RESEARCH　ARTICLE

ProGene 1.0-An *In Silico* Tool for Protein-Gene Analysis

Princy Vijayababu[1]*, Gopinath Samykannu[2], SundaraBaalaji Narayanan[3]

Structural Biology Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, India

*Address for Correspondence: Princy Vijayababu, Research Scholar, Structural Biology Laboratory, Department of Bio-informatics, Bharathiar University, Coimbatore, Tamil Nadu, India

ABSTRACT- Online Bioinformatics tool needs access through the World Wide Web, which requires Internet service. To overcome this we programmed a new offline tool, ProGene1.0 for advanced protein and nucleotide analysis. ProGene 1.0 developed by programming language Microsoft Visual Basic 6.0 designed to predict chemical properties, Molecular mass, Isoelectric point, Functional group variation, Molecular composition, Nucleotide composition and Nucleotide to protein conversion. ProGene 1.0 is a desktop-based tool that allows researchers to get the sequence level information on a novel and unknown or hypothetical protein or nucleotide.

INTRODUCTION

With the number of interdisciplinary subject in life science, Bioinformatics is an emerging field interface of computer science mathematics and biology [1]. It is a foundation of engineering depends of both experimental and derived data. Nucleotide and protein sequence data has considerably increased this pool of data [2,3]. Sequence data and subsequence analysis by a bioinformatics tool is attracts most attention, because it promises a considerable reduction in both time and cost [4]. Protein sequence comparison has become one of the most powerful tools for characterizing protein sequences because of the enormous amount of information that is preserved throughout the evolutionary process [5]. A general approach for functional characterization of unknown proteins is to infer protein functions based on sequence similarity. One of the successful approaches is to define signatures of known families. Signatures usually identify conserved regions among the family of proteins, revealing the importance of their structural or physicochemical properties [6]. Relations between protein sequence and structure can be analyzed by either determining the sequence features of predefined structures and properties [7]. Due to the advancement in Bi-oinformatics, many tools were available for analysis of protein and nucleotide sequences [8].

One of the major drawbacks of those tools is their web based application which needs to be accessed through the World Wide Web [9].

MATERIALS AND METHODS

To overcome the several drawbacks, after a complete analysis of existing protein and nucleotide analysis tools ProGene 1.0 was designed as standalone desktop application. The overall architecture is shown in Fig 1.

Visual Basic is a language for developing graphical user interface [10]. It is an event driven programming language where the program is not based on code, but it was more based on the event given [11]. Because of this, it is mainly used to develop tools and software for Bioinformatics more preferably as offline application. Most standalone Bioinformatics software has been developed with VB only [12]. ProGene 1.0 is also based on this language.

RESULTS

ProGene 1.0–Interface

ProGene 1.0 was basically developed to get the information about a novel, unknown or hypothetical protein and nucleotide sequences. The home page of ProGene 1.0 has horizontal tool bar comprising four options, namely File, Edit, Tool, and Help. Tool menu has two submenus i.e Protein Analyzer and Gene Analyzer. The Protein Analyzer has four submenus like Chemical properties, Functional Group Variation, Amino acid Composition and Molecular Composition as shown in Fig 2. Gene Analyzer has two submenus Sequence converter and Nucleotide Composition.

Access this article online	
Quick Response Code	**Website:** www.ijlssr.com
	crossref **DOI:** 10.21276/ijlssr.2017.3.3.21

