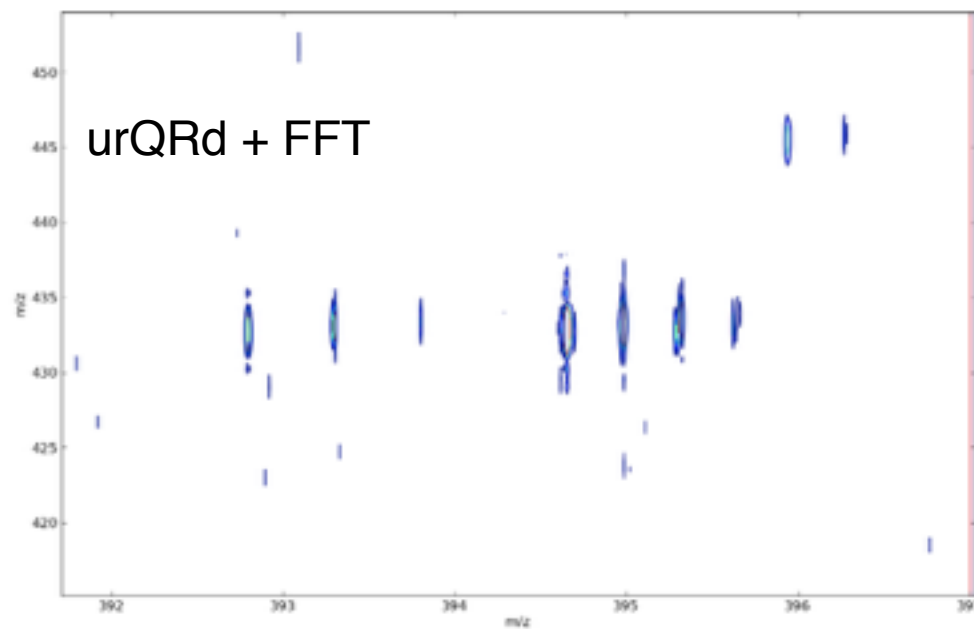
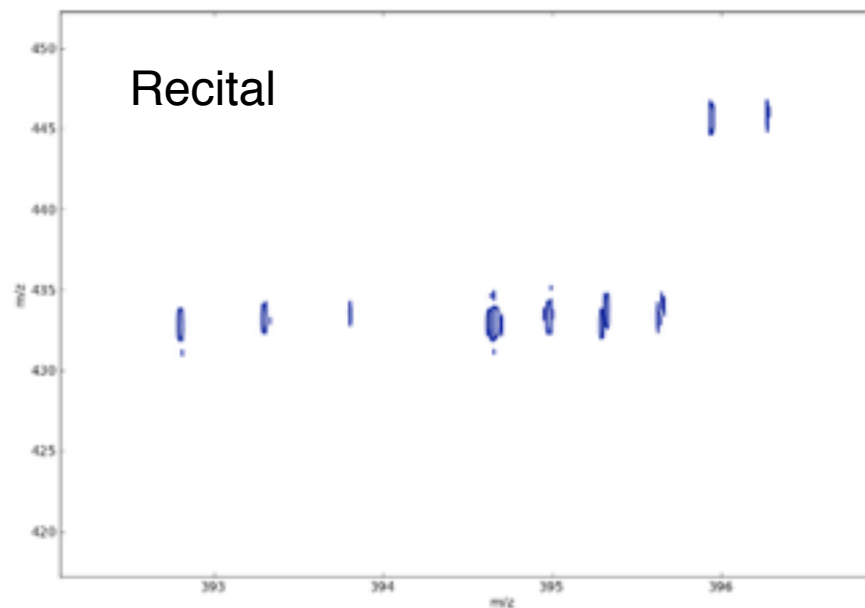
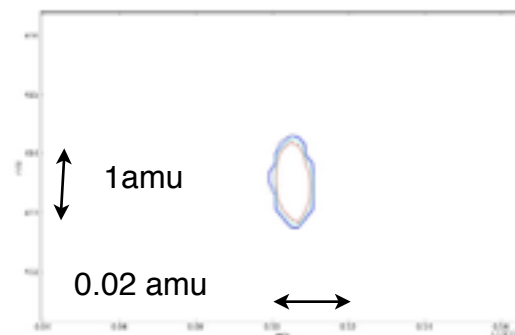
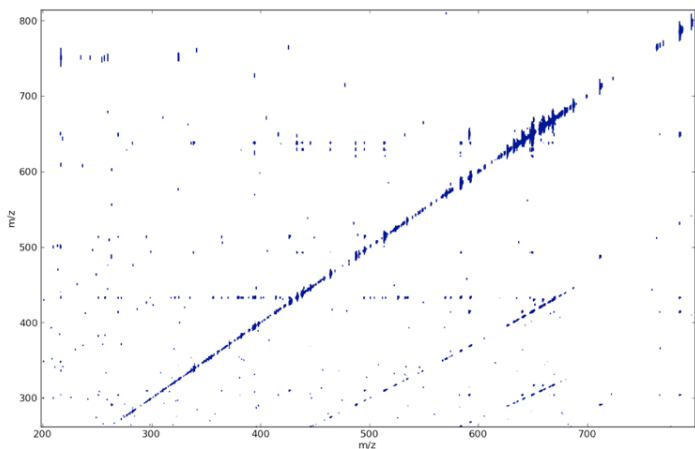
$$T_{\max} \Delta F = 1$$





