Contribution ID : **138** Type : **Oral Presentation**

ML-enhanced integral equation theory for solvent-mediated effective potentials

Solvation effects on thermodynamic properties of chemical systems can be captured by integrating-out the solvent degrees of freedom for a given geometric solute configuration to yield an effective solvent-mediated free energy surface ("potential of mean force", PMF). Based on this philosophy, a plethora of "implicit solvation models" has emerged over the years, from continuum electrostatics approaches for classical and quantum-electronic molecular solutes, which essentially address the solvent integration task iteratively for each spatial solute configuration, up to very efficient non-iterative approaches such as the Generalized Born (GB) model for the electrostatic component of the PMF. However, increasing computational efficiency typically goes hand in hand with reduced accuracy.

An implicit iterative solvation approach that can be parametrized to produce highly accurate free energies is provided by integral equation theory, namely the "3D reference interaction site model" (3D RISM) and its extension to quantum-electronic solutes, "embedded cluster RISM" (EC-RISM), which we developed over the past years for aqueous and non-aqueous solvents.1-6 By accounting for solvent granularity and directional solute-solvent interactions, EC-RISM can be tuned to surpass implicit solvation approaches in terms of quantitative agreement with experiment, as is demonstrated for chemical reaction thermodynamics in solution and spectroscopic observables.

Further progress is possible by using message passing neural networks (MPNN) with special atom-local RISM-derived descriptors, as is demonstrated on predictive calculations of hydration and tautomerization free energies in aqueous solution. This provides a route towards using pure EC-RISM and, as a perspective, MPNN-EC-RISM data as reference for training fast and accurate, non-iterative machine learning free energy potentials (MLFEP), for which first results are presented. A crucial factor in this context is the limited experimental reference data availability and their possibly large uncertainty, particularly for instance for aqueous tautomerization reaction thermodynamics. A purely theory-driven approach for identifying unreliable experimental data points based on a consistency analysis of diverse methods is discussed.

[1] T. Kloss, J. Heil, S. M. Kast, J. Phys. Chem. B 2008, 112, 4337-4343. [2] R. Frach, P. Kibies, …, S. M. Kast, Angew. Chemie Int. Ed. 2016, 55, 8757-8760. [3] N. Tielker, L. Eberlein, …, S. M. Kast, J. Comput.-Aided Mol. Des. 2021, 35, 453-472. [4] B. Sharma, V. A. Tran, …, S. M. Kast, F. Neese, D. Marx, J. Chem. Theory Comput. 2021, 17, 6366-6386. [5] S. Maste, B. Sharma, …, S. M. Kast, Phys. Chem. Chem. Phys. 2024, 26, 6386-6395. [6] K. E. Ebbert, F. Sendzik, …, S. M. Kast, G. H. Clever, Angew. Chemie Int. Ed. 2025, 64, e202416076.

Primary author(s) : KAST, Stefan M. (TU Dortmund University)

Presenter(s) : KAST, Stefan M. (TU Dortmund University)