Recent developments in the identification of natural products by NMR

Jean-Marc Nuzillard and Pedro Lameiras

Institut de Chimie Moléculaire de Reims (ICMR), UMR CNRS 7312, SFR CAP'Santé, Université de Reims Champagne-Ardenne, France

Three topics in 15 minutes

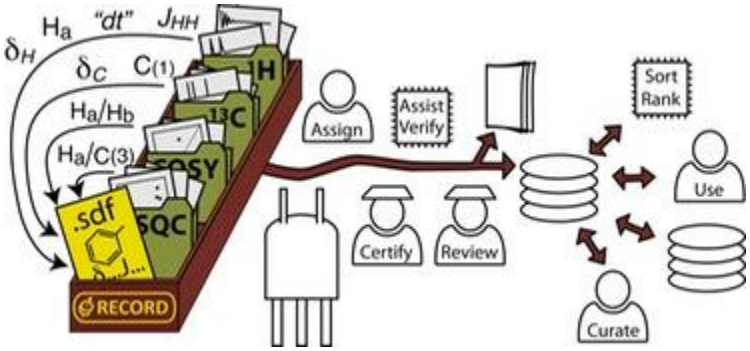
- Making NMR data « public »

- Structure elucidation of pure compounds


- Mixture analysis using viscous solvents


Making NMR data « public »

- The NMReDATA Initiative



- The Raw Data Initiative

- The NMReDATA Initiative

  – Proposed by «the « Associate Editorial Board » of « Magnetic Resonance in Chemistry »

  – Lead by Dr. Damien Jeannerat, University of Geneva

  – Members
    - Individuals
    - **Software Developers**
    - Academic Institutions, including **IUPAC**
    - Journals

  – http://nmredata.org/

  – *Magn Reson Chem.* **2018**, *56*, 703–715.

- The NMReDATA Initiative

"The goal of the NMReDATA initiative is to improve the FAIRness and quality of the NMR data available to the community. We introduced a format for the data, but more importantly, a manner to organize the data in such a way that the assignment data can be stored in a reliable manner on a freely accessible database and allow for their verification against the experimental spectra."

FAIR: Findable, Available, Interoperable, Reusable

- The NMReDATA File Format

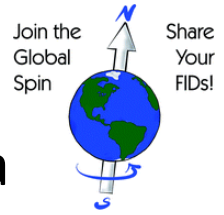
  – NMReDATA: NMR extracted DATA
  Chemical shifts, coupling constants. integrals, correlations, assignments,…

3.23 (dd, 1H, H_2, ^3J
$H_{2'}$, $^3J_{H2'-H1'}$ = 3.1 H
.46 (m, 2H, H_3); 3.
3.49–3.51 (m, 2H,

  – Based on the (MDL/Symyx/Accelrys/Dassault)
  SDF Structure-Data format, derived from the MOL format

  – An NMReDATA file contains a molecular structure and spectroscopic extracted data

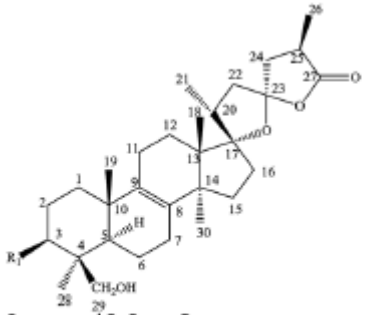
  – Links to original time and/of frequency domain data



- Why ?

NMR DATA

  - Structures in Journals are not readable by computers



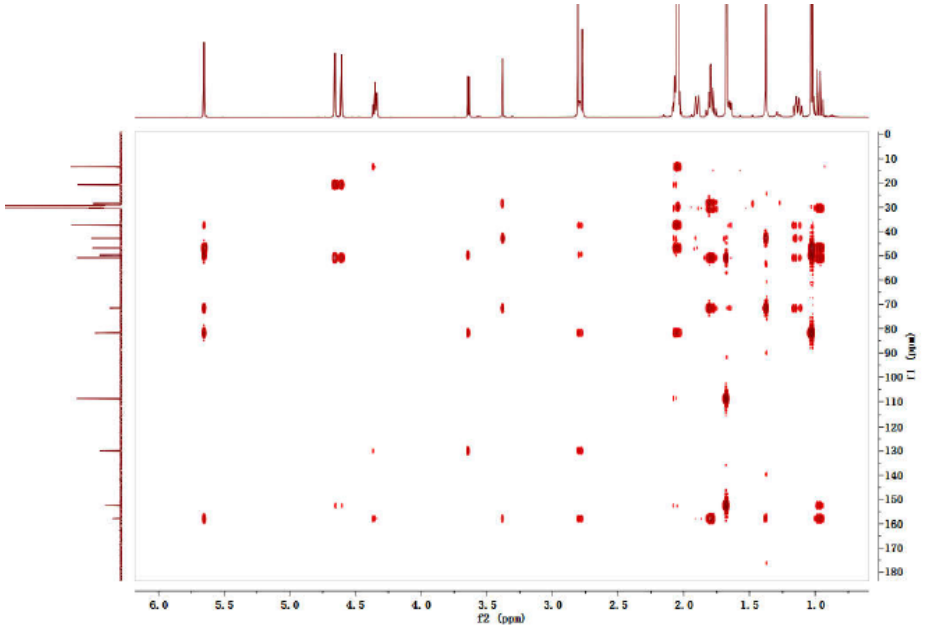
Compound 3: R_1 = =O
Compound 8: R_1 = β-D-glucopyranoside
Compound 9: R_1 = α-L-arabinopyranosyl-(1→6)-β-D-glucopyranoside

