Computational study and design of effective siRNAs to silence structural proteins associated genes of Indian SARS-CoV-2 strains

Premnath Madanagopal[1], Harshini Muthukumar[1], Kothai Thiruvengadam[1, *]

[1]Department of Biotechnology, Anna University, Chennai, India

[*]Corresponding author: Kothai Thiruvengadam, Department of Biotechnology, Anna University, Chennai, India, Email: tkothai@annauniv.edu

Abstract:

SARS-CoV-2 is a highly transmissible and pathogenic coronavirus that first emerged in late 2019 and has since triggered a pandemic of acute respiratory disease named COVID-19 which poses a significant threat to all public health institutions in the absence of specific antiviral treatment. Since the outbreak began in March 2020, India has reported 4.77 lakh Coronavirus deaths, according to the World Health Organization (WHO). The innate RNA interference (RNAi) pathway, on the other hand, allows for the development of nucleic acid-based antiviral drugs in which complementary small interfering RNAs (siRNAs) mediate the post-transcriptional gene silencing (PTGS) of target mRNA. Therefore, in this current study, the potential of RNAi was harnessed to construct siRNA molecules that target the consensus regions of specific structural proteins associated genes of SARS-CoV-2, such as the envelope protein gene (E), membrane protein gene (M), nucleocapsid phosphoprotein gene (N), and surface glycoprotein gene (S) which are important for the viral pathogenesis. Conserved sequences of 811 SARS-CoV-2 strains from around India were collected to design 21 nucleotides long siRNA duplex based on various computational algorithms and parameters targeting E, M, N and S genes. The proposed siRNA molecules possessed sufficient nucleotide-based and other features for effective gene silencing and BLAST results revealed that siRNAs' targets have no significant matches across the whole human genome and hence, siRNAs were found to have no off-target effects on the genome, ruling out the possibility of off-target silencing. Finally, out of 157 computationally identified siRNAs, only 4 effective siRNA molecules were selected for each target gene which is proposed to exert the best action based on GC content, free energy of folding, free energy of binding, melting temperature, heat capacity and molecular docking analysis with Human AGO2 protein. Our engineered siRNA candidates could be used as a genome-level therapeutic treatment against various sequenced SARS-CoV-2 strains in India. However, future applications will necessitate additional validations in vitro and in vivo animal models.

Keywords: COVID-19, siRNA, RNA interference, Gene silencing, Docking

Abbreviations: COVID-19: COrona VIrus Disease 2019, SARS-CoV-2: Severe Acute Respiratory Syndrome-Corona Virus-2, siRNA: Small interfering RNA, PTGS: Post-Transcriptional Gene Silencing, RNAi: RNA interference, Tm: Melting temperature.

1. Introduction:

COVID-19, A viral disease, caused by the new strain of severe acute respiratory syndrome coronavirus (SARS-CoV), known as SARS-CoV-2, has challenged humanity at the beginning of 2020, impacting the lives of billions of people worldwide. Since its outbreak in late December 2019 in Wuhan, China, after a sudden epidemic of atypical pneumonia with unclear illness aetiology, it has caused substantial morbidity and mortality all over the world[1]. Fever, cough, fatigue, dyspnea, and headache[2] are the most common signs of this disease, but it can also be asymptomatic[3]. Currently, molecular techniques based on the real-time reverse transcriptase-polymerase chain reaction (RT-PCR) are considered the gold standard for COVID-19 diagnosis. According to the World Health Organization (WHO), India has recorded 4.77 lakh Coronavirus deaths since the outbreak began in March 2020. Furthermore, 3.47 crore Coronavirus cases were reported in India[4]. Especially the second wave of COVID-19 has resulted in a rise in cases, a decline in crucial treatment supplies, and an increase in deaths, particularly among the young[5]. In severe cases, the patient may die due to massive alveolar damage and progressive respiratory failure[2]. Also, several occurrences of mucormycosis, popularly known as the black fungus, have been reported in patients with diabetes and patients with COVID-19, as well as patients who were recovering from infection[6].

The SARS-CoV-2 is enveloped by single-stranded positive-sense RNA and has 50% and 80% homology with Middle East Respiratory Syndrome virus and SARS-CoV, respectively and it comprises four structural proteins: envelope (E), membrane (M), nucleocapsid (N) and spike (S)[7,8]. The S, M, and E proteins are membrane-bound, while the N protein is located within the virions in complex with the genomic RNA[9]. Coronaviruses are divided into four primary genera based on their genetic makeup: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus[10]. The first two genera primarily infect mammals, whereas the latter two primarily infect birds. Coronaviruses have genomes that are between 26 and 32 kb in size and contain 6 to 11 open reading frames (ORFs)[11]. The transmembrane trimetric spike glycoprotein extends from the viral surface and is made up of two functional subunits (S1 and S2). The S1 subunit aids in the interaction of SARS-CoV-2 with the host cell's angiotensin-converting enzyme 2 (ACE2) receptor, whereas the S2 subunit aids in the fusion of viral and host cell membranes[12]. The M glycoprotein is the most common structural protein in coronaviruses; it spans the membrane bilayer, leaving a short NH2-terminal domain outside the virus and a long COOH terminus (cytoplasmic domain) inside[13]. All other structural proteins can bind to the M protein. Binding with M protein aids in the stabilisation of N proteins and enhances viral assembly completion by stabilising the N protein-RNA complex within the internal virion[14]. Mutations in the M protein, which cooperate with the S protein, may influence virus attachment and entrance into the host cell[15]. The E protein is the smallest and most cryptic of the major structural proteins. It is extensively expressed inside the infected cell during the replication cycle, but only a small proportion of it is incorporated into the virion envelope[16]. In the coronavirus life cycle, the RNA-binding N protein plays two important roles. Its principal function is to form the viral RNA–protein (vRNP) complex with genomic RNA and to mediate vRNP packaging into virions through poorly understood interactions between the N and M proteins. Second, the N protein is hypothesised to recruit host factors and promote RNA template switching in RTCs during the early stages of infection, facilitating viral RNA synthesis and

translation[17]. Since these structural proteins (E, M, N and S) are crucial for the virus survival, these could be the main targets for designing therapeutics.

RNA interference or RNAi, the biological mechanism by which double-stranded RNA (dsRNA) induces gene silencing by targeting complementary mRNA for degradation (Figure 1)[18], is a huge breakthrough in disease therapy and is changing the way scientists analyse transcriptional activity[19] and it is a prospective tool for the control of human viral infections. Small interfering RNA (siRNA) is a double-stranded RNA of around 21 base pairs long RNA duplex bearing overhangs of two nucleotides on the 3′ end[20]. The strand of siRNA which has complementarity with the target gene is the guide strand and the other strand is the passenger strand. Combinations of chemically synthesised siRNA duplexes targeting SARS-CoV genomic RNA result in therapeutic activity of up to 80% inhibition, according to studies[21]. Due to the complementarity of seven nucleotides in the seed region of siRNA with the off-target gene, there is a risk of off-target gene silence or unexpected gene downregulation. The off-target binding impact is mediated by the melting temperature or thermodynamic stability of the duplex of seed siRNA sequence (2-8 nucleotides of siRNA guide strand from 5′ end) and the target gene, according to studies. Thus, a siRNA with perfect complementarity solely for the target gene and low seed-target duplex thermostability (Tm less than 21.5°C) can efficiently remove siRNA off-target binding. Furthermore, selecting a siRNA with at least two mismatches with any other off-target region can lower the likelihood of siRNA binding to the unwanted off-target sequence[22,23]. Along with the seed region, the non-seed region of the guide strand has been shown to be effective in mediating off-target impact, yet there was a negative correlation between Tm value and GC content and off-target effect[24].

Because structural proteins expressed by the E, M, N and S genes are involved in virus survival and infectivity, they can be used as a target for suppressing SARS-CoV-2 infection. *In silico* approaches to discovering innovative therapeutic approaches are becoming increasingly popular, and viruses are no exception. Therefore, in this present study, we have designed siRNAs specific to the various conserved region of the envelope protein, membrane glycoprotein, nucleocapsid and surface glycoprotein genes by analyzing 811 Indian SARS-CoV-2 strains and finally shortlisted 4 effective siRNA molecules for each gene which will inhibit the translations of these target genes and allow the host to discard this infection (Figure 2).

2. Materials and Methods:

2.1. Retrieval of gene sequences

812 gene sequences of the envelope gene(E), membrane gene(M), nucleocapsid phosphoprotein gene(N) and surface glycoprotein gene(S) of various Indian SARS-CoV-2 strains were retrieved from the GenBank available at the National Center for Biotechnology Information (NCBI) (Supplementary Table 1)[25]. Out of 812 gene sequences, 811 were Indian strains and the remaining one was the Wuhan-Hu-1 strain (China, Genbank accession: NC_045512.2) that was used as a reference genome sequence for multiple sequence alignment and site numbering for amino acids.

