



AmiRzyn: PERL Centered Artificial MicroRNA Designing Aid

Joseph Baby¹, Nair Vrundha M^{1,2} and Chandy Susheel G.²

¹Interdisciplinary Research Centre, Malankara Catholic College, Mariagiri, Kaliakkavilai- 629 153, Tamil Nadu, INDIA

²Department of Bioinformatics, Malankara Catholic College, Mariagiri, Kaliakkavilai- 629 153, Tamil Nadu, INDIA

Available online at: www.isca.in

Received 26th June 2012, revised 3rd July 2012, accepted 6th July 2012

Abstract

RNA interference, a gene knock down tool has become a promising technology for researchers. Gene silencing, can be fastened by the process of designing small non- coding RNAs through computational tools. One of the current advancement in this field is artificial microRNA. It exploits the backbone of a miRNA for designing siRNA. The endogenous nature of the miRNA provides more success to this technology. In the current study, a tool named Amirzyn was developed for designing artificial microRNA using PERL language, which is very powerful in string manipulations. The tool facilitates the user to input their DNA sequences and by means of pre- defined parameters, the possible siRNA sequences are predicted as output. Possibilities of occurrence of off- target effects are verified by performing BLAST comparison. For amiRNA designing, AmiRzyn permits the specification of restriction site and miRNA backbone of user's requirement, in addition to the predicted siRNA sequence as the input. AmiRzyn serves as a promising aid for researchers, in the field of gene silencing, to design artificial microRNAs.

Keywords: miRNA, Restriction sites, siRNA prediction, nucleotide sequence, off- target.

Introduction

The new field of RNA interference (RNAi) based genomics has revolutionized studies to determine the role of a gene and for the development of modern therapeutics. Artificial microRNAs (amiRNAs) are single- stranded 21mer RNAs, which are processed from endogenous microRNA (miRNA) precursors¹. The ability of miRNAs to silence target genes was first demonstrated in 1998 by Andrew Fire with his co- workers and has since emerged as a revolutionary strategy to reduce target gene expression². Artificial miRNAs (amiRNAs) that mimic natural miRNA structures have been designed to target endogenous genes and have also proven to be effective and promising technology^{3,4}.

At the time of its discovery in plants this process was called post- transcriptional gene silencing, but after confirming its existence in almost all eukaryotes, by today it is known as RNA interference. Major classes of small RNAs include microRNAs, small interfering RNAs (siRNAs), short hairpin RNAs (shRNAs) and artificial miRNA; differ with respect to their biogenesis^{5,6}. The small RNAs are specifically selected by RNA- induced silencing complex (RISC) which will target the specific mRNAs based on sequence complementarity^{5,7}. Silencing is initiated when the Dicer enzyme processes the double stranded RNA. Only one of the two strands, which is known as the guide strand, binds the argonaute protein and directs gene silencing. The other anti- guide strand or passenger strand is degraded during RISC activation^{8,9}. These small RNAs (sRNAs) are then incorporated into the RISC to guide the degradation or translation repression of cRNA targets¹⁰. These small RNAs then directly interact with Argonaute (AGO)

proteins to form the core of the effector complex, the RISC. The RISC uses the small RNAs as guides to act on their targets based on sequence complementarity, leading to the cleavage of target mRNAs or the repression of translation¹¹. Among all the methods used, amiRNA has become a promising RNAi technology. In recent years there has been an immense interest in using RNAi technology as therapeutic agents for viral infections and in future these molecules can represent an alternative technology to small molecular compounds¹². The potential of RNAi is explored in therapeutic fields, especially in cancer

Artificial miRNAs or miRNA shuttles, mimic naturally occurring pri- miRNAs and are generated by exchanging the miRNA/miRNA* sequence within miRNA precursor genes, while maintaining the pattern of matches and mismatches in the fold back. For amiRNAs the stem loop region or the miRNA replaced by siRNA duplexes, thus creating a miRNA based hairpin which serves to shuttle siRNA sequences into the RNAi pathway. AmiRNA have been generated using a number of naturally occurring pri- miRNAs as scaffolds for siRNA sequences^{13,14}. AmiRNA vectors can be used not only to silence most protein coding genes, but also to knock down the genes encoding the enzymes/proteins of RNAi pathways. The main advantages of amiRNAs when compared to siRNA vectors are their high specificity, fewer off- target effects, tissue specific expression and almost no side effects.

