The LSD Software

A tool for the structure determination of small molecules

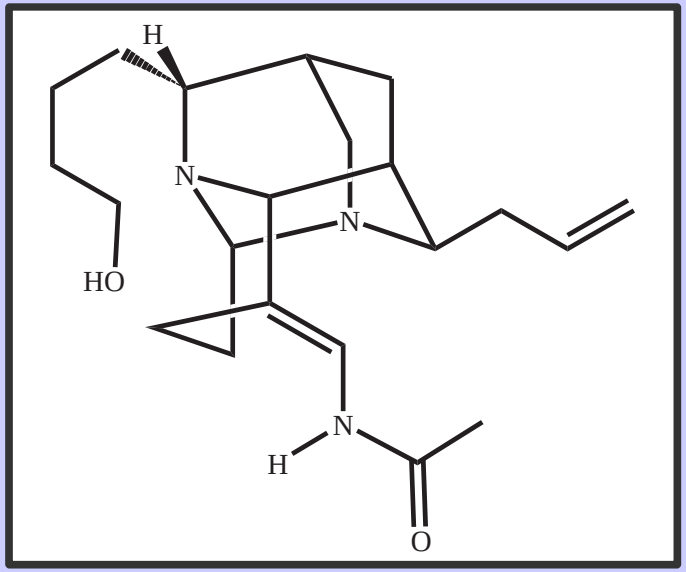
Jean-Marc Nuzillard

University of Reims-Champagne-Ardenne
Molecular Chemistry Institute

Cargèse, 2013, March 23th

Structure Analysis

- **Small Molecules**
  - Secondary metabolites
  - Synthetic molecules



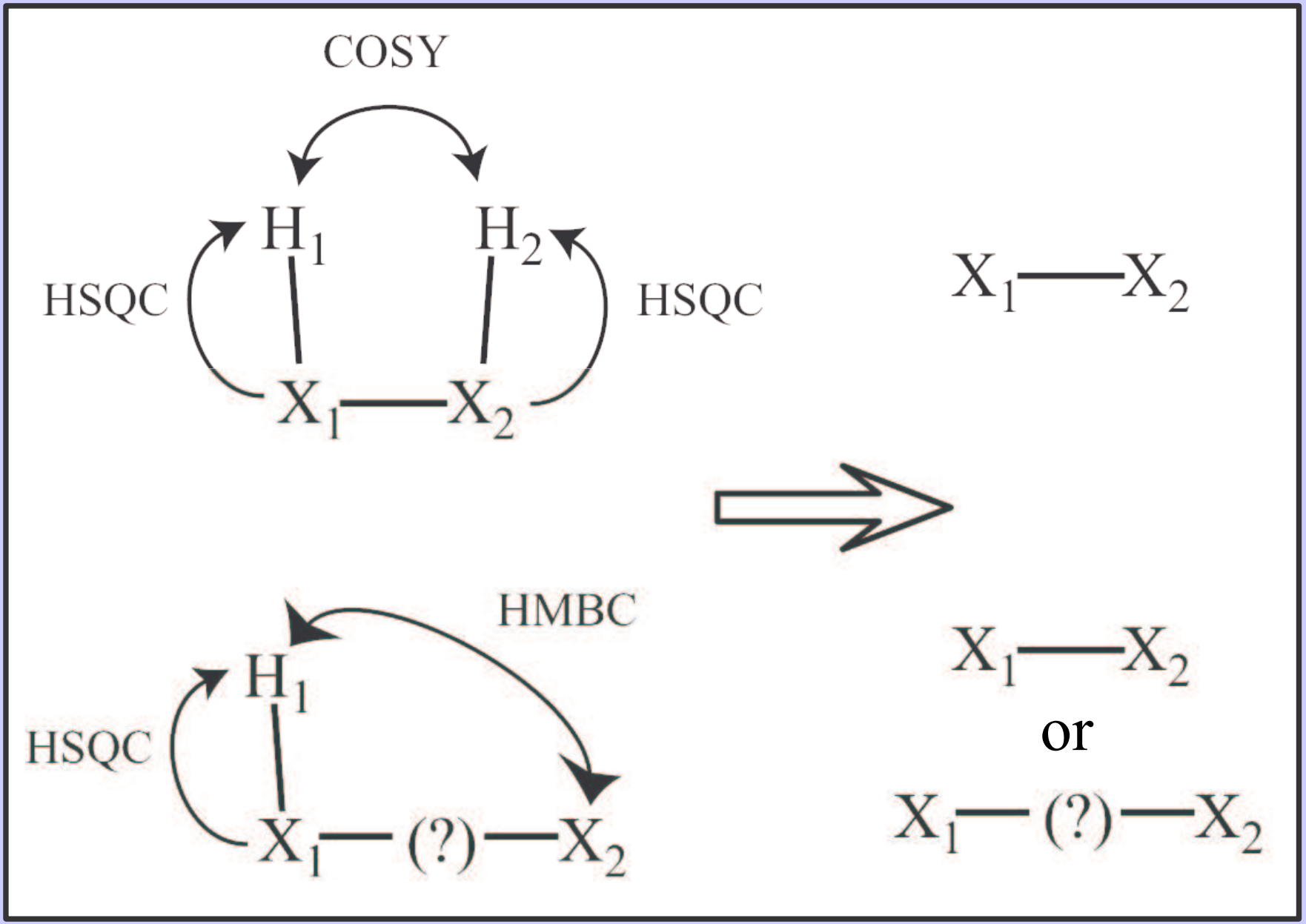
- **Problem**
  - Nothing is known ...
  - ... Everything must be found.

The LSD Software

Logic for Structure Determination

- **The goal of the LSD software is to obtain all possible 2D structures that are compatible with a set of 1D and 2D NMR spectra.**
- **LSD uses the chemist's knowledge of elementary chemical shift rules and/or other structure information like the origin of the molecule.**
- **LSD does not use any database.**

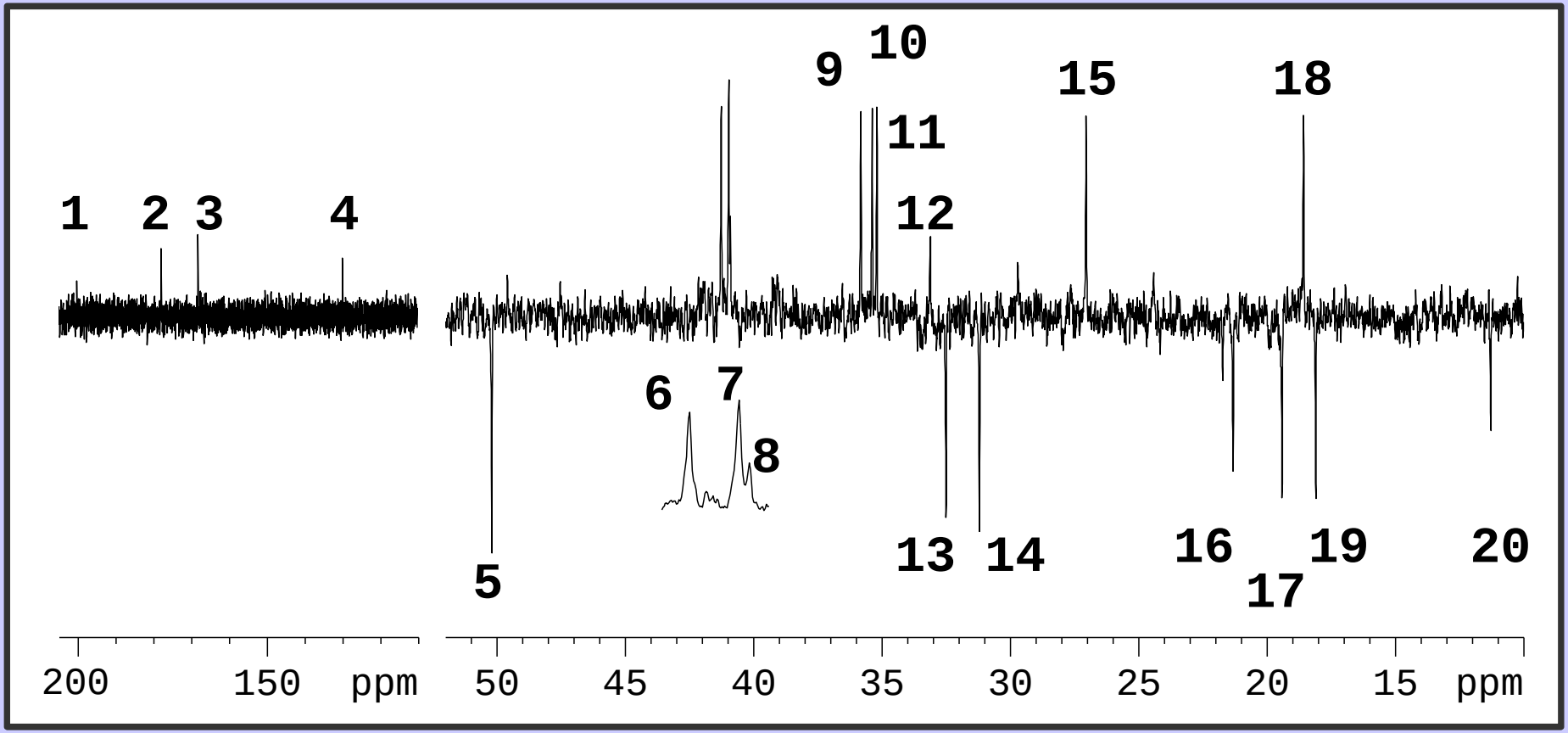
HSQC + COSY + HMBC



A case study

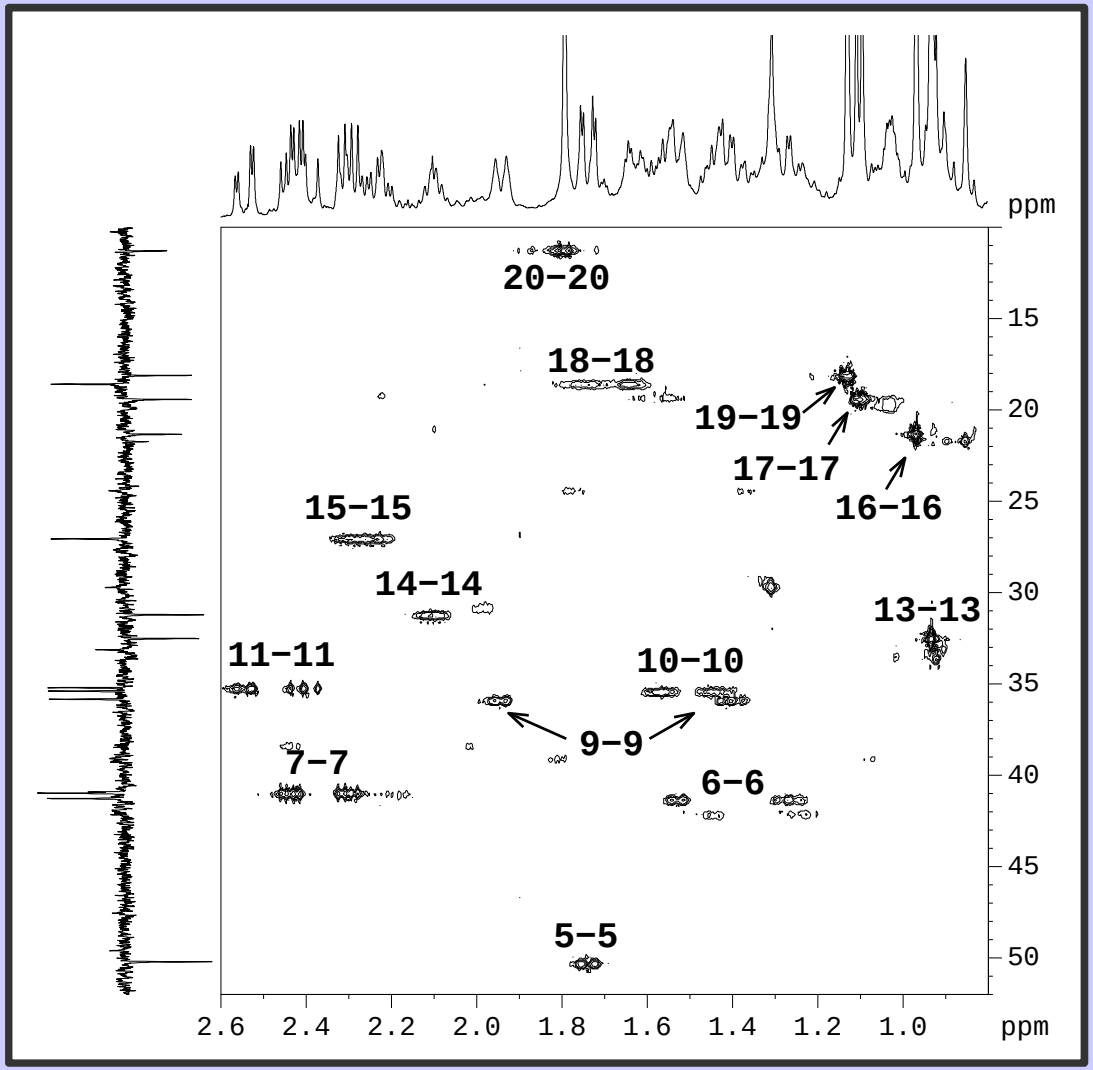
- **The studied molecule was extracted from a particular exotic wood that is well know for its resistance to insect attacks. This work was carried out in collaboration with the French « Research Institute for Development »(IRD)**
- **^{1}H, ^{13}C, COSY, HSQC and HMBC of the sample were recorded.**