2.2. Multiple sequence alignment

Multiple sequence alignment of all 812 gene sequences of E, M, N and S genes were performed using the ClustalW[26] algorithm in the MEGA-X[27] program to find the conserved regions.

2.3. Target recognition and potential siRNA designing

siDirect version 2.0[23] is used to design effective and target-specific siRNA molecules against SARS-CoV-2 E, M, N and S gene sequences. This tool utilizes some rules such as Ui-Tei[28], Amarzguioui[29] and Reynolds[30] algorithms (Table 1) for designing siRNAs and the melting temperature (Tm) of the seed-target duplex was kept below 21.5°C as a default parameter. To avoid off-target silencing, it chooses siRNA sequences with at least two mismatches to any other non-targeted transcripts.

2.4. Off-target investigation using BLAST

The BLAST[31] tool was used to identify the possible off-target matches in a human genomic transcript against the whole GenBank[32] database by applying the expected threshold value 10 and BLOSUM 62 matrix[33] as a parameter.

2.5. GC content calculation

OligoCalc[34] was used to analyze the GC content of predicted siRNA molecules.

2.6. Secondary structure prediction

The secondary structure and the free energy (ΔG) of folding for predicted siRNA molecules were computed using the MaxExpect[35] program of the RNA structure web server[36]. Higher energy values imply that those molecules are less likely to fold and it indicates that they are better candidates.

2.7. Calculation of RNA–RNA Interaction Through Thermodynamics

The thermodynamic interaction between the target strand and the siRNA guide strand was predicted using the DuplexFold[37] program of the RNA structure web server[36]. It folds two sequences of RNA into their lowest hybrid free energy conformation. Higher interaction between the target and the guide strand will aid in a better predictor of siRNA effectiveness.

2.8. Calculation of heat capacity and concentration plot duplex

The heat capacity plot and concentration plot were calculated for the predicted siRNAs using the DINA Melt web server[38]. The ensemble heat capacity (Cp) is plotted as a function of temperature, with the melting temperature Tm (Cp) indicated in Table 2. The detailed heat capacity plot also shows the contributions of each species to the ensemble heat capacity, whereas the concentration plot- Tm (Conc), the point at which the concentration of double-stranded molecules of one-half of its maximum value defines the melting temperature Tm (Conc).

2.9. RNA modelling and Protein preparation

The proper interaction between siRNA duplex and RISC complex, as well as guide siRNA strand and target mRNA inside RISC complex, are required to trigger a sufficient antiviral response via RNAi-mediated viral gene silencing[18]. Hence for interaction pattern analysis, molecular docking was carried over between siRNA (guide strand) and argonaute protein.

The 3D modelled structure of guide siRNA was generated using a 3dRNA v2.0 web server[39]. 3dRNA is an automatic, fast, and high-accuracy RNA tertiary structure prediction method. It uses sequence and secondary structure information to build three-dimensional structures of RNA from template segments. We used an optimization procedure which is an integrated 3dRNA pipeline, to model our predicted siRNAs.

The three-dimensional structure of human argonaute 2 (Ago2) protein (Figure 3a) was retrieved from RCSB Protein Data Bank (PDB ID: 4F3T)[40]. BIOVIA discovery studio[41] was used to prepare the proteins through correction of bonds, removal of unrelated chemical complexes, elimination of water molecules and hetatom groups. The missing side-chain atoms in protein have been reconstructed and then the structure was minimized with the partial implementation of the GROMOS96 force-Field using Swiss-PdbViewer[42]. The quality of the model was checked using Ramachandran Plot generated by PROCHECK[43].

2.10. Molecular Docking and Interaction analysis

The molecular docking between siRNA (guide strand) and argonaute protein was performed using the HDOCK server[44]. This server uses a hierarchical Fast Fourier Transform (FFT) - related docking program. It is a hybrid docking algorithm that uses free docking and template-based modelling and shows high efficacy and robustness. Further docking analysis of protein-siRNA complex was performed using BIOVIA discovery studio[41].

3. Results

3.1. Target specific designing of potential siRNA molecules

siDirect 2.0 web server predicted 7 siRNA for the envelope, 10 siRNA for membrane, 3 siRNA for nucleocapsid phosphoprotein and 137 siRNA for surface glycoprotein (Supplementary Table 6, 7, 8 and 9) that followed algorithms of Ui-Tei, Amarzguioui and Reynolds (Table 1). The seed target duplex stability (Tm) values for all predicted siRNAs were less than 21.5 °C, which shows that predicted siRNAs may avoid nontarget binding.

3.2. Analysis of target and off-target exclusion using blast

The consensus targets obtained by siDirect 2.0 were subjected to NCBI-BLAST for searching similarity against the whole human genome, however, no significant matches were found. This shows that our predicted siRNA may not interact in any place other than the viral target site.

3.3. GC content calculation of predicted siRNA

The sequence's GC content is a vital parameter that influences siRNA functionality. It has been suggested to select sequences with a low to medium proportion (31.6 to 57.9%) of GC content[45]. GC content analysis of the predicted siRNAs was ranged from 21% to 38% for the envelope, 10% to 38% for membrane, 36% to 40% for nucleocapsid phosphoprotein and 10% to 43% for surface glycoprotein (Supplementary Tables 10, 11, 12 and 13).

3.4. Secondary structure determination

The probable folding and minimum free energy of folding of the predicted siRNA guide strands were calculated with the assumption that siRNA with the positive free energy of folding may be more permissible to the target, resulting in better gene silencing[46]. The

calculated free energy of folding ranged from 1.3 to 1.7 for the envelope, from 1.6 to 2 for membrane, from 1.7 to 1.9 for nucleocapsid phosphoprotein and from 1.1 to 2.9 for surface glycoprotein (Supplementary Tables 10, 11, 12 and 13).

3.5. Thermodynamics of target-guide strand interaction

The binding free energy between the target and guide strand was calculated. The values spanned from -25.8 to -34.1 for envelope, -20.9 to -33.5 for membrane, -32.5 to -35.8 for nucleocapsid phosphoprotein and -20.5 to -36.1 for surface glycoprotein (Supplementary Tables 10, 11, 12 and 13).

3.6. Determination of heat capacity and duplex concentration plot

The predicted siRNAs' Tm(Cp) and Tm(Conc) were computed. The greater the values of these two melting temperatures, the more effective the siRNA species. Tm(Conc) values ranged from 76.4 ◦C to 85 ◦C for envelope, 66.8 ◦C to 87.5 ◦C for membrane, 86.2 ◦C to 88.9 ◦C for nucleocapsid phosphoprotein and 63.4 ◦C to 91.1 ◦C for surface glycoprotein. Tm(Cp) values ranged from 77.6 ◦C to 86 ◦C for envelope, 68.4 ◦C to 88.7 ◦C for membrane, 87.6 ◦C to 89.4 ◦C for nucleocapsid phosphoprotein and from 69.0 ◦C to 92.1 ◦C for surface glycoprotein (Supplementary Tables 10, 11, 12 and 13).

3.7. Analysis of molecular modelling and docking of best siRNA (guide strand) & Ago2

According to the Ramachandran plot of Ago2 protein, the percentage of residues that resided in the core region, allowed region, moderately allowed region, and disallowed region was 90.7%, 8.9%, 0.0%, and 0.4% respectively (Figure 3b). Best siRNAs have been filtered for each gene based on the criteria as follows GC content (33% to 43%) and Free energy of binding (>=-31.3) (Table 2). The docking score, interaction statistics and interactive residues were provided in Table 3. The docking score ranged from -290.15 to -358.89 where g82 had the lowest docking score and g80 had the highest binding affinity.

4. Discussion:

COVID-19, respiratory illness caused by the SARS-CoV-2 coronavirus, has recently become pandemic. The pandemic is sweeping through India at a pace that has staggered scientists. To combat this pandemic, promising varied techniques such as gene therapy, as well as other therapeutic tactics such as the creation of vaccines, drugs, monoclonal antibodies, peptides, and other therapeutic strategies are being suggested. The present study introduced an *in silico* approach to design potential siRNA molecules of Indian SARS-CoV-2 strains targeting virus structural genes which included envelope protein (E), membrane protein (M), nucleocapsid phosphoprotein (N) and surface glycoprotein (S).

