# Design of potential siRNA molecules for hepatitis delta virus gene silencing

**Sarita Singh[1, 2]\*, Sunil Kumar Gupta[1, 2], Anuradha Nischal[1], Sanjay Khattri[1], Rajendra Nath[1], Kamlesh Kumar Pant[1] & Prahlad Kishore Seth[2]**

[1]Department of Pharmacology & Therapeutics, Chhatrapati Shahuji Maharaj Medical University, Chowk Lucknow-226003, Uttar Pradesh, India; [2]Bioinformatics Centre, Biotech Park, Sector-G, Jankipuram, Lucknow-226021, Uttar Pradesh, India; Sarita Singh - E-mail: saritasingh.bi@gmail.com; Phone: +91 522 4053010; Fax: +91 522 4012081; \*Corresponding author

**Abstract:**
Hepatitis D is a liable reason of mortality and morbidity worldwide. It is caused by an RNA virus known as Hepatitis Delta Virus (HDV). Genetic studies of HDV have shown that delta antigen protein is responsible for replication of genome and play a foremost role in viral infection. Therefore, delta antigen protein may be used as suitable target for disease diagnosis. Viral activity can be restrained through RNA interference (RNAi) technology, an influential method for post transcriptional gene silencing in a sequence specific manner. However, there is a genetic variability in different viral isolates; it is a great challenge to design potential siRNA molecules which can silence the respective target genes rather than any other viral gene simultaneously. In current study **two** effective siRNA molecules for silencing of HDV were rationally designed and validated using computational methods, which may lead to knockdown the activity of virus. Thus, this approach may provide an insight for the chemical synthesis of antiviral RNA molecule for the treatment of hepatitis D, at genome level.

**Background:**
Hepatitis Delta virus (HDV) is a human sub viral pathogen which is able to infect only such individuals who were previously or are simultaneously infected with Hepatitis B virus [1]. According to Kuo *et. al.* HDV is able to replicate its RNA within cells in the absence of hepatitis B virus but require hepatitis B antigen for packaging and release of HDV virion [2]. HDV currently infects about 20 million people worldwide [3] and is most common among populations using inject-able drugs particularly in countries bordering the Mediterranean Sea. It is least common in Eastern Asia, but is also present in Taiwan, China and India. Most children with HDV have been identified in Italy and Greece, with a few in Northern Africa. Genetic diversity of HDV is related to the geographic origin of the various isolates. Apart from HDV-1, which is ubiquitous, each virus clad is geographically localized: HDV-2 (previously labeled HDV-IIa) is found in Japan [4], Taiwan [5], and Yakutia,

Russia [6]; HDV-4 (previously labeled HDV-IIb) in Taiwan [7] and Japan [8, 9]; HDV-3 in the Amazonian region [10] and HDV-5, HDV-6, and HDV- 7 in Africa. The studies of Radjef *et. al.* shows that more than seven clades exist and characterized a sequence (dFr644) that was not strongly affiliated to any of the seven HDV clades [11]. Frédéric *et. al.* described two HDV isolates (dFr2072 and dFr2736) that have strong phylogenetic relation to dFr644 reported by Radjef *et. al.*, and proposed as 8th clade [12]. The whole genome of HDV was cloned and sequenced in 1986 [13], which is a single stranded circular RNA of about 1,700nt [14]. Genome replication accumulates three RNA species [15, 16]. The genome and its exact complement, the antigenome, are with unit length and exist primarily in a circular conformation but also in a linear conformation [17]. The third RNA species with relatively lower amounts of an 800-nt polyadenylated RNA (antigenome polarity), is translated to produce a 195-amino-acid protein, known as the delta antigen.

Delta antigen is essential for HDV genome replication **[18].** Clinical association of HDV with hepatitis B infection is due to the fact that the outer coat of HDV consists of hepatitis B virus surface glycoprotein **[19]**. Therefore immune prophylaxis against HDV is achieved by vaccination against hepatitis B virus. This mode of prevention is effective only in case of co-infections in hepatitis B virus susceptible individuals **[3]** as it fails to show any significant effect in case of super-infection, which is more serious state of health [**3**]**.** So there is an urgent need to identify the suitable therapeutic molecule for the treatment of HDV.

RNA interference (RNAi), an evolutionary conserved gene silencing mechanism, uses short double-stranded RNA (dsRNA) to trigger degradation or translation repression of homologous RNA targets in a sequence-specific manner. This has been used as alternative antiviral therapy **[20]**. RNAi therapy has been successful in many pathogenic infection as well as genetic disorders, which interferes with expression of disease causing genes. RNAi was first discovered in *Caenorhabditis elegans* **[21]** and plants, **[22]**, but it can also be used to induce gene silencing in a diverse range of organisms including fungi, protozoans and metazoan animals.

Small interfering RNAs (siRNAs), double-stranded RNAs typically of length between 19 and 25 with 2 nucleotide overhangs on the 3' ends, that are capable of inducing sequence-specific, post-transcriptional gene silencing. Naturally occurring siRNAs are cleavage products of long double-stranded RNAs (dsRNAs) by Dicer, a ribonuclease III enzyme **[23, 24]**. The siRNA-induced mRNA degradation is a complicated process involving multiple steps, initiated by the binding of siRNA with RNA induced silencing complex (RISC), followed by RISC's activation, resulting in the recognition of the target mRNA and its degradation **[23, 25, 26].** As a gene knockdown tool used in labs, siRNAs can also be chemically synthesized and introduced into the cells by direct transfection **[27, 28]** or delivered into the cells in forms of hairpin precursors through plasmid or viral vectors **[29, 30]**. The siRNA-based gene knock-down techniques are preferred by several investigators because of their ability to disrupt individual gene's function without affecting related genes **[31]**. The role of RNAi in mammalian innate immunity is poorly understood and relatively little data is available. However, the existences of viruses that encode genes are able to suppress the RNAi response in mammalian cells may be evidence in favor of an RNAi-dependent mammalian immune response **[32, 33].** However, this hypothesis of RNAi-mediated immunity in mammals has been challenged as poorly substantiated **[34]**. But Elbashir *et. al.* in their study found that these techniques are successful for gene silencing studies in mammalian cells because unlike longer double-stranded RNAs, siRNAs do not trigger interferon responses which lead to non-specific mRNA degradation **[27]**. The RNAi technique is being successfully used as antiviral therapy in various known viral diseases **[35-37].** Currently siRNA gene silencing technology is one of the significant aspects to block the expression of disease causing gene in many biological systems and helpful for gene level treatment. However, HDV genomic and antigenomic RNAs can fold into an unbranched rod-like structure with 74% of the bases paired **[2]**, and this folding might interfere with siRNA action. The delta antigen protein has the ability to bind double stranded RNA **[38]** and thus might also interfere. Chang

