Original Article

Computational Approach to Design a Potential siRNA Molecule to Silence the Nucleocapsid Gene of Different Nipah Virus Strains of Bangladesh

Arafat Rahman Oany, Mohammad Uzzal Hossain and Shah Adil Ishtiyaq Ahmad*

Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University, Santosh, Tangail – 1902, Bangladesh.

ABSTRACT: Nipah virus (NiV) is a single stranded negative-strand RNA virus which was first identified in Malaysia and Singapore. It causes fatal respiratory illness and encephalitis in humans. The progression of this infection is very rapid and causes death within 18 days of infection in most of the cases. Scientists are yet to find out a successful therapeutic approach to treat the NiV affected people. In this study, we have tried to utilize different computational methods to design a siRNA molecule to silence the Nucleocapsid gene of Nipah virus. The nucleocapsid protein (N) gene is one of the most suitable targets for the diagnosis and treatment of NIVs. To find out a common siRNA molecule against five different Nipah virus strains, at first the complete nucleocapsid gene sequences of these five strains were collected from NCBI database. SiDirect 2.0 server was used to design a siRNA molecule against these strains. The siRNA molecule was checked for its secondary structure and GC content using Mfold and OligoCalc server respectively. Blast tool was used to identify any off-target similarity with the designed siRNA molecule. In this study we have proposed a duplex siRNA molecule for N gene silencing of five different strains of NiVs. The siRNA molecule was rationally designed and authenticated using different computational methods. It was also confirmed that the designed siRNA does not match with any off-target sequences. There is no effective therapy available at present to cure Nipah virus infections. The designed siRNA molecule might provide an alternative therapeutic approach against the five major Nipah virus strains in Bangladesh.

KEYWORDS: molecular therapy, therapeutic siRNA, siDirect, Reynolds rule.

INTRODUCTION

Nipah virus (NiV), a family of deadly bat-borne *Paromyxoviridae* was first identified in Malaysia and Singapore, when it caused 109 deaths in 283 cases during 1998–1999.[1] They are ssRNA (single stranded RNA) negative-strand enveloped viruses that cause respiratory illness and encephalitis in humans.[2-6] The progression of this infection is very rapid and have a case-fatality ratio of 40%–76%. The disease progression usually starts after 4-18 days of incubation period and starting with mild illness (fever, headache and myalgia) it can reach upto coma and death within 10 days.[7-10] It has been reported that, about 86–93% human NiV encephalitis infections occurred due to the occupational exposure to pigs as an intermediate host for the initial NiV outbreaks.[11-14] Many countries like Malaysia, Bangladesh, Australia, Thailand and Cambodia have reported *Pteropus* bats as a vector of this disease.[15-17] Bangladesh had experienced a high fatality rate of NiV infection during 2001 to 2007. During this period, among 122 identified cases more than 70% (87) patients died.[18] There is not much success in finding a potential treatment approach against NiV infections.[19] With the advancement in sequencing technologies, the whole genome sequence of different strains of this deadly virus are now available on different databases. These resources are opportunities for the researchers to look for an effective treatment against the Nipah virus.

Nucleocapsid proteins of different viruses play a major role in encapsidation, regulating the transcription and replication of the viral genome. For this reason study of the nucleocapsid protein or N gene has become much popular as a diagnostic and therapeutic target.[20]

Protection mechanism against different viruses through designing potential Small Interfering RNA (siRNA) molecules targeting the N gene has already been reported.[21] siRNA or silencing RNA molecules are double stranded RNA containing 20-25 base pairs. It usually works through RNA interference (RNAi) pathway, where it interferes with the expression of specific genes that is complimentary to nucleotide sequences, causing mRNA to be broken down following transcription.[22] Post-transcriptional gene silencing (PTGS) is the main role of these small molecules.[23] The present study is to design a potential siRNA molecule to inhibit the translation of the N gene and thus prevent the encapsidation, transcription and replication process of the NiVs.

MATERIALS AND METHODS

Viral strain selection
ViralZone (http://viralzone.expasy.org/) of the Ex-PASy Bioinformatics Resource Portal was used for the selection of Nipah virus strains and their associated information including their genus, family, host, transmission pattern, disease pathogenicity, genome, proteome and also the available therapeutic agents against them.