Among the 811 SARS-CoV-2 strains from throughout India, 243 conserved regions (10 envelope protein, 26 membrane protein, 67 nucleocapsid phosphoprotein and 140 surface glycoprotein) (Supplementary Tables 2, 3, 4 and 5) were found. Conserved regions that are less than 21 nucleotides were taken out of the further analysis. The siDirect webserver was used to identify potential targets and generate corresponding siRNAs using conserved sequences. The task is completed by siDirect in three steps: highly functional siRNA selection, seed-dependent off-target effects reduction, and near-perfect matched genes removal. There were 157 target regions, 7 envelope proteins, 10 membrane proteins, 3 nucleocapsid phosphoproteins, and 137 surface glycoproteins (Supplementary Tables 6, 7, 8

and 9). To improve the results, the U, R, A (Ui-Tei, Amarzguioui, and Reynolds) guidelines (Table 1) were used to identify the siRNAs. The formation of non-target siRNA bonds was eliminated by lowering the melting temperature (Tm) below 21.5 degrees Celsius.

Despite the fact that the primary goal of siRNA design is to silence specific targets, there is still the possibility of silencing an unknown number of unwanted genes[47]. The off-target activity caused by siRNA can be explained by two mechanisms: mRNA depletion or translational suppression at the protein level[48]. First, siRNA can tolerate several mismatches to the mRNA target while also retaining good silencing capacity with imperfect complementarity[47]. Another possible mechanism for off-target activity is that siRNAs' physical structure, ~21 nucleotides (nt) RNA oligomers, appears to be identical to the related class of microRNAs. MicroRNAs are short, endogenously transcribed RNAs that prevent mRNA from being translated rather than being degraded. MicroRNAs appear to contain RNA oligo and RNA target mismatches built into their structure. Together the mechanisms of siRNA mismatch tolerance and microRNA translation inhibition create a risk of off-target activity when ~21 nt RNAs are introduced into human cells[47,48]. As a result, off-target effects may occur even when siRNAs are generated using different algorithms. To ensure siRNA target specificity, siRNA design processes are often followed by a BLAST search for cross-reactive 21-bp siRNA sequences. BLAST results show that the SARS-CoV-2 target sequences (E, M, N, and S) had no significant hits against the entire genome, ruling out the possibility of off-target silencing.

The GC-content of siRNA influences its functioning, and there is an inverse link between GC-content and siRNA function. siRNA targets with a very high GC content may be more prone to folding, which could weaken target accessibility. For a siRNA to be effective, it needs to have a medium proportion of GC content, which ranges from 31.6 to 57.9%[45]. Therefore, molecules with a GC content of less than ~32% were excluded from the final selection. For each of the 157 predicted species, the GC content ranged from 10% to 43%. The GC content of the filtered siRNAs is greater than or equivalent to 33%. (Table 2).

The RNA structure webserver was used to estimate probable folding structures and corresponding minimal free energy using guide strands of predicted siRNAs. At 37°C, the free energy of folding of the final siRNAs is more than zero (Figure 4, Table 2), implying that the predicted siRNAs are more accessible for efficient binding.

DuplexFold was used to calculate the binding energy of the target mRNA and guide siRNA interactions. Lower binding energy suggests a better interaction, and hence a better likelihood of inhibiting the target. All 157 predicted siRNAs had binding free energy values ranging from -36.1 to -20.5. (Supplementary Tables 10, 11, 12 and 13). We chose the best siRNAs for each E, M, S, and N gene based on GC content and free energy of binding criteria for further docking analysis. The selected siRNAs have a free energy of binding equal to or less than -31.3 (Figure 5, Table 2), implying that predicted siRNAs are more interactive with their targets.

The three-dimensional structure of Human Ago2 protein (PDB ID: 4F3T) that we considered for the docking studies had 90.7% residues in the core region of the Ramachandran plot thereby qualifying as a good quality structure (Figure 3b). On the other hand, we modelled siRNA using the 3dRNA webserver which builds RNA 3D structures from sequences by using the smallest secondary elements (SSEs)[39].

The Ago2 and modelled siRNAs were docked with the help of the HDOCK server. To perform traditional global docking, the server employs a hybrid-docking technique that combines template-based modelling and free docking, as well as follows a hierarchical FFT-based docking algorithm. Resultant Ago2-siRNAs docked complexes were retrieved from the server and manually analysed to determine the optimal docked complex based on docking score, interaction pattern analysis, and placement of siRNA within the Ago2 domains (PAZ, MID, N-terminal and PIWI domain) (Figure 3a).

Based on the interaction pattern analysis of 12 filtered siRNAs guide strand with target mRNA (Table 3), 4 best final siRNAs (e5, m2, n1 and g80) were sorted and hypothesized for each gene that could mediate the post-transcriptional gene silencing of the targeted gene (Table 4). The docking score of final siRNAs (e5, m2, n1 and g80) showed impressive binding with argonaute protein within the docked complex (Figure 6), ranging from -322.34 to -358.89, where g80 had the highest binding affinity facilitated by the formation of 4 salt bridges, 8 H-bonds, 10 electrostatic bonds and 2 hydrophobic bonds involving ARG255, LYS263, ARG280, ARG286, GLY331, GLU333, ARG351, LYS354, ALA369, ARG710, HIS712, ARG714, ILE756, ARG761, TYR790, ARG792 and TYR804 residues. The detailed elucidation about interacting residues is provided in Supplementary Table 14.

Therapeutic applications of siRNAs are quite challenging as they have several issues, such as siRNA instability, limited cellular uptake, and the lack of a competent delivery method[49]. However, an appropriate promoter-controlled vector can aid in the targeting of therapeutic genes to the targeted cell for efficient gene therapy[50]. Vector-based siRNA in plasmid form can also be used to target desired genes within a given cell culture to examine the potentiality of a newly created siRNA[51]. Furthermore, plasmids containing siRNA can be directly delivered into the organs of choice[52].

In this present study, four potential siRNA candidates were hypothesized to be effective at binding and cleaving specific SARS-CoV-2 (Indian strains) mRNA targets of structural proteins (Table 4). The suggested therapeutic molecule may be used to treat the COVID-19 pandemic in India on a broad scale because since the study comprises a large array of 811 SARS-CoV-2 sequences from all over India, however, it needs further validation in *in vitro* and *in vivo* studies.

Conclusions:

siRNA therapy might be a promising tool of the RNAi pathway for controlling viral infections in humans by PTGS of the significant gene in various biological systems. In this study, four siRNA molecules were predicted to be effective against the envelope gene (E), membrane gene (M), nucleocapsid phosphoprotein gene (N) and surface glycoprotein gene (S) of 811 Indian strains of the SARS-CoV-2 virus using various computational tools considering all maximum parameters in prominent conditions molecular modelling and docking analysis. However, future applications will necessitate additional validations *in vitro* and *in vivo* animal models. In the battle against the Covid-19 pandemic, these synthetic molecules may be used as novel antiviral therapy and provide a basis for the researchers and pharmaceutical industry to create antiviral therapeutics at the genome level.

CRediT authorship contribution statement

Premnath Madanagopal: Conceptualization, Methodology, Data Curation, Visualization, Writing - Original Draft, Writing - Review & Editing. **Harshini Muthukumar:** Visualization, Data Curation. **Kothai Thiruvengadam:** Writing - Review & Editing, Supervision

Data availability

All necessary data generated or analyzed during this study are included in this article. Any additional data could be available from the corresponding author upon request.

Acknowledgement

K.T. thank BIC at DoBT, AU (BT/PR40163/BTIS/137/31/2021), Department of Biotechnology, Government of India for computational facilities and manpower support.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References:

[1] S. Muthusamy, H. Krishnasamy Naidu Gopal, T. Manivarma, S. Pradhan, P. Pradhan, Prabhu, P. Prabhu, Virtual Screening Reveals Potential Anti-Parasitic Drugs Inhibiting the Receptor Binding Domain of SARS-CoV-2 Spike protein, (2021). https://doi.org/10.13140/RG.2.2.10250.18883.

[2] 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/ATTACHMENT/D5332CA1-83D8-4C4C-BC57-00A390BF0396/MMC1.PDF.

[3] F. Song, N. Shi, F. Shan, Z. Zhang, J. Shen, H. Lu, Y. Ling, Y. Jiang, Y. Shi, Emerging 2019 novel coronavirus (2019-NCoV) pneumonia, Radiology. 295 (2020) 210–217. https://doi.org/10.1148/RADIOL.2020200274/ASSET/IMAGES/LARGE/RADIOL.2020200274.FIG4D.JPEG.

[4] Coronavirus Pandemic (COVID-19) – the data - Statistics and Research - Our World in Data, (n.d.). https://ourworldindata.org/coronavirus-data?country=~IND (accessed December 18, 2021).

