***S*fold: A Software for Rational Design of RNA-Targeting Nucleic Acids**

Web server URL: http://sfold.wadsworth.org/

Ye Ding
Bioinformatics Lab
Wadsworth Center
Empire State Plaza, Albany NY 12201
Tel: (518) 486 1719; E-mail: yding@wadsworth.org

Abstract

Short interfering RNAs (siRNAs, Elbashir *et al.* 2001), antisense oligonucleotides (oligos), and *trans*-cleaving ribozymes have become increasingly important for high throughput functional genomics, for the validation of therapeutic targets, and for the development of RNA-targeting therapeutics. To a large extent, target accessibility, long known to be important for the potency of antisense oligos and ribozymes, is determined by the secondary structure of the target RNA. Experimental evidence has emerged to suggest that the potency of siRNAs is also determined by target accessibility and secondary structure (Lee *et al.* 2002; Vickers *et al.* 2003). Recently, we have developed novel algorithms for improved prediction of mRNA structures and accessible sites, and for the rational design of RNA-targeting nucleic acids. The design methods have been well validated by preliminary experimental testing both *in vitro* and *in vivo*. A new software named ***S*fold** has been developed for RNA folding and for the rational design of RNA-targeting nucleic acids. A Web server for ***S*fold** 1.0 has been established with URL: **http://sfold.wadsworth.org/**.

Probability profiling for evaluation of target accessibility

Probability profiling. Single-stranded regions in an RNA secondary structure can be important for RNA/DNA, RNA/RNA and RNA/protein interactions. For prediction of accessible target sites, we developed a probability profile approach (Ding & Lawrence 2001; Fig. 1). Accessible sites are predicted by regions with high probabilities on the profile. There is a significant correlation (P=0.0147) between accessibility predictions by probability profiling and the degree of inhibition from experimental data; however, there is a lack of correlation (P=0.567) for the commonly used minimum free energy structure (Ding & Lawrence 2001).

Rationale. Probability profiling is based on a patent-pending RNA folding algorithm for generating a statistical sample of the Boltzmann ensemble of secondary structures. This is appealing for evaluation of target accessibility, because, as noted by researchers from Ribozyme Pharmaceuticals, Inc., " In the prediction of accessible sites, the identification of a single folded structure for a given target mRNA is not of particular interest. Instead, the objective of this exercise is to assess the *likelihood* of unpaired (or substantially unpaired) sites that could be a ribozyme target ... The ambiguities in thermodynamic parameters – and the possibility that each mRNA exists as a population of different structures – suggest that a *stochastic* approach to the

evaluation of accessible sites may be appropriate" (Christoffersen, McSwiggen & Konings 1994).

Validation of target accessibility prediction and rational design of siRNAs

Our method for rational design of siRNAs is based on probability profiling, computation of antisense siRNA binding energy, and computation of GC content. Other emerging empirical rules can be easily incorporated.

Experimental target and computational analysis. The target for experimental testing is exon 3 (nt 1003-1119) of the 6450 nt human estrogen receptor (ER, GenBank Accession No. NM_000125). The wild type cDNA of ER contains eight exons. The regional probability profile for the 117 nt exon 3 is displayed in Figure 1. Two siRNAs, siRNA 1 (47.62% GC) and siRNA 2 (42.86% GC) were designed for two predicted accessible AA(N19) motifs in exon 3.

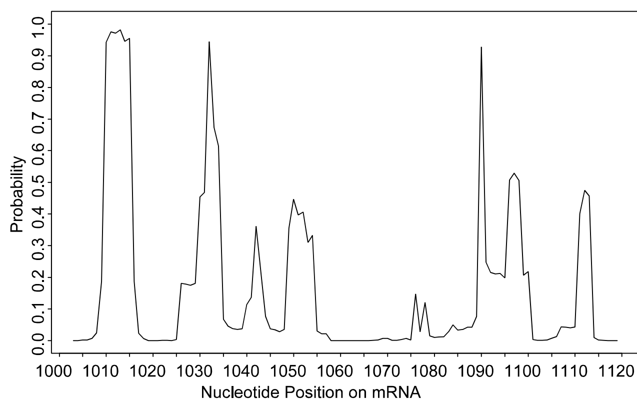

Fig. 1 Regional probability profile for exon 3 (nt 1003-1119) of the ER mRNA.

