Cloud Publications
International Journal of Advanced Bioinformatics and Computational Biology
2014, Volume 1, Issue 1, pp. 31-34, Article ID Med-176

Research Article | Open Access

A Tool for Composition and Representation of Amino Acids of a Protein Sequence

Farhana Sarkhawas and R.K. Kamat

Department of Computer Science, Shivaji University, Kolhapur, Maharashtra, India

Correspondence should be addressed to Farhana Sarkhawas, farhana.ap@gmail.com

Publication Date: 4 October 2014

Article Link: http://medical.cloud-journals.com/index.php/IJABCB/article/view/Med-176

Abstract The tool developed is a graphical representation of protein sequences, constructed based on the composition of amino acids and was displayed in the form of bar graph and line graph. There are many visualization tools available which displays the amino acids in the form of logos, p-h curves, coil etc. which are incomprehensible. The tool described in this paper first calculates the composition of amino acids given a protein sequence and then displays the bar and line graph. The graphical representation displayed by this tool not only avoids decadence, but also has simple and clear visualization no matter how long the sequences are.
Keywords *Amino Acids; Graph; Protein Sequence; Bar Graph; Line Graph*

1. Introduction

Amino acids act as building blocks of proteins and as intermediates in metabolism. The 20 amino acids that are found within proteins convey a vast range of chemical flexibility. The sequence of amino acids of a specific protein and the precise amino acid content is determined by the sequence of the bases in the gene that encodes that protein. Amino acids can be defined and categorized in a number of ways, depending on the perspective they are being examined from each time. Proteins are a combination of 20 different amino acids, which have been selected during evolution from a much larger group of possibilities that exists in nature. Protein sequences are constructed from the residues of 20 amino acids, and most proteins with a sequence length of 200 amino acids or more contain all 20. Several visualization tools exist such as Seq2Logo [1], LogoBar [2], BiasViz [3], IMGT Tool [4] and Composition Profiler [5] which represent the composition of amino acids in the form of sequence logo, bars, plots with prolines, beta strands and bars with error in the form of loops, respectively. These tools mainly take multiple sequence alignment file as input and study the complexity in gaps and also make the visualization difficult to understand.

The current study easily visualizes the range of the residues of a protein sequence through the graph which easily helps the biologists to grasp which bar or which line rises high indicating the maximum of the amino acid at that point, similar with the minimum. Considering the importance of proteins as a

structural component of all living organisms, the significance of a method or tool that could seamlessly manipulate this data would be extremely practical. Indeed, biologists have been using computational tools that enable them to analyze and examine amino acid sequences in a number of ways.

2. Methodology

The tool presented in this paper is an easy-to-use, open source java application used to calculate the composition of amino acids of a protein sequence and displays the graphical representation of the protein sequence in the form of a bar graph and line graph. With this application you can compute and generate composition and representation of amino acids from a protein sequence on your own computer or laptop. The protein sequence is entered in a text area; the composition of amino acids is shown in a JTable whereas the graphs are displayed in separate JPanels. The tool is written using javax.swing functions and requires a java development kit to execute the application. It has been tested on Windows 7 with JDK1.6. The tool runs on any platform provided the JVM is installed.

3. Results and Discussion

The input to the tool is a protein sequence in FASTA format to be entered in a text area which outputs the composition of amino acids (Figure 1) in a JTable which is displayed to the right side of the window frame. The graphical output of the tool is a line graph (Figure 2) and a bar graph (Figure 3) composed of data points, one for each amino acid, where bar heights indicate the frequency of the residues and the lines shows trends in the data clearly. The two dimensions of the graph are the x-axis represents the continuous variable i.e. the residues whilst the y-axis has a scale and indicates the measurement. The maximum number of residues can be found easily by looking at the highest bar height in the bar graph or the top most point in the line graph; it goes same with the minimum number of residues to be known. The window size and the axes are displayed in real time as soon as the check boxes for the graphs are clicked. This graphical representation enables the user to make predictions about the results of data not yet recorded.

