In-Silico Design of a Potent siRNA Molecule for Gene Silencing in Zaire Ebolavirus

Gopaluni Sai Akash
Department of Biotechnology
Sreenidhi Institute of Science and Technology,
Telangana, India - 501301

Jayasree Ganugapati
Department of Biotechnology
Sreenidhi Institute of Science and Technology,
Telangana, India – 501301

ABSTRACT

Zaire ebolavirus (EBOV) is one of the most dangerous and the unknown to humans. This is a filamentous virus of the family Filoviridae. Genetic studies of EBOV have shown that it has a negative-stranded RNA as genetic material and seven genes in its genome. One among these genes is the gene 'L', that code for a protein L (Large structural protein), this protein functions as an RNA dependent RNA polymerase. RNAi (RNA interference) is an influential method for post-transcriptional gene slicing in a specific sequence. This is done with the help of a dsRNA called siRNA. Slicing a target mRNA would mean that the mRNA will not be able to produce any protein and the viral activity can be restrained. This mechanism is observed naturally in the organism to regulate the protein production and also as a defense mechanism against some viruses. In-silico construction of siRNA is possible using computational methods in biology. Construction of siRNA is guided by many parameters and the efficiency of the cleavage of mRNA with siRNA is determined by hybridization thermodynamics. The constructed siRNA is potent of knocking down the activity of the virus. This can lead to the discovery of an effective antiviral drug against EBOV.

Keywords
RNAi, siRNA, EBOV, Zaire Ebola

1. INTRODUCTION

Zaire ebolavirus (EBOV) is one of the most dangerous and most unknown pathogen that humans have ever encountered. This simple organism causes one of the most dangerous diseases in humans called the Ebola hemorrhagic fever (EHF). The EHF has a high fatality rate of 90% and mortality rate of 79% [1]. This deadly viral infection neither has any vaccine nor any cure [2]. 'Fruitbats' are said to be the natural reservoirs of this virus and these bats survive without being affected. This virus enters humans when one consumes bushmeat [3] and hence can be referred as a zoonotic infection. The genus Ebolavirus has five species Bundibugyo ebolavirus, Reston ebolavirus, Sudan ebolavirus, Tai forest ebolavirus and Zaire ebolavirus. All these species of Ebola virus cause hemorrhagic fever with varying mortality and fatality rates. The first recorded infection was in the year 1976 in Sudan by Sudan ebolavirus, and the most recent in 2014 in Liberia and spread to Sierra Leone, Nigeria, Guinea even to USA and Spain [4] [5].

EBOV has a negative-stranded ssRNA as its genetic material and has seven genes in its genome. Genetic studies have shown that the genome is linear and is about 18,959 nt [6]. These seven genes code for seven proteins that perform almost all functions in its life cycle. EBOV like all the viruses uses the host cell mechanism to complete its lifecycle [7]. Various proteins have various functions in the life cycle of

Ebola. This virus gets into a human cell with the help of GP molecule and NPC1 receptor [8]. VP40 is a matrix protein that helps in viral assembly and budding [9]. VP30 and VP35 are proteins that suppress the host immune system and tolerate the host immune system respectively [10] [11]. VP24 is a peripheral viral membrane protein in viral binding that also plays an important role in the suppression of host interferon activity [12] [13]. NP is a nucleoprotein that forms a ribonucleoprotein complex when binds to the viral RNA. The protein L (Large structural protein) helps in the synthesis of mRNA from negative-stranded ssRNA and hence is an RNA-dependent RNA polymerase [14]. If this protein is disrupted or its function is inhibited the virus will not be able to make any more mRNA neither protein for the propagation of the virus. This can be either achieved by specific protein inhibitors or a siRNA that can slice mRNA encoding for the protein L.

RNA interference (RNAi) is a gene silencing mechanism that uses double-stranded RNA (dsRNA) method which is used for post-transcriptional gene silencing. These dsRNA have the ability to degrade the mRNA and hence hindering in the protein production. This mechanism was first discovered in C.elegans and some plants, but this mechanism also exists in some bacterium and animals [15] [16] [17]. These are usually 19 to 23 nt long and have a TT overhanging on its 3' side [18]. In in-vivo conditions this is produced when a dsRNA is cleaved by a ribonuclease III enzyme called 'dicer' [19]. But in in-vitro conditions, when an siRNA is introduced into a cell it binds to RISC and activates it, this activation of RISC leads to the specific cleavage of the target mRNA molecule [20]. With the advancement in biological sciences, this method of gene silencing has been used as antiviral drug [21]. The siRNA can be introduced into target cells by various methods like electroporation, magnetofection and etc [22] [23]. This method along with the given delivery techniques is extensively used in mice to study gene knockdown effects [24]. Though this is not being used in humans at this point in time it can be an effective therapy against viruses with advancement in the sciences.

