



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Original Article

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^{a,1}, Mohammad Umer Sharif Shohan^{a,1}, Kazi Injamamul Hoque^a, Mirza Ashikul Beg^b, Mohammad Kawsar Sharif Siam^c, Mohammad Ali Moni^{d,*}

^a Department of Biochemistry and Molecular Biology, University of Dhaka, Bangladesh

^b Department of Genetic Engineering and Biotechnology, University of Dhaka, Bangladesh

^c Department of Pharmacy, Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh

^d WHO Collaborating Centre for eHealth, School of Public Health and Community Medicine, Faculty of Medicine, University of New South Wales (UNSW), Sydney, Australia

ARTICLE INFO

Keywords:

SARS-CoV-2

Nucleocapsid phosphoprotein

Surface glycoprotein

siRNA

siDirect

ABSTRACT

An outbreak, caused by an 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, efficacy prediction and molecular docking analysis, 8 siRNA molecules were selected which are proposed to exert 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.

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 and 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

* Corresponding author.

E-mail address: m.moni@unsw.edu.au (M.A. Moni).

¹ These authors contributed equally.

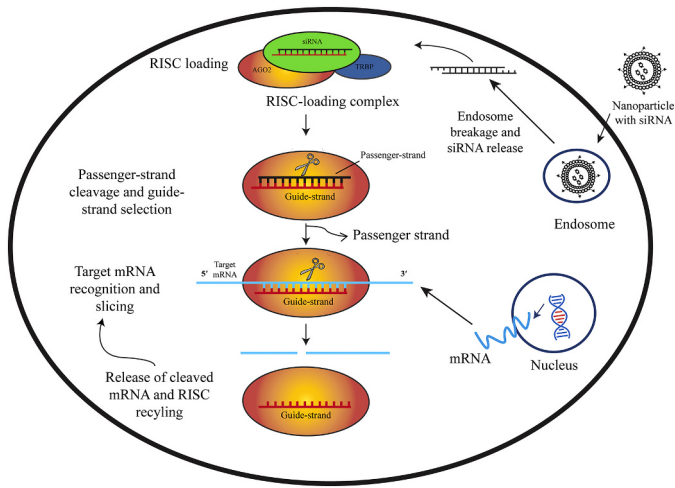


Fig. 1. Graphical representation of the siRNA-mediated gene silencing mechanism. Nanoparticles loaded with siRNA are taken up through endocytosis by the cells. These particles are then trapped into the endosomes. siRNA escape endosomes and release siRNA into the cytoplasm due to pH responsive mechanism or proton sponge effect. Once generated, siRNA is loaded into RNA-induced 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 [74].

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 an 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].

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 (Fig. 1)

As of November 14, 2020, total infected case is 53,521,552 and

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 9 nt
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 ($T_m > -20^\circ\text{C}$) 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

among these patients 1,306,135 people have died which means case fatality rate (CFR) is approximately 2.44%. 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]. Currently, ONPATRO® and GIVLAARI™ RNAi therapeutics are commercially approved on the market for the treatment of polyneuropathy and acute hepatic porphyria respectively [37]. We hope this study will help to develop a similar treatment strategy for SARS-CoV-2.

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 [38] 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 [39] algorithm was employed to perform multiple sequence alignment (MSA). Maximum likelihood phylogenetic trees were constructed with a bootstrap value of 500. Tamura Nei [40] model of evolution was selected while constructing the phylogenetic tree. MEGA-X [41] and MEGA-CC [42] programs were used for alignment formation and phylogenetic tree construction respectively. iTOL online tool [43] (<https://itol.embl.de/>) was used in order to visualize the phylogenetic trees.

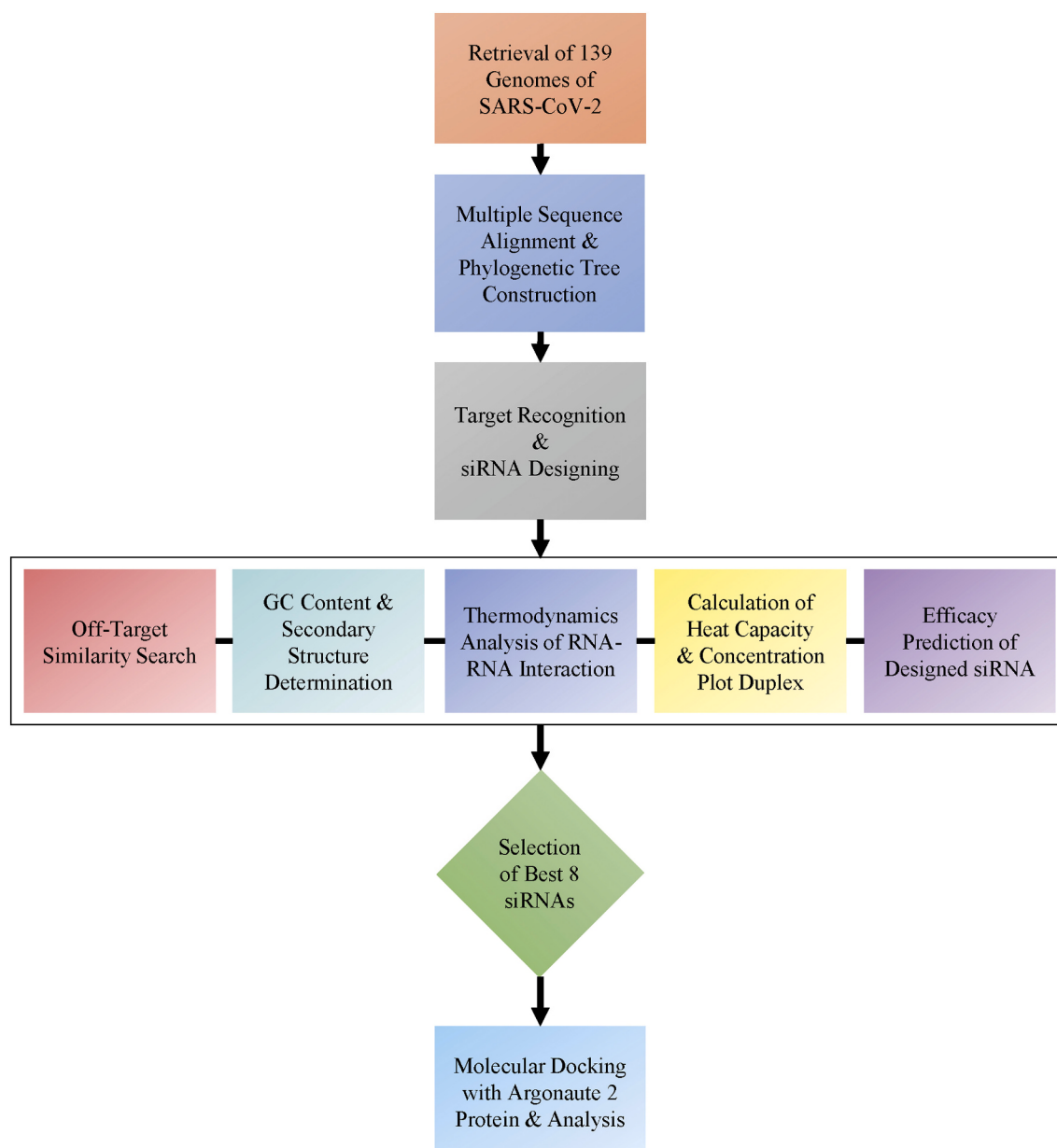


Fig. 2. The graphical workflow of the methodology used in the study.

