Computational prediction of biological activities of peptidic natural products

Alexandra Sadovskaya[1], Alexey Gurevich[2]
[1]Department of Bioinformatics, St. Petersburg State University, St. Petersburg, Russia; [2]Center for Algorithmic Biotechnology, St. Petersburg State University, St. Petersburg, Russia
E-mail: alexandrasad@mail.ru

Introduction

Peptidic Natural Products (PNPs) are a rich source of antibiotics 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, and metabologenomics approaches [2]. These methods produce large amounts of in silico predicted PNPs that require further experimental analysis. In this work, we develop a computational pipeline for predicting biological activities of PNPs based on their chemical structure.

Challenges

- Selection of the necessary software packages, libraries and databases for the search for biological properties of PNPs.
- Bringing the chemical data recording formats to uniform format (SMILES was chosen).
- Search for biological properties of PNPs in various chemical databases and display the results of program execution.

Funding

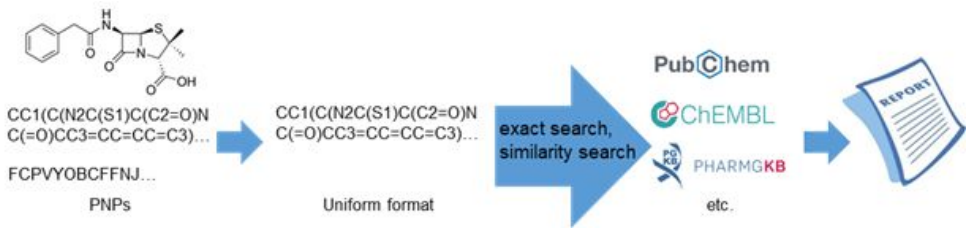
This study was supported by the Russian Science Foundation (grant 20-74-00032).

Methods

RDKit is a collection of cheminformatics and machine-learning software written in C++ and Python [3].

PubChem is an open chemistry database at the National Institutes of Health (NIH). It contains information on chemical structures, identifiers, chemical and physical properties, biological activities, patents, health, safety, toxicity data, and many others [4].

How it works



Results

github.com/Alexsad44/Diploma

Example of report from PubChem:
“Penicillin G Sodium is the sodium salt form of benzylpenicillin, a semi-synthetic, broad-spectrum penicillin antibiotic with bactericidal activity…”

1. Gurevich A, et al. Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. Nat Microbiology, 2018.
2. Behsaz B, et al. Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. Nature communications, 2021.
3. RDKit: Open-source cheminformatics. https://www.rdkit.org.
4. Kim S, Chen J, Cheng T, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res*. 2021;49(D1):D1388–D1395. doi:10.1093/nar/gkaa971