



# Banques de données, bases de connaissance et dépôts de référence en *Métabolomique*

Franck Giacomoni – MetaboHUB – INRA.PFEM

Session « *Analyse et interprétation de données métabolomiques et protéomiques : points communs et spécificités* »

28 mars 2018 – Forum Labo – Lyon



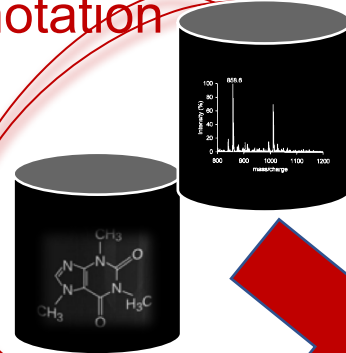
SFEAP - RFMF



# Analyses des données en Métabolomique

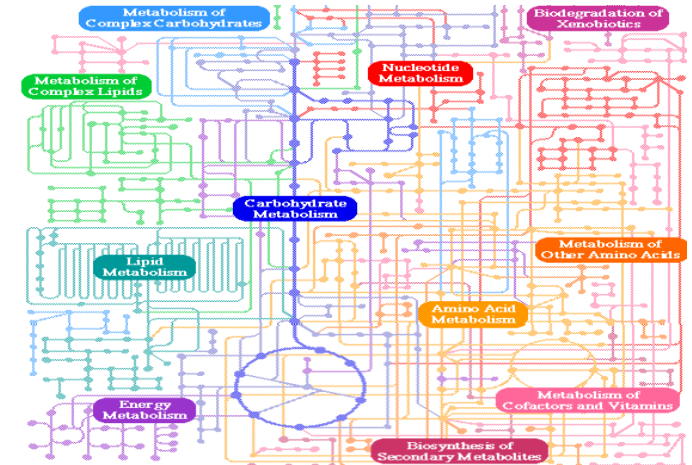
| name      | namecustom       | mz               | mzmin            | mzmax            |
|-----------|------------------|------------------|------------------|------------------|
| M100T183  | M100.07529T183   | 100.075290217332 | 100.073962458109 | 100.075856347639 |
| M101T64   | M101.017412T64   | 101.017411813327 | 101.016363119337 | 101.025598611889 |
| M101T1090 | M101.094497T1090 | 101.094497373251 | 101.093825596321 | 101.094931330369 |
| M102T67   | M102.121467T67   | 102.121466917005 | 102.118620434328 | 102.128045805211 |
| M103T60   | M103.032369T60   | 103.032369394794 | 103.031358949596 | 103.040632681061 |
| M103T61   | M103.120075T61   | 103.120074855461 | 103.118630961785 | 103.129325340649 |
| M104T1162 | M103.950053T1162 | 103.950052934034 | 103.948946676433 | 103.95133690961  |
| M104T60   | M104.037434T60   | 104.037433793914 | 104.036015958492 | 104.046229685927 |
| M104T47   | M104.100665T47   | 104.100665065741 | 104.099665732044 | 104.108679356773 |
| M104T88   | M104.095422T88   | 104.095422407145 | 104.093036006622 | 104.100116358936 |
| M104T42   | M104.195599T42   | 104.195598628802 | 104.187807623186 | 104.197644370905 |
| M105T61   | M105.045229T61   | 105.045228536653 | 105.036480671616 | 105.046334818564 |
| M105T242  | M105.069935T242  | 105.069934637469 | 105.06894296829  | 105.071143736243 |
| M105T48   | M105.103611T48   | 105.103611067744 | 105.102427529644 | 105.111698113321 |

Annotation



Identification

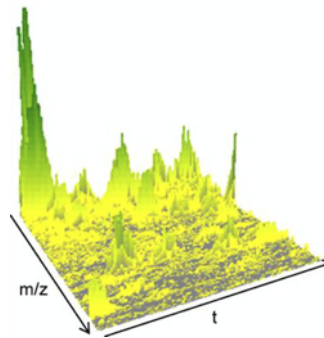
Interprétation



Extraction Data

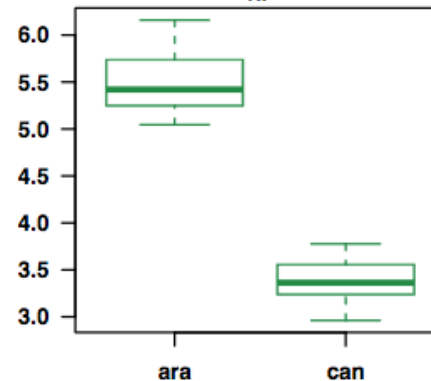
Analyses statistiques

Méthode analytique



M585T579

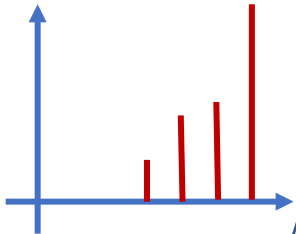
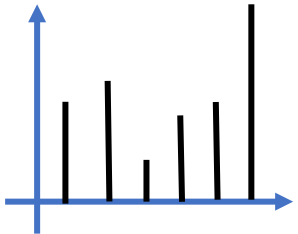
RF



