A Computational Approach to Design Potential siRNA Molecules as a Prospective Tool for Silencing Nucleocapsid Phosphoprotein and Surface Glycoprotein Gene of SARS-CoV-2

Umar Faruq Chowdhury[1], Mohammad Umer Sharif Shohan[1], Kazi Injamamul Hoque[1] Mirza Ashikul Beg[1], Mohammad Ali Moni[2], Mohammad Kawsar Sharif Siam[3, *]

[1]Department of Biochemistry and Molecular Biology, University of Dhaka, Bangladesh
[2]WHO Collaborating Centre for eHealth, School of Public Health and Community Medicine, Faculty of Medicine, University of New South Wales (UNSW), Sydney, Australia
[3]Department of Pharmacy, Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
[*]Corresponding author:
E-mail address: sharif.siam@bracu.ac.bd (Mohammad Kawsar Sharif Siam)

Abstract:

A disease outbreak named COVID-19, caused by a RNA virus, SARS-CoV-2 has become pandemic with a magnitude which is daunting to all public health institutions in the absence of specific antiviral treatment. Surface glycoprotein and nucleocapsid phosphoprotein are two important proteins of this virus facilitating its entry into host cell and genome replication. Small interfering RNA (siRNA) is a prospective tool of the RNA interference (RNAi) pathway for the control of human viral infections by suppressing viral gene expression through hybridization and neutralization of target complementary mRNA. So, in this study, the power of RNA interference technology was harnessed to develop siRNA molecules against specific target genes namely, nucleocapsid phosphoprotein gene and surface glycoprotein gene. Conserved sequence from 139 SARS-CoV-2 strains from around the globe was collected to construct 78 siRNA that can inactivate nucleocapsid phosphoprotein and surface glycoprotein genes. Finally, through a rigorous filtering process 8 siRNA molecules were selected with exerts the best action. These predicted siRNAs should effectively silence the genes of SARS-CoV-2 during siRNA mediated treatment assisting in the response against SARS-CoV-2

Keywords: SARS-CoV-2. Nucleocapsid phosphoprotein, Surface glycoprotein, siRNA, siDirect.

Introduction:

COVID-19, A pandemic affecting lives of billions of people worldwide, has confronted humanity in the commencement of 2020, is caused by a viral pathogen, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or 2019-nCoV. Initial symptoms of this disease mainly include fever, cough, fatigue, dyspnea & headache(Chen *et al.*, 2020; Huang *et al.*, 2020) or it may be asymptomatic(Song *et al.*, 2020). The spike glycoprotein of SARS-CoV-2 binds directly with the surface cell angiotensin converting enzyme II (ACE2) receptor present on alveolar epithelial cells of lung facilitating virus entry, replication and triggers cytokine cascade mechanism(Zhou *et al.*, 2020). In severe cases, patient may die due to massive alveolar damage and progressive respiratory failure(Chan *et al.*, 2020; Huang *et al.*, 2020). The current detection process of SARS-CoV-2 carried out by most countries is using real-time RT-PCR, although several other methods are also being developed(Chu *et al.*, 2020; Corman *et al.*, 2020; Khan *et al.*, 2020). Incubation period for the virus ranges between 2–14 days(Linton *et al.*, 2020) and in some cases, transmission is also reported during asymptomatic period(Yu *et al.*, 2020). Some recent studies suggest that bats are likely reservoir hosts for SARS-Cov-2 but the identity of the intermediate host that might have facilitated transfer to human still remain elusive with some studies indicating pangoline. SARS-CoV-2 is assumed to spread mainly from person-to-person through respiratory droplets produced when an infected person sneezes and coughs or between people who are in close contact(Chan *et al.*, 2020).

Coronaviruses are genetically classified into four main genera*:* Alphacoronavirus*,* Betacoronavirus*,* Gammacoronavirus*, and* Deltacoronavirus(Li, 2016). The first two genera generally infect mammals, while the last two mostly cause disease in birds. The genome size of coronaviruses ranges between approximately 26-32 kb and includes about 6 to 11 open reading frames (ORFs)(Song *et al.*, 2019). Nucleocapsid protein (N), small envelope protein (E), spike surface glycoprotein (S) and matrix protein (M) are the four major structural proteins of coronavirus and all of which are essential to produce a structurally complete virus(Masters, 2006; Wang *et al.*, 2017). The nucleocapsid protein (N) is a multifunctional protein comprising three distinct and highly conserved domains: two structural and independently folded structural regions, namely the N terminal domain and C-terminal domain, which are separated by a intrinsically disordered RNA-binding domain(Huang *et al.*, 2004). The primary role of CoV N protein is to package the genomic viral genome into flexible, long, helical ribonucleoprotein (RNP) complexes called nucleocapsids(de Haan and Rottier, 2005). Apart from these, N protein is essential for viral assembly, envelope formation, genomic RNA synthesis, cell cycle regulation and viral pathogenesis(Chen *et al.*, 2005; Kopecky-Bromberg *et al.*, 2007; Hurst *et al.*, 2010). Spike glycoprotein (S) is a viral fusion protein which forms homotrimers protruding from the viral surface(Tortorici and Veesler, 2019) and mediates virus entry to cell(Howard *et al.*, 2008). S contains two functional subunits: S1 & S2 subunits. The S1 subunit includes the receptor-binding domain(s) and contributes to stabilization of the membrane-anchored S2 subunit that contains the fusion machinery(Walls *et al.*, 2016). As the coronavirus S glycoprotein is surface-exposed and mediates entry into host cells and N nucleocapsid protein are essential for genome replication, these could be the main targets for designing therapeutics(Park *et al.*, 2016).

