Mass spectrometry and Free Software in Debian

Filippo Rusconi, Ph.D.
filippo.rusconi@u-psud.fr

Laboratoire de Chimie Physique
CNRS UMR 8000
Université Paris-Sud 11
F-91405 Orsay

ESRF — June 25th 2012,

Outline

Mass spectrometry:

- Uses in biochemical sciences;
- What does it actually do ?
- Why free software is ever more considered essential;
- Software available in Debian and other packaging work.

Mass spectrometry : uses in biochemical sciences

Measure the molecular (so-to-say) mass of molecules

Measuring the mass of a molecule is useful in:

- Quality control of medicinal substances;
- Characterization of molecules
  $\rightarrow$ polymeric or not, natural or synthetic;
- Basic sciences : biopolymer chemistry
  $\rightarrow$ proteins, nucleic acids (DNA | RNA), sugars
  $\rightarrow$ molecule identification and structural analysis.

Mass spectrometry : how does it work ?

Goal: to measure the mass of ionized molecules

Acquiring one mass spectrum = 3 experiments

1. Ionize the molecule of interest
   (⇒ electrically charged analyte);
2. - Apply forces onto the ion (magnetic | electrostatic)
     ⇒ elicit a movement;
   - Analyze the path of the ion in the vacuum
     ⇒ relate to the $\frac{m}{z}$ ratio of the ion;
3. Count the ions to craft the mass spectrum: $i = f(m/z)$

The path followed by any one ion is related to the "mass-to-charge" m/z ratio of that ion

A mass spectrometer

Source, analyser and detector (ion counter)

