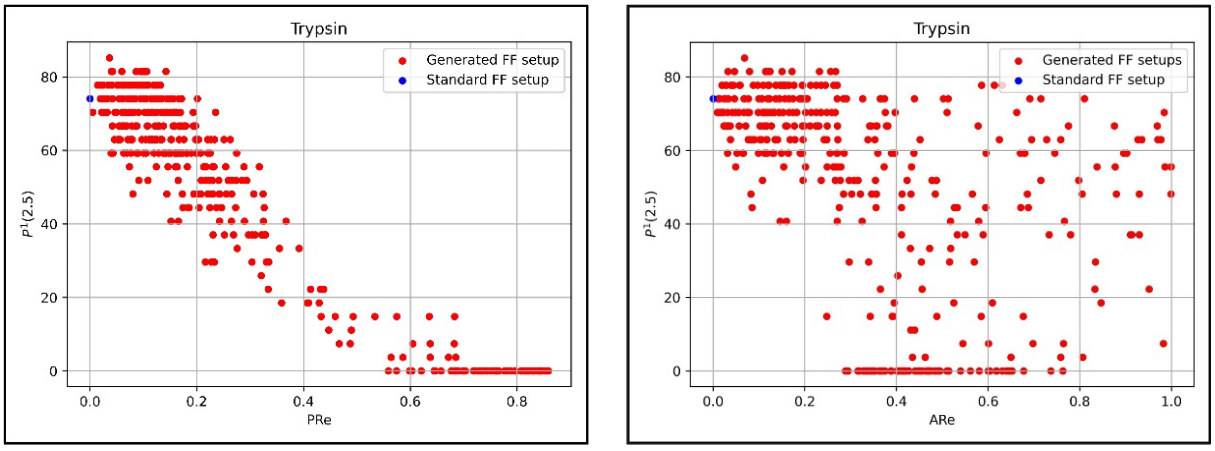
Critical Assessment of the Docking Calibration Problem

Aykhan ISRAFILLI[a], Dr. Dragos HORVATH[a]

[a] *Laboratory of Chemoinformatics, University of Strasbourg, UMR 7140; Institut Le Bel, 4 rue Blaise Pascal, 67081 Strasbourg, France*

Despite the wide-spread use of docking as a virtual screening tool, its performance is often disappointing – a well-known truth acknowledged by practitioners of docking, and yet not visible from literature biased towards selectively publishing success stories. In this contribution, we use our in-house tool S4MPLE[1,2] for an in-depth exploration of its behavior and performances with respect to the tunable parameters of its energy function. Our findings, which likely apply to any other docking protocol, suggest that it is relatively easy to calibrate a docking score achieving state-of-the-art native pose prediction results in redocking. Many different set-ups are seen to lead to near-optimal results in this respect: however, each of these setups may return radically different rankings of poses and ligands in terms of energy/affinity/score (figure 1). Therefore, we suggest that docking calibration and benchmarking should always include, in a multi-objective optimization approach, both pose prediction and active compound retrieval criteria – the latter using as far as possible real activity data rather than artificial active/decoy sets.



(a) (b)

Figure 1. The figure illustrates how variation of force field setup parameters modulates the ranking of sampled poses and ranking of the ligands.

[1] L. Hoffer; C. Chira; G. Marcou; A. Varnek; D. Horvath. Molecules 20 (2015) 8997–9028.
[2] L. Hoffer; D. Horvath. J. Chem. Inf. Model. 53 (2013) 88–102