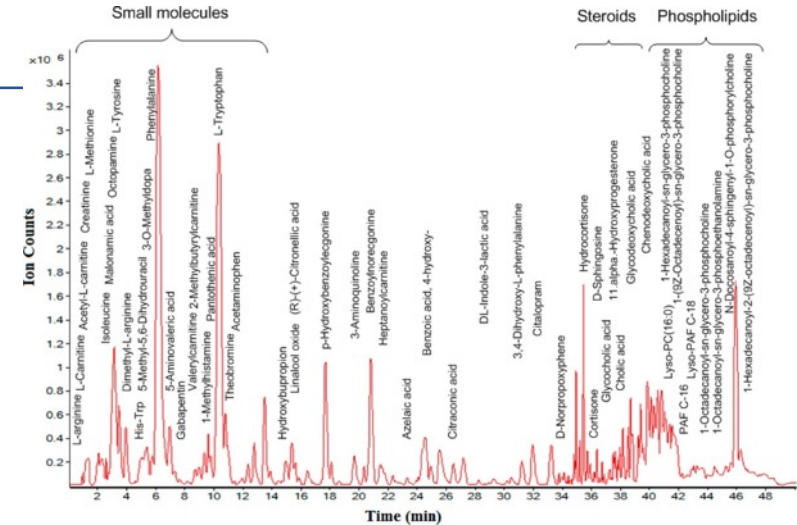
# tification



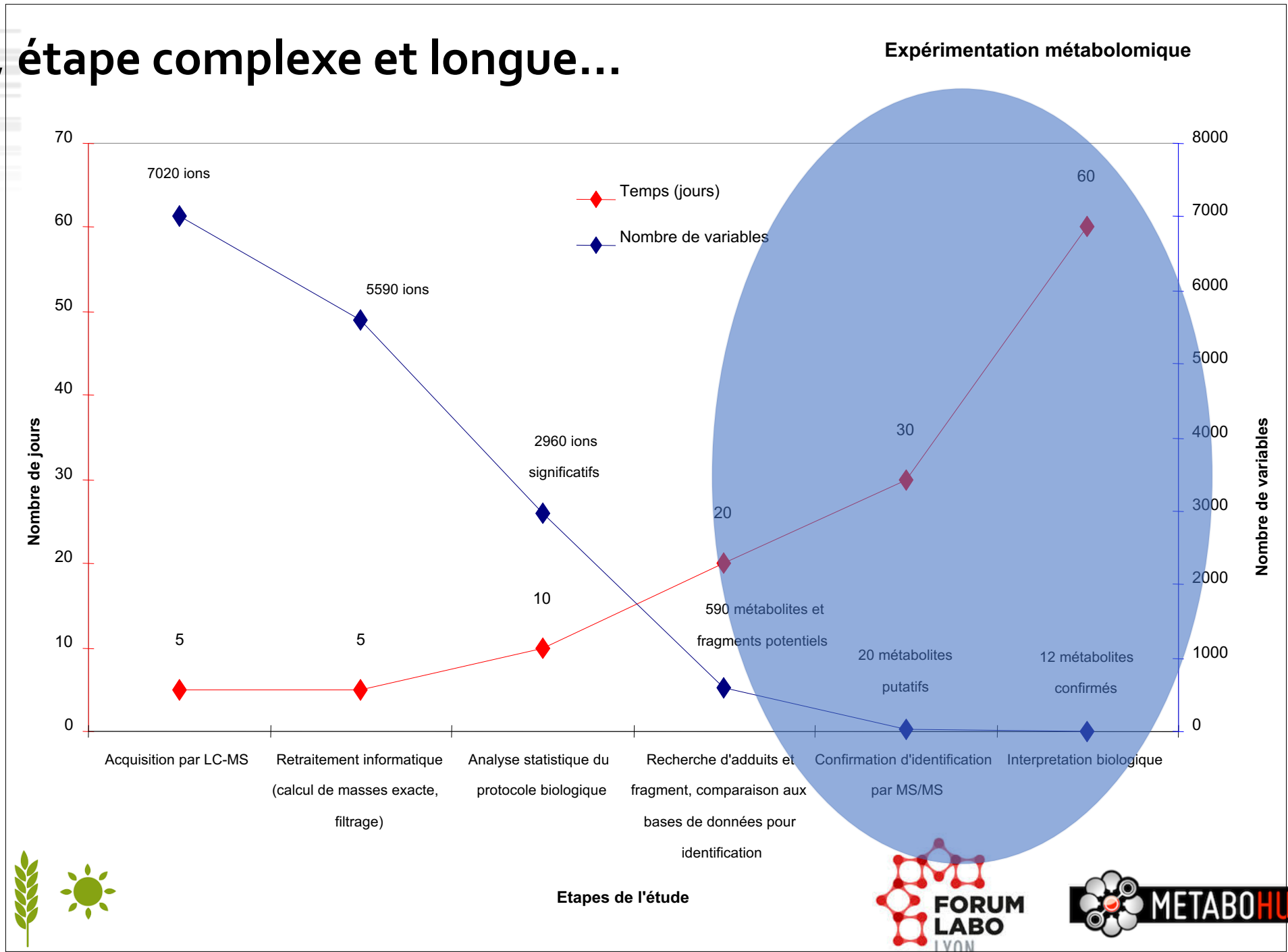
# Technique 02



## Métabolites en différentes concentrations



# L'identification, étape complexe et longue...



# Base de données, banques et librairies : définitions

- *Une base de données* : collection organisée d'information
- *Une banque de données* : représente la donnée seulement
- *Une librairie* : une façon de représenter cette information
