Effect of target secondary structure on RNAi efficiency

YU SHAO,[1] CHI YU CHAN,[1] ANIL MALIYEKKEL,[2,3] CHARLES E. LAWRENCE,[1,4] IGOR B. RONINSON,[2] and YE DING[1]

[1]Wadsworth Center, New York State Department of Health, Albany, New York 12208, USA
[2]Cancer Center, Ordway Research Institute, Albany, New York 12208, USA
[3]Department of Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, USA
[4]Center for Computational Molecular Biology, and Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

ABSTRACT

RNA interference (RNAi) mediated by small interfering RNAs (siRNAs) or short hairpin RNAs (shRNAs) has become a powerful tool for gene knockdown studies. However, the levels of knockdown vary greatly. Here, we examine the effect of target disruption energy, a novel measure of target accessibility, along with other parameters that may affect RNAi efficiency. Based on target secondary structures predicted by the Sfold program, the target disruption energy represents the free energy cost for local alteration of the target structure to allow target binding by the siRNA guide strand. In analyses of 100 siRNAs and 101 shRNAs targeted to 103 endogenous human genes, we find that the disruption energy is an important determinant of RNAi activity and the asymmetry of siRNA duplex asymmetry is important for facilitating the assembly of the RNA-induced silencing complex (RISC). We estimate that target accessibility and duplex asymmetry can improve the target knockdown level significantly by nearly 40% and 26%, respectively. In the RNAi pathway, RISC assembly precedes target binding by the siRNA guide strand. Thus, our findings suggest that duplex asymmetry has significant upstream effect on RISC assembly and target accessibility has strong downstream effect on target recognition. The results of the analyses suggest criteria for improving the design of siRNAs and shRNAs.

Keywords: target structure; RNA folding; RNAi

INTRODUCTION

RNA interference (RNAi) is a sequence-specific gene silencing mechanism that is induced by double-stranded RNA (dsRNA) homologous to the target gene (Fire et al. 1998). RNAi can be mediated either by small interfering RNAs (siRNAs) of about 21 nucleotides (nt) with two-nucleotide 3′ overhang (Elbashir et al. 2001) or by stably expressed short hairpin RNAs (shRNAs), which are processed by Dicer into siRNAs (Brummelkamp et al. 2002; Paddison et al. 2002). During activation of the RNA-induced silencing complex (RISC), the guide (antisense) strand of the siRNA duplex is preferentially assembled into the RISC when the stem formed by the 5′ end and its complement is less stable than the one formed by the 3′ end and its complement (Khvorova et al. 2003; Schwarz et al. 2003); the "passenger" (sense) strand is cleaved by Argonaute2 (Ago2), the catalytic component of RISC (Matranga et al. 2005; Rand et al. 2005). The antisense strand guides Ago2 to cleave mRNA by base-pairing with the complementary site in the target.

Large variation in the efficiency of siRNAs for different sites on the same target is commonly observed (Holen et al. 2002). Usually, only a small proportion of randomly selected siRNAs are potent. Thus, there has been great interest in determining rules for improvement of RNAi design. A number of empirical rules on siRNA duplex features have been reported. These include the asymmetry rule for siRNA duplex ends, which requires that the 5′ end of the antisense strand forms a stem with its complement that is less stable than the stem formed by the 5′ end of the sense strand (Khvorova et al. 2003; Schwarz et al. 2003). The asymmetry rule is strongly related to the requirements of high A/U content at the 5′ end of the antisense strand and high G/C at the 5′ end of the sense strand (Reynolds et al. 2004; Ui-Tei et al. 2004). A number of position-specific nucleotide preferences and other siRNA sequence features have been proposed (Reynolds et al. 2004; Patzel et al. 2005). In addition, the importance of target secondary

Reprint requests to: Ye Ding, Wadsworth Center, New York State Department of Health, 150 New Scotland Avenue, Albany, NY 12208, USA; e-mail: yding@wadsworth.org; fax: (518) 402 4623.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.546207.

structure and accessibility has been suggested by several studies based on computational modeling of target structure and accessibility (Kretschmer-Kazemi Far and Sczakiel 2003; Luo and Chang 2004; Heale et al. 2005; Schubert et al. 2005) and was supported by compelling evidence based on experimentally assessed accessibility (Lee et al. 2002; Bohula et al. 2003; Vickers et al. 2003; Overhoff et al. 2005; Westerhout et al. 2005). Strikingly, it was observed that HIV can escape RNAi-mediated inhibition by a single point mutation that alters the accessibility of the target site (Westerhout et al. 2005). The significance of target structure has long been established for antisense oligonucleotides and *trans*-cleaving ribozymes (Zhao and Lemke 1998; Vickers et al. 2000). For RNAi, however, this has been disputed in several reports, with one based on limited computational analysis (Reynolds et al. 2004; Boese et al. 2005).

Rules for siRNA duplex features are straightforward to quantify and implement for the purpose of rational RNAi design. However, since a messenger RNA (mRNA) is unlikely to have a single stable structure, computational modeling of the target secondary structure and assessment of the effect of secondary structure on target accessibility are much more challenging. To address this challenge, we introduce a novel quantitative measure of target accessibility, target disruption energy, based on structures predicted by the Sfold program, which generates a statistically representative sample from the Boltzmann weighted ensemble of secondary structures (Ding and Lawrence 2003; Ding et al. 2004). We employ this approach with three aims: (1) quantify target structural accessibility, (2) quantitatively assess the net contribution of the target accessibility to RNAi efficiency in the context of the RNAi pathway, and (3) establish a general model for efficient RNAi. We examine target disruption energy along with a number of other parameters that can affect RNAi efficiency. From an analysis of 100 published siRNAs for three endogenous human genes, we found that disruption energy is the most significant parameter for one target and is second in significance level to duplex asymmetry for the other two targets. To quantitatively assess the effects of target accessibility and duplex asymmetry, we utilize an independent data set of 101 shRNAs for 100 endogenous human genes. We found that target accessibility and duplex asymmetry can improve the target knockdown level significantly by nearly 40% and 26%, respectively. These findings suggest that, after RISC assembly, target secondary structure plays an important role in target binding by the guide siRNA strand. Thus, effective silencing by RNAi favors siRNAs with sequence features that facilitate RISC activation, as well as accessible target sites that enable intermolecular base-pairing for target recognition. The results of the analyses suggest criteria for improving the design of siRNAs and shRNAs.

RESULTS

Statistical analyses of siRNA data sets

We first performed weighted regression analyses for the siRNA data sets (see Materials and Methods and also Fig. 1). For lamin A, $\Delta G_{\text{disruption}}$ is the only significant parameter, with a *P*-value of 1.05E−8, and is highly predictive of siRNA activity with a regression R^2 of 0.7656 (Fig. 2; Table 1,). For PTEN and CD54, $\Delta G_{\text{disruption}}$ is the second best predictor, with a *P*-value of 1.80E−16 and a R^2 of 0.6073 (Table 2). $\Delta G_{\text{disruption}}$ is the only parameter that is significantly correlated with siRNA activities for all data sets. *DSSE* is the best predictor for PTEN and CD54, with a *P*-value of 5.24E−36 and a high R^2 of 0.8849 (Table 2). The lack of significance of *DSSE* for lamin A could be due to the siRNA design constraints (Harborth et al. 2003); these could have biased the representation of the nucleotide composition for the duplex ends of the tested siRNAs. S_D is not correlated with siRNA activity for any of the data sets (Tables 1 and 2). ΔG_{hybrid} was found to be significant only for PTEN and CD54 (Table 2). However, its predictive value as measured by R^2 is relatively poor in comparison to $\Delta G_{\text{disruption}}$ and *DSSE*. Thus, in further analyses of effects of target structure on RNAi efficiency, we focused on $\Delta G_{\text{disruption}}$ and *DSSE*.

