International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume 6, Issue 4, July- 2018
DOI: 10.21276/ijircst.2018.6.4.6

Promoters of Dengue Virus Complete Genome

Md. Zakir Hossain, Milton K. Debnath, Ishita R. Das, Anita L. Jackson, Rozina Akter

***Abstract*—Present concerning facts of Dengue virus has proved this as a worldwide emerging threat, which causes Dengue fever.This is an illness not sufficiently covered in medical curriculum and there is no universally acceptable significant diagnostics, vaccines or treatments of Dengue fever. It has become crucial to develop a novel, affordable diagnostic tools, therapeutics and vaccines based on the available clinical experience Dengue complete genome. The promoter is one of the regions of DNA sequence where RNA polymerase first attaches during the initiation of transcription. For gene expression, the promoter acts as an important site. There are severe outbreaks of Dengue in Bangladesh and clearly there is a lack of sufficient anti-viral drug and vaccine against this virus. In this research,we have analyzed of all verified Dengue genome sequences of Bangladesh, available in National Center of Biotechnology Information (NCBI), USA and other authenticdatabases, to look for potential promoters for vaccine and therapeutic applications.**

***Keywords*-Dengue fever, dengue virus (DenV), promoter, bioinformatics, NCBI, Bangladesh**

I. INTRODUCTION

Recently the incidence of dengue has dramatically grown around the world [1]. The World Health organization estimates that about 100 million new infections occur annually more than 100 endemic countries and around 5,00,000 peoples are hospitalized with dengue hemorrhagic fever of which many of the infected are children [1-3].

Manuscript received June 19, 2018

Md. Zakir Hossain, Department of Biomedical Engineering, JUST Medical Center, Jessore University of Science and Technology, Jessore, Bangladesh. (email: mzh.bme@gmail.com)

Milton K. Debnath, Department of Biomedical Engineering, JUST Medical Center, Jessore University of Science and Technology, Jessore, Bangladesh

Ishita R. Das, Department of Biomedical Engineering, JUST Medical Center, Jessore University of Science and Technology, Jessore, Bangladesh

Anita L. Jackson, Department of Pharmacotherapy and Experimental Therapeutics, UNC Eshelman School of Pharmacy, University of North Carolina (UNC), Chapel Hill, NC-27599, USA

Rozina Akte, Department of Biomedical Engineering, JUST Medical Center, Jessore University of Science and Technology, Jessore, Bangladesh

Dengue is an arthropod-born viral disease. Aedes aegypti female mosquitoes are mainly responsible for transmission of dengue virus [1]. It is a member of Flaviviridae family and genus flavivirus with enveloped by lipoprotein and contains single-stranded positive sense RNA virus with four different serotypes, where all four serotypes are responsible for dengue [2, 4]. The dengue genome is consisting of about 11,000 bases of positive-sense single stranded RNA (ssRNA) packed by 40-60 nm virion which comprising with spherical nucleoside core covered by an icosahedral envelope. DenV has three structural proteins (capsid protein C, membrane protein M, envelope protein E) and seven nonstructural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5). Nonstructural proteins NS3 and NS2B are responsible for auto-processing of the virus, mediation of viral genome replication, and virus packaging [5]. Classically dengue caused by fever, headache, myalgia, arthralgia, eye pain, rash but in severe cases dengue virus manifests as defects in plasma leakage and hemostasis which characterized by widely recognized dengue hemorrhagic fever [4]. Currentlyespecially in developing countries, dengue virus (DenV) causes dengue fever. Thus, in this research, we have identified the putative promoters of Dengue complete genome of Bangladesh, a developing country in South-east Asia, using existing and customized Bioinformatics tools for vaccine and therapeutic applications.

II. METHODOLOGY

A. Retrieval of complete Dengue virus genome sequences

The retrieval of complete DENV nucleotide sequences were obtained from the databases of National Centre for Biotechnology Information (NCBI).

B. Putative Promoter determination based on transcription start site

The putative promoters in the Dengue virus complete genome of DENV species were identified using the PROMOTER SCAN (PSCan) version 1.7 suite of programs. The complete genome sequences of Dengue virus type 3 isolate PhMH-J1-97, complete genome, 10,707 bp linear RNA, Accession: AY496879.2, GI: 89242735 was the only complete genome that was showing the putative promoter using the PSCan program[6].

