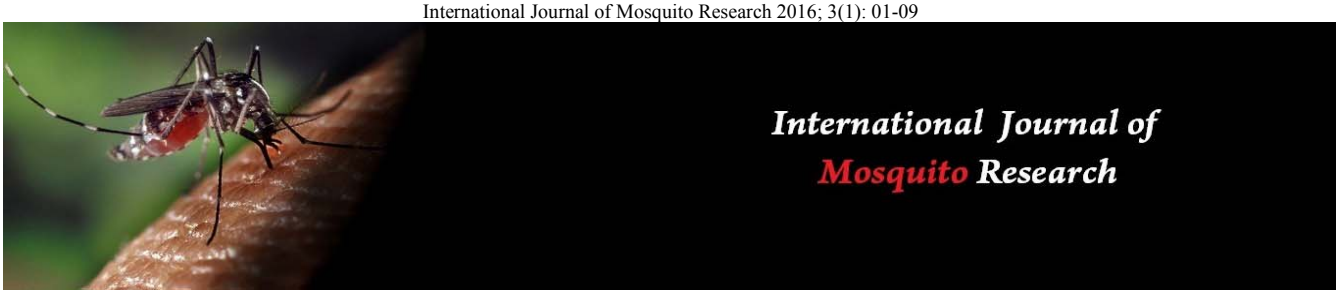

ISSN: **2348-5906**
CODEN: **IJMRK2**
IJMR 2016; 3(1): 01-09

Received: 21-10-2015
Accepted: 22-11-2015

Sujeewa Sampath L.D
Department of Chemistry, University of Colombo, Colombo 00300, Sri Lanka.

Ranil Samantha Dassanayake
Department of Chemistry, University of Colombo, Colombo 00300, Sri Lanka.

Kajan M
Department of Chemistry, University of Colombo, Colombo 00300, Sri Lanka.

Y.I Nilmini Silva Gunawardena
Molecular Medicine Unit, University of Kelaniya, Ragama, Sri Lanka.

Correspondence:
Ranil Samantha Dassanayake
Department of Chemistry, University of Colombo, Colombo 00300, Sri Lanka.

Engineering lhRNA based molecule to interfere replication of dengue virus in transgenic *Aedes aegypti* mosquitoes: Bioinformatics approach

Sujeewa Sampath LD, Ranil Samantha Dassanayake, Kajan M, YI Nilmini Silva Gunawardena

Abstract
As a preliminary step towards the development of transgenic *Aedes aegypti* mosquito's resistance to Sri Lankan dengue viruses (SLDENV), the identification of previously unpublished DNA sequence of the construct, AeCPA/Mnp+/i/Mnp-/svA, was done using bioinformatics tools. Furthermore, a long hairpin RNA (lhRNA) molecule effective against all four SLDENV serotypes was designed. If this lhRNA molecule to be cleaved by dicer at the correct site, the 100% of siRNA molecules produced are going to be effective against each serotype of SLDENV, on the contrary, if it is to be cleaved at random sites, which is the worst case, 52.63% of total siRNA molecules produced are going to be effective against each SLDENV serotype. In addition, an alternative promoter sequence, *Ae. aegypti* vitellogenin A1 (AeVtA), and a construct functionally similar to MosI helper plasmid, pKhsp82MOS (commercially unavailable) used in micro injection procedure of *Ae. aegypti* were also proposed in this study.

Keywords: Dengue, RNA interference, siRNA, long hairpin RNA, bioinformatics

1. Introduction
Sri Lanka is one of the countries, which has the highest suitability for dengue transmission. Furthermore, it holds the 12th position among the 30 most highly endemic countries/territories to dengue disease [1]. The dengue disease was prevalent now in all districts of Sri Lanka [2, 3]. An average of about 34,774 patients were reported each year to Epidemiology Unit of Sri Lanka (EUSL) from the year 2010 to 2013, and the monthly average is about 3035 during the same period [4]. The record-breaking patient number (47502) in 2014 indicated that there is a high risk of hyper endemic transmission of dengue disease and multiple DF/DHF outbreaks, which will cause a lot of human suffering in Sri Lanka [4, 5].
As conventional vector controlling strategies are failing to perform, it is required to find novel strategies to keep the dengue transmission under control. Development of transgenic *Aedes aegypti* mosquitoes with a resistance to the dengue virus is one such strategy, which has not yet been tested in Sri Lanka. As transgenic *Ae. aegypti* mosquitoes with a resistance to dengue type 2 has been already developed, and it successfully passed the dengue challenge experiments [6], the DNA sequence constructed by those researchers needed to be found using bioinformatics tools and constructed.
This construct contains several elements from different organisms. For example, left/right arms of mariner MosI transposable element (TE) (ma. left/ ma. right), *Ae. aegypti* carboxypeptidase A (AeCPA) promoter sequence, cDNA sequence of prM (or Mnp) protein encoding region of the strain of DENV-2 in sense (Mnp+) and antisense (Mnp-) orientations, minor intron of *Ae. aegypti sialokinin I* (i), polyadenylation signal of Simian virus 40 VP1 gene (svA), enhanced green fluorescence protein coding gene (EGFP), and an artificial eye specific promoter for mosquitoes, 3xP3 [6].
The AeCPA promoter of the construct mentioned above is activated upon ingestion of a blood meal, and it strongly drives the expression of Mnp+/i/Mnp- sequence in the midgut of female transgenic *Ae. aegypti* mosquitoes starting from 8 hours (h) post blood meal (pbm) to 48 h pbm [6]. As Mnp+ and Mnp- regions are in sense and antisense orientations respectively, they are complementary to each other. As a result, it forms a short hairpin RNA (shRNA) in the midgut epithelial cell cytoplasm of *Ae. aegypti* resulting in the triggering of the RNA interference (RNAi) mechanism in the midgut cells. Here, dicer enzyme, which is a dsRNA specific endonuclease, cleaves the double-stranded shRNA molecule and generates a pool of

siRNA molecules in which each siRNA has a length of ~21-23 bp. These double-stranded siRNAs are then incorporated into Argonaute-2 (Ago2) protein in the RNA induced silencing complex (RISC). Consequently, RISC is activated. Then, RISC unwinds the double stranded siRNA incorporated into Ago2 and keeps one strand acting as a RISC-targeting co-factor. RISC having single stranded antisense siRNA is capable of binding to the complementary prM (sense) sequence of DENV-2 in the midgut of the mosquitoes if DENV-2 genomic RNA (gRNA) is present in the cytoplasm of the midgut epithelial cells of *Ae. aegypti*. The endonuclease activity of RISC causes a single-site cleavage in the prM region of DENV gRNA and the viral RNA will be degraded [6, 7]. If DENV-2 is ingested with a blood meal, it can be targeted effectively using RNAi mechanism at the time when the virus releases the viral RNA inside the midgut, and the whole process of the viral replication cycle can be knocked out at that spot. The results of the research conducted previously had shown that the resistance to DENV-2 could be seen even after 14 days since having a blood meal [6].

