Title: SiRNA Molecules as Potential RNAi Therapeutics to Silence RdRP Region and N-Gene of SARS-CoV-2: An In Silico Approach

Authors’ name:
Mahedi Hasan\textsuperscript{a,1}, Atiya Tahira Tasnim\textsuperscript{a,1}, Arafat Islam Ashik\textsuperscript{a,1}, Md Belal Chowdhury\textsuperscript{1}, Zakia Sultana Nishat\textsuperscript{1}, Khandaker Atkia Fariha\textsuperscript{1}, Tanvir Hossain\textsuperscript{*1} & Shamim Ahmed\textsuperscript{*1}

\textsuperscript{a}These authors contributed equally.

Affiliation:
\textsuperscript{1}Department of Biochemistry and Molecular Biology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh.

*Corresponding authors:
Tanvir Hossain
Assistant Professor
Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh
E-mail: tanvir-bmb@sust.edu, tanvirain@gmail.com
Telephone: +880 1674622045

Shamim Ahmed, PhD
Professor
Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh
E-mail: shamil174-bmb@sust.edu, shamil174@gmail.com
Telephone: +880 1776196006
COVID-19 pandemic keeps pressing onward and effective treatment option against it is still far-off. Since the onslaught in 2020, 13 different variants of SARS-CoV-2 have been surfaced including 05 different variants of concern. Success in faster pandemic handling in the future largely depends on reinforcing therapeutics along with vaccines. As a part of RNAi therapeutics, here we developed a computational approach for predicting siRNAs, which are presumed to be intrinsically active against two crucial mRNAs of SARS-CoV-2, the RNA-dependent RNA polymerase (RdRp), and the nucleocapsid phosphoprotein gene (N gene). Sequence conservancy among the alpha, beta, gamma, and delta variants of SARS-CoV-2 was integrated in the analyses that warrants the potential of these siRNAs against multiple variants. We preliminary found 13 RdRP-targeting and 7 N gene-targeting siRNAs using the siDirect V.2.0. These siRNAs were subsequently filtered through different parameters at optimum condition including macromolecular docking studies. As a result, we selected 4 siRNAs against the RdRP and 3 siRNAs against the N-gene as RNAi candidates. Development of these potential siRNA therapeutics can significantly synergize COVID-19 mitigation by lessening the efforts, furthermore, can lay a rudimentary base for the in silico design of RNAi therapeutics for future emergencies.
Introduction

SARS-CoV-2, the cause behind the ongoing pandemic COVID-19, is a coronavirus strain under the virus family *Coronaviridae*. The positive sense single-stranded RNA genome (+ssRNA) of SARS-CoV-2 is almost 30 kb long, and the capsid contains four major types of structural proteins. Just like all other coronaviruses, the 5’-terminus of the SARS-CoV-2 has the ORF1ab gene that encodes for non-structural proteins (nsp1 to nsp16), which are essential for viral replication, viral assembly, viral transcription and immune response modulation. The 3’-terminus of the SARS-CoV-2 has open-reading frames (ORFs) that encode the four structural capsid proteins- spike proteins (S), envelope proteins (E), membrane proteins (M), nucleocapsid proteins (N), and also ORF3a, ORF6, ORF7a, ORF7b, ORF8, ORF10 genes.

The SARS-CoV-2 ORF1ab encoded nsp12 is the RNA-dependent RNA polymerase (RdRP) that have vital functions in SARS-CoV-2 pathogenesis. In coronaviruses, the RdRP is an enzyme that treats (+) ssRNA genome of RNA viruses as a template to produce the (-) RNA strand. This (-) RNA strand then guides viral transcription to produce sub-genomic mRNAs. Beside the transcription process, RdRP is also required for viral genome replication. Hence, RdRP plays an indispensable role in the SARS-CoV-2 life cycle and can be an attractive target for novel antiviral therapeutics. Additionally, among the structural proteins of SARS-CoV-2, the N protein mediates the formation of helical ribonucleoproteins during the packaging of the RNA genome, thus plays a crucial role in viral assembly. It has a critical role in viral RNA replication and transcription by regulating viral RNA synthesis. The N protein can modulate host cell metabolism and can suppress host innate immunity, too. Therefore, the N protein can also be a suitable target for potential antiviral therapeutics.

A viral activity can be controlled by any means of silencing non-structural or structural genes. A gene can be silenced at different stages of infection, such as transcriptional, post-transcriptional or translational level. Among the various mechanisms of gene silencing, RNA interference (RNAi) is one of them. The RNAi is a post-transcriptional gene silencing mechanism that either degrades or inactivates the messenger RNA (mRNA) of a corresponding gene with the help of some interfering RNA molecules, which are more or less 20-30 bp in their entire length. Small interfering RNAs (siRNAs) are comparatively short noncoding double-stranded RNAs (dsRNAs) that are generally exogenous and designed experimentally. Although siRNAs are not
encoded in the genome, but can be introduced in cellular cytoplasm with the help of nanoparticle vectors and various insertion techniques, which results in the degradation of homologous mRNA of the target gene\textsuperscript{15}. It is now well-documented that siRNA can be efficient therapeutics because of its high target-specificity by perfect sequence complementarity with its target mRNA region and compatible insertion techniques into the target cell or organ\textsuperscript{16}.