Exon 3 and ER variant. The ER alpha mRNA is known to undergo alternative splicing, generating ER mRNA variants through skipping of single or multiple exons (Murphy *et al.* 1997) during post-transcriptional processing. In TMX 2-11, a cloned cell line derived from ER positive MCF-7, the wild type ER and a variant missing exon 3 (ER△3) mRNAs and proteins are expressed at equivalent levels (Fasco *et al.* 2003). For testing of siRNAs targeted to exon 3 in TMX2 cells, ER△3 serves as an internal control for specificity.

Experimental procedure. SiRNAs were expressed from pSilencer 1.0 under the control of a U6 Pol III promoter (Ambion). TMX2-11 cells were stably transfected using Lipofectamine 2000 (Invitrogen). Cells were plated at 500,000 in 35 mm wells and transfected the next day. Standard transfection conditions for each well were 14 µl ipofectamine 2000, 2 µg of each siRNA construct, 0.1 µg pUR selection vector and total DNA level adjusted with pBluescript to 4 µg. Constructs for siRNA 1 and siRNA 2 were transfected together. Cell were subjected to selection with puromysicn (0.5µg/ml). A positive clone, MCF-7-ERd3, showing suppressed wtER was identified by western immunoblotting using the HC20 (SantaCruz Biotec) antiserum which recognizes a c-terminal epitope shared by wtER and ERΔ3.

Experimental results. The expression of ER was reduced in MCF-7-ERd3 clone, in comparison to parental cell line TMX2-11 without siRNA treatment. In contrast, there is no visible

reduction in the expression of ER△3 (Fig. 2). Cells of the original MCF-7-ERd3 clone were subjected to selection with 4µg/ml puromycin. Silencing by the two siRNAs in the resulting cell population is both potent and highly specific. By serial dilution analysis, we estimate **> 75%** reduction in the level of the wild type ER (Fig. 3).

Conclusion. Potent silencing was observed for two siRNAs rationally designed for two predicted accessible sites. The specificity was demonstrated by the control of no siRNA treatment in parental cell line, and by an internal control of equivalently expressed ER variant, whose mRNA lacks the targeted 117 nt exon, but shares all other 6333 nt (98.2%) of the wild type mRNA.

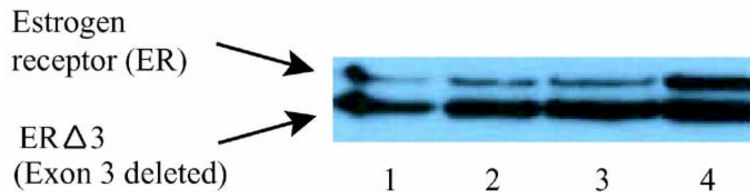

Fig. 2. Western blot analysis reveals maintenance of a suppressed ER phenotype during the growth of MCF-7-ERd3 clone treated with stably transfected siRNA 1 and siRNA 2. Top band: wild type ER; bottom band: ER△3; track 1: clone at the 24-well cluster level on February 12, 2003; track 2: clone at the T25 level on March 14, 2003; track 3: separate culture of the clone at the T25 level on March 14, 2003; track 4: parental cell line TMX2-11, a derivative of MCF-7. Wild type ER and variant ER△3 are equivalently expressed in TMX2-11 (i.e., about one to one ratio), and are detectable by the same C-terminal specific antibody HC20 (SantaCruz Biotec).

