Journal of Global Biosciences
ISSN 2320-1355
Volume 4, Number 3, 2015, pp. 1747-1757
Website: www.mutagens.co.in
E-mail: submit@mutagens.co.in
researchsubmission@hotmail.com

Research Paper

IN SILICO CHARACTERIZATION AND INVESTIGATION OF PUTATIVE PROMOTER MOTIFS IN *EBOLAVIRUS* GENOME

Mohammad Mahfuz Ali Khan Shawan[1], Md. Mozammel Hossain[1], Md. Mahmudul Hasan[1], Afroza Parvin[1], Salina Akter[1], Kazi Rasel Uddin[1], Subrata Banik[1], Mahbubul Morshed[1], Md. Ashraful Hasan[1], Md. Nazibur Rahman[1] and S. M. Badier Rahman[1]

[1]Department of Biochemistry and Molecular Biology,
Jahangirnagar University, Savar, Dhaka-1342,
Bangladesh.

Abstract

The genus *Ebolavirus* comprises with clinically significant viral pathogens, namely *Bundibugyo ebolavirus*, *Reston ebolavirus*, *Sudan ebolavirus*, *Tai Forest ebolavirus* and *Zaire ebolavirus* causing deadly haemorrhagic fever in human with mortality rate approaching 90% and thus posses a potential global health risk. Ebola infection can be characterized by immune suppression results in disintegration of vascular, coagulation and immune system leading to multi organ failure and shock. The recent Ebola epidemic in Africa is spiraling out of control and no approved therapies/vaccines are available yet for its treatment. The genomes of *Ebolavirus* are approximately 18.8-19.0 kb long; encode both structural and non structural proteins. The structural proteins play key role in attachment, entry, stability, gene expression and pathogenesis of virus within the host. Therefore, in this *in silico* approach the complete genome sequences of five different species of *Ebolavirus* were used to characterize and investigate the putative promoter motifs. The promoter sequences were identified in all the species along with their name, sequence, weight and location within gene by PROSCAN Version 1.7. A grand total of 107 promoters were detected within different *Ebolavirus* genome; out of which 31, 29, 33, 10 and 4 promoter sequences were found in *Bundibugyo ebolavirus*, *Reston ebolavirus*, *Sudan ebolavirus*, *Tai Forest ebolavirus* and *Zaire ebolavirus* respectively. The genome size and G+C (%) of different *Ebolavirus* were almost equal. This current study may helps in identifying and analyzing various promoter motifs in *Ebolavirus*, thus understanding their roles in the regulation of gene expression, cell specificity and development. It may also be helpful in designing effective expression vector as well as live attenuated vaccine, hence inventing progressive target specific delivery system and fruitful gene therapy against different *Ebolavirus* species.

Key words: *in silico*, Promoter motifs, *Ebolavirus*, Genome, Haemorrhagic fever and Regulation of gene expression.

INTRODUCTION

The members of the *Ebolavirus* genus are considered as representative pathogen of viral haemorrhagic fever known as Ebola infection, causing initial manifestations include a sharp rise in fever, sore throat, muscle pain, headaches, explosive vomiting and diarrhea [1]. They are considered as a Category-A type infectious pathogen by the U.S. Department of Transportation's Hazardous Materials Regulations (HMR: 49 C.F.R., Parts 171–180) [2]. The genus *Ebolavirus* is inside in the family *Filoviridae* and order *Mononegavirales*. Viruses within this genus are called ebolaviruses and five known species has been identified namely: *Bundibugyo ebolavirus*, *Reston ebolavirus*, *Sudan ebolavirus*, *Tai Forest ebolavirus* and *Zaire ebolavirus* [3]. *Bundibugyo ebolavirus* was first detected as a new species of *Ebolavirus* on November, 2007 in Bundibugyo district, Uganda by World Health Organization (WHO); on the other hand *Reston ebolavirus* was discovered in 1989 during an outbreak of simian hemorrhagic fever virus from Reston, Virginia. On June, 1976 *Sudan ebolavirus* was first identified in Nzara, Sudan while in 1994, *Tai Forest ebolavirus* was discovered in chimpanzees from Tai Forest, Africa. *Zaire ebolavirus* is responsible for 2014, West Africa Ebola outbreak and has highest case fatality rate up to 90% in some epidemics. This species was first identified during an outbreak on August, 1976 in Yambuku [3,4].

The first outbreak caused by *Ebolavirus* was recorded in 1976, but the wild reservoir of this virus is still unknown. Recently researchers isolate antibody from fruit bats against *Reston ebolavirus* and *Zaire ebolavirus*, thus identified the fruit bats as a potential reservoir of this pathogen [5]. From 1976 through 2013, a total of 2000 confirmed cases had been reported by WHO, but as of early August, 2014 the number of present epidemic cases had already surpassed the number of the previous outbreaks combined. Recent outbreak (November, 2014) of the human *Zaire ebolavirus* infection starting in West African countries has resulted in 15351 infected patients. With an increased Ebola death toll, a total of 5459 deaths have been reported in six affected countries (Guinea, Liberia, Mali, Sierra Leone, Spain, and the United States of America) and two previously affected countries (Nigeria and Senegal) [6].

