Poster # 26

A novel machine learning-based protocol for predicting biological activity of chemical compounds

Sabina Smusz*, Stefan Mordalski, Jagna Witek, Krzysztof Rataj, Andrzej J Bojarski
Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences
Smętna 12, 31-343 Krakow, Poland. * e-mail: smusz@if-pan.krakow.pl

A growing popularity of virtual screening (VS) techniques in drug design campaigns has stimulated the development of new tools specific to VS applications [1]. Mostly they are aimed for management and analysis of huge amounts of data received in the screening process [2]. The obvious application of ML is with ligand-based methods, where the biological activity of molecules is evaluated on the basis of structure and properties compared with ones of known ligands of particular target.
In this study, ML algorithms were combined with docking procedure and such protocol was evaluated on compounds active towards serotonergic 5-HT_6 and 5-HT_7 receptors. The approach was tested for its ability to discriminate between actives, true inactives and assumed inactives (generated according to the DUD methodology [3]). Compounds were docked to homology models of the investigated receptor. Then, they were represented in 2 ways – using Structural Interaction Fingerprint [4] (enabling description of interactions occurring between the docked compounds and particular amino acids of the protein) and with the use of Spectrophores [5] from Open Babel package (providing the description of 3-dimensional conformation of a molecule). Five machine learning classification algorithms were applied for compounds evaluation with the above-mentioned representations as an input (individually and using hybrid approach). The final outcome for each of the tested instances was obtained with data fusion methods based on the weighted average strategy, taking into account values of scoring function from the docking program (Glide, Schrödinger Suite 2012) and the quality of particular homology model evaluated by enrichment factor.

Acknowledgements
This study is partially supported by project "Diamentowy Grant" DI 2011 0046 41 financed by Polish Ministry of Science and Higher Education.

References:

[1] A. Breda, L.A. Basso, D.S. Santos, W.F. de Azevedo Jr., *Current Computer - Aided Drug Design*, 4 (2008) 265-272.

[2] A. Schwaighofer, T. Schroeter, S. Mika, G. Blanchard, *Combinatorial Chemistry & High Throughput Screening*, 12 (2009) 453-68.

[3] N. Huang, B.K. Shoichet, J.J. Irwin, *Journal of Medicinal Chemistry*, 49 (2006) 6789-801.

[4] S. Mordalski, T. Kosciolek, K. Kristiansen, I. Sylte, A.J. Bojarski, *Bioorganic & Medicinal Chemistry Letters*, 22 (2011) 6816-6819.

[5] P. Bultinck, W. Langenaeker, R. Carbó-Dorca, J. P. Tollenaere, *Journal of Chemical Information and Computer Sciences*, 43 (2003) 422-428.