Bioinformatics Prediction of siRNA as potential antiviral agents against Hepatitis-C virus

Project Report

Submitted for the fulfillment of the requirement for award of degree of

Master of Technology in Biotechnology

Submitted by:

Divya Khanna (Reg no. 11205246)

Under the supervision of

Mr. Vikas Kaushik

Assistant Professor

SCHOOL OF BIOENGINEERING AND BIOSCIENCES
LOVELY PROFESSIONAL UNIVERSITY
PHAGWARA, PUNJAB-144411

DECLARATION BY CANDIDATE

I hereby declare that the project entitled "Bioinformatics prediction of siRNA as potential antiviral agents against Hepatitis-C virus" submitted by me to Lovely Professional University, Punjab in partial fulfillment of work carried out by me under the guidance **Mr. Vikas Kaushik** (Assistant Professor, domain of Bioinformatics, Lovely Professional University)**.** I further declare that the work reported in this project has not been submitted, and will not be submitted, either in part or in full, for the award of any other degree or diploma of this university or of any other Institute or University.

Place: Punjab

Date: 9th May 2017

Divya Khanna (11205246)

CERTIFICATE

This is to certify that Divya Khanna (11205246), have completed M.Tech Pre-Dissertation titled "Bioinformatics prediction of siRNA as potential antiviral agents against Hepatitis-C virus" under my guidance and supervision. To the best of my knowledge, the present work is the result of their original investigation and study. No part of the dissertation has ever been submitted for any other degree or diploma at any University.

The work is fit for the submission and the partial fulfillment of the conditions for the award of M.Tech Biotechnology.

Mr. Vikas Kaushik

Assistant Professor

Lovely Professional University

Phagwara, Punjab

ACKNOWLEDGEMENT

I want to express my gratitude to my guide **Mr. Vikas Kaushik** (Assistant Professor, domain of Bioinformatics, Lovely Professional University) for his encouraging attitude, affection, inspiration, valuable suggestion and creative guidance for successful completion of this project.

I also want to express my deep sense of gratitude to my parents and friends specially my father **Mr. Yash Pal Khanna**, whose support, care, love, blessings have always strengthened me to carry out my studies unhindered as he bore all loads of life for me and importantly believed in me. He is the man behind me.

.

ABBREVIATIONS

HCV	Hepatitis C Virus
siRNA	Small Interfering Ribonucleic Acid
shRNA	Short Hairpin Ribonucleic Acid
RNAi	Ribonucleic Acid Interference
RNA	Ribonucleic Acid
mRNA	Messenger Ribonucleic Acid
dsRNA	Double-stranded Ribonucleic Acid
RISC	Ribonucleic Acid Induced Silencing Complex
NSP	Non-Structural Protein
UTR	Un-translated Region
NTP	Nucleoside triphosphate
ER	Endoplasmic Reticulum
A	Adenine
G	Guanine
U	Uracil
C	Cytosine
BLAST	Basic Local Alignment Search Tool
MSA	Multiple Sequence Alignment
MSF	Multiple Sequence Format
ORF	Open Reading Frames
WHO	World Health Organization
HIV	Human Immunodeficiency Virus
NCBI	National Center for Biotechnology Information
Nt	Nucleotides

List of Figures

List of tables

CONTENT

CHAPTER -1

INTRODUCTION

1. Introduction

1.1. About HCV

In human and non human primates, Hepatitis is a chronic global health disease. Infection that is thus caused by a virus attacks the liver and inflammation is caused. The virus is thus spread by contact with contaminated blood, from sharing needles to unsterile tattoo equipment. HCV RNA is challenging in designing the therapeutic that target them. Nearly 200 million people are infected worldwide by this chronic disease which thus leads to many problems like cirrhosis, carcinoma and eventually deceases. At present vaccines are not accessible; the at present existing treatment includes interferon with ribavirin which has partial usage because of its undesirable things *(Sagan S et al., 2010).*, although innovative antiviral agents are in progress which should be less contaminated, more efficient and cost effective.

1.2 About siRNA

At the nucleotide level, siRNAs allow the inhibition of genes. Mainly they are RNA molecules duplexes. An exacting power of this knockdown method is, the considered RNA to restrain mRNA expression, and encoding of protein. (*Chalk A et al, 2005&2010*) The approach of knockdown, distinct knockout, allow comprehensive learning of the reducing effect of expression of gene to nothing for a phase of time, and then allowing it to return to normal. This can be established devoid of upsetting related proteins, to make it as a very useful genomics tool. The siRNAs are constructive in Arabidopsis thaliana, Drosophila melanogaster, Caenorhabditis elegans and mammals. *(Watanabe T et al. 2014*)

1.3. Symptoms

Generally, there are no symptoms. Those who face this may have fatigue, nausea, loss of appetite and yellowing of the eyes and skin. Many of the gastrointestinal symptoms include the bleeding, bloating, blood in stool, or nausea with whole body feeling like fatigue, or loss of appetite and weight. (*Modi AA et al, 2008)*

1.4. Mortality rate

Out of all those intensely infected with HCV 85 percent develop unceasing infection. Nearly 70percent of patients with chronic viremia build up constant diseases of liver, 10-20% develops cirrhosis of liver. In 2004, WHO declares that the total death because of liver cancer that HCV caused were 3,08,000 & 7,85,000 respectively. *(Ashfaq U et al. 2011)*

1.5. About the Genome

HCV belongs to *Flaviviridae* which thus separated in these three genera: Flavivirus, Pestivirus, and Hepacivirus. HCV comes under Hepacivirus with at least 7 genotypes: Genotype 1-7 and numerous subtypes. A positive-sense RNA virus that goes through replication through dsRNA transition. The genome thus encodes just about 3010 polyproteins of amino acid which is thus cleave into 3 structural proteins and 7 non-structural proteins. The genome of the HCV RNA serves as template translation, synthesis of negative strand, and packaging in virion. Number of well-defined cis-acting RNA elements has been identified there. Thus the N-terminal of the ORF codes for structural proteins, whereas remaining portion codes for the nsp's. The ORF is flanked in 5' and 3' UTRs of 95-555 and 114-624 nucleotides respectively, which are very vital in polyprotein translation and RNA replication. *(Stéphane Chevaliez and Jean-Michel Pawlotsky: HCV Genome and Life Cycle)*

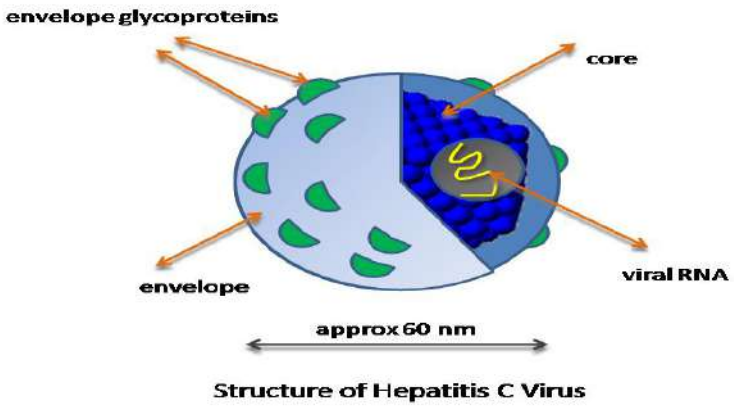


Fig1: Structure of HCV

S.No	Genotype	Subtypes	Countries found in
1	Genotype 1	a,b,c	• Western Europe
2	Genotype 2	a,b,c,i,k	• North America
3	Genotype 3	a,b,k	• India • Nepal • Pakistan
4	Genotype 4	A	• Africa • Middle East
5	Genotype 5	A	• South Africa
6	Genotype 6	a,b,d,f,g,h,k	• Hong Kong • South East Asia
7	Genotype 7	almost a	• South Africa

Table 1: HCV Genotypes

1.6. ABOUT PROTEINS AND THEIR FUNCTIONS

HCV genome encodes approximately 3010 amino acid polyprotein which is processed in structural proteins and non structural proteins. (*Sagan S et al., 2010)* and the functions and characteristics of the proteins are listed below: *(Stéphane Chevaliez and Jean-Michel Pawlotsky: HCV Genome and Life Cycle)*

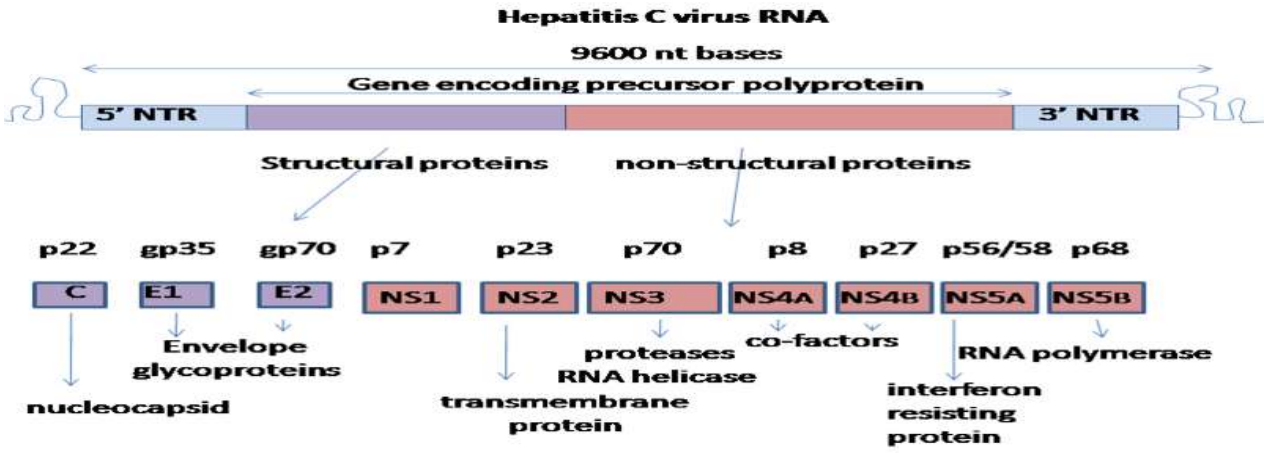


Fig2: HCV Genome

S.No.	Proteins	Characteristics
1.	Core proteins(p22)	• Viral RNA packaging. • Budding of virion promoted. • Initiation of viral translation.
2.	Envelope proteins(E1: gp 35 and E2: gp 70)	• Heterodimer involved in virus attachment is formed by these glycoproteins.
3.	Non structural proteins NS1(p7)	• 63 amino acid polypeptide. • Acts as Heptameric ion channel protein (viroporin).
4.	NS2(p23)	• Non-glycosylated transmembrane protein.
5.	NS3(p70)	• Helicase translocated, the nucleic acid substrate by changing in conformation of protein. • It also utilizes NTP hydrolysis energy during the RNA replication.
6.	NS4A(p8)	• Significant in the lifecycle and pathogenesis of infection.
7.	NS4B(p27)	• Integral protein of membrane of 261 amino acid. • It has ER or ER-derived membrane localization.
8.	NS5A(p56/58)	• A 56–58 kDa phosphorylated zinc-metalloprotein. • Acts as an imperative one in virus replication.
9.	NS5B(p68)	• Belongs to membrane proteins. • It binds the cyclophilin B.

