INTERNATIONAL JOURNAL OF APPLIED BIOLOGY AND PHARMACEUTICAL TECHNOLOGY

www.ijabpt.com Volume-3, Issue-4, Oct-Dec-2012 Coden : IJABPT Copyrights@2012

ISSN : 0976-4550

Received: 12th August -2012 *Revised: 15th August-2012* *Accepted: 05th Oct-2012*

Research article

COMPARATIVE STRUCTURAL AND FUNCTIONAL ANALYSIS OF E6 ONCOGENE OF HUMAN PAPILLOMAVIRUS TYPE 16

Dharmendra Kumar Chaudhary[1†], Indra Mani[1,2†], Vijai Singh[3]*

[1.] National Bureau of Fish Genetic Resources, Canal Ring Road, P.O. Dilkusha, Lucknow-226002, India
[2.] Department of Biochemistry, Faculty of Science, Banaras Hindu University, Varanasi -221005, India
[3] Programme d'Épigénomique, Institute of Systems and Synthetic Biology, Universite d'Evry Val d'Essonne-Genopole-CNRS, 5 rue Henri Desbruères, 91030 Evry, France

Corresponding author: *E-mail: vijaisingh15@gmail.com*

ABSTRACT: Human papillomavirus (HPV) is an etiologic agent of the uterine cervix cancer and several other neoplasias in women globally. E6 protein of HPV type 16 is highly conserved and plays the key role in an inducing cancer via suppressing activity of P^{53}. We have used different bioinformatics tools for generation of phylogenetic tree, modeling of RNA secondary structure, gene designing and codon optimization of HPV E6 gene. The size of E6 gene sequences of nine strains HPV type 16 was estimated to be 456 to 477 bp and G+C content was ranged between 37.06 to 37.94%. We used E6 gene sequences for construction of phylogenetic relationship and these divided into five groups. RNA secondary structures of E6 gene were modeled and analyzed that folding free energy of wild genes was -093.96, -106.21,-040.48, -090.76, -072.68, -092.86, -039.32, -044.78, -047.88 kcal/mol and after codon optimization free energy was -122.70, -107.40, -104.80, -121.40, -127.40, -110.80, -105.20, -122.30, -110.40 kcal/mol respectively. Moreover, gene designing and codon optimization have used to improve the heterologous expression in living organisms by increasing translational efficiency. All strains of HPV16 were used for codon optimization in *E. coli.* Codon adaptation index (CAI) and G+C contents of E6 gene in optimized DNA were enhanced by 3.6 (72.7%) and 1.3 (25.2%) fold, respectively. The present study provides useful insights into phylogenetic and evolution in the cervical cancer causing Human papillomavirus type 16. The optimized DNA can be chemically synthesized and over expressed in *E. coli* as compare to its wild type counterparts. Alternatively, the secondary structure and free energy of E6 were investigated that will be helpful to predict the evolution of primitive and genetically stable HPV type 16 strains. This finding provides new insight in better understanding of cervical cancer.

Keywords: Human Papillomavirus, Cervical cancer, CAI, RNA secondary structure.

INTRODUCTION

Cancer of the uterine cervix is second-most common malignant tumor in world. In India, about 100,000 women suffering from cancer every year (Luthra, et al., 1987) constituting about 16% of the world's annual incidence (WHO, 1986). Some sexually transmitted HPV types may cause genital warts, while others do not cause any noticeable signs of infection. Cervical cancer is a common cancer among women worldwide (CANCER Mondial [http://www-dep.iarc.fr/], 2005). Infection by HPV is one of the primary causes of mortality by cancer in northern Brazil. Sexually vigorous women from Manaus, Amazonas, without cytological alterations; and women with pre-malignant as well as malignant cytological alterations were examined for HPV virus identified using PCR and sequencing (Castro, et al., 2011). Papillomavirus is an ideal model system for the study of DNA virus evolution. On several levels, phylogenetic trees of papillomaviruses reflect the relationship of their hosts (Ong, et al., 1993). Human papillomavirus (HPV) replication occurs in terminally differentiating epithelium and requires the activation of cellular DNA replication proteins. Unscheduled DNA replication can result in the induction of apoptosis, while E6 protein induces the degradation of p53 (Thomas and Banks, 1999; Jakate and Saclarides, 1993). Cervical carcinomas are unfortunately complications of long standing infections with high risk types (Walboomers, et al., 1999). In a meta-analysis of cervical squamous cell carcinomas compared to high grade squamous intraepithelial lesions, HPV16, 18 and HPV45 appeared to display an elevated prevalence in cervical cancer (Clifford, et al., 2003).

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

A second meta-analysis revealed that HPV16 and HPV18 are more prevalent in squamous cell carcinoma (SCC) than in low-grade squamous intraepithelial leison (Clifford, et al., 2005). There is a need to establish a molecular relationship among HPV so that it would be useful further for control of cancer. RNA structure is less complex than protein structure so, it can be well characterized by identifying commonly occurring location of secondary structural elements. Usually, the folded structure of an RNA molecule is determined through X-ray crystallography, but it is time consuming and expensive. The RNA secondary structure solves many problem of the molecular biology. Thermodynamic free energy method has been used to predict secondary structures from a single RNA sequence (Zuker and Stiegler, 1981). The single stranded RNA base pairs within a single RNA molecule, forming the base pair stem regions. Various bioinformatics tools have been developed for predicting the secondary structure of RNA molecule. Many of these methods attempted to minimize the free energy of folded macromolecule, thus searching for most stable structure (Zuker, 1989). The secondary structure of nucleic acids is another well documented factor that affects probe binding for both DNA and RNA molecules (Gamper, et al., 1987). RNA secondary structure model has been predicted for HIV-1 genome and results highlighted that HIV-1 genome and potentially, many coding RNAs are interrupted by previously unrecognized regulatory motifs and that extensive RNA structure constitutes an important component of the genetic code (Watts, et al., 2009). There is a constant need for suitable diagnosis and vaccination for control of cancer. Therefore, we require an enough quantity of protein for diagnosis and vaccination purpose, while expression in host does not have enough quantity of translation system for overproduction of target gene. Recently it has development computational tools for designing and codon optimization of target gene for heterologous expression by increasing the translation efficiency. A potentially important vaccine strategy against virus, bacteria, and protozoa has been recently emerged in the form of DNA immunization. A successful DNA immunization requires high expression of genes derived from micro-organisms in mammalian cells; but may be the main hindrance for gene expression due to the inter-specific difference of codon usage (Koide, et al., 1998). DNA vaccines are plasmids able to express antigenic peptides in host (Babiuk, et al., 2000; Babiuk, et al., 2003). These are used as attractive alternative to over conventional vaccines as they generate both a cellular and a humoral immune response. There are different ways to optimize the DNA vaccine efficiency with the choice of antigen. It has been successfully enhanced by codon optimization (Uchijima, et al., 1998). In the present study, (1) established the phylogentic relationship among different countries isolates of HPV16, (2) predicted the RNA secondary structures and (3) optimized the codon of E6 gene.

