Trends in Peptide and Protein Sciences
Volume 8 (2023): e6 (Supplementary)

DOI: https://doi.org/10.22037/tpps.v8i1.42723

Original Article

In-Silico Prediction of Common siRNA Targeting Protein Coding Sequence of *NS5* Gene of West Nile and Japanese Encephalitis Virus

Rahatul Islam[a,b], Asif Shahriar[a,c]*, Mohd. Faijanur - Rob Siddiquee[d], Nour Fatema[c,e], Muhammad Ramiz Uddin[f], Mrityunjoy Acharjee[c]*, Khosnur Jahan Aurin[g], Md Mukhlesur Rahman Shepon[b], Ashikur Rahaman Bhuiyan[b], Marline Gomes[h]

[a] *Department of Biochemistry and Molecular Medicine, School of Medicine, West Virginia University, Morgantown, WV, USA.*
[b] *Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet, Bangladesh.*
[c] *Department of Microbiology, Stamford University Bangladesh, 51 Siddeswari Road, Dhaka-1217, Bangladesh.*
[d] *Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka 1000, Bangladesh.*
[e] *Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR, USA.*
[f] *Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK, USA.*
[g] *Department of Medicine and Surgery, Ad-din Sakina Medical College, Jashore, Bangladesh.*
[h] *Department of Pharmacy, University of Asia Pacific, Dhaka, Bangladesh.*

Table S1. Multiple sequence alignment comparison of the CDS of West Nile and Japanese Encephalitis Virus

CDS Gene Name (West Nile Virus) Accession no: NC_009942	CDS Gene Name (Japanese Encephalitis Virus) Accession no: NC_001437	MSA result (Identity)
Capsid C	core protein C	70.9%
peptide pr	Protein pr	65.2%
Membrane Glycoprotein M	Membrane Glycoprotein M	68.4%
Envelope E	Envelope E	69.5%
NS1 (Non-Structural protein 1)	NS1 (Non-Structural protein 1)	70.5%
NS2A (Non-Structural protein 2A)	NS2A (Non-Structural protein 2A)	63.5%
NS2B (Non-Structural protein 2B)	NS2B (Non-Structural protein 2B)	66.3%
NS3 (Non-Structural protein 3)	NS3 (Non-Structural protein 3)	70.0%
NS4A (Non-Structural protein 4A)	NS4A (Non-Structural protein 4A)	69.2%
NS4B (Non-Structural protein 4B)	NS4B (Non-Structural protein 4B)	64.0%
NS5 (Non-Structural protein 5)	NS5 (Non-Structural protein 5)	72.4%

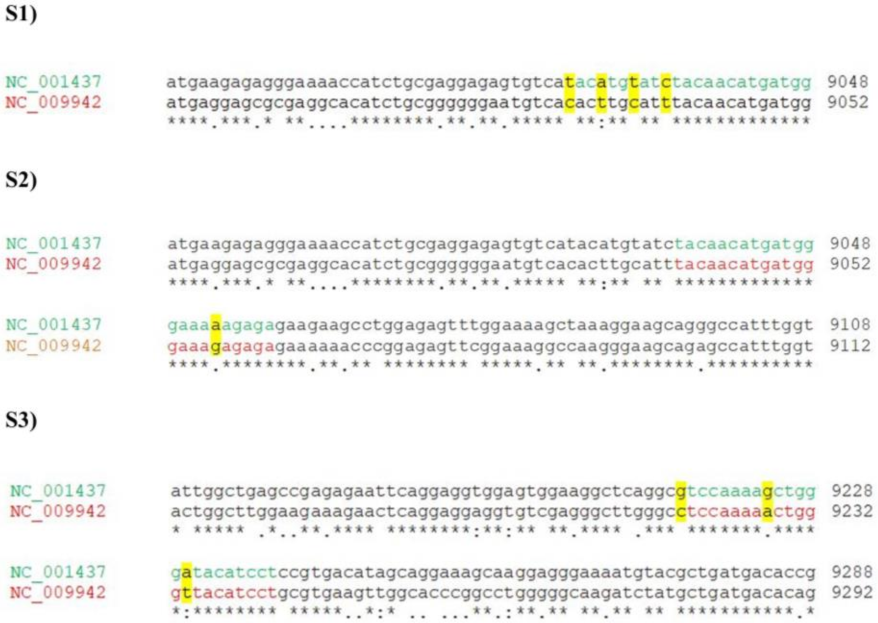
Table S2. siRNA's functionality predicted by siDirect webserver. Only S2 siRNA molecule followed all three algorithms (Ui-Tei, Reynolds & Amarzguioui)

