DOI 10.18699/SBB-2022-193

Predicting the protein-ligand interactions based on ligands' structures and proteins' sequences

Karasev D.[1]*, Sobolev B.[1], Lagunin A.[1, 2], Filimonov D.[1], Poroikov V.[1]

[1] *Laboratory of Structure-Function Based Drug Design, Institute of Biomedical Chemistry, Moscow, Russia*

[2] *Department of Bioinformatics, Pirogov Russian National Research Medical University, Moscow, Russia*

** w.dmitrykarasev@gmail.com*

Key words: proteochemometrics, protein-ligand interaction, drug design

Motivation and Aim: Use of computational methods predicting protein-ligand interactions significantly facilitate finding of biologically active compounds and identification of their molecular targets. Existing ligand-based drug design methods are limited due to the incompleteness of data about ligand-target relationships. Overcoming this limitation, the so-called proteochemometrics methods implement the predictive models including structural features of both the ligands and targets. Currently available methods have a limited applicability domain and are commonly not available online. We propose the proteochemometrics method, providing the wider applicability, and present it online as a web-service.

Methods and Algorithms: Information about protein-ligand interaction was extracted from ChEMBL'28 database. An approach was developed to combine data from heterogeneous experimental sources and evaluate their reliability. We provided prediction options for the most popular scenarios that arise in the computer-aided evaluation of protein-ligand interactions (Fig. 1). The first scenario is designed to assess the ligand interaction with protein by comparing the ligand structures with the PASS approach. In PASS (Prediction of Activity Spectra for Substances), the MNA descriptors (multilevel neighborhoods of atoms) are used as an input for the predictor based on a Naive Bayes classifier [1]. In the second scenario, the protein interaction with ligand is assessed by comparing the amino acid sequences with SPrOS method using the local sequence similarity. The use of the third scenario is required if the interaction spectra are not available for both the ligand and target. In this case, the prediction is based on handling both the ligand structures and protein sequences [2].

Results: We collected training sets for various groups of targets differed in protein target heterogeneity. Two sets included the data on families of homologous targets (protein kinases and G-protein-coupled receptors) and their ligands. The third set united data irrelative to protein families. Testing under the first scenario showed a high prediction accuracy of 0.96–0.99. Testing under the second scenario showed an accuracy of 0.97–0.99 for prediction within specific protein families and sets disordered by protein homology. The third scenario showed the lower but acceptable accuracy of 0.86–0.94. The developed web service allows the user to perform the prediction under the specified scenarios, thereby covering the needs of researchers when planning experimental studies [3].

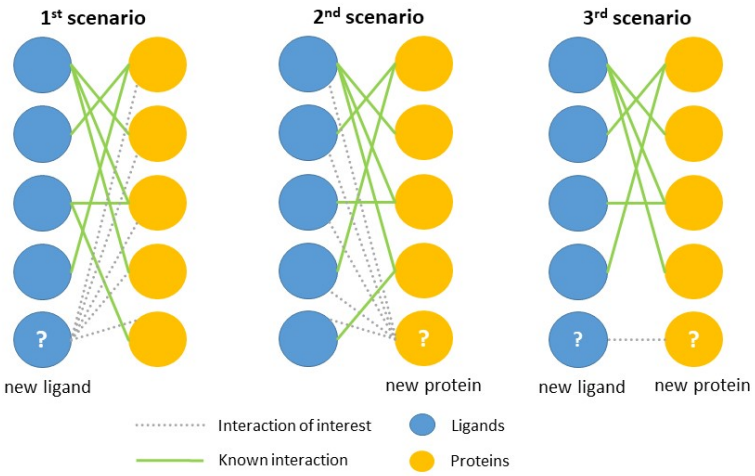

Fig. 1. The most popular scenarios that arise in the computer-aided evaluation of protein-ligand interactions

Web service is available online at the web-site: http://way2drug.com/proteochemometrics/.

Conclusion: The developed freely available web service provides the opportunity to use the proteochemometrics technology to a wide range of users. Researchers can use both classical ligand structure prediction and look for likely ligands for a chosen target. The method makes it possible to predict protein-ligand interactions for popular protein families representing promising drug targets.

Acknowledgements: This study is supported in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030), project No. 122030100170-5.

References

1. Filimonov D.A. et al. Computer-aided prediction of biological activity spectra for chemical compounds: opportunities and limitations. *Biomed Chem Res Method.* 2018;1(1):e00004.
2. Karasev D.A. et al. Prediction of protein-ligand interaction based on sequence similarity and ligand structural features. *Int J Mol Sci*. 2020;21:8152.
3. Karasev D.A. et al. The method predicting interaction between protein targets and small-molecular ligands with the wide applicability domain. *Comp Biol Chem.* 2022;98:107674.