Indian J.Sci.Res.2(3) :45-52, 2011

IN SILICO MICRORNA IDENTIFICATION AND TARGET PREDICTION IN HPV

SHALLU KALIA[a1] AND SARIKA SAHU[b]

[a]Department of Bioinformatics, Hans Raj Mahila Maha Vidyalaya, Jalandhar, Punjab, India
E-mail: shallu.iiit@gmail.com
[b]School of Biotechnology & Center for Bioinformatics, Shobhit University, Meerut, U.P., India
E-mail: sahusarikaiiita@gmail.com

ABSTRACT

Interference RNAs are now a day's one of the most admiring methods for gene silencing. They are of many types but among them microRNAs are very important as they can regulate gene expression by controlling the protein translation mechanism during variety of cell phenomena viz. proliferation of cancer, differentiation of stem cells and neurons etc. MicroRNAs are produced from non-coding DNA region. Since the report of the lin-4 RNA (1993) significant progress has been made in miRNA research. About 300 miRNAs have been identified in different organisms to date. Experimental identification of miRNA is a slow process as they are difficult to isolate. Thus computational identification from genomic sequence is the new approach.
MiRNA related sequences are present in most of the genomes, but this information is unexplored till now in many viral genomes and so there is a need to study it in detail. Computational and experimental microRNA prediction is a challenge for researchers. The basics of computational miRNA prediction are based on few parameters like calculation of optimum free energies (dG), Structural continuity, and number of G: C baces pairing etc. In the present work we have predicted four miRNA molecules in Human Papilloma virus (HPV) through viro-miRNA algorithm which is a computer based program. We have also predicted possible target genes which are inhibited by these predicted microRNAs.

KEY WORDS: MicroRNA, Human Papilloma Virus, Precursors, and Interference RNA

Papillomaviruses are small double-stranded DNA-based viruses. They usually infect the skin and mucous membranes of humans and a variety of animals (Pfeffer et al., 2004). They may cause warts; while others may cause a sub clinical infection that can result in precancerous lesions (Pfeffer et al., 2005). The statistical data show that head and neck squamous cell carcinoma (HNSCC) may be caused by HPV virus, as these patients tend to be younger, nonsmokers and nondrinkers (Wang et al., 2004). Different types of HPV known are about 150 or more. Among them the high risk HPV are types 16 and 18 (Pfeffer et al., 2004). They are of approximately 7900 base pairs (bp).

Current researchers focus themselves towards microRNAs prediction in these economically important viruses, in this connection viral infection mechanism came into the picture to control the host cell defense mechanism through specific RNAi technique. Generally viruses produce mi-RNA, which plays a major role in the disease pathogeny. MicroRNAs are a family of small, non-coding RNAs that regulate gene expression in a sequence-specific manner. MiRNAs are usually small sequences of length ~21-23 nucleotides derived from sequences of length ~70-100 nucleotides called pre-precursors (Enright et al., 2003). These pre-precursor may originate from the UTR, intergenic or intronic regions having a potential stem loop structure (Rodriguez, et al., 2004).