and Taylor study with nonreplicating HDV RNA sequences support the interpretation that neither the potential for intramolecular rod-like RNA folding nor the presence of the delta protein conferred resistance to siRNA. However, the genomic and antigenomic RNAs are resistant to siRNA action. The antigenome is localized in the nucleus and so could be inaccessible for RNAi action **[39]**. But the reason for the resistance of genomic RNA is unclear since a large amount of genomic RNA is cytoplasmic **[40, 41]**. Chang and Taylor subsequently showed that the delta antigen mRNA can be successfully targeted by siRNAs in cell culture and also verified that siRNA cannot target the replicating viral RNA transcripts directly but only indirectly via action on the viral mRNA species **[39]**. Thus, delta antigen protein coding mRNA of HDV is obligatory target to inhibit the RNA processing and may be suitable for antiviral therapy. Therefore, in the current study an attempt has been made to identify potential siRNA molecules for silencing of delta antigen coding mRNA or gene in HDV using computational approach.

**Methodology:**
**Data collection and analysis**
Sixty eight genome isolate sequences of HDV were retrieved from viral GenBank database, available at http://www.ncbi.nlm.nih.gov/. The viral database contains all experimentally identified widespread genome isolates of HDV which were further used for siRNA designing **Table 2 (see supplementary material).**

*Target identification and rational siRNA molecule designing*
siDirect 2.0 **[42]** (http://siDirect2.RNAi.jp/) tool, was used for target identification and designing of potential siRNA molecules. It utilized mixed rule approach of Ui-Tei, Amarzguioui and Reynolds rules **[43]** and melting temperature (Tm) below 21.5°C for siRNA duplex, as parameter. For further verification of predicted molecules GeneScript siRNA Target Finder (http://www.genescript.com/index.html), tool was also applied. Besides these other parameters were taken on the concept of algorithms given in **Table 1** (**see supplementary material).**

*Similarity search*
Blast tool (http://www.ncbi.nlm.nih.gov/blast) **[44]** was used to identify any off target sequence similarity in other non targeted organism's genome against whole Genebank datasets by applying expected threshold value 10 and BLOSUM 62 matrix as parameter. The target sites having similarity of more than 16 adjoining base pair with any other organism were excluded from the consideration.

*GC calculation and siRNA secondary structure prediction*
GC calculator tool www.genomicsplace.com/gc_calc.html was used to calculate the GC content for selected siRNA molecule while secondary structure and free energy of folding was computed through Mfold server http://mfold.rna.albany.edu/?q=mfold/download-mfold.

*Thermodynamics calculation of RNA-RNA interaction*
RNAup program (www.tbi.univie.ac.at/~ulim/RNAup) at Vienna web suit **[45]** was used to study the thermodynamics of interactions between target gene and predicted siRNA molecules. It works on extension of the standard partition

function approach to RNA secondary structures that compute energetic of RNA-RNA interactions **[46]**. Flow chart shows complete methodology used for screening of effective siRNA molecules in this study **(Figure 1)**.
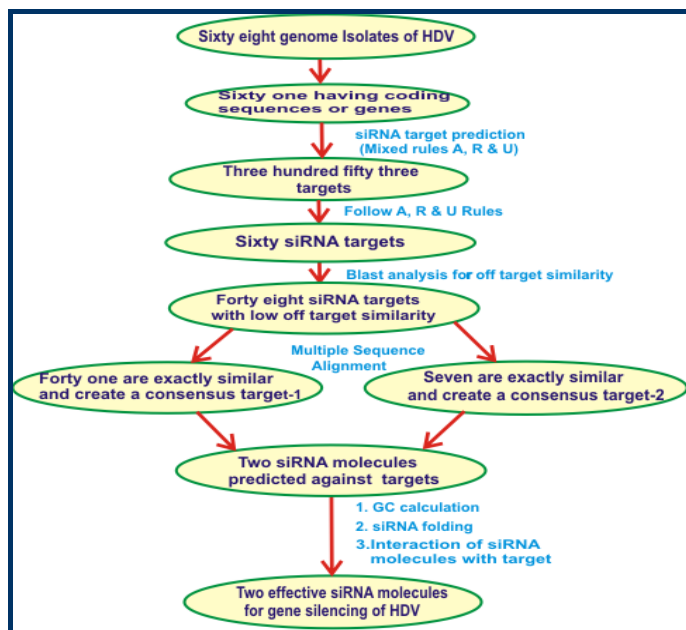