Alias	Target Sequence (21nt target + 2nt overhang)	Target position in mRNA (NC_009942)	Similar Target position in mRNA (NC_001437)	RNA oligo sequences 21nt guide (5′→3′) 21nt passenger (5′→3′)	Functional siRNA selection: Ui-Tei Reynolds Amarzguioui
S1	CACTTGCATTTACAACATGATGG	9029-9052	9026-9048	AUCAUGUUGUAAAUGCAAGUG CUUGCAUUUACAACAUGAUGG	U A
S2	TACAACATGATGGGAAAGAGAGA	9040-9062	9036-9058	UCUCUUUCCCAUCAUGUUGUA CAACAUGAUGGGAAAGAGAGA	U R A
S3	CTCCAAAAACTGGGTTACATCCT	9220-9242	9216-9238	GAUGUAACCCAGUUUUUGGAG CCAAAAACUGGGUUACAUCCU	U A

Table S3. Docking interaction analysis of the best siRNA molecule S2 with human Argonaute-2 protein

Alias	**RNA oligo sequences 21nt guide (5′→3′)**	**Docking score**	**Interacting residues in Ago2 domains**				
			N-terminal (36-166)	L1 (176-226)	PAZ (238-365)	Mid (429-511)	PIWI (517-818)
S2	UCUCUUUCCCAUCAUGUUGUA	-327.03	N/A	ASP218 SER220 ALA221*	ARG286 LYS355 THR361 SER362 ILE365*# ARG366 THR368# ALA369 ARG375#	N/A	LEU522#,PRO523 GLY524#,LYS525# THR526*#,TYR529*# GLN545#,CYS546# VAL547#,GLN548# LYS566*#,THR599# HIS600#,PRO601 PRO602,ALA603 ASP605,LYS607 LYS608,ARG635* LYS709#,ARG710*# ARG714#,GLN757# GLY758#,THR759*# SER760#,ARG761# ARG792*#,TYR804# HIS807,LEU808 PHE811#,ARG812*# ARG814,TYR815# VAL818,ASP819 GLU821,ASP823 ALA859#
Control	UUCACAUUGCCCAAGUCUUU	-1081.52	LYS65 CYS66 PRO67 ARG68 VAL70 ASP95 ARG97 GLY121 GLY123 LYS124 ASP125	PRO176 VAL177 GLY178 ARG179	LYS266 ILE269 ARG277 LYS278 TYR279 ARG280 PHE294 LEU296 VAL308 TYR311 PHE312 ARG315 HIS316 GLN332 LYS335 HIS336 THR337 TYR338 LEU339 PRO340 ARG351 MET364 ILE365 THR368 ARG375	N/A	LEU522,GLY524 LYS525,THR526 TYR529,LYS533 THR544,GLN545 CYS546,VAL547 GLN548,LYS550 ASN551,GLN558 THR559,ASN562 LEU563,LYS566 LYS570,VAL598 THR599,HIS600 GLY670,VAL671 SER672,GLY674 GLN675,LYS709 ARG710,ARG714 HIS753,ALA754 GLY755,ILE756 GLN757,GLY758 THR759,SER760 ARG761,PRO762 TYR790,ARG792 CYS793,ARG795 SER796,VAL797 SER798,TYR804 PHE811,ARG812 TYR815,ALA859

#Residues matched with control; *Residues matched with previous studies



S1)

```
NC_001437      atgaagagagggaaaaccatctgcgaggagagtgtcatacatgtatctacaacatgatgg  9048
NC_009942      atgaggagcgcgaggcacatctgcgggggaatgtcacacttgcatttacaacatgatgg  9052
               ****.***.* **....********.**.**.***** **:** ** *************
```

S2)

```
NC_001437      atgaagagagggaaaaccatctgcgaggagagtgtcatacatgtatctacaacatgatgg  9048
NC_009942      atgaggagcgcgaggcacatctgcgggggaatgtcacacttgcatttacaacatgatgg  9052
               ****.***.* **....********.**.**.***** **:** ** *************

NC_001437      gaaaaagagagaagaagcctggagagtttggaaaagctaaaggaagcagggccatttggt  9108
NC_009942      gaaagagagagaaaaaacccggagagttcggaaaggccaaggaagcagagccatttggt  9112
               ****.********.**.** ******** *****.** **.********.**********
```

