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[2], Mohammad Kawsar Sharif Siam[3], Mohammad Ali Moni[4]*

[1]Department of Biochemistry and Molecular Biology, University of Dhaka, Bangladesh
[2]Department of Genetic Engineering and Biotechnology, Univeristy of Dhaka, Bangladesh
[3]Department of Pharmacy, Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
[4]WHO Collaborating Centre for eHealth, School of Public Health and Community Medicine, Faculty of Medicine, University of New South Wales (UNSW), Sydney, Australia
*Corresponding author:
E-mail address: m.moni@unsw.edu.au (Mohammad Ali Moni)

ABSTRACT

An outbreak, caused by a RNA virus, SARS-CoV-2 named COVID-19 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, based on GC content, free energy of folding, free energy of binding, melting temperature and efficacy prediction process 8 siRNA molecules were selected which are proposed to 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.

1. 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[1,2] or it may be asymptomatic[3]. 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[4]. In severe cases, patient may die due to massive alveolar damage and progressive respiratory failure[1,5]. 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[6–8]. Incubation period for the virus ranges between 2–14 days[9] and in some cases, transmission is also reported during asymptomatic period[10]. 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 pangolins [11]. 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[5].

Coronaviruses are genetically classified into four main genera*:* Alphacoronavirus, Betacoronavirus, Gammacoronavirus, *and* Deltacoronavirus[12]. 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)[13]. 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[14,15]. 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[16]. The primary role of CoV N protein is to package the genomic viral genome into flexible, long, helical ribonucleoprotein (RNP) complexes called nucleocapsids[17]. Apart from these, N protein is essential for viral assembly, envelope formation, genomic RNA synthesis, cell cycle regulation and viral pathogenesis[18–20]. Spike glycoprotein (S) is a viral fusion protein which forms homotrimers protruding from the viral surface[21] and mediates virus entry to cell[22]. 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[23]. 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[24].

Table 1 Rules/Algorithm to construct siRNAs.

Rule Name	Description
Ui-Tei	A or U present at the 5' terminus of the sense strand
	G or C present at the 5' terminus of the antisense strand
	At least 4 A or U residues present in the 5' terminal 7 bp of sense strand
	GC stretch no longer than 9nt
Amarzguioui	Duplex End A or U differential > 0
	No U present at position 1
	Strong binding of 5' sense strand
	Presence of A at position 6
	Weak binding in case of 3' sense strand
Reynolds	1 point for GC content 30–52% (one point)
	1 point for each occurrence of three or more A or U base pair at position 15–19 of sense strand
	1 point for little internal stability at target site (Tm > −20◦C)
	1 point for occupancy of U at position 10 of the sense strand
	1 point for occupancy of A at position 3 of the sense strand
	1 point for occupancy of A at position 19 of the sense strand
	1 point for Absence of G at position 13 of the sense strand
	Threshold for efficient siRNAs score >= 6

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[25–27]. 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[28]. Studies show that, combinations of chemically synthesized siRNA duplexes targeting genomic RNA of SARS-CoV results in therapeutic activity of up to 80% inhibition [29]. siRNAs directed against Spike sequences and the 3′-UTR can inhibit the replication of SARS-CoV.

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 silico approaches are a general trend to discover novel therapeutic approaches [30–33] and for the viruses there is no exception to this [34].Therefore, 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[22,35] and targeting these proteins may cause viral inhibition[29,36].

2. Materials and methods

2.1. Sequence retrieval from NCBI

Coding sequences from 139 genomes of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) were retrieved from NCBI Virus [37] 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.

2.2. Multiple sequence alignment & phylogenetic tree construction

ClustalW [38] algorithm was employed to perform multiple sequence alignment. Maximum likelihood phylogenetic trees were constructed with a bootstrap value of 500. Tamura Nei [39] model of evolution was selected while constructing the phylogenetic tree. MEGA-X [40] and MEGA-CC [41] programs were was used for alignment formation, phylogenetic tree construction. iTOL online tool [42] (https://itol.embl.de/) was used in order to visualize the phylogenetic trees.

2.3. Target recognition & siRNA designing

Target-specific siRNAs were designed with the help of siDirect web server [43]. Rules of Ui-Tei [44], Amarzguioui [45] and Reynolds [46] were used (Table 1) and the melting temperature was kept below 21.5 °C as a default parameter for siRNA duplex.

2.4. Off target similarity search using BLAST

Blast search was performed against human genome and transcriptome using the standalone blast package [47] 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.

2.5. GC content calculation & secondary structure prediction

OligoCalc [48] was used to calculate the GC content. The secondary structure of siRNAs were predicted along with the respective free energy using MaxExpect [49] program in the RNA structure web server [50]. The higher values indicate better candidates as those molecules are less prone to folding.

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	-30.0	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

2.6. Computation 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 [51] program of the RNA structure web server [50].

2.7. Computation of heat capacity & concentration plot

DINA Melt webserver [52] 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).

2.8. 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.

3. Results

3.1. 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%) (Fig 2, 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.

3.2. 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.

3.3. Off-target binding exclusion using blast

Standalone blast [47] 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.

3.4. 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.

3.5. 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).

3.6. 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).

3.7. Validation and selection of best 8 siRNAs

siRNAPred[53] 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).

4. Discussion

COVID-19 is an emerging disease that lays bare the society we have created and its interdependent infrastructure with a surge in cases and deaths since its initial 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, hence it is the focus of our study.

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. Phylogenetic analysis revealed that a small number sequences form significant clades with a bootstrap value greater than 60%. (Fig 2, Supplementary Table 8). 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. siRNA bond formation with non-target sequences were eliminated by optimizing the melting temperature (Tm) below 21.5◦C. The equation to calculate melting temperature (Tm) is below,

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

Here,

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

siRNA's 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, approximately from 31.6 to 57.9%, is ideal for a siRNA to be effective [54]. Therefore, we calculated the GC content of the predicted siRNAs. Molecules that have GC content lower than 32% were not kept in the final selection. Here, GC content ranged from 10% to 43% for all the 78 predicted species. GC content of finally 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 prospective secondary structure and determination of free energy of corresponding folding is crucial. Here, guide strands of predicted siRNAs were subjected to RNA structure web server in order to predict possible folding structures and corresponding minimum free energies. At 37◦C, finally selected siRNAs have free energy of folding greater than zero (Fig 3, Table 2), which suggests the predicted siRNAs are more accessible for efficient binding.

DuplexFold[51] 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 78 predicted siRNAs spanned from -36.6 to -21.6 (Supplementary Table 6 & Supplementary table 7). Finally selected siRNAs have free energy of binding equal or below -30.0 (Fig 4, Table 2), which suggests the predicted siRNAs are more interactive with their corresponding targets.