Members of *Ebolavirus* are filamentous enveloped RNA virus consists of 19 kb linear, non-segmented, single-stranded negative-sense RNA (Group V (- ssRNA)). The negative-sense RNA is complementary to viral mRNA and thus prior translation it must be converted to positive-sense RNA by RNA polymerase [7]. The genome encode seven structural proteins including nucleoprotein (NP), virion protein 35 (VP35), VP40, 3 glycoproteins (sGP/ssGP/GP1,2), VP30, VP24, RNA dependent RNA-polymerase protein (L-polymerase) and two non structural proteins including 3′ non-coding region (lader), 5′ non-coding region [8]. Transmembrane protein GP1 and GP2 is very crucial for viral attachment and pathogenesis. The GP1-GP2 heterodimer assemble as a trimer on viral surface and forms spike on the envelope membrane of mature viral particles. This complex contains a receptor-binding domain (RBD) responsible for attachment, entry and cellular tropism within the host cell and a mucin like domain to protect RBD from humoral and cell mediated immunity [9-12]. The secretary glycoprotein (sGP) synthesized at a large amount during viral infection and constitutes about 80% of total GP. It protects the virus against host humoral immune defense and cytokine-induced cytotoxicity during the early phase of viral infection [13]. The inner ribonucleoprotein complex is associated with NP, VP35, VP30 and RNA-dependent RNA polymerase to the functional transcriptase–replicase system that facilitate the transcription and replication process of *Ebolavirus* [14]. The VP35 and VP24 proteins are involved in hijacking transcription and translation machinery for robust genome replication by

suppressing the host innate immune response [15]. The VP30 is essential in maintaining the balance between transcription and replication process in *Ebolavirus* replication cycle [16]. The matrix proteins VP40 along with VP24 is highly important for structural stability and contribute to the regulation of viral genome replication and transcription [17].

Promoter is one of the most important regulatory regions that regulate the gene expression at transcription level. Transcription begins when specific sequences within the promoter are recognized by proteins, known as transcription factors and proceeds through the coding sequences till reaching the ends at the termination site. The promoter sequences also involved in the regulation of elongation step in a transcription process by means of clearing the RNA polymerase from promoter region. Therefore, it is vital to identify and analyze promoter sequences within the whole genome to regulate gene expression in different species of *Ebolavirus*. So considering the limited data on promoter motifs in *Ebolavirus* genome, current study was designed through *in silico* approach to characterize and investigate the putative promoter motifs in different species of *Ebolavirus*.

MATERIALS AND METHODS

Retrieval of genome sequences

The complete genome sequences of five different species of *Ebolavirus- Bundibugyo ebolavirus*, *Reston ebolavirus*, *Sudan ebolavirus*, *Tai Forest ebolavirus* and *Zaire ebolavirus* were retrieved from the most reliable biological databases such as NCBI (National Center for Biotechnology Information) cited at http://www.ncbi.nlm.nih.gov/genome/viruses/ [18].

Multiple sequence alignment

All of these five complete sequences were aligned with each other using clustalW 2.1 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) at GAP open, GAP extension and GAP distances of 10.0, 0.2 and 5.0 respectively [19].

Sequence similarity search

The sequence similarity for complete genome sequences of different *Ebolavirus* species was evaluated against whole Genebank database by nucleotide BLAST (Basic Local Alignment Search Tool) at expected threshold value 10 and BLOSUM 62 matrix as parameter implemented via the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [19].

Analysis of *Ebolavirus* genomes

The complete genome of five different *Ebolavirus* species were analyzed in FASTA format and the size (kb), total number of genes and proteins were determined by BioProject which is a collection of biological data related to a single initiative, originating from a single organization (http://www.ncbi.nlm.nih.gov/bioproject/). The G+C (%) was analyzed by GC calculator (http://www.genomicsplace.com/cgibin/gc_calculator.pl) [18].

Transcription initiation site

In genetics, transcription of a particular gene is initiated at a region of DNA called promoter. The transcription factor sites are over represented in the promoter region. The putative promoters in the genome of different species of *Ebolavirus* were

determined by using Promoter Scan (PROSCAN Version 1.7 at http://www-bimas.cit.nih.gov/molbio/proscan/) program at promoter cutoff score 53.000000. This program predicts promoter regions in primary sequence data based on scoring homologies with putative eukaryotic Pol II promoter sequences and transcription factor site density. This program recognizes about 70% of primate promoter sequences in the eukaryotic promoter database with a false positive rate of about one in every 14000 single strand bases. The program covers three databases including TF database, promoter database and non promoter set constructed from protein and RNA gene sequences [20]. In this study, we have identified the location and function of putative promoters in different *Ebolavirus* species [21].

RESULTS AND DISCUSSION

In this experiment, complete genome sequences of five different species of *Ebolavirus* named *Bundibugyo ebolavirus, Reston ebolavirus*, *Sudan ebolavirus*, *Tai Forest ebolavirus* and *Zaire ebolavirus* were retrieved from the accession NC_014373.1, NC_004161.1, NC_006432.1, NC_014372.1 and NC_002549.1 respectively. The similarities between these five genome sequences were analyzed by clustalW 2.1 which showed that, they have similarity score 65.26-71.9 between them. A phylogenetic tree with branch length was established to observe the evolutionary relationship among these five species **(Figure)**. BLAST result suggests that, these sequences were similar only with different species of *Ebolavirus*.

