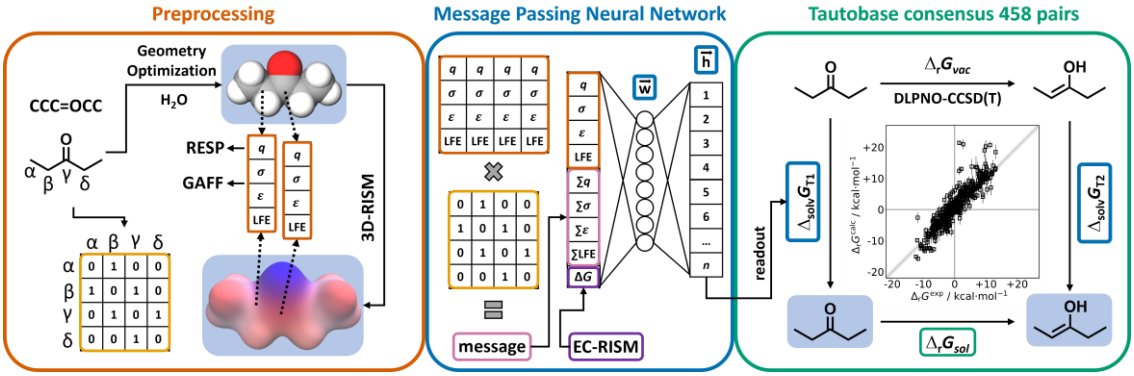
EC-RISM/MPNN-based hydration free energy models with application to tautomer equilibria

Christian Chodun, Yannic Alber, Michael Strobl, Nicolas Tielker, Lukas Eberlein, Florian Mrugalla, Julia Jasper, Jan Sleegers, Stefan M. Kast

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

The accurate prediction of hydration free energies (HFEs) is an important task in different areas, as the thermodynamic cost of solvation is one of the key factors in processes from protein-ligand binding to chemical reactions. Hence, much effort has been spent in predicting the HFE from electronic structure calculations and molecular dynamics simulations as well as, in recent years, machine learning (ML) methods, [1] for which large and reliable datasets are needed.



Training data dependence can be reduced using physics-informed ML, where measured or calculated properties representing the underlying physics of the molecules and their interactions are used as additional input features. [2] Here we present a physics-informed ML method that combines the Embedded Cluster Reference Interaction Site Model (EC-RISM) [3] with a Message Passing Neural Network (MPNN). [4] The solute is represented by a graph, where atoms are described by partial charges and Lennard-Jones parameters together with "local" atomic free energies (LFEs) [5,6] adding up to the total HFE. Augmenting with EC-RISM HFEs, we show state-of-the-art accuracy on independent HFE datasets, including SAMPL challenge data. [7,8]

This model is then used to predict the HFEs of tautomers which are combined with DLPNO-CCSD(T) [9] gas phase data to calculate aqueous tautomerization free energies. Applied to a literature consensus set of more than 450 pairs based on the Tautobase, [10] overall better agreement with experiment than other reported methods is obtained. By filtering according to an upper prediction error criterion, an ordered set with decreasing prediction accuracy is produced, allowing for the identification of suspicious database entries where measurement or annotation errors are likely.

[1] R. E. Skyner, …, J. B. O. Mitchell, *Phys. Chem. Chem. Phys.*, **2015**, *17*, 6174–6191.
[2] G. E. Karniadakis, …, L. Yang, *Nat. Rev. Phys.*, **2021**, *3*, 422–440.
[3] N. Tielker, …, S. M. Kast, *J. Comput. Aided Mol. Des.*, **2018**, *32*, 1151–1163.
[4] J. Gilmer, …, G. E. Dahl, in *Mach. Learn. Meets Quantum Phys.*, Springer International Publishing, Cham, **2020**, pp. 199–214.
[5] S.-H. Chong, S. Ham, *J. Chem. Phys.*, **2011**, *135*, 034506.
[6] Y. Alber, PhD thesis, TU Dortmund (Germany), **2021**.
[7] N. Tielker, … S. M. Kast, .S. Liu, in ACS Symposium Series 1397: *Free Energy Methods in Drug Discovery: Current State and Future Direction*s, **2021**, 67-107.
[8] D. L. Mobley, J. P. Guthrie, *J. Comput. Aided Mol. Des.*, **2014**, *28*, 711–720.
[9] F. Pavošević, …, F. Neese, E. F. Valeev, *J. Chem. Phys.*, **2017**, *146*, 174108.
[10] O. Wahl, T. Sander, *J. Chem. Inf. Model.*, **2020**, *60*, 1085–1089.