ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(8): 661-668 (2022)

BIOINFORMATION
Discovery at the interface of physical and biological sciences
BIOMEDICAL INFORMATICS
www.bioinformation.net
Volume 18(8)
Research Article

Received July 2, 2022; Revised August 31, 2022; Accepted August 31, 2022, Published August 31, 2022
DOI: 10.6026/97320630018661

Edited by P Kangueane
Citation: Prajapati *et al.* Bioinformation 18(8): 661-668 (2022)

Towards the diagnosis of dengue virus and its serotypes using designed CRISPR/Cas13 gRNAs

Archana Prajapati*, AshmitaTandon* & Vikrant Nain#

School of Biotechnology, Gautam Buddha University, Greater Noida-201312, Uttar Pradesh, India; #Corresponding author,*Equal contribution

Institution URL:
https://www.gbu.ac.in/

Author contacts:
Archana Prajapati - E-mail: btphd2018003@gbu.ac.in
Vikrant Nain – E-mail: vikrant@gbu.ac.in

Abstract:
Dengue Virus (DENV) is a mosquito-borne virus that is prevalent in the world's tropical and subtropical regions. Therefore, early detection and surveillance can help in the management of this disease. Current diagnostic methods rely primarily on ELISA, PCR, and RT-PCR, among others, which can only be performed in specialized laboratories and require sophisticated instruments and technical expertise.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(8): 661-668 (2022)

CRISPR-based technologies on the other hand have field-deployable viral diagnostics capabilities that could be used in the development of point-of-care molecular diagnostics. The first step in the field of CRISPR-based virus diagnosis is to design and screen gRNAs for high efficiency and specificity. In the present study, we employed a bioinformatics approach to design and screen DENV CRISPR/Cas13 gRNAs for conserved and serotype-specific variable genomic regions in the DENV genome. We identified one gRNA sequence specific for each of the lncRNA and NS5 regions and identified one gRNA against each of DENV1, DENV2, DENV3, and DENV4 to distinguish the four DENV serotypes. These CRISPR/Cas13 gRNA sequences will be useful in diagnosing the dengue virus and its serotypes for *in vitro* validation and diagnostics.

Keywords: Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), Cas13, Dengue virus, serotypes, diagnosis, gRNA pool, secondary structure, free energy.

Background:

Dengue is a major arthropod-borne viral disease caused by the dengue virus (DENV), which is widespread in tropical and subtropical regions of the world and is spread by the *Aedesaegyptis* and *Aedesalbopictus* mosquitos **[1]**. Its clinical manifestations range from mild Dengue Fever (DF) to severe Dengue Haemorrhagic Fever (DHF) and Dengue shock syndrome (DSS) **[1]**. Dengue virus (DENV) is an enveloped, positive-sense, ss-RNA virus of the Flaviviridae family and the genus *Flavivirus* **[2], [3].** DENV has four well-known serotypes (DENV1, DENV2, DENV3, DENV4) based on antigen cross-reactivity, with each serotype having distinct genotypes **[2][3]**. The DENV genome (~11 kb) comprises one open reading frame (ORF) flanked by 5′ and 3′ UTR regions **[2][4]**. Its ORF encodes a single polyprotein, which is cleaved post-transactionally to generate three structural proteins: capsid (C), pre-membrane/membrane (prM/M), and envelope (E), as well as seven non-structural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5)(**Figure 1**) **[1]**. There are currently no particular antiviral medicines or vaccinations available to treat or prevent dengue infection**[2],[5],[6]**. The sole alternative that relies on an early and accurate dengue diagnosis is symptomatic and supportive therapy of the disease. Dengue infection is diagnosed through the isolation and characterization of the DENV virus, serological tests that detect DENV-related antigens and/or antibodies in patient plasma/serum, and molecular diagnosis **[4], [5], [7]**. The DENV nucleic acid (RNA) is detected using an isothermal nucleic acid sequence-based amplification assay (NASBA), a reverse transcriptase-polymerase chain reaction (RT-PCR) (one-step or nested RT-PCR assay), and a real-time RT-PCR assay; these molecular methods require skilled technicians and specialized laboratories **[4] [5] [8]**. As a result, more precise and point-of-care (POC) molecular approaches that enable improved clinical treatment, proper disease management, surveillance, and prevention of dengue outbreaks are required. Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR), is a highly variable locus consisting of short and identical repeated palindromic sequences that are naturally found in the genome of bacteria, separated by large spacer sequences**[9]**. The CRISPR locus is transcribed to generate pre-crRNA, which is then processed by CRISPR-associated (Cas) proteins to generate mature CRISPR guide RNA (gRNAs) **[10]**. CRISPR-associated (Cas) proteins specifically act as RNA-guided endonucleases which together with gRNA induce indiscriminate cleavage of target nucleic acid thereby protecting the bacterial cell against any invading bacteriophage **[9], [10]**. CRISPR-Cas-based diagnosis relies on the identification of viral nucleic acid sequences by gRNAs followed by non-specific cleavage of viral nucleic acid by Cas **[11], [12]**. This reactionis usually coupled with probe molecules that generate a fluorescent/visual signal which is detected **[11], [12]**. SHERLOCK (Specific High Sensitivity Enzymatic Reporter Unlocking), DETECTR (DNA endonuclease-targeted CRISPR transReporter),CARVER (Cas13-assisted restriction of viral expression and readout), PAC-MAN (Prophylactic Antiviral CRISPR in human cells), SHINE (Streamlined Highlighting of Infections to Navigate Epidemics), All-In-One Dual CRISPR-Cas12a (AIOD-CRISPR)are examples of CRISPR-Cas molecular methods used in disease diagnosis **[13][15]**. In this study, we have designed DENV-specific CRISPR-Cas13 gRNAs for the conserved and variable/hyper variable genomic regions among four DENV serotypes using bioinformatics tools. These CRISPR-Cas13 gRNA sequences serve as molecular detectors that provide point-of-care molecular diagnosis useful in the detection and clinical evaluation of the dengue virus and its serotypes.

