Gene expression

siRNA specificity searching incorporating mismatch tolerance data

Alistair M. Chalk[1] and Erik L. L. Sonnhammer[1,2,*]

[1]Department of Cell and Molecular Biology, Karolinska Institutet, S-171 77 Stockholm and [2]Stockholm Bioinformatics Center, Stockholm University, S-106 91 Stockholm, Sweden

ABSTRACT

Artificially synthesized short interfering RNAs (siRNAs) are widely used in functional genomics to knock down specific target genes. One ongoing challenge is to guarantee that the siRNA does not elicit off-target effects. Initial reports suggested that siRNAs were highly sequence-specific; however, subsequent data indicates that this is not necessarily the case. It is still uncertain what level of similarity and other rules are required for an off-target effect to be observed, and scoring schemes have not been developed to look beyond simple measures such as the number of mismatches or the number of consecutive matching bases present.

We created design rules for predicting the likelihood of a non-specific effect and present a web server that allows the user to check the specificity of a given siRNA in a flexible manner using a combination of methods. The server finds potential off-target matches in the corresponding RefSeq database and ranks them according to a scoring system based on experimental studies of specificity.

Availability: The server is available at http://informatics-eskitis.griffith.edu.au/SpecificityServer.

Contact: Erik.Sonnhammer@sbc.su.se

Supplementary information: Supplementary analysis and figures are available at *Bioinformatics* online.

1 INTRODUCTION

Artificially synthesized short interfering RNAs (siRNAs) are widely used in functional genomics for gene-specific knock-down. A major challenge is ensuring minimal off-target effects. Sequence-specific off-target effects are caused by the siRNA targeting a gene other than the intended target. Computationally, the efforts of a number of groups have focused on developing fast methods of database searching using short sequences. However, the level of similarity required for an off-target effect to be observed remains to be determined. Previously, no scoring schemes had been developed to look beyond the number of mismatches and maximum contiguous bases found.

A number of efforts have endeavoured to clarify the requirements for a non-specific siRNA effect to be observed. Jackson *et al.* (2003) used a microarray approach to identify 10 non-target genes that are consistently affected by a given siRNA. The hits were grouped into two sets; those with high similarity to the target (at least 14 contiguous bases in the positions 3–17, where position 1 refers to the 5′ base of the guide strand) and those only showing similarity at the 3′ sense end (eight contiguous matches in positions 11–19). While these criteria are important for off-target effects, a number of genes with comparable matches do not show any knockdown in expression. An alternative approach was taken by Du *et al.* (2005). who generated a set of targets with all possible single-mismatches to an siRNA. This provided detailed data on the positional and nucleotide-specific importance of a single mismatch's effect on efficacy.

Reproducible off-target effects can be observed for siRNAs with a number of mismatches to the target sequence. In general, terminal bases are tolerant to mismatches, while the central bases are not. There is agreement on the importance of the central region, although exact positions vary. Positions 12 and 13 are absolutely conserved in the genes identified by Jackson *et al.* (2003) while positions 9–11 have the least tolerance for mismatches according to Du *et al.* (2005). The bases at the 5′ half of the siRNA contribute the majority of the energy governing binding to target. Conversely perfect pairing at the 3′ end of the siRNA (approximately positions 1–9) is not a requirement for repression, or even knockdown provided that certain conditions are met (Doench and Sharp, 2004; Haley and Zamore, 2004).

It is important to incorporate knowledge about underlying mechanisms when identifying the most important regions of an siRNA. The cleavage of a target requires recognition by the RISC + siRNA complex, followed by cleavage at the bond between nucleotides at positions 11 and 12 (Elbashir *et al.*, 2001). The target recognition may be partial and not involve the 3′ end (Tomari and Zamore, 2005). In addition, certain mismatches can be more easily tolerated, dependent on the nucleotide (Du *et al.*, 2005). Mismatches in the central region are thought to disturb the RNA–RNA helix structure formed, preventing cleavage. However repression may still occur through an miRNA like mechanism.

*To whom correspondence should be addressed.

Fig. 1. Example results from the specificity server for an siRNA targeting rat accession NM_001000843, target sequence GTGTCCTCCTTGTGG TGAT.

Successful finding of off-target effects requires a sensitive database searching technique and a reliable scoring scheme of the hits found by the database search. We use WU-BLAST (http://blast.wustl.edu) as it is the best blast-style algorithm for dealing with short sequences. We have developed a new specificity scoring scheme based on the results by (Du *et al.*, 2005), which uses experimentally observed off-target effects at each siRNA position. This way the mismatches are weighted by their relative importance (See Supplementary Methods).

2 IMPLEMENTATION

The program was implemented in Java, and utilizes the biojava package (www.biojava.org). All scoring methods described here are available as selectable options in the specificity server online. The user can input two types of data: either an siRNA sequence to be searched against RefSeq, or the results of an external database search in the form query_AC siRNA_query_sequence target_AC target_sequence (e.g. NM_001024 ACGUUAGCUGACUGACUAC NM_205852 GGGUCAGCUGAGUGACUAC). Users can modify the parameters based on their requirements. An example output page is shown in Figure 1, showing summary data for each individual match.

3 DISCUSSION

Specificity of siRNAs is an important aspect of siRNA design. We developed a novel scoring scheme based on experimental data to estimate the likelihood of an siRNA eliciting an off-target effect. Complete elimination of off-target effects is non-trivial, but with sensible design strategies identifying permitted mismatches, it is possible to minimize these problems. Such a strategy utilizes an extensive database search to minimize the potential for off-target effects, combined with scoring methods to identify which matches are most likely to elicit an effect. We here present a server for checking the specificity of siRNAs. Although the specificity score is currently based on data for one gene, we are optimistic that more in-depth experiments will improve our ability to determine genes modulated in an off-target manner.

This siRNA was found to not be specific as it also matches two other genes perfectly. The colour of the row indicates the score of the match. Red, orange and yellow indicate high, low and very low likelihood of causing an off-target effect. A green row indicates a match to a gene that has the same accession number or 100% identity over the entire length. Red indicates other matches with a specificity score >80, orange matches with a score >40 and yellow otherwise. The colour of the mismatches indicates the weighting of that position (green = 0.2, yellow = 0.4, orange = 0.6).

Conflict of Interest: none declared.

REFERENCES

Elbashir,S.M. *et al.* (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO J.*, **20**, 6877–6888.

Doench,J.G. and Sharp,P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.

Du,Q. *et al.* (2005) A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Res.*, **33**, 1671–1677.

Haley,B. and Zamore,P.D. (2004) Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.*, **11**, 599–606.

Jackson,A.L. *et al.* (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.*, **21**, 635–637.

Tomari,Y. and Zamore,P.D. (2005) Perspective: machines for RNAi. *Genes Dev.*, **19**, 517–529.