MATERIALS AND METHODS

Collection and analysis of sequences

The complete E6 genes of HPV16 sequences were retrieved from NCBI-GenBank. The open reading frame(ORF) was validated and changed reverse complementary to proper +1 frame using the different softwares i.e. APEplasmid editor, Generunner and DNAstar (Lasergene7v). Relatedness of each genes sequence was evaluated by BLAST (Altschul, et al., 1997).

Construction and analysis of phylogeny

The different E6 gene of HPV16 nucleotide sequences were retrieved from http://www.ncbi.nlm.nih.gov/. The nucleotide sequences of E6 gene were compared and aligned with sequences deposited in the GenBank database using the BLAST. The sequences were aligned by CLUSTALX (Thompson, et al., 1997). Pair wise evolutionary distances were computed using the Jukes and Cantor equation implemented in the MEGA 4.0 program (Tamura, et al., 2007 and a phylogenetic tree was constructed by the neighbour-joining (NJ) method. The E6 sequences of 9 different isolates of HPV16 were used for construction of phylogenetic tree. A total of 100 bootstrapped values were sampled to determine a measure of the support for each node on the consensus tree.

Modeling of RNA secondary structure

Mfold and NUPACK were used for modeling of RNA secondary structure of E6 gene of HPV16. The widely used algorithms for RNA secondary structure prediction, which are based on a search for minimal free energy state (Zuker, 1989). The genetic algorithm (GA) simulates the natural folding pathway which takes place during RNA synthesis. This is not only enables new stems be added in the growing RNA chain, but also allows structures to be removed at later stages of the simulation if other pairings are found more favorable.

Gene designing and codon optimization

Nucleotide sequences of arranged into triplet (codons) and replace triplets with new one, generated with a given frequency distribution. In this process amino acid will be same, but codon of low frequency of an amino acid will replace with codon of high frequency, according to desired species frequency distribution.

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

Gene designer (https://www.dna20.com/index.php? pageID=220), Optimizer (Puigbo, et al., 2007, CAIcal and MrGene were used for optimization of DNA sequences at maximum suitable threshold level. Codon optimization and simulation of genes were performed on 10-15 % threshold level of host cellular codons. Codon adaptation index (CAI) was also calculated for each gene. It is widely acceptable as an effective measure of potential level of gene expression (Sharp and Li, 1987). Codon optimization is a technique to exploit the protein expression in living organism by increasing the translational efficiency of gene of interest by transforming DNA sequence of one species into DNA sequence of another species like plant sequence to human sequence, human sequence to bacteria or yeast.

Statistical analysis

The CAI, GC and AT of E6 genes of HPV16 were compared using the Wilcoxon matched pairs test. A two-tailed (α=2) probability p<0.05 were considered to be statistically significant. STATISTICA (version 7.0) was used for the analysis of native and optimized DNA sequences.

RESULTS AND DISCUSSION

For comparative analysis, the E6 sequences of HPV16 from different countries were obtained from NCBI-GenBank (Table 1). Homology results show that E6 gene sequences were more similar to the sequences of different HPV16 isolates. The size of different E6 gene was estimated between 456 to 477 bp. Phylogenetic trees was established by aligning the sequences of E6 gene from different isolates. Phylogenetic analysis using the various distance and character methods like neighbor-joining (NJ) showed that the topology is similar among the trees obtained with 100 bootstrap supports for clades. The phylogenetic tree was constructed in MEGA4.0 revealed a close relationship among different isolates of HPV16. The phylogenetic tree in present study showed that HPV16 isolates have divided into five groups (Figure 1). The first group was composed of Germany and Europe. The second group comprised Thailand and Netherlands, third group consisted of East Asia and South Korea while the China represented separate and fifth group consisted of Brazil and Africa. The phylogenetic approach revealed a close relationship with isolates of HPV16 from Brazil and Africa with significant bootstrap values being the closest. Similarly, analyzed the phylogeny and compared 49 isolates of HPV type 44 (HPV-44), subtype HPV-55 and 41 isolates of the subtype pair HPV-68a and -b, sampled from cohorts in four continents (Calleja-Macias, et al., 2005). In another study L1 capsid gene of HPV was targeted from northern Brazilian population. Twenty-three samples have been amplified; sequenced and analyzed 336 bp demonstrated a high incidence of high-risk HPV types in the population of Manaus, identified as HPVs 16, 33, 58, 66, 68. In which HPV type 16 was the most prevalent, presenting two variants similar to the Asian-American (AA) and East-Asian type (As) variants (Castro, et al., 2011). Although, the present phylogenetic tree showed the five groups of nine countries HPV16 isolates.

The genetic distance analyzed using Jukes and Cantor methods showed that the nucleotide sequence difference among isolates from 0.002 to 0.020 (Table 2). In other organism like fish analyzed the intra and inter-specific nucleotide sequence variation of ITS 1 in eight species of the genus *Thunnus* and two out-group species within the same family viz *Katsuwonus pelamis* and *Sarda orientalis*. In their study, the intra-specific nucleotide sequence variation in ITS1 including intra-genomic variation, was low ranging from 0.003 to 0.014, whereas variation between species within the genus *Thunnus* ranged from 0.009 to 0.05 (Chow, et al., 2006). Because of the hydrogen bonding the bases of RNA may form the base pair.

Table 1. Summary of E6 oncogene of HPV type 16.

