



Similarity analysis of protein sequences based on moment of inertia

K. Lakshmi, Mayukha Pal, and P. Manimaran*

C R Rao Advanced Institute of Mathematics, Statistics and Computer Science,
University of Hyderabad Campus, Hyderabad-50046.

Email: pa.manimaran@gmail.com

Abstract

Studying the protein structure is the basis for spying into its function. Tracing back the evolutionary relationship of proteins is a part of functional annotation. In this paper, we study the sequence similarity of Albumin proteins obtained from different organism. The protein sequences are considered as a rigid body with mass and transformed into vectors by the tensor of moment of inertia. From the calculated Eigen vectors, the Euclidean distance between any two sequences were obtained. The closeness between the organisms is inferred from the constructed phylogenetic tree.

Key words: Protein; Phylogeny; Moment of inertia; Similarity Analysis; Sequences.

References

1. Bielińska-Wa, D., Clark, T., Wa, P., Nowak, W. & Nandy, A. (2007). 2D-dynamic representation of DNA sequences, *Chemical Physics Letters*, 442(1), 140-144.
2. Brown, J. R. (1976). Structural origins of mammalian albumin, *Federation proceedings*, 35(10), 2141-2144.
3. Carter, D. C. & Ho, J. X. (1994). Structure of serum albumin, *Advances in protein chemistry*, 45, 153-203.
4. El-Lakkani, A. & El-Sherif, S. (2013). Similarity analysis of protein sequences based on 2D and 3D amino acid adjacency matrices, *Chemical Physics Letters*, 590, 192-195.
5. El-Lakkani, A. & Mahran, H. (2015). An efficient numerical method for protein sequences similarity analysis based on a new two-dimensional graphical representation, *SAR and QSAR in Environmental Research*, 26(2), 125-137.
6. El-Maaty, M. I. A., Abo-Elkhier, M. M. & Elwahaab, M. A. A. (2010). 3D graphical representation of protein sequences and their statistical characterization, *Physica A*, 389(21), 4668-4676.
7. Fiume, L., Busi, C., Mattioli, A. & Spinosa, G. (1988). Targeting of antiviral drugs bound to protein carriers. *Critical reviews in therapeutic drug carrier systems*, 4(4), 265-284.
8. Gorman, G. C., Wilson, A. C. & Nakanishi, M. (1971). A biochemical approach towards the study of reptilian phylogeny: evolution of serum albumin and lactic dehydrogenase, *Systematic zoology*, 20(2), 167-185.
9. Hamori, E. & Ruskin, J. (1983). H curves, a novel method of representation of nucleotide series especially suited for long DNA sequences, *The Journal of Biological Chemistry*, 258(2), 1318-1327.
10. He, P. A., Li, D., Zhang, Y., Wang, X. & Yao, Y. (2012). A 3D graphical representation of protein sequences based on the Gray code, *Journal of Theoretical Biology*, 304, 81-87.
11. He, P. A., Wei, J., Yao, Y. & Tie, Z. (2012). A novel graphical representation of proteins and its application, *Physica A*, 391(1), 93-99.
12. Hou, W., Pan, Q. & He, M. (2016). A new graphical representation of protein sequences and its applications, *Physica A*, 444, 996-1002.
13. Jafarzadeh, N. & Iranmanesh, A. (2012). A novel graphical and numerical representation for analyzing DNA sequences based on codons, *MATCH Communications in Mathematical and in Computer Chemistry*, 68(2), 611-620.
14. Jeffrey, H. J. (1990). Chaos game representation of gene structure, *Nucleic Acids Research*, 18(8), 2163-2170.
15. Jeffrey, H. J. (1992). Chaos game visualization of sequences, *Computers & Graphics*, 16(1): 25-33.
16. Kato, A., Rooney, A. P., Furutani, Y. & Hirose, S. (2010). Evolution of trappin genes in mammals, *BMC evolutionary biology*, 10(1), 31.
17. Kragh-Hansen, U. (1981). Molecular aspects of ligand binding to serum albumin, *Pharmacological reviews*, 33(1), 17-53.
18. Li, Y., Liu, Q., Zheng, X. & He, P. A. (2014). UC-Curve: A highly compact 2D graphical representation of protein sequences, *International Journal of Quantum Chemistry*, 114(6), 409-415.
19. Liao, B. & Wang, T.M. (2004). New 2D graphical representation of DNA sequences, *Journal of Computational Chemistry*, 25(11), 1364-1368.
20. Liao, B., Liao, B., Sun, X. & Zeng, Q. (2010). A novel method for similarity analysis and protein sub-cellular localization prediction, *Bioinformatics*, 26(21), 2678-2683.
21. Liao, B., Zhang, Y., Ding, K. & Wang, T.M. (2005). Analysis of similarity/dissimilarity of DNA sequences based on a condensed curve representation, *Journal of Molecular Structure: THEOCHEM*, 717(1), 199-203.
22. Ma, T., Liu, Y., Dai, Q., Yao, Y. & He, P. A. (2014). A graphical representation of protein based on a novel iterated function system, *Physica A*, 403, 21-28.