In foregoing research work, finding of the sequence of construct AeCPA/Mnp+/i/Mnp-/svA was achieved using bioinformatics tools and a novel effector sequence, long hairpin RNA (lhRNA) effective against all four SLDENV serotypes was designed by identifying the best siRNA target sites in the SLDENV genomes of each serotype and clustering them and adding a loop sequence. Replacing the Mnp+/i/Mnp- sequence of previously mentioned construct [6] with the designed lhRNA sequence and germ-line transformation of the novel construct can produce a transgenic *Ae. aegypti* resistance against all four SLDENV serotypes.

2. Method

The genome sequences derived from Sri Lankan dengue viral strains and other related sequences mentioned elsewhere deposited in various databases from year 1983 to 2014 were retrieved and used in this study.

2.1. Obtaining sequence corresponds to minor intron of *Ae. aegypti* sialokinin I

The complete cDNA sequence (GenBank (gb): AF108099.1) of *Ae. aegypti* salivary vasodilatory protein precursor gene was obtained from GenBank of National Center for Biotechnology Information (NCBI). The region corresponds to minor intron of *Ae. aegypti* sialokinin I was obtained using the forward and reverse primers published [6].

2.2. Obtaining sequences correspond to Mnp+ and Mnp- region of Sri Lankan Dengue virus serotype 2 (SLDENV-2)

All available complete SLDENV-2 genome sequences were retrieved from the NCBI Virus Variation database [8]. The retrieved sequences (gb: FJ882602, GQ252676 and GQ252677) were aligned. Then, the primer pairs designed to amplify the Mnp+ and Mnp- regions of Jamaica 1409 infectious dengue viral clone cDNA [6] (gb: M20558.1) were mapped to the aligned sequences using the Map primer pairs tool of BioEdit software and selected the region corresponding to Mnp+ and Mnp- of SLDENV-2s and corrected mismatches in the primer annealing to obtain the latter region of SLDENV in PCR.

2.3. Obtaining sequence corresponds to the AeCPA promoter of *Ae. aegypti* mosquito

The whole genome sequence of *Ae. aegypti* (gb: AAGE02019343.1), complete cDNA sequence of AeCPA gene (position 1 to 2398 of gb: AF165923.1), final transcript of AeCPA gene (position 690 to 1973 of gb: AF165923.1) as well as the reverse translated amino acid sequence (gb: AAD47827.1) of the AeCPA were aligned. The promoter sequence (-4 to -1126) [6] of *AeCPA* gene was then identified by recognizing the transcription start site (+1) of the whole genome sequence of *Ae. aegypti* (gb: AAGE02019343.1). To obtain this region a primer pair, AeCPAF (forward primer) and AeCPAR (reverse primer), was newly designed.

2.4. Obtaining sequence corresponds to the AeVtA promoter of *Ae. aegypti* mosquito

Complete DNA sequence of *Ae. aegypti* VgA1 gene (gb: L41842.1) and its coding region (adding up the regions of 2091-2121,2192-7405 and 7463-8664 of gb: L41842.1), final transcript sequence (collectively the regions, 2091-2121,2192-7405 and 7463-8664 of gb: L41842.1) and reverse translated amino acid sequence (gb: AAA99486.1) of AeVtA were aligned. A 1.8 kb upstream region to transcription start site (+1) was taken as the AeVtA promoter according to previously published work [9]. A primer pair, AeVtAF (forward primer) and AeVtAR (reverse primer), was designed with the restriction sites of *Sac*I and *Sma*I to replace AeCPA promoter to drive alternate expression of designed shRNA/lhRNA constructs in the fat body of female *Ae. aegypti* mosquitoes [9].

2.5. Obtaining sequence corresponds to svA

The svA primers designed by *Franz et al.* [6] were aligned with pGL3-Basic [10] and pGL3-hsp82 [11] vector sequences found in Addgene and the svA region was identified.

2.6. Obtaining sequence corresponds to MosI transposase

The sequence of MosI transposase was obtained from the coding sequence of MosI transposase gene in the vector, pCFJ601-Peft-3 found in the Addgene website [12]. A primer pair, MosF (forward primer) and MosR (reverse primer), with *Nco*I and *Fse*I restriction recognition sites in forward, and reverse orientations, respectively were designed to obtain this region in PCR.

By digesting the PCR product with *Nco*I and *Fse*I restriction enzymes and cloning it to the vector, pGL3-hsp82 (Addgene) digested with same enzymes, a helper plasmid similar to pKhsp82MOS plasmid (which was not commercially available) can be generated to be used in microinjection procedure described in the previously published work [13], to transform the embryos of *Ae. aegypti* along with the plasmid containing the lhRNA sequence having resistance to DENV, developed by us.

2.7. Development of an lhRNA molecule, which is effective against all four serotypes of Sri Lankan dengue viruses (SLDENV)

The full-length genomic sequences of all four SLDENV serotypes reported from Sri Lanka were downloaded from Virus Pathogen Database and Analysis Resource [14], and they were aligned using ClustalW of Bio Edit software for conserved regions. The minimal length requirement of a siRNA molecule is 21 nucleotides (nts). Since, there was no 21-nt region conserved in all four serotypes of SLDENV, the sequences of each serotype was separately aligned for homologous regions. Due to high sequence identity among the viruses of particular serotypes, the SLDENV genome sequence reported recently in each SLDENV serotype was selected and