**Figure 1:** Flow chart showing complete methodology used for screening of effective siRNA molecules in this study

**Discussion:**
The HDV genome exists as an enveloped negative sense, single-stranded, closed circular RNA. Because of a nucleotide sequence that is 70% self-complementary, the HDV genome forms a partially double stranded RNA structure that is described as rod-like. With a genome of approximately 1700 nucleotides, HDV is the smallest "virus" known to infect animals. On the basis of geographical origin of various isolates of HDV eight major clade were identified till date **[12]**. Sixty eight genome isolates of HDV are available in NCBI Genebank

database which used in current study **Table 2 (see supplementary material).** Out of these sixty eight, nineteen sequences did not belong to any of the above described clade or groups. After phylogenic study of these nineteen sequences, five sequences were found to have strong similarity with group I; four sequences had similarity with group IV, while only one sequence had similarity to group VIII. Rest nine sequences out of these nineteen did not come under any group, but they had strong similarity with each other and so a new group or clad i.e. group IX **(Figure 2)** has been proposed**.** It has been earlier reported that delta antigen protein is responsible for virus replication and RNA processing **[47].** Hence, this protein coding mRNA or gene from all genome isolates have immense potential for molecular diagnosis and used as target for siRNA molecule designing. There are diverse computational algorithms and tools existing for rational designing of siRNA molecules to knock down the activity of genes **Table 1 (see supplementary material).** siDirect 2.0 tool was used in current study to provide functional, target-specific siRNA molecules, which significantly reduces off-target silencing. To avoid off-target effect, Tm for the seed-target duplex was calculated using the nearest neighbor model and the thermodynamic parameters for the formation of RNA duplex were also studied **[48]**. The formula for calculating Tm is:

$$Tm = \{(1000 \times \Delta H) / (A + \Delta S + R \ln (CT/4))\} - 273.15 + 16.6 \log [Na+] \quad \textbf{(Equation 1)}$$

Where $\Delta H$ (kcal/ mol) is the sum of the nearest neighbor enthalpy change, A is the helix initiation constant (-10.8), $\Delta S$ are the sum of the nearest neighbor entropy change **[49].** R is the gas constant (1.987 cal/deg/mol), and CT is the total molecular concentration of the strand (100 μM). [Na+] was fixed at 100 mM. Apart from it, to check the accuracy of result Gene script target Finder was also applied and usage statistical modeling method.
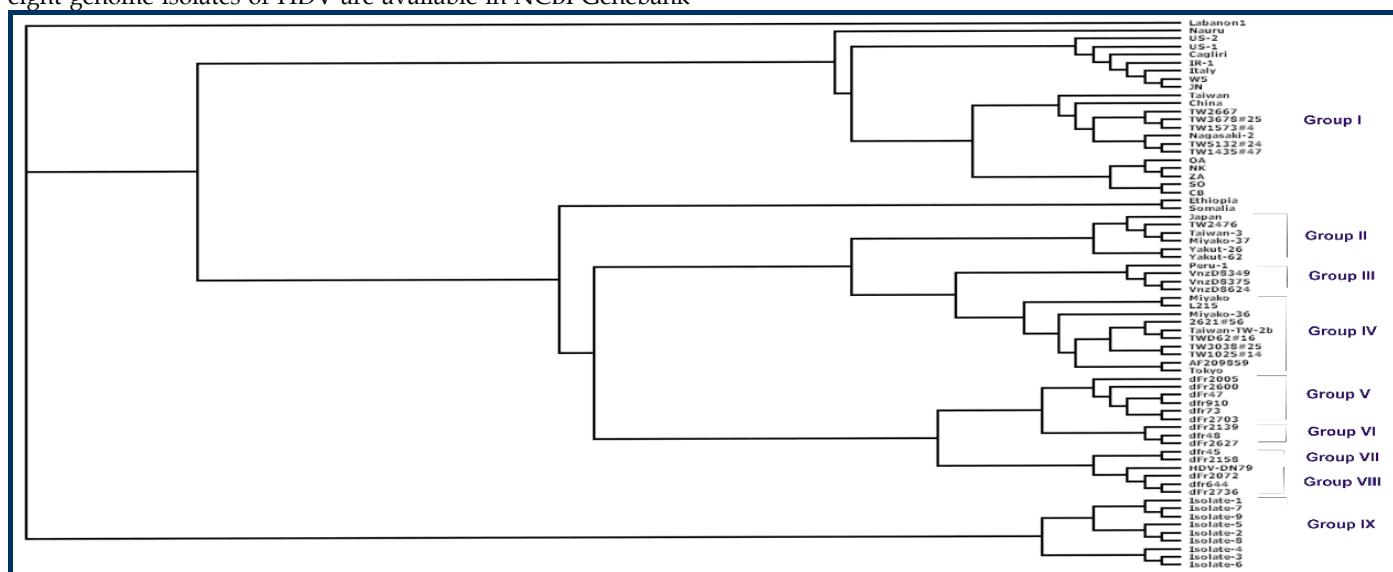


**Figure 2:** Phylogenetic tree for genome isolates of HDV available at NCBI Genebank database showing nine major clades or groups. Eight major clades of HDV were already known but in current analysis ninth one was proposed after phylogenetic study.

# BIOINFORMATION

_open access_

In present study three hundred fifty three siRNA targets were identified for small delta antigen of HDV and potential siRNA molecules against these targets were obtained using mixed rule approach i.e. Ui-Tei, Amarzguioui and Reynolds rule. Out of three hundred fifty three predicted siRNA targets, only sixty were following all three rules. Hence, these sixty siRNA targets were filtered out for further study and considered possible candidates. Consequently these sixty targets were subjected to NCBI Blast tool. Out of theses sixty target only forty eight were selected on the basis of low off target similarity **Table 3 (see supplementary material).** No acceptable siRNA molecule was identified for coding sequences of selected eight isolates (highlighted in **Table 2 (see supplementary material).** All the forty eight effective siRNA molecules selected on the basis of low off target similarity were assessed for target site convenience that can be suitable to knockdown the activity of HDV. MSA of these selected forty eight siRNA targets were depicted that forty one targets have precise similarity **(Figure 3)** while only seven sequences are least similar to them. Subsequently a consensus sequence was executed using EMBOSS tool for these forty one targets and siRNA molecules were designed against these consensus target. Therefore single siRNA might be used to knock down the activity of forty one isolates of HDV while rest seven molecules also create another

