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/](http://nmredata.org/)
• 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,...

  – 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?

  – Structures in Journals are **not readable** by computers

![Chemical Structure](image)

– Spectrum descriptions are **not readable** by computers

<table>
<thead>
<tr>
<th>Atom no.</th>
<th>$^{13}$C</th>
<th>$^1$H</th>
<th>COSY</th>
<th>HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>133.8</td>
<td>9.06 br s</td>
<td></td>
<td>2, 7, 8, 13</td>
</tr>
<tr>
<td>2</td>
<td>53.1</td>
<td>4.41 br s</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>51.5</td>
<td>a: 3.30 br d (11)(^f)</td>
<td>5b, 6a, 6b(^f)</td>
<td>2, 15(^f)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>b: 3.18 br d (11)(^f)</td>
<td>5a, 6a, 6b</td>
<td>3, 7</td>
</tr>
<tr>
<td>5</td>
<td>18.1</td>
<td>a: 3.01 t (15)(^f)</td>
<td>5a, 5b, 6b</td>
<td>2', 5', 7(^f)</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>b: 2.65 dd (15.7, 4.7)</td>
<td>5a(^f), 5b, 6a</td>
<td>2, 5', 7</td>
</tr>
<tr>
<td>7</td>
<td>108.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>128.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>118.7</td>
<td>7.47 d (7.5)</td>
<td>10</td>
<td>7, 8', 11, 13</td>
</tr>
<tr>
<td>10</td>
<td>119.9</td>
<td>7.07 t (7.3)</td>
<td>9, 11</td>
<td>8, 9, 12</td>
</tr>
<tr>
<td>11</td>
<td>122.0</td>
<td>7.11 t (7.3)</td>
<td>10, 12</td>
<td>9, 13</td>
</tr>
<tr>
<td>12</td>
<td>111.4</td>
<td>7.38 d (7.7)</td>
<td>11</td>
<td>8, 10</td>
</tr>
</tbody>
</table>
• 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.

  – 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
• **Structure determination of pure compounds**

  – Obtain an elemental formula \((C_cH_hN_nO_o\ldots)\)
  
  – 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](http://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
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.
- 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 $^{13}\text{C}$ NMR chemical shift predictor (S. Kuhn, C. Steinbeck)

- A score $\Delta$ measures the distance between the experimental and predicted chemical shift series.

$$\Delta = \frac{\sum_{i=1}^{N} |\delta_{i}^{\text{calc}} - \delta_{i}^{\text{exp}}|}{N}$$
EXAMPLE, SOLUTION RANKING

- Ranking:
  - 1. $\Delta = 1.40$ ppm
  - 2. $\Delta = 2.03$ ppm
  - 3. $\Delta = 2.13$ 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 know 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).
• 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


  – Coming soon: a review article, new solvents, new methods.
ACKNOWLEDGEMENTS

- Dr Bertrand Plainchont
- Dr Jane Hubert
- Pr Jean-Hugues Renault
- Romain Reynaud (Soliance-Givaudan)