subjected to search for effective regions for siRNA target sequences using InvivoGen siRNA Wizard v3.1 online software [15]. In doing so, the SLDENV-1 (gb:KJ726662 of year 2012), SLDENV-2 (gb:GQ252677 of 2004), SLDENV-3 (gb: NC_001475 of 2000) and SLDENV-4 (gb:KJ160504 of 2014) full-length gRNA sequences from GenBank were separately submitted to the InvivoGen siRNA Wizard v3.1 software and effective regions to produce siRNA sequences were obtained using standard search option for each serotype. These effective regions identified from each serotype of SLDENV were submitted to *Ae. aegypti* transcript database & EST database of Vector base website [16], separately (the parameter "word size" was set to 11) to check whether siRNA sequences have any effect on the transcriptome of *Ae. aegypti*. The siRNA sequences which had less than 16-nt match with the *Ae. aegypti* sequences in databases mentioned above were selected. Among the chosen sequences, ones having GC nucleotide percentages (GC%) between 30-50 and having AA or A nucleotides at the 3' ends of the sequences were selected as better siRNA candidates. Then, these selected sequences were used to carry out a BLAST search over the SLDENV sequences in databases. Although a contiguous 16-nt match (out of 21-nt long siRNA molecule) with the dengue virus sequences in databases was sufficient for RNAi mechanism, due to high sequence homology within an SLDENV serotype, siRNA sequences having 18-21-nt matching with all known sequences within a particular serotypes were selected as the best sequences to design the lhRNA sequence. A long hairpin RNA (lhRNA) was subsequently designed in the positive sense manner to have 10 siRNA sequences from each of different dengue serotypes and amounting about 40 siRNA, placing them one after the other in a way that each siRNA sequence ends with A or AA nucleotide/s at the 3' end of that siRNA. In designing lhRNA, the sequence of *sialokinin* I intron [6] was used to have a loop structure and the sequence of the design was then submitted to 'Design hairpin insert' option of InvivoGen siRNA Wizard v3.1 online software to obtain the complete sequence of the lhRNA molecule comprising positive sense DNA sequence, loop DNA sequence and negative sense DNA sequence. This was again tested for homology with *Ae. aegypti* contig and transcript databases in the Vector base website [16]. Finally, replacing the Mnp+/i/Mnp- sequence of the previous construct [6] with the new lhRNA's DNA sequence designed in this study, a novel DNA construct, which would give the resistance to the primary dengue vector in Sri Lanka, *Ae. aegypti*, against all serotypes of SLDENVs can be produced.

3. Results & Discussion

3.1. Minor intron of *Ae. aegypti sialokinin* I

The *sialokinin* I intron sequence supports the complementary base pairing of the lhRNA sequence by acting as a loop sequence. The *sialokinin* primers reported previously [6] was found to anneal to the sites of the *Ae. aegypti* salivary vasodilatory protein precursor gene's complete cDNA sequence (gb: AF108099.1) containing minor intron of *sialokinin* I. The forward primer containing *Not*I and *Apa*I restriction sites annealed to 56 bp-74 bp region and the reverse primer containing *Kpn*I and *BamH*I sites annealed to 103 bp-126 bp region of the above-mentioned sequence and the region between these two primers annealing was taken in designing lhRNA. The PCR product resulting from these primers will have the size of 118 bps and the sequence is shown in Fig 01 (see supplementary material).

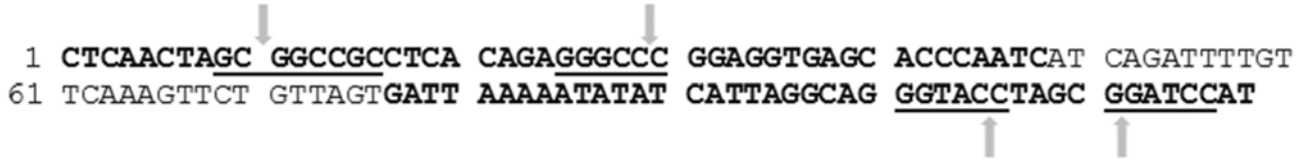

Fig 01: Ae. aegypti sialokinin *I intron sequence (5' to 3').* Forward and reverse primers are shown in bold black letters. The underlined sequences are restriction recognition sites (REs), Not*I*, Apa*I,* Kpn*I and* BamH*I* respectively from the 5' end of the sequence. Each arrow indicates the cleavage site within the *REs.*

3.2. Mnp+ and Mnp- regions of SLDENV-2

The Mnp+ and Mnp- regions of the shRNA sequence [6] act as effector sequence, which is then incorporated into the RISC complex enabling the RISC to recognize the DENV genome in the *Ae. aegypti* cells. The primers designed to amplify the Mnp+ and Mnp- regions of Jamaica 1409 infectious dengue virus [6] were found to match the corresponding regions of Mnp+ and Mnp- of SLDENV-2 having full-length genome sequence gb: GQ252677 in GenBank, helping us to the identification of regions corresponding to Mnp+ and Mnp- from SLDENV-2. Mnp+ and Mnp- forward and reverse primers annealed to 426 bp-445 bp and 695 bp-715 bp regions, respectively to the aligned SLDENV-2 sequences and the length of the sequences of Mnp+ and Mnp- regions with respect to the latest SLDENV-2 strain (gb: GQ252677) are 316 bp and 307 bp, respectively (Fig 02; supplementary material) having significant sequence homology to the said serotype (only five mismatches among GQ252677, FJ882602 and GQ252676), indicating the suitability of latter regions to be included as an effector sequence in engineering the lhRNA molecule in this study [6].

3.3. AeCPA promoter of *Ae. aegypti* mosquito

The AeCPA promoter is activated upon ingestion of a blood meal, and it strongly drives the expression of the effector sequence in the midgut of female transgenic *Ae. aegypti* mosquitoes [6]. Following the multiple sequence alignment of the cDNA sequence of *AeCPA* obtained from GenBank, the translation initiation codon of AeCPA (Fig 03, see supplementary material) could be identified. Identification of the transcription initiation site and the promoter region cannot be achieved directly in this analysis due to the non-availability of sequence beyond the translation initiation site. Therefore, these cDNA sequences were aligned with the *Ae, aegypti* genome sequence (gb: AAGE02019343.1) and identified the potential transcription initiation (+1) site of the AeCPA gene (Fig 03, see supplementary material). Following identification of transcription initiation site, the promoter sequence (-4 to -1126) of the AeCPA gene [6] could be identified and then primers were designed accordingly to obtain the latter region of AeCPA in PCR. They were the forward (AeCPAF) and reverse (AeCPAR) primers having sequences of 5'-CGAGCTCGTTGTAGATAACTTCAAGGAA-3' with a *Sac*I

restriction site (underlined) and 5'-TCCCCCGGGGGGTCACATTGGTCAGTCACAG-3' with a *Sma*I restriction site (underlined), respectively. These forward and reverse primers would anneal to 315 bp-335 bp region and 1419 bp-1438 bp region of *Ae. aegypti* genome (gb: AAGE02019343.1). The sequence obtained using these primers has the size of 1142 bp (Fig 04, see supplementary material).