consensus target and siRNA molecules against that target were also designed. After this analysis only two molecules were found appropriate for silencing of HDV delta antigen gene. However, there are the incompatible results regarding the effect of GC content and secondary structure on siRNA efficiency. Therefore, these parameters cannot be preferred as a primary determinant of siRNA efficiency. Still, it is recommended to choose sequences with low GC content (31- 58%) **[50-52]**, in the present study both predicted siRNA molecules having recommended range of GC content. Furthermore, the possible folding of predicted siRNA molecules for HDV was done with the online MFold package. Mfold follows most widely used algorithms for RNA secondary structure prediction, which are based on a search for the minimal free energy state **[52]**. Here, one siRNA molecule is having more than zero free energy of folding at $37^0$C **Table 4 (see supplementary material).** Earlier studies have recommended that an RNA molecule should have minimum free energy of folding for their stability. Therefore, the molecule with positive energy may be more accessible for target site and have high potential to bind with target and lead to in effective gene silencing. While other molecule is also having less than -1 kcal frees energy of folding.
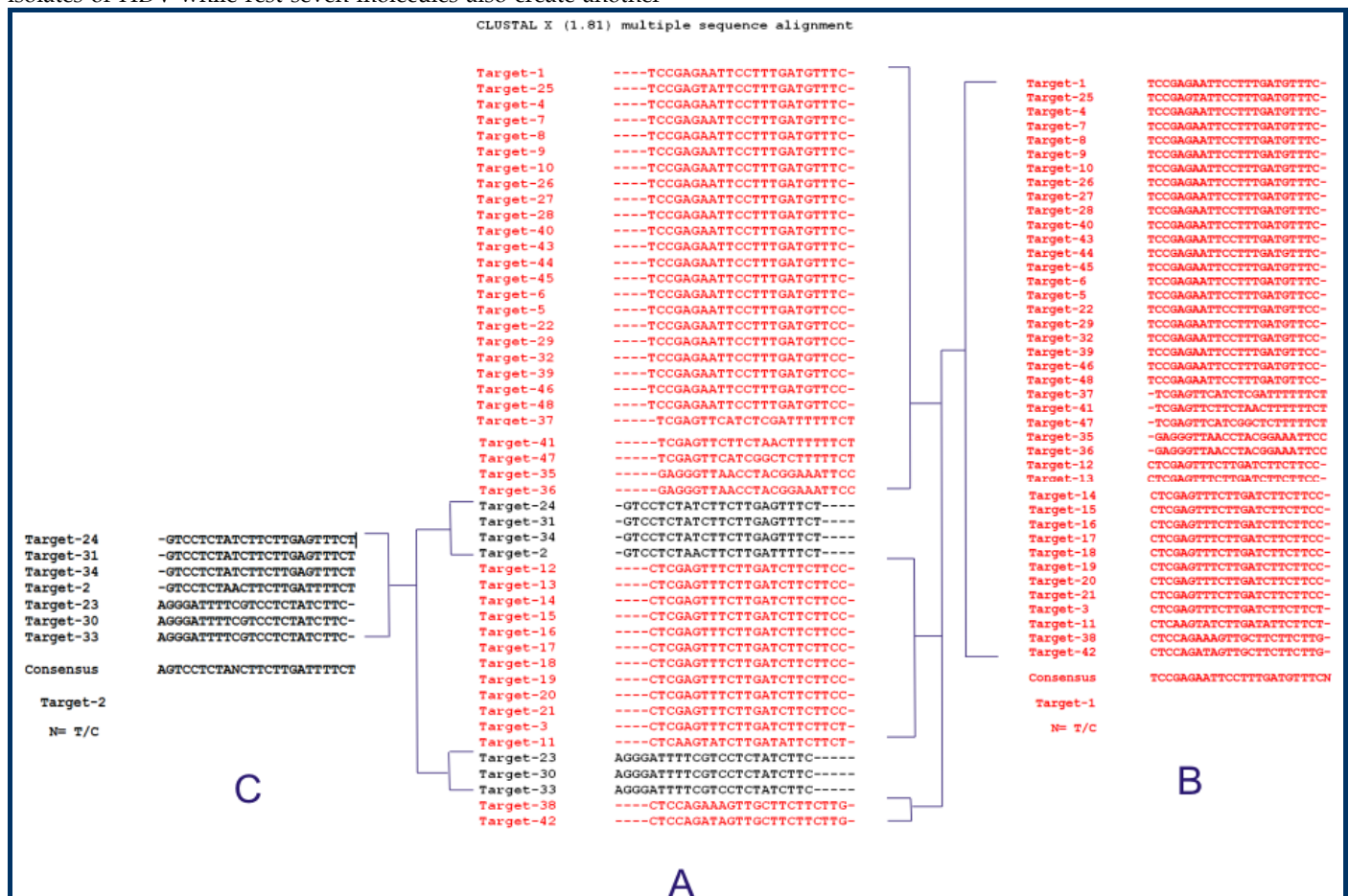


**Figure 3**: Multiple sequence alignment of predicted siRNA target sequences. **(A)** All aligned siRNA target sequence. **(B)** Forty one siRNA targets showing exact similarity and a consensus sequence of them executed using EMBOSS tool. **(C)** Seven target sequences with dissimilarity to forty one sequences, while exact similar to each other and create another good consensus.

