Identifier les metabolites secondaires

connus et inconnus par RMN

Jean-Marc Nuzillard

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

THE LAB

Institute of Molecular Chemistry in Reims

- Five research teams
- Among which is the Natural Product Chemistry team

- Plant chemistry (Pharmacognosy)
- New methods:
  - Chromatography
  - NMR

Mixture Analysis

- Chemicals from **plants**, renewable carbon sources
- Therapeutic drugs from **plants**
- Cosmetic ingredients from **plants**
- **But**: Plants rarely produce pure compounds

- **Mixture analysis plays en central role in plant chemistry**

Natural Products

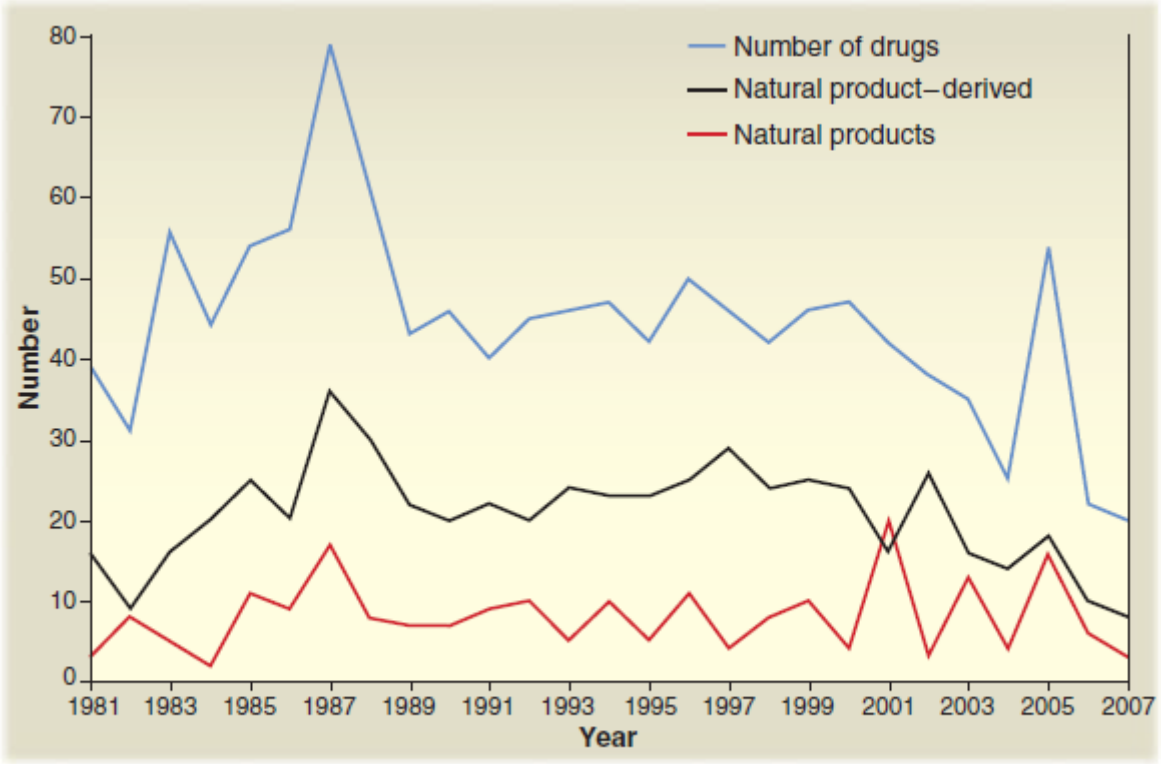


Fig. 1. Number of drugs approved in the United States from 1981 to 2007.

Mixture analysis in the context of cosmetic industry

Development of natural cosmetic ingredients (herbal, marine, microbial extracts)

Why ?

- Growing interest of consumers for natural cosmetics
- Natural products = Source of bioactive compounds
- High chemical diversity from highly diverse sources

Development of natural cosmetic ingredients (herbal, marine, microbial extracts)

NATURAL EXTRACTS

MIXTURES

Complex chemical profile

Development of natural cosmetic ingredients (herbal, marine, microbial extracts)

NATURAL EXTRACTS

MIXTURES
Complex chemical profile

Challenges

- Validation of their biological activity
- Safety (human & environment)
- REACH and precautionary principle
- European directives (76/768/EEC)
- US FDA cosmetics section

Development of natural cosmetic ingredients (herbal, marine, microbial extracts)

NATURAL EXTRACTS

MIXTURES
Complex chemical profile

Challenges

- Validation of their biological activity
- Safety (human & environment)
- REACH and precautionary principle
- European directives (76/768/EEC)
- US FDA cosmetics section

Need for efficient tools to determine the chemical composition of natural ingredients

Chemical profiling of natural cosmetic ingredients
Current strategies

NATURAL EXTRACTS

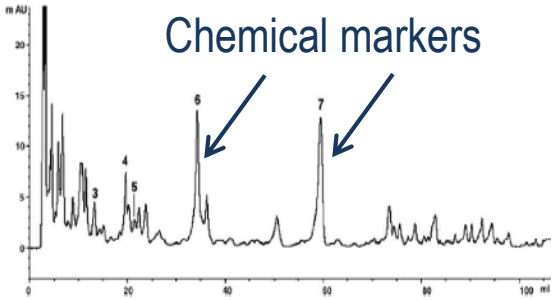
MIXTURES

Complex chemical profile

Chemical profiling of natural cosmetic ingredients
Current strategies

NATURAL EXTRACTS

MIXTURES
Complex chemical profile



Fingerprint analysis

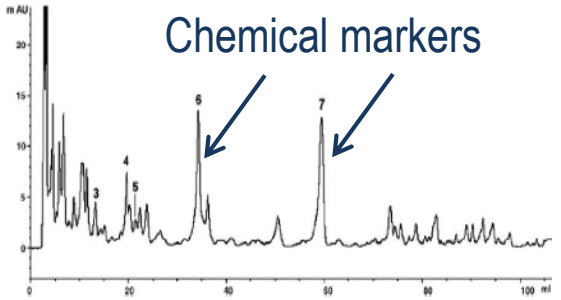
Chemical profiling of natural cosmetic ingredients
Current strategies

NATURAL EXTRACTS

MIXTURES
Complex chemical profile

PURIFICATION

Identification of individual & pure compounds



Fingerprint analysis

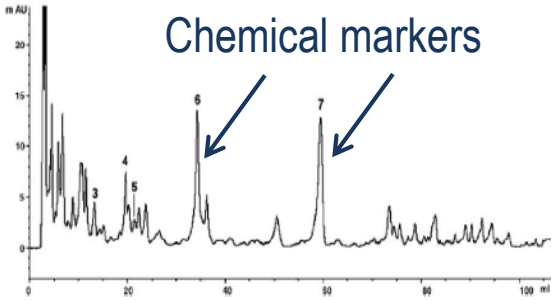
Chemical profiling of natural cosmetic ingredients
Current strategies

NATURAL EXTRACTS

MIXTURES
Complex chemical profile

PURIFICATION

Identification of individual & pure compounds



Many compounds remain unidentified

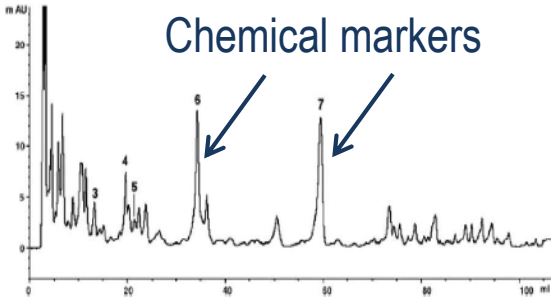
Chemical profiling of natural cosmetic ingredients
Current strategies

NATURAL EXTRACTS

MIXTURES
Complex chemical profile

PURIFICATION

Identification of individual & pure compounds



Many compounds remain unidentified

- **Time-consuming, solvents, materials**
- **Synergy disregarded**
- **Rediscovery of known compounds**

Identification of known compounds

IDENTIFICATION OF METABOLITES WITHIN MIXTURES
DEREPLICATION

analytical chemistry

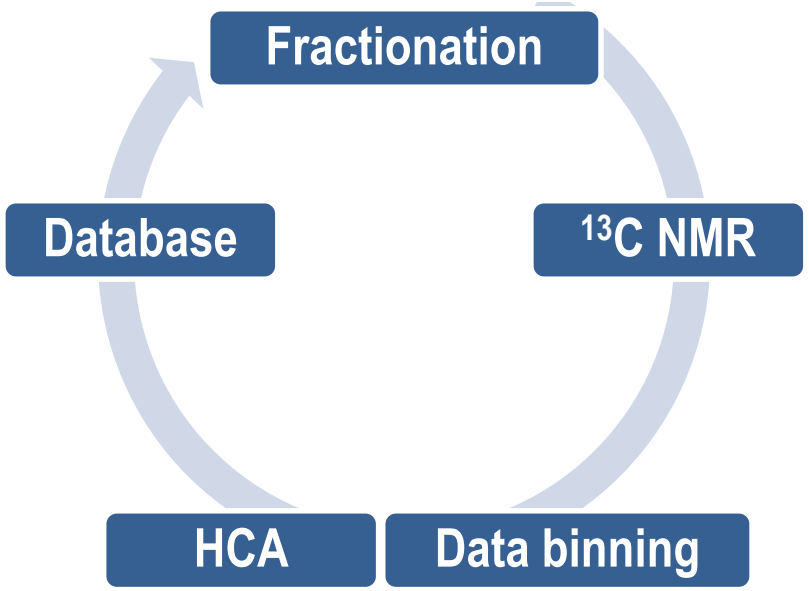
Article

pubs.acs.org/ac

Identification of Natural Metabolites in Mixture: A Pattern Recognition Strategy Based on ^{13}C NMR

Jane Hubert,*[,†] Jean-Marc Nuzillard,[†] Sylvain Purson,[†,‡] Mahmoud Hamzaoui,[§] Nicolas Borie,[†] Romain Reynaud,[‡] and Jean-Hugues Renault[†]

Anal. Chem. 2014, 86, 2955−2962



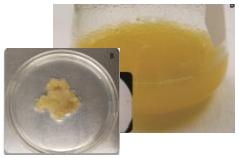
IDENTIFICATION OF METABOLITES WITHIN MIXTURES
DEREPLICATION

Fast identification of the main compounds in a single extract

- ⇒ **Obtain data about the chemical composition of natural ingredients**
- ⇒ **Investigate which compounds are involved in a biological activity**
- ⇒ **Check for the absence of undesirable (toxic) constituents**

STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION

Vitis labrusca



Cell suspension culture developed from calli

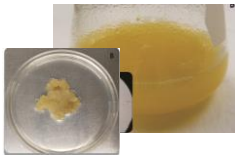
Bioreactor 14 L

Elicitation (10 d)

STILBENOID BIOSYNTHESIS

STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION

Vitis labrusca

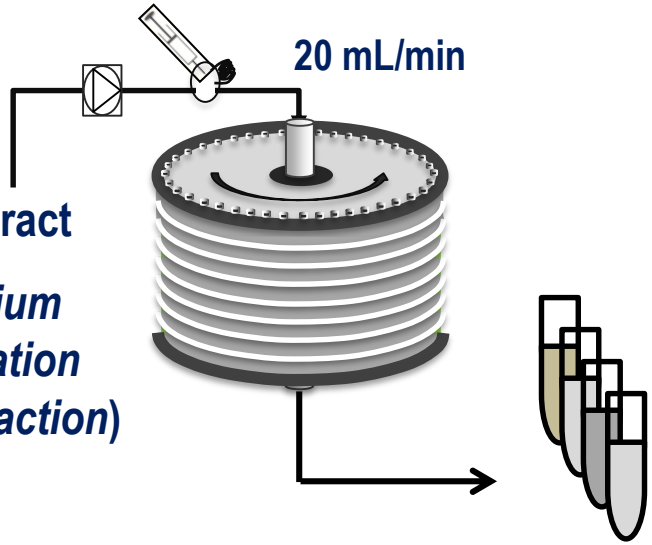


Cell suspension culture developed from calli

Bioreactor 14 L

Elicitation (10 d)

STILBENOID BIOSYNTHESIS



3 g crude extract

(culture medium after cell filtration and EtOAc extraction)

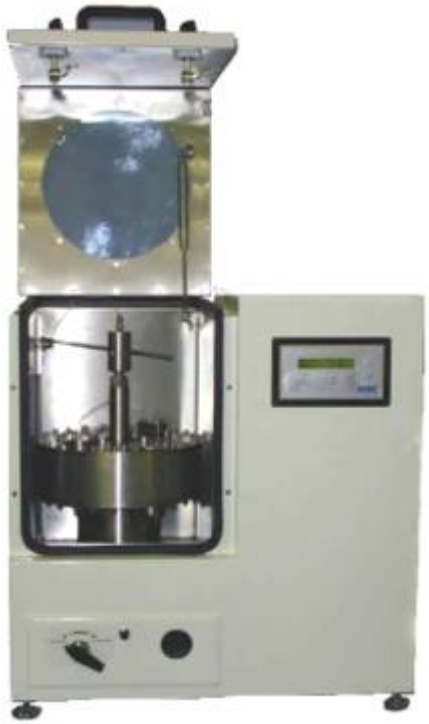
Fractionation by Centrifugal Partition Extraction: CPE

Fraction series

CPE / CPC

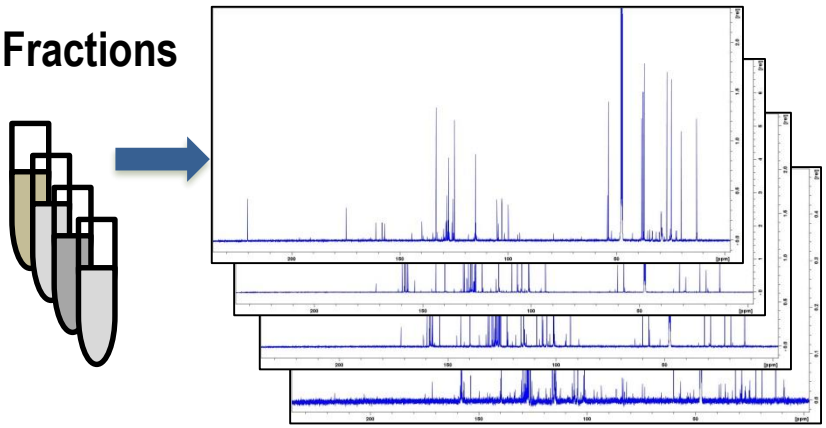
- **C**entrifugal **P**artition **E**xtraction/**C**hromatography
- Based on partition of analytes between two liquid phases
- The “column” is made of hundreds of engraved partition cells
- The stationary phase is kept in place by a centrifugal force field
- The analytes are injected on column head
- The mobile phase percolates through the stationary phase
- No irreversible adsorption on a solid phase
- Elution, graduated elution, displacement mode chromatography
- All you put inside will come out, in one or the other way
- High flow rates, typically 20 mL/mn
- Typically inject 5g in a 200 mL column

- **Preparative technique**



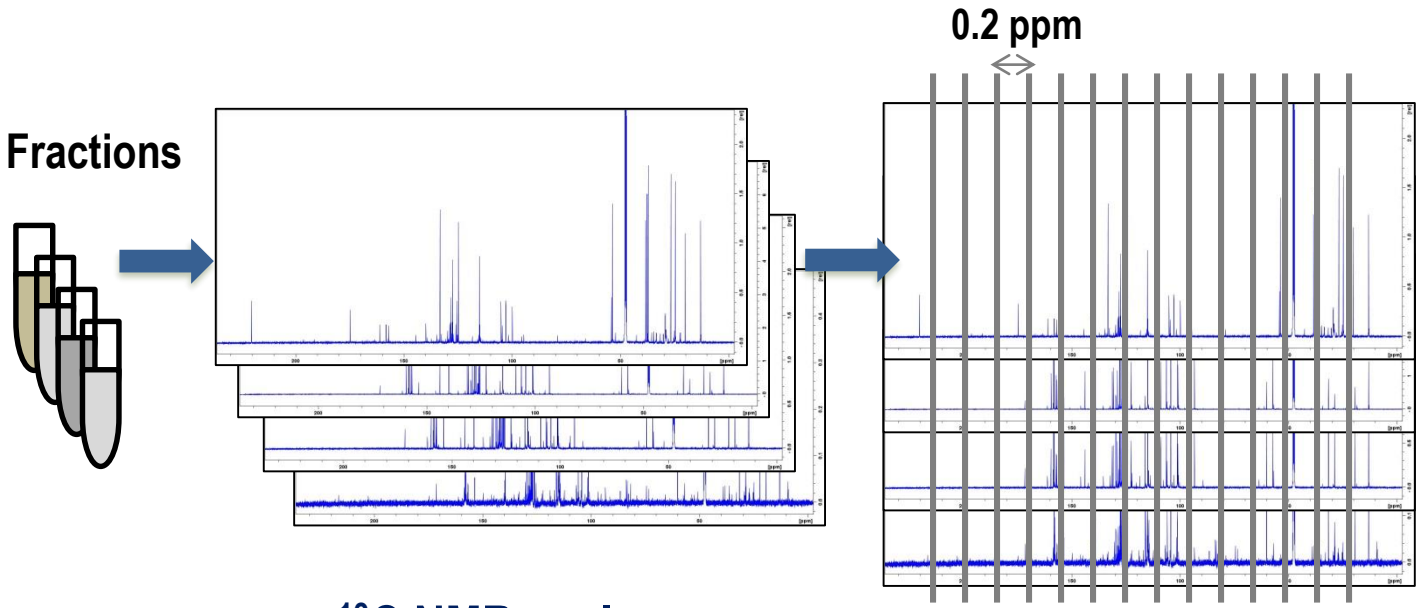
STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION

Fractions



^{13}C NMR analyses

STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION



^{13}C NMR analyses

150 MHz (600 MHz for ^{1}H)
Cryoprobe

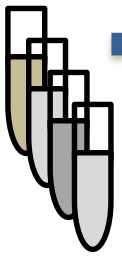
- Automatic peak picking
- File conversion
- Alignment by binning

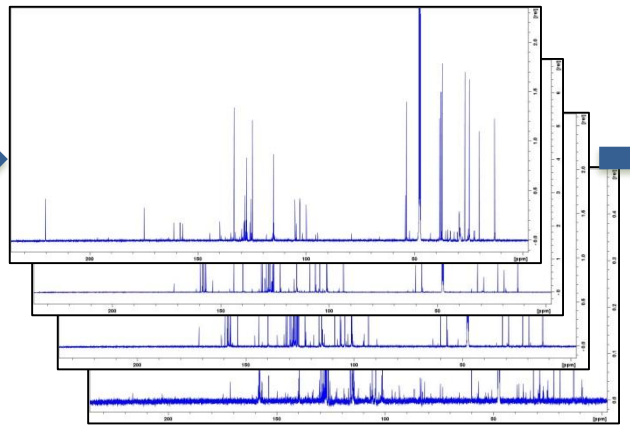
^{13}C NMR

- One carbon, one peak (symmetry disregarded)
- Minimized probability of peak superimposition
- Presumably bad sensitivity
- 600 MHz, cryoprobe, cooled ^{13}C coil
- ^{1}H-detected NMR: you do not always have enough H to observe!
- Other alternatives:
  - Broad-band decoupled ^{1}H 1D NMR (difficult…)
  - HSQC
  - HMBC

STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION

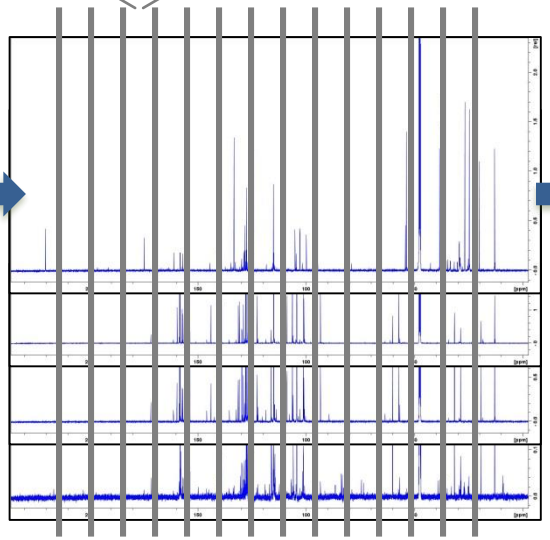
Fractions





^{13}C NMR analyses

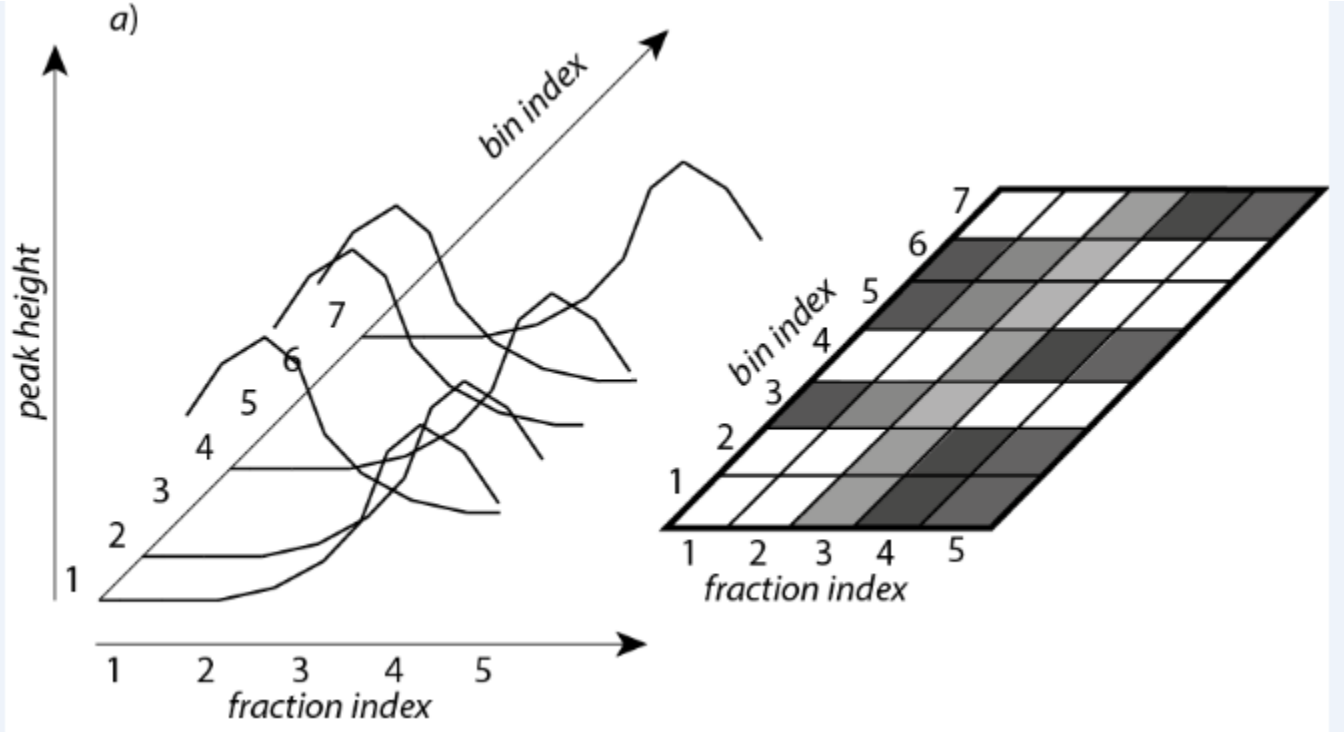
0.2 ppm



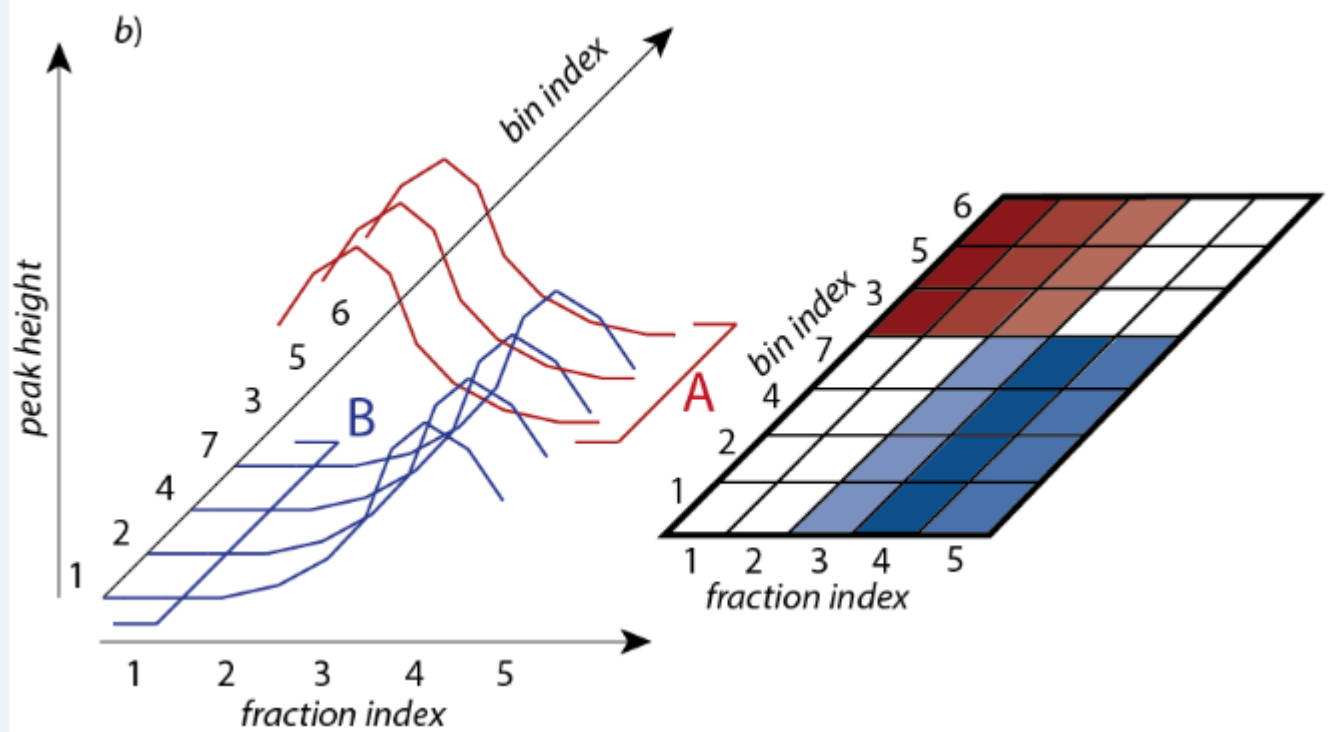
ppm	f1	f2	...	fx
16.3	2E+08	1E+08	...	0
17.5	0	0	...	0
18.7	1E+08	1E+08	...	0
⋮	⋮		NMR signal intensities	
176.1	3E+07	0	...	0
177.7	0	0	...	4E+07
177.9	6E+07	6E+07	...	0
199.5	7E+07	5E+07	...	0

- Automatic peak picking
- File conversion
- Alignment

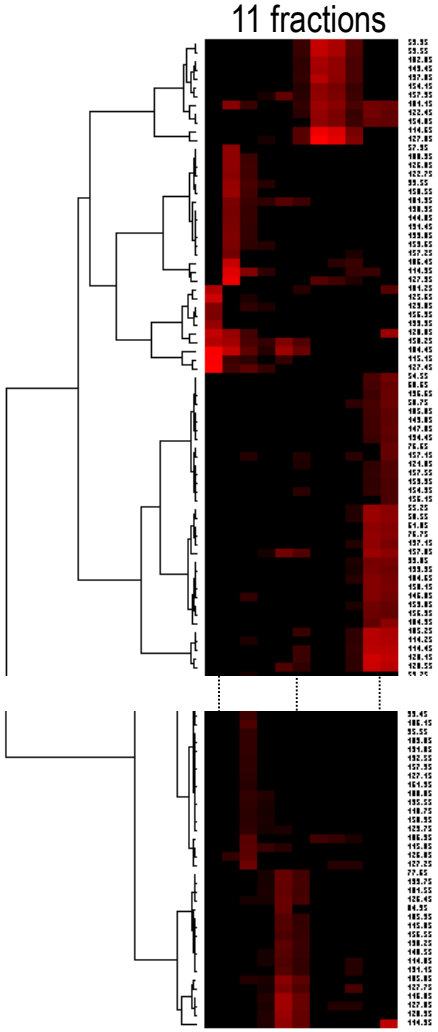
STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION



STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION



STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION

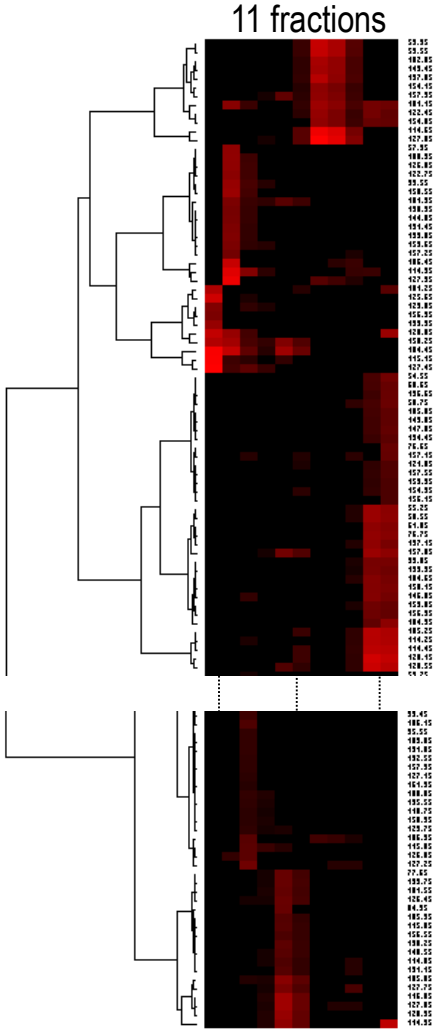


Hierarchical Clustering Analysis of ^{13}C NMR signals

- Clusters of ^{13}C NMR chemical shifts
- Similarity measurement (Euclidean distance)

- **Chemical shift values with similar chromatographic « history » possibly report about the same compound.**

STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION

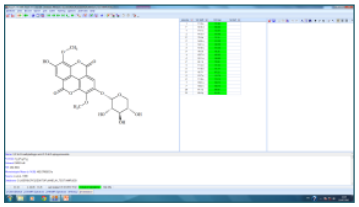


Hierarchical Clustering Analysis of ^{13}C NMR signals

- Clusters of ^{13}C NMR chemical shifts
- Similarity measurement (Euclidean distance)
- Visualization of Carbon skeletons

^{13}C NMR database – Natural metabolites

- Structures + ^{13}C ppm (n ≈ 2200)
- Experimental data, literature or spectra prediction



