20D-dynamic Representation of Protein Sequences

Agata Czerniecka[1], Dorota Bielińska-Wąż[1], Piotr Wąż[2]
[1] Department of Radiological Informatics and Statistics,
Medical University of Gdańsk,
Tuwima 15, 80-210 Gdańsk, Poland
agata.czerniecka@gumed.edu.pl
djwaz@gumed.edu.pl
[2] Department of Nuclear Medicine, Medical Universiy of Gdańsk,
Tuwima 15, 18-210 Gdańsk, Poland
phwaz@gumed.edu.pl

Keywords: Similarity/dissimilarity analysis of protein sequences, Descriptors, Alignment-free methods, Moments of inertia.

A new method of comparison of protein sequences called by us "20D-dynamic Representation of Protein Sequences" is presented [1]. This method belongs to the group of *alignment-free methods*. This class of methods constitute a fast developing branch of bioinformatics. The alignment-free methods reveal different aspects of similarity with low computational effort. They offer both numerical and graphical tools for sequence comparisons. In the 20D-dynamic Representation of Protein Sequences, the sequences of amino acids are represented by a set of point masses in a 20D space. The distribution of points in the space is obtained by applying the method of "a walk" in the 20D space. As a consequence a 20D-dynamic graph is obtained which is a mathematical representation of the protein sequence. The 3D and 2D projections of this graph are convenient graphical tools for sequence comparison. The descriptors, i. e. the numerical values characterizing the 20D-dynamic graphs, are designed using the ideas coming from the classical dynamics (moments of inertia of rigid bodies). Some numerical examples of this method, using NADH dehydrogenase subunit 5 (ND5) protein sequences of nine species and subunit 6 (ND6) of eight species, are presented.

References

[1] A. Czerniecka, D. Bielińska-Wąż, P. Wąż, T. Clark, *20D-dynamic representation of protein sequences*, Genomics **107** (2016) 16–23.