- *Une base de connaissance* : agrégations et enrichissement de banques
- *Un dépôt de référence* : base de données, en accès libre, à usage général pour la recherche sur la métabolomique multiplateformes et inter-espèces.



# Des banques en Métabolomique...

Banques de composés chimiques

Librairies “faites maison”

Librairies de spectres

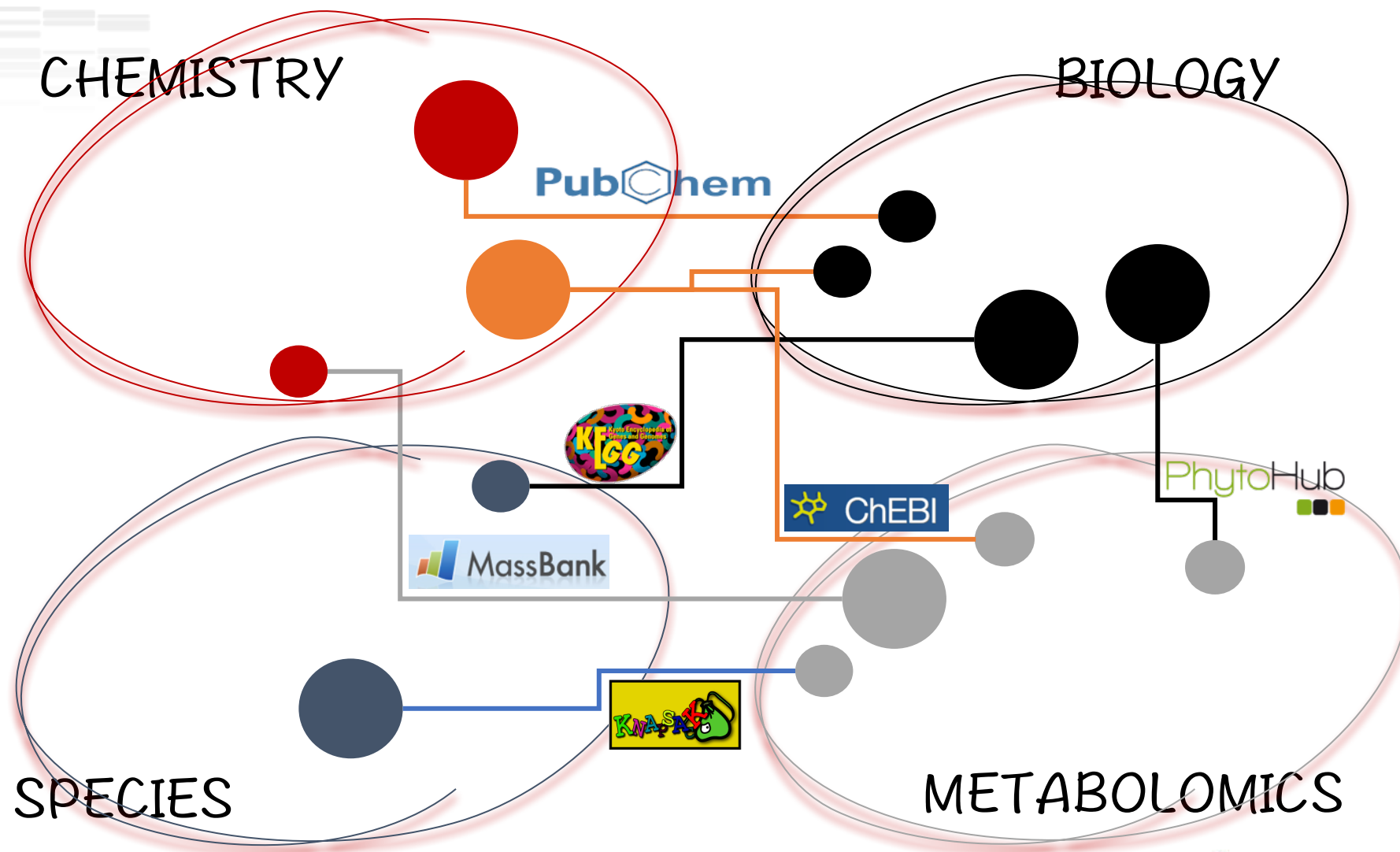
Bases de données de réseaux métaboliques