Sequence retrieval and evolution analysis
Five complete cds of nucleocapsid protein (N) gene of different nipah virus (BDG) strains were collected from the viral gene bank database available at NCBI (http://www.ncbi.nlm.nih.gov/). Multiple sequence alignment of the retrieved sequences was done by T-Coffee server[24] (http://tcoffee.crg.cat/apps/tcoffee/do:regular) and phylogeny was analyzed by jalview 2 tool.[25]

Target identification and rational siRNA molecule designing
SiDirect 2.0[26] (http://siDirect2.RNAi.jp/) web server, was used for efficient and target-specific siRNA design for mammalian RNAi. It utilized combined rules including Ui-Tei, Amarzguioui and Reynolds rules[27-30] and melting temperature (T_m) below 21.5°C for siRNA duplex, as parameter. For further verification of predicted molecules, GenScript[31] siRNA Target Finder (http://www.genescript.com/-index.html) was also applied. Besides these other parameters were taken on the concept of algorithms given in Table 1.

Off target similarity search
Blast tool[32] (http://www.ncbi.nlm.nih.gov/blast) was used to identify off target similarity with any sequence on whole Genebank datasets other than the target sequence by applying expected threshold value 10 and BLOSUM 62 matrix as parameter.

GC calculation and secondary structure prediction
For GC content calculation of predicted siRNA, OligoCalc[33] oligonucleotide Properties Calculator (http://basic.northwestern.edu/biotools/OligoCalc.html) was used while Mfold server[34] (http://www.mfold.-rna.albany.edu/) was used for secondary structure prediction aimed to compute the free energy of folding.

Calculation of RNA-RNA interaction through thermodynamics
To study the thermodynamics of interaction between predicted siRNA and target gene, RNAcofold program[35] (http://rna.tbi.univie.ac.at/cgi-bin/RNAco-fold.cgi) was used. It calculates the hybridization energy and base-pairing form of two RNA sequences. It functions as extension of McCaskill's partition function algorithm to compute probabilities of base pairing, realistic interaction energies and equilibrium concentrations of duplex structures.

Validation of the predicted siRNA
The predicted siRNA was further validated using the siRNAPred[36] (http://imtech.res.in/raghava) server from Imtech. Here we used efficacy prediction for 21 mers. The predicted siRNA was screened against the Main21 dataset using binary pattern.

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.	✓ GC content 30% to 52% (1 points)
✓ G/C at the 5' terminus of the antisense strand	✓ Strong binding of 5' sense strand	✓ Occurrence of 3 or more A/U base pair at position 15-19 of sense strand (Each A/U base pair in this region earns 1 point)
✓ At least 4 A/U residues in the 5' terminal 7 bp of sense strand	✓ No U at position 1.	✓ Low internal stability at target site (melting temperature Tm>-20°c) (1 point)
✓ No GC stretch longer than 9nt	✓ Presence of A at position 6.	✓ Presence of A at position 19 of the sense strand (1 point)
	✓ Weak binding of 3' sense strand.	✓ Presence of A at position 3 of the sense strand (1 point)
		✓ Presence of U at position 10 of the sense strand (1 point)
		✓ Absence of G at position 13 of the sense strand (1 point)
		✓ Threshold for efficient siRNAs score>=6

RESULTS AND DISCUSSION

The nucleocapsid protein (N) gene cds of different Nipah virus strains isolated from different regions of Bangladesh (NIVBGD2010FARIDPUR2, NIVBGD-2004RAJSHAHI, NIVBGD2004RAJBARI2, NIVBG-D2010GOPALGANJ and NIVBGD2004FARIDPUR) were initially selected to design a siRNA molecule. The accession numbers of these five complete cds are JN808859.1, JN808862.1, JN808861.1, JN808860.1 and JN808858.1 respectively. Multiple sequence alignment analyzed by T-Coffee server among the different NiV cds showed that most of the sequences were conserved. Phylogenetic analysis by jalview-2 tool revealed that the sequences had a common ancestor and retained some significant similarity during the course of evolution (Figure 1).

The siDirect web-based online software system computes highly effective small interfering RNA from the retrieved sequences with maximum target

specificity. The proposed consensus target for siRNA with location is shown in Table 2. siDirect predicts siRNA through calculating the Tm of the seed target duplex using the nearest neighbor model and the thermodynamic parameters.
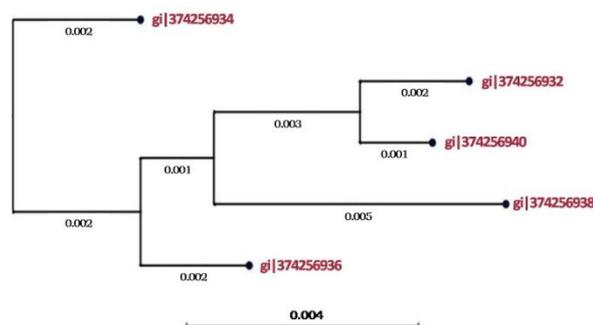
Figure 1. Phylogenetic analysis of the Nucleocapsid (N) gene of different nipah virus (BDG) strains.