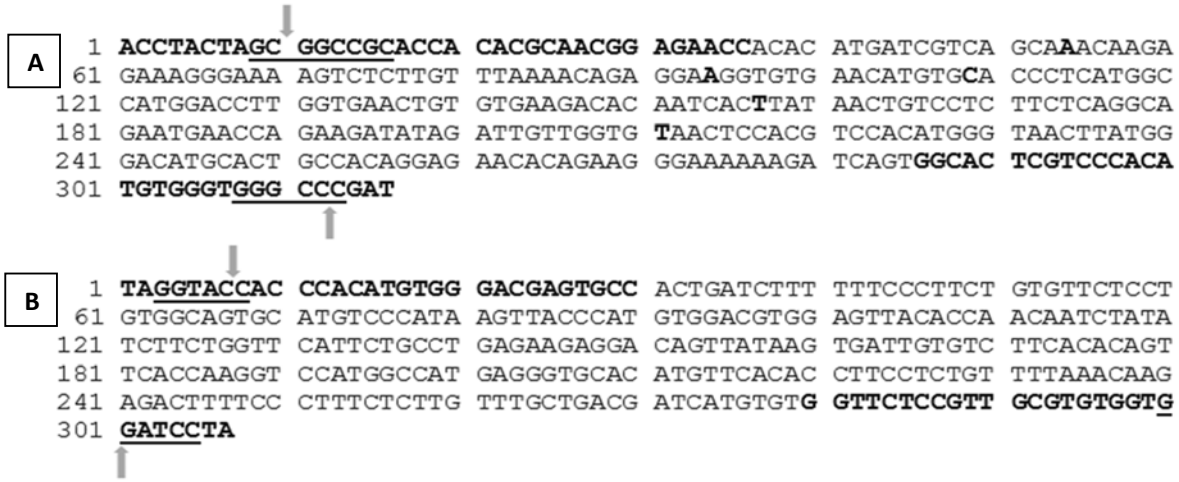

Fig 02: Mnp+ (A) and Mnp- (B) regions (5' to 3') selected from SLDENV-2 (gb: GQ252677). Mnp+ and Mnp- forward and reverse primers are shown in bold black letters. The underlined sequences are REs, *Not*I and *Apa*I respectively from the 5' end of the Mnp+ sequence and *Kpn*I and *BamH*I respectively from the 5' end of the Mnp- sequence. These sequences had a 100% identity with another SLDENV-2 strain (gb: GQ252676). But there were a few mismatches in the sequence when these two sequences were aligned with that of the Mnp+ region of the SLDENV-2 strain found in 1996 (gb: FJ882602). The sites of sequence variations, base 54th position (A), 94th position (A), 109th position (C), 157th position (T) and 211th position (T) are shown in bold black letters between the primer sequences of Mnp+ sequence. Each arrow indicates the cleavage site within the REs.

Fig 03: The multiple sequence alignment showing the transcription initiation site (+1) and the translation initiation site of the AeCPA gene. The alignment of whole genome sequence of *Ae. aegypti,* gb: AAGE02019343.1 (A), complete cDNA sequence of AeCPA gene, gb: AF165923.1 (B), final transcript of AeCPA gene, gb: AF165923.1 (C), and the reverse translated amino acid sequence, gb: AAD47827.1 (D) of the AeCPA is shown. The numerical values at the top of the figure show the position of each nucleotide in the sequence, AAGE02019343.1. Hence, it can be seen that the 1442th position, of genome sequence, AAGE02019343.1 of AeCPA gene is the transcription initiation site of the gene and the 2132th position of the sequence is the translation initiation site.

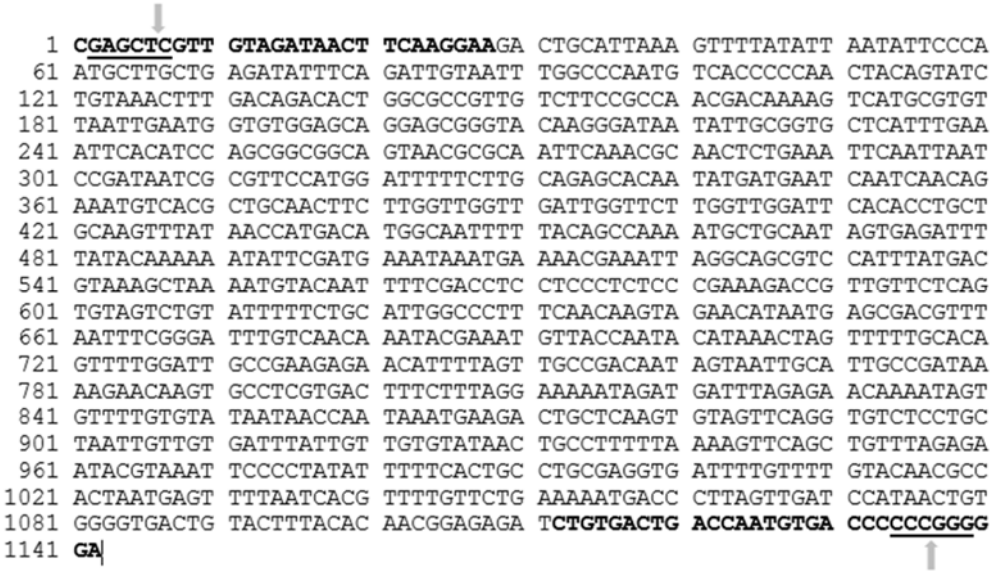

Fig 04: *Ae. aegypti* Carboxypeptidase A promoter sequence from -4 to -1126 (5' to 3'). The AeCPAF and AeCPAR primers are shown in bold black letters. The underlined sequences are REs, *Sac*I and *Sma*I respectively from the 5' end of the sequence. Each arrow indicates the cleavage site within the REs

3.4. AeVtA promoter of *Ae. aegypti* mosquito

The AeVtA promoter is among one of the alternative tissue-specific promoters suggested by the Franz *et al*. to drive the transcription of effector molecules in *Ae. aegypti*. Following the multiple sequence alignment of the cDNA sequence of AeVtA obtained from GenBank, the translation initiation codon of AeVtA (Fig 05, see supplementary material) could be identified. As the transcription initiation site (+1) could not be found in these sequences, the cDNA sequences were aligned with the complete DNA sequence of *Ae. aegypti* VgA1 gene (gb: L41842.1) and identified the potential transcription initiation (+1) site of the AeVtA gene (Fig 05, see supplementary material). Following accurate identification of the transcription initiation site, the promoter sequence (1.8 kb fragment of Vg gene's 5' regulatory region [9]) could be identified and then primers were designed accordingly to obtain the latter region of AeVtA in PCR. They were the forward (AeVtAF), and reverse (AeVtAR), primers having sequences of 5' CGAGCTCGGCCATCTGATGGCTAAATTT-3' with a *Sac*I restriction site (underlined) and 5'-TCCCCCGGGGGACCGTCTGTGGCCATTGGAA-3' with a *Sma*I restriction site (underlined), respectively. These forward and reverse primers would anneal to 216 bp-235 bp region and 1996 bp-2015 bp region of *Ae. aegypti* VgA1 gene sequence (gb: L41842.1) respectively. The sequence (1819 bp) corresponds this region is shown in Fig 06 (see supplementary material).