The submenus interface contains Browse button (browse the input anywhere from personal computer), Clean (reset all information), Save (to save all output in text file format). Users can input the hypothetical protein sequence on "input area" by the help of browse button from PC (Personal Computer) or paste; subsequently select the necessary parameters by selecting check box. Invalid inputs are shown by critical message. Click on "compute" button to get output. Save the output in".txt" format anywhere in the PC using save option. The hierarchy capability of ProGene 1.0 is shown in Fig 3. A sample submenu functional properties output page was illustrated in Fig. 4.
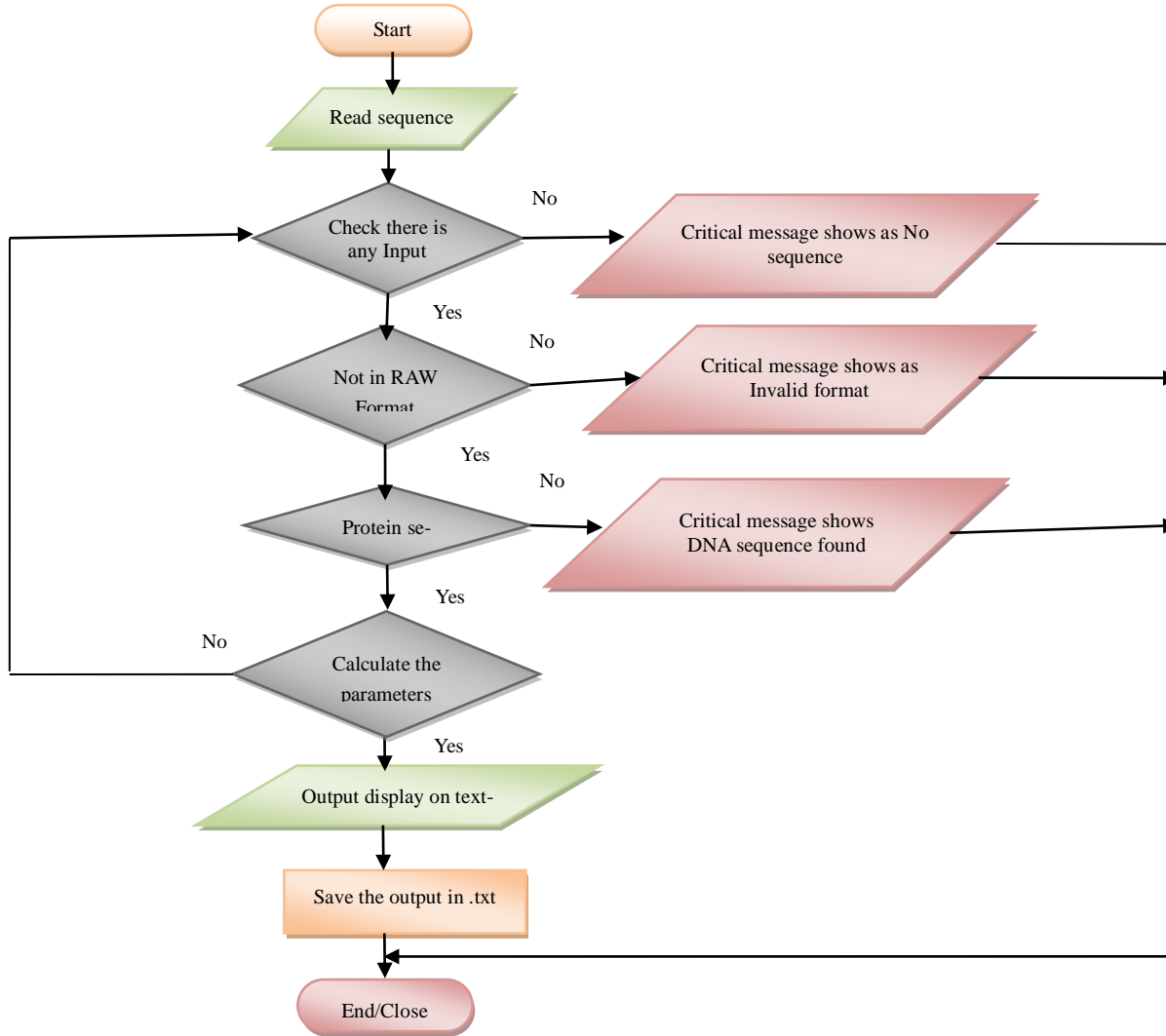

