

In silico development of a novel DNA-directed interfering RNA fragment to treat SARS-CoV-2

Ananya Vittaladevuni¹, Ishya Mukkamala², Saahil Das³, Soumya Suresh

¹Dougherty Valley High School, San Ramon, CA

²Amador Valley High School, Pleasanton, CA

³Irvington High School, Fremont, CA

KEYWORDS: SARS-CoV-2, DNA-directed RNA interference, short interfering RNA, small hairpin RNA, mRNA secondary structures

ABSTRACT: The SARS-CoV-2 pandemic, identified in November 2019 in Wuhan, China¹, has killed more than 250,000 people and has hospitalized countless others². In order to mitigate the virus's effect on the human body, we developed an RNA fragment that may silence a gene in the SARS-CoV-2 genome. ddRNAi, or DNA-directed RNA interference, is a gene-silencing method that uses DNA constructs to hijack the pre-existing animal cell's RNA interference pathways.³ In our project, we designed a novel ddRNAi strand that inhibits the formation of viral copies and the virus' spread is contained. Our siRNA strand does not share any complete similarities with the known human genome. It corresponds to the SARS-CoV-2 virus genome at 9841-9859 bp, which corresponds to the non-structural polyprotein region of the SARS-CoV-2 genome. This region plays a role in forming proteases, which cleave large polypeptides to form core proteins, which form the viral envelope.^{4,5}

The number of deaths related to SARS-CoV-2 is only going up, which is why it is incredibly important to find an effective treatment to combat the virus immediately.

After much research, a vast amount of information has been established about SARS-CoV-2. Through this it is seen that there are numerous advantages to using ddRNAi to combat the virus. The first advantage is that ddRNAi can produce a replenishable supply of shRNAs at a steady level. Short hairpin RNA, called shRNA, is spliced into short interfering RNA, called siRNA, through RNA interference (RNAi) machinery. The siRNA then performs its vital function – cleaving the messenger RNA, or mRNA. This allows siRNAs to bind to the viral RNA and prevent unnecessary folding, which will inhibit the virus from proliferating due to the lack of an essential protein. Since ddRNAi can provide a steady amount of shRNAs, it is an efficient way to combat SARS-CoV-2. Another advantage is that ddRNAi can silence multiple genes in the same area by using DNA constructs to initiate a cell's endogenous RNA interference pathways. DNA constructs are outlined to show self-complementary double-stranded RNA's (usually short hairpin RNAs). The targeted gene will be silenced once the shRNA is

processed. The supplementary segments of the ddRNAi may be used to make copies of normal genes, which are able to restore functions. In this case, the diseased gene's translation is disrupted due to the ddRNAi. ddRNAi can also be used to cure diseases where a defective allele causes negative effects, as in sickle cell disease. Through the copies of normal genes constructed by the supplementary segments, the gene's functions can be restored, thus restoring the function of a healthy cell. ddRNAi can also use shRNA to remove the expression that is emitted from disease protein. The last advantage of using ddRNAi is that one administration is effective for a prolonged period of time since the shRNA fragments are self-proliferating through host cell transcription.^{6,7}

As shown in **Figure 1**, the ddRNAi segment is transported by a vector or transfection agent into the cell, where endogenous RNA transcription mechanisms continuously produce small hairpin RNA fragments, or shRNA. Short hairpin RNA is an artificial RNA molecule and it can be used to silence gene expression using RNAi by delivering plasmids through viral or bacterial

vectors. The shRNA is then spliced into double-stranded siRNA by an enzyme called DICER that is produced by the human cell.⁸ These fragments are used in tandem with an RNA induced silencing complex, or RISC, to silence the target mRNA. ddRNAi has successfully silenced genes in cases including, but not limited to, HIV, Hepatitis A, and Hepatitis B. DNA directed RNAi (ddRNAi) introduces DNA templates to use the cell's endogenous transcriptional machinery to make short hairpin RNAs. This later becomes processed by endogenous RNAi into siRNAs. siRNA is inserted into RNA induced silencing complex (RISC) and the antisense strand-RISC complex binds to the target mRNA. Since the mRNA is then nonfunctional, the protein's function cannot be performed and the virus cannot proliferate. Cells use this natural process to initiate gene expression.⁶

RNA interference, or RNAi, involves designing small interfering RNA fragments, or siRNA, which are double-stranded RNA that contain approximately 19 nucleotides to target genes with a complementary sequence. The siRNA binds to the complementary sequence on the gene, prevents translation, and therefore inhibits the function of that gene. The mRNA secondary structure, also called the mRNA motif, mainly plays a structural role and facilitates target site accessibility, improving efficiency of the siRNA fragment. The mRNA secondary structure also aids in forming the hairpin structure, which aids in the inhibition of translation.

