BIOINFORMATION
Discovery at the interface of physical and biological sciences

BIOMEDICAL INFORMATICS

Open access

www.bioinformation.net
Volume 12(2)

Web Sever

# MFPPI – Multi FASTA ProtParam Interface

**Vijay Kumar Garg[1, 2], Himanshu Avashthi[1], Apoorv Tiwari[1], Prashant Ankur Jain[1] Pramod Wasudev Ramkete[3]*, Arvind Mohan Kayastha[2] &Vinay Kumar Singh[2]***

[1]Department of Computational Biology & Bioinformatics, Jacob School of Biotechnology & Bioengineering, Sam Higginbottom Institute of Agriculture Technology & Sciences, Allahabad-211007, Uttar Pradesh, Bharat (India); [2]Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi-221005, Uttar Pradesh, Bharat (India); [3]Department of Biological Sciences, Sam Higginbottom Institute of Agriculture Technology & Sciences, Allahabad-211007, Uttar Pradesh, Bharat India; Pramod Wasudev Ramkete – Email: pwramteke@yahoo.com; Vinay Kumar Singh – Email: vinaysingh@bhu.ac.in; Phone No.0542-2368364; *Corresponding author

Received January 30, 2016; Revised March 10 2016; Accepted March 12, 2016; Published April 10, 2016

**Abstract:**
Physico-chemical properties reflect the functional and structural characteristics of a protein. The comparative study of the physico-chemical properties is important to know role of a protein in exploring its molecular evolution. A number of online and offline tools are available for calculating the physico-chemical properties of a single protein sequence. However, a tool is not available for a comparative study with graphical visualization of Multi-FASTA sequences. Hence, we describe the development and utility of MFPPI V.1.0 (a web interface developed in JAVA platform) to input each FASTA sequence from Multi-FASTA file into the ProtParam web server for the calculation of physico-chemical properties. MFPPI V.1.0 calculates different physico-chemical properties for a given set of proteins in a single run and saves the data in the MSExcel sheet. Furthermore, it provides a graphical representation of protein physico-chemical properties for analysis and visualization of data in a user-friendly manner. Therefore, the output from the analysis helps to understand compositional changes and functional relationship in evolution among organisms. We have demonstrated the utility of MFPPI V.1.0 using 17 mtATP6 protein sequences from different mammalian species. It is available for free at http://insilicogenomics.in/mfpcalc/mfppi.html.

**Keywords:** Physico-chemical Property, Multi-FASTA Proteins, Amino acid richness, Peptide hydrophobicity, Isoelectric point and Extinction coefficient.

**Background:**
The physicochemical property of proteins is critical for sustainability, efficiency, and stability in a biological system. Various physico-chemical parameters of proteins such as amino acid composition, extinction coefficient **[1],** instability index **[2, 3]**, grand average of hydropathicity (GRAVY)**,** aliphatic index, theoretical pI, atomic composition and molecular weight allows us to understand the stability, activity and nature of protein. There are many web based and standalone softwares available that compute physico-chemical properties of proteins. AACompIdent is a web-based tool at ExPASy that identifies proteins using amino acid composition **[1].**

Protein/Peptide Property Calculator **[4]** is a web-based tool to calculate the peptide chemical formula, molecular weight, net-charge at neutral pH, hydrophilicity, hydrophobicity, iso-electric point and extinction coefficient. It also predicts hydrophobic or hydrophilic region, secondary structure of the protein, trans-membrane region and flexible region of the input protein or peptide sequence of interest. However, it is useful for single sequence analysis.

