METHOD

Dual-targeting siRNAs

KATRIN TIEMANN,^{1,6} BRITTA HÖHN,^{1,6,7} ALI EHSANI,¹ STEPHEN J. FORMAN,² JOHN J. ROSSI,¹
and PÅL SÆTROM^{3,4,5}

¹Department of Molecular Biology, Beckman Research Institute of City of Hope, Duarte, California 91010, USA

²Department of Hematology and Bone Marrow Transplantation, City of Hope National Medical Center, Duarte, California 91010, USA

³Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, NO-7489 Trondheim, Norway

⁴Department of Computer and Information Science, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

⁵Interagon AS, Laboratorienesenteret, NO-7489 Trondheim, Norway

ABSTRACT

We have developed an algorithm for the prediction of dual-targeting short interfering RNAs (siRNAs) in which both strands are deliberately designed to separately target different mRNA transcripts with complete complementarity. An advantage of this approach versus the use of two separate duplexes is that only two strands, as opposed to four, are competing for entry into the RNA-induced silencing complex. We chose to design our dual-targeting siRNAs as Dicer substrate 25/27mer siRNAs, since design features resembling pre-microRNAs (miRNAs) can be introduced for Dicer processing. Seven different dual-targeting siRNAs targeting genes that are potential targets in cancer therapy have been developed including Bcl2, Stat3, CCND1, BIRC5, and MYC. The dual-targeting siRNAs have been characterized for dual target knockdown in three different cell lines (HEK293, HCT116, and PC3), where they were as effective as their corresponding single-targeting siRNAs in target knockdown. The algorithm developed in this study should prove to be useful for predicting dual-targeting siRNAs in a variety of different targets and is available from <http://demo1.interagon.com/DualTargeting/>.

Keywords: RNAi; algorithm; dual-targeting siRNA

INTRODUCTION

For both microRNAs (miRNAs) and short interfering RNAs (siRNAs)—the two major classes of small duplex RNAs characteristic of RNA interference (RNAi)—one strand is generally considered the active or guide strand. This active strand is incorporated into the Argonaute 2 (AGO2) protein active site and acts as a guide for target recognition by the Ago2-containing RNA-induced silencing complex (RISC). The other strand—the passenger strand—is cleaved or removed during this process (Matranga et al. 2005; Rand et al. 2005). Critical parameters for determining the guide strand are the thermodynamic stabilities of the duplex ends, as RISC preferentially incorporates the strand with the least stable 5' end (Khvorova et al. 2003; Schwarz et al. 2003).

This simple distinction between the active guide strand and inactive passenger strand is often blurred as there are examples of both miRNAs and siRNAs where both strands serve as guides in RISC (Hossbach et al. 2006; Stark et al. 2007). When both strands of a siRNA duplex are selected for incorporation into RISC, the intended passenger strand has the potential for down-regulating nontargeted transcripts primarily via miRNA-like targeting (Birmingham et al. 2006; Jackson et al. 2006b), a so-called off-target effect. Such off-target effects can also be caused by the guide strand. Although undesirable in many applications, having two active siRNA strands can be beneficial in several therapeutic settings. The inclusion of two functionally active guide strands in one siRNA could provide redundancy against escape mutations in viral therapy or combinatorial (and synthetic lethal) inhibition of oncogenes in cancer treatment (Boden et al. 2003) as previous results suggest that inhibition of a single oncogenic target is insufficient for the killing of malignant cells (Polo et al. 2004; Ramanarayanan et al. 2004). Although using multiple different siRNAs can create the same redundancy, using dual-targeting siRNAs with two active strands in a therapeutic setting simplifies delivery and can reduce the side effects from passenger strand off-targeting.

⁶These authors contributed equally to this work.

⁷**Present address:** Department of Internal Medicine III, Johannes Gutenberg-University Mainz, 55131 Mainz, Germany.

Reprint requests to: John J. Rossi, Department of Molecular Biology, Beckman Research Institute of City of Hope, Duarte, CA 91010, USA; e-mail: jrossi@coh.org; fax: (626) 301-8271.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.2005710>.

RESULTS AND DISCUSSION

We first attempted to follow a previously published approach and designed dual-targeting siRNAs against Bcl2, Bcl6, Stat3, and MYC by simply searching for partially complementary target sites (Hossbach et al. 2006), but this approach provided an unacceptably small number of potent dual-targeting siRNAs. Six of the seven siRNAs, in which three were tested as 21mer and 25/27mer and four as 21mer only, predicted by this method only gave knock-down of one of the two targeted transcripts (Fig. 1; Supplemental Tables 1, 4). Therefore, we analyzed our initial designs to identify factors that influence the efficacy of dual-targeting siRNAs (Fig. 2). Our goal was to develop a revised algorithm for designing dual-targeting siRNAs.

