How to design a highly effective siRNA

RNA interference (RNAi) is the process of double stranded RNA dependent post-transcriptional gene silencing in plants, invertebrates and mammalian cells. Introduction of double stranded RNA into cells leads to the sequence-specific destruction of endogenous RNAs that match the double stranded RNA. Short interfering RNAs (siRNAs) of 19–23 nt in length can be generated within the cells by using appropriate expression vectors or synthetic double stranded RNAs can be introduced into the cells to obtain silencing of the desired genes. The success of this technique depends on the efficiency with which an endogenous RNA is targeted for degradation. The selection of the target sequence of the mRNA plays an important role in this process. Recent papers describe certain guidelines for selection of target sequences which results in a very high rate of successful RNA interference. These rules are based partly on the recent observations that the internal stability of the siRNA duplexes that function effectively is lower at the 3′-end of the target sequence (Khvorova *et al* 2003; Schwarz *et al* 2003).

A paper by Ui-Tei *et al* (2004) describes simple guidelines for the selection of effective siRNA sequences for mammalian cells. The siRNAs which satisfy the following four conditions are capable of inducing highly effective gene silencing: (i) A/U at the 3′-end of the target sequence (which corresponds to 5′-end of the antisense strand of siRNA). (ii) G/C at the 5′-end of the target sequence. (iii) At least 5 A/Us at the 3′-terminal one third of the target sequence. (iv) The absence of any GC stretch of more than 9 nt in length.

In addition to these rules it is necessary to carry out a Blast search to make sure that only the target mRNA will be affected and activity of other genes will not be affected. This is done by selecting only those sequences which show at least 3 or 4 mismatches with other sequences in the database. These guidelines were arrived at by using 16 siRNAs targeting for the firefly luciferase gene (luc) and their effectiveness was checked in HeLa, Chinese hamster (CHO-K1) and mouse ES cells. In order to test the validity of the guidelines, 15 siRNAs were designed and all of them showed highly effective RNAi (more than 60% inhibition of reporter activity). Ui-Tei *et al* (2004) also designed 5 siRNAs which were not expected to provide efficient RNAi. These five siRNAs showed little or no RNAi. In *Drosophila* cells, surprisingly, all the siRNAs function effectively and no sequence preference was noticed. These rules were arrived at by using chemically synthesized siRNAs but were also found to be applicable to DNA vector-based RNAi in which siRNA is produced via cleavage of hairpin RNA produced and transported from the nuclei.

Reynolds *et al* (2004) have carried out a more extensive analysis for rational design of siRNA. They used 180 siRNAs targeting the mRNA of two genes and arrived at eight characteristics which were associated with functional siRNA. These are: low G/C content (30–52%), a bias towards low internal stability at 3′-end of sense-strand, lack of inverted repeats and sense strand base preferences at positions 3, 10, 13 and 19 (figure 1A). The siRNAs designed using these rules were very efficient in producing RNA interference. They have developed an algorithm which incorporates these eight criteria to evaluate the siRNA test panel. Each siRNA target is assigned a score according to the following logic. One point is earned for each of the following five criteria; (i) low G/C content (30–52%); (ii) absence of internal repeats; (iii) an ‘A’ base at position 19 (sense strand); (iv) an ‘A’ base at position 3 of sense strand; and (v) A ‘U’ base at position 10 of sense strand. In addition to these, for each A/U base at position 15–19, one point was added (potential 5 points). Two negative criteria were used. If a base other than ‘G’ was present at position 13, no point was deducted but a ‘G’ at this position resulted in deduction of one point from the score. Similarly ‘G’ or ‘C’ at position 19 resulted in deduction of one point. A score of 6 or more out of 10 points is likely to produce a functional siRNA. Out of 30 rationally designed siRNAs, 29 were functional which produced more than 50% silencing of six different

genes. These rules were arrived at by using 19 nt target sequence. To minimize potential off-target silencing effects, only sequences with more than 3 mismatches against unrelated sequences were selected. The authors calculate that the selection of targets using these criteria results in over 3.5-fold improvement in achieving 80% knockdown as compared to the random selection.

These two papers (Ui-Tei *et al* 2004 and Reynolds *et al* 2004) describe certain criteria which are common: (i) G/C at the 5′-end of the sense strand of siRNA; (ii) A/U at the 3′-end of the sense strand; and (iii) A/U rich region at the 3′-end of the antisense strand (figure 1A, B). However these two papers differ in some respects. The base preference at position 3 and 10 derived by Reynolds *et al* is not suggested by Ui-Tei *et al*. My analysis of the highly effective siRNA sequences of Ui-Tei *et al* shows that there is no base preference for 'A' at position 3 or for 'U' at position 10 of sense strand of siRNA. Absence of 'G' at position 13 is supported by the results of Ui-Tei *et al*. Another point of apparent difference is that Ui-Tei *et al* do not suggest 30–52% G/C content for highly effective siRNA. However this could be due to the use of highly A/U rich 3′-end of sense strand which will lower the overall G/C content.