Silencing of mRNA or post-transcriptional gene silencing by RNA interference (RNAi) is a regulatory cellular mechanism. RNAi is a prospective tool for the control of human viral infections(Levanova and Poranen, 2018). Small interfering RNAs (siRNAs) and micro RNAs (miRNAs) are involved in the RNA interference (RNAi) pathway, where they hybridize to complementary mRNA molecules and neutralizes mRNA causing suppression of gene expression or translation(Hamilton and Baulcombe, 1999). Studies show that, combinations of chemically synthesized siRNA duplexes targeting genomic RNA of SARS-CoV results in therapeutic activity of up to 80% inhibition (Zheng *et al.*, 2004). siRNAs directed against Spike sequences and the 3′-UTR can inhibit the replication of SARS-CoV

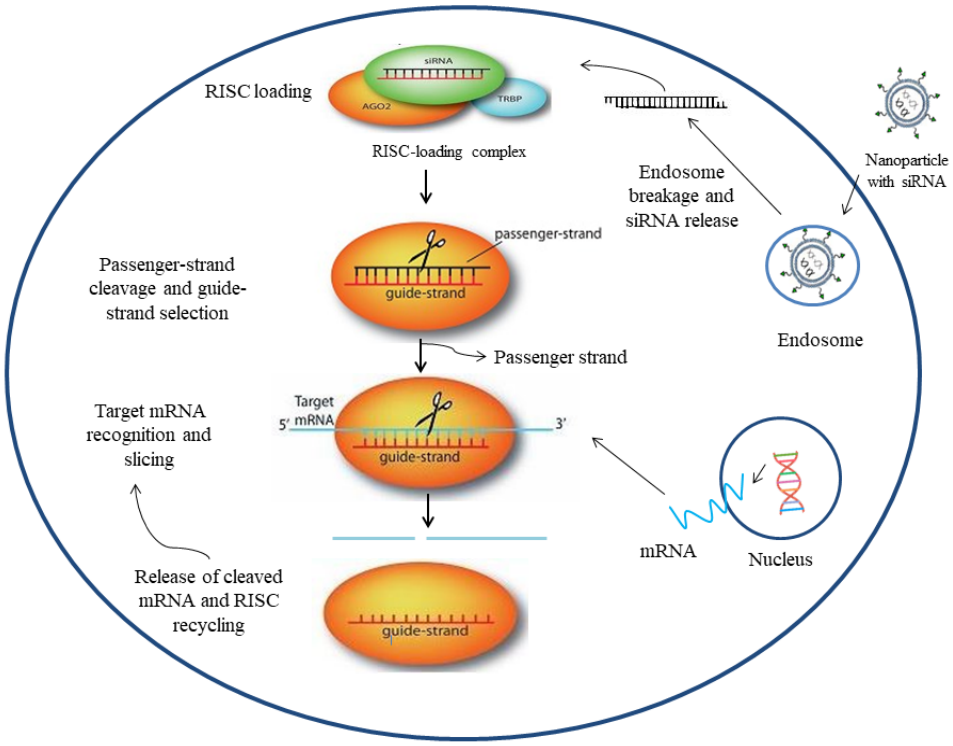

Figure 1: Graphical representation of the siRNA-mediated gene silencing mechanism. Nanoparticles loaded with siRNA are taken up through endocytosis by the cells. These particles are then trapped into the endosomes. siRNA escape endosomes and release siRNA into the cytoplasm due to pH responsive mechanism or proton sponge effect. Once generated, siRNA is loaded into RNA-indiced silencing complex comprising of RNA-binding protein TRBP and Argonaute (AGO2). AGO2 selects the siRNA guide strand, then cleaves and ejects the passenger strand. After that, the guide strand pairs with its complementary target mRNA and AGO2 slices the target. After slicing, the cleaved target mRNA is released and RISC is recycled for another few rounds of slicing using the same guide strand(Figure adapted from Jinek and Doudna, 2009).

As of April 10, total infected case is 1,615,059 and among these patients 96,791 people have died which means case fatality rate (CFR) is approximately 5.99%. The alarming phenomenon is the exponential growth of total infection case and death number (https://www.worldometers.info/coronavirus/). Treatment of this increased number of people is not possible as no antiviral drug is still available specifically for SARS-CoV-2 and there is a lack on appropriate medical response.In this study, we have designed siRNAs specific to various conserved region of nucleocapsid phosphoprotein & surface glycoprotein genes of SARS-CoV-2 and finally predicted 8 universal siRNA molecules against nucleoprotein and glycoprotein genes which will inhibit the translations of these proteins and allow the host to discard this infection. siRNAs are designed against both nucleoprotein and glycoprotein as both are important for the survival of virus(Almazan, Galan and Enjuanes, 2004; Howard *et al.*, 2008) and targeting these proteins may cause viral inhibition(Zheng *et al.*, 2004; Wu *et al.*, 2005).

Method:

Sequence Retrieval from NCBI

Coding sequences from 139 genomes of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) were retrieved from NCBI Virus (Hatcher *et al.*, 2016) portal (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/) (Supplementary Table 1). Nucleocapsid phosphoprotein and surface glycoprotein sequences were manually extracted and curated from the retrieved data using bash scripting in linux computer platform.

Multiple Sequence Alignment & Phylogenetic Tree Construction

ClustalW (Thompson, Higgins and Gibson, 1994) algorithm was employed to perform multiple sequence alignment. Maximum likelihood phylogenetic trees were constructed with a bootstrap value of 500. Tamura Nei (Tamura and Nei, 1993) model of evolution was selected while constructing the phylogenetic tree. MEGA-X (Kumar *et al.*, 2018) and MEGA-CC (Kumar *et al.*, 2012) programs were was used for alignment formation, phylogenetic tree construction. iTOL online tool (Letunic and Bork, 2019) (https://itol.embl.de/) was used in order to visualize the phylogenetic trees.

Target Recognition & siRNA Designing

Target-specific siRNAs were designed with the help of siDirect web server (Naito *et al.*, 2009). Rules of Ui-Tei (Ui-Tei *et al.*, 2004), Amarzguioui (Amarzguioui and Prydz, 2004) and Reynolds (Reynolds *et al.*, 2004) were used (Table 1) and the melting temperature was kept below 21.5 °C as a default parameter for siRNA duplex.

