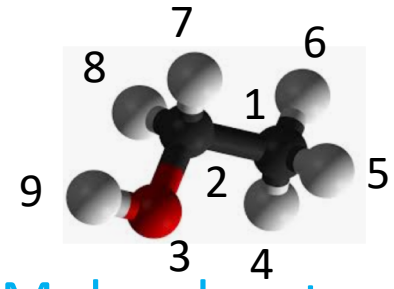
NMReDATA file validation through Computer-Assisted Structure Elucidation

1st NMReDATA Symposium

Porto, 26 September 2019

Jean-Marc Nuzillard and Stefan Kuhn

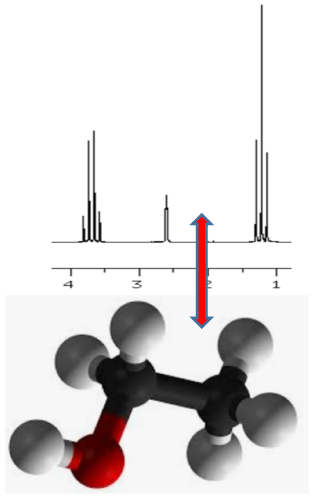
NMReDATA files




```
702
-Xxxxxx-09061911492D

9  8  0     0  0  0  0  0  0  0999 V2000
    3.7320    0.2500    0.0000 O   0  0
    2.8660   -0.2500    0.0000 C   0  0
    2.0000    0.2500    0.0000 C   0  0
    2.4675   -0.7249    0.0000 H   0  0
    3.2646   -0.7249    0.0000 H   0  0
    2.3100    0.7869    0.0000 H   0  0
    1.4631    0.5600    0.0000 H   0  0
    1.6900   -0.2869    0.0000 H   0  0
    4.2690   -0.0600    0.0000 H   0  0
```

Molecular structure

Valid?



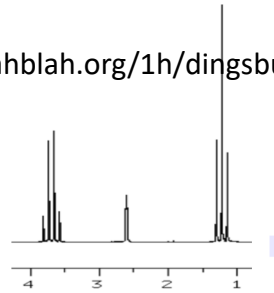
Interpretation

A: 7, 8
B: 9
C: 4, 5, 6

Coherent?

Links to NMR spectra

www.blahblah.org/1h/dingsbums.spec




```
##DATA TABLE=(X++(R..R)), XYDATA
0.000000  8210.441210 15319.203860
0.001200 -3579.394247 -6391.616387
0.002400 -8371.959656 -5958.791078
0.003600  4750.645813  3990.972917
0.004800  -822.396323  -719.694050
0.006000  1951.423817  3084.714990
```

A: δ 3.687; *q*, *J* = 7.1; 2H
B: δ 2.61, *bs*, 1H
C: 1.226; *q*, *J* = 7.1; 3H

Correct?

Description of NMR spectra by spectrum parameter sets

Validate a description

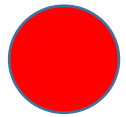
- Q: Is the spectrum description coherent with the spectra?
  - A structure proposal is even not needed to answer this question (but helps!)
  - Generate description again (change method) and compare descriptions
    - Spectrum simulation from NMR parameters
    - Iterative refinement of NMR parameters
    - Not so easy
  - Generate a spectrum from a description and compare spectra
    - Spectrum simulation from NMR parameters
    - Requires a full set of parameters

Validate a structure

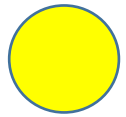
- Q: Does the proposed structure fit with experimental and *a priori* data?
  - Generate descriptions from structures and compare descriptions
    - Database / QM prediction of descriptor values
    - Automatic structure validation
  - Generate spectra from descriptions from structures and compare spectra
    - The same as above, but with spectrum calculation from description
  - Generate structure from description
    - *De novo* structure elucidation, Computer-Assisted Structure Elucidation (CASE).