Bioinformatics is a rapidly developing branch of biology and is highly interdisciplinary. Although its fields of applications are wide, it focuses mainly on research activities. Computational biology is the branch of bioinformatics that deals with creation

of algorithms and tools for the analysis of biological data. Among the various languages used for bioinformatics programming, PERL stands out from the rest as a really strong and highly favored language due to its various features¹⁵.

In the present study, an artificial microRNA prediction tool named "AmiRzyn" is designed using PERL program. The main aim of this program is to design an efficient and best, user friendly and easy prediction of amiRNA sequences, among the vast array of already available bioinformatics tools associated to this field. The text field related advantages of PERL have been exploited in this program to design the amiRNA. The AmiRzyn has been designed to allow the user to choose the restriction sites and the miRNA backbone.

Material and Methods

PERL: PERL is a popular programming language that is extensively used in areas such as Bioinformatics and web programming. PERL has become popular with biologists because it is so well suited to several Bioinformatics tasks. Perl is often used as a glue language, tying together systems and interfaces that were not specifically designed to inter- operate. These combinations make Perl a popular all- purpose language for system administrators, particularly because short programs can be entered and run on a single command line¹⁶.

In the present study, the reason for selecting PERL as the core coding language is the high text manipulating ability, an authoritative feature of PERL. The siRNA prediction part of AmiRzyn is based entirely on PERL. Since an entire DNA sequence, containing thousands of bases, has to be checked for the occurrence of an approximately 21- nucleotide long pattern, the advantage of data- set manipulation ability was utilized. When the amiRNA designing part is considered, an integration of PERL and .NET framework could be explored. The figure-1 shows the workflow performed in AmiRzyn.

siRNA Prediction: The powerful text processing features of PERL language is utilized in order to detect the siRNA sequence. The first step in amiRNA design is the detection of siRNA sequences from the given input sequences. Firstly, the tools uses the pre- defined parameters listed below, in order to predict the possible siRNA sequences.

Parameters for siRNA Prediction: i. Avoid regions within 50-100 base pairs of the start codon and the termination codon, ii. The siRNA targeted sequence usually should be of 19-21 nucleotide in length¹⁷. iii. Avoid stretches of 4 or more bases such as AAAA, CCCC, iv. Avoid regions with GC content <30% or > 60%, v. Tm value should be around -30 to - 60 range. vi. $T_m = 79.8 + 18.5 \cdot \log_{10}([Na^+]) + (58.4 \cdot GC\%/100) + (11.8 \cdot (GC\%/100)^2) - (820/Length)$, vii. Avoid repeats like AA and TT at the start and end of the predicted siRNA sequence.

By experimentally analyzing the silencing efficiency of 180 siRNAs targeting the mRNA of two genes and correlating it with various sequence features of individual siRNAs, Reynolds

and his team at Dharmacon, Inc., identified eight characteristics associated with siRNA functionality. These characteristics are used by rational siRNA design algorithm to evaluate potential targeted sequences and assign scores to them. Sequences with higher scores will have higher chance of success in RNAi. The table- 1 lists the criteria and the methods of score assignment. A sum score of 6 defines the cutoff for selecting siRNAs. All siRNAs scoring higher than 6 are acceptable candidates¹⁸.

Off- Target Effect Verification: BLAST is a rapid sequence comparison tool that uses a heuristic approach to construct alignments by optimizing a measure of local similarity^{19,20}. AmiRzyn provides the facility for the user to check if the predicted siRNA may cause any off- target effects. For this, a web link to the NCBI's BLAST resource is provided. After the similarity search, if BLAST resource retrieves homologous sequences, then the predicted siRNA sequence can cause off-target effects and further proceedings are discouraged. If the BLAST comparison results are negative, then the particular siRNA sequence can be used for amiRNA designing.

amiRNA Design: Concerning amiRNA design in AmiRzyn, the basic requirement is the predicted siRNA sequence. The miRNA backbone is replaced with 21 nucleotide long sense siRNA three nucleotides upstream of loop regions. AmiRzyn accepts the restriction site, miRNA backbone (user's selection) and the predicted siRNA sequence as the input and designs the amiRNA sequence.