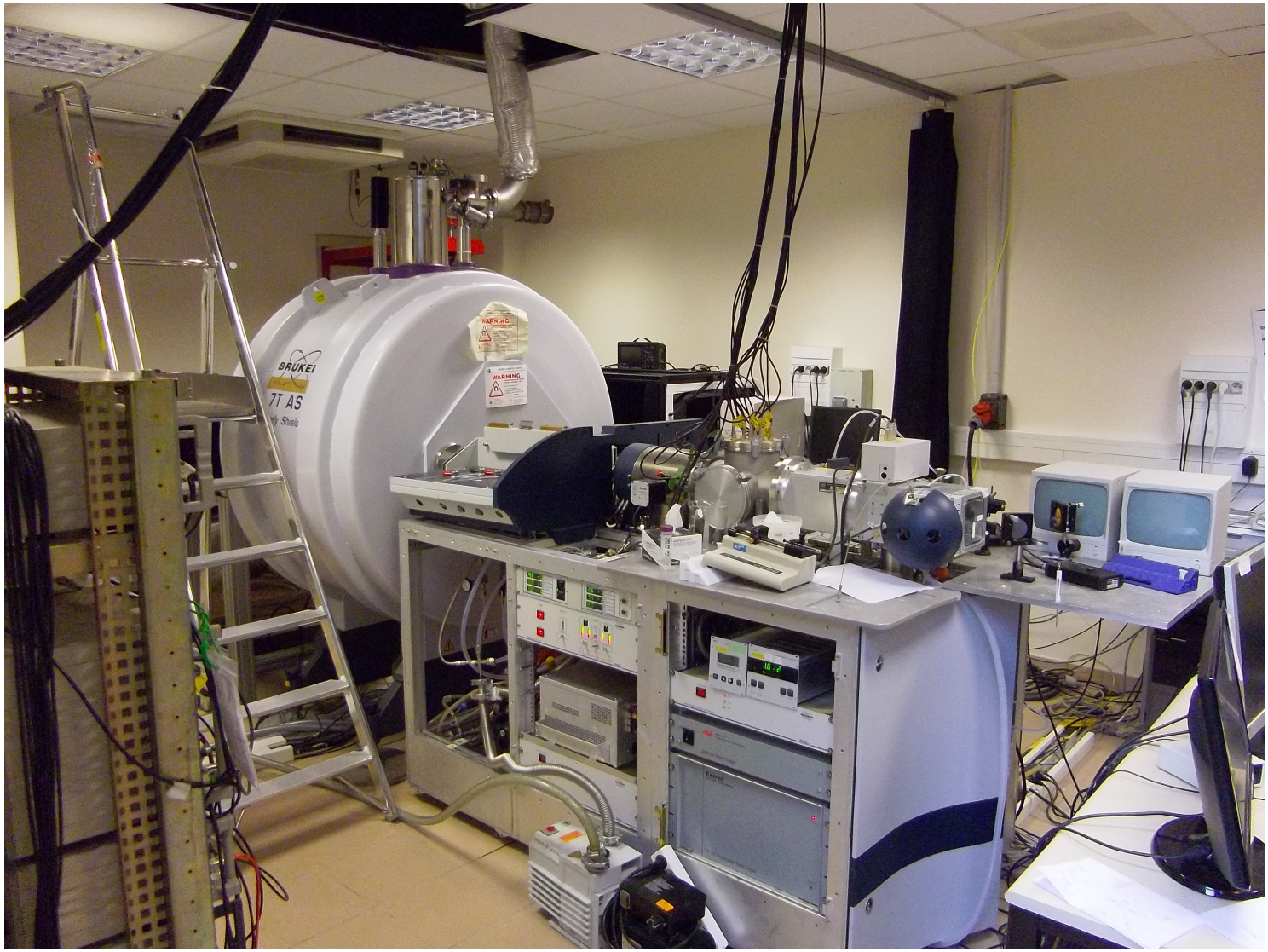
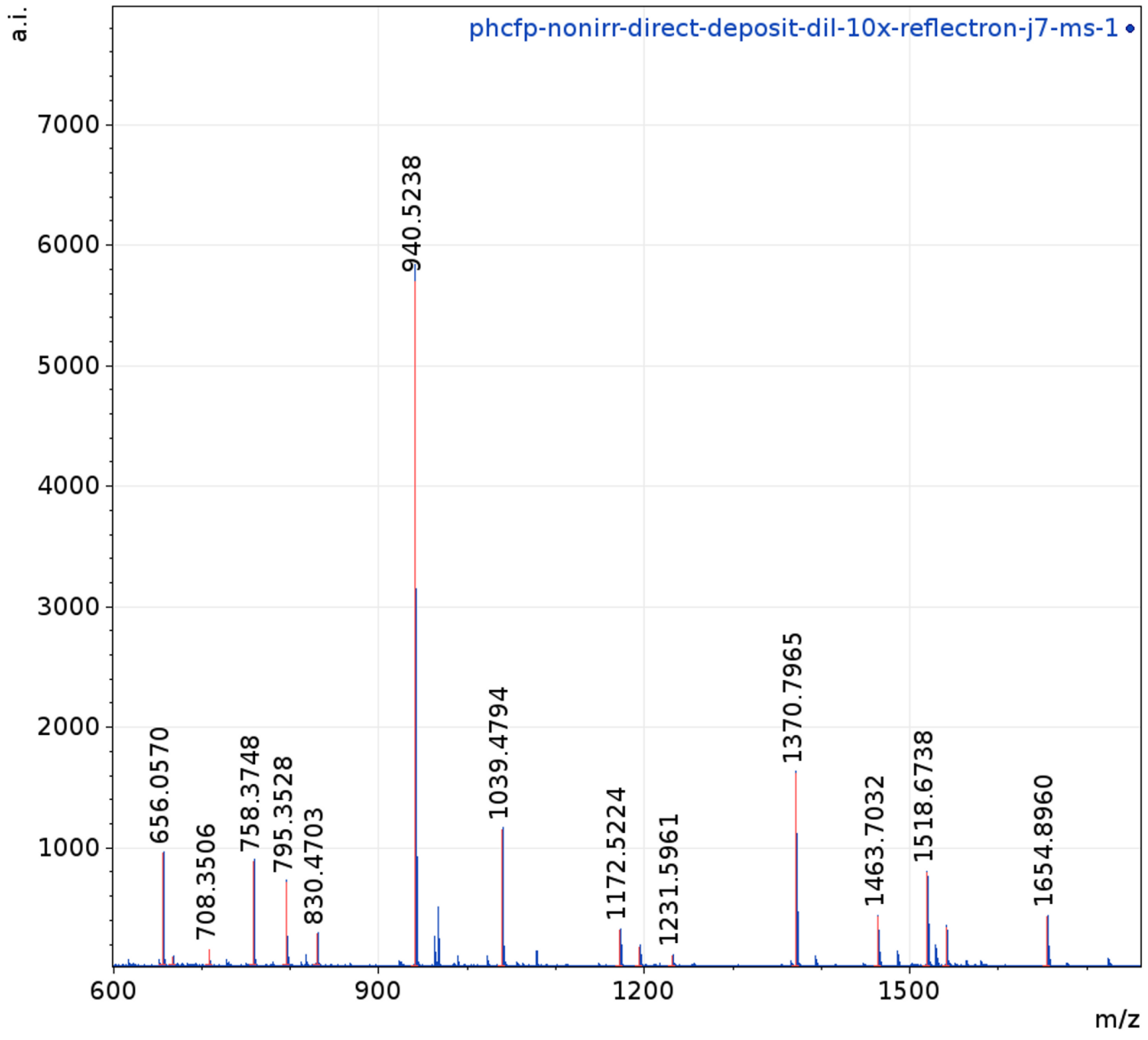