CASE software, for structure validation purpose

- No structure produced or the expected structure is absent
  - Bad description, or bad structure, or both…



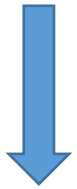
- The expected structure is found as well as other ones
  - Good, try to eliminate alternatives



- The expected structure is the only one to be produced
  - Good for the 2D structure, but other validation criteria (3D geometry) must be considered.

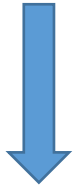
A simple workflow, as already mentioned by Stefan

NMReDATA file



NMReDATA **Javatools**, Stefan Kuhn
github.com/NMReDATAInitiative/javatools

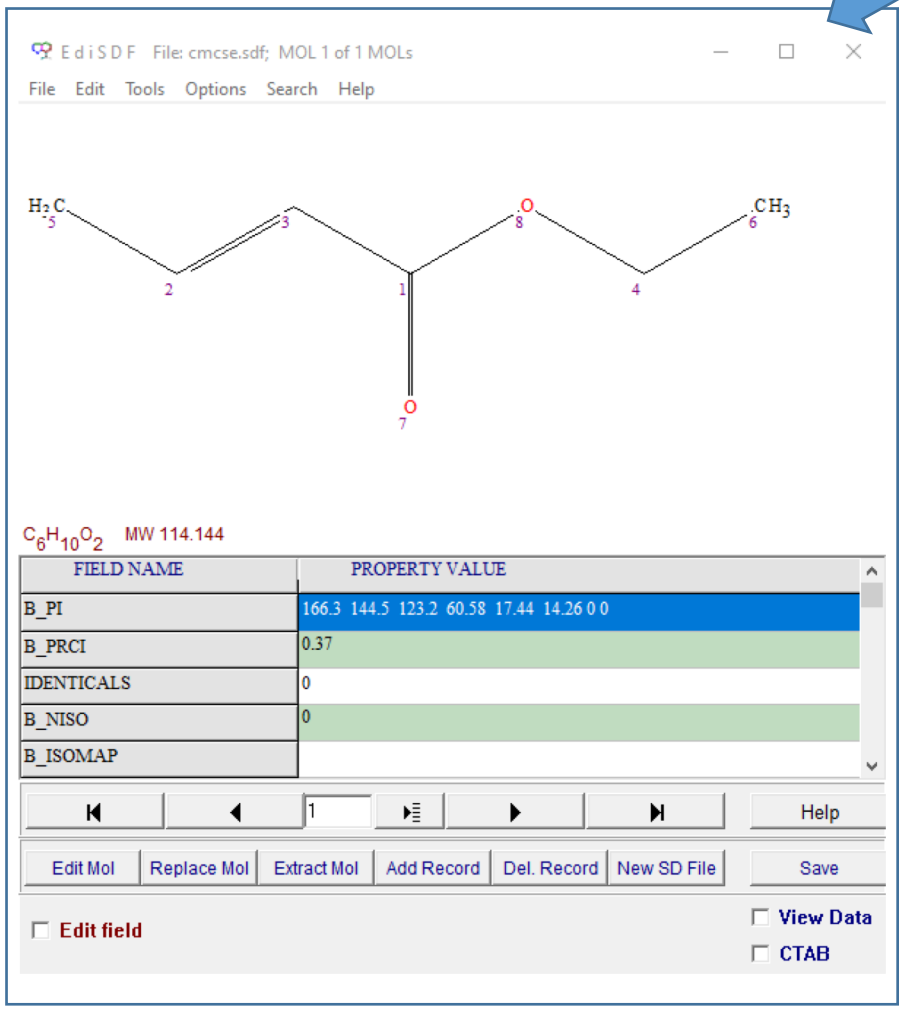
Input file for LSD
CASE software



LSD, Jean-Marc Nuzillard
www.univ-reims.fr/LSD

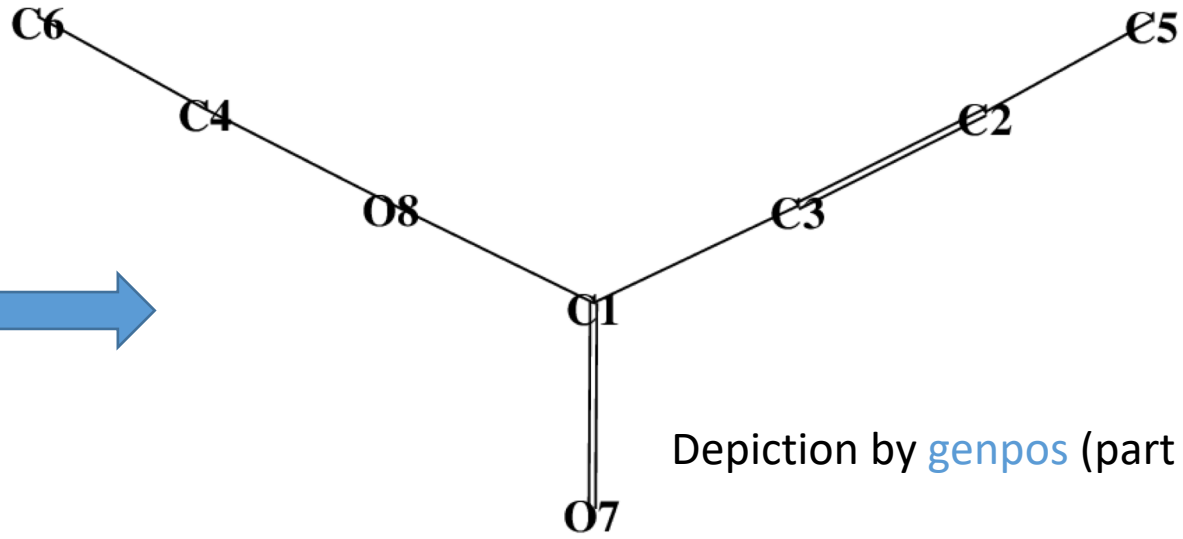
Solution(s)

Example 1



Thanks to **EdiSDF**!

- NMReDATA file, from Bruker CMC-se, generated by the CDK
- Structure elucidation by the LSD software
- SDG by outlsd (part of LSD)



Depiction by genpos (part of LSD)

LSD: J.-M. Nuzillard and G. Massiot, *Tetrahedron* **1991**, *47*, 3655-3664.
Initial development in 1989, 30 YEARS AGO

NMReDATA file, Structure

```
  CDK    1128171657

  8  7  0  0  0  0  0  0  0  0999 V2000
