Adenine in Viral mRNAs Manipulate the Carbon in Proteins

E. Rajasekaran and K. Akila

***Abstract*—Proteins function is determined by the nature of carbon along its sequence. This carbon along the protein is determined by the nucleotides in the gene. The adenine in viral genetic material and carbon distribution along its proteins are the focus of this study. The results reveal that the presence of higher adenine content in viral material add adequate number of large number of hydrophobic residues in its proteins. The natural way of adding different sequences during development and evolution is better understood based on this carbon distribution analysis. Appropriate mutation that changes in carbon content and distribution in viral proteins might improve the functionality of the protein.**

***Index Terms*—Carbon distribution, Viral genome, Adenine, Large hydrophobic residues, Carbon content.**

I. Introduction

Viruses cause diseases from common cold to acquired immunodeficiency syndrome. Viruses are genetic material coated by proteins. It can be either RNA or DNA. The genetic material has instruction for its multiplication. The infected virus instructs the host cell to duplicate it. Some time or the other, all are infected by these viruses.

Matrix proteins are understandably responsible for virus assembly and budding. These structural proteins link the viral envelope and the core. These proteins are proved to be responsible for expelling the genetic material once the virus entered into the host cell. What constitute the matrix protein to assemble and bud a virus? Why and how it expel genetic materials from host once it enters into host?

The matrix proteins of different viruses are analysed for distribution of carbon content and large hydrophobic residues. As uracil plays an important role in translating mRNA into proteins with adequate carbon [1]. The amount and distribution of adenine in viral genome is a factor in deciding the protein stability from hydrophobicity point of view. The genetic material of different viral genome and different segments of Influenza A H1N1 virus are analysed in this study. The role of uracil in mRNA sequence of these matrix proteins is investigated. Particularly the sequences of Chikungunya, Dengue, Rabies, Japanese encephalitis and west nile viruses are analysed.

II. Methodology

The genome and protein sequences of influenza A virus H1N1 are collected from the NCBI. The amino acid compositions of 11 proteins are calculated using in house program. The carbon distributions in these 11 proteins are computed using our CARBANA program [2] available online which uses the principle of 31.45% of carbon. Length of 700 atoms selected for calculation. The results on carbon percentage versus atomic positions are plotted as shown in figures (Fig. 1).

The genomic content of influenza a H1N1 contains 8 segments. The number of bases in each segment is counted and tabulated in Table I.

The role of uracil in mRNA sequence of matrix proteins is investigated by counting number of large hydrophobic residues (FILMV) which are coded by XUX (X=A,U,G or C). The protein sequences of Sendai virus (P06446), Vesicular stomatitis Indiana virus (P03519), Influenza A virus (P05777), Human respiratory syncytial virus (P03419), Zaire ebolavirus (Q05128) and Nipah virus (Q9IK90) are collected from SWISSPROT protein database.

The AUGC content of viral genomes of Chikungunya, Dengue, Rabies, Japanese encephalitis and West nile counted and shown in Table II.

III. Results and Discussion

Carbon is the only element that contributes towards the dominant force, hydrophobic interaction in proteins. Proteins evolve based on carbon content and may influence the code in genes. It is reported that proteins prefers to have 31.45% of carbon for its stability [2],[3]. Depending upon the carbon content, the protein and the corresponding mRNA survive and passed to next generation. The carbon content is determined by the presence of different types of amino acids. The arrangement of these amino acids is instructed in its gene. The adenine in gene is transcribed as uracil in mRNA. The uracil in mRNA is responsible for number of large hydrophobic residues in proteins. The number of uracil in mRNAs is decreased during evolution [4]. Due to this reason the number of large hydrophobic residues decreases which causes production of proteins with less carbon content [5],[6]. This can change the proteins a non functional. On the other side the viruses are stitched into host cell that produces proteins with high carbon content. This is again possible because of adenine in viral genetic materials is transcribed into uracil in the mRNAs.

The influenza A virus H1N1 genome contains 11 genes with eight segments of RNAs, encoding for 11 proteins namely PB2, Polymerase 1(PB1), PB1-F2, Polymerase PA(PA), Haemagglutinin (HA), Nucleocapsid protein(NP), Neuraminidase(NA), Matrix proteins(M2, M1), Nonstructural proteins(NS2, NS1). The results on carbon distribution in these proteins are shown in Fig. 1. PB1-F2 and M2 are very small proteins which are not shown here.