To summarize, \textit{in vivo} siRNA insertion can either be achieved by classic viral delivery method or modern non-viral delivery methods\textsuperscript{17–19}. Delivery of siRNA using viral vector is mediated by several retrovirus vectors, such as mouse stem cell virus (MSCV), the Moloney murine leukemia virus (MoMLV), and lentivirus. Also, adenovirus vectors are used to deliver siRNA into mammalian system\textsuperscript{18,19}. on the other hand, the non-viral siRNA delivery system mainly comprised with five methods, for instance direct injection, cholesterol conjugation, lipid-based transfection, polymers and peptides delivery and delivery with surface modified nanoparticles\textsuperscript{17}. High pressure induces siRNA penetration during direct injection\textsuperscript{20}, cholesterol and lipid-based chemical agent (i.e., lipoplexes, polyplexes, and lipopolyplexes) mediated transfection are facilitated by lipoproteins used in those techniques\textsuperscript{21,22}. The specificity and selectivity of delivery methods have increased with polymers and peptides mediated delivery as it can escape endosomal degradation\textsuperscript{23}. Besides, surface-modified nanoparticle mediated siRNA delivery provides both protection and specificity\textsuperscript{24}.

The introduced siRNA is then cleaved by RNase III type Dicer protein and becomes short double-stranded siRNA. The short antisense guide strand of the double-stranded siRNA is then recognized by cellular RNA-induced silencing complex (RISC). Whereas the sense passenger strand of siRNA is degraded by cellular enzymatic activity. The siRNA guide strand leads RISC to bind perfectly to the complementary mRNA, and the binding causes the cleaving and degradation of the mRNA by the Argonaute protein of RISC\textsuperscript{25}. Recently, investigation of siRNAs is leading to the way of many antiviral defenses against several diseases, such as human immunodeficiency virus-1 (HIV-1)\textsuperscript{26}, Dengue\textsuperscript{27}, Influenza\textsuperscript{28}, hepatitis C (HCV)\textsuperscript{29}, and Zika\textsuperscript{30}. The siRNAs are also being anticipated to treat various genetic disorders. As a result, the first global approval of a siRNA drug, named Patisiran, was announced by the U.S. Food and Drug Administration (FDA) in 2018\textsuperscript{31}, to treat polyneuropathy in people with hereditary transthyretin-mediated amyloidosis. It is delivered into hepatocytes via lipid nanoparticles where it binds the
3′-UTR of both mutant and wild-type transthyretin (TTR) mRNA and causes its degradation, hence reducing TTR protein levels in serum and tissues\textsuperscript{32}. Most recently in 2020, another siRNA drug has been approved by FDA, called Givosiran\textsuperscript{33}, which is used to treat acute hepatic porphyria (AHP). The specific delivery of Givosiran is mediated by covalent linking of siRNA to a ligand and it targets aminolevulinate synthase 1 (ALAS1) mRNA, thus reduces the accumulation of δ-aminolevulinic acid and porphobilinogen successfully\textsuperscript{34}.

In this study, the RdRP (nsp12) region and N gene of the SARS-CoV-2 viral genome were scanned to construct potential siRNA to silence these genes by utilizing some computational tools (Figure 1). Thus, the predicted siRNAs, if developed will be capable of contributing to the fight against SARS-CoV-2 by degrading mRNA regions of RdRP and N. RNAi is going to be the next generation therapeutics because of its available insertion techniques, specificity, and cost-effectiveness\textsuperscript{35}. Therefore, this study provides a benchmark for the establishment of RNAi therapeutics against SARS-CoV-2 as an effective treatment for COVID-19 as well as for any upcoming future pandemic situation.

(Figure 1 will be inserted here in introduction)

**Result**

**Conserved sequence among different variants**

There were more than 97% similarities for RdRP and N among all available strains of the SARS-CoV-2 around the world. Also, the MEGAX integrated MSA derived all conserved sequences having at least 21 nucleotides are given in the Supplementary table S2 and S3 for the RdRP and N, respectively. The conserved sequence positions were used to find the best siRNAs that will be effective against all reported variants of SARS-CoV-2. Figure 2 represents all selected siRNA targets having 100% sequence conservancy among variants via the sequence logos.

**siRNA prediction with siDirect and minimized off-target**

From the siDirect site, 13 siRNAs for RdRP region and 7 siRNAs for N-gene were found (Supplementary tables S4 and S5) to be potential on the basis of checked parameters. To find the T\textsubscript{m} values for all predicted siRNAs of RdRP and N below 21.5 °C (Supplementary tables S4
and S5), the seed-target duplex was predicted, and all the siRNA duplexes were observed using the nearest-neighbor model and the thermodynamic parameters for the formation of siRNA duplexes. Also, there was no significant sequence similarity with the human genomic and transcript sequence database for the predicted siRNAs. Minimization of off-target sequence ensures more specificity for the siRNA target site complementation.