    3.8971   -0.7500    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.2990   -0.7500    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.5981   -0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    6.4952   -0.7500    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    7.7942   -0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    3.8971   -2.2500    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    5.1962   -0.0000    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  6  4  1  0  0  0  0
  5  2  1  0  0  0  0
  2  3  2  0  0  0  0
  4  8  1  0  0  0  0
  3  1  1  0  0  0  0
  1  8  1  0  0  0  0
  1  7  2  0  0  0  0
M  END
```

1
2
3
4
5
6
7
8

Implicit atom #

No explicit H atom!

NMReDATA file, Assignment

```
> <NMREDATA_ASSIGNMENT>
c1, 166.603, 1
c2, 144.493, 2
c3, 122.699, 3
c4, 60.151, 4
c5, 17.980, 5
c6, 14.166, 6
h1,  6.949, H2
h2,  5.844, H3
h3,  4.177, H4
h4,  1.854, H5
h5,  1.255, H6
```

Signal identifier, chemical shift value, **atom identifier**

NMReDATA file, Description (1D)

> <NMREDATA_1D_1H>
Larmor=500.130
Spectrum_Location=file:D:/SciPrograms/Bruker/TopSpin3.5.b.91pl7/data/nes/nmr/nesEX9_CMCse_Test/1/pdata/1
Pulseprogram=zg30
6.94928, L=h1
5.84414, L=h2
4.17706, L=h3
1.85439, L=h4
1.25499, L=h5