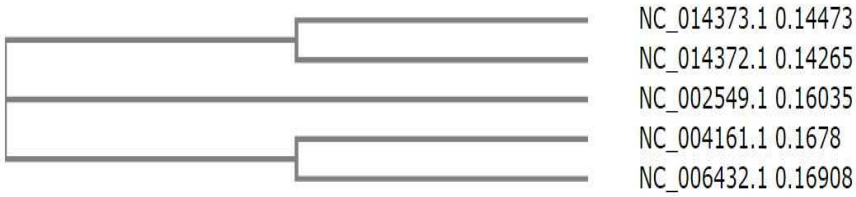

Figure: Phylogenetic tree for the genome sequences of five different species of *Ebolavirus*.

The analysis of different *Ebolavirus* genome showed that, the size of genome was identified to be 18.94 kb, 18.89 kb, 18.87 kb, 18.93 kb and 18.96 kb in *Bundibugyo ebolavirus, Reston ebolavirus*, *Sudan ebolavirus*, *Tai Forest ebolavirus* and *Zaire ebolavirus* respectively. The highest G+C (%) was found 42.3% in *Tai Forest ebolavirus* while lowest was found 40.6% in *Reston ebolavirus*. The highest no. of genes were found 9 both in *Reston ebolavirus* and *Zaire ebolavirus*, while lowest were found 8 in *Bundibugyo ebolavirus, Sudan ebolavirus* and *Tai Forest ebolavirus*. On the other hand, highest no. of proteins were found 9 in *Bundibugyo ebolavirus, Tai Forest ebolavirus* and *Zaire ebolavirus,* while lowest were found 8 in *Reston ebolavirus* and *Sudan ebolavirus* **(Table 1)**. Similarly previous reports suggest that, the genome size of Tick borne encephalitis virus, Kamati River virus and Tamana bat virus was 11.4 kb, 11.1 kb and 10.0 kb respectively, while the G+C (%) content was 50.29, 53.77 and 38.43 respectively [22]. In RNA GC bond have 3 hydrogen bonds while AU have only 2. As a result RNA with high G+C (%) content is much more stable to high temperature than RNA with low G+C (%) content. Moreover large scale systematic gene-centric association analysis revealed the correlation between G+C (%) of certain genomic region with temperature [23].

Table 1: Experimental data for complete genome of different species of *Ebolavirus* used in this experiment

Species	Host	BioProject accession No.	Chrs No.	Length (bases)	G+C (%)	No. of genes	No. of proteins
Bundibugyo ebolavirus	Vertebrates, Human	PRJNA51245	1	18940	42.0	8	9
Reston ebolavirus	Vertebrates, Human	PRJNA15006	1	18891	40.6	9	8
Sudan ebolavirus	Vertebrates, Human	PRJNA15012	1	18875	41.3	8	8
Tai Forest ebolavirus	Vertebrates, Human	PRJNA51257	-	18935	42.3	8	9
Zaire ebolavirus	Vertebrates, Human	PRJNA14703	1	18959	41.1	9	9

For analyzing putative promoter motifs, numerous programs have been available in the public domain along with different algorithms. But still further refinement is required before picking up for investigation in molecular biology laboratory. In this experiment, identification and analysis of putative promoters in the genome of different species of *Ebolavirus* namely *Bundibugyo ebolavirus, Reston ebolavirus*, *Sudan ebolavirus*, *Tai Forest ebolavirus* and *Zaire ebolavirus* was done using PROSCAN Version 1.7 program. Earlier reports suggest that, 72% of putative promoter motifs recognized by PROSCAN have a recognized TATA box and reported Transcription Start Site (TSS) is within +- 10 bases of the actual TSS. No relationship has been exist between the signal quality (score) and signal weight of a particular promoter. The relative weighting of a putative promoter have been determined by estimating the relative frequency with which the signal is found in promoter versus non-promoter sequences [20].

In this study, a total of 107 promoter sequences along with their designation, score, weight, specific location within genome, location of TATA box and estimated TSS were identified in the different *Ebolavirus* genome **(Table 2).** But not a single promoter was found common within these species. TFIID was found in *Bundibugyo ebolavirus, Reston ebolavirus*, *Sudan ebolavirus* and *Zaire ebolavirus*. Promoter silencer_binding_sit was found in both *Tai Forest ebolavirus* and *Zaire ebolavirus* while UCE.2 was common in *Bundibugyo ebolavirus* and *Reston ebolavirus*. Promoter motifs CTF, T-Ag, AP-2 and JCV_repeated_sequence were common in *Bundibugyo ebolavirus, Reston ebolavirus* and *Sudan ebolavirus*. Promoter motif Element_II_rs-4 was common in *Bundibugyo ebolavirus* and *Tai Forest ebolavirus*. On the other hand E2F and CREB were found in both *Sudan ebolavirus* and *Tai Forest ebolavirus*. Promoter motifs Albumin_US2, NF-kB, c-Myc and NRE_Box2_CS were only found in *Bundibugyo ebolavirus, Reston ebolavirus*, *Sudan ebolavirus* and *Tai Forest ebolavirus* respectively **(Figure 1-7)**. Within figures, the bright green lines indicate positive strand and red lines indicate negative strand. The yellow box within positive strand and blue box within negative strand indicates promoter names along with their respective sequences. The number below the box indicates the promoter location while the number above the box (within bracket) indicates their weight.

