SAP: A WEB BASED PACKAGE FOR SEQUENCE ANALYSIS AND MANIPULATION

Sobia Idrees, Shahid Nadeem, Beenish Ehsan
Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan.
Email: sobia_binm@live.com

*ABSTRACT***:** *SAP is an application package developed to publically provide sequence analysis tools at one platform. The ultimate goal of the SAP is to provide online facility of DNA/RNA/Protein sequence analysis tool required by researchers. The analysis of biological information at these three levels requires a lot of package tools simultaneously for the exact interpretation of huge amount of genome sequence data. The functionalities included in package are; calculating number and percentage of nucleotide bases and aminoacids, AT, GC contents, drawing pie and bar charts, transcription, translation, constructing complementary and reverse complementary sequences, performing alignments, molecular weight of protein, isoelectric point, absorbance factor.SAP is freely* available at http://bioanalysis.comli.com/.

INTRODUCTION:

Biological sequence analysis is the heart of computational biology [2].In its early days, the entire field of computational biology revolved almost entirely around biological sequence analysis [1].In the area of sequence analysis, biologists find that existing software developed in the early days of sequencing often falls short of their needs [4]

Many of the software tools used in studying genomes involve sequence analysis, which is one of the many subfields of computational molecular biology. The field of sequence analysis includes pattern and motif searching, sequence comparison, multiple sequence alignment, sequence composition determination, and secondary structure prediction.

The efficiency of bioinformatics programmers can be greatly increased through the provision of readymade software components that can be rapidly combined, with additional bespoke components where necessary, to create finished programs [3].

SAP is designed to provide sequence analysis tools at one platform. The user interactive GUI (Graphical User Interface) of the tools makes it so easy to understand that a novice user can also use it without having troubles. SAP provides flexible framework so it can be extended further.

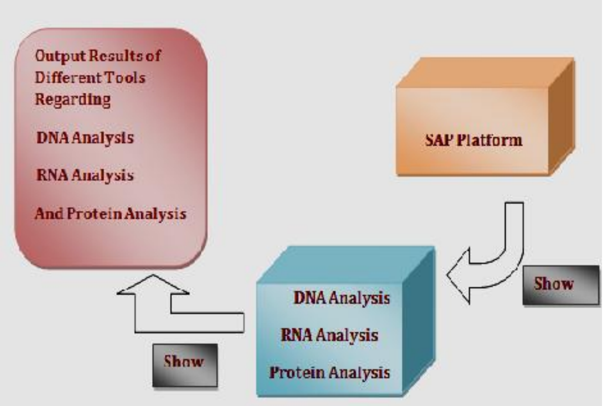

Figure 1.0: Working of SAP

SAP is developed to help the researchers by saving their time in locating different bioinformatics tools from the internet. Working of *SAP* is shown in Figure 1.0. SAP was developed to provide biological community an easy access to different sequence analysis tools and to search, analyze and visualize sequences at one platform rather than learning and using different stand alone tools.

MATERIAL AND METHODS:

We used the latest versions of PHP/Mysql and web development technologies for our work. We developed this using the object oriented mode of programming. The modules used in the current project are regarding to sequence analysis, manipulation and proteomics. List of features added in package is shown in figure 1.2.

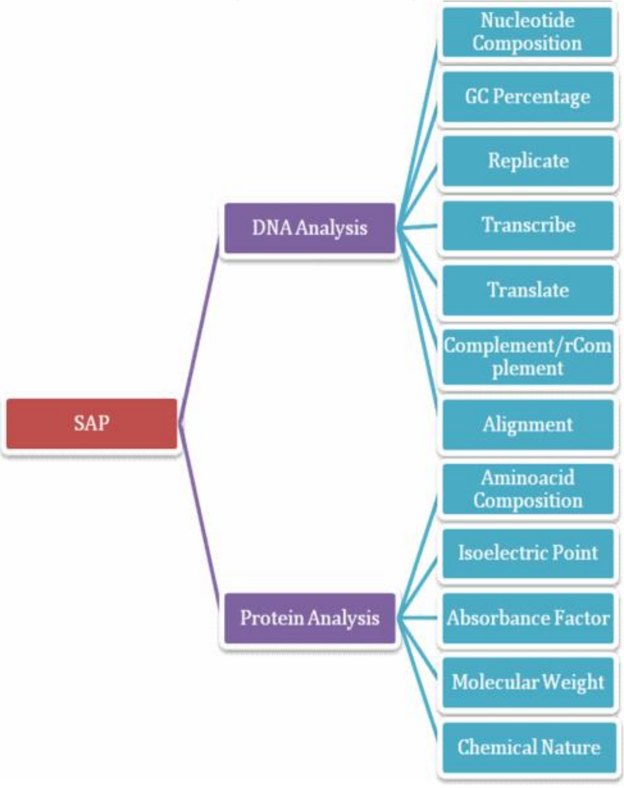

Figure 1.2: List of Tools

RESULTS

The detailed architecture of package is very simple to use and easy to get started with as shown in figure 1.0.

DNA Analysis

Nucleotide Composition: This feature counts the residues in a sequence and finds the percentage and presents it graphically figure 1.3.