[5] P. Asrani, M.S. Eapen, M.I. Hassan, S.S. Sohal, Implications of the second wave of COVID-19 in India, Lancet Respir. Med. 9 (2021) e93–e94. https://doi.org/10.1016/S2213-2600(21)00312-X.

[6] P. Madanagopal, N. Ramprabhu, R. Jagadeesan, In silico prediction and structure-based multitargeted molecular docking analysis of selected bioactive compounds against mucormycosis, Bull. Natl. Res. Cent. 2022 461. 46 (2022) 1–21. https://doi.org/10.1186/S42269-022-00704-4.

[7] A.K. Pandey, S. Verma, C. Anand, K. Pandey, An in silico analysis of effective siRNAs against COVID-19 by targeting the leader sequence of SARS-CoV-2, Adv Cell Gene Ther. 00 (2021) 107. https://doi.org/10.1002/acg2.107.

[8] V. Tiwari, J.C. Beer, N.V. Sankaranarayanan, M. Swanson-Mungerson, U.R. Desai, Discovering small-molecule therapeutics against SARS-CoV-2, Drug Discov. Today. 25 (2020) 1535. https://doi.org/10.1016/J.DRUDIS.2020.06.017.

[9] V. Tiwari, J.C. Beer, N.V. Sankaranarayanan, M. Swanson-Mungerson, U.R. Desai, Discovering small-molecule therapeutics against SARS-CoV-2, Drug Discov. Today. 25 (2020) 1535–1544. https://doi.org/10.1016/J.DRUDIS.2020.06.017.

[10] F. Li, Structure, Function, and Evolution of Coronavirus Spike Proteins, Http://Dx.Doi.Org/10.1146/Annurev-Virology-110615-042301. 3 (2016) 237–261. https://doi.org/10.1146/ANNUREV-VIROLOGY-110615-042301.

[11] 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.

[12] H. Harapan, N. Itoh, A. Yufika, W. Winardi, S. Keam, H. Te, D. Megawati, Z. Hayati, A.L. Wagner, M. Mudatsir, Coronavirus disease 2019 (COVID-19): A literature review, J. Infect. Public Health. 13 (2020) 667–673. https://doi.org/10.1016/J.JIPH.2020.03.019.

[13] L. Mousavizadeh, S. Ghasemi, Genotype and phenotype of COVID-19: Their roles in pathogenesis, J. Microbiol. Immunol. Infect. 54 (2021). https://doi.org/10.1016/J.JMII.2020.03.022.

[14] I. Astuti, Ysrafil, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response, Diabetes Metab. Syndr. 14 (2020) 407–412. https://doi.org/10.1016/J.DSX.2020.04.020.

[15] M. Bianchi, D. Benvenuto, M. Giovanetti, S. Angeletti, M. Ciccozzi, S. Pascarella, Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Characteristics?, Biomed Res. Int. 2020 (2020). https://doi.org/10.1155/2020/4389089.

[16] P. Venkatagopalan, S.M. Daskalova, L.A. Lopez, K.A. Dolezal, B.G. Hogue, Coronavirus envelope (E) protein remains at the site of assembly, Virology. 478 (2015) 75–85. https://doi.org/10.1016/J.VIROL.2015.02.005.

[17] S. Lu, Q. Ye, D. Singh, Y. Cao, J.K. Diedrich, J.R. Yates, E. Villa, D.W. Cleveland, K.D. Corbett, The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein, Nat. Commun. 2021 121. 12 (2021) 1–15. https://doi.org/10.1038/s41467-020-20768-y.

[18] S. MMAK, S. AR, B. M, M. B, A. F, S. MS, H. MM, B. S, L. SS, H. MA, C. C, Designing an effective therapeutic siRNA to silence RdRp gene of SARS-CoV-2, Infect. Genet. Evol. 93 (2021). https://doi.org/10.1016/J.MEEGID.2021.104951.

[19] H. Dana, G.M. Chalbatani, H. Mahmoodzadeh, R. Karimloo, O. Rezaiean, A. Moradzadeh, N. Mehmandoost, F. Moazzen, A. Mazraeh, V. Marmari, M. Ebrahimi, M.M. Rashno, S.J. Abadi, E. Gharagouzlo, Molecular Mechanisms and Biological Functions of siRNA, Int. J. Biomed. Sci. 13 (2017) 48. /labs/pmc/articles/PMC5542916/ (accessed December 26, 2021).

[20] C.Y. Chu, T.M. Rana, Small RNAs: regulators and guardians of the genome, J. Cell. Physiol. 213 (2007) 412–419. https://doi.org/10.1002/JCP.21230.

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

[22] K. Ui-Tei, Optimal choice of functional and off-target effect-reduced siRNAs for RNAi therapeutics, Front. Genet. 4 (2013) 107. https://doi.org/10.3389/FGENE.2013.00107/BIBTEX.

[23] 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 Bioinforma. 2009 101. 10 (2009) 1–8. https://doi.org/10.1186/1471-2105-10-392.

[24] P.J. Kamola, Y. Nakano, T. Takahashi, P.A. Wilson, K. Ui-Tei, The siRNA Non-seed Region and Its Target Sequences Are Auxiliary Determinants of Off-Target Effects, PLOS Comput. Biol. 11 (2015) e1004656. https://doi.org/10.1371/JOURNAL.PCBI.1004656.

[25] Home - Gene - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/gene (accessed September 20, 2021).

[26] 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. https://doi.org/10.1093/NAR/22.22.4673.

[27] K. S, S. G, L. M, K. C, T. K, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, Mol. Biol. Evol. 35 (2018) 1547–1549. https://doi.org/10.1093/MOLBEV/MSY096.

[28] U.-T. K, N. Y, T. F, H. T, O.-H. H, J. A, U. R, S. K, 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.

[29] A. M, P. H, 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.

[30] R. A, L. D, B. Q, S. S, M. WS, K. A, Rational siRNA design for RNA interference, Nat. Biotechnol. 22 (2004) 326–330. https://doi.org/10.1038/NBT936.

[31] A. SF, G. W, M. W, M. EW, L. DJ, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

[32] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, D.L. Wheeler, GenBank, Nucleic Acids Res. 35 (2007) D21–D25. https://doi.org/10.1093/NAR/GKL986.

[33] S. Henikoff, J.G. Henikoff, Amino acid substitution matrices from protein blocks., Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 10915. https://doi.org/10.1073/PNAS.89.22.10915.

[34] W.A. Kibbe, OligoCalc: an online oligonucleotide properties calculator, Nucleic Acids Res. 35 (2007) W43. https://doi.org/10.1093/NAR/GKM234.

[35] L. ZJ, G. JW, M. DH, Improved RNA secondary structure prediction by maximizing expected pair accuracy, RNA. 15 (2009) 1805–1813. https://doi.org/10.1261/RNA.1643609.

[36] 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. https://doi.org/10.1093/NAR/GKT290.

[37] Welcome to the DuplexFold Web Server, (n.d.). https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/DuplexFold/DuplexFold.html (accessed September 20, 2021).

[38] M. NR, Z. M, DINAMelt web server for nucleic acid melting prediction, Nucleic Acids Res. 33 (2005). https://doi.org/10.1093/NAR/GKI591.

[39] J. Wang, J. Wang, Y. Huang, Y. Xiao, 3dRNA v2.0: An Updated Web Server for RNA 3D Structure Prediction, Int. J. Mol. Sci. 20 (2019). https://doi.org/10.3390/IJMS20174116.

[40] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235. https://doi.org/10.1093/NAR/28.1.235.

[41] D.S. BIOVIA, Discovery Studio, (2021).

[42] G. N, P. MC, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis. 18 (1997) 2714–2723. https://doi.org/10.1002/ELPS.1150181505.

[43] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291. https://doi.org/10.1107/S0021889892009944.

[44] Y. Y, T. H, H. J, H. SY, The HDOCK server for integrated protein-protein docking, Nat. Protoc. 15 (2020) 1829–1852. https://doi.org/10.1038/S41596-020-0312-X.

[45] 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 Bioinformatics. 10 (2009) 1–7. https://doi.org/10.1186/1471-2105-10-S1-S33/FIGURES/1.

[46] S. Singh, S.K. Gupta, A. Nischal, S. Khattri, R. Nath, K.K. Pant, P.K. Seth, Design of potential siRNA molecules for hepatitis delta virus gene silencing, Bioinformation. 8 (2012) 749. https://doi.org/10.6026/97320630008749.