J-modulated ^{13}C spectrum



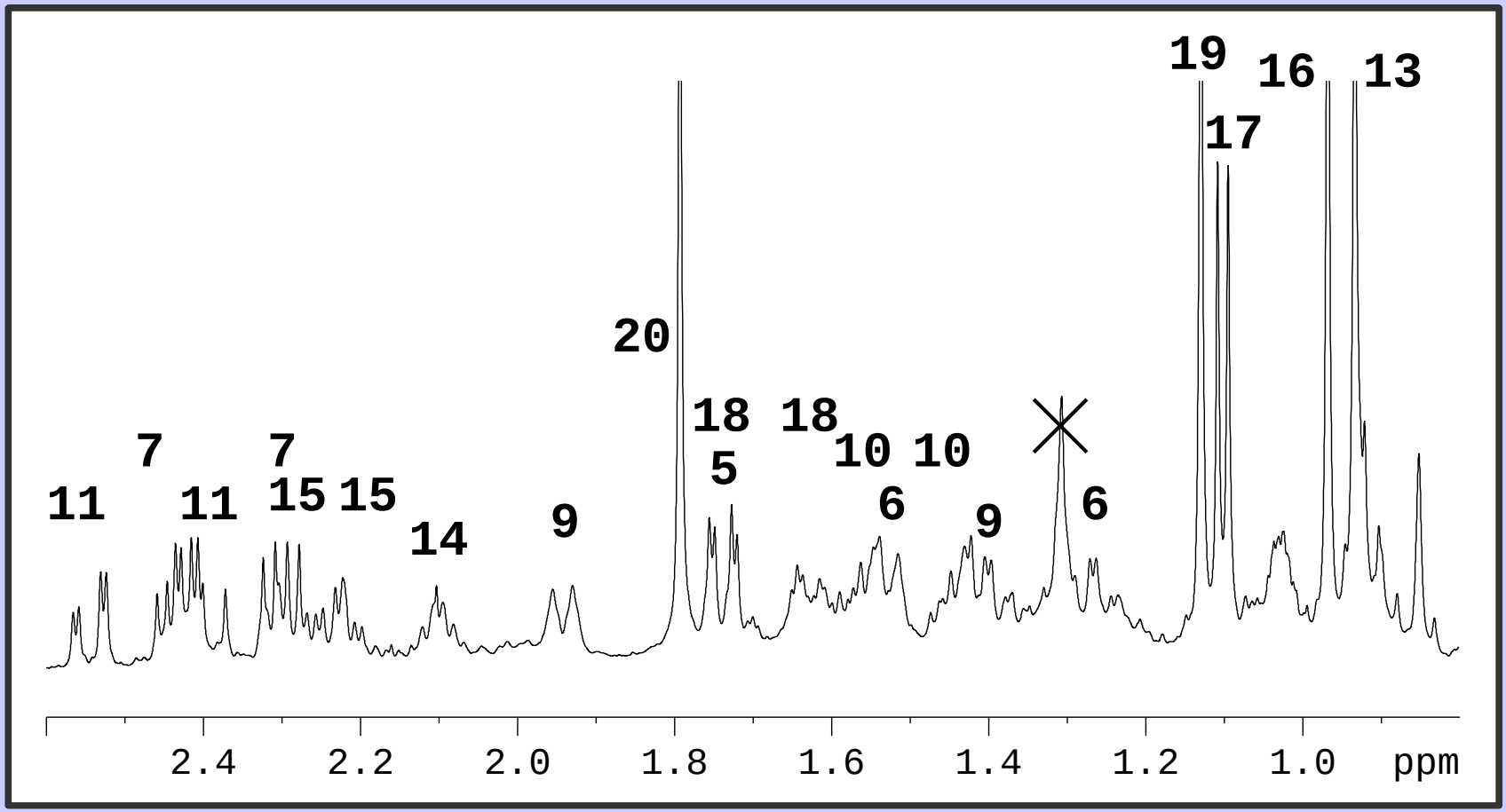
- **Carbon atoms are numbered in decreasing chemical shift order (arbitrarily).**

HSQC, ^{1}J correlations



- **An hydrogen atom gets the number of the carbon atom it is attached to.**

^{1}H spectrum

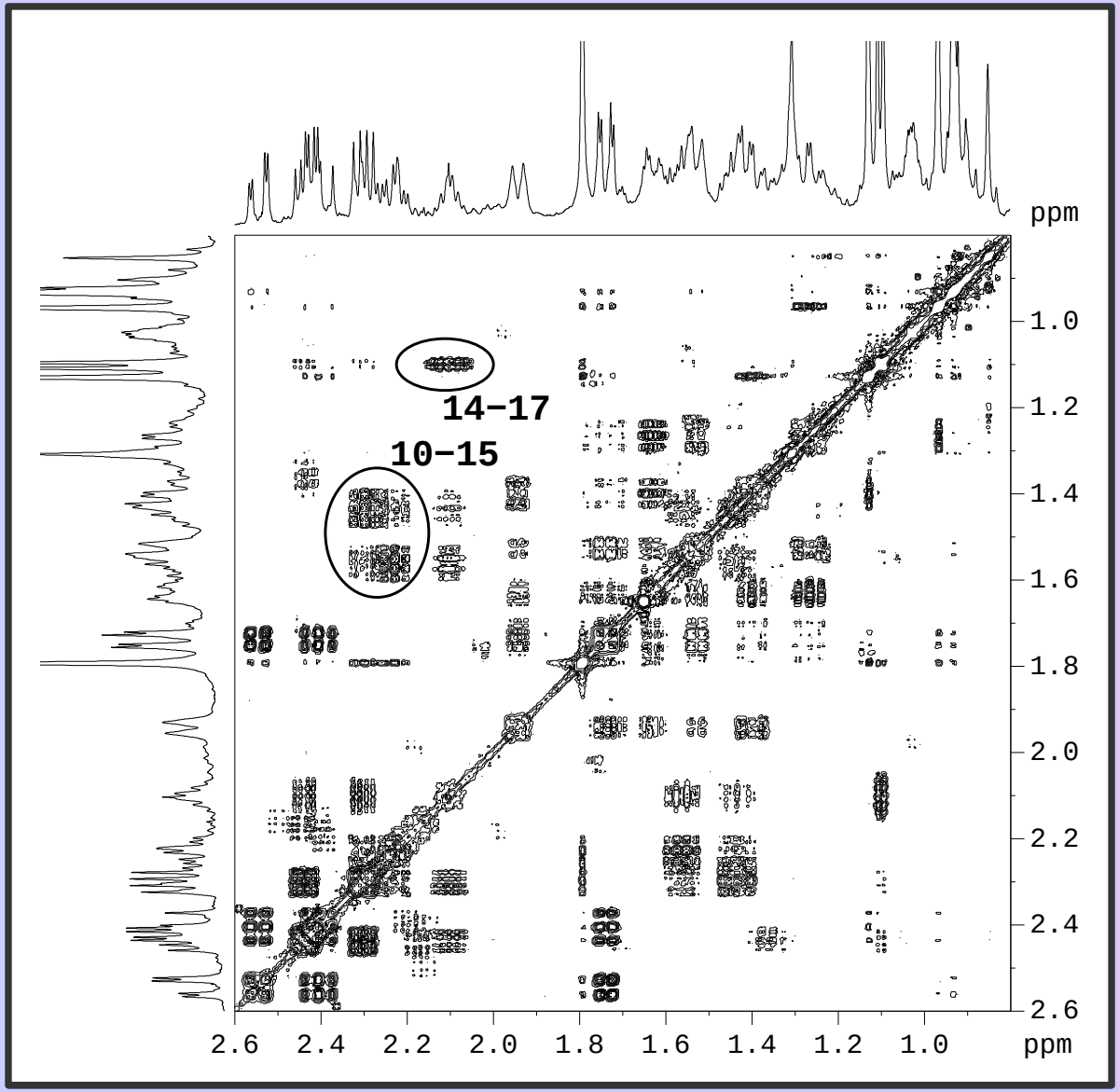


- **The H atoms that are labelled 13, 16, 17, 19 and 20 belong to methyl groups.**

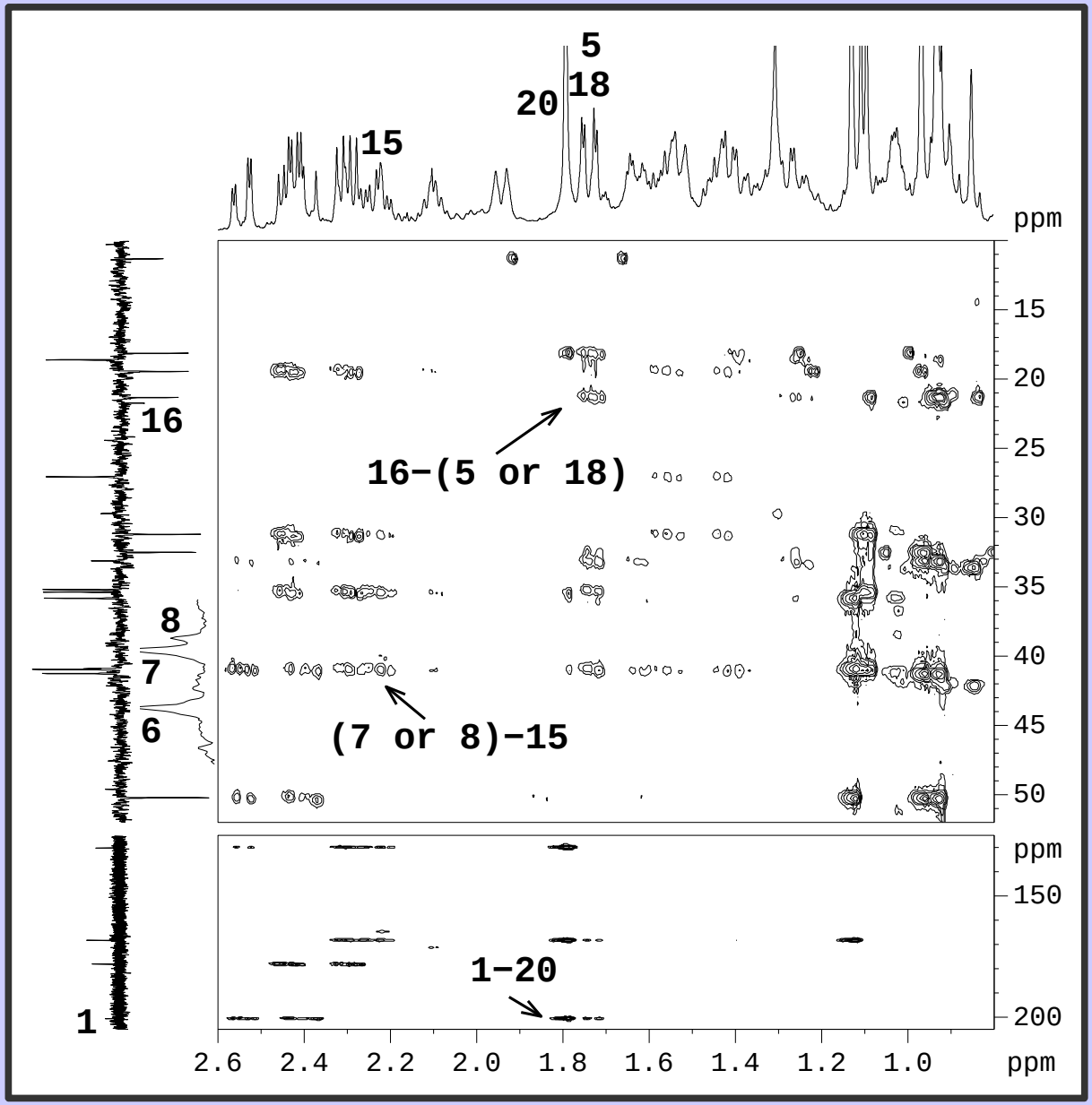
MS and 1D NMR

- **M = 320.**
- **The 13C spectrum shows 6 Q, 2 CH, 7 CH_2 and 5 CH_3, accounting for M = 271.**
- **The 49 missing mass units originate from 3 oxygen atoms (48) and 1 hydrogen atom (1).**
- **A molecular formula $C_{20}H_{32}O_3$ is proposed.**
- **The extra H belongs to an OH group**
- **C 1 and C 2 are doubly bonded to oxygen atoms**
- **C 3 and C 4 are bonded together**

COSY spectrum



HMBC spectrum



The 3 indeterminacy levels

- **Each HMBC correlation gives rise to 3 indeterminacy levels**
  - Chemical shift ambiguity: in case of a HMBC correlation of a group of unresolved nuclei.
  - Each HMBC correlation has to be considered as arising from a ^{2}J, or ^{3}J, or longer-range coupling.
  - In the ^{3}J hypothesis, the exact identity of the intermediate atom (X) is unknown
- *LSD was designed to systematically explore these three indeterminacy levels.*

LSD input file

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

```
BOND 1 21
BOND 2 22
BOND 2 23
BOND 3 4

HSQC 1 1
HSQC 5 5
HSQC 6 6
HSQC 7 7
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
HSQC 19 19
HSQC 20 20
```

```
HMBC 1 11
HMBC 1 20
HMBC (5 18) 1
HMBC 2 7
HMBC 3 15
HMBC 3 10
HMBC 3 19
HMBC 4 15
HMBC 4 20
HMBC 5 11
HMBC 5 19
HMBC 5 16
HMBC 5 13
HMBC (7 8) 11
HMBC (7 8) 15
HMBC (7 8) 18
HMBC (7 8) 9
HMBC (7 8) 19
HMBC (7 8) 17
HMBC 6 16
HMBC 6 13
```

```
HMBC 9 19
HMBC (10 11) 7
HMBC (10 11) 15
HMBC (10 11) 17
HMBC (5 18) 12
HMBC 12 16
HMBC 12 13
HMBC 13 16
HMBC 14 7
HMBC 14 10
HMBC 14 17
HMBC 15 10
HMBC (5 18) 16
HMBC 16 13
HMBC 17 7
HMBC 17 10
HMBC (5 18) 19

COSY 14 17
COSY 10 15
```