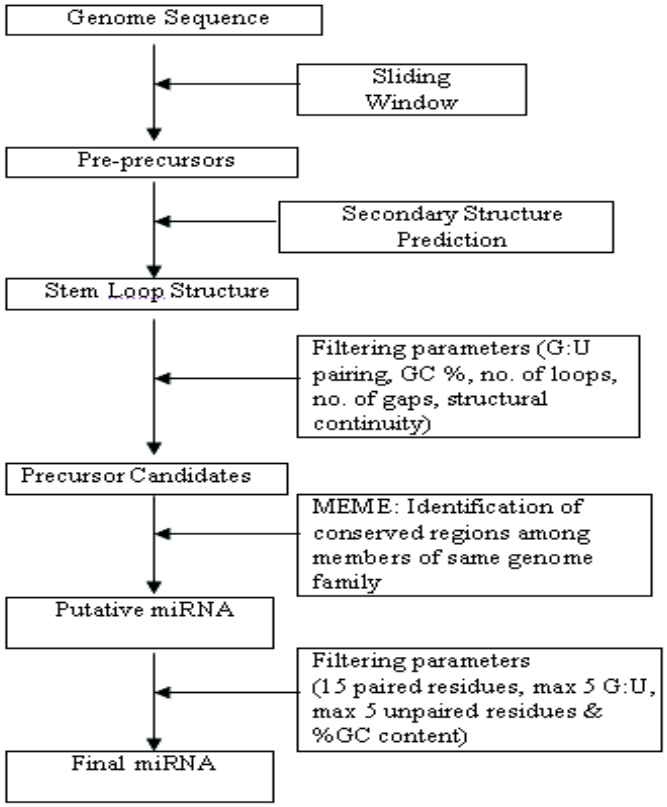
Till now total number of experimentally identified miRNAs is more than 200 and has been in different organisms but in case of viruses this number is very meager (Bonnet et al., 2004). Currently 35 experimentally proved miRNAs are available in mirBase database (Griffiths-Jones, 2004) for different viruses. In miRNA production mechanism, DNA of miRNA gene is transcribed into a single-stranded RNA molecule with self-complementary regions. These regions further bind and form a double stranded RNA hairpin loop; these imperfect hairpin loops called as primary miRNA structures. Drosha, a nuclear enzyme of RNase-III type, cleaves the base of pre-miRNA hairpin to form pre- miRNA. The pre-miRNA molecule is then actively transported into the cytoplasm by Ran GTP based carrier protein Exportin-5 (Yi et al., 2003). There after Dicer enzyme cuts this pre-miRNA into ~21-23 nucleotides length and releases the mature miRNA. These miRNAs are then incorporated into a multi-component RNAinduced silencing complex (RISC). It is facilitating miRNA to bind with specific mRNA targets and mediates degradation of respective mRNA (Enright et al., 2003).

[1]**Corresponding author**

MATERIALS AND METHODS

To predict miRNAs from viral genomes, we have used our in-house developed program Viro-miRNA scan by incorporating standalone module MFold (ver 3.1) and other programs with different filtering parameters; the whole program has been implanted in Perl and BioPerl [http://bioperl.org/.] for connecting all the modules. However, front end was developed in Perl-tk, as it is having graphical user interface with flexibility, by giving user defined filtering parameters. viromiRNA scan is also supported with a web interphase. The complete algorithm and flowchart in detail is given in Figure-1.

Fig.1: Viro-miRNA algorithm



Pre-Precursor Identification

The miRNAs prediction procedure is as follows. First the viral genome (HPV) was downloaded from NCBI [http://www.ncbi.nlm.nih.gov/.] (Ac No.NC_001526) and the sliding window method was adopted for generating subsequences on the genome. In this program, the whole viral genome from 3' to 5' end is divided into 70 nucleotide size of subsequences called as pre-precursors. Each pre-precursor obtained by running the above said program and subjected to MFold secondary structure prediction program.

Filtering Parameters for Predicting miRNA Precursors (Secondary Structures)

Mfold program has generated a number of structures along with information about various parameters including free energy (dG) number of base pairs, number of internal loops, bulges etc (Zuker, 2003). After studying the dG value of all processed precursors of HPV -18 Kcal/mol has been decided as final dG cutoff value.

Predicted secondary structures falls into number of clusters and structures were filtered against following parameters:

a) Structures having dG greater than 18Kcal/mol were eliminated.

b) Number of bulges and internal loops was restricted to obtain the precursor hairpins

c) Discrete folded structures were discarded based on the structural continuity**.**

d) %GC content. After this step, rest of the process done manually to get more accurate results.