Table 2: Identification of putative promoter motifs in different species of *Ebolavirus* genome

Species of *Ebolavirus*	Promoters predicted on (strand)	Range of location (base)	Promoter score	TATA box found at (base)	Estimated Transcription Start Site (base)
Bundibugyo ebolavirus	Forward	5392-5642	69.00	5617	5647
		18061-18311	64.03	18285	18315
		18369-18619	54.20	18578	18608
	Reverse	18175-17925	60.42	-	-
		7864-7614	53.10	7651	7619
Reston ebolavirus	Forward	10424-10674	62.60	-	-
	Reverse	8623-8373	68.47	8389	8357
		7531-7281	55.80	-	-
		4011-3761	54.95	3780	3748
Sudan ebolavirus	Forward	7546-7796	75.65	-	-
		9174-9424	53.79	9405	9435
		10546-10796	59.14	10777	10807
		14416-14666	55.45	-	-
Tai Forest ebolavirus	Forward	6577-6827	90.15	-	-
Zaire ebolavirus	Reverse	13165-12915	56.28	12958	12926

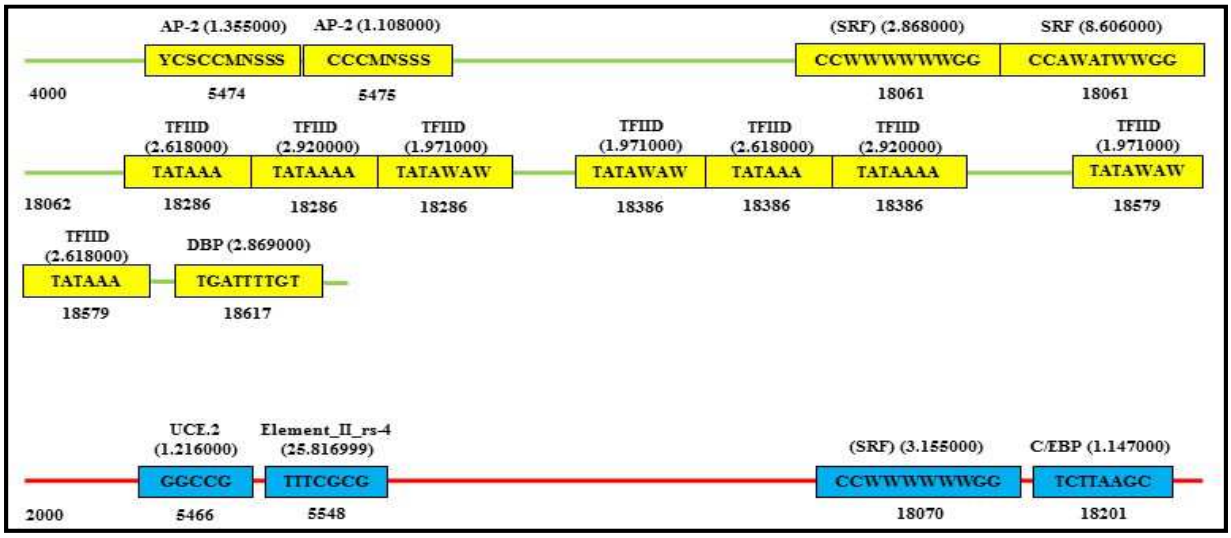

Figure 1: The promoter motifs for the forward strand of *Bundibugyo ebolavirus*.