The Molinspiration server also offers number of chemoinformatics tools to calculate LogP (octanol/water partition coefficient), molecular polar surface area and molecular volume **[5].** ProtParam **[6]** from ExPASy **[7]** server is a reliable algorithm to compute physico-chemical properties. However, it uses single sequence per analysis through the interface. Moreover, current methods do not analyze multiple sequences for comparative analysis. It also does not provide options for downloading results for subsequent analysis. Therefore, it is of interest to develop a novel interface using ProtParam to analyze multiple sequences from a multi-FASTA file producing results for comparative inference with evolutionary insights. It is also of interest to develop methods to download and store results in an ".xls" format for further

analysis. Hence, we describe the development and utility of MFPPI V.1.0 in a JAVA platform version JRE7 (simple, object-oriented, reliable, secure and portable) for this purpose.


**Figure 1:** Multi-FASTA sequence file of different mammalian members. Input file format prepared for Multi-FASTA file to be subjected in the Akriti V.1.0

**Methodology:**

***Sequence retrieval and construction of Multi-FASTA file***

Mitochondrial protein (mtProtein) sequences of 17 different mammalian members were retrieved in FASTA format from National Centre of Biotechnology Information on a single notepad file with ".txt" extension was created. The FASTA format of protein chosen must start with >lcl| then followed by accession number or description. In the end there should be at least one bracket "[ ]" and in this bracket there may be species name or other details, sequence length should start after bracket. The input FASTA file of different mammalian protein has been illustrated in **Figure 1.**

***Script Development***

Java GUI programming involves two packages first the original Abstract Window Toolkit (AWT) and second newer Swing toolkit. Swing is the primary Java GUI widget toolkit. The script of the web interface was developed in four steps.

***Input data***

Multi-FASTA text file of mtProteins were declared as string that contains several sequences in FASTA format separated by greater than (">") symbol.

***Splitting and storing Multi-FASTA sequence into raw sequence***

Each sequence was split and converted into raw format (without any symbol and description line) and then stored into a separate file. To split the sequence from description line, each FASTA sequence was taken into string and then split method was applied from where greater than symbol ">"starts and ends with "]".

***Fetching raw sequence into ProtParam server***

To fetch the sequence into ProtParam server sequentially one by one, a connection was established with ProtParam server using following syntax.

Syntax**:** URL siturl = new URL ("http://web.expasy.org/cgi-bin/ProtParam/ProtParam");
Redirect method was applied to calculate next sequence and then output condition should be "true" to print the results after physico-chemical property calculation compilation.

***Saving data into MS-Excel file***

After compilation of calculated parameters at ProtParam server sequential result was saved in MS-Excel (.xls) file.

***Graphical User Interface***

The graphical user interface was developed very simple and user friendly. Interface contains text field, browse button, submit button and process status. Logo of software with its name in Hindi and English language as well as logo of Banaras Hindu University, Varanasi and Sam Higginbottom Institute of Agriculture Technology & Sciences, Allahabad was also added. MFPPI V.1.0 is fully automated web interface tool for ProtParam to calculate physico-chemical property. Also we divided this software into six different packages for particular calculation.