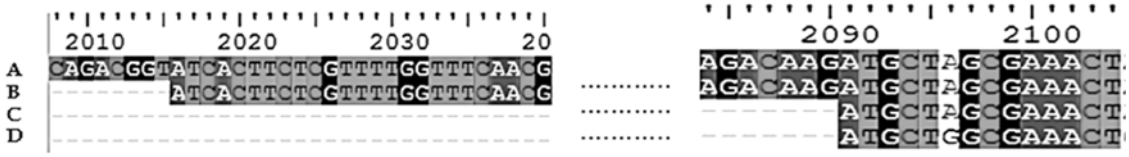

Fig 05: The multiple sequence alignment showing the transcription initiation site (+1) and the translation initiation site of the AeVtA gene. The alignment of complete DNA sequence of *Ae. aegypti* VgA1 gene, gb: L41842.1 (A), the complete cDNA sequence of the very sequence (B), the final transcript of the very sequence (C), and the reverse translated amino acid sequence, gb: AAA99486.1 (D) is shown. The numerical values at the top of the figure show the position of each nucleotide in the sequence, gb: L41842.1. Hence, it can be seen that the 2016th position, of gene sequence, gb: L41842.1 of AeVtA gene is the transcription initiation site of the gene and the 2091st position of the sequence is the translation initiation site.

```
   1 CGAGCTCGGC CATCTGATGG CTAAATTTCG AATTTATCGG CTAGATTAAT GATAGATCTA
  61 AAGCTGCTGA AAAACTTTGC CGAATACGTC ATCTTTCTGA GAGACCAGGA TACTGAGTTA
 121 TTTTTAAAAC AAAGGTTCCG TGCTCACTAA TGCCGCATGG TGGCTAAATG GCTTTTATAC
 181 TGAAAACGCT TGAGCTTTGC CTAACACGTC GTTTTTTTAG GTAGTGAGGA ACCTGATATA
 241 ACAGAATAAA ATCGATACAC CAGCGCCGCC TGGATTCCGC ATATCTATTA CCACCGCATG
 301 TTAGCCCTTT ATCCAAATAA AAAAAATTAC TGAACATCAT GACTCTCTAT GTTGATGACT
 361 TTTCAATAAG CATTTATAAT AACGGTCGTC AATCTCAATT TCATTCGTAT ATATAACCAT
 421 CGTAGTACGA ATGGTCATAA AAAGAGGTTG AAGTGTGTTT GGATTCTCCC CAGACAAAAA
 481 ATCGAAATGA AAAACTGAAA CCCATTTTGA CAATCGTCGG AAAGGTCTTT CTGTTTAGTT
 541 CACTGTAACA AAATGCAATC CAAAGATATG AAAGCTTTGA AAACGGCGAA AAGTTATTAT
 601 ATCTACCGTT TTATTACGGA AAGCTTCTTA TTCCGGACAC TCTACTTTGT ATGAGAAACA
 661 TTTCATATGA GATGTTTCAA TTTTTGCTGT TCAAAAGTTC ACACTTTTGA GGTTCATTTT
 721 AATCAAAATT TTCTATAGAT ATCTATGAAA ATTTATAATA CTCAACTGCT TTAGACGCCT
 781 CTTTAGAGGC TTGCCAATTT CAAATAATGA TTTGAGCTTG TCTTTTTTAT GATTTTCATG
 841 AGCTGTCCGG AATTTGAATC AAAGTGTCCG GAATATGGGG CAAAAGTAAG GAAGCGTCCG
 901 GAATAAGAAT CGTGAAAAGT CCACACATTT CTATTTATTG AAAATCATTC ATGTATCAGA
 961 AACAAAATTT TCACCCACCA TTCGAAAGTT AAGTGTTTTG AAGGCTTAGA GCCGAGGAAT
1021 CATACAGATT GATTTATTTT ATATGCTTCC TGATGGGTTA TACTTCATTG AGGCCTTAAG
1081 TGTCCGTAAT ATGAATCAAA ACGGTACATA TGTCATGGCA ACATTTTCCG AAACCTGGAG
1141 ATAAATTACA GAGCCCACTC AGTATCGCAG ATGCCAACAG TGTGGCTATG GTACCAAACA
1201 TCGTGCAAAA GTTTGTCATT TTACTCAATA ACTGACTTGT ATAATTAGTT TAAGCTTATT
1261 CAAATGCTTT TTACAAGCTG ATGATTTCAA TTTGCCTGTG AAAATTCTGA ACTGTTGCGG
1321 CAAATGAAAT GCAATCGATA GAAACAAATA ATGTAGCAAT AGCAAAAAAT AATCATTTTT
1381 TGCTTATCTT ACTATCTTCA ATTCACATCT GTAGTCTCAA TTGAATAATC TGGAATCCAT
1441 TGCAAGCTAA GTAAATTCAC GTGTGACCTA GCGGGAGGCC AATGGTCGAG TGAATCTTTA
1501 TTTCTTGAAT GGCAGAAACG ATGCCATGAA TCAAATCCAG GATGCGAAAC CGATGCACAA
1561 GAATACCAAT ACGAAAACTA ATCCCAATAC AACGATCACC AGGTGCATCG ATGCGACCGC
1621 GATTACTTTG TTACCATCTG TTGTGCTGCG GTTATCGCTT TTTCGATTAG AAGGCGAACG
1681 CTGAACCGAT CGTATGCTTA TCATCGCGAA CGAAAGGATG CTGTGAATCA CTGCTGATGG
1741 GGGCAAAAAT CTACGAAAAT GTAAGCAATC ACTTTCAAAT ATAAACCCTT CCAATGGCCA
1801 CAGACGGTCC CCCGGGGGA
```

Fig 06: *Ae. aegypti* Vitellogenin promoter sequence *(5' to 3')***.** The AeVtAF and AeVtAR primers are shown in bold black letters. The underlined sequences are REs, SacI and SmaI respectively from the 5' end of the sequence. Each arrow indicates the cleavage site within the REs.