Table 1. Rules/Algorithm to construct siRNAs.	
Rule Name	**Description**
Ui-Tei	A/U at the 5' terminus of the sense strand
	G/C at the 5' terminus of the antisense strand
	At least 4 A/U residues in the 5' terminal 7 bp of sense strand
	No GC stretch longer than 9nt
Amarzguioui	Duplex End A/U differential > 0
	No U at position 1
	Strong binding of 5' sense strand
	Presence of A at position 6
	Weak binding of 3' sense strand
Reynolds	GC content 30–52% (one point)
	Occurrence of three or more A/U base pair at position
	15–19 of sense strand (Each A/U base pair in this region earns one point)
	Low internal stability at target site (Tm > −20◦C) (one point)
	Presence of U at position 10 of the sense strand (one point)
	Presence of A at position 3 of the sense strand (one point)
	Presence of A at position 19 of the sense strand (one point)
	Absence of G at position 13 of the sense strand (one point)
	Threshold for efficient siRNAs score >= 6

Off target similarity search

Blast search was performed against human genome and transcriptome using the standalone blast package (Camacho *et al.*, 2009) to identify the possible off target matches. The e-value was set to 1e-10 to reduce the stringency of the search condition thereby increasing the chances of random matches.

GC content calculation & Secondary Structure Prediction

OligoCalc (Kibbe, 2007) was used to calculate the GC content. The secondary structure of siRNAs were predicted along with the respective free energy using MaxExpect (Lu, Gloor and Mathews, 2009) program in the RNA structure web server (Bellaousov *et al.*, 2013). The higher values indicate better candidates as those molecules are less prone to folding.

Calculation of RNA-RNA interaction through thermodynamics

Higher interaction between the target and the guide strand serves a better predictor for siRNA efficacy. Therefore, the thermodynamic interaction between the target strand and the siRNA guide strand was predicted with the aid of DuplexFold (Piekna-Przybylska *et al.*, 2010) program of the RNA structure web server (Bellaousov *et al.*, 2013).

Calculation of Heat Capacity & Concentration plot

DINA Melt web server was used to generate heat capacity and concentration plot. The ensemble heat capacity (Cp) is plotted as a function of temperature, with the melting temperature Tm (Cp) (Supplementary Table 6 & Supplementary table 7). The contributions of each species to the ensemble heat capacity shown by detailed heat capacity plot. Also, the point at which the concentration of double-stranded molecules of one-half of its maximum value defines the melting temperature Tm(Conc) was shown using the concentration plot- Tm (Conc).

Predicted siRNA Validation

siRNAPred server (http://crdd.osdd.net/raghava/sirnapred/index.html) was used to validate the predicted siRNA species. The predicted siRNAs were evaluated against the Main21 dataset using support vector machine algorithm and the binary pattern prediction approach. siRNAPred score greater than on 1 predicts very high efficacy, score ranging 0.8-0.9 predicts high efficacy and score ranging 0.7-0.8, predicts moderate efficacy. In total, 78 siRNAs were used for efficacy prediction.

Results:

Evolutionary Divergence Analysis shows conserved pattern between strains:

Phylogenetic tree was constructed using 139 sequences for both nucleocapsid phosphoprotein and surface glycoprotein separately. Only a handful number of sequences showed significant divergence (bootstrap value > 60%) (Supplementary Table 8). This suggest that most of the viral sequences have been conserved sequences and therefore used to construct siRNA which will cover a wide range of SARS-CoV-2 strain.

siDirect predicted 78 siRNA

siDirect web server predicted 8 siRNA for nucleocapsid phosphoprotein and 70 siRNA for surface glycoprotein (Supplementary Table 4 & Supplementary table 5) that maintains all the parameters. Seed target duplex stability (Tm) values for all the predicted siRNAs were less than 21.5 °C which suggests the ability of predicted siRNAs to avoid non-target binding.

Off-target Binding Exclusion Using Blast

Standalone blast search against human genome and transcriptome did not reveal any off-target match. This shows that our predicted siRNA would not interact in any other places other than the viral target location.

GC Content Calculation & Secondary Structure Determination

GC content analysis of the predicted siRNAs were ranged 31% to 43% for nucleocapsid phosphoprotein (Supplementary Table 6) and 10% to 40% for surface glycoprotein (Supplementary Table 7). Molecules that have GC content below 31.6% were eliminated. Also, the calculated free energy of folding ranged from 1.4 to 2 for nucleocapsid phosphoprotein (Supplementary Table 6) and from 1.3 to 2 for surface glycoprotein (Supplementary Table 7). The associated secondary structures were also determined.

Thermodynamics of Target-Guide Strand Interaction

Free energy of binding between target and guide strand were calculated. The values spanned from -35.8 to -31 for nucleocapsid phosphoprotein (Supplementary Table 6) and -36.6 to -21.6 for surface glycoprotein (Supplementary Table 7).

Table 2. Best effective siRNA molecules with various parameters.

