



ISSN 2278 – 0211 (Online)

In SILICO Approach to Design Potential Small Interfering RNA (siRNA) against Plant Pathogens and Diseases

Ranjan Sarmah

Research Scholar, Department of Computer Science, Assam University Silchar, Assam, India

Dr. Shahin Ara Begum

Associate Professor, Department of Computer Science, Assam University Silchar, Assam, India

Dr. Mahendra K. Modi

Professor, Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam, India

Abstract:

Among the members of single-stranded and double-stranded DNA and RNA viruses, those belonging to the family of plant genomes, cause various diseases in cereals and other economically important plants and also considered as a major constraint in agricultural production across the globe. These viruses are counteracted by plant defence mechanism through RNAi. There have been increasing interests in the use of siRNA as a tool for conferring resistance against plant-infecting viruses by silencing the target genes in a sequence specific manner. In this study, highly cited bioinformatics tools viz., siRNA Target Finder, Genscript, siDESIGNCenter, siRNA-wizard and siRNA Selection Server, combined with different algorithms namely Tuschl, Reynolds, Uitei and their combinations were used to screen and design putative siRNAs against the genes of five different plant viruses. BLAST was performed to confirm that the designed siRNAs do not have any homology to other plants. In the present study, 18 most promising siRNAs were identified computationally based on GC%, thermodynamic properties and secondary structures having high potential of targeted gene silencing and will be useful for chemical synthesis. This study also reveals that above 70% efficient siRNAs were obtained with bioinformatics tools using machine learning techniques.

Keywords: Plant Pathogens, siRNA, BLAST, Efficacy

1. Introduction

Viruses belonging to the family of plant genome are considered as the most destructive pathogens affecting huge reduction in the yield of legumes, cereals and other economically important crops, such as cotton (*Gossypium hirsutum*), cassava (*Manihot esculenta*), maize (*Zea mays*), common bean (*Phaseolus vulgaris*) and tomato (*Solanum lycopersicum*) in tropical as well as in the subtropical regions and has become a major threat to agriculture across the globe. Considering the significant impact of these plant viruses on crop production, the development of an efficient and effective control strategy against the Gemini viruses and viral plant are of immense importance. Over the last few decades, several means of management, such as physical, biological and chemical strategies have been inducted.

Biotechnological methods such as transgenic approaches [1-3], are also adopted to manage the Gemini viral diseases of plants. Although to some extent, success has been achieved by using these means [4], most of these approaches have their drawbacks and pitfalls, including serious environmental concerns and market acceptability.

RNA interfering (RNAi) is an interesting and widely used tool for the molecular virologists to manage plant viruses. With the increase of high-throughput technologies, it becomes relatively more convenient to characterize the small non-coding RNAs and for RNA silencing. Several plant viruses have been successfully controlled using RNAi-based methods over the last decade [5-7]. The general methods of RNA silencing in plants involves a sequence-specific inhibition of expression of the target genes at a transcriptional or a post-transcriptional levels, known as transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS), respectively. However, in case of antiviral defence, RNAi works at the post-transcriptional level. In PTGS, the 21- 24 nucleotide long small interfering RNAs (siRNA) molecules silence the gene expression by specific cleavage of target RNA which is homologous to target genes [8-10]. The siRNA-mediated RNA silencing, in addition to its function as a natural antiviral defence mechanism, also plays regulatory roles in plants' development process [11-12].

Various studies have shown that PTGS and are good targets for RNA silencing in homology dependent manner [13]. The strategy involves expressing or inducing the short (21-24 nt) siRNA molecules in the plants which are capable of initiating the homology-

dependent gene silencing in a sequence specific manner [14]. Even chemically synthesized siRNA duplexes, when introduced into the plants, have shown great promises in achieving effective target RNA cleavage [15].

Therefore, an *in silico* approach has been used for screening and designing of efficient siRNAs to counter the diseases caused by plant viruses. The present study emphasized on ten viral genes from five different plant viruses such as Srilankan Cassava Mosaic virus-(Ker 20) and Cotton Leaf Curl Rajasthan virus (Family: Geminiviridae), Cauliflower Mosaic Virus Isolate W260 (Family: Caulimoviridae), Cucumber Mosaic Virus (Family: Bromoviridae) and Tobacco Mosaic Virus (Family: Virgaviridae) infecting crops species such as cassava, cauliflower, cotton, cucumber and tobacco respectively. The process of screening and designing of efficient or putative siRNAs consist of series of steps involving several bioinformatics tools accompanied with novel algorithms. Among the bioinformatics tools some are using selection parameter to design effective siRNA and some are using Machine Learning algorithm in the process of improving the success rate effective targeting. Support Vector Machine (SVM), Artificial Neural Network (ANN), etc., are machine learning algorithms which can be a good approach while training a biologically validated siRNA dataset with adequate volume[16]. To cross check the silencing efficiency and the off target effect of putative siRNAs, BLAST [17] search was performed against the known plant genome. To have a stable, broad and effective resistance against these selected plant viruses, screening and design of effective and potential siRNA are of great importance in modern plant-biology prospective.

2. Material and Methods

2.1. Retrieval of Target Viral Gene Sequences

In the present study, two different viral genes from each of the five viral plants have been selected. The complete sequences of the two viral genes namely *coat protein (CP)* and *transcriptional activator protein (TrAP)* (GenBank Accession No: AJ579307) from *SriLankan cassava mosaic virus (SLCMV)*; *CP* and *TrAP* protein from *Cauliflower mosaic virus (CaMV)*(GenBank Accession No: JF809616), *CP* and *TrAP* from *Cotton leaf curl Burewala virus (CLCuBuV)* (GenBank No: JF509750); *2b protein* (GenBank Accession No: EU888910) and *CP* (GenBank Accession No: JQ013954) from *Cucumber mosaic cucumo virus Egyptian isolate (CMV-Gera-EG)*; *CP* and *movement protein (MP)* from *Tobacco mosaic virus (TMV)* (GenBank Accession No:AF273221) were retrieved from the GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) database of NCBI. The information about the various plant viruses and their genes were collected from Plant Virus Online database (<http://www.dpvweb.net>). The complete information of the two genes for each virus, taken in the present investigation, is summarized in Table 1.