  - Spectrum descriptions are not readable by computers

3.12 (dd, 1H, H_2, $^2J_{H2-H1}$ = 3.8 Hz, $^2J_{H2-H3}$ = 9.7 Hz); 3.17–3.18 (m, 4H, H_2); 3.23 (dd, 1H, H_2, $^3J_{H2-H1}$ = 3.2 Hz, $^3J_{H2-H3}$ = 9.7 Hz); 3.27 (dd, 1H, $H_{2'}$, $^3J_{H2'-H1'}$ = 3.1 Hz, $^3J_{H2'-H3'}$ = 8.1 Hz); 3.37 (s, 21H, 3-OMe); 3.46 (m, 2H, H_3); 3.47–3.55 (s, 21H, 2-OMe); 3.49 (m, 1H, H_6); 3.49–3.51 (m, 2H, H_3); 3.51 (m, 1H, H_3); 3.54–3.62 (m, 5H, H_6); 3.57 (dd, 1H, H_3, $^3J_{H3-H2}$ = $^3J_{H3-H4}$ = 9.5 Hz); 3.57 (d, 1H, H_6, $^3J_{H6-H5}$ = 9.7 Hz); 3.58–3.67 (s, 18H, 6-OMe); 3.60 (m, 2H, H_4); 3.61 (m, 1H, H_6); 3.62–3.63 (m, 3H, H_4); 3.63 (dd, 1H, $H_{3'}$, $^3J_{H3'-H2'}$ = $^3J_{H3'-H4'}$ = 7.8 Hz); 3.75 (m, 1H, H_5); 3.78 (dd, 1H, H_4, $^3J_{H4-H3}$ = $^3J_{H4-H5}$ = 9.2 Hz); 3.79–3.80 (m, 2H, H_5); 3.83 (d, 1H, H_6, $^3J_{H6-H5}$ = 10.1 Hz); 3.85 (m, 1H, H_6); 3.88 (d, 1H, H_6, $^3J_{H6-H5}$ = 10.0 Hz); 3.89 (m, 1H, H_5); 3.91 (m, 2H, H_5); 3.96 (d, 1H, H_6, $^3J_{H6-H5}$ = 11.3 Hz); 4.02 (m, 1H, NHCH_αCO); 4.05 (dd, 1H,

Atom no.	$^{13}C^a$	$^1H^b$	COSY[c]	HMBC[d]
1		9.06 *br s*		2, 7, 8, 13
2	133.8			
3	53.1	4.41 br s	14	2, 15[f]
5	51.5	a: 3.30 *br d* (11)[e]	5b, 6a, 6b[f]	3, 7
		b: 3.18 *br d* (11)[e]	5a, 6a, 6b	3, 7[f]
6	18.1	a: 3.01 *t* (15)[e]	5a, 5b, 6b	2[f], 5[f], 7[f]
		b: 2.65 *dd* (15.7, 4.7)	5a[f], 5b, 6a	2, 5[f], 7
7	108.4			
8	128.0			
9	118.7	7.47 *d* (7.5)	10	7, 8[f], 11, 13
10	119.9	7.07 *t* (7.3)	9, 11	8, 9, 12
11	122.0	7.11 *t* (7.3)	10, 12	9, 13
12	111.4	7.38 *d* (7.7)	11	8, 10

- Why ?
  - Spectra in Journals are not readable by computers



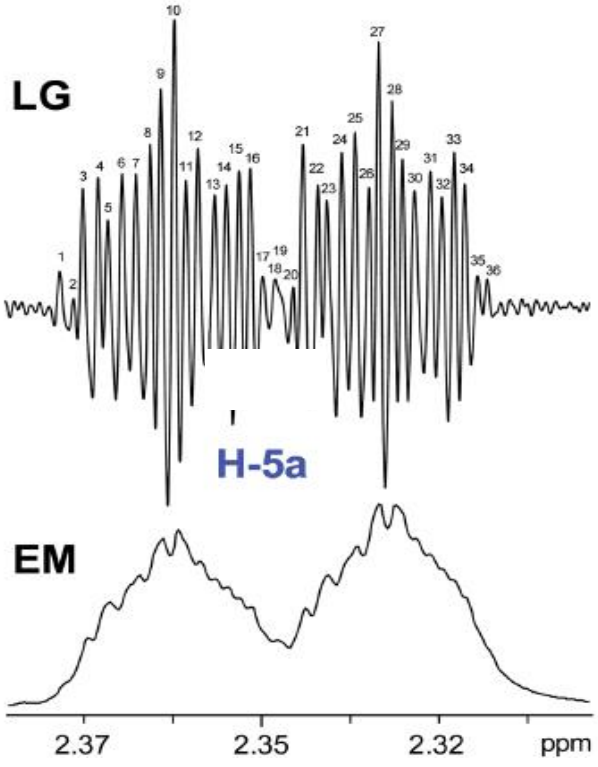
  - **NMReDATA files are readable by humans and computers**
  - **Submission of NMReDATA files is proposed to be part of the spectra/structures publication workflow.**

- Raw data initiative
  - The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research.
  - Nat Prod Rep. 2018 Jul 13. doi: 10.1039/c7np00064b, 73 pp.

McAlpine JB[1], Chen SN[1], Kutateladze A[2], MacMillan JB[3], Appendino G[4], Barison A[5], Beniddir MA[6], Biavatti MW[7], Bluml S[8], Boufridi A[9], Butler MS[10], Capon RJ[10], Choi YH[11], Coppage D[3], Crews P[3], Crimmins MT[12], Csete M[13], Dewapriya P[10], Egan JM[14], Garson MJ[15], Genta-Jouve G[16], Gerwick WH[17], Gross H[18], Harper MK[19], Hermanto P[20], Hook JM[20], Hunter L[20], Jeannerat D[21], Ji NY[22], Johnson TA[3], Kingston DGI[23], Koshino H[24], Lee HW[3], Lewin G[6], Li J[25], Linington RG[14], Liu M[9], McPhail KL[26], Molinski TF[27], Moore BS[17], Nam JW[27], Neupane RP[27], Niemitz M[27], Nuzillard JM[27], Oberlies NH[27], Ocampos FMM[5], Pan G[27], Quinn RJ[9], Reddy DS[2], Renault JH[27], Rivera-Chávez J[27], Robien W[27], Saunders CM[27], Schmidt TJ[27], Seger C[27], Shen B[27], Steinbeck C[27], Stuppner H[27], Sturm S[27], Taglialatela-Scafati O[27], Tantillo DJ[27], Verpoorte R[11], Wang BG[28], Williams CM[15], Williams PG[27], Wist J[27], Yue JM[27], Zhang C[27], Xu Z[27], Simmler C[1], Lankin DC[1], Bisson J[1], Pauli GF[1].