2.3. Target recognition & siRNA designing

Target-specific siRNAs were designed with the help of siDirect web server [44]. Rules of Ui-Tei [45], Amarzguioui [46] and Reynolds [47] 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 [48] 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 [49] was used to calculate the GC content. The secondary

structure of siRNAs was predicted along with the respective free energy using MaxExpect [50] program in the RNA structure web server [51]. Here, the higher values of energy indicate better candidates as those molecules are less prone to folding.

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 [52] program of the RNA structure web server [51].

2.7. Computation of heat capacity & concentration plot

DINA Melt web server [53] was used to generate heat capacity and concentration plot. The ensemble heat capacity (C_p) is plotted as a function of temperature, with the melting temperature T_m (C_p)

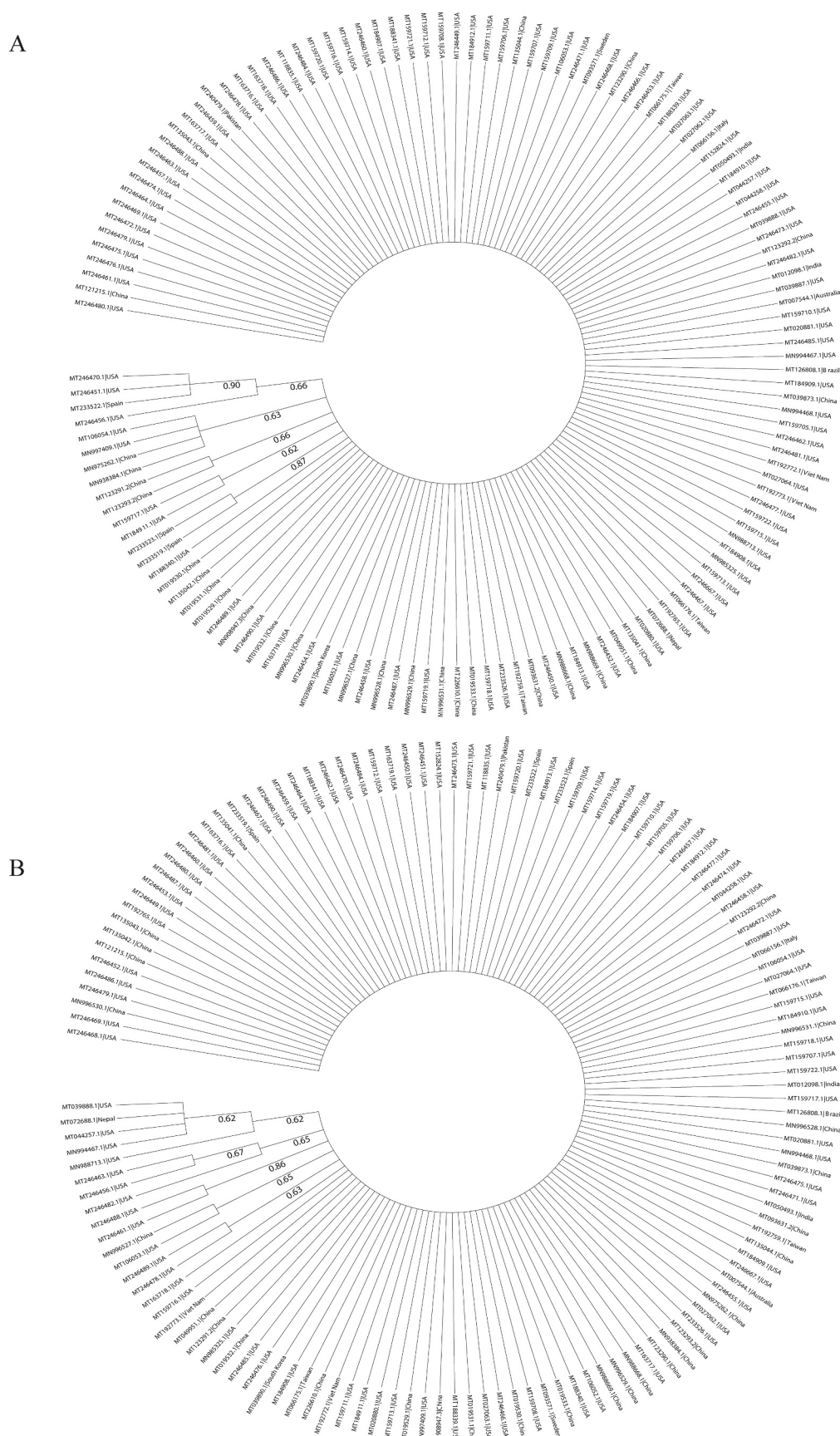


Fig. 3. Radial phylogenetic tree of A. nucleocapsid phosphoprotein and B. surface glycoprotein using 139 strains of SARS-CoV-2 from around the world. The bootstrap value for tree construction was set to 500 and Tamura-Nei model of evolution was used for both trees.

(Supplementary Table 6 & Supplementary Table 7). The contributions of each species to the ensemble heat capacity are 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 T_m (Conc) was shown using the concentration plot- T_m (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 Main 21 datasets using support vector machine algorithm and the binary pattern prediction approach. siRNAPred score greater than 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.

2.9. Modelling & docking of argonaute protein 2 and siRNAs

The three dimensional structure of argonaute 2 (Ago2) protein was formed using Modeler 9.25 [54,55] against the crystal structure of RNA bound argonaute protein (pdb id: 4z4d) [56] via homology modelling. The modelled protein was refined in the GalaxyRefine [57,58] server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>). The quality of the model was checked using Ramachandran Plot generated by PROCHECK [59].