Apart from this, a variety of biologically important RNAs were used for prediction of their function by interacting with other

RNA molecules. Thus, thermodynamics study of RNA-RNA interactions may be an important aspect for siRNA molecule

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 8(16): 749-757 (2012)                    752                              © 2012 Biomedical Informatics

# BIOINFORMATION

efficiency. The predicted siRNA molecules were subjected to RNA-RNA interaction study with their respective targets. The Vienna web suite is a comprehensive collection of tool that offers state of the art algorithms for RNA folding, comparison and prediction of RNA–RNA interactions. RNAup one of the important tools of Vienna web suite was used to predict free energy of RNA-RNA interactions. It models the binding energy for the interaction at a particular site as

(BE) $\Delta G_{binding} = \Delta G_{uA} + \Delta G_{uB} + \Delta G_h \rightarrow$ (Equation 2)

Where $\Delta G_{uAB}$ ($\Delta G_{uA} + \Delta G_{uB}$) is the free energy required to make the binding region in molecule A (target) or B (siRNA) accessible by removing intra-molecular structure. While $\Delta G_h$ denotes the free energy gained from forming the inter-molecular duplex by the partition function over all structures where the short RNA binds to target region. Calculation of the free energy of interaction (binding) between a siRNA molecule and its target was performed by using equation (2) **Table 4 (see supplementary material).**

Schubert *et al.* in their study analyzed that RNAi efficiency correlates well with the binding energies of siRNAs to their respective mRNA target **[53]**. Similarly, Mueckstein *et al.* in their study have chosen the target sites provided by Schubert *et al.* and agreed with the experimental results. Optimal free energy of binding (BE) is highly favorable and the siRNA will bind almost exclusively to the intended target site. The stepwise decrease of the target accessibility is directly correlated to a poor optimal BE and decreased silencing efficiency **[46].** In present analysis, siRNA molecule for consensus target-1 was found to be more effective with BE. Therefore, it may qualify as high quality candidate for silencing the delta antigen gene and be used to cure the delta hepatitis.

RNAi approach is successfully exploited in various cases such as hepatitis B infection **[36]** silencing of endonuclease Argonaute 2 in *Drosophila melanogaster* **[54].** RNAi utilized in HIV-1 infection in human peripheral blood mononuclear cells via best env-specific siRNAs, E7145 targeted to the central region of the V3 loop and E7490 targeted to the CD4 binding site of conserved regions on gp120, significantly inhibited the HIV-1 gene expression. Furthermore, E7145 and E7490 were effective against HIV-1NL4-3 replication in PBMCs for a relatively long time (14 days) **[37]**. In experimental brain cancer pegylated immunoliposomes (PIL) carrying short hairpin RNA expression plasmids driven by the U6 RNA polymerase promoter and directed to target EGFR expression by RNAi. The PIL is comprised of a mixture of known lipids containing polyethyleneglycol (PEG), which stabilizes the PIL structure *in-vivo* in circulation. The tissue target specificity of PILs is given by conjugation of ~1% of the PEG residues to monoclonal antibodies (mAbs) that bind to specific endogenous receptors (i.e., insulin and transferrin receptors) located in the brain vascular endothelium **[55]**. *Trypanosoma cruzi* gp83 ligand, a cell surface trans-sialidase-like molecule that the parasite uses to attach to host cells, increases the level of laminin Y-1 transcript and its expression in mammalian cells, leading to an increase in cellular infection. Stable RNA interference (RNAi) with host cell laminin Y-1 knocks down the levels of laminin Y-1 transcript and protein expression in mammalian cells, causing a dramatic reduction in cellular infection by *Trypanosoma cruzi* **[56]**. This

approach was found to be successful in targeting bovine prion gene PRNP in livestock **[57],** carcinoma of the breast **[58]** and crown gall tumorigenesis in plants **[59].** This technique was also used for silencing of capsid genes of Flavivirus using computational methods **[60]**. However, siRNA is the most influential means to control over gene expression in various organisms and showing antiviral activity too. Therefore, rational siRNA has provided the advancement in the development of experiment based approaches to prevent the HDV infections via gene silencing mechanism.

**Conclusion:**
Using RNAi technology a number of siRNA molecules may be designed for silencing of significant genes in various biological systems. Further their interactions with target can also be calculated, computationally. Therefore, in this study two siRNA molecules were predicted against delta antigen protein as effective candidate using computational approaches. These molecules may lead to a novel antiviral therapy against HDV. Study outcome would also provide a basis to the researchers and pharma industry persons to develop the antiviral therapeutics at genomic level, experimentally.

**References:**
**[1]** Rizzetto M *et al. Proc Natl Acad Sci U S A*. 1980 **77**: 6124 [PMID: 6934539]
**[2]** Kuo MYP *et al. J Virol*. 1988 **62**: 1855 [PMID: 3367426]
**[3]** Hadziyannis SJ, *J Gastroenterol Hepatol*. 1997 **12**: 289 [PMID: 9195369]
**[4]** Imazeki F *et al. J Virol*. 1990 **64**: 5594 [PMID: 2214027]
**[5]** Wu JC *et al. Herpetology*. 1995 **22**: 1656 [PMID: 7489970]
**[6]** Ivaniushina V *et al. J Gen Virol*. 2001 **82**: 2709 [PMID: 11602783]
**[7]** Wu JC *et al. J Gen Virol*. 1998 **79**: 1105 [PMID: 9603325]
**[8]** Sakugawa H *et al. J Med Virol*. 1999 **58**: 366 [PMID: 10421403]
**[9]** Watanabe H *et al. J Gen Virol*. 2003 **84**: 3275 [PMID: 14645909]
**[10]** Casey JL *et al. Proc Natl Acad Sci U S A*. 1993 **90**: 9016 [PMID: 8415646]
**[11]** Radjef N *et al. J Virol*. 2004 **78**: 2537 [PMID: 14963156]
**[12]** Frédéric LG *et al. Emerg Infect Dis*. 2006 **12**: 1447 [PMID: 17073101]
**[13]** Wang KS, *Nature*. 1986 **323**: 508 [PMID: 3762705]
**[14]** Taylor JM, *Annu Rev Microbiol*. 1992 **46**: 253 [PMID: 1444256]
**[15]** Chen PJ *et al. Proc Natl Acad Sci U S A*. 1986 **83**: 8774 [PMID: 2430299]
**[16]** Lai MM, *Annual Rev Biochem*. 1995 **64**: 259 [PMID: 7574482]