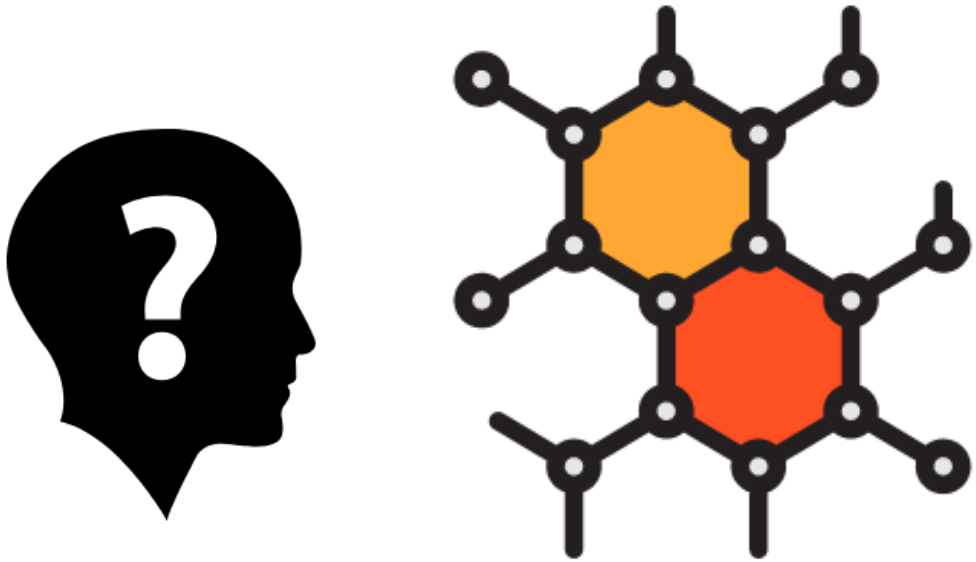
  - Supported by G. Pauli, College of Pharmacy, University of Illinois at Chicago.

- Raw data initiative article
  - The benefit of FID reprocessing
  - Recovery of lost information



Resolution enhancement by Lorentz-Gauss (LG) digital filtering

Standard noise filtering by exponential multiplication (EM) of the FID (LB = 0.3 Hz)



Structure elucidation of pure compounds

- *S*tructure determination of pure compounds

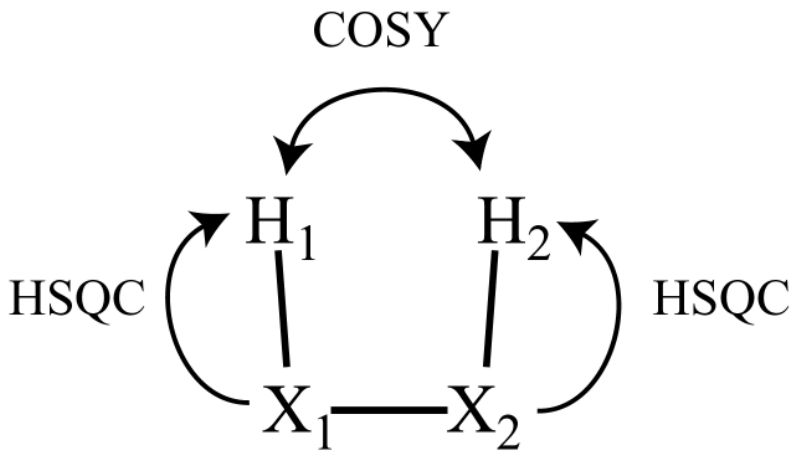
  - Obtain an elemental formula ($C_cH_hN_nO_o$…)
  - Extract pertinent data from 1D and 2D NMR spectra
  - Assemble one (or more) possible structures
  - Be critical on proposed solution(s)

  - Possibly get **assistance from a computer software**
  - Computer-Assisted Structure Elucidation: **CASE software**
  - Example: **Logic for Structure Determination** (LSD)
  - www.univ-reims.fr/LSD

- Academic Software
  - LSD (France)
  - COCON (Germany)
  - SENECA (UK)

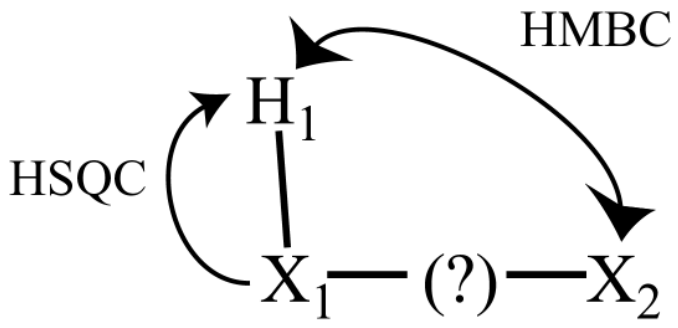
- Commercial software
  - ACD Structure Elucidator
  - Mestrelab Mnova (version >= 12)
  - Bruker CMC-se

