The accuracy limit of aqueous chemical shift predictions

Stefan Maste,[1] Tim Pongratz,[1] Bikramjit Sharma,[2] Ignacio J. Pickering,[3] Jinze Xue,[3] Adrian E. Roitberg,[3] Dominik Marx,[2] Stefan M. Kast [1]

[1]Fakultät für Chemie und Chemische Biologie, Technische Universität Dortmund, 44227 Dortmund, Germany

[2]Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany

[3]Department of Chemistry, University of Florida, Gainesville, FL 32611, USA

Nuclear magnetic resonance (NMR) spectroscopy is one of the main analytical techniques to investigate chemical systems. [1] In addition to experiments, computational methods have been employed to calculate NMR parameters and gain insight into experimentally inaccessible systems. However, reaching quantitative agreement with experiments is still a challenging task due to the high structural sensitivity of NMR obsevables, especially with respect to the accurate description of the solvent influence.

Here we demonstrate that combining state-of-the-art computational methodologies, *ab initio* molecular dynamics simulations (AIMD) for generating locally solvated structural ensembles and Embedded Cluster Reference Interaction Site Model (EC-RISM) [2] calculations allows for predicting accurate NMR response parameters for species in aqueous solution [3], a protocol acievied by extending an earlier approach successfully applied to electron paramagnetic resonance (EPR) parameters. [4] Applied to the reference compound trimethylsilylpropanesulfonate (DSS) and target molecules *N*-methylacetamide (NMA) and trimethylamine *N*-oxide (TMAO), we show that a hybrid solvent system, consisting of a limited number of explicit water molecules in an EC-RISM background, achieves quantitative accuracy for chemical shifts and is considerably more efficient than previous approaches. [5]
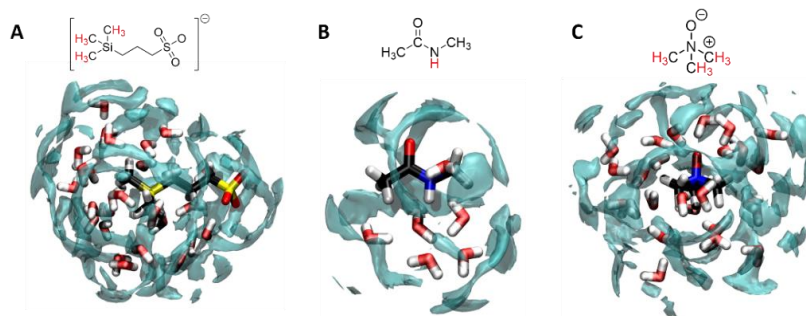
Using the same hybrid solvent system in classical force field molecular dynamics simulations (FFMD) reveals that the approach is transferable. However, deficiencies in FFMD-generated ensembles can be traced back to the inaccurate description of hydrogen bonds as well as the distribution of intramolecular bond lengths.

As a third method to generate molecular ensembles, machine learning potentials such as ANI-2x [6] show an improvement in the intramolecular description of NMA compared to force fields. A modification of the training dataset yields an improved solvent distribution more closely resembling AIMD ensembles compared to FFMD.

[1] J. A. Caro, A. J. Wand, *Methods* **2018**, *148*, 67-80.
[2] T. Kloss, J. Heil, S. M. Kast, *J. Phys. Chem. B* **2008**, *112*, 4337-4343.
[3] S. Maste, B. Sharma, T. Pongratz, …, D. Marx, S. M. Kast, *Phys. Chem. Chem. Phys*. **2024**, DOI: 10.1039/D3CP05471C.
[4] B. Sharma, …, S. M. Kast, …, D. Marx, *J. Chem. Theory Comput.* **2021**, *17*, 6366-6386.
[5] M. Dračínský, H. M. Möller, T. E. Exner, *J. Chem. Theory Comput.* **2013**, *9*, 3806-3815.
[6] C. Devereux, …, A. E. Roitberg, *J. Chem. Theory Comput*. **2020**, 16, 4192.