Table2: HCV proteins and its characteristics

CHAPTER-2

TERMINOLOGY

2. Terminology

siRNA: is the most commonly used RNAi tool for inducing short-term silencing of protein coding genes. It is a synthetic RNA duplex designed, specifically to target a particular mRNA for its degradation.

shRNA: It is an artificial RNA molecule with a tight hairpin turn that can be used to silence target gene expression via RNAi. Expression of **shRNA** in cells is typically accomplished by delivery of plasmids or through the viral or bacterial vectors.

HCV: An infectious disease caused by the C virus that primarily affects the liver. During the initial phase people often bear mild or no symptoms but latterly a fever, dark urine, abdominal pain and yellowness on the skin occurs.

RNAi: It is a biological process in which molecules of the RNA inhibit the expression of gene or the translation, by the neutralization of the targeted mRNA molecules.

mRNA: A large family of RNA molecules that conveys genetic information from DNA to the ribosome, where they specify the amino acid sequence of the protein products of gene expression.

RISC: It is a multiprotein complex, specifically a ribonucleoprotein, that incorporates one strand of a single-stranded RNA (ssRNA) fragment, such as microRNA (miRNA), or double-stranded small interfering RNA (siRNA)

ORF: It is the part of a reading frame that has the potential to be translated. An ATG codon within the ORF (not necessarily the first) indicates the starting of translation. The termination of transcription site is located after the ORF.

CHAPTER-3

REVIEW OF LITRATURE

3. Review of literature

3.1. LIFE CYCLE OF HCV

3.1.1. Mechanism

Specific mechanism of Hepatitis C Virus is inadequately known. But by correlation with other RNA viruses, replication is semi-conservative, asymmetric having two steps that catalyzes by NS5B. Positive-strand RNA is synthesis's template. In next, the negative-strand RNA serve as strands of positive polarity production template that are used for translation of polyprotein. It thus transcribed in a 5-10 fold, thus forms a hybrid of primer to the template of RNA. RNA initiation synthesis at 3′-end involves domain I of 5′ UTR, the 3′ UTR and a element of cis-acting replication having 50 bases.(*You S et al. 2004.*) (*Stéphane Chevaliez and Jean-Michel Pawlotsky: HCV Genome and Life Cycle*)

3.2. SPLICING

RNAi is the preserved, influential gene directive pathway. Thus the RNAi method can be briefly depicted as follows:

- Firstly, introduction of "trigger" dsRNA into the cell's cytoplasm was done.
- Secondly, siRNA pools were generated by the Dicer enzyme.
- Thirdly, the siRNAs processed were delivered to RISC.
- Binding of RISC associated to siRNA to objective mRNA.
- Lastly, endonuclease activity is contained by the RISC complex, attributed to subunit of Argonaut, causes single-site cleavage of target mRNA. Destabilization of fragments of target mRNA occurs through natural endogenous mechanism. *(Ashfaq U et al. 2011)*

3.3. siRNA ANALYSIS

Thus the siRNA analysis on which the efficacy is checked follows the guidelines by (*Reynolds A et al., 2004, Amarzguioui M et al., 2004, Ui-Tei et al., 2004*) as:

- A/U at 5'end of the antisense strand.

- G/C at the 5'end of sense strand.
- At least 5 A/U residues at the 5' terminal, one third of the antisense strand.
- Absence of any GC stretches of more than a nucleotide in length.

Any of these not fulfilled, results in no gene silencing by particular siRNA.

3.4. TREATMENT STRATERGY

3.4.1. Vaccines

Currently to the knowledge there is no successful vaccine available. Dealing with peginterferon-alpha2a and ribavirin may be encited by the genotype. 48 weeks treatment and a usual dosage of ribavirin are given to treat people with HCV genotype 1; whereas low dosage for 24 weeks for patients with genotypes 2 is known. (*Halliday J et al., 2011*). Great levels of viral genetic diversity both in between and within the hosts are needed to be dealt in vaccines. Precisely, to be secure, a booming vaccine is needed to eliminate infection from the liver without inducing the liver immune pathology.

3.4.2. RNAi- An influential approach

Since aligned genomes of humans, there is vital necessity to expand major tools that understand the functions of genes. The requirement for an approach like siRNA is significant for abundant diseases. *(Dallas A et al., 2006; 2012).* The varied siRNA specificity target RNA makes it a useful means to discover the importance of gene and in modification of high throughput use for transitory knockdown. (*Ashfaq U et al. 2011).*

In Table 2 varied studies of RNAi therapeutic effects in rodents have been discussed defining the RNAi formulation with the route of administration along with target.

Tissue	**Disease**	**Target**	**RNAi formulation**	**Administration route**
Liver	Hepatitis B	HBsAg	siRNA	Hydrodynamic(intravenous)

Liver	Hepatitis B	Viral genes	plasmid DNA's shRNA	Hydrodynamic(intravenous)
Liver	Hepatitis B	Viral genes	siRNA stabilized	Hydrodynamic(intravenous)
Liver	Hepatitis B	Viral genes	siRNA stabilized and complexed lipids	Intravenous
Liver	Hepatitis C	Viral genes	shRNA	Hydrodynamic(intravenous)
Liver	Hepatitis C	Viral genes	siRNA	Hydrodynamic(intravenous)
Liver	Autoimmune Hepatitis	Fas	siRNA	Hydrodynamic(intravenous)

Table 3: Studies of rodent's therapeutic efficacy of RNAi.

3.4.3 RNAi- A therapeutic agent

Presently, for the treatment of HCV infection being highly variable no vaccine treatment is available. The present dealing of interferon α in support with ribavirin is expensive, and had poor property to cure the infections. Because of the high modification of HCV in the infected liver, it serves as a striking target meant for RNAi. (*Ashfaq U et al. 2011)*

3.4.4. Combination analysis

Antiviral therapy against HCV, host genes acts as an important target and modulates the viral infection. They are the potential targets for siRNA therapy. Hence, introduction of unique therapeutic siRNAs serve rightly as in the recognition of cellular proteins along properties that targets the several genome sites of HCV along with the factors concerned in replication of HCV. (*Ashfaq U et al. 2011).*

3.5. Design and implication of siRNAs

As verified that the HCV replicon are marked complex. They were considered for convention method *(Reynolds A et al., 2004*) and maintain alike G+C ratio to certify similar strength of binding irrelevant of composition of the siRNAs. They were named on the basis of their nucleotide. For controls, two priories considered HCV specific siRNAs were used: IRES-331 (*Wilson J.A et al. 2003*). GL3 siRNA was considered as the negative control, which does not have sequence complementarily to replicon of HCV RNA (*You S et al. 2004.*). All siRNAs can be considered on the basis of fluorescence. (*Sagan S et al., 2010)* Numerous contribution factors thus leads to the general competence of silencing of RNA. The widely held existing design software of siRNA use following parameters that, enhance the probability of selecting an useful siRNA (*Elbashir et al.2002; Reynolds A et al., 2004).*

- G+C content: 30% and 52%.
- Low internal stability of passenger strand. (*Khvorova A et al. 2003*)
- Lack of internal repeats.
- A base: position 19 (sense strand).
- A base: position 3 (sense strand).
- U base: position 10 (sense strand).
- A base other than G or C: position 19 (sense strand).
- A base other than G: position 13 (sense strand). (*Heale B et al. 2005*). (*Sagan S et al. 2010)*

3.6. Antisense technology

Antisense technology is the most prevailing expertise, uses siRNA that encourage RNAi gene silencing, a regulatory pathway of gene.

Target gene expression inhibition, induced by antisense oligonucleotide agent by exploiting capability to bind to target RNAs in a sequence specific way via Watson-Crick hybridization. The antisense agents disable or induce the degradation of RNA.

Three major categories of molecules of gene-silencing:

(1) Derivatives of Antisense oligonucleotide;

(2) Ribozymes and deoxyribozymes;

(3) siRNA encourages degradation of RNA through a pathway of natural gene-silencing pathway that is RNAi.