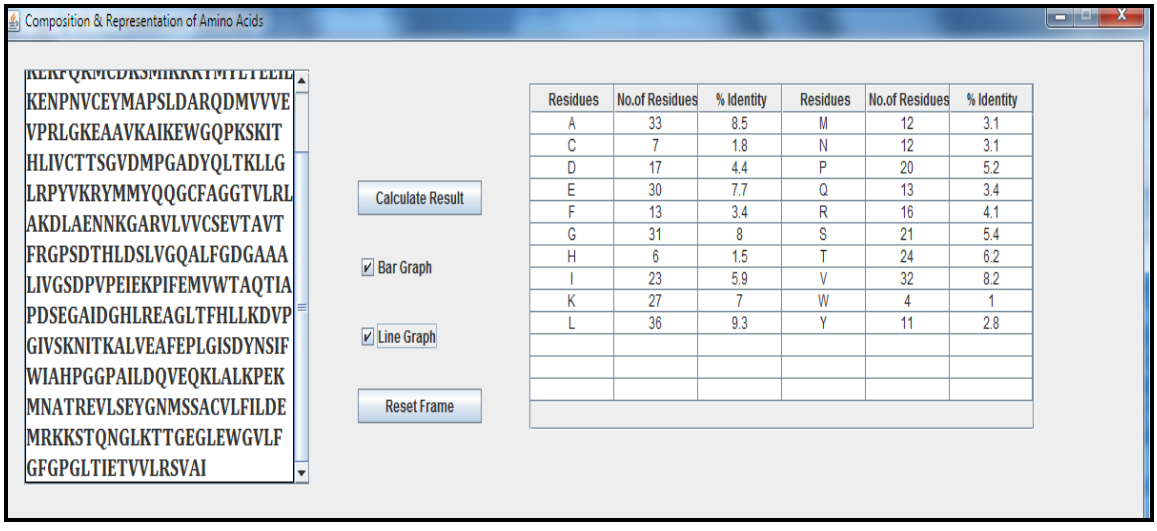

Figure 1: *Composition of Amino Acids*

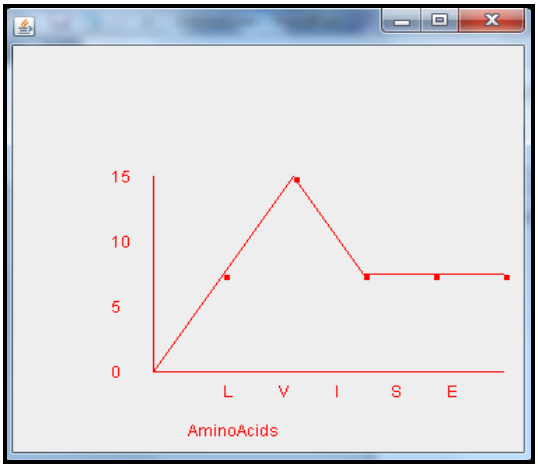

Figure 2: *Line Graph of Amino Acid Composition*

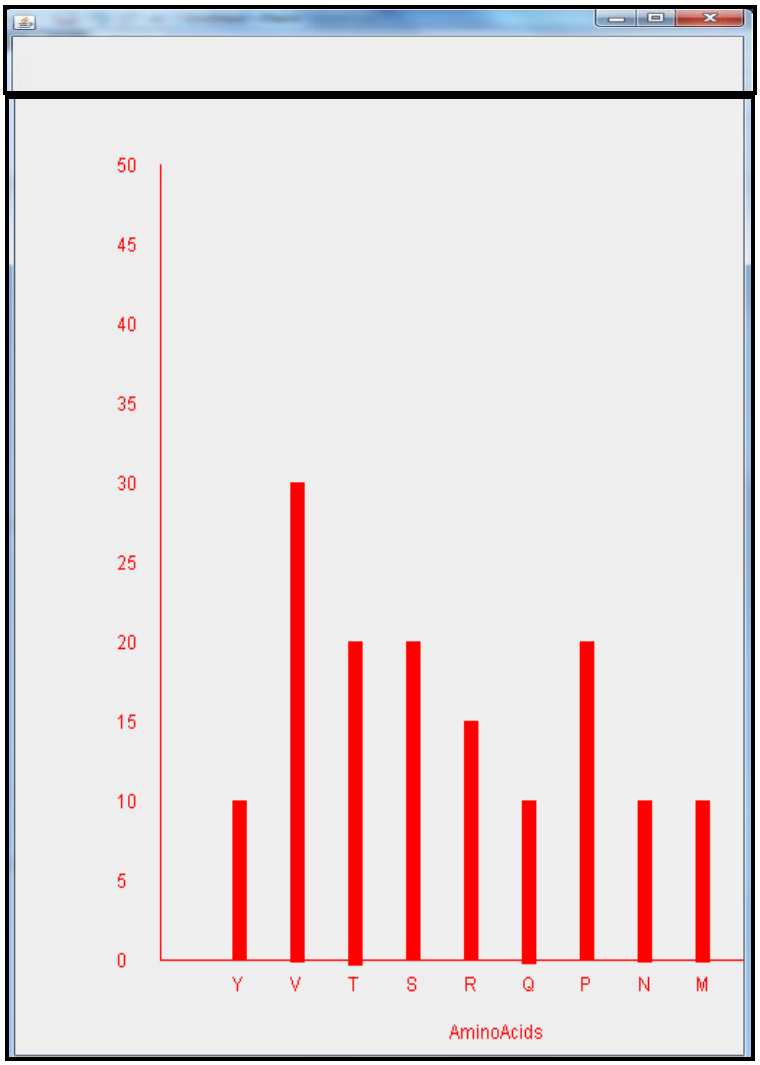

Figure 3: *Bar Graph of Amino Acid Composition*

4. Conclusion

The tool described here displays the bar graph and line graph constructed from the composition of amino acids of a protein sequence. The graphs shown make comparisons between different residues which are easy to examine. They clearly show the data drift, meaning the rise or falls of the residues present in the amino acids. The graph displays information in a way that helps us to make generalizations and conclusions quickly and easily.

References

[1] Martin, C.F.T., and Morten, N. *Seq2Logo: A Method for Construction and Visualization of Amino Acid Binding Motifs and Sequence Profiles Including Sequence Weighting, Pseudo Counts and Two-Sided Representation of Amino Acid Enrichment and Depletion.* Nucleic Acids Research. 2012. 40; 281-287.

[2] Perez-Bercoff, A., Koch, J., and Burglin, T.R. *LogoBar: Bar Graph Visualization of Protein Logos Withgaps*. Bioinformatics. 2006. 22; 112-114.

[3] Huska, M.R., Buschmann, H., and Andrade-Navarro, M.A. BiasViz: Visualization of Amino Acid Biased Regions in Protein Alignments. Bioinformatics. 2007. 23; 3093-3094.

[4] Dimitrios, V., Christos, F., Vasileios, M., and Sophia, K. *IMGT/Collier-de-Perles: A Two-Dimensional Visualization Tool for Amino Acid Domain Sequences.* Theoretical Biology and Medical Modelling. 2013. 10; 10-14.

[5] Vladimir, V., Vladimir, N.U., Keith, D., and Stefano, L. Composition Profiler: A Tool for Discovery and Visualization of Amino Acid Composition Differences. Bioinformatics. 2007. 8; 1471-2105.