Beyond Periodic DFT: Machine Learning of NMR Chemical Shifts in Crystalline Amino Acids with Atomic Environment Descriptors

Misbah Iqbal[1], Kersti Hermansson[1], Jolla Kullgren[1], Shokirbek Shermukhamedov[1]

[1]*Department of Chemistry–Ångström Laboratory, Uppsala University, Box 538, 751 21 Uppsala, Sweden*

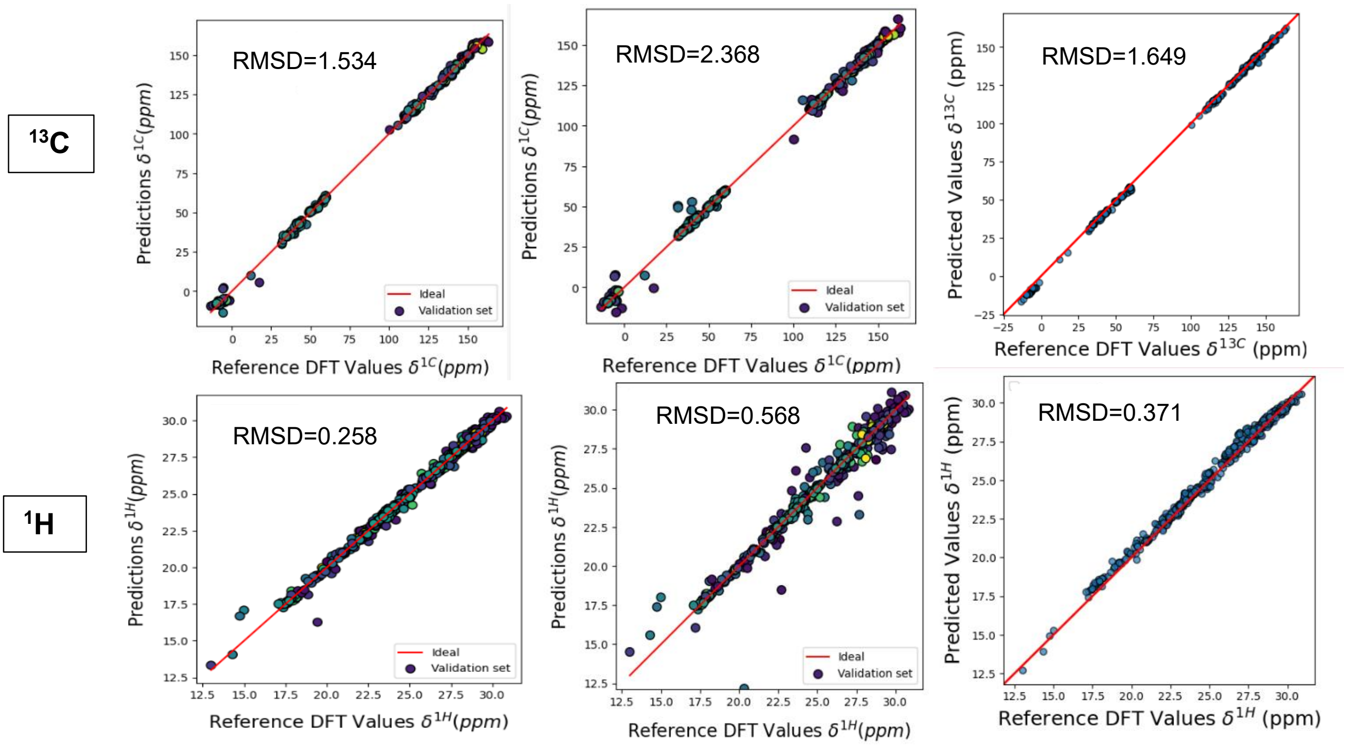
In this work, we present a systematic benchmarking of machine learning (ML) models for the prediction of ^{1}H and ^{13}C NMR chemical shifts*) across a dataset of 36 amino acid crystals containing 492 structurally unique carbon atoms and 764 unique hydrogen atoms. The spectroscopic signals were calculated using density functional theory (DFT) calculations at the PBE level [2]. **The purpose of the study is to examine how well the NMR signals can be predicted from theory using different *descriptors*, i.e. different ways of describing the atomic environment around the ^{1}H and ^{13}C NMR probes**.

Several local atomic environment descriptors were evaluated: from simple ones, with few parameters, to very elaborate ones, like those in Fig. 1. A couple of different regression methods were also used. In the figure, results with the MACE and SOAP descriptors are compared with the ShiftML3 model [1], which is the best performing model in literature, trained on 1.4 million NMR shielding values from 14,000 organic crystals.

We find that for both ^{1}H and ^{13}C, a fine-tuned MACE-MP(0) model (where an additional small neural network was trained to learn which features are most relevant to the target property) outperforms the other descriptors, despite the fact that the original MACE model was pre-trained on inorganic crystal energetics rather than NMR data. **MACE-MP0 thus manages very well to capture the effect of surrounding environment on the chemical shift.**

*) Actually, we calculate "isotropic chemical shielding" but they are proportional to the chemical shifts.

Descriptors MACE-MP0 SOAP ShiftML3



Bibliography:

[1] Chaloupecká, E.; Socha, O.; Dračínský, M. Solid State Nucl. Magn. Reson. 138 (2025) 102019.

[2] Chaloupecká, E.; Tyrpekl, V.; Bártová, K.; Nishiyama, Y.; Dračínský, M. Solid State Nucl. Magn. Reson. 130 (2024) 101921.