DOI 10.18699/MM-HPC-BBB-2018-15

Method of reconstruction of a sequence of non-ribosomal peptides from mass spectra with noise

E. Fomin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: fomin@bionet.nsc.ru

Key words: algorithms, mass spectroscopy, sequences

Motivation and Aim: An important fraction of the peptidoma of bacteria is non-ribosomal peptides (NRP), representing a class of secondary peptide metabolites, usually produced by bacteria and fungi, and having an extremely wide range of biological activity and pharmacological properties. In the overwhelming majority of cases (73 %), NPFs have a complex nonlinear structure [1]. The monomers that make up the NRP have a wide variety of types (~500) and include, apart from 20 proteinogenic amino acids, non-proteinogenic amino acids and modified proteinogenic forms (methylated, glycosylated, D-forms) [2]. In connection with their biosynthesis from the non-bryosomal path, the identification of NPF by classical methods of bioinformatics and genomics is impossible, and is carried out only on the basis of mass spectrometry. At present, the possibilities of de novo reconstruction of the structure of complex NRF from mass spectra are limited. Thus, the development of new bioinformatic methods for the reconstruction of bacterial non-ribosomal peptides is very relevant.
Methods and Algorithms: Previously, we proposed a new method for solving the problem of reconstruction of a sequence of cyclic peptides from mass spectra, based on the removal of redundancy from the spectra [1,2].We made a computer implementation of the method on the assumption that there were no noises or omissions in the spectra. The high efficiency of the proposed method was shown.
Results: In this work, the next step in de novo reconstruction of a sequence of cyclic peptides from mass spectra is made. A generalization of the previously proposed method was constructed by using continuous integral transformations. It is shown that the method makes it possible not only to significantly reduce the additive noise, that is, independent of the signal, in the initial data, but also to restore the omissions in the data.

References

1. Caboche S., Pupin M., Leclère V., Fontaine A., Jacques P., KucherovG. (2008) Norine: a database of nonribosomal peptides. Nucleic Acids Res. 36:D326-D331.
2. Caboche S., Leclère V., Pupin M., Kucherov G., Jacques P. (2010) Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. Journal of bacteriology, 192(19): 5143-5150
3. Fomin E. (2016) A Simple Approach to the Reconstruction of a Set of Points from the Multiset of n^2 Pairwise Distances in n^2 Steps for the Sequencing Problem: I. Theory. J. Comput. Biol, 23(9):769-75;
4. Fomin E. (2016) A Simple Approach to the Reconstruction of a Set of Points from the Multiset of n^2 Pairwise Distances in n^2 Steps for the Sequencing Problem: II. Algorithm J. Comput. Biol. 23(12):934-942.