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. DINA Melt webserver [52] was used to obtain the melting temperatures. All the selected siRNAs have high Tm value (>75◦C) (Table 2).

Lastly, siRNAPred[53] 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. Here, siRNA candidates that have validity score greater than one 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.

5. Conclusions

Computational methods can be employed to design and predict siRNA interaction against specific gene target thereby silencing its expression. In this research, 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.

Author contribution statement

Literature Review: UFC, KIH, MUSS; Data Collection: UFC, MUSS; Data Analysis: UFC, MUSS; Figure: UFC, MKSS; Write-up: UFC, MUSS, KIH, MAB, MKSS, MAM;

Declaration of Competing Interest

The authors declare that they have no competing interests.

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet. 395 (2020) 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.

[2] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet. 395 (2020) 507–513. https://doi.org/10.1016/S0140-6736(20)30211-7.

[3] F. Song, N. Shi, F. Shan, Z. Zhang, J. Shen, H. Lu, Y. Ling, Y. Jiang, Y. Shi, Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia, Radiology. 295 (2020) 210–217. https://doi.org/10.1148/radiol.2020200274.

[4] P. Zhou, X. Lou Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R. Di Jiang, M.Q. Liu, Y. Chen, X.R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F.X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature. 579 (2020) 270–273. https://doi.org/10.1038/s41586-020-2012-7.

[5] J.F.W. Chan, S. Yuan, K.H. Kok, K.K.W. To, H. Chu, J. Yang, F. Xing, J. Liu, C.C.Y. Yip, R.W.S. Poon, H.W. Tsoi, S.K.F. Lo, K.H. Chan, V.K.M. Poon, W.M. Chan, J.D. Ip, J.P. Cai, V.C.C. Cheng, H. Chen, C.K.M. Hui, K.Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, Lancet. 395 (2020) 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9.

[6] V.M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D.K. Chu, T. Bleicker, S. Brünink, J. Schneider, M.L. Schmidt, D.G. Mulders, B.L. Haagmans, B. van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J.L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M.P. Koopmans, C. Drosten, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, Euro Surveill. 25 (2020). https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045.

[7] D.K.W. Chu, Y. Pan, S.M.S. Cheng, K.P.Y. Hui, P. Krishnan, Y. Liu, D.Y.M. Ng, C.K.C. Wan, P. Yang, Q. Wang, M. Peiris, L.L.M. Poon, Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia, Clin. Chem. 66 (2020) 549–555.

https://doi.org/10.1093/clinchem/hvaa029.
[8] S. Khan, R. Nakajima, A. Jain, R.R. de Assis, A. Jasinskas, J.M. Obiero, O. Adenaiye, S. Tai, F. Hong, D.K. Milton, H. Davies, P.L. Felgner, P.S. Group, Analysis of Serologic Cross-Reactivity Between Common Human Coronaviruses and SARS-CoV-2 Using Coronavirus Antigen Microarray, BioRxiv. (2020) 2020.03.24.006544. https://doi.org/10.1101/2020.03.24.006544.
[9] N.M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A.R. Akhmetzhanov, S. Jung, B. Yuan, R. Kinoshita, H. Nishiura, Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data, J. Clin. Med. 9 (2020) 538. https://doi.org/10.3390/jcm9020538.
[10] P. Yu, J. Zhu, Z. Zhang, Y. Han, A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period, J. Infect. Dis. (2020). https://doi.org/10.1093/infdis/jiaa077.
[11] T.T.-Y. Lam, M.H.-H. Shum, H.-C. Zhu, Y.-G. Tong, X.-B. Ni, Y.-S. Liao, W. Wei, W.Y.-M. Cheung, W.-J. Li, L.-F. Li, G.M. Leung, E.C. Holmes, Y.-L. Hu, Y. Guan, Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins, Nature. (2020) 1–6. https://doi.org/10.1038/s41586-020-2169-0.
[12] F. Li, Structure, Function, and Evolution of Coronavirus Spike Proteins, Annu. Rev. Virol. 3 (2016) 237–261. https://doi.org/10.1146/annurev-virology-110615-042301.
[13] Z. Song, Y. Xu, L. Bao, L. Zhang, P. Yu, Y. Qu, H. Zhu, W. Zhao, Y. Han, C. Qin, From SARS to MERS, thrusting coronaviruses into the spotlight, Viruses. 11 (2019). https://doi.org/10.3390/v11010059.
[14] P.S. Masters, The Molecular Biology of Coronaviruses, Adv. Virus Res. 65 (2006) 193–292. https://doi.org/10.1016/S0065-3527(06)66005-3.
[15] C. Wang, X. Zheng, W. Gai, Y. Zhao, H. Wang, H. Wang, N. Feng, H. Chi, B. Qiu, N. Li, T. Wang, Y. Gao, S. Yang, X. Xia, MERS-CoV virus-like particles produced in insect cells induce specific humoural and cellular imminity in rhesus macaques, Oncotarget. 8 (2017) 12686–12694. https://doi.org/10.18632/oncotarget.8475.
[16] Q. Huang, L. Yu, A.M. Petros, A. Gunasekera, Z. Liu, N. Xu, P. Hajduk, J. Mack, S.W. Fesik, E.T. Olejniczak, Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein, Biochemistry. 43 (2004) 6059–6063. https://doi.org/10.1021/bi036155b.
[17] C.A.M. de Haan, P.J.M. Rottier, Molecular Interactions in the Assembly of Coronaviruses, Adv. Virus Res. 64 (2005) 165–230. https://doi.org/10.1016/S0065-3527(05)64006-7.
[18] H. Chen, A. Gill, B.K. Dove, S.R. Emmett, C.F. Kemp, M.A. Ritchie, M. Dee, J.A. Hiscox, 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., J. Virol. 79 (2005) 1164–79. https://doi.org/10.1128/JVI.79.2.1164-1179.2005.
[19] K.R. Hurst, R. Ye, S.J. Goebel, P. Jayaraman, P.S. Masters, An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA., J. Virol. 84 (2010) 10276–88. https://doi.org/10.1128/JVI.01287-10.
[20] S.A. Kopecky-Bromberg, L. Martínez-Sobrido, M. Frieman, R.A. Baric, P. Palese, Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists., J. Virol. 81 (2007) 548–57. https://doi.org/10.1128/JVI.01782-06.
[21] M.A. Tortorici, D. Veesler, Structural insights into coronavirus entry, in: Adv. Virus Res., Academic Press Inc., 2019: pp. 93–116. https://doi.org/10.1016/bs.aivir.2019.08.002.
[22] M.W. Howard, E.A. Travanty, S.A. Jeffers, M.K. Smith, S.T. Wennier, L.B. Thackray, K. V. Holmes, 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, J. Virol. 82 (2008) 2883–2894. https://doi.org/10.1128/jvi.01805-07.