> <NMREDATA_1D_13C>
Larmor=125.758
Spectrum_Location=file:D:/SciPrograms/Bruker/TopSpin3.5.b.91pl7/data/nes/nmr/nesEX9_CMCse_Test/2/pdata/1
Pulseprogram=zgdc30
166.603, L=c1
144.493, L=c2
122.699, L=c3
60.1508, L=c4
17.9798, L=c5
14.1659, L=c6

NMReDATA file, Description (HSQC and COSY)

> <NMREDATA_2D_13C_1J_1H>
Larmor=500.130
CorType=HSQC
Spectrum_Location=file:D:/SciPrograms/Bruker/TopSpin3.5.b.91pl7/data/nes/nmr/nesEX9_CMCse_Test/3/pdata/1
Pulseprogram=hsqcg
c2/h1
c3/h2
c4/h3
c5/h4
c6/h5

> <NMREDATA_2D_1H_NJ_1H>
Larmor=500.130
CorType=COSY
Spectrum_Location=file:D:/SciPrograms/Bruker/TopSpin3.5.b.91pl7/data/nes/nmr/nesEX9_CMCse_Test/4/pdata/1
Pulseprogram=cosygpqf
h1/h2, W2=20.00, W1=20.00
h1/h4, W2=20.00, W1=20.00
h3/h5, W2=20.00, W1=20.00

NMReDATA file, Description (HMBC)

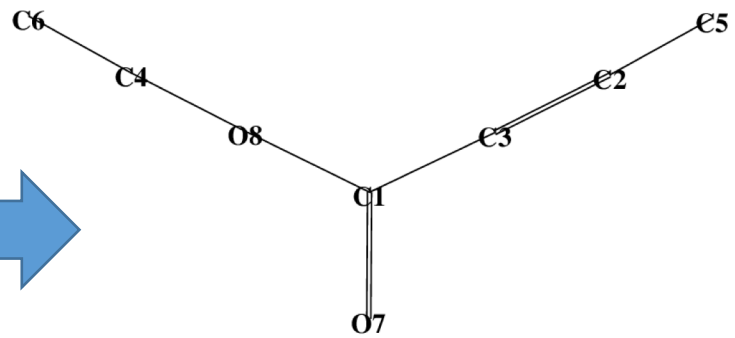
> <NMREDATA_2D_13C_NJ_1H>
Larmor=500.130
CorType=HMBC
Spectrum_Location=file:D:/SciPrograms/Bruker/TopSpin3.5.b.91pl7/data/nes/nmr/nesEX9_CMCse_Test/6/pdata/1
Pulseprogram=hmbcgplpndqf
c1/h1, E=6.70107e+09, W2=98.36, W1=418.65
c1/h3, W2=20.00, W1=80.00
c2/h4, W2=20.00, W1=80.00
c3/h4, W2=20.00, W1=80.00
c4/h5, W2=20.00, W1=80.00
c6/h3, W2=20.00, W1=80.00
c5/h2, W2=20.00, W1=80.00
c5/h1, W2=20.00, W1=80.00

Input file for LSD

```
; file generated by NMReDATA javatools
; Fri Aug 16 16:56:22 BST 2019
MULT 1 C 2 0 0
MULT 2 C 2 1 0
MULT 3 C 2 1 0
MULT 4 C 3 2 0
MULT 5 C 3 3 0
MULT 6 C 3 3 0
MULT 7 O 2 0 0
MULT 8 O 3 0 0

SHIH 6 1.255
SHIH 5 1.854
SHIH 4 4.177
SHIH 3 5.844
SHIH 2 6.949

SHIX 6 14.166
SHIX 5 17.98
SHIX 4 60.151
SHIX 3 122.699
SHIX 2 144.493
SHIX 1 166.603
```

Unused!