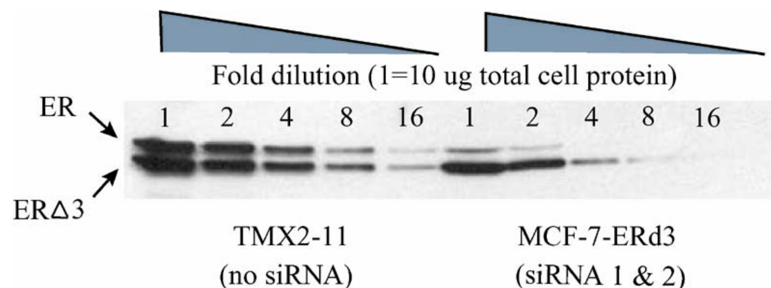

Fig. 3. Serial dilution analysis for both ER and ER△3, in parental cell line TMX2-11 (no siRNA treatment) and in MCF-7-ERd3 clone (siRNA 1 and siRNA 2). The reduction of wtER is estimated to be **> 75%** in the MCF-7-ERd3 clone. Analysis shows ER detected with C-terminal specific HC20 (Santa Cruz) antiserum in 2-fold dilutions.

Validation of target accessibility prediction and rational design of antisense oligos

Experimental target and computational analysis. The target for experimental testing of the design method is *E. coli lac*Z (GenBank Accession No. U00096), which codes for β-galactosidase. Based on probability profiling of the 3113 nt mRNA, and on calculation of antisense oligo binding energy, ten antisense 20-mers were designed to target predicted accessible sites throughout the mRNA.

Experimental conditions and procedure. In total 25 ul *in vitro* reaction: about 80 ng of plasmid pRPC179 with PEndoIV::*lac*Z as template, 7.5 µl S30 extract, 10 µl pre-mix, 2.5 µl complete amino acids, 1 µl oligos of 6 µM concentration or water (for control). S30 extract, pre-mix and complete amino acids were from *E. coli* S30 Extract System for Circular DNA (Promega). The reaction was incubated at 37 °C for 2 hours and then was on ice for 5 minutes. 1 µl of C8FDG (2 mM) was added as substrate to assay β-galactosidase activity in CytoFluo Reader. The experiment for every oligo was carried out in triplicate.

Experimental results. For an oligo, the percentage of inhibition for each of triplicate runs was calculated with readout from the CytoFluo Reader, using the control with no oligo treatment as the baseline. For every oligo, the average and standard deviation for three runs were computed and are shown in Figure 4. The results from different runs are highly stable, as indicated by the small standard deviations. Three of the ten 20-mers achieved nearly **100%** inhibition of β-galactosidase; six other oligos reduced the level of β-galactosidase by **60-80%**; only one oligo was not a potent inhibitor. For these ten oligos, the average inhibition on the protein level is **75.5%** .

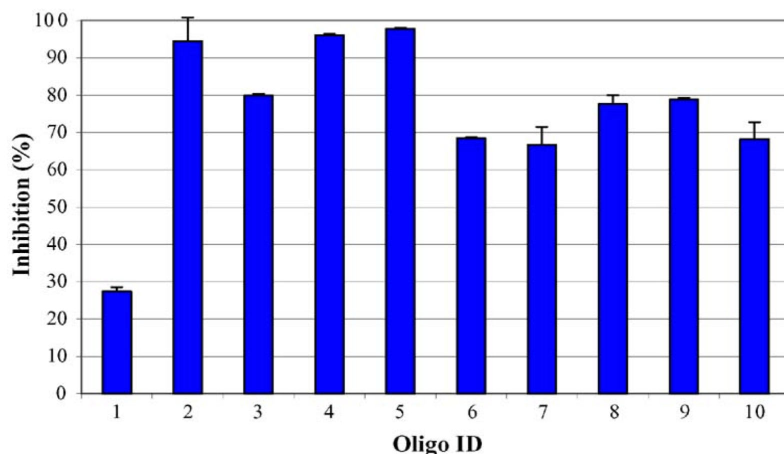

Fig. 4. Potent inhibition of *E. coli lac*Z expression by rationally design antisense 20-mers. The percentage of reduction on the level of β-galactosidase is shown on Y-axis.

Validation of target accessibility prediction and rational design of *trans*-cleaving ribozymes

Design methodology and experimental target. For *trans*-cleaving ribozymes, both the target accessibility and the folding of the ribozyme are important for catalytic efficiency. The two binding arms of a hammerhead ribozyme need to be also accessible for interaction with target sequences flanking the cleavage triplet, e.g., GUC. The flanking sequences of all 23 GUC triplets for the breast cancer resistance protein (BCRP) mRNA (2418 nt, GenBank Accession No. AF098951) were analyzed for accessibility by probability profiling. Three GUC sites with *both* flanking sequences predicted to be accessible are selected and hammerhead ribozymes were designed accordingly.