The guide RNA, bound in the structure of 4z4d, was taken as template to model the siRNA. The sequence of the guide RNA structure was changed by mutating and appending the necessary nucleotides in Discovery Studio Visualizer (version 20.1) thereby, obtaining the particular siRNA structure that is reminiscent to the template structure. Thereafter, the siRNAs were optimized in MMFF94 force field using steepest descent algorithm for 1000 steps (convergence at $10e-7$) in Avogadro v1.2.0. Finally, molecular docking for the siRNAs against the argonaute protein was performed in HDock server (<http://hdock.phys.hust.edu.cn/>) [60,61]. The best results were taken based on the lowest docking energy score and docked complex where siRNA was bound to the binding cavity of Ago2. The structures were analyzed and the necessary images were generated using the PyMOL Molecular Graphics System (v1.8.4). A graphical workflow of the methodology used in this study was provided as well (Fig. 2).

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. 3). This implied that most of the viral sequences have been conserved and therefore could be used to construct siRNA which should cover a wide range of SARS-CoV-2 strains.

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 (T_m) values for all the predicted siRNAs were less than 21.5 °C which suggests the ability of predicted siRNAs to avoid non-target binding.

Table 2
Best effective siRNA molecules with various parameters.

Alias	Conserved position	Location of target within mRNA	siRNA target within mRNA	Duplex candidate at 37 °C	Predicted siRNA (guide): siRNA (passenger)	Seed-duplex (Tm) guide	Seed-duplex (Tm) passenger	GC %	Free energy of folding	Free energy of binding	T_m (Conc)	T_m (Cp)	Validity (binary)
n7	1031-1253	33-55	AAGCATATTGACGCATACAAAAC	GCAUUAUUGAGGCAUACAAAAC	UUUGUAUGCGUCAAUAUGCUU	13.5	5.6	36	1.5	-31.7	75.8	77	1.047
g15	871-1222	20-42	CAGAAACAAAGTGTAGTTGAAA	GAAACAAAGUGUAGGUUGAAA	UCAAAGUACACUUUUUUUCUG	21	5.6	36	1.6	-32.8	79.1	80	1.022
g21	871-1222	136-158	TGCCCTTTTGTGAAGTTTITAA	CCCUUUUGGUGAAGUUUUUAA	AAAAACUUCACCAAAAGGGCA	3.2	16.1	36	1.8	-33.4	83.4	85	1.039
g22	871-1222	138-160	CCCTTTTGTGAAGTTTITTAACG	CUUUUGGUGAAGUUUUUUAACG	UUAAAAACUUCACCAAAAGGG	0	18.8	33	1.9	-30.4	78.6	80	1.01
g44	1842-2389	441-463	CACAAATTAACCGTCTTTAACT	CAAUUAAACCGUGUUUUUAGU	UUAAAGCAGCGUUUAUUUGUG	19.7	-9.7	33	1.9	-30.0	79.5	80	1.102
g46	1842-2389	482-504	GACAAAAACACCAAGAGTTT	CAAAAACACCAAGAGUUUUU	AACUUCUGGUGUUUUUUGUC	17.7	5.6	36	1.6	-32.5	80.6	82	1.076
g59	2790-3822	304-326	GTGTGTACTTGTGACAAATCAAAA	GUGUACUUGGACAAUCAA AAA	UUUGAUUGUCCAGUACACAC	13.8	19	36	1.5	-33.9	80.4	81	1.007
g70	2790-3822	1001-1023	GTGCTCAAAGGAGTCAATTACA	GCUCAAAGGAGUCAAUUACA	UAAUUUGACUCCUUUGAGCAC	7.4	16.6	38	1.9	-33.7	81.7	83	1.012

Table 3
siRNA clusters, docking score and interaction statistics.

Cluster no.	siRNA in Ago2-siRNA complex	HDock docking score	Interaction statistics		
			Hydrogen bonds	Electrostatic bonds	Hydrophobic bonds
Cluster 1	g15	−317.2	45	40	12
	g70	−341.1	27	28	5
	g46	−357.8	50	31	11
	g59	−515.5	48	32	8
Cluster 2	g21	−328.2	17	13	4
	g22	−336.8	18	11	1
	n7	−356.5	18	16	2
	g44	−356.9	21	16	3

3.3. Off-target binding exclusion using blast

Standalone blast [48] search against human genome and transcriptome did not reveal any off-target match. This shows that our predicted siRNA would not interact in any places other than the viral target location.

3.4. GC content calculation & secondary structure determination

GC content analysis of the predicted siRNAs was ranged from 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 was 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 $T_m(C_p)$ and $T_m(Conc)$ were calculated for the predicted siRNAs. The higher values of these two melting temperatures indicate higher effectiveness of the siRNA species. $T_m(Conc)$ 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). $T_m(C_p)$ 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 [62] checked the effectivity of the predicted siRNAs and values greater than 1 are considered highly effective. 2 siRNAs for nucleocapsid phosphoprotein and 32 siRNAs for surface glycoprotein were found to be highly effective. 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) and advanced for molecular docking.

3.8. Analysis of molecular modelling and docking of best siRNAs & Ago2

According to the Ramachandran plot of Ago2 protein, the percentage of residues resided in core region, allowed region, moderately allowed region, and disallowed region were 95.7%, 4.0%, 0.1%, and 0.1%

respectively. The docking score, interaction statistics, interactive residues etc. were provided (Table 3, Supplementary Table 8). Ago2-siRNA complex containing g59 and g15 showed the highest and lowest docking score of −515.5 and −317.2 respectively, keeping rest of the values in between.

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 of sequences form significant clades with a bootstrap value greater than 60%. (Fig. 3). 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 was eliminated by optimizing the melting temperature (T_m) below 21.5 °C. The equation to calculate melting temperature (T_m) is below,

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

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 is represented by ΔS
- The gas constant (1.987 cal/deg./mol) is represented by R
- The total molecular concentration (100 μ M) 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. High GC content in the siRNA corresponds to high content in the target region. siRNA targets with high GC contents might have a propensity for folding which could potentially reduce target accessibility [63]. Usually a low GC content, approximately from 31.6 to 57.9%, is ideal for a siRNA to be effective [63]. 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 is 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

