

AN IN – SILICO APPROACH TO DESIGN POTENTIAL SIRNA MOLECULE TO SILENCE SERPINA -1 GENE IN ALPHA – 1 ANTITRYPSIN DEFICIENCY

Submitted in partial fulfillment of the requirements for the award of
Master of Science in Bioinformatics and Data Science

By

SUPRIYA.S (42739025)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade “A++” by NAAC | 12B Status by UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

April - 2024



SATHYABAMA

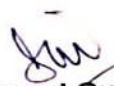
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with "A++" grade by NAAC I 12B Status by UGC I Approved by AICTE
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai-600119
www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **SUPRIYA.S (42739025)**, who carried out the final year project entitled "**AN IN – SILICO APPROACH TO DESIGN POTENTIAL SIRNA MOLECULE TO SILENCE SERPINA – 1 GENE IN ALPHA – 1 ANTITRYPSIN DEFICIENCY**" under our supervision from September 2023 to April 2024.


Internal Guide

Dr. SWETHA SUNKAR


External Guide

Dr. SWETHA SUNKAR


Head of the Department
Dr. H. JEMMY CHRISTY

Submitted for Viva voce Examination held on

20/04/2024


Internal Examiner


External Examiner

DECLARATION

I **SUPRIYA.S (42739025)** hereby declare that the Project Report entitled **“AN IN-SILICO APPROACH TO DESIGN POTENTIAL SIRNA MOLECULE TO SILENCE SERPINA 1 GENE IN ALPHA-1 ANTITRYPSIN DEFICIENCY”** done by me under the guidance of Dr. SWETHA SUNKAR at Sathyabama Institute of Science and Technology, Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai –600 119, Tamil Nadu, India, and is submitted in partial fulfillment of the requirements for the award of Master of Science in Bioinformatics and Data Science.

DATE: 20-04-2024

PLACE: CHENNAI

A handwritten signature in blue ink, appearing to read 'Supriya S.', is shown within a rectangular box.

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to the **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project for completing it successfully. I am grateful to them.

I extend my sincere gratitude to **Dr. REKHA CHAKRAVARTHY, Dean of the School of Science and Humanities**, and I am deeply grateful for her mentorship and leadership, which have been a constant source of inspiration. Her dedication to fostering a conducive academic environment has been instrumental in the successful completion of this thesis.

I convey my thanks to **DR. H. JEMMY CHRISTY, Head of the Department, Department of Bioinformatics**, for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. SWETHA SUNKAR** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Bioinformatics who were helpful in many ways for the completion of the project.

ABSTRACT

Alpha-1 antitrypsin deficiency (AATD) is a genetic disorder characterized by reduced levels or dysfunctional forms of the alpha-1 antitrypsin (AAT) protein. This deficiency results from mutations in the SERPINA1 gene, leading to increased susceptibility to lung and liver diseases. The SERPINA1 gene encodes the AAT protein, and over 100 mutations have been identified in this gene, contributing to the variability and severity of AATD. Targeting the SERPINA1 gene presents a promising therapeutic strategy for AATD treatment, with small interfering RNA (siRNA) molecules designed to silence the expression of mutant SERPINA1 alleles. However, developing a vaccine against AATD remains challenging due to the genetic heterogeneity of the disorder and the intricate interplay of genetic and environmental factors. Small interfering RNAs (siRNAs) play an important role in regulating gene expression by causing the degradation of target mRNAs. In this study, we attempted to rationally design potential siRNA molecules using various software, targeting the *SERPINA 1* gene. In total, 2 siRNAs were found to be potential candidates for SERPINA 1 gene silencing. They showed good target accessibility, optimum guanine-cytosine (GC) content, the least chance of off-target effects, positive energy of folding, and strong interaction with Argonaute2 protein as denoted by a negative docking energy score. In addition, molecular dynamics simulation of the siRNA-Ago2-docked complexes showed the stability of the complexes over 1.5 nanoseconds. These predicted siRNAs have the potential to effectively downregulate the expression of the SERPINA1 gene, which is responsible for producing the faulty alpha-1 antitrypsin (AAT) protein in alpha-1 antitrypsin deficiency (AATD) and thus prove vital in the treatment of Alpha - 1 antitrypsin deficiency

Keywords: SiRNA, SERPINA 1, AATD , AGO2 , Gene silencing, SiDirect, S-FOLD, Hdock, Galaxy refine ,RNA composer.

TABLE OF CONTENTS

S.NO	TOPICS	PAGE NO
	ABSTRACT	i
	LIST OF CONTENTS	ii
	TABLES	iv
	FIGURES	v
	ABBREVIATIONS	vi
1	INTRODUCTION	1
1.1	Alpha-1 antitrypsin deficiency	1
1.2	SERPINA 1	1
1.3	The spectrum of variants in AATD	2
1.4	Epidemiology	3
1.4.1	Frequency: United States:	3
1.4.2	International	3
1.4.3	Race	4
1.4.4	Sex	4
1.4.5	Age	4
1.5	Pathophysiology	4
1.6	Symptoms	5
1.7	Risk factors	5
1.8	Staging	6
1.9	Diagnosis	6
1.10	Treatment	6
1.11	Gene Silencing-RNA Interference	7
2	LITERATURE SURVEY	9
3	AIM AND SCOPE	12
4	MATERIALS AND METHODS	13
4.1	SERPINA 1 sequence retrieval and prediction of sirRNA	13
4.2	Off-target similarity search and filtering siRNAs based	15

	on palindromic sequence	
4.3	Secondary structure prediction and thermodynamics study of the siRNA-target gene interaction	15
4.4	Prediction of siRNA target site accessibility within target mRNA	16
4.5	Modeling and docking of alpha 1 antitrypsin protein and the siRNAs	17
4.6	Md simulation of siRNA – AAT protein Docked complex	17
5	RESULTS AND DISCUSSION	18
5.1	Sidirect 2.0 and i-score designer predicts siRNAs from the retrieved serpin1 target sequence	18
5.2	Filtration of siRNAs for the presence of palindromes, off-target effects, and immunostimulatory action results in 3 potential siRNAs	19
5.3	Secondary structure and thermodynamics study of the candidate siRNAs	19
5.4	Target site accessibility and duplex thermodynamics study of the siRNAs	24
5.5	Molecular docking between Arginase 2 protein and the siRNAs	25
5.6	Molecular Dynamics simulation of the siRNA-AGO2 docked complexes	28
6	SUMMARY & CONCLUSION	35
7	REFERENCES	36

LIST OF TABLES

S.NO.	TABLES	PAGE NO.
4.1	Algorithms/rules applied in siDirect 2.0 server for rational designing of siRNA molecules.	13
5.1	siDirect and i-Score designer predicted overlapped siRNAs along with their different parameters	18
5.2	Displaying the resulting siRNA candidates screened based on Free energy, Binding , T _m (Conc) (°C), T _m (Cp) (°C)	22
5.3	Result of the predicted siRNAs from Sfold server showing target accessibility and various RNA duplex thermodynamics	23
5.4	Displaying the results of Galaxy Refine tool and ERRAT	25
5.5	The docking score, interaction statistics and interactive residues of each of the siRNA docked complexes	27

LIST OF FIGURES

S.NO.	FIGURES	PAGE NO.
1.1	SERPINA1 gene.	1
1.2	RNA interference	8
5.1	Secondary structures of the guide strands and target sequences of siRNAs	20
5.2	Structures of binding of siRNA (guide strand) and target RNA along with their predicted free energy of binding	21
5.3	The top-five ranking models of AGO2 predicted from Galaxy refine tool	25
5.4	S1,S2 AND S3 SiRNA S modelled using RNA composer	26
5.5	Docked structures of siRNAs	28
5.6	RMSF Plot	29

LIST OF ABBREVIATIONS

SIRNA	Small Interfering RNA
AGO2	Argonaute2 protein
SERPINA 1	Serine Protease Inhibitor – A1
AATD	Alpha 1 - antitrypsin deficiency
AAT	Alpha 1 - antitrypsin
COPD	Chronic Obstructive Pulmonary Disease
RNAi	RNA interference
DsRNA	Double Stranded RNA
DCR	Drosophila Dicer
PTGS	Post Transcriptional Gene Silencing
HSV	Herpes simplex virus
ICP22	Infected cell protein 22
HCV	Hepatitis C virus
NiV	Nipah virus
EBOV	Ebolavirus
NCBI	National center of Biotechnology Information
Tm	Melting Temperature
I-SCORE	Inhibitory score
GESS	Genome wide enrichment of seed sequence matches
DSSE	Differential stability of siRNA duplex ends
AIS	Average internal stability
PDB	Protein Data Bank
RMSF	Root Mean square fluctuation
RISC	RNA induced silencing complex
RMSD	Root mean square deviation
PIZ	Piwi/Argonaute/Zwille

CHAPTER 1 : INTRODUCTION

1.1 ALPHA-1 ANTITRYPSIN DEFICIENCY:

Alpha-1 antitrypsin deficiency (A1AD) is a hereditary disorder characterized by low levels of a protein called alpha-1 antitrypsin (A1AT) which is found in the blood. This deficiency may predispose an individual to several illnesses and most commonly manifests as chronic obstructive pulmonary disease and liver disease (especially cirrhosis and hepatoma), or more rarely, as a skin condition called panniculitis. A1AD is also more frequent among individuals with what was previously called Wegener's granulomatosis, now called polyangiitis with granulomatosis (GPA). A deficiency of A1AT allows substances that break down proteins to attack various tissues of the body. (Connolly et al., 2018)

1.2 SERPINA 1:

Serine Protease Inhibitor-A1 (SERPINA1) gene encodes for Alpha1-AntiTrypsin (AAT), the second most abundant circulating serum, is a highly polymorphic gene with three noncoding regions and four protein-coding regions. AAT deficiency (AATD) is caused by variations in the SERPINA1 gene sequence, leading to different changes in the AAT protein and pathological implications. The normal AAT protease inhibitor (PI) is designated with the letter M and is encoded by wild-type SERPINA1 sequences. Individuals with biallelic M alleles are designated as PI*MM individuals. The M allele includes benign subvariants, such as M1, M2, M3, and M4, and other benign allelic variants like E, G, and Z. (Rotondo et al., 2020)

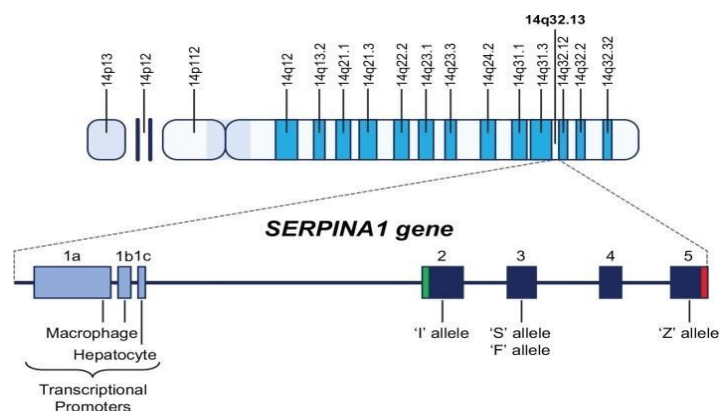


Figure 1.1: SERPINA1 gene.