Known factors that influence siRNA efficacy include the difference in the duplex end thermodynamic stabilities ($\Delta\Delta G$) (Khvorova et al. 2003; Schwarz et al. 2003),

sequence motifs within the guide strand (Reynolds et al. 2004; Saetrom and Snove 2004), the overall duplex stability (Reynolds et al. 2004), and the target site accessibility (Heale et al. 2005; Shao et al. 2007). Of these, only the strand's sequence similarity to highly effective standard siRNAs and $\Delta\Delta G$ showed weak to strong correlations with the observed siRNA activities. Indeed, a single parameter RISC entry (re) that indicated whether a duplex strand was clearly incorporated into ($\Delta\Delta G > 1$ kcal/mol, re = 1) or clearly excluded from ($\Delta\Delta G > -1$ kcal/mol, re = -1) RISC, or whether both strands were likely included into RISC (-1 kcal/mol $\leq \Delta\Delta G \leq 1$ kcal/mol, re = 0), could explain about 40% of the observed variation in siRNA activity (Fig. 2L). We also evaluated each strand's potential to inhibit translation of its target and found a significant correlation with the observed strand activities (Fig. 2M). Together, RISC entry, potential for translational inhibition, and sequence similarity with highly effective standard siRNAs explained 60% of the observed variation in activity (Fig. 2N).

Based on the above analyses, we developed a revised algorithm that uses successive filters to identify all dual-targeting siRNA candidates, where both strands are likely included into RISC and both strands resemble highly effective siRNAs. Given an input of two mRNA sequences, our algorithm uses the following five steps to identify and prioritize dual-targeting siRNA candidates: (1) For each 19mer target sequence in input sequence 1, all 19mer subsequences in input sequence 2 are identified that have a Hamming distance of at most 5 nucleotides (nt) to the 19mer target sequence from sequence 1 (the Hamming distance is the number of single nucleotide mismatches when not allowing insertions or deletions). All such 19mer duplexes are added to the list of candidate dual-targeting siRNAs. (2) All siRNAs are removed from the candidate list that have an absolute difference in the duplex end predicted thermodynamic stabilities ($\Delta\Delta G$) greater than 1 kcal/mol when using 5 nt at the duplex ends. (3) All siRNAs are removed from the candidate list where one or both strands do not have sequence similarity to highly effective known siRNAs. More specifically, this step involves removing all candidates that contain one or more of the AAAA, CCCC, GGGG, or UUUU motifs in either strand; have a GC percentage <25% or >75% in either strand; and have a strand that received a score of <0.1 by the GPboost siRNA efficacy predictor (Saetrom 2004). (4) All siRNAs in the candidate list are removed where both strands have a Hamming distance of <2 for any off-target transcripts or where the strands have a total of >10 potential off-target transcripts with a Hamming distance equal to 2. (5) The remaining candidate siRNAs are sorted based on their potential to direct cleavage and translational suppression of their targets. More specifically, the algorithm assigns a score, which is the sum of the strand's GPboost score and its predicted potential for causing translational suppression

FIGURE 1. Analysis of dual-targeting siRNAs of the first generation. (A) Quantitative real-time PCR performed for dual-targeting siRNAs against BCL6 (black bars) and STAT3 (gray bars) mRNA extracted from HEK293 cells 48 h after transfection. (B) Quantitative real-time PCR performed for dual-targeting siRNAs against Bcl6 (black bars), STAT3 (light gray bars), MYC (dark gray bars), and Bcl2 (open bars) mRNA extracted from HEK293 cells 48 h after transfection.

FIGURE 2. RISC entry, potential for miRNA-like regulation, and sequence similarity to highly effective standard siRNAs explain observed siRNA efficacy. Observed relative target RNA levels for first generation candidate dual-targeting siRNAs (Fig. 1) plotted against features related to siRNA efficacy. Lines are linear regression models; numbers in the lower-left corners show the proportion of variability in the data explained by the model (r^2) and the model's P -value (p). Features are (A,B) target site accessibility ([A] target site accessibility and [B] target site's mean probability of having unpaired nucleotides); (C,D) duplex thermodynamics ([C] duplex stability and [D] target strand stability); (E,F) sequence similarity to highly effective standard siRNAs (scores from [E] GPboost and [F] Reynolds algorithms); (G–K) difference in the duplex ends' thermodynamic stability ($\Delta\Delta G$; $\Delta\Delta G$ computed by considering 1 nt [G] to 5 nt [K] at the ends of the 19mer siRNA strands); (L) RISC entry, which indicates whether the strand is favored (1), unfavored (–1), or undecided (0) for RISC uptake; calculations were based on $\Delta\Delta G$ 5 nt; (M) potential for miRNA-like regulation (miRNA score); and (N) the sum of the RISC entry, GPboost, and miRNA score values.

(see Materials and Methods), to each siRNA strand. The closer these strand scores are to each other, the more likely is their down-regulation potential for both target genes (Table 1). The siRNA's final score (duplex score) is the average minus the standard deviation of the two strands' scores.

Using this algorithm, we designed seven 25/27mer siRNAs against a set of targets involved in various cancers, including Bcl2, Stat3, CyclinD1, EGFR, BIRC5, and CyclinD2 (Table 1; see Materials and Methods for siRNA nomenclature). Dicer was shown to process all seven of the 25/27mer siRNAs (Fig. 3). The exact processing into 21mers by Dicer is hampered, because dual-targeting siRNAs contain bulges due to mismatches in the RNA

duplex. The closer the mismatch is situated to the conventional cut site, the more is the processing influenced resulting in longer products such as 22mers or 23mers (Fig. 3). Nevertheless, dual-targeting 25/27mers were recognized by Dicer and processed in active siRNAs. All seven siRNAs were tested in the human embryonic kidney cell line HEK293 and the colon cancer cell line HCT116. Additionally, four of them were tested in the prostate cancer cell line PC3. One of the seven siRNA candidates, B2S-4543-27, was predicted to be asymmetrically incorporated into RISC as shown by the RISC entry value of 1 (incorporation of the STAT3 strand into RISC) and –1 (no incorporation of the Bcl2 Strand into RISC) (Table 1), and therefore was not predicted to be a useful dual-targeting siRNA by the algorithm. As expected, only the strand the algorithm predicted to be preferentially incorporated into RISC, the strand targeting Stat3 (Table 1), gave functional knockdown in HEK293 and HCT116 cells (Fig. 4A).