**[17]** Chang J *et al. Viral.* 2000 **74**: 9889 [PMCID: PMC102025]

**[18]** Kuo MYP *et al. J Virol.* 1989 **63**: 1945 [PMID: 2649689]

**[19]** Cunha C *et al. RNA.* 1998 **4**: 680 [PMID: 9622127]

**[20]** Yan M *et al. World J Gastroenterol.* 2007 **13**: 5169 [PMID: 17876887]

**[21]** Fire A *et al. Nature.* 1998 **391**: 806 [PMID: 9486653]

**[22]** Jorgensen RA *et al. Plant Mol Biol.* 1996 **31**: 957 [PMID: 8843939]

**[23]** Elbashir SM *et al. Genes Dev.* 2001 **15:**188 [PMID: 11157775]

**[24]** Zamore PD *et al. Cell.* 2000 **101**: 25 [PMID: 10778853]

**[25]** Bernstein E *et al. Nature.* 2001 **409**: 363 [PMID: 11201747]

**[26]** Hammond SM et *al. Nature.* 2000 **404**: 293 [PMID: 10749213]

**[27]** Elbashir SM *et al. Nature.* 2001 **411**: 494 [PMID: 11373684]

**[28]** Caplen NJ *et al. Proc Natl Acad Sci U S A.* 2001 **98**: 9742 [PMID: 11481446]

**[29]** Brummelkamp TR *et al. Science.* 2002 **296**: 550 [PMID: 11910072]

**[30]** Rubinson DA *et al. Nat Genet.* 2003 **33**: 401 [PMID: 12590264]

**[31]** McManus MT & Sharp PA, *Nat Rev Genet.* 2002 **3**: 737 [PMID: 12360232]

**[32]** Berkhout B & Haasnoot J, *FEBS Lett.* 2006 **580**: 2896 [PMID: 16563388]

**[33]** Schütz S & Sarnow P, *Virology.* 2006 **344**: 151 [PMID: 16364746]

**[34]** Cullen B, *Nat Immunol.* 2006 **7:** 563 [PMID: 16715068]

**[35]** Xu J *et al. Antiviral Res.* 2007 **73**: 126 [PMID: 17011052]

**[36]** Chen Y *et al. Pharm Res.* 2008 **25**: 72 [PMID: 18074201]

**[37]** Park WS *et al. Nucleic Acids Res.* 2002 **30**: 4830 [PMID: 12433985]

**[38]** Chao M *et al. Virol.* 1991 **65**: 4057 [PMID: 1906549]

**[39]** Chang J *et al. J Virol.* 2003 **77**: 9728 [MID: 12915586]

**[40]** Gudima S *et al. J Virol.* 2002 **76**: 3709 [PMID:11907210]

**[41]** Macnaughton TB & Lai MMC, *J Virol.* 2002 **76**: 3928 [PMID: 11907232]

**[42]** Naito Y *et al. BMC Bioinformatics.* 2009 **10**: 392 [PMID: 19948054]

**[43]** Gupta SK *et al. J Compur Scie & Syst Biol.* 2010 **3**: 1 [doi:10.4172/jcsb.1000048]

**[44]** Altschul SF *et al. J Mol Biol.* 1990 **215**: 403 [PMID: 2231712]

**[45]** Gruber AR *et al. Nucleic Acids Res.* 2008 **36**: W70 [PMID: 18424795]

**[46]** Mueckstein U *et al. Bioinformatics.* 2006 **22**: 1177 [PMID: 16446276]

**[47]** Monjardino JP & Saldanha JA, *Br Med Bull.* 1990 **46**: 399 [PMID: 2198992]

**[48]** Ui-Tei K *et al. Nucleic Acids Res.* 2008 **36**: 7100 [PMID: 18988625]

**[49]** Freier SM *et al. Proc Natl Acad Sci U S A.* 1986 **83**: 9373 [PMID: 2432595]

**[50]** Amarzguioui M & Prydz H, *Biochem Biophys Res commun.* 2004 **316**: 1050 [PMID: 15044091]

**[51]** Reynolds A *et al. Nat biotech.* 2004 **22**: 326 [PMID: 14758366]

**[52]** Zuker M, *Science.* 1989 **244**: 48 [PMID: 2468181]

**[53]** Schubert SJ *et al. Mol Biol.* 2005 **348**: 883 [PMID: 1584302].

**[54]** Van Rii RP *et al. Genes Dev.* 2006 **20**: 2985 [PMID: 17079687]

**[55]** Boado RJ, *NeuroRx.* 2005 **2**: 139 [PMID: 15717065]

**[56]** Nde PN *et al. Infect immun.* 2006 **74**: 1643 [PMID: 16495535]

**[57]** Golding MC *et al. Proc Natl Acad Sci U S A.* 2006 **103**: 5285 [PMID: 16567624]

**[58]** Stoff-Khalili MA *et al. Cancer Gene Ther.* 2006 **13**: 633 [PMID: 16410823]

**[59]** Escobar MA *et al. Proc Natl Acad Sci U S A.* 2001 **98**: 13437 [PMID: 11687652]

**[60]** Somvanshi P *et al. Interdiscip Sci.* 2009 **1**: 298 [PMID: 20640808]

# BIOINFORMATION

## Supplementary material:

**Table 1:** Algorithms or rules for rational design of siRNA molecules

| Ui-Tei Rules | Amarzguioui Rules | Reynolds Rules |
|---|---|---|
| A/U at the 5' terminus of the sense strand | Duplex End A/U differential > 0. Strong binding of 5'sense strand | Each rule is assigned a score which is summed up to a total duplex score to improve the efficacy of siRNA. |
| G/C at the 5' terminus of the antisense strand | No U at position 1. Presence of A at position 6. | |
| At least 4 A/U residues in the 5' terminal 7 bp of sense strand | Weak binding of 3'sense strand. No G at position 19 | |
| No GC stretch longer than 9nt | | |