Alias	Conserved position	Location of target within mRNA	siRNA target within mRNA	Predicted siRNA siRNA duplex candidate at 37?C	Seed-duplex (Tm) Guide	Seed-duplex (Tm) Passenger	GC %	Free energy of folding	Free energy of binding	Tm (Conc)	Tm (Cp)	Validity (Binary)
n7	1031-1253	33-55	AAGCATATTGACGCATACAAAAC	UUUGUAUGCGUCAAUAUGCUU GCAUAUUGACGCAUACAAAAC	13.5	5.6	36	1.5	-31.7	75.8	77	1.047
g15	871-1222	20-42	CAGAAACAAAGTGTACGTTGAAA	UCAACGUACACUUUGUUUCUG GAAACAAAGUGUACGUUGAAA	21	5.6	36	1.6	-32.8	79.1	80	1.022
g21	871-1222	136-158	TGCCCTTTTGGTGAAGTTTTTAA	AAAAACUUCACCAAAAGGGCA CCCUUUUGGUGAAGUUUUUAA	3.2	16.1	36	1.8	-33.4	83.4	85	1.039
g22	871-1222	138-160	CCCTTTTGGTGAAGTTTTTAACG	UUAAAAACUUCACCAAAAGGG CUUUUGGUGAAGUUUUUAACG	0	18.8	33	1.9	-30.4	78.6	80	1.01
g44	1842-2389	441-463	CACAATTAAACCGTGCTTTAACT	UUAAAGCACGGUUUAAUUGUG CAAUUAAACCGUGCUUUAACU	19.7	-9.7	33	1.9	-30	79.5	80	1.102
g46	1842-2389	482-504	GACAAAAACACCCAAGAAGTTTT	AACUUCUUGGGUGUUUUUGUC CAAAAACACCCAAGAAGUUUU	17.7	5.6	36	1.6	-32.5	80.6	82	1.076
g59	2790-3822	304-326	GTGTGTACTTGGACAATCAAAAA	UUUGAUUGUCCAAGUACACAC GUGUACUUGGACAAUCAAAAA	13.8	19	36	1.5	-33.9	80.4	81	1.007
g70	2790-3822	1001-1023	GTGCTCAAAGGAGTCAAATTACA	UAAUUUGACUCCUUUGAGCAC GCUCAAAGGAGUCAAAUUACA	7.4	16.6	38	1.9	-33.7	81.7	83	1.012

Heat Capacity & Duplex Concentration Plot Determination

The Tm(Cp) and Tm(Conc) were calculated for the predicted siRNAs. The higher values of these two melting temperatures indicate higher effectiveness of the siRNA species. Tm(Conc) values ranged from 71.7◦C to 81.7◦C for nucleocapsid phosphoprotein (Supplementary Table 6) and 66.4◦C to 83.8◦C for surface glycoprotein (Supplementary Table 7). Tm(Cp) values ranged from 72.1◦C to 82.5◦C for nucleocapsid phosphoprotein (Supplementary Table 6) and from 66.3◦C to 85.2◦C for surface glycoprotein (Supplementary Table 7).

Validation and Selection of best 8 siRNAs

siRNAPred checked the effectivity of the predicted siRNAs and values greater than 1 are considered highly effective. 2 siRNAs for nucleaocapsid phosphoprotein and 32 siRNAs for surface glycoprotein were found to be highly effective. Finally, based on all the other criteria, 8 siRNAs were selected as best predicted candidates against the nucleocapsid phosphoprotein and the surface glycoprotein genes of SARS CoV-2 (Table 2).

Discussion:

COVID-19 is an emerging disease that lays bare the society we have created and its interdependent infrastructure with a fast increase in cases and deaths since its first identification. Having no regard for geography, this pandemic has a global reach, and no continent is out of its clutches. Moreover, there is no vaccine available to prevent this disease and no RNAi based treatment is yet in practice or been proposed. So, the next generation medicine, siRNA might be effective in this case.

Here, a total of 34 (15 nucleocapsid phosphoprotein and 19 surface glycoprotein) (Supplementary Table 2 & Supplementary table 3) conserved regions were identified among 139 strains of SARS-CoV-2 from around the world. Conserved portions that are shorter than 21 nucleotides were omitted from further analysis. Conserved sequences were put to siDirect web server to identify possible targets and to generate corresponding siRNAs. siDirect performs the task in three distinct steps, highly functional siRNA selection, seed-dependent off-target effects reduction, near-perfect matched genes elimination. siRNA targets were found in 18 conserved regions, 5 nucleocapsid phosphoprotein (Supplementary Table 4) and 13 surface glycoprotein (Supplementary Table 5). U,R,A (Ui-Tei, Amarzguioui and Reynolds) rules (Table 1) were applied while predicting the siRNAs to obtain better results. Non-target sequences were eliminated by optimizing the melting temperature (Tm) below 21.5◦C. The equation to calculate melting temperature (Tm) is below,