Sl. No	Source Virus Genome/ Family	Gene Name	Gene length (nt)	GenBank Accession number	References
1	Sri Lankan cassava mosaic virus - [Ker20] Family: Geminiviridae (Single Stranded DNA Virus)	Coat protein AV1	771	AJ579307.1	Dutt et al., (2005)
		Transcription activator protein AC2	408	AJ579307.1	Dutt et al., (2005)
2	Cauliflower mosaic virus isolate W260 Family: Caulimoviridae (Double Standed DNA Virus)	Capsid protein	1464	JF809616.1	Froissart,R (2011)
		Transcription activator protein	1563	JF809616.1	Froissart,R (2011)
3	Cotton leaf curl Rajasthan virus Family: Geminiviridae (Single Stranded DNA Virus)	Coat protein V1	771	JF509750.1	Rajagopalan et al.,(2012)
		Transcription activator protein C2	453	JF509750.1	Rajagopalan et al., (2012)
4	Cucumber mosaic virus Family: Bromoviridae (Single Stranded RNA virus)	2b protein	336	EU888910.1	Raj et al., (2008)
		Coat Protein CP	657	JQ013954.1	Sofy et al., (2011)
5	Tobacco mosaic virus Family: Virgaviridae (Single Stranded RNA virus)	Coat Protein	480	AF273221.1	Holt et al., (1990)
		Movement protein	807	AF273221.1	Holt et al., (1990)

Table 1: Viral protein coding genes of different plant viruses selected for designing potential siRNA

2.2. Target Identification and Rational Design of Candidate siRNA

The first and foremost step in design of siRNA is the identification of various target sequence for candidate siRNAs within the viral genes. The various targets were screened based on algorithms incorporated in five highly cited siRNA design tools namely siRNA Target Finder (Ambion, USA) (http://www.ambion.com/techlib/misc/siRNA_finder.html) implements siRNA design guidelines described by Tuschl and colleagues [18], Genscript (<https://www.genscript.com/ssl-bin/app/mna>), Dharmacon DESIGN Center (<http://www.dharmacon.com/designcenter/DesignCenterPage.aspx>), siRNA Wizard v3.1 (<http://www.sirnawizard.com/design.php>) and siRNA Selection Server (<http://sirna.wi.mit.edu/>) [19] using default settings. Several multi stringent criteria as recommended by Birmingham et al., 2007 [20] and Elbashir et al., 2001 [21] was employed in order to screen the best candidate siRNAs which in turn significantly improves the likelihood of identifying functional siRNA. This approach would enhance the target specificity and adapt siRNA designs for more sophisticated experimental design in RNAi mediated gene silencing.

2.3. Cross Validation of Candidate siRNAs by BLAST Search against mRNA Database

The candidate siRNAs were selected to design putative siRNAs for efficient silencing viral genes, where these sequences do not reveal any remote homology to any other gene sequences *i.e.*, target site having more than 16 nt contiguous base pair with any other organism. In addition, to ensure and reduce the off-target cleavage effect of the designed siRNAs in plant gene, BLASTn search was performed against non-redundant (NR) database of NCBI of the selected 21-24 nt long candidate siRNA sequences. Cases where the candidate siRNA showed any homology to any of the plant genes from the database, were discarded from the list of candidate siRNAs, that could be an efficient silencing tool against the target viruses, so as to avoid any unintended effect on endogenous gene expression.

2.4. GC Content, ΔG value and siRNA Secondary Structure Prediction

The short listed siRNAs were further screened for estimation of GC percentage and secondary structures prediction to design the efficient siRNA. The i-Score designer tool (http://www.med.nagoya-u.ac.jp/neurogenetics/i_Score/i_score.html) [22] was used to calculate the GC % and ΔG value of the most stable secondary structure of siRNA strand according to the RNA mfold algorithm [23]. The secondary structure and free energy of folding of each putative siRNA was computed using RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) [24].

2.5. Estimation of the Binding Energy of the Duplex

The thermodynamic stability of the duplex formed between the siRNA and its target is one of the major factors in determining the degree of its efficiency, as well as its possible off-target effects. To estimate thermodynamics properties of the short listed putative siRNA duplexes (*i.e.* between the predicted siRNA molecules and the target genes), the RNAup web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAup.cgi>) [25] was used. RNAup programme uses an extension of the standard partition-function approach to predict the RNA secondary structures and to compute the energy of RNA-RNA interactions.

The detailed work flow for screening and designing of siRNA sequences have been shown in the Figure 1.