**Table 2:** Worldwide, genome isolates of HDV

| Group | Genome isolate | Country | Accession No | Location of coding region in genome | Length (Base pair) |
|---|---|---|---|---|---|
| Group I | Cagliri | Cagliri | X85253 | 1008-1595 | 1676 |
| | China | China | X77627 | 82-666 | 1674 |
| | Ethiopia | Ethiopia | U81989 | 1011-1598 | 1679 |
| | IR-1 | Iran | AY633627 | 1007-1594 | 1676 |
| | Italy | Italy | X04451 | 950-1618 | 1678 |
| | Lebanon1 | Lebanon | M84917 | Not Available | 1678 |
| | Nauru | Nauru | M58629 | 952-1598 | 1678 |
| | Somalia | Somalia | U81988 | 1006-1593 | 1674 |
| | Taiwan | Taiwan | M92448 | 951-1595 | 1682 |
| | TW2667 | Taiwan | AF104263 | 951-1595 | 167 |
| | US-1 | US | D01075 | 1013-1600 | 1682 |
| | US-2 | US | L22066 | 1013-1600 | 1681 |
| | Nagasaki-2 | Nagasaki | AB118849 | 950-1594 | 1675 |
| | OA | Turkey | EF514905 | 950-1594 | 1676 |
| | NK | Turkey | EF514904 | 950-1594 | 1676 |
| | SO | Turkey | EF514906 | 950-1594 | 1676 |
| | ZA | Turkey | EF514907 | 950-1594 | 1676 |
| | CB | Turkey | EF514903 | 950-1594 | 1676 |
| | W5 | Italy | AJ307077 | 954-1598 | 1679 |
| | TW3678#25 | Taiwan | AY648959 | 1006-1593 | 1674 |
| | TW5132#24 | Taiwan | AY648957 | 1007-1594 | 1675 |
| | TW1573#4 | Taiwan | AY648958 | 1006-1593 | 1674 |
| | TW1435#47 | Taiwan | AY648956 | 1007-1594 | 1675 |
| | JN | China | HM046802 | 952-1596 | 1677 |
| Group II | Japan | Japan | X60193 | 1013-1600 | 1683 |
| | Yakut-26 | Russia | AJ309879 | 964-1605 | 1685 |
| | Yakut-62 | Russia | AJ309880 | 964-1605 | 1688 |
| | TW2476 | Taiwan | AF104264 | 956-1600 | 1684 |
| | 3 | Taiwan | U19598 | 949-1593 | 1676 |
| | Miyako-37 | Japan | AB118846 | 952-1596 | 1678 |
| Group III | Peru-1 | Peru | L22063 | 953-1597 | 1677 |
| | VnzD8375 | Venezuela | AB037947 | 1007-1591 | 1672 |
| | VnzD8349 | Venezuela | AB037948 | 1008-1592 | 1673 |
| | VnzD8624 | Venezuela | A B037949 | 1009-1593 | 1674 |
| Group IV | Miyako | Japan | AF309420 | 1011-1595 | 1676 |
| | L215 | Japan | AB088679 | 961-1602 | 1682 |
| | Miyako-36 | Japan | AB118845 | 955-1596 | 1677 |
| | Taiwan TW-2b | Taiwan | AF018077 | 1012-1596 | 1676 |
| | AF209859 | Taiwan | AF209859 | 1013-1597 | 1678 |
| | Tokiyo | Japan | AB118847 | 957-1598 | 1678 |
| | TW3038#25 | Taiwan | AY648955 | 1014-1598 | 1678 |
| | 2621#56 | Taiwan | AY648953 | 1014 -1598 | 1678 |
| | TW1025#14 | Taiwan | AY648954 | 1014-1598 | 1678 |
| | TWD62#16 | Taiwan | AY648952 | 1014-1598 | 1678 |
| Group V | dFr2005 | Guinea | AM183331 | 958-1602 | 1687 |
| | dFr2600 | Togo | AM183326 | 963-1607 | 1691 |
| | dFr47 | Not Available | AX741149 | Not Available | 1697 |
| | dFr910 | Mali | AX741159 | Not Available | 1697 |
| | dFr-73 | Not Available | AX741154 | Not Available | 1697 |
| | dFr-2703 | Senegal | AM183328 | 1017-1604 | 1688 |

| Group VI | dFr48 | Poland | AX741164 | Not Available | 1687 |
|---|---|---|---|---|---|
| | dFr2139 | Central African Republic | AM183332 | 951-1592 | 1677 |
| | dFr2627 | Nigeria | AM183329 | 956-1597 | 1680 |
| Group VII | dFr45 | Not Available | AX741144 | Not Available | 1672 |
| | dFr2158 | Cameroon | AM183333 | 949-1593 | 1674 |
| Group VIII | dFr2072 | Senegal | AM183330 | 1011-1598 | 1681 |
| | dFr644 | Cogo-Brazzaville | AX741169 | Not Available | 1680 |
| | dFr2736 | Cote d' Ivoire | AM183327 | 954-1598 | 1681 |
| | HDV-DN79 | Gabon | GU177114 | 1010-1597 | 1679 |
| Group IX | Isolate-1 | Turkey | HQ005370 | 785-1372 | 1679 |
| | Isolate-2 | Turkey | HQ005371 | 785-1372 | 1679 |
| | Isolate-3 | Turkey | HQ005372 | 774-1361 | 1673 |
| | Isolate-4 | Turkey | HQ005369 | 786-1373 | 1684 |
| | Isolate-5 | Turkey | HQ005364 | 730-1374 | 1681 |
| | Isolate-6 | Turkey | HQ005366 | 786-1373 | 1683 |
| | Isolate-7 | Turkey | HQ005365 | 729-1373 | 1682 |
| | Isolate-8 | Turkey | HQ005367 | 730-1374 | 1685 |
| | Isolate-9 | Turkey | HQ005368 | 786-1373 | 1681 |

