International Journal of Bioinformatics Research, ISSN: 0975–3087, Volume 1, Issue 2, 2009, pp-1-3

What Might be the Difference in Viral Proteins?

Akila K.[1] and Rajasekaran E.*
*Department of Biotechnology, Periyar Maniammai University, Vallam, Thanjavur – 613403, ersekaran@gmail.com
[1]Department of Bioinformatics, Jamal Mohamed College, Trichy – 620020

Abstract- The difference between normal and viral proteins is the way the amino acids organized in some form. Though both proteins use same 20 naturally occurring amino acids, it differs in arrangements. This difference cannot be felt by merely comparing the amino acid sequence but at atom level. This is analysed and compared in terms of carbon distribution as carbon is the only element that contributes towards the dominant force, hydrophobic interaction. For this purpose the arenaviruses have been selected here for study. The protein sequences of 7 arenaviruses are analysed. Our results show that the carbon distribution in viral proteins is different from the normal proteins. Either the carbon content is different or the distribution is not uniform. This carbon distribution analysis uses our earlier report of 31.44% of carbon for its structure and activity.
Keywords: carbon distribution, arenaviruses, theoretical study, sequence analysis

INTRODUCTION

Carbon is the only element contributes towards the dominant force, hydrophobic interaction in proteins. Protein's carbon content evolves in response to carbon availability and may influence the face of duplicated genes [1]. This carbon content and distribution in viral proteins is expectedly different from normal one because the organization of amino acids in viral proteins is different though both use same 20 naturally occurring amino acids. It is reported that proteins prefers to have 31.44% carbon for its stability [2]. This fraction of carbon content can be used as standard of carbon measurement and comparison [3]. The difference in carbon distribution in viral proteins is the focus of this work by taking arenavirus as case study.

The family arenaviridae consists of two antigenic groups, the New World arenaviruses found in the Americas and Old World arenaviruses native to Europe and Africa. The arenaviruses has a tendency to cause silent infections in their natural hosts, rodents and humans. The genome of length 5000-7400 nucleotides consist of two single-stranded RNAs, the small (S) and the large (L), segments. Each segment encodes two different proteins. The S RNA encodes the nucleocapsid protein (NP) and the glycoprotein precursor (GPC) which undergoes post-translational processing to yield two mature proteins (GP1 and GP2). The L RNA encodes the viral RNA-dependent RNA polymerase (L) and a zinc-binding matrix protein (Z). Nucleotide sequences of 3'-terminus largely complementary to similar regions on the 5' end; conserved nucleotide sequences; the same in species of same genus; S on RNA; 19-30 nucleotides long. Encapsulated nucleic acid both genomic and non-viral; including three molecules of host ribosomal RNA. Only seven of the twenty three arenaviruses are associated with human disease which are listed here.

METHODOLOGY

The four different protein sequences of seven arenaviruses are collected from the NCBI. The amino acid compositions are calculated using AACOMP program. The results are discussed and not shown in here. The carbon distribution in these proteins is computed using CARANA program which uses the principle of 31.44% of carbon [2]. The CARANA program can be accessed online [6]. The results on carbon distribution as a function of atomic position are plotted as shown in figures.

RESULTS AND DISCUSSION

We have studied the carbon distribution in protein sequences of 7 arenaviruses containing 4 different proteins. The name of the viruses studied and length of each sequence are given in table 2. As can be seen from the Fig.1, the carbon content in GPC protein is greater than the expected value of 31.44 in entire sequence except in few places. This protein could possibly interact with several other biomolecules that initiate or suppress biochemical reactions. The initial portion of this protein contains higher carbon content that could hold other interacting molecules for sliding or for further reaction. On the other end the carbon content is less than the expected value that may not take part in biochemical reactions. This end may either be silent in water or can have electrostatic interact6ion with other molecules. Similarly, at position around 5250 atom number (or 333 amino acid number) there is a dip in carbon content. The plots for other proteins such as NP-protein, L-protein and Z-protein are not shown but discussed. The Z-protein is a small protein

shows a higher amount of carbon content all along the sequence except at the end of the sequence. There is no single functional site found in this protein. The Z-protein of LCMV seemingly lesser amount of carbon compared to other viruses. This protein's carbon distribution in LCMV and Sabia viruses are quite different from other viruses. This indicates that this protein may not have a specific function in all viruses. NP-protein on the other hand having similar distribution in all viruses. The carbon distribution appears like normal proteins with couple of active sites and adequate carbon content. There is hydrophilic region for length of 99 amino acids between amino acid position 30 and 129. The L-protein is a large protein showing again rich amount of carbon. Overall these viral proteins show up higher amount of carbon. The Large hydrophobic residues (LHR) such as F, I, L, M and V are the major contributors towards this carbon content [4]. These LHRs are coded by codon XTX (where X = A, T, G or C) [5]. We have further investigated the role of thymine in mRNA sequences of these proteins. The results indicate that there is greater amount of thymine in all cases. Adenine is lesser compared to thymine in all cases but in L-genome of LCMV. This difference is also observed in carbon distribution of Z-protein of LCMV. The amount of AT together is always greater than GC in both genome and in all viruses. This is not the case in human mRNA sequences.

