Prediction of acidity constants and pH-dependent microstate populations for drug-like compounds

Nicolas Tielker,[1] Jochen Heil,[1] Lukas Eberlein,[1] Stefan Güssregen,[2] Stefan M. Kast[1]

[1]Fakultät für Chemie und Chemische Biologie, TU Dortmund, Germany
[2]Sanofi-Aventis Deutschland GmbH, Frankfurt am Main, Germany

Reliable yet fast prediction of physicochemical properties of drug-like compounds requires proper theories, as for instance provided by the integral equation approach to fluid phase thermodynamics. [1] Such a method allows for efficient calculations of free energies of solvation of both neutral and ionic molecules in a wide range of solvents. To accurately model the solvation of small molecules in water we here combine such a statistical-mechanical description of the solvent with a quantum-level description of the solute in the form of the "embedded cluster reference interaction site model" (EC-RISM). This combination, optimized with respect to quantitative accuracy, takes both the electronic relaxation and the excess chemical potential governing the insertion into a solvent into account for predicting the free energy of solvation. [2] It is therefore possible to address challenging problems related to drug discovery. [3]

The macroscopic acidity constants of drug-like molecules are difficult to predict since these species often contain functional groups and scaffolds that imply a multitude of tautomeric or ionic states ("microstates") at physiological pH. Moreover, the microstate relevant for a drug's mechanism of action might differ from the highest populated state in bulk solution, which is essential information for predicting binding constants accurately. In this context, the SAMPL6 challenge (Statistical Assessment of the Modeling of Proteins and Ligands [4]) was designed to specifically address such difficulties by requiring the participants to blindly predict the macroscopic pK_a values of kinase inhibitor fragments as well as their local microstate pK_a estimates and the fractional populations as a function of pH.

A workflow is described to accurately calculate the macroscopic pK_a of a given compound by determining the free energy of all tautomers of the deprotonated and the protonated form by EC-RISM calculations. Applied to small organic molecules covering a wide range of functional groups, it is shown that in total at most 5 empirically adjusted parameters (2 or 3 for free energy predictions and 2 for the final pK_a model) are required, ultimately allowing us to predict acidity constants to a root mean square error of about 1.6 pK_a units within the SAMPL6 blind prediction challenge. [4] Unlike other first-principles models there is in our case no need for distinguishing acids and bases or different compound classes. These pK_a values in combination with predicted free energy data of all tautomers within a given protonation state enable us to also predict the fractional microstate populations over the entire pH-range.

[1] E. L. Ratkova, D. S. Palmer, M. V. Fedorov, *Chem. Rev.* **2015**, *115*, 6312-6356.
[2] T. Kloss, J. Heil, S. M. Kast, *J. Phys. Chem. B* **2008**, *112*, 4337-4343.
[3] N. Tielker, D. Tomazic, J. Heil, T. Kloss, S. Erhart, S. Güssregen, K. F. Schmidt, S. M. Kast, J. Comput.-Aided Mol. Des. **2016**, *30*, 1035-1044.
[4] https://drugdesigndata.org/about/sampl6 (last accessed 2018/02/13).