S3)

```
NC_001437      attggctgagccgagagaattcaggaggtggagtggaaggctcaggcgtccaaaagctgg  9228
NC_009942      actggcttggaagaaagaactcaggaggaggtgtcgagggcttgggcctccaaaaactgg  9232
               * ***** .*..**.**** ********:**:** **.**** .*** *******.****

NC_001437      gatacatcctccgtgacatagcaggaaagcaaggagggaaaatgtacgctgatgacaccg  9288
NC_009942      gttacatcctgcgtgaagttggcacccggcctggggcaagatctatgctgatgacacag  9292
               *:******** *****..*:* .. ...**.:**.** **.** ** ***********.*
```

Figure S1: Prediction of three putative siRNAs which have higher similarity in both of the viruses analysed by sequence alignment of Clustal W webserver. Accesion no: “NC_009942” is the genomic sequence of West Nile virus and NC_001437 is denoted for Japanese Encephalitis virus. S1, S2 and S3 denoted as the potential siRNAs target of the final siRNA molecule S1, S2 and S3. The consensus symbol represents (*asterisk) as fully conserved, (: Colon) as identical, (. period) as similar.

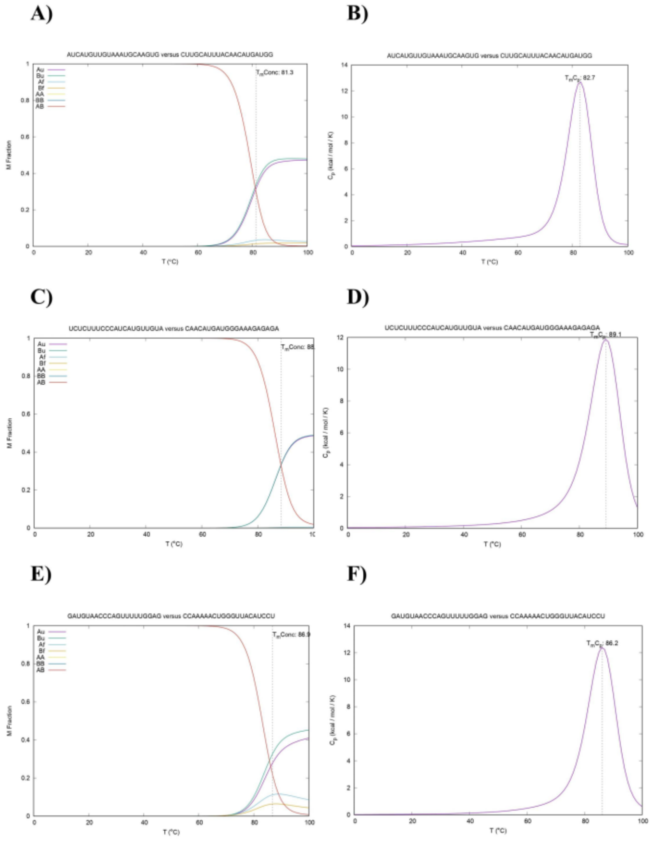
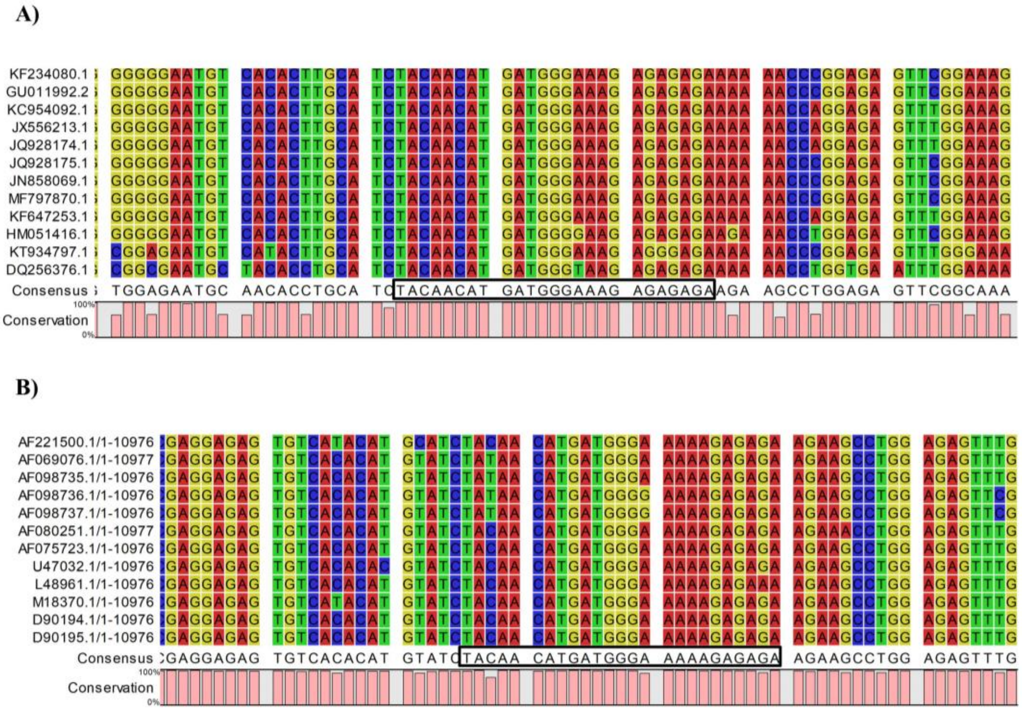