The other six dual-targeting siRNAs were predicted to have both strands incorporated into RISC, and they demonstrated down-regulation of both target genes in HEK293, HCT116, and also in PC3 cells (Fig. 4A–C). The final strand scores of 0.22 for CyclinE2 and 1.23 for CyclinD1 of E2D1-733-27 suggests that CyclinE2 knockdown is less efficient than CyclinD1. Indeed, looking at the real-time PCR results of this particular dual-targeting siRNA in HEK293, HCT116, and PC3 cells the predictions were correct (Fig. 4B). We

can find similar patterns of four other dual-targeting siRNAs, although it depends on the cell line tested. However, looking at the dual-targeting siRNA surD2-2049-27 that was predicted to down-regulate both targets equally well with strand scores of 2.12 and 2.19, we can see that this is clearly not the case in all three cell lines (Fig. 4C). Survivin (BIRC5) is down-regulated ~50%–60%, while CCND2 is down-regulated an average of 90%. Despite some variability in potencies, all the dual-targeting siRNAs the algorithm predicted were indeed able to achieve down-regulation of their two target genes simultaneously.

All seven dual-targeting siRNAs were tested in HEK293 cells for their potential to up-regulate type I interferon genes. Real-time PCR was performed for IFNB1, CDKL2, and

TABLE 1. Rational evaluation of the 25/27mer dual-targeting siRNAs

siRNA ID	Targets	Duplex	Clv. site	$\Delta\Delta G$ 5 nt	RISC				
					entry	miRNA score	Strand score	Duplex score	
B2S-4543-27	BCL2	5' AAUUGAGGUGCUUCCUAAUUGUga	4543	-1.3	0.16	-1	1.5	0.66	0.10
	STAT3	3' GGUUAAACACUACGAAGGACUAAACACU	1490	1.3	0.34	1	2	3.34	
MB2-2019-27	MYC	5' AUUUGAGGCAGUUUACAUUUCGGtg	2019	0.3	0.35	0	1	1.35	1.27
	BCL2	3' CUUGAACUCCUUCACUUGUAAAGCCAC	2086	-0.3	0.26	0	1.5	1.76	
sykB2-700-27	SYK	5' UUCCAUUUGUCUUUGAUCUUUGtt	700	-0.2	0.25	0	1	1.25	1.03
	BCL2	3' CGGAAGGUAAAUAUACACUAGAAACAA	4626	0.2	0.31	0	2	2.31	
E2B2-688-27	CCNE2	5' UUCCUCAAGUUUGGAAGCAGCCUtg	688	-0.6	0.22	0	1	1.22	1.13
	BCL2	3' AAAAGAAGGUCAAUCUUAGUCGGAAC	2875	0.6	0.15	0	1.5	1.65	
E2D1-733-27	CCNE2	5' ACCAUCAGUGACGUAAGCAAAGaa	733	0	0.21	0	0	0.21	0.00
	CCND1	3' GAAGGUAGAGACUGAAUUCGUUUUCUU	1256	0	0.23	0	1	1.23	
surD2-2049-27	BIRC5	5' UCCAUUUACAGACUGACACUUUCca	2049	0.6	0.32	0	1.8	2.12	2.11
	CCND2	3' UGAGAUAAAUGUCAGACAUUGAAAGGU	4267	-0.6	0.19	0	2	2.19	
ED1-2880-27	EGFR	5' AUUCCAUGCCAUCACUUUCAGtc	2880	-0.3	0.21	0	1	1.21	1.20
	CCND1	3' GGUAAGGUAAAGGUUCGUGAAAGUCAG	2097	0.3	0.24	0	1	1.24	

The table shows the dual-targeting siRNA names, its targets and duplex as well as the cleavage site ("Clv. site"). The $\Delta\Delta G$ illustrates the difference in predicted thermodynamic end stability in the duplex considering 5 nt at the ends. "RISC entry" indicates whether the strand is favored (1), unfavored (-1), or undecided (0) for RISC uptake. "miRNA score" predicts the likelihood of the strand causing miRNA-like translation inhibition. The strand score (final score) takes all the other factors together. If the two strand scores of two strands of one dual-targeting are both positive and close together (ED1-2880-27), it indicates that both strands have the capacity of down-regulating their target equally.

OAS1. As a positive control, HEK293 cells were incubated with 5 nM tri-phosphate containing RNA obtained after in vitro T7-transcription or 50 ng poly I:C for 24 h. In contrast to the high levels of type I IFN gene response seen with the controls, we did not observe any interferon responses with the dual-targeting siRNAs in this setting (Fig. 5).