Methodology:

The dataset of complete genome sequences of four Dengue virusserotypesDENV1 (NC_001477), DENV2 (NC_001474), DENV3 (NC_001475), DENV4 (NC_002640) were compiled from the Dengue Virus Variation Resource Centre at National Centre for Biotechnology Information (NCBI) **[16]**. The Multiple Sequence Alignment (MSA) of these sequences was performed using Bio Edit v 7.2 **[17]**. The CRISPR-Cas13 gRNAs were designed the using CHOPCHOP v2 CRISPR designing web tool **[18]**. These gRNA sequences were visually inspected for the number of mismatches and number of bases conserved to obtain potentialCRISPR-Cas13 gRNA sequences showing high specificity (maximum target annealing) and low off-target selectivity (minimum off-target binding) for the conserved and variable portions in the DENV genome. The secondary structures and free energies of gRNA sequences were predicted using the default parameters in RNA fold web-server **[19]**.

ISSN 0973-2063 (online) 0973-8894 (print)

Bioinformation 18(8): 661-668 (2022)

Table 1: Free energies of DENV specific gRNA sequences

S. No.	Genomic region	CRISPR/Cas13 gRNA	Start position	End position	No. of mismatches	Mismatch in seed region(11-18 bp)	Self-complementarity	Free Energy (ΔG in kcal/mol)
1	**lncRNA**	**TAGAGGAGACCCCCCCGAAACAAAAAAC**	**293**	**321**	**1**	**no**	**0**	**-0.7**
2	lncRNA	AGAGGAGACCCCCCCGAAACAAAAAACA	294	322	1	no	0	-0.7
3	lncRNA	CTGGGAAAGACCAGAGATCCTGCTGTCT	333	361	1	no	1	-6.1
4	lncRNA	AGATCCTGCTGTCTCCTCAGCATCATTC	347	375	1	no	2	-6.2
5	lncRNA	CAGAGATCCTGCTGTCTCCTCAGCATCA	344	372	1	no	3	-6.2
6	lncRNA	AGAGATCCTGCTGTCTCCTCAGCATCAT	345	373	1	no	3	-6.2
s7	lncRNA	TTAGAGGAGACCCCCCCGAAACAAAAAA	292	320	2	no	0	-0.7
8	lncRNA	AGGAGACCCCCCCGAAACAAAAAACAGC	296	324	2	no	0	-0.7
9	lncRNA	ACGCTGGGAAAGACCAGAGATCCTGCTG	330	358	2	no	1	-5.9
10	lncRNA	CGCTGGGAAAGACCAGAGATCCTGCTGT	331	359	2	no	1	-5.9
11	lncRNA	ATCCTGCTGTCTCCTCAGCATCATTCCA	349	377	2	no	2	-6.2
12	**NS5**	**ATGTATGCCGATGACACCGCAGGATGGG**	**1588**	**1616**	**4**	**Yes, 1**	**0**	**-0.9**
13	NS5	TGTATGCCGATGACACCGCAGGATGGGA	1589	1617	5	Yes, 1	0	-2
14	NS5	ATGACACCGCAGGATGGGATACAAGAAT	1598	1626	5	Yes, 1	0	-1.3
15	NS5	TGACACCGCAGGATGGGATACAAGAATC	1599	1627	5	Yes, 1	0	-1.3