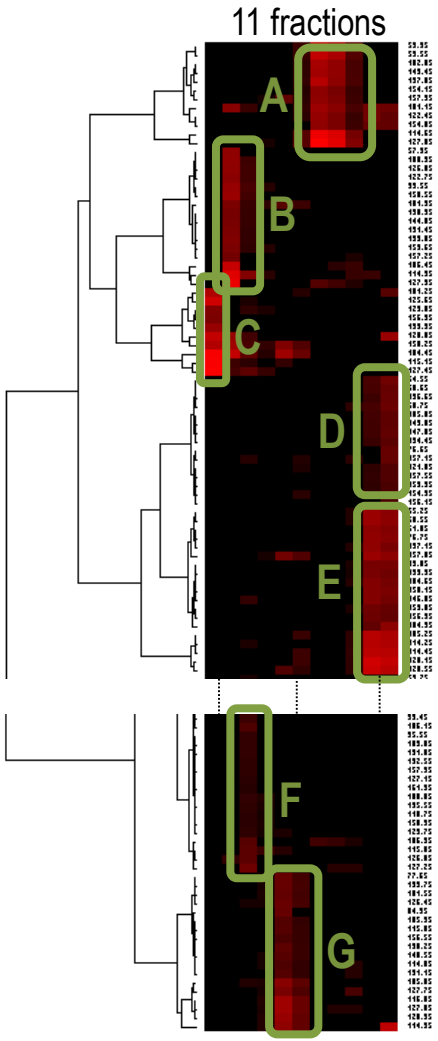
ACD/Labs

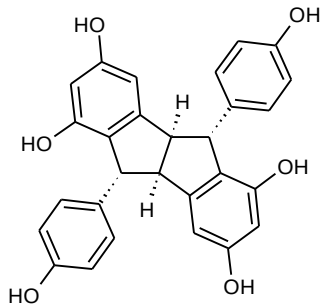
HIERARCHICAL CLUSTEING ANALYSIS

- Carried out by the Permutmatrix software
- Free sotware
- It rearranges lines (or columns) of a matrix in order to group together the most similar (correlated) lines (or columns)
- Result shows up as “heat map” and “hierarchy tree”

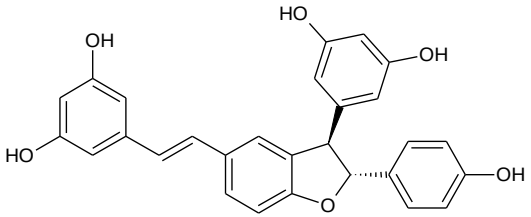
- http://www.atgc-montpellier.fr/permutmatrix/

STILBENOIDS BIOPRODUCED BY GRAPEVINE CELL ELICITATION

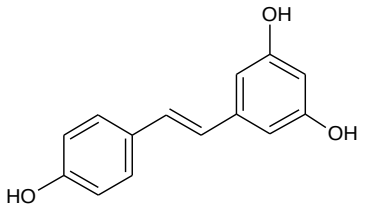




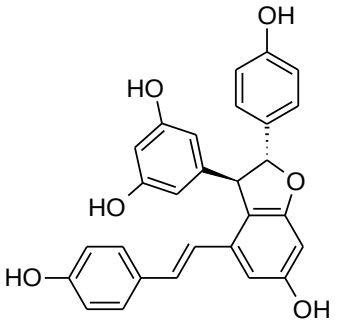
A: Pallidol



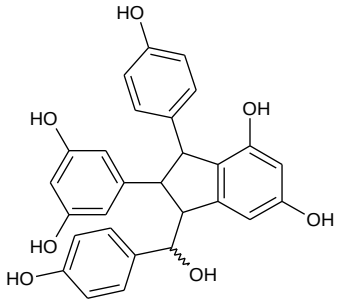
B: δ-viniferin



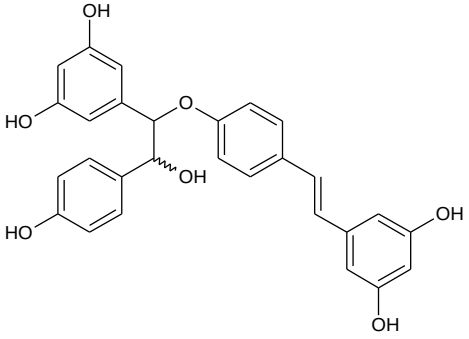
C: *trans*-resveratrol



F: ε-viniferin



D, E: Leachianol isomers



G: **New dimer**

BRUTE FORCE APPROACH, WITHOUT FRACTIONATION

Peumus boldus

Leaf extract

JOURNAL OF NATURAL PRODUCTS

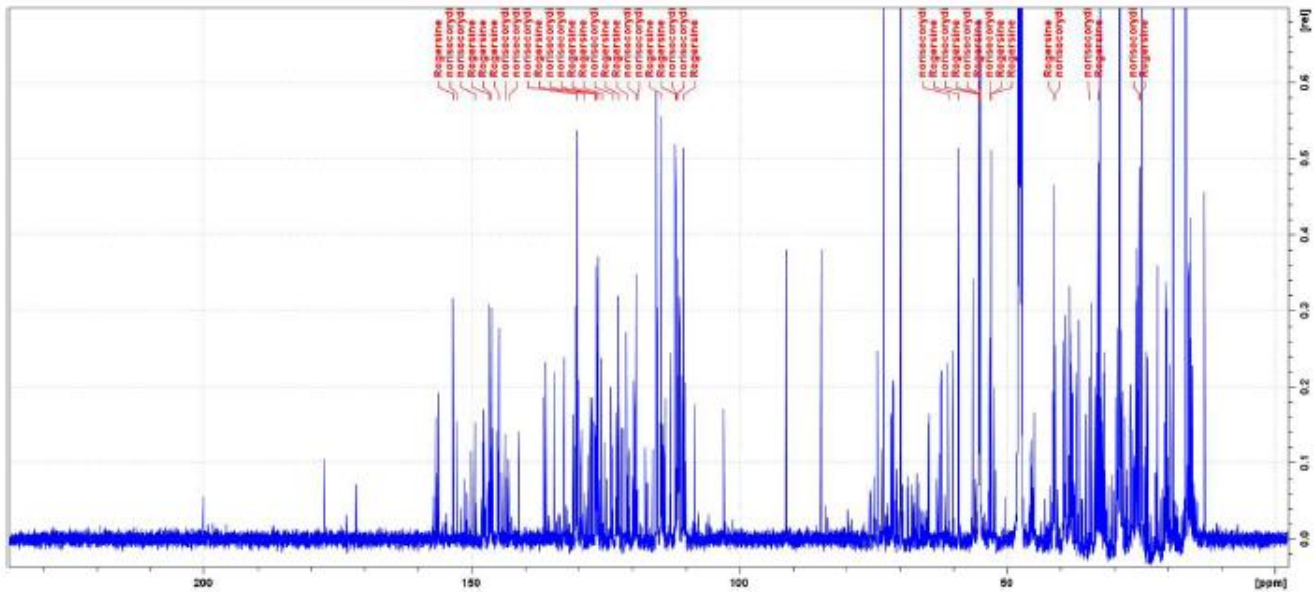
Article

pubs.acs.org/jnp

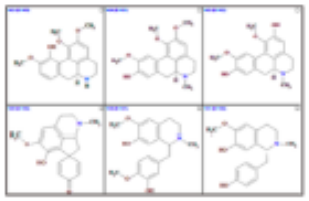
Computer-Aided ^{13}C NMR Chemical Profiling of Crude Natural Extracts without Fractionation

Ali Bakiri,[†,§] Jane Hubert,[*,†] Romain Reynaud,[§] Sylvie Lanthony,[†] Dominique Harakat,[†] Jean-Hugues Renault,[†] and Jean-Marc Nuzillard[†]

J. Nat. Prod., **2017**, *80* (5), pp 1387–1396



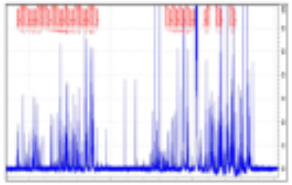
BRUTE FORCE APPROACH, WITHOUT FRACTIONATION



Genus under examination
Known metabolites

1

Spectral database
- Structures drawing
- ^{13}C NMR prediction



Crude extract ^{13}C NMR analysis

2

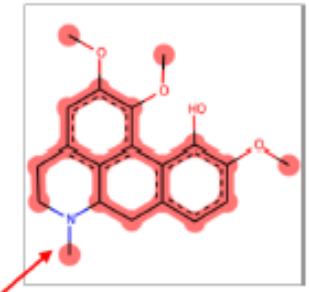
No.	(ppm)	Intensity
1	18.25	0.0011
2	21.48	0.0008
3	22.33	0.0008
4	23.07	0.0009
5	24.13	0.0010
6	27.13	0.0015
7	29.07	0.0009
8	29.34	0.0033
9	29.79	0.0007
10	38.03	0.0010
11	55.00	0.0029

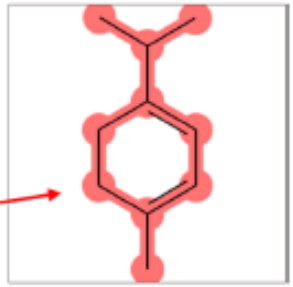
Peak list to text file

3

Searching Algorithm

Linalol	1.0
alpha-terpineol	1.0
alpha-pinene	1.0
Terpinen-4-ol	1.0
p-cimene **	1.0
caryophyllene	1.0
Eucalyptol **	1.0
beta-Pinene	1.0
Phellandrene	1.0
3-carene	1.0
Thymol	1.0
Isocorydine **	1.0
Norisocorydine **	1.0
Rogersine **	1.0
Boldine **	1.0
Reticuline **	1.0
N-methylcoclaurine **	1.0
Coclaurine **	1.0
alpha-terpinene **	1.0
2-carene	1.0
alpha-thujene	1.0
Dehydro-1,8-cineole	1.0
Sabinene hydrate	1.0
trans-*p*-menth-2-en-ol	1.0
Ascaridole	1.0
beta-elemene	1.0
beta-caryophyllene oxyde	1.0
Laurotetanine **	1.0
Pronuciferine	0.947
alpha-copaene	0.933
beta-oploplenone	0.933
delta-cadinene	0.928
Bornyl acetate	0.916





Input data

Output data (list of results)

SUMMARY, FOR DEREPLICATION

STRENGTHS OF THE STRATEGY

- Identification of natural compounds in mixture
- Rapid chemical profiling of crude extracts
  (≈ 80% w/w of the material is characterized)
- Substantial time, solvents and cost savings

