

Top-down venomomics: a *de novo* sequencing approach

K. Vyatkina^{1, 2, 3, 4*}, D. Petras⁵

¹ Saint Petersburg State University, Saint Petersburg, Russia

² Saint Petersburg Academic University of the Russian Academy of Sciences, Saint Petersburg, Russia

³ ITMO University, Saint Petersburg, Russia

⁴ Saint Petersburg State Electrotechnical University "LETI", Saint Petersburg, Russia

⁵ University of California, San Diego, CA, USA

* e-mail: k.vyatkina@spbu.ru

Key words: proteomics, top-down mass spectrometry, snake venomomics, toxins, *de novo* sequencing

Motivation and Aim: Understanding of the composition of snake venoms can significantly contribute to drug discovery research, and is important for the development of effective antidotes. However, often the lack of comprehensive genome or transcriptome databases, as well as alternative splicing events and PTMs substantially complicate their analysis [1, 2]. The aim of this research is to develop an algorithmic framework for identifying new proteoforms of known toxins from top-down mass spectrometry through *de novo* sequencing. Efficiency of the proposed approach is illustrated for the venoms of green and black mamba.

Methods and Algorithms: The venom samples of green and black mamba were reduced with TCEP and analyzed by LC-MS/MS using an Agilent 1260 HPLC system coupled to a Thermo Orbitrap LTQ XL mass spectrometer, as described in [2]. The obtained top-down MS/MS spectra were processed with a modified version of the Twister *de novo* sequencing algorithm [3] to derive a number of highly accurate sequence tags of length 3. Those 3-tags were applied to match their underlying spectra against the 157 mamba toxin sequences available in the NCBI database, with a goal of identifying novel proteoforms of the known venom proteins. For each suggested identification, the spectra witnessing for it were annotated against the putative proteoform sequence.

Results: For the green mamba and black mamba venom samples, putative novel proteoforms were proposed for 17 and 27 toxins from the NCBI database, respectively. Upon a more thorough examination, appearance of 6 and 8 out of those, respectively, was attributed to misinterpretation of common PTMs. Among the remaining suggestions, 8 and 10, respectively, were confidently confirmed by the annotated mass spectra. A decision on the remaining hypotheses could presumably be made from native mass spectrometry of the venom samples.

Conclusion: The proposed technique for analyzing toxin families based on *de novo* sequencing of top-down tandem mass spectra has proven to be a handy tool for studying snake venoms.

Acknowledgements: The first author was supported by the RFBR (grant No. 16-54-21006).

References

1. Petras D. (2015) Venom proteomics of Indonesian king cobra, *Ophiophagus hannah*: Integrating top-down and bottom-up approaches. *J. Proteome Res.* 14(6):2539-2556.
2. Petras D. et al. (2016) Top-down venomomics of the East African green mamba, *Dendroaspis angusticeps*, and the black mamba, *Dendroaspis polylepis*, highlight the complexity of their toxin arsenals. *J. Proteomics.* 146:148-164.
3. Vyatkina K. et al. (2015) *De novo* sequencing of peptides from top-down tandem mass spectra. *J. Proteome Res.* 14(11):4450-4462.