2D NMR provides proximity relationships between non-H atoms



Simplified version

X_1—X_2



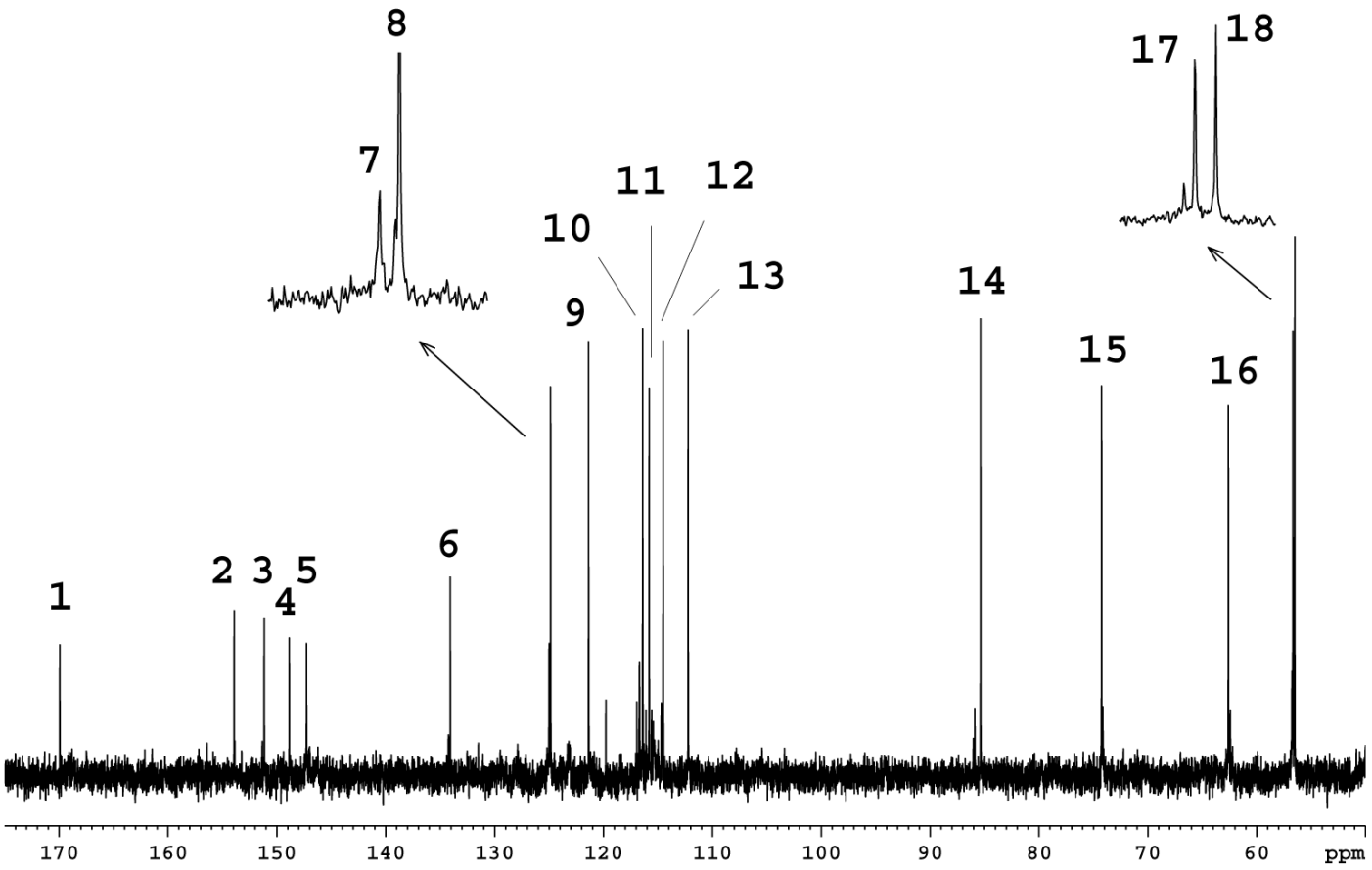
or

X_1—X_2

ou

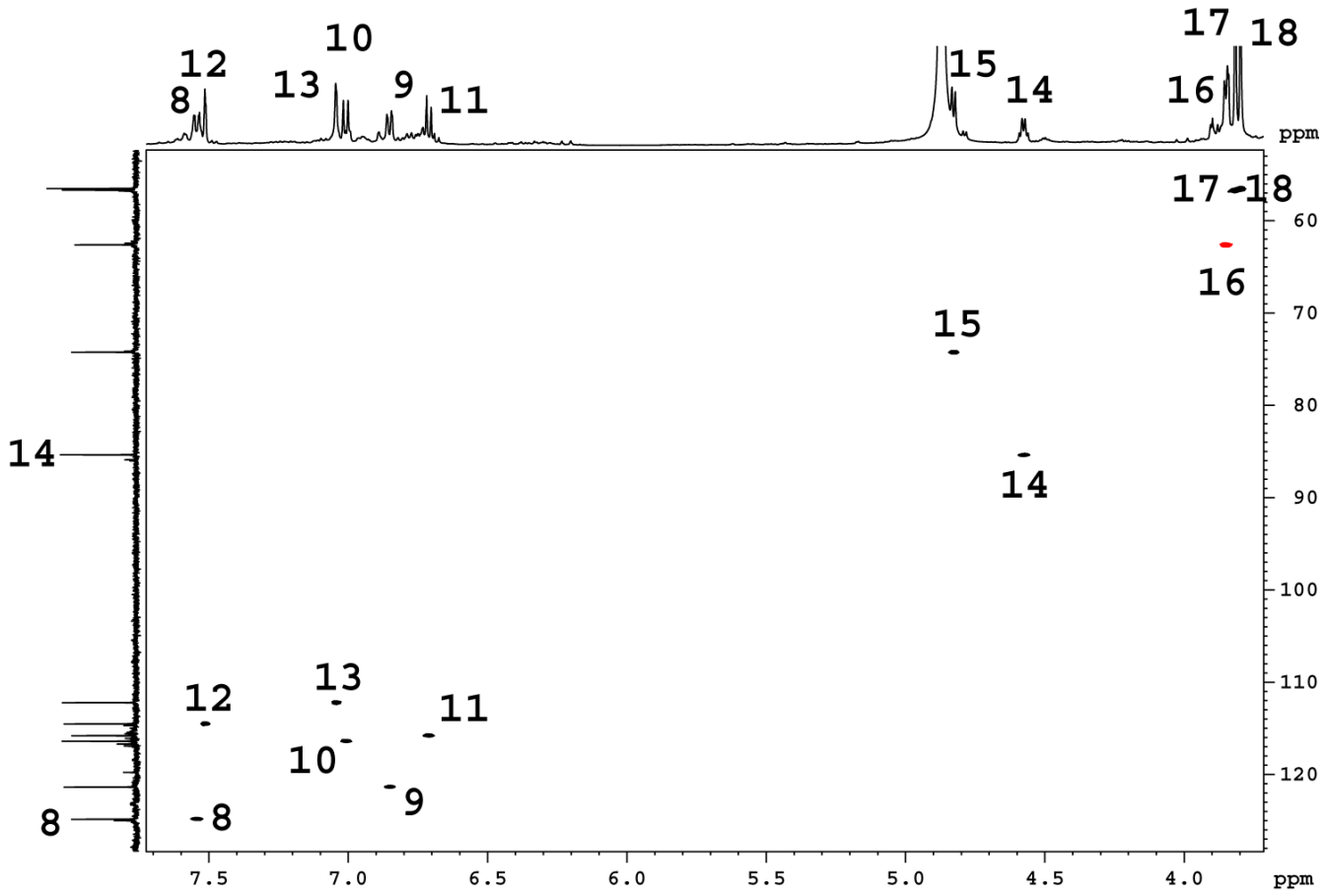
X_1— (?) —X_2

EXAMPLE, ^{13}C NMR SPECTRUM