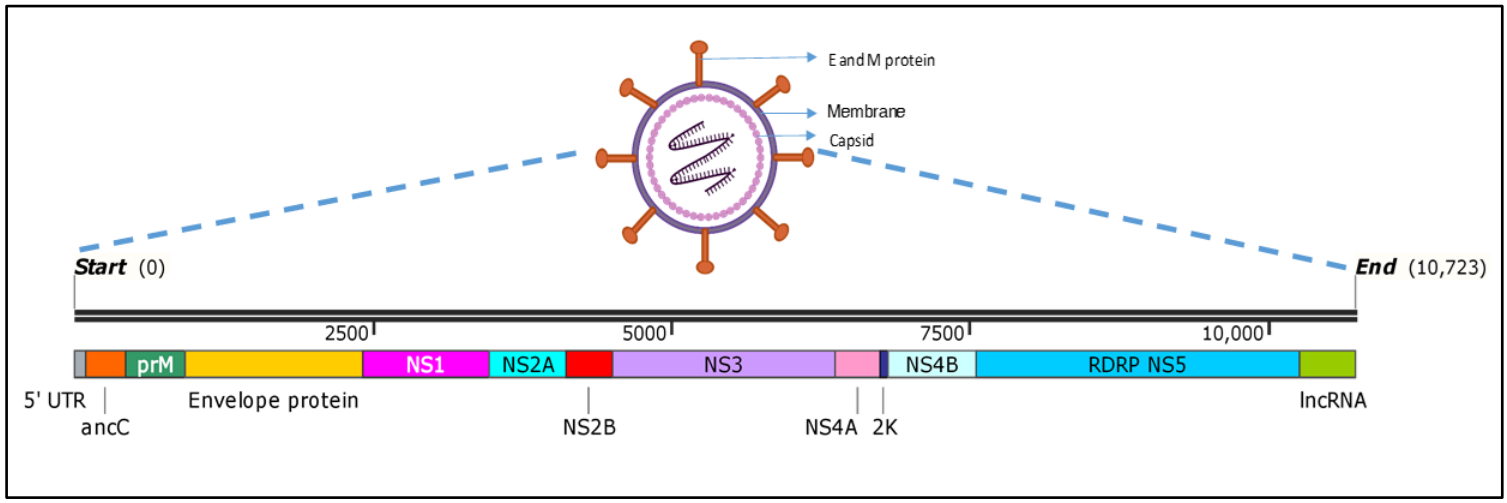

Figure 1: Genome organization of dengue virus

Results and Discussion:

We identified the conserved and variable sequences across the genome of four common DENV serotypes. Subsequently, these selected genomic regions were used to design CRISPR-Cas13gRNAsequences.

Sequence retrieval and multiple sequence alignment:

The extensive whole-genome Multiple sequence alignment (MSA) of four DENV reference serotypes DENV1 (Gen Bank: NC_001477.1), DENV2 (Gen Bank: NC_001474.2), DENV3 (Gen Bank: NC_001475.2), and DENV4 (Gen Bank: NC_002640.1) revealed lncRNA and NS5 genomic regions that are conserved across all DENV serotypes, while variable sequences found in the NS2A and NS2B gene. The NS2A gene is involved in coordination during RNA packaging and replication whereas the NS2B gene serves as a co-factor in the structural activation of DENV serine protease NS3 **[1]**. NS2B gene also assists in viral replication and blocks IFN- induced signal transduction **[1]**. Gene NS5 codes for

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(8): 661-668 (2022)

methyltransferase domain (located at residues 1-269 amino acids) and RNA-dependent RNA polymerase (located at residue 270-900 amino acids) **[1]**.