Prediction of Putative miRNA Sequences from miRNA Precursors

Generally miRNA prediction in Eukaryotes is based on sequence conservation in the related genome families. But in the case of viruses the conserved regions (Motifs) (Can Cui, et al., 2006) predicted by using precursor sequences for alignment with in the same clusters. To find out the motifs from the clusters multiple sequence alignment based MEME tool was used, where miRNAs generally consist of 21-23 nucleotide length.

After prediction of miRNA from Human papilloma virus, target gene prediction in oral cancer in humans was done by using data from Upstate Medical University. We have set parameters as follows score for each 20 nucleotide is 3, G: C wobble pair 6, indels 1 and 3 mismatches are allowed.

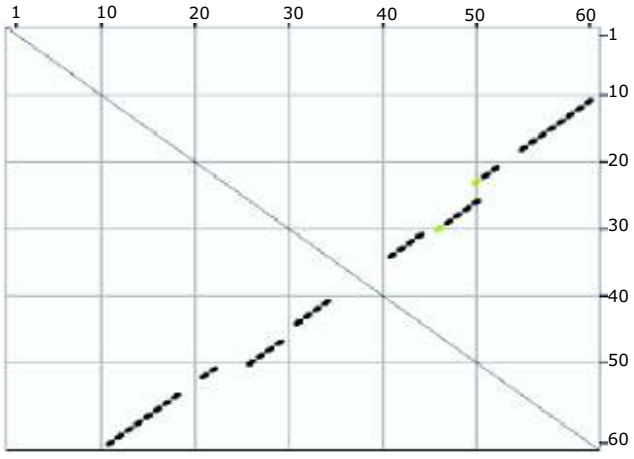
RESULTS AND DISCUSSION

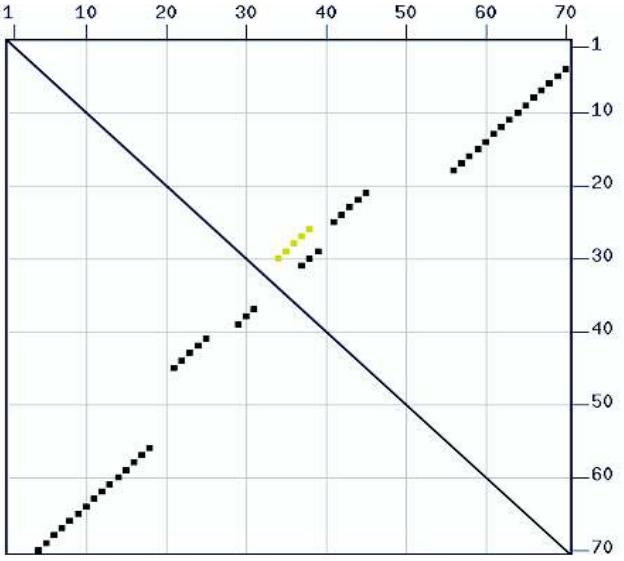
Through viro-miRNA program initially 2095 pre-precursors of length 70 nucleotides were obtained from Human papilloma virus genome (HPV). Large numbers of secondary structures were obtained after submitting the pre-precursors to MFold program. Secondary structures were passed through number of filtering parameters as mentioned in the methodology (like dG, continuity of structure etc.).

In the above said one of the crucial filtering parameter is dG value, but there is a major problem to decide dG cutoff value for the precursors. Hence, the predicted precursor structures dG value of HPV were observed and structures having dG less than -18 kcal/mol were observed to be having optimal folding, finally -18 kcal/mol was accepted as cutoff value and which have yielded 209 precursors. Out of these, only 7 potential candidates were selected later only four of them were finalized based on high structural continuity and energy dot plots (Fig. 2). In this study energy dot plot for four precursors gave detail idea about RNA folding. Energy plot with black dotted diagonal lines showed optimal folding, while yellow lines indicated base pairs suboptimal folding plots, in the case of precursor 2187, black dotted lines and yellow dotted lines plotted together on a diagonal line, which shows that the precursor is having structural continuity (Fig. 2A). Similar results were also observed in precursor 232. While precursors 1963, 510, 1992 have straight black dotted graph but the base pair suboptimum (yellow) line got deviated, it shows that it is having very low structural continuity (Fig. 2B). Finally four precursors were selected with dG values of all around -18kcal/mol. Though precursor 1993 has been observed with high dG value -20.6 but having many loops and continuity of structure is also very less. Due to presence of buldges and internal loops P-1993 was discarded. P-881 showed two miRNA, but one was discarded due to large bulges. Finally four precursors qualified for prediction of miRNAs (HPV-miR) and their reverse complement (HPV-miR*) sequences.