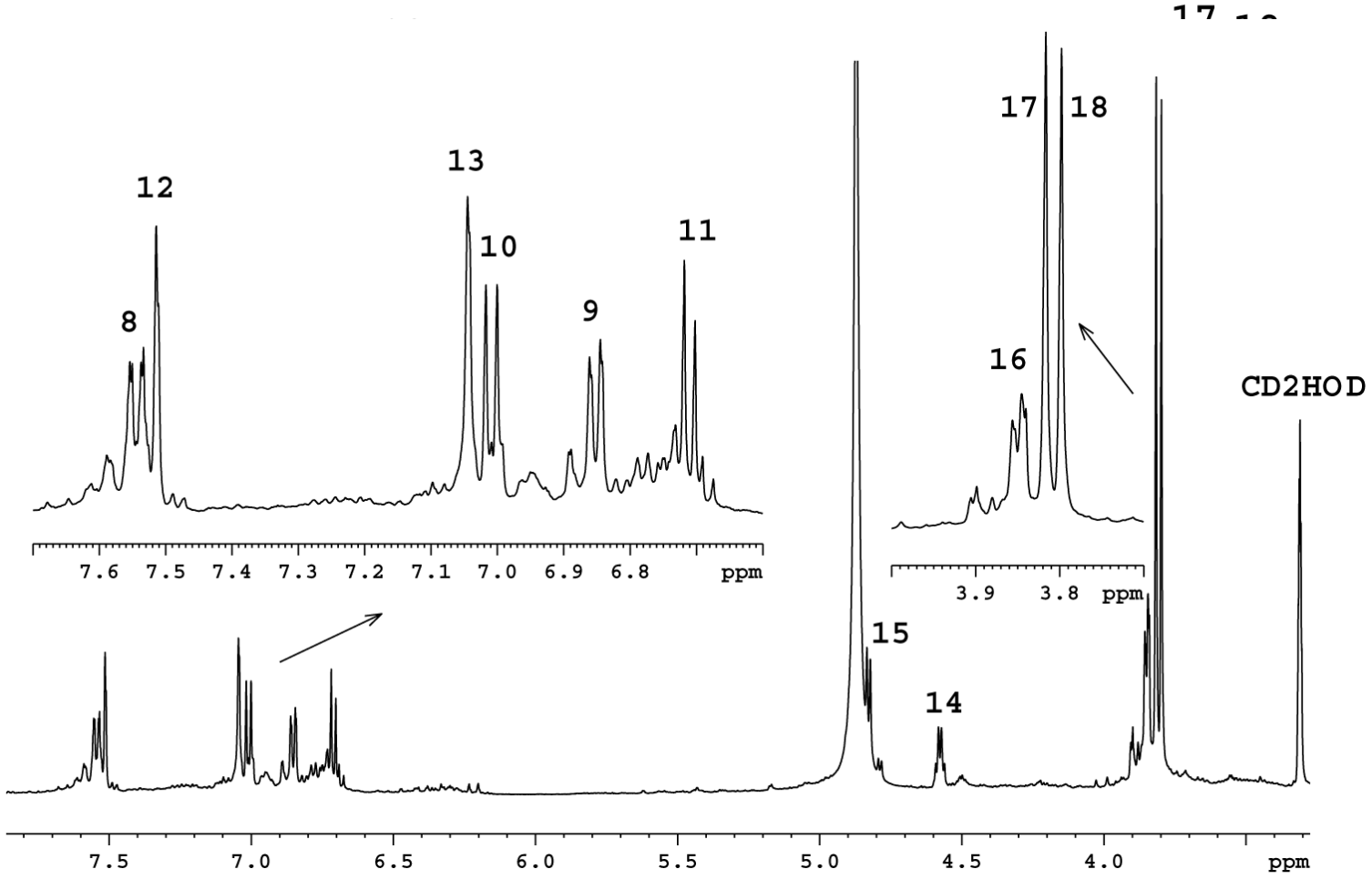
18 C, 1 C=O, 6 aromatic Q, 6 aromatic CH, 5 aliphatic C bound to an O atom

EXAMPLE, 2D HSQC SPECTRUM



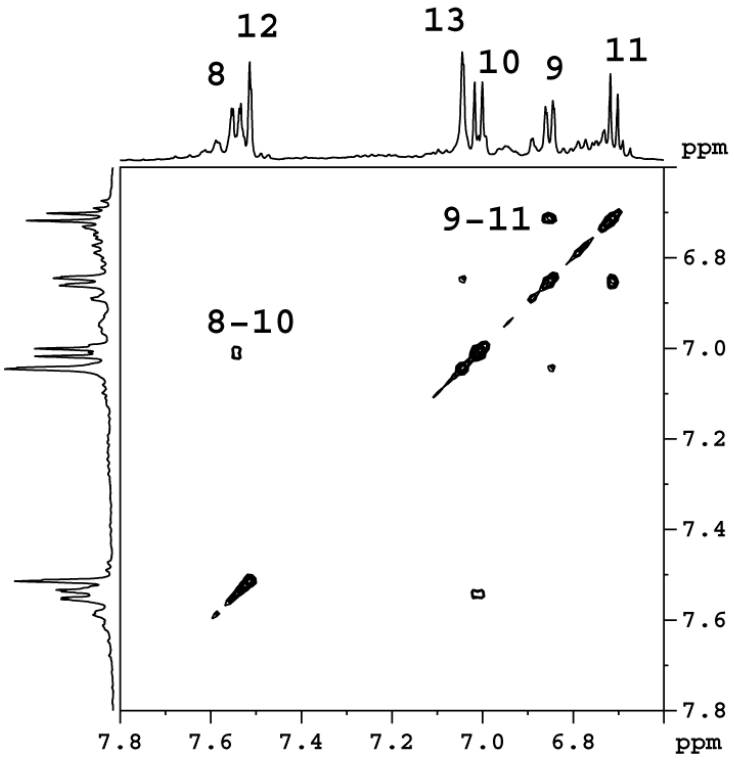
Aliphatic C: 2 CH-O, 1 CH_2-O, 2 CH_3-O-Ar