Hence, newest addition to antisense technologies family is RNAi. (*Dallas A et al. 2006; 2012)*

3.7. About RNAi

RNAi was first reported; a powerful silencing of gene consequence after dsRNA was injected in to the C. elegans. Silencing mechanism of RNA, dsRNA is thus processed by RNase III family component a dicer (19–21 nucleotides) double stranded siRNA which has 2-nucleotide with 30 overhangs which are not phosphorylated. Guide or antisense strand and a passenger or sense strand are two components of double stranded siRNA. Argonaut catalyses the unwinding of siRNA duplex. After the guide strand was included into the RISC while passenger strand was set free. Thus by usage of the antisense strand RISC targets complementary mRNA that results in the cleavage of the latter required. *(Thakur N et al. 2011)*

CHAPTER-4

RATIONALE AND SCOPE OF THE STUDY

4. Rationale and scope of the study

Nowadays, Hepatitis C virus causes persistent diseases related to liver. Thus it also leads to hepatocellular carcinoma, liver damage and eventually death. Currently, the available cure with interferon plus ribavirin, has partial benefits because of the unfavourable side effects like anaemia, depression and "flu-like" symptom. Mostly, the usefulness of Pegylated-interferon treatment is partly, partial to the 3a and 3b virus kind. So, due to the restricted effectiveness of this treatment, RNAi can be said to as the regulatory and potent silencing mechanism for the therapeutics throughout sequence specific approaches by multiple alignments. Numerous reports and researches have depicted the effectiveness and also the specificity of siRNAs. Thus it is being noted grouping of siRNAs will behave as best notion for treatment of the HCV. *(Ashfaq U et al. 2011).* For the knockdown of gene in a sequence-specific manner RNAi is extensively used, it not only use as best tool for gene function but also for many treatment aids, it also includes the treatment of antivirals. Currently, replication of virus extensive series, it can be productively inhibited using siRNA and also vectors that are expressed. (*Naito Y et al. 2006)*

CHAPTER-5

OBJECTIVE OF THE STUDY

5. Objective of the study

- To predict and design siRNA and designing its shRNA as potential antiviral agents against HCV.
- To provide new approach for antiviral therapy against HCV.

CHAPTER-6

RESEARCH METHODOLOGY

6. Research Methodology

6.1. Databases and Softwares used

6.1.1. siVirus Database

It is a web-based software that provides the siRNA design for the antiviral RNAi. It acts as a tool for siRNAs designing that target pathogens, HIV, HCV, etc. all pose several fear to global health.

F: target sites
G: position sites
H: efficacy predictions of siRNA.
I: results of off target matches.
J: Degrees of conservation.
K: Conservation in the sequences of the user.

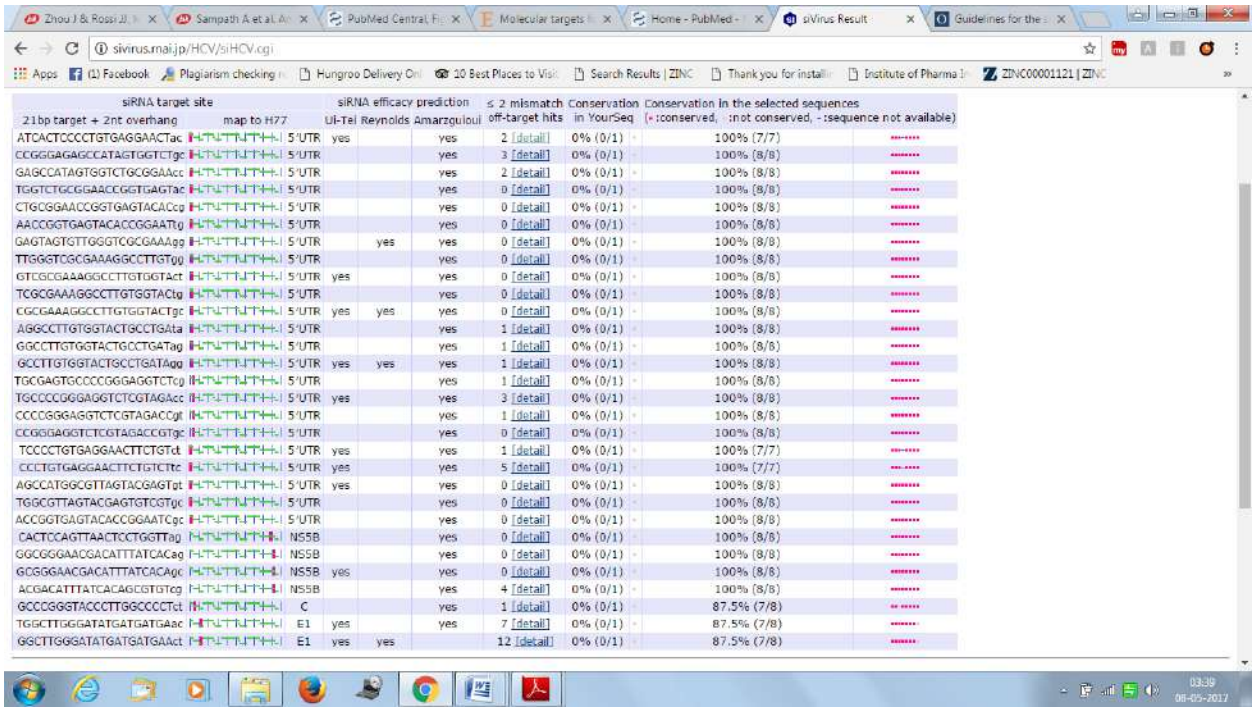


Fig 3: GUI of siVirus Database

6.1.2. siDirect Database

It is used to provide competent as well as targeted design for siRNA for RNAi. It works on the basis of the restructured siRNA algorithm for design based on features combined with the consistent algorithms reported previously.

Thus in the screen image of the interface it is being depicted as the:

A. Target positions of siRNA.
B. Target sequences of siRNA.
C. Efficacy predictions of siRNA. (Ui-Tei *et al.* 2004, Reynolds A *et al., 2004* and Amarzguioui M *et al., 2004*).
D. Homology results.
E. Graphical view of designed siRNAs.
F. Tab-delimited siRNA list.

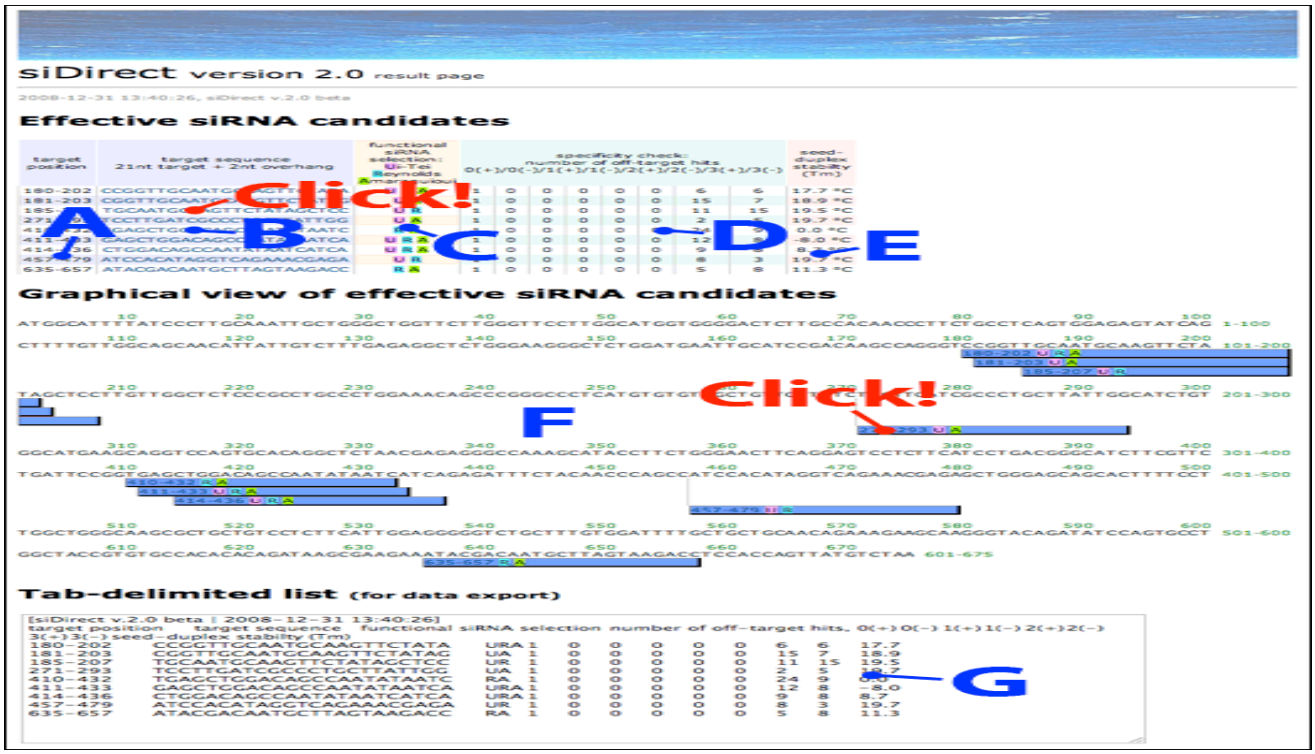


Fig 4: GUI of siDirect Database

6.1.3. Gendock Software

Gendock is software used for obtaining consensus sequence of the multiple alignment thus obtained after aligning the sequences in CLUSTALX or CLUSTALW. It thus provides us with the conserved region thus present in the sequences. It supports MSF files. It also has several view modes. It is thus useful in visualization purposes by applying the information of an alignment to the molecule.