Two dual-targeting siRNAs, E2D1-733-27 and E2B2-688-27, which were efficient in all three cell lines, were chosen for comparisons with their mono-functional equivalents. E2B2-688-27 gave 65% down-regulation of both targets, while the mono-functional achieved 80% for CyclinE2 and 60% for Bcl2 (Fig. 6). The E2D1-733-27 dual-targeting siRNA demonstrated near equivalent target knock-downs in HEK293 cells (62% for CyclinE2 and 74% for CyclinD1) as

compared with 60% for Cyclin E2 and 70% for CyclinD1 by its mono-functional equivalents (Fig. 6).

Both dual-targeting and mono-functional siRNAs can induce off-target effects through sequence-specific miRNA-like targeting or by triggering immune responses. We therefore used microarrays to assess the off-targeting effects of the E2D1-733-27 dual-targeting siRNA compared with those of the corresponding mono-functionals. To determine the role of sequence-specific off-target effects, we searched the 3' UTRs of down-regulated genes for sites with perfect complementarity to nucleotides 2–7 and 2–8 from both strands of the siRNAs. We then determined whether genes down-regulated in treated versus mock-transfected HEK293 cells were significantly enriched for

FIGURE 3. Dicer processing of dual-targeting siRNAs. Dicer substrates are processed into smaller molecules by recombinant Dicer. The indicated siRNAs (A) B2S-4543-27, MB2-2019-27, sykB2-700-27, (B) E2D1-733-27, ED1-2880-27, and (C) E2B2-688-27 and surD2-2049-27 were radioactively 5' labeled either at the first antisense strand (1) or at the second antisense strand (2) and incubated in the presence (+) or absence (–) of recombinant human Dicer. The sequence shown was incubated with either NaOH for the alkali ladder or with RNase T1 under denaturing conditions for the G-ladder. The products were resolved on a 20% denaturing polyacrylamide gel.

such 6mer or 7mer miRNA-like seed sites (see Materials and Methods). None of the down-regulated genes in HEK293 cells transfected with E2D1-733-27 were significantly enriched for 6mer and 7mer seed sites ($P \geq 0.2$), which indicated that the dual-targeting siRNA caused few miRNA-like off-target effects. In contrast, the genes down-regulated in the cells transfected with the corresponding mono-functionals were significantly enriched for 6mer seed sites of the Cyclin E2 mono-functional siRNA's passenger strand ($P \leq 5E-15$). We saw the same trend of overrepresentation of 6mer seed sites for the Cyclin E2 passenger strand when comparing the genes down-regulated in the cells transfected with the pooled mono-functionals with the cells transfected with the single dual-targeting siRNA ($P \leq 1E-7$). Consequently, the Cyclin E2 mono-functional siRNA's passenger strand appeared to be active and a source of off-target effects.

To further assess potential off-target effects, we performed gene set enrichment analysis (Subramanian et al. 2005). No gene sets were significantly enriched in the

HEK293 cells transfected with E2D1-733-27 compared with mock-transfected HEK293 cells (false discovery rate $q \geq 0.05$). In contrast, 23 gene sets related to interferon alpha and beta induction were significantly positively enriched in the HEK293 cells transfected with the mono-functionals (Supplemental Table 6). Even more gene sets related to type I interferon genes (33 sets) were significant when comparing the cells transfected with the mono-functionals with the cells transfected with the dual-targeting siRNA (Supplemental Table 7). Moreover, toll-like receptor 3 (TLR3), which recognizes double-stranded RNA, was up-regulated in the HEK293 cells transfected with mono-functionals. Thus, in addition to the pool of mono-functionals inducing significant sequence-based off-targeting, the pool also resulted in significant interferon responses—possibly through TLR3 recognizing the double-stranded RNAs in the pool. The dual-targeting siRNA, in contrast, neither induced significant sequence-dependent nor immune response related off-target effects.

FIGURE 4. Validation of the dual-targeting rational design algorithm. Quantitative real-time PCR performed for dual-targeting siRNAs B2S-4543-27, MB2-2019-27, and sykB2-700-27 (A) against BCL2 (black bars), STAT3 (medium gray bars), MYC (light gray bars), and SYK (open bars), E2B2-688-27 and E2D1-733 (B) against CCNE2 (light gray bars), BCL2 (black bars), CCND1 (medium gray bars), surD2-2049-27 and ED1-2880-27 (C) against BIRC5 (survivin) (black bars), and CCND2 (medium gray bars), EGFR (dark gray bars), and CCND1 (light gray bars). The mRNA was extracted from HEK293, HCT116, and PC3 cells 48 h after transfection. Experiments were performed in independent triplicates.

Dual-targeting siRNAs require two sites with high complementarity. This requirement strongly restricts the number of available target sites and may make genome-wide design of dual-targeting siRNAs impossible. To test the algorithm's applicability in genome-wide design, we designed dual-targeting siRNAs against randomly selected pairs of human protein-coding transcripts from the RefSeq

database (Pruitt et al. 2005). Based on this randomization experiment, we estimated that with default settings, the algorithm will find dual-targeting siRNA candidates for 55% of the possible pairs of human protein coding genes (standard deviation [SD] of 2.2% based on 10 randomizations of 500 pairs each). Lowering the GPboost threshold in step 3 to 0.0 increased the targetable pairs to 96% (SD of 0.75%; 10×500 random pairs). In contrast, increasing the end stability threshold ($\Delta\Delta G$) in step 2 to 2.0 kcal/mol increased the targetable pairs to 67% (SD of 2.4%; 10×200 random pairs). As increasing the $\Delta\Delta G$ threshold will largely increase the number of failed designs, whereas lowering the GPboost threshold to 0 still results in that 60% of the returned sites are highly effective single-targeting siRNAs (data from Reynolds et al. 2004; Saetrom and Snove 2004), lowering the GPboost threshold is likely the better approach for designing dual targeting siRNAs against arbitrary transcript pairs. Alternatively, as we saw no clear correlation between duplex stability and dual targeting (Fig. 2C), allowing additional mismatches between the two strands will likely increase the number of genes that can be targeted without adversely affecting the dual-targeting siRNAs' efficacy. Indeed, by allowing up to six mismatches in step 1, most pairs ($87\% \pm 2.8\%$) of protein coding genes could be targeted by at least one dual-targeting siRNA.

