Discovering peptidic natural products by integrating computational mass spectrometry and genome mining

Liu Cao,[1*] Yi-Yuan Lee,[1] Bahar Behsaz,[1] and **Hosein Mohimani**[1]

[1]Computational Biology Department, Carnegie Mellon University, Pittsburgh

http://mohimanilab.cbd.cmu.edu/

Project Goals

Peptidic natural products (PNP) are a major source of signal molecules and drug leads [1]. The existing techniques for PNP discovery require isolation of bioactive molecules and structure elucidation, which are time consuming and expensive. Recent advances in high-throughput mass spectrometry (MS) and next generation sequencing has resulted in large MS/genomic datasets, which are gold mines for PNP discovery. However, currently there is no efficient algorithm to mine these datasets. We have developed computational tools to integrate MS/genomic data for automated discovery of PNPs from environmental isolates/communities.

HypoNPAtlas is a database of hypothetical natural products that is readily searchable against MS [2]. Seq2ripp predicts the structure of ribosomally synthesized and post-translationally modified peptides from microbial genome [2]. NRPminer integrates MS/genomic data to discover non-ribosomal peptides (NRP) [3]. MolDiscovery is a probabilistic model that efficiently searches small molecules MS [4]. These tools has enabled discovery of various novel PNP from public datasets. One of the NRP showed anti-parasite activity [3].

References

[1] Gordon M. Cragg, and David J. Newman. “Natural products: a continuing source of novel drug leads.” Biochimica et Biophysica Acta (BBA)-General Subjects 1830.6 (2013): 3670-3695.

[2] Yi-Yuan Lee, Haodong Liu, Neel Mittal, Liu Cao, Alexey Gurevich, and Hosein Mohimani. “hypoNPAtlas: an atlas of hypothetical natural product for mass spectrometry database search”. In preparation.

[3] Bahar Behsaz, Edna Bode, Alexey Gurevich, Yan-Ni Shi, Florian Grundmann, Deepa Archarya, Andrés Mauricio Caraballo-Rodríguez, Amina Bouslimani, Morgan Panitchpakdi, Annabell Linck, Changhui Guan, Julia Oh, Pieter C. Dorrestein, Helge B. Bode, Pavel A. Pevzner, and Hosein Mohimani. “Integrating Genomics and Metabolomics for Scalable Non-Ribosomal Peptide Discovery”. Provisionally accepted by Nature Communications.

[4] Liu Cao, Mustafa Guler, Azat Tagirdzhanov, Yiyuan Lee, Alexey Gurevich, Hosein Mohimani. “MolDiscovery: Learning Mass SpectrometryFragmentation of Small Molecules”. Submitted to Nature Communications, under revision.