Fold of gi|9627100|ref|NC_001526.1|Precursor-2187 Human Pap deltaG in plot file = 0.8Kcal/mole

Fig. 2A: Fold of gi|9627100|ref|NC_001526.1| Precursor-1992 Human Pap deltaG in plot file = 1.0 Kcal/mole

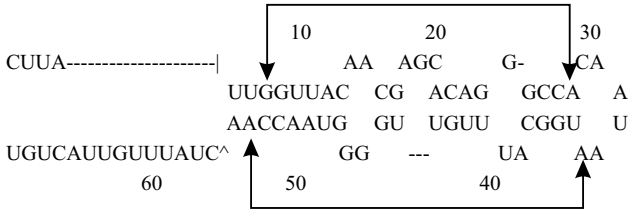




Lower Triangle Shows	Optimal energy: -18.9
Optimal Energy	-18.9< energy <= -18.6
Upper Triangle	-18.6< energy<=-18.4
Basepairs Plotted: 28	-18.4< energy<=-18.1

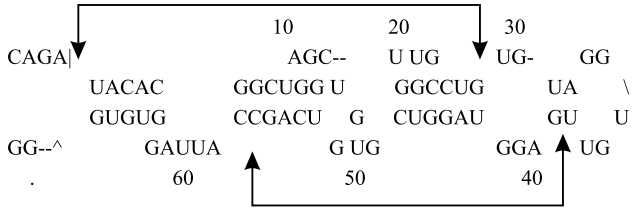
Fig. 2B: Energy plot format of predicted miRNA in precursor no.2187 and 1992

Linear RNA folding at 37° C. [Na+] =1.0 M, [Mg++] =0.0 M.
Structure 1
Folding bases 1 to 70 of gi|9627100|ref|NC_001526.1| Precursor-2187 Human pa
dG=-18.1 dH=-124.8 dS=-344.0 Tm=89.6



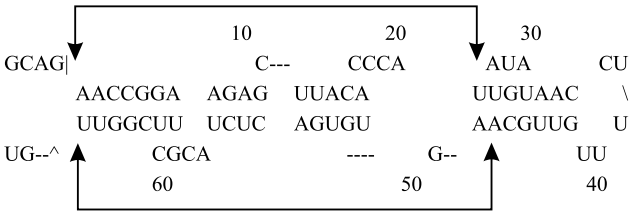
Precursor-2187

Linear RNA folding at 37° C. [Na+] =1.0 M, [Mg++] = 0.0 M.
Structure 1
Folding bases 1 to 70 of gi|9627100|ref|NC_001526. 1|Precursor-1972 Human pa
DG = -18.6 dH = -183.5 dS = -531.7 Tm = 72.0



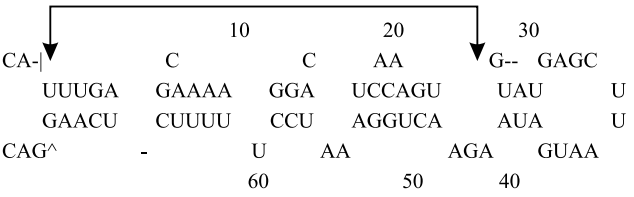
Precursor-1972

Linear RNA folding at 37° C. [Na+] = 1.0 M, [Mg++] = 0.0 M.
Structure 1
Folding bases 1 to 70 of gi|9627100|ref|NC_001526.1| Precursor-232 Human pap
dG = -18.5 dH = -180.9 dS = -523.6 Tm =72.3