S. No.	Countries	Accession Number	Size (bp)	Wild type Gene		Codon Optimized Gene	
				GC Content (%)	Free energy (Kcal/mol)	GC Content (%)	Free energy (Kcal/mol)
1.	Germany	EU118173	477	37.73	-093.96	49.3	-122.70
2.	Brazil	HM057182	477	37.73	-106.21	49.1	-107.40
3.	Africa	AF472508	456	37.06	-040.48	51.1	-104.80
4.	East Asia	AF534061	477	37.73	-090.76	50.7	-121.40
5.	Thailand	FJ610152	477	37.73	-072.68	50.3	-127.40
6.	China	Eu918764	477	37.94	-092.86	49.9	-110.80
7.	Europe	AY112663	456	37.50	-039.32	49.6	-105.20
8.	South Korea	AF187866	477	37.94	-044.78	50.3	-122.30
9.	Netherlands	AJ388066	456	37.28	-047.88	52.2	-110.40

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

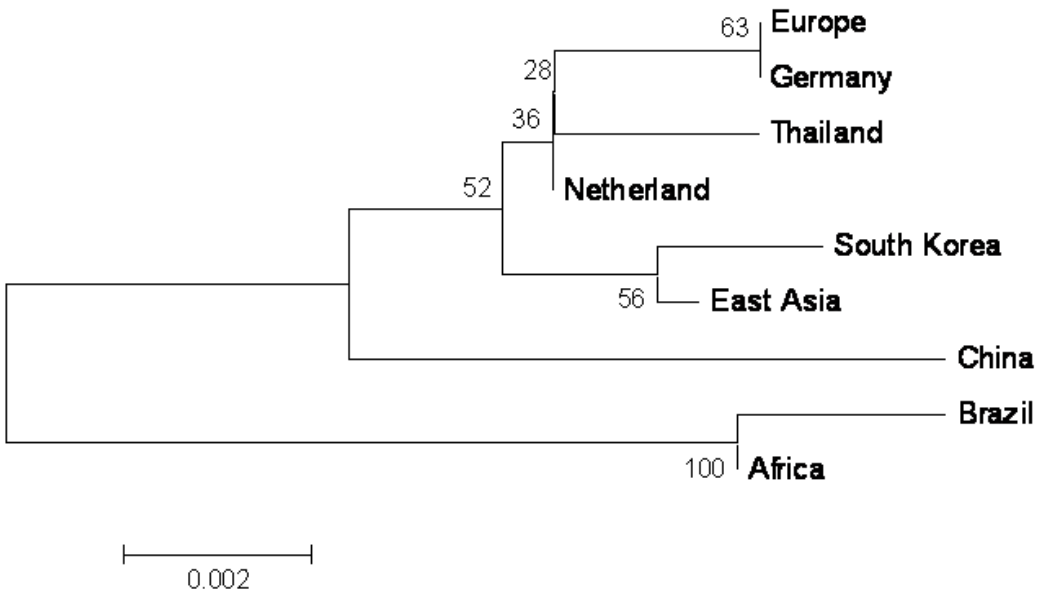

Figure 1. Phylogenetic analysis of E6 of different countries strains of HPV type 16. Bar 0.002 nucleotide changes per site.

Table 2. Jukes and Cantor genetic distances among different countries E6 gene of HPV16

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
[1]	0.000								
[2]	0.000	0.000							
[3]	0.004	0.004	0.000						
[4]	0.002	0.002	0.002	0.000					
[5]	0.004	0.004	0.004	0.002	0.000				
[6]	0.007	0.007	0.007	0.004	0.002	0.000			
[7]	0.011	0.011	0.011	0.009	0.011	0.009	0.000		
[8]	0.018	0.018	0.018	0.016	0.018	0.020	0.020	0.000	
[9]	0.016	0.016	0.016	0.013	0.016	0.018	0.018	0.002	0.000

Note: 1- Germany, 2- Europe, 3- Thailand, 4- Netherlands, 5- East Asia, 6- South Korea, 7- China, 8- Brazil, 9- Africa.

The G≡C is formed by a triple-hydrogen bond, A=U is formed by a double-hydrogen bond (Watson-Crick base pairs) and the G−U is formed by a single hydrogen bond (Wobble base pairs). In the secondary structure of rRNA, the Watson-Crick and Wobble base pairs occur in the RNA fold. Watson-Crick base pairs are more stable than wobble base pairs. The base pairs increase the structural stability, but the unpaired bases decrease the structural stability. In the present study, we maximized GC content while low GC content reduces the expression level whereas high AT cleaved by RNase E and degraded the mRNA. Therefore, it is an important that codon optimized gene with reference to availability of tRNA codon exists in host for better expression. It indicates the optimize gene demonstrated improve the level of expression in host (Mani, et al., 2010; Mani, et al., 2011).

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

The GC content of E6 gene was ranged from 37.1 to 37.90% in wild type sequences and increased after codan optimization ranged between 49.1% to 52.2%. (Table 3). The GC-rich sequences were analyzed within the 5' untranslated region of human basonuclin mRNA .The ability of this GC-rich sequence to form a large and stable secondary structure was suggested by experimental results from primer extension, RNase resistance, and computer analysis of the sequence (Tang and Tseng, 1999).

In the present investigation, we predicted the RNA secondary structure of wild and codon optimized genes that could be enhance the folding free energy. Nine of each wild and codon optimized genes were predicted RNA secondary structures with gibbs free energy of different isolates of HPV16 to provide the basic information for phylogenetic analysis. The secondary structure of E6 gene as shown (Figures 2-10) were analyzed based on conserved stems and loops. The folding free energy of wild gene of E6 were analyzed as -093.96, -106.21,-040.48, -090.76, -072.68, -092.86, -039.32, -044.78, -047.88 kcal/mol and after codon optimization free energy were -122.70, -107.40, -104.80, -121.40, -127.40, -110.80, -105.20, -122.30, -110.40 kcal/mol respectively. By the thermodynamic hypothesis, the actual secondary structure of a RNA sequence is the one with the minimum free energy (http://alg.csie.ncnu. edu.tw/course/biology/slidef2_rna.pdf). It was observed that similarities at the secondary structure are reflected at energy level. The only difference in their topology because of differences in nucleotide sequences.