Photo: Vincent Steinmetz

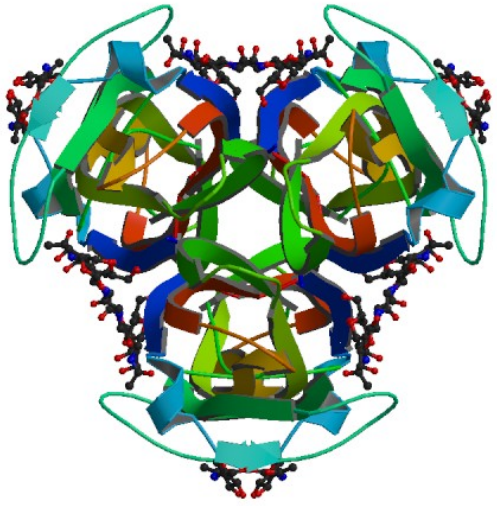
A mass spectrum

Detected ion masses *versus* the count of the ions



Protein structural analysis

The very best way to get insights into the structure of a polymer: cristallography



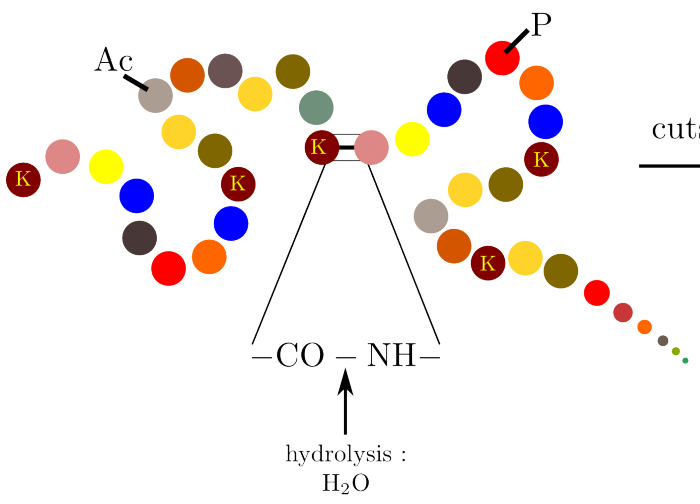
Helix pomatia (snail) agglutinin : a glycoprotein

Requirements:

- Large amounts of material (tens or hundreds of milligrams);
- Purified or highly enriched material (difficult);
- Cristal growing is a difficult/empirical process.

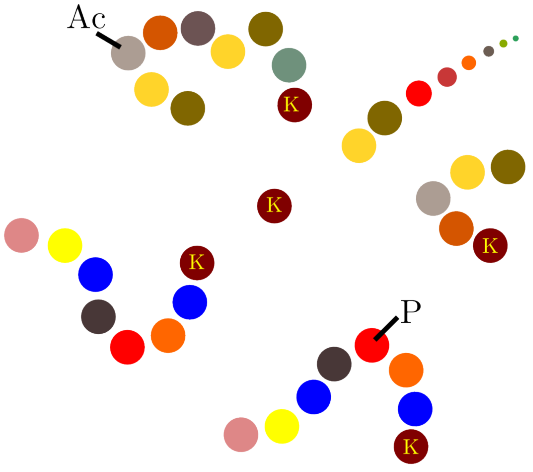
Mass spectrometry : “replacement” for cristallography

Requirement for some protein chemistry



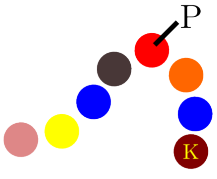
Analyzed protein

Trypsin
cuts right of ‘K’



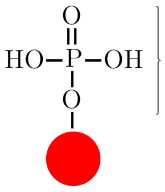
Peptides (protein fragments)

Mass computation
(*in silico*)