3.5. The sequence of svA terminator

The svA is a terminator sequence of Simian virus 40 VP1 gene and it functions as a transcription termination signal. Engineering this sequence to the 3' end of lhRNA molecule ensures the correct transcription termination of the effector sequences of the latter molecule. The sequence corresponding to this region was obtained from the vector (pGL3-hsp82 and pGL3-Basic) sequences deposited at Addgene [10, 11], using the primers published in *Franz et al*. The forward primer containing *BamH*I restriction site anneals to 2736 bp-2756 bp region and the reverse primer containing *Xba*I site anneals to 2935 bp-2957 bp region of the above-mentioned vector sequences and the size of the fragment that could result from these primers would be 240 bp (Fig 07; supplementary material). The advantage of using pGL3-hsp82 vector over pGL3-Basic vector used in previous work [6] is that the former vector could provide both svA region and the hsp82 promoter. The svA is required for the termination of the transcription of lhRNA molecule while hsp82 promoter is required to drive the expression of MosI transposase coding region in helper plasmid. If the latter vector is used in such attempt, it could only provide the svA sequence. Thus, usage of pGL3-hsp82 vector for the construction will have a clear advantage in designing lhRNA molecule.

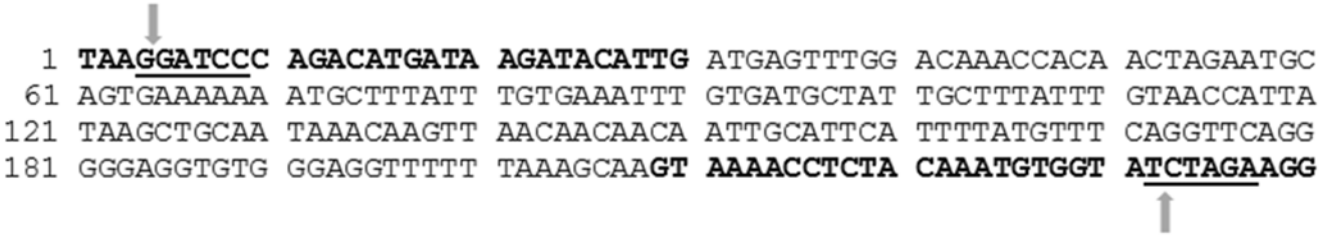

Fig 07: svA sequence (5' to 3'). svA forward and reverse primers are shown in bold black letters. svA region of both vectors, pGL3-hsp82 (Addgene) and pGL3-Basic (Addgene) could be amplified by the primers indicated. The underlined sequences are REs, *BamH*I and *Xba*I respectively from the 5' end of the sequence. Each arrow indicates the cleavage site within the REs.

3.6. The sequence of MosI transposase

The MosI transposase enzyme derived from *Drosophila mauritiana* mediates integration of gene cassette containing MosI transposable element (TE) to *Ae. aegypti* genome. MosI TE have a higher transformation efficiency than the PiggyBac TEs and these TEs have been used for the construction of shRNA/lhRNA sequence. However, the MosI transposase coding helper plasmid needed for the transformation of the *Ae. aegypti* embryos is not commercially available. Hence, designing a helper plasmid with MosI transposase was necessary. In doing so, the primers were newly designed to amplify MosI transposase coding gene and the sequence of the forward (MosF), and reverse (MosR), primers were 5'-CATGCCATGGCATGATGGGAAAGATTCAGAAGGTC -3' with an *Nco*I restriction site (underlined) and 5'-GTCGGCCGGCCATACTATTCAAAGTATTTGCCGTC-3', with an *Fse*I restriction site (underlined), respectively. These restriction sites were engineered to fuse MosI transposase coding sequence of vector, pCFJ601-Peft-3 to the vector, pGL3-hsp82 digested with the same restriction enzymes. The restriction digestion of the vector, pGL3-hsp82 will result in the removal of luciferase coding region under the control of hsp82 promoter in the vector, pCFJ601-Peft-3 (Addgene). Fianlly, the cloning of the PCR product containing MosI transposase sequence, cleaved with restriction enzyme mentioned above, to this vector, a novel vector that expresses MosI transposase under the control of hsp82 promoter can be produced. The primers MosF and MosR will anneal to 2637 bp-2658 bp and 3328 bp-3350 bp region, respectively, of the relevant sequence of pCFJ601-Peft-3 MosI transposase vector (Addgene) and the sequence that can be obtained in PCR is shown in Fig 08 (see supplementary material).

Fig 08: MosI transposase enzyme coding sequence. The MosF and MosR primers are shown in bold black letters. The underlined sequences are REs, *Nco*I and *Fse*I respectively from the 5' end of the sequence. Each arrow indicates the cleavage site within the REs.

3.7. The lhRNA molecular design

The Fig 09 (see supplementary material) shows the lhRNA molecule, having siRNA sequences selected from each SLDENV serotype.

The minimal number of nucleotide matches required for siRNA and target for the cleavage is 16-nt and some SLDENV-2 specific siRNA sequences had 17 and 20 numbers of matches with the target, dengue virus 2 isolate DENV-2/LK/BID-V2416/1996, complete genome (gb: FJ882602.1) sequence. However, these numbers are greater than 16-nts. Hence, the siRNA molecules selected from the above genome will be able to trigger RNAi pathway against such SLDENV-2s as mentioned elsewhere.

In designing the lhRNA molecule, the effective regions for siRNA sequences in 5' untranslated region (UTR) and 3' UTR of serotypes were not taken into account because these regions might contain binding sequences for regulatory proteins, which might affect the accessibility of the RNA target sequence to the RISC complex [17]. The size of each siRNA used in designing the lhRNA was 21 nts because the double-stranded lhRNA to be cleaved by dicer protein at 21 to 23 nucleotide intervals [18, 19]. The purpose of selecting siRNA sequences to have A or AA nucleotides at the 3' end of the potential siRNA sequences generated by InvivoGen siRNA wizard is the sequence preference of dicer protein to cleave the lhRNA molecule at U nucleotide [18]. As each siRNA in the lhRNA molecules designed to have A or AA nucleotide at 21-nt intervals, it can be assumed that this molecule will get cleaved by dicer protein of *Ae. aegypti* at these sites in generating each assembled siRNA sequence (Fig 09) of the lhRNA molecule to exert RNA interference to dengue virus genome. Each siRNA molecule generated in this manner will have at least 90% homology to the sequences of particular SLDENV serotype. Further, if the dicer enzyme to cleave the lhRNA molecule in a random fashion, base pair homology of resulted siRNA molecule with the target SLDENV genome would vary and the sequences having base pair homology at a stretch equal to or greater than 76.19% (at least 16 nt homology of a 21 nt long siRNA) to the SLDENV genome would exert RNAi effect [17]. However, even in the worst-case scenario, it could be calculated that there would be 52.63% probability that siRNA molecules having base pair homology at a stretch at least with 76.19% to the target site of the SLDENV genome of the corresponding serotype. Hence, RNAi mediated dengue viral genome destruction would still be possible even if the dicer enzyme cleaves the lhRNA molecule randomly.