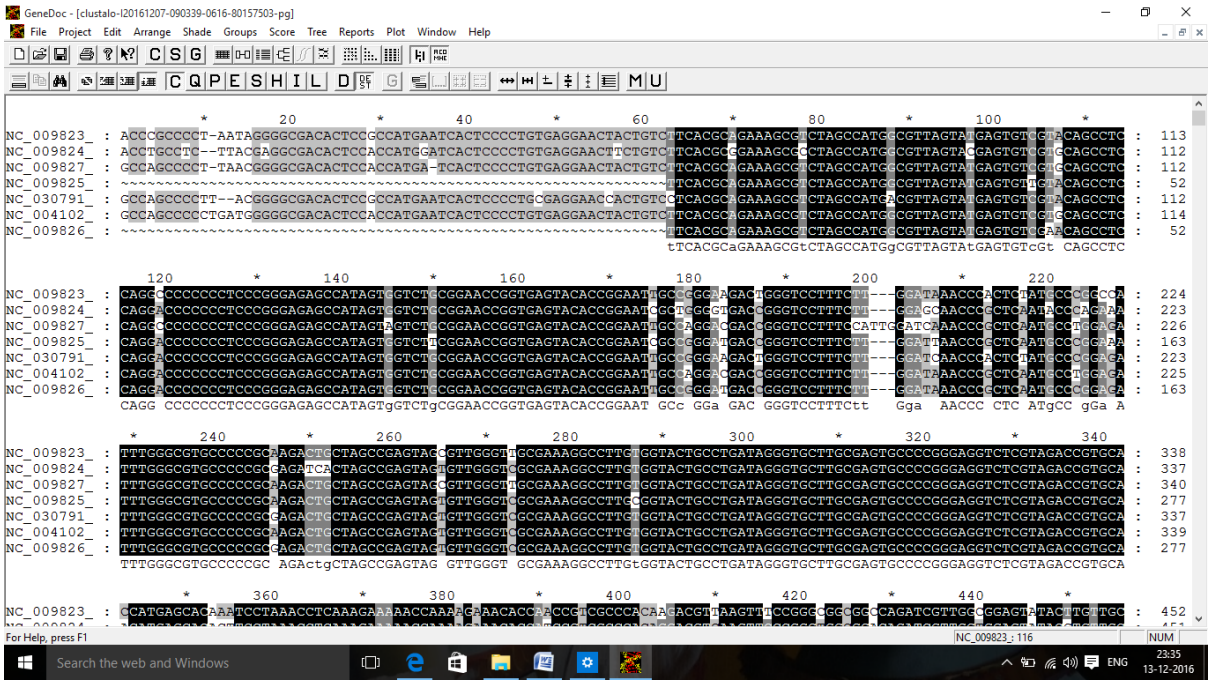


Fig5: GUI of Gendock Software

6.1.4. siWizard Software Online Resources

The siRNA Wizard algorithm allows to select effective and specific siRNAs/shRNA against your gene of interest based on thermodynamic and sequence-related criteria. It will help you find the best siRNA sequences on your target gene. The siRNA Wizard tool will also design the pair of oligonucleotide needed to generate shRNA using InvivoGen's psiRNA plasmids.

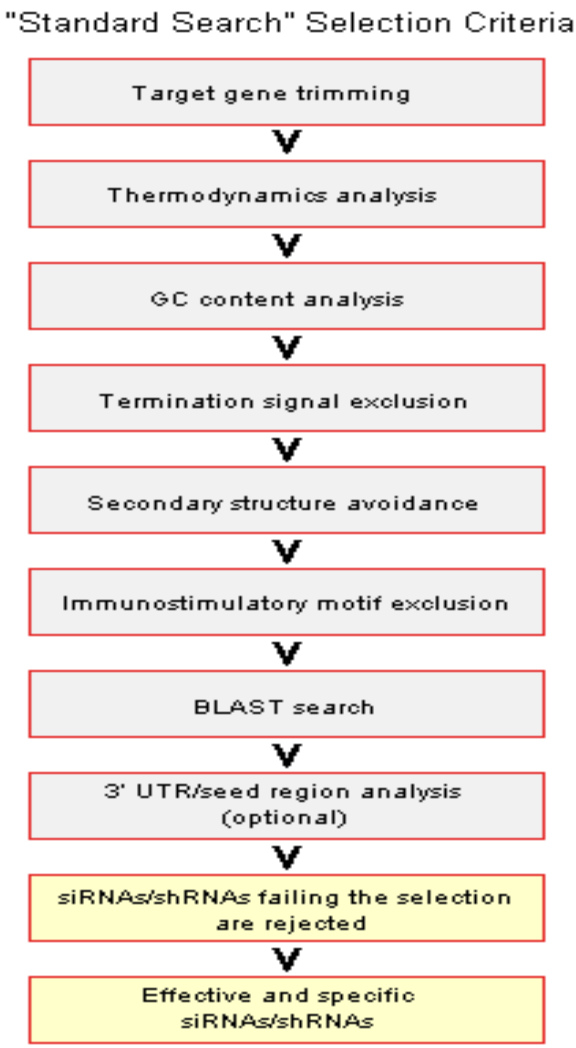


Fig6: Flowchart of siRNA wizard

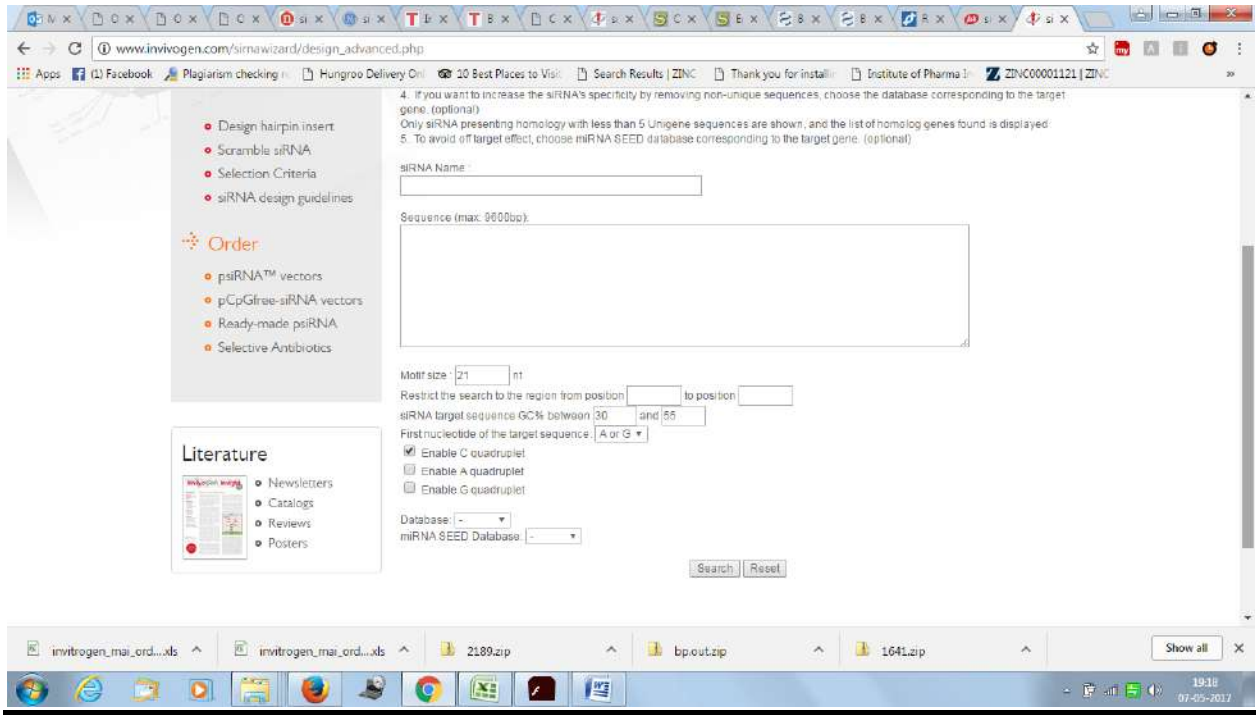


Fig7: GUI of siRNA wizard software for siRNA/shRNA design

6.1.5. i-Score Designer

It thus helps in calculating different scores and thus regarded as siRNA scores based upon the second generation algorithms based upon the ΔG values throughout the siRNA stretches.

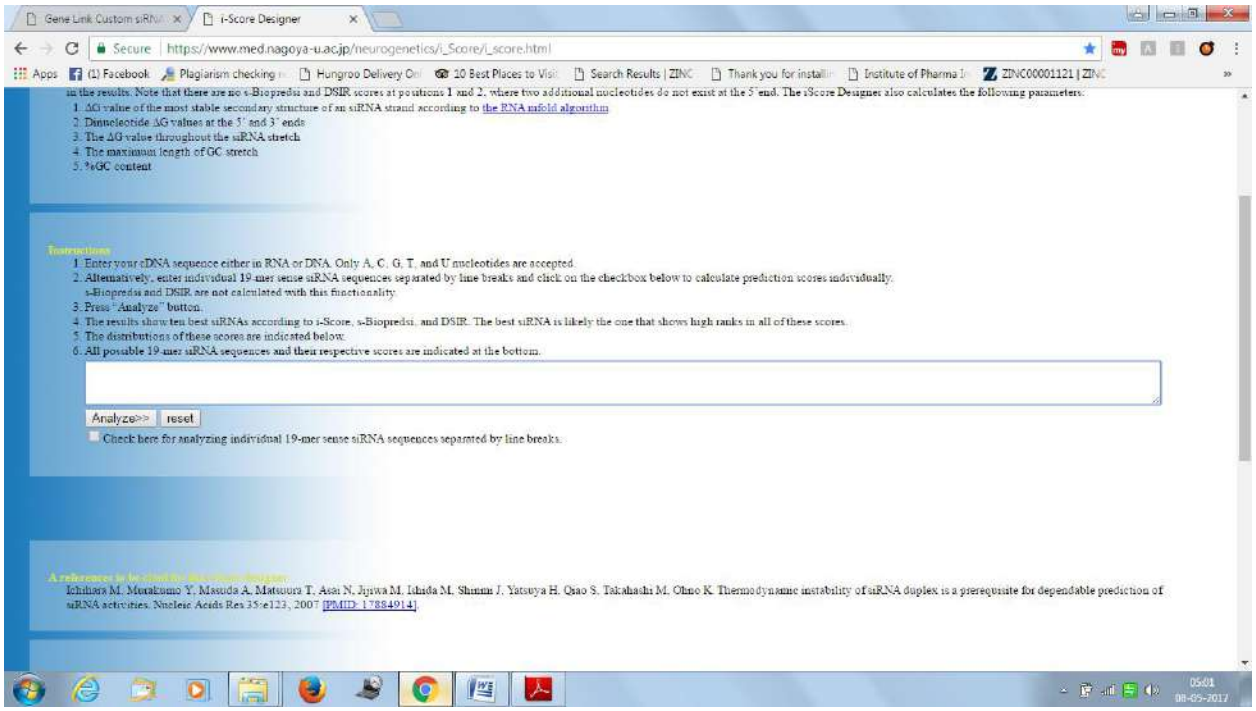


Fig 8: GUI of i-score designer

6.1.6. siRNAmap Database

It helps in mapping the virSirna with the user provided sequences hence helping in the validation of the siRNA targets that are thus designed from those sequences.