- Located on the long arm of chromosome 14 at position 14q32.13, the SERPINA1 gene consists of three transcriptional promoters (1a–c) upstream of the transcriptional start codon and four protein-coding regions (2–5). The S and F mutations occur in the third protein-coding region, and the I and Z mutations occur in coding regions 2 and 5, respectively. The green bar in coding region 2 represents the AUG transcriptional start codon and the red bar in coding region 5 represents the TAA stop codon.
- The Z allele is the most common variant associated with severe AATD, with individuals homozygous for it at high risk of lung and liver disease. Other alleles associated with severe AATD include MMalton, MHeerlen, select other rare variants, and the family of null alleles. Some alleles, such as the S variant, are associated with less severe deficiency and intermediate forms of AATD, resulting in reduced serum AAT levels but higher than PI*ZZ levels. A serum AAT level of <57 mg/dL, or <11 µM, conventionally defines AATD, but attention to moderate AAT levels and corresponding genotypes will continue to elucidate the full spectrum of AATD-associated risks and phenotypes
- Nomenclature of SERPINA1 alleles is unconventional, with AAT protein variants originally identified and named based on their electrophoretic properties long before the encoding SERPINA1 gene was identified. Null alleles, which produce no AAT protein, were originally collectively named 'null' (or –) and later individually named with the prefix 'Q0' and place of origin suffix. A wide spectrum of SERPINA1 genetic variants that cause AATD, but currently there is only clear clinical guidance for the management and treatment of the most common (e.g. PI*ZZ) or severe (PI*Znull) forms of the disease.

1.3 THE SPECTRUM OF VARIANTS IN AATD:

- AATD is most commonly associated with single-nucleotide polymorphisms (SNPs); for example, the Z and S allelic variants result from the base pair substitutions c.1096G>A and c.863A>T leading to the amino acid changes p.Glu366Lys and p.Glu288Val, respectively, in the AAT protein. Small changes to the AAT amino acid sequence can affect the structure/folding of the protein and reduce its secretion from hepatocytes, and/or reduce the binding affinity of AAT for neutrophil elastase (NE), with or without protein polymerization. However,

AATD can also be caused by null variants which can arise from missense mutations, nonsense mutations, and complete absence of the SERPINA1 gene.

- The SERPINA1 gene has three transcriptional promoters and four protein-coding regions. Most of the SERPINA1 transcript variants are produced by alternative splicing between the non-coding exons 1A, 1B, and 1C. The 1A transcripts are mainly expressed in leukocytes and lung tissue, while the 1B transcripts are highly restricted to leukocytes. The 1C transcripts are specifically found in lung, liver, kidney, and pancreas. The expression of transcripts is affected by deficient and null variants of the SERPINA1 gene. For example, the S allele and some other SERPINA1 gene variants lead to the production of an abnormal form of alpha-1 antitrypsin that is quickly broken down in the liver. (Foil et al., 2021)
- Based on this, variants of AAT can be categorized into four basic groups:
 - Normal that is associated with normal levels of AAT and normal function. The family of normal alleles is referred to as M, and the normal genotype is MM.
 - Deficient that is associated with plasma AAT levels less than 35% of the average normal level. The most common deficient allele associated with emphysema is the Z allele, which is carried by about two to three percent of the Caucasian population in the United States.
 - Null alleles that lead to no detectable AAT protein in the plasma. Individuals with the null genotype are the least common and are at risk for the most severe form of associated lung disease but not liver disease.
 - Dysfunctional alleles produce a normal quantity of AAT protein, but the protein does not function properly. (Abboud et al., 2011)

1.4 EPIDEMIOLOGY:

1.4.1 Frequency: United States:

- Alpha1-antitrypsin deficiency (AATD) is 1 of the 3 most common lethal genetic diseases among adult white persons, affecting 1 per 3000-5000 individuals. Severe AATD affects an estimated 70,000-100,000 individuals, and approximately 25 million people carry of at least 1 deficient gene. However, less than 10% of severely deficient individuals are currently identified.

1.4.2 International:

- AATD has been identified in all populations, but it is most common in individuals of Northern European (1 in 1600) and Iberian descent. Similar rates are found among white persons worldwide, with an estimated 117 million carriers and 3.4 million affected individuals.

1.4.3 Race:

- White persons constitute an estimated 117 million carriers and 3.4 million affected individuals. Racial groups other than whites are affected less frequently.

1.4.4 Sex:

- Women and men are affected in equal numbers

1.4.5 Age:

- The enzyme deficiency is congenital and has a bimodal distribution with respect to symptoms. It can be seen in neonates as a cause of neonatal jaundice and hepatitis. It can present in infants as cholestatic jaundice and in children as hepatic cirrhosis or liver failure. AATD is also the leading underlying condition requiring liver transplantation in children.
- In adults, AATD leads to chronic liver disease in the fifth decade of life. As a cause of emphysema, it is seen in nonsmokers most commonly in the fifth decade of life and during the fourth decade of life in smokers. (Stoller et al., 2012)

1.5 PATHOPHYSIOLOGY:

Emphysema in AAT deficiency is considered due to an imbalance between neutrophil elastase in the lung. This destroys elastin, and the elastase inhibitor AAT, which protects against proteolytic degradation of elastin. This mechanism is known as a “toxic loss of function.” Specifically, cigarette smoking and infection increase elastase production in the lung, therefore increasing lung degradation. Also, the polymers of “Z” antitrypsin are chemotactic for neutrophils, which may contribute to local inflammation and tissue destruction in the lung. The pathogenesis of the liver disease is different and is called a “toxic gain of function.” The liver disease results from accumulation within the hepatocyte of un-secreted variant AAT protein. Only those genotypes associated with pathologic polymerization of AAT within the

endoplasmic reticulum of hepatocytes produce disease. Most patients are homozygous for the Z allele (i.e., PI*ZZ); liver disease does not occur in null homozygotes who have severe AAT deficiency. (Hersh et al., 2019)

1.6 SYMPTOMS:

The most common symptom of A1AD is shortness of breath after physical activity. This can limit the patient's ability to exercise and can cause wheezing when breathing. Other common symptoms include:

- Unintentional weight loss
- Vision changes
- Fatigue
- Repeated respiratory infections
- Rapid heartbeat when standing
- Barrel-shaped chest

A1AD also can cause other conditions and patients may experience symptoms associated with conditions such as:

- Cirrhosis
- Jaundice
- COPD
- Emphysema

1.7 RISK FACTORS:

Risk factors associated with development of alpha 1-antitrypsin deficiency are as follows:

- First degree relatives of patients with known AAT deficiency are at an increased risk of inheriting the disorder.
- Smoking is risk factor for development of serious disease in patients with AAT deficiency.
- Risk for lung disease also increases with exposure to dust, fumes, or other toxic substances.

1.8 STAGING:

- No specific grading system exists for alpha1-antitrypsin deficiency (AATD), but the severity of the emphysema that it creates can be staged using the body mass index, airflow obstruction, dyspnea, and exercise capacity (BODE) index. This 4-step evaluation of patients with chronic obstructive lung disease appears to identify a population with limited survival who might benefit from intensified therapy. The index has not been evaluated in a population of individuals with AATD. (Patrias et al., 2007)

1.9 DIAGNOSIS:

- Your healthcare provider may test you for AAT deficiency if you have relatives who have AAT deficiency or a lung or liver disease or after you develop a lung or liver disease that is related to the condition.
- A blood test can check the level of AAT protein in your blood. If the level is lower than normal, it is likely that you have AAT deficiency.
- A genetic test is the most certain way to check for AAT deficiency and should be done to confirm the results of the blood test and find the mutation in the AAT gene. A genetic counselor can help you understand what to expect from a genetic test and what your results mean.
- A pulmonary test to see how well your lungs are working may be recommended by your healthcare provider if you have COPD related to AAT deficiency.

1.10 TREATMENT:

- There is no cure for AAT deficiency, but there are treatments to slow the lung damage it causes. If you have emphysema and AAT, standard COPD treatments - bronchodilators, inhaled steroids, antibiotics, oral corticosteroids, regular vaccinations, pulmonary rehabilitation, oxygen therapy, and (for severe cases) surgery — may be given.
- If there is a sign of emphysema, a lifelong treatment is needed which is called an augmentation therapy. This treatment involves raising the AAT protein to acceptable amounts. This helps slow down lung damage. Side effects of this treatment are rare and may include a mild fever, headaches, nausea, and dizziness.

- If you have COPD, you may also need medicines or other treatments. Talk to your healthcare provider about ways to help prevent or delay lung damage, such as quitting smoking and avoiding second hand smoke, dust, or air pollution. (Feitosa et al., 2023)

1.11 GENE SILENCING-RNA INTERFERENCE:

- The RNA interference is the latest practical approach to silence the expression of a targeted gene by means of small interference RNA. Silencing of genes carried out by means of RNAi to control the products of genes at the transcriptional level. By means of RNAi double-stranded RNA suppress the expression of genes through small interfering RNA comprising on 21–22 nucleotide sequence. RNAi is responsible for controlling vital processes of the cell including growth, heterochromatin formation, cell proliferation and tissue differentiation. Because double-stranded siRNA have the ability to silence the genes.
- Each strand of siRNA has phosphate at 5' end and hydroxyl at 3' ends and 2 to 3 nucleotide at 3' overhangs, the exogenous insertion promotes the degradation of RNA by means of siRNA. siRNAs are found in many organisms and in mammalian cells and may be synthetically developed for a specific mRNA target. Rendering small RNAs silencing machinery could be controlled and predicted, and the pathways that have been associated with it show significance in the practice area. The corresponding regulatory mechanisms of siRNAs are, submitted by RNA silencing, showing inhibitory effects on gene expression.
- Gene silencing is a complex process involving various components, including the RNase III family members, Dicer (DCR), and other proteins. The Dicer enzyme, found in worms, flies, fungi, plants, and mammals, digests dsRNA into uniformly sized small RNAs (siRNA) and is involved in the initiation of RNAi. It has four distinct domains, including an amino-terminal helicase domain, dual RNase III motifs, a dsRNA binding domain, and a PAZ domain. The *C. elegans* Dicer homologue, K12H4.8, is the functional ortholog of the *Drosophila* Dicer protein. DCR1 mutants of *C. elegans* showed defects in RNAi of germ line-expressed genes but no effect on the RNAi response of somatic genes. CAF1 has been identified as a Dicer homologue in *A. thaliana*, but it is not involved

in PTGS activity. The crystal structure of the RNase III catalytic domain has led to the model for generation of 23- to 28-mer diced siRNA products. (Jiang et al.,2019)