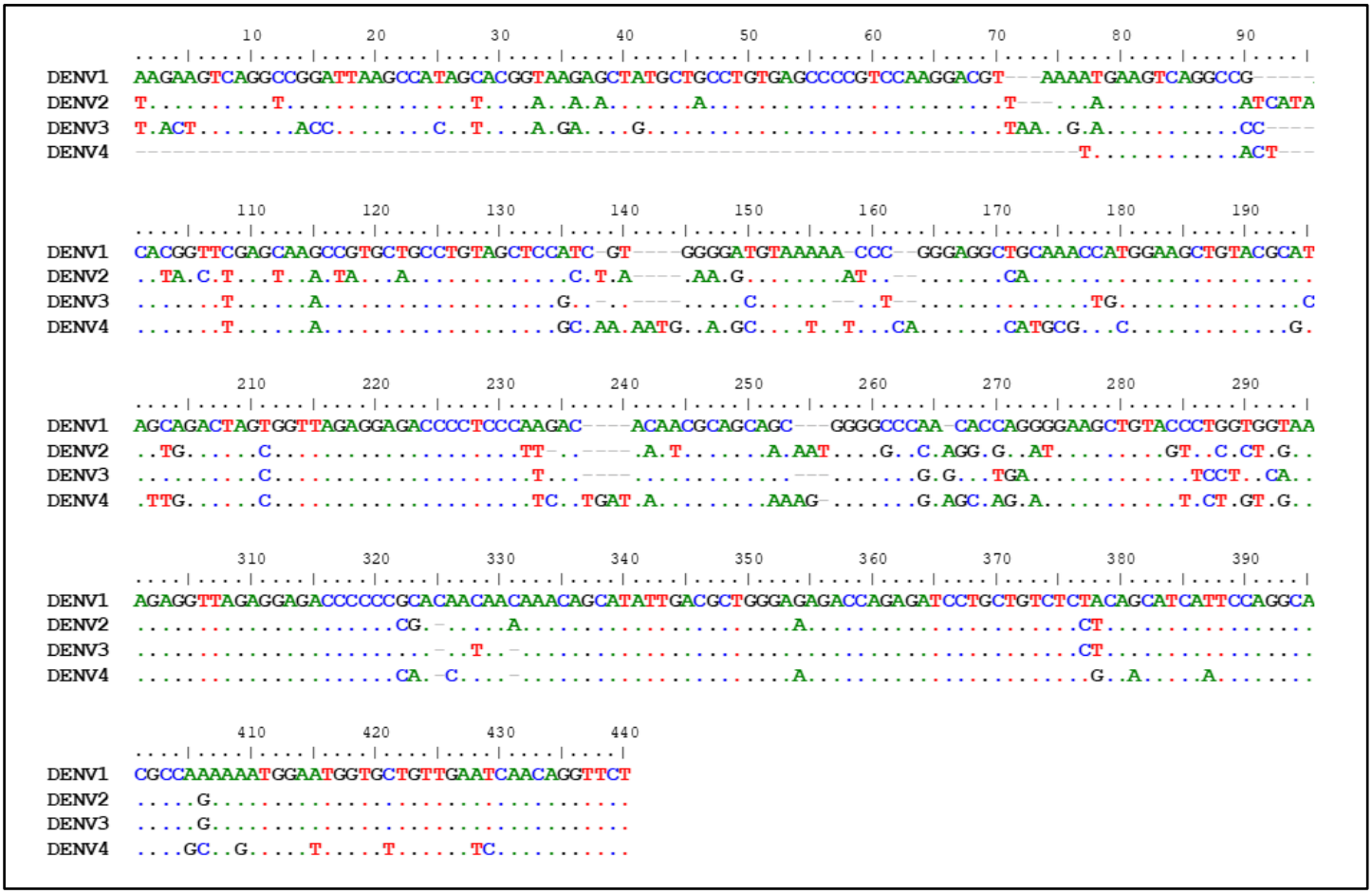

Figure 2: Multiple sequence alignment of lncRNA in four dengue serotypes

Designing of CRISPR/Cas13 gRNAs against conserved regions in DENV genome:

MSA of reference DENV serotypes revealed that the most conserved sequences are in the NS5 and long non-coding RNA (lncRNA) regions (**Figures 2 and 3**). As a result, genomic sequences of the DENV2 (Gen Bank: NC_001474.2) lncRNA (0.425 kb) and NS5 (2.7 kb) regions were used as a query in the CHOPCHOP tool to design CRISPR/Cas13 gRNAs for Dengue diagnosis. The CHOPCHOP online tool returned 293 gRNA sequences for DENV lncRNA and 1973 gRNAs for NS5. These gRNA sequences were visually inspected for the number of conserved nucleotides in the MSA, and the sequences with the highest number of conserved nucleotides among four DENV serotypes were selected, yielding a total of 251 gRNA sequences for lncRNA and 268 gRNA sequences for NS5. To narrow the search even further and to improve gRNA specificity against the target, gRNA sequences were re-examined for the number of sequence mismatches in the lncRNA and NS5 regions, and the gRNA sequences with the least mismatches were selected. We selected a total of 17 lncRNA-specific sequences by selecting gRNA sequences with a mismatch of less than or equal to two nucleotides and a total of 10 NS5 specific gRNA sequences by selecting gRNA sequences with a mismatch of less than or equal to five nucleotides because lncRNA was found to be more conserved among DENV serotypes than NS5, as indicated by MSA. Furthermore, gRNA sequences with no mismatches in the core region, i.e. within a stretch of 11-18nts, were highly target-specific and anneal quickly with complementary sequences in the target genome **[20]**. As a result, we visually inspected the lncRNA and NS5 specific gRNA sequences for mismatches and eliminated those with mismatches within the core region. Our findings showed that 11 of the 17 lncRNA gRNA sequences have no mismatches within an 8-base sequence stretch spanning 11-18 nucleotides, while 4 of 10 NS5 gRNA sequences have only one mismatch within the central seed region (**Table 1**).

