100 A cross-platform format to associate NMR-extracted data (NMReDATA) to chemical structures

Julien Wist[1], Jean-Marc Nuzillard[2], Nils E. Schloerer[3], Stefan Kuhn[3], Marion Pupier[4], Mate Erdelyi[5], Christoph Steinbeck[6], Patrick Giraudeau[7], Roberto R. Gil[8], Antony J. Williams[9], Craig Butts[10], Kessler Pavel[11], Fabrice Moriaud[11], Mikhail Elyashberg[12], Dimitris Argyropoulos[12], Manuel Perez[13], Gary Martin[14], Paul Trevorrow[15], and **Damien Jeannerat**[4]

1. University of Valle, Cali, Colombia
2. CNRS and University of Reims-Champagne-Ardenne, Reims, France
3. IDNMR, University of Cologne, Cologne, Germany
4. Department of organic chemistry, University of Geneva, Geneva, Switzerland
5. Department of Chemistry and Molecular Biology, University of Gothenburg, Sweden
6. Cheminformatics and Chemometrics, Friedrich-Schiller-University Jena, Germany
7. CEISAM, University of Nantes, Nantes, France
8. Department of Chemistry, Carnegie Mellon University, Pittsburgh, USA
9. National Center for Computational Toxicology, EPA, USA
10. Craig Butts, University of Bristol, United Kingdom
11. Bruker BioSpin GmbH, Rheinstetten, Germany
12. ACD/Labs, Toronto, Canada
13. Mestrelab Research, Santiago de Compostela, Spain
14. Merck and co. New-Jersey, USA
15. Magnetic Resonance in Chemistry, Wiley, Chichester, UK

An open initiative involving cross-disciplinary contributors of computer-assisted structure elucidation (CASE), including methodology specialists, software and database developers and the editorial board of Magnetic Resonance in Chemistry, is addressing the old problem of reporting and sharing the assignment of 1D and 2D NMR spectra of organic molecules. [1]

Our approach aims to solve some of the problems encountered with the “full analysis” of organic compounds. Usually, they are reported in chemistry journals using an image of the chemical structure, a text-based assignment of the 1D 1H and 13C spectra and a set of tables listing the correlations found in 2D spectra such as COSY, HSQC and HMBC. In the best case, images of the spectra (of uneven quality and resolution) can be found as Supplementary material. This is unsatisfactory.[2,3]

We introduced a data format to associate the data extracted from the spectra and the structure of the identified compound. The file uses the SD format, a type of files that is compatible with .mol files (a quite commonly format used to draw chemical structures). The NMR-extracted data (e.g., chemical shift, coupling and assignment) are encoded as so-called “tags”, that are included in the .sdf files. These “tags” are not visible when displaying the molecules but can be accessed by specialized software such as CASE software and analyzed by the database during the importation of the data. These .sdf files including NMReDATA will be generated by future releases of computer-assisted structure elucidation software and have multiple roles:

1) They make the link between the atoms of the structure and the signals found in the spectra (assignment).
2) They list, for each 1D and 2D spectrum, the spectral parameters in a defined format (chemical shifts, couplings, integral, 2D correlations)
3) They combine the data extracted from the spectra into an aggregated table (list of chemical shifts, coupling network, etc.).
4) They include links from the spectral data to the original files of the spectra (located in a local files folder in a database).

These files will be uploaded on a database together with the associated spectra as embargoed NMR data. A link to the data will be included in the manuscript submitted to scientific journals. Reviews will be facilitated by the fact that the spectra, the extracted data and the structure will be accessible in a usable electronic format. The reviewers will use their favorite CASE software to assess the assignment. Once the paper is accepted, the spectra and the extracted spectral data become openly accessible to everybody.

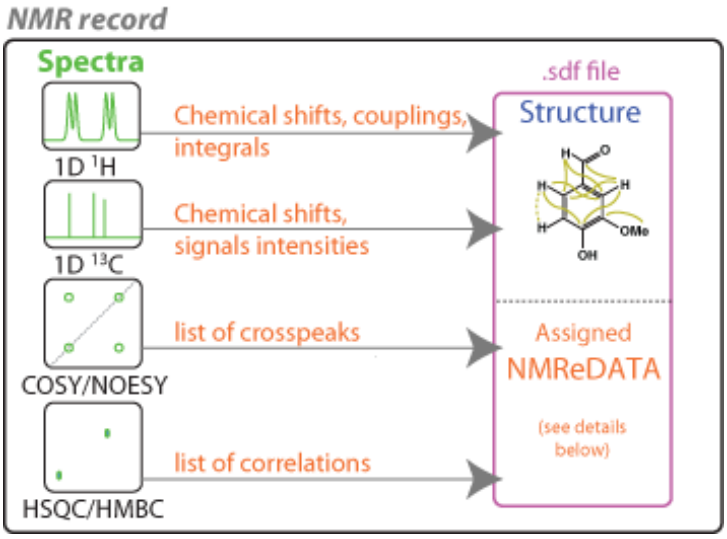

Fig. 1 Schematic representation of an NMR record including the files of the NMR spectra, and the SDF file containing the structure and the assigned NMR data extracted from the spectra.

1. www.nmredata.org
2. Jeannerat, D. Magn. Reson. Chem. 55(1) 7-14, 2017
3. Castillo, A. M., Bernal, A., Dieden, R., Patiny, L., Wist, J., J. Cheminform., 8(1), 26, 2016