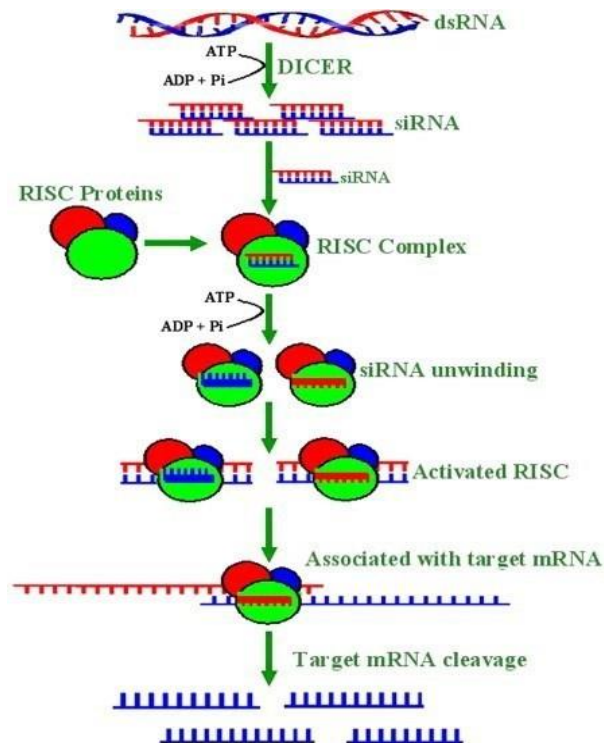


Figure 1. 2 : RNA interference (Ali et al.,2008)

CHAPTER 2 : LITERATURE SURVEY

- Islam, R,et al.,(2023)embarked on an innovative study to utilize in silico methods for predicting siRNAs capable of silencing the SARS-CoV-2 Omicron variant's spike protein, a crucial step in halting viral infection and spread. Their work highlights the successful identification of bioactive compounds in *M. Oleifera* for wound healing. By targeting the spike protein, this research opens new avenues for precise antiviral therapies. This approach underscores the potential of computational biology in accelerating the discovery and optimization of targeted treatments for COVID-19 and its variants.
- Nur SM,et al.,(2013)conducted a study targeting the ICP22 (US1) gene of HSV-1, a key player in viral replication and infection. Utilizing RNA interference technology and computational tools for siRNA design, they developed two effective siRNA molecules against seven different HSV-1 strains. This innovative approach, combining target alignment, secondary structure prediction, and RNA interaction evaluation, offers new avenues for HSV-1 treatment at the genomic level, suggesting potential for the chemical synthesis of antiviral RNA molecules.
- Shawan ,et al.,(2021) investigated the potential of small interfering RNA (siRNA)- based therapy targeting the RNA-dependent RNA polymerase (RdRp) gene of SARS- CoV-2. Utilizing advanced siRNA designing tools, they identified candidate siRNAs with optimal features for efficient gene silencing without off-target effects. Molecular docking studies and molecular dynamics simulations confirmed the stability and efficacy of the selected siRNA molecule suggesting its potential as an effective therapeutic agent to combat COVID-19 outbreaks.
- ElHefnawi ,et al., (2016) demonstrated the effectiveness of an integrated approach for designing small interfering RNAs (siRNAs) targeting the highly conserved regions of the hepatitis C virus (HCV) genome. Through meticulous consideration of various factors such as target RNA variations, thermodynamics, and accessibility, coupled with rigorous in silico analysis and filtration using the MysiRNA-Designer pipeline, two potent siRNAs (HCV353 and HCV258) were identified.

- Sohrab,et al.,(2022) present an innovative approach to combatting COVID-19 through the development and validation of siRNAs targeting the SARS-CoV-2 spike protein. Utilizing in silico prediction and experimental validation, they identified potent siRNAs with no cellular toxicity, paving the way for effective antiviral therapeutics. Notably, the 3rd siRNA exhibited superior antiviral efficiency, underscoring the potential of this strategy in inhibiting SARS-CoV-2 proliferation across multiple cell lines.
- Mahfuz,et al.,(2022)in their study, address the urgent need for effective interventions against the Nipah virus (NiV), a deadly zoonotic pathogen responsible for severe infections in both humans and animals. The researchers explore the potential of small interfering RNAs (siRNAs) as molecular therapeutics. By targeting the nucleocapsid protein gene of NiV, they designed ten prospective siRNAs. These designed siRNAs hold promise in silencing the NiV nucleocapsid gene, offering a potential avenue for mitigating the severity of NiV infections.
- Asiedu et al.,(2020) investigated the therapeutic potential of targeting human gremlin-1, a versatile signaling molecule implicated in various diseases, including cancer. Gremlin-1's role in inducing fibrosis and promoting angiogenesis in organs highlights its significance as a target for cancer therapy. Leveraging RNAi- based approaches, the study employed rigorous siRNA design strategies and target site analysis to identify highly effective candidates with minimal off-target effects. Through computational predictions and comparison with experimentally validated siRNAs, siRNA-6, siRNA-7, and siRNA-47 emerged as promising candidates for gremlin-1 silencing, offering potential avenues for RNAi-based cancer therapy.
- Gopaluni et al.,(2015)review highlights the potential of RNA interference (RNAi) using small interfering RNA (siRNA) to target the L gene of Zaire ebolavirus (EBOV), essential for RNA polymerase function. In silico design of siRNA guided by hybridization thermodynamics demonstrates a computational approach to identify effective antiviral strategies, offering hope for the development of novel therapies against this deadly pathogen.
- Asokan et al.,(2012) conducted a study to investigate sequence polymorphism in five crucial genes of the fruit borer, *Helicoverpa armigera*, to facilitate RNA interference (RNAi) applications in pest management. By identifying off-target

minimized regions and analyzing nucleotide variations, the study aimed to design common siRNAs for each target gene. This approach could enhance the efficacy of RNAi-based strategies by overcoming potential limitations posed by sequence variations in target pest populations.

- The papers discussed primarily delve into the realm of viral diseases, offering comprehensive insights and methodologies into the application of RNA interference (RNAi) for combating viral infections. However, beyond their immediate scope, these studies shed light on the broader applicability of in-silico siRNA design techniques. By extrapolating the principles and strategies elucidated in these papers, it is possible to harness the power of RNAi in addressing genetic disorders, thereby showcasing the versatility and immense potential of in-silico siRNA design in the field of biomedicine. These findings underscore the importance of interdisciplinary approaches in biomedical research and highlight the cross-disciplinary transferability of innovative methodologies and concepts.

CHAPTER 3 : AIM AND SCOPE

3.1 AIM:

The aim of this study is to design a computationally optimized siRNA molecule that can effectively silence the SERPINA1 gene, which is associated with Alpha-1 Antitrypsin Deficiency (AIATD), through an in-silico approach.

3.2 SCOPE:

This study pioneers the design of siRNA molecules to silence a gene implicated in genetic diseases, notably alpha 1 antitrypsin deficiency (AATD). Despite strides in diagnostic modalities and therapeutic options like intravenous augmentation therapy, AATD remains profoundly underdiagnosed, attributable in part to its rarity and the logistical hurdles of conducting extensive clinical trials. While previous investigations predominantly centered on viral ailments, this research leads the way in adapting a systematic computational framework to target the SERPINA 1 gene associated with AATD. The study utilizes a comprehensive suite of computational tools encompassing siRNA design algorithms, RNA structure prediction models, and protein modeling techniques. Through meticulous analysis of siRNA sequences predicted from the target gene's mRNA, the study endeavors to refine two siRNA candidates meeting stringent criteria for efficacy and specificity. By championing this innovative approach in the realm of genetic disorders, the research aims to catalyze the development of tailored precision medicine strategies specifically aimed at addressing the intricacies of AATD.

CHAPTER 4 : MATERIALS & METHODS

4.1 SERPINA 1 SEQUENCE RETRIEVAL AND PREDICTION OF siRNA

The coding sequence of the Human SERPINA1 was retrieved from the National Center for Biotechnology Information (NCBI) protein portal (<http://www.ncbi.nlm.nih.gov/>). The entire sequence was used for designing functional and target-specific siRNAs, using the siDirect 2.0 web server (<http://siDirect2.RNAi.jp/>) which is an updated version of the web-based software siDirect. The study of Naito et al has concluded that when the candidate functional siRNAs could form seed-target duplexes with T_m values below 21.5°C, and their 19-nt regions spanning positions 2 to 20 of both strands have at least 2 mismatches to any other nontargeted transcripts, siDirect 2.0 can design at least one qualified siRNA for >94% of human mRNA sequences in RefSeq. Therefore, the maximum melting temperature (T_m value) for the formation of the seed duplex was kept below 21.5°C (the default) for reducing off-target effects and guanine-cytosine (GC) content was set between 30% and 64% (Birmingham et al.,2007) In the server, 3 rules including Ui-Tei (Ui-Tei et al.,2004), Amarzguioui (Amarzguioui et al.,2004) and Reynolds (Reynolds et al.,2004) based algorithms were selected as parameters for siRNA selection.

TABLE 1.

Algorithms/rules applied in siDirect 2.0 server for rational designing of siRNA molecules.

Uei -Tei rules	Reynolds rules	Amarzguioui rules
According to Ui-Tei rules, highly effective RNA interference can occur in mammalian cells if siRNA satisfying the 4 following sequence conditions at the same time is used.	According to Reynolds rules, a score of 6 out of an 8-component algorithm defines the cutoff for selecting siRNAs. The 8 criteria are the following conditions.	According to Amarzguioui rules, a score of 3 represents the lower-end cut-off for selecting siRNAs. The algorithm incorporates the following features.
1. The presence of A/U base at the 5' end of the antisense strand	1. 30%-52% of G/C content	1. A/U content differential between the terminal 3

		nucleotides of both ends
2. Presence of G/C at the 5'end of the sense strand	2. The presence of at least 3 A or U bases at positions 15-19	2. The absence of U at position 1 of the sense strand
3. At least 5 A/U residues in the 5' terminal one-third of the antisense strand	3. The absence of internal repeats	3. The presence of A at position 6 of the sense strand
4. Absence of any GC stretch of more than 9 nucleotides in length	4. The presence of an A base at position 19 of the sense strand	4. Strong attachment of 5' sense strand
	5. The presence of an A base at position 3 of the sense strand	5. Weak binding of 3' sense strand
	6. The presence of a U base at position 10 of the sense strand	6. The absence of G at position 19 of the sense strand
	7. The absence of G or C at position 19 of the sense strand	7. The presence of G or C at position 1 of the sense strand
	8. The absence of G at position 13 of the sense strand	8. The presence of A or U at position 1 of the sense strand

The predicted siRNAs were cross-validated by the i-Score (inhibitory-Score) Designer server in which siRNA prediction scores were calculated from 9 different algorithms including both first- generation algorithms and second-generation algorithms. The second-generation algorithms include i-Score (Ichihara et al.,2007), s-Biopredsi (Huesken et al.,2005), and DSIR(Vert et al.,2006) which are considered more reliable than the other algorithms. In this server, the siRNAs predicted from siDirect were used as input and those which ranked in the top 10 predictions made by any of the following algorithms (s-Biopredsi, i-Score, and DSIR) in i-Score Designer analysis, were selected for further analysis.

In this server, the GC content of individual siRNAs was also given as part of the prediction results. Afterwards, the GC content of each of the candidate siRNAs was calculated in OligoCalc.(Kibbe et al.,2007)

4.2 OFF-TARGET SIMILARITY SEARCH AND FILTERING SIRNAS WITH PALINDROMIC SEQUENCE

- Both sense and antisense strands of the siRNAs should avoid any secondary structures like palindrome that can be prohibitive to the formation of the duplex properly (Fakhr et al.,2016). Therefore, the predicted siRNA sequences were screened for palindromic sequences using the EMBOSS palindrome online tool in which the minimum length of the palindrome was set at 4 nucleotides. Next, both strands of the via blastN at the NCBI database against the Refseq-RNA database of humans to reduce the risk of silencing unintended genes. While the parameters were adjusted for short-input sequences, word size was set at 7, and the expected threshold was set at 100.
- The guide and passenger strands of the siRNAs were also screened by the Genome-wide Enrichment of Seed Sequence matches (GESS) online tool (www.flyrnai.org/gess/) (Yilmazel et al.,2014) to further check for any potential off-target effects. The seed regions of the siRNAs were analyzed to identify potential off-targeted transcripts that could produce false positive results. The input file contained only active RNAi reagents, the region was selected 3' UTR and the reference database was the built-in human database.