Manuscript received June 16, 2011; revised November 7, 2011.

E. Rajasekaran is with the Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Karunya Nagar, Coimbatore 641114, Tamil Nadu, India (corresponding author: phone: +91-9524825876; e-mail: ersekaran@gmail.edu).

K. Akila is with the Periyar Maniammai University, Thanjavur, 613403.

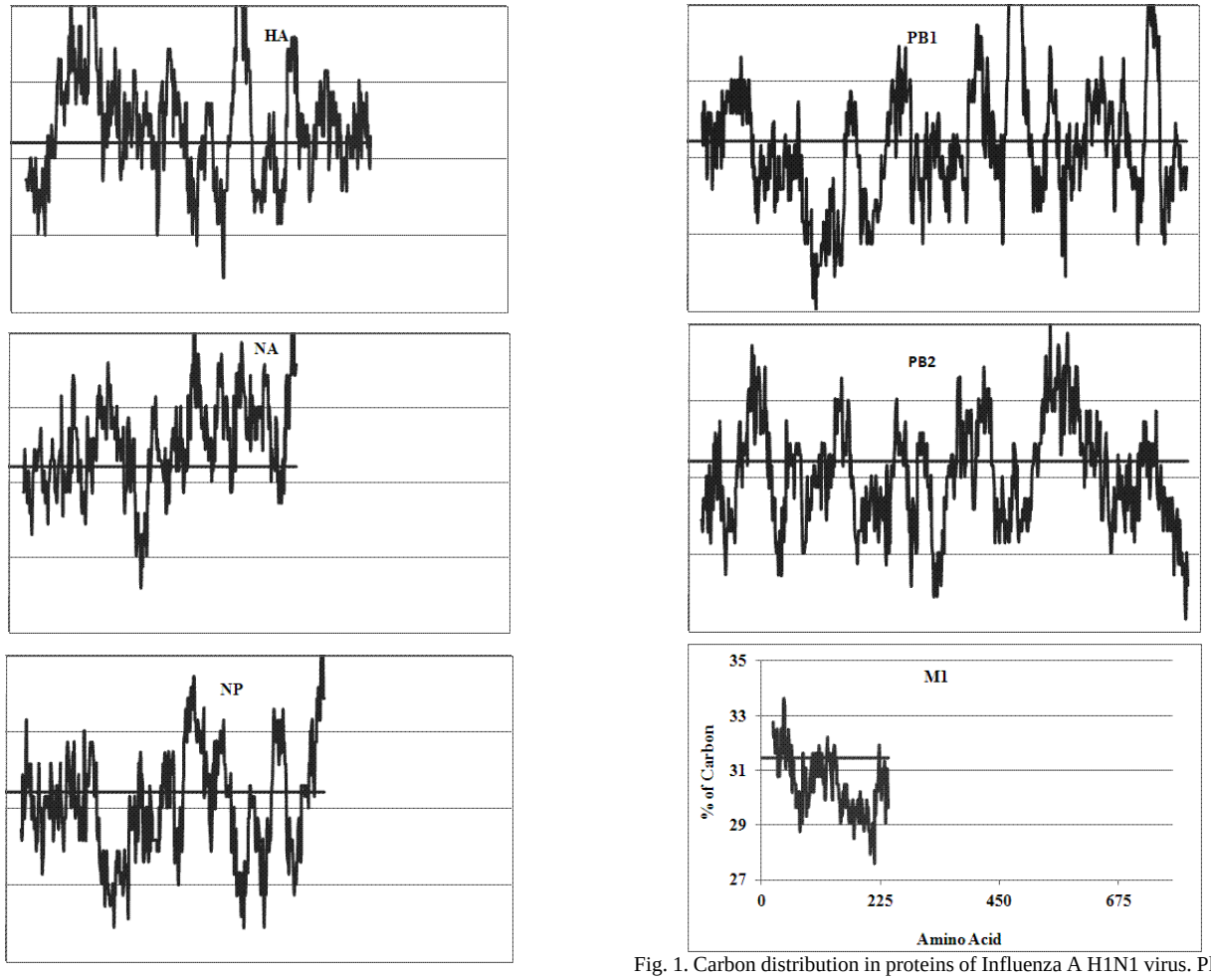

Fig. 1. Carbon distribution in proteins of Influenza A H1N1 virus. Plot of atom number (X-axis) versus % of carbon (Y-axis).