C. Analysis of complete DENV genomes G - C content

The Dengue virus type 3 isolate PhMH-J1-97, complete genome-size of DENV species was analyzed in FASTA format which were extracted from Genbank database. The GC content was compared and the complete Dengue virus

type 3 isolate PhMH-J1-97, complete genome was analyzed.

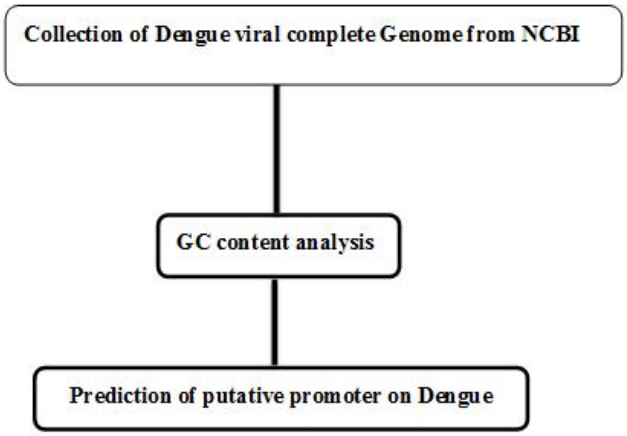

Fig 1: Flowchart of prediction of putative promoter on Dengue virus complete genome.

II. RESULTS AND DISCUSSIONS

Using the search "Dengue complete genome Bangladesh" at NCBI databases, we have found nine complete Dengue genome nucleotide sequences (Table 1). The putative promoters in the Dengue virus complete genome were identified using the PROMOTER SCAN (PSCan) program [7]. The complete genome sequences ofDengue virus type 3 isolate PhMH-J1-97, complete genome, 10,707 bp linear RNA, Accession: AY496879.2, GI: 89242735 was the only complete genome that was showing the putative promoter.

Table 1: Nine complete Dengue genome nucleotide sequences

Dengue virus type 3 complete genome	Accession	GI	10,707 bp linear RNA
Isolate mutant BDH02-01	DQ401689.1	89274855	No promoter found
Isolate mutant BDH02-03	DQ401691.1	89274859	
Isolate mutant BDH02-04	DQ401692.1	89274861	
Isolate mutant BDH02-07	DQ401693.1	89274863	
Isolate BDH02-1	AY496871.2	89242727	
Isolate BDH02-4	AY496874.2	89242731	
Isolate BDH02-7	AY496877.2	89242733	
Isolate BDH02-3	AY496873.2	89242729	
Isolate PhMH-J1-97	AY496879.2	89242735	Promoter

III. ANALYSIS OF G-C CONTENT

The GC base pairs are more stable than AU base pairs, due to having three hydrogen bonds and AU on the contrary have two. This makes high GC-content RNA structures more tolerant to high temperatures. The genomic comparisons of Dengue virus type 3 isolate PhMH-J1-97, complete genome G-C % of the 46.9 % [7].

Table 2: Genomic comparisons of Dengue virus type 3

Total count, all bases in Dengue virus type 3 isolate PhMH-J1-97, complete genome	10707	% G ~ C content
Adenine (A) count:	3441	
Thymine (T) count:	2244	
Guanine (G) count:	2772	46.9
Cytosine (C) count:	2250	

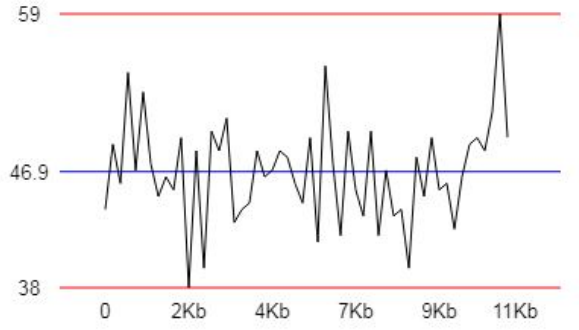

Fig 2: Analysis of G-C content dengue virus type 3 isolate PhMH-J1-97, complete genome.

IV. PREDICTION OF PUTATIVE PROMOTER ON DENGUE VIRAL GENOME

Analysis of dengue virus isolates Dengue virus type 3 isolate PhMH-J1-97, 10,707 bp linear RNA, Accession: AY496879.2, GI: 89242735complete genome using Proscan, Version 1.7[7]. The promoter region predicted on forward strand in 1042 to 1292 using the promoter Score: 138.54 (Promoter Cutoff = 53.000000). Thus, the significant signals of Dengue virus type 3 isolate PhMH-J1-97, complete genome, positive strand are depicted.