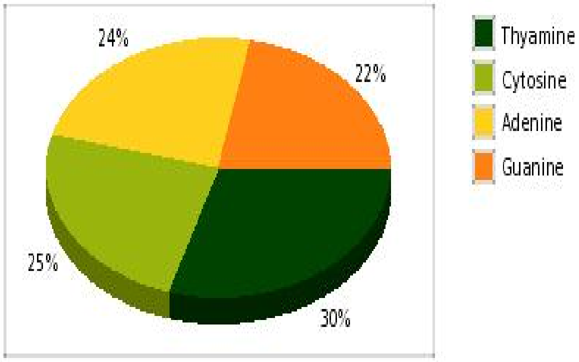

Figure 1.3**:** Nucleotide Composition Pie Chart

GC content: This feature will find the percentage of GC contents in a sequence and present them graphically figure 1.4.

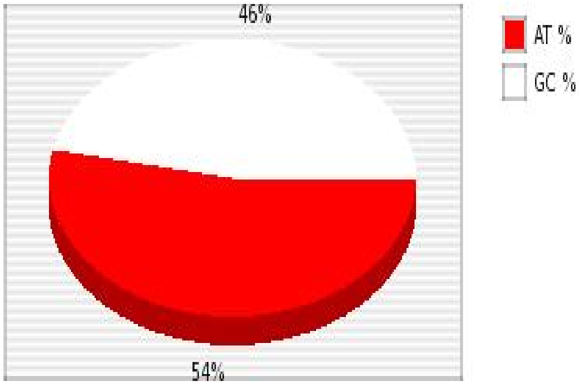

Figure 1.4: GC/AT Percentage Pie Chart

Replication, Transcription and Translation: The package provides the feature of replication, transcription and translation of the input sequences. Thus it covers central dogma of molecular biology along with other features.

Complement and reverse complement: In molecular biology, complementarity is a property of double-stranded nucleic acids such as DNA and RNA as well as DNA: RNA duplexes. Each strand is complementary to the other in that the base pairs between them are non-covalently connected via two or three hydrogen bonds. The package also provides the facility to do this.

Alignment: A very common task in bioinformatics is the alignment of two sequences also known as a 'pair-wise alignment'. Two common algorithms to generate pairwise alignments are the Needleman-Wunsch and Smith-Waterman algorithms which generate global and local alignments respectively. The resulted alignment will be shown in the output Area.

Protein Analysis:

Calculate Mass and pl of peptide: In a proteomics project it is important to know what the approximate mass and pI of any putative gene is. You can give any sequence of you choice and calculations will be made with a blink of eye.

Calculate Absorbance factor: This feature can calculate the absorbance factor of protein.

Calculate Molecular weight: This feature is used for calculating molecular weight of protein in Dalton.

Amino acid Composition: This feature is used to graphically present the aminoacid composition figure 1.5.

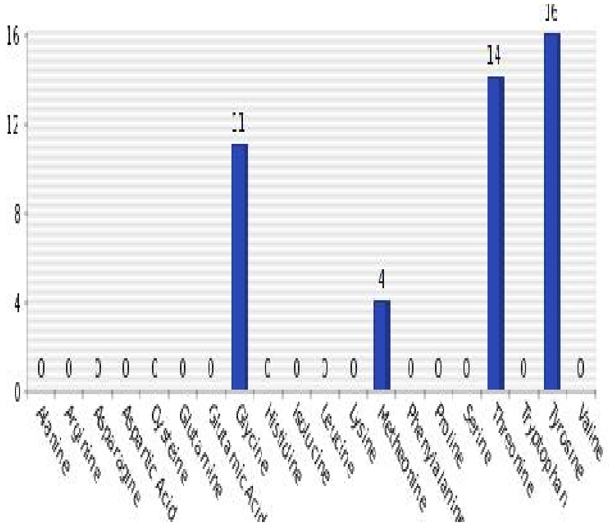

Figure 1.5: Amino acid Composition Bar Graph

Chemical Nature of proteins: This feature is designed to find the chemical nature of protein.

DISCUSSION

SAP is developed to provide all current sequence analysis tools on one platform and making easier to find and use required tools. Sequence analysis tools have been developed using PHP. This package provides an interactive and user friendly platform. SAP is designed to be fast, memory efficient and flexible. Two categories are made in the package each having its set of tools.

SAP has a simple Graphic User Interface (GUI) in which user can provide information and presses submit button and resulted information can be visualized in output area. It provides detail description of the tools available to enhance the knowledge of the user about the usage of the system.

To summarize it, SAP provides combination of tools that are never assembled together. Package has list of tools that can be important for different researchers. The aim to develop this kind of package is to provide unique set of tools at one platform in a more efficient and better way than the software or web tools available on Internet. Due to its flexible design it can be extended further in future.

Future Developments:

SAP is a software package for bioinformatics community. In future we will provide more sequence analysis features so a larger community can have its benefits.

Acknowledgments

The development of SAP is supported by Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan. We are thankful to Dr. Shahid Nadeem who not only helped in laying the foundations of Database and Software Engineering Research Group in the Department of Bioinformatics and Biotechnology but also actively leading the Group.

REFERENCES:

1. CALIFANO, A. Advances in sequence analysis. *Curr Opin Struct Biol.*, **11**(3):330-333(2001)
2. DORING, A., WEESE, D., RAUSCH, T. AND REINERT, K. SAP An efficient, generic C++ library for sequence analysis. *BMC Bioinformatics*, **9**(1):1(2008)
3. PITT, R., WILLIAMS, W., STEVEN, M. A., SWEENEY, M., BLEASBY, B.AND MOSS, S., D. The Bioinformatics Template Library—generic components for biocomputing. *Bioinformatics*, **17** (8):729-737(2000)
4. RICE, P., LONGDEN, AND I., BLEASBY, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*,**16**(6):276-277(2000)