**Functional significance of GC content and RNA-RNA interaction by thermodynamics**

The resultant GC content analysis using the GC content calculator showed an exact GC content percentage from 33.33% to 42.86% for the predicted siRNAs of both RdRP and N sequences (Supplementary tables S4 and S5), which is a suitable range for hybridization with a specific target\(^{41,54}\). Furthermore, the hybrid RNA structure of the predicted siRNAs and target sequences with their corresponding minimum free energy of hybridization (Supplementary tables S4 and S5) were measured using the RNA structure webserver’s bifold section, which is displayed in Figure 3 for the selected siRNAs only.

**Heat capacity and concentration plots of RNA-duplex**

The higher values of \(T_mC_p\) and \(T_m(Conc)\) indicate higher effectiveness of siRNAs. These melting temperatures for predicted siRNAs of both RdRP and N were measured using the DNAmelt Server, where \(T_mC_p\) and \(T_m(Conc)\) values respectively for siRNAs of RdRP ranged from 79.9 °C to 85.1 °C and 79.0 °C to 83.7 °C (Supplementary table S4), and for siRNAs of N ranged from 80.7 °C to 85.5 °C and 80.1 °C to 84.4 °C (Supplementary table S5).

**Validation of predicted siRNAs**

The validity and effectiveness of the predicted siRNA molecules for both RdRP and N regions were measured by operating siRNAPred, and the siRNAs having validity (binary) values greater than 1 are considered as highly effective. Among 13 siRNA molecules for RdRP, 5 siRNAs were found to have value greater than 1, and among 7 siRNA molecules for N-gene, 1 siRNA had a value greater than 1. Besides, all other predicted siRNAs had validity (binary) values very close to 1 (Supplementary tables S4 and S5).

**Secondary structure prediction of siRNA guide strands**

The positive values of minimum free energy of folding (value > 0) for siRNAs indicate better candidates because those siRNA molecules undergo least folding among themselves. To
visualize the possible folding of the selected siRNA guide strands for both RdRP and N with their corresponding minimum free energy of folding (Figure 4) the RNA structure webserver was used. The values of minimum free energy of folding were in the range of 1.7–2.0 for RdRP siRNAs and of 1.4–1.8 for N siRNAs. The selected siRNAs showed unpaired terminal ends and less degree of folding for both RdRP region and N-gene.

Finally, 5 siRNAs for RdRP and 3 siRNAs for N were selected as the best possible candidates against the SARS-CoV-2 based on different parameters, such as sequence conservation, GC percentages, Tm values, the free energy of hybridization, prediction validity, and free energy of folding of the siRNAs, and the Tm values for these selected siRNAs are given in Table 1.

3D modeling and energy minimization of selected siRNAs

The 3D structures and energy-minimized structures of selected siRNA guide strands were modelled using ModeRNA and YASARA, respectively and visualized in PyMOL molecular graphics visualization software V.2.0 (Figure 5). The starting and ending energy values from YASARA energy minimization are reported in the Table 2.

Macromolecular docking analyses

The HDOCK docking scores were in between -625.91 Kcal/mol to -313.37 Kcal/mol for the siRNAs n1 and r1, respectively. The docking model 1 given by the HDOCK server was the best docking model for all siRNAs, except the r2 siRNA, for which the best model was the HDOCK docking model 2. All of these docking performances showed effective Ago2-mediated mRNA silencing possibilities according to the Ago2-siRNA interaction pattern. Among them, the interaction patterns between Ago2 and r1, r3, n1, n2, n3 siRNAs were better than the remaining according to the interaction pattern. The HDOCK docking scores along with the PAZ and the MID domain interacting residues are given in the Table 2. The interacting residue-pair list is given in the Supplementary table S6. The docking results were finally visualized by PyMOL molecular graphics visualization software V.2.0 (Figure 6).

Discussion
The ongoing global crisis, COVID-19, caused by the infection of SARS-CoV-2. According to World Health Organization (WHO) COVID-19 dashboard (https://covid19.who.int/), more than 500 million human respiratory illnesses, including more than six million deaths caused by this disease till this year. Sustainable and effective treatment regimen against COVID-19 along with proper confinement strategy against its causal agent is still beyond the reach, calling the need for potential therapeutics to fight the pandemic. The present study was carried out using the RdRP and N sequences of the SARS-CoV-2 to design siRNA molecules as potential antivirals against COVID-19. The RdRP is a non-structural protein (nsp12) encoded in the ORF1ab gene, which is required for the replication, transcription, and increasing the copy number of the SARS-CoV-2 genome. Whereas the N-gene encodes nucleocapsid phosphoprotein, which is required for viral RNA genome packaging, assembly, and interaction with host metabolism or immunity. Thus, the RdRP and N sequences are suitable targets for silencing via siRNAs, thereby disrupting SARS-CoV-2 life cycle.