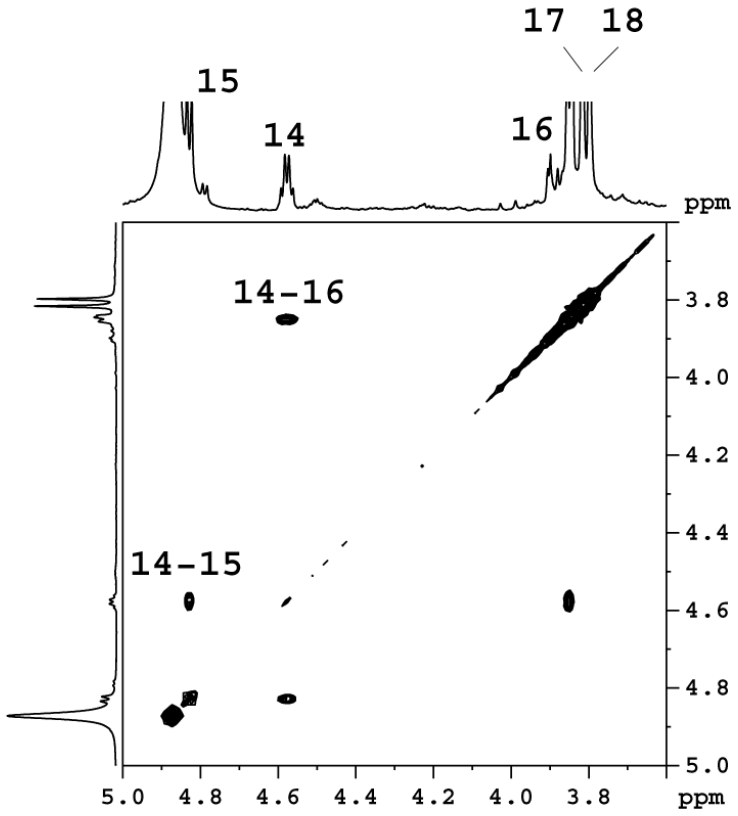
EXAMPLE, ^{1}H NMR SPECTRUM



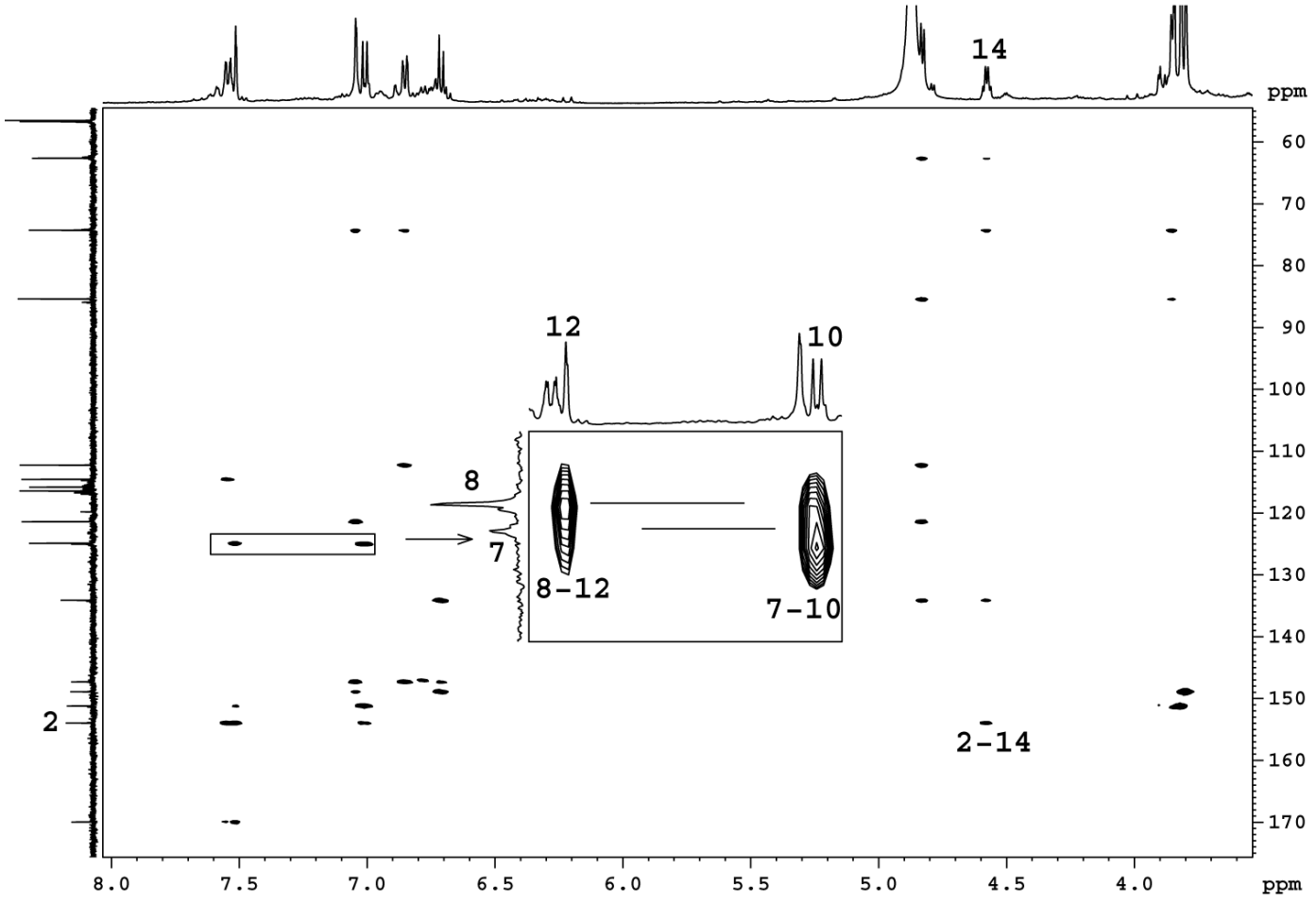
16 H atoms bound to C atoms

EXAMPLE, 2D COSY SPECTRUM





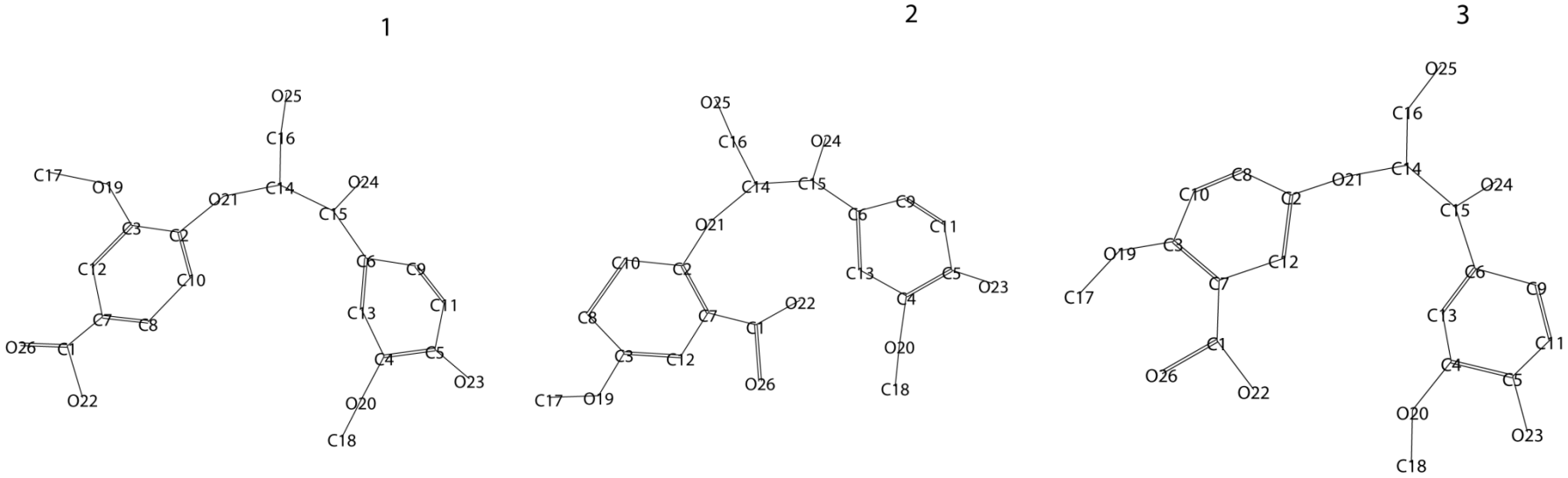
EXAMPLE, 2D HMBC SPECTRUM



EXAMPLE, ELEMENTAL FORMULA

- $[M + Na]^+ = 387$ u. Low resolution MS
- M = 364 u.
- 18 C + 16 H = 232 u.
- Remain: 364 – 232 = 132 u.
- 132 = 128 + 4 = 8 x 16 + 4
- 8 O + 4 H bound to O
- Resulting in $C_{18}H_{20}O_8$
- Not always so simple… Ask for HR-MS and MF

EXAMPLE, SOLUTIONS BY LSD



- Structures generated by **lsd**, 2D coordinates by **outlsd**, depictions by **genpos**. Drawing improvements by **m_edit**.
- **solve** chains all the process.
- Structure generation in less than 1 second on a PC.
- Too many solutions: ranking wanted

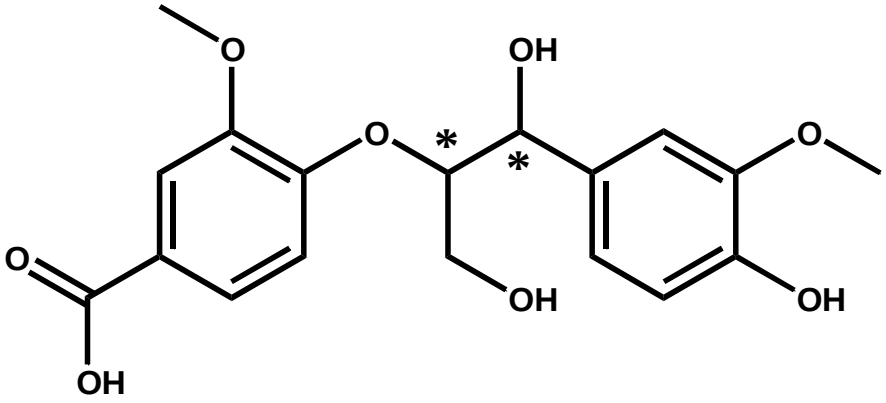
EXAMPLE, SOLUTION RANKING

- A score is assigned to each solution using a standalone version of the **nmrshiftdb2** ^{13}C NMR chemical shift predictor (S. Kuhn, C. Steinbeck)