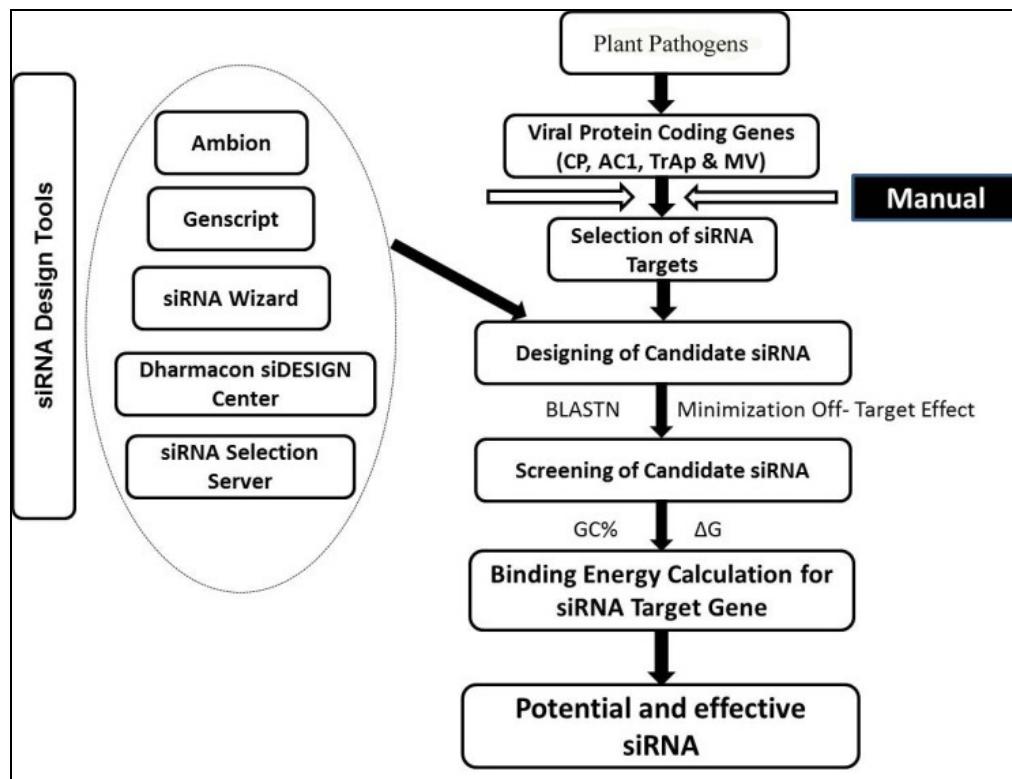


Figure 1: Various phases for designing siRNA for plant pathogens with high efficiency and sensitivity

3. Results and Discussions

Viral genes are multifunctional in nature, and effect on various endogenous processes for the successful infection. The genome of many Gemini viruses consists of two circular genomes of similar size but of different nucleotide sequence, known as DNA-A and DNA-B [25]. Among the viral proteins, replication initiation protein (AC1/ Rep) plays an essential role in replication [26-27] and also acts as an elicitor of the hypersensitive response (HR) which is detrimental to the virus life cycle [28]. Transcriptional activator protein called TrAP/ AC2 is a multifunctional protein involved in regulation of expression of other viral proteins [29], and also possibly involved in suppression of the endogenous post-transcriptional gene-silencing [30]. In fact most of the viral proteins such as TrAp, AC1, CP (coat protein) and MP (movement protein) are required for virus infection including virus movement, viral replication,

and can be recognised in plants and elicit HR [31-32].

The viral gene sequences of different plant viruses infecting various crops were downloaded from GenBank of NCBI are reported in Table 1. Therefore, to select various target sequence for siRNA within the viral genes, a stretch of nucleotides with 5'AA dinucleotide or triplet AAG/C with 18/19 nucleotides were scanned manually. In addition to this a stretch of 4 T's or A's in the target sequence were avoided as it may act as a termination signal for RNA polymerase III [33].

For the selection of the siRNA, following optimal parameters such as oligonucleotides should have 30-50% GC content, no G residues in the overhang and extra UU or dTdT dinucleotide at the 3'-ends were considered for best and accurate results.

Several rules for designing rational siRNA proposed by different group's such as Tuschl et al., [18], Reynolds et al., [34], Chalk et al., [35] and Amarzguioui et al., [36], provides a strong impact on the first generation of siRNA, mostly based on the information of GC content, preference or avoidance of specific nucleotides at specific positions and siRNA sequence motifs. So to assess the correctness of the manually selected oligonucleotides, they were subjected to various siRNA target finder and design softwares which implements the different rules described above. Five highly accessed and efficient siRNA design tools viz., siRNA Target Finder (Ambion), Genscript, Dharmaconsi DESIGN Center, siRNA Wizard and siRNA Selection Server were used in the present study. A number of filtering procedure described by Birmingham et al., 2007 [20] and Elbashir et al., 2001 [21] was employed to screen the best possible candidate siRNA from the number of siRNA generated by different softwares. The 562 siRNA sequences satisfying the mixed rules of Ui-Tei[37], Hsieh [38] , s-Biopredsi[39],i-Score [22] and Reynolds [34] were considered for further analyses.

A cross homology search of 562 candidate siRNA against non-redundant database of NCBI through BLAST was performed and 60 putative siRNAs were screened out. Cross homology search revealed that the designed siRNAs showed a similarity of 95-100% to considered plant viruses but not to any other gene of the plant genome, thus assured silencing of only selected viral plant genes can be achieved through the designed siRNA. Also the target sites of the designed siRNAs were scanned against various completed genomes databases (rice and Arabidopsis).

Inconsistent results may be obtained with regards to the effects of GC content and secondary structure on siRNA efficiency and functionality. However, it is recommended to consider siRNA sequences with low GC content ranging from 31%-58% [39, 40, 34, 36]. Furthermore, the 60 putative siRNA were assessed for their possible folding pattern using RNAfold server, a widely used server to predict minimum free energy (MFE) structures and base pair probabilities of RNA molecules. These structures are predicted using a loop-based energy model and the dynamic programming algorithm reported by Zuker et al., 1981 [41].

Out of the 60 putative siRNA, only 18 siRNA sequences showed zero free energy of folding at room temperature (37°C) [Table 2]. Previous study by Shao et al., 2007 [42], it has been reported that the RNA molecule should have minimum free energy of folding for their stability. Therefore, the RNA molecule with positive energy may be more accessible for target site and have high potential to bind with target site and lead to an effective gene silencing. Whereas, the rest of 42 siRNA sequences are also having less than -1 kcal free energy of folding. The effective 18 siRNA sequences with GC%, free energy of folding and free energy of binding with target has been shown in Table 2.