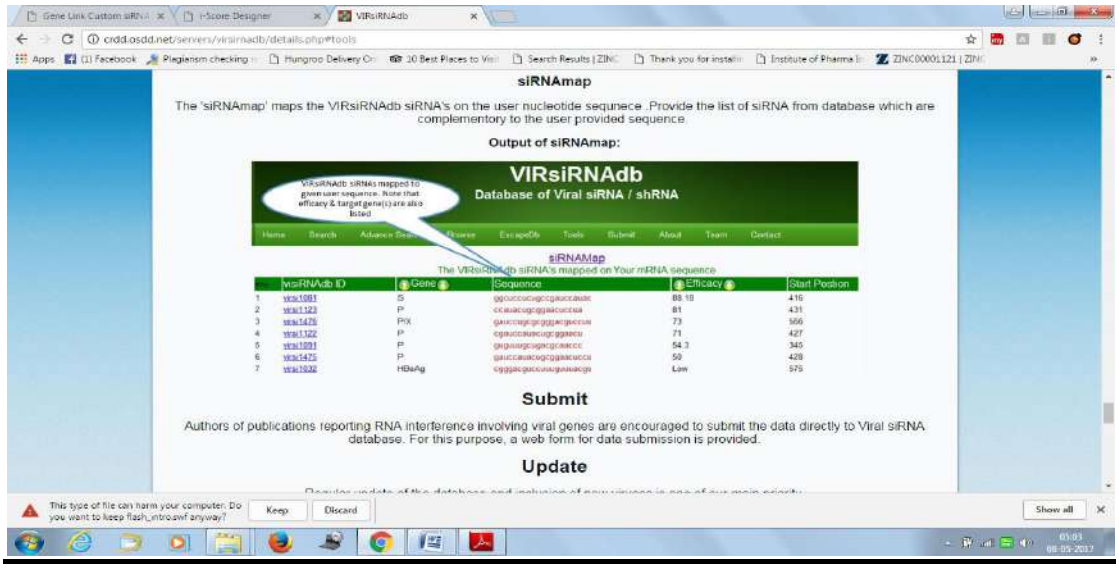
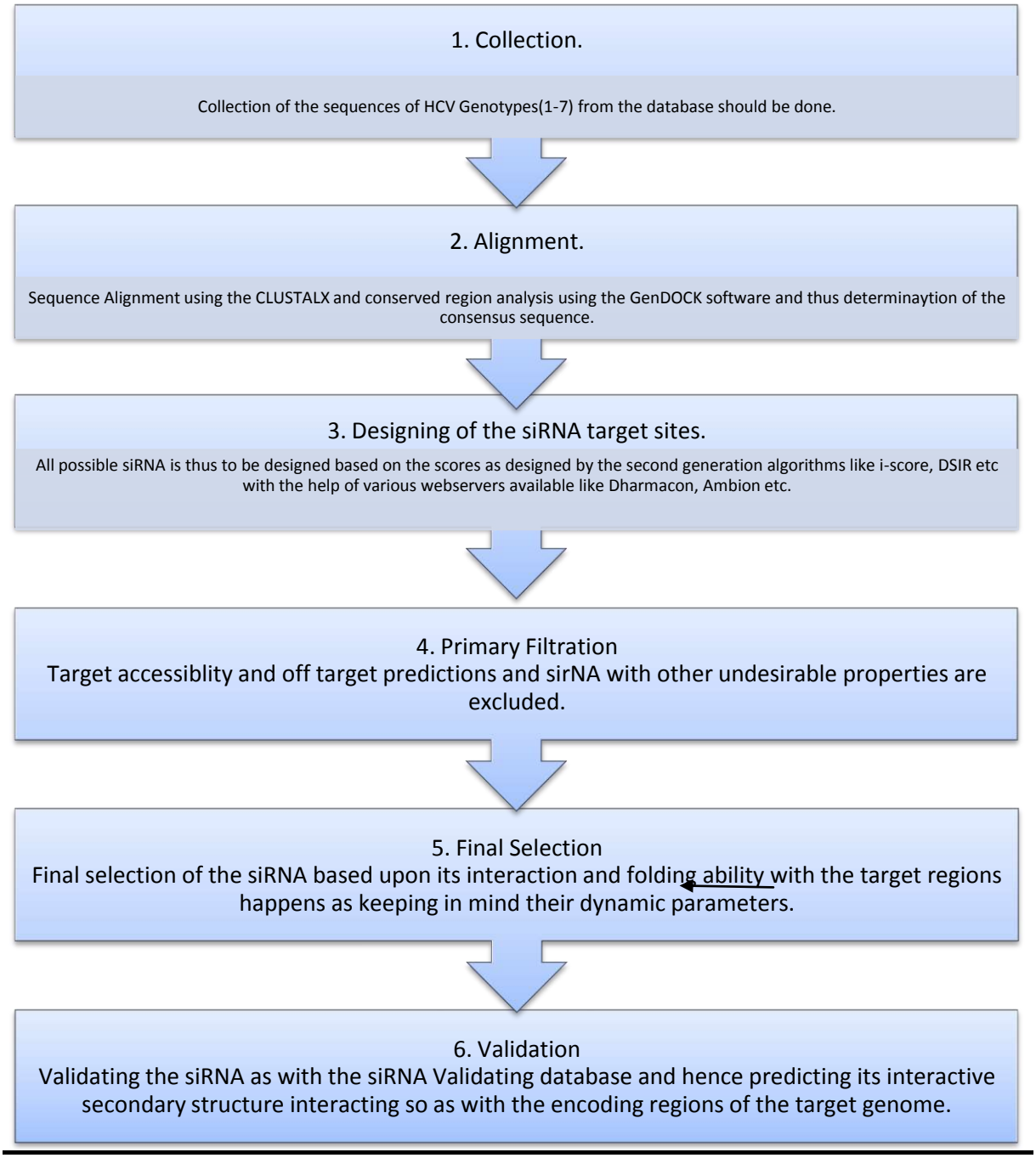


Fig 9: siRNAmap GUI

6.2. Research Methodology

Following is the best yet formulated research methodology to be followed for the best designing of the siRNA.



6.2.1. Collection of the HCV Whole Genome Sequences

To design siRNA, all the sequences of the Genotypes 1-7 of the HCV and the HCV subtypes sequences already present in the database were initially selected. Sequence data for gene segments of HCV were collected from past few years from NCBI Database.

6.2.2. BLAST and Multiple Alignments

Multiple sequence alignment of the extracted sequences is thus being done to align the sequence. Thus CLUSTALX is being used for multiple alignments to observe conservational and functional sites in the genotypes at the amino acid levels using the Gendock software. BLAST program was thus used later on for the similar sequence searches. *(Raza A et al., 2011)*

6.2.3. Designing and Prediction of siRNA targets

Interaction of all genotypes 1 to 7 with the subtype 3 all already present in the siVirus Database within the target ranges is thus made for obtaining the siRNA target sites(21bp+2nt overhangs) with the conservation in selected and your own sequences. For the designing of the siRNA target sites are thus analyzed by the various webservers available online like Dharmacon, Ambion, Qiagen, siWIZARD, siDirect etc.

6.2.4. Primary Filtration

siRNAs were thus selected based on the following stringency parameters:

- Initially 95% conservation is put into consideration for selection of the target sites.
- Minimum or no off targets or mismatches are thus prescribed.
- Lastly the stringency parameters were decreased to 90% and 85% for the selection of the target sites.

Scoring of the siRNAs using the second generation algorithms using databases like i-score, DSIR etc and also assigning of the threshold scores. siRNA with the least undesirable properties are thus excluded and siRNA are thus filtered upon the threshold scores, target accessibility and off-target matches.

6.2.5. Final Selection

Prediction and final selection of the siRNA with the best dynamic parameters. The selection of siRNAs was optimized by choosing siRNAs with the best overall parameters for siRNA ΔG and differential-end instability as a significant thermodynamic feature. The siRNAs with the highest predicted efficacy, specificity, and ability to target the HCV region and adjacent sequences were selected.