LSD input file

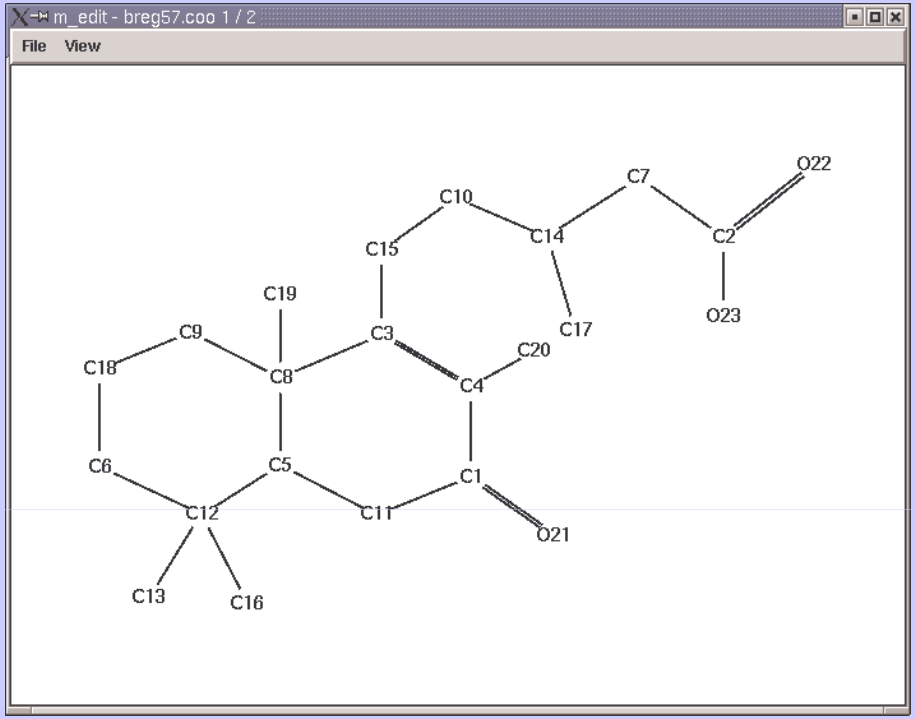
```
QUAT L1
LIST L2 13 16 19 20
PROP L2 1 L1
```

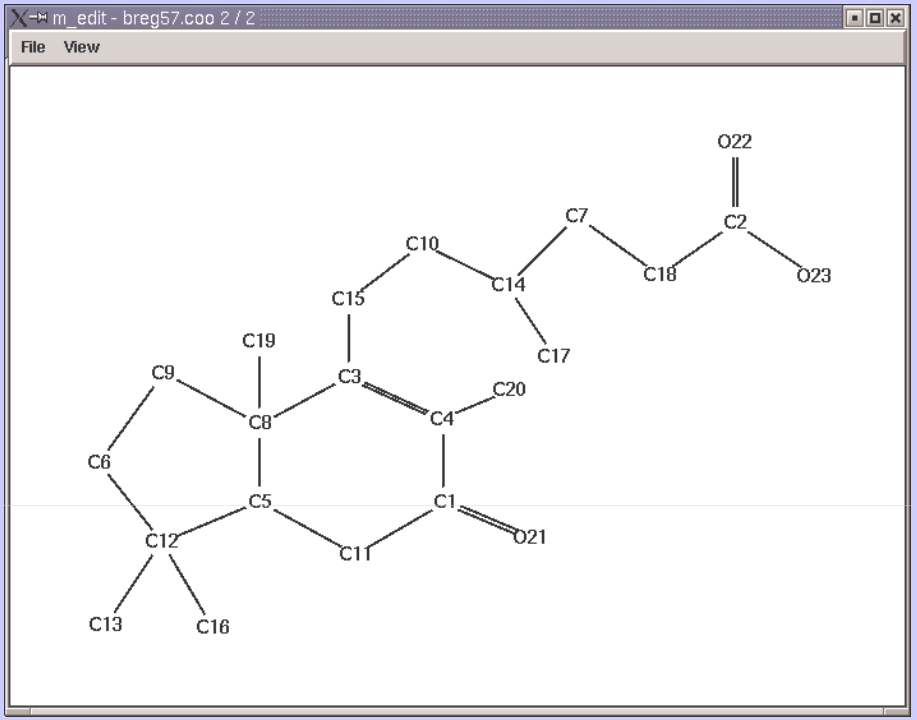
- **Each skeletal (non-hydrogen) atom is described by its status : resonance #, chemical element, hybridization state, number of attached hydrogens.**
- **Bonds, correlations and additional atom properties are then reported**
- **A sub-structure information can be provided (for presence or absence).**

Execution control

- File checking
- Step by step operation
- Verbosity level
- Resolution history
- Sub-structure handling
- Automatic elimination of HMBC correlations
- Production of uncomplete solutions
- Elimination of duplicated solutions
- Elimation of non-connected solution
- Limits on resolution time and number of solutions

Results





- The molecular drawings can be edited by m_edit
- The left hand side solution corresponds to an already known molecule (labdane family)

LSD graphical output

- **The structures that are produced by LSD are written as text in a « local » format (.sol).**
- **The « outlsd » program converts them as**
  - Bond lists
  - 2D coordinates (.coo)
  - 3D coordinated (for Macromodel)
  - SMILES chains
  - 2D and 3D « Moldraw » files (.mol)
- **The « genpos » programs produces Postscript files for structure printout**

The solution set...

- **...is empty**
  - The data set is logically not self-consistent
  - The biggest built fragment can be visualized
- **...contains one or few structures**
  - The simulation of chemical shifts and their comparison with experimental values allows to keep or to reject solutions.
- **...contains many (too many) solutions**
  - Supplementary data need to be considered

How it works...

- **Phase 0 : initialization**
  - Input file reading (syntax analysis)
  - Semantic analysis
  - Structure initial state from COSY and BOND
- **Phase 1 : HMBC correlations → structure**
  - Repeat
    - Correlation selection
    - Correlation analysis
  - While a correlations can be selected

How it works…

- **Phase 2 : systematic atom pairing**
  - … of the atoms that do not have all their neighbors

- **Phase 3 : validation**
  - Double bond placement
  - Bredt's rule checking
  - Sub-structure information
  - Uniqueness: identity/isomorphism detection

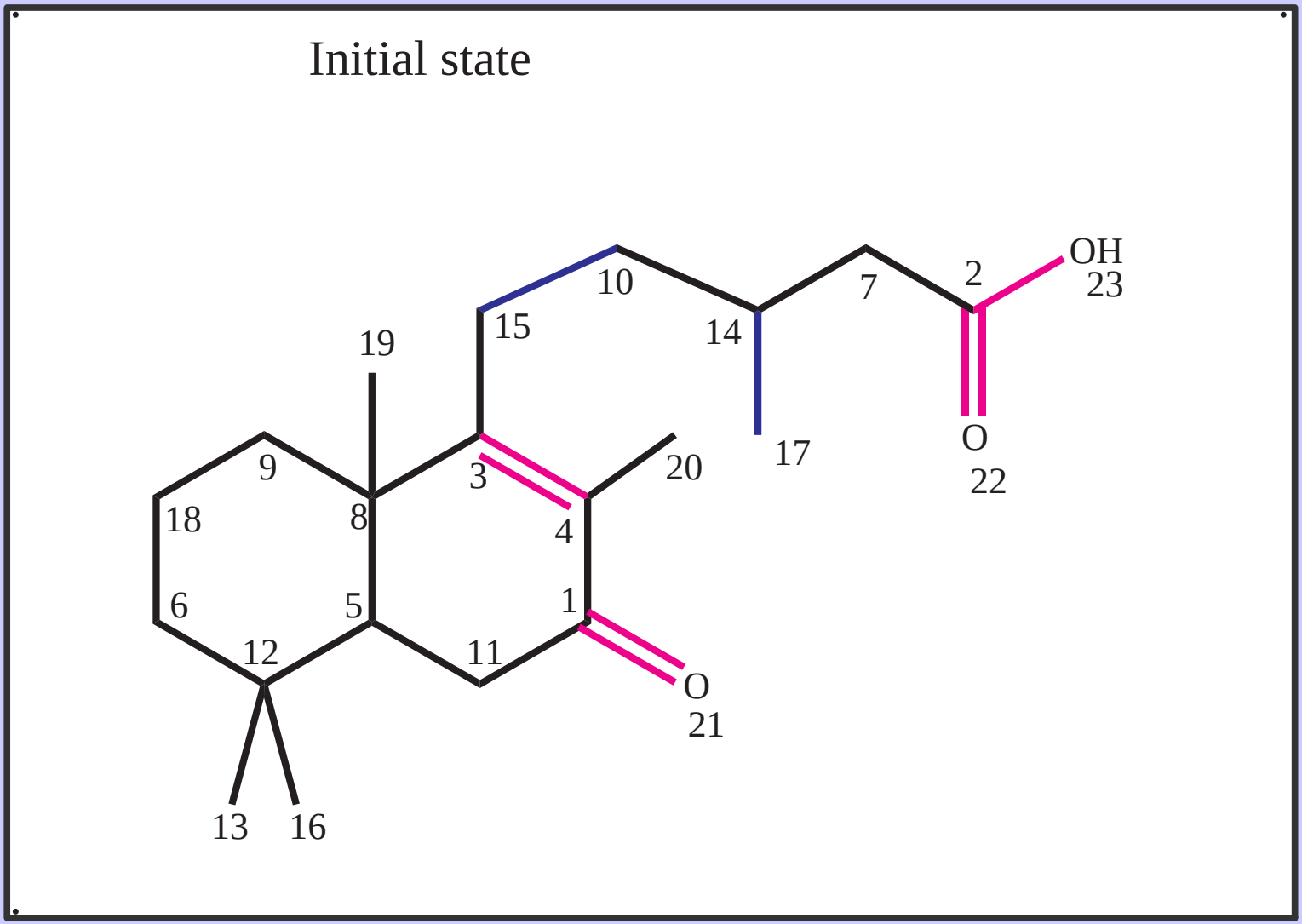
From Correlations to Structure

- **The order in which correlations are considered has a strong impact on resolution times.**
- **Correlations that open a minimum number of hypothesis must be chosen first.**
- **Selection order follows the following rules:**
  - The correlations of the atoms that are the closest to get all their neighbors are selected first.
  - The growing structure has to be as « in one piece » as possible.

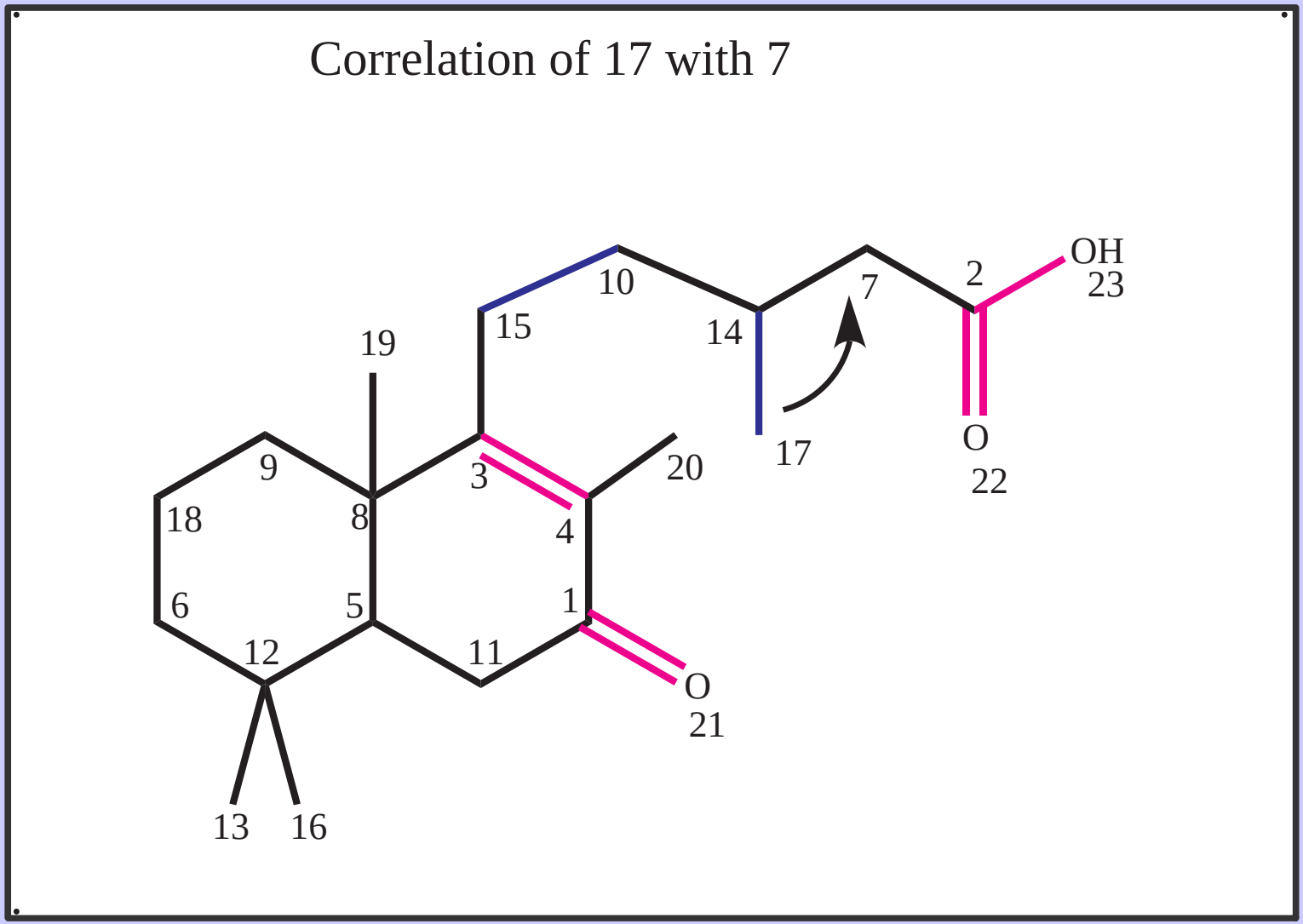
HMBC correlation analysis

- **Make an hypothesis if it is a correlation of a group, like in “HMBC (7 8) 9”.**
- **Test the ^{2}J hypothesis (establish one bond).**
- **Test the ^{3}J hypothesis (establish one or two bonds).**
- **Test the ^{n}J (n > 3) hypothesis, if possible.**
- **Eliminate the correlations that become useless because they can be interpreted from the already existing bonds.**