Major findings from testing in cultured cells. All three ribozymes substantially reduced the expression level of BCRP. In comparison with a forth ribozyme designed with the minimum free energy structure, there is an increase of 125%, 100% and 50%, respectively, in the amount of reduction on the protein level. This presents experimental evidence to support our design rationale that is based on a statistical sample of probable structures rather than a single structure.

*S*fold software for rational design of RNA-targeting nucleic acids

5A **5B**

Fig. 5. Front page (5A) and module *S*irna input page (5B) for ***S*fold** Web server with URL: **http://sfold.wadsworth.org/**

Sample text output from module *S*irna

File *sirna1.out* for (AAN19) targets:

~~~~~~~~~~~~~~~ Output for design of siRNAs for AA(N19) motifs ~~~~~~~~~~~~~~~
Line 1: ID of AA(N19) motif; target position; target sequence;
GC content; antisense siRNA binding energy (kcal/mol)
Line 2: sense siRNA (5' → 3'); antisense siRNA (5' →3')
~~~~~~~~~~~~~~~

Note: dTdT at the 3'ends of sense and antisense siRNA can be replaced by UU

```
 1  18 − 38   AACUGUGUUUACUUGCAAUCC      38.10%         − 13.8652
              CUGUGUUUACUUGCAAUCCTT      GGAUUGCAAGUAAACACAGTT
 2  34 − 54   AAUCCCCCAAAACAGACAGAA      42.86%         − 16.4761
              UCCCCCAAAACAGACAGAATT      UUCUGUCUGUUUUGGGGGATT
 3  42 − 62   AAAACAGACAGAAUGGUGCAU      38.10%         − 17.9349
              AACAGACAGAAUGGUGCAUTT      AUGCACCAUUCUGUCUGUUTT
 ......
```

File *sirna.out* for target sequences with NN dinucleotide leader:

~~~~~~~~~~~~~~~~~~~~ Output for design of siRNAs ~~~~~~~~~~~~~~~~~~~~
Column 1: target position (starting − ending)
Column 2: sense siRNA (5' → 3')
Column 3: antisense siRNA (5' → 3')
Column 4: GC content of target sequence
Column 5: antisense siRNA binding energy (kcal/mol)
Note: sense siRNA=target sequence+3' dTdT overhang;
dTdT overhangs for both sense and antisense siRNAscan be replaced by UU

```
 1 − 21  ACACUUGCUUUUGACACAACUTT   AGUUGUGUCAAAAGCAAGUGUTT  38.10%  −14.4131
 2 − 22  CACUUGCUUUUGACACAACUGTT   CAGUUGUGUCAAAAGCAAGUGTT  42.86%  −12.6306
 3 − 23  ACUUGCUUUUGACACAACUGUTT   ACAGUUGUGUCAAAAGCAAGUTT  38.10%  −10.8443
 ......
```

References

Christoffersen, R.E., McSwiggen, J.A., Konings, D. (1994) *J. Mol. Structure* (*Theochem*) **311**, 273-284.

Ding, Y., and Lawrence, C.E. (2001) *Nucleic Acids Res.* **29**, 1034-1046.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T. (2001) *Nature.* **411**, 494-8.

Fasco, M.J., Amin, A., Pentecost, B.T., Tang, Y., Gierthy, J.F. (2003) (manuscript).

Lee, N.S., Dohjima, T., Bauer, G., Li, H., Li, M.J., Ehsani, A., Salvaterra, P., Rossi, J. (2002) *Nat. Biotechnol.* **20**, 500-5.

Murphy, L.C., Dotzlaw, H., Leygue, E., Douglas, D., Coutts, A., Watson, P.H. (1997) *J. Steroid Biochem. Mol. Biol.* **62**, 363-72.

Vickers, T.A., Koo, S., Bennett, C.F., Crooke, S.T., Dean, N.M., Baker, B. (2003) *J. Biol. Chem.* **278**, 7108-18.
~~~~~~~~~~~~~~~~~~~~