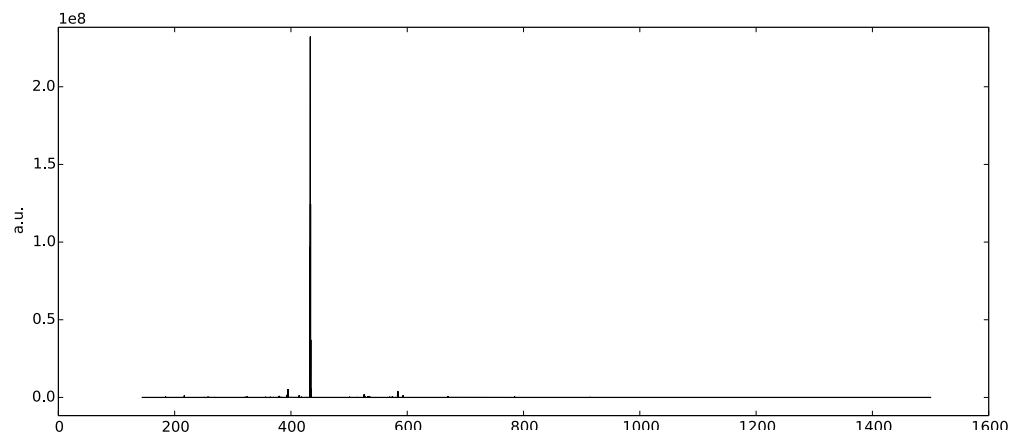
# 2D Recital

R1=500  
R2=50.000

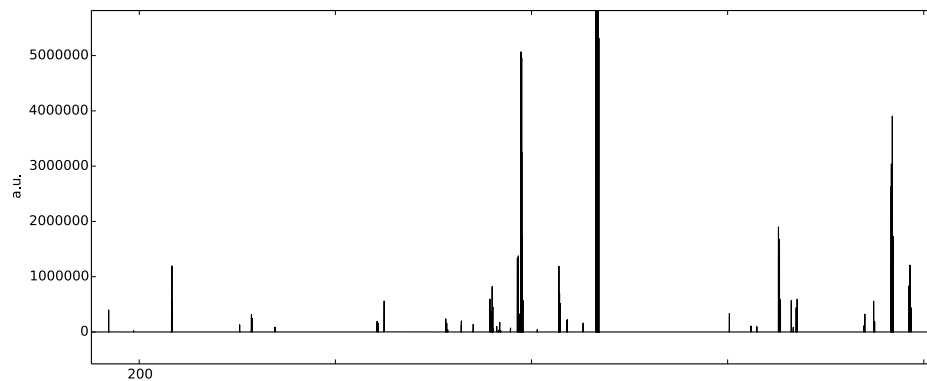




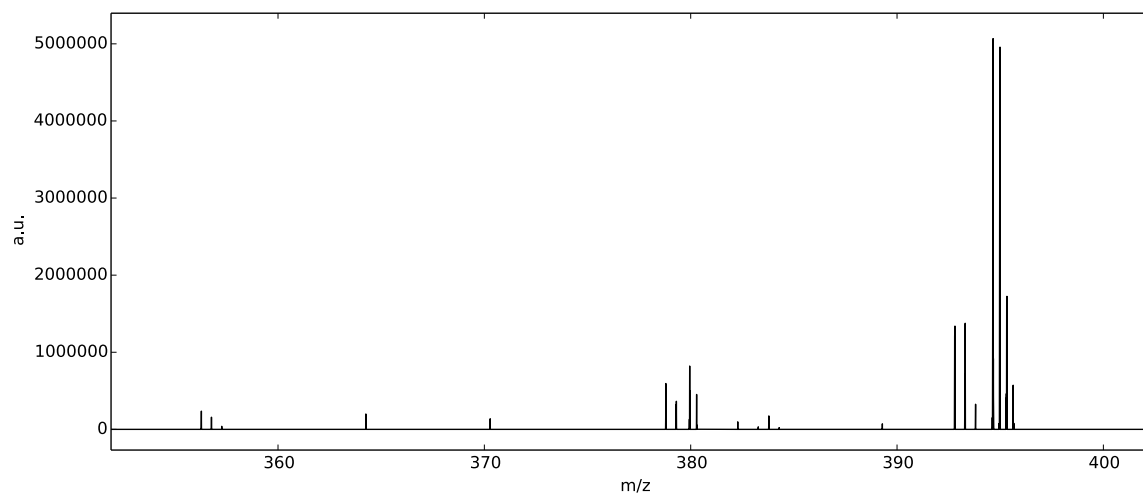
# Looking along F2



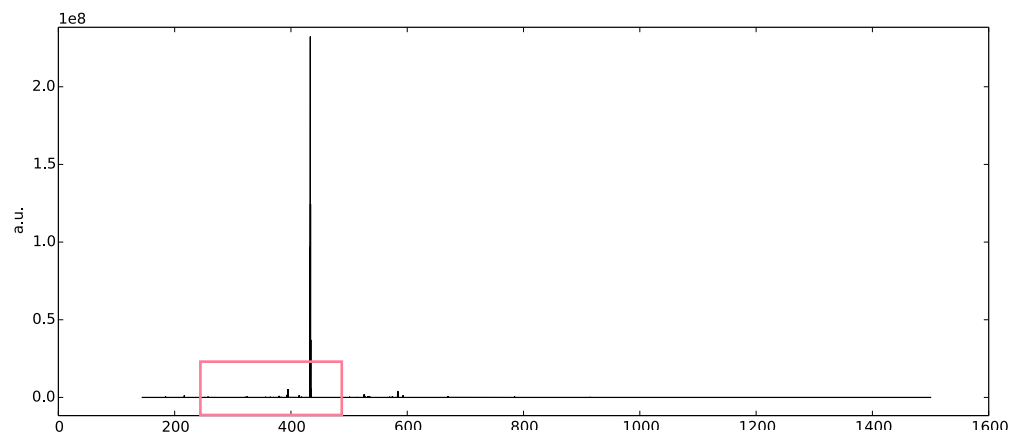
F2 dimension  
horizontal spectra  
fragments



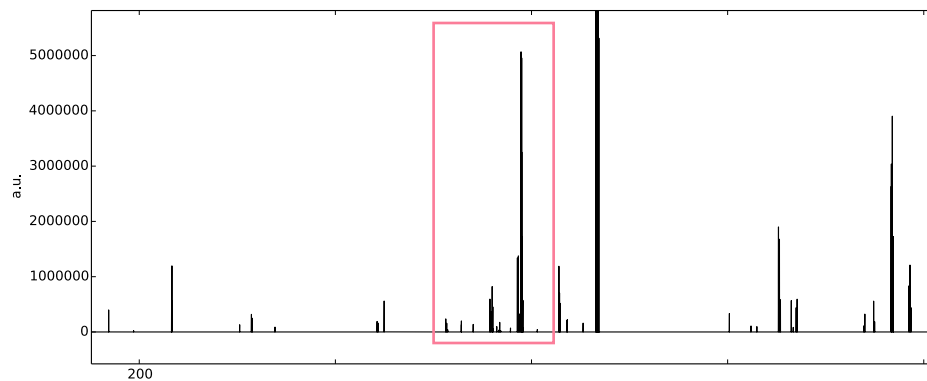
row at  
m/z F1 = 433.3



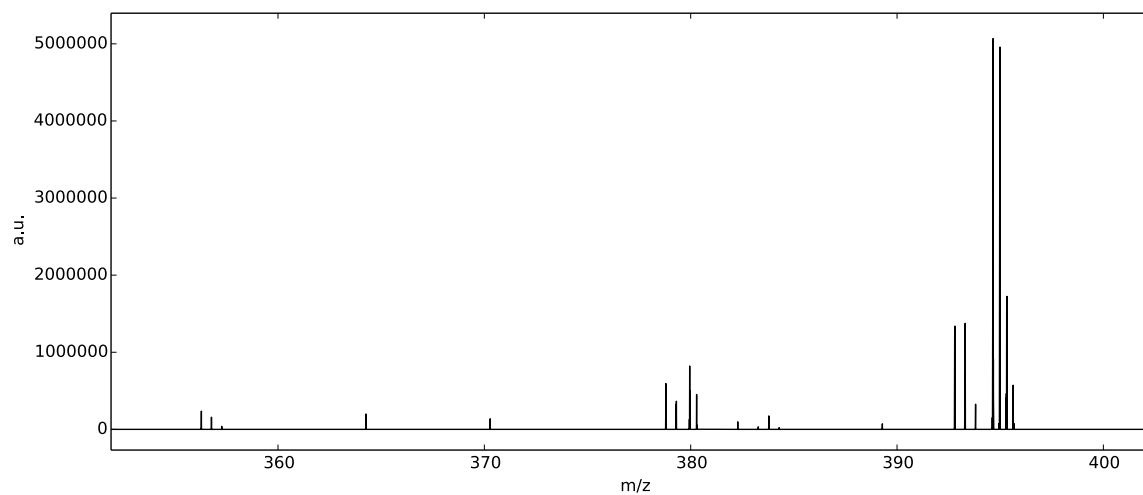
# Looking along F2



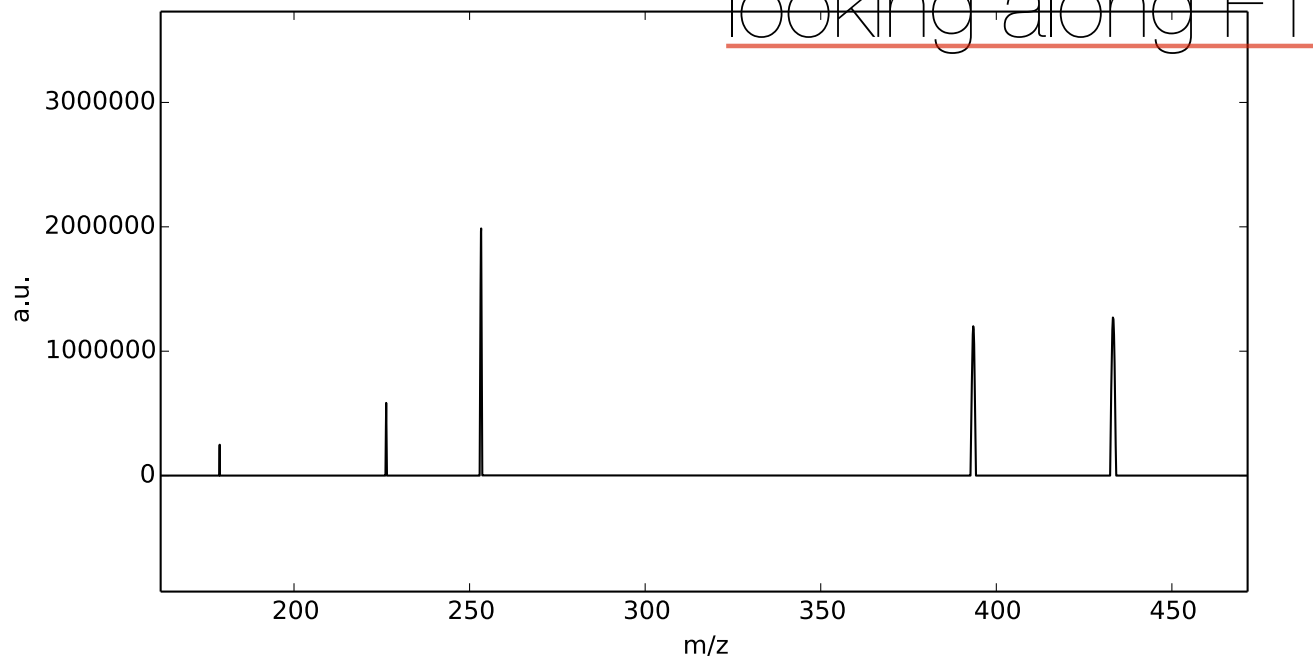
F2 dimension  
horizontal spectra  
fragments