AmiRzyn provides five restriction sites: EcoRI, EcoRII, BamHI, HindIII and TaqI (table- 2). It allows the user to select any one as per user's interests.

miRBase: The miRBase database is a searchable database of published miRNA sequences and annotation. The miRBase database was established in 2002 to provide microRNA researchers with stable and unique gene names for their novel microRNA discoveries and an archive of all microRNA sequences^{21,22}. The miRNAs used in the present study were retrieved from the miRBase.

AmiRzyn also provides the facility for the user to select a miRNA backbone (figure- 2) of interest to design amiRNAs (table- 3).

.NET: The .NET Framework is a software framework that runs primarily on Microsoft Windows. It includes a large library and provides language interoperability across several programming languages. Programs written for the .NET framework execute in a software environment known as the common language runtime (CLR), an application virtual machine that provides important services such as security, memory management and exception handling. The class library and the CLR together constitute the .NET Framework²³.

In the present study, this portability feature is utilized and thus integrated both PERL and .NET for creating AmiRzyn. While PERL is used in the core programming of AmiRzyn, the user interface has been developed using the .NET framework.

Table-1
Criteria for scoring siRNA Evaluation

Criteria	Description	Score	
		Yes	No
1.	Moderate to low (30%-50%) GC content	1 point	
2.	At least 3 A/Us at positions 15- 19 (sense)	1 point /per A or U	
3.	Lack of internal repeats (Tm<20)	1 point	
4.	A at position 19 (sense)	1 point	
5.	A at position 3 (sense)	1 point	
6.	U at position 10 (sense)	1 point	
7.	No G/C at position 19 (sense)		- 1 point
8.	No G at position 13 (sense)		- 1 point
9.	G/C at position 1,2,9 (sense)	1 point	
10.	A/U at position 3,6,7,13,18	1 point	
11.	A/C at position 11		- 1 point

Where., A- Adenine, U- Uracil, G- Guanine, C- Cytosine, Tm- Melting Temperature

Table-2
Restriction Sites and its features catalogued in AmiRzyn

S. No.	Restriction Site	Source	Length	Sequence
1.	EcoRI	<i>Escherichia coli</i>	12	5' GAATTC 3' CTTAAG
2.	EcoRII	<i>Escherichia coli</i>	10	5' CCWGG 3' GGWCC
3.	BamHI	<i>Bacillus amyloliquefaciens</i>	12	5' GGATCC 3' CCTAGG
4.	HindIII	<i>Haemophilus influenzae</i>	12	5' AAGCTT 3' TTCGAA
5.	TaqI	<i>Thermus aquaticus</i>	8	5' TCGA 3' AGCT