Dépôts de références

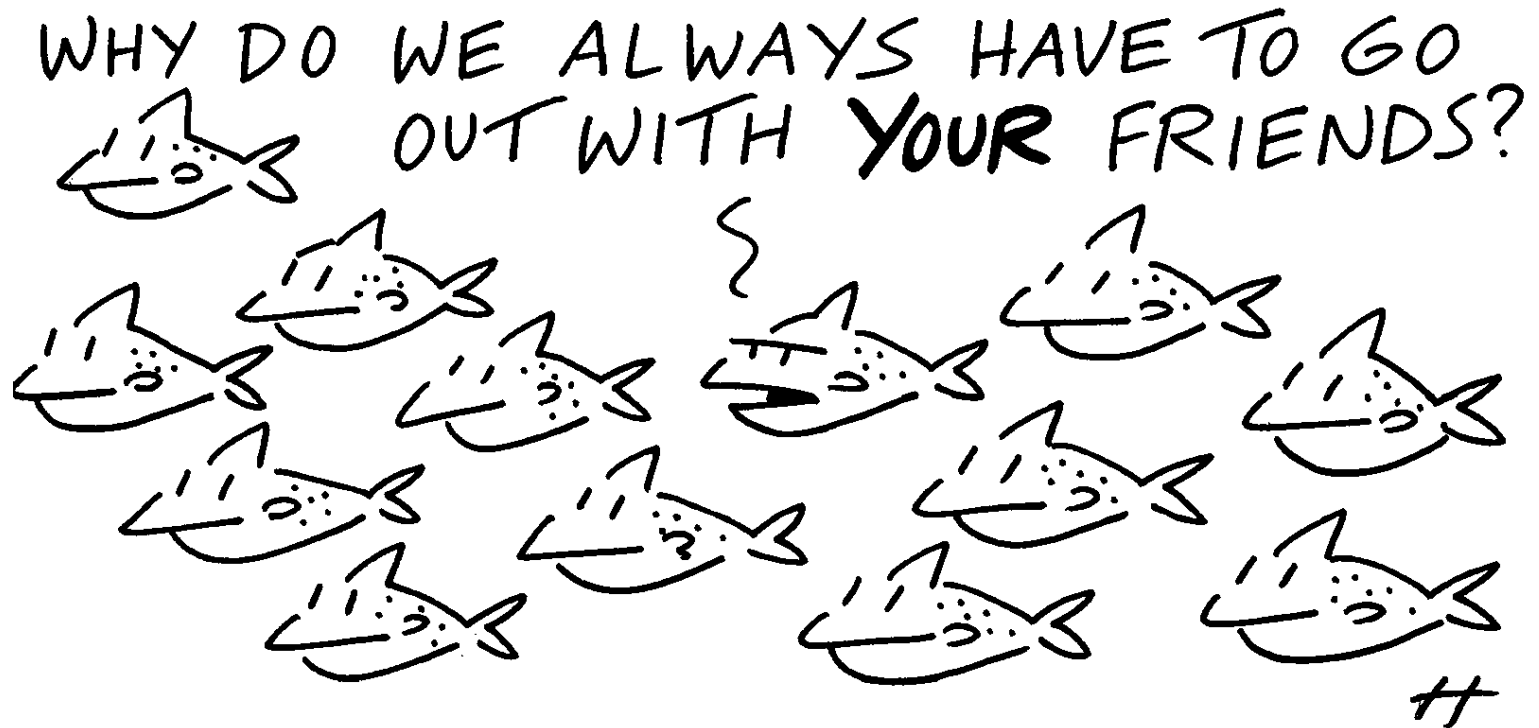


PSM V43 D169 Alfalfa fields in the desert, 1893

# Des banques en Métabolomique...



# Comment choisir votre banque ?



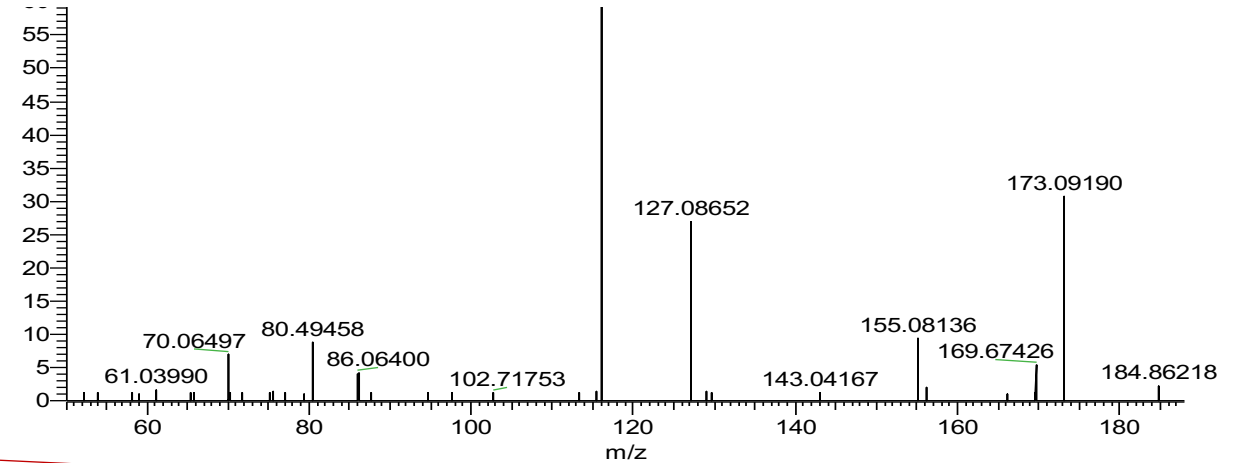
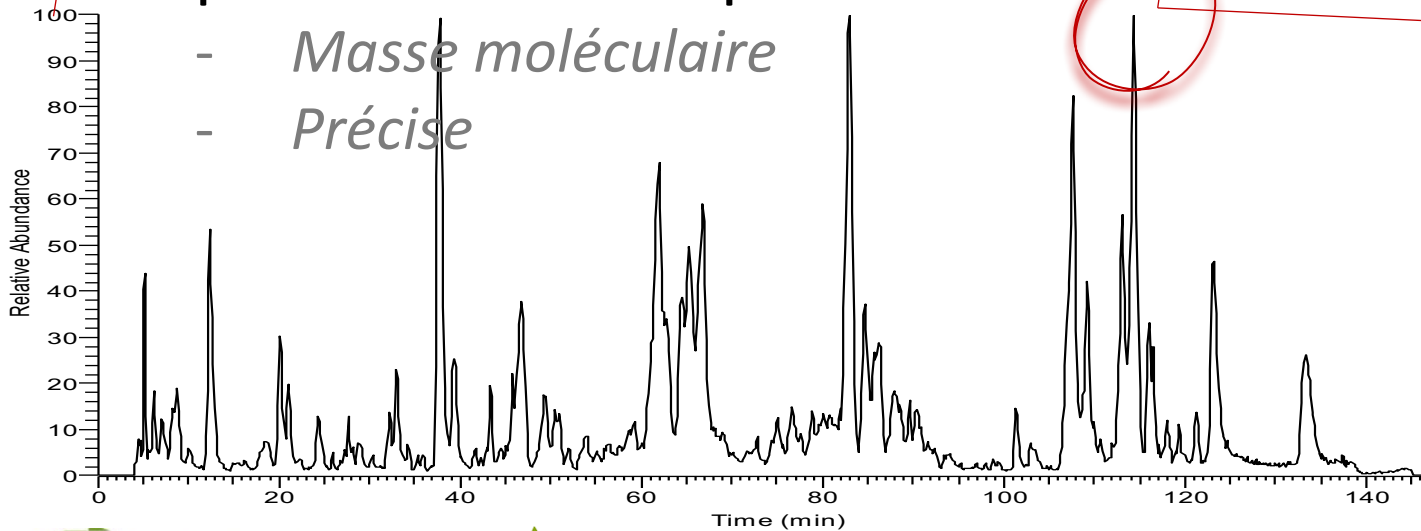
# Des empreintes métaboliques à annoter