row at  
 $m/z$  F1 = 433.3

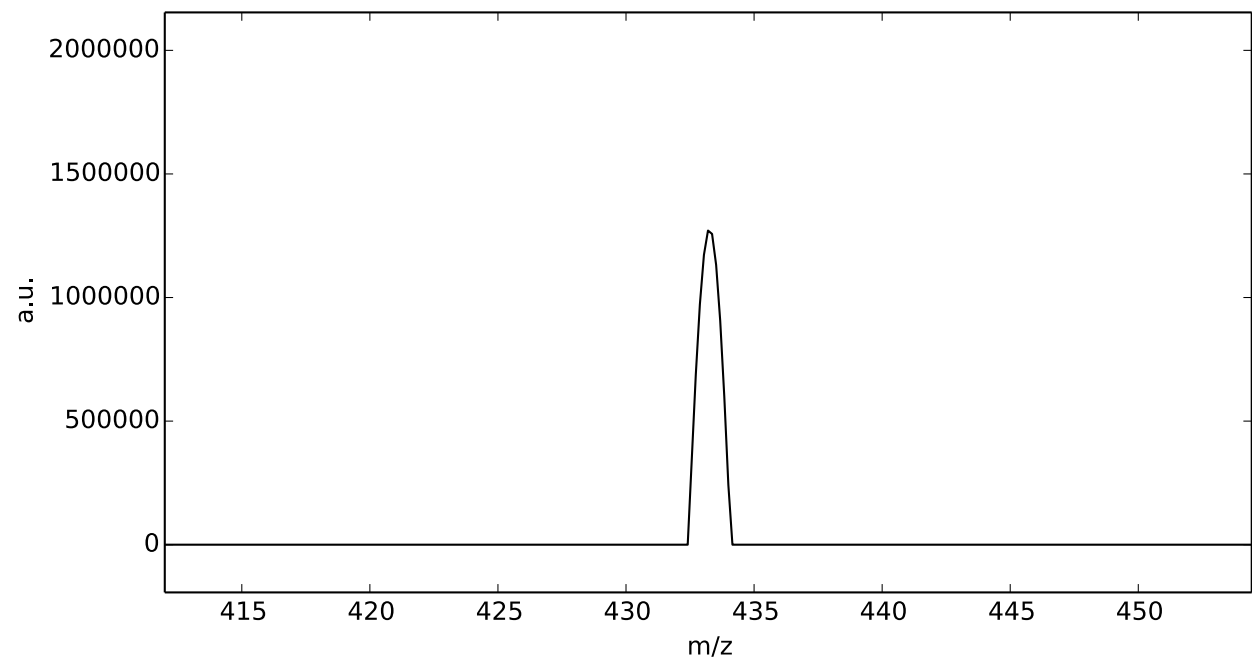


looking along F1

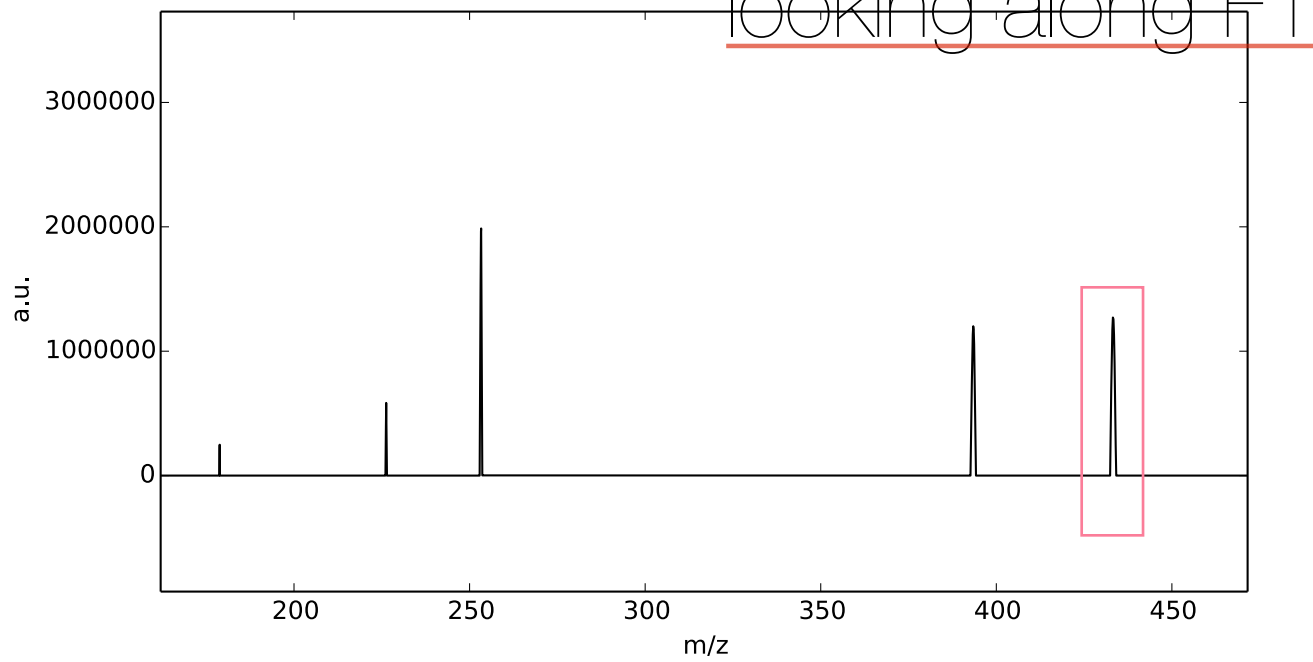


F1 dimension  
vertical spectra  
parent ions

column at  
m/z F2 = 393.717

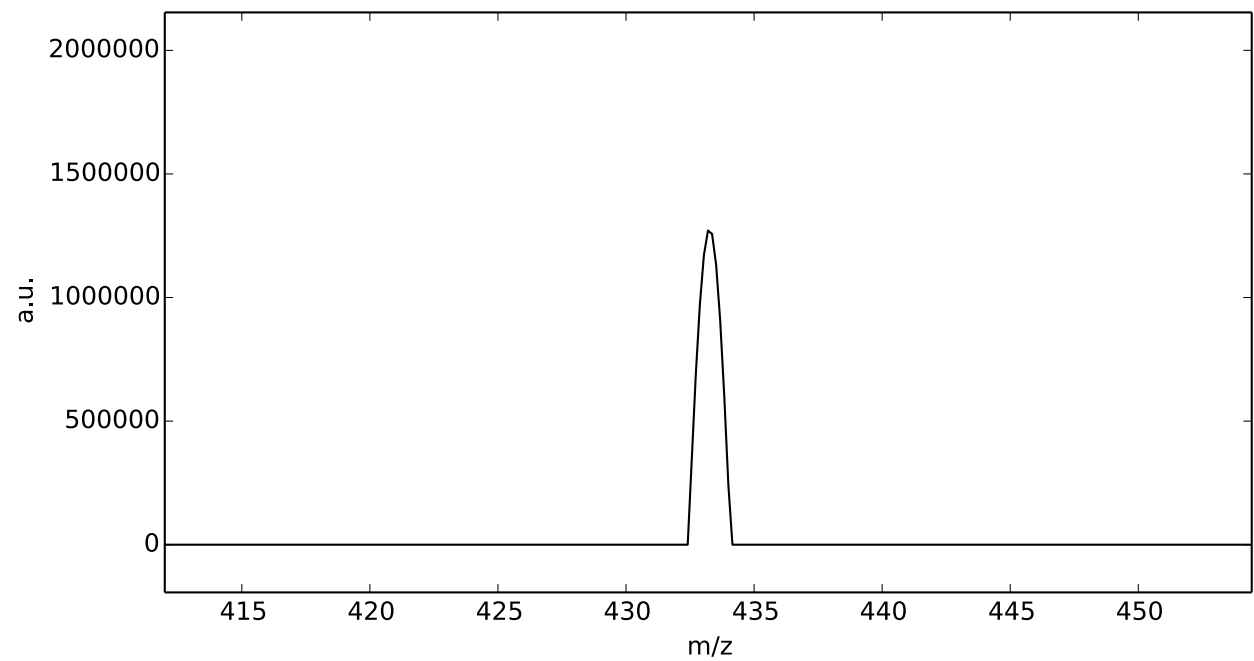


looking along F1



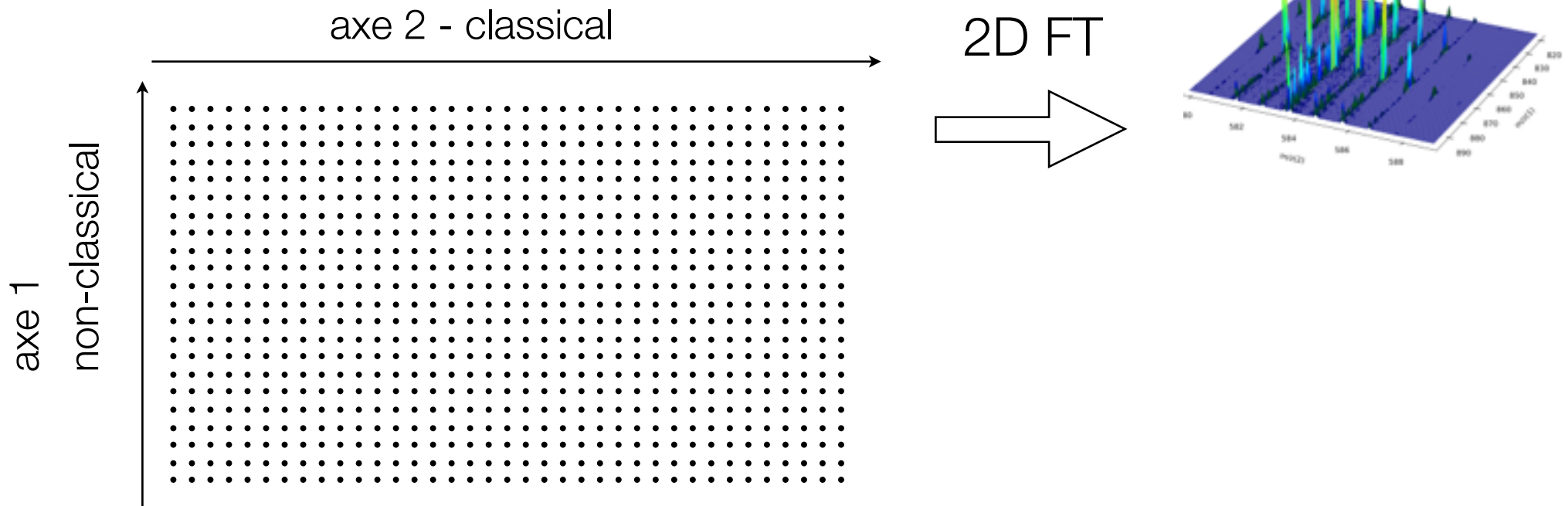
F1 dimension  
vertical spectra  
parent ions