Table 3(a): Complete genome positive strand

Name	TFD	# Strand	Location	Weight
E4F1	S01249	+	1134	3.764000
APRT-CHO_US	S00215	+	1155	1.860000
AP-2	S01936	+	1156	1.108000
EIIF	S00659	+	1281	50.000000

Table 3(b): Complete genome positive strand

TFD Site ID	Factor(s)	Site	System	Sequence
S01249	E4F1	E4F1 CS	MAMM	ACGTMAC
S00215	Unknown	APRT-CHO US	MAMM	GCCCCACC
S01936	AP-2	AP-2 CS6	MAMM	CCCMNSSS
S00659	EIIF	EIIF-EIIaE1	MAMM	GCGCGAAA

ISSN: 2347-5552, Volume 6, Issue 4, July- 2018
DOI: 10.21276/ijircst.2018.6.4.6

Fig 3: Schematic representation of prediction of putative promoter on Dengue viral genome positive strand [Dengue virus type 3 isolate PhMH-J1-97, complete genome].

The significant signals of Dengue virus type 3 isolate PhMH-J1-97, complete genome negative strand are as follows:

Table 5(a): Complete genome negative strand

Name	TFD	# Strand	Location	Weight
myosin-specific	S00608	-	1098	1.115000
EivF	S00399	-	1140	3.227000
EivF/CREB	S00104	-	1140	1.564000
PuF	S02016	-	1163	1.391000
JCV_repeated_seq	S01193	-	1163	1.658000
AP-2	S01936	-	1239	1.091000
E2F	S00147	-	1288	50.000000
element_II_rs-4	S01507	-	1288	25.816999

Table 5(b): Complete genome negative strand

TFD Site ID	Factor(s)	Site	System	Sequence
S00608	**Myosin-specific**	a-actin US	AVIAN/ MAMM	GTCGCC
S00399	**EivF**	EivF CS	MAMM	GTKACGT
S00104	**EivF/CREB**	EivF/CRE CS	MAMM	GTKACGW
S02016	**PuF**	PuF RS	MAMM	GGGTGGG
S01193	**unknown**	JCV repeated sequence	MAMM	GGGNGGRR
S01936	**AP-2**	AP-2 CS6	MAMM	CCCMNSSS
S00147	**E2F**	E2F CS.1	MAMM	TTTCGCGC
S01507	Unknown	element II rs-4	MAMM	TTTCGCG

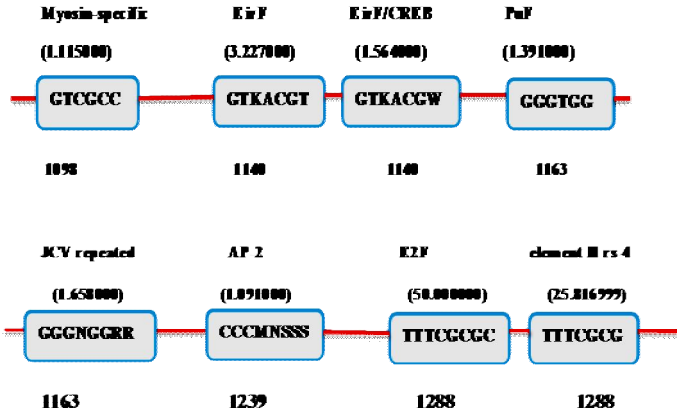

Fig 4: Schematic representation of prediction of putative promoter on Dengue viral genome, negative strand [Dengue virus type 3 isolate PhMH-J1-97, complete genome].

The significant signals of Dengue virus type 3 isolate PhMH-J1-97, complete genome positive and negative strands are described with their associated functions. Myosin specific genes encode for a family of protein called myosins [8]. Myosin is a hexameric protein that composed of two pairs of non-identical light chains and a pair of myosin heavy chains (MYH). The genes that code for myosin are member of MYH family. It has contractile properties. Chemical energy is converted into mechanical energy through the hydrolysis of adenosine tri phosphate (ATP) [9]. It causes motility by binding with another protein named “actin” as like as lock and key [8-9]. Almost all eukaryotic cells comprise myosin isoforms. Certain isoforms perform specialized functions in specific cell types, while other isoforms are universal [8]. These genes encode a protein with a myosin head like domain and an IQ domain [10]. The other genes that translate myosin are MYH3, MYO5A, MYO5B, MYO7A MYH6, MYH7, MYH9, MYH11 [11].