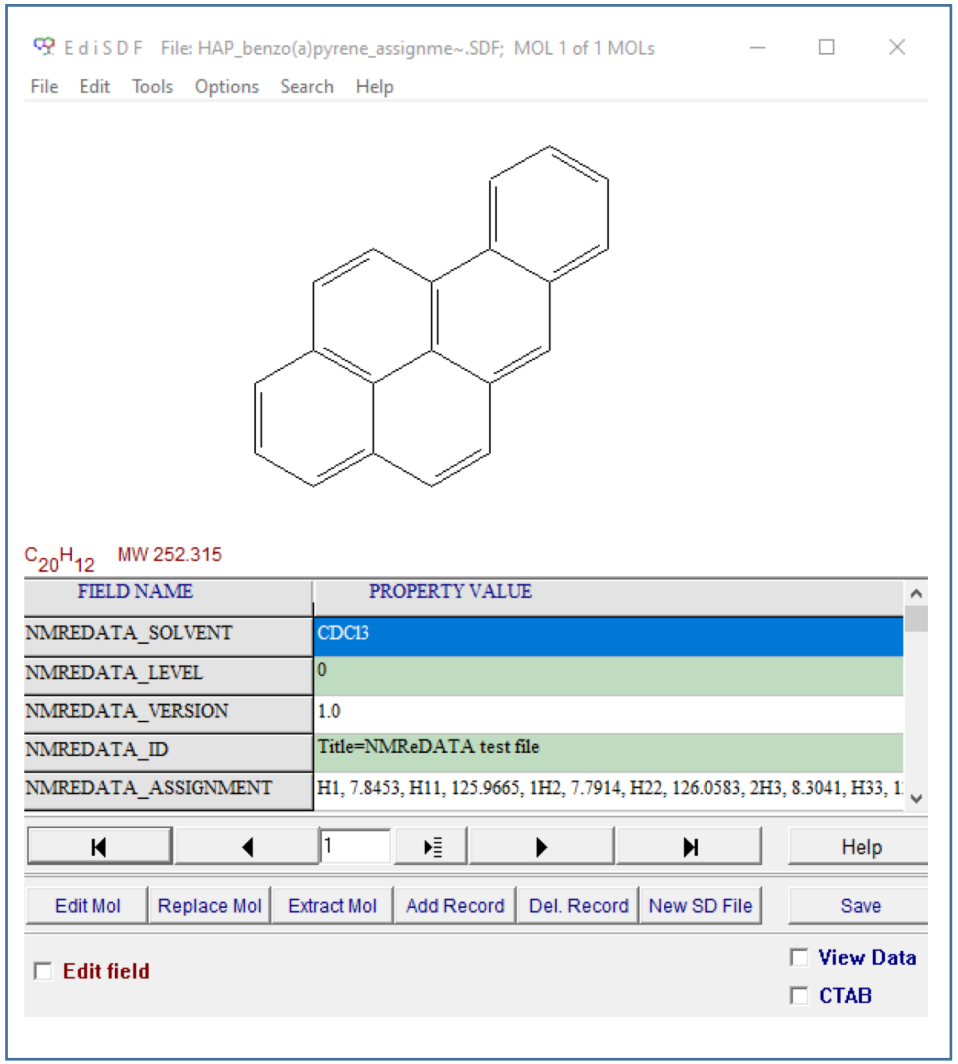
```
COSY 2 3
COSY 2 5
COSY 4 6

HSQC 2 2
HSQC 3 3
HSQC 4 4
HSQC 5 5
HSQC 6 6

HMBC 1 2
HMBC 1 4
HMBC 2 5
HMBC 3 5
HMBC 4 6
HMBC 6 4
HMBC 5 3
HMBC 5 2
```

