ISSN: 0976-3104
REGULAR ISSUE
Gaur.

THE IIOAB JOURNAL

SHORT COM | OPEN ACCESS

AMINO ACID FREQUENCY DISTRIBUTION AMONG EUKARYOTIC PROTEINS

Rajneesh Kumar Gaur*
Department of Biotechnology, Ministry of Science and Technology, Lodhi Road, New Delhi, INDIA

ABSTRACT

Eukaryotes constitute one of the major kingdoms. The comparative analysis of amino acid frequency distribution among proteins of Fungi, Protozoa, Invertebrates, Vertebrates and Plants is carried out to determine the compositional similarities or differences. The results indicate that, in general, all eukaryotic proteins exhibit significant similarity in their amino acid frequency distribution. All eukaryotes possess equal frequency distribution of polar uncharged residues in membrane and non-membrane proteins. Protozoa and Invertebrate follows the similar trajectory of amino acid frequency distribution. Fungal proteins show significant similarity to plant proteins. Plant proteins possess highest frequency of 'A and G' amino acids among all the eukaryotes. These results can be used in the development of prediction algorithms.

Received on: 02-Aug-2013
Revised on: 16th-Nov-2013
Accepted on: 26th– Nov-2013
Published on: 25th –Apr-2014

KEY WORDS

Eukaryotes, Membrane Proteins; Non-membrane proteins; Amino acid frequency

***Corresponding author:** Email: meetgaur@gmail.com; Tel: +91 9990290384

[I] INTRODUCTION

Amino acids are building blocks of proteins and responsible for their characteristic features. The amino acid composition of proteins has been used for different purposes such as to determine the trend among full genomes of species of different kingdoms [1], to determine the biological evolution [2], discrimination of intra- and extra cellular proteins [3], prediction of sub-cellular location [4] and to identify the contrasting features of prokaryotic and eukaryotic integral membrane and non-membrane proteins [5-6]. The various types of eukaryotes evolved from the same root [7] and adapted to different ecological environment. Their common origin indicates that amino acid composition of proteins of different kinds of eukaryotes may be similar. However, it is interesting to observe the contrasting features introduced due to their local ecological adjustment. The focus of this study is the comparative analysis of amino acid composition of membrane and non-membrane proteins of eukaryotes belongs to different phyla.

[II] MATERIALS AND METHODS

The three databases PSORT [8], eSLDB [9] and RefSeq [10] were used for manually curating the experimental sequence dataset. Refseq is a non-redundant database of DNA, RNA and Protein sequence. RefSeq database sequence release files were used to prepare eukaryotic experimental dataset. Protein sequences flagged as uncharacterized, similar to the predicted protein and the repetitive and sequences having repeats of single amino acid are deleted from the initially downloaded RefSeq sequences. The eukaryotic sequences of PSORT, eSLDB and Refseq database were pooled to prepare the final experimental dataset.

The entire dataset used for computing the composition of 20 amino acid residues comprised of eukaryotic Membrane Proteins (MPs) (28280) and non-membrane proteins (nMPs) (338146) sequences. The amino acid composition was computed using the number of amino acids of each type and the total number of residues (Eq. 1). The residue composition is calculated as follows -

$$(\%)\ (r) = (\sum nr/N) \times 100 \qquad (1)$$

where 'r' stands for any one of the 20 amino acid residue. $\sum nr$ is the total number of residue of each type and N is the total number of residues in the dataset.

[III] RESULTS AND DISCUSSION

3.1. Membrane proteins (MPs)

In general, the order of percentage composition of hydrophobic amino acids in eukaryotic MPs is L>I>V>F>A>G>P>M>C>W, while hydrophilic amino acids are arranged in S>T>E>D>N>K>R>Y>Q>H order **[Table–1** and **Figure–1]**. The high frequency distribution of non-polar hydrophobic amino acids especially 'L, I, A, V' in MPs is due to their involvement in membrane integration [11]. In contrast to other eukaryotes, Plant possess significantly high occurrence of short hydrophobic residues 'A and G'. The high frequency distribution of 'A' residue suggests that most of the plant MPs exhibits preferential helical conformation [12]. Furthermore, the short side chains of 'A and G' play significant role in helix-helix packing [13]. The composition of 'A' is lowest in Invertebrates and Protozoans. Vertebrates and plants have slightly low frequency of 'Y' residue in their MPs. Protozoa and Invertebrates possess high frequency of 'N and K' residues,

THE IIOAB JOURNAL
ISSN: 0976-3104

which suggest their close evolutionary relationship. In comparison to other eukaryotes, Vertebrates-mammals have significantly high proportion of 'R' amino acid in their MPs. In Plants, the low and high occurrence of 'Y' and 'R' respectively is in accordance with earlier study [14]. The high 'R' distribution is related to the rich occurrence of 'R' in trafficking signals [15]. All eukaryotic MPs have low abundance of 'W, M and H'.

