Decoding Solid-State NMR Descriptors of Group(VI)-Metal and Ligand Nuclei using Machine Learning

M. Lederbauer[1], Z. J. Berkson[1], Y. Kakiuchi[1], J. Roudin[1], M. Seidel[1], K. Jorner[1]*, C. Copéret[1]*

[1]ETH Zurich, D-CHAB, Vladimir–Prelog-Weg 1–5, 8093 Zurich

Understanding the structure-reactivity relationship in catalytic species requires spectroscopic techniques capable of accessing and rationalizing this connection. In solid-state NMR spectroscopy, chemical shift tensors (CSTs) provide valuable information about local chemical environments and their connection to frontier molecular orbitals.[1] However, current methods relying on orbital analysis to connect experimentally accessible NMR descriptors of ligands[2] and metal nuclei[3] to reactivity face limitations due to computational expense and time-consuming solid-state NMR measurements. To overcome these challenges, we employ data-driven methods to uncover trends between NMR descriptors (CST and electric field gradient tensor for quadrupolar nuclei) and electronic structure in a broader chemical space. Our goal is to develop a machine learning model that maps global and local descriptors of ligands and metal complexes[4] to the chemical shieldings of d^0 group(VI) Mo and W complexes, an important class of precatalysts for olefin conversion processes such as metathesis and oligomerization.

We generated a dataset of 29'000 penta- and hexacoordinated complexes combinatorically with a d^0 metal (Mo or W) and various ligands, with each molecule bearing a dianionic ligand (oxo, imido, sulfido, selenido), four monoanionic ligands (e.g. halogenides, alkoxides, alkyls) and up to one neutral ligand (e.g. THF). Using semiempirical geometry optimization and high-level DFT calculations, we obtained the CST values for the dataset. Featurization of the dataset involved 125 local and global descriptors, including steric and electronic parameters. Through hierarchical clustering, we reduced the feature space to 15-30 uncorrelated features suitable for regression algorithms. By training a Random Forest Regressor on a subset of the data, we achieved accurate predictions (R^2 of 0.7-0.9) for the metal and dianionic ligand isotropic chemical shieldings, with mean absolute errors of 200-400 ppm for the metal (^{95}Mo, ^{183}W) and 20-40 ppm for the ligand (^{15}N, ^{17}O).

Analysis of the dataset revealed trends consistent with previous case studies, such as the linear correlation of ^{95}Mo and ^{183}W chemical shifts for metals in the same coordination environment and the influence of ligand electronegativity on metal nuclei chemical shifts. Additionally, we uncovered unexplored trends related to the influence of metal σ- and π-bonding with the dianionic ligand and the impact of ligand steric bulk on NMR parameters. This approach demonstrates the potential of machine learning in providing novel insights into NMR-based descriptors and advancing our understanding of the relationship between NMR spectroscopy and electronic structure across a broad composition space.

[1] Christopher P. Gordon, Lukas Lätsch, Christophe Copéret, *J. Phys. Chem. Lett.* **2021**, 12, 8, 2072–2085
[2] Christopher P. Gordon, Christophe Raynaud, Richard A. Andersen, Christophe Copéret, Odile Eisenstein, *Acc. Chem. Res.* **2019**, 52, 8, 2278–2289
[3] Zachariah J. Berkson, Lukas Lätsch, Julius Hillenbrand, Alois Fürstner, Christophe Copéret, *J. Am. Chem. Soc.* **2022**, 144, 33, 15020–15025
[4] Aditya Nandy, Chenru Duan, Michael G. Taylor, Fang Liu, Adam H. Steeves, Heather J. Kulik, *Chem. Rev.* **2021**, 121, 16, 9927–10000