column at  
m/z F2 = 393.717



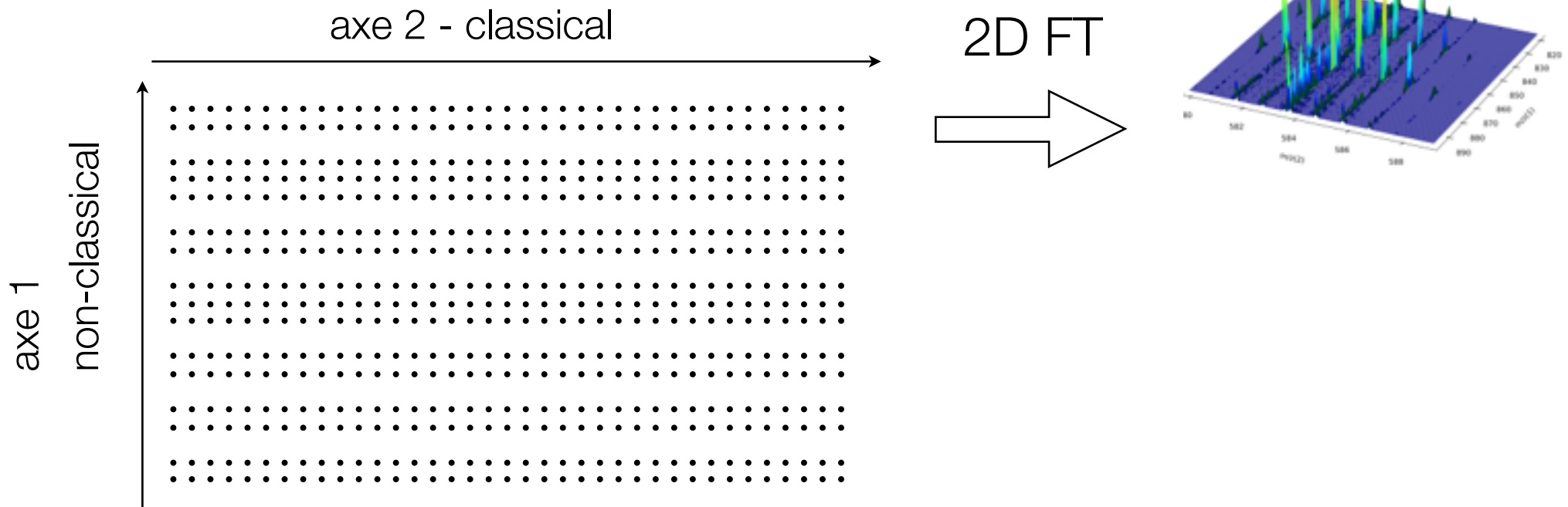
# Non Uniform Sampling in 2D

- each axes are very different in nature  
each measure on axe 1 has to be acquired separately  
high cost in time.
- Recital allows to drop some acquisition  
maintain  $\Delta t$   
faster acquisition



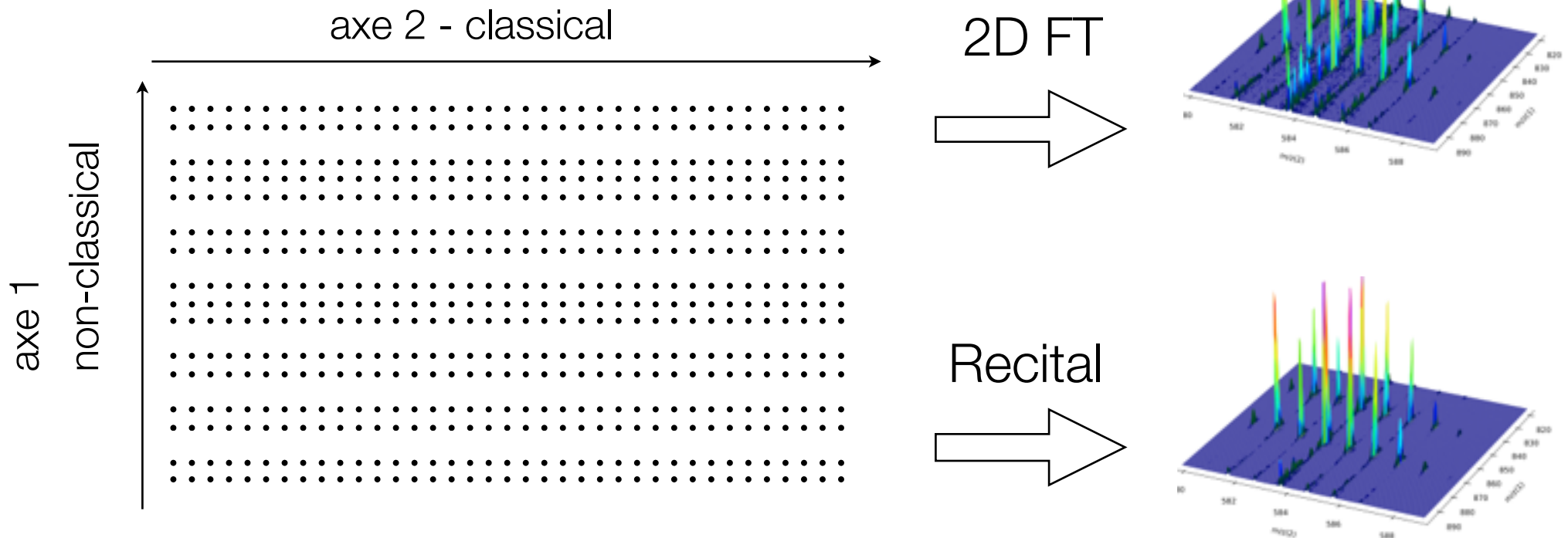
# Non Uniform Sampling in 2D

- each axes are very different in nature  
each measure on axe 1 has to be acquired separately  
high cost in time.
- Recital allows to drop some acquisition  
maintain  $\Delta t$   
faster acquisition



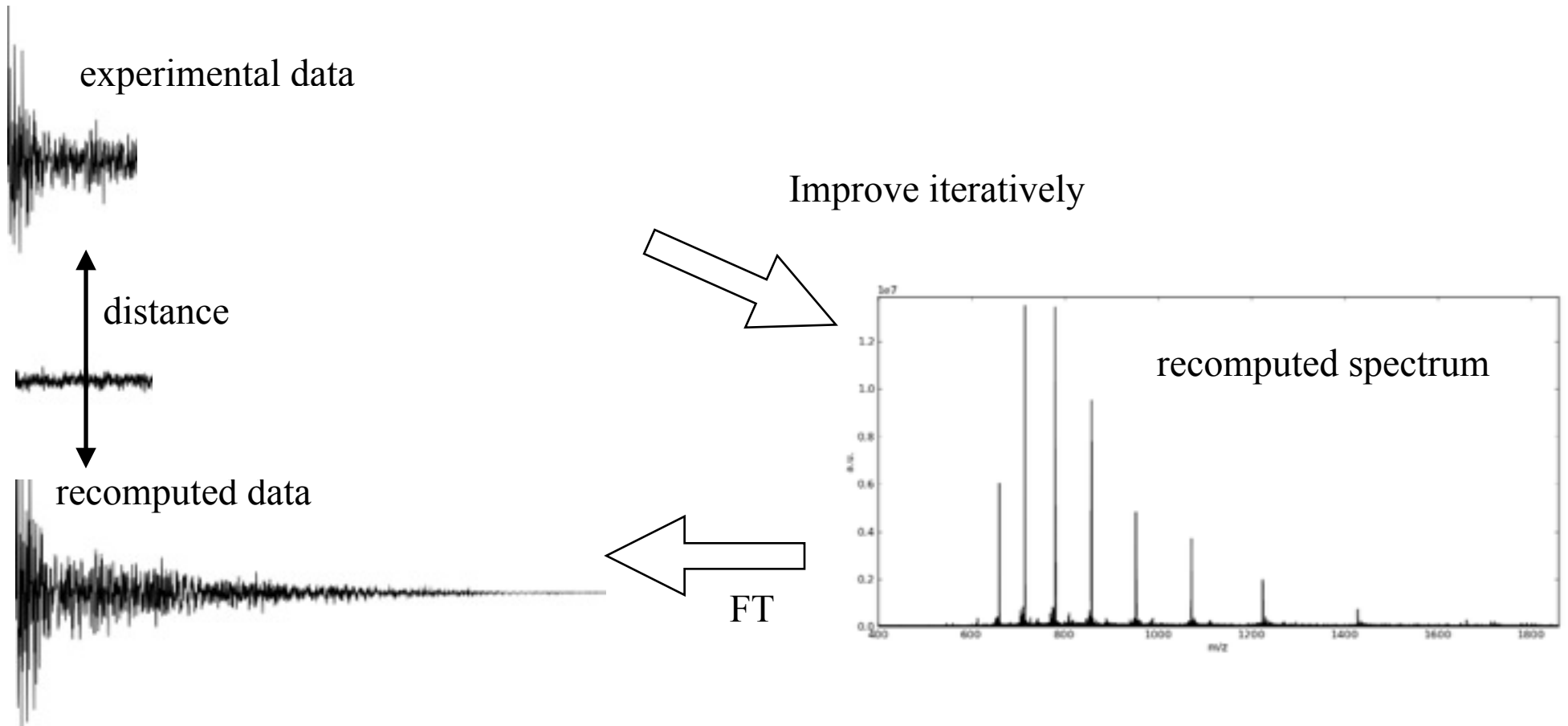
# Non Uniform Sampling in 2D

- each axes are very different in nature  
each measure on axe 1 has to be acquired separately  
high cost in time.
- Recital allows to drop some acquisition  
maintain  $\Delta t$   
faster acquisition