4.3 SECONDARY STRUCTURE PREDICTION AND THERMODYNAMICS STUDY OF THE SIRNA-TARGET GENE INTERACTION

- The function of a regulatory RNA molecule is determined by its structure mainly. RNA tertiary interactions involve secondary structure elements and are substantially weaker than secondary interactions. Thus, to a large extent, the free energies in the secondary structure represent the thermodynamics of RNA folding (Ding et al.,2014). In this study, these secondary structures of the guide strand of the siRNAs and their target site as well as their free energy of folding were determined using MaxExpect program (Lu et al.,2009) of the RNA structure webserver.

Here, the RNAi structures with higher values of energy were considered better candidates as those molecules had less chance of folding.

- The thermodynamics of RNA-RNA interaction is another major parameter for siRNA efficiency where stronger interaction denoted by lower binding energy indicates better efficacy. For this, the interaction between the guide strand and the target sequence was analyzed by the DuplexFold program of the RNAstructure web server in which 2 RNA sequences are folded into their lowest hybrid free energy conformation along with their free energy of binding being calculated.
- Afterwards heat capacity and concentration plots for the predicted sirnas were generated using m fold server (Markham et al.,2005).In the heat capacity plot, the collective heat capacity of the RNA duplex (C_p) is plotted as a function of temperature, and the melting temperature, T_m (C_p) is denoted by the local maximum of the heat capacity curve.In the concentration plot, the relative concentration of 3 species is plotted as a function of temperature, and T_m (conc) indicates the point at which the concentration of a double-stranded molecule is one-half of its maximum value.

4.4 PREDICTION OF SIRNA TARGET SITE ACCESSIBILITY WITHIN TARGET MRNA

The siRNA application module of S-fold (<https://sfold.wadsworth.org/cgi-bin/sirna.pl>) was applied to determine the target site accessibility of the predicted siRNA molecules. The target sequences of the siRNAs were given as input and the accessibility of target sites within the target mRNA for specific siRNAs was then assessed by the server through multiple parameters. Several siRNA parameters are being provided by S-fold, including target site accessibility score, differential stability of siRNA duplex ends (DSSE), average internal stability at the cleavage site (AIS), total stability of the siRNA duplex, and so on.

4.5 MODELING AND DOCKING OF ALPHA 1 ANTITRYPSIN PROTEIN AND THE SIRNAS

- The 3-dimensional structure of argonaute 2 (Ago2) protein was predicted using SWISS- MODEL <https://swissmodel.expasy.org/interactive> (WaterhouseA et al.,2018) based on the crystal structure of RNA- bound human AGO2 protein (Protein Data Bank [PDB] identifier: 4Z4D) via homology modeling. The model was further refined by the GalaxyRefine (<http://galaxy.seoklab.org/refine>) server (Heo L et al.,2013) which

can improve structure quality beyond the scope of template information, through MD simulation. Next, the quality of the models was checked using ERRAT(Colovos et al,1993) based on the quality scores from ERRAT and GalaxyRefine, and one of the refined models was finally selected.

- On the contrary, the 3D structure of the guide RNA was modeled RNA composer in which the RNA structure bound to the Ago2 protein (PDB identifier: 4Z4D) was taken as a template. Finally, molecular docking between the Ago2 and the siRNAs was performed in the HDock server (<http://hdock.phys.hust.edu.cn/>) which is a hybrid docking algorithm employing both free docking and template-based modeling with high efficacy and robustness(Yan et al.,2017). The resulting complexes with the lowest docking energy scores and closest siRNA binding position to the template were finally selected and visualized for further analysis

4.6 MD SIMULATION OF siRNA (GUIDE STRAND) - A1AT PROTEIN DOCKED COMPLEX:

The study used MD simulation to study the stability of siRNA-A1AT protein docked complexes. The simulation was performed using the NAMD simulation in Discovery studio, the simulations were evaluated using RMSF analysis using Cabs Flex server.

CHAPTER 5 : RESULTS AND DISCUSSION

5.1 SIDIRECT 2.0 AND I-SCORE DESIGNER PREDICTS siRNAs FROM THE RETRIEVED SERPINA 1 TARGET SEQUENCE.

The coding sequence of the human SERPINA 1 was retrieved from the NCBI-Genbank database, and it consists of 3006 nucleotides. siDirect 2.0 predicted in total 46 siRNAs based on the algorithms of Ui-Tei, Reynolds, and Amarzguioui as well as seed-duplex stability. Out of these 46 siRNAs, 7 followed all 3 rules (URA). To cross-check, the predicted siRNAs were also analyzed by i-Score Designer, and 5 of them were found to rank among the top 10 predictions made by s-Biopredsi, i-Score, or DSIR prediction tools and so these siRNAs (Table 2) were selected for further filtration.

TABLE 2.

siDirect and i-Score designer predicted overlapped siRNAs along with their different parameters

Target position	Target sequence	Guide strand	Passenger strand	Seed-duplex stability T _m (°C)		GC content
				Guide strand	Passenger strand	
935-957	CACCAAGTTCCT GGAAAATGAAG	UCAUUUCCAG GAACUUGGUG	CCAAGUCCUG GAAAAUGAAG	7.4 °C	19.2 °C	36.8%
504-526	AAGTTTTTGGAG G ATGTTAAAAA	UUUAAAUCCU CCAAAAACUU	GUUUUUGGAGG AUGUUAAAAA	6.9 °C	11.5 °C	31.6
2016- 2038	CTCACATATCTC C GTTTAGAATC	UUCUAAACGGA GAUAUGUGAG	CACAUUCUCC GUUUAGAAUC	14.6 °C	16.0 °C	42.1
784-806	AGCGTTTAGGCA TGTTTAACATC	UGUUAACAUG CCUAAACGCU	CGUUUAGGCAU GUUUAACAUC	8.2 °C	18.6 °C	42.1
575-597	GGCCAAGAAACA GATCAACGATT	UCGUUGAUCUG UUUCUUGGCC	CCAAGAAACAGA UCAACGAUU	21.1 °C	12.2 °C	47.4

5.2 FILTRATION OF SIRNAS FOR THE PRESENCE OF PALINDROMES, OFF-TARGET EFFECTS, AND IMMUNOSTIMULATORY ACTION RESULTS IN 3 POTENTIAL SIRNAS

All 5 did not have any palindromic sequences in their guide and passenger strands. These 5 siRNAs were blasted against the reference RNA database of humans, and all of them showed higher than 78% query coverage with other genes which are considered tolerable. For further confirming this observation, when these RNAs were analyzed by GESS, the results showed that the off-targets were not significant by any of the tests including Fisher, Bonferroni, Bonferroni step-down (Holm) correction, and Benjamini and Hochberg methods. It suggests that these siRNAs would not create undesirable silencing of any other genes.

5.3 SECONDARY STRUCTURE AND THERMODYNAMICS STUDY OF THE CANDIDATE SIRNAS

The predicted secondary structures of the guide siRNA strands are presented in (Figure 1). Their free energy of folding ranged between 1.75 to 1.7 kcal/mol (Table 3). All of the candidate siRNAs displayed promising outcomes, which is consistent with the theory that siRNAs with positive free energies of folding may encourage increased permissibility to the target, improving gene silencing. On the contrary, the binding between the siRNAs and their target gene was also analyzed (Figure 2), and the resulting binding energy ranged between -33.4 and -36.0 kcal/mol .

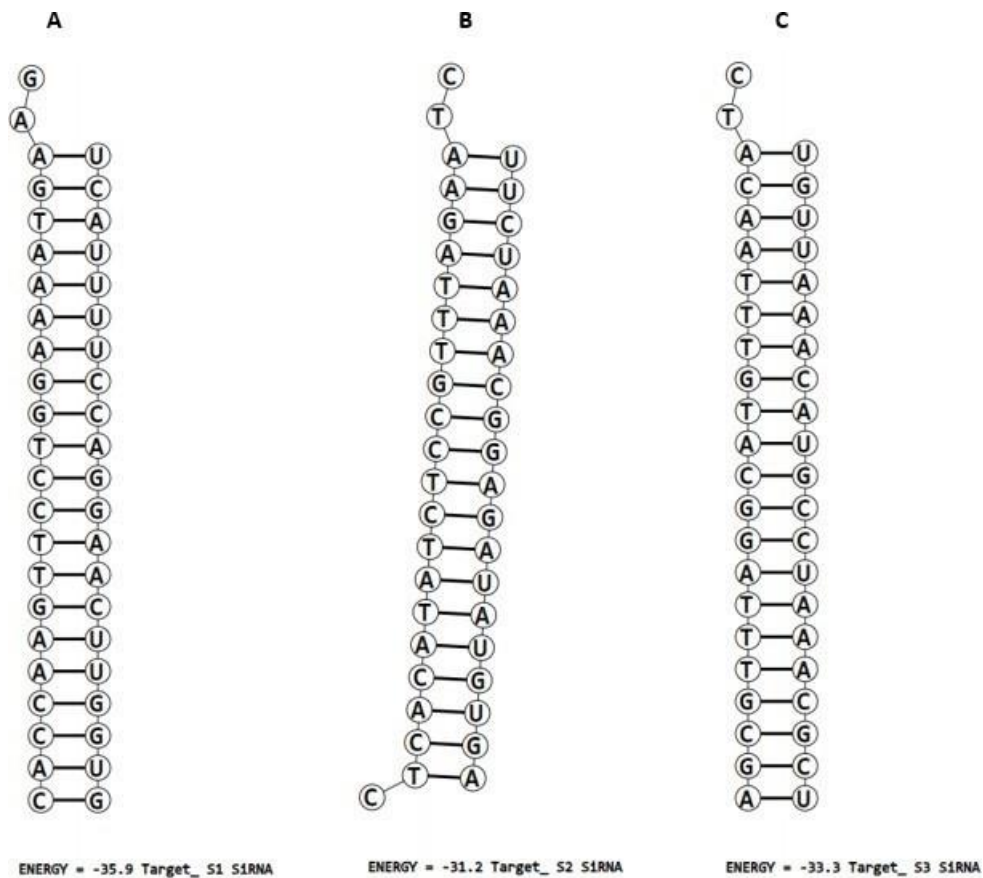


Figure 4 : Structures of binding of siRNA (guide strand) and target RNA along with their predicted free energy of binding

Structure of binding of siRNA (guide strand) and target RNA along with their predicted free energy of binding: (A) structure of interaction between guide strand and target sequence for S1 siRNA, (B) structure of interaction between guide strand and target sequence for S2 siRNA, and (C) structure of interaction between guide strand and target sequence for S3 siRNA. siRNA indicates small interfering RNA.