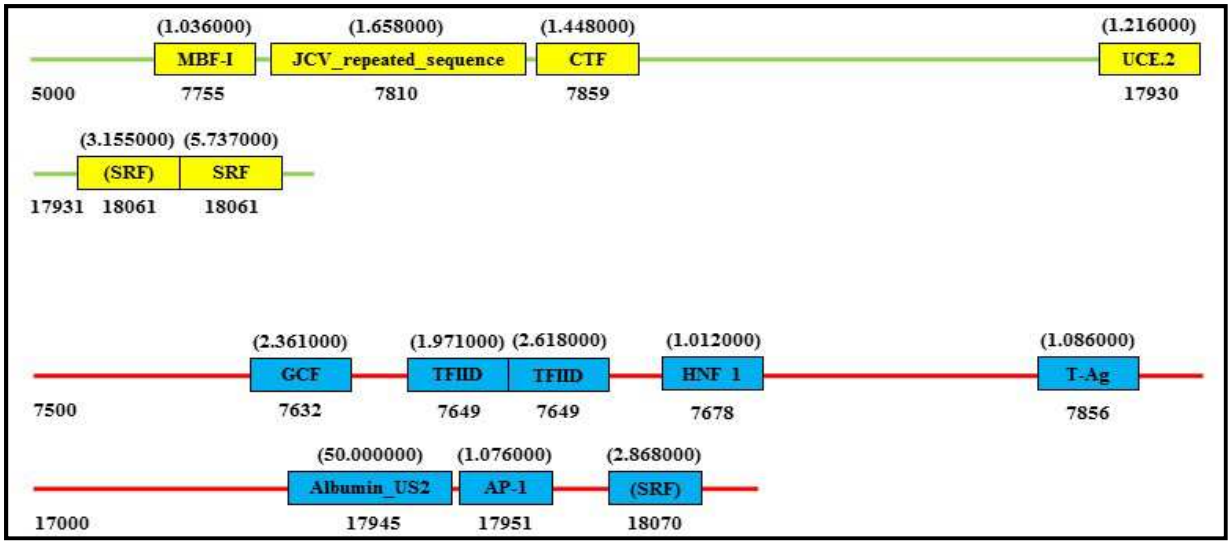

Figure 2: The promoter motifs for the reverse strand of *Bundibugyo ebolavirus*.

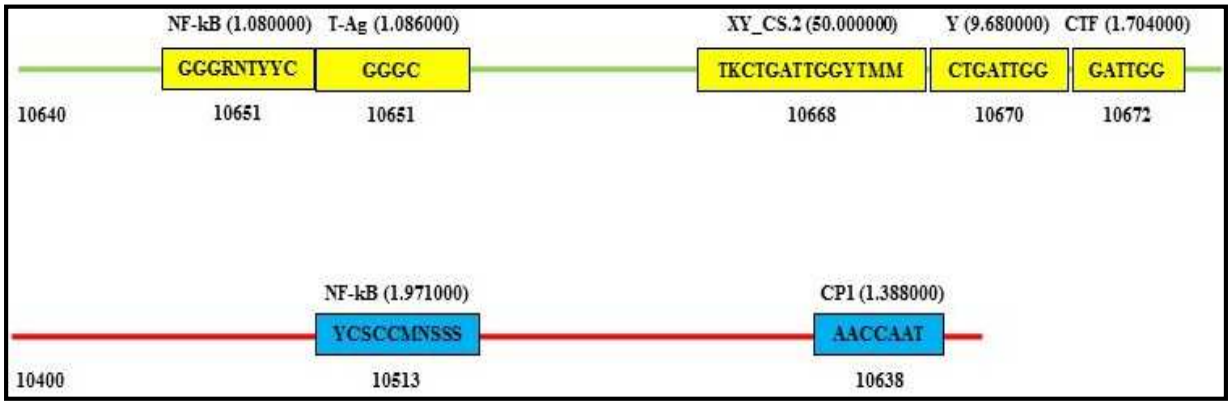

Figure 3: The promoter motifs for the forward strand of *Reston ebolavirus*.

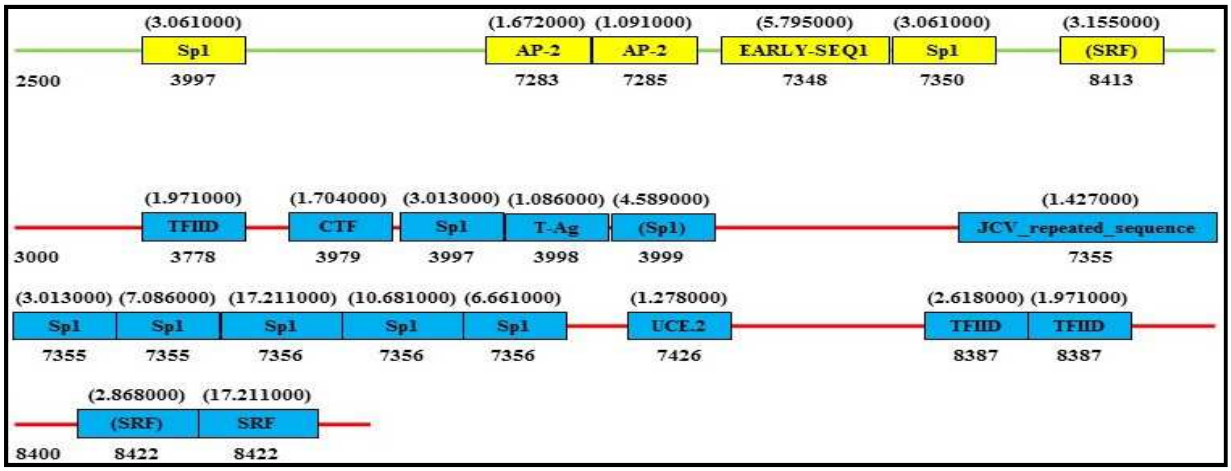

Figure 4: The promoter motifs for the reverse strand of *Reston ebolavirus*.