The siRNA can be constructed in-silico with the help of computational methods in biology [25]. There are some tools like RNA wizard, Sfold and many others that can design an effective siRNA against a given sequence. These softwares identify specific target sites on the mRNA that are susceptible to cleavage by siRNA. An antisense strand is created against these specific target sites. The construction of siRNA is guided by a number of parameters and the produced siRNA should satisfy all the parameters. Once an siRNA is constructed it undergoes another set of tests that determine the hybridization thermodynamics of the siRNA mRNA duplex which determine the stability of the complex and the efficiency of cleavage [26]. Hence, using these methods and

tools a potent siRNA can be created that can cleave the required mRNA.

2. METHODOLOGY

2.1 Data and Material

The genome sequence of EBOV was retrieved from GenBank database of NCBI that is available at (http://www.ncbi.nlm.nih.gov/). This database contains all genome sequences collected from EBOV and can be used for finding an siRNA target site. The protein sequence can be retrieved from NCBI 'protein' database or nay other relevant databases. ORF finder a tool on NCBI can be used to find the start codon

2.2 Target Location, siRNA Selection based on Thermodynamic Calculations

Sirna (software) was used from Sfold server to produce a graph which shows the probability of finding target sequences in the given mRNA. Computer softwares like siRNA wizard or Sfold can be used to locate the siRNA target sites. This can be produced with the help of the target sequence and the tool Sirna on Sfold (http://sfold.wadsworth.org/cgi-bin/sirna.pl). Thermodynamic calculations of the siRNA duplex- mRNA interactions and cleavage were studied using the same tool Sirna (http://sfold.wadsworth.org/cgi-bin/sirna.pl).

2.3 Similarity Search

The selected siRNA was checked for any off-target interactions, for this BLAST (NCBI) can be used to find all similar sequences to the sense strand of the siRNA (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi). The target sites having similarity of more than 16 nucleotides were avoided though they satisfy all the screening criteria.

3. RESULTS AND DISCUSSION

The genomic sequence of EBOV retrieved from NCBI database was 18,959 nt long and the gene of interest L is 6,781nt long. This gene produces a protein that has 2212 amino acid residues long and functions as an RdRp (RNA-dependent RNA polymerase). This is found in all negative-sense, single-stranded RNA viruses as none of the host enzymes can make an mRNA copy of negative sense ssRNA. The function of this protein is very crucial in its life cycle. Designing a potent siRNA that can silence the gene producing protein L would mean that the protein L cannot be produced. In the absence of protein L the viral genome can be transcribed and infection can be controlled. The tool Sirna was used in the current study to find a target-specific siRNA molecule.

3.1 Probability of finding the siRNA Target Site

This (Sirna) tool uses standard criteria for the selection of ideal siRNA. The target site of siRNA is a region in the gene where it is susceptible to cleavage. The probability of finding an siRNA target site in the given mRNA sequence is indicated in **Fig 1**

This graph gives us a clear idea that the gene has a good probability of finding a target site at approximate positions 10, 30, 80-110, 140 and 160-180 where the probability is almost near to unity. The siRNA found in this region can be considered for further evaluations based on other factors such as GC%, location, and thermodynamic stability. The sharp hike from 170 to 190 shows the probability of finding such a sequence is near to 1.
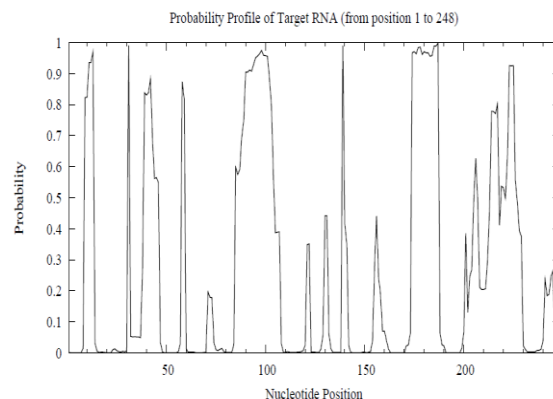
Fig 1: Graph indicating siRNA target site