- Using the DINAMelt web server (Markham et al., 2005), the heat capacity plot and concentration plot were also calculated for the RNAi molecules. As temperature increases, the heat capacity of the siRNAs also increases, and the temperature at which the collective heat capacity (C_p) reaches its maximum value is denoted as the melting temperature, T_m (C_p). In the concentration plot, the concentration of the double-stranded molecule decreases with increasing temperature, and the concentration of folded and unfolded single strands increases with increasing temperature.

- The higher values of these 2 melting temperatures can be considered as an indication of higher efficacy of the siRNAs. From **Table 3**, it can be seen that among the 3 siRNAs, S1 siRNA showed the highest T_m (C_p) and T_m (Conc) values (75.2°C and 45.6°C, respectively). Therefore, these 3 siRNAs were finally selected as potentially most effective siRNAs which were then subjected to further analysis.

TABLE 3.

siRNA candidates screened based on Free energy, Binding, T_m (Conc) (°C), T_m (C_p) (°C)

Alias and guide strand sequence	Free energy Of folding (kcal/mol)	Free energy Of binding (kcal/mol)	T _m (Con) (°C)	T _m (C _p) (°C)
S1 (UCAUUUCCAGGAACUUGGUG)	1.8	-35.9	75.2	45.6
S2 (UUCUAAACGGAGAU AUGUGAG)	1.5	-31.2	47.7	43.0
S3 (UGUUAACAUGCCUAAACGCU)	1.7	-33.3	50.8	44.2

5.4 TARGET SITE ACCESSIBILITY AND FURTHER RNA DUPLEX THERMODYNAMICS STUDY OF THE siRNAs

- The parameters obtained from Sfold output for the 3 finally selected siRNAs are given in Table 4. Through multiple experimental approaches, the significance of target accessibility for the function of siRNAs has been proven (Tian et al., 2021). In this study from Sfold analyses, it was found that S1 siRNA had a target accessibility score of 3, whereas S2 had a score of 4 and S3 had a target accessibility score of 6.
- It has also been stated that functional siRNA duplexes tend to have lower stability at the cleavage site as well as on the 5'-antisense end than on the 5'-sense end and the relative stability of the 2 ends influences the degree to which each strand participates in the gene silencing pathway (Meister et al., 2004). This difference is shown by the DSSE value which should be more than 0 kcal/mol as the guide strand gets assembled

with RNA-induced silencing complex (RISC). The results demonstrate that the values for the proposed siRNAs ranged between 1 and 3 kcal/mol, demonstrating that ineffective and undesirable mRNA targeting that can occur from the assembly of passenger strands with RISC is avoided. Another score which is siRNA duplex feature score for S1 was 9, S2 and S3 was found to be 7 which can be as low as -2 and as high as 10 and values greater than 6 are acceptable.

- Results from Sfold also showed that the AIS values (AIS at the cleavage site) for all the predicted siRNAs were around -7.2 kcal/mol and according to cleavage site instability rule this value should be greater than -8.6 kcal/mol. Average internal stability computes the average internal energy for positions 9 to 14 in the antisense strand. In addition, the duplex thermodynamics score of the 3 siRNAs was found to be 2 in which 1 point was attributed to their DSSE values being greater than 0 kcal/mol, and another point was due to their AIS values being greater than -8.6 kcal/mol. Finally, the total score for these siRNA duplexes were found between 13 and 15 which surpassed the Sfold-imposed threshold of 12. This score sums up target accessibility score, duplex feature score, and duplex thermodynamics score and its maximum value can be 20. It can be seen that all Sfold filtration criteria have been fulfilled by these 3 siRNAs.

TABLE 4.

Sirna result of the predicted siRNAs from Sfold server showing target accessibility and various RNA duplex thermodynamics

Si RNA	Total score for siRNA duplex	Target accessibility score	Duplex feature score	Duplex thermodynamics score	DSSE kcal / mol	(AIS) kcal/mol	Total stability of siRNA duplex Kcal/mol
S1	14	3	9	2	2.4	-8.1	-36.1
S2	15	6	7	2	1.3	-7.2	-33.9
S3	13	4	7	2	0.4	-8.2	-33.0

5.5 MOLECULAR DOCKING BETWEEN AGO-2 PROTEIN AND THE siRNAs

- The predicted 3D structure of Ago2 protein was refined by GalaxyRefine server which generated 5 models and calculated multiple parameters for each of the model through which the quality of the models could be assessed including MolProbity score, clash score, poor rotamers, Ramachandran favored region, and so on. In addition, the models were checked by ERRAT. All the results are shown in Table 5.
- Clash score represents the number of atom pairs in a protein model that are unusually close to each other and may result in steric clashes and therefore lower numbers are considered better. Poor rotamers indicate the number of amino acids in the protein with poor side chain geometry and a high percentage of poor rotamers (>1.5%) can be concerning. Another important factor is Ramachandran favored region within the protein which refers to the percentage of amino acids having favored phi/psi angles in the protein backbone. Ideally, this number should be higher than 98%. The next value is MolProbity score which combines all scores regarding atom contacts, geometry, and conformation into a single value to validate structure quality and once again comparatively lower values are desired.
- Finally, the models were also run through ERRAT program which determines whether the distribution of nonbonded atomic interactions in a candidate structure correlates with the distribution established from a database of reliable, high-resolution structures. Thus, ERRAT can be considered an overall quality factor for nonbonded interactions, with higher scores suggesting higher quality. From the results shown in Table 5, it can be seen that model 4 had highest ERRAT score (93.89%), closest MolProbity score (1.58) to the original model and lowest clash score (11.9) among the 5 models. In addition, it showed less than 1.5% poor rotamers (0.4) and greater than 98% residues in Ramachandran favored region. Therefore, model 4 of Ago2 protein structure was selected for docking with the candidate siRNAs.

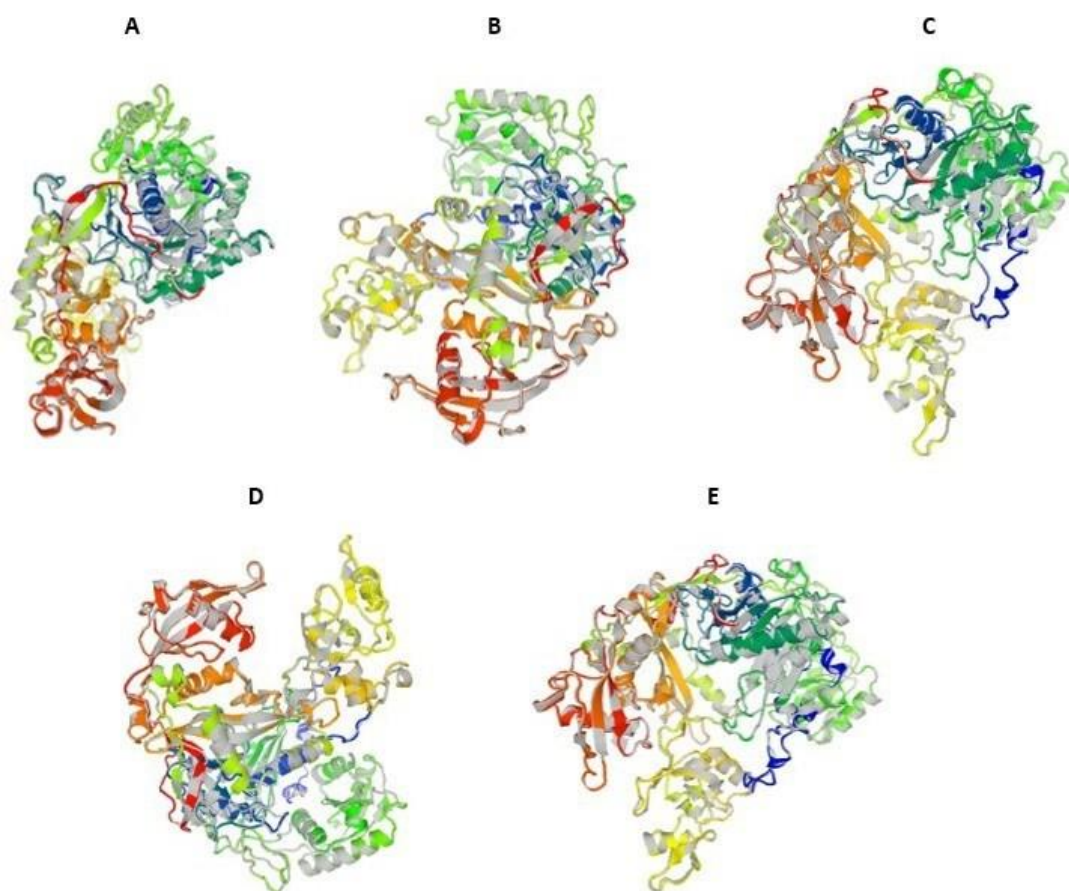


Figure 5. The top-five ranking models of AGO2 predicted from Galaxy refine tool.

The five top-ranking models are shown in static images. The structure changes relative to the initial model in terms of RMSD and MolProbity score are presented in a separate table. Three components of the MolProbity score, namely, the number of atomic clashes per 1000 atoms, the percentages of rotamer outliers and Ramachandran favored backbone torsion angles, are also reported in the Table 5.

TABLE 5.

Top five ranking models of AGO2 predicted by Galaxy Refine tool and ERRAT

Model	MolProbity	Clash score	Poor rotamers	Rama favored	ERRAT
Initial	2.152	5.8	4.3	94.9	
MODEL 1	1.586	11.8	0.1	98.1	91.354
MODEL 2	1.653	14.0	0.1	98.1	91.667
MODEL 3	1.582	11.7	0.1	98.2	91.5935

MODEL 4	1.589	11.9	0.4	98.2	93.8903
MODEL 5	1.610	12.6	0.5	98.3	92.6434

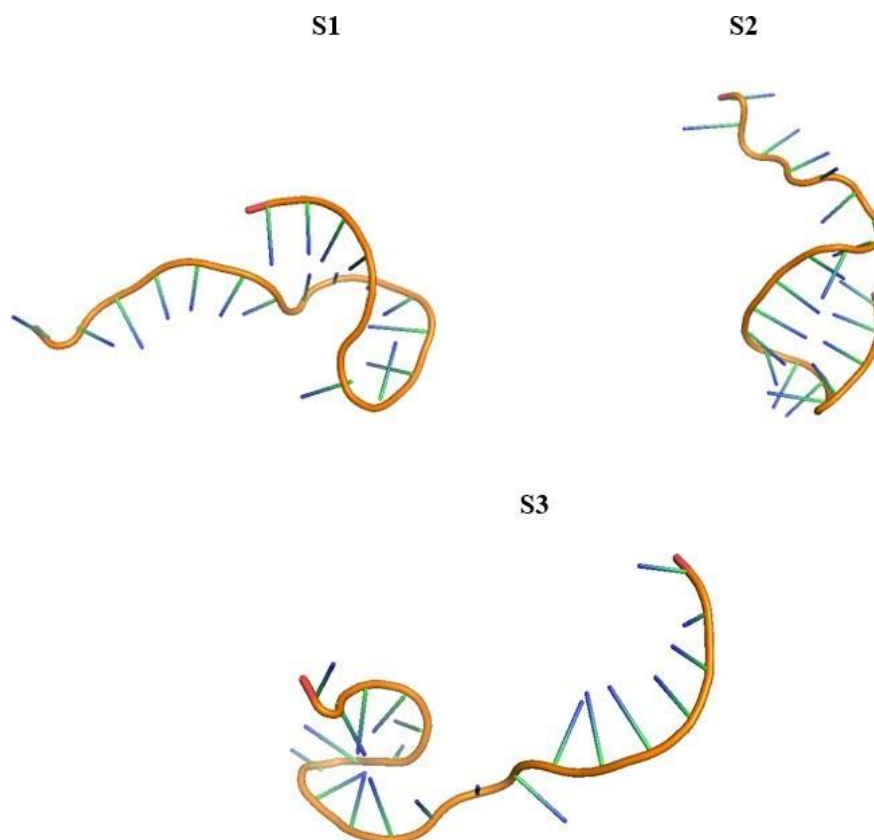


Figure 6: S1,S2 AND S3 SiRNAs modelled using RNA composer

The docked complexes between Ago2 protein and our candidate siRNAs were screened to find out in which complexes the siRNA was placed inside the PAZ and MID domains of the Ago2 protein thus maintaining a similar conformation to the PDB 4z4d structure. It was found that for S2 and S3 siRNAs, only one docked complex of each docking analysis showed the involvement of PAZ and MID domains with the siRNAs. On the contrary, for S2 siRNA 2 docked complexes had similar conformation to the 4z4d structure and based on the docking energy score, the complex with the lower docking energy was selected. The docking score, interaction statistics and interactive residues of each of the siRNA docked complexes are presented in Table 6 and the docked complexes are shown in Figure 4.