## Exemple de la LCMS



### Empreinte métabolique

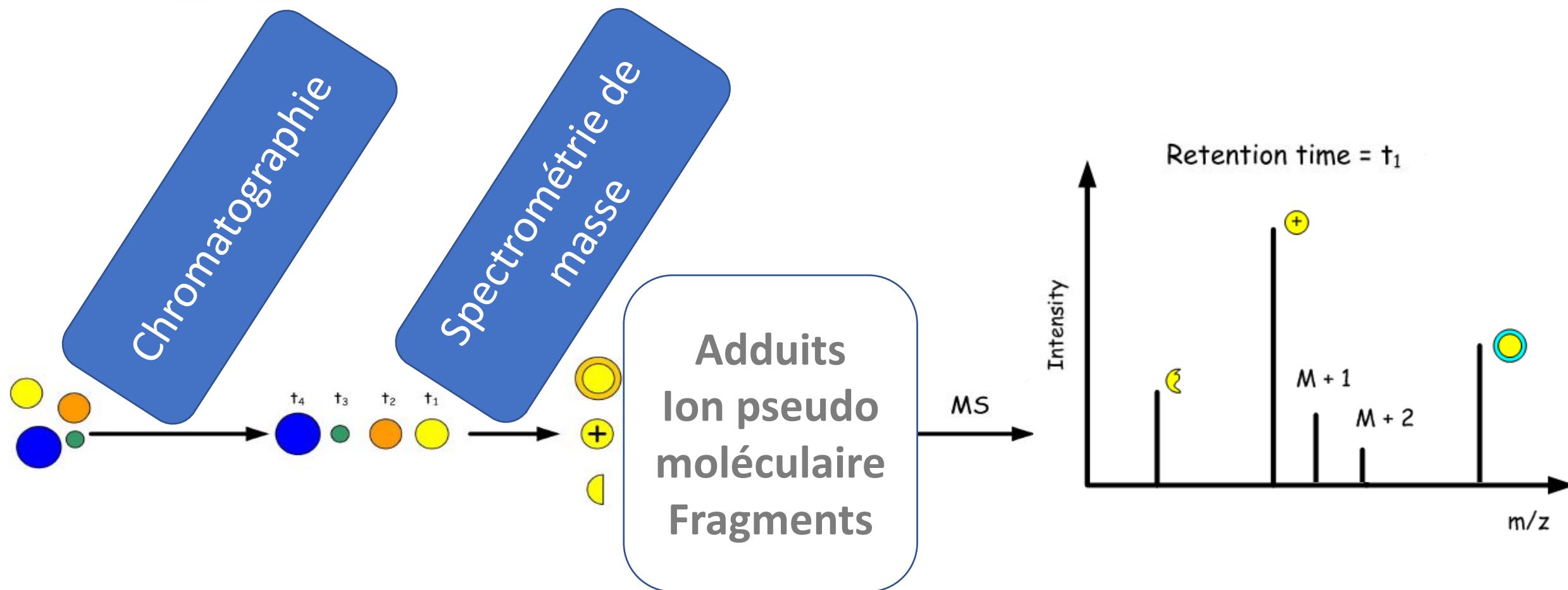
- *Masse moléculaire*
- *Précise*



### Spectre

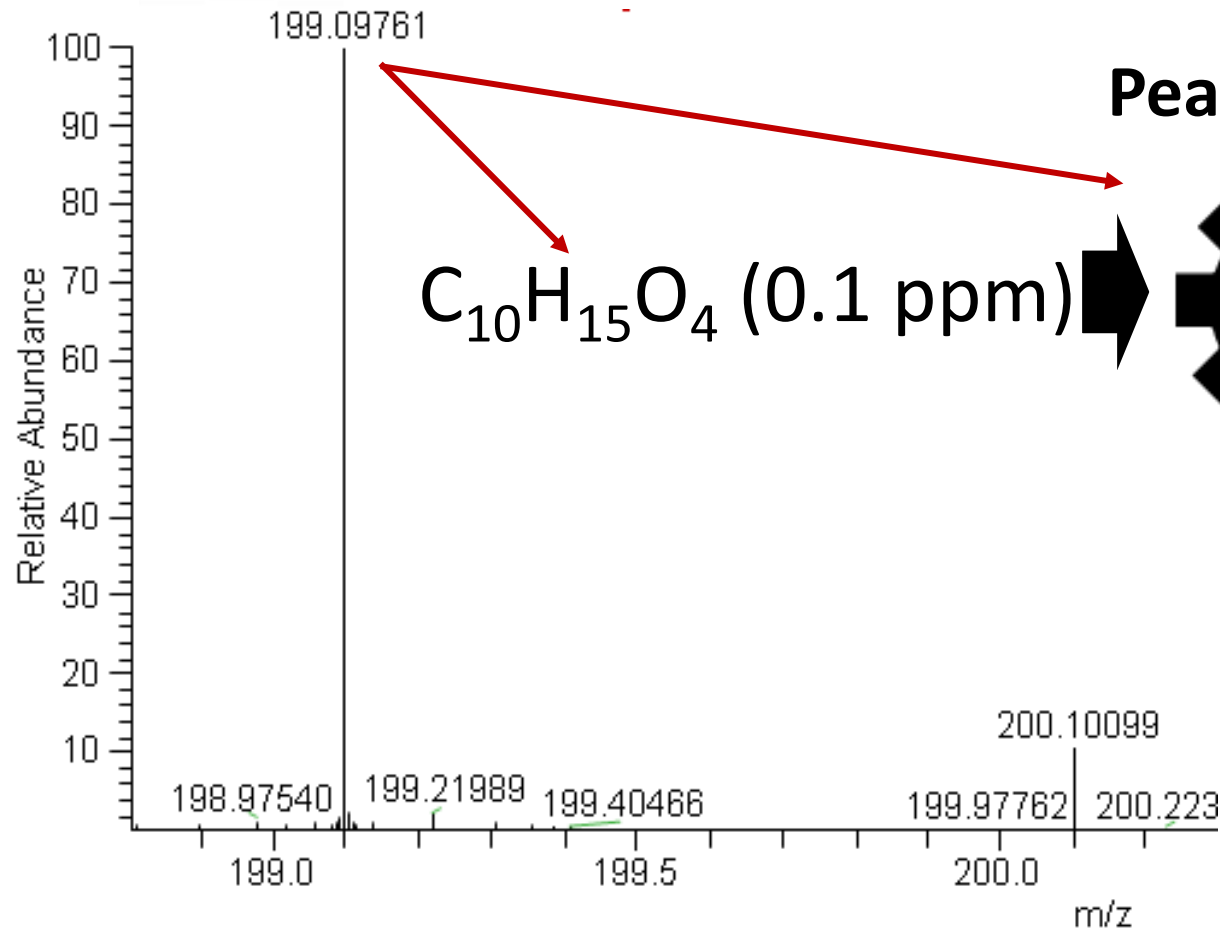
- *M/Z*
- *Intensité*
- *À une temps de rétention donné*

# Une molécule, c'est plusieurs ions...

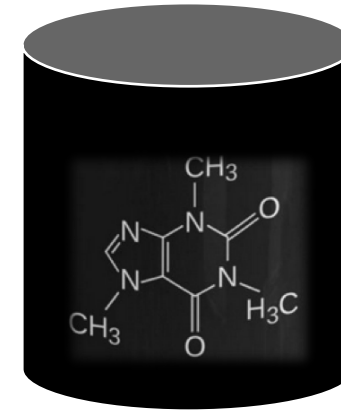
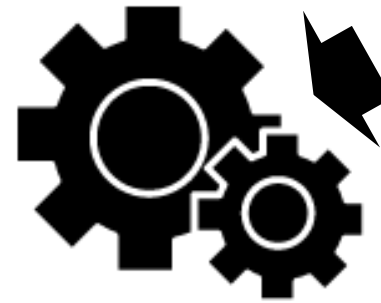


# Des banques... des stratégies d'annotations

Précision en masse



Peak matching



Banques de composés chimiques

# Des banques... des stratégies d'annotations

PubChem

A public repository of information on small molecules and their biological activities

157,000,000 entries

VS



ChEBI

## Chemical databank

*Name*

Chemical Entities of Biological Interest

Freely available dictionary of molecular entities focused on 'small' chemical compounds.

*Address*

EMBL – EBI

[www.ebi.ac.uk/chebi](http://www.ebi.ac.uk/chebi)

*Entries*

41,000 entities fully annotated



# Des banques... des stratégies d'annotations



Ressource de référence et publique

Statut des données

- « Manually annotated by the ChEBI team »
- « Manually annotated by a third party »
- « Marked as deleted or obsolete »

Search for

- ☐ ★★ only
- ☒ All in ChEBI
- ☐ ChEBI and Partner Databases

Systèmes d'Ontologies : structure moléculaire, rôles et informations sur les particules subatomiques.