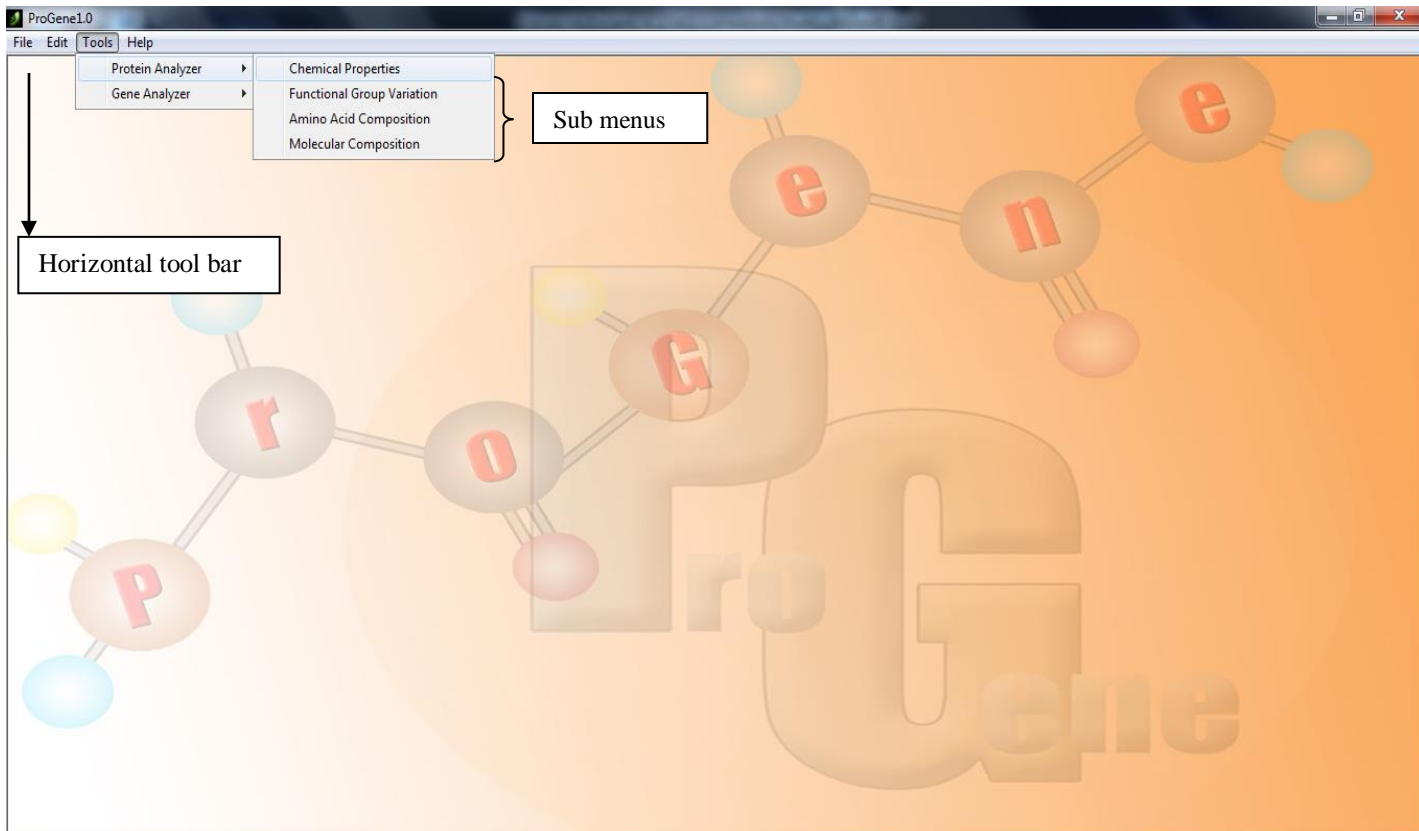**Fig. 1:** Architecture of ProGene 1.0