CHAPTER-7

RESULT AND DISCUSSION

7. Results and Discussion

7.1. Results

The development of infection treatment was much required. siRNAs prediction and properties, various mechanism of life cycle of HCV have given novel antiviral therapy method. Few attempts are to be made in course of our work, so both the sequence and the length of siRNA are important in designing potent siRNAs *(Moon J.S et al, 2016)*. Thus, it is determined that grouping of siRNA alongside virus and host genes enhanced alternative to treat HCV. siRNA prediction with better efficacy and specificity by high throughput screening are to be expected. *(Rosales PM et al. 2012).*

S.no	**Genotypes**	**Accession No**
1	Genotype 1	NC_004102
2	Genotype 2	NC_009823
3	Genotype 3	NC_009824
4	Genotype 4	NC_009825
5	Genotype 5	NC_009826
6	Genotype 6	NC_009827
7	Genotype 7	NC_030791

S.no	**HCV Subtype**	**Country Code**	**Accession No**
1	3a		AF046866
2	3a		D17763
3	3a		D28917
4	3a	DE	X76918
5	3b	JP	D49374
6	3b		E10841
7	3b		E10839
8	3k	ID	D63821

Table 4: Sequence selection and retrieval

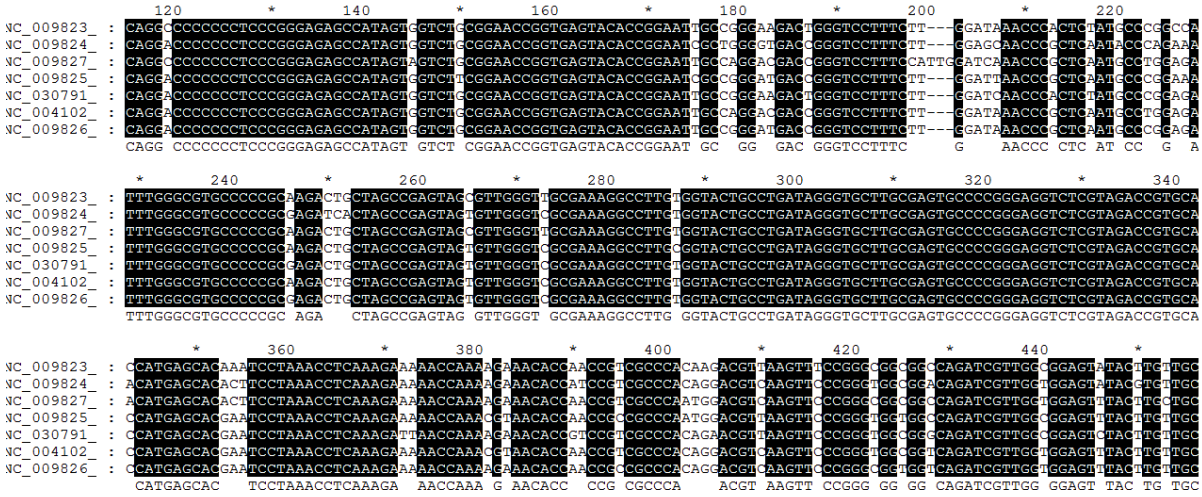


Fig 10: Alignment using Gendock Software

	Genotype 1	Genotype 2	Genotype 3	Genotype 4	Genotype 5	Genotype 6	Genotype 7
5 UTR	9	4	12	3	9	3	9
Core	7	2	7	3	0	3	0
E1	0	0	7	0	0	0	0
E2	0	0	1	0	0	0	0
Ns1	No	No	No	No	No	No	No
Ns2	0	0	3	0	0	0	0
Ns3	0	0	8	0	0	0	0
Ns4a	No	No	No	No	No	No	No
Ns4b	0	0	11	0	0	0	0
Ns5a	No	No	No	No	No	No	No
Ns5b	2	1	11	0	0	0	0
3 UTR	0	0	0	No	0	0	0

Table 5: siRNA analysis using siVirus Database of Genotypes and Subtype 3 all

siRNA target site			siRNA efficacy prediction			≤ 2 mismatch	Conservation	Conservation in the selected sequences
21bp target + 2nt overhang	map to H77		Ui-Tei	Reynolds	Amarzguioui	off-target hits	in YourSeq	(•:conserved, :not conserved, -:sequence not available)
ATCACTCCCCTGTGAGGAACTac		5'UTR	yes		yes	2 [detail]	0% (0/1)	100% (7/7)
CCGGGAGAGCCATAGTGGTCTgc		5'UTR			yes	3 [detail]	100% (1/1) •	100% (8/8)
GAGCCATAGTGGTCTGCGGAAcc		5'UTR			yes	2 [detail]	100% (1/1) •	100% (8/8)
TGGTCTGCGGAACCGGTGAGTac		5'UTR			yes	0 [detail]	100% (1/1) •	100% (8/8)
CTGCGGAACCGGTGAGTACACcg		5'UTR			yes	0 [detail]	100% (1/1) •	100% (8/8)
AACCGGTGAGTACACCGGAATtg		5'UTR			yes	0 [detail]	0% (0/1)	100% (8/8)
GAGTAGTGTTGGGTCGCGAAAgg		5'UTR		yes	yes	0 [detail]	100% (1/1) •	100% (8/8)
TTGGGTCGCGAAAGGCCTTGTgg		5'UTR			yes	0 [detail]	100% (1/1) •	100% (8/8)
GTCGCGAAAGGCCTTGTGGTAct		5'UTR	yes		yes	0 [detail]	100% (1/1) •	100% (8/8)
TCGCGAAAGGCCTTGTGGTACtg		5'UTR			yes	0 [detail]	100% (1/1) •	100% (8/8)
CGCGAAAGGCCTTGTGGTACTgc		5'UTR	yes	yes	yes	0 [detail]	100% (1/1) •	100% (8/8)
AGGCCTTGTGGTACTGCCTGAta		5'UTR			yes	1 [detail]	100% (1/1) •	100% (8/8)
GGCCTTGTGGTACTGCCTGATag		5'UTR			yes	1 [detail]	100% (1/1) •	100% (8/8)
GCCTTGTGGTACTGCCTGATAgg		5'UTR	yes	yes	yes	1 [detail]	100% (1/1) •	100% (8/8)
TGCGAGTGCCCCGGGAGGTCTcg		5'UTR			yes	1 [detail]	100% (1/1) •	100% (8/8)
TGCCCCGGGAGGTCTCGTAGAcc		5'UTR	yes		yes	3 [detail]	100% (1/1) •	100% (8/8)
CCCCGGGAGGTCTCGTAGACCgt		5'UTR			yes	1 [detail]	100% (1/1) •	100% (8/8)
CCGGGAGGTCTCGTAGACCGTgc		5'UTR			yes	0 [detail]	100% (1/1) •	100% (8/8)
TCCCCTGTGAGGAACTTCTGTct		5'UTR	yes		yes	1 [detail]	100% (1/1) •	100% (7/7)
CCCTGTGAGGAACTTCTGTCTtc		5'UTR	yes		yes	5 [detail]	100% (1/1) •	100% (7/7)
AGCCATGGCGTTAGTACGAGTgt		5'UTR	yes		yes	0 [detail]	100% (1/1) •	100% (8/8)
TGGCGTTAGTACGAGTGTCGTgc		5'UTR			yes	0 [detail]	100% (1/1) •	100% (8/8)
ACCGGTGAGTACACCGGAATCgc		5'UTR			yes	0 [detail]	100% (1/1) •	100% (8/8)
TGTGAGGAACTTCTGTCTTCAcg		5'UTR		yes		3 [detail]	100% (1/1) •	87.5% (7/8)
TGACCGGGTCCTTTCTTGGAAca		5'UTR		yes		1 [detail]	0% (0/1)	87.5% (7/8)
GACCGGGTCCTTTCTTGGAACaa		5'UTR			yes	2 [detail]	0% (0/1)	87.5% (7/8)
ACCGGGTCCTTTCTTGGAACAac		5'UTR	yes	yes	yes	0 [detail]	0% (0/1)	75% (6/8)
CCGGGTCCTTTCTTGGAACAAcc		5'UTR	yes		yes	2 [detail]	0% (0/1)	75% (6/8)
TTCTTGGAACAACCCGCTCAAta		5'UTR			yes	1 [detail]	0% (0/1)	75% (6/8)
TTGGGCGTGCCCCCGCGAGATca		5'UTR			yes	0 [detail]	100% (1/1) •	75% (6/8)

Fig 11: siRNA analysis prediction using siVirus Database.

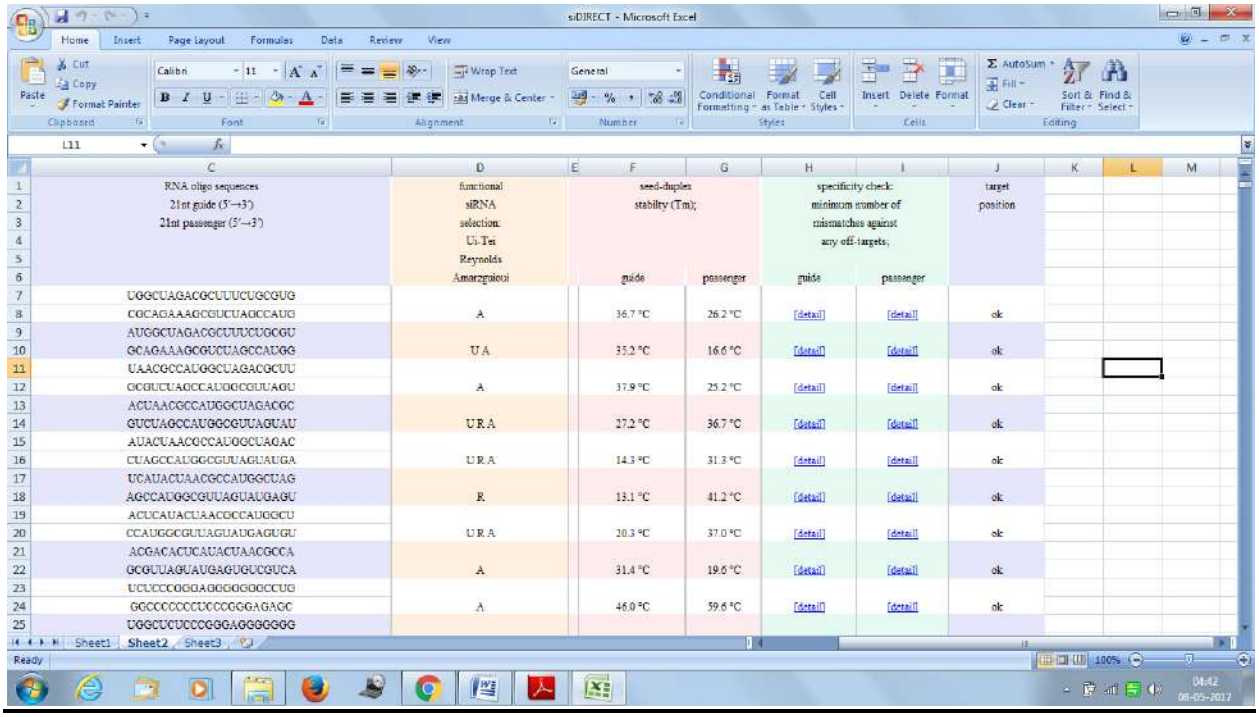


Fig 12: Compilation of the siRNA Design by the siDirect software

After the designing part the filtration and selection processes was done as available in the Supplementary Material and thus finally the below sequences were analysed.