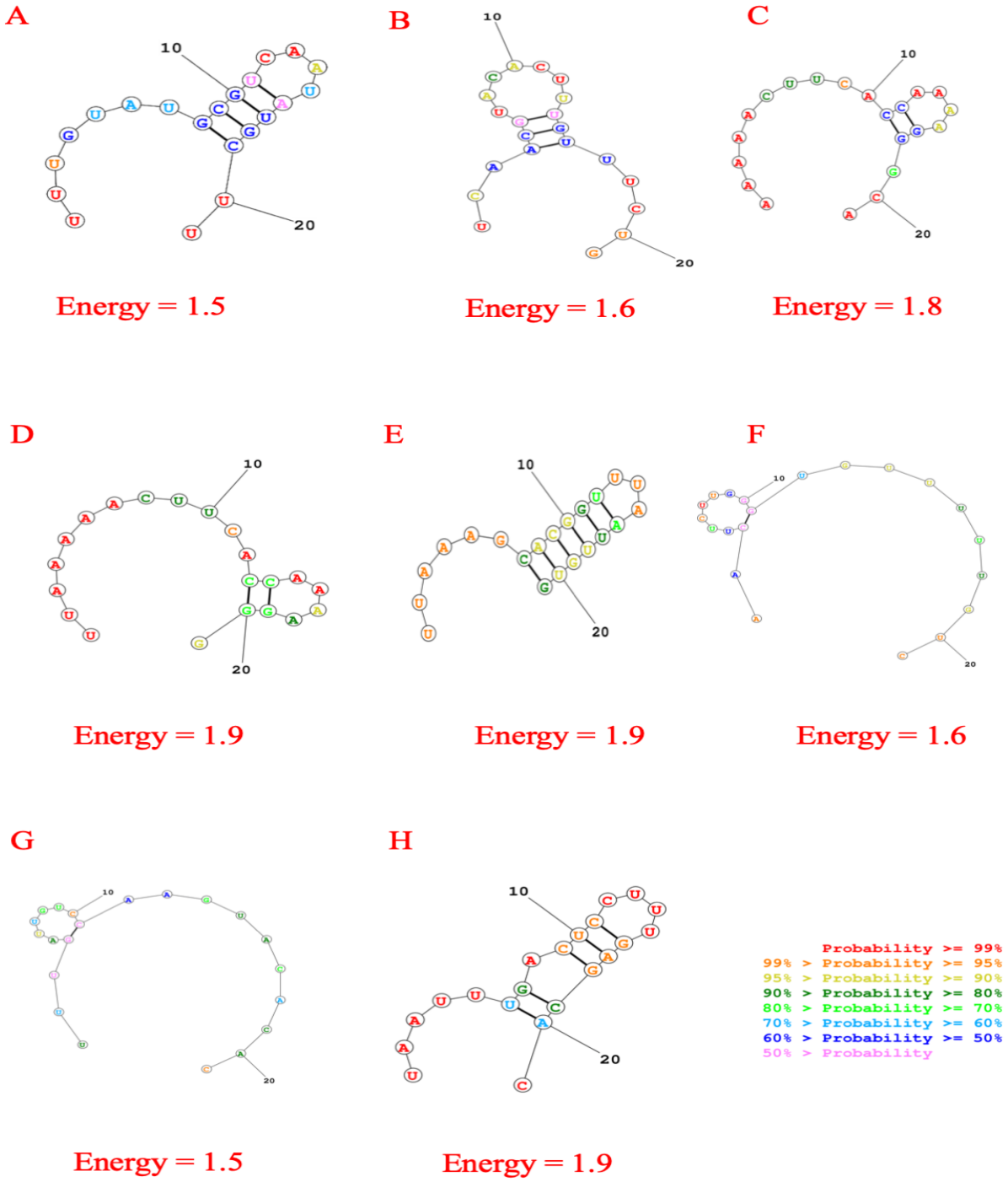

Figure 2. Secondary structures of best eight predicted siRNA with possible folding and minimum free energy for consensus sequence. The structures are A. n7 B.g15 C. g21 D. g22 E. g44 F. g46 G. g59 H. g70 siRNAs.

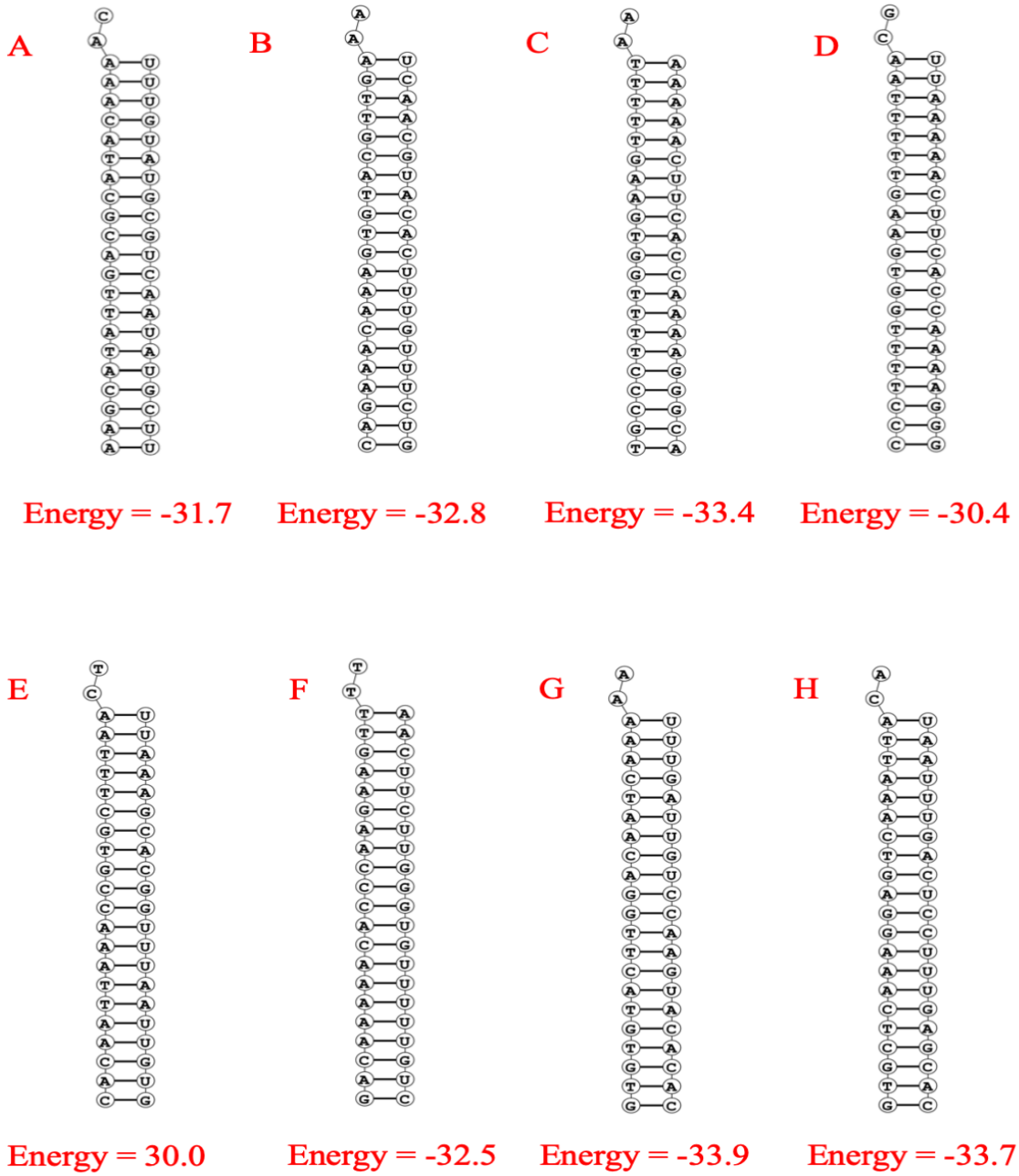

Figure 3. Predicted lowest free energy structure for binding siRNA (guide strand) and target RNA. The structures are for A. n7 B.g15 C. g21 D. g22 E. g44 F. g46 G. g59 H. g70 siRNAs and their corresponding targets.