23. Mendez, C. M., Mc Clain, C. J. & Marsano, L. S. (2005). Albumin therapy in clinical practice, *Nutrition in clinical practice*, 20(3), 314-320.
24. Nandy, A. (1994). A new graphical representation and analysis of DNA sequence structure: I. Methodology and application to globin genes, *Current Science*, 309-314.
25. Pal M., Satish, B., Srinivas, K., Madhusudana Rao P., & Manimaran, P. (2015). Multifractal detrended cross-correlation analysis of coding and non-coding DNA sequences through chaos-game representation, *Physica A*, 436, 596-603.
26. Pal M., Satya Kiran, V., Madhusudana Rao P., & Manimaran P. (2016). Multifractal detrended cross-correlation analysis of genome sequences using chaos-game representation, *Physica A*, 456, 288-293.
27. Peters, T. (1985). Serum albumin, *Advances in protein chemistry*, 37, 161-245.
28. Randić, M., Vračko, M., Lerš, N. & Plavšić, D. (2003). Novel 2-D graphical representation of DNA sequences and their numerical characterization, *Chemical Physics Letters*, 368(1), 1-6.
29. Randić, M. (2004). 2-D graphical representation of proteins based on virtual genetic code, SAR and QSAR in *Environmental Research*, 15(3), 147-157.
30. Randić, M., Butina, D. & Zupan, J. (2006). Novel 2-D graphical representation of proteins, *Chemical Physics Letters*, 419(4), 528-532.
31. Waz, P. & Bielińska-Waz, D. (2013). Moments of inertia of spectra and distribution moments as molecular descriptors, *MATCH Communications in Mathematical and in Computer Chemistry*, 70(3), 851-865.
32. Wąż, P. & Bielińska-Wąż, D. (2014). 3D-dynamic representation of DNA sequences, *Journal of Molecular Modeling*, 20(3), 2141.
33. Wąż, P., Bielińska-Wąż, D. & Nandy, A. (2014). Descriptors of 2D-dynamic graphs as a classification tool of DNA sequences, *Journal of Mathematical Chemistry*, 52(1), 132-140.
34. Wen, J. & Zhang, Y. (2009). A 2D graphical representation of protein sequence and its numerical characterization, *Chemical Physics Letters*, 476(4), 281-286.
35. Xu, S. C., Li, Z., Zhang, S. P. & Hu, J. L. (2014). Primary structure similarity analysis of proteins sequences by a new graphical representation, SAR and QSAR in *Environmental Research*, 25(10), 791-803.
36. Yao, Y., Yan, S., Xu, H., Han, J., Nan, X., He, P. A. & Dai, Q. (2014). Similarity/dissimilarity analysis of protein sequences based on a new spectrum-like graphical representation, *Evolutionary Bioinformatics*, 10, 87-96.
37. Yau, S. S. T., Wang, J., Niknejad, A., Lu, C., Jin, N. & Ho, Y. K. (2003). DNA sequence representation without degeneracy, *Nucleic Acids Research*, 31(12), 3078-3080.
38. Yau, S. S. T., Yu, C. & He, R. (2008). A protein map and its application, *DNA and Cell Biology*, 27(5), 241-250.