TABLE 6.

Docking energy scores and Interactive residues predicted using HDOCK and Pymol

siRNA in Ago2- siRNA complex	HDOCK docking energy score (kcal/mol)	Interactive residues within PAZ and MID domain of Ago2 protein
S1	-288.62	PAZ: VAL70, ARG69, ASN43, ALA322, PHE44, ALA42, PHE45, PRO50, ILE49, ASP48, MET46, GLU46, GLN41, LEU40, LYS39, ILE38, THR37, ARG36, GLY35, MID: PRO107, ILE108, GLY109, LEU106, PRO105, MET104, ALA109, LYS124, GLY121, PRO120, ASP125, VAL117, PHE128, LYS129, SER131, LYS129, SER131, LYS112, ARG150, PRO155, PHE156, GLU167, VAL135, LEU140, SER121
S2	-303.25	PAZ: ARG36, GLY35, SER34, THR33, GLY32, PHE31, ASP30, PRO29, ARG28, PRO27, PRO26, PRO25, ALA22, LYS24, PHE23, THR37, ILE38, LYS39, PRO63, VAL75, GLN35, PHE87. MID: PRO152, GLY149, ARG150, PRO107, PRO120, GLY121, GLU122, ARG126, GLY123, GLY178, GLY187, ASP111, SER153, PRO152, LEU151, ARG150, ILE108, ALA184, ASN190, PRO191, ARG170, MET166
S3	-308.49	PAZ: THR84, LYS73, LYS62, PRO63, LYS65, ILE54, ASP89, PHE89, LYS91, ARG90, ILE61, CYS66, VAL.93, MET47, PRO50, ARG36, PRO29, PRO26, ALA22, PHE23, ASP30, ILE52. MID: ASP125, GLU122, ILE127, LEU115, VAL113, SER131, ILE132, LYS112, THR163, THR173, LEU169, ARG179, GLY178, GLU97, PHE200, MET166.

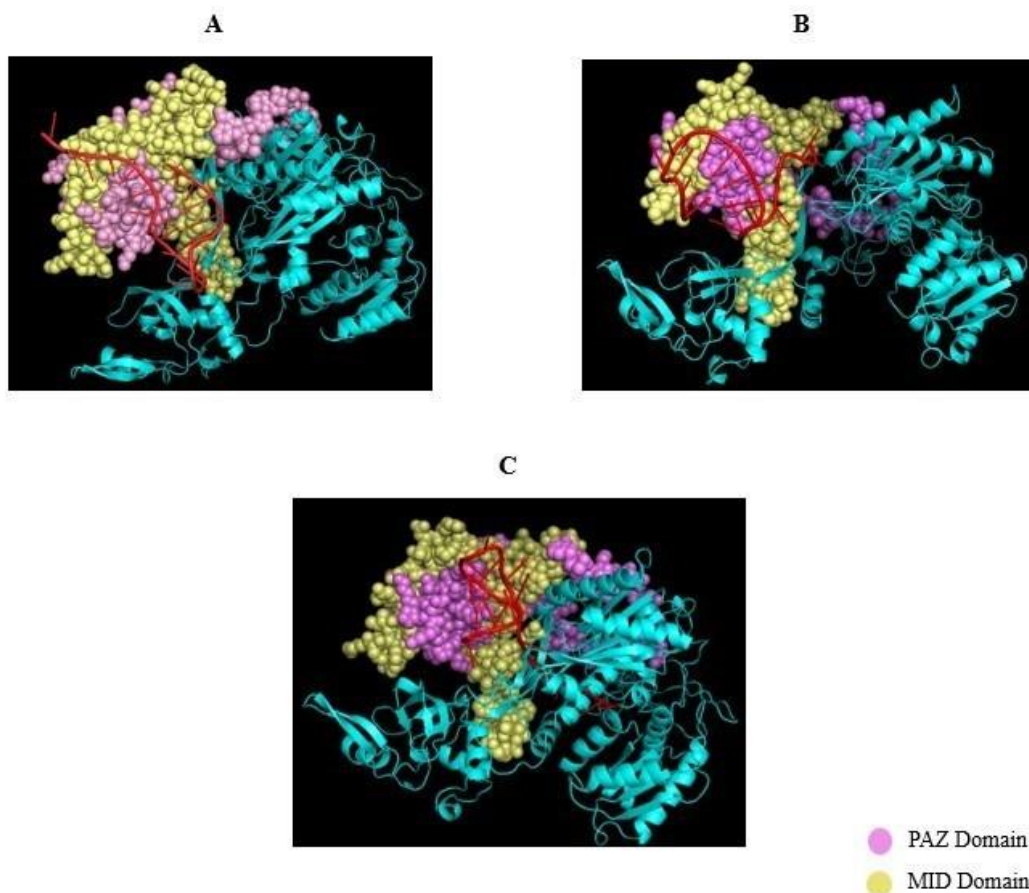


Figure 7. Docked structures siRNAs

Docked structures of S1, S2, and S3 siRNAs (cartoon view) with Ago2 proteins. (A) S1 siRNA docked with Ago2 protein (B) S2 siRNA docked with Ago2 protein and (C) S3 siRNA docked with Ago2 protein. The PIWI/Argonaute/Zwille (PAZ) domain and the Middle (MID) domain are colored as purple and Pale yellow spheres, respectively, and the rest of the protein is denoted by blue color ribbon structure. The siRNAs are represented by red in colour (cartoon view).

5.6 MD SIMULATION OF THE SIRNA-AGO2 DOCKED COMPLEXES

S1,S2 and S3 docked complexes were examined through MD simulation.NAMD Simulation was conducted Using Discovery studio visualizer to observe the deviation of structural dynamics between the S1,S2 and S3 siRNA-Ago2 docked complexes in terms of RMSF analysis.RMSF represents the portions of a protein structure that are fluctuating from their mean structure by analyzing the average deviation of each protein residue over a period of time from a reference position.

The RMSF graphs are shown in Figure 8 , and it can be seen that S2 docked complex was comparatively more stable than S1 docked complex as in S2 complex the RMSF values only reached 4.8 Å and in S1 complex, the RMSF values went up to almost 5 Å.

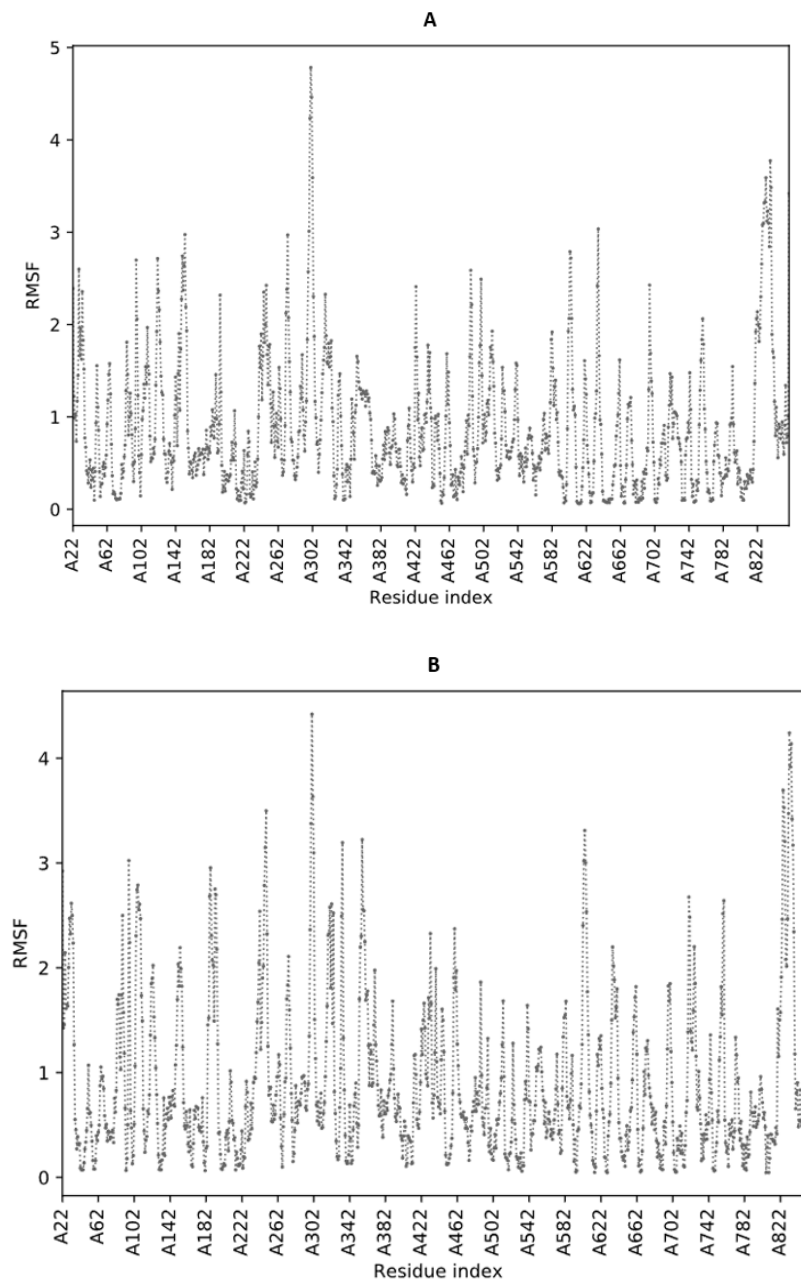


Figure 8: RMSF of each residue of the Ago2 protein in siRNA-Ago2 complexes (A) RMSF plot of S1 siRNA-Ago2 complex and (B) RMSF plot of S2 siRNA-Ago2 complex.