We have designed a fragment that can participate in ddRNAi to inhibit nucleotides 9841-9859 on the sense strand of the SARS-CoV-2 genome, which corresponds to a region on the non-structural polyprotein 1ab gene in the SARS-CoV-2 genome. This will prevent the proliferation and formation of new viral copies within the body and so may prevent life-threatening infection.

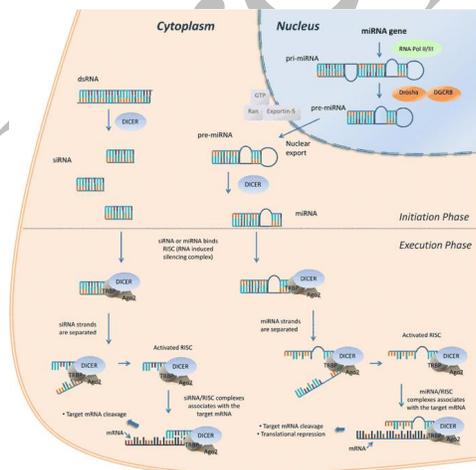


Figure 1:

The mechanism of action of ddRNAi in the animal cell.⁹

Methods:

The Eurofins Genomics siRNA Design Tool was utilized to generate possible siRNA fragments and their respective mRNA secondary structures. The mRNA secondary structures are used to find the corresponding siRNA target site for the ddRNAi fragment. There are five options as to which nucleotides are used in the mRNA secondary structure, labeled "mRNA secondary structure" on the Eurofins Genomics digital page. The "All secondary structures" option was chosen to generate possible mRNA secondary structure sequences. GC content is directly related to stability under high temperatures, the threshold of which is 43 for most molecules involved in the human body.¹⁰ All other options and settings were set to the default given by Eurofins Genomics.

In order to verify that the siRNA binds to a region in the known SARS-CoV-2 genome, the fragment sequence was analyzed in tandem with the SARS-CoV-2 genome as recorded by NCBI as of May 5, 2020. One complete similarity was found in the non-structural polyprotein 1ab region, which corresponds with nucleotides 9841-9859 on the sense strand of the SARS-CoV-2 genome.

In order to verify that the ddRNAi fragment does not share any complete similarities with the known human genome, the fragment sequence was analyzed in tandem with the human genome as recorded by NCBI as of May 5, 2020. No complete similarities were found with human chromosomes 1-23 and the X and Y chromosomes. We also verified that the siRNA fragment produced by cellular degradation of the original fragment was present in the known SARS-CoV-2 genome, as recorded by NCBI. The NCBI accession numbers used are shown below in **Figure 2**.¹¹

Human Chromosome	Accession Number
1	NC_000001.11
2	NC_000002.12
3	NC_000003.12
4	NC_000004.12
5	NC_000005.10
6	NC_000006.12

7	NC_000007.14
8	NC_000008.11
9	NC_000009.12
10	NC_000010.11
11	NC_000011.10
12	NC_000012.12
13	NC_000013.11
14	NC_000014.9
15	NC_000015.10
16	NC_000016.10
17	NC_000017.11
18	NC_000018.10
19	NC_000019.10
20	NC_000020.11
21	NC_000021.9
22	NC_000022.11
X	NC_000023.11
Y	NC_000024.10

Table 1. Table of all NCBI accession numbers corresponding with the human chromosomes.

The i-Score algorithm, developed by Ichihara, Murakumo, et al., was utilized to determine the most potent siRNA fragment. The sequence GUUGCGUAGUGAUGUGCUA was found to be the highest-scoring siRNA fragment of the ten fragments we analyzed. Multiple other potency and stability tests were conducted and the results are shown below in **Figure 2**.

