Asian Journal of Convergence in Technology *Volume III, Issue III*
ISSN No.:2350-1146, I.F-2.71

An Android Java Application for Evaluating Physico-chemical Properties of a Protein Sequence

Farhana Desai
Department of Computer Science, Symbiosis Institute of Computer Studies and Research, Pune, Maharashtra, India

R. K. Kamat
Department of Electronics, Shivaji University, Kolhapur, Maharashtra, India

Abstract— **The study describes several physicochemical properties of amino acids which can help to study biological functions, profiles and folding of proteins. The java based tool calculates seven different statistical parameters namely Isoelectric Point, Molecular Weight, Aliphatic index, GRAVY (Grand average of hydropathicity), Extinction Coefficient, Aromaticity Score and Instability index for fasta input protein sequence. The tool has the option to save the result in an auto generated filename and also finds the time required to calculate the result. The tool runs on android phones by using the JBlend application. The tool is available at https://usegalaxy.org/SequenceStatistics.**

Keywords: amino acid, java, protein, mobile

I. INTRODUCTION

Proteins occupy a prominent place in all biological systems, both qualitatively and quantitatively. Regardless of their structural or functional role, they all are formed from the fundamental building blocks, called amino acids. These amino acids sequences (primary structures) determine the secondary and tertiary structures of proteins and specify their biological properties. Combining all the physicochemical properties of amino acids will help to solve the problems related protein sequence comparison , macroscopic properties of proteins (such as aggregation or understand the conserved regions in protein sequences. The tools such as PepStats and SAPS are described in EMBL-EBI [1, 2] bioinformatics web and programmatic tools framework. These are online tools which require an internet connection with a compatible web browser to run.

This paper describes a very user-friendly and platform independent java tool for your android mobile to study the statistical analysis of a protein sequence. Since most of the users are smartphone users, the current tool is implemented on android phones so that the user can access at any point of time. The tool is built using java swing components and tested on windows and Ubuntu as well. The current tool is capable of calculating the amino acid composition, various physicochemical properties namely Isoelectric Point, Molecular Weight, Aliphatic index, GRAVY (Grand average of hydropathicity), Extinction Coefficient, Aromaticity Score and Instability index, saving the result in the current directory and calculating the time for result generation.

II. METHODOLOGY

The current tool is an easy-to-use, open source java application used to determine the statistical properties of a protein sequence. The tool is written in java programming language and requires a java development kit (Jdk1.6) to execute on windows, whereas there are no external files to be installed on Ubuntu to execute the tool. It makes use of the strong components of javax.swing package. The java development kit is available at http://www.oracle.com/technetwork/java/javase/downloads/index.html. It has been tested on Windows 7 and Ubuntu. The tool to work on any android mobile requires a JBlend application.

Algorithms

There are several standard algorithms derived to calculate the physicochemical properties of proteins. The various physicochemical properties evaluated by this tool includes Isoelectric Point [3], Molecular Weight, Extinction Coefficient [4,5], GRAVY [6], Aliphatic Index [7], Aromaticity Score[8] and Instability index[9] and Amino acid Composition with percent residues [10].

Graphical User Interface

The graphical user interface of the tool is designed using the Java swing components and the layout manger. The Java Foundation Classes (JFC) helps to build a rich graphical user interface using its strong features like swings. The JFC provides the GUI with strong graphical functionality and interactivity feature. Since the swing components are considered to be light weight components the look and feel of the screen is very smooth. The Swing GUI components include everything from buttons, tabbed panes to tables. The swing package in java is flexible and powerful. The layout manager of awt package helps to position the components in a container (i.e window) and adjusts when the window resizes. The tool is designed using JPanels, JButton, JLabes, JTextField, JTextArea and JTables of swings. The components are arranged on the window by using the BorderLayout and GridLayout managers. As seen in Figure 1, the tool workflow is divided into four parts : 1) TextArea to display protein

sequence 2) Buttons to select result 3) Textboxes to view result and 4) Table to display amino acid composition.

Implementation

The tool named as SequenceStatistics.jar (https://usegalaxy.org) is freely available to any user. The tool requires a JDK (http://www.oracle.com/technetwork/java/javase/downloads) environment for its smooth functioning. The users using Ubuntu can directly run the SequenceStatistics.jar file by double click, whereas, to access the tool on your smartphones a JBlend (https://sourceforge.net/projects/jblend/) application need to be installed or a java manager which executes the jar fileAs soon as you open the JBlend it has option as "Run the jar file", which propmts you to select the jar file. The result protein statistics is calculated is calculated and displayed in text boxes and the amino acid composition is shown in JTable .

III. RESULTS AND DISCUSSION

PropCalc[11] is one of the tools that calculate four of the physicochemical properties as compared to seven properties calculated by the current tool along with the option to save the result. The tool computes the time required for the output which the other tools don't have the option for it. The GUI of the tool is build using the JDK environment which makes its quite user friendly as compared to the other tools such as ProPas[12] and SeqCalc [13] which is developed using Perl programming. The online tools have disadvantage that they need an internet connection with a compatible web-browser. These tools don't have the option to save the result.

