[L16] Using Machine Learning for Structure-Based Predictions of Protein-Ligand Interactions

Sergei Grudinin[1], Maria Kadukova[1]

[1] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble, 38000, France.*

Development of new computational methods for the prediction of protein-ligand interactions is stimulated by the growing demand in drug discovery for the efficiency and accuracy of virtual screening. To computationally analyze numerous compounds, extremely fast techniques are required. Therefore, improving their speed and accuracy is an active research field in structural chemoinformatics. Machine-learning and statistical potentials are thus very often employed for the selection of the most probable binders to a target protein. In this talk I will demonstrate how machine-learning can be used for structural predictions in general, and, more specifically, how to use it for the parametrization of small molecules [1,9] and for the training of a free-shape distance-dependent protein-ligand potential [7,8].

Unlike knowledge-based methods based on Boltzmann statistics, or physics-based methods, in our approach, called Convex-PL, we do not impose any functional form of the interaction potential. Instead, we use an optimization technique, accepting that the target binding energy value is decomposable into a polynomial basis with unknown expansion coefficients. These are then deduced from the structural data collected from protein-ligand complexes using a convex formulation of the optimization problem, similar to our protein-protein interaction potentials [2,3]. The training set consists of the complexes taken from the PDBBind database. We generate false poses with constant RMSD rigid-body deformations of the ligands inside the binding pockets. This allows the obtained potential to be generally unbiased towards other molecular docking methods, which are often used for decoys generation.

We first tested Convex-PL in the CSAR 2013-2014 and D3R 2015-2016 blind assessments [4,5], where it successfully predicted the binding poses. For a more general validation, we assessed it using data from D3R Grand Challenge submissions and the CASF 2013 study [6], which includes the docking, scoring, ranking, and screening tests. Our docking and ranking test results outperform the other 20 methods previously assessed in CASF 2013. Also, Convex-PL performs better than average in the scoring test. Finally, I will discuss the current challenges in virtual screening and protein-ligand docking. Specifically, I will put attention into the importance of modelling protein flexibility and multiple conformational states upon binding to small molecules [10].

Bibliography:

[1] Kadukova, M.; Grudinin, S. J. Chem. Inf. Model. 56 (2016) 1410-1419.
[2] Popov, P.; Grudinin, S. J. Chem. Inf. Model. 55 (2015) 2242-2255.
[2] Neveu, E.; Ritchie, D.W.; Popov, P.; Grudinin, S. Bioinformatics 32 (2016) i693-i701.
[4] Grudinin, S.; Popov, P.; Neveu, E.; Cheremovskiy, G. J. Chem. Inf. Model. 56 (2015) 1053-1062.
[5] Grudinin, S.; Kadukova, M.; Eisenbarth, A.; Marillet, S.; Cazals, F. J. Comput. Aided Mol. Des. 30 (2016) 791-804.
[6] Li, Y.; Han, L.; Liu, Zh.; Wang, R. J. Chem. Inf. Model. 54 (2014) 1700–1716.
[7] Kadukova, M. & Grudinin, S. J Comput Aided Mol Des 31 (2017) 943.
[8] https://team.inria.fr/nano-d/convex-pl/
[9] https://team.inria.fr/nano-d/software/knodle/
[10] Kadukova, M. & Grudinin, S. J Comput Aided Mol Des 32 (2018) 151.