The two membrane proteins (HA and NA) clearly show the rich of carbon contents all along the sequences. Normally a threshold value of 31.45% is expected along the sequences. NP is a nucleoprotein which wrap around genomic RNA forming a ribonucleoprotein complex. This protein contains a mix of both hydrophobic and hydrophilic regions along the sequence. The PA and PB1 are viral RNA polymerase proteins which show a normal carbon distribution. PA protein has slightly higher carbon content. The two small non-structural proteins, NS1 and NS2 are shown with normal carbon content which is not significant. The matrix protein, M1 has less amount of carbon content. There are regions of hydrophilic in nature. The matrix proteins and haemagglutinin proteins play an important role in the viral entry to the host cell. So this carbon distribution study will certainly help to modify these proteins for efficient function.

TABLE I: THE AUGC CONTENT IN DIFFERENT SEGMENTS OF INFLUENZA A H1N1 VIRUS. NOTE THAT HIGHER NUMBER OF ADENINE (A) IN EACH SEGMENTS.

Segment no.	No. of Proteins(GENE)	AUGC count				
		A	U	G	C	Total
1	1 (PB2)	786	531	597	427	2341
2	2 (PB1, PB1-F2)	811	545	526	459	2341
3	1 (PA)	751	545	521	416	2233
4	1 (HA)	621	417	409	331	1778
5	1 (NP)	504	335	412	314	1565
6	1 (NA)	428	381	351	253	1413
7	2 (M2, M1)	294	248	268	217	1027
8	2 (NS2, NS1)	285	211	215	179	890

The genomic particulars of the Influenza A H1N1 virus are given in Table I. That is the number of total and individual bases are given in this table. Number of proteins and its gene identifiers are given in the second column. Note that the number of adenine is higher in all segments. Because of this high adenine content the host cell produces gene with higher amount uracil that is responsible for producing proteins with higher number large hydrophobic residues (including F, I, L, M and V) which will make the protein with higher carbon content. Analysis of mRNA sequences reveals that the presence of higher adenine content that might add appropriate number of large hydrophobic residues in the proteins. In particular, the adenine as second nucleotide in frame 1 is important to have appropriate uracil as second nucleotide in frame 4 in back chain. Framing these adenine (uracil at the back chain) is important for producing highly functional and long lasting proteins that will be passed on to the next generation. This is the classical example of adding more hydrophobic proteins to the host cell naturally. In human mRNAs the adenine content is less. The above principle can be adapted for adding genomic content that will give adequate carbon distribution in the proteins.

The carbon distribution studies on viral proteins reveal that the viral proteins show up with higher carbon content. The atomic composition plays a role in evolution of proteins. The carbon distribution along the protein chain is dominated by presence of uracil in frame 1 of its mRNA. The adenine in viral genome is responsible for uracil in the corresponding mRNA. The amount of adenine is always greater than other bases in all viruses (see Table II). Probably nature is adding more hydrophobic elements in the proteins through viruses.

All large hydrophobic residues such as phenylalanine (F), Isoleucine (I), leucine (L), methionine (M) and valine (V) are coded by codons XUX (where X = A, U, G or C). Earlier investigation on this codon distribution in mRNA sequences of different species reveals that, the frame 1 prefers to maintain some degree of this codon for maintaining sufficient hydrophobicity in its proteins. The matrix protein of different species are analysed for the presents of XUX in frame 1. That is the number and percentage of large hydrophobic residues is calculated. It reveals the presence of excess uracil in the coding sequences.