$Tm = (1000*\Delta H) / (A + \Delta S + R \ln (CT/4)) - 273.15 + 16.6 \log [Na^+]$

Here,

- ΔH (kcal/ mol) represents the sum of the nearest neighbor enthalpy change
- A is the helix initiation constant (-10.8)
- ΔS represents the sum of the nearest neighbor entropy change
- R is the gas constant (1.987 cal/deg/mol)
- CT represents the total molecular concentration of the strand (100 µM) and
- Concentration of Sodium, $[Na^+]$ was fixed at 100 mM

siRNAs functionality is influenced by the GC-content and there is an inverse relationship of the GC-content with the function of siRNA. Usually a low GC content (between 31.6 and 57.9%) is ideal for a siRNA to be effective (Chan *et al.*, 2009). So, siRNA species that have GC content lower than 31.6 were not kept in the final selection. In the study, GC content ranged from 10% to 43% for all the predicted species. GC content of all the eight selected siRNAs are greater than or equal to 33% (Table 2). Formation of secondary structure of siRNA may inhibit the RISC mediated cleavage of target. So, the prediction of secondary structure and determination of free energy of folding is crucial. So, guide strands of siRNA were subjected to RNA structure web server in order to predict possible folding structure and corresponding minimum free energy. At 37◦C, finally selected siRNAs have free energy of folding greater than zero, which suggests the predicted siRNAs are more accessible for efficient binding. DuplexFold was used to determine the target and guide siRNA interaction and their corresponding binding energy. Lower binding energy indicate better interaction therefore better chance of target inhibition. The values of free energy of binding of all the predicted siRNAs spanned from -36.6 to -21.6 (Supplementary Table 6 & Supplementary table 7). The collective heat capacity, denoted as Cp, is plotted as a function of temperature and the melting temperature, denoted as Tm (Cp), was determined. The contribution of each molecules to the collective heat capacity was demonstrated using the inclusive heat capacity plot where melting temperature Tm (Conc), indicates the temperature at which the concentration of double-stranded molecules of becomes one-half of its maximum value. All the selected siRNAs have high Tm value (>75◦C) (Table 2).

siRNAPred was used to determine the inhibition efficacy of the predicted siRNAs. siRNAPred uses Main21 dataset which consist of 2182 siRNAs (21mer) derived from a homogeneous experimental condition to predict the actual efficacy of 21mer siRNAs with high accuracy using the support vector machine based method. So, siRNA candidates that have validity score greater than 1 were chosen for the final selection.

In this study, eight prospective siRNA molecules were proposed to be efficient at binding and cleaving specific mRNA targets of SARS-CoV-2 (Table 2). As the study contain a large array of 139 sequences of SARS-CoV-2 from around the world, the predicted therapeutic agent can be employed to large scale treatment of COVID-19 pandemic.

Conclusion:

Computational methods can be employed to design and predict siRNA interaction against specific gene target thereby silencing its expression. In this study, eight siRNA molecules were predicted to be effective against nucleocapsid phosphoprotein and surface glycoprotein gene of 139 strains of SARS-CoV-2 virus using computational method considering all maximum parameters in prime conditions. In order to decelerate the COVID-19 pandemic and recover the affected individuals the development of siRNA therapeutic approaches could be a promising alternative to traditional vaccine designing.

Limitations: This study employed extensive computational approaches to assess the quality of the designed siRNA sequences and need to be explored further in vitro lab settings.

Financial and Competing Interest Disclosure: The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.

References:

Almazan, F., Galan, C. and Enjuanes, L. (2004) 'The Nucleoprotein Is Required for Efficient Coronavirus Genome Replication', *Journal of Virology*. American Society for Microbiology, 78(22), pp. 12683–12688. doi: 10.1128/jvi.78.22.12683-12688.2004.

Amarzguioui, M. and Prydz, H. (2004) 'An algorithm for selection of functional siRNA sequences', *Biochemical and Biophysical Research Communications*. Academic Press, 316(4), pp. 1050–1058. doi: 10.1016/j.bbrc.2004.02.157.

Bellaousov, S. *et al.* (2013) 'RNAstructure: web servers for RNA secondary structure prediction and analysis', *Nucleic Acids Research*, 41(W1), pp. W471–W474. doi: 10.1093/nar/gkt290.

Camacho, C. *et al.* (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics*. BioMed Central, 10(1), p. 421. doi: 10.1186/1471-2105-10-421.

Chakraborty, C. *et al.* (2017) 'Therapeutic miRNA and siRNA: Moving from Bench to Clinic as Next Generation Medicine', *Molecular Therapy - Nucleic Acids*. Elsevier Inc, pp. 132–143. doi: 10.1016/j.omtn.2017.06.005.

Chan, C. Y. *et al.* (2009) 'A structural interpretation of the effect of GC-content on efficiency of RNA interference', *BMC Bioinformatics 2009 10:1*. BioMed Central, 10(1), pp. 1–7. doi: 10.1186/1471-2105-10-s1-s33.

Chan, J. F. W. *et al.* (2020) 'A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster', *The Lancet*. Lancet Publishing Group, 395(10223), pp. 514–523. doi: 10.1016/S0140-6736(20)30154-9.

Chen, H. *et al.* (2005) 'Mass spectroscopic characterization of the coronavirus infectious bronchitis virus nucleoprotein and elucidation of the role of phosphorylation in RNA binding by using surface plasmon resonance.', *Journal of virology*, 79(2), pp. 1164–79. doi: 10.1128/JVI.79.2.1164-1179.2005.

Chen, N. *et al.* (2020) 'Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study', *The Lancet*. Lancet Publishing Group, 395(10223), pp. 507–513. doi: 10.1016/S0140-6736(20)30211-7.

