[P18] Knowledge-based prediction of protein-ligand interactions

Maria Kadukova[1], Sergei Grudinin[1]

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

New methods for accurate and computationally simple description of protein-ligand interactions are still demanded by industry and academia to elucidate the activity of biological targets and find their new binders modulating functions of the living systems. Here we present a machine learning approach to first parameterize small molecules [1] and second to train a free-shape distance-dependent protein-ligand potential called Convex-PL. Unlike knowledge-based methods based on Boltzmann statistics, we do not impose any functional form of the potential. Instead, we use an optimization approach, accepting that the target binding energy value is decomposed 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. We have already implemented this principle for successful predictions of protein-protein binding [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 unbiased towards other molecular docking methods, which are often used for training decoys generation. Convex-PL performed successfully in the CSAR 2013-2014 and D3R 2015-2016 competitions [4,5]. For a more general validation, we assessed it using data from 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, but produces rather average results in the screening test. The parametrization is available at https://team.inria.fr/nano-d/software/knodle/. The protein-ligand potential is available at https://team.inria.fr/nano-d/software/Convex-PL. The GUI will be made available as a part of the SAMSON software platform at https://www.samson-connect.net.

Bibliography:

[1] Kadukova, M.; Grudinin, S., Knodle: A Support Vector Machines-Based Automatic Perception of Organic Molecules from 3D Coordinates, J. Chem. Inf. Model., 2016
[2] Neveu, E.; Ritchie, D.W.; Popov, P.; Grudinin, S. PEPSI-Dock: a detailed data-driven protein-protein interaction potential accelerated by polar Fourier correlation. Bioinformatics, 2016
[3] Popov, P.; Grudinin, S. Knowledge of Native Protein-Protein Interfaces is Sufficient to Construct Predictive Models for the Selection of Binding Candidates. J. Chem. Inf. Model., 2015
[4] Grudinin, S.; Popov, P.; Neveu, E.; Cheremovskiy, G. Predicting Binding Poses and Affinities in the CSAR 2013-2014 Docking Exercises Using the Knowledge-Based Convex-PL Potential. J. Chem. Inf. Model., 2015
[5] Grudinin, S.; Kadukova, M.; Eisenbarth, A.; Marillet, S.; Cazals, F., Predicting binding affinities for protein - ligand complexes in the 2015 D3R Grand Challenge using a physical model with a ridge regression parameter estimation, J. Comput.-Aided Mol. Des., 2016
[6] Li, Y.; Han, L.; Liu, Zh.; Wang, R. Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. J. Chem. Inf. Model., 2014