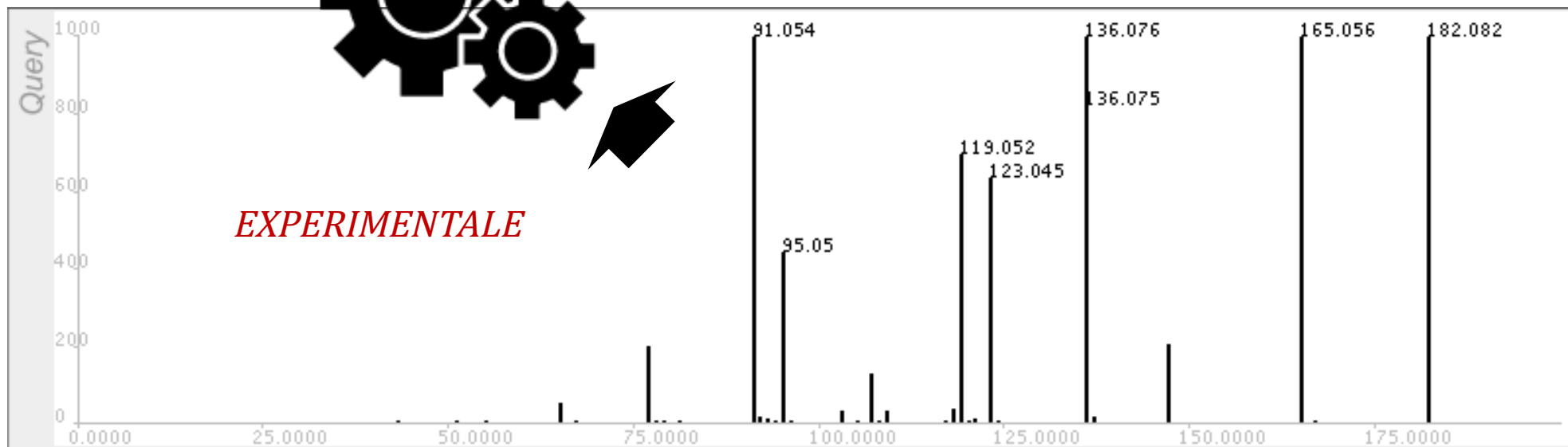
Recherche manuelle par masse et formule

Web services

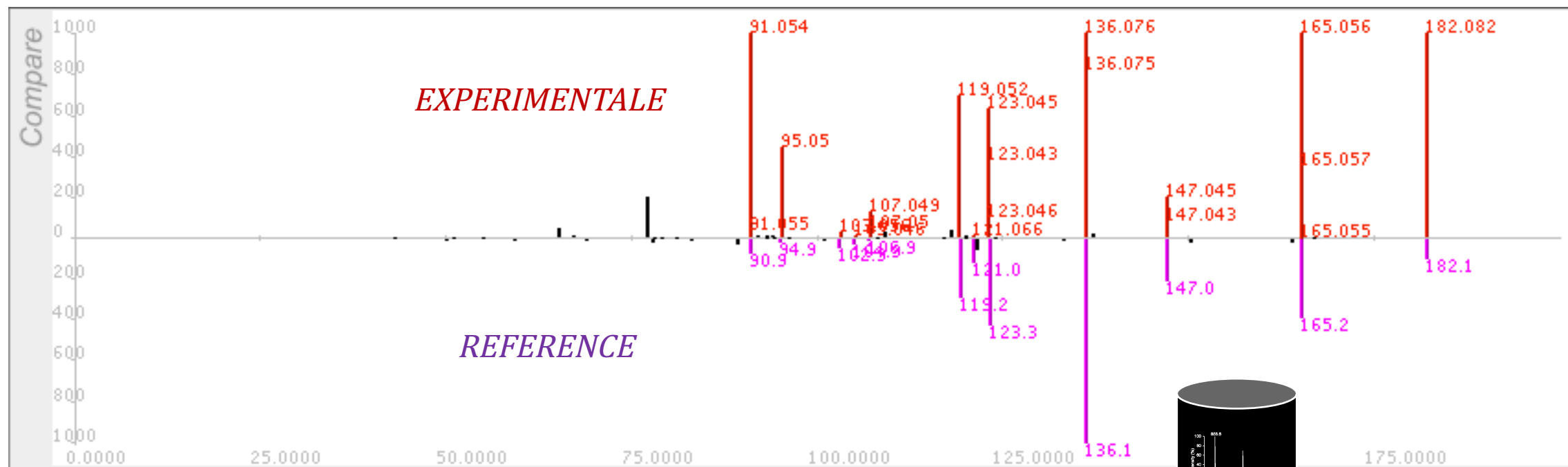


# Des banques... des stratégies d'annotations

## Spectral matching



# Des banques... des stratégies d'annotations



**Spectral matching**  
**Score of 0.83630...**

## Forest – c



## Indicateurs

- 6 000 profils métaboliques (MS & NMR) and 2 500 métabolites
- *Contributeurs*: > 20 (4 nœuds FR MetaboHUB)

## Améliorations en cours

- Ajout de métabolomes complets d'organisme modèles
- **Interopérabilité** avec W4M, MetExplore & dépôts européens.
- Devenir un dépôt de référence