# Recital for NUS Data

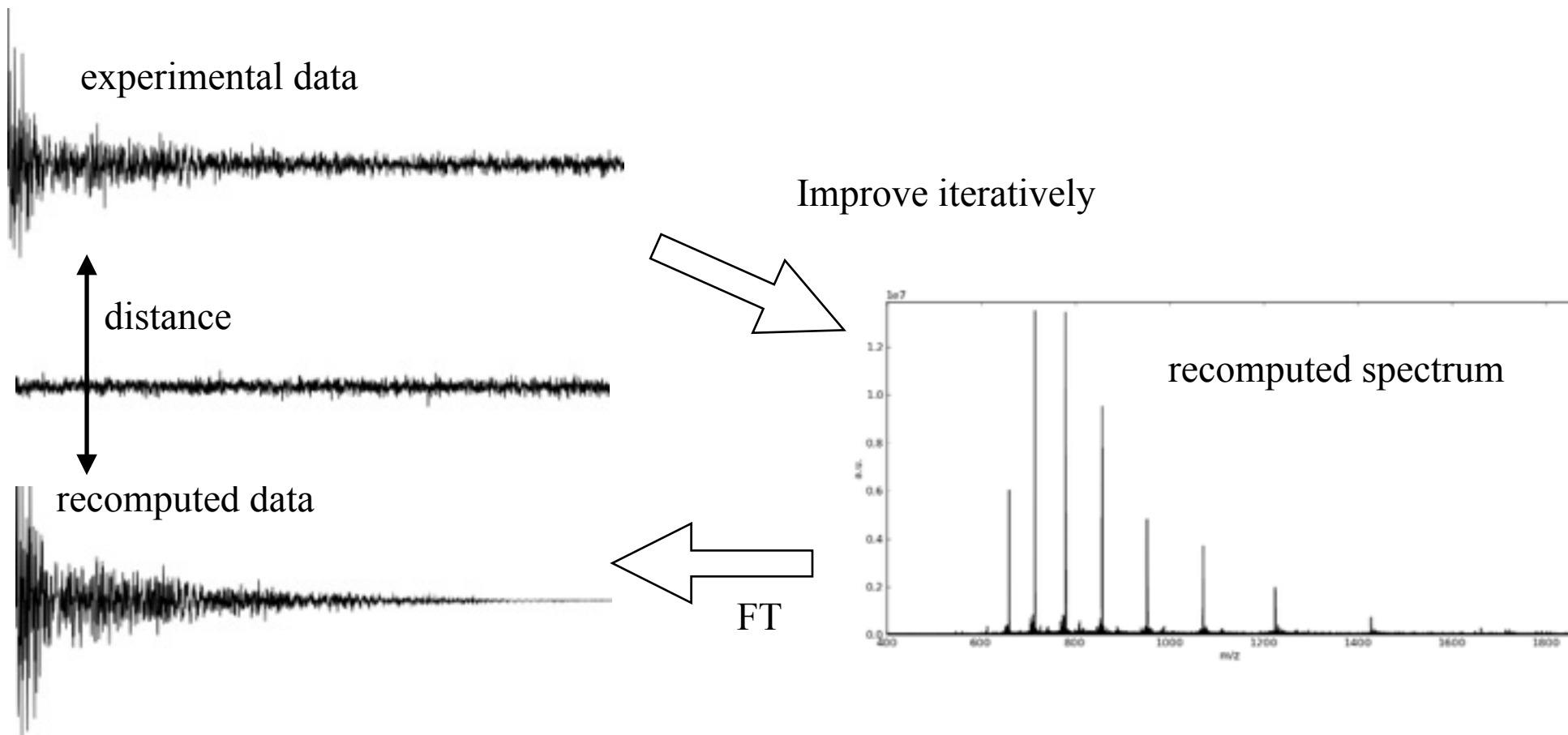
- Inverse approach





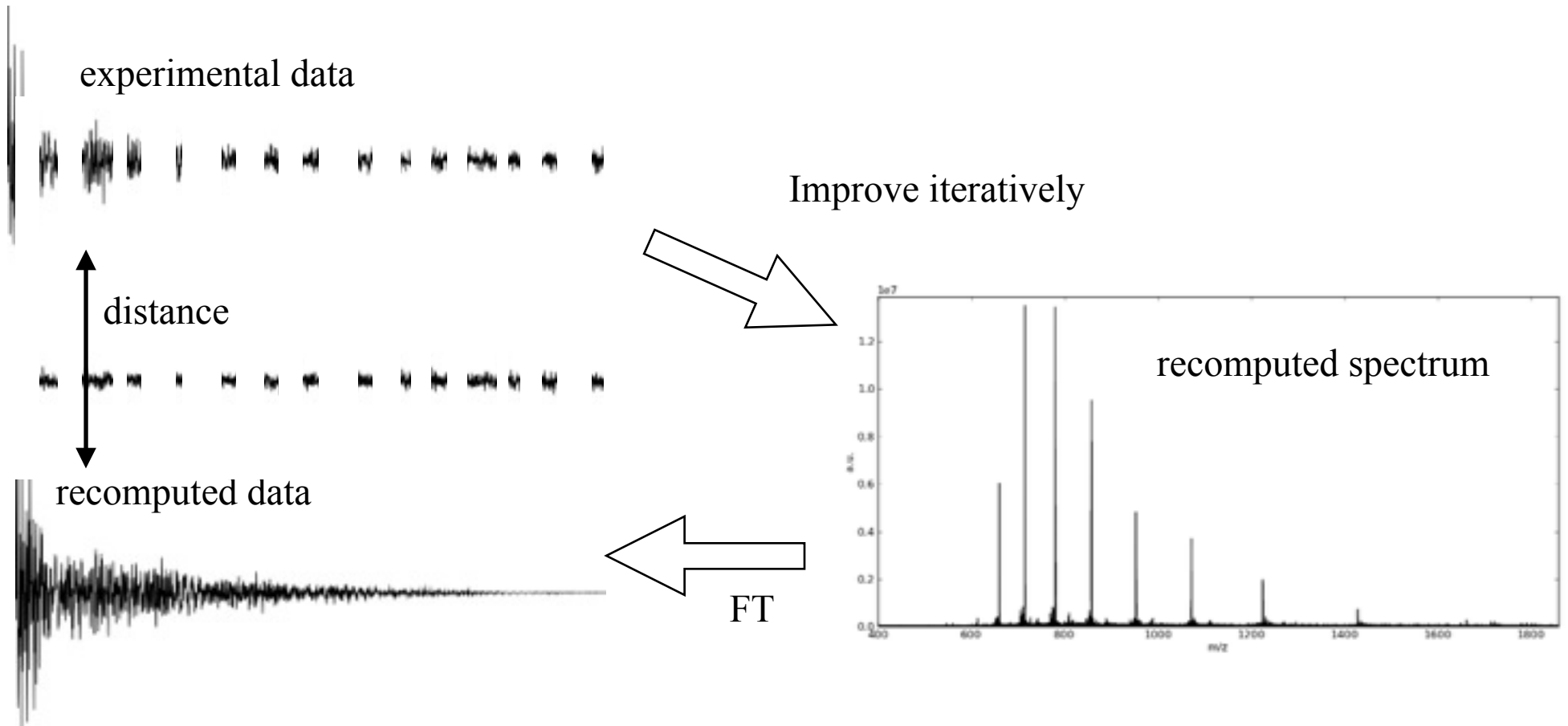
# Recital for NUS Data

- Inverse approach



# Recital for NUS Data

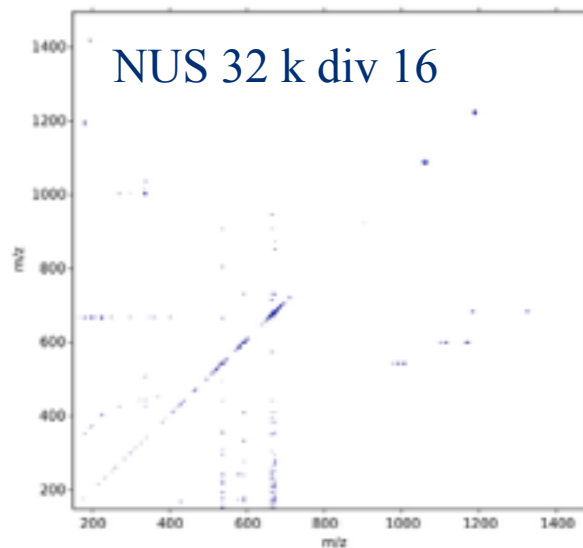
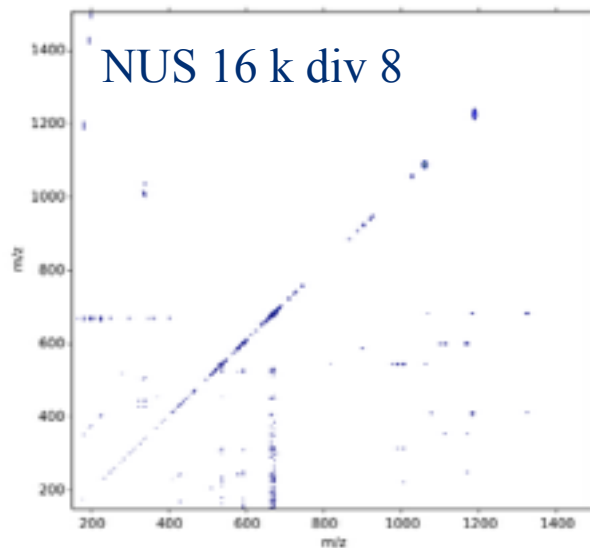
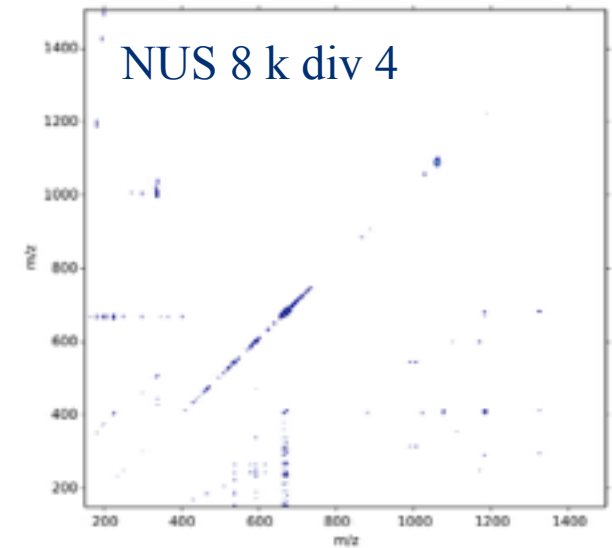