DISCUSSION

- RNA interference is a remarkable endogenous regulatory pathway that can bring about sequence-specific gene silencing. If harnessed effectively, RNAi could result in a targeted therapeutic approach with applications ranging from viral diseases to cancer that are particularly hard to address with existing drugs. The pathways are guided by small RNAs including small interfering RNAs (siRNAs). In addition to translational inhibition and transcript degradation of target mRNAs that are poorly complementary, these effector molecules also guide sequence-specific cleavage of perfectly complementary mRNAs. Theoretically, they can be designed for any gene of interest based on its mRNA sequence alone.
- Moreover, by using nanoparticles like PEI and other approaches, they can be introduced into the cell for specific knockdown of a gene of interest. Such unlimited potential has particularly made RNAi a favorite gene knockdown strategy in mammalian cells. In this study, an attempt was made through computational methods to rationally design potential siRNA molecules against SERPINA 1 in humans which could lead to new treatment strategies for Alpha 1 antitrypsin deficiency.
- SERPINA1, the gene responsible for encoding Alpha-1 Antitrypsin (AAT), emerges as a pivotal target for therapeutic strategies aimed at addressing the pathophysiology of A1ATD. Mutations in SERPINA1 lead to the production of dysfunctional AAT, resulting in diminished AAT activity in the lungs and the accumulation of misfolded AAT in the liver, contributing to the disease's manifestations. Current therapeutic approaches for A1ATD primarily focus on managing lung disease through serum protein replacement therapies. However, the treatment landscape for liver disease associated with A1ATD remains limited. In this context, exploring novel therapeutic avenues targeting SERPINA1 is crucial for developing effective interventions that can mitigate the impact of A1ATD on both lung and liver function. One promising strategy under investigation is mRNA therapy directed at enhancing SERPINA1 expression. Studies have demonstrated the potential of mRNA therapy to increase SerpinA1 protein levels in AAT-deficient patient cells, suggesting a promising avenue for addressing the underlying molecular defects in A1ATD.

- By targeting SERPINA1, researchers aim to advance precision medicine approaches that could lead to improved outcomes and quality of life for individuals affected by A1ATD, particularly focusing on mitigating lung and liver complications associated with the condition.
- In this study, to narrow the search space and improve prediction accuracy, various criteria for predicting and filtering potential siRNA molecules were applied. For instance, siDirect 2.0 predicts off-target minimized siRNA based on an algorithm that takes into consideration the fact that the capability of siRNA to induce off-target effect is highly correlated to the thermodynamic stability of its seed-target duplex. For this, the server first selected highly functional siRNA sequences that follow Ui-Tei, Amarzguioui and Reynolds rules. For instance they had A/U nucleotides at the 5' end of the guide strand; G/C nucleotides at the 5' end of the passenger strand, and at least 4 A/U residues in the 5' terminal 7 base pair of the guide strand. After that, siRNAs whose seed-target T_m was below 21.5°C for both guide and passenger strands, were regarded almost off-target-free seed sequences. Finally, these siRNAs were checked whether they had at least 2 mismatches to any other nontargeted but near-perfect match transcripts, thus further ensuring that no nontargeted gene silencing can take place. Through these selection steps, siDirect generated and filtrated potential siRNAs.
- While these first-generation algorithms heavily depend on base preferences at specific positions, thermodynamic stability, target mRNA secondary structure and several other siRNA features, they can be sometimes not very effective due to their small dataset coverage. However, with the introduction of Huesken's data set that included 2431 experimentally validated siRNAs, several second-generation algorithms came into the scenario that were based on machine learning models and considered more reliable. Three algorithms were included in this work because they were thought to be the most efficient: s-Biopredsi based on artificial neural network, DSIRand i-score based on linear regression models. Altogether 5 siRNAs were found to give good results in both generations of algorithms and hence predicted to be more efficient during our study.
- The GC content of siRNAs is a significant parameter which influences siRNA functionality to a large extent. It has been argued that a high GC content may be a factor that negatively affects the functionality of an siRNA, inhibiting the unwinding of

the siRNA duplex, which is necessary for RISC loading (Liu Q et al.,2012).It might also be associated with a prohibitive secondary structure of the target mRNA. On the contrary, several studies have also indicated that a low GC content may lead to decreased functionality, presumably due to reduced target affinity and specificity of the siRNA(Liu Q et al.,2012). As a result, different ranges of GC content have been suggested by different algorithms such as 31.6% to 57.9% and 36% to 52%.In this work, all the candidate siRNAs showed between 33% and 43% of GC content.

- Although a low T_m was set to minimize seed-dependent off-target effects, siRNAs designed by different algorithms might still induce off-target effects. These small RNAs can tolerate several mismatches at the mRNA target and thus can silence those targets with imperfect complementarity. Also as they are nearly identical to the related class of miRNAs, they can recognize and degrade mRNAs that are complementary to their seed region and the rest of the nucleotides' complementarity is insignificant. Therefore, the predicted siRNAs were further checked through BlastN and GESS online tool to ensure their target specificity. BlastN revealed that all 5 of the siRNAs higher than 80% query coverage with SERPINA 1gene and However, because no 3'UTR regions were discovered to be over-represented for seed region matches, GESS analysis revealed that none of the off-targeted transcripts were significant using the Fisher exact test and Yates chi-square test. On the contrary, the presence of palindromic motifs within the guide strands were checked, again all 5 siRNAs followed the required criteria.
- RNA secondary structure significantly influences the efficiency of a siRNA. Self-structure in either the siRNA sequence or the target mRNA at the binding site may prevent gene silencing by inhibiting RISC-mediated cleavage of target. Therefore, predicting secondary structure of the siRNAs and determining free energy of corresponding folding are crucial for designing effective siRNAs and were conducted in this study. Free-energy minimization approach to RNA structure prediction assumes that the lowest free-energy structure will form. All siRNA molecules were discovered to have positive free energies for folding the guide strands (more than 1.5 kcal/mol), which increased the likelihood that the siRNAs would interact with their targets since they would be less likely to form secondary structures (Chowdury UF et al.,2021)To further estimate the efficiency of the predicted siRNAs, their free energy of binding with the target mRNA molecules were also measured. Lower binding energy indicates stronger interaction, and therefore well correlates

with a better chance of target inhibition.

- The effectiveness of the siRNAs is also indicated by the higher values of T_m (Cp) and T_m (Conc) as the thermal stability of the RNA molecules is an important parameter of RNAi activity in its conjugated form (Berezhnoy A et al.,2012). Since the T_m values of all the RNAi molecules were between 20-60°C, it suggested that the molecules were thermally stable and thus efficient. The temperature of the cells may also influence the effectiveness of a particular siRNA since the secondary or local structure of an mRNA can change with temperature changes. One study has found varying effects of the temperature over siRNA efficacy (Russell P et al.,2006) .
- While some siRNAs showed no significant changes or slightly increased silencing at lower temperatures, others showed significantly reduced or no silencing when the cells were grown at lower temperatures like 33°C and 35°C. In addition to the liver, SERPINA1 can also be expressed in other tissues and cells, including Lung epithelial cells, Monocytes and macrophages, Skin fibroblast. Regarding exposure, SERPINA1 is primarily exposed to the bloodstream, where it circulates and exerts its anti-protease function. However, in tissues where it is locally produced, such as the lung epithelium, it may also act locally to protect against protease-mediated damage.
- Higher temperatures may enhance the efficiency of certain gene silencing mechanisms, such as RNA interference (RNAi), by increasing the kinetics of RNA degradation or the activity of RNAi machinery. It is assumed that the predicted siRNAs from our study would be able to retain their silencing efficiency in the human body. The downregulation potential of a candidate siRNA crucially depends on its accessibility of the target site within the mRNA molecule. (Gilmore IR et al.,2004). Hence, this benchmark was also studied in detail in this work using the Sirna module of S-fold. Target site accessibility score ranges between 0 and 8, where 0 indicates antisense siRNA binding energy is greater than -2 kcal/mol and 8 indicates the binding energy is less than -16 kcal/mol. The high target site accessibility score for every siRNA in this study, therefore, indicated stronger target binding.
- The key component of RNAi machinery is the Ago2 protein which plays an important role in the functioning of siRNAs. The interactions between the candidate siRNA molecules and Ago2 protein were analyzed in this study. Human Ago2 protein

consists of 4 domains including N- terminal (Asp53-Ser139), PAZ (Pro229-Val347), MID (Gly445-Pro580), and catalytic PIWI (Gln581-Ala859). The PAZ domain holds the 3'-end of siRNA while the MID domain binds to the 5'-end. It was found that in each of the docked complexes, ARG714, ARG792, ARG68, SER220, TYR529, CYS546, HIS600, ALA603, SER798, GLN757, LYS566, THR559 residues among others formed hydrogen bonds with the siRNAs. On the contrary, ARG68, ARG97, ARG375, LYS709, ARG761, ARG792, and HIS600 formed electrostatic interactions while LYS525 and CYS66 residues formed hydrophobic bonds with all the siRNAs. Some of these residues including TYR529, Arg761, SER798, ARG792, LYS566, and CYS546 have also been previously reported to interact with other siRNAs and miRNAs.

- S1 showed the strongest interaction with the Ago2 protein as revealed by the lowest docking energy score as well as a large number of hydrogen bonds. The score reflects the potential energy change when the protein and ligand come together. A positive scoring function represents higher potential affinity. Since in molecular docking analysis of a complex, the energy of the complex is subtracted from its components, a low and negative score indicates that the complex is more stable than its components by themselves. Therefore, they would have a favorable binding interaction. This means that a very negative score corresponds to a strong binding and a less negative or even positive score corresponds to a weak or nonexisting binding.
- MD simulations were performed on S1 and S2 docked complexes. It was revealed from the RMSF analyses that compared to S1, S2 siRNA formed a more stable complex with Ago2 protein which also well correlated with its binding energy and hydrogen bond measurements.

Overall, in this study with the help of multiple computational tools, the most effective and potential siRNA molecules have been designed against the mutated SERPINA 1 in humans. In the next step, the efficacy of these predicted siRNAs needs to be experimentally validated. The candidate siRNAs may prove instrumental in the downregulation of the genetic activity in liver hepatocytes through transient silencing of the mutated SERPINA1 and therefore may lead to a promising new therapeutic approach against Alpha 1 antitrypsin deficiency.

CHAPTER 6 : SUMMARY AND CONCLUSION

RNA interference molecules were designed through in silico methods against the SERPINA 1 of humans to silence their expression. 2 siRNAs were selected as final candidates because they showed good target accessibility, positive energy of folding, and strong binding with Argonaute2 protein. The approach may pioneer a vision for the chemical synthesis of novel I therapeutics for the treatment of A1ATD at the genomic level.