As alternatives to Sfold, we also computed $\Delta G_{\text{disruption}}$, using target structures predicted by other RNA folding programs, and performed the weighted regression analysis. We found that Sfold is by far the best performer. For any of the other programs, either there is a lack of statistical significance or the R^2 is rather poor in the case of statistical significance (Table 3).

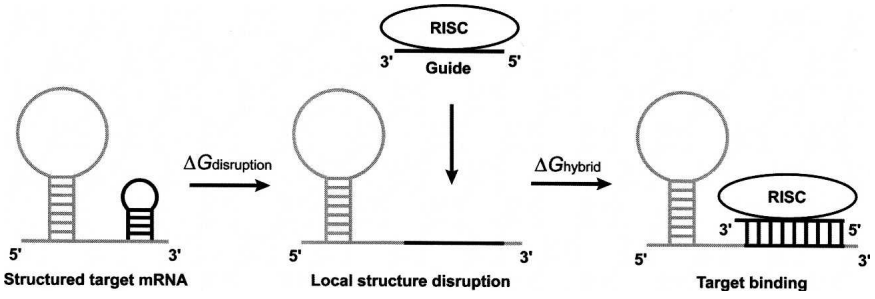

FIGURE 1. Energetic exchanges for local target disruption by binding of guide siRNA. Target disruption energy, $\Delta G_{\text{disruption}}$, a measure of target accessibility, is the free energy cost for opening the local secondary structure at the target site; ΔG_{hybrid} is the free energy gain due to the hybridization between the guide siRNA strand and the target site.

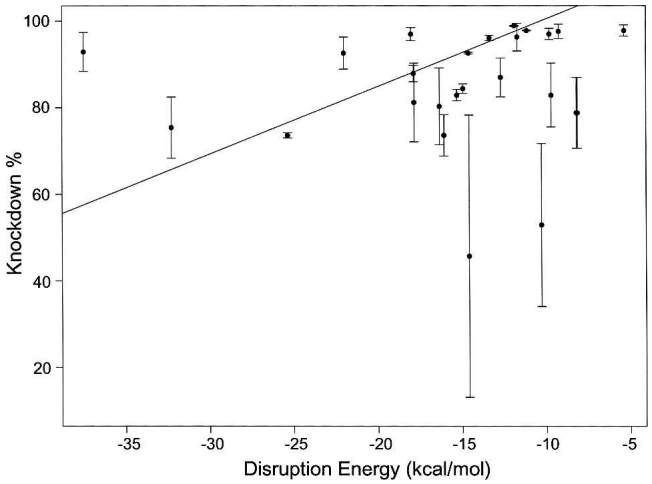

FIGURE 2. Weighted linear regression analysis of siRNA knockdown data for the lamin A siRNA data set using disruption energy, with a regression $R^2 = 0.766$ and a *P*-value of 1.05E−8. Error bars represent standard deviations from at least three independent experiments. Larger weights are assigned to data points that have smaller standard deviations. The fitted regression line is thus dictated by data points with small standard deviations.

Quantitative assessment of the effects of target structure and duplex asymmetry

Separating the effect of target structure from effects of upstream factors

We were most interested in obtaining a quantitative estimate of the net effect of target structure on RNAi efficiency, by taking advantage of the relatively large independent shRNA data set. Because target structure is only relevant for the target recognition step of the RNAi pathway, factors that can have negative effects on the upstream steps of the RNAi pathway must be considered in such a quantitative analysis. In other words, the overall effects of upstream factors and target structure are convoluted in the knockdown data. For example, a siRNA or shRNA targeted to an accessible region will not necessarily be functional if the guide strand could not be successfully assembled into the RISC. To address this issue of convolution, we adopted the approach of using data filters to control for the upstream effects of nonstructure factors.

For the shRNAs in the cDNA library, a low or high GC content and the occurrence of a AAAA, TTTT, GGGG, or CCCC motif were observed to have negative impacts on RNAi activity. The AAAA motif or the TTTT motif has the tendency to cause the premature termination of transcription of shRNAs from the RNA polymerase III promoter (Geiduschek and Kassavetis 2001). A GC-rich sequence can promote formation of quadruplex structures (Hardin et al. 1992), and GGGG can form tetraplex structures (Laughlan et al. 1994), or it can cause potential nonspecific effects through its interaction with heparin-binding proteins (Stein 1999). Thus, to remove these potential adverse effects, we consider two filters: (1) 30%≤GC%≤70% and (2) absence of AAAA, TTTT, GGGG, and CCCC motifs. In addition, to separate the downstream effect of target structure from the upstream effect of duplex asymmetry, we enforce the rule of asymmetry (*DSSE* > 0.0 kcal/mol; see Materials and Methods) for estimating the effect of target structure.

The siRNAs resulting from shRNA cleavage by Dicer are mostly 19 base pairs (bp) or 20 bp in length (with additional 2-nt 3′ overhang), at comparable yields (Rose et al. 2005). Because the computational results are highly similar for both lengths, we focus on reporting the results for the length of 19 bp (the guide strand sequences are given in Supplemental Table 1).

Assessing the net effect of target accessibility

From weighted regression analysis, we found that $\Delta G_{\text{disruption}}$ is the most important parameter, with a *P*-value of 1.52E−8. To assess the net effect of target accessibility, we make a two-group comparison between accessible sites and inaccessible sites. Because $\Delta G_{\text{disruption}}$ is a quantitative measure of accessibility and is positively correlated with RNAi efficiency, we consider a target site accessible if its $\Delta G_{\text{disruption}} > M$ kcal/mol, and the site inaccessible if $\Delta G_{\text{disruption}} < N$ kcal/mol, where M and N are two threshold values for defining accessibility and inaccessibility. For the shRNA data set, we found that the average knockdown level for accessible sites is maximized at $M = -10$, and that the average knockdown level for inaccessible sites is minimized at $N = -19$. Using these two threshold values, the difference in average knockdown levels is 39.7%, and this improvement by accessibility is highly significant with a *P*-value of 0.0004 by the *t*-test and 0.0007 by the nonparametric Wilcoxon rank sum test. For several alternative pairs of the thresholds, the difference in the knockdown levels is over 30% (Table 4), and the average improvement for all five pairs of thresholds is 34.62%. When the rule of asymmetry is not enforced and the two filters are not applied, the degree of improvement is substantially reduced, but still significant with an average of 14.18% (see Supplemental Table 2). These results indicate that the net effect of target structure is substantially underestimated if upstream factors were not taken into consideration.

TABLE 1. Weighted regression results for lamin A siRNA data set

Parameter	Coefficient	*P*-value	R^2
$\Delta G_{\text{disruption}}$	0.0156	1.05E−08	0.7656
ΔG_{hybrid}	0.0023	0.0896	0.1202
DSSE	−0.0031	0.5370	0.0168
S_D	−0.0054	0.7067	0.0063

TABLE 2. Weighted regression results for PTEN and CD54 siRNA data set

Parameter	Coefficient	*P*-value	R^2
$\Delta G_{disruption}$	0.0586	1.80E−16	0.6073
ΔG_{hybrid}	0.0431	9.49E−09	0.3651
DSSE	0.1393	5.24E−36	0.8849
S_D	−0.0530	0.3265	0.0132

Assessing the effect of duplex asymmetry

From a weighted regression analysis of the shRNA data set, *DSSE* was found to be significant with a *P*-value of 0.00096. We next compared the average knockdown level for those shRNAs that meet the rule of asymmetry (*DSSE* > 0.0 kcal/mol) and the average for those that do not (*DSSE* ≤ 0.0 kcal/mol). We found that the improvement by enforcing the rule of asymmetry is 10.21% (*P*-value of 0.0197 by the one-sided *t*-test and 0.0284 by the one-sided Wilcoxon rank sum test). For 19 shRNAs that pass the two filters and have accessible target sites ($\Delta G_{disruption} > -10$ kcal/mol), the improvement by duplex asymmetry is 25.99% (*P*-value of 0.0082 by the *t*-test, and 0.0086 by the Wilcoxon rank sum test; also see Table 5). Because the two filters appear to be associated with adverse events upstream of the RISC assembly, 25.99% may be a more accurate estimate of the effect of duplex asymmetry for the shRNA data set.