Precursor-232

Linear RNA folding at 37° C. [Na+] = 1.0 M, [Mg++] = 0.0 M.
Structure 1
Folding bases 1 to 70 of gi|9627100|ref|NC_001526.1|Precursor-881 Human pap
dG = -18.4 dH = -202.1 dS = -592.3 Tm = 68.1

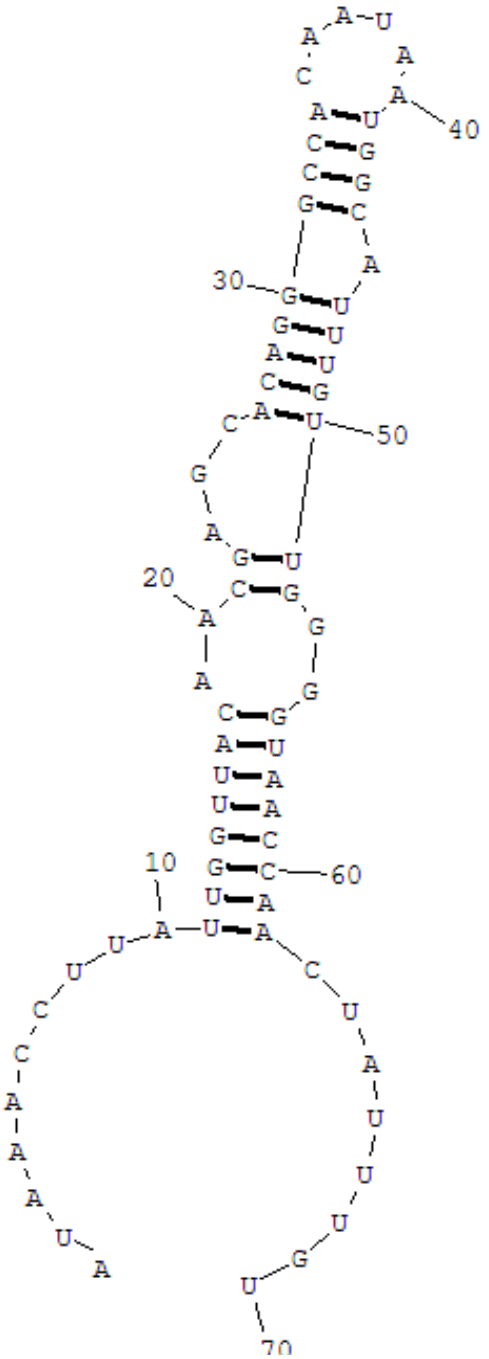


Precursor-881

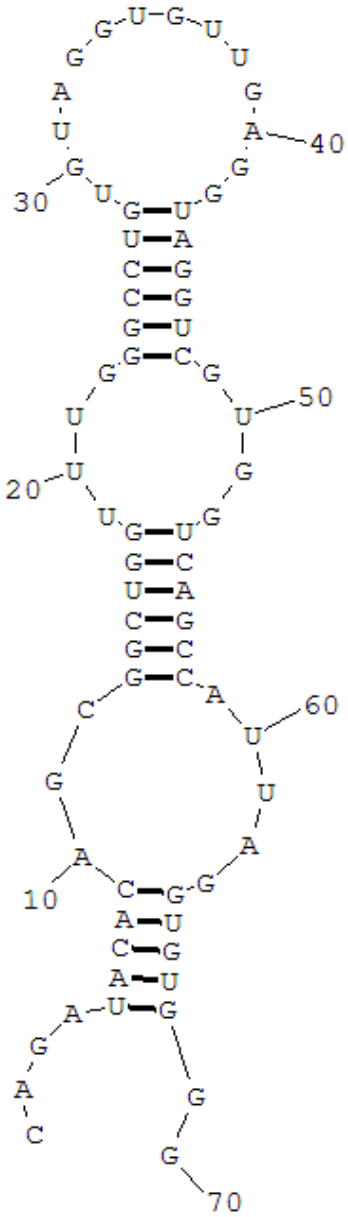
Fig. 3: Text format of Predicted folded secondary structures of P-2187, 1972, 232 and 881 genome of HPV (with continuous stalks)