Chen, Y., Cheng, G. and Mahato, R. I. (2008) 'RNAi for treating hepatitis B viral infection', *Pharmaceutical Research*, pp. 72–86. doi: 10.1007/s11095-007-9504-0.

Chu, D. K. W. *et al.* (2020) 'Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia', *Clinical Chemistry*. Oxford Academic, 66(4), pp. 549–555. doi: 10.1093/clinchem/hvaa029.

Corman, V. M. *et al.* (2020) 'Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR', *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*. NLM (Medline), 25(3). doi: 10.2807/1560-7917.ES.2020.25.3.2000045.

de Haan, C. A. M. and Rottier, P. J. M. (2005) 'Molecular Interactions in the Assembly of Coronaviruses', *Advances in Virus Research*. Academic Press, pp. 165–230. doi: 10.1016/S0065-3527(05)64006-7.

Hamilton, A. J. and Baulcombe, D. C. (1999) 'A species of small antisense RNA in posttranscriptional gene silencing in plants', *Science*, 286(5441), pp. 950–952. doi: 10.1126/science.286.5441.950.

Hatcher, E. L. *et al.* (2016) 'Virus Variation Resource – improved response to emergent viral outbreaks', *Nucleic Acids Research*, 45(D1), pp. D482–D490. doi: 10.1093/nar/gkw1065.

Howard, M. W. *et al.* (2008) 'Aromatic Amino Acids in the Juxtamembrane Domain of Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein Are Important for Receptor-Dependent Virus Entry and Cell-Cell Fusion', *Journal of Virology*. American Society for Microbiology, 82(6), pp. 2883–2894. doi: 10.1128/jvi.01805-07.

Huang, C. *et al.* (2020) 'Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China', *The Lancet*. Lancet Publishing Group, 395(10223), pp. 497–506. doi: 10.1016/S0140-6736(20)30183-5.

Huang, Q. *et al.* (2004) 'Structure of the N-terminal RNA-binding domain of the SARS CoV

nucleocapsid protein', *Biochemistry*, 43(20), pp. 6059–6063. doi: 10.1021/bi036155b.
Hurst, K. R. *et al.* (2010) 'An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA.', *Journal of virology*, 84(19), pp. 10276–88. doi: 10.1128/JVI.01287-10.
Jinek, M. and Doudna, J. A. (2009) 'A three-dimensional view of the molecular machinery of RNA interference', *Nature*, pp. 405–412. doi: 10.1038/nature07755.
Khan, S. *et al.* (2020) 'Analysis of Serologic Cross-Reactivity Between Common Human Coronaviruses and SARS-CoV-2 Using Coronavirus Antigen Microarray', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.03.24.006544. doi: 10.1101/2020.03.24.006544.
Kibbe, W. A. (2007) 'OligoCalc: an online oligonucleotide properties calculator', *Nucleic Acids Research*, 35(suppl_2), pp. W43–W46. doi: 10.1093/nar/gkm234.
Kopecky-Bromberg, S. A. *et al.* (2007) 'Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists.', *Journal of virology*, 81(2), pp. 548–57. doi: 10.1128/JVI.01782-06.
Kumar, S. *et al.* (2012) 'MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis', *Bioinformatics*, 28(20), pp. 2685–2686. doi: 10.1093/bioinformatics/bts507.
Kumar, S. *et al.* (2018) 'MEGA X: Molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution*. Oxford University Press, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.
Letunic, I. and Bork, P. (2019) 'Interactive Tree Of Life (iTOL) v4: recent updates and new developments', *Nucleic Acids Research*, 47(W1), pp. W256–W259. doi: 10.1093/nar/gkz239.
Levanova, A. and Poranen, M. M. (2018) 'RNA interference as a prospective tool for the control of human viral infections', *Frontiers in Microbiology*. Frontiers Media S.A., p. 2151. doi: 10.3389/fmicb.2018.02151.
Li, F. (2016) 'Structure, Function, and Evolution of Coronavirus Spike Proteins', *Annual Review of Virology*. Annual Reviews, 3(1), pp. 237–261. doi: 10.1146/annurev-virology-110615-042301.
Linton, N. M. *et al.* (2020) 'Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data', *Journal of Clinical Medicine*. MDPI AG, 9(2), p. 538. doi: 10.3390/jcm9020538.
Liu, Z. *et al.* (2020) 'Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2', *Journal of Medical Virology*. John Wiley and Sons Inc., p. jmv.25726. doi: 10.1002/jmv.25726.
Lu, Z. J., Gloor, J. W. and Mathews, D. H. (2009) 'Improved RNA secondary structure prediction by maximizing expected pair accuracy', *RNA*. Cold Spring Harbor Laboratory Press, 15(10), pp. 1805–1813. doi: 10.1261/rna.1643609.
Masters, P. S. (2006) 'The Molecular Biology of Coronaviruses', *Advances in Virus Research*, pp. 193–292. doi: 10.1016/S0065-3527(06)66005-3.
Naito, Y. *et al.* (2009) 'SiDirect 2.0: Updated software for designing functional siRNA with reduced seed-dependent off-target effect', *BMC Bioinformatics*. BioMed Central, 10(1), p. 392. doi: 10.1186/1471-2105-10-392.
Park, J. E. *et al.* (2016) 'Proteolytic processing of middle east respiratory syndrome coronavirus spikes expands virus tropism', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(43), pp. 12262–12267. doi: 10.1073/pnas.1608147113.
Piekna-Przybylska, D. *et al.* (2010) 'A sequence similar to tRNA3Lys gene is embedded in HIV-1 U3–R and promotes minus-strand transfer', *Nature Structural and Molecular Biology*. Nature Publishing Group, 17(1), pp. 83–90. doi: 10.1038/nsmb.1687.
Reynolds, A. *et al.* (2004) 'Rational siRNA design for RNA interference', *Nature Biotechnology*. Nature Publishing Group, 22(3), pp. 326–330. doi: 10.1038/nbt936.
Song, F. *et al.* (2020) 'Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia', *Radiology*. Radiological Society of North America Inc., 295(1), pp. 210–217. doi: 10.1148/radiol.2020200274.