Combined effects of target accessibility and duplex asymmetry

To examine the combined effects of target accessibility and duplex asymmetry, we also assessed the improvement by one parameter under the negative condition specified by the other parameter (Table 5). The two filters were also used in this assessment to minimize effects of upstream factors. For shRNAs that failed the asymmetry test, we found that target accessibility can still improve the knockdown level by 16.03%. For shRNAs targeted to inaccessible sites ($\Delta G_{disruption} < -19$ kcal/mol), however, duplex asymmetry did not make an appreciable difference. In the RNAi pathway, duplex asymmetry is concerned with the upstream step of RISC assembly, while target accessibility presumably governs target recognition. Our results as summarized in Table 5 show that duplex asymmetry is not a rate-limiting factor and that target accessibility is the more influential factor. Furthermore, the combination of both duplex asymmetry and accessibility was found to yield the highest level of improvement.

Software availability

$\Delta G_{disruption}$, *DSSE*, and other tools for structure-based rational RNAi design are available through the application module Sirna of the Sfold software for the folding and design of nucleic acids. Sfold is available through Web server at http://sfold.wadsworth.org.

DISCUSSION

In this work, we have studied the effects of a number of siRNA duplex features and the effect of predicted target secondary structure on the efficiency of RNAi. We have introduced a novel measure of target structural accessibility $\Delta G_{disruption}$; we found this measure to be the most important predictor for RNAi activity. *DSSE*, an implementation of the asymmetry rule (Khvorova et al. 2003; Schwarz et al. 2003), was found to be an important duplex sequence feature. By taking advantage of a shRNA data set and by controlling for

TABLE 3. Comparison of results of weighted regression for predicting silencing efficiency by $\Delta G_{disruption}$ computed with target structures predicted by various RNA folding programs

		MFE structure			1000 suboptimal structures by Mfold[c]	1000 lowest energy structures by RNAsubopt[d]	1000 suboptimal structures by RNAstructure[e]	1000 structures sampled by Sfold
		by Mfold	by RNAfold[a]	by RNAstructure[b]				
lamin A	R^2	0.0748	0.3324	0.0056	0.0793	0.3322	0.0091	0.7656
	P-value	0.1860	0.0026	0.7230	0.1730	0.0026	0.6500	1.05E−08
PTEN and CD54	R^2	0.0231	0.0908	0.0158	0.0021	0.0673	0.0141	0.6073
	P-value	0.1930	0.0086	0.2819	0.6977	0.0246	0.3094	1.80E−16

[a]Implementation of free energy minimization by the Vienna RNA package (Hofacker 2003).
[b]Implementation of free energy minimization for Windows platform (Mathews et al. 2004).
[c]A set of 1000 suboptimal structures by Mfold (Zuker 2003) was generated using the following settings for folding parameters: percent suboptimality *P* = 100, window parameter *W* = 0, and maximum number of foldings MAX = 1000 (a much smaller set of structures was generated for default parameter settings).
[d]Complete suboptimal folding implemented by the Vienna RNA package (Wuchty et al. 1999; Hofacker 2003).
[e]A set of 1000 suboptimal structures by RNAstructure (Mathews et al. 2004) was generated using the following settings for folding parameters: max % energy difference = 100, max number of structures = 1000, and window size = 0 (a much smaller set of structures was generated for default parameter settings).

TABLE 4. Net improvement in knockdown level by target accessibility for various $\Delta G_{disruption}$ thresholds[a]

Energy threshold in kcal/mol for accessible sites (number of data points)	Energy threshold in kcal/mol for inaccessible sites (number of data points)	Net improvement in average knockdown %[b]	*t*-test *P*-value[c]	Wilcoxon rank sum test *P*-value[c]
$\Delta G_{disruption} > -9$ (5)	$\Delta G_{disruption} < -20$ (6)	38.72	0.0066	0.0152
$\Delta G_{disruption} > -10$ (10)	$\Delta G_{disruption} < -19$ (8)	39.72	0.0004	0.0007
$\Delta G_{disruption} > -11$ (12)	$\Delta G_{disruption} < -18$ (12)	31.21	0.0003	0.0004
$\Delta G_{disruption} > -12$ (14)	$\Delta G_{disruption} < -17$ (17)	32.11	9.34E−05	0.0002
$\Delta G_{disruption} > -13$ (16)	$\Delta G_{disruption} < -16$ (19)	31.32	6.77E−05	0.0002

[a]Estimated by using 42 of 101 shRNAs in the cDNA library that meet the rule of duplex asymmetry (DSSE > 0 kcal/mol) and satisfy the two filtering criteria: (1) 30% ≤ GC% ≤ 70% and (2) absence of AAAA, TTTT, GGGG, and CCCC motifs.
[b]Net improvement in average knockdown % due to target accessibility = average knockdown % for accessible sites − average knockdown % for inaccessible sites.
[c]One-tailed test of significance for average knockdown % for accessible sites being greater than that for inaccessible sites.

factors that may negatively affect upstream steps in the RNAi pathway, we found that target accessibility and duplex asymmetry can improve the target knockdown level significantly by nearly 40% and 26%, respectively. These percentages are far greater than the degree of improvement by any of the single sequence features reported in a previous study (Reynolds et al. 2004). Our qualitative findings are consistent with a previous report based on alternative calculations (Heale et al. 2005).

For efficient gene silencing, a number of studies reported the significance of siRNA sequence features (Khvorova et al. 2003; Schwarz et al. 2003; Reynolds et al. 2004), and other studies reported the importance of target structure (Bohula et al. 2003; Kretschmer-Kazemi Far and Sczakiel 2003; Vickers et al. 2003; Yoshinari et al. 2004; Overhoff et al. 2005; Schubert et al. 2005; Westerhout et al. 2005). Based on our findings, we propose a simple model for efficient RNAi that combines both perspectives in the context of the RNAi pathway (Fig. 3). The asymmetry of siRNA duplex ends is important for RISC assembly, whereas target accessibility is important for the downstream step of target recognition in the RNAi pathway. A siRNA designed for an accessible target site will not necessarily be functional, if it does not have a favorable *DSSE* for effective assembly of the guide strand into RISC. Likewise, a siRNA with a favorable *DSSE* will not necessarily yield potent silencing when the guide strand cannot effectively bind to the highly structured target site. Thus, the combination of favorable *DSSE* and target accessibility can greatly improve the efficiency of RNAi. Because the two factors operate sequentially in the RNAi pathway, their effects on the efficiency of RNAi are heavily convoluted. Deconvolution is necessary to tease apart the individual effects, particularly the net effect of target structure. We expect the model to be generally valid in many experimental systems. However, exceptions are likely due to system-specific factors. For example, in a viral system, a number of siRNAs targeted to experimentally identified accessible sites were not effective (Das et al. 2004), and some of them do have favorable duplex asymmetry.