Sl. No	Location of siRNA within gene	Length of siRNA	siRNA sequences	Gene	Virus	Software Used	Minimum energy (kcal/mol)	GC%	Total free energy of binding (kcal/mol)	Secondary Structure
1	364-382	19	ATAAGGAGCGCGATAAATA	coat protein	Tobacco Mosaic virus	Genscript/ML	0.00	37.0	-5.68	Fig 2 A
2	995-1013	19	TCAAGAACACAGAGAAAGA	TAV protein	Cauliflower Mosaic virus	Genscript/ML	0.00	37.0	-7.06	Fig 2 B
3	352-370	19	GGTAAGATATGGATGGATG	coat protein	Cotton Leaf Curl Virus	Genscript/ML	0.00	42.0	-6.55	Fig 2 C
4	144-164	21	TCTCAGACTATTCCGCTTCCT	2b protein	Cucumber Mosaic Virus	Genscript/ML	0.00	47.62	-8.36	Fig 2 D
5	477-495	19	AGTCCAAGCCAACAAACAAA	coat protein	Cucumber Mosaic Virus	siDesign	0.00	42.0	-5.14	Fig 2 E
6	162-182	21	AAACCTTCACCACAAGTAAC T	coat protein	Tobacco Mosaic Virus	Genscript /Statistical Methd	0.00	38.1	-4.75	Fig 2 F
7	379-397	21	GACCAAGAUCAUACGAU	coat protein	Cassava Mosaic Virus	Genscript/ML	0.00	37.0	-5.31	Fig 2 G
8	156-174	19	CCGCTTCCTACCGTTCAAT	2b protein	Cucumber Mosaic Virus	Ambion	0.00	52.64	-11.26	Fig 2 H
9	154-172	19	CCGCUUCCUACCGUUCAAU	2b protein	Cucumber Mosaic Virus	Genscript/ML	0.00	53.0	-11.26	Fig 2 I
10	491-509	19	GCAGCAGAAUCGGAUUGA	coat protein	Cauliflower Mosaic virus	Genscript/ML	0.00	47.0	-6.89	Fig 2 J
11	391-409	19	GTAGAATTGATCAGAGGAA	coat protein	Tobacco Mosaic Virus	Genscript/ML	0.00	37.0	-6.37	Fig 2 K
12	353-373	21	GCAAGATATGGATGGATGAA A	coat protein	Cassava Mosaic Virus	siRNA_wizard	0.00	38.1	-8.21	Fig 2 L
13	819-839	21	AAGACAGAACTGGCGGATTTC	coat protein	Cauliflower Mosaic virus	Genscript/ML	0.00	47.62	-8.71	Fig 2 M
14	120-140	21	AAACACAGGCATGGACAA AC	coat protein	Cotton Leaf Curl Virus	Genscript/ML	0.00	42.86	-9.84	Fig 2 N
15	382-400	19	ACGAAGAACATCACACGAATA	TAV protein	Cotton Leaf Curl Virus	siDesign	0.00	37.0	-7.60	Fig 2 O
16	378-398	21	TAAGACGAAGAACACACGA AA	coat protein	Cotton Leaf Curl Virus	Genscript/ML	0.00	38.1	-6.11	Fig 2 P
17	381-401	19	AAGACGAAGAACACACGA AT	coat protein	Cotton Leaf Curl Virus	Genscript/ML	0.00	38.1	-6.11	Fig 2 Q
18	957-975	19	GGTTAAAGATGCAGTCAAA	TAV protein	Cauliflower Mosaic virus	Genscript/ML	0.00	37.1	-6.92	Fig 2 R

Table 2: List of high potential siRNA sequences developed for the viral coding genes of plantviruses, with their sequence features (%GC, Minimum energy, Free energy of binding)

The thermodynamics study of RNA-RNA interactions might be one of the important aspects for siRNA efficiency. The 60 putative siRNA sequences were subjected to RNA-RNA interaction study with their respective targets to cross validate the results. RNA up tool of Vienna website was used to predict free energy of RNA-RNA interactions; where the Vienna suite offers various tools with the state-of-the-art algorithms for RNA folding, prediction and comparison of RNA-RNA interactions. The free energy of interaction (binding) between each siRNA sequence and its target was calculated. It was reported previously by Schubert et al., 2005 [43] that the efficiency RNAi correlates well with the binding energies of siRNAs and their respective mRNA target. Muecksteinet. al.,[44] reported that if the optimal free energy of binding (BE) is highly favourable then the RNA molecule will bind almost exclusively to the intended target site. It has also been reported that the poor stepwise decrease of the target accessibility is directly correlated to a poor optimal BE and decreased silencing efficiency [19]. In this case a total of 18 siRNA sequences of their respective genes were found to be more promising with BE [Table 2]. The secondary structure of the mRNA target site plays a vital role in siRNA efficiency prediction. Thus secondary structures as listed in Table 2 are shown in Figure 2 by using the RNAfold program of the Vienna RNA package.