Further, the siRNA molecules having GC content of 30-50% were selected out of the siRNA sequences generated by siRNA Wizard software, as these siRNA molecules were reported to be more active in the cleavage than those with a higher GC content [19]. A BLAST search was carried out using the sequences that encode for each siRNA molecule selected against *Ae. aegypti* transcript and EST databases to choose the siRNA sequences that would minimally interfere with the *Ae. aegypti* transcripts and any hit having greater than 16-17 base pairs of homology with transcripts at a stretch to query siRNA sequence was excluded [17] and selected only the sequences having less than such a homology assuming that these siRNA would not interfere with host cells' gene expression.

The second BLAST search was then carried out against the genomic sequences of all reported DENV strains of serotypes in databases using siRNA sequences selected in above step to check whether these DENV strains could be silenced using designed siRNA molecules. Analysis indicated some of these siRNA molecules have 100% sequence homology with the complementary regions genomes of DENV in a particular serotype. These siRNA sequences together with siRNA sequences that showed at least 85.7% (18 nt homology out of 21nt present in siRNA sequence) homology within the strains of serotype were selected in designing the effector lhRNA sequence as 16 contiguous base homology was sufficient to carry out RNAi mediated gene silencing.

The third BLAST search was carried out with the siRNAs identified to be effective for particular serotype against the genomic sequences of DENV strains of other serotypes to check whether there are any siRNA molecules that would be effective against more than one DENV serotype. Analysis revealed that none of the siRNAs effective against other serotypes.

Finally, a search was carried out against the sequences in the Vector Base website containing *Ae aegypti* coding and genome sequences using lhRNA sequence (Fig 09) as a query sequence to check whether the final lhRNA molecule to have any region with 16 contiguous base pair homology to the transcriptome of *Ae. aegypti* as results of compiling individually designed siRNA sequences into lhRNA molecule. This analysis indicated that there is no sequence identity having over more than 16 nt at a stretch between the lhRNA and *Ae. aegypti* sequences other than the region that corresponds to *Ae. aegypti sialokinin* I intron sequence (to 799-859bp of the lhRNA sequence) that was included in designing the loop region of lhRNA, which does not have any effect on silencing gene expressions. Therefore, it could be assumed that lhRNA designed in this study is effective only for silencing DENV transcripts but not for silencing *Ae. aegypti* transcripts.

The size of lhRNA molecule designed in this study is 2117 bp and require synthesis. However, just by synthesizing the bases 1 to 798 (Fig 09) and obtaining complementary region of this sequence in PCR by designing primers appropriately together with PCR amplifying AeCPA, and svA regions and combining them restriction digestion followed by ligation can cost-effectively generate IhRNA molecule.

In this study, attempts were not made to find the DNA sequences corresponding to 3xP3/EGFP/svA region [6] because the sequences for these regions could directly be obtained from the plasmid, pMos-3xP3DsRed-attP available in Addgene. The only difference that was found between DNA construct published [6] and the region above-mentioned was that former encodes for a protein which emits green fluorescence and the latter encodes a protein that emits red fluorescence.

