

Prokaryotic and Eukaryotic Non-membrane Proteins have Biased Amino Acid Distribution

Rajneesh Kumar Gaur

Bioinformatics Infrastructure Facility, Jamia Hamdard (Hamdard University), Hamdard Nagar, New Delhi, India – 110062

Abstract

Proteins constitute the important constituent of the cellular machinery. The comparative analysis of non-membrane proteins (nMPs) between prokaryotes and eukaryotes carried out to determine the biasedness in amino acid distribution. On comparison, the results revealed that 'Ala' is the dominant amino acid in prokaryotic nMPs while 'Lys, Ser and Cys' are the dominant amino acids in eukaryotic nMPs.

Keywords: Non-membrane proteins; Amino acid composition; Prokaryotes; Eukaryotes

Abbreviations: MPs: Membrane Proteins; nMPs: non-membrane proteins

Introduction

Proteins constitute about 50% of the dry weight of most cells and are the most structurally complex macromolecules known. Proteins can be classified in different manner but for the purpose of this study we classified them as membrane (part of either cellular or organelle membrane; MPs) and non-membrane (located outside the membrane; nMPs) proteins. Amino acids are the building block of a protein and their composition determines the overall properties and stability of a protein. Many previous studies have shown how amino acid composition can be successfully applied to protein sequence analysis, including prediction of structural class (Zhang et al., 1992), discrimination of intra- and extra cellular proteins (Nakashima et al., 1994), prediction of sub-cellular location (Cedano et al., 1997). It was suggested that composition differences are a consequence of different requirements for protein folding, stability and transportation.

The recent increase in the number of whole genome sequences has made the analysis of the corresponding proteomes possible. So far the amino acid composition of both the prokaryotic and eukaryotic proteomic databases have been explored separately for different purposes such as determination of sequence length (Gerstein, 1998a), identification of conserved sequences (Sobolevsky et al., 2005); elucidation of simple sequences (Subramanyam et al., 2006) etc. However, till now the comparative analysis of their non-membrane proteins (nMPs) have not been carried out to determine the overall amino acid compositional differences. This computational study is performed to develop the amino acid distribution of proteins as a tool to identify the proteins frequently undergo mutations and largely responsible for the pathogenicity of the organism.

Methodology

The dataset was curated manually from the sequences extracted from PSORT (Rey et al., 2005), eSLDB (Pierleoni et al., 2007) and RefSeq (Pruitt et al., 2005) databases. Only the experi-

mentally annotated entries were extracted from PSORT database. From the RefSeq database, we used microbial (microbial1.protein.faa.gz; 05/11/2009) and eukaryotic (vertebrate_mammalian1.protein.faa.gz; 05/11/2009 & vertebrate_other1.protein.faa.gz; 05/10/2009) sequence release files for construction of the experimental dataset. Protein sequences flagged as putative, hypothetical, potential, uncharacterized, similar to the predicted protein, membrane, porin, receptor are deleted from the initially downloaded RefSeq sequence release files in the preparation of experimental dataset. The prokaryotic sequence dataset was created by merging the sequence entries from PSORT db and refseq dataset after appropriate deletions. Similarly, the eukaryotic dataset was prepared after deleting and merging the sequence entries from eSLDB and refseq dataset.

The entire dataset used for computing the composition of 20 amino acid residues comprised of prokaryotic (63644) and eukaryotic (88400) nMP sequences. The amino acid composition for the prepared datasets was computed using the number of amino acids of each type and the total number of residues. It is defined as Residue composition (%) (r) = $\frac{\sum n_r}{N} \times 100$ (1) where 'r' stands for any one of the 20 amino acid residue. $\sum n_r$ is the total number of residue of each type and N is the total number of residues in the dataset.

Results and Discussion

The amino acid compositional distribution between prokaryotic and eukaryotic nMPs was computed using eq. (1). The prokaryotic nMPs shows the dominant occurrence of a non-polar amino acid 'Ala' ($\sigma = 0.45$) while the eukaryotic nMPs predominantly possess the polar amino acids 'Lys' ($\sigma = 0.66$), 'Ser' ($\sigma = 0.60$) and 'Cys' ($\sigma = 0.29$) (Figure 1). In prokaryotic nMPs, the high frequency of short side-chained non-polar aliphatic amino acid 'Ala' may be due to various possibilities such as its over-representation in highly expressed proteins (Tats et al., 2006), its role in determining the cleavage of N-terminal formyl methionine (Solbiati et al., 1999), its role in assisting the entrance of the nascent peptide chain into the ribosomal tunnel (Tenson et al., 2002) and in helix-helix packing (Eyre et al.,

***Corresponding author:** Rajneesh Kumar Gaur, Bioinformatics Infrastructure Facility, Jamia Hamdard (Hamdard University), Hamdard Nagar, New Delhi, India – 110062, Tel: +91 9990290384; E-mail: meetgaur@gmail.com