[23] A.C. Walls, M.A. Tortorici, B.-J. Bosch, B. Frenz, P.J.M. Rottier, F. DiMaio, F.A. Rey, D. Veesler, Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer, Nature. 531 (2016) 114–117. https://doi.org/10.1038/nature16988.

[24] J.E. Park, K. Li, A. Barlan, A.R. Fehr, S. Perlman, P.B. McCray, T. Gallagher, Proteolytic processing of middle east respiratory syndrome coronavirus spikes expands virus tropism, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 12262–12267. https://doi.org/10.1073/pnas.1608147113.

[25] A. Levanova, M.M. Poranen, RNA interference as a prospective tool for the control of human viral infections, Front. Microbiol. 9 (2018) 2151. https://doi.org/10.3389/fmicb.2018.02151.

[26] M.U. Sharif Shohan, A. Paul, M. Hossain, Computational design of potential siRNA molecules for silencing nucleoprotein gene of rabies virus, Future Virol. 13 (2018) 159–170. https://doi.org/10.2217/fvl-2017-0117.

[27] F.T. Chowdhury, M.U.S. Shohan, T. Islam, T.T. Mimu, P. Palit, A Therapeutic Approach Against Leishmania donovani by Predicting RNAi Molecules Against the Surface Protein, gp63, Curr. Bioinform. 14 (2019) 541–550. https://doi.org/10.2174/1574893613666180828095737.

[28] A.J. Hamilton, D.C. Baulcombe, A species of small antisense RNA in posttranscriptional gene silencing in plants, Science (80-.). 286 (1999) 950–952. https://doi.org/10.1126/science.286.5441.950.

[29] B. Zheng, Y. Guan, Q. Tang, C. Du, F.Y. Xie, M.-L. He, K.-W. Chan, K.-L. Wong, E. Lader, M.C. Woodle, P.Y. Lu, B. Li, N. Zhong, Prophylactic and therapeutic effects of small interfering RNA targeting SARS-coronavirus., Antivir. Ther. 9 (2004) 365–374.

[30] E.R. Kabir, M.K.S. Siam, N. Mustafa, Scaffold of N-(2-(2-(tosylcarbamoyl)hydrazinyl)ethyl)isonicotinamidereveals anticancer effects through selective inhibition of FAP, in: ACM Int. Conf. Proceeding Ser., Association for Computing Machinery, New York, NY, USA, 2019: pp. 1–11. https://doi.org/10.1145/3365953.3365963.

[31] E.R. Kabir, N. Mustafa, M. Kawsar, S. Siam, S.M. Kabir, Molecular docking reveals pitavastatin and related molecules antagonize 1DHF and its pseudogene DHFR2 in cancer treatment., in: ACM Int. Conf. Proceeding Ser., Association for Computing Machinery, New York, New York, USA, 2018: pp. 1–9. https://doi.org/10.1145/3291757.3291763.

[32] E.R. Kabir, M.K.S. Siami, S.M. Kabir, A. Khan, S.A. Rajib, Drug repurposing: Targeting mTOR inhibitors for anticancer activity, in: ACM Int. Conf. Proceeding Ser., Association for Computing Machinery, New York, New York, USA, 2017: pp. 68–75. https://doi.org/10.1145/3156346.3156359.

[33] M.K.S. Siam, M.S. Hossain, E.R. Kabir, S.A. Rajib, In Silico structure based designing of dihydrofolate reductase enzyme antagonists and potential small molecules that target DHFR protein to inhibit the folic acid biosynthetic pathways, in: ACM Int. Conf. Proceeding Ser., Association for Computing Machinery, New York, New York, USA, 2017: pp. 62–67. https://doi.org/10.1145/3156346.3156358.

[34] R. Rahman, S.M.M. Rashid, M. Sayeem, I. Sharif, K. Sharif, Surface proteins, potential drug target for antiviral therapy against Nipah virus and in silico drug design, Clin. Biochem. 44 (2011) S34. https://doi.org/10.1016/j.clinbiochem.2011.08.1035.

[35] F. Almazan, C. Galan, L. Enjuanes, The Nucleoprotein Is Required for Efficient Coronavirus Genome Replication, J. Virol. 78 (2004) 12683–12688. https://doi.org/10.1128/jvi.78.22.12683-12688.2004.

[36] C.J. Wu, H.W. Huang, C.Y. Liu, C.F. Hong, Y.L. Chan, Inhibition of SARS-CoV replication by siRNA, Antiviral Res. 65 (2005) 45–48. https://doi.org/10.1016/j.antiviral.2004.09.005.

[37] E.L. Hatcher, S.A. Zhdanov, Y. Bao, O. Blinkova, E.P. Nawrocki, Y. Ostapchuck, A.A. Schäffer, J.R. Brister, Virus Variation Resource – improved response to emergent viral

