**A cross-platform format to associate**

**NMR-extracted data (NMReDATA)**

**to chemical structures**

Jean-Marc Nuzillard1, Nils E. Schloerer2, Stefan Kuhn2, Patrick Giraudeau3,

Roberto R. Gil4, Julien Wist5, Kessler Pavel6, Manuel Perez7,

Paul Trevorrow8, Marion Pupier9, Damien Jeannerat9

1CNRS and University of Reims-Champagne-Ardenne, Reims, France, 2Department of Chemistry, University of Cologne, Cologne, Germany, 3CEISAM, University of Nantes, Nantes, France, 4Department of Chemistry, Carnegie Mellon University, Pittsburgh, USA, 5University of Valle, Cali, Colombia, 6Bruker BioSpin GmbH, Rheinstetten, Germany, 7Mestrelab Research, Santiago de Compostela, Spain, 8Wiley, Chichester, United Kingdom, 9Department of organic chemistry, University of Geneva, Geneva, Switzerland

An open initiative involving some of the major players 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.

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 many reasons.[1]

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 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 can in principle be generated by any computer-assisted structure elucidation routine, available in most NMR related 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. The reviewing will be facilitated by the fact that the spectra, the extracted data and 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.

|  |
| --- |
| Macintosh HD:Users:jeannerat:Dropbox:mrc_working_group:euromar_poster:fig_abs.png |
| ***Fig. 1*** *Schematic representation of the .sdf file containing a structure and the extracted NMR data (top) and the link to the files containing the NMR spectra (bottom).* |

Jeannerat, D. Human- and computer-accessible 2D correlation data for a more reliable structure determination of organic compounds. Future roles of researchers, software developers, spectrometer managers, journal editors, reviewers, publisher and database managers toward artificial-intelligence analysis of NMR spectra**.** *Magn Reson Chem. 55(1) 7-14, 2017.*

2 Castillo, A. M., Bernal, A., Dieden, R., Patiny, L., & Wist, J. “Ask Ernö”: a self-learning tool for assignment and prediction of nuclear magnetic resonance spectra. *J. Cheminform., 8(1), 26,* *2016.*