Table: 1. Amino acid frequency distribution of different eukaryotic membrane proteins (MPs). The amino acids are arranged in decreasing order of hydrophobicity [24]

Amino acid residues	Membrane proteins (MPs) - amino acid composition (%)					
	Fungi	Protozoa	Invertebrate	Vertebrate		Plant
				Mammalian	Others	
L	11.79	10.54	9.97	11.40	9.45	10.69
I	8.36	8.95	8.28	5.58	6.46	4.63
F	6.42	6.55	5.99	4.79	4.93	4.43
W	1.56	1.29	1.58	1.57	1.54	1.69
V	6.45	5.77	6.75	7.10	7.40	7.77
M	2.69	2.47	2.15	2.45	2.69	2.74
A	6.63	5.15	4.91	6.77	6.16	12.42
G	6.43	6.06	5.46	6.69	6.10	9.24
P	4.13	3.37	3.53	5.68	4.53	5.12
C	1.15	1.55	2.00	2.61	2.59	1.38
Y	4.11	4.36	4.15	3.08	3.57	2.75
T	5.73	5.04	5.12	5.70	6.73	5.55
E	4.29	5.24	5.18	4.73	4.91	4.13
S	8.63	7.82	8.13	8.61	8.51	7.15
Q	2.89	3.00	3.08	3.43	3.48	3.22
D	3.92	3.95	4.75	3.92	4.52	3.95
H	1.94	2.05	1.84	2.38	2.06	1.94
N	4.73	6.74	6.77	3.64	5.06	3.00
K	4.55	6.72	6.96	4.24	4.88	3.20
R	3.60	3.38	3.38	5.63	4.40	4.99

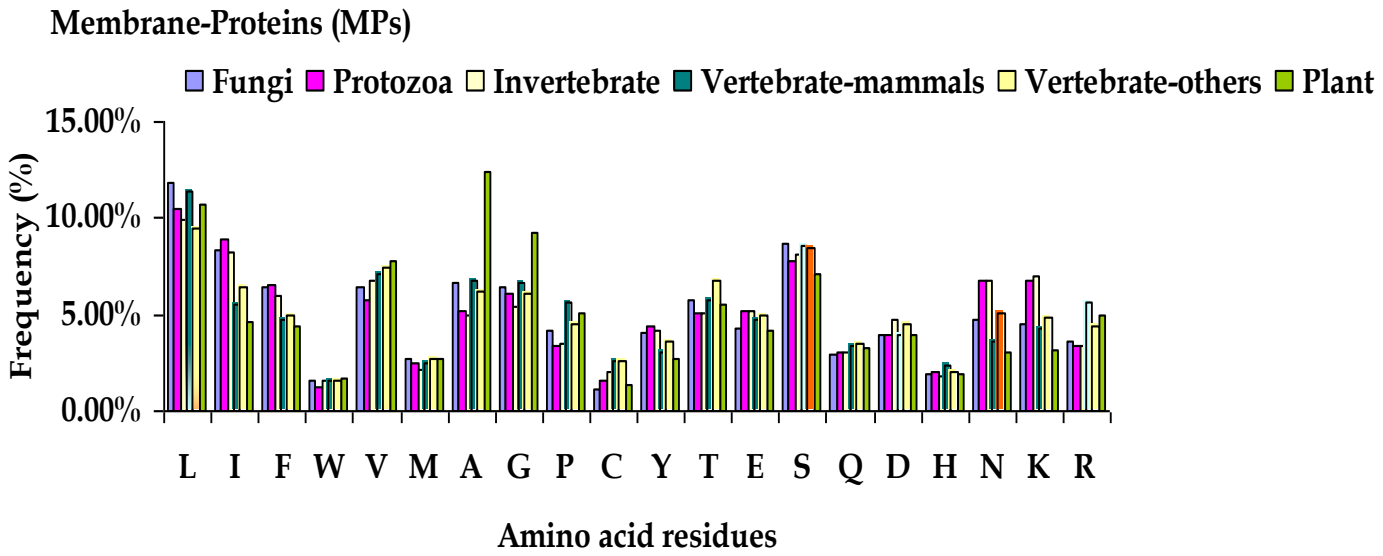

Fig: 1. Histogram showing the amino acid frequency distribution of membrane proteins (MPs) of different types of eukaryotic organisms