Figure S2: Heat capacity and concentration plot analysis of putative siRNAs. For siRNA molecule S1, the Tm (Conc) is shown in (A) and Tm (Cp) in (B). For S2 molecule, Tm (Conc) is (C) and Tm (Cp) in (D). For S3 siRNA, Tm (Conc) is (E) and Tm (Cp) is shown in (F).



```
KF234080.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCCGGAGA GTTCGGAAAG
GU011992.2  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCCGGAGA GTTCGGAAAG
KC954092.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCAGGAGA GTTTGGAAAG
JX556213.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCAGGAGA GTTTGGAAAG
JQ928174.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCAGGAGA GTTTGGAAAG
JQ928175.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCCGGAGA GTTCGGAAAG
JN858069.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCCGGAGA GTTCGGAAAG
MF797870.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCCGGAGA GTTCGGAAAG
KF647253.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGAAAG AGAGAGAAAA AACCAGGAGA GTTTGGAAAG
HM051416.1  GGGGGAATGT CACACTTGCA TCTACAACAT GATGGGGAAG AGAGAGAAGA AACCTGGAGA GTTCGGAAAG
KT934797.1  CGGAGAATGT CATACTTGCA TCTACAACAT GATGGGAAAG AGGGAGAAAA AACCCGGAGA GTTTGGGAAA
DQ256376.1  CGGCGAATGC TACACCTGCA TCTACAACAT GATGGGTAAG AGAGAGAAAA AACCTGGTGA ATTTGGAAAA
Consensus   TGGAGAATGC AACACCTGCA TCTACAACAT GATGGGAAAG AGAGAGAAGA AGCCTGGAGA GTTCGGCAAA
```

```
AF221500.1/1-10976  GAGGAGAG TGTCATACAT GCATCTACAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
AF069076.1/1-10977  GAGGAGAG TGTCACACAT GTATCTATAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
AF098735.1/1-10976  GAGGAGAG TGTCACACAT GTATCTATAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
AF098736.1/1-10976  GAGGAGAG TGTCACACAT GTATCTATAA CATGATGGGG AAAAGAGAGA AGAAGCCTGG AGAGTTCG
AF098737.1/1-10976  GAGGAGAG TGTCACACAT GTATCTATAA CATGATGGGG AAAAGAGAGA AGAAGCCTGG AGAGTTCG
AF080251.1/1-10977  GAGGAGAG TGTCACACAT GTATCTACAA CATGATGGGA AAAAGAGAGA AGAAACCTGG AGAGTTTG
AF075723.1/1-10976  GAGGAGAG TGTCACACAT GTATCTACAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
U47032.1/1-10976    GAGGAGAG TGTCACACAC GTATCTACAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
L48961.1/1-10976    GAGGAGAG TGTCACACAT GTATCTACAA CATGATGGGA AAAAGAGAAA AGAAGCCTGG AGAGTTTG
M18370.1/1-10976    GAGGAGAG TGTCATACAT GTATCTACAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
D90194.1/1-10976    GAGGAGAG TGTCACACAT GTATCTACAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
D90195.1/1-10976    GAGGAGAG TGTCACACAT GTATCTACAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
Consensus           GAGGAGAG TGTCACACAT GTATCTACAA CATGATGGGA AAAAGAGAGA AGAAGCCTGG AGAGTTTG
```

Figure S3: Multiple sequence alignment (MSA) analysis against all other strains of both viruses, e.g., 91 strains for West Nile and 96 strains for the Japanese Encephalitis Virus. siRNA S2 molecule showed 100% conservancy in MSA analysis.