Monomeric composition
with relative masses

	57
	99
	$101 \times 2 = 202$
	128
	129
K	186
	113
P	80
	$\sum$ 914 : without chemical modification
	994 : addition of the phosphoryl mass (80 Da)

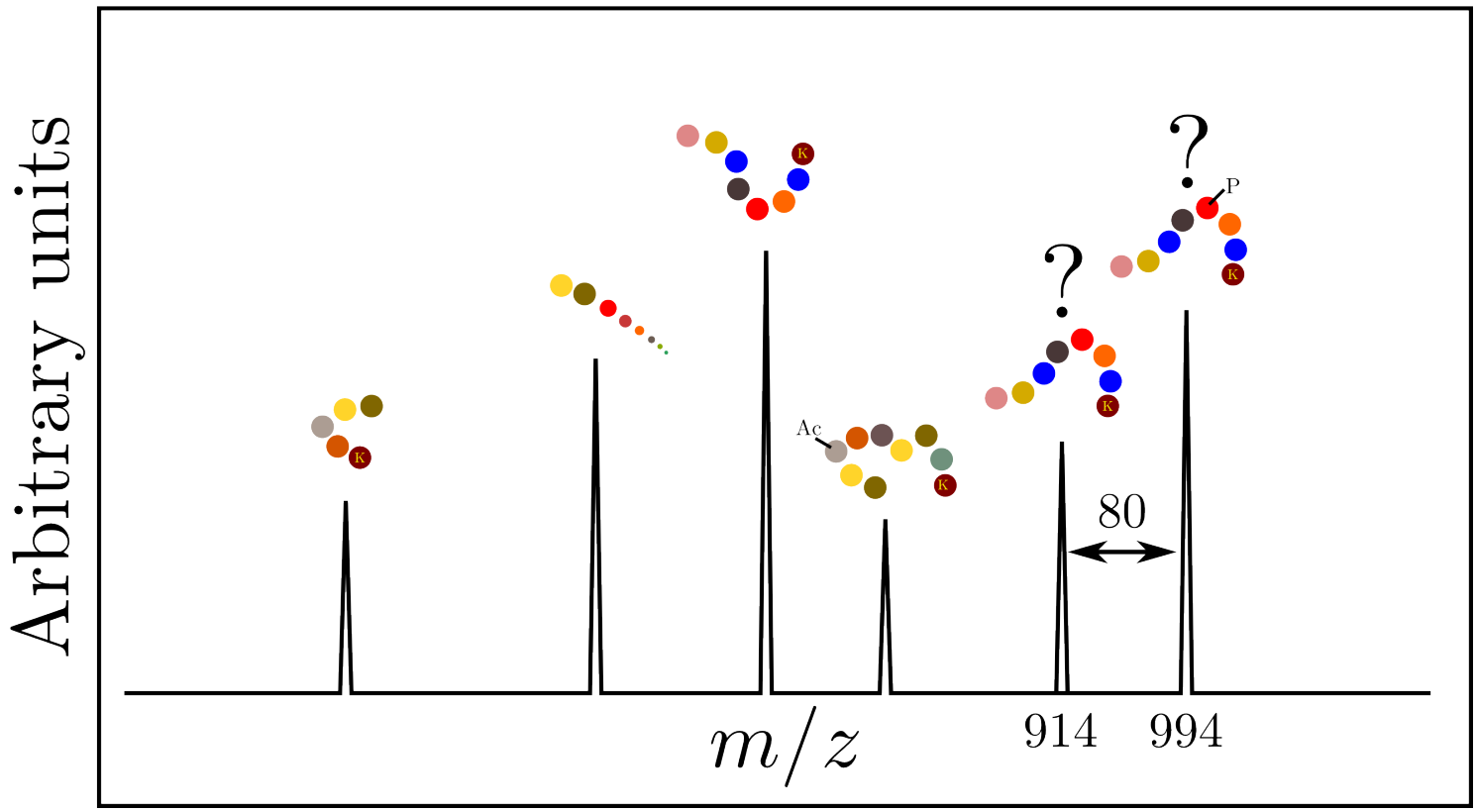


Mass change + 80 u

Mass spectrometry : “replacement” for cristallography

First-level protein structure characterization

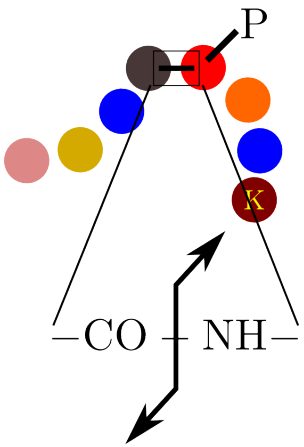
Peptide mass fingerprint (PMF)