THE IIOAB JOURNAL
ISSN: 0976-3104

3.2. Non-membrane proteins (nMPs)

In eukaryotic nMPs, the average percentage of hydrophobic and hydrophilic residues is 48.31% and 51.69% respectively **[Table–2 and Figure–2]**.Plants have lowest proportion of 'I and F' residues in their nMPs, while similar to their MPs, the frequency of 'A and G' residues is highest among all the eukaryotes. Protozoa and Invertebrates nMPs possess lowest and highest frequency distribution of 'P' and 'N and K' residues respectively among all the eukaryotes. Vertebrates-mammals and Plants possess highest frequency distribution of 'R' residues in their nMPs.

Table: 2. Amino acid frequency distribution of different eukaryotic non-membrane proteins (nMPs). The eukaryotes are arranged in increasing level of complexity

Amino acid residues	non-Membrane Proteins (nMPs) - amino acid composition (%)					
	Fungi	Protozoa	Invertebrate	Vertebrate		Plant
				Mammalian	Others	
L	9.49	8.54	9.47	9.79	8.71	9.78
I	5.84	8.43	7.05	4.38	5.26	2.66
F	4.03	4.82	4.67	3.66	3.82	2.64
W	1.22	0.65	1.08	1.22	1.14	1.39
V	5.84	4.66	6.05	6.02	6.62	6.91
M	2.12	2.14	2.19	2.03	2.47	2.13
A	6.83	3.43	4.99	6.94	5.98	14.37
G	5.87	4.52	4.71	6.86	5.63	9.24
P	5.41	2.73	3.75	6.51	4.86	6.40
C	1.19	1.87	2.19	2.14	2.25	1.65
Y	3.11	4.82	3.53	2.60	3.08	2.17
T	5.65	4.58	5.18	5.16	6.26	5.08
E	6.53	7.08	6.61	7.05	6.43	5.19
S	8.93	7.35	8.12	8.15	8.30	7.03
Q	3.74	3.08	3.78	4.48	4.28	4.23
D	5.51	5.45	5.51	4.93	5.59	4.66
H	2.28	2.30	2.30	2.50	2.56	2.21
N	4.80	9.56	6.37	3.66	5.11	2.44
K	6.36	10.23	8.26	5.68	6.47	3.24
R	5.23	3.75	4.19	6.22	5.20	6.56