outbreaks, Nucleic Acids Res. 45 (2016) D482–D490. https://doi.org/10.1093/nar/gkw1065.
[38] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680. https://doi.org/10.1093/nar/22.22.4673.
[39] K. Tamura, M. Nei, Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees., Mol. Biol. Evol. 10 (1993) 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023.
[40] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549. https://doi.org/10.1093/molbev/msy096.
[41] S. Kumar, G. Stecher, D. Peterson, K. Tamura, MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis, Bioinformatics. 28 (2012) 2685–2686. https://doi.org/10.1093/bioinformatics/bts507.
[42] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, Nucleic Acids Res. 47 (2019) W256–W259. https://doi.org/10.1093/nar/gkz239.
[43] Y. Naito, J. Yoshimura, S. Morishita, K. Ui-Tei, SiDirect 2.0: Updated software for designing functional siRNA with reduced seed-dependent off-target effect, BMC Bioinformatics. 10 (2009) 392. https://doi.org/10.1186/1471-2105-10-392.
[44] K. Ui□Tei, Y. Naito, F. Takahashi, T. Haraguchi, H. Ohki□Hamazaki, A. Juni, R. Ueda, K. Saigo, Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference, Nucleic Acids Res. 32 (2004) 936–948. https://doi.org/10.1093/nar/gkh247.
[45] M. Amarzguioui, H. Prydz, An algorithm for selection of functional siRNA sequences, Biochem. Biophys. Res. Commun. 316 (2004) 1050–1058. https://doi.org/10.1016/j.bbrc.2004.02.157.
[46] A. Reynolds, D. Leake, Q. Boese, S. Scaringe, W.S. Marshall, A. Khvorova, Rational siRNA design for RNA interference, Nat. Biotechnol. 22 (2004) 326–330. https://doi.org/10.1038/nbt936.
[47] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: Architecture and applications, BMC Bioinformatics. 10 (2009) 421. https://doi.org/10.1186/1471-2105-10-421.
[48] W.A. Kibbe, OligoCalc: an online oligonucleotide properties calculator, Nucleic Acids Res. 35 (2007) W43–W46. https://doi.org/10.1093/nar/gkm234.
[49] Z.J. Lu, J.W. Gloor, D.H. Mathews, Improved RNA secondary structure prediction by maximizing expected pair accuracy, RNA. 15 (2009) 1805–1813. https://doi.org/10.1261/rna.1643609.
[50] S. Bellaousov, J.S. Reuter, M.G. Seetin, D.H. Mathews, RNAstructure: web servers for RNA secondary structure prediction and analysis, Nucleic Acids Res. 41 (2013) W471–W474. https://doi.org/10.1093/nar/gkt290.
[51] D. Piekna-Przybylska, L. DiChiacchio, D.H. Mathews, R.A. Bambara, A sequence similar to tRNA3Lys gene is embedded in HIV-1 U3–R and promotes minus-strand transfer, Nat. Struct. Mol. Biol. 17 (2010) 83–90. https://doi.org/10.1038/nsmb.1687.
[52] N.R. Markham, M. Zuker, DINAMelt web server for nucleic acid melting prediction, Nucleic Acids Res. 33 (2005) W577–W581. https://doi.org/10.1093/nar/gki591.
[53] M. Kumar, S. Lata, G.R.-P. of the First, U. 2009, siRNApred: SVM based method for predicting efficacy value of siRNA, CSIR-IMTECH. (2009).
[54] C.Y. Chan, C.S. Carmack, D.D. Long, A. Maliyekkel, Y. Shao, I.B. Roninson, Y. Ding, A structural interpretation of the effect of GC-content on efficiency of RNA interference, BMC Bioinforma. 2009 101. 10 (2009) 1–7. https://doi.org/10.1186/1471-2105-10-s1-s33.
[55] M. Jinek, J.A. Doudna, A three-dimensional view of the molecular machinery of RNA

interference, Nature. 457 (2009) 405–412. https://doi.org/10.1038/nature07755.

Figure captions:

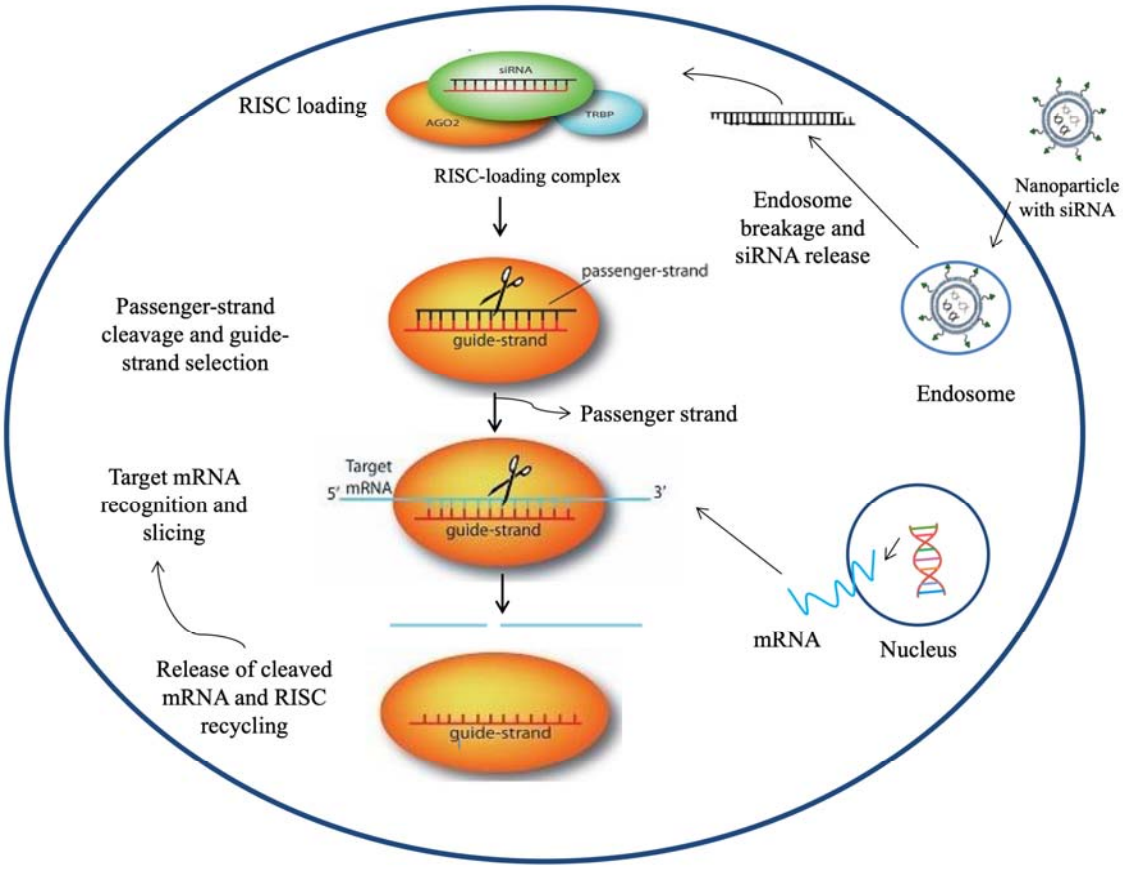


Fig. 1. Graphical representation of the siRNA-mediated gene silencing mechanism. Nanoparticles loaded with siRNA are taken up through endocytosis by the cells. These particles are then trapped into the endosomes. siRNA escape endosomes and release siRNA into the cytoplasm due to pH responsive mechanism or proton sponge effect. Once generated, siRNA is loaded into RNA-indiced silencing complex comprising of RNA-binding protein TRBP and Argonaute (AGO2). AGO2 opts the siRNA guide strand, then excises 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 [55].