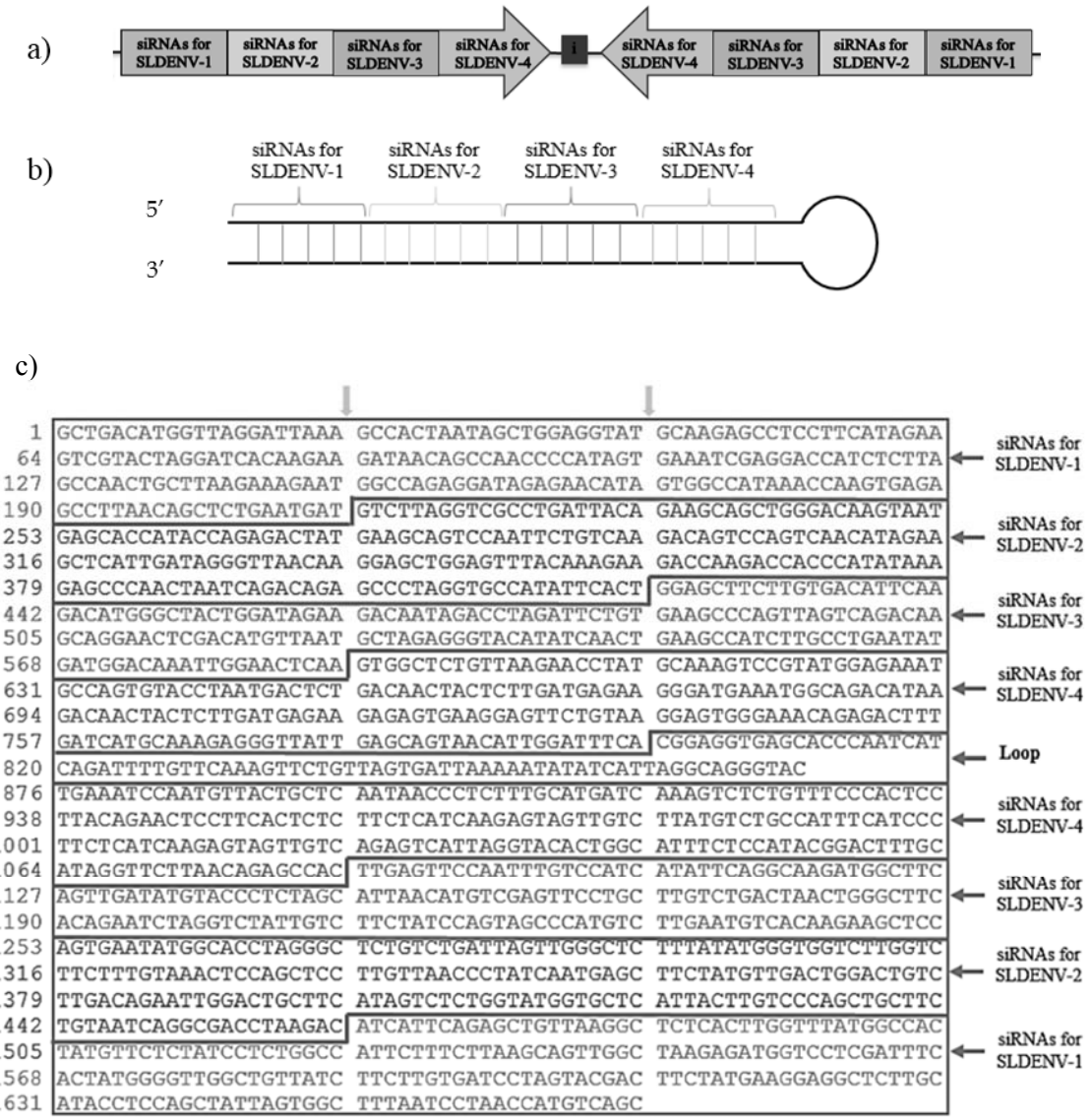

Fig 09: The designed effector lhRNA molecule in this study. The construct showing different regions of lhRNA made up of siRNAs specific to different dengue virus serotypes (a). The stem and loop structure of lhRNA molecule formed by complementary base paring of its sense and antisense regions (b). The sequence of designed lhRNA molecule (C), and the position 1 to 798-nt of the sequence is the region that corresponds to positive (+) strand of the stem structure, the position 799 to 874 is the region that corresponds to loop structure and the position 874 to 1672 is the region that corresponds to negative (-) strand sequence of the stem structure of lhRNA molecule. The spaces in the figure indicated by two arrows (at the top of the figure) are the sites of lhRNA molecule that are cleaved by the dicer enzyme and the regions between spacer are the designed siRNA sequences containing 21 nucleotides.

4. Conclusion

In this study, the unpublished sequences of previously published construct, AeCPA/Mnp+/i/Mnp-/svA and alternative promoter, *Ae. aegypti* vitellogenin A1 (AeVtA) promoter were found using bioinformatics tools. The promoter, AeVtA is an alternative promoter to AeCPA promoter as it also gets activated only after a blood meal in fat body of female *Ae. aegypti* [9] in contrast to AeCPA promoter which gets activated in the midgut epithelium of *Ae. aegypti* [6].

Since in this study the promoter sequences of AeCPA and AeVtA were identified, the effect of these promoters on siRNA mediated blocking of DENV replication in fat body and midgut epithelium of *Ae. aegypti* could be ascertained in developing transgenic mosquitoes resistant to DENV transmission. Further, in this study a method to construct a helper plasmid functionally similar to the pKhsp82MOS plasmid was described. Furthermore, in this study, the development of a new molecule, lhRNA that could be effective for all four SLDENV serotypes was described and even if this lhRNA molecule are to be cleaved by dicer enzyme in a random manner without sequence preference, which is not the case, will have a probability of 52.63% of siRNA molecules from the lhRNA to have homology that is required for the cleavage of target sites of the SLDENV genome of the respective serotype. Finally, it can be concluded lhRNA sequence designed in this study could trigger RNAi mechanism against every SLDENV serotype reported so far in the best and worst scenario of cleavage.

5. Acknowledgements

This work was supported by a grant (AP/3/2011/CG/03) awarded to Prof. R.S. Dassanayake by the University of Colombo, Sri Lanka.

6. References

1. World Health Organization. Treatment, Prevention and Control Global Strategy for Dengue Prevention and Control 2012-2020. WHO Press, 2012, 1-5.
2. Current Situation & Epidemiology of Dengue in Sri Lanka. http://www.healthedu.gov.lk/web/images/pdf/msp/current_situation_epidemiology_of_dengue.pdf. 28 Jul, 2014.
3. National Dengue Control Unit. Dengue Update. http://www.dengue.health.gov.lk/index.php/updates. 10 Aug, 2014.
4. Epidemiology Unit-Ministry of Health. Trends. http://www.epid.gov.lk/web/index.php?option=com_casesanddeaths&Itemid=448&lang=en#. 1 Dec, 2014.
5. Sirisena PDNN, Noordeen F. International Journal of Infectious Diseases Evolution of Dengue in Sri Lanka-Changes in the Virus, Vector, Int. J Infect Dis. 2014; 19:6-12.
6. Franz AWE, Sanchez-vargas I, Adelman ZN, Blair CD, Beaty BJ, James AA *et al*. Engineering RNA Interference-Based Resistance to Dengue Virus Type 2 in Genetically Modified *Aedes Aegypti* PNAS 2006; 103(11):4198-4203.
7. Idrees S, Ashfaq UA. RNAi: Antiviral Therapy against Dengue Virus. Asian Pac. J Trop Biomed. 2013; 3(3):232-236.
8. National Center for Biotechnology Information. Virus Variation - NCBI. http://www.ncbi.nlm.nih.gov/genomes/VirusVariation/Database/nph-select.cgi?tax_id=12637. 24 feb, 2015.
9. Kokoza V, Ahmed A, Shin SW, Okafor N, Zou Z, Raikhel AS. Blocking of *Plasmodium* Transmission by Cooperative Action of Cecropin A and Defensin a in Transgenic *Aedes Aegypti* Mosquitoes. PNAS. 2010; 107(18):8111-8116.
10. Addgene Addgene. Analyze Sequence: pGL3-Basic. http://www.addgene.org/browse/sequence_vdb/2929/. 28 feb, 2015.
11. Addgene Addgene. pGL3-hsp82 Sequences. https://www.addgene.org/52909/sequences/#depositor-full. 28 feb, 2015.
12. Addgene Addgene. pCFJ601 - Peft-3 Mos1 transposase - Analyze Sequence https://www.addgene.org/browse/sequence/42175/. 28 feb, 2015.
13. Moreira LA, Edwards MJ, Adhami F, Jasinskiene N, James AA, Jacobs-lorena M. Robust Gut-Specific Gene Expression in Transgenic Aedes Aegypti Mosquitoes. 2000; 97(20):10895-10898.
14. (ViPR) V.P.R. Virus Pathogen Database and Analysis Resource (ViPR) - Flaviviridae - Genome database with visualization and analysis tools: http://www.viprbrc.org/brc/home.spg?decorator=flavi_dengue (accessed Nov 8, 2014).
15. InvioGen. siRNA Wizard - Standard search of siRNA target sequences – InvivoGen: http://www.sirnawizard.com/design.php. 8 nov, 2014.
16. Vector Base-Bioinformatics Resource for Invertebrate Vectors of Human Pathogens. VectorBase: https://www.vectorbase.org/blast?organism=Aedes aegypti. 8 nov, 2014.
17. InvivoGen. siRNA Wizard - siRNA design guidelines http://www.sirnawizard.com/guidelines.php. 9 Nov, 2014.
18. Zamore PD, Tuschl T, Sharp PA, Bartel DP. RNAi : Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell 2000; 101:25-33.
19. Life technologies. Using siRNA for Gene Silencing Is a Rapidly Evolving Tool in Molecular Biology. www.lifetechnologies.com/lk/en/home/references/ambion-tech-support/rnai-sirna/general-articles/-sirna-design-guidelines.html. 30 Oct, 2014.