

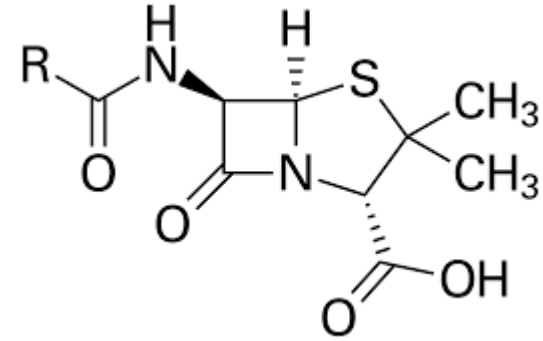
SMMole: Pipeline for searching biological properties of secondary metabolites based on their molecular structures

Alexandra Sadovskaya, Alexey Gurevich
Center for Algorithmic Biotechnology, St. Petersburg State University,
St. Petersburg

PNP

Peptidic Natural Products (PNPs):

- small bioactive peptidic molecules
- nonribosomal peptides (NRPs) and ribosomally synthesized and post-translationally modified peptides (RiPPs)
- promising compounds in the drug research (including antibiotics, antitumor agents, ...)
(Dang & Süssmuth, 2017).



Genome-based MS search of PNPs

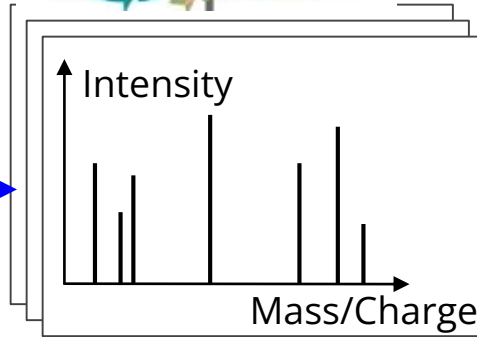
Genome sequencing &
Genome mining



Sample



Mass spectrometer



Spectral Dataset

Genome-based MS search of PNPs

Genome sequencing &
Genome mining



Structure prediction



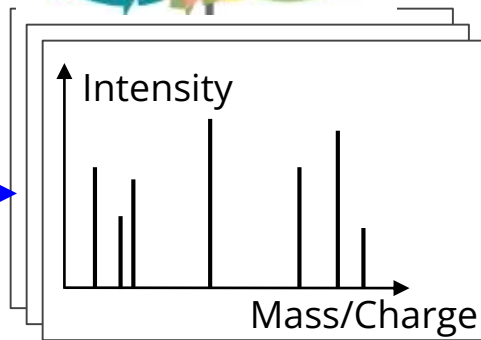
Database of
putative PNPs



Sample



Mass
spectrometer

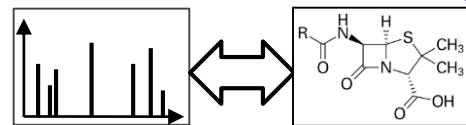


Spectral Dataset



Many predicted
molecules

Identification



MetaMiner (Cao et al., 2019)

Current approaches & their limitations

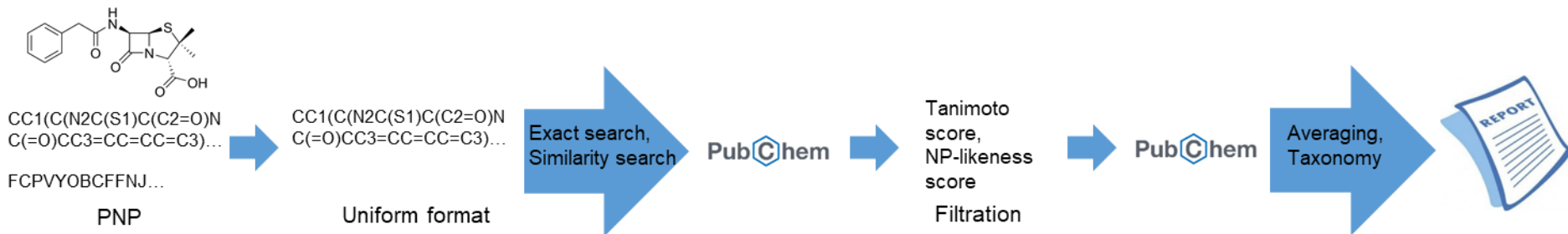
PASS (Filimonov et al., 2014)

- Commercial (or online only, one query per time, limited functionality)
- Should be retrained with new data available
- No taxonomy information

PubChem (Kim S et al., 2020)