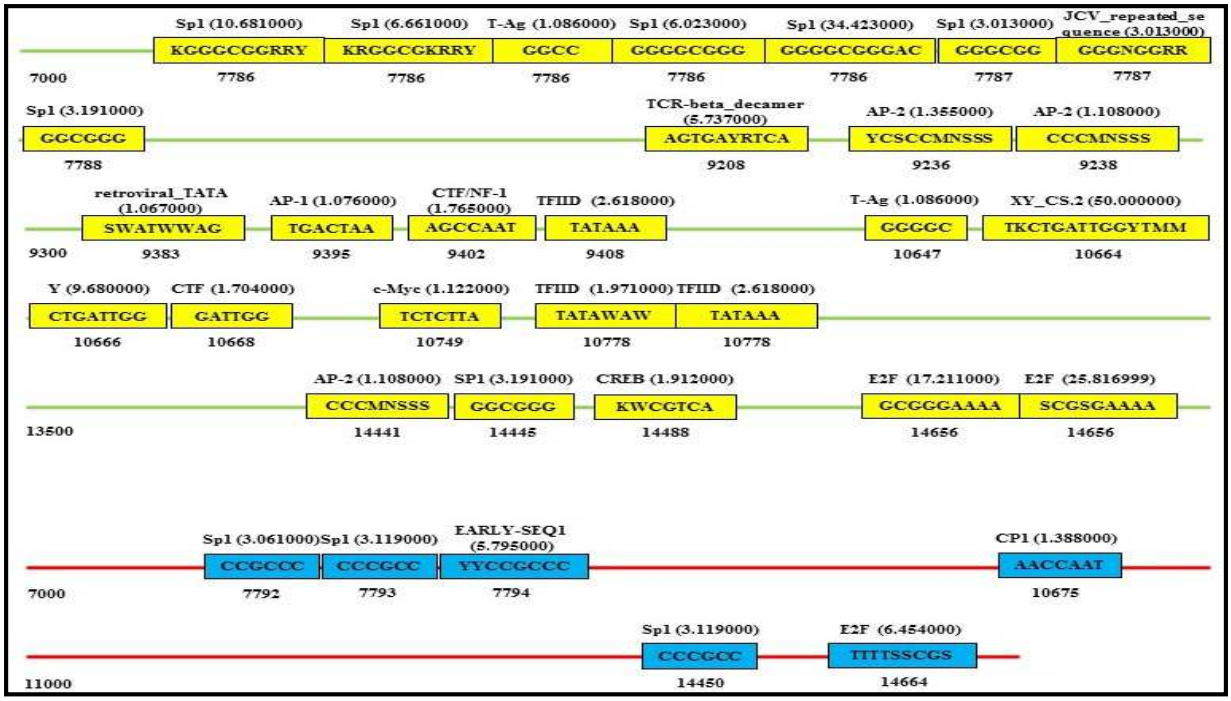

Figure 5: The promoter motifs for the forward strand of *Sudan ebolavirus*.

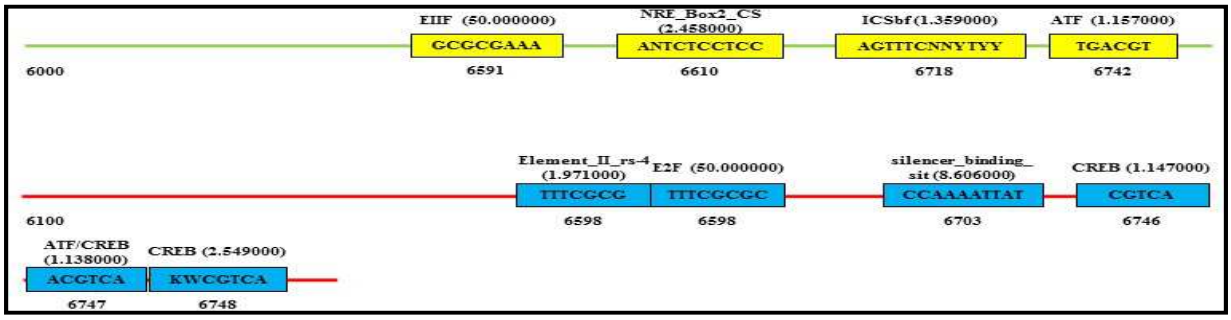

Figure 6: The promoter motifs for the forward strand of *Tai Forest ebolavirus*.

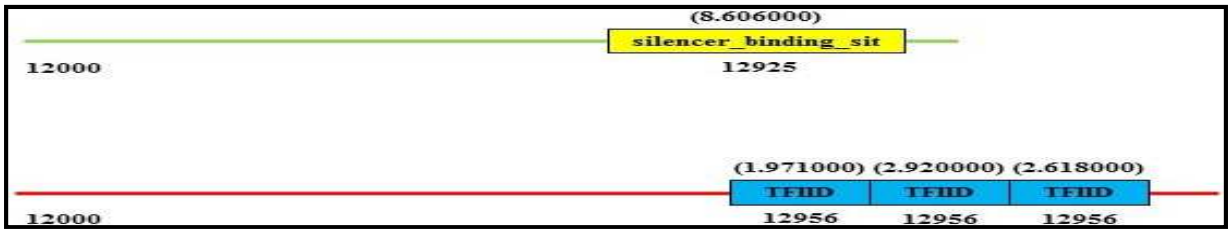

Figure 7: The promoter motifs for the reverse strand of *Zaire ebolavirus*.