Sense	Antisense	Sec. dG	5' end
GUUGCGUAGUGAUGUGCUA	UAGCACAUACACUACGCAAC	-2.7	-2.2

Takasaki	s- Biopredsi	i- Score	Reynolds	Katoh	DSIR
8.2		70.8	6	73.1	

3' end	Whole dG	GC stretch	%GC	Ui- Tei	Amarzguioui	Hsieh
-1.3	-37.6	3	47.4	1b	2	2

Figure 2. Results from the i-Score algorithm^{12,13}

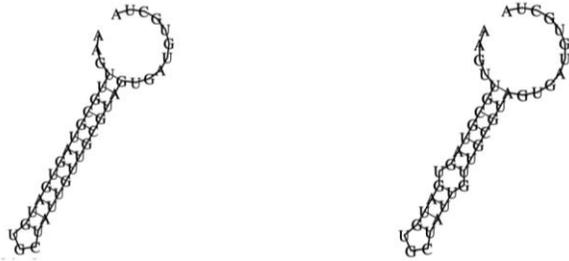
Its from the i-Score algorithm^{12,13}

Results:

The final sequence of the ddRNAi fragment is AAGUUGCGUAGUGAUGUGCUA where AAGUUGCGUAGUGAUGUGCUA is the mRNA secondary structure region and GUUGCGUAGUGAUGUGCUA is the siRNA region.

There are two main ways to show the probability of certain mRNA secondary structures based on the primary structure, as shown in **Figure 3**, where the discussed siRNA is shown. The centroid secondary structure predicts a more accurate structure than the MFE (minimal free energy) secondary structure because of three factors: sensitivity, positive predictive values (PPV), and base-pair distance. The sensitivity and PPV percentages are analogous to one another and measure the number of similar base pairs between the predicted structure and the one produced by comparative sequence analysis. The only difference between these percentages is that the PPV focuses on the accuracy of each base pair in the structure while sensitivity focuses on predicting base pairs. The higher these percentages are, the closer the secondary structure prediction is to the comparative sequence analysis model and the fewer differences between base pairs are present. Base pair distance refers to the distance between the base pairs of the secondary sequence and the comparative sequence analysis model. Looking at sensitivity, the centroid secondary structure predicts slightly improved, if not comparable, structures. However, observing the PPV, the centroid secondary structure predicts structures with 46.5% and 30% fewer errors for the best centroid and ensemble centroid, respectively. The best centroid is the cluster centroid, or a group of similar centroids, with the shortest base pair distance to the reference structure. An ensemble centroid is an entire collection of structures sampled from a weighted ensemble. The centroid secondary structure also predicts a structure with less base-pair distance with the comparative sequence analysis model than the MFE. Additionally, MFE predictions become unreliable when they aren't in the best cluster. When the MFE was located outside of the best cluster, the prediction had a sensitivity percentage of 31.4% and a PPV percentage of 62.5%. However, when the MFE is located inside the best cluster, the sensitivity and PPV percentages improve and the distance between base pairs decreases. This is important because the MFE structure falls outside of this cluster for over half of the studied sequences. This further proves

that the centroid secondary structure predicts better and



MFE secondary structure
more accurate structures.¹⁴

centroid secondary structure

Figure 3. Two representations of probable conformations of the discussed siRNA and mRNA secondary structure.¹⁵

The mRNA secondary structures are used to find the corresponding siRNA target site for the ddRNAi fragment. In Eurofins Genomics, there are five options as to which nucleotides are used in the mRNA secondary structure. The “All motifs” option was chosen to generate possible mRNA secondary structure sequences. GC content is directly related to stability under high temperatures, the threshold of which is 43 for most molecules involved in the human body. All other options and settings were set to the default given by Eurofins Genomics.

In Nucleotide BLAST, the “Align two or more sequences” option was chosen. The siRNA fragment sequence was entered into the ‘Query’ section and the SARS-CoV-2 NCBI accession number was entered into the ‘Subject’ section. The ‘megablast’ setting was used as, according to the NCBI website, it compares a query to a closely related sequence and functions optimally if the target percent identity is equal to or greater than 95%. We found that the siRNA inhibits function by binding to base pairs 9841-9859 on the sense strand of SARS-CoV-2, which is a section of the non-structural polyprotein 1ab gene.^{16,17}

To verify that the siRNA doesn’t bind to any site in the human genome, Nucleotide BLAST was used. One Nucleotide BLAST was run for each chromosome, so a total of 24 BLASTs were run to cross-reference the siRNA fragment and the human genome. The “Align two

or more sequences” option was chosen. The siRNA fragment sequence was entered into the ‘Query’ section. The NCBI accession number for each chromosome was entered into the ‘Subject’ section. The ‘megablast’ setting was used as, according to the NCBI website, it compares a query to a closely related sequence and functions optimally if the target percent identity is equal to or greater than 95%. We found that no chromosome has a segment completely similar to the siRNA fragment.¹⁸