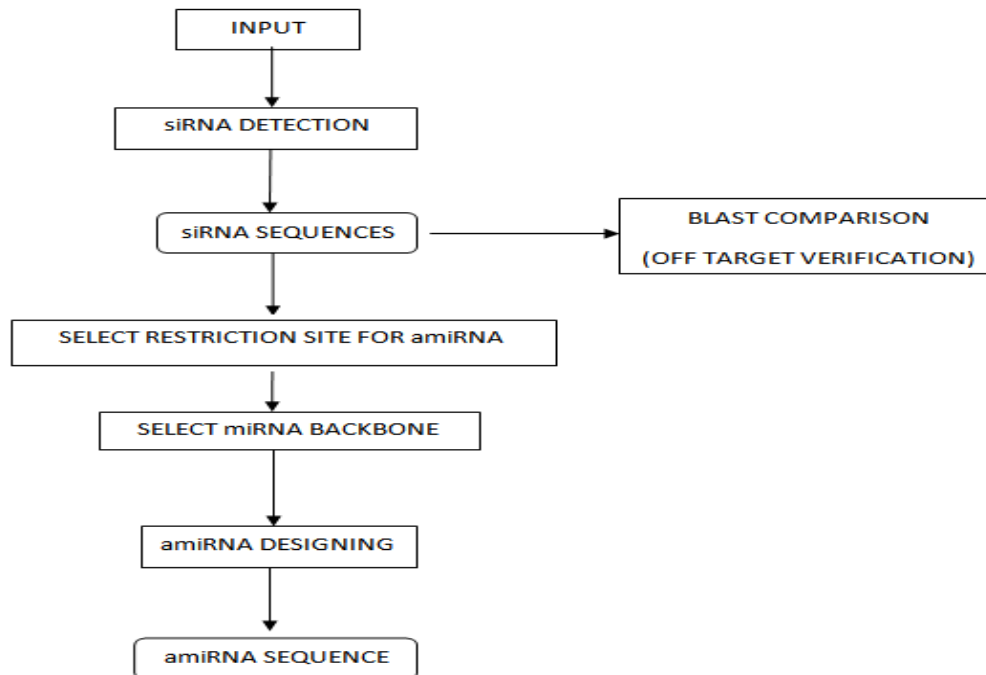


Figure-1
Pipeline of AmiRzyn

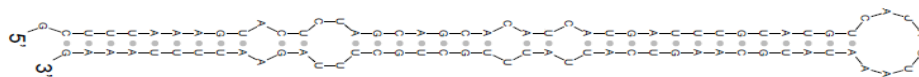


Figure-2(a)
Stem and Loop Structure of Mir15

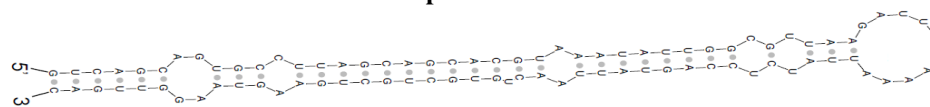


Figure-2(b)
Stem and Loop Structure of Mir16

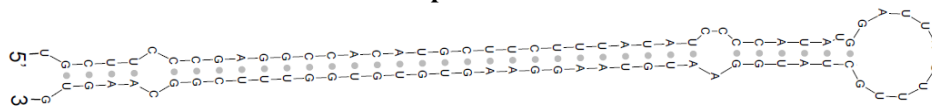


Figure-2(c)
Stem and Loop Structure of Mir 206

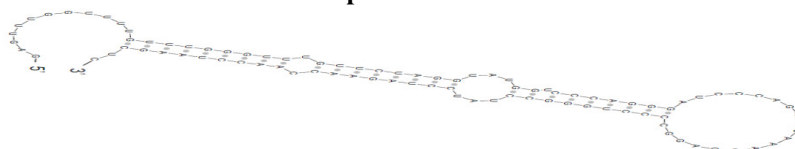


Figure-2(d)
Stem and Loop Structure of Mir331

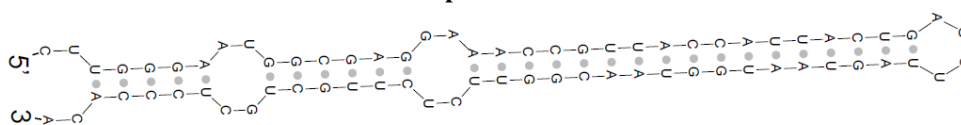


Figure-2(e)
Stem and Loop Structure of Mir451

Figure-2
Structure of the miRNAs used as backbone for amiRNA design

Results and Discussion

AmiRzyn, the artificial microRNA designing aid was created using PERL and .NET. It takes a DNA sequence from the user as the input and predicts the siRNA sequences using PERL codes. To avoid off- target effects, a web link to the online BLAST resource is provided. The user can select the Restriction Site and the miRNA backbone of interest.

The amiRNAs are designed in the following pattern: Restriction site (sense) – miRNA Backbone (sense) – siRNA (sense) – Loop – siRNA (Antisense) – miRNA Backbone (Anti sense) – Restriction site (Anti sense).

As shown in the figure- 3, the user can either paste or upload a sequence of interest as the input for AmiRzyn. It is not necessary that user should provide the accession number or other details of the sequence. When the Submit button is clicked, the sequence entered is taken as the input and the siRNA prediction begins. Clear button helps the user to remove the given sequence and enter another sequence of interest.