We adopted a population approach to modeling of mRNA secondary structure by employing the Sfold

TABLE 5. Improvement in knockdown level for 72 of 101 shRNAs in the cDNA library[a]

Criterion for assessing improvement	Control condition	Improvement in average knockdown %	*t*-test *P*-value[b]	Wilcoxon rank sum test *P*-value[b]
Target accessibility[c]	DSSE > 0 kcal/mol	39.72	0.0004	0.0007
	DSSE ≤ 0 kcal/mol	16.03	0.0888	0.0836
Duplex asymmetry[d]	$\Delta G_{disruption} > -10$ kcal/mol	25.99	0.0082	0.0086
	$\Delta G_{disruption} < -19$ kcal/mol	2.29	0.4197	0.3605

[a]Satisfying the two filtering criteria: (1) 30% ≤ GC% ≤ 70% and (2) absence of AAAA, TTTT, GGGG, and CCCC motifs.
[b]One-tailed test for significance of improvement.
[c]For sites that meet the control condition on DSSE, improvement by target accessibility = (average knockdown % for accessible sites, i.e., sites with $\Delta G_{disruption} > -10$ kcal/mol) − (average knockdown % for inaccessible sites, i.e., sites with $\Delta G_{disruption} < -19$ kcal/mol); for DSSE > 0 kcal/mol, there are 10 accessible sites and 8 inaccessible sites; for DSSE ≤ 0 kcal/mol, there are nine accessible sites and eight inaccessible sites.
[d]For sites that meet the control condition on $\Delta G_{disruption}$, improvement by duplex asymmetry = (average knockdown % for sites passing the asymmetry test, i.e., DSSE > 0 kcal/mol) − (average knockdown % for sites failing the asymmetry test, i.e., DSSE ≤ 0 kcal/mol); for $\Delta G_{disruption} > -10$ kcal/mol, there are 10 sites passing the asymmetry test and 9 sites failing the test; for $\Delta G_{disruption} < -19$ kcal/mol, there are eight sites passing the asymmetry test and eight sites failing the test.

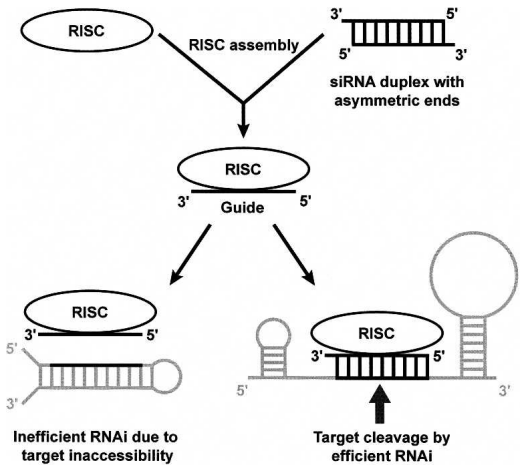

FIGURE 3. A proposed simple model for efficient RNAi. RISC assembly is facilitated by asymmetric ends of siRNA duplex; target recognition via intermolecular base-pairing is aided by structural accessibility at the target site. The combination of the upstream effect of duplex asymmetry and the downstream effect of target accessibility is generally essential for potent gene silencing.

program. This approach has been found to perform better than the minimum free energy method for prediction of the activities of antisense oligonucleotides (Ding and Lawrence 2001). In the current application, the sampling approach was found to outperform other established programs for RNA secondary structure predictions (Table 3). In a recent work on predicting microRNA–target interactions, Sfold was used for extensive structural analyses (Long et al. 2007). These analyses were based on probabilistic accessibility profiling, statistics of open nucleotide blocks, and a two-step hybridization model that involves energy calculations specific to microRNA–target interactions. A potent effect of target structure on microRNA function was observed from both data analyses and in vivo experimental testing (Long et al. 2007). The effect of target structure appears to be stronger for microRNAs than for siRNAs/shRNAs. There are two potential reasons for this observation. First, the rule of duplex asymmetry is not relevant for mature microRNAs that are single-stranded. Second, this may be due to mechanistic differences between target cleavage by RNAi and translation repression by microRNAs. For future research, the applicability of the two-step hybridization model to modeling siRNA–target interactions warrants investigation.

In the calculation of $\Delta G_{\text{disruption}}$, we assumed that the binding of target mRNA by the siRNA guide strand induces only a local structural alteration at the target site. It is likely that in some, if not all, cases, nucleotides outside the target site will also contribute to the energy change owing to siRNA binding. An alternative to the local disruption model is a global disruption model, which assumes that the rest of the target mRNA molecule can completely refold after siRNA binding. For this model, $\Delta G_{\text{disruption}}$ can be recalculated by constraining the target site to be single stranded and refolding the rest of the target mRNA. However, the predictability by $\Delta G_{\text{disruption}}$ is rather low (for lamin A, P-value = 0.0223, R^2 = 0.2069; for PTEN and CD54, P-value = 2.05E−05, R^2 = 0.2213). This suggests that target cleavage occurs rapidly after target binding by the siRNA guide strand such that global refolding of the target before cleavage is unlikely. While partial refolding is a possibility, it is highly uncertain what region of the target may be involved in refolding. Thus, it is difficult to construct a computational model that may represent a reasonable compromise between the local model and the global model. For assessing improvement in predictions, the performance of any intermediate model will need to be compared with that for the local model.

It has been reported that the 5′ bases of the siRNA are more important than the 3′ bases for the strength of target binding (Haley and Zamore 2004), and that nucleotides 2–8 of the 5′ end of microRNAs are important for target recognition (Lewis et al. 2005). We thus statistically tested the hypothesis that functional RNAi requires good accessibility for the 3′ end of the target site. For each of all target sites for siRNAs in our study, we computed the average accessibility for the first five bases from the 3′ end of the target site and also the average accessibility for nucleotide positions 2–8 from the 3′ end of the target site. The probability that a base is unpaired is computed by the Sfold structure sample (Ding and Lawrence 2001). The average accessibility for a block of nucleotides is computed by the sum of the unpaired probabilities divided by the number of bases in the block, i.e., the average unpaired probability for the block. To facilitate a two-group statistical comparison, a block is considered to be accessible if the average accessibility is $\geq$0.5 and inaccessible if the average accessibility is <0.5. All siRNA target sites were partitioned into two groups: group 1 with the 3′ end of the target site being accessible and group 2 with an inaccessible 3′ end. Here we consider the first five bases from the 3′ end and nucleotide positions 2–8 from the 3′ end separately. A one-sided t-test was then performed to determine if the RNAi activity for target sites in group 1 is significantly higher than that for target sites in group 2. For the lamin A data set, the P-value for the first five bases from the 3′ end of the target site is 0.3163, and the P-value for nucleotide positions 2–8 is 0.0973; for the PTEN and CD54 data set, the P-value for the first five bases is 0.1458, and the P-value for nucleotide positions 2–8 is 0.2024. Thus, we did not find statistical support for the hypothesis. This suggests that nucleation of siRNA–target hybridization can occur anywhere within the target site, not necessarily the 3′ end of the target site. The same conclusion was reached for microRNA–target hybridization in a recent study (Long et al. 2007).

Our study seeks to improve the potency of gene silencing by RNAi. Based on our findings, we recommend selecting

siRNAs or shRNAs with $\Delta G_{\text{disruption}} > -10$ kcal/mol and *DSSE* > 0.0 kcal/mol, in addition to applying the filters of balanced GC content and the absence of nucleotide repeat tracts. The issue of the specificity of RNAi is also important (Jackson et al. 2003; Semizarov et al. 2003; Pei and Tuschl 2006), particularly for high-throughput RNAi screening. Recent studies suggest that some off-target effects are associated with "seed" matches in the 3′ UTRs (Birmingham et al. 2006; Jackson et al. 2006). These findings and the findings in our study can be useful for improving the design of RNAi experiments, by addressing both the issue of potency and the issue of specificity.

