4D-Dynamic Representation of DNA/RNA Sequences - A New Bioinformatics Method

Dorota Bielińska-Wąż[a], Piotr Wąż[b]

[a] Department of Radiological Informatics and Statistics, Medical University of Gdańsk, 80-210 Gdańsk, Poland
[b] Department of Nuclear Medicine, Medical University of Gdańsk, 80-210 Gdańsk, Poland

A new method, *4D-Dynamic Representation of DNA/RNA Sequences*, aiming at similarity/dissimilarity analysis of biological sequences, has been formulated. It belongs to a group of non-standard bioinformatics approaches called alignment-free methods. In the new method, sequences are represented by sets of material points in a 4D space - *4D-dynamic graphs*. The method is a generalization of our previous 2D and 3D approaches [1, 2]. In particular, 2D and 3D methods have been applied for the characterization of the complete genome sequences of viruses [3, 4, 5]. We call the methods dynamic, because the graphs are characterized by some values analogous to the ones used in the classical dynamics. In this work 4D moments of inertia are proposed as numerical characteristics (descriptors) of the 4D-dynamic graphs representing the sequences. 2D and 3D projections of the 4D-dynamic graphs are proposed as graphical representations of the sequences. 4D-Dynamic Representation of DNA/RNA Sequences has been applied to an analysis of the complete genome sequences of the 2019 novel coronavirus, available in March 2020 in GenBank. The proposed descriptors of the 4D-dynamic graphs proved to be very good. 4D moments of inertia correctly classify the sequences. The descriptors representing complete genome sequences of Deltacorovirus and of Betacoronavirus are located in different parts of the classification maps. The detailed classification of Betacoronaviruses to Embevovirus and the 2019 novel coronavirus is also recognized by the method. The descriptors representing Embevovirus and the 2019 novel coronavirus are also located in different parts of the maps.

References

[1]. Bielińska-Wąż, D.; Clark, T.; Wąż, P.; Nowak, W.; Nandy, A. 2D-dynamic representation of DNA sequences. Chem. Phys. Lett. **2007**, *442*, 140–144.

[2]. Wąż, P.; Bielińska-Wąż, D. 3D-dynamic representation of DNA sequences. J. Mol. Model. 2014, 20, 2141.

[3]. Panas, D.; Wąż, P.; Bielińska-Wąż, D.; Nandy, A.; Basak, S.C. 2D–Dynamic Representation of DNA/RNA Sequences as a Characterization Tool of the Zika Virus Genome. MATCH Commun. Math. Comput. Chem. **2017**, *77*, 321–332.

[4]. Panas, D.; Wąż, P.; Bielińska-Wąż, D.; Nandy, A.; Basak, S.C. An Application of the 2D-Dynamic Representation of DNA/RNA Sequences to the Prediction of Influenza A Virus Subtypes. MATCH Commun. Math. Comput. Chem. **2018**, *80*, 295-310.

[5]. Bielińska-Wąż, D.; Panas, D.; Wąż, P. Dynamic Representations of Biological Sequences. MATCH Commun. Math. Comput. Chem. **2019**, *82*, 205-218.