DOI 10.18699/SBB-2022-201

Computational approach for searching bioactivity of natural products based on their 2D chemical structure

Sadovskaya A.*, Gurevich A.
Center for Algorithmic Biotechnology, St. Petersburg State University, St. Petersburg, Russia
** alexandrasad04@gmail.com*

Key words: peptidic natural products, bioinformatics, bioactivity, chemical structure

Motivation and Aim: The object of this study is peptidic natural products (PNPs), such as ribosomal and nonribosomal peptides. PNPs are an important group of natural products that include many antibiotics, anticancer drugs, and other pharmaceuticals. Recent breakthroughs in mass spectrometry (MS) and genome sequencing enabled high-throughput PNP discovery. Emerging computational methods allowed identification of novel PNPs via modification-tolerant database search of MS data [1], *de novo* MS sequencing [2], and metabologenomics approaches [3]. These methods produce large amounts of *in silico* predicted PNPs that require further experimental analysis. Creating a robust strategy for selecting the most promising compounds out of many computationally predicted molecules remains an open problem.
Methods and Algorithms: We handle PNPs structures in standard chemical formats (currently SMILES, MDL MOL and SDF) and use the RDKit library to convert them into the SMILES chemical format. The resulting SMILES are queried in PubChem via the exact and similarity search modes [4]. We filter found molecules using the Tanimoto coefficient and the NP-likeness score to retain likely natural products having chemical structure similar to the query molecule. Key compound information, such as bioassay tests results and known producers, is summarized into a short report, while more detailed metadata for each identified molecule is saved into an extended report.
Results: In this work, we developed a computational pipeline for predicting biological activities of PNPs based on searching their chemical structure online in the PubChem database. Our tool is suitable for high-throughput analysis that distinguishes it from the alternatives: the PubChem information search and PASS, the leading bioactivity prediction software [5]. In addition, our method works with compounds absent in the PubChem database and also predicts taxonomic data missed by PASS. The created pipeline was tested on 550 PNPs from the MIBiG database. For 84 % of the PNPs, the pipeline found compounds identical in 2D structure in PubChem and generated a report with the biological activities described for them.
Conclusion: We develop a computational approach for predicting bioactivity of PNPs based on their 2D chemical structure. We plan to integrate the developing pipeline with the state-of-the-art approaches to PNPs discovery [1-3] to complement their output with the tentative biological activity of the identified compounds and likely taxonomy of their producers. We anticipate our pipeline will facilitate the search for new therapeutic agents by providing researchers with biologically relevant information. This data will help to prioritize *in silico* predicted PNPs for experimental validation and testing.
Acknowledgements: The study is supported by RSF (20-74-00032).

References

1. Gurevich A. et al. Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. *Nat Microbiol.* 2018;3:319-327.
2. Behsaz B., Mohimani H., Gurevich A., Prjibelski A., Fisher M., Vargas F. et al. De novo peptide sequencing reveals many cyclopeptides in the human gut and other environments. *Cell Syst*. 2020;10:99-108.
3. Cao L. et al. MetaMiner: a scalable peptidogenomics approach for discovery of ribosomal peptide natural products with blind modifications from microbial communities. *Cell Syst.* 2019;9:600-608.
4. Kim S., Chen J., Cheng T., Gindulyte A., He J. et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res*. 2020;49:D1388-D1395.
5. Filimonov D.A. et al. Prediction of the biological activity spectra of organic compounds using the pass online web resource. *Chem Heterocycl Compd.* 2014;50:444-457.