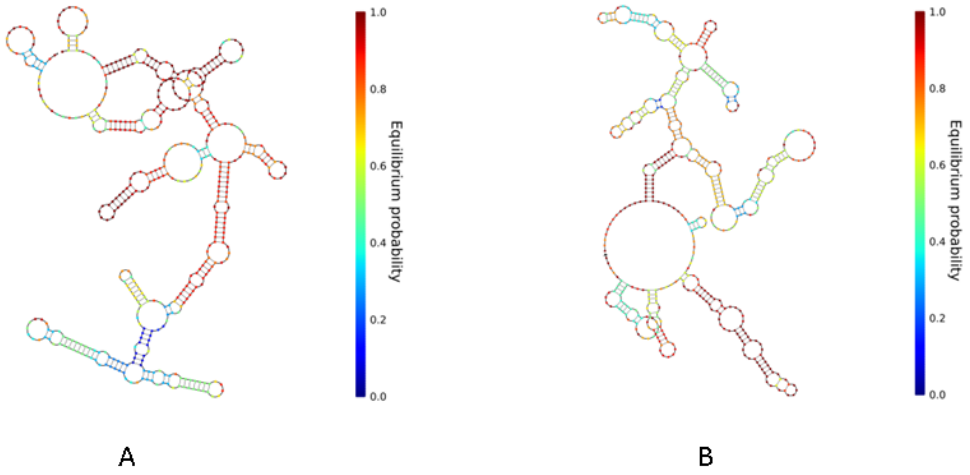

Figure 2. RNA secondary structure of wild (A) and optimized (B) E6 gene from Germany strain of HPV.

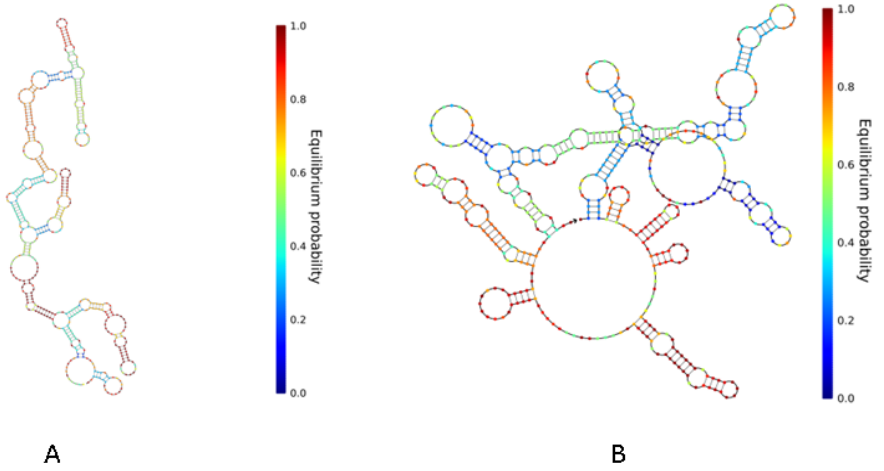

Figure 3. RNA secondary structure of wild (A) and optimized (B) E6 gene from Brazil strain of HPV.

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

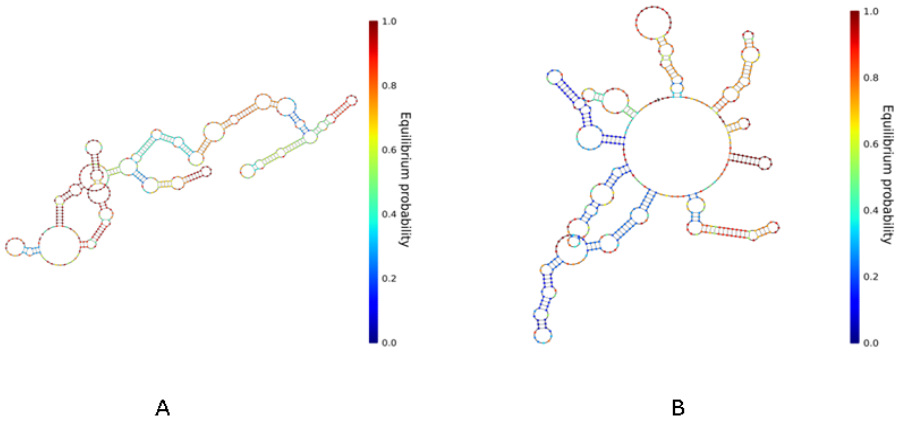

Figure 4. RNA secondary structure of wild (A) and optimized (B) E6 gene from Africa strain of HPV.

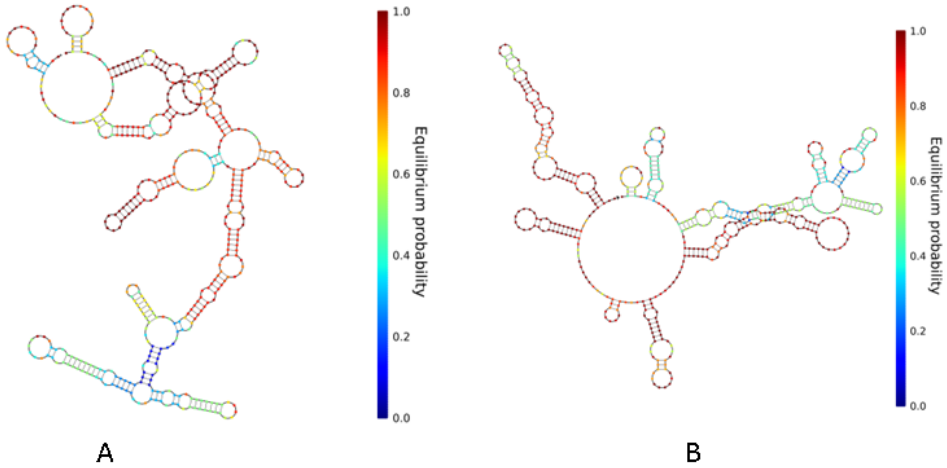

Figure 5. RNA secondary structure of wild (A) and optimized (B) E6 gene from East Asia strain of HPV