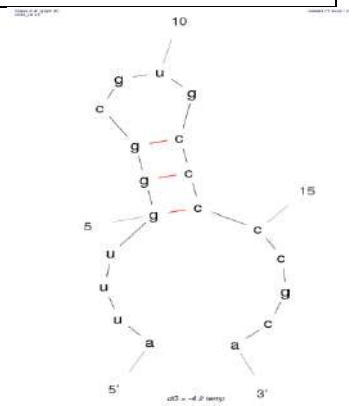
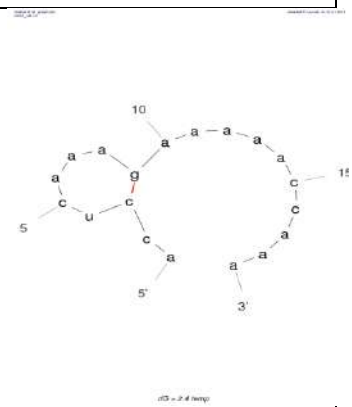
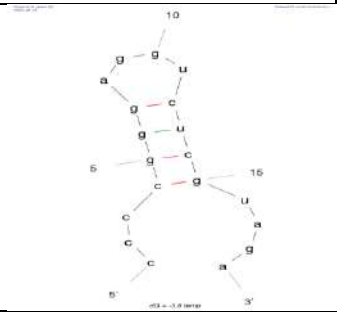
siRNA no.	Sense Sequence	Length	GC %	Virus Name	Target Gene	Cell-line	Target Object	Silencing Efficacy	Structure
1	auuugggcgugccccgca	19	68	HCV	5'UTR	Huh-7	Protein	100	
2	accucaaagaaaaaccaaa	19	32 Tm: 53.8	HCV	C	Huh-7	mRNA	70	
3	ccccgggaggucucguaga	19	68	HCV	5 UTR'	Huh-7	Protein	47	

Table 6: Predicted and Validated siRNA that serves as best potential antiviral agents against HCV.

To design the siRNAs for HCV genomes silencing, a consensus sequence demonstrating all the genotypes of HCV genomes available was proposed. For that, the 7 reference HCV genome genotypes were analyzed by a multiple alignment procedure, finding a highly conserved region among all the genotypes. The viral genomes were depurated to eliminate identical genomes or redundant sequences. A consensus sequence was thus proposed using the GenDock software aligning the conserved regions. The consensus sequence was thus then used using various softwares to design the siRNA. Initially a total of 2632 siRNAs were obtained using the siDirect software. Then after scoring by the second generation algorithms like i-score designer of all the base pairs 122 of the potent siRNAs were discovered as primary filtered, out of which final selection which turns to be the best sequences crosses along the 57 siRNAs for the consensus sequences of HCV genome to silence it for the prediction of siRNA to be potential antiviral agents. Later on 3 siRNA proved to be targeting the 5UTR and Core regions of the HCV genome and hence validate the potency and efficacy to silence the genome based upon the Ambion webserver algorithm which is thus a part of Dharmacon software that we got as siRNAmap to our genome consensus sequence.

7.2. Discussion

Thus here in the prerequisite step in the results, Table 4 depicts the viral sequences selected of the Genotype 1-7 and the other subtypes of the HCV initially present in the database. Fig no 6. describes about the visualization of the conserved region by the Gendock Software using the multiple aligned sequences from CLUSTALX .Also Table no.5 summarizes the number of siRNA target sites obtained after the interaction of genotypes1-7 with the subtype 3(all) of HCV targeting the sites of various non structural proteins. Fig. 9 also provides a representation of the user interface of the siVirus database underlying all the siRNA predicted analysis being done, it thus provides a complete overview of the siRNA analysis using siVirus Database formed by the interaction of the viral sequence of the genotypes and the HCV subtype 3(all). It do tells about the siRNA target sequences, on which parameters it is based and also the mismatch targets depicting the user and the selected conservation sequences. Thus in Fig 11, siDirect software helps us to predict and analyse the siRNA with minimized off targets to serve as better and efficient siRNA target sites. And thus in the Table 6,it was summed up as our predicted and siRNAmap viral siRNA to be configured as the potential antiviral targets that work best against the HCV genome for better silencing.

CHAPTER-8

CONCLUSION AND FUTURE SCOPE

HCV in the present time is the most widespread and life threatening disease. Thus it is not possible to design the sufficient number of vaccines or the drugs due to the diversity in the genotype of the virus. As siRNAs were to be predicted, to be capable to silence the HCV genome serotypes by targeting the non structural sequences, forming double-chain structures, which are necessary to activate the silencing method. Therefore, all the pre-requisite work is thus being done to analyse and benefits the ease in the in-vitro validation steps of the prediction of the siRNAs. In this siRNAs will be synthesized and their inhibitory outcome will be tested *in vitro* aligned with the HCV genome based on their various insertion activities like transfection with the different cell lines and cultures in-vitro in labs.

CHAPTER-9

REFERENCES

1. *Amarzguioui M. and Prydz H. An algorithm for selection of functional siRNA sequences., 2004*
2. *Ashfaq U, Yousaf M, Aslam M, Ejaz R, Jahan S and Ullah O. siRNAs: Potential therapeutic agents against Hepatitis C Virus, Virology Journal. 2011*
3. *Chalk A, Warfinge R, Hemming PG and Sonnhammer E. siRNAdb: a database of siRNA sequences. 2005*
4. *Dallas A, Vlassov A.V. RNAi: A novel antisense technology and its therapeutic potential. 2006; 2012*
5. *Elbashir S.M., Harborth J., Weber K. and Tuschl T. (2002) Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods, 26*
6. *Halliday J, Klenerman P, and Barnes E. Vaccination for hepatitis C virus: closing in on an evasive target. 2011*
7. *Heale B, Soifer H, Browers C, Rossi J. siRNA target site secondary structure predictions using local stable substructures. 2005*
8. *Khvorova A., Reynolds A. and Jayasena S.D.. Functional siRNAs and miRNAs exhibit strand bias. 2003.*
9. *Modi AA & Liang TJ., Hepatitis C: a clinical review. 2008*
10. *Moon J.S, , Lee S.H, Kim E.J, Cho H, Lee W, Kim G.W, Park H.J, Cho S.W, Lee C.G, Oh J.W Inhibition of Hepatitis C Virus in Mice by a Small Interfering RNA Targeting a Highly Conserved Sequence in Viral IRES Pseudoknot. 2016*
11. *Naito Y., Ui-Tei K., Nishikawa T., Takebe Y. and Saigo K. siVirus: web-based antiviral siRNA design software for highly divergent viral sequences, 2006*
12. *Raza A, Shareef H, Salim H, Khushal R, Bokhari H. Selection of predicted siRNA as potential antiviral therapeutic agent against influenza virus.2011*
13. *Reynolds A., Leake D., Boese, Q., Scaringe S., Marshall W.S. and Khvorova A. Rational siRNA design for RNA interference. Nat. Biotechnol.,2004*
14. *Rosales PM, Tenorio AM, Soto E.O, Barrón B.L. Bioinformatics prediction of siRNAs as potential antiviral agents against dengue viruses. 2012*

15. *Sagan S, Nasheri N, Luebbert C, and Pezacki J. The Efficacy of siRNAs against Hepatitis C Virus Is Strongly Influenced by Structure.,2010*
16. *Stéphane Chevaliez and Jean-Michel Pawlotsky: HCV Genome and Life Cycle*
17. *Thakur N, Qureshi A and Kumar M. VIRsiRNAdb: a curated database of experimentally validated viral siRNA/shRNA. 2011*
18. *Ui-Tei K., Naito Y., Takahashi F., Haraguchi T., Ohki-Hamazaki H., Juni A., Ueda R. and Saigo K. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference,2004*
19. *Watanabe T , Hatakeyama H, Yasui CM, Sato Y, Sudoh M, Takagi A, Hirata Y, Ohtsuki T,Arai M, Inoue K, Harashima H & Kohara M. In vivo therapeutic potential of Dicer-hunting siRNAs targeting infectious hepatitis C virus, 2014*
20. *Wilson JA, Jayasena S, Khvorova A, Sabatinos S, Rodrigue-Gervais IG, Arya S, Sarangi F, Harris-Brandts M, Beaulieu S, Richardson CD:. RNA interference blocks gene expression and RNA synthesis from hepatitis C replicons propagated in human liver cells, 2003*
21. *You S, Stump D.D, Branch AD, Rice C.M. A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication, 2004.*

WEB ADDRESSES

1. *http://siVirus.RNAi.jp*
2. *http://si.Direct2.jp*
3. *http://www.genelink.com/custom_sirna/addsiRNA.asp*