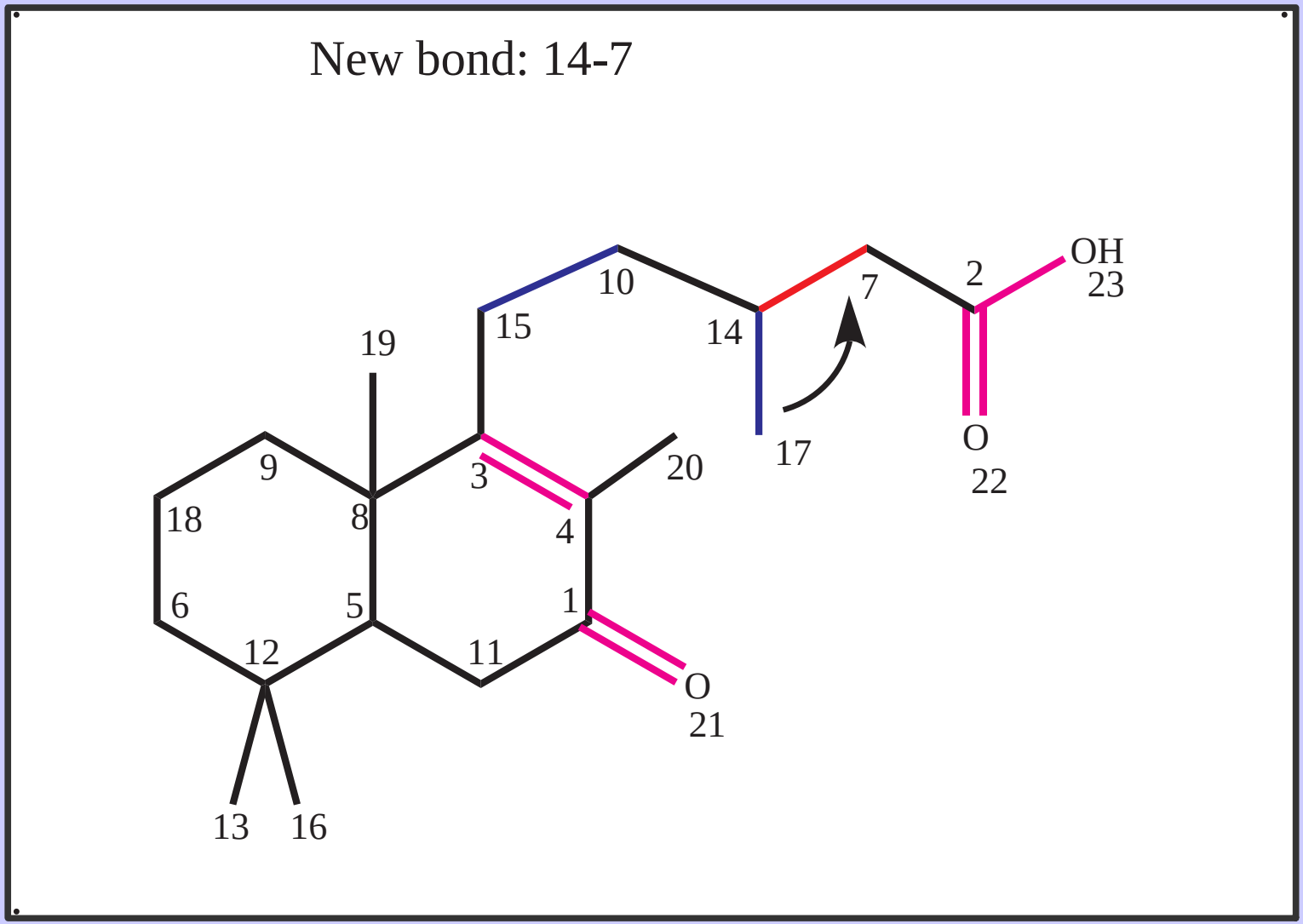
Resolution (1)



Resolution (2)

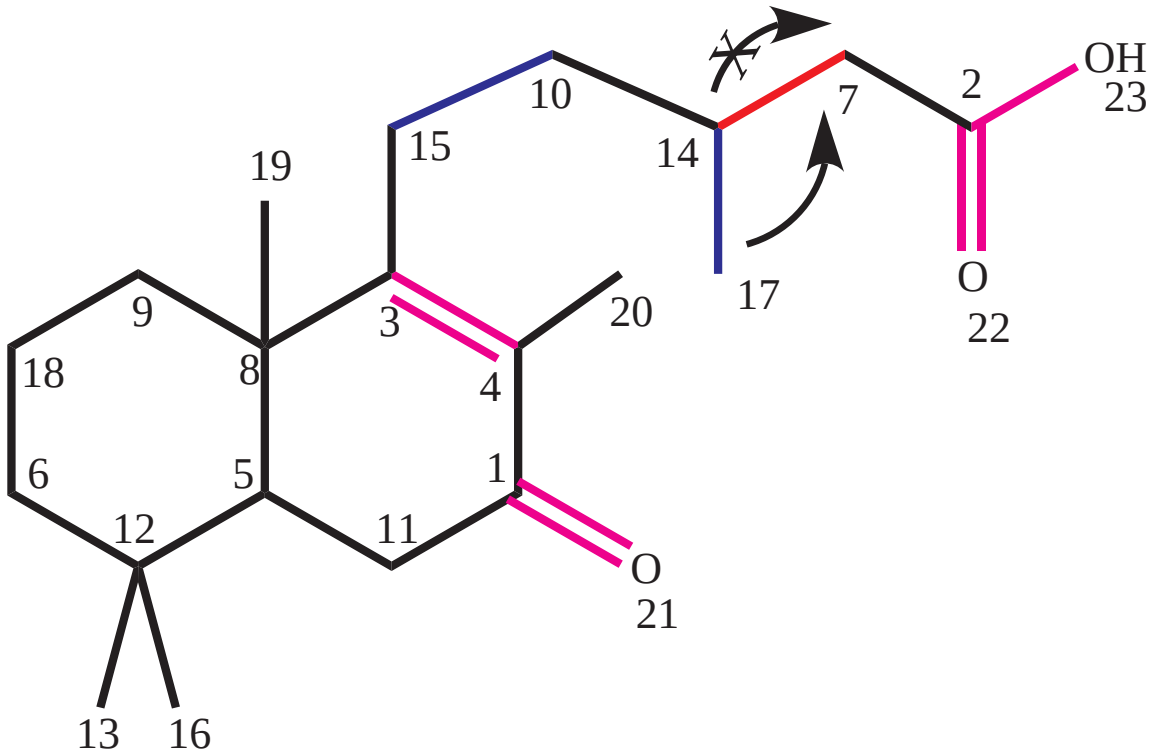


Resolution (3)

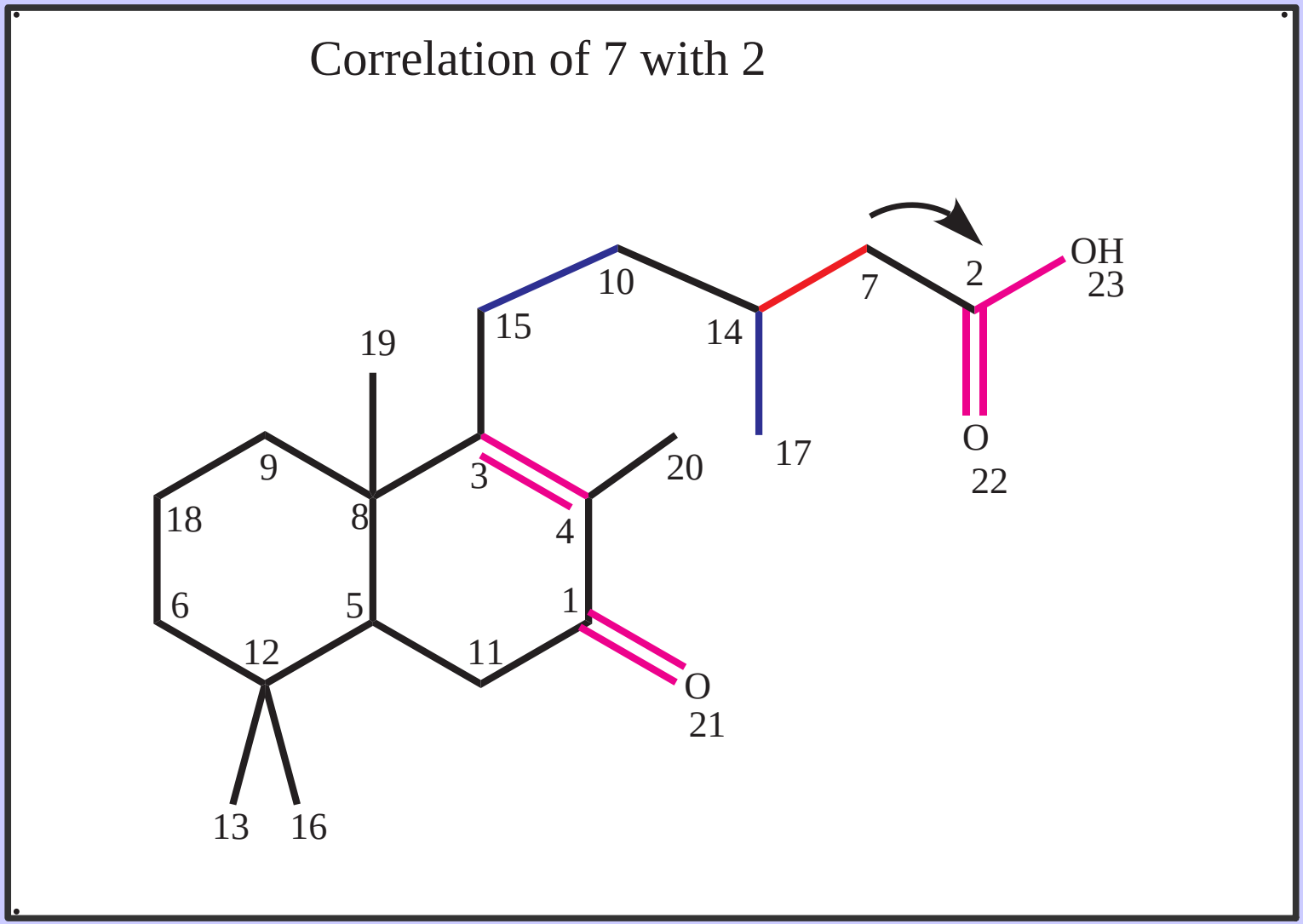


Resolution (4)

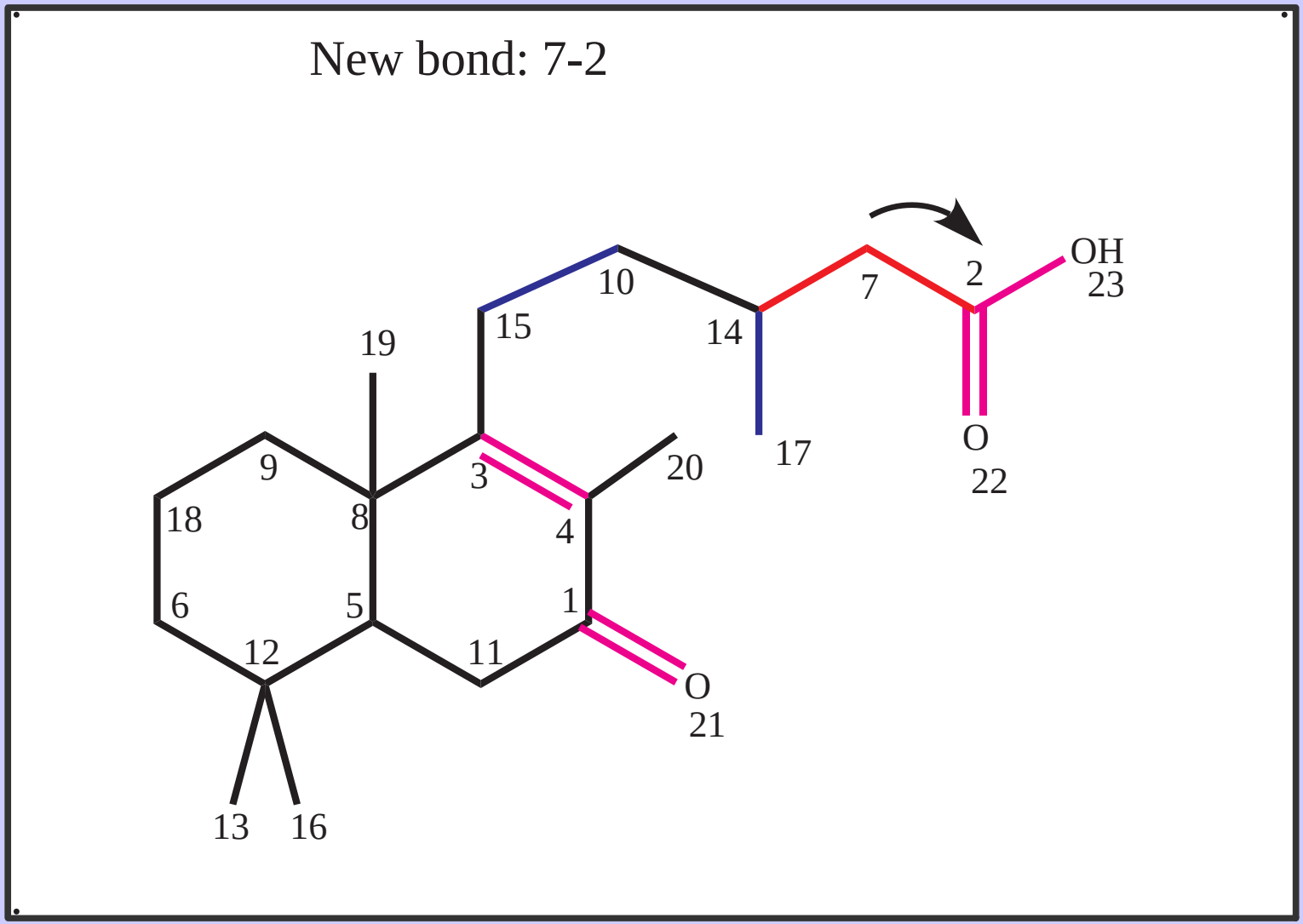
Useless correlation of 14 with 7



Resolution (5)