E4F1 (E4F Transcription Factor 1) is a protein coding gene. P53 is a related pathway of this gene. Gene Ontology (GO) represents that this gene involves in nucleic acid binding and ligase activity [12]. E4F1 gene uniquely expresses zinc finger protein. Zinc finger is a tiny protein. In order to stabilize the fold, this protein is coupled with one or more zinc ions (Zn2+) [12-14]. DNA binding activity is regulated by the action of E1A. E1A induces phosphorylation and regulates the 50-kDa amino-terminal created by proteolytic cleavage. In the absence of E1A, full length gene transcription by the E4 promoter is blocked where the 50-kDa product performs as a transcriptional activator [12].

An early gene iv factor (EivF) binds to an element that involves in cAMP induction and EIa-dependent activation during transcription. Using a functional transcription assay, the factor EivF was purified fromHeLa cells and measured 72,000 to 65,000 dalton. Transcriptionally active and purified EivF also binds with DNA sequence promoter and shows responsiveness to cAMP [15].Transcription factor cAMP response element binding protein (CREB) has ability to bind with certain DNA sequences called cyclic adenosine mono phosphate response elements (cAMP-RE) and regulate gene expression [16-17]. In 1987, CREB was narrated as a transcription factor that was cAMP responsive and controlled the somatostatin gene [18]. CREB is intimately related in structure and function with ATF-1 (acting transcription factor-1) and CREM (cAMP response element modulator) proteins. CREB are evolved with many animals as well as human [19]. This protein has a role in long term memory establishment and neuronal plasticity in the brain that seems to be effective in developing spatial memory [19].

Adenine phosphoribosyl transferase (APRT) locus arises spontaneously at high frequencies in CHO (Chinese hamster Ovary) cells. The functional APRT gene is identified on Z7 chromosome and Z4 APRT allele is missed or deactivated [20]. APRT gene acts like microRNA (miRNA) and stands for limiting the transcription process of RNA [21-23]. Activating enhancer binding protein 2 alpha (AP-2 alpha) is a transcription

factor that is also known as TFAP2A. This protein is encoded in humans by the TFAP2A gene [24]. The protein performs as a transcription factor that attaches with the DNA at a specific sequence and initiates transcription cascades. Its attaching spot is GC-rich cis-regulatory gene regions of some cells and virus [25].

The founding member of PuF protein family is Pumilio protein which attaches with RNA. It consists of eight repeated PuF domains that plays significant role in post-embryogenesis and embryogenesis period of eukaryotes [26]. This protein is also present in other species. Distinct features and numbers of the PuF domains in each species play certain roles in each species [26]. PuF proteins attach with the 3' un-translated region (3'UTR) of specific mRNAs and block the translation process. This results in unequal cell division and thereby modulating cell fate specification [27-28].JCV repeated sequence promoter may influences early viral gene expression got from the cells of human brain tissue. Cellular transcription factors and non-coding viral regulatory domain restricts JCV gene expression [29].

Cell specific tropism has been identified in JCV for the large tumor antigen (T antigen) that is encoded by the transcription of early viral gene [30]. The mechanism by which nuclear oncogene EIA activates transcription of several viral and host promoters is a significant matter in the conduction of eukaryotic gene expression and virus host cell interactions. Identification of similar host transcription factors and cis acting portion of the promoters that target for EIA action is important for the EIA-mediated control of coordinately regulated genes. Strong transactivation of the EII early promoter by several EIA sensitive portions suggests that the EII early promoter is stringently regulated in virus-infected cells [31].E2F is a group of genes. In higher eukaryotes, E2F encode a family of proteins that function as transcription factors. Among all the genes, three are activators and six are suppressors. The suppressors are named as E2F3b, E2F4-8 and activators are named as E2F1, 2, E2F3a [32]. Small DNA tumor virus produces some transforming proteins which aim at the E2F family proteins. E2F family proteins also help in the action of tumor suppressor proteins and regulation of cell cycle [33]. E2F transcription factors are perhaps well known for their function in cell cycle advancement, cellular proliferation, DNA synthesis and transition into S phase [34].