Prediction of gRNA secondary structure:

There is a link between the secondary structure of gRNA and gene editing efficiency. The formation of secondary structures in gRNA has a significant impact on cleavage efficiency. More specifically, when the Gibbs free energy (ΔG) for the formation of the most

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(8): 661-668 (2022)

stable RNA structure is between -2 and 0 kJmol^{-1}, cleavage efficiencies are highest **[21]**. As a result, the gRNA can invade the target strand without the aid of the gRNA scaffold. RNA fold Web Server was used to evaluate the secondary structures and minimum free energy of lncRNA and NS5 specific gRNA sequences (**Table 1 and Figure 4**). We selected two DENV-specific gRNAs based on the minimum free energy, least number of mismatches among DENV serotypes, zero self-complementarity of gRNA, GC content, and no mismatch within the central region and gRNA secondary structure. One sequence **(TAGAGGAGACCCCCCCGAAACAAAAAAC)** is specific for targeting the lncRNA and the other **(ATGTATGCCGATGACACCGCAGGATGGG)** is specific for targeting the NS5 gene (**Table 1 and Figure 4**). These two CRISPR-Cas13 gRNA sequences have a low possibility of self-annealing; allowing them to recognize and bind to the target site within the DENV genome's conserved regions.

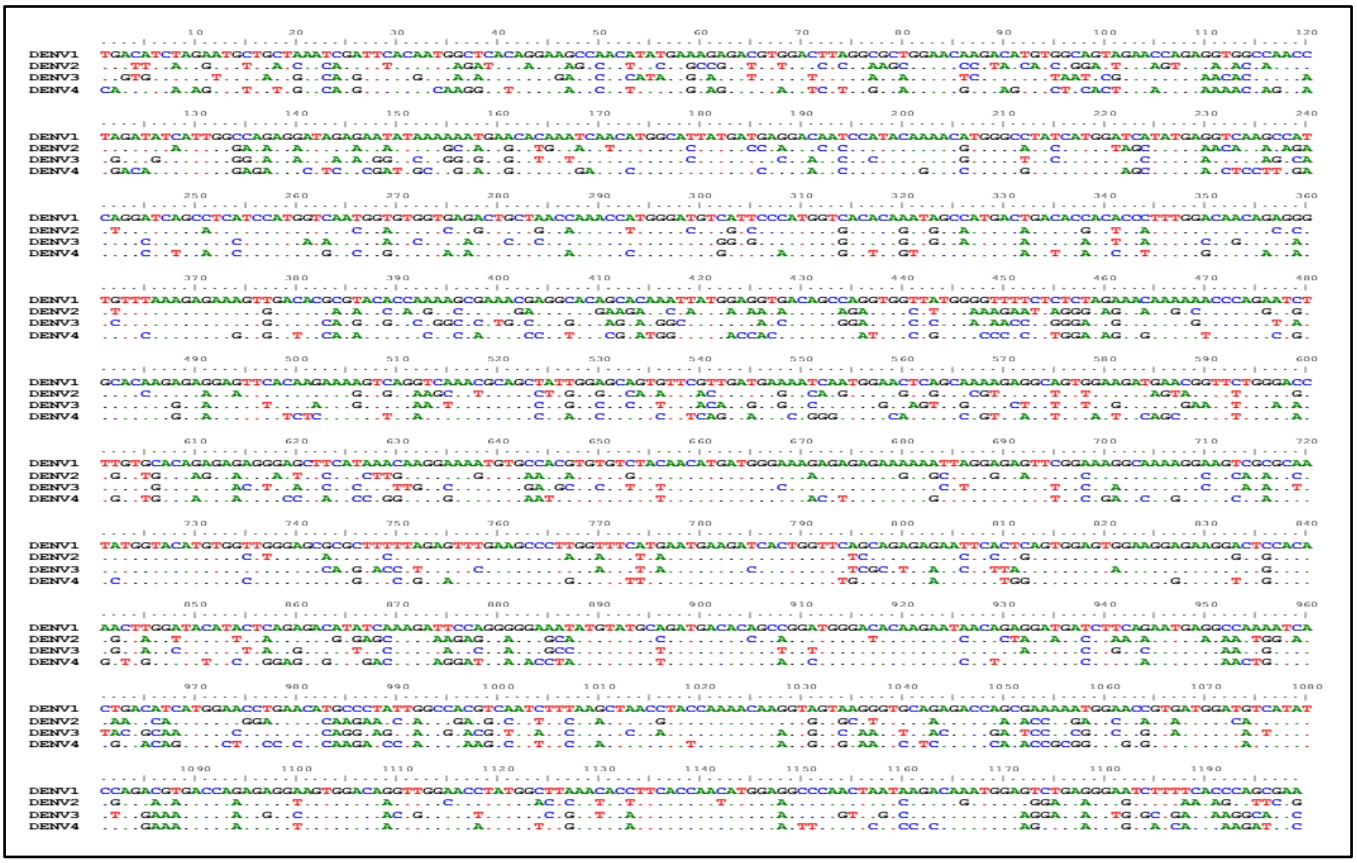

Figure 3: Multiple sequence alignment of NS5 (700-1200 nt positions) in four dengue serotypes