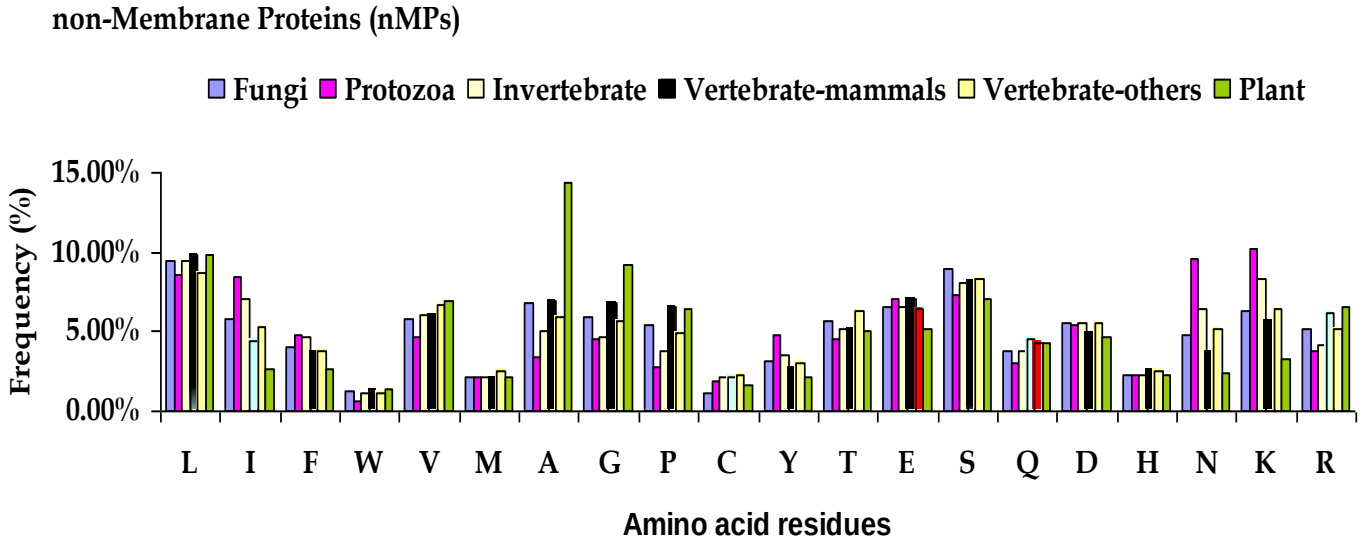

Fig: 2. Histogram showing the amino acid frequency distribution of non-membrane proteins (nMPs) of different types of eukaryotic organisms

BIOINFORMATICS

THE IIOAB JOURNAL
ISSN: 0976-3104

3.3. MPs and nMPs comparison

The frequency distribution difference is calculated by subtracting amino acid percentage of nMPs from MPs separately for each eukaryotic type **[Figure–3]**. In general, eukaryotic MPs possess higher proportion of four hydrophobic amino acids (L, I, F, V, G) in comparison to nMPs. In contrasts to MPs, eukaryotic nMPs possess preferential occurrence of six polar residues 'E, Q, D, K, R and H'.

The amino acid composition difference between fungal MPs and nMPs shows the highest percentage of hydrophobic residues namely 'L, I and F' among all the eukaryotes. These three residues constitute 26.57% of total amino acid. Slightly high occurrence of 'Y' is probably due its catalyzing role in MPs such as channels or transporters. Fungal nMPs possess high proportion of five charged polar residues 'E, Q, K, D and R' and a single non-polar residue 'P' in comparison to MPs. The polar and non-polar residues constitute 23.63% and 5.41% of the total nMPs amino acid composition. In general, fungal proteins show significant similarity to plant MPs and nMPs in their overall amino acid frequency distribution with the exception of 'A and G' **[Figure– 3]**.