V. CONCLUSION

In this research, for the first time ever reported, we have identified the putative promoters of complete dengue genomes using bioinformatics tools. We also analyzed and discussed putative promoters' motifs of complete dengue genomes of Bangladesh. These results can be useful for species specific dengue viral diagnostics, treatment such as gene therapy and future vaccines against dengue virus.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with animals performed by any of the authors.

ACKNOWLEDGEMENT

The author has acknowledged the contribution of the University Grants Commission (UGC) of Bangladesh for funding the research work.

REFERENCES

[1] Anne TuiskunenBäck,ÅkeLundkvist. Dengue viruses – an overview; 2013; v3; PMC3759171. Doi10.3402/iee.v3i0.19839.

[2] David W. Vaughn Sharone Green SiripenKalayanarooj ; Dengue Viremia Titer, Antibody Response Pattern, and Virus Serotype Correlate with Disease Severity ; The Journal of Infectious Diseases, Volume 181, Issue 1, 1 January 2000, Pages 2–9,https://doi.org/10.1086/315215.

[3] Yen-Lung Lo,#1,3,4 Gunn-GuangLiou; Dengue Virus Infection Is through a Cooperative Interaction between a Mannose Receptor and CLEC5A on Macrophage as a Multivalent Hetero-Complex; J PLoS One; 11(11); 2016; PMCID: PMC5104462; doi: 10.1371/journal.pone.0166474.

[4] Dengue and severe dengue; World Health Organization; April 2017; http://www.who.int/mediacentre/factsheets/fs117/en/

[5] Global Strategy for dengue prevention and control, 2012–2020; WHO/Department of Control of Neglected Tropical Diseases; vi,43p; August2012;ISBN: 9789241504034; WHO reference number: HO/HTM/NTD/VEM/2012.

[6] Carlo Amorin Daep,1,2 Jorge L. Muñoz-Jordán,3 and Eliseo Alberto Eugenin1,2,*; Flaviviruses, an expanding threat in public health: focus on Dengue, West Nile, and Japanese encephalitis virus ; J Neurovirol. 2014 Dec; 20(6): 539–560. PMCID: PMC4331079;NIHMSID: NIHMS633656; . doi: 10.1007/s13365-014-0285-z.

[7] Md. Zakir Hossain, Rozina Akter; Putative Promoters of Two Complete Zika Virus Genome; IJBBE 2016; 2(5), 59-64; files.aiscience.org/journal/article/pdf/70070109.pdf

[8] Dantzig JA, Liu TY, Goldman YE. Functional studies of individual myosin molecules. Ann N Y Acad Sci. 2006 Oct;1080:1-18. Review. PubMed: 17132771.

[9] Lodish H, Berk A, Zipursky SL, et al. Myosin: The Actin Motor Protein. J Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Bookshelf ID: NBK21724

[10] Toydemir RM, Bamshad MJ. Sheldon-Hall syndrome. Orphanet J Rare Dis. 2009 Mar 23;4:11. doi: 10.1186/1750-1172-4-11. PMID: 19309503 PMCID: PMC2663550

[11] Syamaladevi DP, Spudich JA, Sowdhamini R. Structural and functional insights on the Myosin superfamily. Bioinform Biol Insights. 2012;6:11-21. doi: 10.4137/BBI.S8451. Epub 2012 Feb 1. PubMed: 22399849. Free full-text available from PubMed Central: PMC3290112.

[12] https://www.genecards.org/cgi-bin/carddisp.pl?gene=E4F1 (accessed on 21.05.2018)

[13] Hanas JS, Hazuda DJ, Bogenhagen DF, Wu FY, Wu CW (December 1983). "Xenopus transcription factor A requires zinc for binding to the 5 S RNA gene". The Journal of Biological Chemistry. 258 (23): 14120–5. PMID 6196359.

[14] Berg JM (April 1990). "Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules". The Journal of Biological Chemistry. 265 (12): 6513–6. PMID 2108957.

[15] Cortes P, Buckbinder L, Leza MA, Rak N, Hearing P, Merino A, Reinberg D. EivF, a factor required for transcription of the adenovirus EIV promoter, binds to an element involved in EIa-dependent activation and cAMP induction. Genes Dev. 1988 Aug;2(8):975-90. PMID: 2844626

[16] Bourtchuladze; et al. "Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein". Cell. 79 (1): 59–68. doi:10.1016/0092-8674(94)90400-6. PMID 7923378.