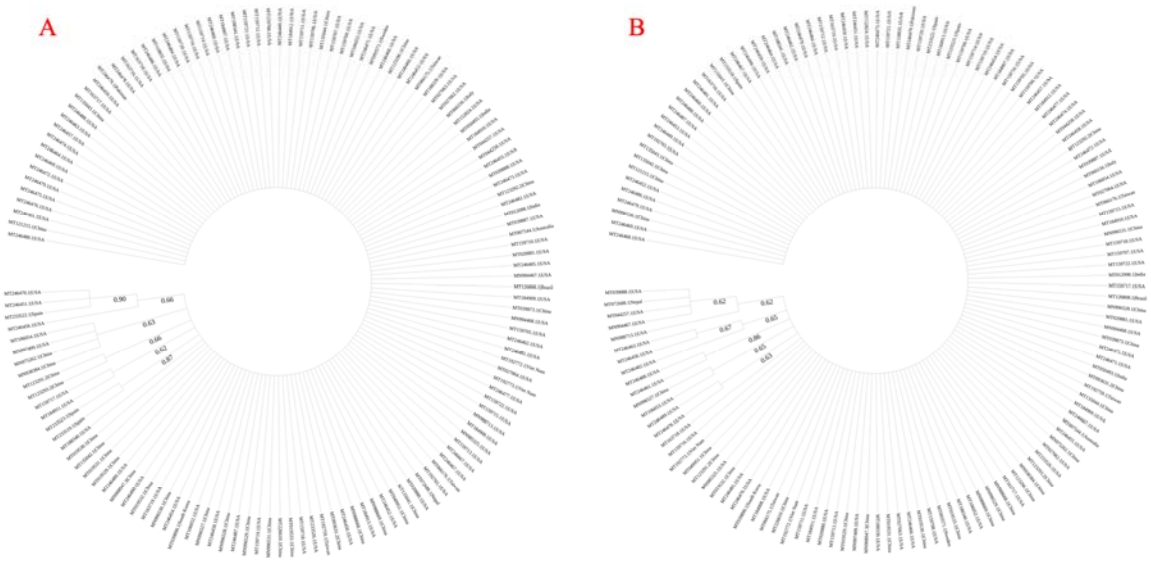


Fig. 2. Radial phylogenetic tree of A. nucleocapsid phosphoprotein and B. surface glycoprotein using 139 strains of SARS-CoV-2 from around the world. The bootstrap value for tree construction was set to 500 and Tamura-Nei model of evolution was used for both trees.

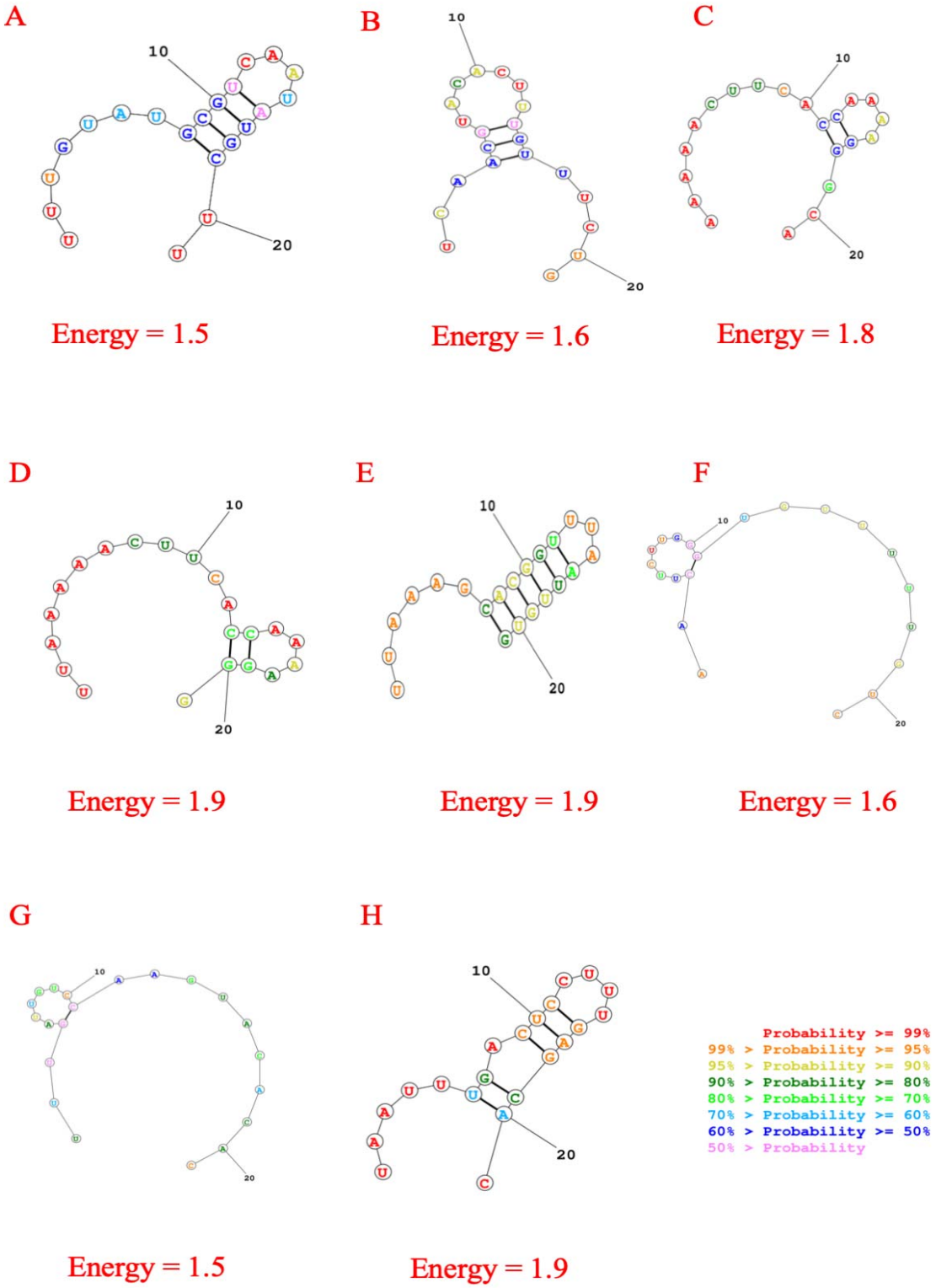


Fig. 3. Secondary structures of best eight predicted siRNA with probable folding and lowest free energy for consensus sequence. The structures are for A. n7 B.g15 C. g21 D. g22 E. g44 F. g46 G. g59 H. g70 siRNAs.