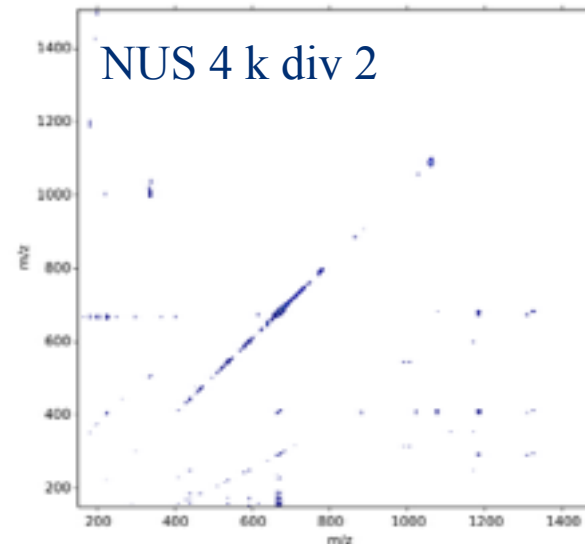
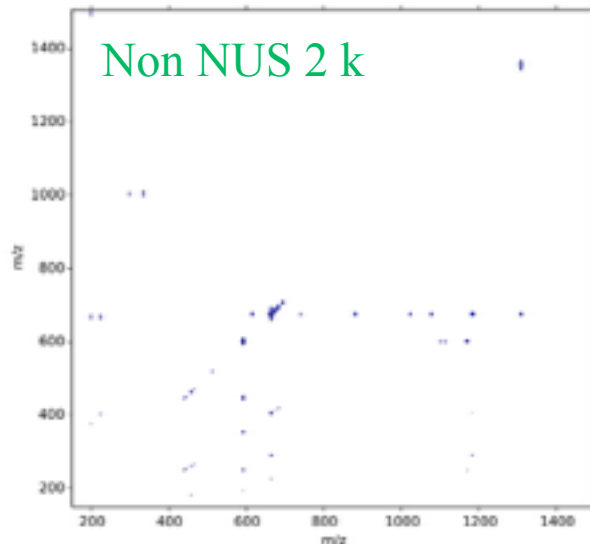
- Inverse approach



# Non Uniform Sampling & RECITAL & 2D

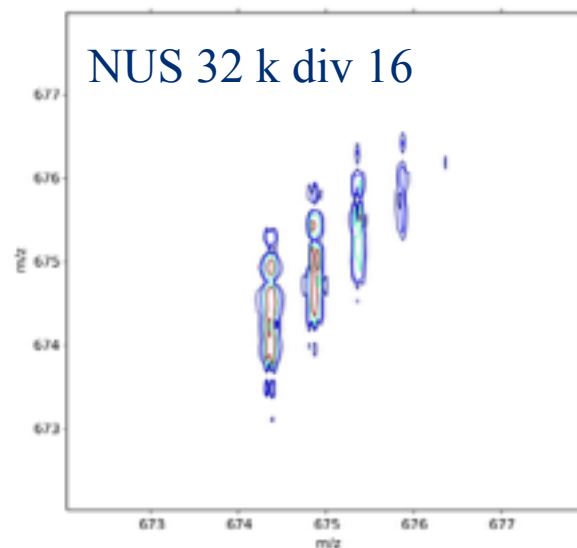
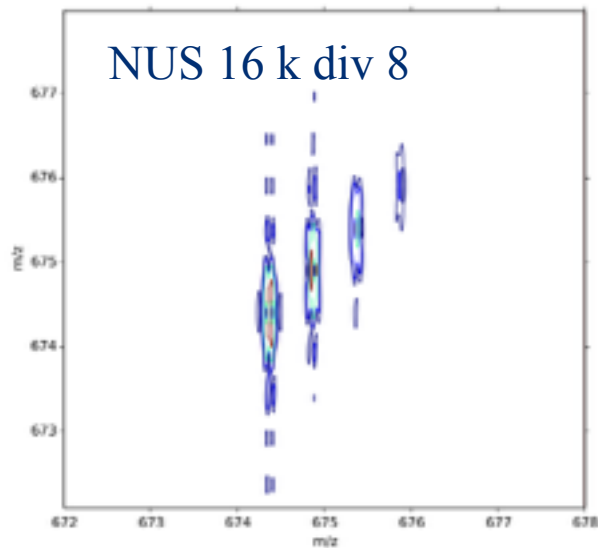
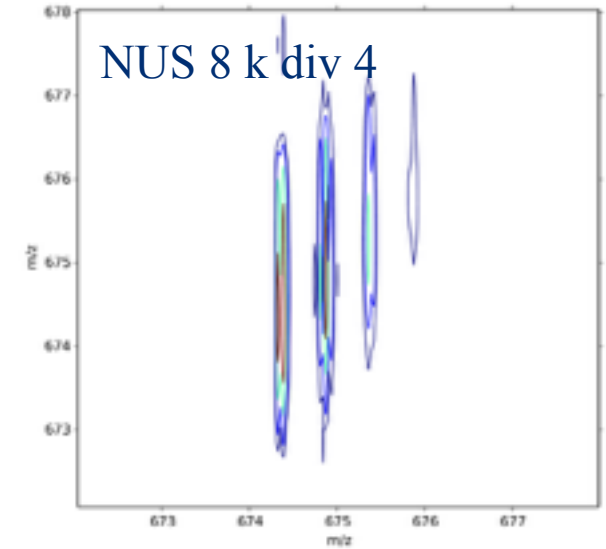
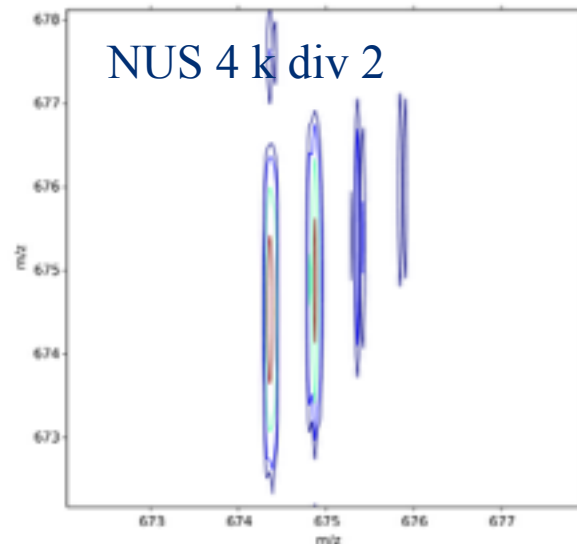
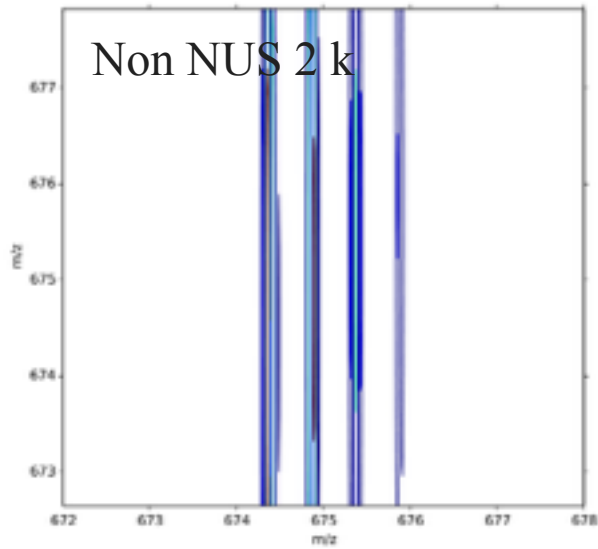
- Substance P