**Table 1: Amino Acid composition (%) of 17 mammalian mitochondrial ATP 6 encoded protein**

| Species | A | R | N | E | C | Q | D | G | H | I | L | K | M | F | P | S | T | W | Y | V |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Bos taurus* | 6.6 | 1.8 | 4.4 | 0.4 | 0.0 | 4.0 | 1.3 | 4.9 | 2.7 | 9.7 | 19.5 | 1.8 | 5.3 | 5.8 | 5.3 | 7.1 | 11.9 | 1.3 | 0.9 | 5.3 |
| *Canis lupus* | 8.8 | 2.2 | 4.4 | 0.4 | 0.0 | 4.0 | 1.3 | 4.9 | 2.7 | 11.9 | 18.6 | 1.8 | 4.9 | 5.8 | 5.8 | 6.2 | 9.3 | 1.3 | 1.3 | 4.4 |
| *Cavia porcellus* | 7.1 | 1.8 | 4.0 | 0.4 | 0.0 | 3.1 | 1.3 | 4.4 | 3.1 | 12.8 | 19.5 | 2.2 | 6.2 | 4.9 | 6.2 | 5.8 | 10.6 | 1.3 | 1.3 | 4.0 |
| *Cricetulus griseus* | 6.6 | 2.2 | 3.5 | 0.9 | 0.0 | 3.1 | 1.3 | 4.9 | 3.5 | 13.3 | 17.7 | 2.7 | 6.6 | 5.8 | 5.8 | 6.2 | 9.7 | 1.3 | 0.9 | 4.0 |
| *Equus caballus* | 8.0 | 1.8 | 4.4 | 0.4 | 0.0 | 4.0 | 1.3 | 4.9 | 3.1 | 11.9 | 17.7 | 1.8 | 6.2 | 6.2 | 5.8 | 6.6 | 9.3 | 1.3 | 0.9 | 4.4 |
| *Felis catus* | 8.0 | 1.8 | 4.9 | 0.4 | 0.0 | 4.0 | 1.3 | 4.9 | 3.5 | 10.6 | 19.0 | 1.8 | 5.8 | 5.3 | 5.8 | 6.2 | 9.7 | 1.3 | 0.9 | 4.9 |
| *Gorilla gorilla gorilla* | 9.7 | 1.8 | 4.9 | 0.0 | 0.0 | 3.5 | 1.8 | 3.5 | 2.7 | 10.6 | 19.9 | 2.2 | 5.8 | 3.5 | 6.2 | 5.8 | 11.9 | 1.3 | 1.3 | 3.5 |
| *Homo sapiens* | 8.4 | 1.8 | 4.9 | 0.4 | 0.0 | 3.1 | 1.3 | 3.5 | 2.7 | 12.8 | 19.5 | 2.7 | 5.3 | 4.0 | 6.2 | 5.8 | 11.5 | 1.3 | 1.3 | 3.5 |
| *Loxodonta africana* | 6.8 | 2.3 | 3.6 | 0.0 | 0.0 | 3.2 | 2.3 | 4.1 | 2.7 | 12.2 | 19.4 | 1.8 | 4.1 | 4.1 | 5.4 | 5.9 | 13.1 | 1.8 | 2.3 | 5.4 |
| *Mus musculus* | 6.6 | 2.2 | 4.0 | 0.4 | 0.0 | 2.7 | 1.3 | 4.4 | 4.0 | 12.8 | 17.3 | 2.7 | 6.2 | 6.2 | 6.2 | 6.6 | 9.7 | 1.3 | 0.9 | 4.4 |
| *Ovis aries* | 7.1 | 1.8 | 5.3 | 0.4 | 0.0 | 4.0 | 1.3 | 5.3 | 2.7 | 9.70 | 19.5 | 1.8 | 5.8 | 5.8 | 5.3 | 6.2 | 10.6 | 1.3 | 0.9 | 5.3 |
| *Pan paniscus* | 8.8 | 1.8 | 4.4 | 0.4 | 0.0 | 3.5 | 1.3 | 3.5 | 3.1 | 11.1 | 19.5 | 2.2 | 4.9 | 4.9 | 6.2 | 5.8 | 11.9 | 1.3 | 0.9 | 4.4 |
| *Pan troglodytes* | 9.3 | 1.8 | 4.4 | 0.4 | 0.0 | 3.5 | 1.3 | 3.5 | 3.1 | 11.1 | 19.9 | 2.2 | 4.9 | 4.4 | 6.2 | 5.3 | 11.5 | 1.3 | 1.3 | 4.4 |
| *Pongo abelii* | 8.4 | 2.2 | 4.0 | 0.4 | 0.0 | 3.1 | 1.3 | 3.1 | 2.7 | 11.9 | 22.1 | 2.2 | 4.4 | 3.5 | 6.6 | 5.8 | 11.9 | 1.3 | 1.3 | 3.5 |
| *Rattus norvegicus* | 6.6 | 2.2 | 3.5 | 0.9 | 0.0 | 2.7 | 1.8 | 4.4 | 4.0 | 12.8 | 18.1 | 2.2 | 6.2 | 5.8 | 6.2 | 6.6 | 9.3 | 1.3 | 0.9 | 4.4 |
| *Saimiri boliviensis* | 5.8 | 1.8 | 5.3 | 0.0 | 0.0 | 4.0 | 0.9 | 4.0 | 2.2 | 11.5 | 21.2 | 1.3 | 5.8 | 4.0 | 5.3 | 7.1 | 11.9 | 1.3 | 1.8 | 4.9 |
| *Sus scrofa* | 7.5 | 1.8 | 4.9 | 0.4 | 0.0 | 4.4 | 1.3 | 4.4 | 2.7 | 11.9 | 17.7 | 2.2 | 5.8 | 6.2 | 5.3 | 5.8 | 11.5 | 1.3 | 1.3 | 3.5 |