In summary, we have designed an algorithm that can accurately predict dual-targeting siRNAs, which work as efficiently as their mono-functional equivalents. Our approach is general and can be used to target most pairs of human protein coding genes. An advantage of the dual-targeting design compared with two mono-functional siRNAs is that there are two fewer strands that compete

FIGURE 5. Influence of dual-targeting siRNAs on the interferon response. Real-time PCR was performed for IFNB1 (black bars), CDKL2 (light gray bars), and OAS1 (dark gray bars) mRNAs extracted from HEK293 cells 48 h after transfection. As a positive control, HEK293 cells were incubated with 5 nM RNA obtained after in vitro T7-transcription or 50 ng poly I:C for 24 h. Data were normalized to RPLP0 and are shown relative to the mock transfected control. Experiments were done in triplicate.

FIGURE 6. Comparison of dual-targeting siRNAs E2D1-733-27 and E2B2-688-27 with their single-targeting equivalents. HEK293 cells were transfected with the dual-targeting E2D1-733-27 and E2B2-688-27 and their single-targeting equivalents E2D1-E2-27, E2D1-D1-27, E2B2-E2-27, and E2B2-B2-27. Real-time PCR was performed with RNA extracted 48 h after transfection. RPLP0 was used as an internal control. Bcl2 is shown in black, Cyclin E2 in medium gray, and Cyclin D1 in light gray. Experiments were performed in independent triplicates.

for RISC entry, thereby minimizing potential off-targeting (Khan et al. 2009). Moreover, dual-targeting siRNAs typically form imperfect duplexes, which likely help the siRNAs avoiding innate immune responses (Sano et al. 2008). This explains why our dual-targeting siRNA E2D1-733-27 did not trigger the interferon response, whereas the pair of corresponding mono-functional siRNAs apparently did. Although chemical modifications can reduce off-target effects (Jackson et al. 2006a; Behlke 2008), such modifications are not possible when using short hairpin (sh) RNA or miRNA expression cassettes to deliver siRNAs. Designing dual-targeting shRNAs should therefore be a promising alternative to standard shRNAs. Moreover, the dual-targeting approach should be applicable to targets where synthetic lethality is required, such as in cancer applications, or in targeting viral transcripts such as HCV and HIV for which genetic variability often results in escape for a single RNAi trigger. This algorithm should prove useful for the design of clinically relevant siRNAs in which two or more transcripts need to be targeted simultaneously, such as in triggering synthetic lethality in cancer treatment. Furthermore, by combining two or more dual-targeting siRNAs it should be possible to down-regulate four or more targets simultaneously. Compared with using pools of multiple standard siRNAs, such pools of dual-targeting siRNAs may be a more effective approach involving fewer side effects.

MATERIALS AND METHODS

Sequence data

We downloaded target sequences and annotations from the University of California Santa Cruz (UCSC) Table Browser's

(Karolchik et al. 2004) refGene table and used them to identify the targets' 3' UTR. The target genes and their corresponding Reference Sequence (RefSeq) database identifiers are STAT3 (NM_003150), BCL6 (NM_001706), MYC, (NM_002467), BCL2 (NM_000633), SYK (NM_003177), CCNE2 (NM_057749), CCND1 (NM_053056), BIRC5 (NM_001168), CCND2 (NM_001759), EGFR (NM_005228), and TGFA (NM_003236).

Identifying partially complementary dual-targeting siRNA candidates

The number of base pairs determines duplex stability. We used a filter on the number of mismatches to identify stable candidate 19mer dual-targeting siRNA duplexes with five or less mismatches as dual-targeting siRNA candidates. Formally, given two mRNA sequences s_1 and s_2 as input, the set of dual-targeting siRNA candidates will be all pairs of 19mer subsequences $s_{1,i}$ in s_1 and $s_{2,j}$ in s_2 such that $m(s_{1,i}, s_{2,j}) \leq 5$, where $m(s_{1,i}, s_{2,j})$ calculates the number of mismatches when the reverse complement of $s_{1,i}$ is aligned to $s_{2,j}$. Most endogenous miRNAs have symmetric duplexes with few bulges. Asymmetric siRNA duplexes may therefore be suboptimal substrates for RISC. Consequently, to filter out such asymmetric candidates, we used the Hamming distance measure to identify dual-targeting siRNA candidates in our final design algorithm.

Our search strategy does not take into account mismatch positions or thermodynamic parameters, such as base stacking and internal loop and bulge instabilities. Dual-targeting siRNA candidate duplexes identified by our algorithm can therefore appear to have more than five mismatches if aligned by an RNA secondary structure prediction program such as RNAhybrid.