- A score Δ mesures the distance between the experimental and predicted chemical shift series.

$$\Delta = \frac{\sum_{i=1}^{N} \left| \delta_i^{calc} - \delta_i^{exp} \right|}{N}$$

EXAMPLE, SOLUTION RANKING



- Ranking:
  - **1. Δ = 1.40 ppm**
  - 2. Δ = 2.03 ppm
  - 3. Δ = 2.13 ppm
- The ranking order may depend on the predictor
- The predictor depends on the database it makes use of and on the prediction algorithm
- So, never blindly trust a predictor.

IN PROGRESS…

- PyLSD
  - Software layer above LSD, written in Python
  - Removes some LSD limitations
  - Automates solution ranking

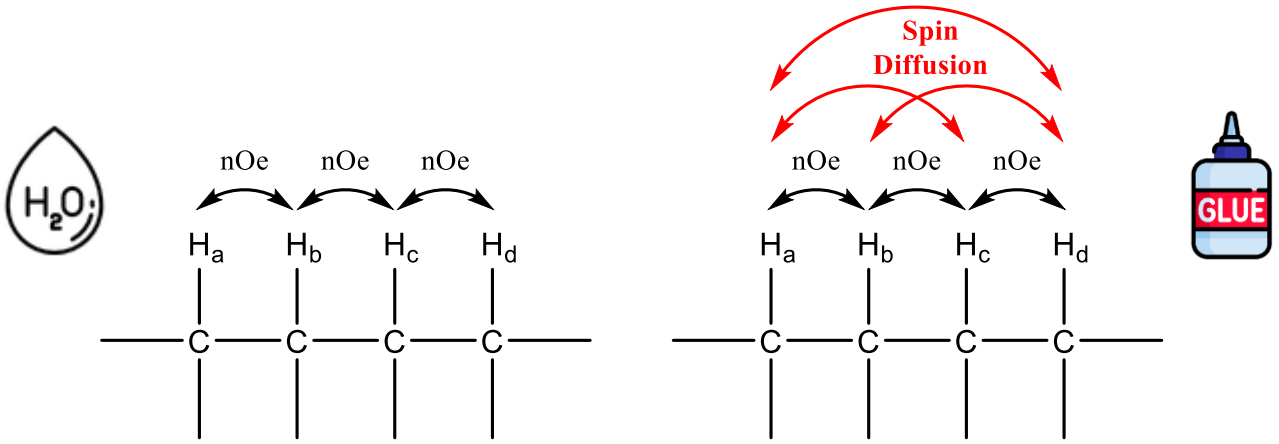
- Learning to use LSD
  - Not easy but possible.  I kow people who did it.
  - A collection of examples is in preparation and will be publicly available on GitHub

- Convert from NMReDATA to LSD
  - Check for structure elucidation exhaustivity before publication.
  - Next article about NMReDATA (S. Kuhn)