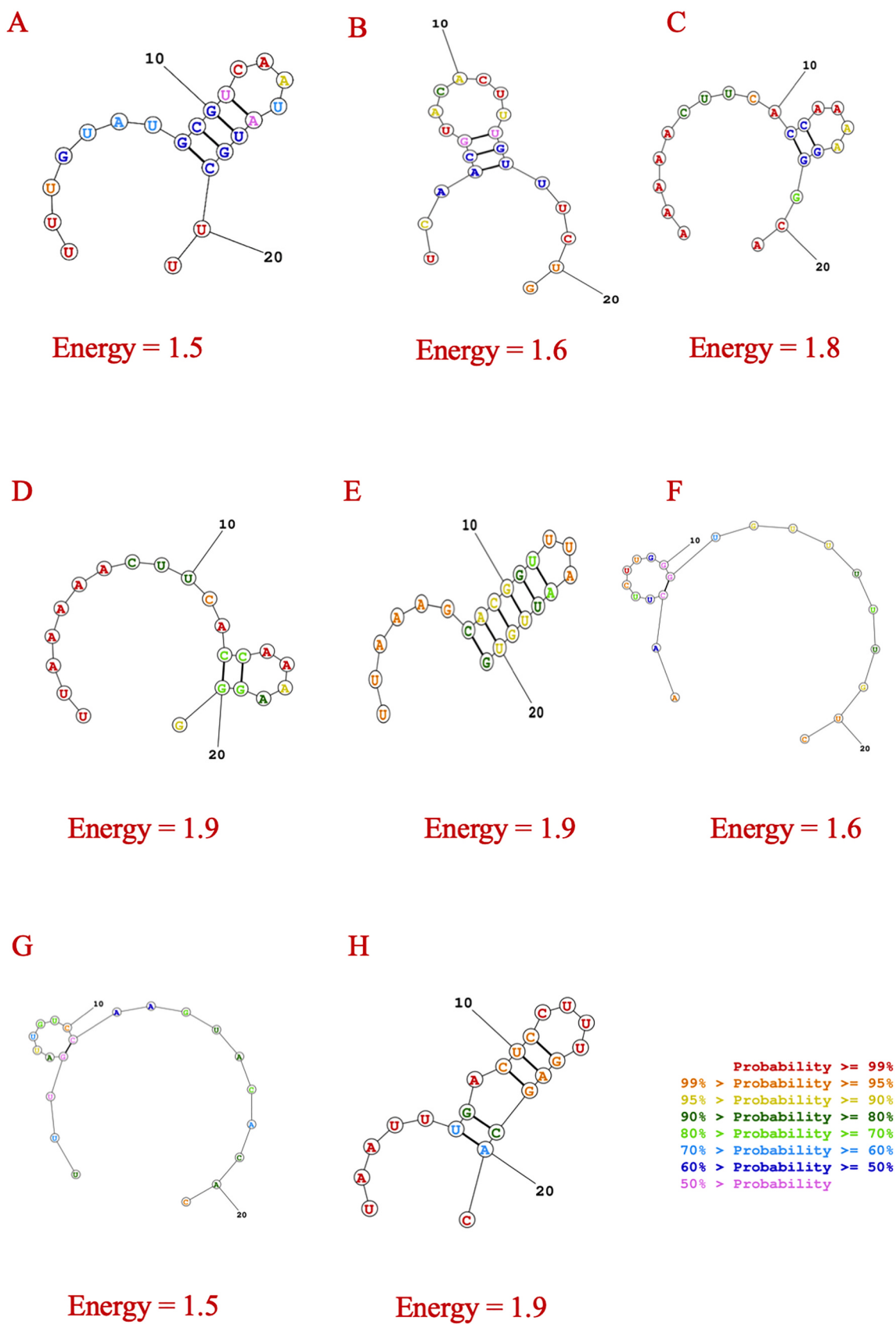


Fig. 4. Secondary structures of best eight predicted siRNA with probable folding and lowest free energy for consensus sequence. The structures are for A. n7 B.g15 C. g21 D. g22 E. g44 F. g46 G. g59 H. g70 siRNAs.