To identify the miRNAs sequences from the selected precursors multiple alignment method was used. To find out the conserved regions (Can Cui, et al., 2006), predicted precursor sequences from each cluster was used for alignment. Precursors corresponding to the same clusters like P-2185-2190, 232-233, 1972-1973 (Figure 3) were submitted to MEME program. First miRNA motif of precursor 2187, 1972 located within the stem region and it was selected, while second motif located in the loop region, due to this reason second motif was discarded. Similarly, precursor 232-233 yielded one raw miRNA and one reverse complementary raw miRNA. Precursor 881 yielded single miRNA with no complementary. Raw miRNAs were further

processed through different filtering parameters (Rodriguez, A., et al., 2004*)* (minimum base paired, maximum unpaired residue, G: U pairs and %GC content etc.). After passing through these filtering parameters location of the miRNA (1020) was shifted and final miRNA were predicted. Finally P-2187,1972 and 232 yielded HPV-1miR, HPV-1miR* and P-881 yielded HPV-1miR as final miRNA (miR) and their reverse complement (miR*). Similarly in the year 2001 *C. elegans* miRNAs were first predicted by using both bioinformatics and cDNA cloning techniques (Lee and Am bros, 2001).



Precursor- 2187



Precursor-1972

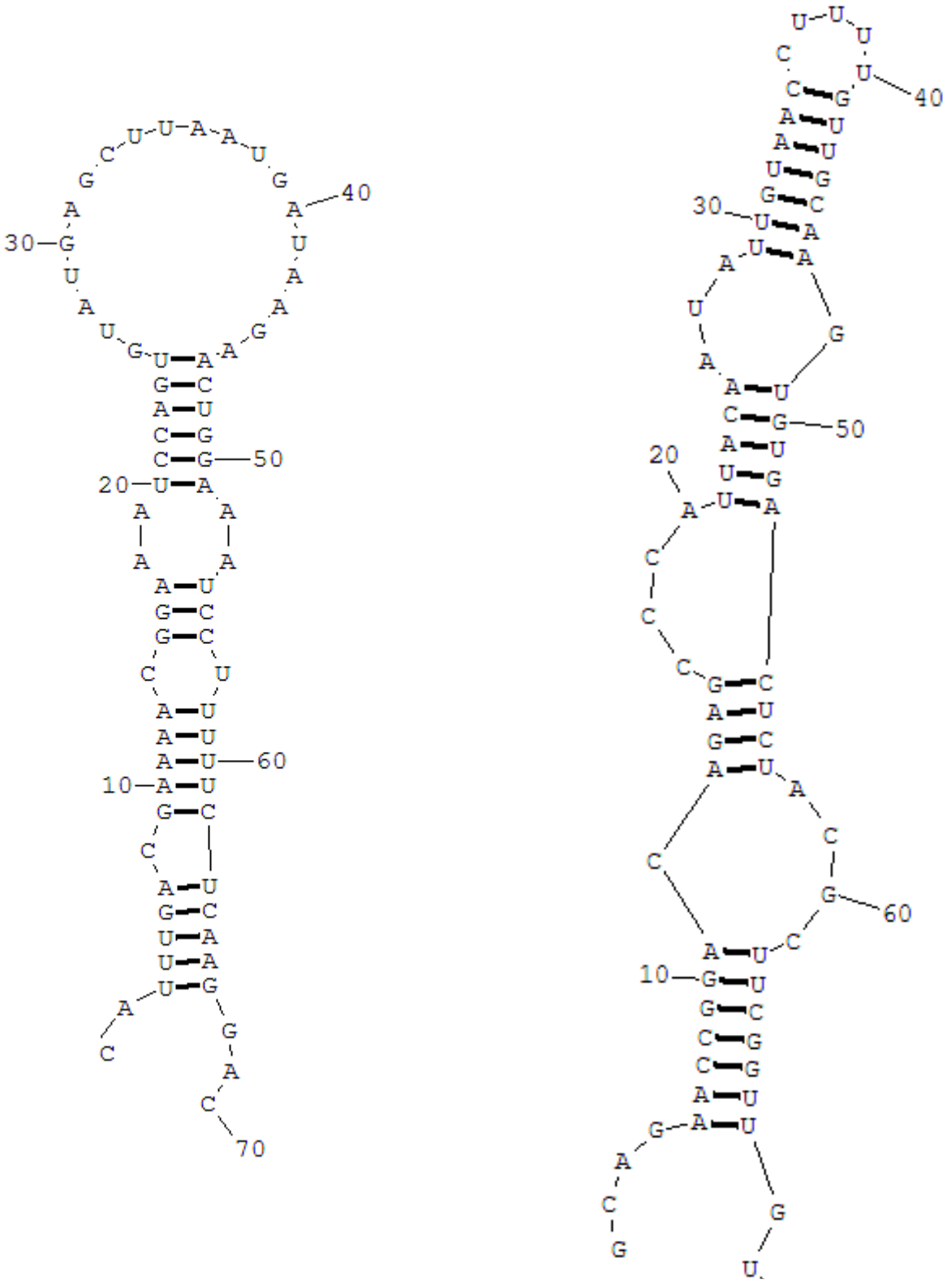


Fig. 4: Secondary structure of miRNA

First miRNA motif of precursor 2187, 1972 located within the stem region and it was selected, while second motif located in the loop region, due to this reason second motif was discarded. Similarly, precursor 232-233 yielded one raw miRNA and one reverse