Calculating thermodynamic features

We used RNAfold (Hofacker 2003), RNAhybrid (Rehmsmeier et al. 2004), and sfold (Shao et al. 2007) to compute duplex end and intrastrand stability, overall duplex stability, and target site accessibility. To compute the stability of a given duplex end without overhangs, we placed 1–5 nt from each strand on opposite ends of an artificial hairpin sequence cassette (5'-CCLLLL LGG-3'; L denotes a loop nucleotide) and used RNAfold to predict the stability of the resulting hairpin sequence.

To compute intrastrand stability and overall duplex stability, we used individual duplex strands and both duplex strands as input to RNAfold and RNAhybrid.

To compute the target site accessibility of each strand's cleavage site, we used a sequence consisting of the strand's cleavage site and the flanking 80 nt within the target gene. We then used the site's target site accessibility and the site's mean probability of having unpaired nucleotides within the resulting 179 nt sequence, as predicted by the Sirna application module in sfold.

Evaluating sequence similarity to highly effective standard siRNAs

We used the two algorithms described by Reynolds et al. (2004) and Sætrom (2004) to measure a strand's sequence similarity to highly effective standard siRNAs. The algorithms differ in that the Reynolds algorithm looks for specific sequence features associated with highly effective siRNAs, whereas Sætrom's GPboost algorithm

is a machine learning-based classifier that uses the weighted sum of sequence motifs automatically generated to recognize highly effective siRNAs.

Identifying siRNAs that have a high probability of cleavage-based off-target effects

Short interfering RNAs can induce cleavage of transcripts that have a limited number of mismatches to the siRNA guide strand (Dahlgren et al. 2008). To identify a siRNA's potential off-target transcripts, we used Hamming distance-based searches to screen both strands in the siRNA candidate against the Ensembl (Flicek et al. 2008) human transcriptome database (Snove et al. 2004).

Predicting siRNA's potential for causing translational suppression

Short interfering RNAs cause miRNA-like translational suppression by binding seed sites in 3' UTRs (Birmingham et al. 2006). Multiple seed sites increase the siRNAs' regulatory potential and multiple optimally spaced seed sites give synergistic regulation (Saetrom et al. 2007). To model a siRNA's potential for causing translational suppression, we therefore used a model that takes the number of seed sites and the distance between these seed sites into account. In addition, we assumed that the siRNA's maximum number of consecutively optimally spaced seed sites mostly determines the siRNA's regulatory potential, that additional seed sites have a smaller impact on regulatory potential, and that this impact depends on their individual distance.

Our model uses a concept called seed site modules, which are sets of seed sites where the maximum distance between consecutive seed sites is 35 nt. To score the regulatory potential of a siRNA, we (1) identify the siRNA's seed sites, (2) structure the seed sites into seed site modules, and (3) score individual seed site modules to identify the module with the highest regulatory potential. The score for individual seed site modules is simply the module's number of seed sites. As seed sites that are less than 17 nt apart have either reduced or no cooperative effect, we subtract the number of closely spaced sites when computing the module score. The score for the highest scoring seed site module plus a distance-dependent contribution from the other seed site modules then give the siRNA's regulatory potential.

More formally, given a sequence of n seed sites $S = \{s_1, \dots, s_n\}$ and a sequence of $n - 1$ distances $D = \{d_1, \dots, d_{n-1}\}$, which correspond to the distances between consecutive seed sites such that d_i is the distance between seeds s_i and s_{i+1} . A seed site module M_i is then defined as $M_i = \{s_p, \dots, s_j\}$, where $\forall i \in \{1, \dots, j-1\} d_i \leq 35$. In other words, the distance between all consecutive seed sites in a seed site module is at most 35 nt. Consequently, for a set of seed sites S , there is a corresponding set of seed site modules $SM = \{M_1, \dots, M_k\}$, where $k \leq n - 1$. In other words, any set of seed sites can be structured into a smaller or equal sized set of seed site modules. The score f for seed site module M_i is $f(M_i, D) = 1 + \sum_{l=i}^n g(d_l)$, where $g(d_l) = \{0 \text{ if } d_l \leq 13; 1 \text{ if } d_l \geq 17; \text{ and } (d_l - 13)/4 \text{ otherwise}\}$. To illustrate, a seed site module that consists of four seed sites with distances $\{10, 30, 15\}$ has a score $f = 1 + g(10) + g(30) + g(15) = 1 + 0 + 1 + 2/4 = 2.5$. Finally, the score F for a siRNA with seed sites S , distances D , and seed site modules SM is $F(S, D, SM) = f(M_m, D) + \sum_{Mi \in \{M_1, \dots, M_m\}} f(M_i, D) \times h(d_{i-1}) + \sum_{Mi \in \{M_{m'}, \dots, M_k\}} f(M_i, D) \times h(d_{i-1})$, where $m = i$, $\max \{f(M_i, D)\}$,

$M_i \in SM\}$ (the index of the highest scoring seed module), m' is the index of the seed module following m , and $h(d_{i-1}) = \{0.25 \text{ if } d_{i-1} \geq 70; (70 - d_{i-1})/35 \text{ otherwise}\}$. In other words, the score is given by the highest scoring seed site module M_m plus the scores of the remaining seed site modules weighted by the distance between the individual seed site modules. To illustrate, an siRNA that has three seed site modules with scores $\{2, 4, 3\}$ and individual distances $\{1000, 50\}$ has a score $F = 4 + 2 \times h(1000) + 3 \times h(50) = 4 + 2 \times 0.25 + 3 \times 20/35 \approx 6.21$.