Song, Z. *et al.* (2019) 'From SARS to MERS, thrusting coronaviruses into the spotlight', *Viruses*. MDPI AG. doi: 10.3390/v11010059.
Tamura, K. and Nei, M. (1993) 'Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.', *Molecular Biology and Evolution*. Oxford University Press (OUP), 10(3), pp. 512–526. doi: 10.1093/oxfordjournals.molbev.a040023.
Tang, Q. *et al.* (2015) 'Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition', *Scientific Reports*. Nature Publishing Group, 5(1), pp. 1–8. doi: 10.1038/srep17155.
Tang, X. *et al.* (2020) 'On the origin and continuing evolution of SARS-CoV-2', *National Science Review*. doi: 10.1093/nsr/nwaa036.
Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) 'CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Research*, 22(22), pp. 4673–4680. doi: 10.1093/nar/22.22.4673.
Tortorici, M. A. and Veesler, D. (2019) 'Structural insights into coronavirus entry', in *Advances in Virus Research*. Academic Press Inc., pp. 93–116. doi: 10.1016/bs.aivir.2019.08.002.
Ui-Tei, K. *et al.* (2004) 'Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference', *Nucleic Acids Research*, 32(3), pp. 936–948. doi: 10.1093/nar/gkh247.
Walls, A. C. *et al.* (2016) 'Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer', *Nature*. Nature Publishing Group, 531(7592), pp. 114–117. doi: 10.1038/nature16988.
Wang, C. *et al.* (2017) 'MERS-CoV virus-like particles produced in insect cells induce specific humoural and cellular imminity in rhesus macaques', *Oncotarget*. Impact Journals LLC, 8(8), pp. 12686–12694. doi: 10.18632/oncotarget.8475.
Wilson, R. C. and Doudna, J. A. (2013) 'Molecular Mechanisms of RNA Interference', *Annual Review of Biophysics*. Annual Reviews, 42(1), pp. 217–239. doi: 10.1146/annurev-biophys-083012-130404.
Wu, A. *et al.* (2020) 'Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China', *Cell Host and Microbe*. Cell Press, 27(3), pp. 325–328. doi: 10.1016/j.chom.2020.02.001.
Wu, C. J. *et al.* (2005) 'Inhibition of SARS-CoV replication by siRNA', *Antiviral Research*, 65(1), pp. 45–48. doi: 10.1016/j.antiviral.2004.09.005.
Yu, P. *et al.* (2020) 'A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period', *The Journal of Infectious Diseases*. doi: 10.1093/infdis/jiaa077.
Zhang, L. *et al.* (2020) 'Origin and Evolution of the 2019 Novel Coronavirus', *Clinical Infectious Diseases*. doi: 10.1093/cid/ciaa112.
Zhang, T., Wu, Q. and Zhang, Z. (2020) 'Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak', *Current Biology*. Cell Press, 30(7), pp. 1346-1351.e2. doi: 10.1016/j.cub.2020.03.022.
Zheng, B. *et al.* (2004) 'Prophylactic and therapeutic effects of small interfering RNA targeting SARS-coronavirus.', *Antiviral therapy*. England, 9(3), pp. 365–374.
Zhou, P. *et al.* (2020) 'A pneumonia outbreak associated with a new coronavirus of probable bat origin', *Nature*. Nature Research, 579(7798), pp. 270–273. doi: 10.1038/s41586-020-2012-7.

Figure captions:

Figure 1: Graphical representation of the siRNA-mediated gene silencing mechanism. Nanoparticles loaded with siRNA are taken up through endocytosis by the cells. These particles are then trapped into the endosomes. siRNA escape endosomes and release siRNA into the cytoplasm due to pH responsive mechanism or proton sponge effect. Once generated, siRNA is loaded into RNA-indiced silencing complex comprising of RNA-binding protein TRBP and Argonaute

483 (AGO2). AGO2 selects the siRNA guide strand, then cleaves and ejects the passenger strand. After
484 that, the guide strand pairs with its complementary target mRNA and AGO2 slices the target. After
485 slicing, the cleaved target mRNA is released and RISC is recycled for another few rounds of slicing
486 using the same guide strand(Jinek and Doudna, 2009).
487
488 **Figure 2:** Secondary structures of best eight predicted siRNA with possible folding and minimum
489 free energy for consensus sequence.
490
491 **Figure 3:** Predicted lowest free energy structure for binding siRNA (guide strand) and target RNA.
492
493 **Table Captions:**
494
495 **Table 1:** Rules/Algorithm to construct siRNAs
496 **Table 2:** Best effective siRNA molecules with various parameters
497
498 **Supplement Figure Caption:**
499
500 **Figure S1:** Radial phylogenetic tree of A. nucleocapsid phosphoprotein and B. surface glycoprotein
501 using 139 strains of SARS-CoV-2 from around the world. The bootstrap value for tree construction
502 was set to 500 and Tamura-Nei model of evolution was used for both trees.
503
504 **Supplement Table Captions:**
505 **Supplementary Table 1:** Accession number, Length and Location of the SARS-CoV-2 Stains Used
506 in the Study.
507 **Supplementary Table 2:** Conserved sequence of nucleocapsid phosphoprotein gene
508 **Supplementary Table 3:** Conserved sequence of surface glycoprotein gene
509 **Supplementary Table 4:** Tm values of predicted siRNA (guide strand and passenger strand)
510 against nucleocapsid phosphoprotein.
511 **Supplementary Table 5:** Tm values of predicted siRNA (guide strand and passenger strand)
512 against surface glycoprotein.
513 **Supplementary Table 6:** Effective siRNAs against nucleocapsid phosphoprotein with GC%, free
514 energy of folding and free energy of binding with target.
515 **Supplementary Table 7:** Effective siRNAs against surface glycoprotein with GC%, free energy of
516 folding and free energy of binding with target.
517 **Supplementary Table 8:** Bootstrap values of significant clades in the radial phylogenetic tree.