With C.Rolando



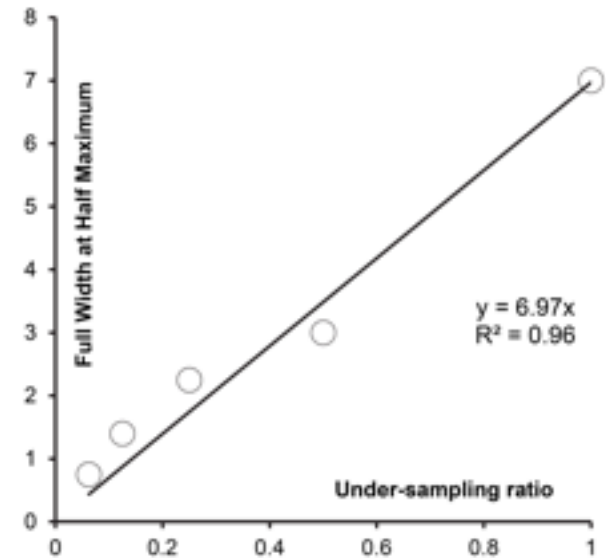
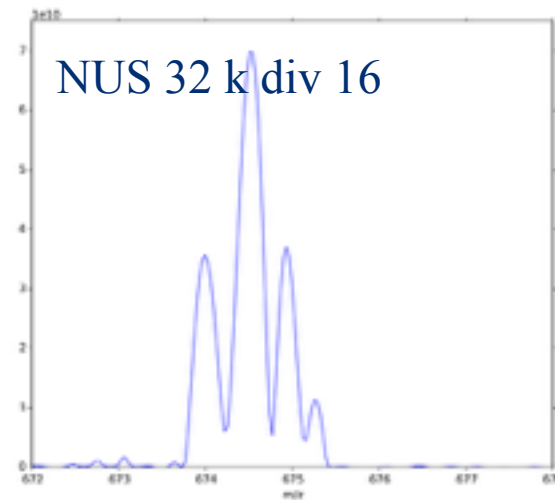
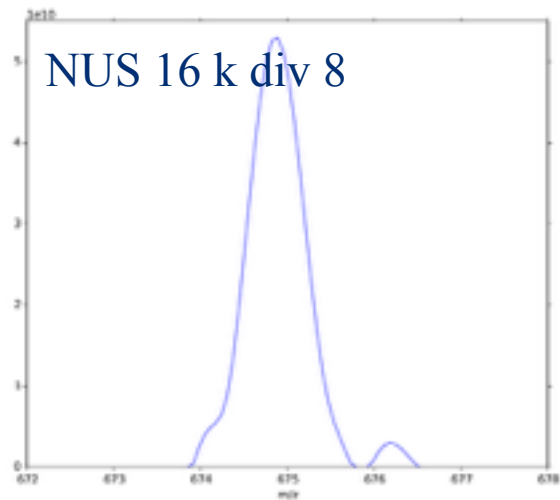
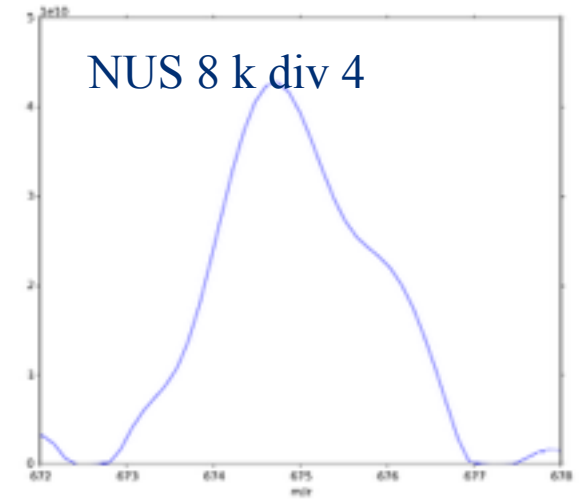
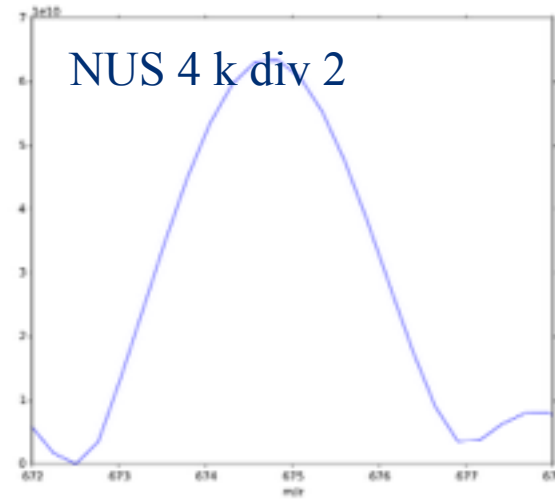
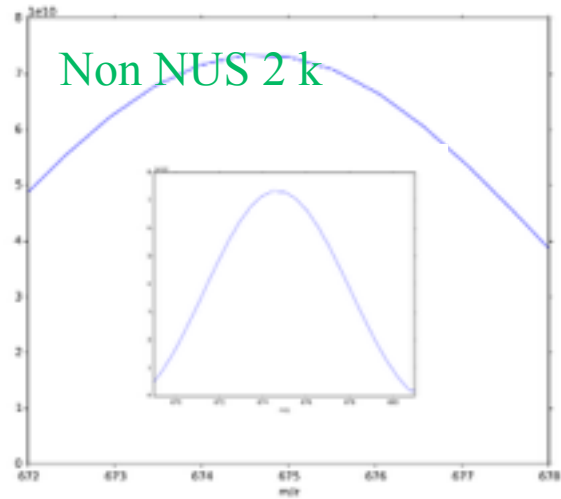
The overall aspect of the 2D spectrum is maintained.

# Zooming on a off-diagonal peak



Precursor resolution is increasing with undersampling ratio as expected.

# parent precursor profile



**Precursor FWHM is decreasing proportionally to the undersampling ratio.**

# NPK v2 program

- New program

  - ▶ Extension of a previous NMR processing program

- multi FT spectroscopies

  - FT-ICR Orbitrap

  - NMR FT-IR soon

- Implements standard 1D, 2D and even 3D processing

- Advanced technologies and algorithms

  - Uses multi-core / multi-processor architectures through MPI

  - Unlimited data-size though HDF5 (we do regularly 130Gb spectra )

  - FFT, Hilbert, Laplace, Recital, MaxEnt, etc...

  - real - complex - quaternionic (hypercomplex) values

- Programmatic interface

  - ▶ K is for Kernel !

  - python ipython as interactive shell standard graphic output

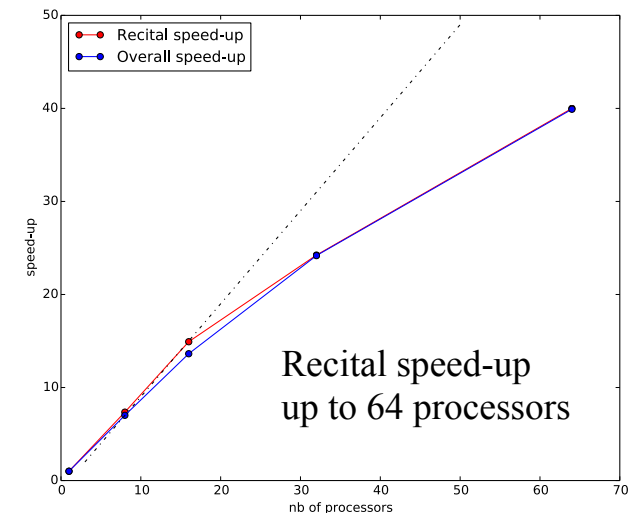
  - configuration files for involved operations (2D, NUS, ...)

- Open-Source

  - available on bitbucket.org

  - not fully-documented yet !

  - contact-me !