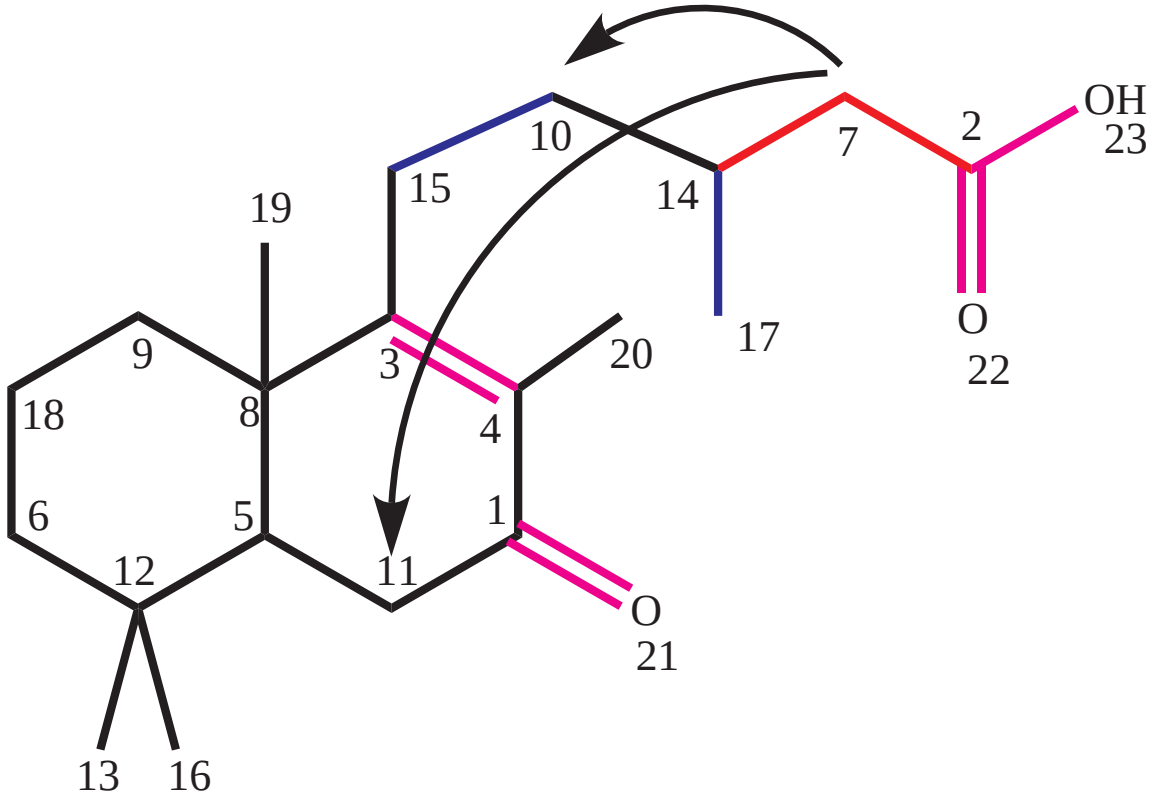


Resolution (6)



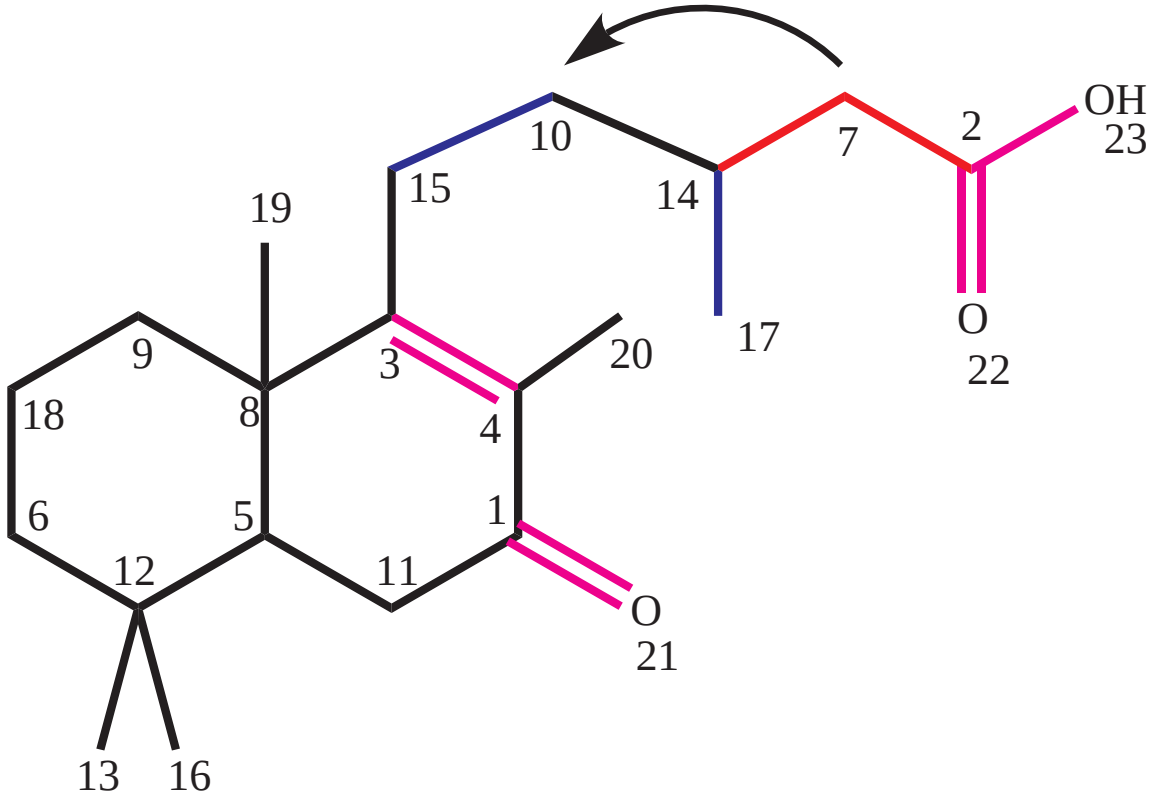
Resolution (7)

Correlation of 7 with either 10 or 11

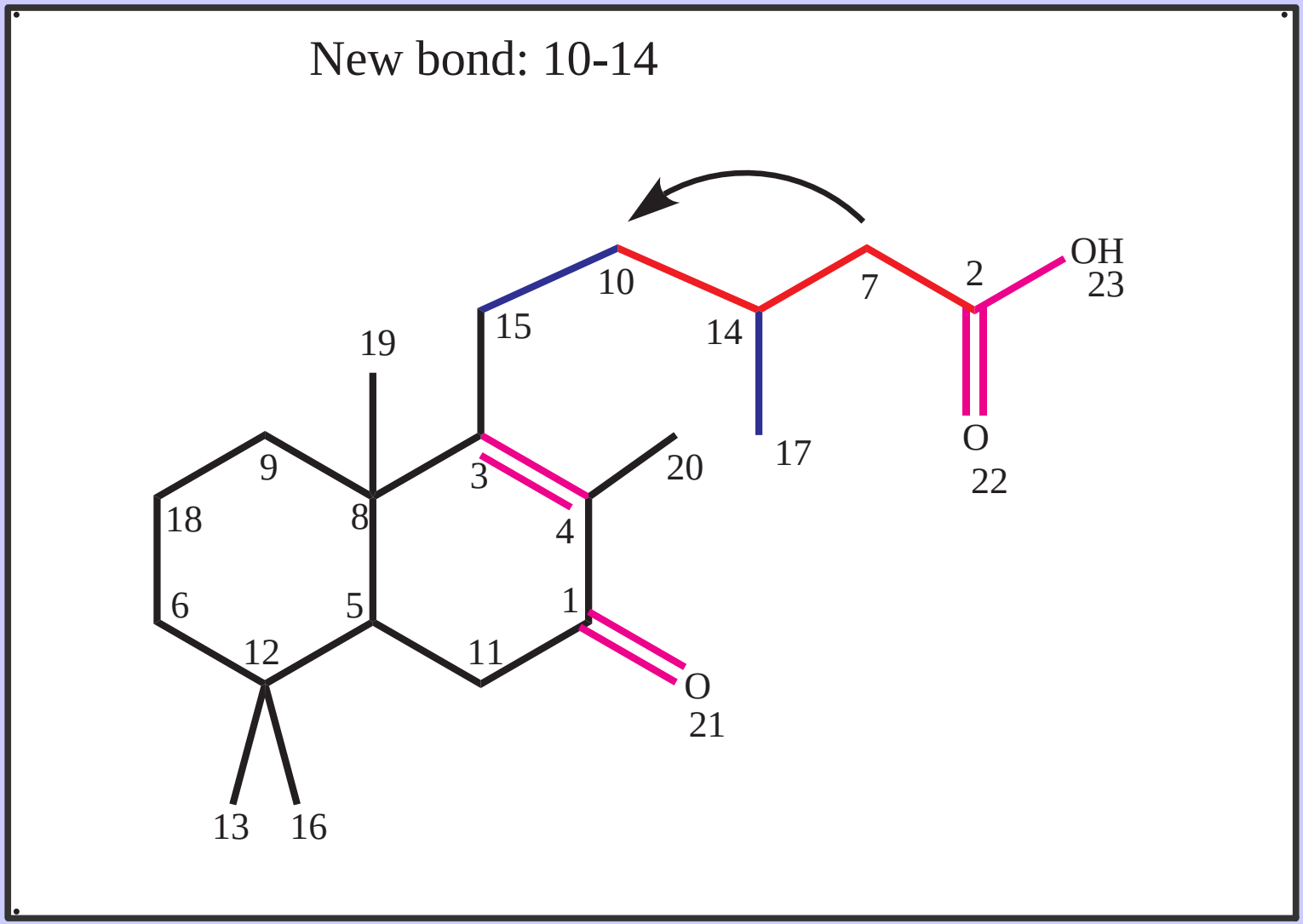


Resolution (8)

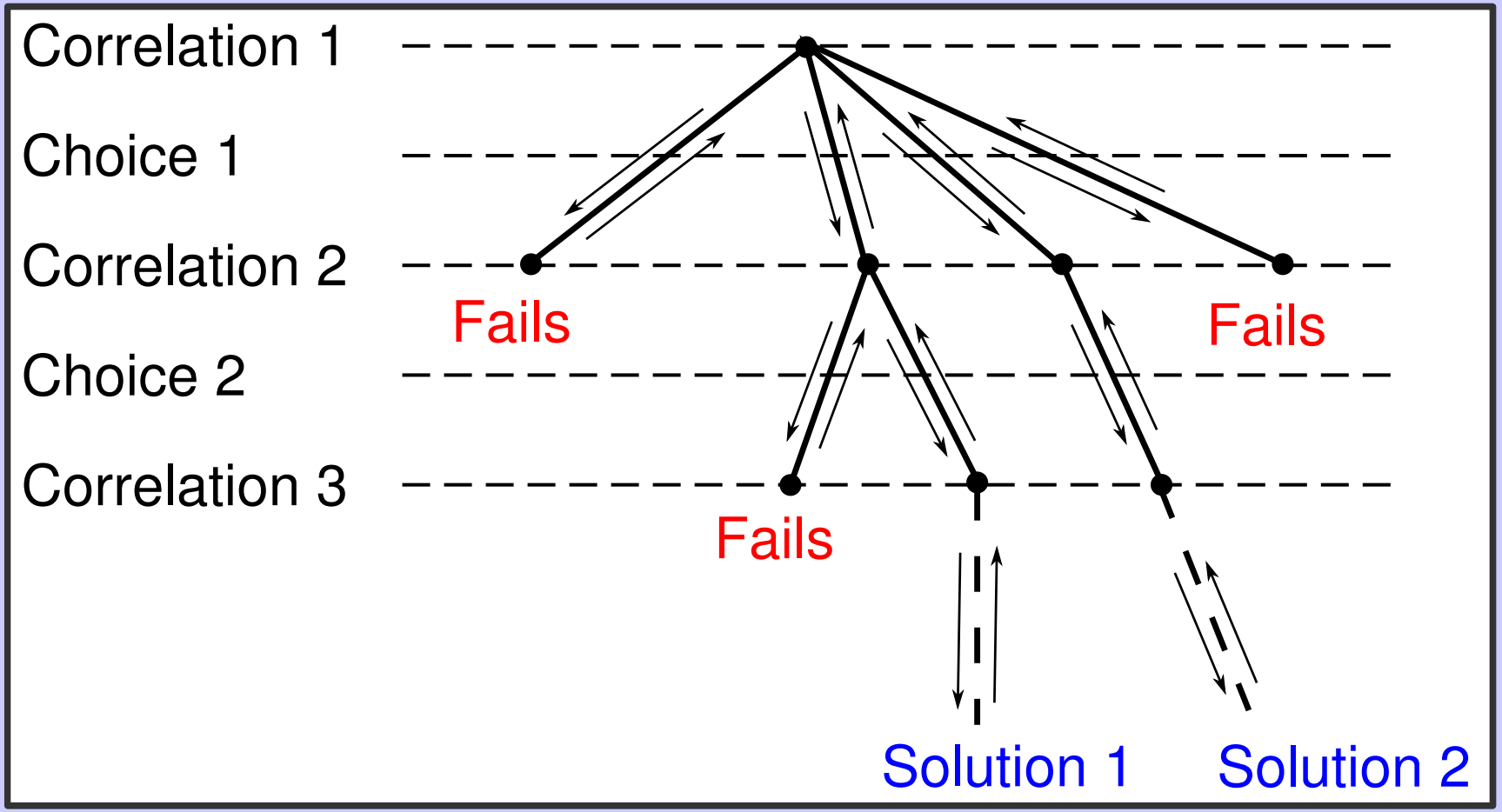
Hypothesis : 7 correlates with 10



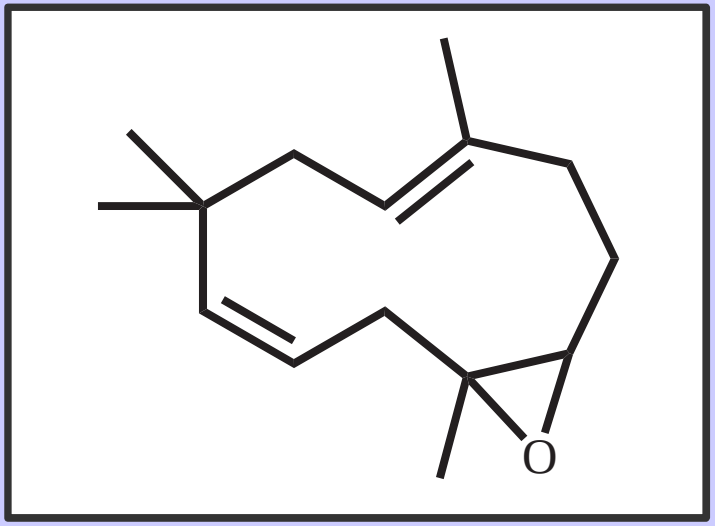
Resolution (9)