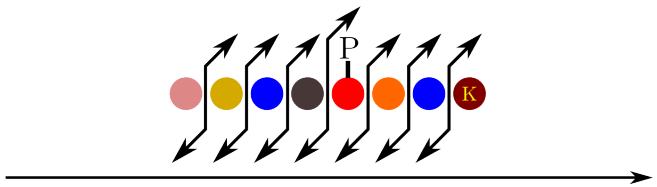
Mass spectrometry : “replacement” for cristallography

In-depth protein structure characterization

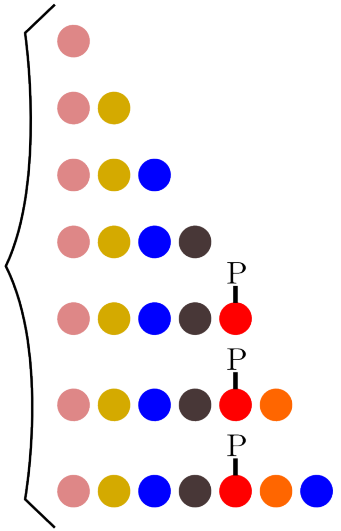
Gas-phase fragmentation (in the mass spectrometer)

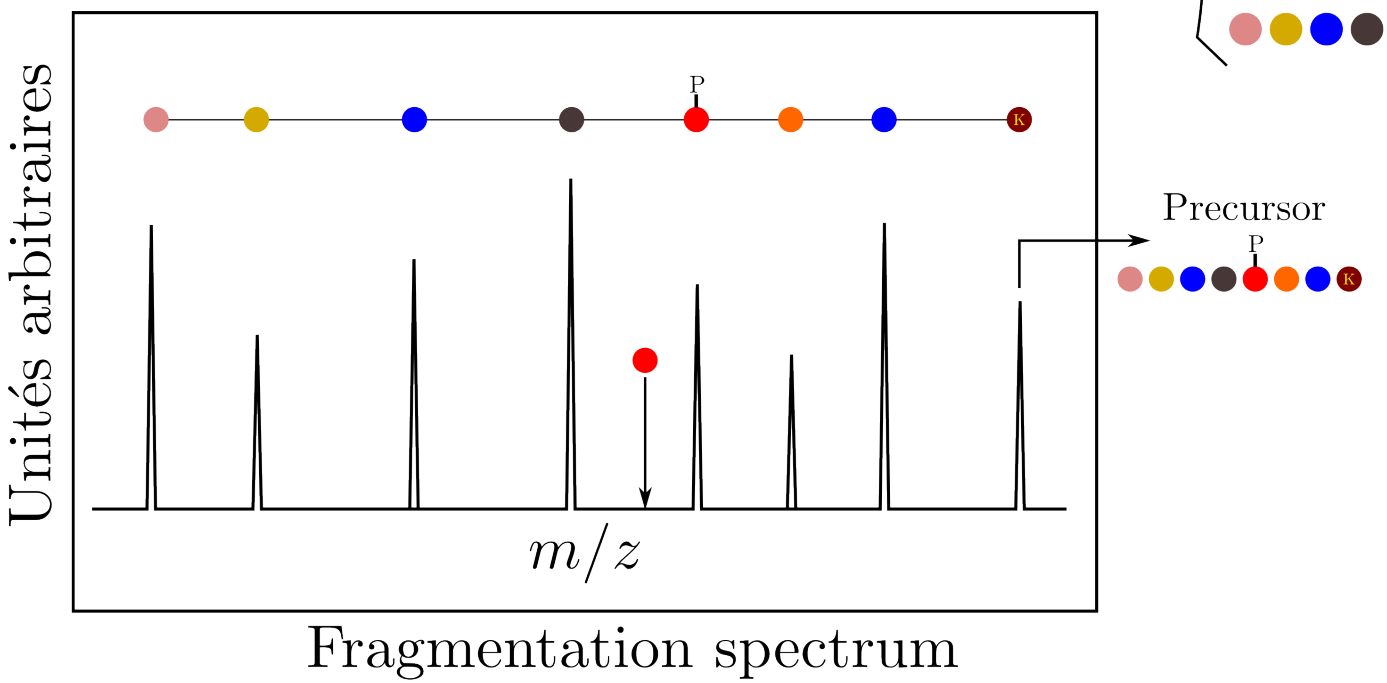


Precursor molecule



Fragmentation !