Input Data

The sequence to be analyzed for its properties is typed or pasted in the left window of the frame. The sequence needs to be an only amino acid sequence and the format supported is fasta. The text area accepts only a single sequence. If it's a fasta file the line containing the sequence name with ">" character is skipped and starts reading the sequence from the second line. The length of the input sequence depends on the size and the speed of one's computer. A default protein sequence can also be used to analyze the result which is displayed in the textfield as soon as the tool executes. The JTextArea component used to display the protein sequence automatically includes the horizontal and vertical scroll bars as the length of the protein sequences increases the default number of rows of the JTextArea Component.

Output Data

The result of the calculated statistical properties of the given sequence is displayed in the respective textboxes at the right side of the frame. A dialog box is popped up if any of the values is not calculated due to some missing amino acid. Further, the time required to calculate the properties is displayed in milliseconds. The System class of java has the method currentTimeMillis() to calculate the time in milliseconds. The HasMap of Collections interface is used to save the default amino acid composition where the key is the residue name and the value is its molecular weight. The amino acid composition along with its percent identity is shown in the table which is at the bottom of the frame. The concept of streams in java helps to write the contents in the new file as well as save it with a new name every time. The given protein sequence and the result are saved in the current folder with an auto-generated filename. The filename is generated as "ProtStat" with a random number postfixed (eg. ProtStat25). Once the "Save" button is clicked a dialog box is popped saying "Your Output is Saved in the file ProtStat25".If the user wants to test for another sequence, a reset button is available which clears all the data on the frame, so that a new protein sequence can be entered.

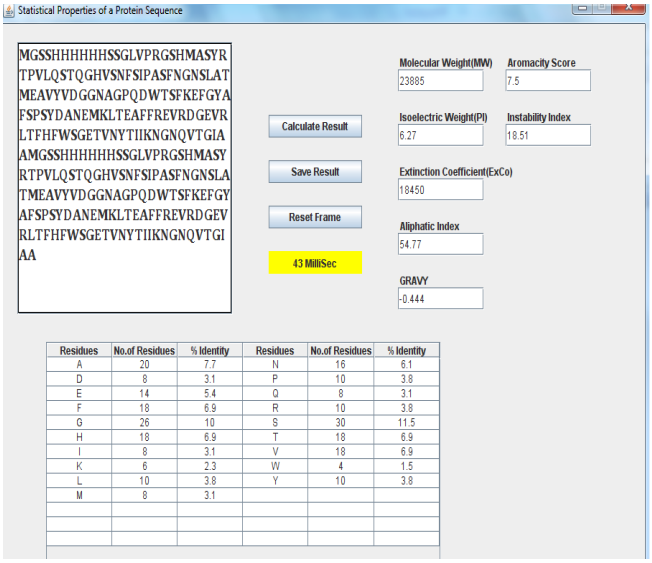

Figure 1: The Graphical User Interface of the tool

CONCLUSION

We present here a new tool with an efficient and generic design to calculate maximum physicochemical properties of a protein sequence. The biologists and researchers can install and execute the tool easily on windows, Ubuntu and android phones. Whereas running the tool on your android phone can save time and allow anywhere access. It is under active development to add more statistical analysis related to a protein and work for multiple protein formats. Besides, we are planning to integrate the current tool with STRAP.

ACKNOWLEDGEMENT

I thank Dr. Haridas Acharya for his guidance and support.

REFERENCES

[1] "The EMBL-EBI bioinformatics web and programmatic tools framework.," Nucleic acids research, vol. W143, pp. 580-4, 2015 .
[2] "Analysis Tool Web Services from the EMBL-EBI. Nucleic acids research ," 41, pp. 597-600, 2013.
[3] B. Bjellqvist, K. Ek, P.G. Righetti, E. Gianazza, A. Gorg, R. Westermier, W. Postel, "Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications.," J.Biochem.Biophys. Methods, vol. 6, pp. 317-339, 1982.
[4] S.C. Gill and P.H. Von Hippel, "Calculation of protein extinction coefficients from amino acid sequence data," Anal. Biochem., vol. 182, pp. 319-326, 1989.
[5] H. Edelhoch, "Spectroscopic determination of tryptophan and tyrosine in proteins.," Biochemistry, vol. 6, pp. 1948-1954, 1967.
[6] J. Kyte, R.F. Doolittle, "A simple method for displaying the hydropathic character of a protein.," J.Mol. Biol. , vol. 157, pp. 105-32, 1982.
[7] A.J. Ikai, " Thermostability and aliphatic index of globular proteins.," J. Biochem., vol. 88, pp. 1895-1898, 1980.
[8] K. Guruprasad, B. V. Reddy , "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence." Protein Eng vol. 4, pp. 155-61, 1990.
[9] J.R. Lobry and C. Gautier, "Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 Escherichia coli chromosome-encoded genes.," Nucleic Acids Research, vol. 22, pp. 3174-3180, 1994.
[10] C. Oliviero, " Amino acid composition and protein dimension.," Protein Science, vol. 17, pp. 2187–2191, 2008.
[11] C. Boobalan and N. Bharathi, " ProPCalc–a tool for calculating the protein properties," International Journal of Advanced Computer and Mat hematical Sciences ,vol. 3, pp. 205-208, 2012.
[12] W. Songfeng and Yunping Z. "ProPAS: standalone software to analyze protein properties.," Bioinformation, vol. 8, pp. 167-169, 2012.