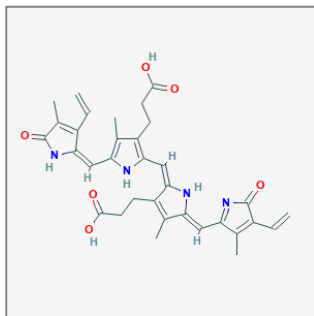
- Online only, thus not suitable for high-throughput analysis
- No predictions
- General purpose (both natural and synthetic compounds)

SMMole pipeline

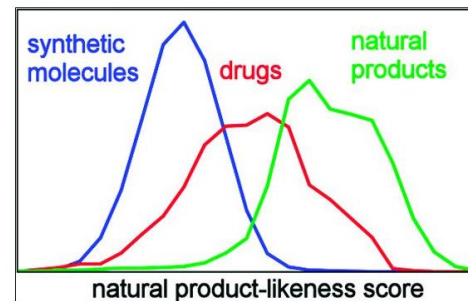
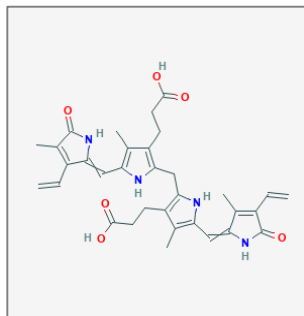


Python-based, with the use of

- RDKit
- PubChemPy
- Biopython
- requests



Tanimoto similarity (0.655)
(Kim et al., 2016)



NP-likeness
(Ertl et al., 2007)

Report example

EXACT match (Tanimoto == 100%)

Name	Azithromycin
Active (items):	310/4792
Tanimoto score	100.0 %
NP-likeness	1.85
Producer taxonomy	<i>Streptomyces lavendulae</i> (100.0 %)

SIMILAR compounds (100% > Tanimoto >= Threshold)

Name	Many hits (2)
Active (items):	809/10309 (min:110, max:699)
Tanimoto score	83.0±0.486 % (min:82.0 %, max:83.0 %)
NP-likeness	1.83±0.06 (min:1.77, max:1.89)
Producer taxonomy	<i>Streptomyces</i> (81.82 %)

Taxonomy processing

Producers of similar compounds:

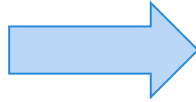
Streptomyces hygrosopicus

Streptomyces lydicus

Streptomyces albidoflavus

Streptomyces platensis

Aeromicrobium erythreum



Predicted taxonomy:

Genus: *Streptomyces* (80%)

Results

- We tested SMMole on PNPs from MIBiG (Kautsar et al., 2019).

Total	Identified structure	Identified bioactivity	Running time (avg/total)
541	85.8 % (464)	55.8% (259 from 464)	4 min / 36 hours

- We plan to integrate SMMole with approaches to PNPs discovery: MetaMiner and NRPminer (Cao et al., 2019, Behsaz et al., 2021).
- SMMole output will help to prioritize *in silico* predicted PNPs for experimental validation and testing.

Thank you!

This study was supported by RSF (20-74-00032)

References

1. Behsaz B, Bode E, Gurevich A, Shi Y, Grundmann F, Acharya D, et al. Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. *Nat Commun.* 2021; 12: 3225.
2. Cao L, Gurevich A, Alexander K, Naman C, Leão T, Glukhov E, 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.
3. Dang, T., & Süßmuth, R. D. Bioactive Peptide Natural Products as Lead Structures for Medicinal Use. *Accounts of Chemical Research*, 2017, 50: 1566–1576.
4. Ertl, P., Roggo, S., & Schuffenhauer, A. (2007). Natural Product-likeness Score and Its Application for Prioritization of Compound Libraries. *Journal of Chemical Information and Modeling*, 48(1), 68–74. <https://doi.org/10.1021/ci700286x>
5. Filimonov, D. A., Lagunin, A. A., Glorizova, T. A., Rudik, A. v., Druzhilovskii, D. S., Pogodin, P. v., & Poroikov, V. v.. Prediction of the biological activity spectra of organic compounds using the pass online web resource. *Chemistry of Heterocyclic Compounds*, 2014, 50: 444–457.
6. Gurevich A, Mikheenko A, Shlemov A, Korobeynikov A, Mohimani H, Pevzner P. Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. *Nat Microbiol.* 2018; 3: 319-327.
7. Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., Van Der Hoof, J. J. J., Van Santen, J. A., Tracanna, V., Suarez Duran, H. G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S. L., Lund, G., Epstein, S. C., Sisto, A. C., Charkoudian, L. K., Collemare, J., Linington, R. G., . . . Medema, M. H. (2019). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkz882>
8. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2020; 49: D1388-D1395.
9. Kim, S., Bolton, E. E., Bryant, S. H. Similar compounds versus similar conformers: complementarity between PubChem 2-D and 3-D neighboring sets. *Journal of Cheminformatics*: 2016; 8: 1–17.