MATERIALS AND METHODS

Prediction of mRNA secondary structure

An mRNA is likely to exist as a population of structures (Christoffersen et al. 1994). This view has been supported by the experimental elucidation of multiple equilibrium conformations (Altuvia et al. 1989; Betts and Spremulli 1994). Thus, the use of a single structure, e.g., the minimum free energy (MFE) structure, is not well suited to structure prediction for mRNAs. An alternative ensemble-based method has been developed (Ding and Lawrence 2003). In this approach, a statistically representative sample from the Boltzmann-weighted ensemble of probable RNA secondary structures is generated, in a manner to faithfully and reproducibly capture the statistical features of the structure ensemble of enormous size. In comparison with MFE predictions, this method has been shown to substantially improve predictions for structural RNAs (Ding et al. 2005) and to better represent the likely population of mRNA structures (Ding et al. 2006). A sample size of 1000 structures is sufficient to guarantee statistical reproducibility in sampling statistics (Ding and Lawrence 2003; Ding et al. 2006). The structure sampling method has been implemented in the Sfold software package (Ding et al. 2004) and is applied here to mRNA folding.

Parameters for analyses

For the statistical analyses of knockdown data from RNAi experiments, we consider several empirical rules in the literature. In addition, we introduce a novel measure of target accessibility. The stability of the target:guide strand duplex is also considered. The parameters included in our analysis are defined below.

DSSE: Differential stability of siRNA duplex ends

For the 5′ end of the antisense (guide) siRNA strand, 5′-antisense stability (*AntiS*, in kcal/mol) is computed by the summation of the free energies for four base-pair stacks (involving five consecutive base pairs) and the 3′ dangling base (overhang), with a penalty for a terminal A-U pair. Similarly, 5′-sense stability (*SS*, in kcal/mol) is the sum for the 5′ end of the sense siRNA strand. The differential stability of siRNA duplex ends (*DSSE*, in kcal/mol) is the difference between the 5′-antisense stability and the 5′-sense stability, i.e., $DSSE = AntiS - SS$. These calculations are based on the established RNA thermodynamic rules and parameters (Xia et al. 1998; Mathews et al. 1999). Because *DSSE* measures the difference in stability between the two siRNA duplex ends, a siRNA duplex meets the rule of asymmetry when *DSSE* > 0.0 kcal/mol.

S_D: Dharmacon score

This score was proposed by Dharmacon scientists for rational siRNA design (Reynolds et al. 2004). As the sum of eight component scores for various sequence features of siRNA duplex, the Dharmacon score ranges between −2 and 10.

$\Delta G_{disruption}$: A measure of target site accessibility

$\Delta G_{\text{disruption}}$ is the energy cost of disruption of the mRNA structure so that the binding site becomes completely single stranded (Fig. 1). Given the small size of the antisense siRNA, we adopt a local disruption model, i.e., the alteration of target structure due to siRNA binding is local rather than global. Specifically, we assume that only the binding site is involved in structural alternation (Fig. 1). Under this assumption, $\Delta G_{\text{disruption}}$ is the energy cost for breaking those target intramolecular base pairs at the binding site and is given by the energy difference between ΔG_{before}, the free energy of the original mRNA structure, and ΔG_{after}, the free energy of the new, locally altered structure, i.e., $\Delta G_{\text{disruption}} = \Delta G_{\text{before}} - \Delta G_{\text{after}}$. For 1000 structures predicted by Sfold, we calculate ΔG_{before} by the average energy of the original 1000 structures and ΔG_{after} by the average energy of all the 1000 locally altered structures. A largely single-stranded (i.e., structurally accessible) site does not require substantial structure alteration for the guide siRNA strand to bind to the target. The disruption energy $\Delta G_{\text{disruption}}$ is a quantitative measure of the structural accessibility at the target site.

ΔG_{hybrid}: Stability of hybrid formed by siRNA guide strand and target

ΔG_{hybrid} is the energy gain due to the hybridization at the binding site (Fig. 1). This parameter measures the stability of the hybrid formed by the siRNA guide strand and the nucleotides at the target site. ΔG_{hybrid} is calculated as the sum of the stacking energies for the siRNA guide:target duplex, with the penalty of an initiation energy:

$$\Delta G_{\text{hybrid}} = \Delta G_{\text{initiation}} + \sum \Delta G_{\text{stacking}},$$

where $\Delta G_{\text{initiation}}$ = 4.1 kcal/mol (Mathews et al. 1999), and the sum is over RNA/RNA stacking energies (Xia et al. 1998).

Statistical analyses

Weighted least-squares regression

To assess the contribution by each of the above parameters to RNAi efficiency, we employ weighted least-squares regression for prediction of target knockdown level by each of the parameters. For each of the siRNA data sets as described below, there exists a large variation in the standard deviations of the measured knockdown levels. In a statistical analysis, a data point with smaller standard deviation should carry more weight than one

with larger standard deviation. This consideration can be addressed by the use of weights in the least-squares regression (Weisberg 2005). In other words, the square term in the sum of squares for a data point is multiplied by a weight. When the standard deviation of the knockdown level from multiple measurements is available for every data point, 1/(standard deviation)^2 can be used as the weight. The *P*-value and R^2 of the regression analysis for a parameter are, respectively, the measures of the statistical significance of the parameter and the degree of variability in silencing activity that is attributed to the parameter.

Statistical tests for two-group comparison

The unpaired *t*-test was used for comparing data for two independent groups. The corresponding nonparametric test, the Wilcoxon rank sum test (also known as the Mann–Whitney U-test or the Wilcoxon–Mann–Whitney test), was also used to confirm the results by the *t*-test, which relies on the assumption of the normality of the data. All of the statistical analyses in this study were performed with the statistical package R (http://www.r-project.org).

Selection of siRNAs data sets

For selection of RNAi data sets from the literature for analysis, we employed two criteria: (1) at least 10 siRNAs for the same target must have been tested and (2) target sites on the same mRNA must not have substantial overlap. Due to the high costs of synthetic siRNAs, usually only a few siRNAs are tested for one target. Experimental variation between different RNAi knockdown experiments is difficult to account for in statistical analysis. Thus, we focus on data sets that have a sufficient number of siRNAs for the same target or for multiple targets tested by the same experimental system. Heavy overlap of target sites can introduce an autocorrelation bias that is difficult to assess. It is conceivable that a local region of the target could be highly susceptible to RNAi, e.g., due to high steric accessibility. Criterion 2 aims to avoid such likely region bias. We identified published siRNA data sets for lamin A (Harborth et al. 2003), PTEN, and CD54 (Vickers et al. 2003), and included 25, 36, and 39 siRNAs, respectively, in our analysis. The GenBank accession number and sequence length of the target are NM_170707 and 3181 nt for lamin A, U92436 and 3160 nt for PTEN, and J03132 and 2986 nt for CD54. For lamin A, another 19 siRNAs were tested for a short distance walk through single-base shift in the target site. In the light of criterion 2, however, these 19 siRNAs were not included here. Inclusion of these 19 siRNAs would enhance support for our computational approach, because these siRNAs are highly functional, and their target region is highly accessible by our prediction. For lamin A, siRNA activity was measured at the protein level using Western blot. For PTEN and CD54, the siRNA activity was measured at the mRNA level using RT-PCR. For these data sets, every siRNA was tested at least twice, in triplicate, with standard error available for measured activity.