complementary raw miRNA. Precursor 881 yielded single miRNA with no complementary. Raw miRNAs were further processed through different filtering parameters (Rodriguez, et al., 2004) (minimum base paired, maximum unpaired residue, G: U pairs and %GC content etc.). After passing through these filtering parameters location of the miRNA (1020) was shifted and final miRNA were predicted. Finally P-2187, 1972 and 232 yielded HPV-1miR, HPV-1miR* and P-881 yielded HPV-1miR as final miRNA (miR) and their reverse complement (miR*). Similarly in the year 2001 *C. elegans* miRNAs were first predicted by using both bioinformatics and cDNA cloning techniques (Lee and Ambros ,2001). Later computational miRNA prediction methods has been evolved and found to be feasible approach (Lai et al., 2003). Computationally predicted miRNAs were predicted in *Arabidopsis thaliana* genome and the predicted miRNAs were also proved experimentally (Wang et al., 2004). For the first time in the year 2004 miRNAs detected in the EBV viral genome (Pfeffer et al., 2004). Several miRNAs were also predicted from the *Herpes* virus family and reported by Pfeffer et al., in 2005 (Pfeffer et al., 2005). Peffer *et al*., has introduced a novel miRNA prediction method based on defined set of properties from known miRNA precursor stems and subsequently trained a Support Vector Machine (SVM) to separate known pre-miRNAs from stem-loops and unlikely to code for miRNAs, these miRNAs were also proved experimentally (Pfeffer et al., 2004)**.** Hence, we can say our program works in similar with above said programs. The predicted miRNAs are in the process of experimental evaluation.

