
siRNA Target Finding Strategy ^[1]

Machine learning algorithm

Based on a large number of experimental data of RNAi and advanced data mining technology, GenScript now develops a newest machine learning algorithm to predict siRNA. Compared with our first generation design tool, the accuracy of new algorithm, especially for the design of shRNA, has greatly enhanced by 90%. The siRNA design algorithm developed by GenScript, along with its siRNA Genome-Wide Bio-Reagent Bank (GenPool™) library, represents a major improvement over our previous design approach.

Empirical rules

To maximize the prediction accuracy of RNAi design, GenScript scientists incorporate the following empirical rules into the novel generation RNAi design software successfully:

1. Thermodynamic property:

Sequences were filtered to remove candidates with unfavored thermodynamic property ^[2, 4, 5].

2. Length of the siRNA target site:

Default is 21 mer ^[3].

3. GC content of the target site:

Default GC content range is 30-60%. Sequences with low GC content (< 60%) have a better chance to be functional siRNAs site than those with high GC content (> 60%). For GC rich genes, the upper limit of the GC content range may be adjusted to a higher level.

4. Sequence region:

We recommend ORF (open reading frame) as the target selecting region, preferably 50–100 nt downstream of the "ATG" start codon, and you may specify your desired region in our design tool.

5. Organism:

The human, mouse, and rat genomes are currently available. The specified genome is used to remove non-unique target sequences via BLAST/SmithWaterman search.

6. RNA secondary structure:

We calculated the RNA secondary structure and minimum free energy (MFE) for each target sense and anti-sense sequences. Sequences with relatively strong internal structures will be removed.

7. Repeat and low complex sequence:

Sequences with tandem repeats and other low complex features will be removed.

8. Filtering sequences:

Filtering sequences are the sequences that trigger mammalian immune response ^[6, 7].

9. SNP:

Sequences overlapping with a SNP site will be marked and filtered. This is done only when an accession/gene id is provided.

10. ΔE and ranking:

The candidates are ranked based on a GenScript proprietary algorithm and using the parameter ΔE . This parameter measures the specificity of a target site.

11. Positive motif:

Positive motifs is a group of motifs which can significantly influence the *in vivo* application of siRNA owing to off target effects and toxicities associated with innate immune stimulation. Therefore, siRNA containing positive motifs will be suppressed; in the other hand, siRNA containing negative motifs will be given a bonus.

Controlled off-target efficiency

Off-target of RNAi operation is an important feature of GenScript RNAi design software, which stands for the latest developing trend of RNAi technology and application. Controlling off-target efficiency is performed:

1. BLAST/SmithWaterman search:

The imperfectly matched target mRNA with siRNA is a major factor for occurrence of off-target unintended genes. Therefore, each candidate target sequence will be searched against a unique and comprehensive EST/mRNA collection for the specified organism. Sequences overlapping 16 bp or more with other transcripts are removed from the final list.

2. Off-target of seed match of siRNA:

2.1 Seed region

Previous investigations and GenScript experiment data show that siRNA "off target" transcript silencing mediated by seed region sequence is widespread^[8]. "Seed region" of siRNA is analogous to seed region of microRNA. GenScript researchers found that size of seed region plays a major part in off target prediction. Size of seed region of siRNA we proposed is 6-8 bases.

2.2. Seed match:

Depending on the proprietary off-target prediction database, threshold of seed match for the specific species in siRNA design is a unique characteristic in GenScript's siRNA design software. This cutoff value is computed by a probability model which has general biological significance.

3. Functional alignment:

Besides general sequence alignment, GenScript siRNA design tool incorporates a novel alignment approach, functional alignment. This idea for functional alignment derives from asymmetry of siRNA in the assembly of the RNAi enzyme complex.

Announcement:

GenScript guarantees that 95% of off target effect will be filtered if all filtering conditions are chosen. However, some candidate siRNAs with highest knockdown efficiency are filtered owing to carry out a very strict screening condition. Therefore, we should make a tradeoff between off-target effect and gene knockdown efficiency.

References:

1. Wang L, Mu FY. A Web-based design center for vector-based siRNA and siRNA cassette. *Bioinformatics*. 2004; 20(11): 1818–1820.
2. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. *Cell*. 2003; 115(2): 209–216.
3. Harborth J, Elbashir SM, Vandeburgh K, *et al*. Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Antisense Nucleic Acid Drug Dev*. 2003; 13(2): 83–105.
4. Reynolds A, Leake D, Boese Q, *et al*. Rational siRNA design for RNA interference. *Nat Biotechnol*. 2004; 22(3): 326–330.
5. Ui-Tei K, Naito Y, Takahashi F, *et al*. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res*. 2004; 32(3): 936–948.
6. Adam DJ, Vandana S, Janet RS, *et al*. Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat Biotechnology*. 2005; 23: 457–462
7. Hornung V, Guenther-Biller M, Bourquin C, *et al*. Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat Medicine*. 2005; 11(3): 263–270
8. Jackson LA, Burchard J, Schelter J, *et al*. Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. *RNA*. 2006; 12(7): 1179–1187.