Current uses

- Strategy optimization on a range of natural cosmetic ingredients (plants, marine…)
- Application within past and ongoing academic European projects (NatProtec, Microsmetics)

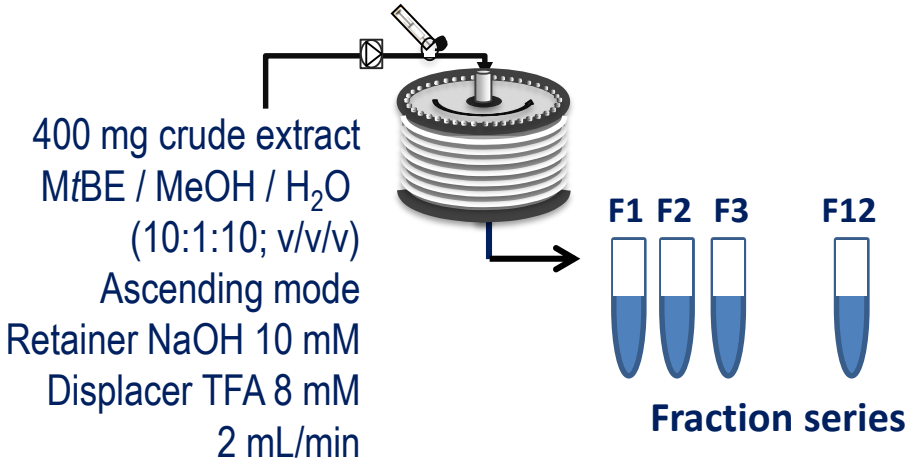
Still a lot of work to be done …

- Identification of minor compounds
- What about unknown metabolites?
  ⇒ MS, 1D & 2D NMR
  ⇒ Computer tools

UNKNOWN METABOLITES

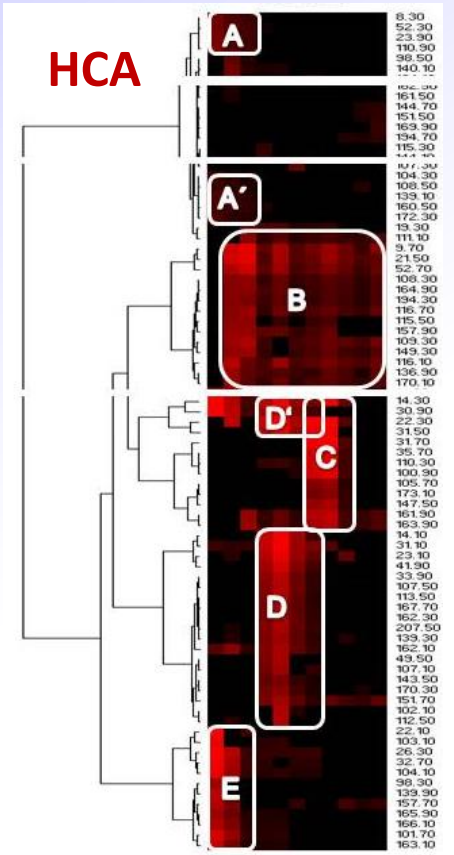
Pseudevernia furfuracea

Dereplication, as usual…



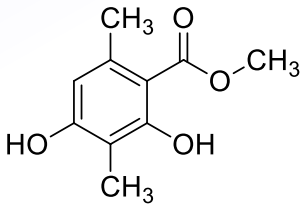
UNKNOWN METABOLITES

Successive fractions

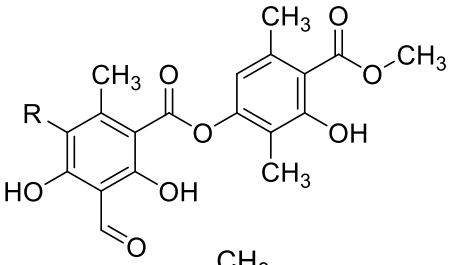


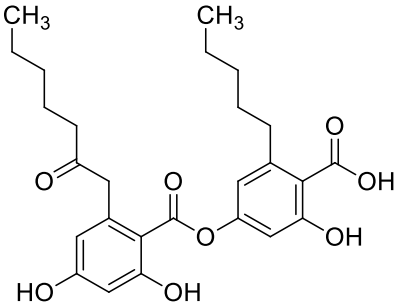
Metabolite identification + UNKNOWNS

Cluster A + A´
Methyl β-orcinolcarboxylate

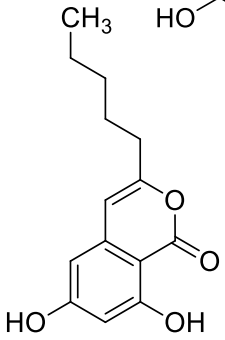


Cluster B
- Atranorin (R=H)
- 5-Chloroatranorin (R=Cl)

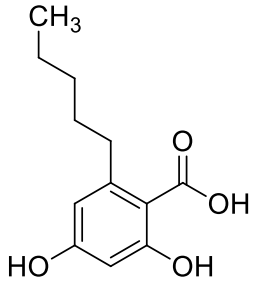




Cluster D + D´
Olivetoric acid



Cluster E
Olivetonide



Cluster C
Olivetolcarboxylic acid

Identification of unknown compounds

DE NOVO STRUCTURE DETERMINATION WORKFLOW

- 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
- Example: Logic for Structure Determination (LSD)

CASE: COMPUTER-ASSISTED STRUCTURE ELUCIDATION

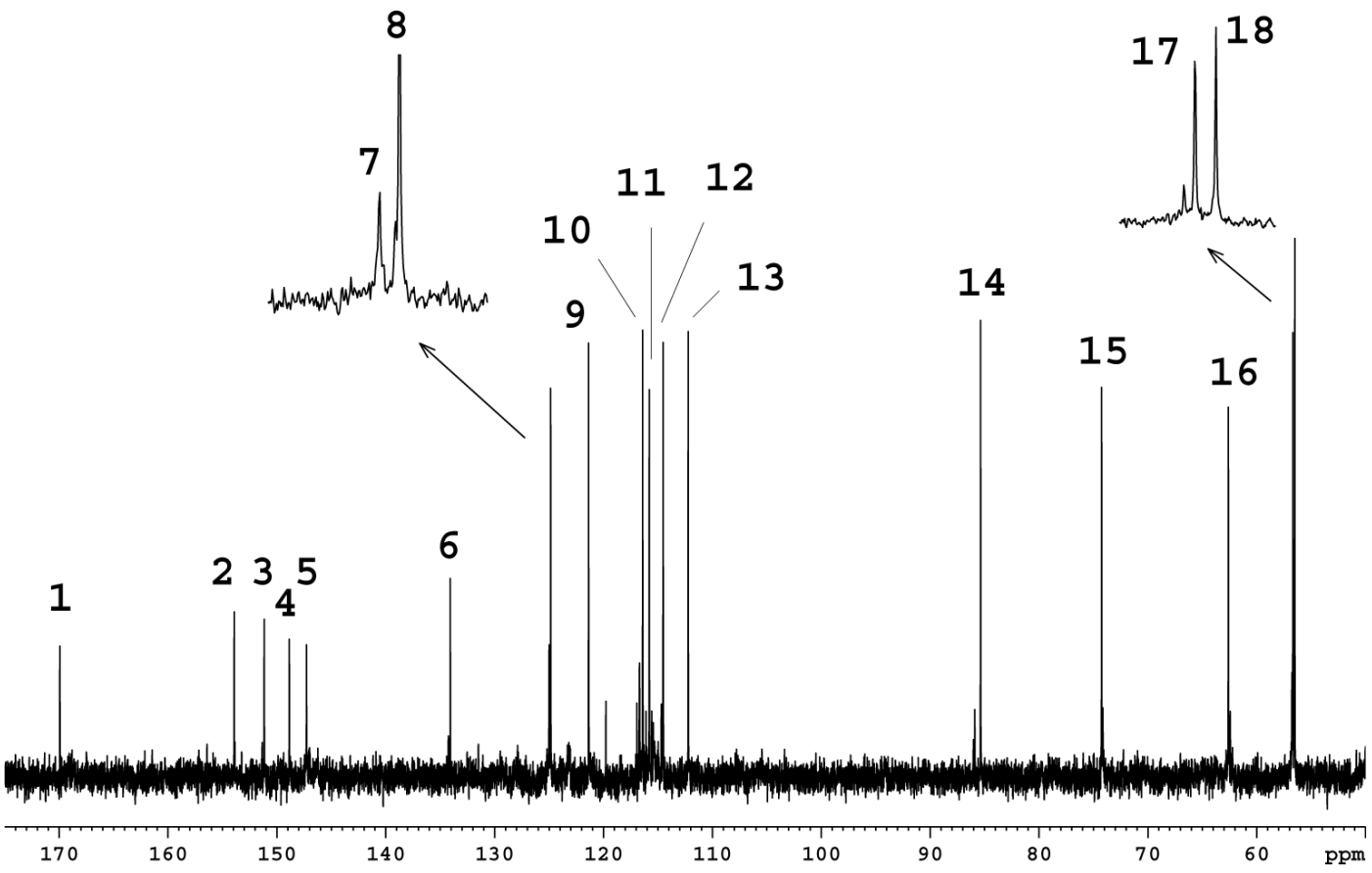
- Academic Software
  - LSD (France)
  - COCON (Germany)
  - SENECA (UK)

- Commercial software
  - ACD Structure Elucidator
  - Bruker CMC-se
  - MestreNova
  - ScienceSoft AssembleIt (?)