# Un exemple de réseau

- Il faut :
  - des mathématiques
  - de l'algorithmique
  - de l'informatique
  - de la physique
  - de la biologie
  
- Réseau MeasureHD (CNRS - Mastodons)
  - mathématicien
    - ▶ P.L. Combette lab JLL
  - algorithmiciens
    - ▶ JC Pesquet - E Chousezoux lab GM
  - informaticiens / GPU
    - ▶ P.Collet lab iCube
  - physiciens de la mesure
    - ▶ FT-IR : P.Roy Soleil
    - ▶ FT-ICR : J.Chamot-Rooke Pasteur C.Rolando MSAP
    - ▶ NMR : B.Kieffer MA Delsuc ESBS/IGBMC

# Le consortium

Introduction Projections Proximity Algorithms

## Notation 1

- $\mathcal{H}, \mathcal{G}_1, \dots, \mathcal{G}_m$ : real Hilbert spaces
- $\mathcal{B}(\mathcal{H}, \mathcal{G})$ : bounded linear operators from  $\mathcal{H}$  to  $\mathcal{G}$
- $\rightharpoonup$ : weak convergence;  $\rightarrow$ : strong convergence
- $\Gamma_0(\mathcal{H})$ : lower semicontinuous convex functions  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  with  $\text{dom } f = \{x \in \mathcal{H} \mid f(x) < +\infty\} \neq \emptyset$ .
- The conjugate of  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  is
 
$$f^*: \mathcal{H}^* \rightarrow ]-\infty, +\infty]: x^* \mapsto \sup_{x \in \mathcal{H}} (\langle x \mid x^* \rangle - f(x))$$
- The subdifferential of  $f: \Gamma_0(\mathcal{H})$  is
 
$$\partial f: x \mapsto \{u \in \mathcal{H} \mid f(x) + f^*(u) = \langle x \mid u \rangle\}$$
- $f$  differentiable at  $x \Rightarrow \partial f(x) = \{\nabla f(x)\}$ ;
- Infimal convolution of  $f \in \Gamma_0(\mathcal{H})$  and  $g \in \Gamma_0(\mathcal{H})$ :
 
$$f \square g: \mathcal{X} \rightarrow ]-\infty, +\infty]: x \mapsto \inf_{y \in \mathcal{X}} (f(y) + g(x - y))$$

Patrick L. Combettes Algorithms proximaux 2/41

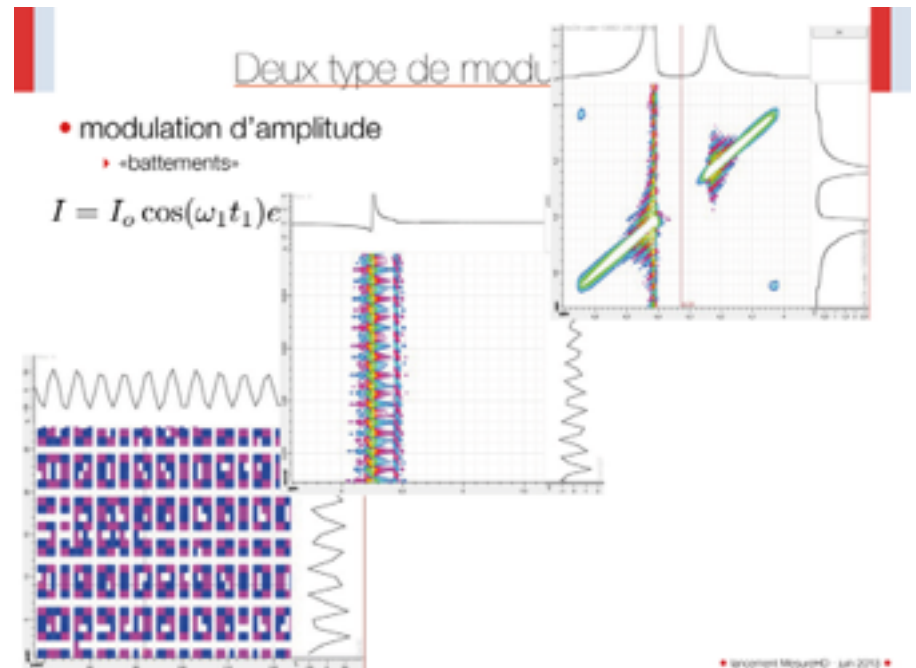
## Proposed algorithm

### Majorize-Minimize Memory Gradient (3MG) algorithm

For  $k = 1, 2, \dots$

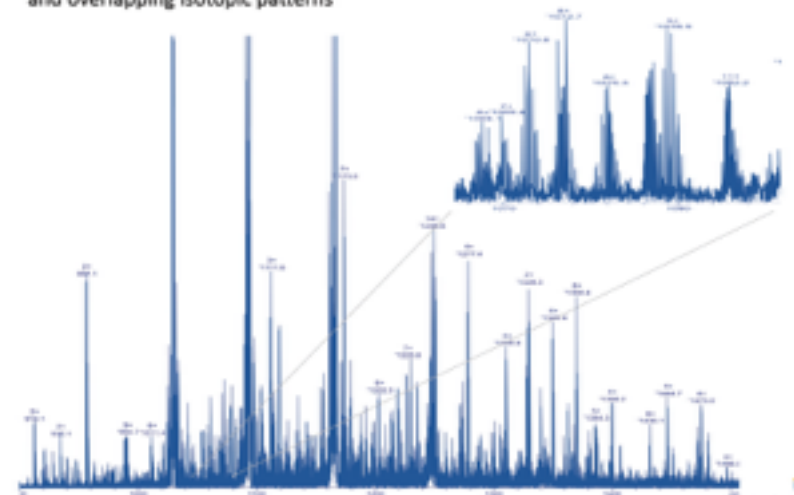
- 1 Compute the set of directions  $D_k = [-g_k, x_k - x_{k-1}]$
- 2  $u_k^0 = 0$
- 3 For  $j = 1, \dots, J$ 
  - $B_{u_k^j} = D_k^\top A(x_k + D_k u_k^j) D_k$
  - $u_k^{j+1} = u_k^j - (B_k^j)^{-1} \nabla f(u_k^j)$
- 4  $u_k = u_k^J$
- 5 Update  $x_{k+1} = x_k + D_k u_k$

$\rightsquigarrow$  Converges to a critical point of  $F$  [Chouzenoux13a]



## ECD Fragmentation spectrum of 17+

- Presence of multiple fragment ions with various charge states (from 1+ to 17+) and overlapping isotopic patterns





# Le consortium

Introduction Projections Proximity Algorithms

## Notation 1

- $\mathcal{H}, \mathcal{G}_1, \dots, \mathcal{G}_m$ : real Hilbert spaces
- $\mathcal{B}(\mathcal{H}, \mathcal{G})$ : bounded linear maps from  $\mathcal{H}$  to  $\mathcal{G}$
- $\rightharpoonup$ : weak convergence
- $f: \mathcal{H} \rightarrow [-\infty, +\infty]$ : convex function
- $f^*$ : conjugate of  $f$
- The subdifferential of  $f: \Gamma_0(\mathcal{H})$  is
 
$$\partial f: \mathcal{H} \rightarrow \{u \in \mathcal{H} \mid f(x) + f^*(u) = \langle x \mid u \rangle\}$$
- $f$  differentiable at  $x \Rightarrow \partial f(x) = \{\nabla f(x)\}$
- Infimal convolution of  $f \in \Gamma_0(\mathcal{H})$  and  $g \in \Gamma_0(\mathcal{H})$ :
 
$$f \square g: \mathcal{X} \rightarrow [-\infty, +\infty]: x \mapsto \inf_{y \in \mathcal{X}} (f(y) + g(x - y))$$

Patrick L. Combettes Algorithms proximaux 2/41

**INSMI**

**P-L Combettes**  
LJLL-Paris  
Théorie

Proposed algorithm

**INSB**

M-A Delsuc  
IGBMC-Strasbourg  
RMN - Développements

**INC**

C. Rolando  
MSAP - Lille  
FTICR - Métabonomique

P. Roy  
Soleil - Paris  
FTIR

ECD Fragmentation spectrum of 17+

Presence of multiple fragment ions with various charge states (from 1+ to 17+) and overlapping isotopic patterns

Majorize-Minimize Memory C

For  $k = 1, 2, \dots$

- 1 Compute the set of directions
- 2
- 3
- 4  $u_k = u_k^j$
- 5 Update  $x_{k+1} = x_k + D_k u_k$

$B_{u_k^j} = D_k^T A(x_k + D_k u_k^j) D_k$

$u_k^{j+1} = u_k^j - (B_k^j)^{-1} \nabla f(u_k^j)$

$\rightsquigarrow$  Converges to a critical point

Emile Chouzenoux Séminaire Systèmes Complexes 6 juin 2013 9 / 15

**INS2I**

**J-C Pesquet**  
LIGM-Paris  
Algorithmique

**P. Collet**  
iCube-Strasbourg  
GPGPU

**INSB**

B. Kieffer  
IGBMC - Strasbourg  
RMN - Interactions

A. Ourjoustev  
IGBMC - Strasbourg  
RX - Biol. Structurale

J. Chamot-Rooke  
B. Schwikowski  
Pasteur Paris  
MS - protéomique

# nombreux projets - en cours

- algorithmiques

- appliqués à la déconvolution des patterns en MS
  - massifs isotopiques
  - motifs multicharge
  - combinaison des deux

- développement

- déployer les algorithmes vers les laboratoires d'analyse
- plateforme de traitement intégrée dans le data-flow "bioinformatique"

- Symposium

- Chemical Complexity & Biology      Strasbourg 19-20 Janvier 2015

- Renforcer le consortium

- modélisation des interactions complexes
- start-up
  - ▶ outils d'intégration logiciel

- Recherchons Post-Doc !!!

- et doctorants

# Acknowledgments

- Lille - MSAP
  - C.Rolando
  - M. van Agthoven
  - F. Bray
  - O.Plechakova
- Paris - I.Pasteur
  - J.Chamot-Rooke
  - E.Vorontsov
- Paris - ENS
  - G. Bodenhausen
  - A.Sehgal
  - P.Pelupessy
- Strasbourg IGBMC
  - B.Kieffer
  - L.Chiron
- Strasbourg NMRTEC
  - M-A Coutouly
- Strasbourg CASC4DE
  - J-P Starck
  - L.Chiron

- fundings
- ANR
  - grant 2010 FT-ICR-2D
- CNRS
  - TGIR
  - grant 2013 Mesure-HD
- NMRTEC S.A.S