CHAPTER 8 : REFERENCES

1. Abboud, R. T., Nelson, T. N., Jung, B., & Mattman, A. (2011). Alpha1-antitrypsin deficiency: a clinical-genetic overview. *The application of clinical genetics*, 2(6), 55–65. <https://doi.org/10.2147/TACG.S10604>.
2. Akash, G. S., & Ganugapati, J. (2015). In-Silico Design of a Potent siRNA Molecule for Gene Silencing in Zaire Ebolavirus. *International Journal of Computer Applications*, 113(4), 618-635 <https://doi.org/10.5120/19991-7255>
3. Ali, I., Husnain, T., & Riazuddin, S. (2008). RNA interference: the story of gene silencing in plants and humans. *Biotechnology advances*, 26(3), 202-209. <https://doi.org/10.1016/j.biotechadv.2007.12.002>
4. Amarzguioui M, Prydz H. (2004) . An algorithm for selection of functional siRNA sequences. *Biochem Biophys Res Commun.*;35(8), 1050-1058. <https://doi.org/10.1016/j.bbrc.2004.02.157>
5. Asokan, R., Nagesha, S. N., Manamohan, M., Krishnakumar, N. K., Mahadevaswamy, H. M., Prakash, M. N., & Ellango, R. (2012). Common siRNAs for various target genes of the fruit borer, *Helicoverpa armigera* Hubner (Lepidoptera:Noctuidae). 117(5) , 303-689 <https://doi.org/10.1007/s10822-012-9563-2>
6. Berezhnoy A, Brennehan R, Bajgelman M, Seales D, Gilboa E. (2012) Thermal stability of siRNA modulates aptamer- conjugated siRNA inhibition. *Mol Ther Nucleic Acids*. 212(8) ,505-789 <https://doi.org/10.1038/mtna.2012.39>
7. Birmingham A, Anderson E, Sullivan K, et al.(2007) A protocol for designing siRNAs with high functionality and specificity. 116(8), 2068-2078. <https://doi.org/10.1038/nprot.2007.252>
8. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded

- atomic interactions. Protein Sci. (1993); 6(2), 1511-1519
<https://doi.org/10.1002/pro.5560020916>
9. Connolly B, Isaacs C, Cheng L, Asrani KH, Subramanian RR. SERPINA1 mRNA as a Treatment for Alpha-1 Antitrypsin Deficiency. J Nucleic Acids. (2018) Jun 13(1), 650-793. <https://doi.org/10.1155/2018/8247935>.
 10. Chowdhury UF, Sharif Shohan MU, Hoque KI, Beg MA, Sharif Siam MK, Moni MA. (2021) A computational approach to design potential siRNA molecules as a prospective tool for silencing nucleocapsid phosphoprotein and surface glycoprotein gene of SARS-CoV-2. Genomics.;113(1), 331-343. <https://doi.org/10.1016/j.ygeno.2020.12.025>
 11. Ding Y. RNA secondary structure prediction and gene regulation by small RNAs. In: Jianfeng Feng WF, Sun F, eds.(2010) Frontiers in Computational and Systems Biology. London: Springer- Verlag London Limited; 21(7), 20-40 https://doi.org/10.1007/978-1-84996-292-8_2
 12. Ebenezer, A. (2020). Designing Effective small interfering RNA for Post-Transcriptional Silencing of Human GREM1: A Comprehensive Bioinformatics Approach. bioRxiv, 11(5) , 201-311 <https://doi.org/10.1101/2020.05.29.123646>
 13. ElHefnawi, M., Kim, T., Kamar, M. A., Min, S., Hassan, N. M., El-Ahwany, E., ... & Windisch, M. P. (2016). In silico design and experimental validation of siRNAs targeting conserved regions of multiple hepatitis C virus genotypes. PloS one, 11(7), 12(8),159-211. <https://doi.org/10.1371/journal.pone.0159211>
 14. Fakhr E, Zare F, Teimoori-Toolabi L. Precise and efficient siRNA design: a key point in competent gene silencing. Cancer Gene Ther.116(8), 73-82
<https://doi.org/10.1038/cgt.2016.4>
 15. Feitosa P. H. (2023). Diagnosis and augmentation therapy for alpha-1 antitrypsin deficiency: current knowledge and future potential. Drugs in context, 12(7), 23-170
<https://doi.org/10.7573/dic.2023-3-1>

16. Foil, K. E. (2021). Variants of SERPINA1 and the increasing complexity of testing for alpha-1 antitrypsin deficiency. *Therapeutic Advances in Chronic Disease*, 6(12), 2040-5954. <https://doi.org/10.1177/20406223211015954>
17. Gilmore IR, Fox SP, Hollins AJ, Sohail M, Akhtar S. (2004) The design and exogenous delivery of siRNA for post-transcriptional gene silencing. *J Drug Target.*;12, 315-340 <https://doi.org/10.1080/10611860400006257>
18. Hersh CP, Campbell EJ, Scott LR, Raby BA. (2019) Alpha-1 Antitrypsin Deficiency as an Incidental Finding in Clinical Genetic Testing. *Am J Respir Crit Care Med.* Jan 15;199(2), 246-248. <https://doi.org/10.1164/rccm.201809-1679LE>
19. Heo L, Park H, Seok C. (2006) GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.*, 41(6), 384-388. <https://doi.org/10.1093/nar/gkt458>
20. Huesken D, Lange J, Mikanin C, et al. (2005) Design of a genome-wide siRNA library using an artificial neural network. *Nat Biotechnol.*, 23(8), 995-1001 <https://doi.org/10.1038/nbt1118>
21. Ichihara M, Murakumo Y, Masuda A, et al. (2019) Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic Acids* 35(5), 110-123 <https://doi.org/10.1093/nar/gkm699>
22. Islam, R., Shahriar, A., Fatema, N., Uddin, M. R., Acharjee, M., Shepon, M. M. R., ... & Aurin, K. J. (2023). In silico prediction of siRNA to silence the SARS-CoV-2 omicron variant targeting BA. 4, BA. 5, BQ. 1, BQ1. 1. and XBB: an alternative to traditional therapeutics. *Future Journal of Pharmaceutical Sciences*, 9(1), 63-75 <https://doi.org/10.1186/s43094-022-00251-1>
23. Kibbe WA. (2006) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids* 71(6), 30-46. <https://doi.org/10.1093/nar/gkm234>

24. Lu ZJ, Gloor JW, Mathews DH.(2009) Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA*.15(6), 1805-1813. <https://doi.org/10.1261/rna.1643609>
25. Liu Q, Zhou H, Cui J, Cao Z, Xu Y.(2012) Reconsideration of in-silico siRNA design based on feature selection: a cross-platform data integration perspective. *PLoS ONE*. 138(8) , 2500-3779. <https://doi.org/10.1371/journal.pone.0037879>
26. Liu Q, Zhou H, Zhang K, et al. (2007) In silico target-specific siRNA design based on domain transfer in heterogeneous data. *PLoS ONE*. 215(7) ,156-567 <https://doi.org/10.1371/journal.pone.0050697>
27. Markham NR, Zuker M. (2007) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res*. 33(6) ,577- 581 <https://doi.org/10.1093/nar/gki591>
28. Mahfuz, A. M. U. B., Khan, M. A., Sajib, E. H., Deb, A., Mahmud, S., Hasan, M., ... & Rahaman, M. M. (2022). Designing potential siRNA molecules for silencing the gene of the nucleocapsid protein of Nipah virus: A computational investigation. *Infection, Genetics and Evolution*, 102(7),1053-1267 <https://doi.org/10.1016/j.meegid.2022.105310>
29. Meister, G., & Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature*, 431(7006), 343–349. <https://doi.org/10.1038/nature02873>
30. Naito Y, Yoshimura J, Morishita S, Ui-Tei K. siDirect 2.0: (2009) updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics*.;10(4) ,280-392 <https://doi.org/10.1186/1471-2105-10-392>
31. Nur, S. M., Al Amin, M., Alam, R., Hasan, M. A., Hossain, M. A., & Mannan, A. (2013). An in silico approach to design potential siRNA molecules for ICP22 (US1) gene silencing of different strains of human herpes simplex 1. *Journal of Young Pharmacists*, 5(2), 46-49. <https://doi.org/10.1016/j.jyp.2013.05.001>
32. Patrias K, author; Wendling D, editor. (2007) Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers [Internet]. 2nd edition. Bethesda (MD): National

Library of Medicine (US) 4(8) ,456-890 <https://doi.org/10.4103/0975-1483.11373>

33. Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W. S., & Khvorova, A. (2004). Rational siRNA design for RNA interference. *Nature biotechnology*, 22(3), 326–330. <https://doi.org/10.1038/nbt936>
34. Rotondo, J. C., Oton-Gonzalez, L., Selvatici, R., Rizzo, P., Pavasini, R., Campo, G. C., Lanzillotti, C., Mazziotta, C., De Mattei, M., Tognon, M., & Martini, F. (2020). SERPINA1 Gene Promoter Is Differentially Methylated in Peripheral Blood Mononuclear Cells of Pregnant Women. *Frontiers in cell and developmental biology*, 8(1), 550- 543. <https://doi.org/10.3389/fcell.2020.550543>.
35. Russell P, Walsh E, Chen W, Goldwich A, Tamm ER. (2006) The effect of temperature on gene silencing by siRNAs: implications for silencing in the anterior chamber of the eye. *Exp Eye Res* ;82(1), 1011-1016. <https://doi.org/10.1016/j.exer.2005.12.003>
36. Shawan, M. M. A. K., Sharma, A. R., Bhattacharya, M., Mallik, B., Akhter, F., Shakil, M. S., & Chakraborty, C. (2021). Designing an effective therapeutic siRNA to silence RdRp gene of SARS-CoV-2. *Infection, Genetics and Evolution*, 93(1), 104- 951. <https://doi.org/10.1016/j.meegid.2021.104951>
37. Stoller JK, Aboussouan LS. (2102) A review of a1-antitrypsin deficiency. *Am J Respir Crit Care Med*. 185(3) ,246-359. <https://doi.org/10.1164/rccm.201108-1428CI>
38. Sohrab, S. S., El-Kafrawy, S. A., & Azhar, E. I. (2022). In silico prediction and experimental evaluation of potential siRNAs against SARS-CoV-2 inhibition in Vero E6 cells. *Journal of King Saud University-Science*, 34(4), 102-449. <https://doi.org/10.1016/j.jksus.2022.102049>
39. Tian Z, Liang G, Cui K, et al.(2012) Insight into the prospects for RNAi therapy of cancer. *Front Pharmacol* ;12(1) 644-718 <https://doi.org/10.3389/fphar.2021.644718>
40. Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R., & Saigo, K. (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic acids research*, 32(3), 936–948.

<https://doi.org/10.1093/nar/gkh247>

41. Vert J-P, Foveau N, Lajaunie C, Vandenbrouck Y. An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics*. (2006); 21(8) 520-700
<https://doi.org/10.1186/1471-2105-7-520>
42. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(1), 296–303. <https://doi.org/10.1093/nar/gky427>
43. Xu, W., Jiang, X., & Huang, L. (2019). RNA Interference Technology. *Comprehensive Biotechnology*, 16(7) 560–575. <https://doi.org/10.1016/B978-0-444-64046-8.00282-2>.
44. Yan, Y., Zhang, D., Zhou, P., Li, B., & Huang, S. Y. (2017). HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic acids research*, 45(1), 365–373. <https://doi.org/10.1093/nar/gkx407>
45. Yilmazel B, Hu Y, Sigoillot F, et al. (2104) Online GESS: prediction of miRNA-like off-target effects in large-scale RNAi screen data by seed region analysis. 15(1) ,16-117.
<https://doi.org/10.1186/1471-2105-15-192>