USING LSD

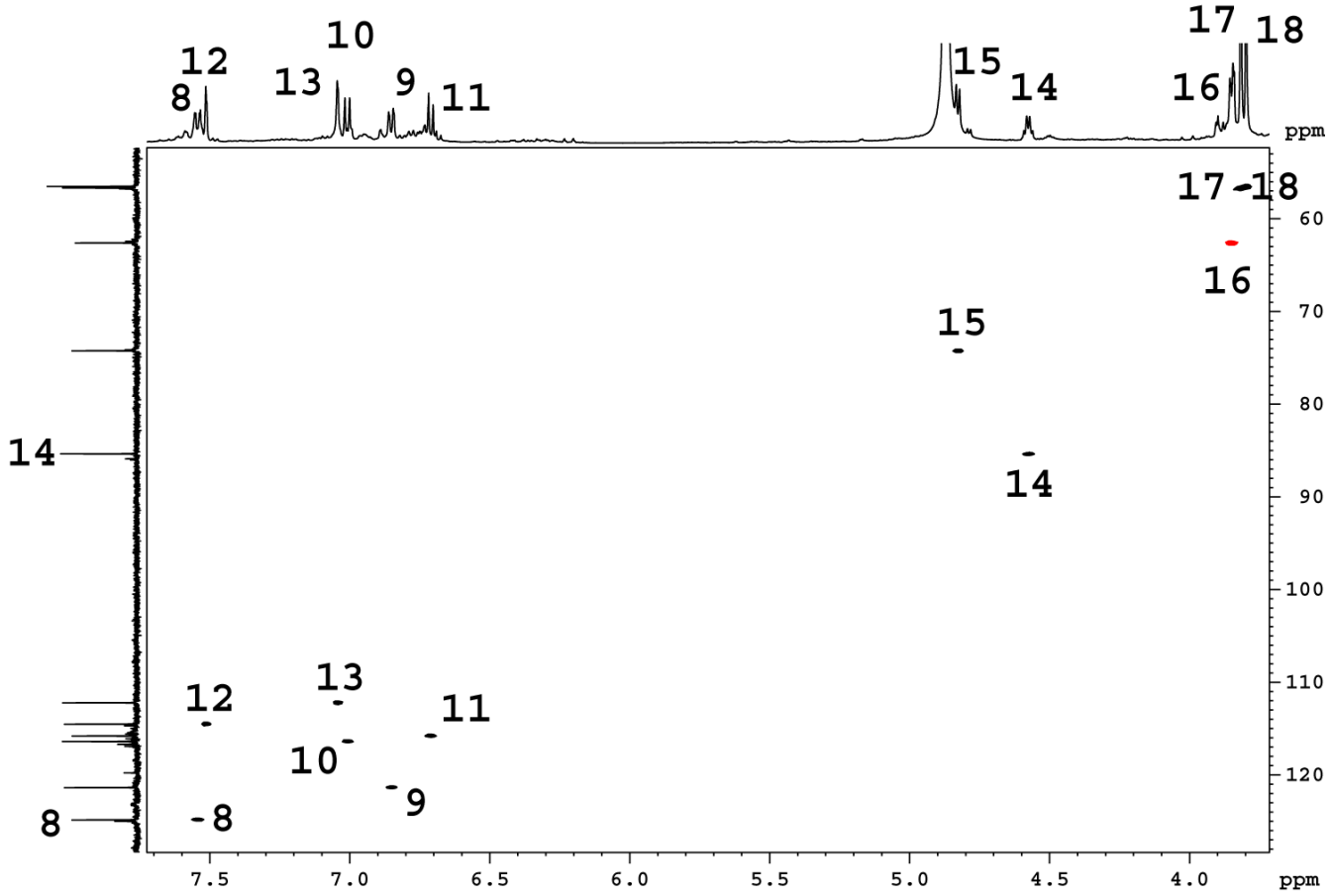
- Assign a numerical index to each non-H atom
  - Any non-H atom is « heavy »
  - For C atoms, assign indexes in the decreasing order of chemical shifts
- Assign a numerical index to H atoms
  - A H bound to a C receives the index of the C atom
- Translate correlations in 2D spectra into pairs of numbers
- Impose neighborhood constraints to atoms
  - From chemical shift values or coupling patterns
- Impose substructural constraints, if any
- Provide software execution control options, if necessary

EXAMPLE, ^{13}C NMR SPECTRUM



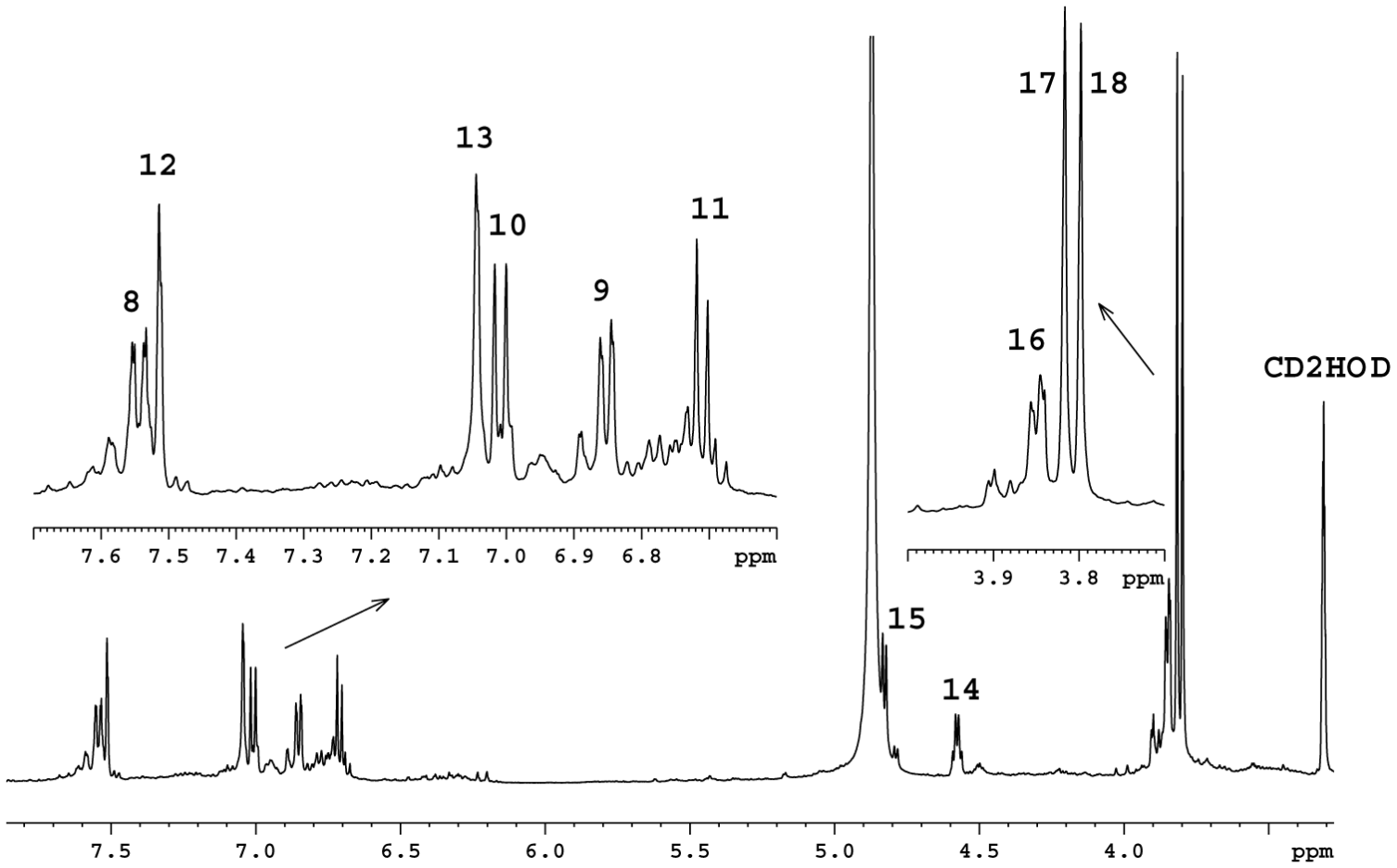
18 C, 1 C=O, 6 aromatic Q, 6 aromatic CH, 5 aliphatic C bound to an O atom

EXAMPLE, 2D HSQC SPECTRUM



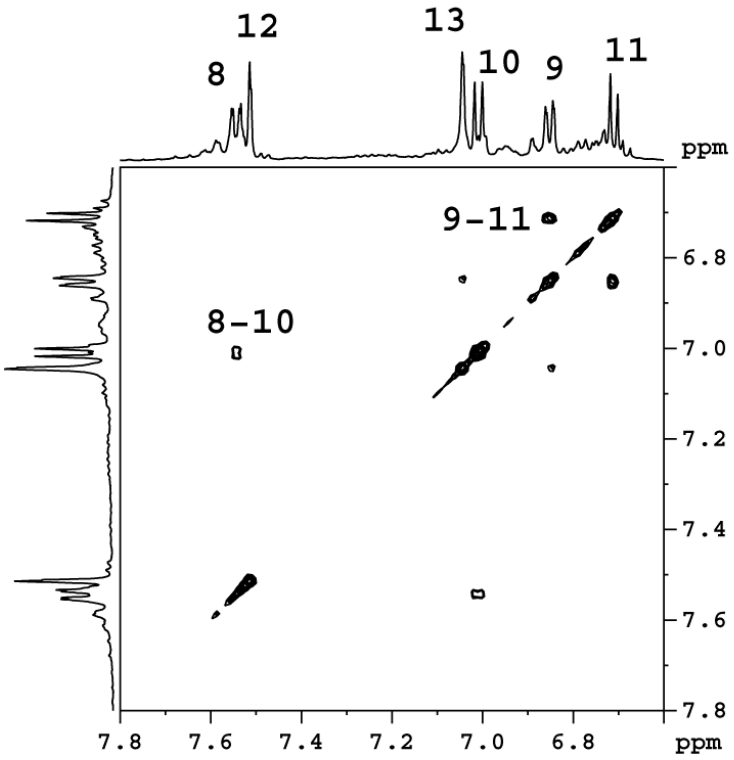
Aliphatic C: 2 CH-O, 1 CH_2-O, 2 CH_3-O-Ar