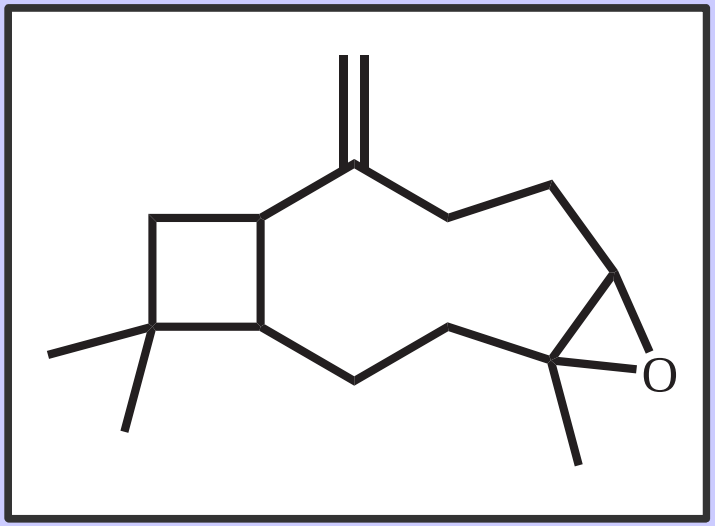
Hypothesis analysis

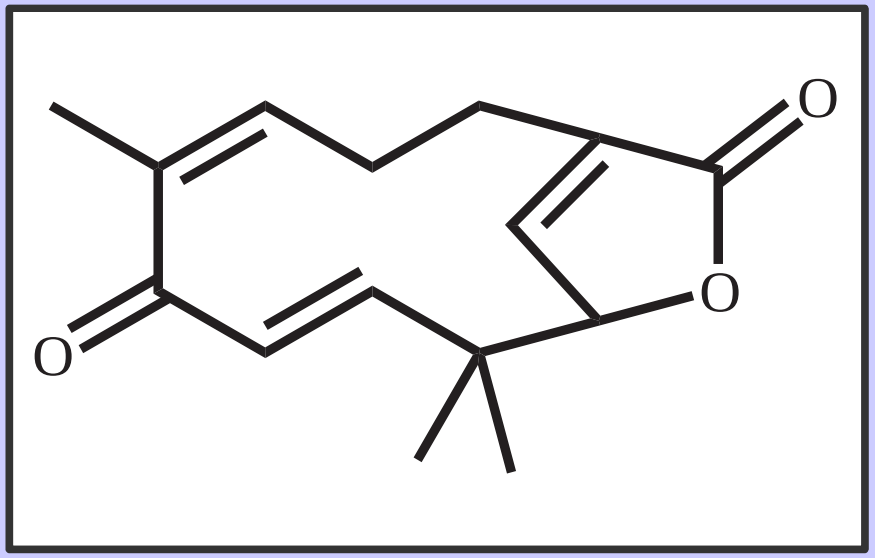


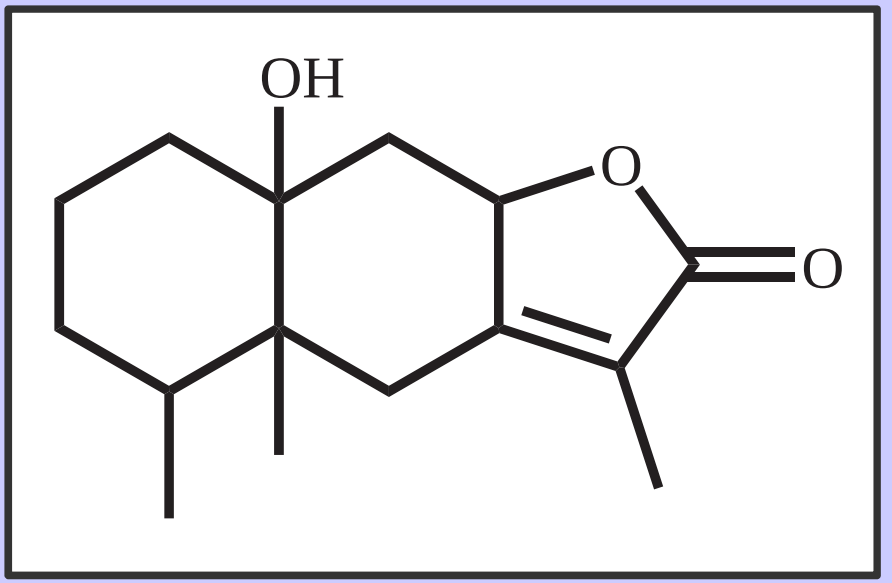
A few solved problems... (1)



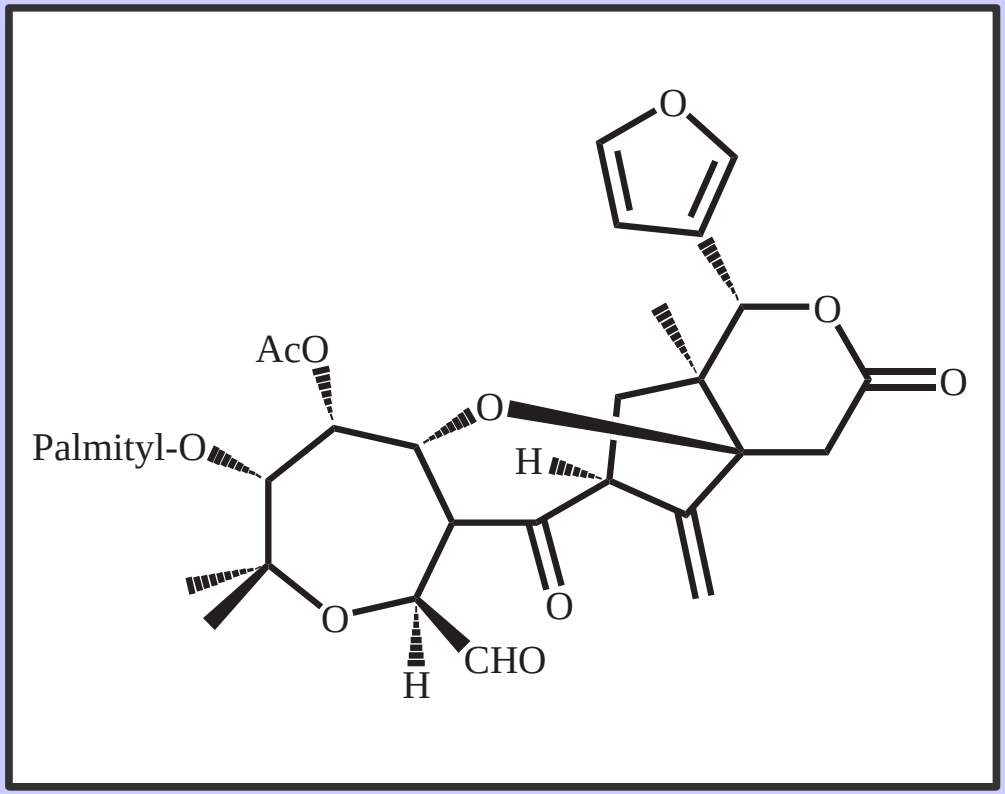
and

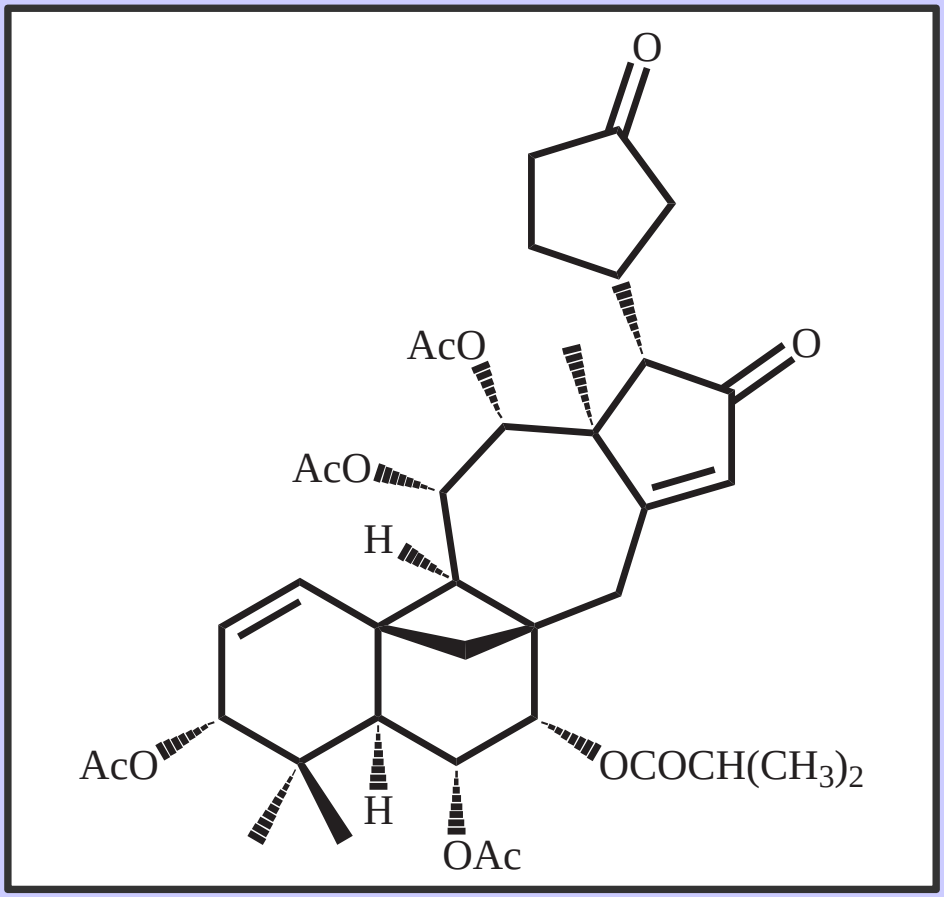




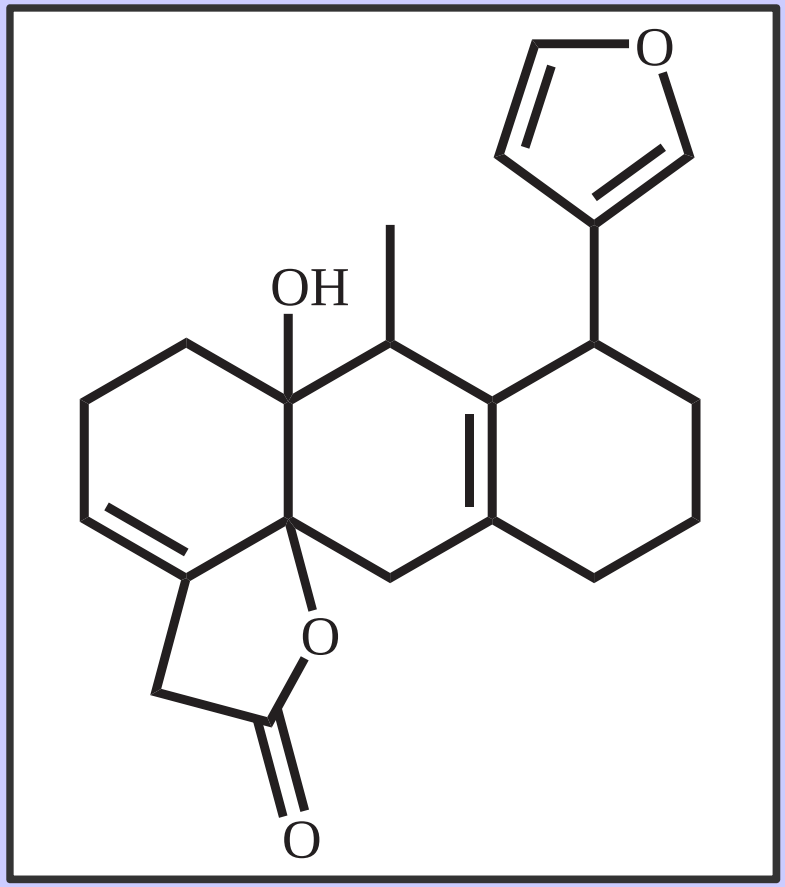


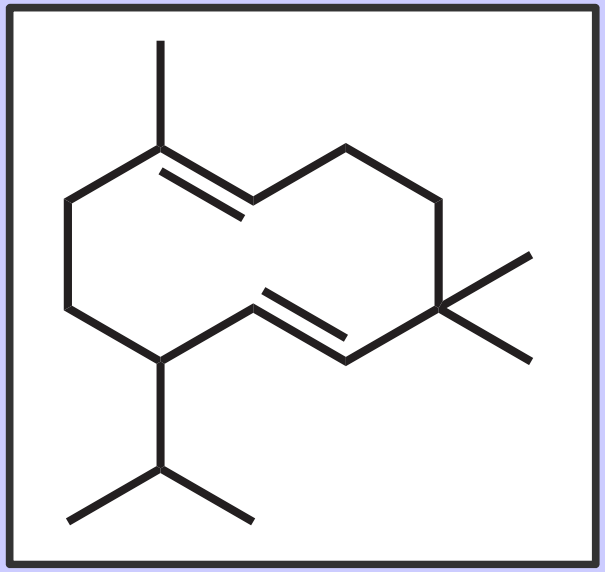
A few solved problems... (2)