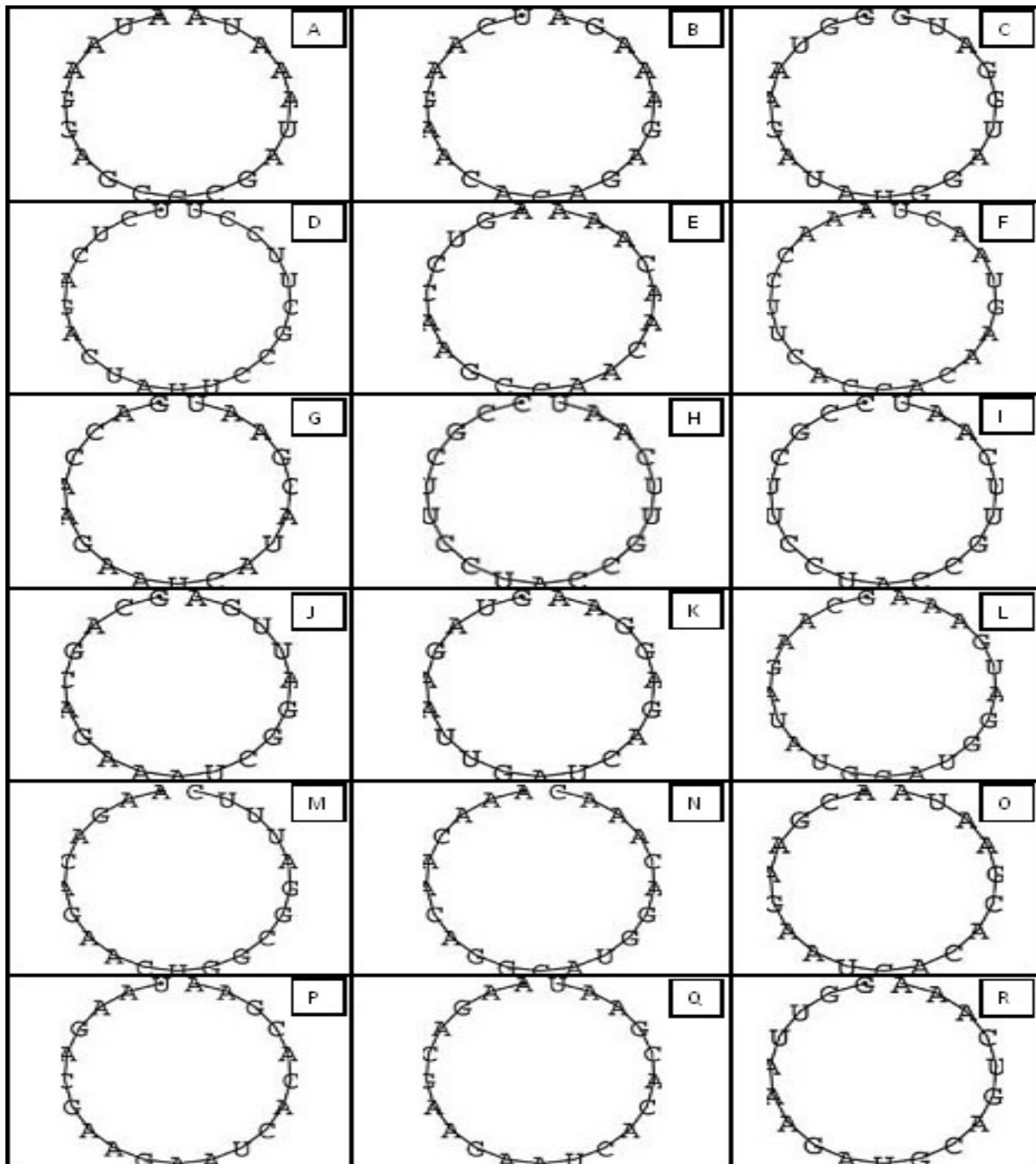


Figure 2: (A-R).siRNA secondary structures.

The other 42 numbers of putative siRNA along with their features have been reported in Table3. Therefore, these potential siRNA may qualify as high quality candidate for silencing the viral protein coding genes and may be used for effective silencing of plant viruses.

Sl. No.	Location of siRNA within gene	Length of siRNA	DNA sequence for targeted siRNA	Minimum energy (kcal/mol)	GC%	Total free energy of binding with the target
1	121-139	19	CAACAGGCATGGACAAACA	-0.28	47	-9.92
2	1240-1260	21	ATCTGCAATATCGAAGGCCAT	-0.31	42.86	-11.32
3	12-30	19	CACCGAGAUUGUUCAGAU	-0.31	42	-7.03
4	188-206	19	CGGAACGTAGAGATGTA	-0.33	42	-5.41
5	569-587	19	CAGAAGAACCTCGTTGATGA	-0.38	42	-8.67
6	140-158	19	CCTCTCTACCTAGTGATAA	-0.39	42.11	-6.31
7	558-576	19	CATGGAACCTACAGAAGAA	-0.40	37	-9.32
8	1259-1279	21	ATTACGCCAACGAATGTCCTA	-0.41	42.86	-6.60
9	810-830	21	GTTCACCACGTATAATGAGAA	-0.42	38.1	-6.46
10	247-267	21	AAGACGAAAGCGATTAGGAG	-0.43	47.62	-5.35
11	341-361	21	GACCATCAAGATACAATGAGA	-0.43	38.1	-5.34
12	54-72	19	GGCCGACCCAATAGAGTTA	-0.43	53	-6.24
13	52-70	19	GGCCGACCCAUAUAGAGUUA	-0.43	53	-6.71
14	200-218	19	ACTTTAAGGTGTACAGGTA	-0.44	37	-5.93
15	53-71	19	TGGTGGAGGTGAAGAGACA	-0.45	53	-8.03
16	424-444	21	GATCGTAGACCTGTTGATAAA	-0.46	38.1	-10.60
17	54-72	19	CGATCTAGTAAGAGCAAAA	-0.55	37	-5.67
18	79-97	19	CGCGGAUGCUAACUUUAGA	-0.58	47	-8.48
19	66-84	19	GATCTTACCGTCGATGTTT	-0.59	42.11	-7.12
20	545-563	19	GCUGACGCAUAUACAACAA	-0.61	42	-7.49
21	13-31	19	GCTGCCGATATCGTCATT	-0.62	47.37	-8.19
22	212-230	19	GCUCCUGGUAAAGAAUCAA	-0.63	42	-9.55
23	259-179	21	GGTAAGGTATGTGCATCT	-0.65	47.62	-11.57
24	475-495	21	GGAGTCCAAGCCAACAACAA	-0.65	47.62	-7.42
25	547-567	21	GCTGACCCATATACAACAAGT	-0.68	42.86	-7.45
26	316-334	19	GGTAAGAGATTTGTGTCA	-0.70	37	-8.51
27	75-93	19	GCAGGUCUCACAAGAAGAA	-0.75	47	-5.97
28	285-303	19	AGTATGAGCAGCCGAATT	-0.77	47	-8.15
29	11-29	19	ACGCAGGCGCAATGACAAA	-0.77	53	-8.00
30	483-503	21	AATGAACCTAGTACAGCTACG	-0.78	42.86	-6.70
31	582-602	21	AAGGAACAGGCTTAGTTAGG	-0.79	42.86	-9.27
32	544-564	21	GCAGCTGACGCATATACAACA	-0.80	47.62	-7.30
33	4-22	19	GAATTGAACGCAGGCGCAA	-0.81	52.64	-10.37
34	523-543	21	GCTGATATAGGTGACATGAGA	-0.83	42.86	-9.25
35	93-111	19	GGUUGUGGUUGAAGCGUAU	-0.87	47	-9.12
36	159-177	19	GCUGGAUGAUCCUGAAUA	-0.89	42	-7.63
37	641-658	18	GAGGCUGGCAAGUAUGAA	-0.90	47	-9.45
38	103-123	21	TTGAAGCGTATCTGGATGGTT	-0.91	42.86	-12.45
39	6-26	21	TTACAGTATCACTACTCCATC	-0.93	38.1	-3.69
40	7-27	21	AATTGAACGCAGGCGCAATGA	-0.96	47.62	-10.23
41	307-325	19	GCCACTCTCGGATCTTACT	-0.97	52.64	-6.73
42	1040-1058	19	GGAGAACGUUCAGUAGAAU	-1.03	42	-6.23