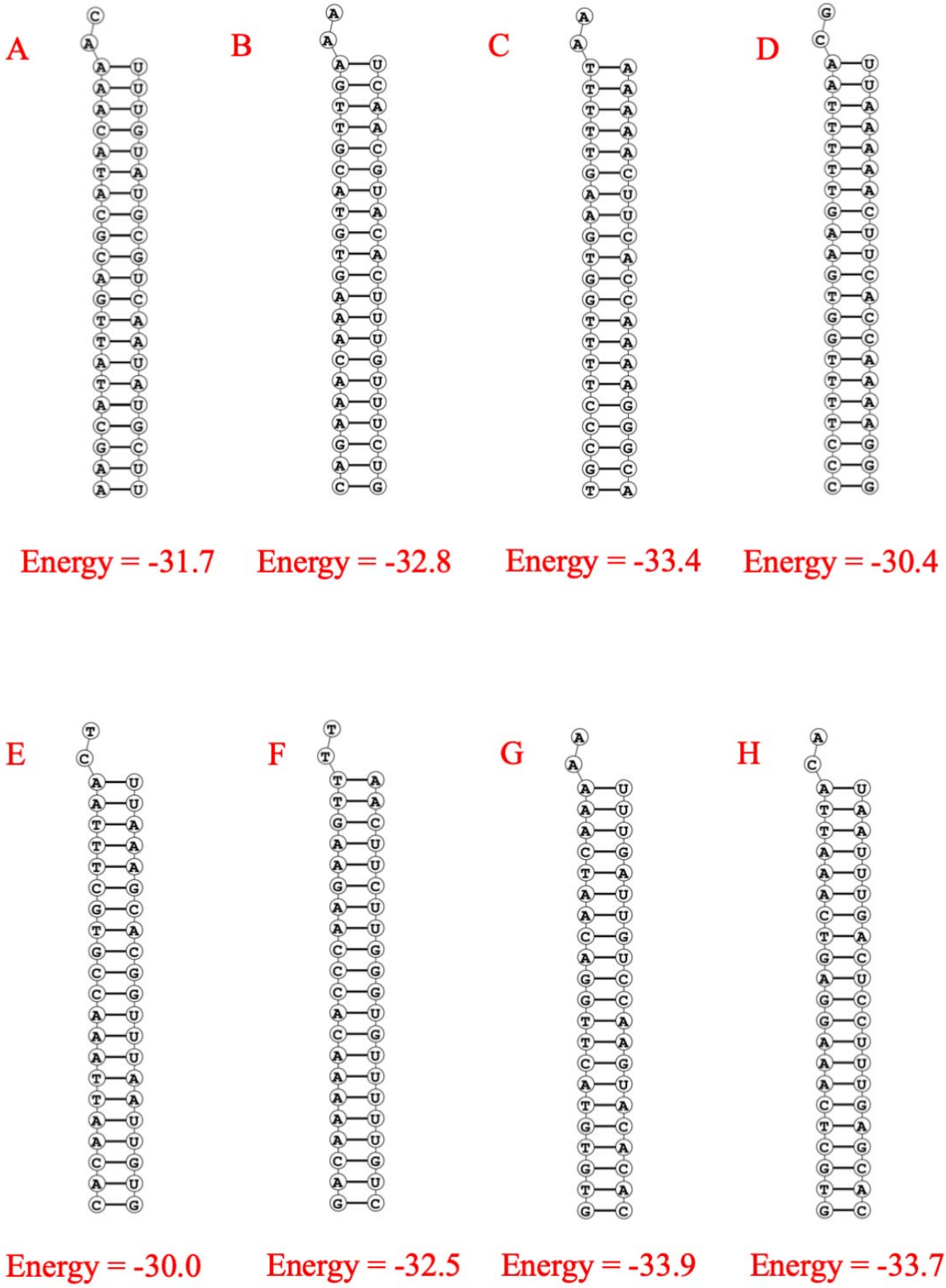
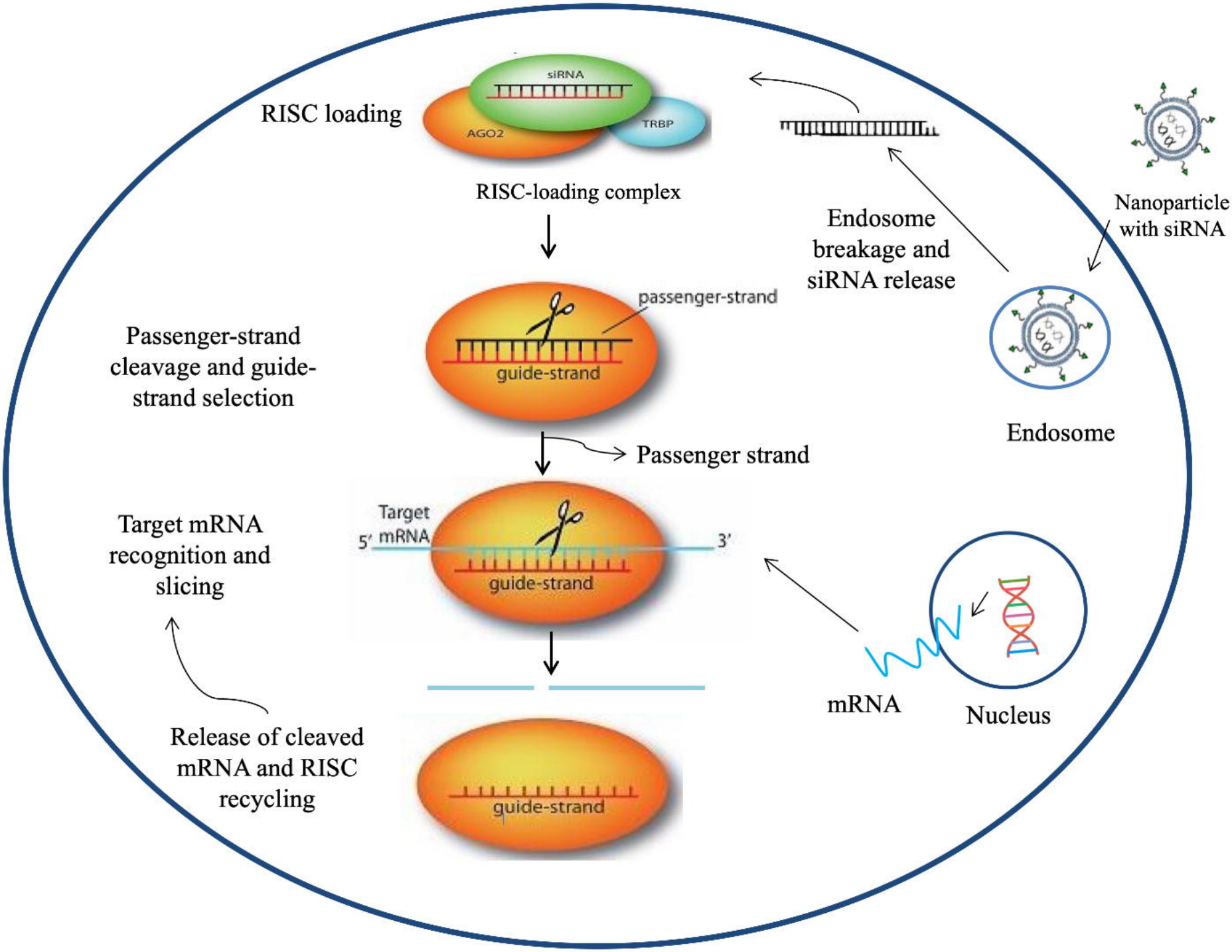