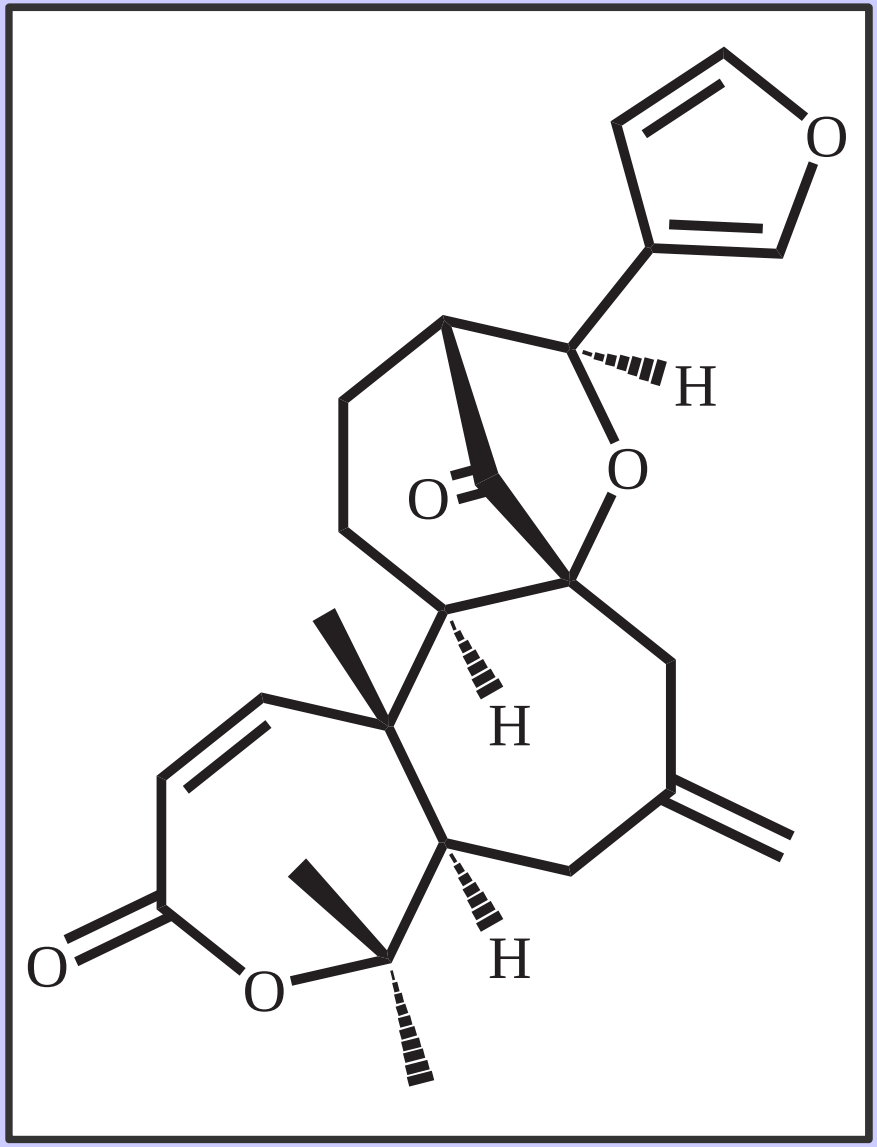


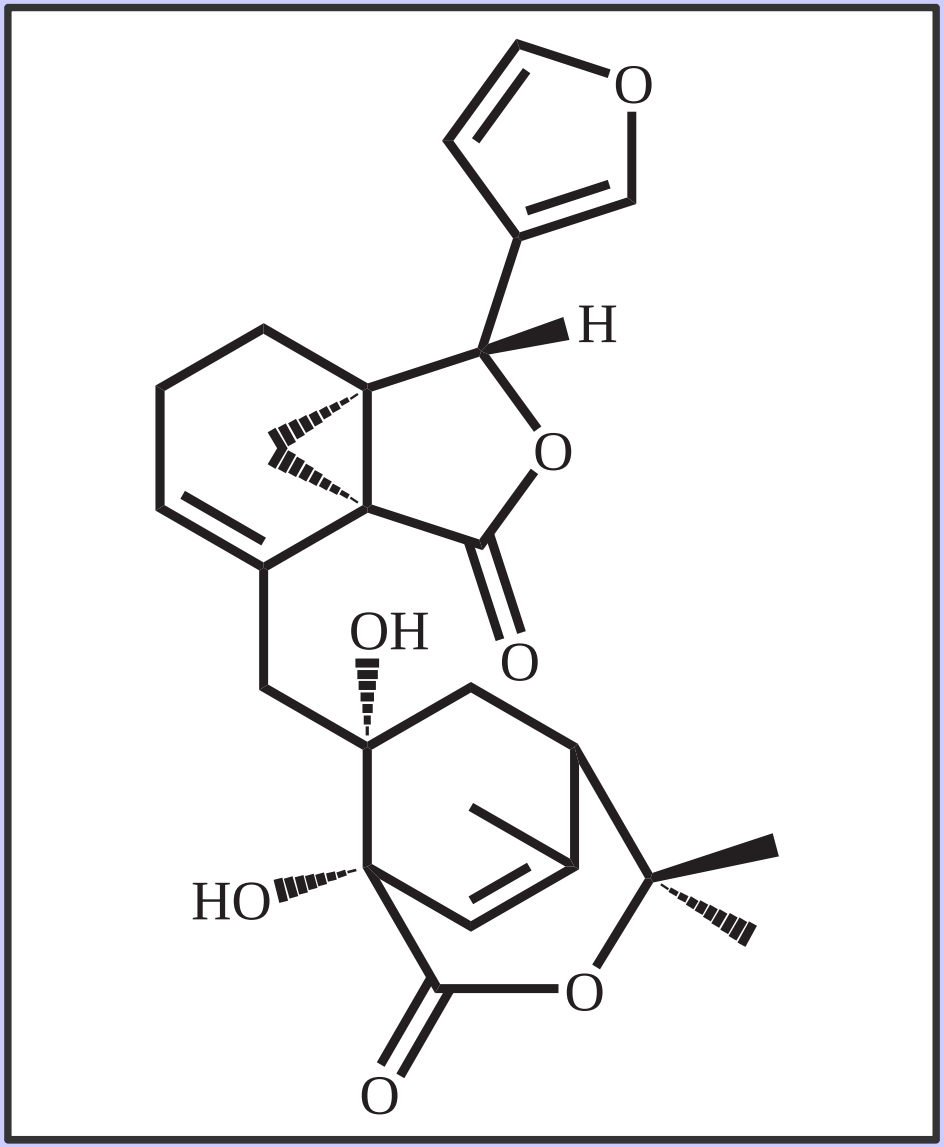
A few solved problems... (3)



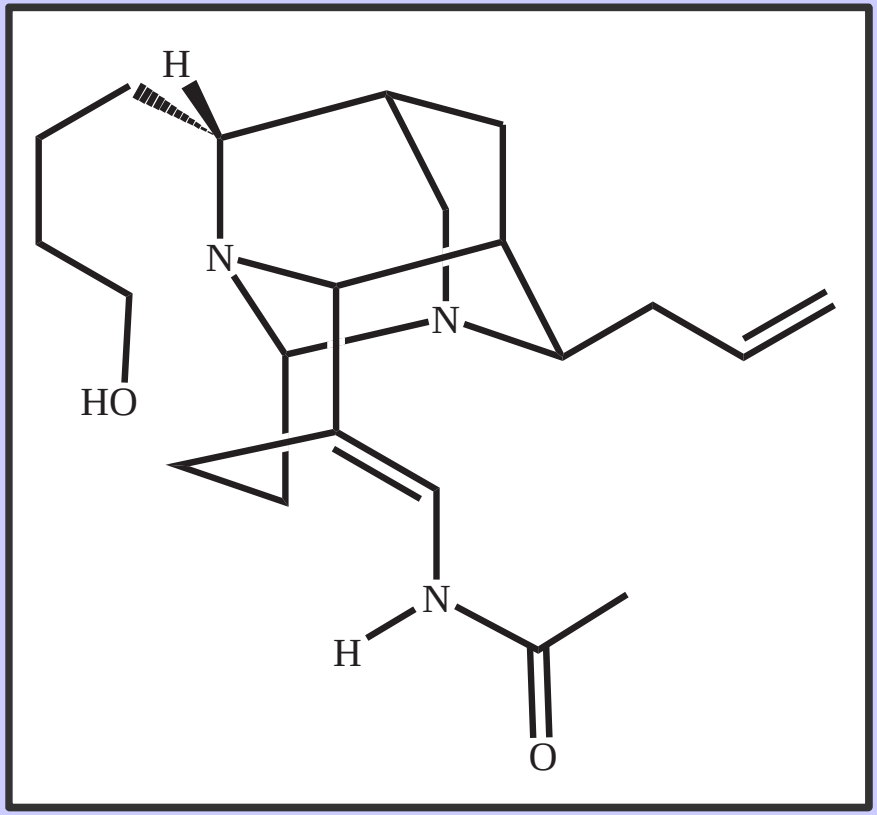


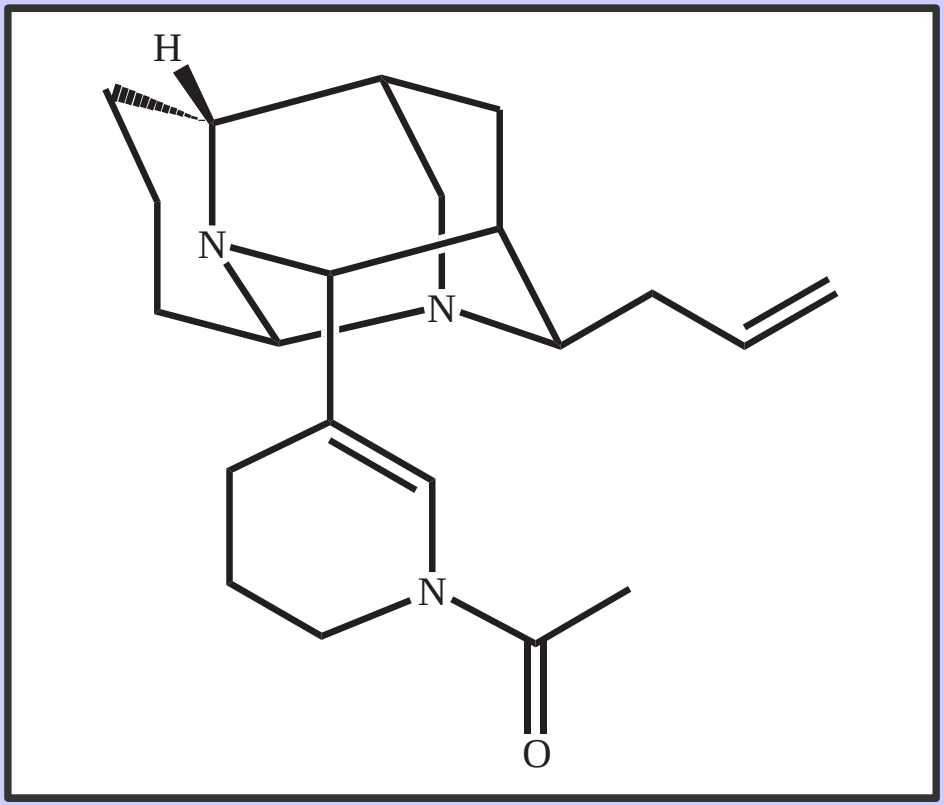
A few solved problems... (4)



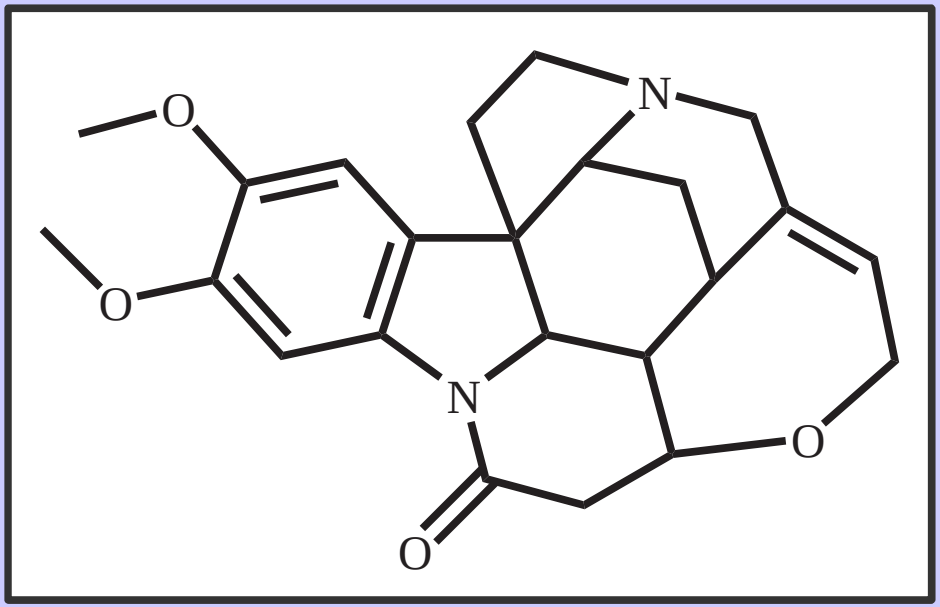


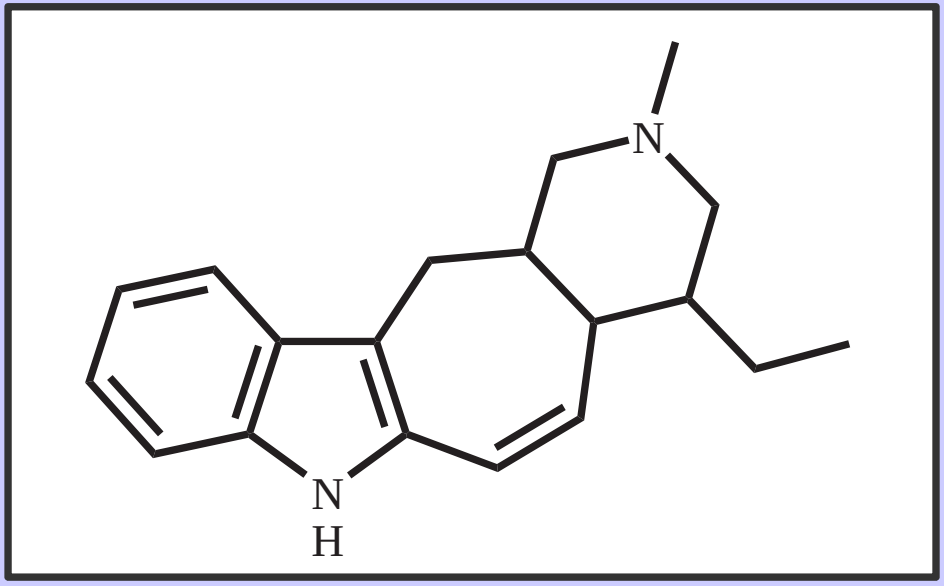
A few solved problems... (5)