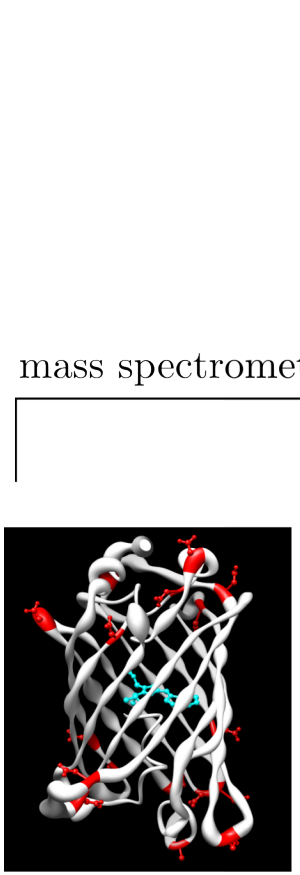




Fragmentation spectrum

Fluorescent proteins: how stable are these proteins ?

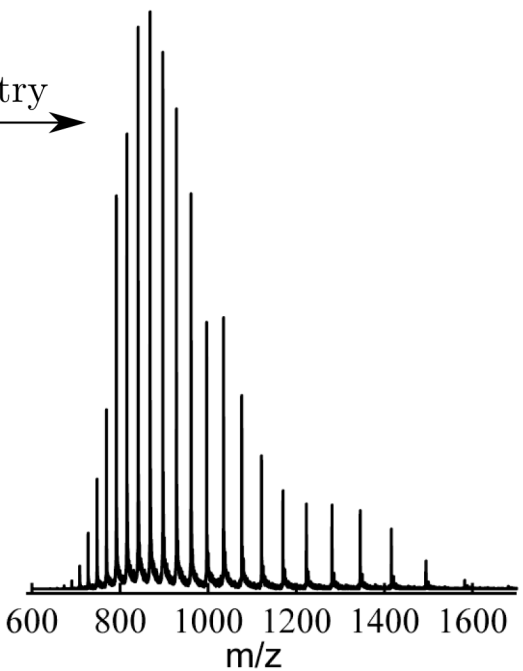
. . . In particular under γ radiation conditions ? Multiple oxidation events detected !

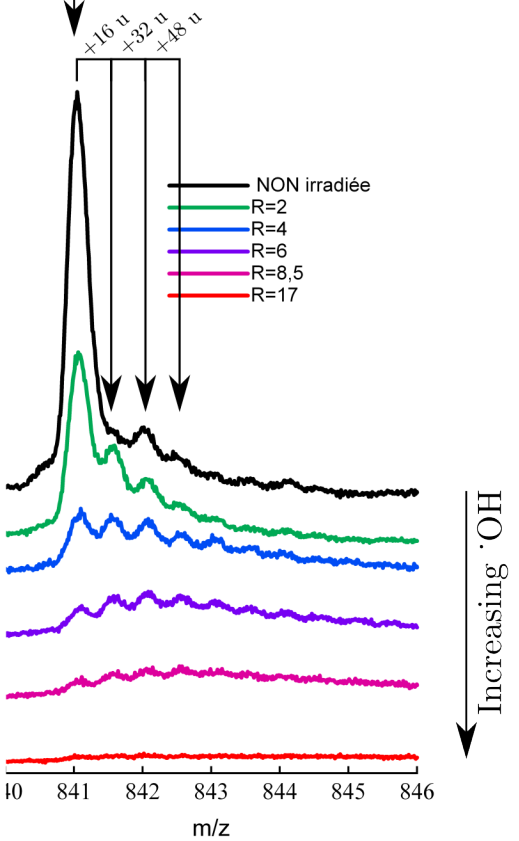


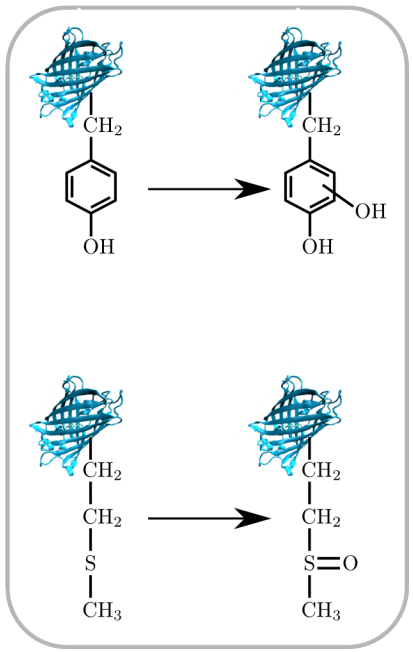
CFP
cyan fluorescent protein

mass spectrometry

Effect of γ-irradiation







Collaboration Dr Marie Erard — Univ. Paris Sud — Orsay

There are at least five different events of oxidation

Importance of Free Software in mass spectrometry

Hardware manufacturers are fiercely struggling to gain exclusive control...