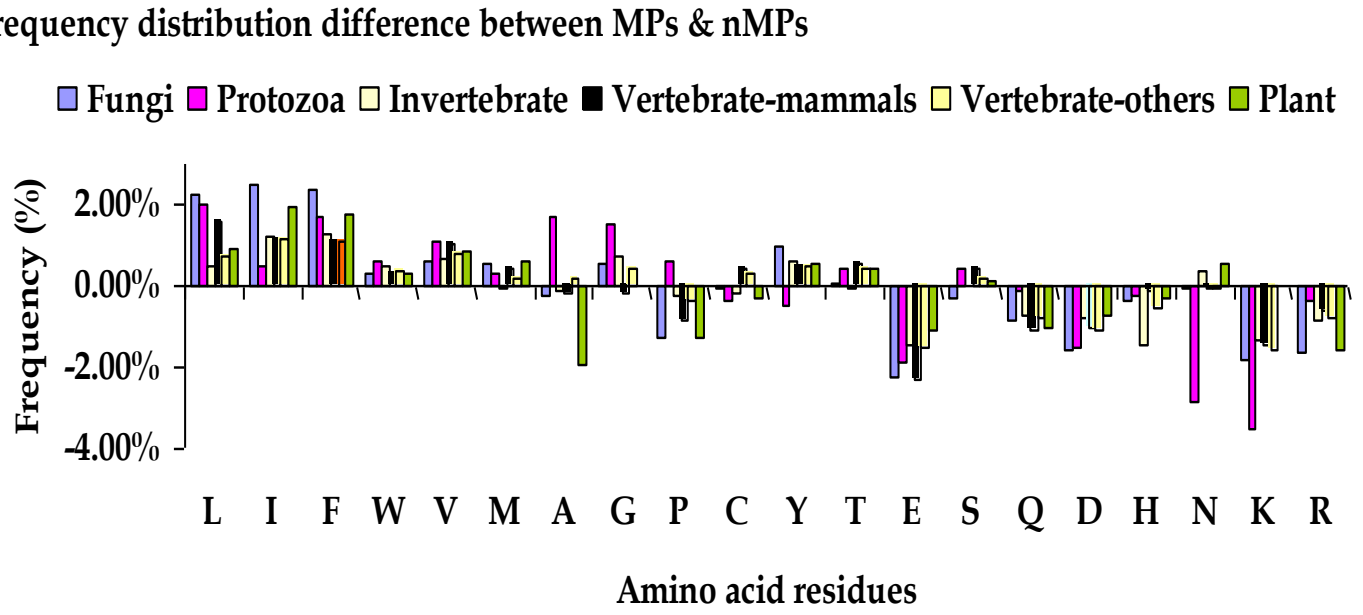

Fig: 3. Difference of amino acid frequency distribution. The difference is calculated after subtracting the amino acid percentage frequency of nMPs from MPs for each class of organism. MPs: positive scale; nMPs: negative scale

Protozoa MPs possess proportionately high occurrence of 'L, F, G, V, A, P and S' (45.26% of total amino acids) hydrophobic residues in comparison to their nMPs. The high proportion of two small non-polar residues 'A and G' in Protozoa MPs indicates the higher occurrence of alpha-helices in comparison to nMPs [16]. With respect to MPs, protozoa nMPs possess high proportion of three charged ('E, D and K') and one uncharged ('N') polar residue. The protozoa nMPs possess uncharged polar residue 'N' in highest frequency among all the eukaryotes.

With respect to their nMPs, plant MPs have high proportion of hydrophobic ('I and F') and hydrophilic ('Y, T and N') residues. High frequency of hydrophobic residues in MPs is expected. On comparison between plant MPs and nMPs, plant nMPs possess high proportion of 'A, P and R' (27.33% of the total amino acids).

Invertebrate nMPs possess high proportion of uncharged polar residue 'H' with respect to their MPs. Vertebrate-mammals nMPs possess high proportion of E, D, K and P (24.17 % of the total amino acids) with respect to their MPs. The higher occurrence of these charged amino acids is related to the nMPs functionality required to accomplish the wide array of complex cellular processes.

In order to find out the contrasts in amino acid frequency distribution among the eukaryotic MPs and nMPs, the amino acid composition of each class of organism is superimposed separately **[Figure–4]**. For few amino acids, the highest deviation from the average percentage is observed in Protozoans, Invertebrates and Plants. Protozoa and Invertebrate MPs and nMPs possess slightly high concentration of 'N and K' with respect to other eukaryotes [17]. In fact, Protzoans MPs and nMPs possess highest concentration of 'N and K', which is

THE IIOAB JOURNAL
ISSN: 0976-3104

probably related to their infectivity [18-19]. Invertebrate and Protozoa follows the similar trajectory of amino acid frequency distribution, which suggests their evolution in the same line [20]

Plant MPs and nMPs possess highest frequency of 'A and G' amino acids among all the eukaryotes. Plant nMPs have highest proportion of 'A' (14.37%), which is approximately two times more than the rest of the eukaryotic nMPs. High distribution of 'A' is commonly observed in green eukaryotes (www.computational-genomics.net/case_studies/chloroplast_demo.html). In addition to 'K', our results show that plant MPs and nMPs also possess low proportion of 'N and I' among all the eukaryotes. The low frequency of 'N' is observed depending upon the tissue and the stress conditions of the plant [21].