Description of shRNA data set

We have also analyzed a data set of shRNA activities obtained from the analysis of a library of shRNA sequences generated from randomly fragmented cDNA of normalized (reduced-redundance) cDNA of all of the genes expressed in the MCF-7 human breast carcinoma cells. The generation and testing of the library will be described in detail elsewhere (A. Maliyekkel, Y. Shao, N. Warholic, K. Cole, Y. Ding, and I.B. Roninson, in prep.). Briefly, DNAseI-generated fragments of normalized cDNA were converted into shRNA templates by the procedure of Shirane et al. (2004), with some modifications. The shRNA templates were cloned into lentiviral vector LLCEP TU6LX (Maliyekkel et al. 2006) that expresses shRNA from RNA polymerase III promoter, which is positively regulated by tetracycline/doxycycline via the tTR-KRAB repressor. cDNA was cut by MmeI to produce 19–21-bp cDNA fragments. These fragments were then ligated with a hairpin adaptor to produce a hairpin–stem with a length of 27–29 bp. The hairpin–stems are very efficiently processed by Dicer to generate either 19- or 20-bp siRNA, with the adaptor sequences removed. The positions of the 19-nt and 20-nt Dicer cleavage sites are known precisely.

Individual sequenced shRNAs were matched with the corresponding human genes, and randomly selected shRNA sequences were transduced into MCF-7 cells expressing tTR-KRAB. shRNA activity was determined by measuring the levels of each target mRNA by real-time reverse-transcription PCR, in triplicate. Percent knockdown was calculated from the ratio of mRNA levels with and without doxycycline. The data for 101 shRNA sequences targeting 100 different genes (i.e., two shRNAs for only one gene) were used for the analysis. The list of these genes and their lengths are given in Supplemental Table 1.

Strategy of data analyses

The siRNA data sets were first analyzed to identify parameters that are important for RNAi efficiency. These parameters were then further examined with the independent shRNA data set, and their contributions to RNAi knockdown levels were quantitatively assessed. Such a statistical assessment was made possible by the relatively large size of the shRNA data set.

SUPPLEMENTAL DATA

Supplemental Tables 1 and 2 are available at http://sfold.wadsworth.org/Shao_RNA07_supp.pdf.

ACKNOWLEDGMENTS

The Computational Molecular Biology and Statistics Core at the Wadsworth Center is acknowledged for providing computing resources for this work. This work was supported in part by National Science Foundation grant DMS-0200970 and National Institutes of Health grant R01 GM068726 to Y.D. and National Institutes of Health grants R33 CA95996, R01 CA62099, and R01 AG17921 to I.B.R.

Received March 7, 2007; accepted July 10, 2007.

REFERENCES

Altuvia, S., Kornitzer, D., Teff, D., and Oppenheim, A.B. 1989. Alternative mRNA structures of the cIII gene of bacteriophage λ determine the rate of its translation initiation. *J. Mol. Biol.* **210:** 265–280.

Betts, L. and Spremulli, L.L. 1994. Analysis of the role of the Shine–Dalgarno sequence and mRNA secondary structure on the efficiency of translational initiation in the *Euglena gracilis* chloroplast atpH mRNA. *J. Biol. Chem.* **269:** 26456–26463.

Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-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.

Boese, Q., Leake, D., Reynolds, A., Read, S., Scaringe, S.A., Marshall, W.S., and Khvorova, A. 2005. Mechanistic insights aid computational short interfering RNA design. *Methods Enzymol.* **392:** 73–96.

Bohula, E.A., Salisbury, A.J., Sohail, M., Playford, M.P., Riedemann, J., Southern, E.M., and Macaulay, V.M. 2003. The efficacy of small interfering RNAs targeted to the type 1 insulin-like growth factor receptor (IGF1R) is influenced by secondary structure in the IGF1R transcript. *J. Biol. Chem.* **278:** 15991–15997.

Brummelkamp, T.R., Bernards, R., and Agami, R. 2002. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296:** 550–553.

Christoffersen, R.E., McSwiggen, J.A., and Konings, D. 1994. Application of computational technologies to ribozyme biotechnology products. *J. Mol. Struct. THEOCHEM* **311:** 273–284.

Das, A.T., Brummelkamp, T.R., Westerhout, E.M., Vink, M., Madiredjo, M., Bernards, R., and Berkhout, B. 2004. Human immunodeficiency virus type 1 escapes from RNA interference-mediated inhibition. *J. Virol.* **78:** 2601–2605.

Ding, Y. and Lawrence, C.E. 2001. Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res.* **29:** 1034–1046.

Ding, Y. and Lawrence, C.E. 2003. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.* **31:** 7280–7301. doi: 10.1093/nar/gkg938.

Ding, Y., Chan, C.Y., and Lawrence, C.E. 2004. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* **32:** W135–W141. doi: 10.1093/nar/gkh449.

Ding, Y., Chan, C.Y., and Lawrence, C.E. 2005. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **11:** 1157–1166.

Ding, Y., Chan, C.Y., and Lawrence, C.E. 2006. Clustering of RNA secondary structures with application to messenger RNAs. *J. Mol. Biol.* **359:** 554–571.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:** 494–498.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391:** 806–811.

Geiduschek, E.P. and Kassavetis, G.A. 2001. The RNA polymerase III transcription apparatus. *J. Mol. Biol.* **310:** 1–26.

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

Harborth, J., Elbashir, S.M., Vandenburgh, K., Manninga, H., Scaringe, S.A., Weber, K., and Tuschl, T. 2003. 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.* **13:** 83–105.

Hardin, C.C., Watson, T., Corregan, M., and Bailey, C. 1992. Cation-dependent transition between the quadruplex and Watson–Crick hairpin forms of d(CGCG3GCG). *Biochemistry* **31:** 833–841.

Heale, B.S., Soifer, H.S., Bowers, C., and Rossi, J.J. 2005. siRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res.* **33:** e30. doi: 10.1093/nar/gni026.

Hofacker, I.L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31:** 3429–3431.

Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E., and Prydz, H. 2002. Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res.* **30:** 1757–1766.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21:** 635–637.

Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L., and Linsley, P.S. 2006. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12:** 1179–1187.

Khvorova, A., Reynolds, A., and Jayasena, S.D. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115:** 209–216.

Kretschmer-Kazemi Far, R. and Sczakiel, G. 2003. The activity of siRNA in mammalian cells is related to structural target accessibility: A comparison with antisense oligonucleotides. *Nucleic Acids Res.* **31:** 4417–4424.

Laughlan, G., Murchie, A.I., Norman, D.G., Moore, M.H., Moody, P.C., Lilley, D.M., and Luisi, B. 1994. The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* **265:** 520–524.

Lee, N.S., Dohjima, T., Bauer, G., Li, H., Li, M.J., Ehsani, A., Salvaterra, P., and Rossi, J. 2002. Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat. Biotechnol.* **20:** 500–505.

Lewis, B.P., Burge, C.B., and Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120:** 15–20.

Long, D., Lee, R., Williams, P., Chan, C.Y., Ambros, V., and Ding, Y. 2007. Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.* **14:** 287–294.

Luo, K.Q. and Chang, D.C. 2004. The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region. *Biochem. Biophys. Res. Commun.* **318:** 303–310.

Maliyekkel, A., Davis, B.M., and Roninson, I.B. 2006. Cell cycle arrest drastically extends the duration of gene silencing after transient expression of short hairpin RNA. *Cell Cycle* **5:** 2390–2395.

Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288:** 911–940.

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci.* **101:** 7287–7292.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. 2005. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123:** 607–620.

Overhoff, M., Alken, M., Far, R.K., Lemaitre, M., Lebleu, B., Sczakiel, G., and Robbins, I. 2005. Local RNA target structure influences siRNA efficacy: A systematic global analysis. *J. Mol. Biol.* **348:** 871–881.

Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. 2002. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes & Dev.* **16:** 948–958.

Patzel, V., Rutz, S., Dietrich, I., Koberle, C., Scheffold, A., and Kaufmann, S.H. 2005. Design of siRNAs producing unstructured guide-RNAs results in improved RNA interference efficiency. *Nat. Biotechnol.* **23:** 1440–1444.

Pei, Y. and Tuschl, T. 2006. On the art of identifying effective and specific siRNAs. *Nat. Methods* **3:** 670–676.

Rand, T.A., Petersen, S., Du, F., and Wang, X. 2005. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123:** 621–629.

Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S., and Khvorova, A. 2004. Rational siRNA design for RNA interference. *Nat. Biotechnol.* **22:** 326–330.

Rose, S.D., Kim, D.H., Amarzguioui, M., Heidel, J.D., Collingwood, M.A., Davis, M.E., Rossi, J.J., and Behlke, M.A. 2005. Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res.* **33:** 4140–4156. doi: 10.1093/nar/gki732.

Schubert, S., Grunweller, A., Erdmann, V.A., and Kurreck, J. 2005. Local RNA target structure influences siRNA efficacy: Systematic analysis of intentionally designed binding regions. *J. Mol. Biol.* **348:** 883–893.

Schwarz, D.S., Hutvagner, 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.

Semizarov, D., Frost, L., Sarthy, A., Kroeger, P., Halbert, D.N., and Fesik, S.W. 2003. Specificity of short interfering RNA determined through gene expression signatures. *Proc. Natl. Acad. Sci.* **100:** 6347–6352.

Shirane, D., Sugao, K., Namiki, S., Tanabe, M., Iino, M., and Hirose, K. 2004. Enzymatic production of RNAi libraries from cDNAs. *Nat. Genet.* **36:** 190–196.

Stein, C.A. 1999. Two problems in antisense biotechnology: In vitro delivery and the design of antisense experiments. *Biochim. Biophys. Acta* **1489:** 45–52.

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 effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.* **32:** 936–948. doi: 10.1093/nar/gkh247.

Vickers, T.A., Wyatt, J.R., and Freier, S.M. 2000. Effects of RNA secondary structure on cellular antisense activity. *Nucleic Acids Res.* **28:** 1340–1347.

Vickers, T.A., Koo, S., Bennett, C.F., Crooke, S.T., Dean, N.M., and Baker, B.F. 2003. Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J. Biol. Chem.* **278:** 7108–7118.

Weisberg, S. 2005. *Applied linear regression*, 3rd ed. John Wiley and Sons, New York.

Westerhout, E.M., Ooms, M., Vink, M., Das, A.T., and Berkhout, B. 2005. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucleic Acids Res.* **33:** 796–804. doi: 10.1093/nar/gki220.

Wuchty, S., Fontana, W., Hofacker, I.L., and Schuster, P. 1999. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49:** 145–165.

Xia, T., SantaLucia Jr., J., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., and Turner, D.H. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* **37:** 14719–14735.

Yoshinari, K., Miyagishi, M., and Taira, K. 2004. Effects on RNAi of the tight structure, sequence, and position of the targeted region. *Nucleic Acids Res.* **32:** 691–699. doi: 10.1093/nar/gkh221.

Zhao, J.J. and Lemke, G. 1998. Rules for ribozymes. *Mol. Cell. Neurosci.* **11:** 92–97.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31:** 3406–3415.

Supplemental Materials

Supplemental Table 1. Sequence and target information for the 101 shRNAs in the MCF-7 cDNA library

shRNA number	Target gene GenBank accession number	Sequence length (nt)	Target position (for 19-bp Dicer cleavage)	Guide strand sequence including 2-nt 3′ overhang in lower case (5′→3′)	Target mRNA knockdown %
1	NM_002799	1012	888-906	GGGCCTCAATGCTCACCACct	-5.86
2	NM_003618	4141	418-436	ATATGCAATTTGCAGTTCTtt	-0.48
3	NM_004359	1462	354-372	AAGTAGCCGCCCTCGTAGTag	39.85
4	NM_003129	2989	2216-2234	CTTTAGCAGTTTTCTCCATtt	30.93
5	NM_014671	5160	1273-1291	AATAATTTTCGCCAACAGTtt	-21.32
6	NM_015231	5383	3549-3567	TTCCTCTTAGGGGATGCTCtt	45.44
7	NM_003600	2346	1886-1904	CTTTCCCCACAGCCAGGCTtt	47.69
8	NM_004526	3453	1085-1103	GGTTTCACCTCCTGGTTCTtt	45.56
9	NM_031946	3152	2056-2074	GCAGCTCAGGCGGCCAGTCtt	23.47
10	NM_005745	1324	914-932	GCACGTGCCAGGTGGAAGCtt	-6.44
11	NM_004707	2359	476-494	GGATGGTTCGTGTTCGCTCtt	68.07
12	NM_014330	2349	941-959	CTTCTTTCTGTTCTTTTATtt	4.12
13	NM_052844	1818	1352-1370	GCCCGTCAGTCCCAGCGCTcg	-30.24
14	NM_006513	1942	1787-1805	TGGGTCCCTATGCCCATGCtt	45.36
15	NM_004860	2895	2457-2475	AGAGAGGCTCCCCTTACCCtt	12.86
16	NM_152787	3798	2483-2501	CCTGTACTTTGGAGGTCACtt	19.08
17	NM_004336	3486	996-1014	GGGATGTCTCCACCACCTGat	-5.54
18	NM_004716	3497	798-816	ACCCTCAGGGCTATAGTTGgg	8.43
19	NM_005415	3297	449-467	ACTGGTAATCAGCGTTGCCat	35.75
20	NM_005463	3514	1147-1165	AAGCACAAATCCAAATCCTtt	44.63
21	NM_004356	1497	1026-1044	ACCGGAAACGTTATATACAcg	15.37
22	NM_000098	3090	1815-1833	TTTCATGGTGGCATCAAACtt	33.12
23	NM_001015	641	556-574	CGGGCCGATGTCCAGCCTCtt	13.03
24	NM_021212	2952	1452-1470	AATAGTCACATGTTATCATtt	14.34
25	NM_001537	1919	319-337	TAAATTATTAGCAACCTTCtt	69.4
26	NM_006640	3929	1533-1551	TGAGTGTGTCCGCCTTGGCtt	-1.36
27	NM_005733	2972	1658-1676	TCAATATCATCATCAAGGCtt	70.39
28	NM_002791	1035	248-266	GTGTGACAATTACTGCACAgt	18.59
29	NM_017811	4457	798-816	ATTTTGGTCAAGAATCTGAag	-14.38
30	NM_152441	2164	606-624	GTTAAGGCTCTTGAGGCGCtt	10.23
31	NM_015213	4965	2706-2724	ATGTAACAGGTGGGACCATtt	27.49
32	NM_006908	2355	246-264	ACCACACACTTGATGGCCTgc	48.74
33	NM_012432	4420	3208-3226	TGTCTCCCTCATCCTTGATtt	36.68
34	NM_002018	4176	3106-3124	GCTGCTGCGTCATGCGTACtt	52.12
35	NM_004939	2706	942-960	GTTTTTCATATGTGGTGGTtt	40.14