Statistical analysis

We used the R statistical framework (www.r-project.org) and F-tests to create and test linear regression models between siRNA-related features and target gene relative RNA levels.

To determine enrichment of seed sites in genes down-regulated in treated versus mock-transfected cells, we assumed that the number of seed sites in down-regulated genes was binomially distributed, with parameters p equal to the probability of a seed site occurring at random and n equal to the number of down-regulated genes. For a particular seed sequence, we estimated p from the number of RefSeq transcripts containing the seed site in their 3' UTR divided by the total number of RefSeq transcripts considered. We then used the standard-normal approximation of the binomial distribution to calculate a Z-score for x —the observed number of down-regulated genes that contained the seed site. We considered this approximation to be adequate, as np and $n(1 - p)$ were greater than 35 for most tests. We used a binomial test to confirm the results of the Z-test for the other cases.

siRNA nomenclature

All dual-targeting siRNA duplex names consist of three parts: the first describes the two target genes (antisense strand 1 is listed first; S = STAT3, B2 = BCL2, B = BCL6, M = MYC, syk = SYK, E2 = CCNE2, D1 = CCND1, D2 = CCND2, sur = BIRC5 [survivin], and E = EGFR). The second part refers to the first nucleotide of the 5' end position of the first target mRNA, and the last part refers to the Dicer-substrate 27mer design. Single-targeting siRNA duplexes include the dual-targeting siRNA name and their target gene.

siRNA sequences

Supplemental Tables 1 and 2 list mono-functional and dual-targeting siRNA sequences. All siRNAs were purchased from IDT-DNA.

Cell culture and transfection

HEK293 and HCT116 cells were grown in P100 dishes and maintained at 37°C and 5% CO₂ in DMEM supplemented with 10% FCS, 1 mM sodium pyruvate, and 2 mM glutamine. PC3 cells were cultured in RPMI supplemented with 10% FCS and 2 mM glutamine. For transfection, a 25 pmol siRNA duplex (50 nM end concentration) was used following the RNAimax lipofectamine (Invitrogen) manufacturer's protocol. RNA was extracted after 48 h incubation.

Quantitative real-time PCR

RNAStat60 reagent was used to extract total RNA according to the manufacturer's instructions (Tel-Test). Residual DNA was

digested using the DNA-free kit following the manufacturer's instructions (Ambion). The cDNA was produced using 2 µg of total RNA, Moloney murine leukemia virus reverse transcriptase, and random primers in a 20 µL reaction according to the manufacturer's instructions (Invitrogen). Expression of human mRNAs encoding STAT3, BCL2, MYC, SYK, CCNE2, CCND1, CCND2, BIRC5 (survivin), EGFR, and RPLP0 were determined by quantitative real-time PCR (iCycler, Bio-Rad) using 2X iQ SyberGreen Mastermix (Bio-Rad) and specific primer sets (Supplemental Table 3) at final concentrations of 400 nM. RPLP0 expression was used as an internal control.

Interferon response assay

HEK293 cells were transfected with dual-targeting siRNAs as described above. For the positive controls, HEK293 cells were transfected with 5 nM RNA obtained after in vitro T7-transcription or 50 ng poly I:C for 24 h. RNA extraction, cDNA synthesis, and quantitative real-time PCR were performed as described above; see Supplemental Table 3 for primers. RPLP0 expression was used for normalization of the real-time PCR data. Experiments were done in triplicate.

In vitro Dicer assay

One hundred-picomole single strands of dual-targeting siRNAs were radioactively labeled at the 5' end and purified with Micro-SpinTM G-25 Columns (Amersham Biosciences) according to the manufacturer's instructions. Radioactively labeled strands were annealed to 100 pmol of cold corresponding strands. Approximately 400,000 cpm of each annealed duplex was incubated with and without recombinant Dicer (Ambion) at 37°C for 1 h and products were resolved on a 20% denaturing polyacrylamide gel. The 5' labeled sequence 5'-AACAAAGATCACATATAAATGGAA GGC-3', incubated in the presence of NaOH at 95°C for 10 min, resulted in the alkali ladder. For the G-ladder, the same sequence was incubated with RNase T1 (Ambion) under denaturing conditions at room temperature for 15 min.

Microarray

HEK293 cells were transfected with Lipofectamin RNAimax (Invitrogen) either with 50 nM E2D1-733-27 or E2D1-D1-733-27 and E2D1-E2-733-27. As a negative control Mock transfected HEK293 cells were used. Total RNA of all samples was extracted with RNeasy60 (Tel-Test) according to the manufacturer's protocol and the microarray analysis was performed by the City of Hope Micro Array Core Facility using Affymetrix human gene 1.0 array.

The statistical analyses of the microarrays were performed by City of Hope Micro Array Core Facility preprocessing of the probes includes IterPlier normalization, quality assessment, and filtering.

For filtering, probe sets with log2 intensity < 4 in more than two arrays were considered uninformative and removed. Regression and cutoff for fold change (FC) were set to 1.5. We performed final statistical analyses for off-target effects.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at <http://www.rnajournal.org>.