[47] G. Ozcan, B. Ozpolat, R.L. Coleman, A.K. Sood, G. Lopez-Berestein, Preclinical and clinical development of siRNA-based therapeutics, Adv. Drug Deliv. Rev. 87 (2015) 108–119. https://doi.org/10.1016/J.ADDR.2015.01.007.

[48] O. Snøve, T. Holen, Many commonly used siRNAs risk off-target activity, Biochem. Biophys. Res. Commun. 319 (2004) 256–263. https://doi.org/10.1016/J.BBRC.2004.04.175.

[49] K. Tanaka, T. Kanazawa, T. Ogawa, Y. Takashima, T. Fukuda, H. Okada, Disulfide crosslinked stearoyl carrier peptides containing arginine and histidine enhance siRNA uptake and gene silencing, Int. J. Pharm. 398 (2010) 219–224.

https://doi.org/10.1016/J.IJPHARM.2010.07.038.

[50] J.C. Glorioso, N.A. Deluca, D.J. Fink, DEVELOPMENT AND APPLICATION OF HERPES SIMPLEX VIRUS VECTORS FOR HUMAN GENE THERAPY, Https://Doi.Org/10.1146/Annurev.Mi.49.100195.003331. 49 (2003) 675–710. https://doi.org/10.1146/ANNUREV.MI.49.100195.003331.

[51] M. ElHefnawi, T.K. Kim, M.A. Kamar, S. Min, N.M. Hassan, E. El-Ahwany, H. Kim, S. Zada, M. Amer, M.P. Windisch, In Silico Design and Experimental Validation of siRNAs Targeting Conserved Regions of Multiple Hepatitis C Virus Genotypes, PLoS One. 11 (2016) e0159211. https://doi.org/10.1371/JOURNAL.PONE.0159211.

[52] H. Giladi, M. Ketzinel-Gilad, L. Rivkin, Y. Felig, O. Nussbaum, E. Galun, Small interfering RNA Inhibits Hepatitis B virus replication in mice, Mol. Ther. 8 (2003) 769–776. https://doi.org/10.1016/S1525-0016(03)00244-2.

Figure legends

Figure 1: Graphical representation of the siRNA-mediated gene silencing mechanism.

Figure 2. Flowchart depicting the workflow of the methodology used in the study.

Figure 3. A. The cartoon representation of the structure of Human Argonaute 2 (Ago2) protein. B. The Ramachandran plot of Human Ago2 protein.

Figure 4. The possible folding and minimum free energy of the guide strands of the predicted siRNA molecules. The structures are for A. e5 B. m1 C. m2 D. n1 E. n2 F. g36 G. g82 H. g83 I. g105 J. g134 K. g80 L. g109 siRNAs.

Figure 5. Structure of binding of target RNA and siRNA (guide strand) with corresponding predicted minimum free energy. The structures are for A. e5 B. m1 C. m2 D. n1 E. n2 F. g36 G. g82 H. g83 I. g105 J. g134 K. g80 L. g109 siRNAs.

Figure 6. Docked structures of final siRNA candidates (cartoon view) with Human Ago2 protein (surface view). The PAZ domain and the MID domain are coloured as magenta and cyan respectively and rest of the protein denoted by orange colour (including N-terminal domain and PIWI domain). The siRNA was designated by yellow colour. The structures are for A. e5 B. m2 C. n1 D. n1 E. g80 siRNAs.

Figure 1. Graphical representation of the siRNA-mediated gene silencing mechanism.

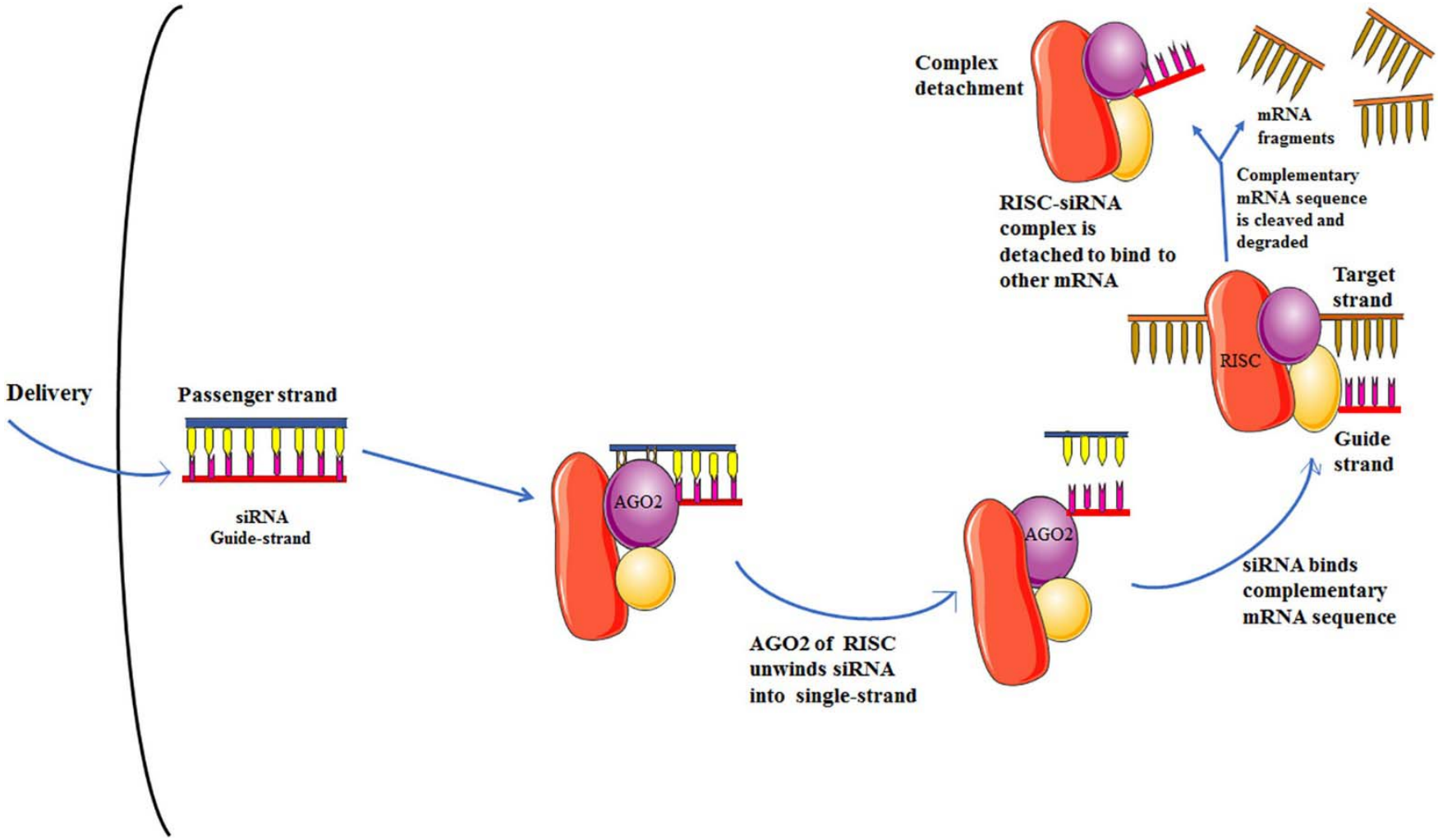

Figure 2. Flowchart depicting the workflow of the methodology used in the study.