**Table 2: Physico-chemical properties of 17 mammalian mitochondrial ATP 6 encoded protein calculated by MFPPI V.1.0.**

| Species | MW | EC | II | AI | GRAVY |
|---|---|---|---|---|---|
| *Bos taurus* | 24787.9 | 19480 | 36.15 | 135.93 | 0.924 |
| *Canis lupus* | 24789 | 20970 | 32.34 | 140.75 | 0.977 |
| *Cavia porcellus* | 24952.5 | 20970 | 36.95 | 144.6 | 1.025 |
| *Cricetulus griseus* | 25071.6 | 19480 | 35.14 | 138.98 | 0.965 |
| *Equus caballus* | 24866.1 | 19480 | 40.64 | 136.42 | 0.973 |
| *Felis catus* | 24805 | 19480 | 40.85 | 137.7 | 0.920 |
| *Gorilla gorilla gorilla* | 24676.9 | 20970 | 35.9 | 139.07 | 0.888 |
| *Homo sapiens* | 24817.2 | 20970 | 34.74 | 144.65 | 0.952 |
| *Loxodonta africana* | 24575.7 | 29450 | 32.01 | 145.41 | 0.963 |
| *Mus musculus* | 25095.5 | 19480 | 31.88 | 136.81 | 0.943 |
| *Ovis aries* | 24797.9 | 19480 | 34.15 | 136.37 | 0.924 |
| *Pan paniscus* | 24758 | 19480 | 31.82 | 140.75 | 0.939 |
| *Pan troglodytes* | 24770 | 20970 | 32.19 | 142.92 | 0.953 |
| *Pongo abelii* | 24801.2 | 20970 | 30.49 | 151.55 | 1.004 |
| *Rattus norvegicus* | 25075.5 | 19480 | 28.24 | 140.27 | 0.969 |
| *Saimiri boliviensis* | 24925.3 | 22460 | 37.5 | 147.57 | 1.019 |
| *Sus scrofa* | 25039.2 | 20970 | 34.58 | 133.41 | 0.881 |

GRAVY = Grand average of hydropathicity; MW = Molecular weight; AI = Aliphatic Index; EC = Ext. coefficient; II = Instability Index

**Utility and application:**

***General features***

The MFPPI V.1.0 graphical user interface of tool has only two buttons, browse and submit **(Figure 2).** The server is able to calculate total number of amino acid, molecular weight, theoretical pI, number of each amino acid residue and their percentage, total number of negatively charged residues (D + E), instability index, aliphatic index, and grand average of hydropathicity (GRAVY) for several protein sequences simultaneously.


**Figure 2**: Graphical User Interface of MFPPI V.1.0. web-interface for MULTI-FASTA PROT-PARAM interface

***Special features***

Multiple FASTA format (>lcl|Sequence ID or description of protein [sequence source or any other information]) sequences in a file are used as input for analysis. The result is saved in an excel file format for further analysis and inference.