The BLAST result between all available SARS-CoV-2 strains in NCBI database showed more than 97% similarities for RdRP and N sequences. Also, the conservancy analysis between different variants of SARS-CoV-2, such as the alpha variant, beta variant, gamma variant and delta variant showed highly conserved sequence similarities for these aforementioned genes. The RdRP region and N-gene were 99.21% and 97.14% conserved respectively (Supplementary tables S2 and S3) for all stated variants that was confirmed by MEGAX multiple sequence alignment program. Later, the conservancy of screened siRNA targets were visualized in Figure 2 as two bits long sequence logos. A sequence logo is a 2D graph having base position in x-axis and bits in Y-axis. Base frequency is the measurement of how frequent a specific nucleotide is present in each position of an MSA. This base frequency is used to measure the entropy of that position and subsequently the height (Bits - Entropy) of all bases of that corresponding position. The most frequent base in a particular position has least entropy, hence highest height. Thus, bases having height equals to bits are 100% conserved. According to Figure 2, all base position of every siRNA target sequences had complete 2 bits long base heights, which indicates 100% sequence conservancy among all variants of SARS-CoV-2 in these particular positions.

The siRNA molecules are short noncoding exogenous RNA molecules that can effectively silence any specific gene by degrading its corresponding mRNA via perfect complementarity.
between the siRNA guide strand and the target mRNA. The siDirect 2.0 software predicted possible siRNAs for the target sequences and distinguished off-target-free seed sequences properly from the off-target-positive ones because the Tm of 21.5 °C was marked while running the program. A total of 13 siRNAs for the RdRP sequence and 7 siRNAs for the N-gene sequence were found from the siDirect site for further analysis. Although the siDirect 2.0 software can minimize the off-target for designing siRNAs, the NCBI BLASTn similarity search was executed for all predicted siRNAs, and no significant sequence similarities were found with the human genomic plus transcript database.

The siRNAs with high GC% can interfere with siRNA duplex unwinding and effective degradation of mRNA targets. Also, too low GC content can decrease target mRNA recognition and hybridization with the target. Hence, the suitable GC range is an essential parameter for siRNA’s proper functioning. In this present study, an exact GC content percentage from 33.33% to 42.86% was found using the GC content calculator for the predicted siRNAs of both RdRP and N, which is within a suitable range to support the hybridization of those siRNAs with their specific target mRNAs. This range was also maintained by similar previous studies, for instance the study of Chowdhury U.F. et al (2021) and Shohan M.U.S. et al (2018). Also, the hybridization and minimal free energy of binding (Kcal/mol) with target for the predicted siRNAs were -32.7, -33.6, -33.9, -32.9, -31.7, -32.5, -35.2, -33.5, -34.5, -32.1, -35.9, -32.7, -34.3 for RdRP and -34.5, -34.2, -35.8, -34.5, -32.5, -33.1, -31.7 for N, which were found using the bifold section of RNA structure webserver (Supplementary tables S4 and S5) as well as their corresponding hybridization secondary structures for selected siRNAs are illustrated in Figure 3. The more negative free energy of binding value, the more strong binding of the predicted siRNAs with their target sites, and therefore the predicted siRNAs can also suitably bind to their targets.

Moreover, heat capacity plot and concentration plot are crucial benchmarks to predict the siRNA thermodynamics interactions. In the heat capacity plot, the heat capacity (Cp) is plotted as a function of temperature, and when it is a function of melting temperature (Tm) it is called Tm(Cp), which is the local maximum of the heat capacity curve. On the other hand, in the concentration plot, molar fractions are plotted as a function of temperature, and concentration as a function of Tm is Tm(Conc), which is the point where the concentration of the siRNA duplex is one-half of
its maximum value\textsuperscript{47}. The DNAMelt web server was run to find the $T_mC_p$ and $T_m(\text{Conc})$ values, and the higher values lead to the best siRNAs as found in some similar study, for example the study of Nur S.M. \textit{et al} (2015)\textsuperscript{61} and Shohan M.U.S. \textit{et al} (2018)\textsuperscript{59}. In this study, the predicted siRNAs showed high melting profiles, as shown in Table 1, hence regarded as the best candidates as potential antivirals.

Additionally, primarily selected siRNA molecules were validated to measure their efficacy, and siRNAs were selected as potential antiviral candidates for their high validity score (value $\sim$1 or $\geq$ 1) (Table 1)\textsuperscript{58,62}. As crystal structures of RNAs are not available, using computational methods to design RNA secondary structures is much valuable\textsuperscript{63}. The MaxExpect program determined the predicted siRNAs’ secondary molecular structures by computational methods in the RNA structure web server, and the secondary folding structure of selected siRNAs are shown in Figure 4. An unstructured guide strand of siRNA can easily bind with the target mRNA and silence it. Nevertheless, a base-paired terminal of the siRNA guide strand is less available to bind with the complementary mRNA and becomes inactivated\textsuperscript{64}. Hence, if the degree of guide RNA folding structure decreases, the gene silencing by siRNA becomes much more efficient. In these analyses, the selected siRNAs also displayed the availability of nucleotides in the terminal regions and less folding degree (Figure 4). Finally, 5 siRNAs for RdRP region and 3 siRNAs for the N-gene were selected as suitable antiviral candidates for effective post-transcriptional gene silencing of SARS-CoV-2 based on different measured parameters.