[17] Purves, Dale; George J. Augustine; David Fitzpatrick; William C. Hall; Anthony-Samuel LaMantia; James O. McNamara & Leonard E. White (2008). Neuroscience (4th ed.). Sinauer Associates. pp. 170–6. ISBN 978-0-87893-697-7.

[18] Montminy MR, Bilezikjian LM. "Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene." J Nature. 1987 Jul 9-15;328(6126):175-8.

[19] Silva; et al. "CREB and Memory" (PDF). Annual Review of Neuroscience. 21: 127–148. doi:10.1146/annurev.neuro.21.1.127

[20] Adair GM, Siciliano MJ, Brotherman KA, Nairn RS. "Preferential loss or inactivation of chromosome Z4 APRT allele in CHO cells." J Somat Cell Mol Genet. 1989 Jul;15(4):271-7. PMID: 2762931

[21] Ambros V (September 2004). "The functions of animal microRNAs". J Nature; 431 (7006): 3505. doi:10.1038/nature02871. PMID 15372042

[22] Bartel DP (January 2004). "MicroRNAs: genomics, biogenesis, mechanism, and function". Cell. 116 (2): 281–97. doi:10.1016/S0092-8674(04)00045-5. PMID 14744438

[23] "Extracellular/Circulating MicroRNAs: Release Mechanisms, Functions and Challenges". Achievements in the Life Sciences. 10: 175 – 186. doi:10.1016/j.als.2016.11.007

[24] https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowD etailView&TermToSearch=7020 (accessed on 23.05.2018)

[25] Williams T, Tjian R (Apr 1991). "Analysis of the DNA-binding and activation properties of the human transcription factor AP-2". Genes & Development. 5 (4): 670 - 82. doi:10.1101/gad.5.4.670.

[26] Zhang CD, Pan MH, Tan J, Li FF, Zhang J, Wang TT, Lu C. Characteristics and evolution of the PUF gene family in Bombyx mori and 27 other species. J Mol Biol Rep. 2012 Jan;39(1):675-83. doi: 10.1007/s11033-011-0785-3. Epub 2011 May 20. PMID: 21598114. DOI: 10.1007/s11033-011-0785-3

[27] Spassov DS, Jurecic R. The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? J IUBMB Life. 2003 Jul;55(7):359-66. PMID: 14584586. DOI: 10.1080/1521654031000160309 3

[28] Tara Quenault, Trevor Lithgow, Ana Traven.PUF proteins: repression, activation and mrna localization. J Cell. Volume 21, Issue 2, p104–112,February2011. DOI: https://doi.org/10.1016/j.tcb.2010.09.013

[29] Marshall LJ, Dunham L, Major EO. Transcription factor Spi-B binds unique sequences present in the tandem repeat promoter/enhancer of JC virus and supports viral activity. J Gen Virol. 2010 Dec;91(Pt 12):3042-52. doi: 10.1099/vir.0.023184-0. Epub 2010 Sep 8. PMID: 20826618. PMCID: PMC3052566. DOI: 10.1099/vir.0.023184-0

[30] Chang CF, Tada H, Khalili K. The role of a pentanucleotide repeat sequence, AGGGAAGGGA, in the regulation of JC virus DNA replication. J Gene. 1994 Oct 21;148(2):309-14. PMID: 7958960.

[31] C F Manohar, J.Kratochvil and B Thimmapaya. The Adenovirus Eii Early Promoter Has Multiple Eia-Sensitive Elements, Two Of Which Function Cooperatively In Basal And Virus-Induced Transcription. J. Virol. June 1990 vol. 64 no. 62457-2466

[32] Zheng N, Fraenkel E, Pabo CO, Pavletich NP (1999). "Structural Basis of DNA Recognition By The Heterodimeric Cell Cycle Transcription Factor E2F-DP". Genes Dev. 13 (6): 666-74. doi:10.1101/gad.13.6.666. PMC 316551 . PMID 10090723.

[33]https://www.ncbi.nlm.nih.gov/gene/1869(accessed on 24.05.2018)

[34]Chen HZ, Tsai SY, Leone G. Emerging roles of E2Fs in Cancer: an exit from cell cycle control. J Nat Rev Cancer. 2009 Nov;9(11):785-97. DOI: 10.1038/nrc2696. PMID:19851314. PMCID: PMC3616489.