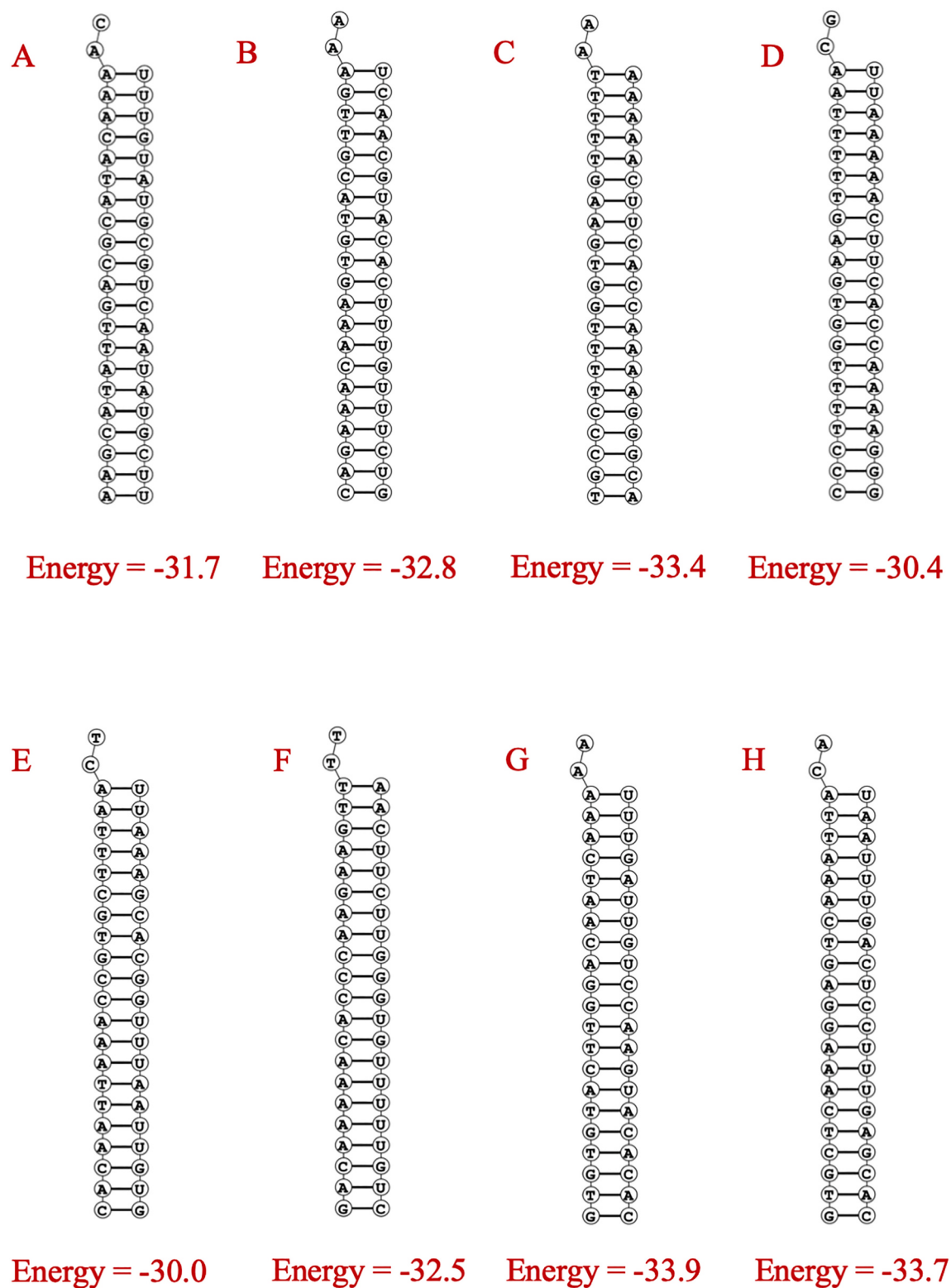


Fig. 5. Structure of binding of siRNA (guide strand) and target RNA with corresponding predicted minimum free energy. The structures are for A. n7 B.g15 C. g21 D. g22 E. g44 F. g46 G. g59 H. g70 siRNAs and their corresponding targets.

have free energy of folding greater than zero (Fig. 4, Table 2), which suggests the predicted siRNAs are more accessible for efficient binding. DuplexFold [52] was used to determine the target and guide siRNA interaction and their corresponding binding energy. Lower binding energy indicates 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. 5, 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

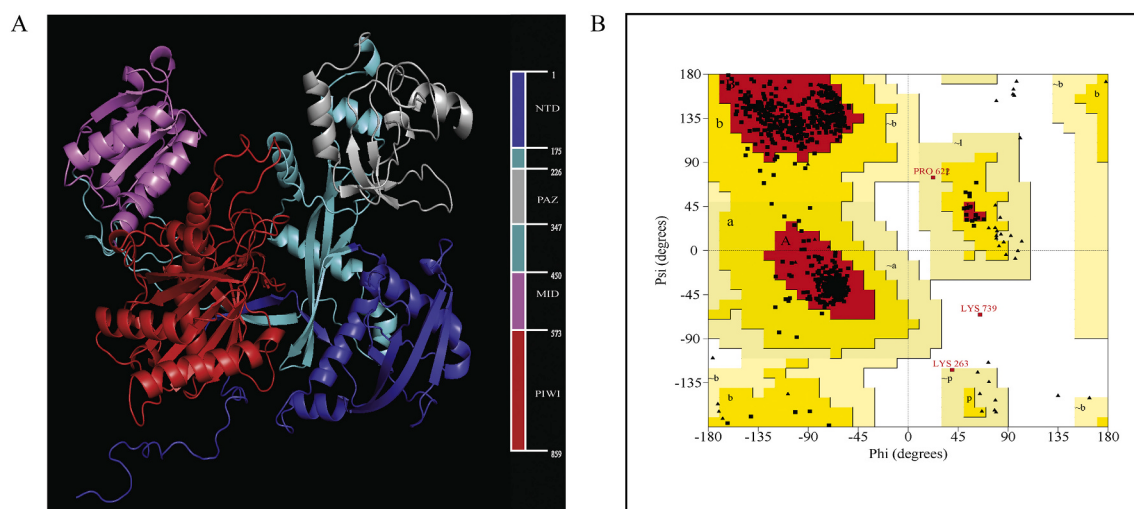


Fig. 6. A. The cartoon representation of the structure of modelled Argonaute 2 (Ago2) protein. The N-terminal domain (NTD), PAZ domain, MID domain and PIWI domain corresponds to blue, gray, pink and red color respectively. Rest of the protein was denoted by Aquamarine color. The sidebar corresponds to range of residues possessed by each domain. B. The Ramachandran plot of modelled Ago2 protein. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

determined. The contribution of each molecule to the collective heat capacity was demonstrated using the inclusive heat capacity plot where melting temperature T_m (Conc), indicates the temperature at which the concentration of double-stranded molecules becomes one-half of its maximum value. DINA Melt web server [53] was used to obtain the melting temperatures. All the selected siRNAs have high T_m value ($>75^\circ\text{C}$) (Table 2).

siRNAPred [62] 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. Considering optimal parameters from the results of GC content calculation, DuplexFold, DINA Melt, and siRNAPred webserver 8 best siRNA candidates were selected (Table 2). These siRNAs are expected to perform better than all the other predicted siRNAs in the study. Afterwards, siRNAs were considered for molecular docking analysis for better understanding their interaction with Ago2 protein which is the key enzyme in RISC.

The complete structure of Ago2 was predicted in Modeler 9.25 standalone package based on homology modelling which is essentially built on target template alignment. The process further comprises steps such as fold designation, model construction and quality control [54]. The missing residues of the protein in the crystal structure were successfully predicted, thus providing the opportunity of covering the maximum interactions between protein and siRNA complex. Afterwards, the model was refined in GalaxyRefine web server to improve the global as well as the local structure quality. The method initially reconstructs side chains, executes side-chain repacking and thereafter performs an overall relaxation by small molecular dynamics simulation in two steps namely, mild and aggressive relaxations. The resultant model had 95.7% residues inside the core region Ramachandran plot thereby surpassing the 90% cut-off and qualifying as a good quality structure.

In the cellular environment, there are multiple factors that are involved in the mechanism of siRNA mediated gene silencing, and one needs to account for all the factor to effectively simulate the interaction patterns and the structural status of all the stakeholders involved in the process which is virtually impossible to this date. De-novo prediction of the siRNA structure will try to obtain a conformation with the lowest energy which may not necessarily be the exact confirmation that will

take place while it interacts with Ago2. Molecular docking will also try out other possible conformations of the siRNA, which will further reduce the chance to have a clearer look at the actual event. Therefore considering aforementioned statements, unlike protein structure prediction, a different approach was taken for the prediction of siRNA structure that had similarity with other research groups. The guide RNA bound to 4z4d structure was taken as template and mutated in Discovery Studio Visualizer (version 20.1) to model the siRNA structures. This approach ensured a better conformational reminiscence and a more appropriate depiction of the actual event.

After modelling, molecular docking of the Ago2 and siRNAs were done with the aid of HDock server (<http://hdock.phys.hust.edu.cn/>) [60,61]. The server uses a hybrid-docking algorithm of template-based modelling and free docking and a hierarchical FFT-based docking program, HDockLite, to execute traditional global docking. The resultant docking complexes were downloaded from the server and inspected manually to identify best docked complex based on docking score, visual similarity of the complex with 4z4d structural composite and placement of siRNA inside PAZ domain and MID domain of the Ago2. Such criteria were enforced to reduce the spurious results and identify the complex that is most similar to experimental evidence [64].