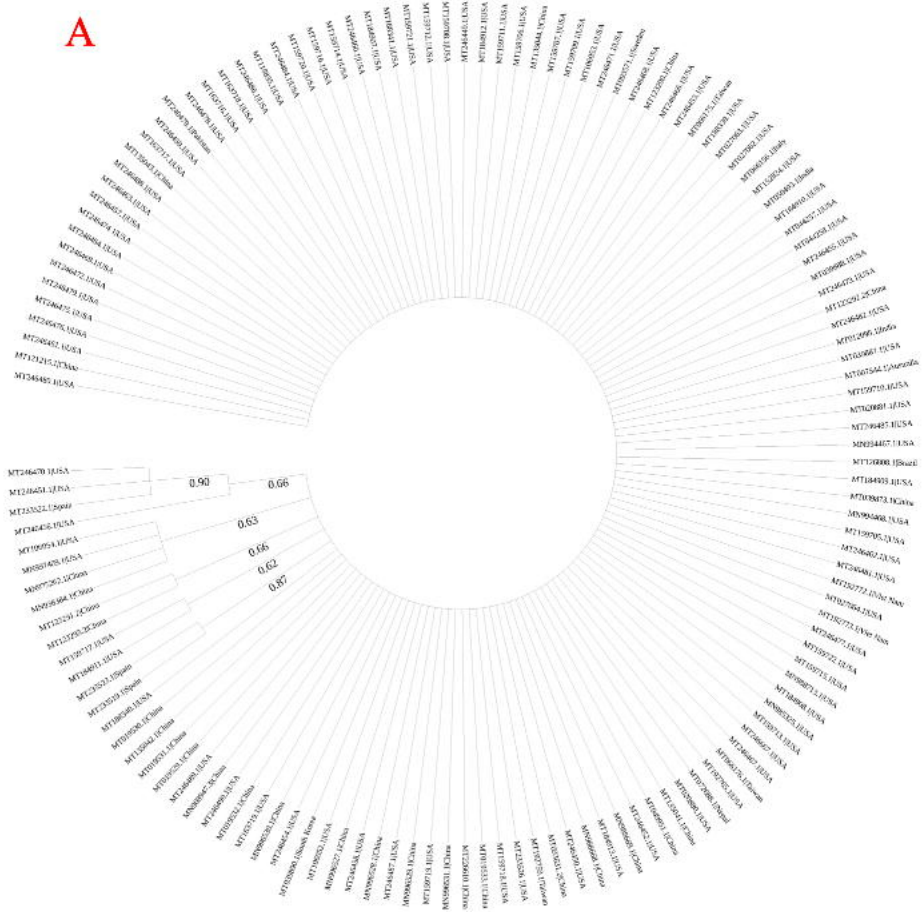
Fig. 4. Structure of binding of siRNA (guide strand) and target RNA with corresponding predicted minimum free energy. The structures are for A. n7 B.g15 C. g21 D. g22 E. g44 F. g46 G. g59 H. g70 siRNAs and their corresponding targets.

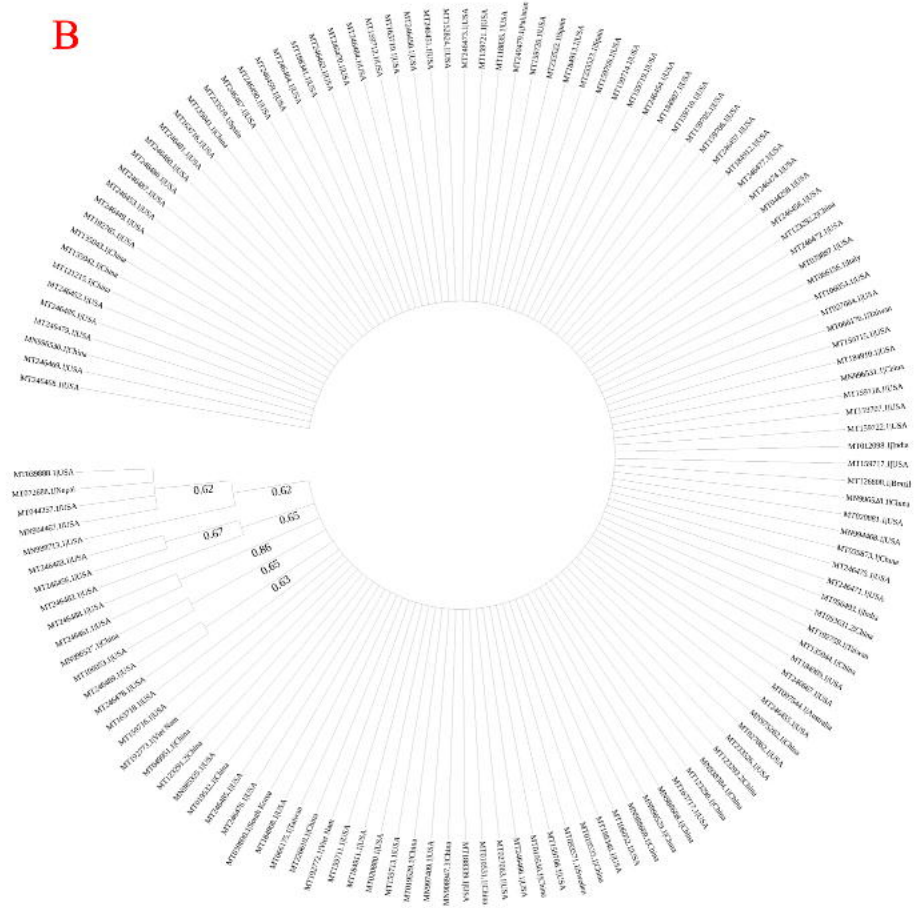
Table Captions:

Table 1: Rules/Algorithm to construct siRNAs.
Table 2: Best effective siRNA molecules with various parameters.