Table 3: Predicted putative siRNA sequences along with their features

Numerous algorithms have been developed to determine the siRNA silencing activity. However, a very few have sustained to a level of specificity and sensitivity. These algorithms can be divided into first generation and second generation. In the “first generation”, the siRNA efficacy is determined on the basis of thermodynamic stability, secondary structure of mRNA and target positions. On the other hand, the second generation tools were developed by using eminent data mining techniques to interpret glossed records of siRNA with their experimental inhibition. The second generation models (whether neural network or linear regression based or hybrid) perform significantly better than the first generation models.

In the present study, it has been observed that out of 18 effective siRNA sequences, 13 siRNA sequences were obtained from the bioinformatics software using machine learning algorithms [Table 2].

4. Conclusions

Plant pathogens are considered as eminent threat to agricultural practices across the globe and one of the major constrain to the development of virus resistant crop plants. The economically important cereals crops are prone to various plant diseases, resulting million dollar loss every year thereby affects the global economy. To feed the ever increasing population, development of virus resistance crop plants is of immense importance and needs much more attention in Agriculture. As siRNA mediated gene silencing of viral genes have been proven to become a pivotal tool in the development of virus resistance crop plants. Thus, it is of interest to design effective and efficient siRNA to silence the most casual agents of virus infection and replication in cereal crops. The availability of high-throughput computational tools along with optimised algorithms has heralded to design and characterise efficient siRNA against these viral genes.

In the present scenario the viral genes of plant viruses from four different families are considered as the most important targets for silencing through RNAi technology. The present study involves screening and design of promising siRNA against five different selected viruses considering two genes from each virus. The target and putative siRNA were selected from these genes manually and also with five different approaches by satisfying all the criteria of siRNA designing with high functionality and specificity. A list of 18 most promising potential siRNA have been obtained by satisfying all biosafety and structural criteria and that will be a valuable repository for chemical synthesis and insect feeding assay. As these siRNA represent the 18 most potential from 562 possible siRNA in terms bioactivity and safety, it will considerably reduce the cost of chemical synthesis and time complexity. It is also found that out of 18 efficient siRNA sequences, more than 70% siRNA were obtained with the Bioinformatics tool using Machine Learning Techniques.

With the great potential of RNAi techniques in the continuous development of high throughput RNAi screening, it is important to improve the siRNA designing rules as well as to construct reliable siRNA efficacy prediction model. Machine learning algorithms like Support Vector Machine (SVM) and Artificial Neural Network (ANN) can outperform when trained with sufficient volume of biologically validated siRNA dataset.

5. Acknowledgements

The author is grateful to Department of Biotechnology, Assam Agricultural University for providing constant support, necessary facilities and encouragement during the study.

6. References

- i. Vanderschuren, H., Stupak, M., Futterer, J., Gruissem, W., Zhang, P.(2007). Engineering resistance to geminiviruses-review and perspectives. *Plant Biotechnol Journal*, 5(2), 207–220.
- ii. Zrachya, A., Kumar, P.P., Ramakrishnan, U., Levy, Y., Loyter, A., Arazi, T., Lapidot, M., Gafni, Y.(2007).Production of siRNA targeted against TYLCV coat protein transcripts leads to silencing of its expression and resistance to the virus. *Transgenic Research*, 16 (3), 385–398.
- iii. Fuentes, A., Ramos, P.L., Fiallo, E., Callard, D., Sanchez, Y., Peral, R., Rodriguez, R., Pujol M. (2006). Intron-hairpin RNA derived from replication associated protein C1 gene confers immunity to Tomato Yellow Leaf Curl Virus infection in transgenic tomato plants. *Transgenic Research*, 15 (3), 291–304.
- iv. Lucioli, A., Sallustio, D.E., Barboni, D., Berardi, A., Papacchioli, V., Tavazza, R., Tavazza, M. (2008).A cautionary note on pathogen- derived sequences. *Nature Biotechnol*, 26: 617–619.
- v. Vanitharani, R., Chellappan, P.,Fauquet, C.M.(2003). Short interfering RNA-mediated interference of gene expression and viral DNA accumulation in cultured plant cells. *Proceedings of the National Academy of Science of USA*,100(16),9632–9636.
- vi. Bonfim, K., Faria, J.C., Nogueira, E., Mendes, E.A., Aragao, F.J.L.(2007). RNAi-mediated resistance to Bean golden mosaic virus in genetically engineered common bean (*Phaseolus vulgaris*). *Molecular Plant- Microbe Interactions*,20(6), 717–726.
- vii. Tenllado, F., Llave, C., Diaz-Ruiz,J.R.(2004).RNA interference as a new biotechnological tool for the control of virus diseases in plants, *Virus Research*,102(1),85–96.
- viii. Baulcombe, D.C. (2000). Unwinding RNA Silencing, *Science*,290 (5494) ,1108-1109.
- ix. Vaucheret, H., Fagard, M.(2001). Transcriptional gene silencing in plants: targets, inducers and regulators, *Trends in Genetics*, 17 (1),29-35.
- x. Hamilton, A.J., Baulcombe, D.C. (1999).A species of small antisense RNA in posttranscriptional gene silencing in plants, *Science* 286 (5441): 950–952.
- xi. Beclin, C., Boutet,S., Waterhouse, P., Vaucheret, H. (2002).A branched pathway for transgene-induced RNA silencing in plants, *Current Biology*,12(8),684-688.
- xii. Ding, S.W. (2010).RNA-based antiviral immunity, *Nature Reviews Immunology*,10(9), 632–644.
- xiii. Elbashir, S.M., Lendeckel, W., Tuschl, T. (2001).RNA interfernce is mediated by 21 and 22nucleotide RNAs, *Genes and Development*, 15,188 – 200.
- xiv. Praveen, S., Mishra, A.K., Antony, G.(2006).Viral Suppression in Transgenic Plants Expressing Chimeric Transgene from Tomato Leaf Curl Virus and Cucumber Mosaic Virus, *Plant CellTissue Organ Culture*, 84(1) ,49–55.
- xv. Tenllado, F., Diaz-Ruiz, J.R. (2001). Double-Stranded RNA-Mediated Interference with Plant Virus Infection, *Journal of Virology*,75(24),12288–12297.
- xvi. Ghosal, S., Das, S.,Chakrabarti, J. (2012). Computational approaches for Designing Efficient and Specific siRNAs, *International J.*