Table 1: Putative targets of predicted miRNAs and the genes down regulated

Putative miRNA sequence	Target Genes
>gi|9627100|ref|NC_001526.1|Precursor-972 Human papillomavirus – 16 (AGGUAGGUCGUGGUCAGCCA)	Bax protein, Cytoplasmic isoform Delta, Transforming Protein P21/H-RAS-1 (C-H-RAS), Tumor Necrosis Factor Receptor, Apoptosis mediating surface Antigen FAS (APO-1 Antigen).
>gi|9627100|ref|NC_001526.1|Precursor-972(complementary)Human apillomavirus – 16 (AUACACAGCGGCUGGUUUGGGCC)	Transforming Protein P21/H-RAS-1 (C-H-RAS), Caspase 10 precursor, Apoptotic protease MLH-4, FAS associated Death Domain protein Interleukin-1B-converting enzyme (FLICE2).
>gi|9627100|ref|NC_001526.1|Precursor-187 Human papillomavirus – 16 (CAACGAGCACAGGGCCACAAUA)	Bax protein.
>gi|9627100|ref|NC_001526.1|Precursor-32 Human papillomavirus – 16 (AACCGGACAGAGCCCAUUACAA)	Phosphatase and Tensin Homolog 2 (FRAGMENT), Transforming protein P21/H-RAS-1 (C-H-RAS).
>gi|9627100|ref|NC_001526.1|Precursor-32(complementary)Human papillomavirus-16 (CAAGUGUGACUCUACGCUUCGG)	Transforming protein P21/H-RAS-1 (C-H-RAS)
>gi|9627100|ref|NC_001526.1|Precursor-81 Human papillomavirus – 16 (UGACGAAAACGGAAAUCCAGU)	Bax protein

REFERENCES

Bonnet E., Wuyts J., Rouze P. and de Peer Y. V., 2004. Detection of 91 potential conserved microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes. PNAS, **101**(31): 11511-11516.

Cui C, Griffiths A, Li. G, Silva L.M, Kramer M.F, Gaasterland T, Wang X.J, Coen D.M., 2006. Prediction and Identification of Herpes Simplex Virus 1-Encoded MicroRNAs. J Virol., **80**:5499-5508.

Enright A.J., B. John T., Tuschl U. ,Gaul C., Sander D.S. Marks., 2003. MicroRNA targets in Drosophila. Genome Biology, **5**(1): R1.

Griffiths-Jones S., 2004. The microRNA registry. Nucleic Acids Res., **32**:109111.
http://bioperl.org/.
http://www.ncbi.nlm.nih.gov/.

Lai E.C.,Tomancak P.,Williams R.W., Rubin G.M., 2003. Computational identification of *Drosophila* microRNA genes. Genome Biology, **4**:R42.

Lee R.C and Ambros V., 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. Science, **294** (5543): 862-864.

Pfeffer S., Sewer A., Lagos-Quintana M., Sheridan R., Sander C., Grasser F.A. van Dyk L.F., Ho C.K., Shuman S., Chien M., Russo J.J., Ju J., Randall G., Lindenbach B.D., Rice C.M., Simon V., Ho D.D., Zavolan M., Tuschl T., 2005. Identification of microRNAs of the herpesvirus family. Nat Methods, **2**:269-276.

Pfeffer S., Zavolan M., Grasser F.A., Chien M., Russo J.J., Ju J., John B., Enright A.J., Marks D., Sa nder C., Tuschl T., 2004. Identification of virus encoded microRNAs. Science, **304**:734-736.

Rodriguez. A., Jones, S.G., Ashurst, J. L., and Bradley, A., 2004. Identification of Mammalian microRNA Host Genes and Transcription Units. Genome Reearch, **14**(10A): 1902-1910.

Wang X.J., Reyes J.L., Chua N.H., Gaasterland T.,2004. Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biology **5**(9):R65.

Yi R., Qin Y., Macara I.G., Cullen B.R., 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev., **17**: 3011-3016.

Zuker M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic sAcids Res., **31** (13): 3406-15.