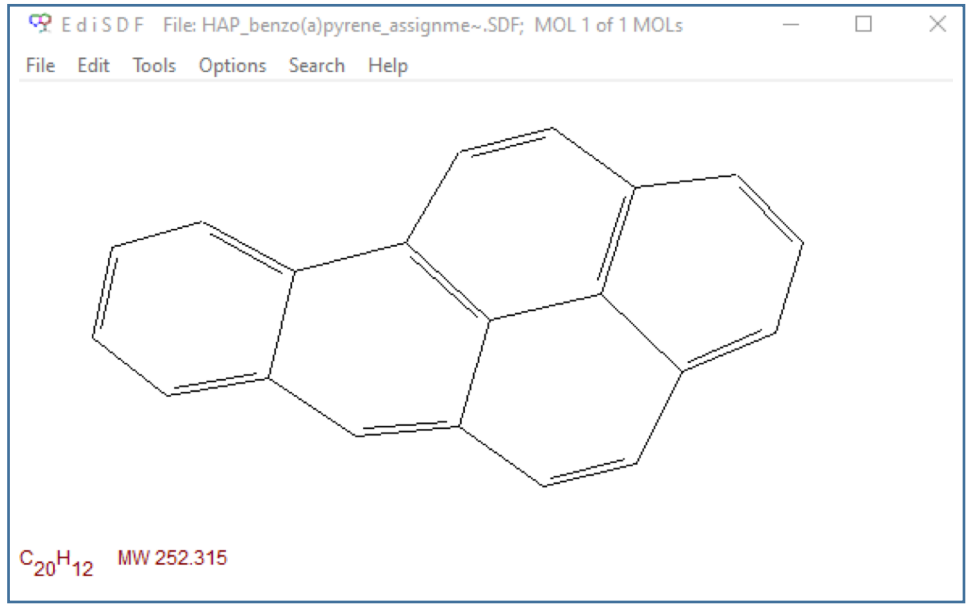


LSD produces a single structure out of this input file

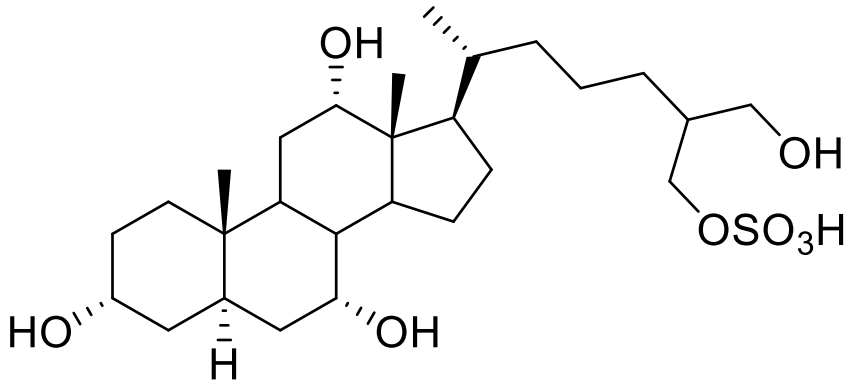
Example 2

- Data from Damien Jeannerat
- NMReDATA file from Mnova
- Structure generation by the LSD software. 1 solution.
- SDG by outlsd (part of LSD)





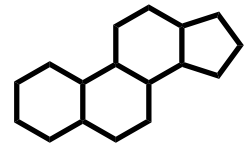
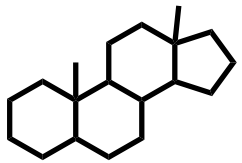
Example 3



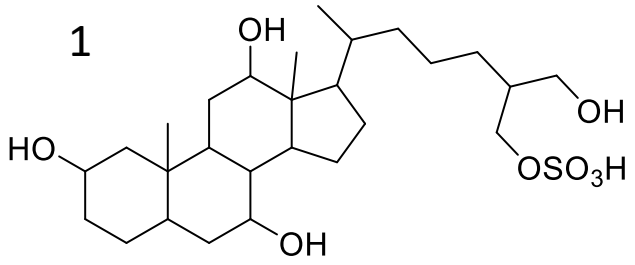
- 5α-Cyprinol sulfate
- MRC 2018, 56, 1201–1207
- Data from Nils Schlörer