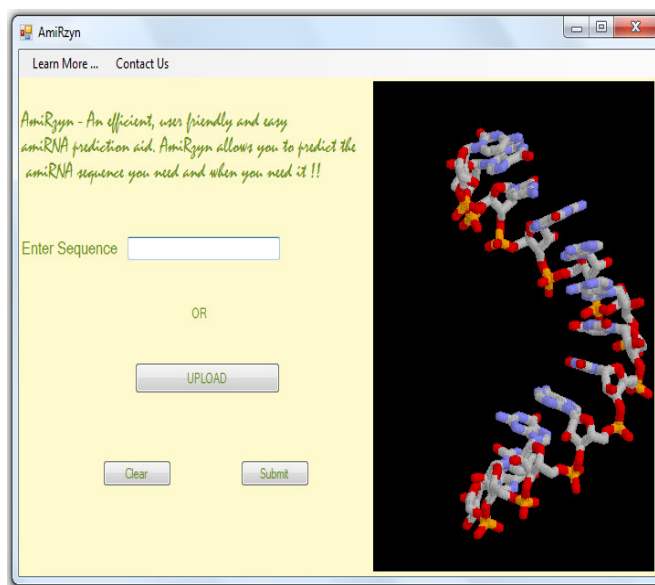


Figure-3
Homepage of AmiRzyn

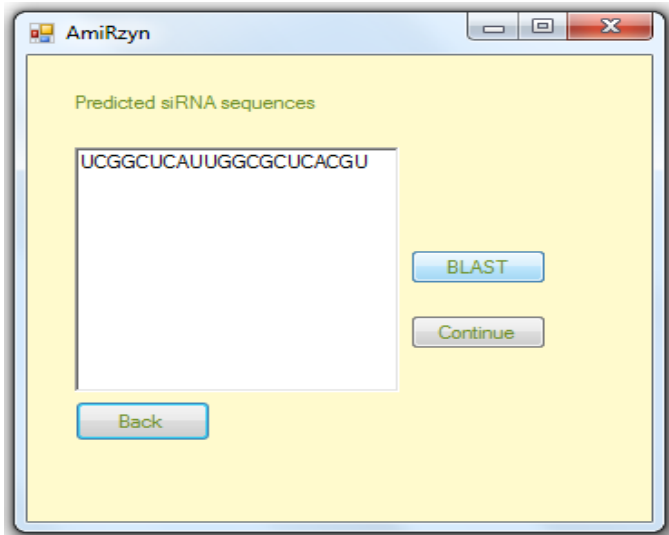


Figure-4
 siRNA result and Off- Target verification page

The page in figure-4 shows the predicted 21 nucleotide long siRNA sequence in the textbox. After obtaining the siRNA sequence, the user can perform off- target verification by using the BLAST resource. For this, AmiRzyn is provided with a web link to the NCBI's BLAST homepage. By clicking on the "BLAST" button, the user is taken to the relevant web page using the default browser. By clicking the "Continue" button, the user is taken to the next page to proceed with the amiRNA designing part. By clicking the "Back" button, the user is taken back to the homepage for performing another siRNA prediction. This is meant to be done when off target effect occurs.

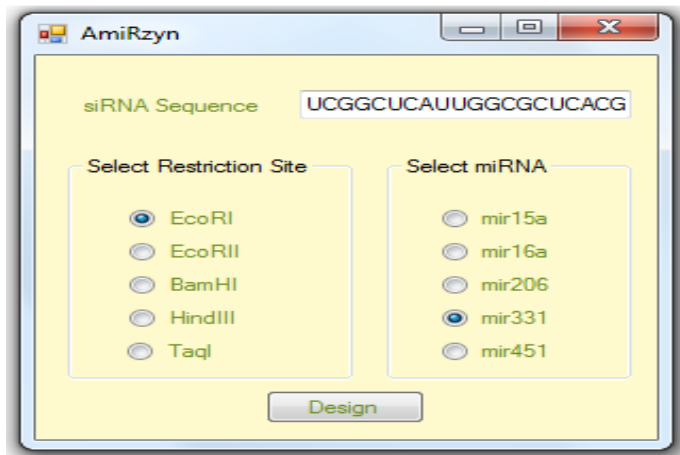


Figure-5
 amiRNA Design Page

Here, AmiRzyn permits the user to design the amiRNA using his/her restriction site and miRNA backbone of interest. The text box named siRNA sequence contains the predicted siRNA sequence. User can select one among the five restriction sites and one among the 5 miRNA backbones. After doing so, by clicking "Design" button, AmiRzyn accepts these inputs and starts designing the amiRNA (figure- 5).