The carbon content of matrix proteins are further analysed with CARBANA tool. The results for sendai virus is shown in Fig. 2. The carbon content in matrix protein is greater than the expected value of 31.45% in entire sequence except in few places. The first half of the sequence appears higher carbon content than the next half. The number of large hydrophobic residues in the first half is 55 and the other half has 75. Caution must be note that residues like tryptophan, tyrosine, proline, histidine, glutamate and aspartate also contributing the carbon content. Similarly residues such as arginine, cystein, lysine, serine, glycine, aspatamine, threonine, glutamine and alanine reduce the carbon amount.

TABLE II: THE AUGC CONTENT (ADENINE RICH) IN DIFFERENT VIRUSES.

S. No.	Virus	AUGC count				
		A	U	G	C	Total
1	Chikungunya virus	3517	2391	2971	2947	11826
2	Dengue virus type 1	3426	2299	2770	2240	10735
3	Rabies	3421	3130	2736	2645	11932
4	Japanese encephalitis virus	3047	2285	3113	2531	10976
5	West Nile virus	3017	2370	3172	2470	11029
6	Yellow fever virus	2963	2498	3087	2314	10862

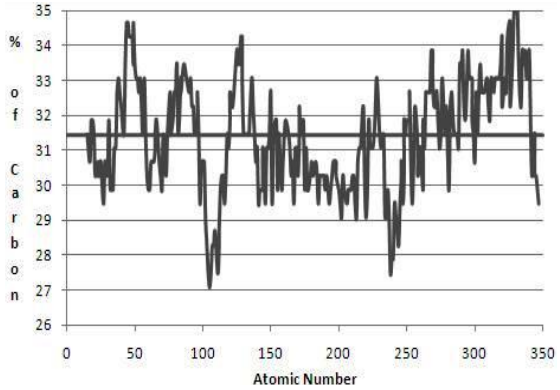

Fig. 2. Carbon distribution in matrix protein of sendai virus.

IV. CONCLUSION

The carbon distribution studies on viral proteins reveal that the carbon content and distribution along the sequences is vital for its function. A difference in carbon distribution pattern is noticed in most of the H1N1 proteins. The distribution is not normal. The difference in carbon distribution in proteins causes diseases. The carbon content and distribution plays a role in evolution of proteins. The carbon distribution study along the protein chain is the most significant step towards understanding the biological reactions. Higher number of adenine in mRNAs is noticed to be playing an important role in producing proteins with higher number of large hydrophobic residues which is responsible for richness in carbon content. Further analysis of different viral genome reveals that the presence of higher adenine content might add appropriate number of large hydrophobic residues in the proteins. In particular, the adenine as second nucleotide in frame 1 is important to have appropriate uracil as second nucleotide in frame 4 in back chain. Framing these adenine (uracil at the back chain) is important for producing highly functional and long lasting proteins that will be passed on to the next generation. But otherwise in human mRNAs the adenine content is less that needs to be worked out for disease free living. This is the classical example of adding more hydrophobic proteins to the host cell naturally.

Apart from large hydrophobic residues, there are other residues which can contribute to the total carbon content of a protein for floating which needs to be taken into account.

REFERENCES

[1] E. Rajasekaran and P. Anandagopu, "Reduction of thymine in mRNA sequence of tumor protein", *J. Adv. Biotech.*, 9(7), 9-10. 2010

[2] E. Rajasekaran and M. Vijayasarathy, "CARBANA: Carbon analysis program for protein sequences", *Bioinformation*, 5(10), 455-457, 2011.

[3] E. Rajasekaran, C. S. Vinobha, M. Vijayasarathy, and R. Senthil and P. Shankarganesh, "The nature of proteins", Proc. (IACSIT 2009) IEEE CS, 452-453, 2009.

[4] P. Anandagopu, S. Suhanya, V. Jayaraj, and E. Rajasekaran, "Role of thymine in protein coding frames of mRNA sequences", *Bioinformation*, 2(7), 304-307, 2008.

[5] V. Jayaraj, R. Suhanya, M. Vijayasarathy, P. Anandagopu, and E. Rajasekaran, "Role of large hydrophobic residues in proteins", *Bioinformation*, 3(9), 409-412, 2009.

[6] C. S. Vinobha and E. Rajasekaran, "Comparative analysis on large hydrophobic residues and small hydrophobic resides in different organisms", Int. *J. Bioinfo. Res*., 3(1), 115-117, 2011.