In the RNAi pathway, siRNA molecules are intrinsically double-stranded that eventually separate, and only the guide strand interacts with RISC\textsuperscript{65,66}. As the guide strand plays a prominent role in this pathway, it was inevitable to focus on the guide strand’s energy minimization. The energy minimization is one of the crucial steps to measure the compatibility of a guide strand during \textit{in vivo} functions\textsuperscript{67–69}. If the potential energies of the guide strand’s 3D models would much higher, they would not be compatible with the living system. The energy minimization process can resolve improper structures and stereochemistry, can add ions and water molecules, thus subsequently can relax the system and re-position atoms’ arrangement in molecules at its local minima using the principle of Molecular Mechanics (MM)\textsuperscript{70,71}. In this study, the YASARA force field was used for all residues, and RNA atom type description was
included in the parameter file. The lower ending potential of each strand indicates that the structure has minimized (Table 2), and favors the in vivo functioning.

Macromolecular docking between human Ago2 protein and energy minimized modelled siRNAs shows perfect interaction pattern, specially, with the PAZ and MID domain of Ago2 (Table 2). Generally, both ends of an siRNA should be anchored to exert an effective silencing, the PAZ domain interacts in a weaker binding manner with the 3′-end of siRNAs, and the MID domain strongly interacts with the 5′-end of the siRNAs. Because, the 3′-end of an siRNA needs to be switched between a bound and unbound manner during the nuclease activity of Ago2 protein. In this study, among the selected 8 siRNAs, 4 siRNAs namely r1, r3, n2, n3 have shown effective binding with both PAZ and MID domain. Whereas another 3 siRNAs namely r2, r4, and n1 have shown interaction only with the MID domain. But an exceptional case is observed for r5, which only interacts with the PAZ domain of Ago2.

According to the surface residue interaction patterns from this macromolecular docking (given in Supplementary table S6 and visualized in Figure 6) the siRNAs, namely r1, r3, n1, n2 and n3 might be the best siRNAs since they have shared interacting residues with some wet lab interaction studies. All these 5 siRNAs contain stably interacting arginine (ARG) residues, such as the ARG179, ARG635, ARG710, ARG761, ARG792, and 4 of them have ARG812 interacting residue (except n1) too. Although these residues are PAZ and MID domain interface residues, they are verified by the studies mentioned above. Except the r1 siRNA, the other stated 4 siRNAs have TYR804 interaction with Ago2, which was also found in the aforementioned studies. Another 2 siRNAs r2 and r4 also have shown ARG792 interaction but the interaction pattern for r5 is, however, poor in manner and can be negligible.

Among the PAZ domain interacting residues, ARG277, TYR279 are the most common and for the MID domain interacting residues GLN548 interaction is the most common, which is observed in r1, r2, r3, r4, n1, n2, n3. Some other MID domain interactions include GLN548, GLN558, THR559, ASN562, LEU563, LYS566 that have been observed for 6 aforementioned siRNAs except the n1 and the LEU522, GLY524, LYS525, TYR529, CYS546, VAL547 interactions are found in other 6 aforementioned siRNAs except the r4. All of these interaction are common and verified by other studies stated above. As shown in the Supplementary table S6, the r5 siRNA has an unusual interaction pattern and becomes insignificant. But the
interaction patterns for the other 7 siRNAs cannot be neglected, hence it can be said that they hold promising potential to effectively silence their corresponding target RdRP and N mRNA sequences of SARS-CoV-2.

The HDOCK docking scores for siRNAs targeting RdRP region range from -609.29 Kcal/mol to -313.37 Kcal/mol that correspondence to r3 and r1, respectively (Table 2). Also, the HDOCK docking scores for siRNAs targeting N-gene range from -625.91 Kcal/mol to -336.96 Kcal/mol that respectively corresponding to n1 and n3 (Table 2). These scores are much favorable for all of these selected siRNAs and indicate their strong interaction with the human Ago2 protein. Although the interaction patterns are similar to previously mentioned verified studies\textsuperscript{73,76,77}, it leaves further confirmation about the effectiveness of these siRNAs via further \textit{in vitro} and \textit{in vivo} studies.

From our previous study, both miRNAs and siRNAs were predicted against the SARS-CoV-2 ORF1ab gene\textsuperscript{78}. Another recent study\textsuperscript{58} had shown an almost similar strategy targeting the N and S protein gene of SARS-CoV-2. As this current study targets the most crucial protein of SARS-CoV-2, the RdRP, to identify and design best siRNAs against it, as well as identified 3 potential N-gene siRNAs, this study can also add some valuable inputs in the race of applying therapeutics against SARS-CoV-2. Also, inclusion of most recent delta variant genome and other variants of SARS-CoV-2 makes these predicted siRNAs relatively more effective against SARS-CoV-2 in comparison with other previous studies.