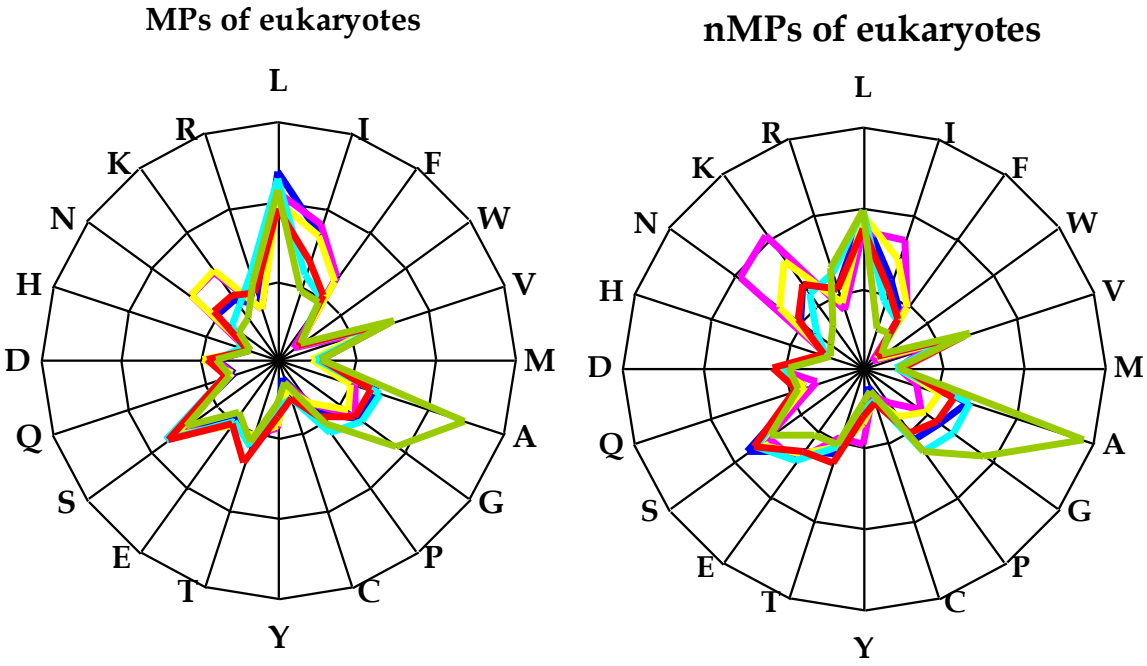

Fig: 4. Superimposition of amino acid composition of membrane and non-membrane proteins (MPs and nMPs). Blue: Fungi; Purple: Protozoa; Yellow: Invertebrates; Sky-blue: Vertebrate-mammals; Red: Vertebrate-others; Green: Plants

Table: 3. Frequency distribution of non-polar and polar amino acid residues in eukaryotic membrane (MPs) and non-membrane proteins (nMPs).

Amino acid residue groups	Amino acid composition (%)					
	Fungi MPs (nMPs)	Protozoa MPs (nMPs)	Invertebrate MPs (nMPs)	Vertebrate-mammals MPs (nMPs)	Vertebrate-others MPs (nMPs)	Plants MPs (nMPs)
Non-polar (L, I, F, V, M, A, G, P)	52.9 (45.43)	48.86 (39.14)	47.04 (42.88)	50.46 (46.19)	47.77 (43.35)	57.24 (54.13)
Polar -uncharged (S, T, N, Q, H, W, C, Y)	30.74 (30.92)	31.85 (34.21)	32.67 (32.63	31.02 (29.91)	33.54 (32.98)	26.68 (26.2)
Polar -charged (D, E, K, R)	16.36 (23.63)	19.29 (26.52)	20.27 (24.57)	18.52 (23.88)	18.71 (23.69)	16.27 (19.65)

3.4. General

In general, all eukaryotic MPs and nMPs exhibits similar basic pattern of amino acid frequency distribution though there are minor individual differences [2]. The minor individual differences are the result of increased eukaryotic complexity during the course of evolution and reflect their sustained adaptation to the constantly changing surrounding environment [22]. In general, the average amino acid composition indicates the higher occurrence of hydrophobic amino acids like 'L, I, F' in MPs [23] and hydrophilic amino acids in nMPs [6]. The high

THE IIOAB JOURNAL
ISSN: 0976-3104

occurrence of hydrophobic amino acids expected in MPs as these amino acids play an important role in embedding and retention of MPs within the lipid bilayer, while polar amino acids in nMPs execute their function in a hydrophilic cell protoplasm. Eukaryotes possess more or less equal distribution of polar uncharged residues in MPs and nMPs. The lower eukaryotes i.e. fungi and protozoa have higher occurrence of charged polar residues in their nMPs with respect to MPs, while plants have least difference of distribution of polar charged amino acids in their MPs and nMPs **[Table– 3]**.

FINANCIAL DISCLOSURE

The work is not supported by any funding agency.

CONFLICT OF INTEREST

Author does not have any competing financial or any other interest.