36	NM_001419	6075	2061-2079	GAAGGGATGCGAGAAATACTtt	-17.62
37	NM_004941	4201	1548-1566	ATCTCAGCTTCCCGCTGGGcc	-17.49
38	NM_001001520	2344	1485-1503	GGAGGGCTCTTTCTTCTTCtg	-4.75
39	NM_014267	1637	564-582	TTCTTCAGCAGATTCTTCTtt	59.13
40	NM_001067	5698	829-847	TGACCATTAGTGCAACAATtt	17.2
41	NM_020948	4972	1699-1717	TTTTCATCAAAATCATTTTtt	-4.48
42	NM_182931	7139	3858-3876	GCCTCTAGTATCATCCTGCat	1.14
43	NM_000122	2751	1751-1769	AATCTTGTCATTCCTCCTTtc	31.13
44	NM_021259	2561	2356-2374	TCGTTCTTGCAGATCTGATtt	51.8
45	NM_004421	2941	2793-2811	ACGTGGGGGCCAGAGAAGCtt	-1.67
46	NM_006098	1125	81-99	TCATGGCGGCGGCGAGAGCtt	-0.39
47	NM_015326	6305	2544-2562	ATAAATATCGGCTACATGAcg	29.42
48	NM_001386	4567	2573-2591	TATTAATATGGAATTAAATtt	10.66
49	NM_014891	2383	550-568	CTGTTTCCGGATGATGGCCtt	46.74
50	NM_001029835	2140	770-788	TGGGCACCCTGAGGATGATtt	2.24
51	NM_021259	2561	2223-2241	AGTAGTTGTCGCTAGTCATtt	69.2
52	NM_001001438	2658	1873-1891	AGCAGGAAGTCACAGGCCCtt	26.7
53	NM_002032	1245	739-757	GTGCTTGTCAAAGAGATATtt	51.97
54	NM_001802	2702	571-589	AGTTCCCTTGCTGTGACGTtt	-17.74
55	NM_003903	2196	1159-1177	TTTTCTCAAGTGTTGTGGCtt	44.78
56	NM_002623	1383	143-161	GCCTTCTTCATGGTGGCCGcg	-26.56
57	NM_172020	6014	4754-4772	GCTTGAAACATACAATCCTtt	44.62
58	NM_004788	6138	1310-1328	GGAGAGAGCTGGAGTAAGTtt	-0.93
59	NM_005973	2142	1766-1784	GCAATCAGTTCCAGAGCCCta	-5.81
60	NM_004504	4878	569-587	AAAAGAGATTTCAGTGGTTtg	8.35
61	NM_012099	3286	1433-1451	GGCTTCCTCCAGTGGCTCCac	-0.47
62	NM_020753	5005	1515-1533	TGGCTCTCGTGGATGTGGCtt	-21.95
63	NM_173823	5633	2966-2984	CAGCTAAGCTTTTCCTTATtt	-28.84
64	NM_002264	6887	828-846	TCACTGACAAACAGCAACCtt	40.29
65	NM_139280	2109	1221-1239	GGAACCAAGCCATCTACACtc	37.07
66	NM_012479	3747	785-803	GGCGTCGTCGAACGCGGTCtt	25.6
67	NM_015954	1561	1454-1472	AAAATGAAAGATGAGTTTTtt	0.04
68	NM_006284	762	686-704	AAGTACATTTAGGTTGGGTtt	37.22
69	NM_001943	3516	3386-3404	GCTAATTAAACTCTGGGTCag	22.67
70	NM_001122	2010	372-390	TCTGGATGATGGGCAGAGCtt	-15.94
71	NM_020198	3294	863-881	CATTCAGTAAGTCTTGTCTtt	14.25
72	NM_006135	2758	415-433	CATAAACAGTACAGAAGCCtt	-4.6
73	NM_003318	2984	148-166	TCTTCATTTTTAAACTTATtt	14.54
74	NM_017978	2149	1884-1902	AAAATAGTCTTCATGTCCTtt	-6.52
75	NM_014077	1465	1217-1235	ACTGCAGCCTTGTCTGTTTtt	25.42
76	NM_003674	1913	613-631	TGGATAATGAAGTTCCTGTtt	23.62
77	NM_032139	4501	3596-3614	GAATGCGGTGGTGAAACCTtt	30.62
78	NM_021188	2207	1921-1939	AACAGTCAGAATGAAGAACtt	15.24
79	NM_002627	2628	1891-1909	TCTCCGTCAGGTGCTCCACtt	8.72
80	NM_006471	1243	393-411	TTCTTTTGCTCGACATGGTtt	31.99

81	NM_005312	6121	3228-3246	ATAATGTCGTCGTTCCTCCgc	58.25
82	NM_004652	9018	5232-5250	TAAAAGATTCTTCACATTCtt	2.03
83	NM_001430	5186	1931-1949	GTCTTCAGGGCTATTGGGCgt	52.19
84	NM_006601	1490	1293-1311	TGACAGTGCAAATACAAATtt	26.24
85	NM_003168	1499	767-785	GAGAATGGAAGAGGGAGGCtt	-14.24
86	NM_005720	1520	277-295	GTCACAATACGGTTACTCTcg	-16.01
87	NM_007126	3198	1017-1035	TGATCTCAGGACCATTGATtt	41.46
88	NM_015906	8339	2106-2124	CAGCGATGGAAGGTTTTCTtt	14.15
89	NM_002067	1540	921-939	GCTCTCCTCCATCCGGTTCtg	-8.96
90	NM_013433	4901	864-882	GTCAATCCGCACTTCCAGAag	6.16
91	NM_024811	3524	2969-2987	ACCCAGAAGGGTCTCTAGTtt	-22.86
92	NM_004423	5062	3200-3218	CTGGACATCCTTGGGAAGCtt	-1.93
93	NM_001024	418	272-290	GAGACGATGCCATCGGCCTtt	-6.45
94	NM_178148	2057	1806-1824	CCGCTTGTGCTTTTCTCACtt	-6.59
95	NM_006342	2788	2666-2684	AGAATCAGACAGGACAGACtt	46.39
96	NM_005628	2856	2596-2614	CAACTACAGCCGCCAAAATtt	-10.15
97	NM_014400	1698	930-948	GCCTCGTGTTCTACTCCCTgt	2.6
98	NM_002806	1590	36-54	AGTCCTGAAGCGCCTTATCtc	25.42
99	NM_003025	2460	1062-1080	AGCTCCCCGTCGTTCTCGGgc	32.63
100	NM_006803	3472	2973-2991	ATCTGGTTGTGTACAGTACac	24.74
101	NM_021953	3527	630-648	CTGGTTTGGGTTTGAGGCCtt	38.67

Supplemental Table 2. Improvement in knockdown level by target accessibility for various $\Delta G_{disruption}$ thresholds, as estimated by using all of 101 shRNAs in the cDNA library

Energy threshold in kcal/mol for accessible sites (number of data points)	Energy threshold in kcal/mol for inaccessible sites (number of data points)	Improvement in average knockdown % [a]	*t*-test *p*-value [b]	Wilcoxon rank sum test *p*-value [b]
$\Delta G_{disruption} > -9$ (20)	$\Delta G_{disruption} < -20$ (18)	14.13	0.0437	0.0384
$\Delta G_{disruption} > -10$ (29)	$\Delta G_{disruption} < -19$ (23)	17.19	0.0072	0.0067
$\Delta G_{disruption} > -11$ (31)	$\Delta G_{disruption} < -18$ (30)	15.62	0.0057	0.0066
$\Delta G_{disruption} > -12$ (37)	$\Delta G_{disruption} < -17$ (41)	12.41	0.0117	0.0151
$\Delta G_{disruption} > -13$ (41)	$\Delta G_{disruption} < -16$ (46)	11.57	0.0133	0.0198

[a] Improvement in average knockdown % = average knockdown % for accessible sites – average knockdown % for inaccessible sites

[b] One-tailed test of significance for average knockdown % for accessible sites being greater than that for inaccessible sites