Currently, evidence has suggested that \textit{in silico} assessment might be a novel and feasible approach for viral therapeutics development. In this study, 07 siRNA molecules were predicted as the credible candidates against RNA-dependent RNA polymerase and nucleocapsid phosphoprotein gene of SARS-CoV-2 using all maximum parameters in optimum conditions including their binding interaction analysis with argonaute protein. Additionally, these systematic analyses will provide a guideline for the future establishment of a computational model to anticipate siRNA antiviral therapeutics. Moreover, the compatible siRNA insertion techniques (via lipid-based transfection, nanoparticle, etc.) into the target cells or organs\textsuperscript{79}, RISC-mediated target specific mRNA degradation, and cost-effective manner\textsuperscript{35} warrants the siRNA therapeutics an effective treatment for the ongoing pandemic, COVID-19. Besides, the workflow of this study will guide the \textit{in silico} design of RNAi therapeutics and also will inspire
researchers for developing RNAi therapeutics to control any upcoming pandemic situation caused by infectious agents faster.

**Materials and methods**

**Sequence retrieval and analysis**

The reference sequence of SARS-CoV-2 from the National Center for Biotechnology Information (NCBI), GB: NC_045512.2, was used to obtain the target sequences “RdRP or nsp12” and “N”. The NCBI graphics database was used to retrieve the nucleotide sequence of 13442-16236 bp for RdRP and of 28274-29533 for N ([https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?report=graph](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?report=graph)). After that, the nucleotide sequences were utilized to perform the BLAST (Basic Local Alignment Search Tool) ([https://blast.ncbi.nlm.nih.gov/Blast.cgi](https://blast.ncbi.nlm.nih.gov/Blast.cgi)) to achieve their sequence similarities among all available strains of the SARS-CoV-2 around the world in GenBank, the National Institutes of Health (NIH) genetic sequence database.

**Sequence conservancy analysis among variants**

Nearly 50 recent completely sequenced SARS-CoV-2 genome from all 6 geographical continents around the globe for the early pangolin lineages, A and B, was derived from NCBI Virus SARS-CoV-2 resources ([https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#!/sars-cov-2](https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#!/sars-cov-2)) and subsequently the RdRP and N sequences were retrieved from them. Meanwhile, each geographic continent that had at least one complete sequence of SARS-CoV-2 variants, such as the alpha variant, the beta variant, the gamma variant, and the delta variant sequences having pangolin lineages B.1.1.7, B.1.351, P.1, and B.1.617.2, respectively were retrieved by using the reported website for the RdRP and N sequences of these variants. Later, evolutionary analysis by maximum likelihood method were performed by using the MEGAX software integrated multiple sequence alignment (MSA) program for 60 different sequences. Finally, the conservancy was visualized as sequence logos for selected RdRP and N siRNA target sequences that was constructed by the WebLogo webserver.

**siRNA prediction and off-target minimization**
To predict potential siRNAs for target nucleotide sequences, RdRP and N, an online web server siDirect version 2.0\textsuperscript{40} was operated. Several parameters, including melting temperature below 21.5 °C, GC content 31.6%-57.9%\textsuperscript{41,42} and Ui-Tei\textsuperscript{43,44}, Renold\textsuperscript{45}, Amarzguioui\textsuperscript{41} combined rules (algorithms have given in Supplementary table S1) were checked in the siDirect site. For the formation of a stable RNA duplex the seed-target duplex stability (T\textsubscript{m}) was also calculated. Consequently, nucleotide BLAST (BLASTn) program was executed for siRNA target sequences using the human genomic plus transcript database by default parameters to check off-target sequences in the genome of any non-targeted sites for more specific siRNA selection.

GC content calculation and measurement of thermodynamic interaction

GC Content Calculator (http://www.endmemo.com/bio/gc.php) was operated to calculate the exact GC content percentages in predicted siRNAs. Afterwards, the thermodynamics interaction between the predicted viral siRNAs and the target sequences was measured by the RNA structure web server’s bifold section (http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/bifold/bifold.html).

Calculation of the heat capacity and concentration plot

The melting temperature T\textsubscript{m}C\textsubscript{p}, heat capacity curve’s local maximum, and T\textsubscript{m}(Conc), one-half concentration of maximum value of a double-stranded molecule were measured by operating the hybridization of two different strands option (http://unafold.rna.albany.edu/?q=DINAMelt/Hybrid2) of the DNAmelt Server\textsuperscript{46,47} with default values and in simple form checking the RNA option for the predicted siRNAs.

Validation of predicted siRNA molecules

The predicted siRNA molecules were screened using the Main21 dataset with the support vector machine algorithm and by the binary pattern prediction approach to measure the actual efficacy of the siRNA molecules and their validation via operating an online free accessible server siRNApred (http://crdd.osdd.net/raghava/sirnapred/).