Designing of serotype-specificCRISPR/Cas13 gRNAs:
Using reference genomic sequences of the NS2A and NS2B genes from four DENV serotypes, the CHOPCHOP online tool identified a total of 3,072 CRISPR-Cas13 gRNA sequences, with 776, 772, 763, and 761 gRNA sequences for the NS2A and NS2B regions of DENV1, DENV2, DENV3, and DENV4 respectively. The number of conserved bases in the NS2A and NS2B regions of these gRNA was visually inspected in MSA. The gRNA sequences showing the lowest number of conserved bases among the serotypes was selected to provide the most variable serotype-specific targets. It narrowed down serotype-specific gRNAs to 143, 173, 136, and 275 for DENV1, DENV2, DENV3, and DENV4 respectively. To further reduce the number of gRNA sequences we identified gRNA sequences with the highest specificity and selected only those gRNAs that show the least conserved bases not more than 3 bases among serotypes. This screening further reduced the gRNA sequences to 5, 19, 14, and 11for DENV1, DENV2, DENV3, and DENV4 respectively (**Supplementary Table 1**, selected gRNA sequences).

Prediction of secondary structure in gRNA:
The secondary structures and minimum free energy of DENV serotype-specific gRNAs were predicted to aid in the selection of gRNAs that are more efficient and specific to target RNA. For the targets DENV1, DENV2, DENV3, and DENV4, we obtained 1, 5, 5, and 2 potential serotype-specific gRNA sequences with zero or nearly zero free energy, respectively (**Figure 5**). These gRNAs were then screened for minimum free energy, secondary structure, and the fewest number of conserved bases across DENV serotypes. We got one gRNA for each of the DENV serotypes, which were TCAAAACAACTTTTTCATTGCACTATGC, AACCCTCTCAAGAACCAGCAAGAAAAGG, TTAGCTTGAAAGACACACTCAAAAGGAG, and AACAGCACTCATCCTAGGAGCCCAAGCT for DENV1, DENV2,

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(8): 661-668 (2022)

DENV3, and DENV4 respectively (**Figure 5**, gRNA in bold). These gRNA sequences have the potential for efficient binding with the genome of selected DENV serotypes. CRISPR-based diagnostics are next generation biosensing techniques for detection of viral and bacterial infections. SHERLOCK has been used for detection of Dengue, Zika, bacterial infections, and tumour changes in cell-free DNA using CRISPR-Cas13a **[13], [22]**. These *In-silico* designed28nt long gRNAs sequences can detect complementary sequences in the DENV genome. These DENV-specific gRNA sequences coupled with Cas13 enzyme can be used in sequence-specific cleavage of the target genome and reporter molecules which can be developed into molecular test useful in identification of dengue virus and its serotypes. Molecular tests like these provide rapid detection and quick identification of virus in clinical samples from dengue infected patients providing an early and accurate disease diagnosis.

Conclusions:

Early and serotype specific diagnosis is important in management of spread of the any diseases and treatment of patients. CRISPR-Cas13 based diagnostics are highly sensitive and specific as well as cost-effective, fast, and can be used as point of care diagnostics. The present study gRNA sequences specific for DENV were designed to provide molecular diagnosis. These gRNA will be valuable for *in vitro* evaluation the sensitivity and efficiency of CRISPR-Cas13-based POC diagnostics.

List of Abbreviations:

DENV: Dengue virus; ZIKV: Zika virus; ELISA: Enzyme-linked Immunosorbent Assay; PCR: Polymerase Chain Reaction; RT-PCR: reverse transcriptase-polymerase chain reaction; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeat; Cas: CRISPR associated proteins; gRNA: guide RNA; lncRNA: Long non-coding RNA; DF: Dengue Fever; DHF: Dengue Haemorrhagic Fever; DSS: Dengue shock syndrome; ORF: Open Reading Frame; UTR: Untranslated region; C: Capsid; prM/M: pre-membrane/membrane; NS: Non-structural; E: envelope; NASBA: Nucleic acid sequence-based amplification assay; SHERLOCK: Specific High Sensitivity Enzymatic Reporter Unlocking; DETECTR: DNA endonuclease-targeted CRISPR transReporter; CARVER: Cas13-assisted restriction of viral expression and readout; PAC-MAN: Prophylactic Antiviral CRISPR in human cells; SHINE : Streamlined Highlighting of Infections to Navigate Epidemics; AIOD-CRISPR:All-In-One Dual CRISPR-Cas12a; FELUDA: FNCas9 Editor Limited Uniform Detection Assay; MSA: Multiple sequence alignment; IFN: Interferons; ΔG: Gibbs free energy; RAA: recombinase aided amplification; STOP: SHERLOCK testing in one pot; CREST: Cas13-based, rugged, equitable, scalable testing; HPV: Human papillomavirus.