3.2 In-silico design of siRNA

Finding a target site does not indicate that a potent siRNA molecule can be produced. An siRNA is designed using a set of rules. Keeping in mind these sets of rules in the selected sequences from the previous study can be further evaluated. For an efficient silencing, the target sequence should be 50-100 nt downstream the start codon. Using ORF finder the location of the start codon was found at the position 81. **(Fig 2)** According to the results obtained from the ORF finder the target site should be between the nucleotides 131 and 181. Referring the probability graph and the results from ORF finder, the selected sequences of siRNA can be further shortlisted.
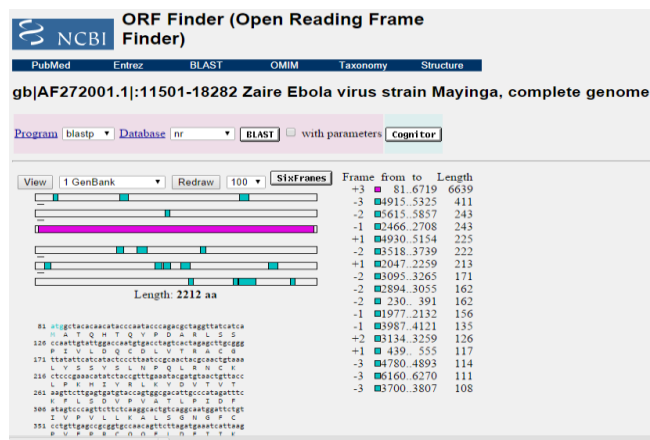

Fig 2: NCBI ORF FINDER Result

Based on the mixed approach of Ui-Tei *et al* and Reynolds *et al*.

1. Targets should be located 50 – 100 nt downstream of the start codon.

2. The target sequence should have G+C content between 35% - 60%.

3. Avoid stretches of four or more nucleotide repeats.

4. Avoid 5' UTR and 3' UTR although siRNA targeting UTRs have been shown to successfully induce gene silencing.

5. Avoid sequence that shares a certain degree of homology with other related or unrelated genes.

Using the design tool 'Sirna' a sequence of siRNA is found out at the position **166 (Fig 3)** which satisfies all the parameters of siRNA construction. The selected sequence is 19 nt long and with dTdT over hangings on the 3' side. From the shortlisted siRNA sequences, only one sequence at the position 166 has met all the criteria.
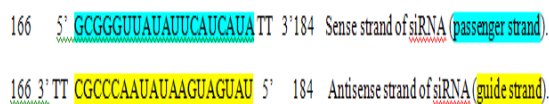
The calculated GC% is 36.8%, which is well within the required range (35% to 60%). A set of hybridization thermodynamic tests was carried out on the final sequence.

3.3 Thermodynamics studies of selected siRNA

Hybridization thermodynamics studies play a vital role in determining the stability of producing siRNA and siRNA-mRNA interactions. On the basis of these two studies, a score for the given siRNA molecule is generated. Thermodynamics of siRNA duplex gives the stability of siRNA and overall quality of siRNA. The results **(Fig 4)** obtained by the software indicate that the constructed siRNA is stable internally as well as at the ends.Since the siRNA molecule is found to be stable internally and externally further studies can be carried out to find the stabilities of sense strand stability, anti-sense strand stability and average internal stability of the cleavage site. Average internal stability (AIS) is determined as per the average cleavage site instability rule and the value of this should be > -8.6 Kcal/ Mol [27]. Differential stability of the siRNA duplex ends is also a major parameter that determines the stability of the siRNA and is guided by asymmetry rule which is enforced by DSSE > 0. DSSE = (stability on 5'-antisense end of 4 base pairs – stability on 5'-sense end of 4 base pairs). **(Table 1)**

Table 1: Thermodynamics of siRNA duplex

PROPERTIES	CONSTRUCTED siRNA
Antisense stability (Antis, in K Cal / Mol)	-7.0
Sense stability (SS, in K Cal / Mol)	-13.6
Differential stability of siRNA duplex ends (DDSE, in K Cal / Mol)	6.6
Average internal stability at cleavage site (AIS, in K Cal / Mol)	-7.2