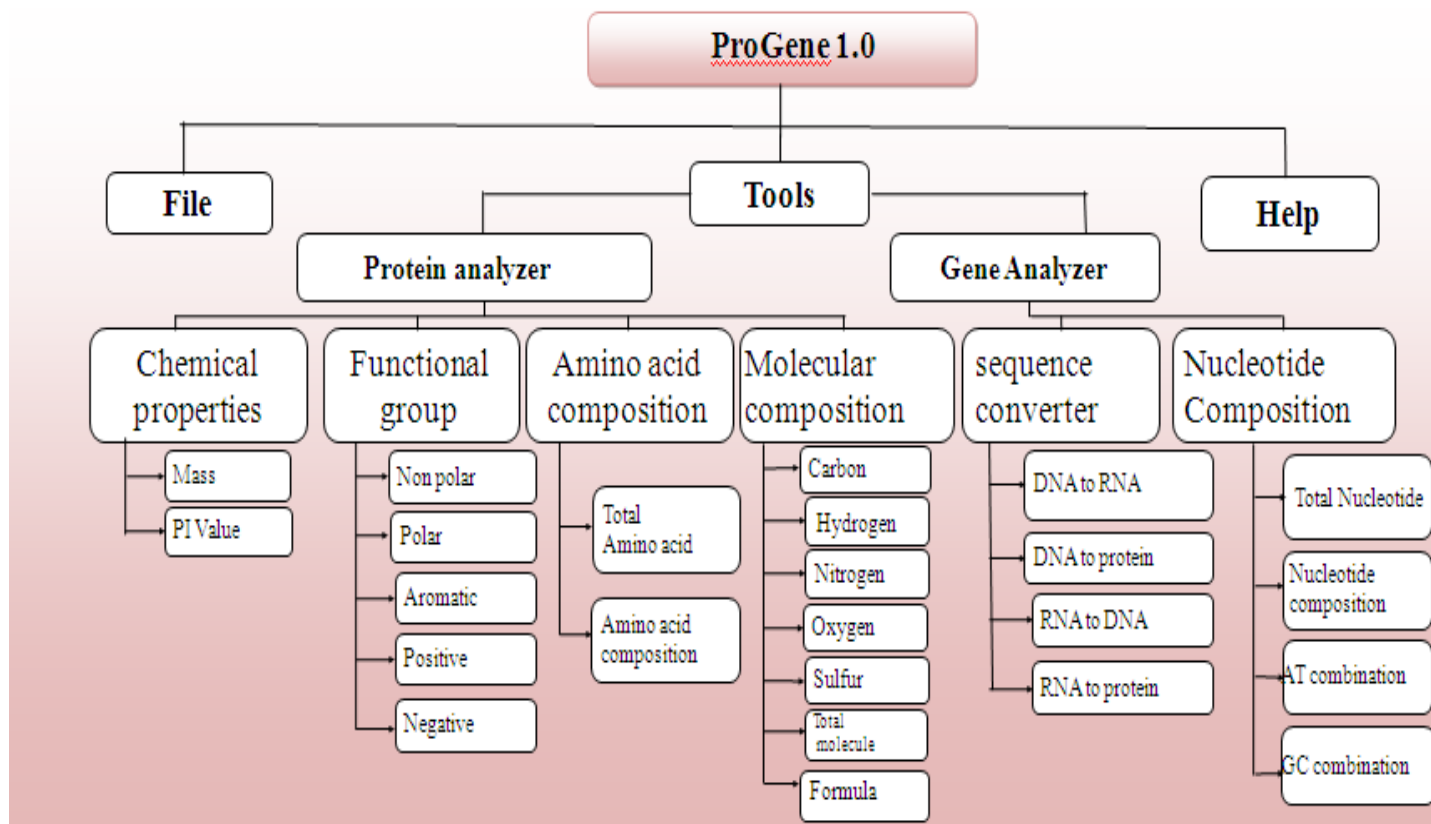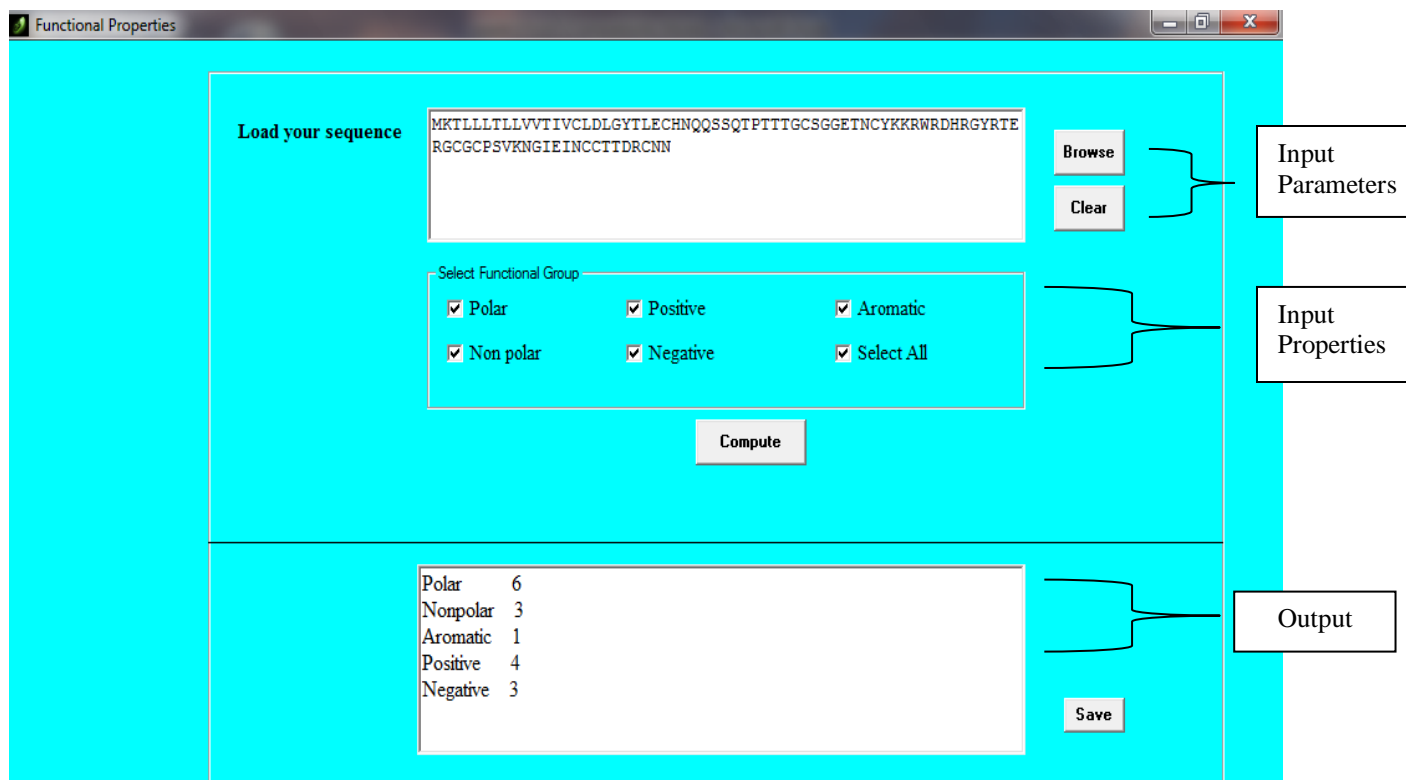
Fig. 2: Home page of ProGene 1.0

