P. Pogodin [1]
A. Lagunin [1,2]
D. Filimonov [1]
A. Rudik [1]
V. Poroikov [1]

HOW SHOULD ONE USE AVAILABLE DATA TO BE SUCCESSFUL IN COMPUTER-AIDED SEARCH FOR NOVEL KINASE INHIBITORS?

[1] Institute of Biomedical Chemistry, 10 Bldg. 8, Pogodinskaya Str. Moscow, 119121, Russia;
[2] Medico-biological Faculty, Pirogov Russian National Research Medical University, Ostrovitianov str. 1, Moscow, 117997, Russia;
pogodinpv@gmail.com

Protein kinases represent an ubiquitous and large (more than 500 members) group of enzymes, which is known as kinome. Because of the tight involvement in many pathological processes, kinome is of particular interest as a set of therapeutic targets. Moreover, protein kinases were claimed as the major drug targets of the XXI century and, to date, efforts in kinase drug discovery have led to the successful approval of 28 kinase inhibitors by FDA. Accumulated experimental data on kinase inhibitory activity for chemical compounds are broadly available in the public domain.

Utilizing these data for training of (Q)SAR models, it is possible to predict single or multiple targets for chemical compounds across kinome, i.e. perform computer-aided search for novel kinase inhibitors with the desirable profile of the inhibitory activities. Given the fact that particular kinase inhibitory spectrum contributes to the safety profile of compound and to its efficacy; computational prediction of the spectrum may assist in accomplishing many tasks related to the drug discovery. Some of these tasks are: searching for the novel actives in still unallocated chemical space; planning a rational safety assessment; selection of effective mono- or combinational therapy among approved drugs for the treatment of diseases characterized by the well-defined kinase-related alterations on the molecular level.

Since kinase inhibitors often affect more than one target due to similar structure of ATP binding domain, which is a main target for kinase inhibitors; we have to deal with similar properties of the training data to utilize them efficiently. In our study, we compared different approaches to handling data for (Q)SAR modeling to reveal one, that provides best results for the task of the early detection of inhibitors which are active against a particular kinase. As long as the discovery of new actives via virtual screening is the most realistic applications for the (Q)SAR modeling and a commonplace in the interests of researchers from both industry and academia. Furthermore, reliable predictive models for searching novel inhibitors of distinct kinases will allow researchers to integrate them to search for inhibitors active against the desirable kinase spectrum.

In our study, we used: PASS [1] software, in which the term «biological activity spectrum» was introduced, - to build classifiers; ChEMBL [2] database, one of the reliable and most comprehensive sources among publicly available ones, - to get the activity data; 5-fold cross-validation procedure and bootstrap approach to assess quality of classifiers; MySQL and PHP to handle the data; R language for statistical evaluation.

Our results indicate, that, in general, the most efficient way to build classifiers for the search of novel kinase inhibitors is to create them using the data on actives and inactives against particular target separately from compounds tested against other targets (build individual classifiers).

However, we elucidated a number of kinases for which classifiers, built on the actives against multiple kinases (or actives and inactives together), outperform individual classifiers. We found this promising, since it may mean, that there are possibilities to find kinase inhibitors with modes of action and/or structures, which differ significantly from typical ones.

1. Filimonov D. et al. *Chemistry of Heterocyclic Compounds*, 2014, **50** (3): 444-457.
2. Bento A. et al. *Nucleic Acids Research*, 2014, **42**: 1083-1090.

This research is supported by the RFBR grant № 16-34- 01243.