Fig 3: Sequence of siRNA

Fig 4 Thermodynamics of siRNA

3.4 Net Results

Here the selected siRNA was evaluated on the basis of some other important features and compared them against the ideal values. Target accessibility score correlates with the accessibility of siRNA [28]. The duplex feature score is calculated by the algorithm as per Reynolds, it has a range of [-2, 10] but for an ideal siRNA it should be =>6 [29]. The duplex thermodynamic score has a range between [0, 2] here one point stands for AIS and another for DSSE. If both are satisfied two points are awarded. GC% or GC content is another important criterion on which the efficiency of produced siRNA depends on; GC% should be ideally between 35 to 60%. It is speculated that low GC% facilitates unwinding of 5' end and if more designed siRNA will form secondary structures [30]. Antisense siRNA binding energy is the energy, with which the guide strand/ antisense strand binds to the complementary target site, stronger binding is indicated by a lower binding energy and ideally antisense siRNA binding energy should be <= -10Kcal/Mol. The total siRNA duplex score is the sum of target accessibility scores, duplex feature score, duplex thermodynamics score and other thermodynamic scores.

The overall score produced for this siRNA based on collective results is 14 **(Table 2)**

Properties	Constructed siRNA	Ideal siRNA
Target accessibility scores	7	Between 0 to 8
Duplex feature score	6.6	=>6
Duplex thermodynamic score	2	Between 0 to 20
GC%	36.8%	35% to 60%
antisense siRNA binding energy (kcal/Mol)	-11.1 K Cal/Mol	<= -10 Kcal/ Mol
Differential stability of siRNA duplex ends (DSSE, in kcal/Mol)	6.6 Kcal/Mol	> 0 Kcal/Mol
Average internal stability of the cleavage site	-7.2 Kcal/ Mol	> -8.5 Kcal / Mol
The overall score produced for this siRNA based on collective results is 14.		

3.5 Similarity Search

The designed siRNA must have a target sequence only on the targeted mRNA of the target organism any off-site interactions may lead to unnecessary cleavage and silencing of other biologically important pathways. To check any off-target interactions, the target site of siRNA was run on a BLAST search. Here the target sequence of siRNA was compared with all other sequences in NCBI database; results show all the sequences that are either very similar to the target sequence or partially similar to the target sequence. Any off target sequence, which has a match of 16 nt or more, can cause off-target interactions and in that case, the chosen siRNA is a potential threat. Selected siRNA has similarities only with the target sequence and no other sequence in the database matches 16 nt with the target sequence. Hence, a conclusion can be made from this search that designed siRNA will not show an off-target interactions.

The use of RNAi in humans was targeted to the VEGF protein (vascular endothelial growth factor) that helps in angiogenesis and to KSP in cancer patients with liver involvement which was illustrated in Tabernero *et al*. Other prominent RNAi studies on humans were conducted to treat hepatitis B infection, hepatitis delta virus infection [31] [32] and carcinoma of the breast [33]. But this is being used as a therapeutic in nonhuman primate for Marburg infection and for gene knockdown in many organisms like *D.melanogaster, C.elegans, M.musculus* and many others [34]. All these studies indicate that this technology can be a breakthrough in medicine if an efficient delivery system is used. In mice, magnetofection has been used for successful delivery of siRNA into various target cells. Not only in mice, magnetofection has been also been used as a delivery system for oligonucleotides in human cells both in-vitro and in-vivo [35].

4. CONCLUSION

The aim of the study was to find a siRNA target sequence site for mRNA coding for protein L in Zaire ebolavirus. The mRNA sequence and other information regarding the nucleotide sequence were obtained from NCBI database and tool Sfold was used to find the target sequence, design a siRNA, finding the stability of siRNA and the efficiency of cleavage. One such site was found on the mRNA which shows required qualities and a siRNA was created accordingly. The antisense strand of the siRNA was subjected to BLAST search to check for off target activities. The BLAST search revealed that the siRNA generated was unique and had no similarities with any other target other than Zaire Ebolavirus.