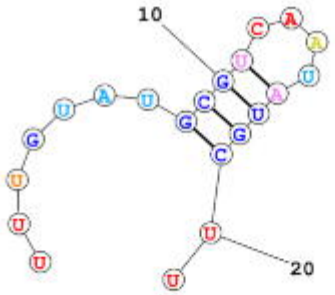
Supplement Table Captions:
Supplementary Table 1: Accession number, Length and Location of the SARS-CoV-2 Stains Used in the Study.
Supplementary Table 2: Conserved sequence of nucleocapsid phosphoprotein gene.
Supplementary Table 3: Conserved sequence of surface glycoprotein gene.
Supplementary Table 4: Tm values of predicted siRNA (guide strand and passenger strand) against nucleocapsid phosphoprotein.
Supplementary Table 5: Tm values of predicted siRNA (guide strand and passenger strand) against surface glycoprotein.
Supplementary Table 6: Effective siRNAs against nucleocapsid phosphoprotein with GC%, free energy of folding and free energy of binding with target.
Supplementary Table 7: Effective siRNAs against surface glycoprotein with GC%, free energy of folding and free energy of binding with target.
Supplementary Table 8: Bootstrap values of significant clades in the radial phylogenetic tree.


RISC loading
siRNA
AGO2
TRBP
RISC-loading complex
Endosome breakage and siRNA release
Nanoparticle with siRNA
Endosome
Passenger-strand cleavage and guide-strand selection
passenger-strand
guide-strand
Passenger strand
Target mRNA recognition and slicing
Target mRNA
5′
3′
guide-strand
mRNA
Nucleus
Release of cleaved mRNA and RISC recycling
guide-strand


A
0.90
0.66
0.63
0.66
0.62
0.87

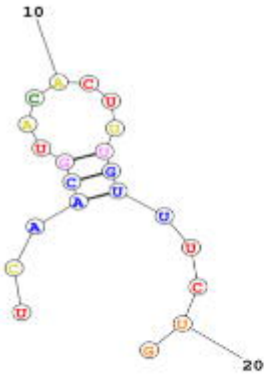

B
0.62
0.62
0.65
0.67
0.66
0.65
0.63

A



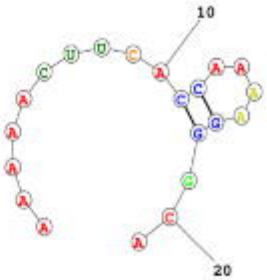
Energy = 1.5

B



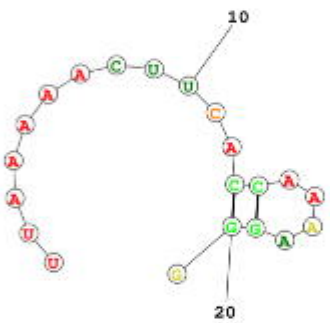
Energy = 1.6

C



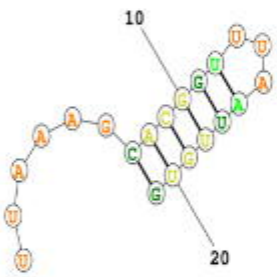
Energy = 1.8

D



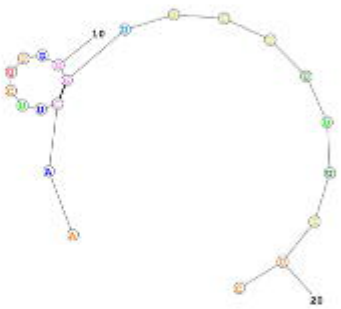
Energy = 1.9

E



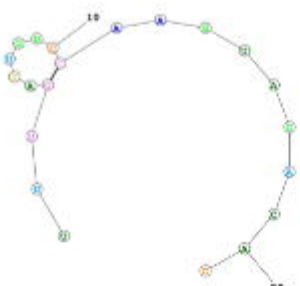
Energy = 1.9

F



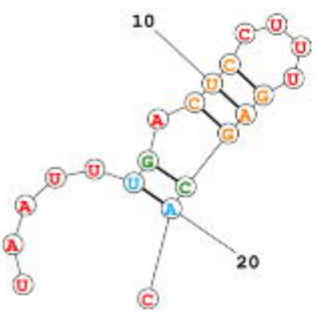
Energy = 1.6

G



Energy = 1.5

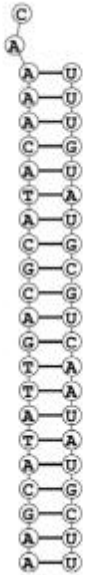
H



Energy = 1.9

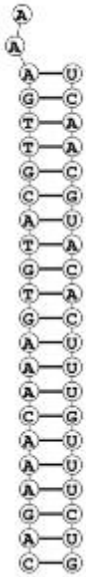
Probability >= 99%
99% > Probability >= 95%
95% > Probability >= 90%
90% > Probability >= 80%
80% > Probability >= 70%
70% > Probability >= 60%
60% > Probability >= 50%
50% > Probability

A



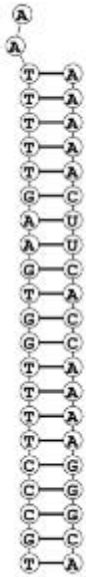
Energy = -31.7

B



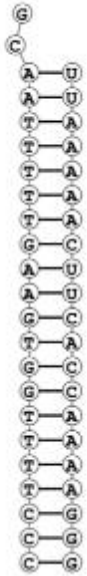
Energy = -32.8

C



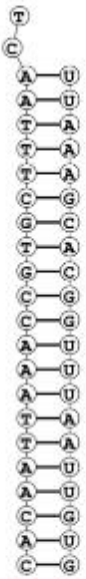
Energy = -33.4

D



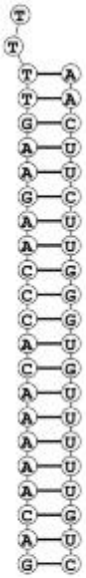
Energy = -30.4

E



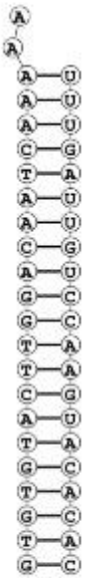
Energy = -30.0

F



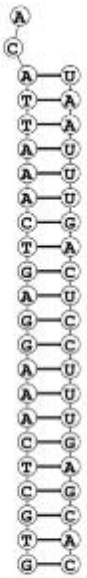
Energy = -32.5

G



Energy = -33.9

H



Energy = -33.7