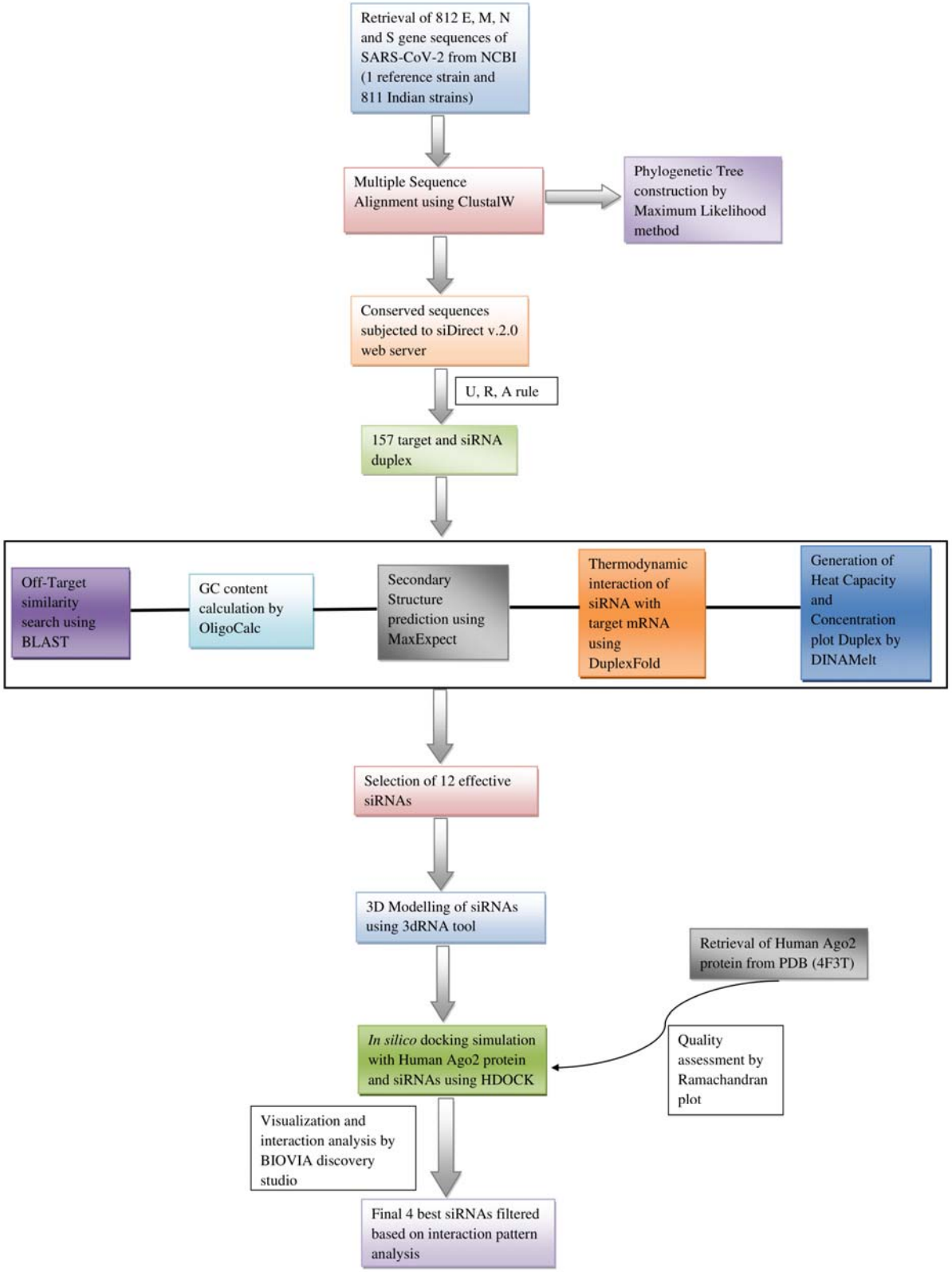

Figure 3. A) The cartoon representation of the structure of Human Argonaute 2 (Ago2) protein. B) The Ramachandran plot of Human Ago2 protein.

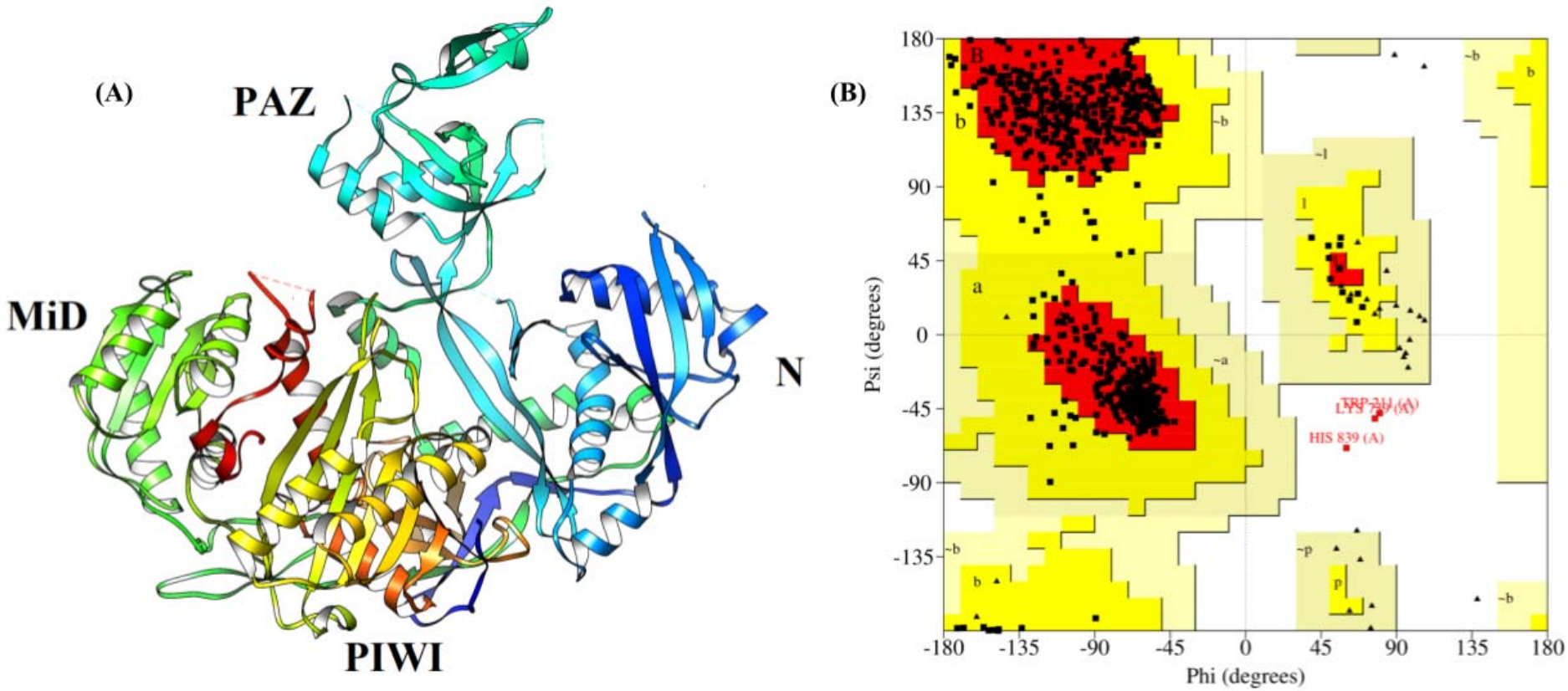

Figure 4. The possible folding and minimum free energy of the guide strands of the predicted siRNA molecules. The structures are for A. e5 B. m1 C. m2 D. n1 E. n2 F. g36 G. g82 H. g83 I. g105 J. g134 K. g80 L. g109 siRNAs.