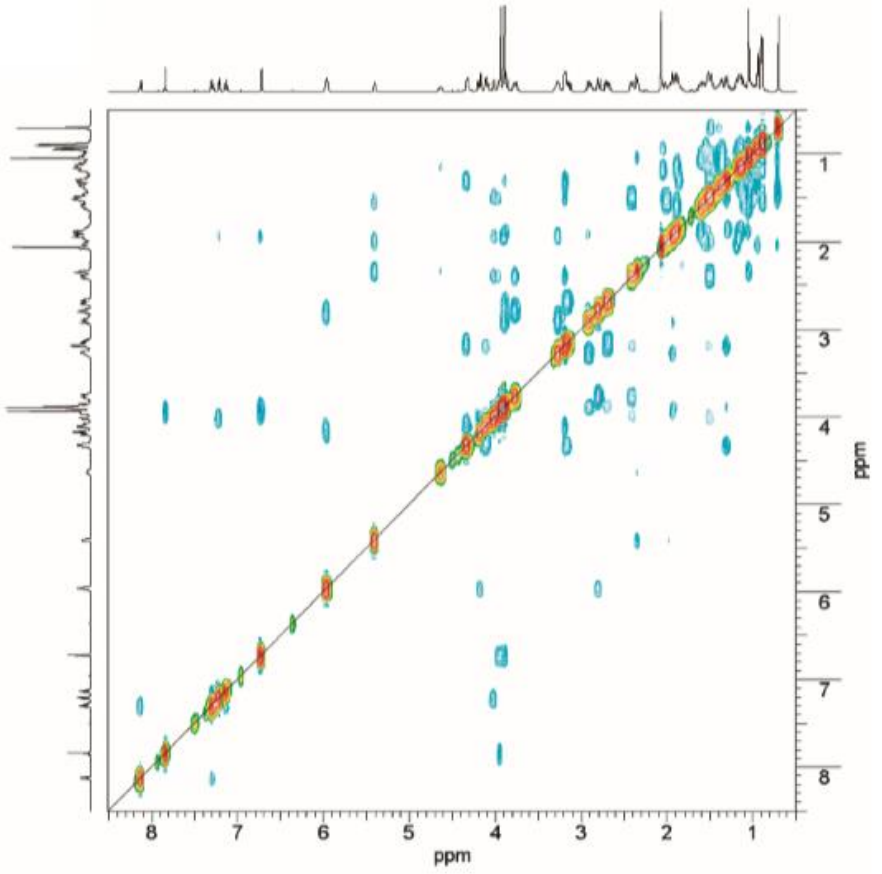
Mixture analysis using viscous solvents

- Mixture analysis by NMR
  - Compound identification without purification
  - Alternative to Diffusion-Ordered Spectroscopy (DOSY)

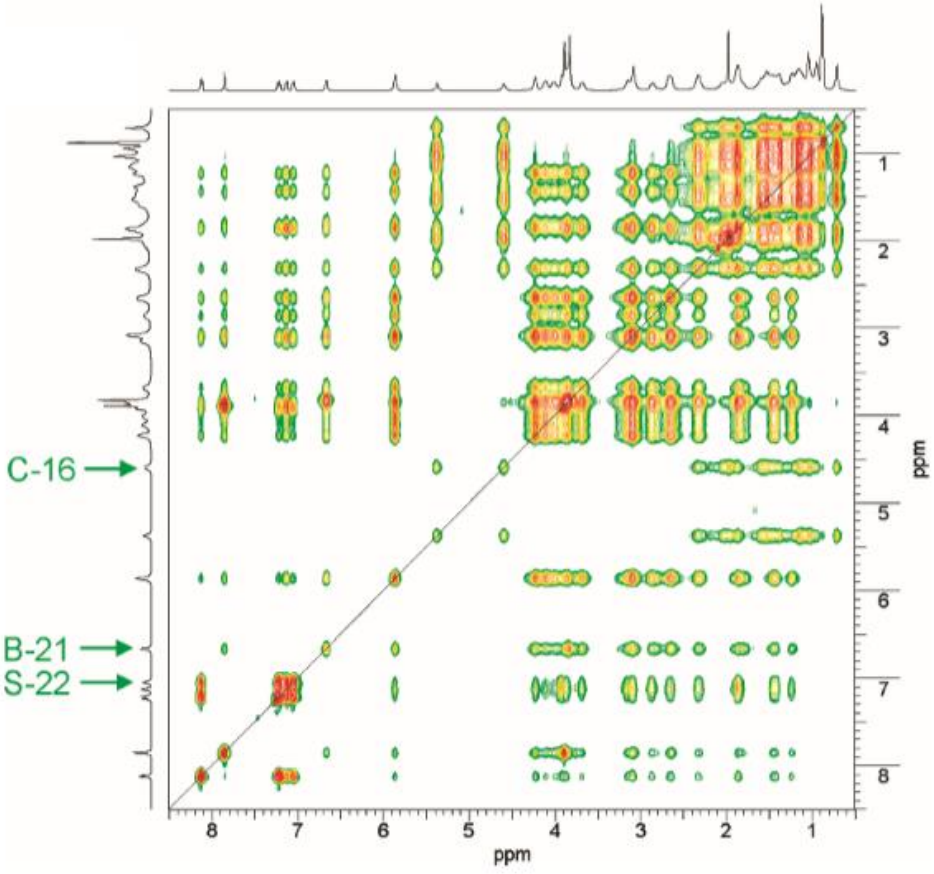
- Simpson *et al.*, *Anal. Chem.* **2008**, *80*, 186−194
  - 2D NOESY spectrum in spin diffusion conditions



NOESY: mixture of Strychnine (S), Brucine (B), and cholesteryl acetate (C).



In CDCl3



In a perfluorinated polymer

- Our contributions
  - Viscous solvents for polar compounds
    - Glycerol, glycerol carbonate
    - Water/glycerol blends
    - Water/DMSO blends

  - New NMR pulse sequences, other nuclei than ^{1}H
    - ^{13}C, ^{15}N, ^{19}F
  - See: *Chem. Eur. J* **2017**, *23*, 4923–4928.
  - Coming soon: a review article, new solvents, new methods.

ACKNOWLEDGEMENTS

- **Dr Bertrand Plainchont**
- **Dr Jane Hubert**

- **Pr Jean-Hugues Renault**

- **Romain Reynaud** (Soliance-Givaudan)