In order to exert the silencing function, the 5' and 3' ends of the siRNA need to be anchored in the MID and PAZ domain of the Argonaute protein respectively [64], although there is a room for flexibility in PAZ domain due to its dynamic nature [65,66]. On the other hand, strong binding of siRNA with Ago2, especially PAZ domain, corresponds to weaker nuclease/RNAi activity [67]. This is expected as 3' end of siRNA needs to dynamically toggle in between bound and unbound states during the nuclease activity therefore a strong binding with PAZ domain may interrupt this process [67].

Based on the placement of siRNA inside the Ago2 protein two distinct clusters were found. Four out of the eight siRNA namely, g15, g59, g46, g70, showed anchoring in both MID and PAZ domain forming the Cluster 1 (Fig. 6). Although rest of the four siRNAs (n7, g21, g22, g44), constituting Cluster 2, protruded outside of the protein near the 5' end, they successfully anchored to the PAZ domain of the protein. Cluster 1 showed a consistent bonding pattern with 11 residues of protein where the bonding category and contributing group matches for all the four-member of the group. Ago2 residues such as ARG179, THR526, THR759 formed hydrogen bond, ARG277, LEU522, TYR529 formed hydrophobic interactions, and ARG277 formed an electrostatic

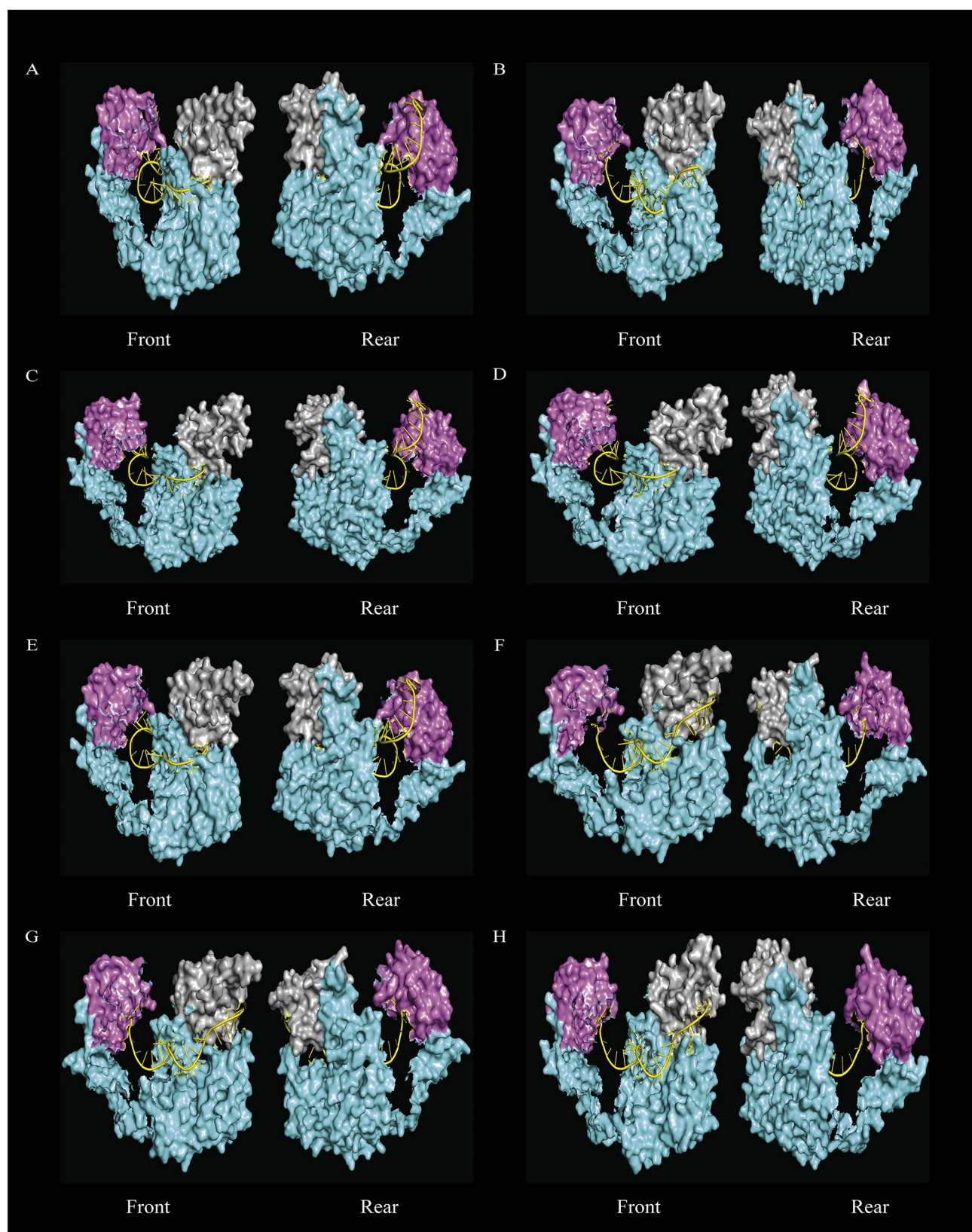


Fig. 7. Docked structures of siRNA (cartoon view) with Ago2 proteins (surface view). The PAZ domain and the MID domain are colored as gray and pink respectively and rest of the protein denoted by aquamarine color (including N-terminal domain and PIWI domain). The siRNA was designated by yellow color. Front and rear view (180° opposite to Front view along x-axis) were provided and some residues of the N-terminal domains were omitted in order to clearly depict the anchoring of siRNA within Ago2. The structures are for A. n7 B. g15 C. g21 D. g22 E. g44 F. g46 G. g59 H. g70 siRNA complexes respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4

Previously reported interactive Argonaute protein residues involved in interactions with siRNAs of the current study.