***Example analysis***

The results from MFPPI V.1.0 for 17 mtATP6 protein **[8]** sequences from different mammalian species are given in **Table 1 &Table 2.** A graph drawn using **Table 1** is shown in **Figure 3**. This is an example of comparative analysis of multiple sequences. The sequences are amino acid C poor and L rich. Low frequency of D was found across the species and absent in Saimiri boliviensis and Gorilla gorilla gorilla. The amino acid residues R, E, K, W and Y were also present in low frequency in comparison to higher frequencies of N, Q, G, H, M, F, P and V. Residues A, I, S and T frequency was found relatively higher among all species.

***Other features***

The interface also provides values for molecular weight, extinction coefficient, instability index, aliphatic index and grand average of hydropathycity (GRAVY) **[9]** for the protein sequences **(Table 2)** in a comparative manner among 17 mammalian species. This provides insight for functional analysis and molecular evolution.

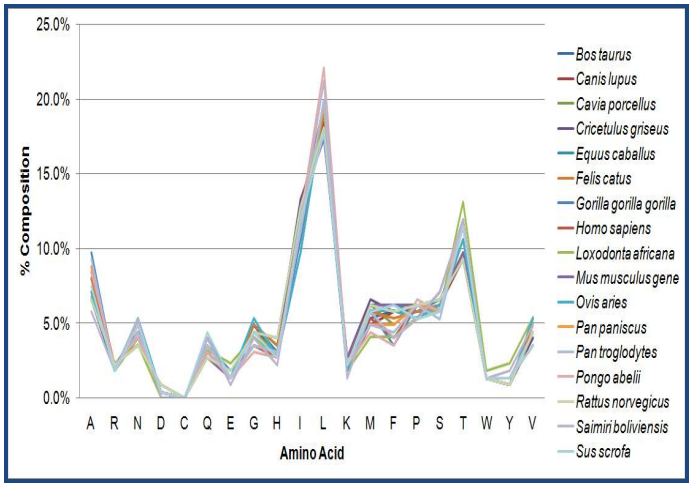


**Figure 3:** The relationship between amino acid and their percent composition in mtATP6 among different species is shown. The composition graph shows mtATP6 is rich in amino acid L and poor in C.

**Conclusion:**

The added feature in MFPPI V.1.0 interface is its ability to calculate physico-chemical properties of multiple protein sequences along with comparative analysis of several physiochemical parameters using the Expasy's ProtParam server. The interface provides output in Excel sheet format for further useful statistical analysis and graph generation for further visualization analysis. MFPPI V.1.0 finds utility in understanding compositional changes and functional relationship in evolution among organisms. We have demonstrated this using 17 mtATP6 protein sequences from different mammalian species.

**Acknowledgment:**

Authors are grateful to Centre for Bioinformatics, Institute of Science, Banaras Hindu University, Varanasi, Bharat (India) for

providing necessary infrastructure facility to carry out this work.

**Disclosure:**
The authors report no conflict of interest regarding this work.

**References:**
[1] Gill SC *et al. Anal Biochem.* 1989 **182:** 319 [PMID: 2610349]
[2] Guruprasad K *et al. Protein Eng.* 1990 **4:** 155 [PMID: 2075190]
[3] Ikai A *et al. J Biochem.* 1980 **88:** 1895 [PMID: 7462208]
[4] http://lifetein.com/peptide-analysis-tool.html
[5] Ertl P *et al. J Med Chem.* 2000 **43:** 3714 [PMID: 11020286]
[6] Gasteiger E *et al. Humana Press.* 2005 pp.571-607.
[7] Wilkins MR *et al. Methods Mol Biol.* 1999 **112:** 531 [PMID: 10027275]
[8] Mitchell P *et al. Biol Rev Camb Philos Soc.* 1966 **41:** 445 [PMID: 5329743]
[9] Kyte J & Doolittle RF, *J MolBio.* 1982 **157:** 105 [PMID: 7108955]

**Edited by P Kangueane**
**Citation**: **Garg *et al.*** Bioinformation 12(2): 74-77 (2016)