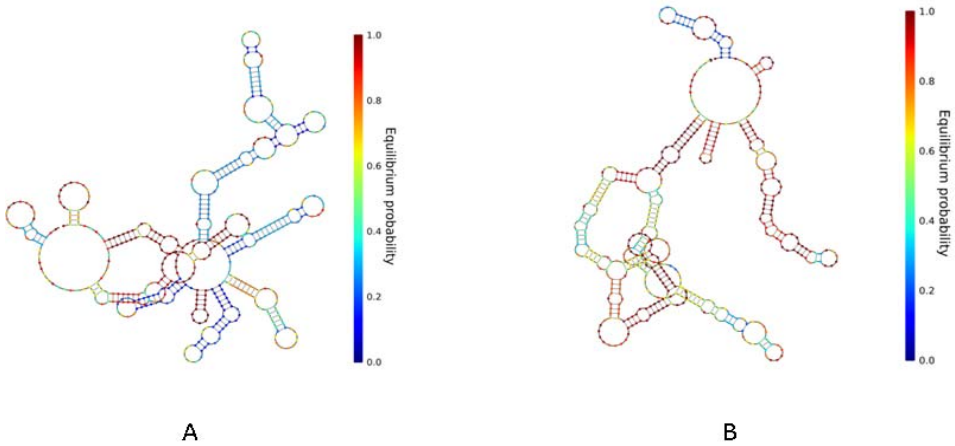

Figure 6. RNA secondary structure of wild (A) and optimized (B) E6 gene from Thailand strain of HPV.

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

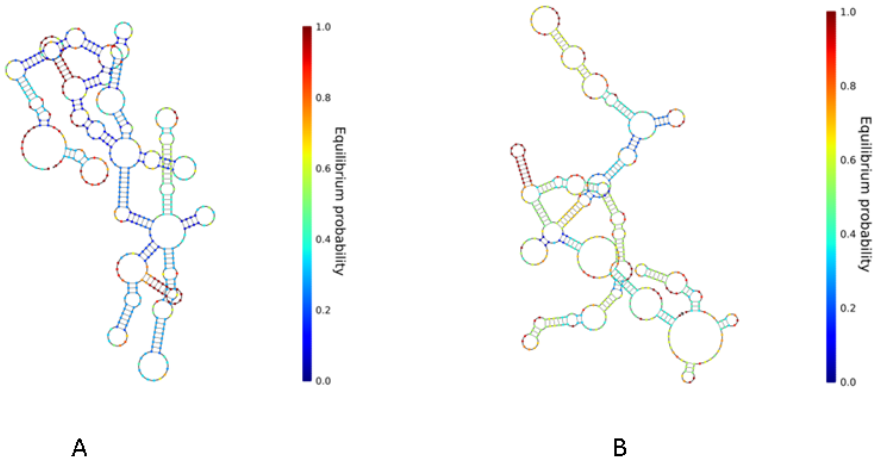

Figure 7. RNA secondary structure of wild (A) and optimized (B) E6 gene from China strain of HPV.

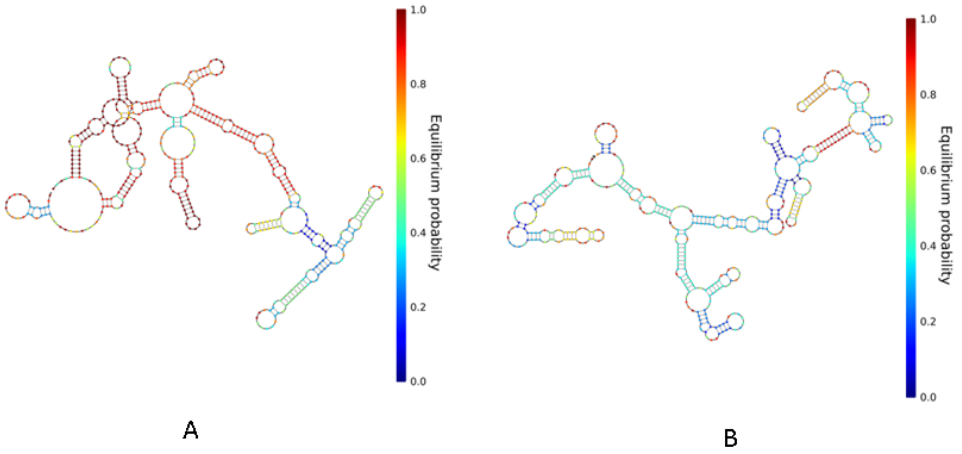

Figure 8. RNA secondary structure of wild (A) and optimized (B) E6 gene from Europe strain of HPV.

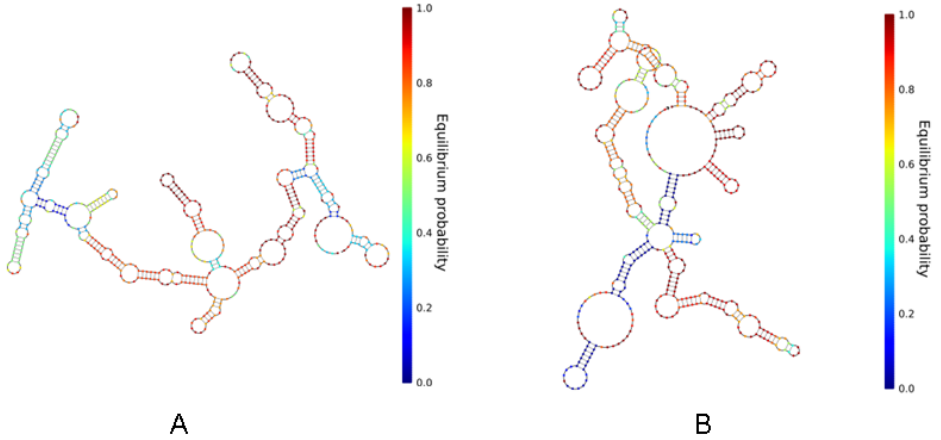

Figure 9. RNA secondary structure of wild (A) and optimized (B) E6 gene from South Korea strain of HPV

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

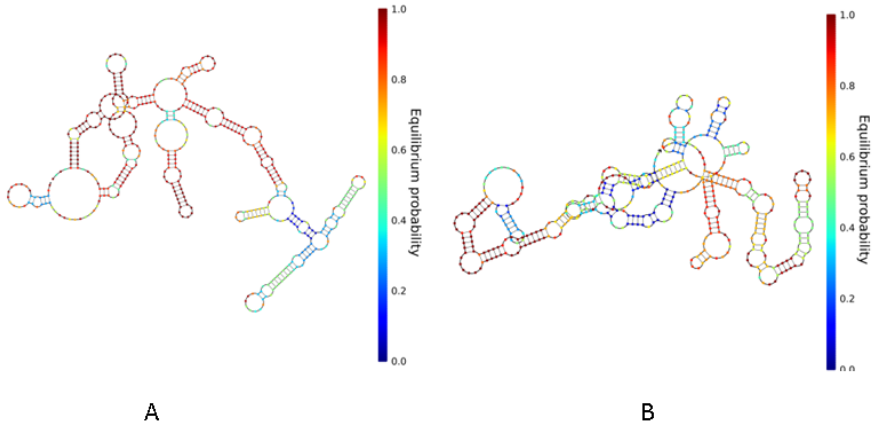

Figure 10. RNA secondary structure of wild (A) and optimized (B) E6 gene from Netherland strain of HPV.