Table 2. Predicted siRNA consensus target of the Nucleocapsid (N) gene cds.

Accession No.	Target	Location of target within the gene	siRNA target sequence within gene
JN808859.1		675-697	GTCAGTAAGAAAGTTTATGGTTG
JN808862.1		675-697	GTCAGTAAGAAAGTTTATGGTTG
JN808861.1	Consensus Target	675-697	GTCAGTAAGAAAGTTTATGGTTG
JN808860.1		675-697	GTCAGTAAGAAAGTTTATGGTTG
JN808858.1		675-697	GTCAGTAAGAAAGTTTATGGTTG

Table 3. The proposed siRNA molecule with GC%, free energy of folding and free energy of binding with target.

Target	Predicted duplex siRNA candidate at 37°C	GC%	Free energy of folding with target	Free energy of binding (Kcal/mol)
Consensus Target	ACCAUAAACUUUCUUACUGAC CAGUAAGAAAGUUUAUGGUUG	33.33	0.00	−28.17

The formula for calculating the T_m is:
$$T_m = \{(1000 \times \Delta H) / (A + \Delta S + R\ln (CT/4))\} -273.15+16.6 \log [Na+] \quad (1)$$

Where ΔH (kcal/ mol) is the sum of the nearest neighbor enthalpy change, A is the helix initiation constant (-10.8), ΔS are the sum of the nearest neighbor entropy change.[37] 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.

The GC content of our targeted siRNA was calculated 33.3% by the OligoCalc Oligonucleotide Properties Calculator. Mfold server predicts RNA secondary structure through widely used algorithms, which are based on a search for the minimal free energy state.[38] Here, siRNA molecule is having zero free energy of folding at 37° C (Table 3). The RNAcofold server from Vienna RNA web services calculated the hybridization energy and base-pairing pattern of the RNA sequences (Table 3). The GenScript server output also supported the siDirect output. siRNAPred server assessed the 21mer siRNA through the Support

vector machine based methods with high accuracy. The Main21 dataset of the siRNAPred server consists of 2182 siRNAs (21mer) derived from homogeneous experimental conditions. The binary pattern was selected for the validation which results the prediction score of 0.872. This score lays within the range of high efficacy prediction score for siRNAPred server which is 0.8-0.9.

Therapeutic approaches based on siRNA have huge potential to provide treatment for diseases that are caused by a single or a known set of genes, such as inherited genetic defects, viral diseases, autoimmune disorders and cancers.[39] The potentiality of the siRNA to silence specific gene is already approved against many viruses like hepatitis C virus,[40] HIV-1 infection[41] and herpes simplex virus 2 infections.[42] Designing an effective siRNA molecule targeting a specific gene have some challenges. Apart from finding an effective delivery method other challenges are siRNA stability, off-target silencing and activation of immune response.[43] Scientists have developed some rules to predict the efficacy of a designed siRNA molecule prior to in vivo analysis – namely Ui-Tei

rules, Amarzguioui rules and Reynolds rule.[44] The siRNA molecule designed in this study follows all these rules. The threshold score for Reynolds rules covered by the proposed siRNA is 6, which indicates an efficient siRNA. The GC content of the proposed siRNA was also found to be within the supported range 30-52%. [45] The prediction of thermodynamics of RNA-RNA interaction which is another important parameter for siRNA efficiency was predicted by the RNAcofold server. Here the two RNA sequences are concatenated and the point of concatenation is specified by an ampersand. In addition to ΔGAB, the free energy of the heterodimer of sequence A and sequence B can be calculated using the equation,

$$\Delta G_{binding} = \Delta G_{AB} - \Delta G_A - \Delta G_B \quad (2).[46]$$

So, for the interaction of consensus target and its predicted siRNA, free energy of binding is

$$\Delta G_{binding} = \Delta G_{AB} - \Delta G_A - \Delta G_B$$
$$= -28.86 - (-0.141582) - (-0.548420)$$
$$= -28.17 \text{ Kcal/mol.}$$