...both on the mass data and on the users themselves...
("vendor lock-in")

- Software is used as a sales pitch (particularly, LIMS†);
- Proprietary formats (terabytes of data stored !);
- Always ask for an "Export to mzML" [or to (x, y)] feature;
- What are the software offerings for mass spectrometry facilities ?
- massXpert: experiment preparation in biopolymer chemistry
  simulation and analysis of mass data;
- mMass: powerful mass spectrum display and analysis.

† Laboratory Information Management System

Typical workflow

From the experiment design to the biological interpretation *via* mass data analysis

1. Sample conditioning :
   - Choose the proper chromatographical technique to purify the protein of interest (massXpert);
   - Choose the proper enzyme to cleave the protein into peptides (massXpert);
2. Acquire mass data (proprietary software);
3. Spectrum display and data analysis (mMass);
4. Spectrum interpretation (massXpert).

Mass spectrometry-related Debian packages

Already available. . .

DebiChem Polymer editors and mass spectrometry packages

Official Debian packages with high relevance

Lutefisk
de novo interpretation of peptide CID spectra
http://www.hairyfatguy.com/lutefisk/
Maintainer: The Debichem Group (Filippo Rusconi)

Popcon: 3 users (1 upd.)*
Versions and Archs
Debtags

License: DFSG free
Official Debian package
Git

Lutefisk performs a de novo interpreation of CID spectra, providing the user with a file containing all the possible sequence candidates corresponding to the CID data.

Please cite: Richard S. Johnson and J. Alex Taylor: *Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry.* (PubMed,eprint) Molecular Biotechnology 22(3):301-315 (2002)

Upload screenshot

Massxpert
linear polymer mass spectrometry software
http://www.massxpert.org
Maintainer: The Debichem Group (Filippo Rusconi)

Popcon: 24 users (12 upd.)*
Versions and Archs
Debtags

License: DFSG free
Official Debian package
Git

massXpert is a program to simulate and analyse mass spectrometric data obtained on linear (bio-)polymers. It is the successor of GNU polyxmass.
Four modules allow:

- making brand new polymer chemistry definitions;
- using the definitions to perform easy calculations in a desktop calculator-like manner;
- performing sophisticated polymer sequence editing and simulations;
- perform m/z list comparisons;

Chemical simulations encompass cleavage (either chemical or enzymatic), gas-phase fragmentations, chemical modification of any monomer in the polymer sequence, cross-linking of monomers in the sequence, arbitrary mass searches, calculation of the isotopic pattern...

Please cite: F. Rusconi: *massXpert 2: a cross-platform software environment for polymer chemistry modelling and simulation/analysis of mass spectrometric data.* (2009)

Mmass
Mass spectrometry tool for proteomics
http://www.mmass.org/
Maintainer: The Debichem Group (Filippo Rusconi)

Popcon: 7 users (4 upd.)*
Versions and Archs
Debtags

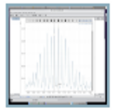
License: DFSG free
Official Debian package
Git

mMass is a free mass spectrum viewer/analyzer in which the following proteomics-related tasks can be performed:

- Open raw text, mzXML and mzData mass spectra;
- Define peak lists;
- Powerful mass spectrum viewer (zoom, cursor...);
- Data recalibration;
- Protein-only simulations;
- Online Mascot searches.

The software can be easily extended by additional Python modules. This package contains the platform-independent parts of the software.