Prediction of secondary structure indicate that domains base pair to form a central core region to several stem features implying that conserved is more important for appropriate RNA folding pattern. The orders of preference were interior loop, bulge loop, multiple branch loop, hairpin loop and exterior loop in all the isolates. Synonymous mutations do not alter the encoded protein, but they can influence gene expression. The previous study supports our work in reference to the free folding energy of RNA secondary structure of gene (Kudla, et al., 2009). The majority of RNA secondary structures were investigated that they are highly conserved among the HPV isolates. Sequence variability in the base paired regions are structurally conserved and therefore involves patterns of nucleotide substitution that retain base-pairing such as co-ordinated changes in paired bases to maintain binding (covariance) or alternation between C and U pairing to G residues (semi-covariance). Other study suggested the lowest free energy of 5S rRNA may divulge the most primitive bacteria and slow changes occurs throughout the evolution, while higher free energy indicates less stability during the evolution (Singh and Somvanshi, 2009). Our finding also suggested low free energy isolates will be evolutionary more stable and highly pathogenic and higher free energy isolates may be less stable and slow induce carcinoma in humans.

Moreover, secondary structures of internal transcribed spacer regions were reported in Sponge: Porifera (Itskovich, et al., 2008), Gyrodactylus: Platyhelminthes (Carey, et al., 2000), *Ixodes holocyclus*: Arthopds (Hlinka, et al., 2002) and Indian liver fluke, *Fasciola*: Fasciolidae (Prasad, et al., 2009). The secondary structure models also have been proposed for 16S rDNA from *Escherichia coli* and *Zea mays* chloroplast ribosome, the 18S rRNA from *Saccharomyces cerevisiae* and *Xenopus laevis* cytoplasmic ribosome, and the 12S rRNA from human and mouse mitochondrial ribosomes. Computational methods have improved the structure prediction accuracy by compiling information from mapping experiments, including pseudo knotted base pairs and by finding a secondary structure common to a set of homologous sequences. Further, investigation may also require for role of secondary structure in favoring the evolutionary conservation of the E6.

The CAI of E6 gene from wild type of DNA sequences was selected and its codon was optimized with reference to *E. coli* as it is an accepted host for heterologous gene expression. In the present study, we have considered CAI, GC and AT frequencies in all 9 strains of wild type HPV 16 ranged from 0.207 to 0.212, 37.10 to 37.90 and 62.10 to 62.20 respectively with an average (± SD) of 0.209±0.002, 37.58±0.26 and 62.41±0.26 respectively. The respective frequencies of these in optimized DNA ranged from 0.708 to 0.8, 49.1 to 52.2 and 47.8 to 50.9 respectively with an average (± SD) of 0.766±0.027, 50.27±0.96 and 49.72±0.96 respectively (Table 3).

The revolution of molecular biology gains momentum, as different number and variety of natural genes have been re-designed at the nucleotide level and synthesized in attempts to improve protein yields (Gustafsson, et al., 2004). Codon distribution acts in respect to GC content of genome and the changes in codon usage are at least partly explained by mutation-selection equilibrium between the different synonymous codons in each organism (Knight, et al., 2001). On comparing the mean, CAI, GC and AT of all HPV16 isolates of optimized DNA was found to be significantly (p <0.01) different and higher than the respective values of wild type. The mean CAI, GC and AT in optimized DNA were 3.6 (72.7%) and 1.3 (25.2%) fold, higher than the respective mean values of wild type. In this study, we have observed that the CAI values of E6 genes of the studied isolates of HPV16 were more as compared to wild type sequences. The CAI value was shown at Y-axis, while the number of strains on X-axis (Figure 11). Recently, the expression of aerolysin (Singh, et al., 2010a) and hemolysin of *A. hydrophila* have been reported in *E. coli* (Singh, et al., 2009).

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

Table3. E6 gene of HPV type 16 from different countries strains for heterologous expression in *E. coli* of wild type and codon optimized sequences.

E6 Gene	Wild type gene			Codon optimized gene		
	CAI	GC %	AT %	CAI	GC %	AT %
Germany	0.212	37.7	62.3	0.763	49.3	50.7
Brazil	0.207	37.5	62.5	0.781	49.1	50.9
Africa	0.212	37.1	62.9	0.800	51.1	48.9
East Asia	0.207	37.7	62.3	0.781	50.7	49.3
Thailand	0.210	37.7	62.3	0.763	50.3	49.7
China	0.212	37.9	62.1	0.772	49.9	50.1
Europe	0.212	37.5	62.5	0.708	49.6	50.4
South Korea	0.207	37.9	62.1	0.746	50.3	49.7
Netherlands	0.208	37.3	62.7	0.787	52.2	47.8
N	9	9	9	9	9	9
Min	0.207	37.1	62.1	0.708	49.1	47.8
Max	0.212	37.9	62.9	0.8	52.3	50.9
Mean ± SD	0.209±0.002	37.58±0.26	62.41±0.26	0.766±0.027**	50.27±0.96**	49.72±0.96**