REFERENCES

[1] Bogatyreva NS, Finkelstein AV, Galzitskaya OV. [2006] Trend of amino acid composition of proteins of different taxa. *J Bioinfor Comput Biol* 4(2):597−608.
[2] Sorimachi K. [1999] Evolutionary changes reflected by the cellular amino acid composition. *Amino Acids* 17:207−226.
[3] Nakashima H, Nishikawa K. [1994] Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. *J Mol Biol* 238:54−61.
[4] Bhasin M, Raghava GPS. [2004] ESL pred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. NAR, 32(web server issue), W414−419
[5] Gaur RK, Natekar G. [2010] Prokaryotic and eukaryotic integral membrane proteins have similar architecture. *Mol Biol Rep* 37(3):1247−1251.
[6] Gaur RK. [2009] Prokaryotic and eukaryotic non-membrane proteins have biased amino acid distribution. *J Comput Sci Syst Biol* 2 (6):298−299.
[7] Roger AJ, Simpson AGB. [2009] Evolution: revisiting the root of the eukaryote tree. *Current Biol* 19(4):R165−167.
[8] Rey S, Acab M, Gardy JL, Laird MR, deFays K.et al. [2005] PSORTdb: A Database of Subcellular Localizations for Bacteria. *Nucleic Acids Res* 33: D164−168.

[24] Monera OD, Sereda TJ, Zhou NE, Kay CM, Hodges RS. [1995] Relationship of side chain hydrophobicity and alpha-helix propensity on the stability of the single-staranded amphipathic alpha-helix. *J Pep Sci* 1(5):319–329.

[9] Pierleoni A, Martelli PL, Fariselli P, Casadio R. [2007] eSLDB: Eukaryotic subcellular localization databse. *Nucleic Acids Res* 35: D208−212.
[10] Pruitt KD, Tatusova T, Maglott DR. [2005] NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33: D501−504.
[11] Ulmschneider MB, Sansom MSP. [2001] Amino acid distribution in integral membrane protein structures. *Biochem Biophys Acta* 1512:1−14.
[12] Pace CN, Scholtz JM. [1998] A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* 75:422−427.
[13] Eyre TA, Partridge L, Thronton JM. [2004] Computational analysis of alpha-helical membrane protein structure: implications for the prediction of 3D structural models. *Protein Eng Des Sel* 17:613−624.
[14] Pirman T, Stibily V, Stekar JMA, Combe E. [2001] Amino Acid Composition of Beans and Lentil. *Zb Biotech Fak Univ Ljubl Kmet Zootech*78 (1):57 – 68.
[15] Zerangue N, Malan MJ, Fried SR, et al. [2001] Analysis of endoplasmic reticulum trafficking signals by combinatorial screening in mammalian cells. *Proc Natl Acad Sci* USA 98(5):2431−2436.
[16] Javadpour MM, Eilers M, Groesbeek M. [1999] Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. *Biophysical J* 77:1609−1618.
[17] Wti C, Hogg JF. [1956] Free and Nonprotein amino acids of Tetruhymena pyriformis. *Archives of Biochemistry and Biophysics* 62:707−707.
[18] Singh GP, Chandra BR, Bhattacharya A, Akhouri RR, Singh SK, Sharma A. [2004] Hyper expansion of asparagines correlates with an abundance of proteins with prion-like domains in Plasmodium falciparum. *Mol Biochem Parasitol* 137(2):307−319.
[19] Mai Z, Samuelson J. [1998] A new gene family (ariel) encodes asparagines-rich Entamoeba histolytica antigens, which resemble the amebic vaccine candidate serine-rich E. histolytica protein. *Infect Immun* 66(1):353−355.
[20] Tartar A, Boucias DG, Adams BJ, Becnel JJ. [2002] Phylogenetic analysis identifies the invertebrate pathogen Helicosporidium sp. as a green algae (Chlorophyta). *International J Systematic and evolutionary Microbiology* 52:273−279.
[21] Yang X, WangX, Wei M, Hikosaka Sh, Goto E. [2011] Changes of glutamine and asparagine content in cucumber. *Int J Plant Prouction* 5(1):1C8.
[22] Copley RR, Letunic I, Bork P. [2002] Genome and protein evolution in eukaryotes. *Curr Opin Chem Biol* 6(1):39−45.
[23] Tourasse NJ, Li WH. [2000] Selective constraints, amino acid composition and the rate of protein evolution. *Mol Biol Evol* 17(4):656−664.