Please cite: M. Strohalm, D. Kavan, P. Novak, M. Volny and V. Havlicek: *mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data.* (2010)



Crystallography-related Debian packages

DebiChem Crystallography packages

Official Debian packages with high relevance

Drawxtl
crystal structure viewer
http://www.lwfinger.com/drawxtl/index.html
Maintainer: Debichem Team (Daniel Leidert (dale))

Popcon: 31 users (100 upd.)*
Versions and Archs
Debtags

License: DFSG free
Official Debian package
Svn

DRAWxtl reads a basic description of the crystal structure, which includes unit-cell parameters, space group, atomic coordinates, thermal parameters or a Fourier map, and outputs a geometry object that contains polyhedra, planes, lone-pair cones, spheres or ellipsoids, bonds, iso-surface Fourier contours and the unit-cell boundary.
Four forms of graphics are produced:

- an OpenGL window for immediate viewing
- the Persistence of Vision Ray Tracer (POV-RAY) scene language for publication-quality drawings
- the Virtual Reality Modeling Language (VRML) for dissemination across the Internet
- a Postscript rendering of the OpenGL window for those who want high-quality output but do not have POV-RAY installed.

File formats DRAWxtl can read include CIF, FDAT, FullProf (pcr), GSAS, SCHAKAL, SHELX, DISCUS and WIEN2k.

Upload screenshot

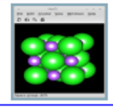
Please cite: Larry W. Finger, Martin Kroeker and Brian H. Toby: *DRAWxtl, an open-source computer program to produce crystal-structure drawings.* (eprint) J. Appl. Cryst. 40:188-192 (2007)

Gcrystal
lightweight crystal structures visualizer
http://www.nongnu.org/gchemutils/
Maintainer: Debichem Team (Daniel Leidert (dale))

Popcon: 30 users (109 upd.)*
Versions and Archs
Debtags

License: DFSG free
Official Debian package
Svn

GNOME Crystal is a light model visualizer for crystal-structures. It is based on the GNOME Chemistry Utils and should display models of all sorts of crystal microscopic structures using OpenGL.



Shelxle
graphical user interface for SHELXL
http://ewald.ac.chemie.uni-goettingen.de/shelx/
Maintainer: Debichem Team (Daniel Leidert (dale))

Popcon: 5 users (3 upd.)*
Versions and Archs
Debtags

License: DFSG free
Official Debian package
Svn

ShelXle combines an editor with syntax highlighting for the SHELXL-associated .ins (input) and .res (output) files with an interactive graphical display for visualization of a three-dimensional structure including the electron density (Fo) and difference density (Fo-Fc) maps.
http://dx.doi.org/10.1107/S0021889811043202

Upload screenshot

Please cite: Christian B. Hübschle, George M. Sheldrick and Birger Dittrich: *ShelXle: a Qt graphical user interface for SHELXL.* (eprint) J. Appl. Cryst. 44(6):1281-1284 (2011)

Xcrysden
Crystalline and Molecular Structure Visualizer
http://www.xcrysden.org/
Maintainer: Debichem Team (Anton Kokalj)

Popcon: 7 users (5 upd.)*
Newer upstream!
Go tagging

License: DFSG free
Official Debian package

XCrySDen is a crystalline and molecular structure visualisation program, which aims at display of isosurfaces and contours, which can be superimposed on crystalline structures and interactively rotated and manipulated. It can run on most UNIX platforms, without any special hardware requirements.

Upload screenshot

Please cite: Anton Kokalj: *Computer graphics and graphical user interfaces as tools in simulations of matter at the atomic scale.* (eprint) Comp. Mater. Sci. 28(2):155-168 (2003)

Debian packages

Some more work to do (no ITPs yet)... at `http://open-ms.sourceforge.net/`

1. `massXpert`:

   Requested feature : scriptability (Python ?);

2. `OpenMS`: (from the site)

   "OpenMS is an open-source software C++ library for LC/MS data management and analyses. It offers an infrastructure for the development of mass spectrometry-related software. OpenMS is free software available under the LGPL."

3. `TOPP` :

   - Whole suite of programs to perform piped mass data processing;
   - Makes use of `OpenMS`;
   - Highly complex set of libraries and tools.

When Debian will be a full-featured software solution?

What kind of software would a turn-key software solution need?

- First and central: a relational database system to store all the data (mass spectrometric data and chromatographic data);
- A set of (pipe) |-able software pieces aimed at the fine-grained analysis of mass data (TOPP, for example);
- Connectors to external databases: most analysis work is by trying to perform matches between external data (often theoretical data and own data);
- Powerful reporting tools.

Challenges. . .

When things can hardly be distributed. . .

- Non-free but highly useful software;
- Databases of natural data;

(pedagogy and diplomacy help sometimes)

Thank you for listening