Insights from the first *euroSAMPL* blind prediction challenge

Nicolas Tielker[1], Michel Lim[2], Patrick Kibies[1], Juliana Gretz[3], Björn Hein-Janke[4], Ricardo A. Mata[4], Paul Czodrowski[2], Stefan M. Kast[1]

[1]Fak. Chemie und Chemische Biologie, Technische Universität Dortmund, [2]Dept. Chemie, Johannes-Gutenberg-Universität Mainz, [3]Fak. Chemie und Biochemie, Ruhr-Universität Bochum, [4]Inst. Phys. Chemie, Georg-August-Universität Göttingen

The FAIR (Findable, Accessible, Interoperable, and Reusable) principles for sustainable research data management (RDM) are vital for maximizing reproducibility of computational chemistry data and results, and to easily verify and use newly developed methods. This includes methods such as the automated or manual annotation of generated research data with relevant author- and domain-specific metadata, shared indexing between experimental and computational data, persistent storage, and the transparent and automated analysis of raw computational data.

To advance and encourage the adoption of RDM methods, we organized a pK_a blind prediction challenge as a community task for testing models in the spirit of the SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) series of challenges. [1,2,3] The first *euroSAMPL* challenge [4] was based on experimental aqueous pK_a measurements of 35 small, drug-like molecules done in-house, and allowed participants to submit computational predictions over a timeframe of three months before the experimental values were published. Compounds were selected according to novelty, existence of only one protonation equilibrium in the experimental pH range between 2 and 12, no significant populations of additional tautomeric microstates, and chemical diversity.

Processing and initial analysis of participants' data was automated, with final scores weighting not only the accuracy of the predicted values but also the submissions' RDM quality to generate the final ranking. The latter was judged by questionnaire-based peer evaluation by all other participants. Results of the challenge reveal good agreement between predicted and experimental values for some of the methods, including QM-based and ML-based methods, but the quality of the RDM still has significant room for improvement. Here, one of the submissions received the clear first place with respect to its "FAIRness", with three other submissions on a shared, but distant, second place, indicating a need for more standardization and awareness of proper research data management. Individual prediction results are compared to a "consensus prediction" from either all, or a subset of the submissions, and discussed with a perspective on best practices in the application of pK_a predictions in industry and academia.

All data and statistical analyses are collected in a GitLab repository, [5] to be published as part of the "SAMPL Special Collection" in Phys. Chem. Chem. Phys. Collectively, insights gathered in terms of challenge design and technical implementation stimulate ideas for future rounds.

[1] https://www.samplchallenges.org (last visited 2025/02/19)
[2] N. Tielker, L. Eberlein, O. Beckstein, S. Güssregen, B. I. Iorga, S. M. Kast, S. Liu, Perspective on the SAMPL and D3R Blind Prediction Challenges for Physics-Based Free Energy Methods. In K. A. Armacost, D. C. Thompson (eds.), *ACS Symposium Series Vol. 1397: Free Energy Methods in Drug Discovery: Current State and Future Directions*, **2021**, 67-107.
[3] M. Işık, A. S. Rustenburg, A. Rizzi, M. R. Gunner, D. L. Mobley, J. D. Chodera, *J. Comput.-Aided Mol. Des.*, **2021**, *35*, 131-166.
[4] https://qmbench.net/eurosampl/ (last visited 2025/02/19)
[5] https://gitlab.tu-dortmund.de/kast_ccb/eurosampl/challenge (last visited 2025/02/19)