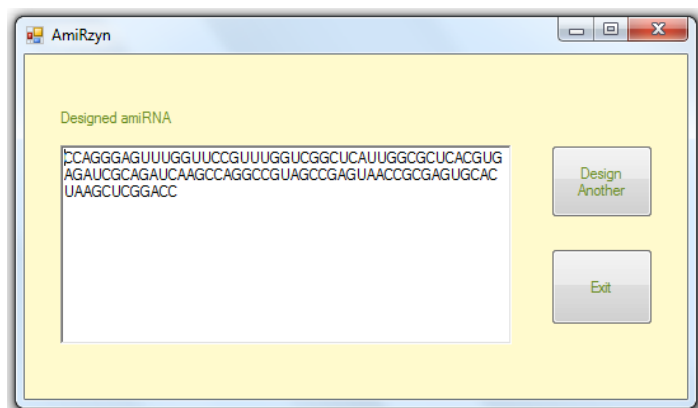


Figure-6
 Output of AmiRzyn

The figure- 6 depicts the final output page of AmiRzyn. As per the inputs given in the previous page, AmiRzyn designs the amiRNA sequence and displays it in this page. By clicking the "Design Another" button, the user can go back to the AmiRzyn homepage and start creating another amiRNA sequence. By clicking the "Exit" button, the user can exit AmiRzyn.

There are various tools available for designing siRNA provided by private and public organizations. siDirect²⁴, was a software system for computing siRNA with maximum target-specificity for mammalian RNA interference. AsiDesigner²⁵, a functional and medical genomics research tool in considering the alternative splicing it was designed for mRNA isoforms. There is a filtering tool for siRNA which combines some of the existing tools like Ambion, Dharmacon, Genscript, Qiagen, Invitrogen etc. The algorithm filters out the ineffective siRNA based on some new observations on the secondary structure²⁶. WMD 2 (Stephan, Joffrey, Rebecca, Markus and Detlef, personal communication) is the only artificial miRNA designing tool but it is solely for plant kingdom. AmiRzyn uniqueness lies in the integration of siRNA in to miRNA backbone. This software avoids off-target effects which may arise during experimentation by rapid BLAST search.

Conclusion

Research areas throughout the world are experiencing a surge because of the advent of the state of the art technologies. The consequence of development in such technological advancements is the vast amounts of biological data. Here arises the problem, how to analyze these data? Computational biology, a wing of bioinformatics, finds solution to this question by developing programs and software for biological data analysis. Gene silencing, a current area of research can be fastened by the process of designing small non- coding RNAs through computational tools. The present study put forward a tool named AmiRzyn for designing artificial miRNA using PERL language. It predicts the possible siRNA sequences as output when user submits their DNA sequences through a set of pre- defined parameters. Possibilities of occurrence of off- target effects are verified by performing BLAST comparison. For amiRNA

designing, AmiRzyn permits the specification of restriction site and miRNA backbone of user's requirement, in addition to the predicted siRNA sequence as the input. The designed miRNA can be used for several downstream experiments.

Acknowledgement

The authors gratefully thank the management of Malankara Catholic College for their encouragement in promoting research. We also thank Mr. Raj Kumar for his technical support.