Discussion:

A hairpin, or loop-like shape, is formed when an unpaired mRNA strand folds or forms hydrogen bonds between complementary nucleotides within the mRNA strand or with another section of the same strand. This hairpin can regulate interactions in a ribozyme, act as a substrate for enzymatic reactions, serve as a secondary structure for RNA binding proteins, protect the mRNA from degradation, and guide RNA folding.^{19,20}

The impact of an mRNA secondary structure on translation depends on the position and stability of the hairpin, which is located near the 5’ end of the mRNA sequence. The stability of the hairpin and its proximity to the 5’ end may play a role in determining translational initiation, and therefore translational inhibition. For example, a very unstable hairpin intensified translation initiation when it was located 14 nucleotides downstream of a AUG codon. Meanwhile, moderate to very stable hairpins are able to prevent translation by binding the mRNA to the preinitiation complex, a complex needed for protein production.²¹

The sequence provided by the i-Score algorithm had no similarities to any of the human chromosomes and corresponded to a section of the viral genome. It would be able to attach to that part of the virus and the siRNA would remove the mRNA from the rest of the process. This would stop the virus from continuing to spread and successfully combat it. This sequence is unable to be tested *in vivo* because of restrictions due to SARS-CoV-2. The algorithm provided the most potent siRNA possible, so there is a high probability this sequence could interfere with the translation process and combat the virus.

The siRNA fragment binds to the SARS-CoV-2 virus genome at the 9841-9859 bp sense strand. This site is a section of the gene necessary to form the non-structural polyprotein 1ab. This polyprotein forms proteases, which are utilized to cleave large proteins into smaller polypeptides that will assemble the mature virus. Since our siRNA inhibits the expression of this gene, the

proteases will not form properly and the mature viruses will not be assembled. The non-structural polyprotein 1ab may also inhibit translation of host RNA, which normally prevents survival of the host cell. However, since the polyprotein is not formed, host RNA translation will continue, further countering proliferation of the virus.^{16,17}

The character of dinucleotide overhangs on the 3' end of the siRNA sequence determines the duration of gene silencing. The 3' overhang on the guide strand interacts with the PAZ domain, an RNA binding molecule that can bind to the 3' end of both siRNA and miRNA, of RISC in the RNA-binding pocket, suggesting that there might be a correlation between the character of the overhang and gene silencing. Thymidine overhangs seem to be detrimental to obtaining the maximum duration of silencing from the siRNA. A possible theory for why this occurs is that the siRNA might be exposed to DNAses in the presence of thymidine. Deoxythymidine overhangs have been tested and are proven to affect the duration of gene silencing negatively. Knockdown efficiency was not compromised when thymidine was substituted with uridine in the 3' overhangs to reduce the cost of siRNA synthesis and to protect siRNA from nuclease degradation. In other words, one way to protect the siRNA from degradation without affecting its ability to attach to the virus is by replacing thymidine with uridine in the overhangs.^{22,23}

In regards to whether the siRNA can survive in human body conditions, the GC content of the siRNA segment is 47.4%, which is above the average 41% for the human genome. So, the siRNA is capable of functioning in normal human body conditions. There is also some proof that higher GC content is directly related to a stable secondary structure, which allows the ddRNAi fragment to be sturdier under higher temperatures.²⁴

In order to properly and effectively administer the siRNA fragment, it must first overcome the barriers of each administration route. The basic endocytosis of an ddRNAi vector and the simple mechanism of ddRNAi is shown in **Figure 4**. In cutaneous administration, the stratum corneum, or outermost layer of the skin, imposes a barrier. The morphology and anatomy of the lung prevents potent pulmonary administration. Ocular siRNA injection poses many difficulties, including physical barriers and possible degradation of the siRNA due to metabolic barriers. In order to successfully be transported in the central nervous system, the siRNA must cross the blood-brain barrier using endogenous receptors. In addition to transporting the siRNA, the carrier must also bind and condense the siRNA, protecting it from degradation. The carrier must transport

the siRNA to the target cells and facilitate endocytosis of the siRNA. The siRNA requires this aid from the carrier due to its large molecular weight and strong anionic charge. Viral vectors, though effective, pose a threat regarding oncogenicity, immunogenicity, and cytotoxicity. Non-viral vectors, including but not limited to polyplexes, lipoplexes and peptide-based systems, are promising tools for gene delivery because they incorporate ligand systems to target specific cell types and are reasonably safe.²⁵

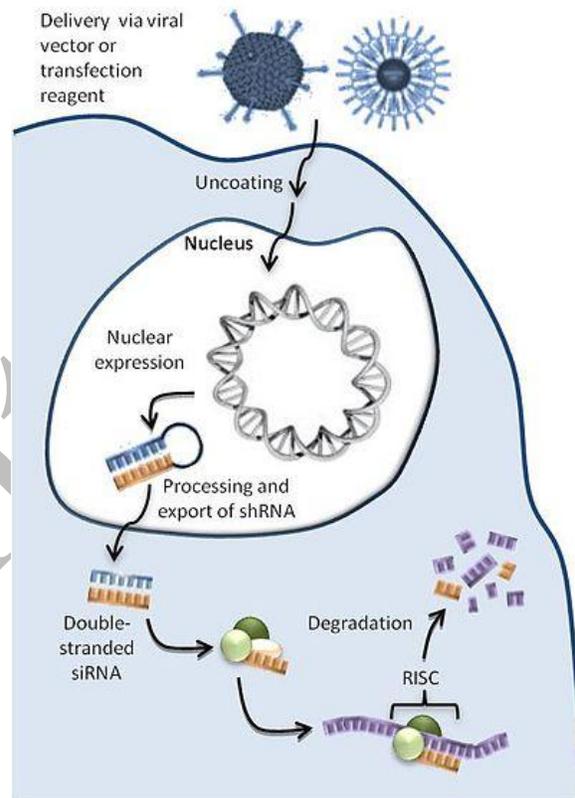


Figure 4. Basic transmission and mechanism of ddRNAi in an animal cell (simplified)²⁶

ASSOCIATED CONTENT

N/A

AUTHOR INFORMATION

Corresponding Author

Ananya Vittaladevuni. Email address: ananyav0607@gmail.com. Phone number: (925)-389-8515.

Author Contributions

The manuscript was written through contributions of all authors. The authors have no conflict of interest.

Funding Sources

Olive Children Foundation and the Center for Advanced Study funded this project.

ACKNOWLEDGMENT

We would like to acknowledge our advisor, Soumya Suresh, who helped us through every aspect of our research. We would also like to thank Olive Children's Foundation and the congregation of sponsors who made this research possible. Lastly, a big thank you to the Center for Advanced Study program and all the Fremont Stem faculty, who supported us in all lab-associated work.

ABBREVIATIONS

SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus-2; MERS, Middle East Respiratory Syndrome; SARS, Severe Acute Respiratory Syndrome; RNA, ribonucleic acid; DNA, deoxyribonucleic acid; mRNA, messenger RNA; RNAi, RNA interference; ddRNAi, DNA-directed RNA interference; shRNA, short hairpin RNA; siRNA, short interfering RNA; RISC, RNA-induced silencing complex; HIV, human immunodeficiency virus; NCBI, National Center for Biotechnology Information; MFE, minimal free energy; PPV, positive predictive values; BLAST, Basic Local Alignment Search Tool

REFERENCES

Andersen, Kristian G., Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes, and Robert F. Garry. "The Proximal Origin of SARS-CoV-2," March 17, 2020. <https://www.nature.com/articles/s41591-020-0820-9>.

"COVID-19 Situation Reports," May 9, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.

"DNA-Directed RNA Interference," April 13, 2020. https://en.wikipedia.org/wiki/DNA-directed_RNA_interference.

Lu, Roujian, Xiang Zhao, Juan Li, Peihua Niu, Bo Yang, Honglong Wu, Wenling Wang et al. "Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding," February 22, 2020.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7159086/>.

Johnston, Sarah L. "Biologic Therapies: What and When?," January 2007. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1860592/>.

"Gene Silencing," April 19, 2020. <https://benitec.com/our-science/gene-silencing/>.

Haussecker, Dirk. "The Business of RNAi Therapeutics in 2012," February 7, 2012. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3381602/>.

O'Keefe, Erin P. "SiRNAs and ShRNAs: Tools for Protein Knockdown by Gene Silencing," April 11, 2020. <https://www.labome.com/method/siRNAs-and-shRNAs-Tools-for-Protein-Knockdown-by-Genes-Silencing.html>.

Chan, Chi Yu, C Steven Carmack, Dang D Long, Anil Maliyekkel, Yu Shao, Igor B Roninson, and Ye Ding. "A Structural Interpretation of the Effect of GC-Content on Efficiency of RNA Interference," January 30, 2009. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2648742/>.

