Bioinformatic Approach for the Identification of Hepatitis B Viral Insert in The Exon Region of Human Genome

P. Pandarinath[1]*, M. Shashi[1] and A. Appa Rao[1]

Department of Computer Sciences, College of Engineering (A) Andhra University, Visakhapatnam, India

Abstract

Hepatitis B viral sequence was downloaded from NCBI Tax Browser and scanned against complete genome of Homo sapiens for the presence of possible viral inserts in human genome. The alignments which showed more than 25-30 residues, 90-100 % identities and localization in the exon regions were only considered. The results from the computational analysis revealed that Hepatitis B virus resulted in viral segment inserted in the exon region of human genome.

Introduction

Humans have been carrying unwanted viral gene segments since many years and reports suggests that approximately 3-8 % of the human genome has been comprised of viral DNA. During the course of a viral infection, some viruses insert their DNA into the host's genome (proviruses) and direct the host cellular machinery to make the proteins and genetic material needed to make / assemble more viruses. If this gene insertion takes place in a cell, the host's offspring will have a copy of the virus in every single cell. These and other parasite, self-replicating pieces of nucleic acids have evolved with us over millions of years after being inserted into the human DNA by the virus that infected our ancestors.

Because of these viral gene insertion events, genetic material from inactive viruses accounts for roughly 3 percent of the human genome. Nearly 30-50 copies of HERV-K exist in the human genome, and that some of the copies appear to be active at a low level in normal testicular and placental tissue (Jin et al., 1999). Earlier Japanese researchers found copies of the Bornavirus N (nucleoproteins) gene inserted in at least four separate locations in the human genome and clearly provided a fossil record of bornavirus (Tina, 2010).

It has been since decades that many viruses have been infecting vertebrates, humans in particular. Experimental evidence suggests that HIV, Epstein-Barr viral segments were inserted in few regions of the human genome. Recent developments in the field of Bioinformatics owed to make an attempt to identify the insertions of various viral gene segments in human genome and computational analysis made previously revealed that dengue virus resulted in viral segments inserted in the intron regions of human genome and exon region insertions were observed with polio and simian enterovirus (Pandarinath et al., 2010).

Hepatitis B virus (HBV) is a hepatotropic DNA virus belongs to *Hepadnaviridae*. The full-length of the viral genome is about 3.2kb. HBV infection can cause acute and chronic type B hepatitis, and eventually leads to serious consequence, such as hepatic cirrhosis and primary hepatocellular carcinoma (PHC) etc. Following similar approaches, the aim of the present study was to identify the Hepatitis B viral segment if any in human genome. Keeping this aspect into consideration, Hepatitis B virus sequence was downloaded from NCBI Tax Browser and scanned against complete genome of *Homo sapiens* for the presence of any inserted segments.

Material and Methods

The genomic sequence of Hepatitis B virus *was* downloaded from the NCBI Genbank database. *Blastn* algorithm was used to perform simlarity search using default parameters and scanned against the complete genome of *Homo sapiens*. The option, somewhat similar sequences was selected in order to obtain probable matches. The program compares the query sequence with the human genome sequence and calculates the statistical significance of matches (Altschul et al., 1990).

Ensembl consists of genomes of all vertebrates and eukaryotic species; the query genome is compared to know whether any genes are present within the exon region of complete human genome. PDBsum provides an at-a-glance overview of every macromolecular structure deposited in the Protein Data Bank (PDB), giving schematic diagrams of the molecules in each structure and of the interactions between them. It also gives the information about secondary structure of respective protein. Sequence tools are used for basic and advanced analysis of nucleotide and protein sequence. The hydropathy plot was used to find cluster of hydrophobic amino acids, which indicates the polypeptide in question is a transmembrane protein.

A commercial software Molegro Virtual Docker was used to analyze organic and inorganic structures, proteins, DNA/RNA, and crystals. The structures of High Mobility Group Box 2 protein were downloaded from the PDB and used in the Molegro Virtual Docker to display secondary structure view and docking view and many other features of the protein. Molecular Genetic Evolutionary Analysis

TYPE	AMINOACIDS BEARING (EXON REGION)
6 helices	E,E,E,E,E,E,E,E,D,D
9 helix-helix turns	E, E
2 beta turns	P, N
1 gamma turn	Nil

Table 1: PDBsum secondary structural information of the protein.

(MEGA 4) was used to analyze sequence alignments to estimate evolutionary distances. In the MEGA 4 the viral sequence and the human genome sequence were compared to highlight conserved sights, nucleotide comparison, nucleotide pair frequencies from the statistics option.

Results and Discussion

The blast result of the Hepatitis B viral nucleotide sequence and the human genome shows that the sequence was present in the exon region of High-Mobility Group box 2 (chromosome 4) sequence of *Homo sapiens* with a match of 42 residues (Supplementary Figure 1). The accuracy of exon repeats in human sequence was confirmed by comparing it with human genome in ENSEMBL. The sequence in fasta format of the High-Mobility Group box 2 (subject sequence region) and Hepatitis B viral nucleotide (query sequence region) were aligned using bl2Seq. The encoded amino acid matching for both the sequences i.e., DEEEEEEEEDEPEN was visible in the sequence view of the graphics page (Supplementary Figure 2). The results of SEQ tools show the codon usage, the GC content of the viral genome 22 (42.3%), and the AT content of the viral genome 30 (57.7%) (Supplementary Figure 3 (a)); and for human genome the GC content is 23 (46.9%) and the AT content is 26 (53.1%) (Supplementary Figure 3 (b)). The protein sequence shows the properties and the hydropathy plot i.e., protein length was 90 aa and the molecular weight was 10,196 (Supplementary Figure 4). The secondary structure and the related information of the protein were retrieved by using PDBsum (Table 1). Docking view (Supplementary Figure 5 (a)) and Secondary structure view (Supplementary Figure 5 (b)) of High-mobility Group box 2 protein were viewed using Molegro Virtual Docker. The identical pairs (42), transitional pairs (1) and transversional pairs (6) were also obtained (Supplementary Figures 6 (a) and 6 (b)) using MEGA4 program.

Conclusions

From computational analysis, it was observed that Hepatitis B virus resulted in viral segment inserted in the exon region of human genome. Nearly 42 nucleotide residue segment (80% identities) of Hepatitis B virus was found to be inserted in the exon region of High-Mobility Group box 2 (chromosome 4) sequence of *Homo sapiens*. The present work indicates that with few computational efforts viral inserts in human genome can be identified and represents that further analysis needs to be carried out to study the influence of such inserts on the structural and functional features of respective genes.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410.

2. Hong C, Liang P, Zhenhua D, Zhihua L, Xiaolu Liu (2008) Computer-based siRNA Design to Hepatitis B Virus and Its Expression Vector Construction. International Journal of Advanced Science and Technology 1: 99-108.

3. Jin Y, Hal PB, Sheila P, Heather W, Ray T, et al. (1999) An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein. Proc Natl Acad Sci U S A 96: 13404-13408.

4. Pandarinath P, Shashi M, Allam A R (2010) Computational study of viral segments inserted within the regions of human genome. J Comput Sci Syst Biol 3: 74-75.

5. Tina HS (2010) Bornavirus genes found in human DNA. Sci News 177: 10.