Fig. 3: Hierarchy of capability of ProGene 1.0

Fig. 4: Functional properties interface of ProGene 1.0

DISCUSSION

The basic purpose of this tool is to get primary level information of the protein. One of the major application of ProGene 1.0 is prediction of Chemical properties like Molecular weight [13] and Isoelectric point (PI) used in protein identification. Proteins have an amazing range of structural and catalytic properties as a result of their varying amino acid composition [14]. Thus, calculating amino acid composition is another important application of ProGene 1.0. Functional group variation module capable to calculate number of Polar, Non polar, Aromatics, Positive and Negative amino acids. Molecular Composition application can compute five major chemical compounds of protein molecule resembling Carbon, Hydrogen, Nitrogen, Oxygen, and Sulfur [15] and total number of atoms present in proteins. It also calculates Molecular formula of the protein sequence. Gene converter can convert nucleotide sequences to their respective transcribed reverse transcribed and translated sequences. Another important application of ProGene1.0 is to provide detailed information about nucleotide sequences via Nucleic acids Composition. It predicts total number of nucleotide sequences, total composition of Adenine (A) and Thiamine (T) combination as well as Guanine (G) and Cytosine (C) combination [16,17], to know the stability of DNA structure [18] and the GC content percentage especially for the optimized primer design [19]. As a result ProGene 1.0 enables more flexibility and expandability compared with other online tools.