Received September 30, 2009; **Accepted** December 27, 2009; **Published** December 27, 2009

Citation: Gaur RK (2009) Prokaryotic and Eukaryotic Non-membrane Proteins have Biased Amino Acid Distribution. J Comput Sci Syst Biol 2: 298-299. doi:10.4172/jcsb.1000045

Copyright: © 2009 Gaur RK. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

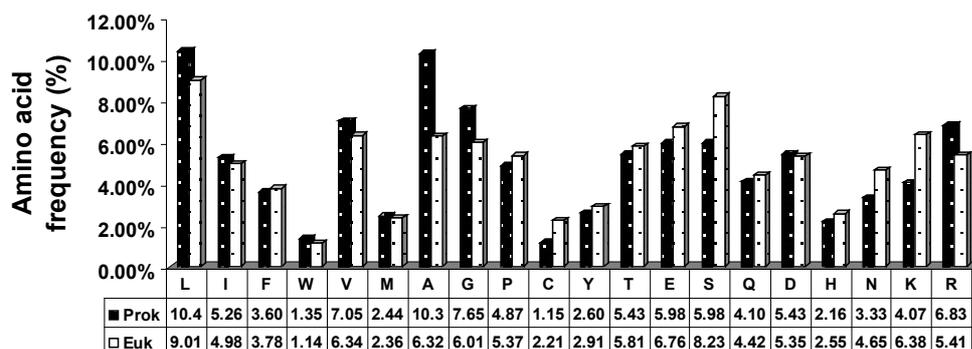


Figure 1: Histogram showing the overall amino acid composition of prokaryotic (black bars) and eukaryotic (white bars) nMPs. The amino acids are arranged in decreasing order of hydrophobicity. Pro: Prokaryotic nMPs; Euk: Eukaryotic nMPs

2004). Though ‘Ala’ might perform the similar functions in both prokaryotic and eukaryotic nMPs but its higher frequency in nMPs probably related to the higher proportion of prokaryotic helical nMPs.

The eukaryotes show the high occurrence of positively charged polar residue ‘Lys’ in their nMPs repertoire. This positively charged residue helps in the secretion of proteins through the membrane via interaction with export machinery and signal recognition particles (vonHeijne, 1984). The overabundance of ‘Ser’ in eukaryotic nMPs may be due to their ability to form H-bonds and stabilizing the helices (Subramaniam et al., 2006). In particular, the two-fold higher ‘Cys’ of eukaryotic nMPs compared to prokaryotic nMPs most probably compensates for their lower hydrophobicity (D’Onofrio et al., 1999).

Acknowledgement

I express my gratitude to the Council of Scientific and Industrial Research (CSIR), New Delhi, India for granting me the Senior Research Associateship. I am also thankful to Dr. Sayeed Ahmed, Faculty of Pharmacy, Jamia Hamdard University, New Delhi, India for extending his computational facility.

References

- Cedano J, Aloy P, Perez-Pons JA, Querol E (1997) Relation between amino acid composition and cellular location of proteins. *J Mol Biol* 266: 594-600. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- D’Onofrio G, Jabbari K, Musto H, Bernardi G (1999) The correlation of protein hydrophobicity with the base composition of coding sequences. *Gene* 238: 3-14. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Eyre TA, Partridge L, Thornton 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. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Gerstein (1998a) How representative are the known structures of the proteins in a complete genome? A comprehensive structural census. *Fold Des* 3: 497-512. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- 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. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Pierleoni A, Martelli PL, Fariselli P, Casadio R (2007) eSLDB: Eukaryotic subcellular localization database. *Nucleic Acids Res* 35: D208-212. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- 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. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- 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. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Sobolevsky Y, Trifonov EN (2005) Conserved sequences of prokaryotic proteomes and their computational age. *J Mol Evol* 61: 591-596. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Solbiati J, Chapman-Smith A, Miller JL, Miller CG, Cronan JEJ (1999) Processing of the N termini of nascent polypeptide chains requires deformation prior to methionine removal. *J Mol Biol* 290: 607-614. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Subramaniam S, Henderson R (2000) Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature* 406: 653-657. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Subramanyam MB, Gnanamani M, Ramachandran S (2006) Simple sequence proteins in prokaryotic proteome. *BMC Genomics* 7: 141. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Tats A, Remm M, Tenson T (2006) Highly expressed proteins have an increased frequency of alanine in the second amino acid position. *BMC Genomics* 7: 28. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Tenson T, Ehrenberg M (2002) Regulatory nascent peptides in the ribosomal tunnel. *Cell* 108: 591-594. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- vonHeijne G (1984) Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. *EMBO J* 3: 2315-2318. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
- Zhang CT, Chou KC (1992) An optimization approach to predicting protein structural class from amino acid composition. *Protein Sci* 1: 401-408. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)