- xvii. Bioautomation, Bioinformatics (Book), Chapter 11.
- xviii. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.(1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Research*,25(17), 3389-3402.
- xix. Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., Sharp, P.A. (1999).Targeted mRNA degradation by double-stranded RNA in vitro,*Genes & Development*,13(24), 3191-3197.
- xx. Yuan, B., Latek, R., Hossbach, M., Tuschl, T., Lewitter, F.(2004). siRNA Selection Server: an automated siRNA oligonucleotide prediction server, *Nucleic Acids Research*,32(2), W130-W134.
- xxi. Birmingham, A., Anderson, E., Sullivan, K., Reynolds, A., Boese, Q., Leake, D., Karpilow, J., Khorova, A. (2007). A protocol for designing siRNAs with high functionality and specificity, *NatureProtocols*, 2(9), 2068–2078.
- xxii. Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T. (2001).Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate, *EMBOJournal*,20(23),6877-6888.
- xxiii. Ichihara, M., Murakumo, Y., Masuda, A., Matsuura, T., Asai, N., Jijiwa, M., Ishida M., Shinmi, J., Yatsuya, H., Qiao, S., Takahashi, M., Ohno, K. (2007). Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities, *Nucleic Acids Research*,35(18), e123.
- xxiv. Zuker, M., Mathews, D.H., Turner, D.H.(1999). Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide, *RNA Biochemistry and Biotechnology*,70, 11-43.
- xxv. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., Hofacker, I.L. (2008). The Vienna RNA websuite,*Nucleic Acids Research*,36(Web Server issue), W70-W74.
- xxvi. Stanley, J., Gay, M.R.(1983). Nucleotide sequence of cassava latent virus DNA, *Nature*, 301, 260-262,
- xxvii. Chatterji, A., Beachy, R.N., Fauquet, C.M. (2001). Expression of the oligomerization domain of the replication-associated protein (Rep) of tomato leaf curl New Delhi virus interferes with DNA accumulation of heterologous geminiviruses, *The Journal of Biological Chemistry*,276, 25631-25638.
- xxviii. Elmer, J.S., Brand, L., Sunter, G., Gardiner, W.E., Bisaro, D.M., Rogers, S.G.(1988).Genetic analysis of the tomato golden mosaic virus. II. The product of the AL1 coding sequence is required for replication,*Nucleic Acids Res.*,16(14B),7043-7060.
- xxix. Hussain, M., Mansoor, S., Iram, S., Zafar, Y.,Briddon, R.W. (2007).The hypersensitive response to tomato leaf curl New Delhi virus nuclear shuttle protein is inhibited by transcriptional activator protein, *Mol Plant Microbe Interact*,20, 1581–1588.
- xxx. Trinks, D., Rajeswaran, R., Shivaprasad, P.V., Akbergenov, R., Oakeley, E.J., Veluthambi, K., Hohn, T., Pooggin, M.M.(2007). Suppression of RNA silencing by a geminivirus nuclear protein, AC2, correlates with transactivation of host genes,*Journal of Virology*,79, 2517–2527.
- xxxi. Raja, P., Wolf, J.N., Bisaro, D.M. (2010).RNA silencing directed against geminiviruses: post-transcriptional and epigenetic components,*BiochimicaetBiophysicaacta*, 1799(3-4),337–351.
- xxxii. Garrido-Ramirez, E.R., Sudarshana, M.R., Lucas, W.J., Gilbertson, R.L.Bean dwarf mosaic virus BV1 protein is a determinant of the hypersensitive response and avirulence in *Phaseolus vulgaris*,*Molecular plant- microbe interactions*, 13(11), 1184–1194.
- xxxiii. Erickson, F.L., Holzberg, S., Calderon-Urrea, A., Handley, V., Axtell, M., Corr, C., Baker, B. (1999).The helicase domain of the TMV replicase proteins induces the N-mediated defence response in tobacco, *Plant Journal*,18(1),67–75.
- xxxiv. Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T. (2001).Functional anatomy of siRNA for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate,*EMBOJournal*,20 (23),6877-6888.
- xxxv. Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S., Khvorova, A. (2004).Rational siRNA design for RNA interference, *Nature Biotechnology*,22(3), 326–330.
- xxxvi. Chalk, A.M., Wahlestedt, C., Sonnhammer, E.L. (2004).Improved and automated prediction of effective siRNA, *Biochemical and Biophysical Research Communications*,319(1), 264–74.
- xxxvii. Amarzguioui, M., Prydz, H. (2004).An algorithm for selection of functional siRNA sequences, *Biochemical and Biophysical Research Communications*,316(4), 1050–1058.
- xxxviii. Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R., Saigo, K. (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference,*Nucleic Acids Research*,32(3), 936-948.
- xxxix. Hsieh, A.C., Bo, R., Manola, J., Vazquez, F., Bare, O., Khvorova, A., Scaringe, S., Sellers, W.R. (2004). A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens,*Nucleic Acids Research*, 32(3), 893-901.
- xl. Huesken, D., Lange, J., Mickanin, C., Weiler, J., Asselbergs, F., Warner, J., Meloon, B., Engel, S., Rosenberg, A., Cohen, D., Labow, M., Reinhardt, M., Natt, F., Hall, J. (2005). Design of a genome-wide siRNA library using an artificial neural network, *Nat Biotechnology*,23(8), 995-1001.
- xli. Choudhary, M., Sahi, S.(2011). In silico designing of insecticidal small interfering RNA (siRNA) for *Helicoverpaarmigera* control, *Indian journal of Experimental Biology*,49,469-474.
- xlii. Zuker, M., Stiegler, P. (1981).Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information, *Nucleic Acid Research*, 9, 133-148.
- xliii. Shao, Y., Chan, C.Y., Maliyekkel, A.(2007). Lawrence CE, Roninson IB, Ding Y, Effect of target secondary structure on RNAi efficiency, *RNA*13, (10), 1631-1640.
- xlv. Schubert, S., Grünweller, A., Erdmann, V.A., Kurreck, J. (2005).Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions, *Journal of Molecular Biology*,348(4),883-893.
- xlii. Muckstein, U., Tafer, H., Hackermüller, J., Bernhart, S.H., Stadler, P.F., Hofacker, I.L. (2006).Thermodynamics of RNA-RNA binding, *Bioinformatics*,22(10), 1177-82.