siRNA	Residues matching the reported residues
n7	ARG351, GLN548, ASN551, ARG635, ARG710, ARG761, ARG792, SER798, TYR804
g15	ALA221, ARG277, ARG315, ARG351, ILE365, THR526, TYR529, LYS533, GLN548, LYS566, ARG635, LYS709, ARG710, GLN757, THR759, ARG761, ARG792, TYR804, ARG812
g21	ARG351, ARG635, ARG710, TYR790, ARG792, CYS793
g22	ARG635, ARG710, ARG761, TYR790, ARG792, CYS793, SER798, TYR804
g44	ARG351, GLN548, ASN551, ARG635, ARG710, ARG761, ARG792, SER798, TYR804
g46	ALA221, THR222, HIS271, ARG277, ARG315, HIS316, ARG351, ILE365, THR526, TYR529, LYS533, GLN545, CYS546, GLN548, ASN551, THR559, ASN562, LYS566, ARG635, LYS709, ARG710, GLN757, THR759, ARG761, ARG792, SER798, TYR804, ARG812
g59	THR222, HIS271, ARG277, ARG315, ARG351, ILE365, GLY524, THR526, TYR529, LYS533, GLN545, CYS546, GLN548, ASN551, THR559, ASN562, LYS566, ARG635, LYS709, ARG710, GLN757, THR759, ARG761, ARG792, TYR804, ARG812
g70	ASP125, THR222, ARG277, ARG315, ARG351, GLY524, THR526, TYR529, LYS533, ASN551, LYS566, ARG635, ARG710, THR759, ARG761, ARG792, TYR804, ARG812

interaction with the bases and LYS566 formed a hydrogen bond and ARG68, ARG315, LYS533, ARG812 formed electrostatic with the phosphate backbone of all the Cluster1 siRNAs. TYR529, LYS533 residues of MID domain interacted with first U/A of the siRNA which is also reported by others [68]. Some residues that interacted with all Cluster 1 siRNAs such ARG277, ARG315, THR526, TYR529, LYS533, LYS566, THR759, ARG812 are previously reported to form interactions with other siRNA as well [68,69]. (See Fig. 7.)

In contrast to Cluster 1, Cluster 2 showed fewer consistent bonding pattern where only two residues of the protein (MID domain) interacted (ARG460 with base and ARG554 with phosphate backbone) with all four siRNAs of the group. This was not surprising, as siRNAs in Cluster2 were not completely encapsulated inside the Ago2 resulting in fewer interactions. ARG635, ARG792 ubiquitously formed electrostatic interaction with phosphate backbone of all the siRNAs of Cluster 1 and Cluster 2. As both ends placed within the Ago2 protein, Cluster 1 siRNAs more likely to perform better than Cluster 2.

Previously reported interactive residues of Ago2 such as ARG635, ARG710, ARG792 interact with all eight siRNAs [68,69]. Other reported residues do not encapsulate any particular cluster e.g. ARG635, ARG710, ARG792 interacts with n7, g15, g21, g22, g44, g46, g59, g70 whereas ARG761, TYR804 interacts with the aforementioned siRNAs except g21. Table 4 contains previously reported [68,69] protein residues involved in siRNA-Ago2 interactions that also participated in the interaction with siRNAs of the current study.

The highest and lowest docking score of Cluster 1 ranged from −317.2 to −515.5 for g15 and g59 respectively. Although Cluster 2 siRNAs did not place effectively in MID domain, the docking scores were in between −328.2 to −356.9 that may confer better efficacy during RNAi. Higher hydrogen bonding is associated with higher RNAi whereas electrostatic interaction shows a moderate correlation with RNAi activity [67]. The total number of hydrogen bonds in the study ranged from 17 to 50 which was, in general, predominant than electrostatic and hydrophobic interactions. Electrostatic and hydrophobic interactions ranged 11–40 and 1–12 respectively.

From the structural perspective, Cluster 1 complexes undoubtedly obtained conformation close to actual event, therefore they are expected to outperform Cluster 2 complexes in RNAi activity. However, from the perspective of docking score the results are inconclusive as g21 & g22 have higher energy than g70 and n7 & g44 have higher docking energy than g46. Therefore, some Cluster 2 siRNAs are expected to show greater RNAi activity than some Cluster 1 siRNAs. Considering the siRNAPred data adds further complexity to the analysis as all of the Cluster 2 siRNAs

were predicted to show better efficacy than g15 which is the best candidate from both structural and docking perspective. Finally, due to the inherent difficulty in effective modelling of complex cellular processes such as RISC, molecular docking of siRNA and Ago2 was used a proxy which, although provided some great insights, did not help to further filter out siRNA candidates. Therefore, all the eight siRNAs are proposed to be effective in SARS-CoV-2 treatment and should be further validated in in-vivo for therapeutic purpose.

Various research groups also proposed similar siRNA treatment strategy for COVID 19 [70,71]. Chen et al. [72], for example, took a different computational approach to identified 9 siRNAs for 5 different sites of SARS-CoV-2 genome (length 21 – 25nts) using single reference genome obtained from NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide/MN908947>) and analyzing single nucleotide polymorphisms in the target sites. Parikesit & Nurdiansyah [73], alternatively, took 24 sequences of a single gene, performed MSA and used the conserved sequences for *in-silico* development of siRNA. Unlike aforementioned research groups, this study considered nucleocapsid phosphoprotein and surface glycoprotein genes of 139 genomes of SARS-CoV-2 which provided scope for a more detailed analysis. Interaction between the siRNA and argonaute 2 protein was also taken into consideration in the current study which was previous studies overlooked. Pharmaceutical companies such as Siranomics, Vir Biotechnology & Alnylam Pharmaceuticals (United States), OilX Pharmaceuticals (South Korea) have also identified various RNAi targets and corresponding siRNA agents of SARS-CoV-2 [70]. We believe our study will be an interesting addition to this landscape.

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 and state of art molecular modelling and docking analysis. 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.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2020.12.021>.

CRedit authorship contribution statement

Umar Faruq Chowdhury: Methodology, Writing - original draft, Visualization. **Mohammad Umer Sharif Shohan:** Conceptualization, Data curation, Methodology, Writing - original draft, Visualization. **Kazi Injamamul Hoque:** Visualization, Writing - original draft. **Mirza Ashikul Beg:** Writing - review & editing. **Mohammad Kawsar Sharif Siam:** Conceptualization, Conceptualization, Writing - original draft, Visualization, Supervision. **Mohammad Ali Moni:** Supervision.