**Table 3:** Predicted siRNA target for genome isolates of HDV, worldwide

| Accession No | Target | Location of target within gene | siRNA target sequence within gene |
|---|---|---|---|
| X85253 | Target-1 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |
| | Target-2 | 451-473 | GTCCTCTAACTTCTTGATTTTCT |
| U81989 | Target-3 | 454-476 | CTCGAGTTTCTTGATCTTCTTCT |
| AY633627 | Target-4 | 408-430 | TCCGAGAATTCCTTTGATGTTTC |
| X04451 | Target-5 | 472-494 | TCCGAGAATTCCTTTGATGTTCC |
| M92448 | Target-6 | 469-491 | TCCGAGAATTCCTTTGATGTTTC |
| AF104263 | Target-7 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |
| | Target-8 | 469-491 | TCCGAGAATTCCTTTGATGTTTC |
| D01075 | Target-9 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |
| L22066 | Target-10 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |
| AB118849 | Target-11 | 511-533 | CTCAAGTATCTTGATATTCTTCT |
| EF514905 | Target-12 | 454-476 | CTCGAGTTTCTTGATCTTCTTCC |
| | Target-13 | 511-533 | CTCGAGTTTCTTGATCTTCTTCC |
| EF514904 | Target-14 | 454-476 | CTCGAGTTTCTTGATCTTCTTCC |
| | Target-15 | 511-533 | CTCGAGTTTCTTGATCTTCTTCC |
| EF514906 | Target-16 | 454-476 | CTCGAGTTTCTTGATCTTCTTCC |
| | Target-17 | 511-533 | CTCGAGTTTCTTGATCTTCTTCC |
| EF514907 | Target-18 | 454-476 | CTCGAGTTTCTTGATCTTCTTCC |
| | Target-19 | 511-533 | CTCGAGTTTCTTGATCTTCTTCC |
| EF514903 | Target-20 | 454-476 | CTCGAGTTTCTTGATCTTCTTCC |
| | Target-21 | 511-533 | CTCGAGTTTCTTGATCTTCTTCC |
| AJ307077 | Target-22 | 412-434 | TCCGAGAATTCCTTTGATGTTCC |
| | Target-23 | 441-463 | AGGGATTTTCGTCCTCTATCTTC |
| | Target-24 | 451-473 | GTCCTCTATCTTCTTGAGTTTCT |
| AY648959 | Target-25 | 412-434 | TCCGAGTATTCCTTTGATGTTTC |
| AY648957 | Target-26 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |
| AY648958 | Target-27 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |
| AY648956 | Target-28 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |

| | | | |
|---|---|---|---|
| HM046802 | Target-29 | 412-434 | TCCGAGAATTCCTTTGATGTTCC |
| | Target-30 | 441-463 | AGGGATTTTCGTCCTCTATCTTC |
| | Target-31 | 451-476 | GTCCTCTATCTTCTTGAGTTTCT |
| | Target-32 | 469-491 | TCCGAGAATTCCTTTGATGTTCC |
| | Target-33 | 498-520 | AGGGATTTTCGTCCTCTATCTTC |
| | Target-34 | 508-530 | GTCCTCTATCTTCTTGAGTTTCT |
| AB118845 | Target-35 | 48-70 | GAGGGTTAACCTACGGAAATTCC |
| AB118847 | Target-36 | 48-70 | GAGGGTTAACCTACGGAAATTCC |
| AM183332 | Target-37 | 551-573 | TCGAGTTCATCTCGATTTTTTCT |
| HQ005370 | Target-38 | 237-259 | CTCCAGAAAGTTGCTTCTTCTTG |
| | Target-39 | 412-434 | TCCGAGAATTCCTTTGATGTTCC |
| HQ005371 | Target-40 | 410-432 | TCCGAGAATTCCTTTGATGTTTC |
| | Target-41 | 494-516 | TCGAGTTCTTCTAACTTTTTTCT |
| HQ005372 | Target-42 | 237-259 | CTCCAGATAGTTGCTTCTTCTTG |
| | Target-43 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |
| HQ005364 | Target-44 | 469-491 | TCCGAGAATTCCTTTGATGTTTC |
| HQ005366 | Target-45 | 412-434 | TCCGAGAATTCCTTTGATGTTTC |
| HQ005365 | Target-46 | 469-491 | TCCGAGAATTCCTTTGATGTTCC |
| | Target-47 | 554-576 | TCGAGTTCATCGGCTCTTTTTCT |
| HQ005367 | Target-48 | 469-491 | TCCGAGAATTCCTTTGATGTTCC |

**Table 4**: Two effective siRNA molecule with GC%, free energy of folding and free energy of binding with target

| Target | Location of target within mRNA | siRNA target within in consensus target | Predicted siRNA duplex | GC% | Free energy of folding of siRNA candidate at $37^0$C | Free energy of Binding with target |
|---|---|---|---|---|---|---|
| Consensus Target-1 | 478-501 | TCCGAGAATTCCTTTGATGTTTCT | AACAUCAAAGGAAUUCUCGGA CGAGAAUUCCUUUGAUGUUUC | 47.1 | 0.15 | -21.21 |
| Consensus Target-2 | 508-530 | GTCCTCTATCTTCTTGATTTTCT | AAAAUCAAGAAGAUAGAGGAC CCUCUAUCUUCUUGAUUUUCU | 36.8 | -0.82 | -9.41 |