NMR Prediction, Calculation and Molecular Interpretation Complexity

Sunilkumar Ashok,[a*] Thomas Cauchy,[a] Arnaud Fihey,[b] Benoit Da Mota [c], Narcis Avarvari,[a]

a. Univ Angers, CNRS, MOLTECH-ANJOU, SFR MATRIX, F-49000 Angers. b. Institut des Sciences Chimiques de Rennes, UMR CNRS 6226, Université de Rennes, Campus de Beaulieu, Rennes, France. c. Univ Angers, LERIA, SFR MATRIX, F-49000 Angers, France

**Email : sunilkumar.ashok@univ-angers.fr*

Nuclear Magnetic Resonance spectroscopy is a central tool for molecular structure elucidation. While the forward problem of predicting chemical shifts from a known molecular structure is well studied, the inverse question remains insufficiently quantified.[1,2] To what extent does a given set of ^1H and ^{13}C chemical shifts constrain the underlying molecular graph? The objective of this study is to quantify the structural information content and intrinsic ambiguity of experimental NMR data using curated entries from NMR chemical shifts databases.[3] Atomic environments are represented through HOSE descriptors, enabling systematic analysis of how defined ppm intervals correspond to distinct local structural motifs. By evaluating this mapping across multiple spectral resolutions and structural radii, we assess how uniquely NMR chemical shifts specify atomic environments and how this specificity evolves with increasing structural context.

The analysis presents both methodological and theoretical challenges. Public NMR datasets contain duplicate records, inconsistent assignments, solvent-dependent variations and outliers that can artificially inflate structural diversity and bias ambiguity estimates. A controlled curation and validation workflow was therefore implemented, complemented by density functional theory (DFT) and density functional tight binding (DFTB) calculations to assess assignment consistency and evaluate theoretical chemical shift variability. Structurally, substantial overlap exists among chemical shift regions, rendering the inverse mapping inherently non-unique. The poster will therefore present quantitative mappings between ppm intervals and distinct HOSE-based environments, entropy-based measures of ambiguity across descriptor radii, sensitivity analyses evaluating the impact of data curation and spectral binning, and molecular-level complexity estimates derived from branching formulations and co-occurrence statistics.

Bibliography :

(1) Guan, Y.; et. al. Chem. Sci. **36**, 12 (2021), 12012–12026. https://doi.org/10.1039/D1SC03343C.
(2) Elyashberg, M.; Argyropoulos, D. Magnetic Resonance in Chemistry **7**, 59 (2021), 669–690. https://doi.org/10.1002/mrc.5115.
(3) Kuhn, S.; et al. Magnetic Resonance in Chemistry **2**, 62 (2024), 74–83. https://doi.org/10.1002/mrc.5418.