## Fonctionnalités

## Dépôts de métabolomes

# SOP d'acquisition des données

# Annotation de matrices biologiques

## Curation de données, visualisation



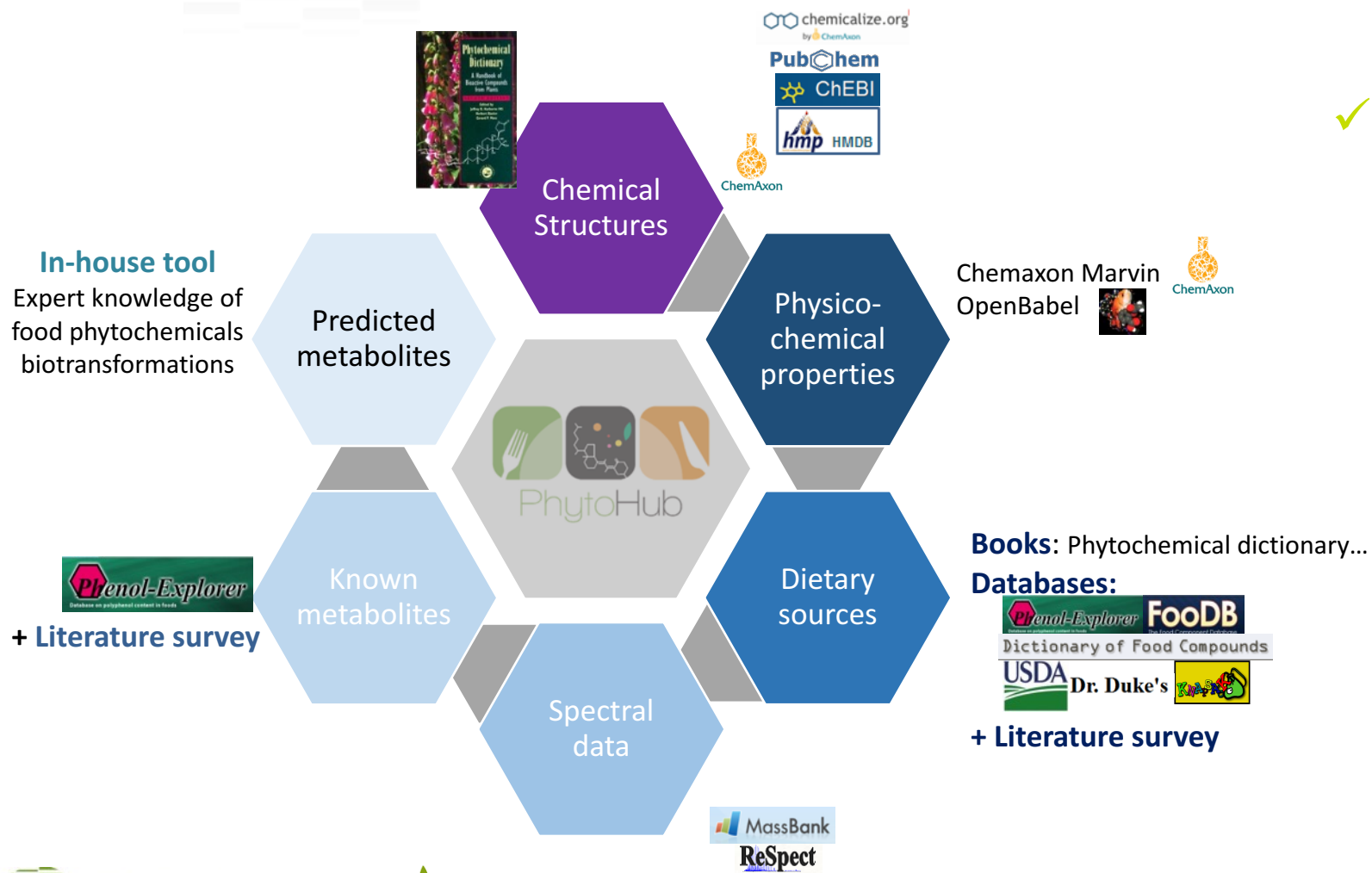
PI : Franck Giacomoni



<https://peakforest.org>



# Des banques... des stratégies... des résultats... Consolidons nos connaissances



- ✓ Base de connaissances sur le « Food Metabolome »
- ✓ Une base de données en ligne et en « open access »
  - ✓ Dédiée aux phytomicronutriments et leurs métabolites



PI : Claudine Manach



[www.phytohub.eu](http://www.phytohub.eu)



# Des banques... des stratégies... des résultats... des connaissances... Référez SVP !

Galaxy / 4 / Metabolomics

Tools

search tools

Upload File from your computer

LC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

COMMON TOOLS

Data Handling

Text Manipulation

Filter and Sort

Join, Subtract and Group

Statistics

Graph/Display Data

Deprecated Tools

New

Workflows

All workflows

4 Wm Workflow4metabolomics

Current version : 3.0

Publication: Franck Giacomoni, Gildas Le Corguillé, Mishari Monsoor, Marion Landi, Pierre Pericard, Mélanie Pétéra, Christophe Duperier, Marie Tremblay-Franco, Jean-François Martin, Daniel Jacob, Sophie Goulitquer, Etienne A. Thévenot and Christophe Caron (2014). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. Bioinformatics doi:10.1093/bioinformatics/btu813

Help and support: support@workflow4metabolomics.org

Latest news

- 10/05/2017 - LC-MS: A new tutorial video explain how to run xcmsSet in parallel on single files [link]
- 20/04/2017 - Workflow4Metabolomics v3.0 starts today - Check the changelog section below

Changelog

Tutorials

LC/MS

GC/MS

NMR

Common

History

search datasets

Sacuri35 -> CAMERA

30 shown, 180 hidden

3.58 GB

- 210: xcms.summary.html
- 209: xcms.summary.html
- 208: xset.merged.group.retcor.group.fillPeak.sannotateDiffreport (sample-vs-b.tsv)
- 207: xset.merged.group.retcor.group.fillPeak.sannotateDiffreport (pool-vs-sample.tsv)
- 206: xset.merged.group.retcor.group.fillPeak.sannotateDiffreport (pool-vs-b.tsv)
- 205: xset.merged.group.retcor.group.fillPeak.sannotateDiffreport (sample-vs-b-eic.pdf)
- 204: xset.merged.group.retcor.group.fillPeak.sannotateDiffreport (sample-vs-b-box.pdf)
- 203: xset.merged.group.retcor.group.fillPeak.sannotateDiffreport (pool-vs-sample-eic.pdf)
- 202: xset.merged.group.retcor.group.fillPeak.sannotateDiffreport (pool-vs-sample-box.pdf)
- 201: xset.merged.group.retcor.group.fillPeak

- **W4M** participe à l'effort de l'« open science »
- **W4M** fournit des DOI afin de référencer les historiques de ses utilisateurs (de la données brute aux résultats statistiques et d'annotation, les outils utilisés, leurs paramètres et les workflows)
- Utilisable dans les publications



*Guillon et al (2017). IJBC, doi:10.1016/j.biocel.2017.07.002*

[http://workflow4metabolomics.org/referenced\\_W4M\\_histories](http://workflow4metabolomics.org/referenced_W4M_histories)



# Des banques... des stratégies... des résultats... des connaissances...

## Référenciez SVP !



### MetaboLights: An Open-Access Database Repository for Metabolomics Data

Namrata S. Kale,<sup>1</sup> Kenneth Haug,<sup>1</sup> Pablo Conesa,<sup>1</sup> Kalaivani Jayseelan,<sup>1</sup>  
Pablo Moreno,<sup>1</sup> Philippe Rocca-Serra,<sup>2</sup> Venkata Chandrasekhar Nainala,<sup>1</sup>  
Rachel A. Spice  
Julian L. Griffin

Published online 29 October 2012

*Nucleic Acids Research*, 2013, Vol. 41, Database issue **D781–D786**  
doi:10.1093/nar/gks1004

### MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data

Kenneth Haug<sup>1</sup>, Reza M. Salek<sup>1,2,3</sup>, Pablo Conesa<sup>1</sup>, Janna Hastings<sup>1</sup>, Paula de Matos<sup>1</sup>,  
Mark Rijnbeek<sup>1</sup>, Tejasvi Mahendraker<sup>1</sup>, Mark Williams<sup>1</sup>, Steffen Neumann<sup>4</sup>,  
Philippe Rocca-Serra<sup>5</sup>, Eamonn Maguire<sup>5</sup>, Alejandra González-Beltrán<sup>5</sup>,  
Susanna-Assunta Sansone<sup>5</sup>, Julian L. Griffin<sup>2,3</sup> and Christoph Steinbeck<sup>1,\*</sup>