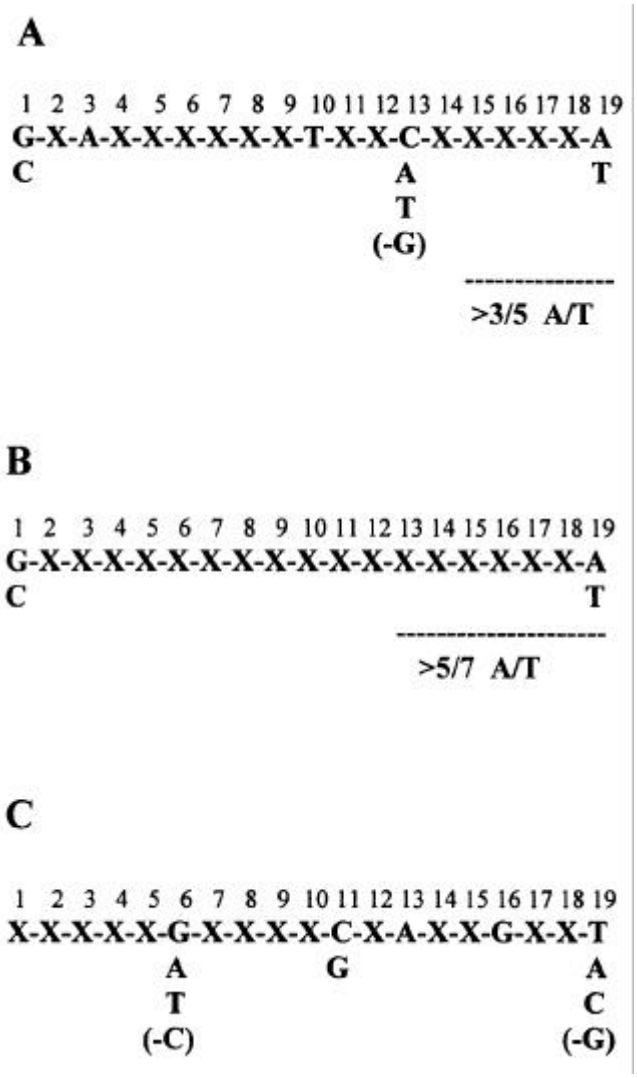

Figure 1. Nucleotide preferences in the target sequence for the design of effective siRNA based on the data of Reynolds *et al* (2004) **(A)**, Ui-Tei *et al* (2004) **(B)**, and Hsieh *et al* (2004) **(C)**. X indicates any nucleotide. Negative selection is indicated by a minus sign, for example (-G) indicates a selection against nucleotide G.

In an extensive study, Hsieh *et al* (2004) have used 148 siRNA duplexes (19 mer), to target 30 genes involved in phosphatidylinositol-3-kinase pathway. Their analysis showed that: (i) siRNA duplexes targeting the middle of the coding region were significantly poorer in gene silencing; (ii) silencing of 3′-UTR targets was as good as of the coding sequence; and (iii) there was base preference among siRNAs that achieved more than 70% knockdown of mRNA. G or C was preferred at position 11 and T was preferred at position 19. There was selection against G at position 19 and selection against A and A/T at position 11. There was preference for G at position 16, for A at position 13 and a selection against C at position 6 (figure 1C). These results of Hsieh *et al* (2004) with respect to base preferences at position 6, 11, 16 are not in agreement with those of Reynolds *et al* (2004). Even at the 3′-end of target sequence Hsieh *et al* (2004) show a preference for T as opposed to preference for A by Reynolds *et al* (2004).

Overall it appears that the four guidelines provided by Ui-Tei *et al* (2004) are simple to follow, do not require an algorithm and are in agreement to some extent with the results of Reynolds *et al* (2004). Therefore it may be advantageous to design siRNAs which follow the four guidelines of Ui-Tei *et al* and also give a score of more than 6 when the criteria of Reynolds *et al* are used. The effectiveness of siRNAs can be further increased by incorporating 1–2 mismatches at the 3′-end of the sense strand (Hohjoh 2004). An independent confirmation of the effectiveness and general applicability of these guidelines will be eagerly awaited, specially for vector based expression of short hairpin RNAs.

Wang and Mu (2004) have developed a web-based design center (http://www.genscript.com) to facilitate the designing process of siRNA. The program ‘siRNA target finder’ automates homology filtering, minimizes non-specific cross reactions and filters target sites based on internal stability (of siRNA duplexes) and secondary structure. This programme also takes into account single nucleotide polymorphism (the regions of target sequence showing single nucleotide polymorphism should be avoided). The practical usefulness of this tool in designing efficient siRNAs is yet to be tested by experiments but it may provide a useful starting point from which one can further select targets using the guidelines described by others.

References

Hsieh A C, Bo R, Manola J, Vazquez F, Bare O, Khvorova A, Scaringe S and Sellers W R 2004 A library of siRNA duplexes targeting phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens; *Nucleic Acids Res.* **32** 893–901

Hohjoh H 2004 Enhancement of RNAi activity by improved siRNA duplexes; *FEBS Lett.* **557** 193–198

Khvorova A, Reynolds A and Jayasena S 2003 Functional siRNA and miRNA exhibit strand bias; *Cell* **115** 209–216

Reynolds A, Leake D, Boese Q, Saringe S, Marshal W S and Khvorova A 2004 Rational siRNA design for RNA interference; *Nature Biotechnol*. **22** 326–330

Schwarz D S, Hutvágner G, Du T, Xu Z, Aronin N and Zamore P D 2003 Asymmetry in the assembly of the RNAi enzyme complex; *Cell* **115** 199–208

Ui-Tei K, Naito Y, Takahashi F, Haraguchi T, Ohki-Hamazaki H, Juni A, Ueda R and Saigo K 2004 Guidelines for the selection of highly efficient siRNA sequences for mammalian and chick RNA interference; *Nucleic Acids Res*. **32** 936–948

Wang L and Mu F Y 2004 A web-based design center for vector based siRNA and siRNA cassette; *Bioinformatics* (Advance Access published on 4 March 2004); DOI; 101093/bioinformatics/bth164

GHANSHYAM SWARUP
Centre for Cellular and Molecular Biology,
Hyderabad 500 007, India
(Email, gshyam@ccmb.res.in)