Secondary structure prediction

The MaxExpect\textsuperscript{48} program in the RNA structure web server\textsuperscript{49} was operated to anticipate the secondary structure of siRNA guide strands with their respective free energy of folding.
Modeling the 3D structures of predicted siRNAs and Energy Minimization

For validating predicted siRNAs’ function in vivo, it is indispensable to construct the 3D structure of siRNAs and minimize energy. Construction of a three-coordinated arrangement of guide strand was done using template-based sequence alignment followed by exporting the 3D model in PDB file format via an advanced online tool for modeling RNA 3D structures, the ModeRNA server\textsuperscript{50}. For this purpose, the guide RNA strand bound to the human Argonaute2 (Ago2) protein (PDB id: 4W5Q) was used as a template. In searching for the lowest potential conformation (local minima), the energy was further minimized using YASARA Energy Minimization Web Server\textsuperscript{51}. YASARA Energy Minimization Web Server uses YASARA force field. The YASARA force field includes bonded parameters (bond length, bond angle, and dihedrals) and non-bonded parameters (electrostatic interactions and Lennard-Jones potentials) for corresponding atom types of RNA molecules\textsuperscript{52}.

Macromolecular docking of Ago2 and predicted siRNAs

The macromolecules for conducting an effective RNAi silencing, the human Ago2 protein and predicted siRNAs, were docked by using the HDOCK server (http://hdock.phys.hust.edu.cn/) to investigate their interaction pattern with each other. HDOCK is an online server that perform any macromolecular docking based on a hybrid algorithm of template-based modeling and ab-initio free docking system\textsuperscript{53}. The receptor protein Ago2 was derived from RCSB Protein Data Bank (https://www.rcsb.org/) (PDB id: 4W5Q) and refined to extract the Ago2 protein from the PDB cluster. Whereas the ligands were YASARA energy-minimized siRNAs in PDB format. Based on the best docking scores and the protein-siRNA interaction pattern, the best docking performances were evaluated.

Data availability statement

We have used publicly available free databases as the main data source that are mentioned in method section with proper referencing. The supplementary data file contains all produced data during analyses.

References


69. Thompson, E., Thackray, T., Kalthoff, C., Rapoport, R. & Jagodzinski, F. Using Energy-Minimization Profiles to Measure Protein Resistance to Drugs. in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health*


Acknowledgements

The authors would like to acknowledge the Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet for laboratory facilities and logistic supports.

Author contributions


Additional Information

The authors declare no competing interests among them.
Figure Legends

**Figure 1: Graphical representation.** Schematic diagram of siRNA prediction and optimization from the gene sequence.

**Figure 2: Conservancy analysis of selected siRNA targets for RdRP region and N gene.** The RdRP and N sequences retrieved for MEGAX multiple sequence alignment are analyzed for conserved sequence of those two genes. The target sequences of screened siRNAs displayed 100% conservancy that has visualized using two bits long sequence logos constructed with WebLogo. The serial number below the logo indicates base position in RdRP and N separately. All bases of target sequences represent height of two bits long, hence producing 100% conservancy of corresponding bases in corresponding positions. rN, r = RdRP region targeting siRNAs, N = serial number of siRNAs; nN, n = N-gene targeting siRNAs, N = serial number of siRNAs.

**Figure 3: Secondary structures of potential target–siRNA duplex.** The secondary structures of RdRP and N targeting siRNAs hybridized with their corresponding targets with their minimum free energy of hybridization (Kcal/mol). rN, r = RdRP region targeting siRNAs, N = serial number of siRNAs; nN, n = N-gene targeting siRNAs, N = serial number of siRNAs.

**Figure 4: Secondary structures of siRNA guide strands.** The values of minimum free energy of guide strands’ folding with visualization for both RdRP and N targeting siRNAs are illustrated and retrieved from RNA structure webserver. Color coding indicates the probability of predicted secondary structure positioning of that base. The probabilities with their corresponding colors are shown in inset. rN, r = RdRP region targeting siRNAs, N = serial number of siRNAs; nN, n = N-gene targeting siRNAs, N = serial number of siRNAs.

**Figure 5: Three-dimensional (3D) structure visualization of YASARA energy minimized siRNAs.** 3D secondary structure of siRNAs at starting energy (yellow color) in comparison with their energy-minimized versions (blue color), displayed using PyMol for both RdRP and N targeting siRNAs. rN, r = RdRP region targeting siRNAs, N = serial number of siRNAs; nN, n = N-gene targeting siRNAs, N = serial number of siRNAs.

**Figure 6: Visualization of argonaute protein and screened siRNAs binding interaction.** Docking illustrations of siRNA (yellow cartoon view) with human Ago2 protein (cyan surface...
view), where the PAZ domain and the MID domain are colored as deep teal and greenish cyan, respectively. There siRNA-Ago2 docking interaction is visualized for potential siRNAs targeting the RdRP and N sequences of SARS-CoV-2. rN, r = RdRP region targeting siRNAs, N = serial number of siRNAs; nN, n = N-gene targeting siRNAs, N = serial number of siRNAs.
### Tables

Table 1. Effective siRNA molecules for RdRP and N regions with $T_m$ values of both guide strand and passenger strand, guide strands’ GC%, free energy of binding (MEF) with target, $T_m$Cp, $T_m$(Conc), and validity (binary). rN, N = serial number of RdRP targeting siRNAs; nN, N = serial number of N targeting siRNAs.