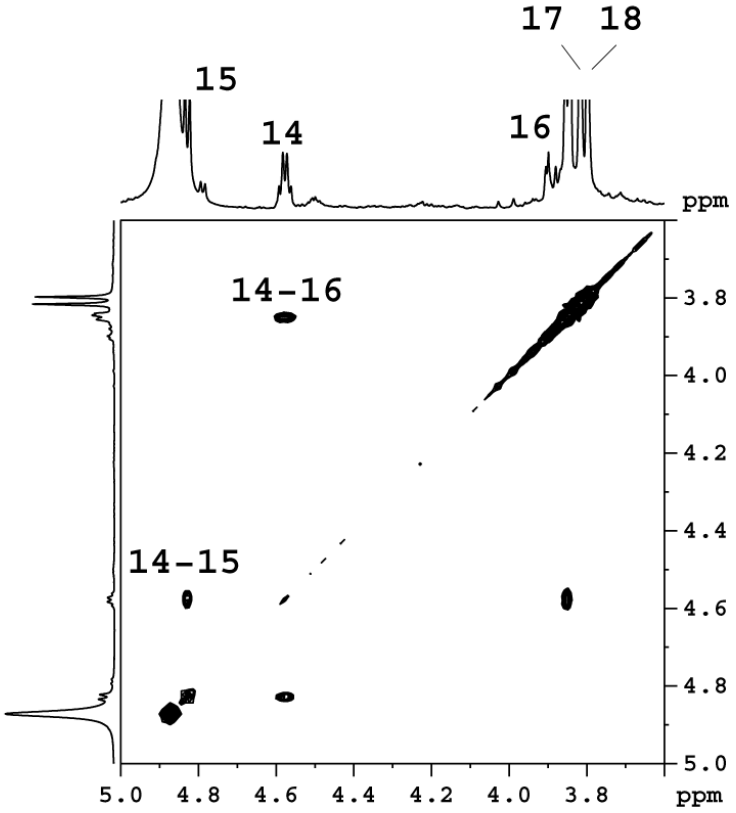
EXAMPLE, ^{1}H NMR SPECTRUM



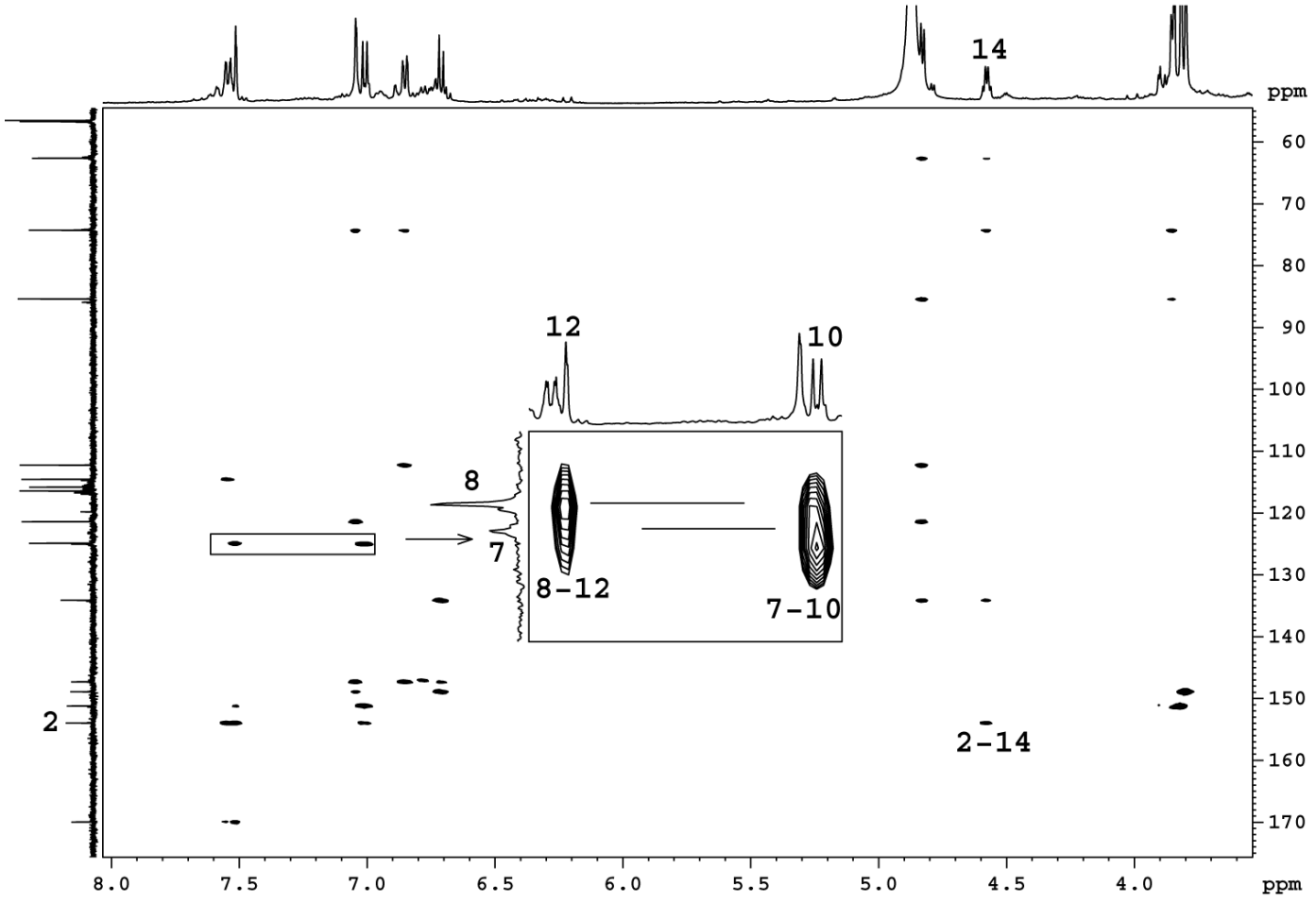
16 H atoms bound to C atoms

EXAMPLE, 2D COSY SPECTRUM

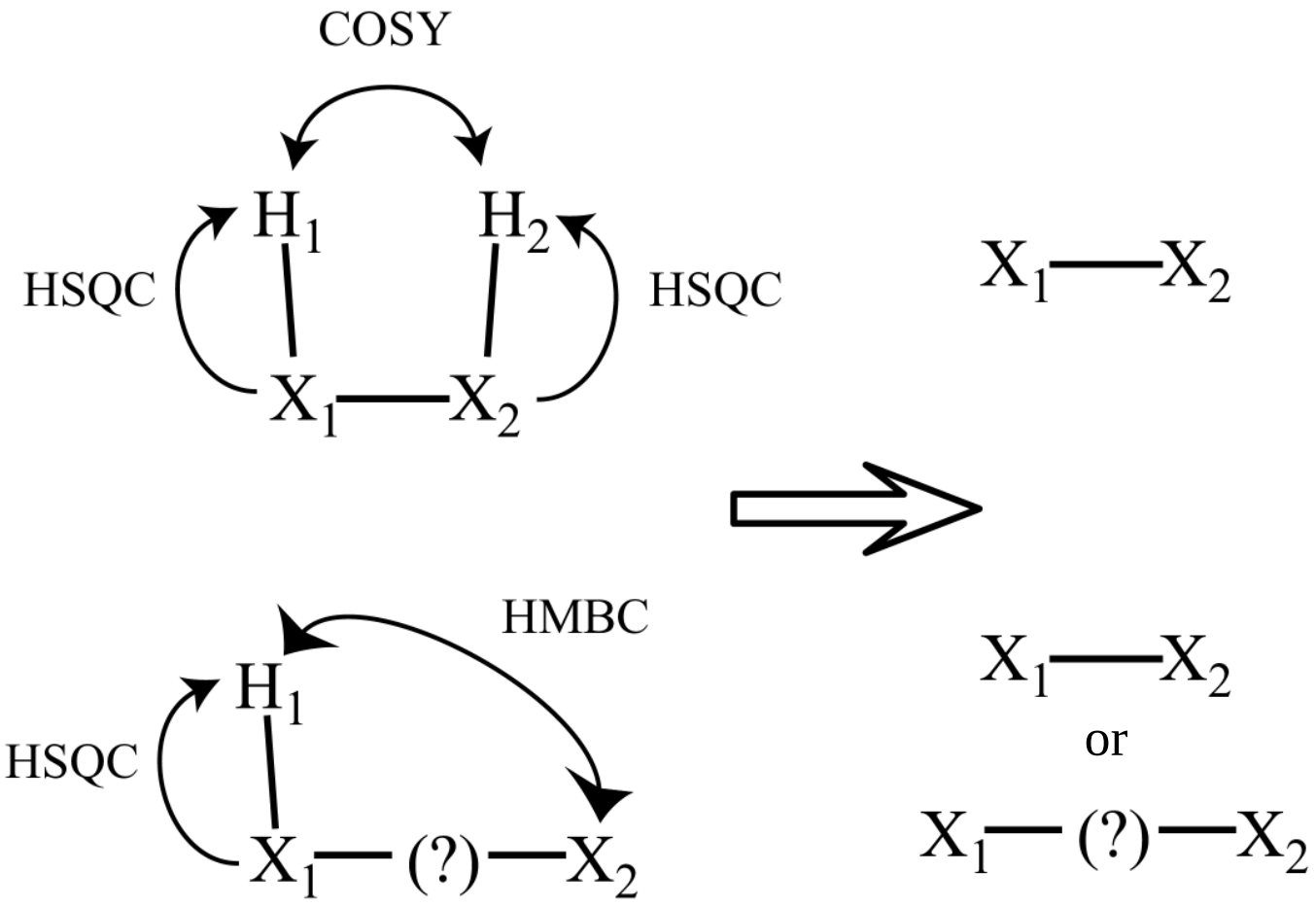




EXAMPLE, 2D HMBC SPECTRUM



WHY DO WE USE 2D NMR?



2D NMR provides proximity relationships between non-H atoms

EXAMPLE, ELEMENTAL FORMULA

- $[M + Na]^+ = 387$
- M = 364
- 18 C + 16 H = 232
- Remain: 364 – 232 = 132
- 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, LSD INPUT FILE, ATOM STATUS

MULT Index Symbol Hybridization Multiplicity [Charge]

```
MULT 1  C 2 0
MULT 2  C 2 0
MULT 3  C 2 0
MULT 4  C 2 0
MULT 5  C 2 0
MULT 6  C 2 0
MULT 7  C 2 0
MULT 8  C 2 1
MULT 9  C 2 1
MULT 10 C 2 1
MULT 11 C 2 1
MULT 12 C 2 1
MULT 13 C 2 1
MULT 14 C 3 1
MULT 15 C 3 1
MULT 16 C 3 2
MULT 17 C 3 3
MULT 18 C 3 3
MULT 19 O 3 0
MULT 20 O 3 0
MULT 21 O 3 0
MULT 22 O 3 1
MULT 23 O 3 1
MULT 24 O 3 1
MULT 25 O 3 1
MULT 26 O 2 0
```

-O-

-OH

=O

C atoms

C1 (169.9 ppm) is assigned to a carbonyl group bound to O26

EXAMPLE, LSD INPUT FILE, BONDS

```
COSY 14 15
COSY 14 16
COSY 8 10
COSY 9 11

BOND 17 19 ; Me-O-Ar
BOND 18 20 ; Me-O-Ar
BOND 1 26  ; C=O
```

- **COSY** : ^{3}J by default.
- **COSY** 9 13 3 4 : ^{3}J or ^{4}J coupling of H9 and H13