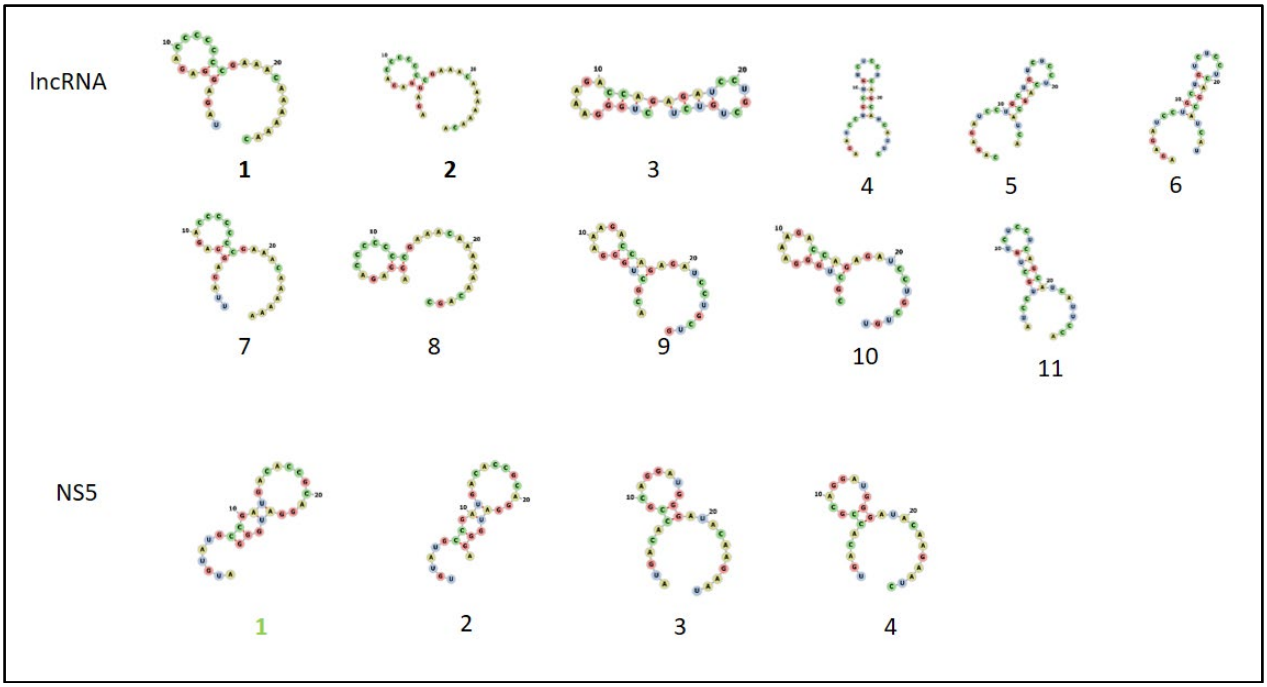

Figure 4: Secondary structure of selected gRNA designed against conserved regions of DENV lncRNA and NS5 regions

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(8): 661-668 (2022)