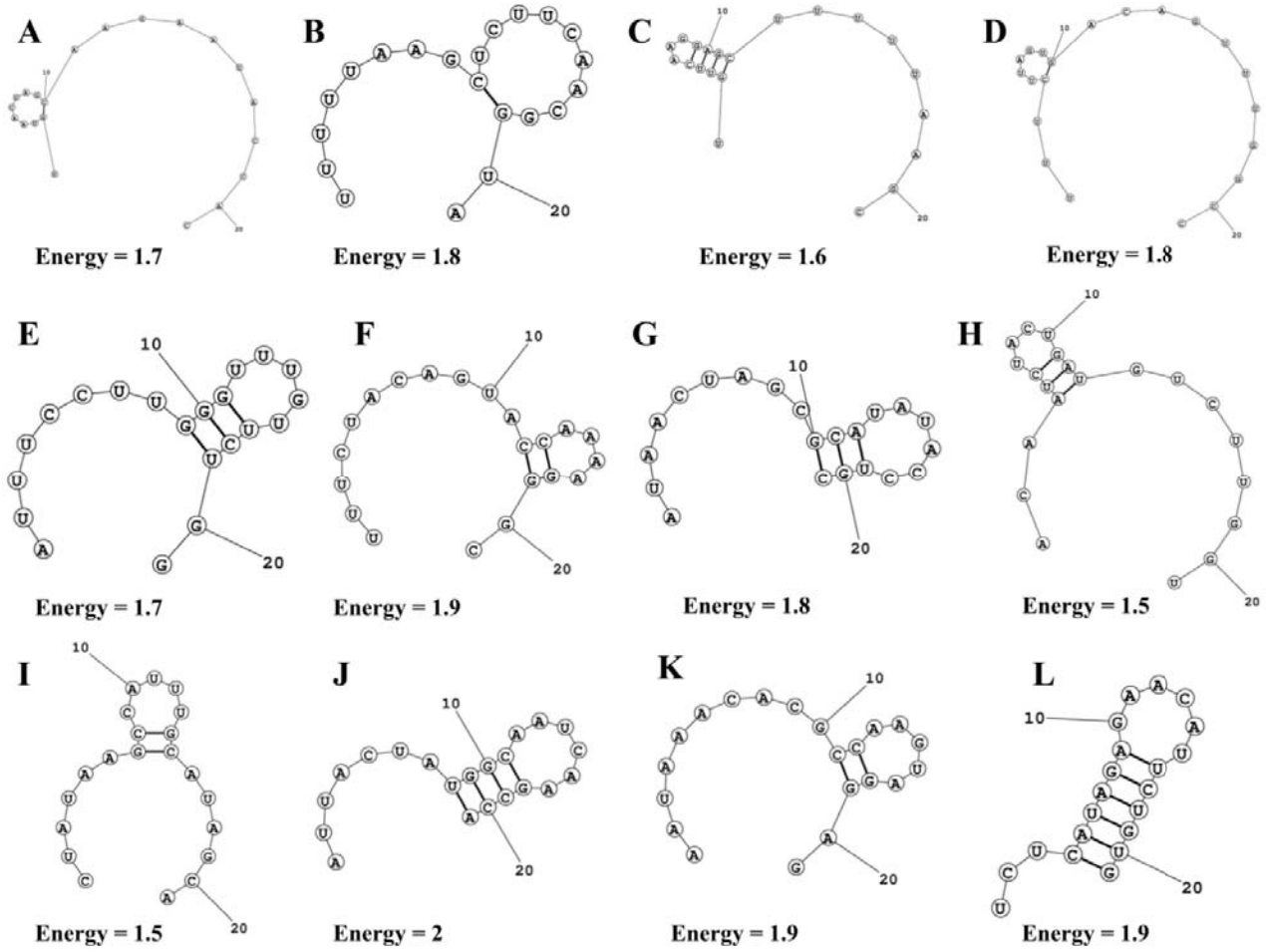

Figure 5. Structure of binding of target RNA and siRNA (guide strand) with corresponding predicted minimum free energy. The structures are for A. e5 B. m1 C. m2 D. n1 E. n2 F. g36 G. g82 H. g83 I. g105 J. g134 K. g80 L. g109 siRNAs.

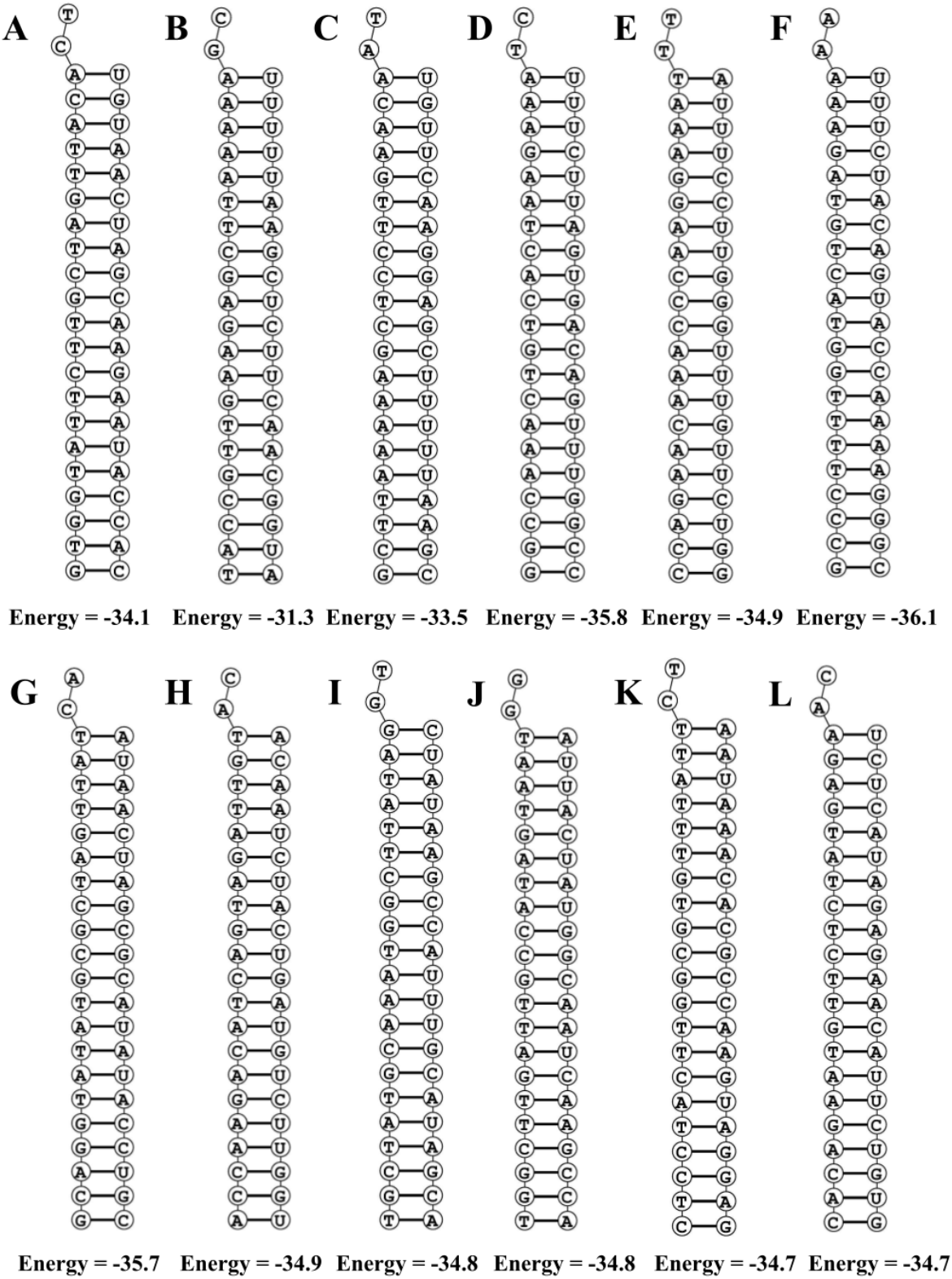

Figure 6. Docked structures of final siRNA candidates (cartoon view) with Human Ago2 protein (surface view). The PAZ domain and the MID domain are coloured as magenta and cyan respectively and rest of the protein denoted by orange colour (including N-terminal domain and PIWI domain). The siRNA was designated by yellow colour. The structures are for A. e5 B. m2 C. n1 D. g80 siRNAs.