EXAMPLE, LSD INPUT FILE, HSQC AND SHIX

```
HSQC 8 8
HSQC 9 9
HSQC 10 10
HSQC 11 11
HSQC 12 12
HSQC 13 13
HSQC 14 14
HSQC 15 15
HSQC 16 16
HSQC 17 17
HSQC 18 18
```

```
; CD3OD at 49.2

SHIX 1  169.9
SHIX 2  153.9
SHIX 3  151.2
SHIX 4  148.9
SHIX 5  147.3
SHIX 6  134.1
SHIX 7  125.0
SHIX 8  124.8
```

```
SHIX 9  121.4
SHIX 10 116.4
SHIX 11 115.8
SHIX 12 114.5
SHIX 13 112.2
SHIX 14  85.4
SHIX 15  74.3
SHIX 16  62.6
SHIX 17  56.7
SHIX 18  56.5
```

- **HSQC** 8 8 : C8 bound to H8
- **SHIX** 1 169.9 : δ(C1) = 169.9 ppm
  - Used by pyLSD, ignored by LSD

EXAMPLE, LSD INPUT FILE, HMBC

```
HMBC 1 12
HMBC 2 8
HMBC 2 12
HMBC 2 14
HMBC 3 10
HMBC 3 17
HMBC 4 11
HMBC 4 18
HMBC 5 13
HMBC 5 9
HMBC 6 11
HMBC 6 15
HMBC 7 10
HMBC 8 12
HMBC 9 13
HMBC 9 15
HMBC 12 8
HMBC 13 9
HMBC 13 15
HMBC 14 15
HMBC 15 13
HMBC 15 14
HMBC 15 16
HMBC 16 15
```

- HMBC 1 12 : C1 and H12 are separated by 2 or 3 bonds.
- The ^{n}J hypothesis with n>3 is taken into account only if an ELIM command is present.
- HMBC (X Y) Z : HMBC X Z or HMBC Y Z

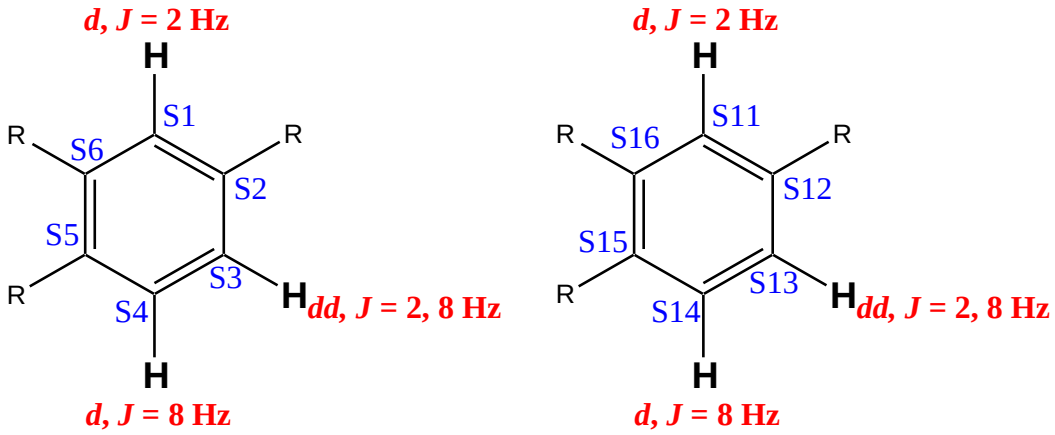
EXAMPLE, LSD INPUT FILE, ATOM PROPERTIES

```
ELEM  L1  O
LIST  L2  14 15 16 17 18
PROP  L2  1  L1
PROP  1  2  L1
```

```
CARB  L3
LIST  L4  10 11 12 13
PROP  L4  0  L3
PROP  L1  0  L3
```

- L1 is the list of all Oxygen atoms
- L2 contains the atoms C14 to C18: from 55 to 85 ppm
- Each atom in L2 is bound to exactly of atom in list L1:
  C13 to C18 are bound to a single Oxygen atom
- **C1 is bound to 2 Oxygen atoms**: 169.9 ppm
- L3 is the list of all Carbon atoms
- L4 is the list of atoms C10 to C13: δ from 112.2 to 116.4 ppm
- Each atom in L4 is bound to Carbon atoms only (0 means all !)
- **Each Oxygen is bound to Carbon atoms only** (no peroxydes)

EXAMPLE, LSD INPUT FILE, SUBSTRUCTURE




```
SSTR S1 C 2 1
SSTR S2 C 2 1
SSTR S3 C 2 0
SSTR S4 C 2 1
SSTR S5 C 2 0
SSTR S6 C 2 0
SSTR S11 C 2 1
SSTR S12 C 2 1
SSTR S13 C 2 0
SSTR S14 C 2 1
SSTR S15 C 2 0
SSTR S16 C 2 0
```

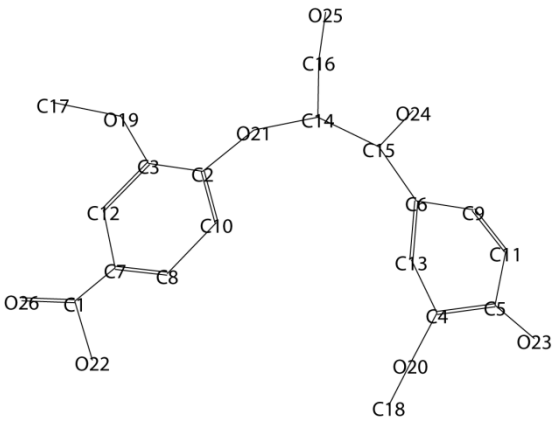
Type assignment of substructure atoms

```
LINK S1 S2
LINK S2 S3
LINK S3 S4
LINK S4 S5
LINK S5 S6
LINK S6 S1
LINK S11 S12
LINK S12 S13
LINK S13 S14
LINK S14 S15
LINK S15 S16
LINK S16 S11
```

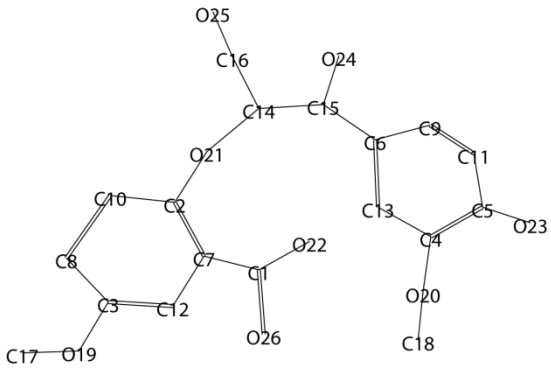
Bonds between substructure atoms

EXAMPLE, SOLUTIONS

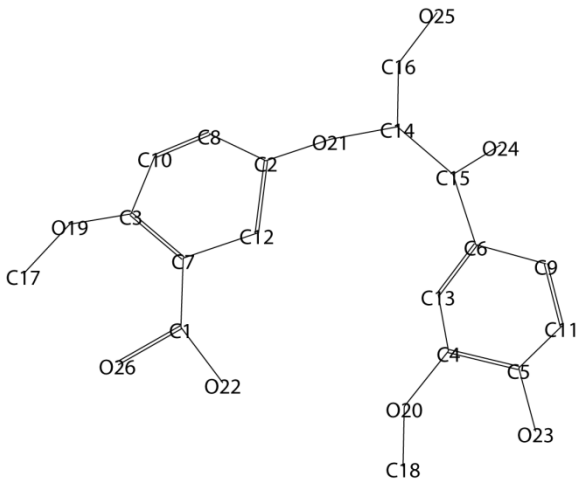
1



2



3



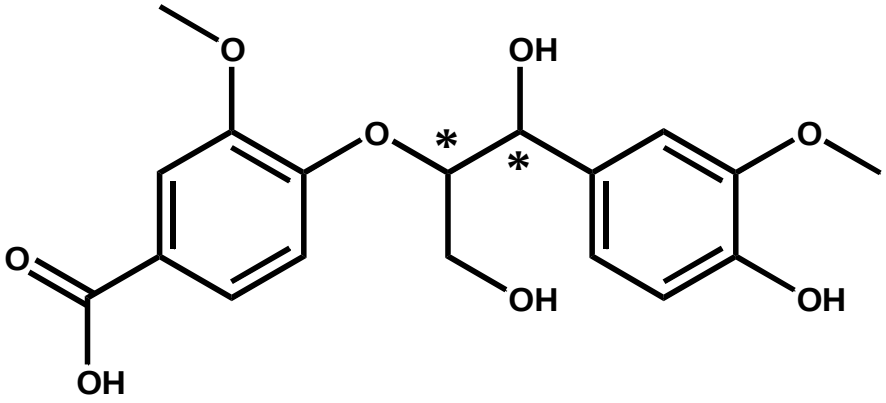
- Structures generated by **lsd**, 2D coordinates by **outlsd**, drawings by **genpos**. Drawing improvements by **m_edit**.
- **solve** chains all the process.
- Structure generation in less than 1 second on a PC.

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 difference 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 depends on the predictor
- The predictor depends on the database it makes use of and on the prediction algorithm
- Never blindly trust a predictor

FEEDBACK FROM AN LSD USER

... The advantage of the LSD is that it will determine all possibilities that should be considered, without preconceptions.

Thus, we have used it to confirm that our proposed structures were indeed consistent with the data and that other possible structures (as suggested by LSD) could be disregarded based on chemical arguments; this gives us further confidence that the structures we are proposing are correct."

Tim Claridge, Université d'Oxford

Auteur de « High-Resolution NMR Techniques in Organic Chemistry »

www.univ-reims.fr/LSD

Natural Product Chemistry Research Team in Reims

ACKNOWLEDGEMENTS

- **Bertrand Plainchont** (ICMR)
- **Jean-Hugues Renault** (ICMR)
- **Jeanne Hubert** (NatExplore)
- **Nicolas Borie** (ICMR)
- **Sébastien Chollet** (ICMR)

- **Romain Reynaud** (Soliance-Givaudan)
- **Ali Bakiri** (CIFRE Soliance-Givaudan/ICMR)

- **Eric Courot** (URVVC Reims)
- **Christophe Clément** (URVVC Reims)

- **Leandros Skaltsounis** (NKU Athens)
- **Maria Halabalaki** (NKU Athens)
- **Eirini Kouloura** (NKU Athens)

- **Jean-Baptiste Gallé** (Strasbourg university)
- **Catherine Vonthron-Sénéchau** (Strasbourg university)