P. Vassiliev[1,2,3]

CONSENSUS APPROACH IN QSAR

[1] Research Institute of Pharmacology, [2] Chair of pharmacology,
[3] Laboratory of Molecular Modeling and Computer Search for Drug Substances, Volgograd State Medical University, Pavshikh Bortsov square, 1, Volgograd, Russia

pvassiliev@mail.ru

The theoretical basics of using a consensus approach for the analysis of "chemical structure - biological activity" relationships are described. Consensus is a decision-making method based on the estimates obtained with several approaches.

According to the type of prediction dependencies, following consensus types are recognized: 1) ensemble prediction, when several similar decision rules are applied; 2) actual consensus, when prediction methods with different mathematical formalism are used in conjunction; 3) committee approach, which summarizes the estimates obtained by several groups of different methods. According to the method of prediction estimates generalization (voting), we distinguish full, selective, supremal, unweighted/weighted, simple and average consensus. According to the architecture of the predictive systems, consensus can be single-level, multi-level and hierarchical.

When a chemical compound interacts with a living system, a biological response arises, which could be described as a multi-dimensional discrete value with very high fluctuating variability. A reasonably accurate calculated assessment of that response is only possible within the framework of the classification model. The sets of estimates, which are generalized in process of consensus, do not have clear boundaries. In that case, the metric of assignment of compound to active class is the membership function.

The presentation contains examples of apllications of different types of consensus in QSAR.

QSAR model for the prediction of plasma protein binding was constructed by means of an ensemble method of boosted regression trees [1]. A prediction of acute toxicity of environmentally hazardous organic compounds was carried out by consensus PLS-RF-CBM-GBM model using SiRMS-SPCI program [2]. Search for antimalarial compounds was performed using ISIDA program and a two-level mixed consensus of 17 families (ensembles) SVM-models [3].

In IT Microcosm, a hierarchical multi-level mixed consensus based on various voting procedures is used to predict the biological activity of chemical compounds. Search for substances with high pharmacological activity among 1700 condensed azoles derivatives showed the enrichment coefficient for 29 activity types varied from 3 to 503 times, with an average of 47 times. The accuracy of the search for highly active compounds ranged from 44% to 100%, with an average of 66%. The effectiveness of the consensus search for highly active substances ranged from 1.6 to 3.9 times, with an average of 2.9 times, in comparison with intuitive non-computer search. Out of the 281 predicted and tested compounds 143 (51%) had the activity equal or greater than the reference drugs. Of these 143 substances, 62 were comparable or exceeded the activity of substances which were found before as a result of the "human" prediction [4].

A consensus search for substances with antidiabetic activity using IT Microcosm, PASS and AutoDock Vina was performed in a focused library of 2500 structurally unrelated compounds. The enrichment coefficient for 5 activity types ranged from 1.3 to 2.5 times, with an average of 2.0 times. The accuracy of the search for active compounds ranged from 40% to 79%, or 60% on average. Out of the 739 predicted and tested compounds, 441 (60%) were active, and 86 compounds (12%) were more active than the reference drugs [5].

1. Ghafourian T. et al. *BioImpacts*, 2013, **3(1)**: 21-27.
2. Tin'kov O.V. et al. *Pharmaceutical Chemistry Journal*, 2015, **49(2)**: 104-110.
3. Viira B. et al. *Molecules*, 2016, **21**: 853-870.
4. Vassiliev P.M. *In: Application of Computational Techniques in Pharmacy and Medicine*, Springer, 2014: 369-431.
5. Vassiliev P.M. et al. *In: Target-oriented search for antidiabetic agents*, VSMU, 2016: 126-181.

The work is funded by the Russian Science Foundation, 14-25-00139 project.