References

1. Du G., Yonekubo J., Zeng Y., Osisami M. and Frohman M.A., Design of expression vectors for RNA interference based on miRNAs and RNA splicing, *FEBS J.*, **273**, 5421-5427 (2006)
2. Fire A., Xu S., Montgomery M., Kostas S., Driver S. and Mello C., Potent and specific gene interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature*, **391**, 806-811 (1998)
3. Alvarez J.P., Pekker I., Goldshmidt A., Blum E., Amsellem Z. and Eshed Y., Endogenous and synthetic microRNAs stimulate simultaneous, efficient and localized regulation of multiple targets in diverse species, *Plant Cell*, **18**, 1134-1151 (2006)
4. Schwab R., Ossowski S., Riester M., Warthmann N. and Weigel D., Highly specific gene silencing by artificial microRNAs in *Arabidopsis*, *Plant Cell*, **18**, 1121-1133 (2006)
5. Bartel D.P., MicroRNAs: Genomics, biogenesis, mechanism and function, *Cell*, **116**, 281-297 (2004)
6. Chapman J.E and C.J. Carrington, Specialization and evolution of endogenous smallRNA pathways, *Nat. Rev. Genet.*, **8**, 884-896 (2007)
7. Filipowicz W., RNAi: The nuts and bolts of the RISC machine, *Cell*, **122**, 17-20 (2005)
8. Martinez J., Patkaniowska A., Urlaub H., Luhrmann R. and Tuschl T., Single stranded antisense siRNA guide target RNA cleavage in RNAi, *Cell*, **110**, 563- 574 (2002)
9. Sijen T., Fleenor J., Simmer F., Thijssen K.L. and Parish S. *et al.*, On the role of RNA amplification in dsRNA triggered gene silencing, *Cell.*, **107**, 465- 476 (2001)
10. Baulcombe D., RNA silencing, *Trends Biochem.*, **30**, 290-293 (2005)
11. Meister G. and T. Tuschl, Mechanisms of gene silencing by double-stranded RNA, *Nature*, **431**, 343-349 (2004)
12. Zhang J. and Hattori T., Small RNA molecules as Therapeutic Agents for Viral Infectious Diseases, *Journal of Pharmacology and Toxicology*, **2(2)**, 103-113 (2007)
13. Zeng Y., Wagner E.J. and Cullen B.R., Both natural and designed microRNAs can inhibit the expression of cognate mRNAs when expressed in human cells, *Mol. Cell*, **9**, 1327-1333 (2002)
14. Chung K.H., Hart C.H., Al-Bassam S., Avery A. and Taylor J. *et al.*, Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155, *Nucleic Acids Res.*, **34**, e53 (2006)
15. Wall L., Christiansen T. and Orwant J., Programming Perl. Third Edn., O'Reilly Media, USA (2000)
16. Schwartz R., Christiansen T. and Wall L., Learning perl. Second Edn., O'Rielly and Associates, USA (1997)
17. Elbashir S.M., Harborth J., Lendeckel W., Yalcin A., Weber K. and Tuschl T., Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature*, **411**, 494-498 (2001)
18. Reynolds A., Leake D., Boese Q., Scaringe S., Marshall W.S. and Khvorova A., Rational siRNA design for RNA interference, *Nat Biotechnol.*, **3**, 326-330 (2004)
19. Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J., Basic local alignment search tool, *J. Mol. Biol.*, **215**, 403-410 (1990)
20. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J., Gapped BLAST and PSI- BLAST: A new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389-3402 (1997)
21. Griffiths-Jones S., The microRNA Registry, *Nucleic Acids Res.*, **32**, D109-D111 (2004)
22. Griffiths-Jones S., Saini H.K., van Dongen S. and Enright A.J., miRBase: Tools for microRNA genomics, *Nucleic Acids Res.*, **36**, D154-D158 (2008)
23. Jeffrey R., Applied Microsoft: NET Framework Programming, Microsoft Press, Washington, 21-34 (2002)
24. Naito Y., Yamada T., Ui-Tei K., Morishita S. and Saigo K., siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference, *Nucleic Acids Res.*, **32(2)**, W124-W129 (2004)
25. Park Y., Park S.M., Choi Y.C., Lee D., Won M. and Kim Y.J., AsiDesigner: exon-based siRNA design server considering alternative splicing, *Nucleic Acids Research*, **36**, W97-W103 (2008)
26. Yiu S.M., Wong P.W.H., Lam T.W., Mui Y.C., Kung H.F., Lin M. and Cheung Y.T., Filtering of Ineffective siRNAs and improved siRNA Design Tool, *Bioinformatics*, **21 (2)**, 144-151(2005)