To check for any off-target silencing effect of the proposed siRNA the NCBI blast program was used and no such effects were found. The other challenge for siRNA designing – the activation of immune response, cannot be measured without in vivo experimentation. From all the above analysis it can be said that the proposed siRNA molecule meets all the desired criteria to be considered a potential siRNA and might play an important role in silencing the nucleocapsid gene of different NiVs. The multiple sequence alignment showed that the target region for siRNA in the N gene cds predicted by siDirect server was completely conserved (Figure 2). This result supports the probable use of the proposed siRNA against the five strains that were used in this study.

```
JN808859.1 GTCAGTAAGAAAGTTTATGGTTG
JN808862.1 GTCAGTAAGAAAGTTTATGGTTG
JN808861.1 GTCAGTAAGAAAGTTTATGGTTG
JN808860.1 GTCAGTAAGAAAGTTTATGGTTG
JN808858.1 GTCAGTAAGAAAGTTTATGGTTG
           ***********************
```

Figure 2. Multiple sequence alignment of the consensus target within the Nucleocapsid (N) gene from five different strains of Nipah virus. The MSA shows an entirely conserved target.

CONCLUSION

Molecular therapy is a promising alternative therapeutic approach where conventional therapeutic approaches fail. SiRNAs have already proved its potential to provide specific and effective cure for many diseases which was otherwise hard to cure. In this study, we have proposed a siRNA molecule to be used as a therapeutic against Nipah virus. We have also calculated different parameters to support the stability and functioning of the proposed siRNA molecule. Since molecular therapies are specific to DNA sequences, such therapies are sometimes non-functional against different strains. Our designed

siRNA are expected to overcome this issue as it has targeted a conserved sequence found in five major Nipah virus strains. Although proper *in vitro* and *in vivo* validation is still required, we expect that this siRNA molecule will provide an effective treatment approach against the targeted Nipah virus strains.

REFERENCES

1. Chua, K. B. et al. 2000. Nipah virus: a recently emergent deadly paramyxovirus. *Science*. **288**(5470), 1432–1435.
2. Field, H. et al. 2001. The natural history of Hendra and Nipah viruses. *Microbes Infect*. **3**(4), 307–314.
3. Gurley, E. S. et al. 2007. Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerg Infect Dis*. **13**(7), 1031–1037.
4. Harcourt, B. H. et al. 2000. Molecular characterization of Nipah virus, a newly emergent paramyxovirus. *Virology*. **271**(2), 334–349.
5. Parashar, U. D. et al. 2000. Case-control study of risk factors for human infection with a new zoonotic *paramyxovirus*, Nipah virus, during a 1998-1999 outbreak of severe encephalitis in Malaysia. *J Infect Dis*. **181**(5), 1755–1759.
6. Wong, K. T. et al. 2002. Nipah virus infection: pathology and pathogenesis of an emerging *paramyxoviral* zoonosis. *Am J Pathol*. **161**(6), 2153–2167.
7. Chua, K. B. 2003. Nipah virus outbreak in Malaysia. *J Clin Virol*. **26**(3), 265–275.
8. Goh, K. J. et al. 2000. Clinical features of Nipah virus encephalitis among pig farmers in Malaysia. *N Engl J Med*. **342**(17), 1229–1235.
9. International Centre for Diarrheal Disease Research. Bangladesh. Nipah encephalitis outbreak over wide area of western Bangladesh. *Health and Science Bulletin*. 2004. **2**, 7–11.
10. World Health Organization. Nipah virus outbreak(s) in Bangladesh, January – April 2004. *Wkly Epidemiol Rec*. 2004. **79**, 168-171.
11. Tambyah, P. A., Tan, J. H., Ong, B. K. C., Ho, K. H. and Chan, K. P. 2001. First case of Nipah virus encephalitis in Singapore. *Intern Med J*. **31**(2), 132–133.
12. Chan, K. P. et al. 2002. A survey of Nipah virus infection among various risk groups in Singapore. *Epidemiol Infect*. **128**(1), 93–98.
13. Chew, M. H. L. et al. 2000. Risk factors for Nipah virus infection among abattoir workers in Singapore. *J Infect Dis*. **181**(5), 1760–1763.
14. Paton, N. I. et al. 1999. Outbreak of Nipah-virus infection among abattoir workers in Singapore. *Lancet*. **354**(9186), 1253–1256.
15. Yob, J. M. et al. 2001. Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia. *Emerg Infect Dis*. **7**(3), 439–441.
16. Reynes, J. M. et al. 2005. Nipah virus in Lyle's flying foxes, Cambodia. *Emerg Infect Dis*. **11**(7), 1042–1047.
17. Olson, J. G. et al. 2002. Antibody to Nipah-like virus in bats (*Pteropus lylei*), Cambodia. *Emerg Infect Dis*. **8**(9), 987–988.
18. Luby, S. P. et al. 2009. Recurrent Zoonotic Transmission of Nipah Virus into Humans, Bangladesh, 2001–2007. *Emerg Infect Dis*. **15**(8), 1229-1235.
19. Hulo, C. et al. 2011. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res*. **39** (suppl 1), D576-D582.
20. Eshaghi, M. et al. 2005. Purification and characterization of Nipah virus nucleocapsid protein produced in insect cells. *J Clin Microbiol*. **43**(7), 3172-3177.
21. Ge, Q. et al. 2003. RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription. *Proc Natl Acad Sci USA*. **100**(5), 2718–2723.
22. Agrawal, N. et al. 2003. RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev*. **67**(4), 657-685.