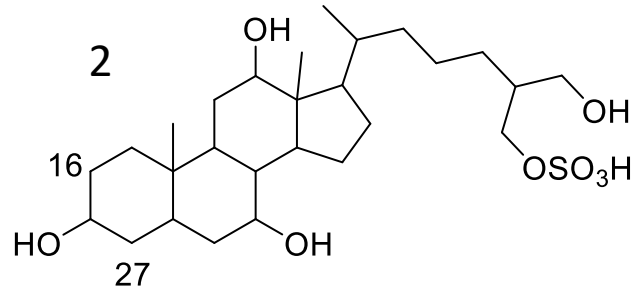
- Planar structure validation.
- NMReDATA →lsd input file
- Manual edition before running lsd

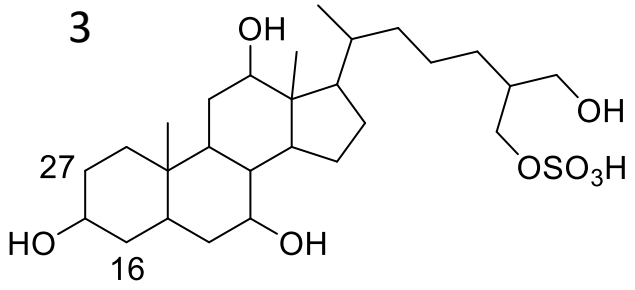
Example 3, approach 1

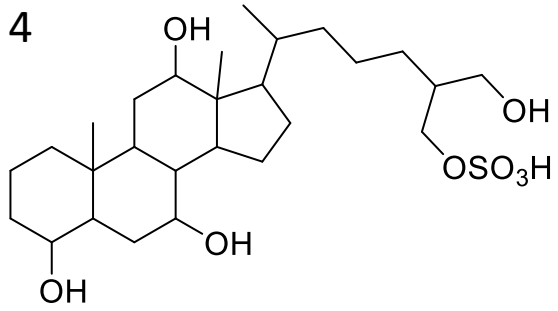
- 2D NMR data from HSQC and HMBC only, no COSY, no H2BC
- “Give” the sulfate group
- Allow for 1 ^{4}J HMBC correlation
  - → 5325 solutions
- Indicate the tetracyclic ring system as a sub-structure 
  - → 140 solutions
- 5 carbon atoms, from 61.7 to 72.6 ppm, are bound to 1 oxygen atom
  - → 50 solutions
- Indicate the 2 methyl groups attached to the tetracyclic ring system 
  - → 20 solutions
- Attach the sulfate group at one end of the side-chain (2 CH_2, 62 and 68 ppm)
  - → 4 solutions

Example 3, approach 1, 4 solutions









- The sulfate group can be attached anywhere (5 positions)
- Structures 2 and 3 and are the most likely ones (assignment isomers)

Example 3, approach 2

- 2D NMR data from HSQC and HMBC only, no COSY, no H2BC
- “Give” the sulfate group
- Allow for 1 ^{4}J HMBC correlation
  - → 5325 solutions, as for approach 1.

- Rank solutions according to similarity between predicted and experimental ^{13}C chemical shifts.
- Prediction uses nmrshiftdb
- PyLSD
  - → The solution ranked first is the published one for 5α-cyprinol sulfate

Conclusion

- NMReDATA Javatools transforms NMReDATA files into LSD input files
- LSD searches for 2D alternatives to the structure in the NMReDATA file
- Nmrshiftdb provides chemical shift prediction for structure ranking

- NMReDATA Javatools, LSD, nmrshiftdb are free software
- NMReDATA file validation through computer-assisted structure elucidation is possible but still may require human intervention
- Software bricks are there but still need some integration to improve user-friendliness.