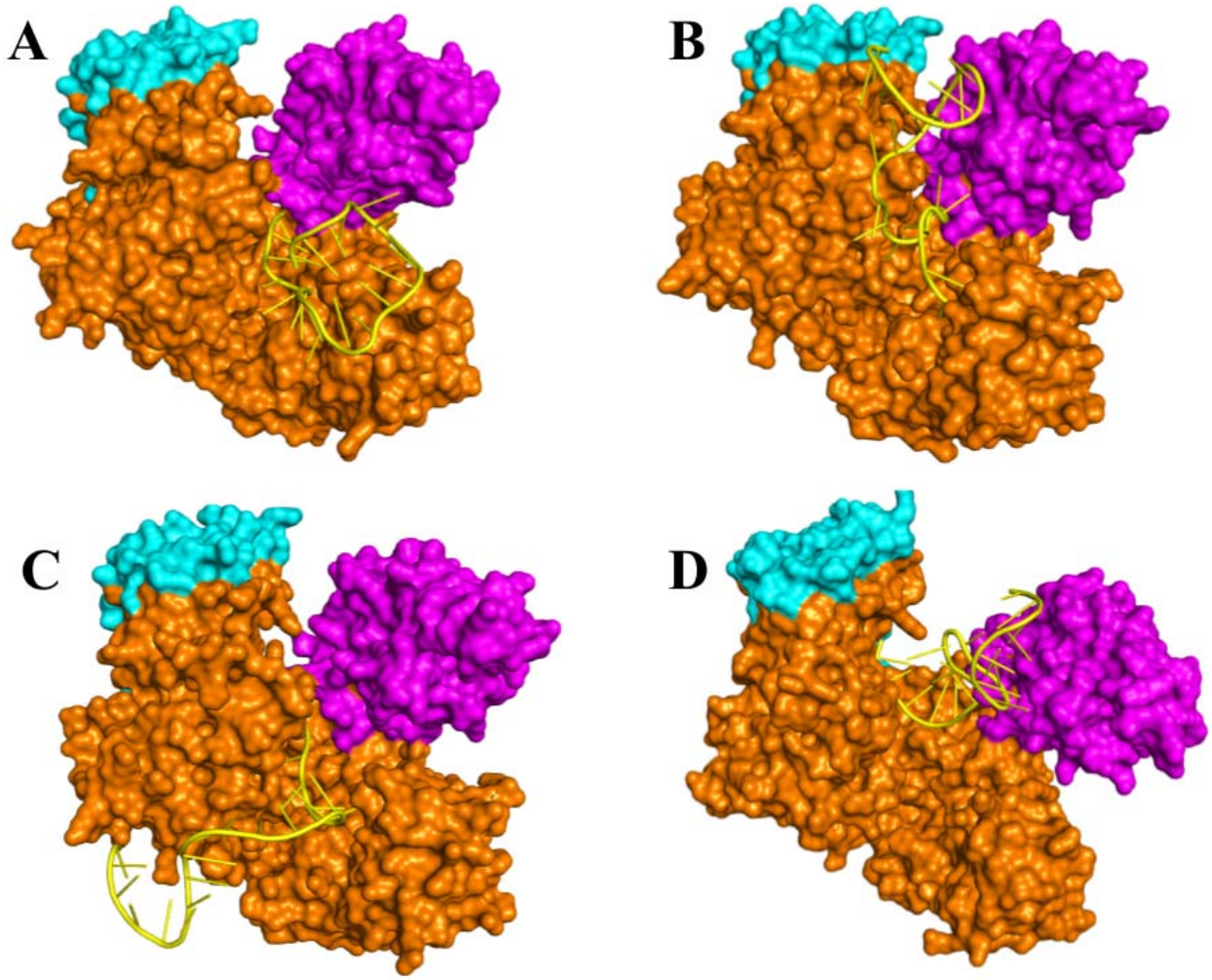

Table 1. Rules/Algorithms for designing of effective siRNA molecules.

Ui-Tei rules	Amarzguioui rules	Reynolds ruless
• A/U at the 5′ end of the antisense strand • G/C at the 5′ end of the sense strand • At least five A/U residues in the 5′ terminal one□third of the antisense strand • The absence of any GC stretches of more than 9 nt in length	• The A/U differential of the duplex end should be >0 • Strong binding of 5 sense strand • Position 1 must have any bases other than U • Position 6 must have A constantly • Weak attachment of 3′ sense/passenger strand.	• The designed siRNA must maintain a GC content between 30% to 52% (1 point) • Occurrence of three or more A/U base pair at position 15–19 of sense strand (Each A/U base pair in this region earns one point) • The Tm (melting temperature) of designed siRNA must be greater than −20 °C (1 point) • Position 19 of sense strand must contain A (1 point) • Position 3 of sense strand must include A (1 point) • Position 10 of the sense strand should have U (1 point) • Position 13 of the sense strand must contain any bases other than G (1 point) • Threshold for efficient siRNAs score = 6

Table 2. Effective siRNA molecules with GC%, free energy of folding and free energy of binding with target.

S.no	Alias	Conserved position	Location of target within mRNA	siRNA target within mRNA	Predicted siRNA siRNA duplex candidate at 37◦C; RNA oligo sequences 21nt guide (5' →3') 21nt passenger (5' →3')	Functional siRNA selection: Ui-Tei, Reynolds and Amarzguioui	Seed duplex (Tm) guide	Seed duplex (Tm) Passenger	GC%	Free energy of folding	Free energy of binding	Tm (Conc)	Tm (Cp)
1	e5	70-95	1-23	GTGGTATTCTTGCTAGTTACACT	UGUAACUAGCAAGAAUACCAC GGUAUUCUUGCUAGUUACACU	U R A	14.3 °C	14.5 °C	38%	1.7	-34.1	85	86
2	m1	16-83	9-31	TACCGTTGAAGAGCTTAAAAAGC	UUUUUAAGCUCUUCAACGGUA CCGUUGAAGAGCUUAAAAAGC	U R A	-3.8 °C	21.1 °C	38%	1.8	-31.3	87.5	88.7
3	m2	16-83	21-43	GCTTAAAAAGCTCCTTGAACAAT	UGUUCAAGGAGCUUUUUAAGC UUAAAAAGCUCCUUGAACAAU	R	19.2 °C	-3.8 °C	33%	1.6	-33.5	86.3	86.9
4	n1	717-749	11-33	GGCCAAACTGTCACTAAGAAATC	UUUCUUAGUGACAGUUUGGCC CCAAACUGUCACUAAGAAAUC	U R A	11.7 °C	16.7 °C	40%	1.8	-35.8	86.2	87.6
5	n2	809-860	27-49	CCAGAACAAACCCAAGGAAATTT	AUUUCCUUGGGUUUGUUCUGG AGAACAAACCCAAGGAAAUUU	R A	18.7 °C	14.9 °C	38%	1.7	-34.9	88.9	90.3
6	g36	919-1022	38-60	GCCCTTTTGGTACTGTAGAAAAA	UUUCUACAGUACCAAAAGGGC CCUUUUGGUACUGUAGAAAAA	UA	20.3 °C	16.1 °C	38%	1.9	-36.1	88.9°C	89.3°
7	g82	1964-2024	39-61	GCAGGTATATGCGCTAGTTATCA	AUAACUAGCGCAUAUACCUGC AGGUAUAUGCGCUAGUUAUCA	A	11.3 °C	15.2 °C	40%	1.8	-35.7	87.4°C	88.2°C
8	g83	2194-2339	1-23	ACCAAGACATCAGTAGATTGTAC	ACAAUCUACUGAUGUCUUGGU CAAGACAUCAGUAGAUUGUAC	UR	13.4 °C	19.2 °C	38%	1.5	-34.9	: 83.5°C	84.7°C
9	g105	2676-2720	19-41	TGCTATGCAAATGGCTTATAGGT	CUAUAAGCCAUUUGCAUAGCA CUAUGCAAAUGGCUUAUAGGU	R	14.9 °C	21.1 °C	38%	1.5	-34.8	85.1°C	86.0°C
10	g134	3649-3713	18-40	TGGCTTGATTGCCATAGTAATGG	AUUACUAUGGCAAUCAAGCCA GCUUGAUUGCCAUAGUAAUGG	UA	6.3 °C	12.0 °C	40%	2	-34.8	86.1°C	87.1°C
11	g80	1881-1925	9-31	CTCCTACTTGGCGTGTTTATTCT	AAUAAACACGCCAAGUAGGAG CCUACUUGGCGUGUUUAUUCU	UA	6.9 °C	16.4 °C	43%	1.9	-34.7	89.7°C	91.0°C
12	g109	2722-2785	14-36	CACAGAATGTTCTCTATGAGAAC	UCUCAUAGAGAACAUUCUGUG CAGAAUGUUCUCUAUGAGAAC	UA	17.8 °C	19.2 °C	38%	1.9	-34.7	84.4°C	84.0°C

Table 3. siRNAs docking score and interaction statistics.

S.no	siRNA in Ago2-siRNA complex	HDock Docking score	Interaction Category				Total number of interacting residues
			Salt bridge (Hydrogen Bond+Electrostatic)	Hydrogen bonds	Electrostatic bonds	Hydrophobic bonds	
1	e5	-322.34	7	16	8	5	23
2	m1	-324.56	3	13	11	6	27
3	m2	-348.34	3	15	8	3	20
4	n1	-335.64	1	15	7	0	16
5	n2	-294.72	1	10	9	5	14
6	g36	-354.19	1	12	7	2	16
7	g82	-290.15	1	21	10	3	19
8	g83	-326.1	1	12	4	5	12
9	g105	-312.45	2	14	3	3	17
10	g134	-338.34	0	14	3	2	12
11	g80	-358.89	4	8	10	2	17
12	g109	-319.69	3	12	7	2	12

Table 4. Final siRNA candidates with respect to each gene.

S.no	Alias	Predicted siRNA siRNA duplex candidate at 37◦C; RNA oligo sequences 21nt guide (5' →3') 21nt passenger (5' →3')	Functional siRNA selection: Ui-Tei, Reynolds and Amarzguioui	Seed duplex (Tm) guide	Seed duplex (Tm) Passenger	GC%	Free energy of folding	Free energy of binding	HDock Docking score
1	e5	UGUAACUAGCAAGAAUACCAC GGUAUUCUUGCUAGUUACACU	U R A	14.3 °C	14.5 °C	38%	1.7	-34.1	-322.34
2	m2	UGUUCAAGGAGCUUUUUAAGC UUAAAAAGCUCCUUGAACAAU	R	19.2 °C	-3.8 °C	33%	1.6	-33.5	-348.34
3	n1	UUUCUUAGUGACAGUUUGGCC CCAAACUGUCACUAAGAAAUC	U R A	11.7 °C	16.7 °C	40%	1.8	-35.8	-335.64
4	g80	AAUAAACACGCCAAGUAGGAG CCUACUUGGCGUGUUUAUUCU	UA	6.9 °C	16.4 °C	43%	1.9	-34.7	-358.89