545	180	GCTTATTGATAGTGGATACGT	38.1	-0.73 kcal/mol	30.41 %	4.12	-6.95 kcal/mol	-12.40 kcal/mol	siRNA_wizard	Tobacco Mosaic Virus	Movement Protein
546	306	GGCCACACTCGGATCATACTA	52.38	-1.16 kcal/mol	55.44 %	1.48	-7.52 kcal/mol	-16.82 kcal/mol	siRNA_wizard	Tobacco Mosaic Virus	Movement Protein
547	361	GTCGTTCCAATTATGCTATA	38.1	-0.34 kcal/mol	57.26 %	1.79	-6.78 kcal/mol	-9.99 kcal/mol	siRNA_wizard	Tobacco Mosaic Virus	Movement Protein
548	528	GAAGATTACAAACCGTGAGAGA	38.1	-1.17 kcal/mol	76.47 %	1.3	-3.90 kcal/mol	-14.98 kcal/mol	siRNA_wizard	Tobacco Mosaic Virus	Movement Protein
549	594	GGAGAGATGTCCTATGTCAT	42.86	-1.33 kcal/mol	35.84 %	4.1	-8.98 kcal/mol	-13.59 kcal/mol	siRNA_wizard	Tobacco Mosaic Virus	Movement Protein
550	609	GTCAATCAGGGTTGCAAGIT	42.86	-1.19 kcal/mol	27.76 %	4	-6.96 kcal/mol	-14.60 kcal/mol	siRNA_wizard	Tobacco Mosaic Virus	Movement Protein
551	689	GTCAGTGGCGAACAGAACCT	52.38	-1.79 kcal/mol	62.69 %	2.43	-5.55 kcal/mol	-14.28 kcal/mol	siRNA_wizard	Tobacco Mosaic Virus	Movement Protein
552	694	GTGCCGAACAAAGAACATAGA	42.86	-0.17 kcal/mol	76.13 %	0.81	-5.80 kcal/mol	-12.67 kcal/mol	siRNA_wizard	Tobacco Mosaic Virus	Movement Protein
553	478	GUUGGUUGUGUAUUGUUUAU	37	-0.22 kcal/mol	69.51 %	1.71	-7.02 kcal/mol	-9.40 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
554	692	GUGCCGAACAAGAACUUAUA	42	-0.17 kcal/mol	76.15 %	0.81	-5.80 kcal/mol	-12.67 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
555	553	GCCCCAUUGGAACUUCAGAAA	47	-0.32 kcal/mol	59.90 %	1.31	-6.45 kcal/mol	-9.03 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
556	552	GGCCC AUUGGAACUUCAGAGA	53	-0.33 kcal/mol	58.28 %	1.4	-6.43 kcal/mol	-9.03 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
557	548	GGAGGGCCAUGGGAACUUA	58	-3.47 kcal/mol	75.75 %	0.62	-6.31 kcal/mol	-13.87 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
558	448	GGGUUUCUGUUCGGCUUUCU	53	-1.55 kcal/mol	29.75 %	2.99	-8.21 kcal/mol	-12.20 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
559	355	CAAGGUCGUUCCCCAUUAU	42	-1.03 kcal/mol	58.58 %	1.69	-7.56 kcal/mol	-11.62 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
560	782	GUGGCCGAUUCGAUUCGU	58	-2.45 kcal/mol	29.39 %	4.56	-8.36 kcal/mol	-17.52 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
561	150	CAGAGGGUAGACCUUCUUA	42	-3.40 kcal/mol	32.25 %	4.41	-6.45 kcal/mol	-12.91 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein
562	1	GGCUCUAGUUGUUAAAAGGA	42	-0.47 kcal/mol	46.76 %	2.42	-8.80 kcal/mol	-11.42 kcal/mol	whitehead	Tobacco Mosaic Virus	Movement Protein