<sup>1</sup>European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD,  
<sup>2</sup>MRC HNR, Elsie Widdowson Laboratory, Fulbourn Road, Cambridge CB1 9NL, <sup>3</sup>Department of Biochemistry  
and Cambridge Systems Biology Centre, University of Cambridge, Cambridge CB2 1GA, UK, <sup>4</sup>Department of  
Stress- and Developmental Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle, Germany  
and <sup>5</sup>Oxford e-Research Centre, University of Oxford, 7 Keble Road, Oxford OX1 3QG, UK

Received August 15, 2012; Revised September 22, 2012; Accepted October 1, 2012





## MetaboLights

MetaboLights is a database for Metabolomics experiments and derived information. The database is cross-species, cross-technique and covers metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from metabolic experiments. MetaboLights is the recommended Metabolomics repository for a number of [leading journals](#).

## Download

 **Pre-packaged ISAcceptor download.** To make it easy for new users, please download and just unzip our pre-packaged ISAcceptor with plugin and configurations.

 **Experiments.** All public MetaboLights experiments can be downloaded from our public [ftp archive](#). Please find zip archives under the "studies" folder. Each public study can be found in the corresponding MTBLS-id folder. Complete experiments can be opened with [ISAcceptor](#) or you can extract the archives using your normal unzip program.

## Tweets by @MetaboLights

**MetaboLights** @MetaboLights

MTBLS570: Metabolic profiling of polarised macrophages determines cont... [ebi.ac.uk/metabolights/M...](http://ebi.ac.uk/metabolights/M...)



4h

**MetaboLights** @MetaboLights

MTBLS616: A lipidomic human population and translational feeding study... [ebi.ac.uk/metabolights/M...](http://ebi.ac.uk/metabolights/M...)



5h

**MetaboLights** @MetaboLights[Embed](#)[View on Twitter](#)

## Submit a new study

Use this option if your study has not been submitted before

## Update an existing study

Use this option if you like to update a previously submitted study



This service is part of the ELIXIR infrastructure  
MetaboLights is an ELIXIR Recommended Deposition Database [Learn more](#)

EMBL-EBI

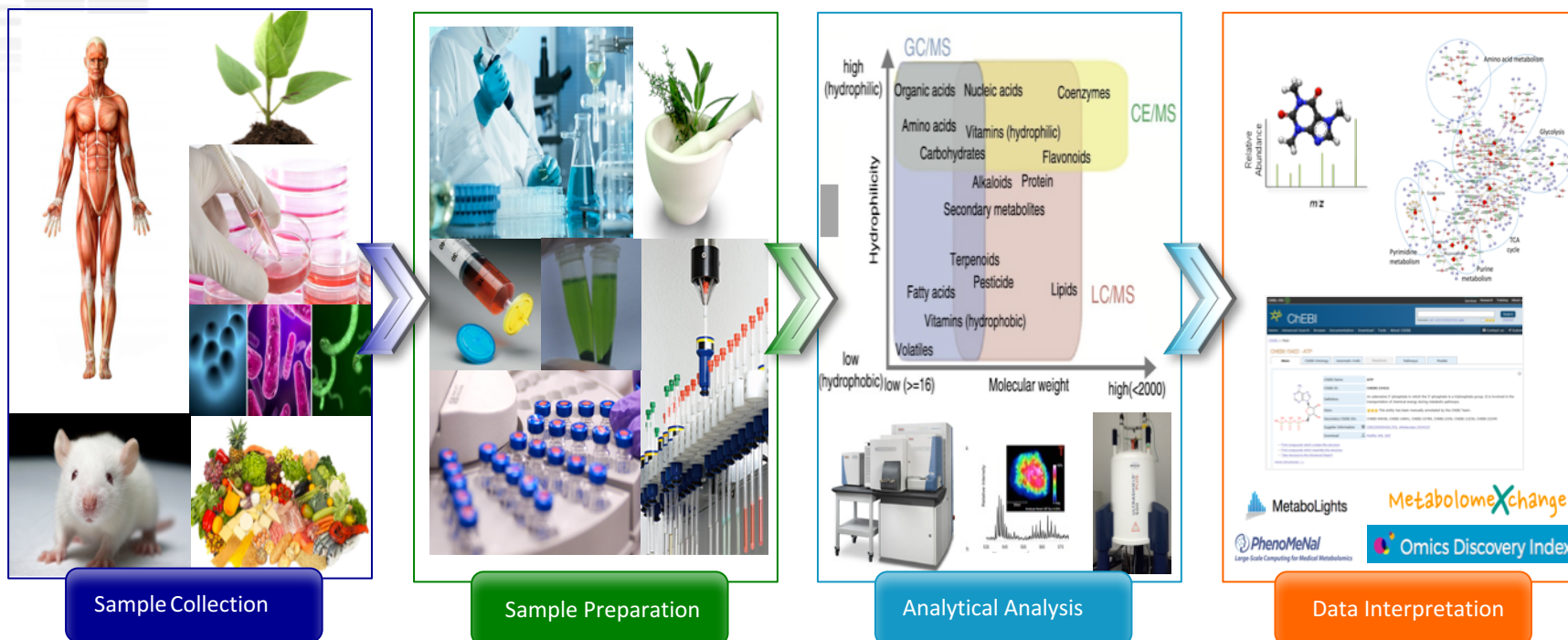


<http://www.ebi.ac.uk/metabolights>

Email: [metabolights-help@ebi.ac.uk](mailto:metabolights-help@ebi.ac.uk)



# MetaboLights - Open Source Database



- « Cross » espèces, « cross » techniques
- Collection d'information la plus complète possible
- « Open access », accessible via différents portails

# Conclusions

- *L'annotation est une étape critique et délicate – rôle premier des banques de données*
- *Echanges + efforts entre chimistes / biologistes / bioinformaticiens*
- *La constitution et le maintien de banques de données pour la métabolomique, de qualité est primordiale*
- *Challenge(s) à relever :*
  - *Construire collections de métabolomes de références d'organismes modèles*
  - *Définition d'un composé chimique (nom(s), représentation 2D / 3D)*
  - *Développer les aspects « interopérabilité » avec les autres ressources de référence du domaine (Workflow4Metabolomics, MetExplore ;-) ... )*