A very limited resource has been available for the identification and characterization of promoter motifs in the viral genome. Transcription factor IID (TFIID) is one of the important prerequisite for RNA pol. II preinitiation complex [24]. Within core promoter of a particular gene TFIID binds up with TATA box and regulates the transcription initiation by RNA pol. II. SRF is a serum response element-binding transcription factor that plays an important role in cell cycle regulation, apoptosis, cell growth and differentiation. This is done by regulating the activity of many immediate early genes viz C-fos. T-Ag is a protein product of proto-oncogene and is involved in viral genome replication and regulation of host cell cycle [25]. The transcription factor Sp1 has multiple binding sites within major late promoter of adenovirus. The binding site

consists of -18GC rich sequences which helps Sp1 to stimulate gene expression, enhanced by adenovirus E1A protein [26].

JCV_repeated_sequence have been found within two 98 bp tandem repeats in human neurotropic papovirus and plays an important role in glial specific transcription of early and late stages of viral promoter sites [27]. AP1 (Activating Promoter 1) is a transcription factor that binds up with common DNA sites by Fos, ATF and Jun subunits. Different AP1 factors regulate different target genes and thus execute distinct biological functions [28]. CREB have been originally recognized as a transcription factor mediating transcriptional responses to cAMP and intracellular Ca^{2+} concentration via phosphorylation by PKA. It is also found that, the ATF/CREB family of transcription factors, i.e. ATF-1 and ATF-2 in conjugation with CREB is most crucial transcription factors involved in the initiation of bovine leukemia virus genome transcription [29].

E2F is a group of genes that encode another family of transcription factor which binds TTTCCCGC consensus binding site in the target promoter sequence. It has undoubtedly play important roles in cell cycle control, tumor development and metastasis in mouse mammary tumor virus (MMTV)-polyomavirus mouse model in breast cancer [30]. Several viruses including HIV and AIDS virus has a binding site for NF-kB transcription factor within their genome. This binding controls the expression of viral genes resulting viral replication, viral pathogenicity and activation of virus from latent or inactive state [31]. Transcription factor TCR-beta binds up with highly conserved decamer, homologous to cAMP response element (CRE). TCR-beta expression immediately increases the cAMP inducibility of CAT activity, mediated by V beta-CRE sequence [21].

CONCLUSION

In current study computational approaches have been made to find out the significant number of promoter motifs within the genome of five different species of *Ebolavirus*. A total of 107 promoters were determined out of which, 31 promoters were from *Bundibugyo ebolavirus,* 29 from *Reston ebolavirus*, 33 from *Sudan ebolavirus*, 10 from *Tai Forest ebolavirus* and 4 from *Zaire ebolavirus*. These promoters play vital role in regulating viral gene expression, thus *in silico* identification and analysis of putative promoter is important for understanding viral gene expression pattern, regulation network, cell specificity and development. The findings of this experiment may be the riding lamp for next level antiviral therapeutics by designing live attenuated vaccine candidate with the help of site directed mutagenesis in promoter regions. These results may also contribute to design expression vector and target specific delivery systems in gene therapy and thus combat against Ebola infection.

ACKNOWLEDGEMENTS

The author would like to appreciate and gratefully acknowledge all the members of Dept. of Biochemistry and Molecular Biology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh for their tremendous support.

REFERENCES

1. Sanchez A, Geisbert TW and Feldmann H. Filoviridae: Marburg and Ebola viruses. In: Knipe DM, Howley PM, eds. Fields virology. Philadelphia: Lippincott Williams & Wilkins, 2006: 1409–48.
2. Bausch DG, Towner JS, Dowell SF, Kaducu F, Lukwiya M, Sanchez A, et al. Assessment of the risk of Ebola virus transmission from bodily fluids and fomites. *J Infect Dis*. 2007; 196S: S142–7.