**- $p<0.01$: in comparison with wild type.

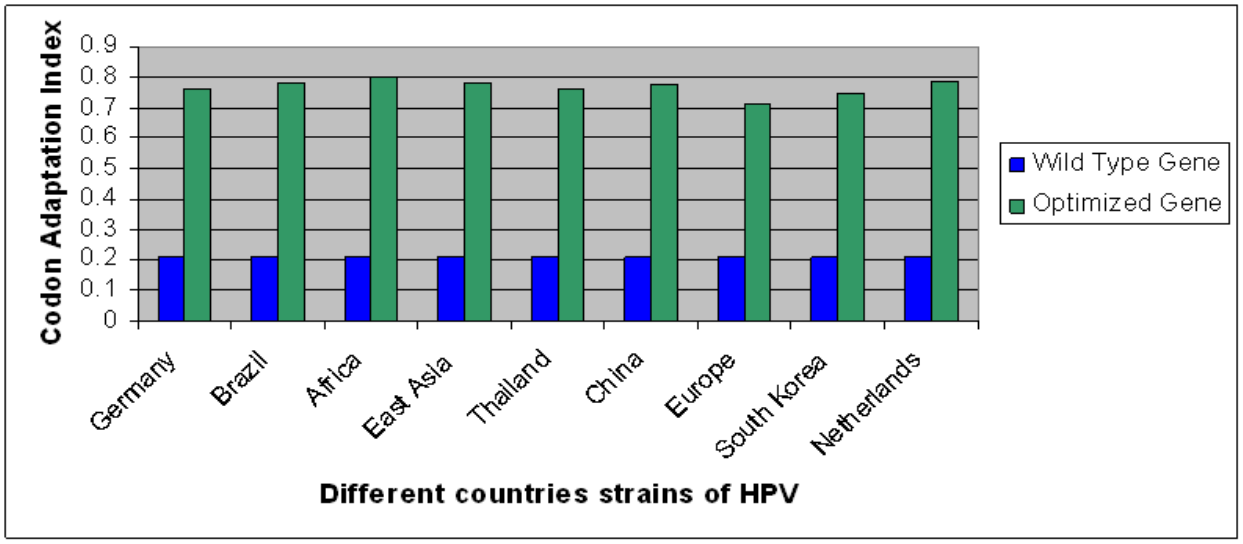

Figure 11. Graphical representation of wild and codon optimized E6 gene of DNA sequences of different countries strains of HPV type 16.

Furthermore, another study demonstrates that the codon optimization of virulence proteins such as hemolysin, aerolysin and lipase of *A. hydrophila* for over expression in *E. coli* without codon bias. The CAI values of these genes were 2-fold greater than the wild type of genes (Singh, et al., 2010b). The level of gene expression of eukaryotic genes introduced into mammalian cells depends on many factors viz. gene copy number, transcriptional control elements, site of chromosomal integration, mRNA stability and translational efficiency (Gross and Hauser, 1995). The human papillomavirus type 16 (HPV-16) of E5 protein which is codon optimized and showed the 6-9 folds higher expression than wild-type HPV-16 (Disbrow, et al., 2003). The 505 amino acid L1 protein of the HPV 11 is the major capsid polypeptide that has been shown to self-assemble into virus-like particles (VLPs) in vivo and in vitro. Codon optimized L1 gene expression in mammalian cells the protein levels that were 100-fold higher than wild-type HPV 11 L1, while no obvious differences were seen in the level of mRNA (Mossadegh, et al., 2004).

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

CONCLUSION

The present study provides useful insights into phylogenetic and evolution in the cervical cancer causing Human papillomavirus type 16. The wild type frequencies of E6 gene sequences of HPV16 showed significantly (p <0.01) different codons and a good correlation between the isolates based on statistical analysis was observed. These optimized DNA will be chemically synthesized and over expressed in *E. coli* as compare to its wild type counterparts. Furthermore, this study provides relevant information on the evolutionary significance of E6 sequences, which play a significant role in the inter-specific variation among HPV isolates. These nine isolates can be considered as a different haplotypes. The expression of the different haplotypes of E6 gene could be more due to the high CAI value as compare to wild type DNA. These have several applications like to remove the stop codons, to clone, in custom design of synthetic genes, to improve the functionality of genes, to increase protein expression level and for lower production costs, in drug development. On the other end, the secondary structure and free energy of E6 were investigated that will be helpful to predict the evolution of primitive and genetically stable HPV isolates. Furthermore, studies will be needed for *in vitro* validation of codon optimized DNA as the studies provide more information of high level of expression and that could be used in adequate amount for vaccine production.

Competing interests: There is no competing interest.

Authors' contributions: DKC, IM and VS designed the research work and performed experiments. All authors analyzed the data, wrote the paper and approved the final manuscript.

ACKNOWLEDGMENTS

The authors are thankful to A.K. Singh, Pritee Singh and Reena for providing the suggestions, fruitful discussion and encouragement during preparation of the manuscript.

REFERENCES

Altschul, S.F., Thomas, L.M., Alejandro, A.S., Jinghui, Z. and Zheng, Z. (1997). Gapped BLAST and PSIBLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Babiuk, L.A., Babiuk, S.L., Loehr, B.I. and van Drunnen Littel-van den Hur (2000). Nucleic acid vaccines: Research tool or commercial reality. Vet Immunol Immunopathol 76:1-23.

Babiuk, L.A., Pontarollo, R., Babiuk, S., Loehr, B. and van Drunnen Littel-van den Hur (2003). Induction of immune responses by DNA vaccines in large animals. Vaccine 21:649-658.

Calleja-Macias, I.E., Kalantari, M., Allan, B., Williamson, A.L., Chung, L.P., et al. (2005). Papillomavirus Subtypes Are Natural and Old Taxa: Phylogeny of Human Papillomavirus Types 44 and 55 and 68a and –b. J Virol. 79:6565-6569.

Carey, O.C., Holly. A. and Catherine, M.C. (2000). Sequence and secondary structure variation in the Gyrodactylus (Platyhelminthes: Monogenea) Ribosomal RNA gene array. J Parasitol. 86:567-576.

Castro, M.M., Farias, I.P., Borborema-Santos, C.M., Correia, G. and Astolfi-Filho, S. (2011). Prevalence of human papillomavirus (HPV) type 16 variants and rare HPV types in the central Amazon region. Genet Mol Res. 10:186-196.

Chow, S., Nakagawa, T., Suzuki, N., Takeyama, H. and Matsunaga, T. **(**2006**).** Phylogenetic relationships among *Thunnus* species inferred from rDNA ITS1 sequence. J Fish Biol. 68:24-35.