Xu, S., Montgomery, Kostas, Elbashir, J. Harborth, W. Lendeckel, A. Yalcin et al. "Delivery Systems and Local Administration Routes for Therapeutic siRNA," January 1, 1998. <https://link.springer.com/article/10.1007/s11095-013-0971-1>.

Ichihara, M, Y Murakumo, A Masuda, T Matsuura, N Asai, M Jijiwa, M Ishida et al. "i-Score Designer," 2007. https://www.med.nagoya-u.ac.jp/neurogenetics/i_Score/i_score.html.

Masatoshi, Murakumo, Yoshiki, Masuda, Akio, Matsuura, Toru et al. "Thermodynamic Instability of siRNA Duplex Is a Prerequisite for Dependable Prediction of siRNA Activities," September 20, 2007. <https://academic.oup.com/nar/article/35/18/e123/2402822>.

"Human Genome Assembly GRCh38.p13 - Genome Reference Consortium." Accessed May 9, 2020. <https://www.ncbi.nlm.nih.gov/grc/human/data>.

Ding, Ye, Chi Yu Chan, and Charles E Lawrence. "RNA Secondary Structure Prediction by Centroids in a Boltzmann Weighted Ensemble," August 2005. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1370799/>.

"RNAfold Web Server." Accessed May 9, 2020.
<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>.

Lu, Roujian, Xiang Zhao, Juan Li, Peihua Niu, Bo Yang, Honglong Wu, Wenling Wang et al. "Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding," February 22, 2020.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7159086/>.

"Replicase Polyprotein 1ab (SARS Coronavirus)." Accessed May 9, 2020.
<https://pubchem.ncbi.nlm.nih.gov/protein/POC6X7>.

"Nucleotide BLAST: Search Nucleotide Databases Using a Nucleotide Query." Accessed May 9, 2020.
<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast>.

Svoboda, P, and A Di Cara. "Hairpin RNA: a Secondary Structure of Primary Importance," April 2006.
<https://www.ncbi.nlm.nih.gov/pubmed/16568238>.

"Hairpin Loop (mRNA)." Accessed May 9, 2020.
<https://www.nature.com/scitable/definition/hairpin-loop-mrna-314/>.

Wang, L, and S R Wessler. "Role of mRNA Secondary Structure in Translational Repression of the Maize Transcriptional Activator Lc(1,2)," March 2001.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC65616/>.

Strapps, Walter R, Victoria Pickering, Gladys T Muiru, Julie Rice, Stacey Orsborn, Barry A Polisky, Alan Sachs, and Steven R Bartz. "The siRNA Sequence and Guide Strand Overhangs Are Determinants of in Vivo Duration of Silencing," August 2010.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2919711/>.

Wrighton, Katharine H. "Small Interfering RNAs Silence Genes in Mammals," November 14, 2019.
<https://www.nature.com/articles/d42859-019-00083-3>.

Chan, Chi Yu, C Steven Carmack, Dang D Long, Anil Maliyekkel, Yu Shao, Igor B Roninson, and Ye Ding. "A Structural Interpretation of the Effect of GC-Content on Efficiency of RNA Interference," January 30, 2009.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2648742/>.

Xu, S., Montgomery, Kostas, Elbashir, J. Harborth, W. Lendeckel, A. Yalcin et al. "Delivery Systems and Local Administration Routes for Therapeutic siRNA," January 1, 1998.
<https://link.springer.com/article/10.1007/s11095-013-0971-1>.

"DNA-Directed RNA Interference," April 13, 2020.
https://en.wikipedia.org/wiki/DNA-directed_RNA_interference.

SYNOPSIS The SARS-CoV-2 pandemic, identified in November 2019 in Wuhan, China, has killed more than 250,000 people and has hospitalized countless others. In order to mitigate the virus's effect on the human body, we developed an RNA fragment that may silence a gene in the SARS-CoV-2 genome. ddRNAi, or DNA-directed RNA interference, is a gene-silencing method that uses DNA constructs to hijack the pre-existing animal cell's RNA interference pathways. In our project, we designed a novel ddRNAi strand that inhibits the non-structural polyprotein 1ab in SARS-CoV-2. By inhibiting this gene, the viral copies cannot properly form and the virus' spread is contained. Our siRNA strand does not share any complete similarities with the known human genome and corresponds to the SARS-CoV-2 virus genome at 9841-9859 bp.

Preprint Only