3. Kuhn JH, Becker S, Ebihara H, Geisbert TW, Johnson KM, Kawaoka Y, Lipkin WI, Negredo AI, Netesov SV, Nichol ST, Palacios G, Peters CJ, Tenorio A, Volchkov VE and Jahrling PB. Proposal for a revised taxonomy of the family Filoviridae: Classification, names of taxa and viruses, and virus abbreviations. *Archives of Virology*. 2010; 155 (12): 2083–2103.
4. Suzuki Y and Gojobori T. The origin and evolution of Ebola and Marburg viruses. *Mol Biol Evol*. 1997; 14(8):800–6.
5. Olival KJ, Islam A, Yu M, Anthony SJ, Epstein JH, Khan SA, Khan SU, Crameri G, Wang LF, Lipkin WI, Luby SP and Daszak P. Ebola Virus Antibodies in Fruit Bats, Bangladesh. *Emerg Infect Dis*. 2013; 19(2): 270–273.
6. World Health Organization Ebola response roadmap Situation reports: 1 October 2014. http://apps.who.int/iris/bitstream/10665/144117/1/roadmapsitrep_21Nov2014_eng.pdf?ua=1 (Accessed 22/11/2014).
7. Kiley MP, Bowen ET, Eddy GA et al. Filoviridae: a taxonomic home for Marburg and Ebola viruses? *Intervirology*. 1982; 18: 24–32.
8. Hoenen T, Groseth A, Falzarano D and Feldmann H. Ebola virus: unraveling pathogenesis to combat a deadly disease. *Trends Mol Med*. 2006, 12: 206–215.
9. Feldmann H, Volchkov VE, Volchkova VA and Klenk HD. The glycoproteins of Marburg and Ebola virus and their potential roles in pathogenesis. *Arch Virol Suppl*. 1999; 15: 159–169.
10. Volchkov VE, Volchkova VA, Slenczka W and Klenk HD, Feldmann H. Release of viral glycoproteins during Ebola virus infection. *Virology*. 1998; 245: 110–119.
11. Cook JD and Lee JE. The secret life of viral entry glycoproteins: moonlighting in immune evasion. *PLoS Pathog*. 2013; 9: e1003258.
12. Takada A and Kawaoka Y. The pathogenesis of Ebola hemorrhagic fever. *Trends Microbiol*. 2001; 9: 506–511.
13. Escudero-Pérez B, Volchkova VA, Dolnik O, Lawrence P and Volchkov VE. Shed GP of Ebola Virus Triggers Immune Activation and Increased Vascular Permeability. *PLoS Pathog*. 2014; 10: e1004509.
14. Muhlberger E, Weik M, Volchkov VE, Klenk HD and Becker S. Comparison of the transcription and replication strategies of Marburg virus and Ebola virus by using artificial replication systems. *J Virol*. 1999; 73: 2333–42.
15. Kühl A and Pöhlmann S. How Ebola virus counters the interferon system. *Zoonoses Public Health*. 2012; 59(Suppl 2):116–131.
16. Martínez MJ, Biedenkopf N, Volchkova V, Hartlieb B, Alazard-Dany N, Reynard O, Becker S and Volchkov V. Role of Ebola virus VP30 in transcription reinitiation. *J Virol*. 2008; 82: 12569–12573.
17. Hoenen T, Jung S, Herwig A, Groseth A and Becker S. Both matrix proteins of Ebola virus contribute to the regulation of viral genome replication and transcription. *Virology*. 2010; 403: 56–66.
18. Saha A, Mazumder N, Alam S, Farook KB, Tamanna NAT, Rahman R, Ali Y, HossaiN MA and Mannan A. Prediction and Analysis of Putative Promoter in Fowl Adenovirus Genome by *in silico* Approach. *Int J Biol Med Res*. 2013; 4(3): 3432-3436.
19. Shawan MMAK, Hossain MM, Hasan MA, Hasan MM, Parvin A, Akter S, Uddin KR, Banik S, Morshed M, Rahman MN and Rahman SMB. Design and Prediction of Potential RNAi (siRNA) Molecules for 3'UTR PTGS of Different Strains of Zika Virus: A Computational Approach. *Nat. Sci*. 2015; 13(2): 37-50.

20. Prestridge DS. Predicting Pol II Promoter Sequences Using Transcription Factor Binding Sites*. J. Mol. Biol*. 1995; 249: 923-32.
21. Kurjogi MM, Sanakal RD, Kaliwal BB. Identification and analysis of putative promoter motifs in bovine herpes virus. *Bioinformation*. 2012; 8(23): 1167-70.
22. Somvanshi P, Singh V and Seth PK. Identification and analysis of putative promoter motifs in *Flavivirus* genome. *Bioinformation*. 2008; 3(4): 162–167.
23. Zheng H and Wu H. Gene-centric association analysis for the correlation between the guanine-cytosine content levels and temperature range conditions of prokaryotic species. *BMC Bioinformatics*. 2010; 11: S7.
24. Lewin Benjamin. Genes VIII. 2004; pp. 636-637. ISBN 0-13-144946-X.
25. Ali SH and DeCaprio JA. Cellular transformation by SV40 large T antigen: interaction with host proteins. *Semin Cancer Biol*. 200; 11(1): 15-23.
26. Carthew RW, Chodosh LA and Sharp PA. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell*. 1985; 43(2 Pt 1): 439-48.
27. Hiroomi T. Regulation of JCV promoter function: Evidence that a pentanucleotide "silencer" repeat sequence AGGGAAGGGA down-regulates transcription of the JC virus late promoter. *Virology*. 1991; 180: 327.
28. Hess J., Angel P. and Schorpp-Kistner M. AP-1 subunits: quarrel and harmony among siblings. *J Cell Sci*. 2004; 117(Pt 25): 5965-73.
29. Adam E, Kerkhofs P, Mammerickx M, Burny A, Kettmann R and Willems L. The CREB, ATF-1, and ATF-2 transcription factors from bovine leukemia virus-infected B lymphocytes activate viral expression. *J Virol*. 1996; 70(3): 1990-9.
30. Hollern DP, Honeysett J, Cardiff RD and Andrechek ER. The E2F transcription factors regulate tumor development and metastasis in a mouse model of metastatic breast cancer. *Mol Cell Biol*. 2014; 34(17): 3229-43.
31. Hiscott J, Kwon H and Genin P. Hostile takeovers: viral appropriation of the NF-kB pathway. *J. Cli. Invest*. 2001; 107(2): 143-51.