Clifford, G.M., Rana, R.K., Franceschi, S., Smith, J.S., Gough, G., et al. (2005). Human papillomavirus genotype distribution in low-grade cervical lesions: comparison by geographic region and with cervical cancer. Cancer Epidemiol Biomarkers Prev. 14:1157-1164.

Clifford, G.M., Smith, J.S., Aguado, T. and Franceschi, S. (2003). Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. Br J Cancer 89:101-105.

Disbrow, G.L., Sunitha, I., Baker, C.C., Hanover, J. and Schlegel, R. (2003). Codon optimization of the HPV-16 E5 gene enhances protein expression. Virology 311(1):105-14.

Gamper, H.B., Cimino, G.D. and Hearst, J.E. (1987). Solution hybridization of crosslinkable DNA oligonucleotides to bacteriophage M13 DNA. Effect of secondary on hybridization kinetics and equilibria. J Mol Biol. 197:349-362.

Gross, G. and Hauser, H. (1995). Heterologous expression as a tool for gene identification and analysis. J Biotechnol. 41:91-110.

Gustafsson, C., Govindarajan, S. and Minshull, J. (2004). Codon bias and heterologous protein expression. Trends Biotechnol. 22:346-353.

Hlinka, O., Murrell, A. and Barker, S.C. **(**2002). Evolution of the secondary structure of the of the rRNA internal transcribed spacer 2 (ITS2) in hard ticks (Ixodidae, Arthropoda). Heredity 8:275-279.

Itskovich, V., Gontcharov, A., Masuda, Y., Nohno, T. and Belikov, S. **(**2008). Ribosomal ITS Sequences allow resolution of freshwater sponge phylogeny with alignments guided by secondary structure prediction. J Mol Evol. 67:608-620.

Jakate, S.M. and Saclarides, T.J. (1993). Immunohistochemical detection of mutant P^{53} protein and human papillomavirus-related E6 protein in anal cancers. Dis colon rectum. 36:1026-1029.

Knight, R.D., Freeland, S.J. and Landweber, L.F. (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol 2:research0010.1-0010.13.

Koide, Y., Nagata, T., Yoshida, A. and Uchijima, M. (1998). Optimization of codon usage of plasmid DNA vaccine is required for the effective MHC class I restricted T cell responses against an intracellular bacterium. J Immunol. 161:5594-5599.

Kudla, G., Murray, A.W., Tollervey, D. and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. Science 324:255-258.

Luthra, U.K., Prabhakar, A.K., Seth, P., Aggarwal, S.S., Murthy, N.S., et al. (1987). Natural history of precancerous and early cancerous lesions of the uterine cervix. Acta Cytol. 31(3):226-236. PMID: 3035845.

Mani, I., Chaudhary, D.K., Somvanshi, P. and Singh, V. (2010). Codon optimization of the potential antigens encoding genes from *Mycobacterium tuberculosis*. IJABPT 1:292-301.

Mani, I., Singh, V., Chaudhary, D.K., Somvanshi, P. and Negi, M.P.S. (2011). Codon optimization of the major antigen encoding genes of diverse strains of Influenza A virus. Interdiscip Sci Comput Life Sci. 3:36-42.

Mossadegh, N., Gissmann, L., Müller, M., Zentgraf, H., Alonso, A. and Tomakidi, P. (2004). Codon optimization of the human papillomavirus 11 (HPV 11) L1 gene leads to increased gene expression and formation of virus-like particles in mammalian epithelial cells. Virology 326(1):57-66.

Ong, C.K., Chan, S.Y., Campo, M.S., Fujinaga, K. and Mavromara-Nazos, P. (1993). Evolution of Human Papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with Human ethnic groups. J Virol. 67:6424-6431.

Prasad, P.K., Tandon, V., Biswal, D.K., Goswami, L.M. and Chatterjee, A. (2009). Use of sequence motifs as barcodes and secondary structures of internal transcribed spacer 2 (ITS2, rDNA) for identification of the Indian liver fluke, *Fasciola* (Trematoda: Fasciolidae). Bioinformation. 3:314-320.

Puigbo, P., Guzman, E., Romeu, A. and Garcia-Vallve, S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 35:1-6.

Sharp, P.M. and Li, W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281-1295.

Singh, V. and Somvanshi, P. (2009). Computational modeling analyses of RNA secondary structures and phylogenetic inference of evolutionary conserved 5S rRNA in the prokaryotes. J Mol Graph Model. 27:770-776.

Singh, V., Chaudhary, D.K., Mani, I., Somvanshi, P., Rathore, G., et al. (2010b) Molecular identification and codon optimization analysis of major virulence encoding genes of *Aeromonas hydrophila.* African J Microbiol Res. 4:952-957.

Singh, V., Somvanshi, P., Rathore, G., Kapoor, D. and Mishra, B.N. (2010a). Gene cloning, expression and characterization of recombinant aerolysin from *Aeromonas hydrophila.* Appl Biochem Biotechnol. 160:1985-1991.

Singh, V., Somvanshi, P., Rathore, G., Kapoor, D. and Mishra, B.N. (2009). Gene cloning, expression and homology modeling of hemolysin gene from *Aeromonas hydrophila*. Protein Expr Purif. 65:1-7.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.00. Mol Biol Evol. 24:1596-1599.

Tang, T. and Tseng, H. (1999). A GC-rich sequence within the 5" untranslated region of human basonuclin mRNA inhibits its translation. Gene 237:35-44.

Coden : IJABPT Copyrights@2012 ISSN : 0976-4550

Thomas, M. and Banks, L. **(**1999). Human papillomavirus (HPV) E6 interactions with Bak are conserved amongst E6 proteins from high and low risk HPV types. J Gen Virol. 80:1513-1517.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.

Uchijima, M., Yoshida, A., Nagata, T. and Koide, Y. (1998). Optimization of codon usage of plasmid DNA vaccine is required for the effective MHC class I-restricted T cell responses against an intracellular bacterium. J Immunol. 161:5594-5599.

Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X. and Kummer, J.A. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 189:12-19.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., et al. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460:711-716.

World Health Organization. Control of cancer of the cervix uteri: review article based on a report of a WHO meeting, November 1985, Geneva. Bull World Health Organ 1986, 64: 607-618. PMID: 3490930.

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. Science 244:48-52.

Zuker, M. and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 9:133-148.

.