23. Hamilton, A. and Baulcombe, D. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*. **286**(5441), 950–952.

24. Notredame, C., Higgins, D. G. and Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. **302**(1), 205-217.

25. Waterhouse, A. M. et al. 2009. Jalview Version 2 a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **25**(9), 1189–1191.

26. Naito, Y., Yoshimura, J., Morishita, S. and Ui-Tei, K. 2009. siDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics*. **10**(1), 392.

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

28. Reynolds, A. et al. 2004. Rational siRNA design for RNA interference. *Nat Biotechnol*. **22**(3), 326-330.

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

30. Ui-Tei, K., Naito, Y., Nishi, K., Juni, A. and Saigo, K. 2008. Thermodynamic stability and Watson-Crick base pairing in the seed duplex are major determinants of the efficiency of the siRNA-based off-target effect. *Nucleic Acids Res*. **36**(22), 7100-7109.

31. Wang, L. and Mu, F. Y. 2004. A Web-based design center for vector-based siRNA and siRNA cassette. *Bioinformatics*. **20**(11), 1818-1820.

32. Johnson, M. et al. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res*. **36** (suppl 2), W5-W9.

33. Kibbe, W. A. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res*. **35** (suppl 2), W43-W46.

34. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. **31**(13), 3406-3415.

35. Gruber, A. R. et al. 2008. The Vienna RNA websuite. *Nucleic Acids Res*. **36** (suppl 2), W70-W74.

36. Kumar, M., Lata, S. and Raghava, G.P.S. 2009. siRNApred: SVM based method for predicting efficacy value of siRNA. *Proceedings of the OSCADD-2009: International Conference on Open Source for Computer Aided Drug Discovery: 22-26 March 2009; IMTECH, Chandigarh.*

37. Freier, S. M. et al. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci USA*. **83**(24), 9373-9377.

38. Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science*. **244**(4900), 48-52.

39. Aagaard, L. and Rossi, J.J. 2007. RNAi therapeutics: principles, prospects and challenges. *Adv Drug Deliv Rev*. **59**(2-3), 75-86.

40. Zhang, J. et al. 2004. Down-regulation of viral replication by adenoviral-mediated expression of siRNA against cellular cofactors for hepatitis C virus. *Virology*. **320**(1), 135-143.

41. Qin, X. F. et al. 2003. Inhibiting HIV-1 infection in human T cells by lentiviral-mediated delivery of small interfering RNA against CCR5. *Proc Natl Acad Sci USA*. **100**(1), 183-188.

42. Palliser, D. et al. 2006. An siRNA-based microbicide protects mice from lethal herpes simplex virus 2 infection. *Nature*. **439**(7072), 89-94.

43. Gavrilov, K. and Saltzman, W.M. 2012. Therapeutic siRNA: principles, challenges, and strategies. *Yale J Biol Med*. **85**(2), 187-200.

44. Taxman, D. J. et al. 2006. Criteria for effective design, construction, and gene knockdown by shRNA vectors. *BMC Biotechnol*. **6**(1), 7.

45. Chan, C.Y. et al. 2009. A structural interpretation of the effect of GC-content on efficiency of RNA interference. *BMC Bioinformatics*. **10** (suppl 1), S33.

46. Bernhart, S. H. et al. 2006. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol*. **1**(1), 3.