CONCLUSION

To conclude, carbon content and distribution are different in viral proteins. The carbon distribution studies on viral proteins reveal that the viral proteins show up with higher amount of carbon. The atomic composition plays a role in evolution of proteins. The difference in carbon distribution in proteins causes disease. The carbon distribution study along the protein chain is the most significant step towards understanding the biological reactions. The mRNA sequences of these proteins show up with greater amount of thymine all cases. The amount of AT together is always greater than GC in both L and S genome of all viruses. This is not the case in human mRNA sequences. That is to say nature is adding more hydrophobic proteins to the host cell through viruses.

REFERENCES

.

[1] Bragg J.G. and Wanger A. (2007), *Proc. R. Soc*. B, 274, 1063-1070.

[2] Rajasekaran E., Vinobha C.S., Vijayasarathy M., Senthil R. and Shankarganesh P. (2009) IACSIT-SC, IEEE 452-453.

[3] Senthil R. and Rajasekaran E. (2009), *J. Adv. Biotech*. 8(9), 30-31.

[4] Jayaraj V., Suhanya R., Vijayasarathy M., Anandagopu P. and Rajasekaran E. (2009) *Bioinfo*. 3(9) 4.09-412.

[5] Anandagopu P., Suhanya S., Jayaraj V. and Rajasekaran E., (2008) *Bioinfo*., 2(7), 304-307.

[6] www.rajasekaran.net.in/tools/carana.html

Table 1: The Arenavirus Disease and its geographic distribution

Virus	Disease	Geographic Distribution
Lymphocytic Choriomeningitis virus	Meningitis	Europe, Americas
Lassa virus	Hemorrhagic fever	West Africa
Junin Virus	Argentine hemorrhagic fever	Argentina
Machupo Virus	Bolivian hemorrhagic fever	Bolivia
Guanarito	Hemorrhagic fever	Venezuela
Sabia virus	Brazilian hemorrhagic fever	Brazil
White water arroyo virus	Hemorrhagic Fever	North America

Table 2: Number of amino acids in all 4 types of proteins and 7 arenaviruses

Name of the Virus	GPC-protein	NP- Protein	L-Protein	Z-Protein
Guanarito virus (GV)	479	560	2198	95
Junin virus (JV)	485	564	2210	94
Mopeia Lassa reassortant 29 (LV)	491	569	2220	99
Lymphocytic choriomeningitis virus (LCMV)	498	558	2209	90
Machupo virus (MV)	496	564	2209	94
Whitewater Arroyo virus (WWAV)	480	562	2219	95
Sabia virus (SV)	488	562	2212	100

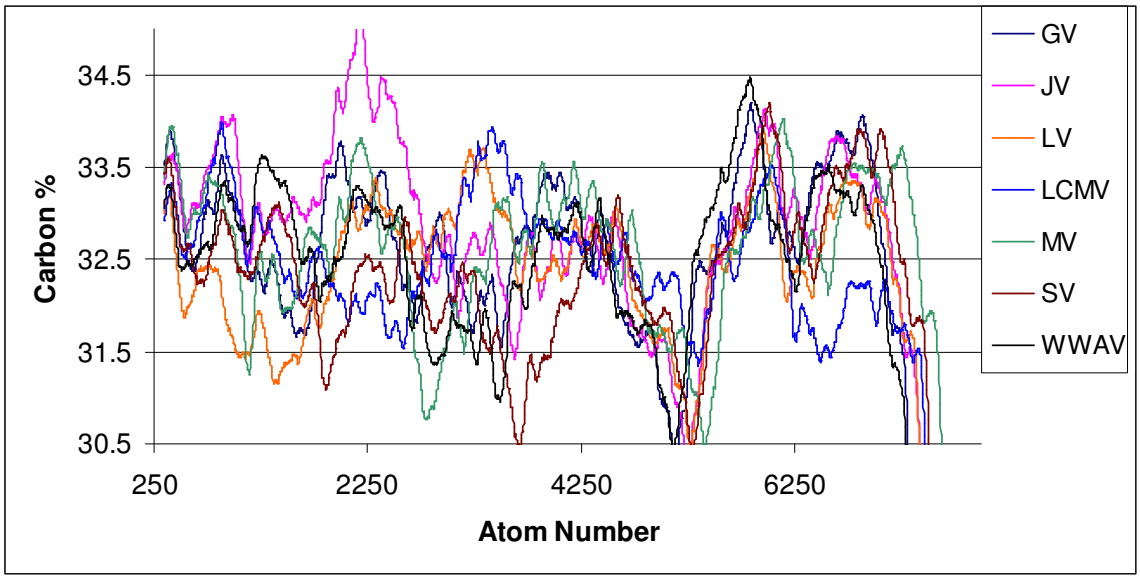

Fig 1: Carbon distribution in the GPC protein in all 7 viruses.