Hence, a conclusion can be derived from the given results that the produced siRNA may be efficient in cleaving the required mRNA and further wet lab studies can show the extent of its efficiency of these studies.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Goeijenbier M et al. 2014 72: 442. Ebola virus disease: a review on epidemiology, symptoms, treatment and pathogenesis. [PMID:25387613]

[2] Hoenen T et al. 2012 12: 859. Current Ebola vaccines. [PMCID: PMC3422127]

[3] Lroy EM et al. 2005 438: 575. Fruit bats as reservoirs of Ebola virus. [PMID:16319873]

[4] No authors listed. 1978 56: 247. Ebola hemorrhagic fever in Sudan, 1976. Report of a WHO/International Study Team. [PMID: 307455]

[5] WHO Ebola response team. The New England journal of medicine. 2014 371:1481 [PMID:25244186]

[6] Bukreyev A et al. 1993 322: 41. The VP35 and VP40 proteins of filoviruses. Homology between Marburg and Ebola viruses. [PMID: 8482365]

[7] http://www.ncbi.nlm.nih.gov/gene/?term=zaire+ebolavir us

[8] Elke M et al. 1999 73: 2333. Comparison of the Transcription and Replication Strategies of Marburg Virus and Ebola Virus by Using Artificial Replication Systems. [PMCID:PMC104478]

[9] Silva LP et al. 2012 7: e39978. Assembly of Ebola virus matrix protein VP40 is regulated by latch-like properties of N and C terminal tails. [PMID:22792204]

[10] Sinu P John et al. 2007 81: 8967. Ebola Virus VP30 Is an RNA Binding Protein. [PMCID: PMC1951390]

[11] Cardenas WB et al. 2006 80: 5168. Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. [PMID:16698997]

[12] Ziying Han et al. 2003 77: 1793. Biochemical and Functional Characterization of the Ebola Virus VP24 Protein: Implications for a Role in Virus Assembly and Budding. [PMCID:PMC140957]

[13] Basler CF et al. 2009 29: 511. Evasion of interferon responses by Ebola and Marburg viruses. [PMID:19694547]

[14] Watanabe S et al. 2006 80: 3743. Functional mapping of the nucleoprotein of Ebola virus. [PMID:16571791]

[15] Fire A et al. 1998 391:806. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. [PMID:9486653]

[16] Lu R et al. 2005 436: 1040. Animal virus replication and RNAi-mediated antiviral silencing in C elegans. [PMCID:1388260]

[17] Ahlquist P et al. 2002 296: 1270. RNA-dependent RNA polymerases, viruses, and RNA silencing. [PMID:12016304]

[18] Carette JE et al. 2011 477: 340. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. [PMID:21866103]

[19] Zamore PD et al. 2000 101: 25. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. [PMID:10778853]

[20] Hammond SM et al. 2000 404:293. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. [PMID: 10749213]

[21] Ahn J et al. 2012 56: 3516. Antiviral effects of small interfering RNA simultaneously inducing RNA interference and type 1 interferon in coxsackievirus myocarditis. [PMID:22508300]

[22] Prechtel AT et al. 2006 311:139. Small interfering RNA (siRNA) delivery into monocyte-derived dendritic cells by electroporation. [PMID:16556448]

[23] Mykhayiyk O et al. 2008 10: 493. siRNA delivery by magnetofection. [PMID:18830925]

[24] Kleinhammer A et al. 2010 477:387. Gene knockdown in the mouse through RNAi. [PMID:20699152]

[25] Gupta et al. 2010 10: 4. A Novel Algorithm to Design an Efficient siRNA by Combining the Pre Proposed Rules of siRNA Designing.

[26] Lu ZJ et al. 2008 36: 640. Efficient siRNA selection using hybridization thermodynamics. [PMID:18073195]

[27] Y Seong et al. 2014 42: 12806. Global identification of target recognition and cleavage by the Microprocessor in human ES cells. [PMCID: PMC4227787]

[28] Tafer H et al. 2008 26: 578. The impact of target site accessibility on the design of effective siRNAs. [PMID:18438400]

[29] Reynolds A et al. 2004 22:326. Rational siRNA design for RNA interference. [PMID:14758366]

[30] Chi Y C et al. 2009 10: S33. A structural interpretation of the effect of GC-content on efficiency of RNA interference.

[31] Van Rii RP et al. 2006 20: 2985. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. [PMID: 17079687]

[32] S Singh et al. 2012 8: 749. Design of potential siRNA molecules for hepatitis delta virus gene silencing. [PMCID:PMC3449391]

[33] Khalili MA et al. 2006 13:636. Gene therapy for carcinoma of the breast. [PMID: 16410823]

[34] Thi EP et al. 2014 20: 250. Marburg virus infection in nonhuman primates: Therapeutic treatment by lipid-encapsulated siRNA. [PMID:25143366]

[35] Krotz F et al. 2003 7:700. Magnetofection--a highly efficient tool for antisense oligonucleotide delivery in vitro and in vivo. [PMID:12718913]