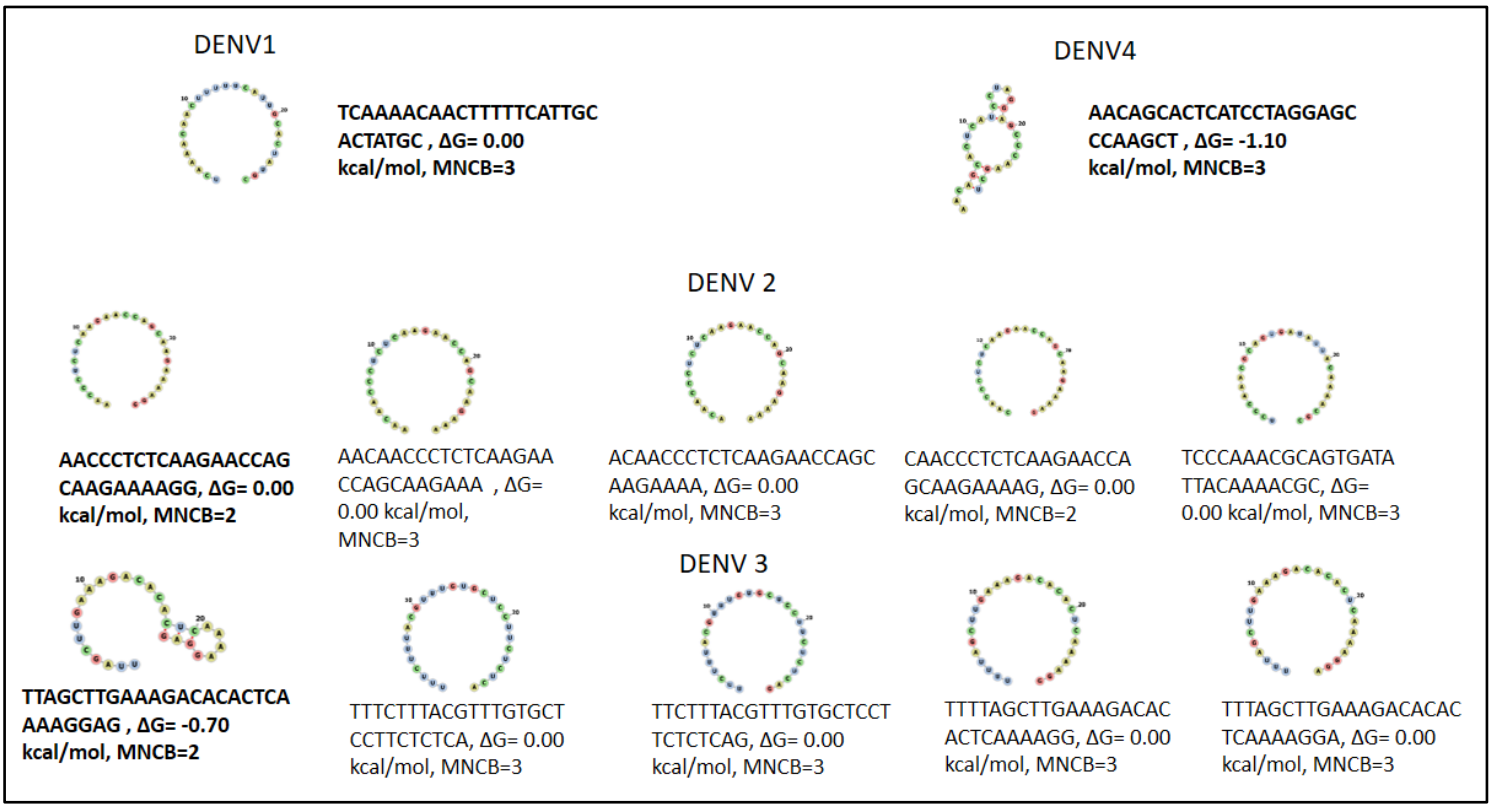

Figure 5: DENV Serotype specific gRNAs after designing and screening, Abbreviation: MNCB= minimum number of conserved bases found among reference serotypes in alignment

Acknowledgments:
This work was supported by the Senior Research Fellowship awarded to Archana Prajapati by the Council of Scientific and Industrial Research (CSIR), Human Resource Development group Government of India is highly acknowledged.

References:

[1] Harapan H *et al.Viruses* 2020 **12**:829. [PMID: 32751561]
[2] Waman VP *et al. PeerJ* 2016 **4**:e2326. [PMID: 27635316]
[3] Laiton-Donato K *et al.Virol. J.* 2019 **16**:62. [PMID: 31068191]
[4] Cecchetto J *et al. Biosens. Bioelectron.* 2020 **151**:111972. [PMID: 31999580]
[5] Mardekian SK and Roberts AL, *Biomed Res. Int.* 2015 **2015.** [PMID: 26509163]
[6] Wong PF *et al.J. Infect. Public Health* 2020 **13**:198. [PMID: 31405788]
[7] Shu P and Huang J, *Clin. Diagn. Lab. Immunol.* 2004 **11**:650. [PMID: 15242935]
[8] Muller DA, *J. Infect. Dis.* 2017 **215**:S89. [PMID: 28403441]
[9] Prajapati A and Nain V, *Bioinformation* 2021 **17**:645. [PMID: 35173386]
[10] Jolany Vangah S *et al. Biol. Proced. Online* 2020 **22**:14. [PMID: 32939188]
[11] Kumar P *et al. Front. Cell. Infect. Microbiol.* 2020 **10**:576875. [PMID: 33251158]
[12] Rahimi H *et al. ACS Sensors* 2021 **6**:1430. [PMID: 33502175]
[13] Gootenberg JS *et al. Science* 2017 **356**: 442. [PMID: 28408723]
[14] Ding X *et al. Nat. Commun.* 2020 **11**:10. [PMID: 32948757]
[15] Arizti-Sanz J *et al. Nat. Commun.* 2020 **11. [**PMID: 33219225]
[16] Benson DA *et al. Nucleic Acids Res.* 2013 **41**:42. [PMID: 23193287]
[17] Hall T *GERF Bull. Biosci.* 2011 **2**:61.
[18] Labun K *et al. Nucleic Acids Res.* 2016 **44**:W276. [PMID: 27185894]
[19] https://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi.
[20] Bandaru S *et al. Sci. Rep.* 2020 **10**:12. [PMID: 32665590]
[21] Jensen KT *et al. FEBS Lett.* 2017 **591**:1901. [PMID: 28580607]
[22] Vatankhah M *et al. Crit. Rev. Clin. Lab. Sci.* 2021 **58**:241. [PMID: 33245685]

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(8): 661-668 (2022)