ACKNOWLEDGMENTS

This work was supported by a grant from the Keck Foundation to S.J.F. and the NIH/LBI to J.J.R., as well as SPORE; P.S. received support from the Norwegian Functional Genomics Program (FUGE) of the Norwegian Research Council. P.S., B.H., K.T., and J.J.R. conceived and designed the experiments. B.H. and K.T. performed all the experiments and analyzed the data. P.S. wrote the algorithm and analyzed the microarray off-target effects. P.S. and K.T. wrote the manuscript. S.J.F. advised the target mRNAs for testing.

Received November 13, 2009; accepted February 15, 2010.

REFERENCES

- Behlke MA. 2008. Chemical modification of siRNAs for in vivo use. *Oligonucleotides* **18**: 305–319.
- Birmingham A, Anderson EM, Reynolds A, Ilesley-Tyree D, Leake D, Fedorov Y, Baskerville S, Maksimova E, Robinson K, Karpilow J, et al. 2006. 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods* **3**: 199–204.
- Boden D, Pusch O, Lee F, Tucker L, Ramratnam B. 2003. Human immunodeficiency virus type 1 escape from RNA interference. *J Virol* **77**: 11531–11535.
- Dahlgren C, Zhang HY, Du Q, Grahn M, Norstedt G, Wahlestedt C, Liang Z. 2008. Analysis of siRNA specificity on targets with double-nucleotide mismatches. *Nucleic Acids Res* **36**: e53.
- Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, et al. 2008. Ensembl 2008. *Nucleic Acids Res* **36**: D707–D714.
- Heale BS, Soifer HS, Bowers C, Rossi JJ. 2005. siRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res* **33**: e30. doi: 10.1093/nar/gni026.
- Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**: 3429–3431.
- Hossbach M, Gruber J, Osborn M, Weber K, Tuschl T. 2006. Gene silencing with siRNA duplexes composed of target-mRNA-complementary and partially palindromic or partially complementary single-stranded siRNAs. *RNA Biol* **3**: 82–89.
- Jackson AL, Burchard J, Leake D, Reynolds A, Schelter J, Guo J, Johnson JM, Lim L, Karpilow J, Nichols K, et al. 2006a. Position-specific chemical modification of siRNAs reduces 'off-target' transcript silencing. *RNA* **12**: 1197–1205.
- Jackson AL, Burchard J, Schelter J, Chau BN, Cleary M, Lim L, Linsley PS. 2006b. Widespread siRNA 'off-target' transcript silencing mediated by seed region sequence complementarity. *RNA* **12**: 1179–1187.
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**: D493–D496.
- Khan AA, Betel D, Miller ML, Sander C, Leslie CS, Marks DS. 2009. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol* **27**: 549–555.
- Khvorova A, Reynolds A, Jayasena SD. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209–216.
- Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. 2005. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**: 607–620.
- Polo JM, Dell'Oso T, Ranuncolo SM, Cerchietti L, Beck D, Da Silva GF, Prive GG, Licht JD, Melnick A. 2004. Specific peptide interference reveals BCL6 transcriptional and oncogenic mechanisms in B-cell lymphoma cells. *Nat Med* **10**: 1329–1335.
- Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): A curated nonredundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **33**: D501–D504.

- Ramanarayanan J, Hernandez-Ilizaliturri FJ, Chanan-Khan A, Czuczman MS. 2004. Pro-apoptotic therapy with the oligonucleotide Genasense (oblimersen sodium) targeting Bcl-2 protein expression enhances the biological anti-tumour activity of rituximab. *Br J Haematol* **127**: 519–530.
- Rand TA, Petersen S, Du F, Wang X. 2005. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**: 621–629.
- Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. 2004. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**: 1507–1517.
- Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. 2004. Rational siRNA design for RNA interference. *Nat Biotechnol* **22**: 326–330.
- Saetrom P. 2004. Predicting the efficacy of short oligonucleotides in antisense and RNAi experiments with boosted genetic programming. *Bioinformatics* **20**: 3055–3063.
- Saetrom P, Snove O Jr. 2004. A comparison of siRNA efficacy predictors. *Biochem Biophys Res Commun* **321**: 247–253.
- Saetrom P, Heale BS, Snove O Jr, Aagaard L, Alluin J, Rossi JJ. 2007. Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res* **35**: 2333–2342.
- Sano M, Li H, Nakanishi M, Rossi JJ. 2008. Expression of long anti-HIV-1 hairpin RNAs for the generation of multiple siRNAs: Advantages and limitations. *Mol Ther* **16**: 170–177.
- Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199–208.
- Shao Y, Chan CY, Maliyekkel A, Lawrence CE, Roninson IB, Ding Y. 2007. Effect of target secondary structure on RNAi efficiency. *RNA* **13**: 1631–1640.
- Snove O Jr, Nedland M, Fjeldstad SH, Humberset H, Birkeland OR, Grunfeld T, Saetrom P. 2004. Designing effective siRNAs with off-target control. *Biochem Biophys Res Commun* **325**: 769–773.
- Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ, Kellis M. 2007. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res* **17**: 1865–1879.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102**: 15545–15550.

Dual-targeting siRNAs

Katrin Tiemann, Britta Höhn, Ali Ehsani, et al.

RNA 2010 16: 1275-1284 originally published online April 21, 2010
Access the most recent version at doi:[10.1261/rna.2005710](https://doi.org/10.1261/rna.2005710)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2010/03/30/rna.2005710.DC1>

References

This article cites 29 articles, 7 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/16/6/1275.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>