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](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](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.2020020274>.
- [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](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. Godebski, 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), <https://doi.org/10.1101/2020.03.24.006544>, 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 Malaysian 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](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 humoral and cellular immunity 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](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) 1164–1179, <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) 10276–10288, <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) 548–557, <https://doi.org/10.1128/JVI.01782-06>.
- [21] M.A. Tortorici, D. Veeler, Structural insights into coronavirus entry, in: *Adv. Virus Res.*, Academic Press Inc., 2019, pp. 93–116, <https://doi.org/10.1016/b.s.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. Veeler, 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, *Futur. 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. Mim, P. Palit, A therapeutic approach against leishmania donovani by predicting RNAi molecules against the surface protein, gp63, *Curr. Bioinforma.* 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. Zhang, 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-(tosylcarbonyl)hydrazinyl)ethyl)isonicotinamide reveals 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. Siam, 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, *Antivir. Res.* 65 (2005) 45–48, <https://doi.org/10.1016/j.antiviral.2004.09.005>.
- [37] B. Hu, L. Zhong, Y. Weng, L. Peng, Y. Huang, Y. Zhao, X.-J. Liang, Therapeutic siRNA: the state of the art, *Signal Transduct. Target. Ther.* 5 (2020) 101, <https://doi.org/10.1038/s41392-020-0207-x>.
- [38] E.L. Hatcher, S.A. Zhdanov, Y. Bao, O. Blinkova, E.P. Nawrocki, Y. Ostapchuk, 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>.
- [39] 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>.
- [40] 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>.
- [41] 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>.
- [42] 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>.
- [43] 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>.

- [44] 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>.
- [45] 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>.
- [46] 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>.
- [47] 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>.
- [48] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications, *BMC Bioinform.* 10 (2009) 421, <https://doi.org/10.1186/1471-2105-10-421>.
- [49] W.A. Kibbe, OligoCalc: an online oligonucleotide properties calculator, *Nucleic Acids Res.* 35 (2007) W43–W46, <https://doi.org/10.1093/nar/gkm234>.
- [50] 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>.
- [51] 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>.
- [52] D. Piekna-Przybylska, L. DiChiacchio, D.H. Mathews, R.A. Bambara, A sequence similar to tRNA^{3Lys} 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>.
- [53] 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>.
- [54] B. Webb, A. Sali, Comparative protein structure modeling using MODELLER, *Curr. Protoc. Bioinformatics* 2016 (2016) 5.6.1–5.6.37, <https://doi.org/10.1002/cpbi.3>.
- [55] M.A. Martí-Renom, A.C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Šali, Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct.* 29 (2000) 291–325, <https://doi.org/10.1146/annurev.biophys.29.1.291>.
- [56] N.T. Schirle, J. Sheu-Gruttadauria, S.D. Chandradoss, C. Joo, L.J. MacRae, Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets, *Elife* 4 (2015), <https://doi.org/10.7554/eLife.07646>.
- [57] L. Heo, H. Park, C. Seok, GalaxyRefine: protein structure refinement driven by side-chain repacking, *Nucleic Acids Res.* 41 (2013), <https://doi.org/10.1093/nar/gkt458>.
- [58] G.R. Lee, L. Heo, C. Seok, Effective protein model structure refinement by loop modeling and overall relaxation, *Proteins Struct. Funct. Bioinforma.* 84 (2016) 293–301, <https://doi.org/10.1002/prot.24858>.
- [59] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, IUCr, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291, <https://doi.org/10.1107/S0021889892009944>.
- [60] Y. Yan, D. Zhang, P. Zhou, B. Li, S.-Y. Huang, HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy, *Nucleic Acids Res.* 45 (2017) W365–W373, <https://doi.org/10.1093/nar/gkx407>.
- [61] S.-Y. Huang, X. Zou, A knowledge-based scoring function for protein–RNA interactions derived from a statistical mechanics-based iterative method, *Nucleic Acids Res.* 42 (2014) e55, <https://doi.org/10.1093/nar/gku077>.
- [62] M. Kumar, S. Lata, G.R.-P. of the First, U, siRNApred: SVM Based Method for Predicting Efficacy Value of siRNA, *CSIR-IMTECH*, 2009, 2009.
- [63] 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.* 101 (10) (2009) 1–7, <https://doi.org/10.1186/1471-2105-10-s1-s33>.
- [64] M. Müller, F. Fazi, C. Ciaudo, Argonaute proteins: from structure to function in development and pathological cell fate determination, *Front. Cell Dev. Biol.* 7 (2020) 360, <https://doi.org/10.3389/fcell.2019.00360>.
- [65] A. Boland, F. Triteschler, S. Heimstädt, E. Izaurralde, O. Weichenrieder, Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein, *EMBO Rep.* 11 (2010) 522–527, <https://doi.org/10.1038/embor.2010.81>.
- [66] A. Boland, E. Huntzinger, S. Schmidt, E. Izaurralde, O. Weichenrieder, Crystal structure of the MID-PIWI lobe of a eukaryotic argonaute protein, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 10466–10471, <https://doi.org/10.1073/pnas.1103946108>.
- [67] M. Kandeel, Y. Kitade, Computational analysis of siRNA recognition by the Ago2 PAZ domain and identification of the determinants of RNA-induced gene silencing, *PLoS One* 8 (2013), <https://doi.org/10.1371/journal.pone.0057140>.
- [68] E. Elkayam, C.D. Kuhn, A. Tocilj, A.D. Haase, E.M. Greene, G.J. Hannon, L. Joshua-Tor, The structure of human argonaute-2 in complex with miR-20a, *Cell* 150 (2012) 100–110, <https://doi.org/10.1016/j.cell.2012.05.017>.
- [69] V. Bhandare, A. Ramaswamy, Structural dynamics of human argonaute2 and its interaction with siRNAs designed to target mutant tdp43, *Adv. Bioinforma.* 2016 (2016), <https://doi.org/10.1155/2016/8792814>.
- [70] H. Uludağ, K. Parent, H.M. Aliabadi, A. Haddadi, Prospects for RNAi therapy of COVID-19, *Front. Bioeng. Biotechnol.* 8 (2020) 916, <https://doi.org/10.3389/fbioe.2020.00916>.
- [71] S. Habtemariam, I. Berindan-Neagoe, C.A. Cismaru, D. Schaafsma, S.F. Nabavi, S. Ghavami, M. Banach, S.M. Nabavi, Lessons from SARS and MERS remind us of the possible therapeutic effects of implementing a siRNA strategy to target COVID-19: shoot the messenger!, *J. Cell. Mol. Med.* 24 (2020) 10267–10269, <https://doi.org/10.1111/jcmm.15652>.
- [72] W. Chen, P. Feng, K. Liu, M. Wu, H. Lin, Computational identification of small interfering RNA targets in SARS-CoV-2, *Viro. Sin.* 35 (2020) 359–361, <https://doi.org/10.1007/s12250-020-00221-6>.
- [73] A.A. Parikesit, R. Nurdiansyah, The predicted structure for the anti-sense siRNA of the RNA polymerase enzyme (RdRp) gene of the SARS-CoV-2, *Ber. Biol.* 19 (2020) 97–108, <https://doi.org/10.14203/beritabiologi.v19i1.3849>.
- [74] 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>.