<table>
<thead>
<tr>
<th>Alias</th>
<th>Location of target within mRNA</th>
<th>siRNA target within mRNA</th>
<th>Predicted siRNA duplex candidates at 37°C (21nt guide (5′→3′): 21nt passenger (5′→3′))</th>
<th>Seed-duplex stability ($T_m$) °C</th>
<th>GC% (Kcal/mol)</th>
<th>$T_m$Cp °C</th>
<th>$T_m$(Conc) °C</th>
<th>Validity (binary)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>228-250</td>
<td>TTCTCTAACTACCAACATGAAGA</td>
<td>CUCUAACUACCAACAUAGAAGA</td>
<td>20.5°C</td>
<td>18.9°C</td>
<td>33.33</td>
<td>-32.7</td>
<td>81.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UUCAUGUAGUAGUAGAGAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>585-607</td>
<td>ATGCAGAGATTGCTGTTTGGTGG</td>
<td>GCGAAAUGCUGGUAGUUGUGG</td>
<td>14.6°C</td>
<td>14.1°C</td>
<td>38.1</td>
<td>-33.6</td>
<td>84.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AACAAUAACAGCAUUUGCAU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GCGAAUGCUGGUAGUUGUGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>1267-1289</td>
<td>TCTGTCTCTGTTTCTTTAAG</td>
<td>UAAAGAAACCCUCUAGACAGG</td>
<td>5.5°C</td>
<td>20.3°C</td>
<td>38.1</td>
<td>-33.9</td>
<td>83.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GUGUCUAAGGUUCCUUUAGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>2414-2436</td>
<td>TACTAACAGCTCATGAATTTT</td>
<td>AAUCAAGGAGUAGUUAGUAGA</td>
<td>13.6°C</td>
<td>19.9°C</td>
<td>33.33</td>
<td>-32.1</td>
<td>82.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CUAAGGACCUCUAGAAUUUU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r5</td>
<td>2622-2644</td>
<td>TAGAATGTATGATGCTTTCA</td>
<td>AAAGACAGACAGCAUCACUGG</td>
<td>19.2°C</td>
<td>20.3°C</td>
<td>42.86</td>
<td>-35.9</td>
<td>85.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGAGUAGCUAGUAGCUUUCUUCUCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n1</td>
<td>314-336</td>
<td>GTTAAAGATGTATTATTACTAC</td>
<td>AGUAGAAAUCACACUUUGGAC</td>
<td>14.6°C</td>
<td>18.1°C</td>
<td>38.1</td>
<td>-34.5</td>
<td>80.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCAAGAUAGUUUCUUACUAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n2</td>
<td>727-749</td>
<td>GGCCAAGCTGTACTAAGAAAATC</td>
<td>UUUCUAGUGACAGUUUGGCC</td>
<td>11.7°C</td>
<td>16.7°C</td>
<td>42.86</td>
<td>-35.8</td>
<td>83.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCAAACGUACUAGAAAUC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n3</td>
<td>892-914</td>
<td>TACAAACATTTGCCGCAAATTGC</td>
<td>AAVUGGGGCAUUUGUUGGA</td>
<td>20.6°C</td>
<td>5.3°C</td>
<td>38.1</td>
<td>-33.1</td>
<td>85.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAAAGCAUGUAGUGUUGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

License: [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).
Table 2. YASARA energy minimization values for screened siRNAs and the human Ago2 protein PAZ and MID domains interacting residues after docking with energy minimized siRNA molecules including their docking scores. rN, N = serial number of RdRP targeting siRNAs; nN, N = serial number of N targeting siRNAs.

<table>
<thead>
<tr>
<th>siRNAs</th>
<th>Energy minimization values</th>
<th>Docking score and interacting residues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PAZ Domain</td>
</tr>
<tr>
<td></td>
<td>Starting energy (kJ/mol)</td>
<td>Ending energy (kJ/mol)</td>
</tr>
<tr>
<td>r1</td>
<td>2004174.3</td>
<td>-67729.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>265228.7</td>
<td>-57795.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>-44569.7</td>
<td>-75266.0</td>
</tr>
<tr>
<td>r4</td>
<td>166805899.4</td>
<td>-63556.9</td>
</tr>
<tr>
<td>r5</td>
<td>-32458.0</td>
<td>-66374.0</td>
</tr>
<tr>
<td>n1</td>
<td>-35609.6</td>
<td>-70222.5</td>
</tr>
<tr>
<td>n2</td>
<td>286466.1</td>
<td>-62300.3</td>
</tr>
</tbody>
</table>

The copyright holder for this preprint available under a CC-BY-NC-ND 4.0 International license.
| n3 | 756245.5 | -67784.7 | -336.96 | ARG277, TYR279, ARG315, GLN332, LEU522, GLY524, LYS525, THR526, TYR529, LYS533, GLN545, CYS546, VAL547, GLN548, ASN551, GLN558, THR559, ASN562, LEU563, LYS566, LYS570 |
Figure 1:
Figure 4:
Figure 5: