A computational approach to design putative potential siRNAs for silencing the nucleoprotein and surface glycoprotein gene of MARV

By

Tasnia Iffat Suraiya

ID- 18136032

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of B.Sc. in Biotechnology

Department of Mathematics and Natural Sciences

Brac University

June 2023

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing a degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. I have acknowledged all main sources of help.

Student's Full Name and Signature :

_____

Tasnia Iffat Suraiya

Student ID : 18136032

Approval

The thesis/project titled "[Thesis/Project Title]" submitted by

1. Tasnia Iffat Suraiya (18136032) of summer 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B. Sc in Biotechnology on

Examining Committee:

Supervisor:

(Member) _____

Mohammed Sayem

Lecturer, Department of Mathematics and Natural Science, Brac University

Program Director:

(Member) _____

Dr. Munima Haque,

Associate Professor, Department of Mathematics and Natural Sciences, Brac University

Departmental Head:

(Chair) _____

A F M Yusuf Haider

Professor and Chairperson, Department of Mathematics and Natural Sciences, Brac University

Abstract

Human fatality rates for the highly lethal hemorrhagic fever caused by the Marburg virus (MARV), a member of the Filoviridae family, range from 23% to 90% [1]. This virus is native to central Africa and often causes periodic outbreaks with a high mortality toll [1, 2]. Despite the seriousness of the sickness brought on by filoviruses, there is no approved vaccination or treatment. A nonsegmented, negative-sense, single-stranded RNA genome called MARV is made up of seven genes, of which the nucleoprotein and surface glycoprotein mediate entry into host cells and genome replication, respectively [3,4,5]. These genes are good candidates for the development of therapies due to their essential involvement in the viral life cycle.By reducing viral gene expression by hybridization and neutralization of complementary target mRNA, small interfering RNA (siRNA) is a promising antiviral therapeutic approach that uses the RNA interference (RNAi) pathway [6,7,8]. In this study, we use RNA interference technology to create siRNA molecules that target particular MARV target genes.This virus's surface glycoprotein and nucleocapsid phosphoprotein are two crucial proteins that aid in genome replication and host cell entrance. The RNA interference (RNAi) pathway's small interfering RNA (siRNA) is a potential strategy for controlling human viral infections by inhibiting viral gene expression by hybridization and neutralization of the target complementary mRNA. In order to create siRNA molecules against two specific target genes, nucleocapsid phosphoprotein and surface glycoprotein, the power of RNA interference technology was used in this study. Finally, three siRNA molecules were chosen that are thought to have the best effect based on GC content, free energy of folding, free energy of binding, melting temperature, efficacy prediction, and molecular docking study.

Acknowledgement

I thank Allah for guiding me throughout my academic careers, including this particular endeavor, in my work and in my studies. This research would not have been possible without the help of others, who are respectfully recognized here.

I would like to express my gratitude to Professor A F M Yusuf Haider, Chairperson, Department of Mathematics and Natural Sciences for allowing me to finish my undergraduate thesis.

My supervisor, Mohammed Sayem, Lecturer, Department of Mathematics and Natural Sciences, has been incredibly supportive and dedicated throughout this process. He has also given me helpful ideas and direction whenever I needed it, and for that I am incredibly grateful to him and owe him a great deal of appreciation.

Contents

Chapter 4

List of Tables

List of Figures

List of Acronyms

MVD Marburg virus disease

GP Glycoprotein

NP nucleoprotein

VP Virion Protein

L Polymerase

BLAST Basic Local Alignment Search Tool

FASTA Fast- all

HF Hemorrhagic fever

CDS Coding Sequence

NCBI National Center for Biotechnology Information

MSA Multiple sequence alignment

DSSE Different stability of siRNA ends

DSIR Dept. of Scientific and Industrial Research

GESS Genome-wide enrichment of seed sequence matches

AGO2 Argonaute 2 protein

AIS Almost Invariant Sets

MD Molecular docking

siRNA Small interfering RNA

RNAi RNA Interference

PDB Protein Data Bank

CDC Centers for Disease Control and Prevention

Chapter 1

Background

The Marburg virus was discovered for the first time in 1967 when lab staff and researchers at institutions in Germany and the former Yugoslavia got affected by imported monkeys from the Lake Kyoga region of Uganda. Since its discovery in 1968, MARV has been responsible for occasional infections in a small number of people in Africa. Two significant outbreaks occurred in the Democratic Republic of the Congo (DRC) from 1998 to 2000, with a lethality of 83%, and in Angola both 2004 and 2005, with a lethality of 90%. As a result, the public's health is seriously threatened by MARV. Research on MARV is critically required because there are no vaccinations or medications for treating MARV infections [9].

People and also mammals are both prone to a rare yet very severe hemorrhagic fever termed Marburg virus disease (MVD). Marburg virus or Ravn virus infection, both belonging to the genus Marburgvirus, is the cause of it. Marburg viruses and Ebola viruses share a tight genetic relationship while being genetically diverse. [51] Research has revealed that the marburg viruses are shed in the urine, feces, and oral secretions of infected Egyptian rousette bats. It is uncertain how the marburg viruses, first transferred from the animal host to humans. During the two incidents in 2008 , tourists while visiting to Uganda probably caused by improper contact with infectious bat stool or urine droplets. Following this, transmission happens when two individuals come into touch. The virus can spread through touch. Fever, chills, headaches, and myalgia are the first symptoms to appear after an incubation span of 2–21 days. There are numerous signs and symptoms of MVD, including those of endemic viral hemorrhagic fevers like Lassa fever or Ebola or parasitic diseases like malaria or typhoid fever. [52]. Case mortality rates of Marburg outbreaks in previous years have varied from 23 to 83%[11].Recently, The minister of health and social welfare of Equatorial Guinea just declared the Marburg Virus Disease outbreak in that nation on February 13, 2023..Moreover,Tanzanian

government officials announced the first-ever Marburg virus disease epidemic in their nation on March 21, 2023. At this point, we don't know anything about any further suspicious cases. The CDC is actively monitoring the issue.

The filoviruses Marburg and Ebola are both members of the family *Filoviridae*. The two illnesses share clinical similarities despite being brought on by distinct viruses. Both illnesses are rare and have the potential to spread in large-scale outbreaks with a high mortality rate. Since then, there have been numerous outbreaks of the virus around the globe, and because of its peculiar character, there has been a lot of research done on its structure and modes of transmission. The Marburg virus is a single species, however various lineages might differ from one another at the nucleotide level by up to 21%.The Marburg virus is quite slow to mutate and has a small range of variation when compared to other viral species. The Marburg virus has an odd shape. Due to their versatile nature, they can assume a number of shapes, such as rod- or ring-like, crook- or six-shaped, or those with branched structures, 37% are spherical, 37% of viral particles are six-shaped, and 30% are filamentous. In contrast, another study found that, although Marburg virus particles varied greatly in length, they were uniformly 80 nm in diameter.[52]

Genome organizations

Marburg virus (MARV) genome at the top and the minigenome at the bottom showing in the diagram below. The seven MARV genes coding areas are depicted in the genome as boxes. Above the genome are references to the corresponding proteins of various NNS RNA viruses. The nucleoprotein (NP) gene's nontranslated region, which contains the gene start (GS) signal, and the large protein (L) gene's nontranslated region, which contains the gene end (GE) signal, surround a single reporter gene that serves as the minigenome's place-holder for the viral genes.

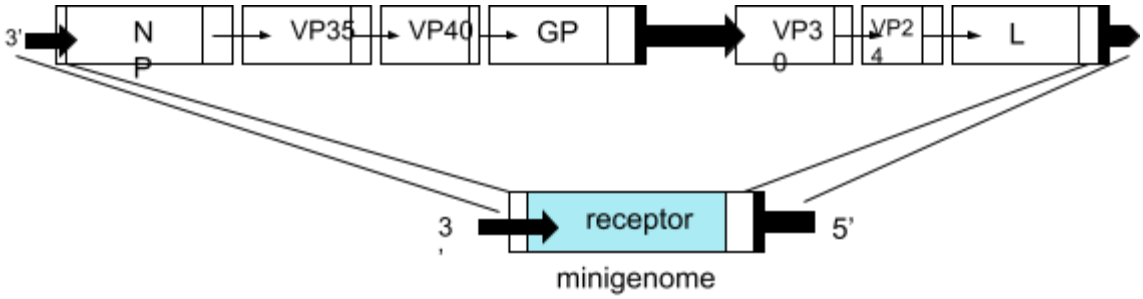

Figure 1: The genome organization of the Marburg virus

Viral proteins

The MARV genome contains the codes for seven structural proteins. The NP gene's role in transcription and replication, nucleocapsid production, and budding. NP is a very crucial part of the replication machinery. Vp35 involves in polymerase cofactor, an essential part of nucleocapsids, and an IFN antagonist. Therefore Vp40 works as an antagonist of IFN signaling, consequently functions in budding. Gp is in responsibility for fusion, tetherin antagonist, receptor binding, and attachment. Moreover, Vp30 essentials for nucleocapsid formation. Also Vp 24 considered to be maturation of nucleocapsids, budding. Additionally, The role of L is to catalyze the RNA-dependent RNA polymerase's catalytic domain.

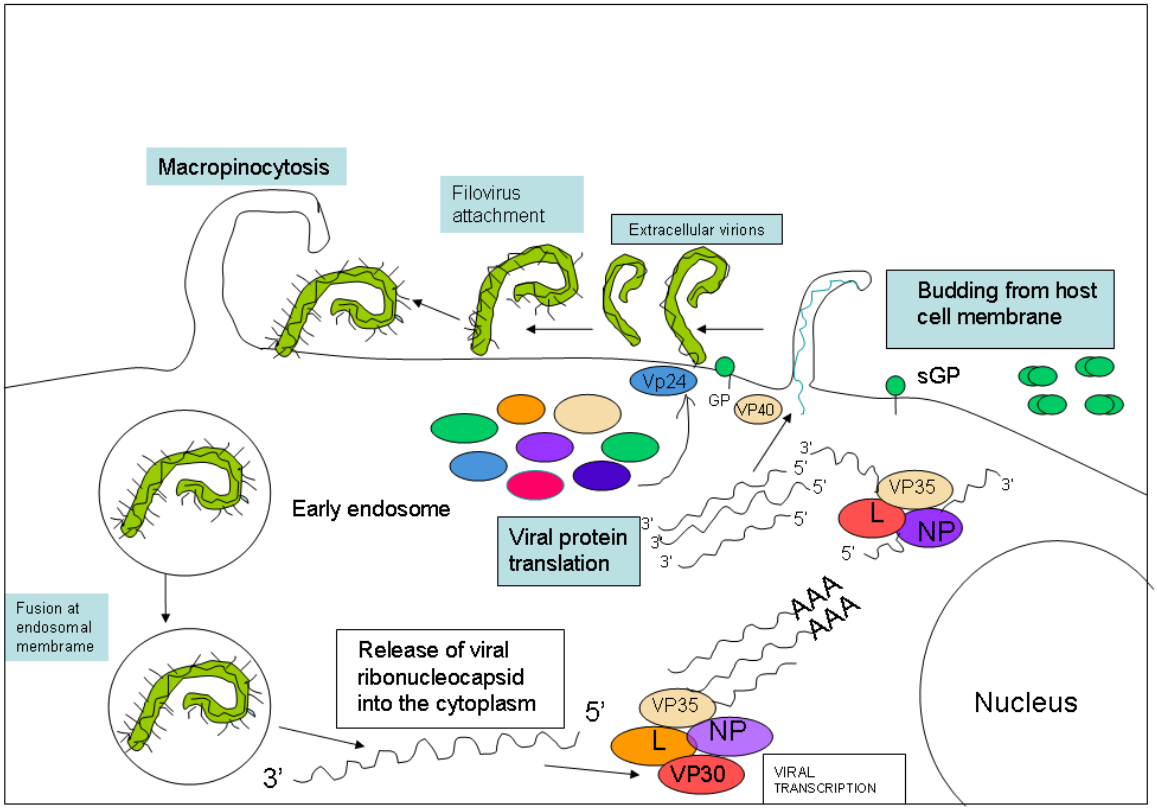

Figure 2: Replication Cycle of marburg virus. Retrieved from [60]

MARV's viral life cycle is shown. Firstly, surface GP protein binds to the cell membrane. A technique known as macropinocytosis is used to take up viruses. The viral ribonucleocapsid is released into the cytoplasm where the negative strand RNA genome proceeds through transcription and replication thanks to GP's role in mediating the fusion of the viral and endosomal membranes. Individual viral genes produce 5′-capped, 3′-polyadenylated mRNAs, and then the genome replicates, copying the genomic RNA template into a full-length complement. For the synthesis of new negative-sense genomes, the full-length complement acts as a template. The NP, VP35, and L proteins are necessary for the RNA production reactions. The VP30 protein is necessary for the marburg virus to begin transcription, but it is not necessary for the virus to replicate. Virus mRNAs are translated into viral proteins like NP, VP35, VP40, GP, VP30, VP24, and L. The plasma membrane is where new virus particles are assembled. As the viral

matrix protein, the VP40 protein guides the budding of particles from the cell surface1. The type I transmembrane protein GP and viral nucleocapsids holding the viral DNA are both integrated into the budding particles. [62]

After nearly 18 years, two instances of the Marburg virus disease were discovered in the Ashanti region of Ghana in west Africa in July 2022.The fact that there are now no antiviral medications available that are especially for the marburg virus makes it impossible to treat this growing number of patients. There is no exception to the overall trend of discovering novel therapeutic techniques in-silico when it comes to viruses[13].

Silencing RNA or short interfering RNA are other names for small interfering RNA (siRNA).RNA with two strands (dsRNA). In the somatic tissues of mammals siRNA, a double-stranded RNA with a length that ranges from 21–25 nucleotides, has the power to inhibit the production of particular genes.. As its capacity to mute gene expression was discovered, siRNA has provided a cutting-edge way to treat diseases with a genetic basis [53]. siRNAs typically consist of two 21-mer oligonucleotide strands with 19 complementary nucleotides (nt) at either 3′ end, a 2-nt overhang [54]. When this siRNA molecule reaches the cell cytoplasm, it will be incorporated into a multiprotein complex called the RNA-induced silencing complex (RISC), which contains nucleases. The siRNA duplexes will be split in the RISC complex. The more stable 5′ end of the strand is combined with the active RISC complex. The RISC complex is then guided to locate and cleave the target by the catalytic RISC protein, an argonaute family member (Ago2).. This siRNA molecule will be integrated into a nuclease-containing multi-protein complex called the RNA-induced silencing complex (RISC) after it has reached the cell cytoplasm. In the RISC complex, the siRNA duplexes will be divided. The active RISC complex is merged with the strand with the more stable 5′ end. The catalytic RISC protein, a member of the argonaute family (Ago2), is then responsible for directing the RISC complex to seek out and cleave the target. [56,57]

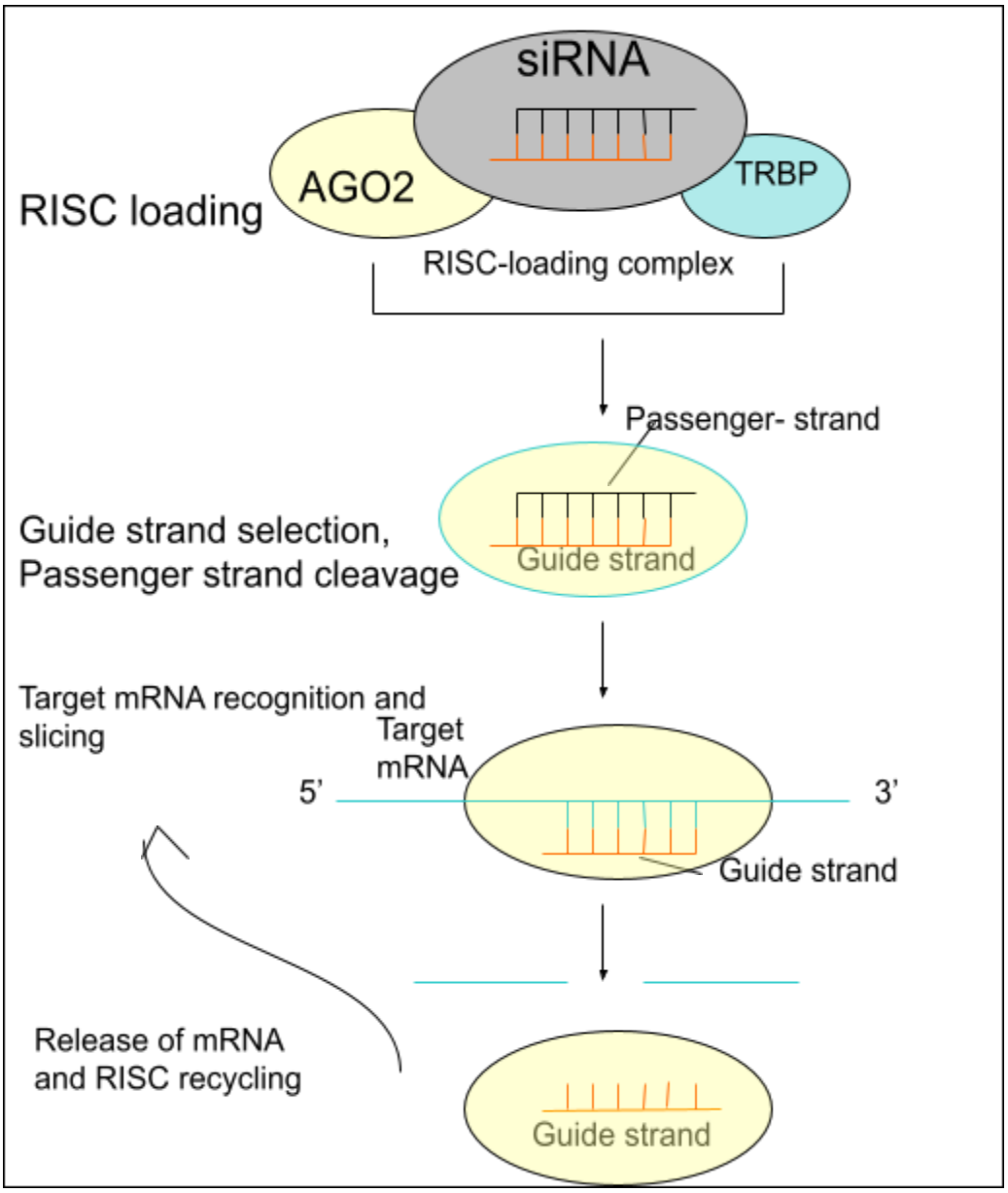

Figure 3: Mechanisms of siRNA.

SiRNA's ability to suppress genes linked to disease makes RNAi a promising mechanism for modern therapy. We aim to create siRNAs that are specific to different consensus regions of the marburg virus's nucleocapsid phosphoprotein and surface glycoprotein genes. Both nucleoproteins and glycoproteins are targeted by siRNAs as they are

necessary for viral survival and because doing so may prevent viral replication.[14,15].Currently, acute hepatic porphyria and polyneuropathy are now being treated with the help of the RNAi therapies ONPATTRO® and GIVLAARITM. This study should help in the creation of a Marburg virus treatment strategy, as expected[16].

Chapter 2

Methods

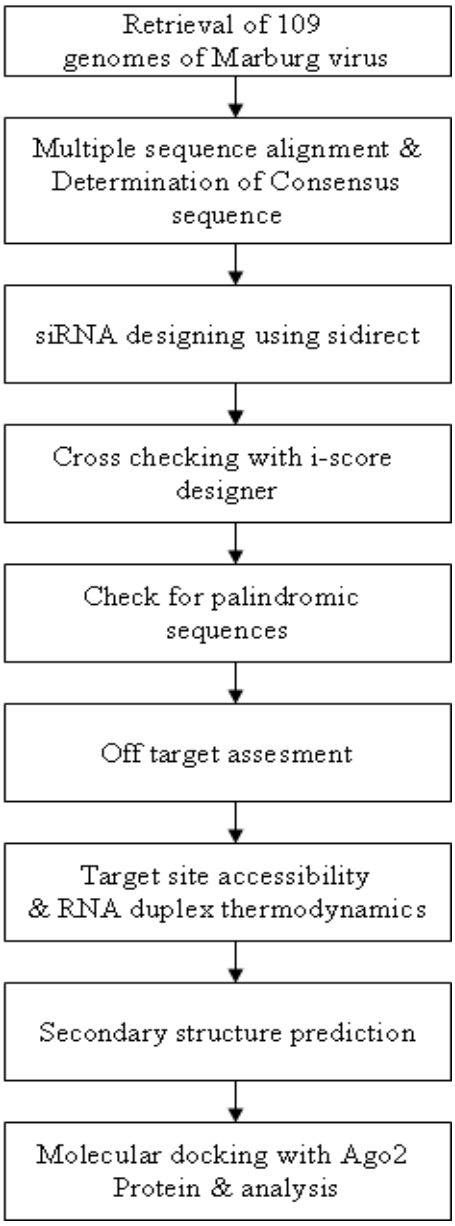

Figure 4: Illustrated the methodology in flowchart.

2.1 Retrieval oof sequences from NCBI

NCBI or The National Center for Biotechnology Information a bioinformatics website www.ncbi.nlm.nih.gov/labs/virus/vssi/#/ was used to find the marburg protein's coding sequences (CDSs). Numerous databases pertaining to biotechnology and biomedicine can be found at the NCBI, which is a major resource for bioinformatics tools and services. "Marburg virus," was set to in all databases of nucleotides, and the sequence length ranged from 15000 to 50000.

2.2 Alignment of multiple sequences

Using the online clustal omega platform, CDSs (coding sequences) align to the specified marburg nucleotide sequences for a particular protein. Multiple sequence alignment, or MSA has become essential in the investigation of biological sequences' comparative function and structure as well. It frequently results in important biological insights into the links between the sequence, structure, and function of families of nucleotide or protein sequences.

2.3 Determination of consensus sequence

MSA output used as input to find the consensus sequences on Clustalw https://www.ebi.ac.uk/Tools/msa/clustalo/. This tool is used for pairwise alignment. After that, using emboss needle https://www.ebi.ac.uk/Tools/psa/emboss_needle/, aligned the consensus sequence with the gp and np.

2.4 siRNA target-specific prediction

The siDirect 2.0 web server was used to create target-specific siRNAs based on consensus areas of the CDSs of the seven proteins. A web-based software program called siDirect http://design.RNAi.jp/ is used to produce siRNA sequences for mammalian RNA interference (RNAi) that are extremely effective and have the highest level of target selectivity.

The siRNAs were predicted by siDirect 2.0 then compared to the results from i-Score Designer, which assigns scores to nine different siRNA designing algorithms, including Ui-Tei, s-Biopredsi, i-Score and Reynolds, and DSIR [22]. I-Score Designer https://www.med.nagoya-u.ac.jp/neurogenetics/i_Score/i_score.html used for the additional target sequence analysis. Last but not least, the siRNAs predicted by siDirect 2.0 that matched with siRNAs predicted by i-Score Designer and had a rank of 1 in all three methods (s-Biopredsi, i-Score, and DSIR) were chosen for additional analysis. The iScore Designer computes 9 different siRNA designing scores . i-Score [28], s-Biopredsi [27], and DSIR [29] are three of the nine scores.

2.5 Palindromic sequence filtering and siRNA off-target evaluation

Using the EMBOSS palindrome online tool, guide and passenger strands both can be examined www.bioinformatics.nl/cgi-bin/emboss/palindrome .To satisfy the needs of the molecular biology and bioinformatics user community, a software analysis toolkit known as EMBOSS was developed as free and open source software. The palindrome had to be at least four letters long.

Three stages were taken to evaluate how well siRNAs performed off-target. First, siRNAs' guide and passenger strands were compared using NCBI BLAST to the database of reference transcript sequences.

The BLASTN program was selected. The word size was set at seven. The anticipated threshold for a complete specificity check was set at 1000. The elimination of siRNAs with query coverage 80% or lower than Query coverage is the proportion of the counting length that corresponds to an NCBI hit. Only a very small portion of the contig is

aligning, as shown by a low query coverage %. In order to achieve perfect alignment with the short query sequences, the BLAST settings were changed. Mismatch penalty, gap opening, and gap extension were set to 2, 2, and -1, respectively.When it comes to the second stage (seed-dependent off-target or miRNA-like off-target), a partial homology assessment should be done. Although utilizing siRNAs with melting temperatures (Tm) of 21.5°C has already reduced seed-dependent off-target. It offers the Genome-wide Enrichment of Seed Sequence Matching program www.flyrnai.org/gess/ , which directly analyzes the outcomes of primary screening to identify potential off-target transcripts. Individual seed-dependent off-target evaluations were performed on the sense and antisense strands of siRNAs.

2.6 Prediction of RNA duplex thermodynamics and accessibility of target sites

Target accessibility and siRNA duplex end stability can be predicted using the siRNA application S-fold module https://sfold.wadsworth.org/cgi-bin/sirna.pl . This tool predicts potential RNA secondary structures and evaluates viral RNA target accessibility. Also,Internal stability at the cleavage point on average, and Antisense siRNA binding energy kcal/mol, Internal stability at the cleavage point on average are some of the siRNA parameters that are provided by S-fold.

2.7 Evaluation of the secondary structure and its interaction with the target region and the guide strand

The secondary structures and folding free energies of the target site and the guide strand of the siRNA were calculated using the MaxExpect program of the RNA structure web server https://rna.urmc.rochester.edu/RNAstructureWeb/ Lower folding probability is indicated by a higher free energy of folding. The DuplexFold tool of the RNA structure online service was used to observe the interaction among target sequence and the guide strand. The binding free energy is also estimated using DuplexFold.

2.8 Molecular Docking (MD)

In order to molecularly bind the anticipated siRNAs with AGO2, it was first necessary to determine the 3D structure of the receptor and ligand molecules. Using the HDOCK server http://hdock.phys.hust.edu.cn/ , the AGO2 protein and predicted siRNA will be docked. [50]. Using a hybrid template-based or free template modeling method, the HDOCK server performed protein-RNA docking and predicted AGO2-RNA complexes in addition to docking energy scores. The HDOCK generates several models from the input structure, including confidence scores, docking scores, and ligand RMSDs. The possibility of binding models will increase with a high negative score. Additionally, if the confidence score is greater than 0.7, the two molecules will have the greatest chance of binding; but, if the confidence level is between 0.5 and 0.7, binding may not be possible; and if the confidence level is less than 0.5, binding may not be possible at all.

Chapter 3

Result

3.1 Retrieval of the Marburg virus's nucleotide sequences

Based on the date of collection, the database was accessed and 109 viral genome sequences by using NCBI's https://www.ncbi.nlm.nih.gov/ for subsequent analysis.

3.2 MSA (Multiple sequence alignment)

We performed MSA of all the genomes with this tool https://www.genome.jp/tools-bin/clustalw and generated a consensus sequence using the MSA file in the next step.

3.3 Determination of consensus sequence

In this step, we found a consensus sequence from the whole genome which is the representative of 109 genome sequences. This tool first reads two input sequences, then EMBOSS Needle saves the file with the best global sequence alignment. Here, we extracted the gene sequences of gp and np after the pairwise alignment were done.

3.4 Target-specific siRNA prediction

Following that, gene sequences in fasta format were identified by siDirect 2.0. All three algorithms Ui-Tei, Amarzguioui and Reynolds were followed by the output of sidirect sequences. 17 of these were found to be of a rank of 1 across all 3 methods such as i-score, dsir and s-biopredsi and to match with predicted siRNAs that were obtained using i-Score Designer.

Table 1: sidirect2.0 Output

Sl no	Target position	Target sequence	RNA (oligo guide,	Passenger (sense)	Functional siRNA selection	seed-duplex stabilty (Tm), guide	passenger

			antisense)				
gp1	344-366	CCGTGACTATCCTAAATGCAAAA	UUGCAUUUAGGAUAGUCACGG	GUGACUAUCCUAAAUGCAAAA	UA	20	20.3
gp2	576-598	CACATGAATCTGACTTCTACTAA	AGUAGAAGUCAGAUUCAUGUG	CAUGAAUCUGACUUCUACUAA	UA	18.9	16.2
gp3	1351-1373	TACTCAACATCTTATATATTTCA	AAAUAUAUAAGAUGUUGAGUA	CUCAACAUCUUAUAUAUUUCA	URA	-8.6	20.5
gp4	1728-1750	CGCTGGTTTAATTAAAATCAAA	UGAUUUUUAAUUAAACCAGCG	CUGGUUUAAUUAAAAAUCAAA	URA	0.4	20
gp5	1733-1755	GTTTAATTAAAATCAAAACAAT	UGUUUUGAUUUUUAAUUAAAC	UUAAUUAAAAAUCAAAACAAU	R	14.9	-7.5

gp6	1833-1855	GGAAAGAACATTTTCCTTAATCA	AUUAAGGAAAAUGUUCUUUCC	AAAGAACAUUUUCCUUAAUCA	R	19.9	19.2
gp7	2112-2134	ATCCTGTATTTGTCGTATCTTTA	AAGAUACGACAAAUACAGGAU	CCUGUAUUUGUCGUAUCUUUA	UA	20.9	11.6
np1	58-80	CGTAATAAGAAGGTGATATTATT	UAAUAUCACCUUCUUAUUACG	UAAUAAGAAGGUGAUAUUAUU	R	8.7	6.9
np2	109-131	AATCAGATAATAGATGCAATAAA	UAUUGCAUCUAUUAUCUGAUU	UCAGAUAAUAGAUGCAAUAAA	R	19.9	13.4
np3	172-194	CTGACTTTGTGTGTTGAACATTA	AUGUUCAACACACAAAGUCAG	GACUUUGUGUGUUGAACAUUA	URA	20.5	17.8
np4	544-566	TGAAAGTAATTTTTG	AAAUACCAAAAAUUA	AAAGUAAUUUUUGGU	R	20	4.6

		GTATTTTG	CUUUCA	AUUUUG			
np5	545-567	GAAAGTAATTTTTGGTATTTTGA	AAAAUACCAAAAAUUACUUUC	AAGUAAUUUUUGGUAUUUUGA	A	13.9	4.6
np6	641-663	ATGACAGTATCATCAGTAATTCA	AAUUACUGAUGAUACUGUCAU	GACAGUAUCAUCAGUAAUUCA	UA	11.6	19
np7	712-734	GAGTTCATCTTGCAAAAAACTGA	AGUUUUUUGCAAGAUGAACUC	GUUCAUCUUGCAAAAAACUGA	U	-1.8	20.4
np8	964-986	GTGTCAATGTCGGCGAACAATAT	AUUGUUCGCCGACAUUGACAC	GUCAAUGUCGGCGAACAAUAU	UA	21.1	20.5
np9	1463-1485	AGCACAACATCGAGAGAATTTCA	AAAUUCUCUCGAUGUUGUGCU	CACAACAUCGAGAGAAUUUCA	URA	14.8	19.3

np10	1783-1805	GTGAT GTAAA TGGTG ATATCT TA	AGAUA UCACC AUUUA CAUCA C	GAUGU AAAUG GUGAU AUCUU A	UA	17.4	6.9

3.5 siRNA off- target evaluation and palindromic sequence filtering

Only six siRNA exhibited query coverage less than or equal 80% for both strands among 17 siRNAs. After off-target evaluation, those sequences were checked for a palindromic pattern. There were only three siRNAs found which had no palindrome in both guide and passenger strands out of 6 siRNAs. These 3 were gp2, np3 and np7.

Table 2 : off-target evaluation by using blastN criteria 80% or less than that of 80% of query coverage

Name of siRNA	**Highest query coverage matched protein**	**Query coverage(%)**
gp2	Chondroitin sulfate proteoglycan 5 (CSPG5)	80%
gp7	Golgin A8	80%
np3	Tousled like kinase 2 (TLK2)	80%
np7	PHD Finger Protein 3 (PHD3)	80%
np8	Plastin 3 (PLS)	80%
np9	Steap	80%

3.6 Utilizing the S-fold to predict RNA duplex thermodynamics and target accessibility

On the target site accessibility scale, Gp2 earned an 8, along with an 8 for np3, and np7. DSSE was 1, 2.9 & 3.1 kcal/mol for gp2, np3, and np7. Gp2, np3, and np7 had an AIS of -7.2 , -7.1 & -7.4 kcal/mol. The overall siRNA duplex scores for np3 and np7 were both 16. These three siRNAs met all S-fold filter requirements for functional siRNA. Table 3. lists the parameters for the three siRNAs that were ultimately chosen based on S-fold results.

Table 3 : S-fold output for the three siRNAs that were ultimately chosen.

Alias	Passenger (Sense) strand	sum of the siRNA duplex scores	Score for the target accessibility	Score for duplex thermodynamics	Antisense siRNA binding energy kcal/mol	variable stability of siRNA duplex ends (DSSE, kcal/mol)	Internal stability at the cleavage point on average (AIS, kcal/mol)	Score for the duplex feature
gp2	CAUGAAUCUGACUUCUACU	15	8	2	-28.3	1	-7.2	5
np3	GACUUUGUGUGUUGAACAU	16	7	2	-14.3	2.9	-7.1	7
np7	GUUCAUCUUGCAAAAAACU	16	8	2	-22	3.1	-7.4	6

3.7 Identifying secondary structure and interacting with the target region of the guide strand

For the gp target site and guide strand, the free energy of folding was 1.6 kcal/mol and 1.9kcal/mol, respectively. For the np3 target site and the np3 guide strand, it was 1.5 and 1.6, and for the np7 target site and np7 guide strand, it was 1.6 and 1.7. The designs of binding for gp2, np3, and np7 between their target sequences, and guide strands, are shown in (Figure 5)

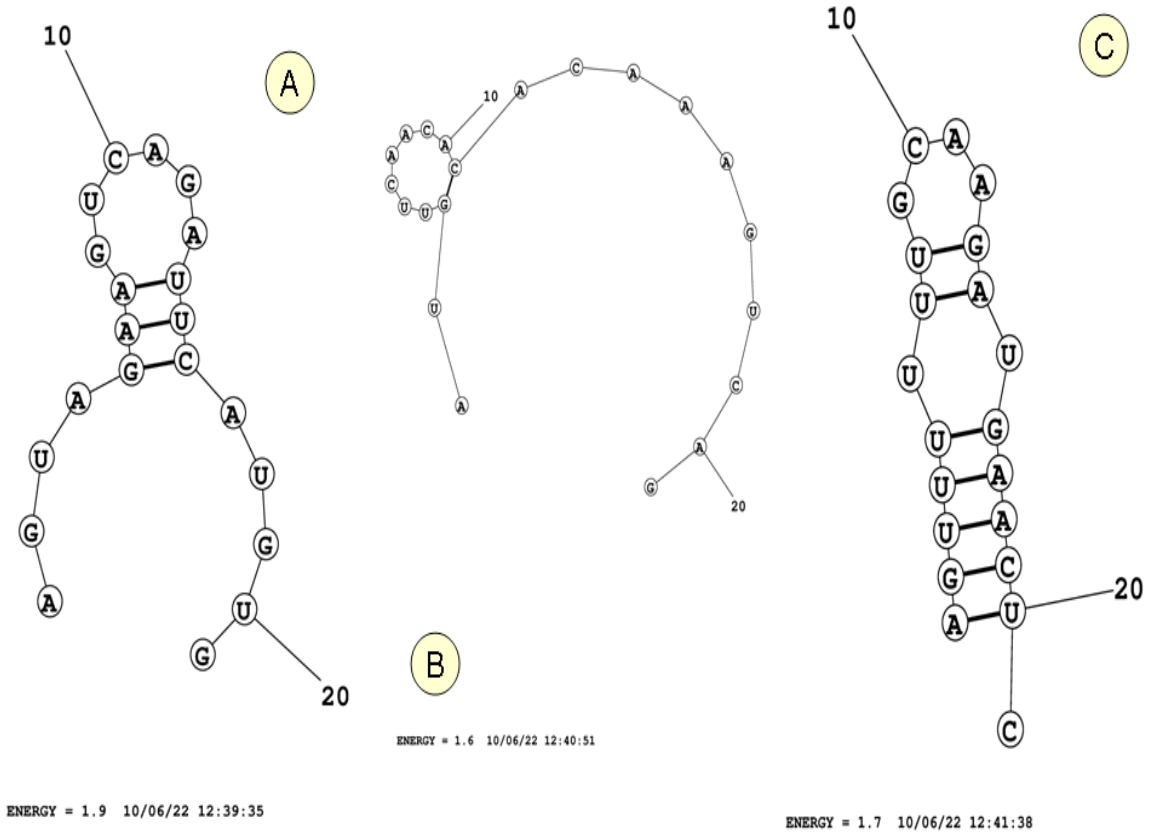

Figure 5 : Three potent siRNAs have secondary structures in their guide strands and target sequence. Also the free energy of folding of (a)gp, (b) np3, and (c) np7.

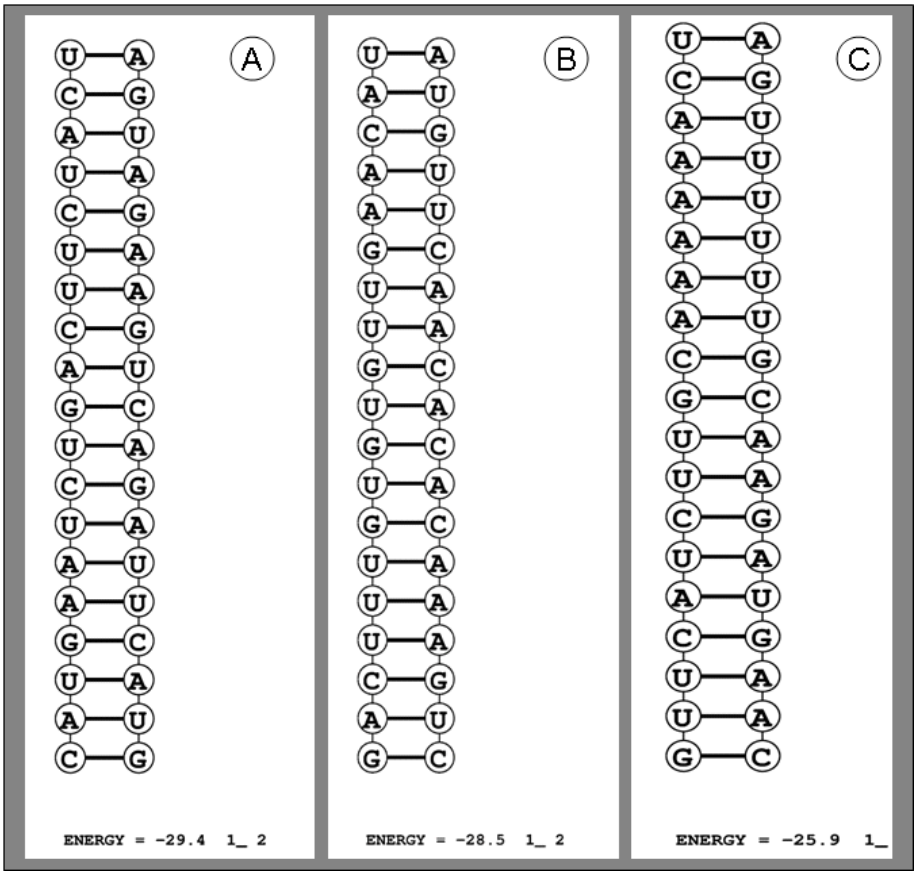

Figure 6: Three efficient siRNAs (a) gp2, (b) np3, and (c) np7 structures of interactions with their target sequences and binding free energies are shown. The strand on the left is the target sequence and the strand on the right is the guide strand.

3.8 Molecular docking (MD) among the predicted siRNA strands and AGO2 Protein

The HDOCK tool predicts the binding complexes between two molecules (receptor and ligand) The HDOCK method's Docking Candidates went through a lot of testing such as docking scores and confidence scores, (Table 4). We identified and visualized the complexes with the least energy values and the nearest siRNA binding locations. Figure 7 illustrates the flow of the current study from sequence retrieval through docking. The Ago2-siRNA complex comprising np3 and np7 demonstrated the greatest and lowest docking scores, respectively, of -455.11 and -317.85, with the remaining values falling in between.

Table 4: docking score, confidence score, and siRNA complex.

Sl no.	Docking Score	Confidence Score
Gp1	-352.77	0.9830
Np3	-455.11	0.9978
Np7	-317.85	0.9663

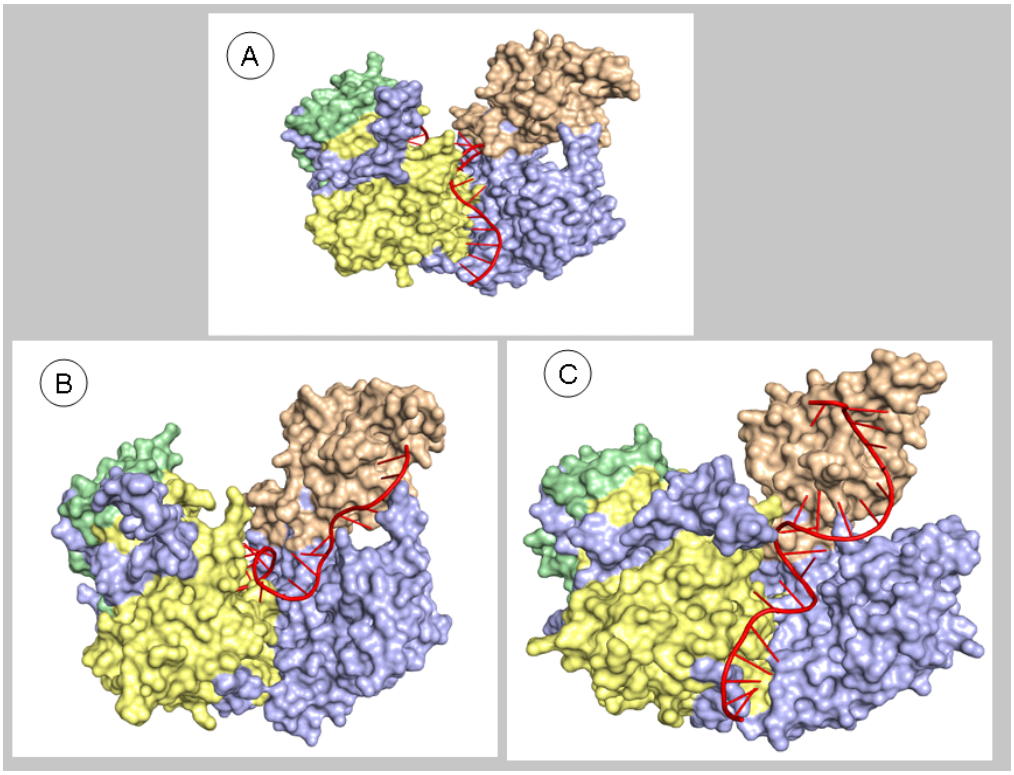

Figure 7 : Molecular docking of siRNA guide strands with the AGO2 protein. Docking structures of the AGO2 protein with the(A) gp2, (B) np3 and (C) np7 sirna guide strand. Here, MID defines pale green, PIWI pale yellow and PAZ wheat-colored. The remaining protein is purple in hue and Red-colored siRNA guide strands are used.

Chapter 4

Discussion

In this study, siRNAs against the marburg virus were predicted using in silico approaches. The sequences were first determined using the sample collection date, which is related to the strain of the marburg virus that was in use at the time.Using consensus sequences of the marburg virus, which include variants of interest and concern, the siRNAs were predicted.

In this case, seven different proteins are present in total gp, np, vp30,vp35,vp24,L were found in 109 Marburg virus strains from throughout the world with consensus areas The siDirect web service was utilized to identify potential targets and create appropriate siRNAs using consensus sequences. Seven gene of MARV, among them 2 of which are nucleocapsid phosphoproteins and surface glycoprotein were identified as siRNA targets.

To get better results and to predict the siRNAs, Ui-Tei, Amarzguioui, and Reynolds guidelines were used by i-score designer. An artificial neural network serves as the foundation for BIOPREDsi by Huesken [31]. DSIR and i-Score have been generated using a linear regression method [29,37] and only six were discovered to overlap with projected siRNAs (determined using i-Score Designer) and achieved a score of 1 in three models those are- i-Score, dsir and s-Biopredsi.. In the next step, NCBI transcript reference sequences database was utilized to compare BLASTN against in order to forecast mRNAs that would be perfectly or almost perfectly complementary to undesirable mRNAs. Short sequences were precisely aligned using modified BLAST parameters. GESS was used to examine the effects in order to boost the expected siRNAs' safety, even though Tm lowers seed-dependent off-target effects.

Target site accessibility and the thermodynamics of RNA duplex were addressed using the siRNA module of S-fold. Based on Turner thermodynamic parameters, the thermodynamic indices utilized in the siRNA module of the S-fold infer siRNA duplex stability [40,41].Target site accessibility is necessary for RISC-mediated mRNA cleavage both in vitro and intracellularly.[42]. Both siRNAs (np7 & gp2) achieved strong target site accessibility scores of 8, indicating effective target binding. The DSSE, which is

measured in kcal/mol, is determined by the variation in 5 end instability between the sense and antisense strands.[43,44].

Furthermore, the DSSE and AIS values were calculated through s-fold. DSSE and AIS values for GP2 were 1 AND -7.2 . NP3 and NP7 both showed DSSE values of 2.9 & 3.1 and an AIS value of -7.1, -7.4 respectively. According to S-fold, Gp2, np3, and np7 all had AIS values more than -8.6 kcal/mol. For these three, a DSSE score >0 indicates a greater likelihood of the guide strand joining with RISC. Target accessibility, duplex features, and duplex thermodynamics are added to create the siRNA duplex score overall. The siRNA duplex score can go up to 20. The Reynolds algorithm is used to calculate the duplex feature score[28]. Three of each kind were above the S-fold threshold of 12 and had a high siRNA duplex score overall of 17.Both np3 and np7's guide strands and target sequences exhibited higher free energy values, clearly demonstrating that they were less likely to fold. The highest accessibility ratings for the target site attained using the siRNA S-fold module were consistent with the free energy of connecting among the target sequences and guide strands are quite low for the gp2(-28.3), np3(-28.3), and np7(-28.3), respectively.

The siRNA's secondary structure may prevent the target from being cleaved by RISC. Therefore, it is critical to forecast future secondary structure and determine the associated folding's free energy. Thereby, the estimated siRNA guide strands' possible folding structures and minimum lowest free energies were calculated using the RNA structure online server (figure 6). The target and guide siRNA interaction was then estimated using DuplexFold [58], along with the associated binding energies. Lower binding energy is a sign of better interaction as well as a greater chance of target inhibition.

Utilizing the siRNAPred, DuplexFold, , the top three siRNA candidates were selected after taking into account the ideal variables derived from the calculation (Table 3).In addition , these siRNAs are anticipated to function better. Following that, to understand how siRNAs interact with the critical RISC enzyme Ago2, molecular docking studies using siRNAs were considered. HDOCK server http://hdock.phys.hust.edu.cn/ was used to dock the Ago2 and siRNA molecules after modeling . In order to select the best docked complex based on the docking score, the resultant docking complexes were

manually reviewed after being downloaded from the server. After docking, among the 10 models , took the most lowest docking score and confidence score of gp2, np3 and np7 (table 3).Compared to gp2 and np7, np3 guide strand AGO2 complex scored higher in terms of docking energy. As we know the more negative docking score , the sirna will be stronger. Additionally, the complexes visualization revealed that the np3 guide strand was entirely bound in the AGO2 chamber, than the gp2 and np7 complex. It can be said that, np3 gains a potential performance over gp2 and np7.

Conclusion

Using computational methods, siRNA interactions can be built and predicted towards a particular gene target to silence the expression of it. In this study, the effectiveness of three siRNA molecules against the genes encoding the outside glycoprotein and nucleoprotein of 109 Marburg virus strains was predicted using a computational method that took into account all the maximum parameters under ideal circumstances and cutting-edge molecular modeling and docking analysis. The creation of therapeutic siRNA methods may offer a viable alternative to conventional vaccine design for slowing the Marburg outbreak and healing the affected people. To sum up, using computational and bioinformatics approaches, three siRNAs were predicted utilizing the consensus areas of the coding sequences of 7 marburg virus genes . These three siRNAs match all the strict and robust filtering criteria. To demonstrate their true therapeutic efficacy, however, additional in vivo research is required.

References

1. World Health Organization. (2021). Marburg virus disease–Guinea. 2021− 08− 09)[2021− 09− 17].https://www. who. int/emergencies/disease-outbreak-news/item/2021-DON331.

2. World Health Organization. (2012). Ebola haemorrhagic fever, Uganda—update. Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire, 87(49-50), 493-493.

3. Kiley MP, Bowen ET, Eddy GA, et al. Filoviridae: a taxonomic home for Marburg and Ebola viruses?. Intervirology. 1982;18(1-2):24-32. doi:10.1159/000149300

4. Miller EH, Obernosterer G, Raaben M, et al. Ebola virus entry requires the host-programmed recognition of an intracellular receptor. EMBO J. 2012;31(8):1947-1960.doi:10.1038/emboj.2012.53

5. Feldmann, H., Geisbert, T. W., Jahrling, P. B., Klenk, H. D., Netesov, S. V., Peters, C. J.& Volchkov,

V. E. (2004). Virus taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. Filoviridae, 645-53.

6. Sharif Shohan, M., Paul, A., & Hossain, M. (2018). Computational design of potential siRNA molecules for silencing nucleoprotein gene of rabies virus. Future Virology, 13(3), 159-170. doi: 10.2217/fvl-2017-0117

7. Chowdhury, F., Shohan, M., Islam, T., Mimu, T., & Palit, P. (2019). A Therapeutic Approach Against Leishmania donovani by Predicting RNAi Molecules Against the Surface Protein, gp63. Current Bioinformatics, 14(6), 541-550.

8. Wu CJ, Huang HW, Liu CY, Hong CF, Chan YL. Inhibition of SARS-CoV replication by siRNA. Antiviral Res. 2005;65(1):45-48. doi:10.1016/j.antiviral.2004.09.005

9. Marburg Virus - an overview | ScienceDirect Topics. (n.d.).

10.Marburg Virus - an overview | ScienceDirect Topics. (n.d.).

11. E.R. Kabir, M.K.S. Siam, N. Mustafa Scaffold of N-(2-(2-(tosylcarbamoyl)hydrazinyl)ethyl)isonicotinamidereveals anticancer effects through selective inhibition of FAPACM Int. Conf. Proceeding Ser, Association for Computing Machinery, New York, NY, USA (2019), pp. 1-11

12.Neumeier, J., & Meister, G. (2021). siRNA Specificity: RNAi Mechanisms and Strategies to Reduce Off-Target Effects. *Frontiers in Plant Science*, *11*.