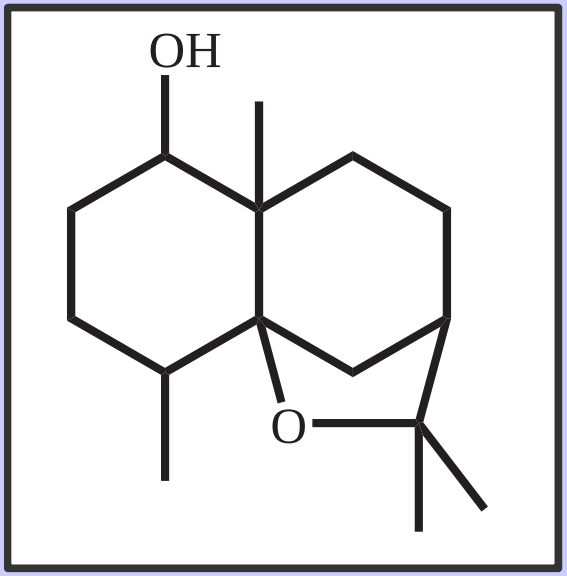


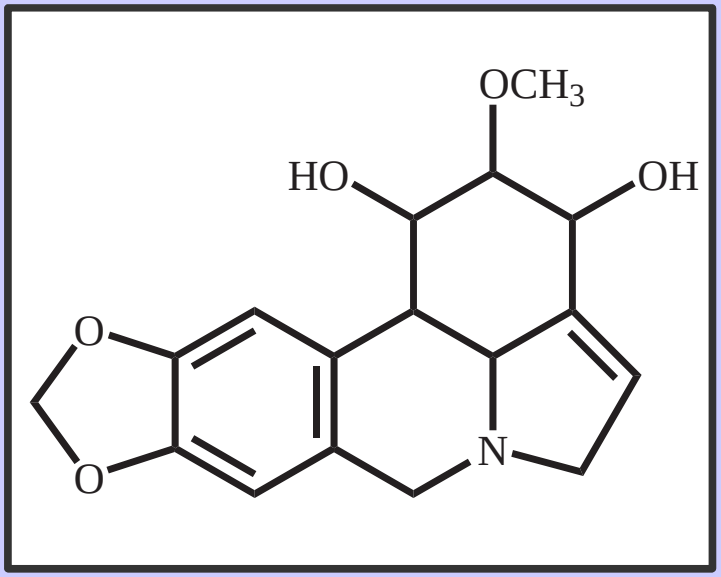
A few solved problems... (6)

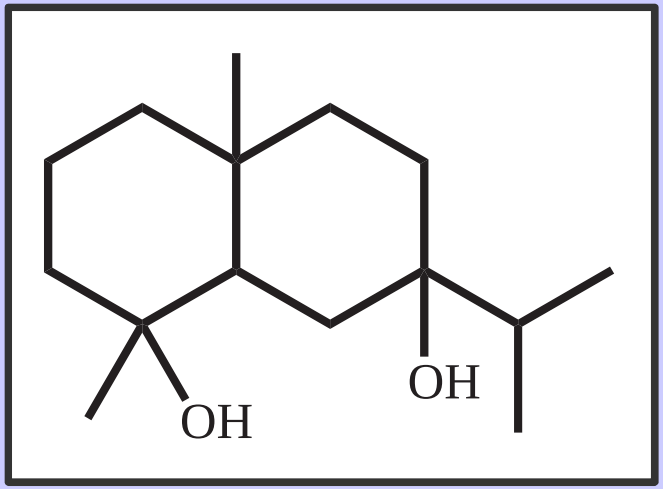


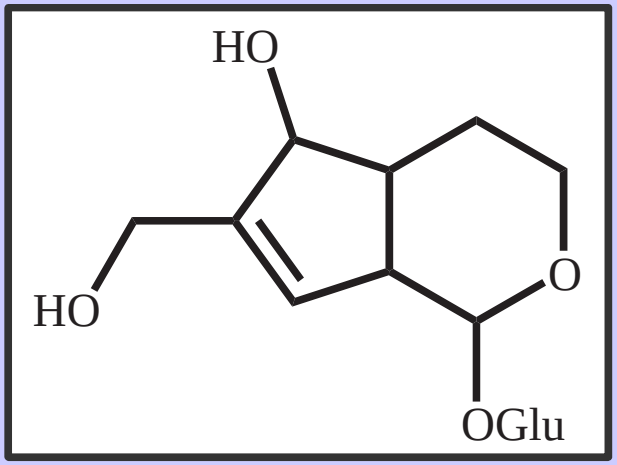


A few solved problems... (7)

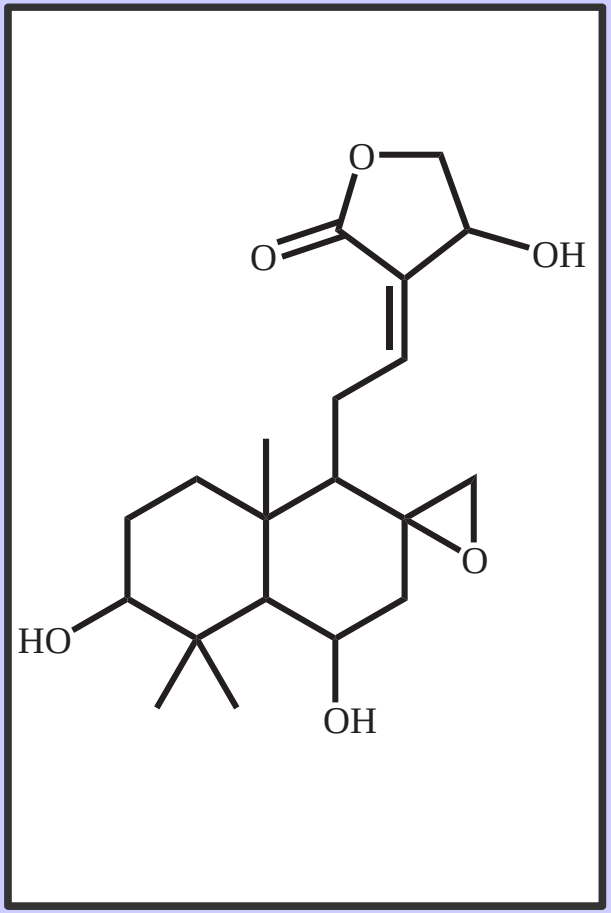


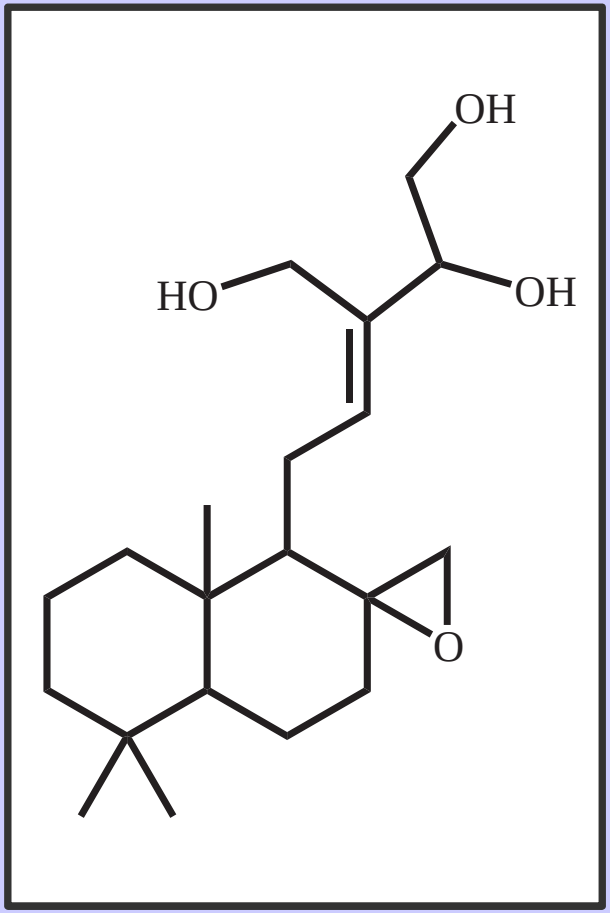


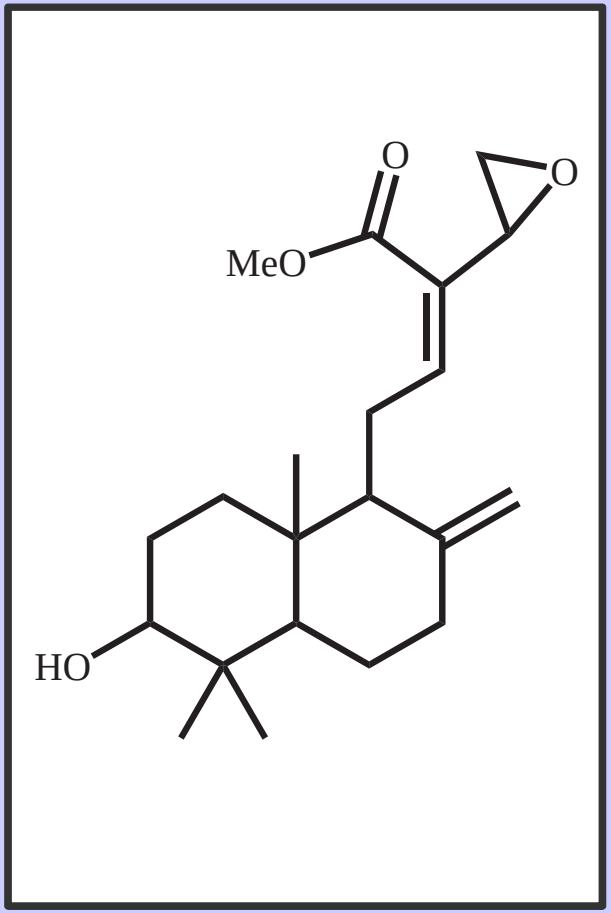




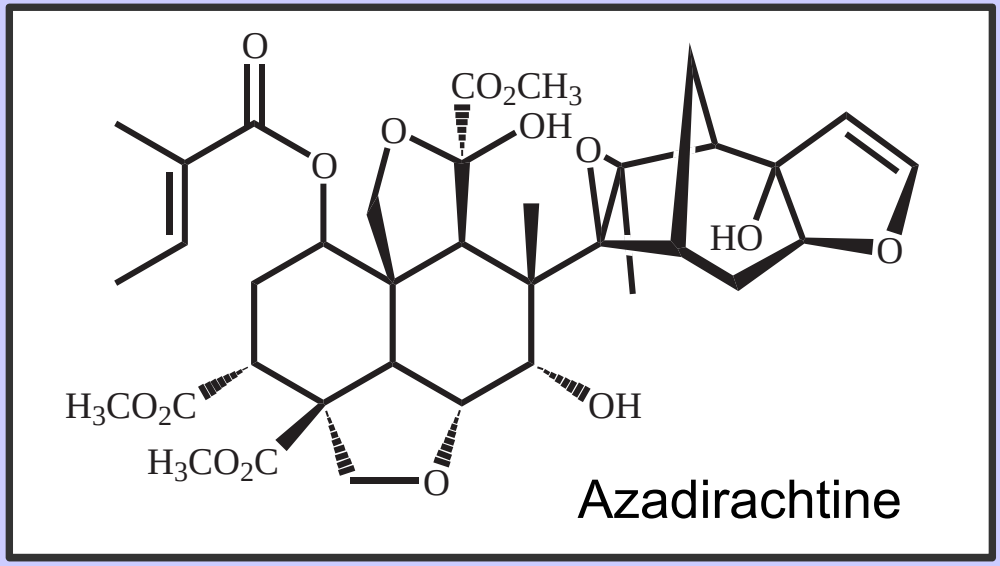
A few solved problems... (8)



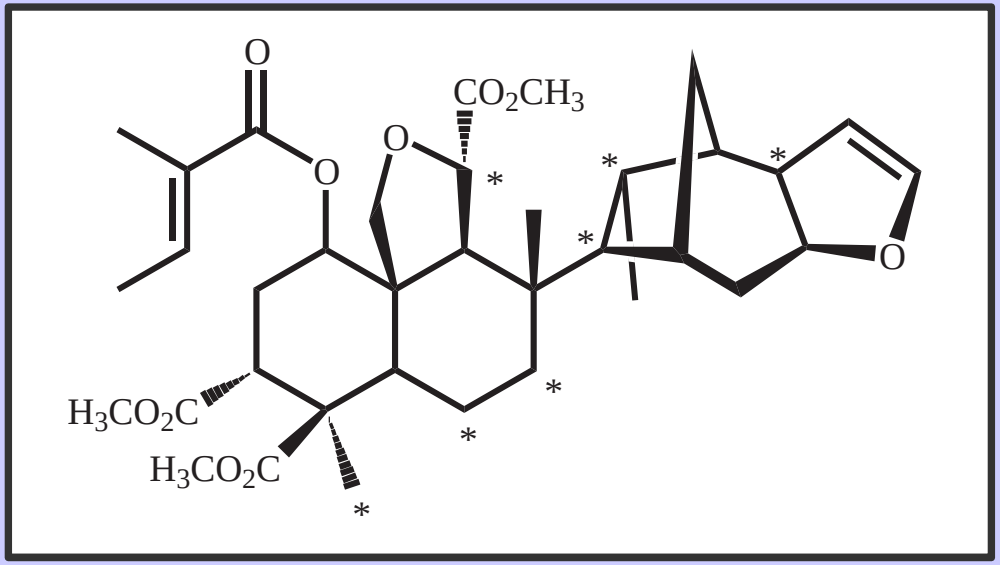




A few solved problems... (9)



Azadirachtine



The future

- **Data input interface**
  - Pavel Kessler (Bruker) wrote such an interface
- **Struture generator**
  - Non-rigid atom status
  - Connection with a chemical shift database
  - Simultaneously use substructure information and NMR data
- **Solution output interface**
  - Better 2D drawings needed...
- **Ultimate goal**
  - ***Do you want to look at the spectra or at the structure ?***

The feedback of a user

- **“I would say the real value of the LSD program in such cases is its ability to propose structures consistent with the NMR data without prior bias or expectations that us humans can have.**
- **When piecing together structures based principally on HMBC data it is possible to arrive at a structure that is consistent with the data, but it is often difficult then to think of others that could also be considered as the correct structure....**

- **... 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 »

How to get LSD

- **LSD is a free software, under GPL licence.**
- **www.univ-reims.fr/LSD**
- **Available for :**
  - Linux (recommended)
  - Windows through cygwin (obsolete)
  - Windows (for DOS indeed, but is OK)
  - Mac OS X (maybe better to recompile from sources)