CONCLUSIONS

ProGene 1.0 is a standalone desktop application and developed with Microsoft Visual Basic 6.0 to calculate physio-chemical properties like Molecular mass, Iso-electric point, Functional group residues, Molecular composition of the sequence, Nucleotide composition of the gene. This is a user friendly applications even a non bioinformaticians. ProGene 1.0 was runs on Microsoft Windows (XP, 7, 8). It is available on http://www.b-u.ac.in/projects/progene.zip.

AUTHOR'S CONTRIBUTION

NSB identified the need had input on tool design and features. SG designs the work and also contributed to manuscript writings VP develop the tool and manuscript. All the authors participated in the writing and approval of the manuscripts. All authors read and approved the final manuscript.

REFERENCES

[1]   Benton D. Bioinformatics- Principles and potential of a new multidisciplinary tool. Trends Biotechnol., 1996; 14(8): 261-72.

[2]   Wang X, Slebos RJ, Wang D, Halvey PJ, Tabb DL, et al. Protein identification using customized protein sequence databases derived from RNA-Seq data. J. Proteome Res., 2012; 11(2), 1009–17.

[3]   Smalheiser NR. Linking investigators. A centralized linking facility for data sharing and coordination of samples in tissue banks. EMBO Rep., 2003; 4(2):108-10.

[4] Moore A, Brailsford T. Unified Hyperstructures for Bioinformatics: Escaping the Application Prison. J. Digital Information, 2006; 5(1): 254.

[5] Dayhoff MO, Eck RV, Park CM. Atlas of protein sequence and structure. Washington. Nat. Biomed. Res. Foundation, 1972.

[6] Hulo N, Sigrist CJ, Le Saux V, Langendijk-Genevaux PS, Bordoli L, et al. Recent improvements to the PROSITE database. Nucleic Acids Res., 2004; 32:134-37.

[7] Bystroff C, Simons KT, Han KF, Baker D. Local sequence-structure correlations in proteins. Curr Opin Biotechnol., 1996; 7(4): 417-21.

[8] Xiong J. Essential Bioinformatics. New York, Cambridge University Press, 2006.

[9] Afzal M1, Shahid AA, Shehzadi A, Nadeem S, Husnain T. RDN Analyzer: A tool for DNA secondary structure prediction and sequence analysis. Bioinformation, 2012; 8(14): 687-90.

[10] Haggard G, Hutchison W, Shibata C. Introduction: Visual BASIC 6.0. 1st ed, Bookboon, 2013.

[11] Schneider ID. An Introduction to Programming Using Visual Basic 6.0. 6th ed, USA, University of Maryland, Prentic Hall publishers, 2006.

[12] Jung SK, McDonald K. Visual gene developer: A fully programmable bioinformatics software for synthetic gene optimization. BMC Bioinformatics, 2011; 16; 12: 340.

[13] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins RM, et al . Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, Humana Press, 2005; 571-607.

[14] Wade LG. Organic chemistry. 6th ed, India, Pearson Education publisher. 2005.

[15] Cohn EJ. Some Physical-Chemical Characteristics of Protein Molecules. Chemistry Rev., 1939; 24 (2): 203–32.

[16] Lercher MJ, Urrutia AO, Pavlícek A, Hurst LD. A unification of mosaic structures in the human genome. Hum. Mol. Genet. 2003; 12(19): 2411-15.

[17] Raymond A, Lovell S, Lorimer D, Walchli J, Mixon M, et al. Combined protein construct and synthetic gene engineering for heterologous protein expression and crystallization using Gene Composer. BMC Biotechnol., 2009; 09-37.

[18] Lercher JM, Urrutia OA, Pavlíček A, Laurence D, Hurst DL. A unification of mosaic structures in the human genome. Hum. Mol. Genet., 2003; 12 (19): 2411-15.

[19] Kampke T, Kieninger M, Mecklenburg M. Efficient primer design algorithms. Bioinformatics, 2001; 17 (3): 214-25.