13.J. S., Khristova, M. L., Sealy, T. K., Vincent, M. J., Erickson, B. R., Bawiec, D. A., Hartman, A. L., Comer, J. A., Zaki, S. R., Ströher, U., Gomes da Silva, F., del Castillo, F., Rollin, P. E., Ksiazek, T. G., & Nichol, S. T. (2006). Marburgvirus Genomics and Association with a Large Hemorrhagic Fever Outbreak in Angola. *Journal of Virology*, *80*(13), 6497–6516.

14. 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), pp. 365-374

15.C.J. Wu, H.W. Huang, C.Y. Liu, C.F. Hong, Y.L. Chan Inhibition of SARS-CoV replication by siRNA Antivir. Res., 65 (2005), pp. 45-48, 10.1016/j.antiviral.2004.09.005

16. B. Hu, L. Zhong, Y. Weng, L. Peng, Y. Huang, Y. Zhao, X.-J. Liang Therapeutic siRNA: state of the art Signal Transduct. Target. Ther., 5 (2020)

17. Bawono, P., Dijkstra, M. J. J., Pirovano, W., De Bree, R., Abeln, S., & Heringa, J. (2008). Multiple sequence alignment. *PubMed*, *452*, 143–161.

18. Ui-Tei K, Naito Y, Takahashi F *et al.* Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. Nucleic Acids Res. 32(3), 936–948 (2004).

19. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. Nat. Biotechnol. 22(3), 326–330 (2004).

20. Amarzguioui M, Prydz H. An algorithm for selection of functional siRNA sequences. Biochem. Biophys. Res. Commun. 316(4), 1050–1058 (2004).

21. Birmingham A, Anderson E, Sullivan K *et al.* A protocol for designing siRNAs with high functionality and specificity. Nat. Protoc. 2(9), 2068–2078 (2007).

22.Vert JP, Foveau N, Lajaunie C, Vandenbrouck Y. An accurate and interpretable model for siRNA efficUi-Tei K, Naito Y, Takahashi F, Haraguchi T, Ohki-Hamazaki H, Juni A, Ueda R, Saigo K. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. Nucleic Acids Res 32: 936-948, 2004.

23Amarzguioui M, Prydz H. An algorithm for selection of functional siRNA sequences. Biochem Biophys Res Commun 316: 1050-1058, 2004.

24. Hsieh AC, Bo R, Manola J, Vazquez F, Bare O, Khvorova A, Scaringe S, Sellers WR. A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. Nucleic Acids Res 32: 893-901, 2004.

25.Takasaki S, Kotani S, Konagaya A. An effective method for selecting siRNA target sequences in mammalian cells. Cell Cycle 3: 790-5, 2004.

26. Huesken D, Lange J, Mickanin C, Weiler J, Asselbergs F, Warner J, Meloon B, Engel S, Rosenberg A, Cohen D, Labow M, Reinhardt M, Natt F, Hall J. Design of a genome-wide siRNA library using an artificial neural network. Nat Biotechnol 23: 995-1001, 2005.

27. Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, Ishida M, Shinmi J, Yatsuya H, Qiao S, Takahashi M, Ohno K. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. Nucleic Acids Res 35: e123, 2007.

28. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. Nat Biotechnol 22: 326-330, 2004.

29.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

30. 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

31. Schirle NT, Sheu-Gruttadauria J, Chandradoss SD, Joo C, MacRae IJ. Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. Elife 4, e07646 (2015).

32. Waterhouse A, Bertoni M, Bienert S *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296–W303 (2018).

33. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res. 41, 384–388 (2013)

34. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2(9), 1511–1519 (1993).

35. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminform. 4(1), 17 (2012).

36. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res. 45(W1), W365–W373 (2017).Huesken D, Lange J, Mickanin C *et al.* Design of a genome-wide siRNA library using an artificial neural network. Nat. Biotechnol. 23(8), 995–1001 (2005).

37. ElHefnawi M, Kim TK, Kamar MA *et al.* In silico design and experimental validation of sirnas targeting conserved regions of multiple hepatitis C virus genotypes. PLoS ONE 11(7), 1–26 (2016).

38. Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J. 20(23), 6877–6888 (2001).Crossref, Medline, CAS, Google Scholar

39.Naito Y, Yoshimura J, Morishita S, Ui-Tei K. SiDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. BMC Bioinformatics 10(1), 1–8 (2009).

40. Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. Cell 115(2), 199–208 (2003).

41. Zhang Y, Li T, Fu L *et al.* Silencing SARS-CoV spike protein expression in cultured cells by RNA interference. FEBS Lett. 560(1–3), 141–146 (2004).

42. Ameres SL, Martinez J, Schroeder R. Molecular basis for target RNA recognition and cleavage by human RISC. Cell 130(1), 101–112 (2007)

43.Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell 115(2), 209–216 (2003)

44. ElHefnawi M, Kim TK, Kamar MA *et al.* In silico design and experimental validation of sirnas targeting conserved regions of multiple hepatitis C virus genotypes. PLoS ONE 11(7), 1–26 (2016).

45. M. Müller, F. Fazi, C. Ciaudo Argonaute proteins: from structure to function in development and pathological cell fate determination Front. Cell Dev. Biol., 7 (2020), p. 360

46. Schirle NT, Sheu-Gruttadauria J, Chandradoss SD, Joo C, MacRae IJ. Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. Elife 4, e07646 (2015).

47. Waterhouse A, Bertoni M, Bienert S *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296–W303 (2018).

48. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res. 41, 384–388 (2013).

49. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminform. 4(1), 17 (2012)

50. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res. 45(W1), W365–W373 (2017).

51. *Marburg Virus Disease Outbreaks | Marburg (Marburg Virus Disease) | CDC*. (2023, April 19). Marburg Virus Disease Outbreaks | Marburg (Marburg Virus Disease) | CDC.

52. *Marburg Virus Structure and Transmission*. (2020, April 1). News-Medical.net.

53. Trends Pharmacol Sci, 25 (1) (2004), pp. 22-28

54. de Fougerolles, A., Manoharan, M., Meyers, R., and Vornlocher, H.P. (2005). RNA interference in vivo: toward synthetic small inhibitory RNA-based therapeutics. Methods Enzymol 392, 278–296.

55. Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells.

56. Nair, J.K., Willoughby, J.L.S., Chan, A., Charisse, K., Alam, M.R., Wang,

Q., Hoekstra, M., Kandasamy, P., Kel'in, A.V., Milstein, S., et al.(2014). Multivalent N-acetylgalactosamine-conjugated siRNA Localizes in hepatocytes and elicits robust RNAi-mediated gene silencing.

57. Ryther, R.C.C., Flynt, A.S., Phillips, J.A., and Patton, J.G. (2005). siRNA therapeutics: big potential from small RNAs.

58. D. Piekna-Przybylska, L. DiChiacchio, D.H. Mathews, R.A. Bambara A sequence similar to tRNA3Lys gene is embedded in HIV-1 U3–R and promotes minus-strand transfer

59. Brauburger, K., Hume, A. J., Mühlberger, E., & Olejnik, J. (2012). Forty-Five Years of Marburg Virus Research. *Viruses*, *4*(10), 1878–1927.

60. Messaoudi, I., Amarasinghe, G. K., & Basler, C. F. (2015). Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus. *Nature Reviews Microbiology*, *13*(11), 663–676.

61. Saw, P. E., & Song, E. (2020). siRNA therapeutics: a clinical reality. *Science China-life Sciences*, *63*(4), 485–500.

62. Messaoudi, I., Amarasinghe, G. K., & Basler, C. F. (2015c). Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus. *Nature Reviews Microbiology*, *13*(11), 663–676. https://doi.org/10.1038/nrmicro3524