CHAPTER-10

APPENDIX

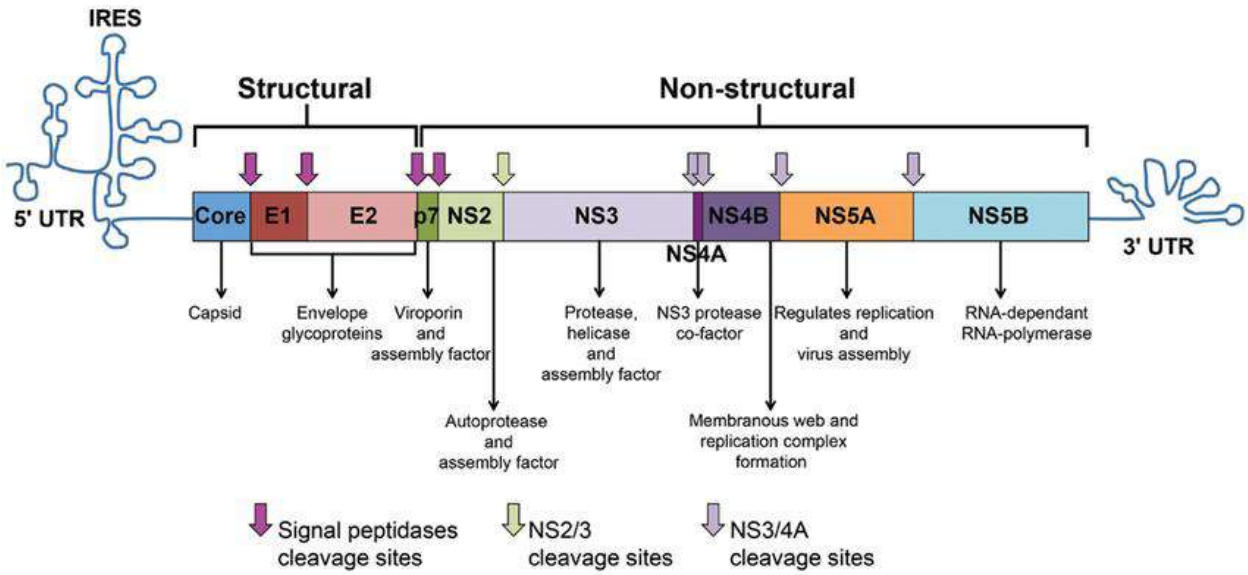


Figure 13: HCV Genome Structure

Source: https://www.researchgate.net/figure/263585596_fig1_Figure-1-HCV-genome-and-polyprotein-The-HCV-genome-is-composed-of-an-open-reading-frame

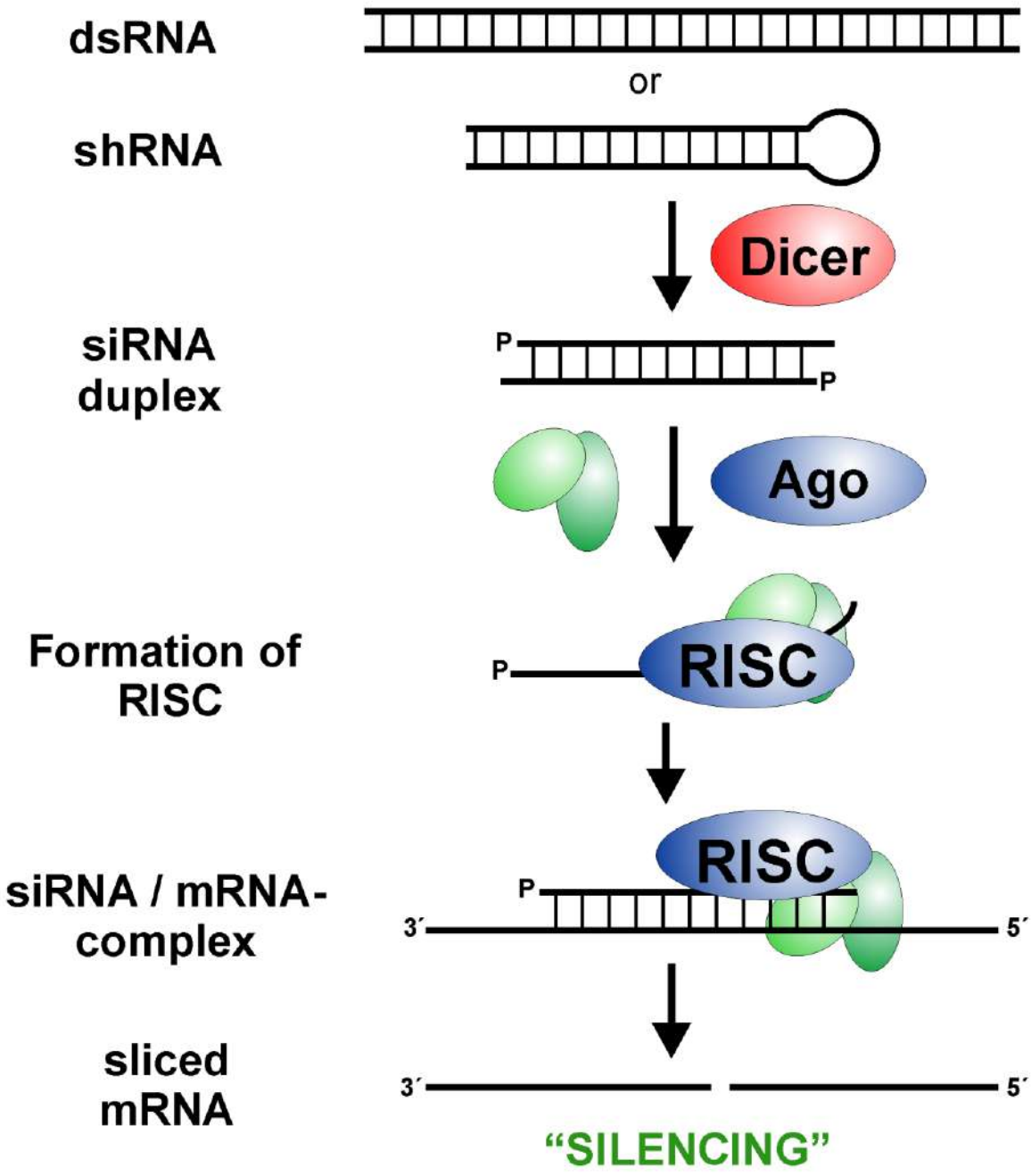


Figure 14: siRNA generation

Source: http://www.gene-quantification.de/siRNA-mechanism.png

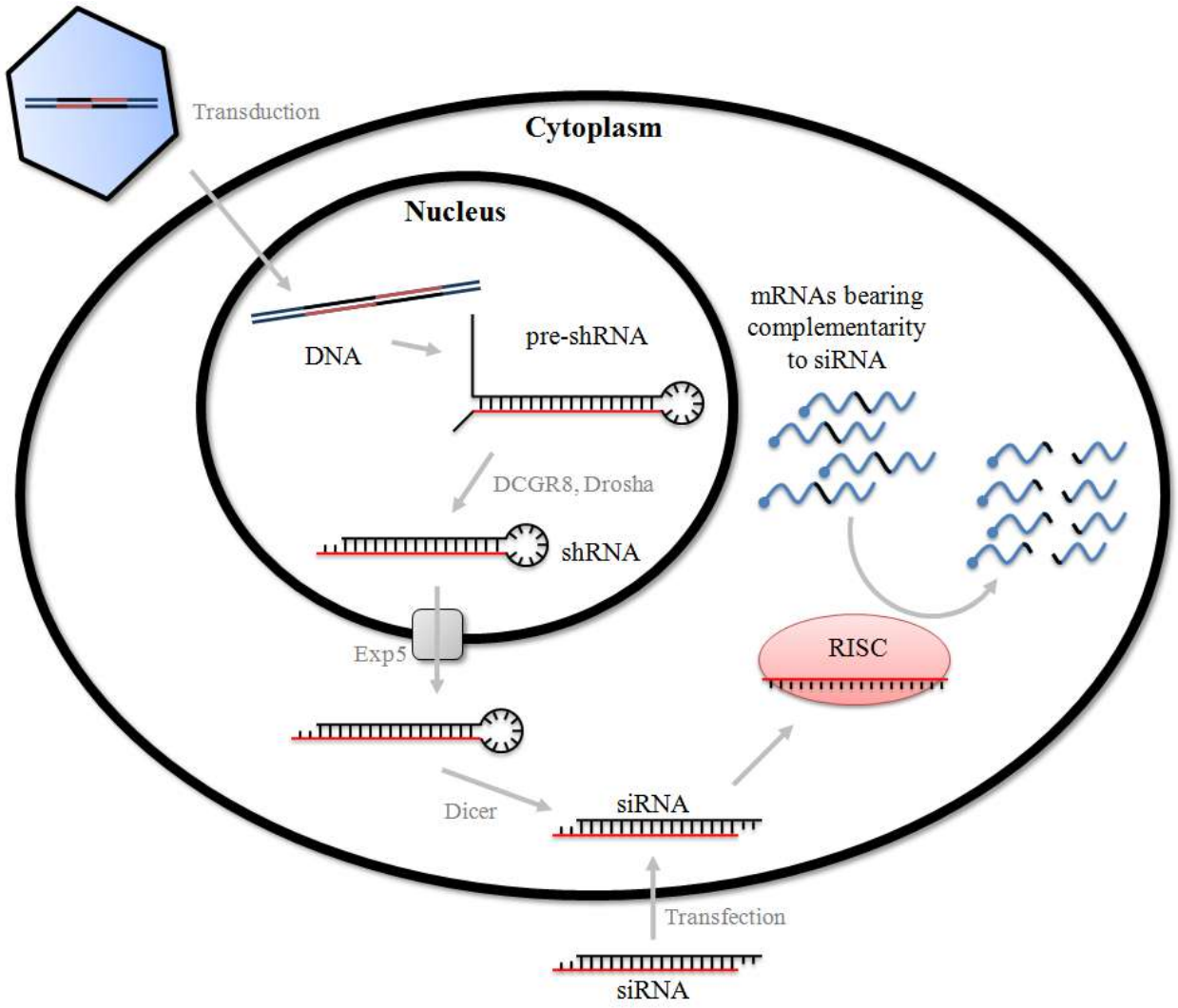


Figure 15: siRNA knockdown mechanism

Source: https://aws.labome.com/figure/te-197-2.png