Published in final edited form as:

Biotechnol Bioeng. 2008 July 1; 100(4): 744–755. doi:10.1002/bit.21798.

Impact of target mRNA structure on siRNA silencing efficiency: a large-scale study

Joseph A. Gredell, Angela K. Berger, and S. Patrick Walton

Department of Chemical Engineering and Materials Science, Michigan State University, East Lansing, MI 48824-1226, USA

Abstract

The selection of active siRNAs is generally based on identifying siRNAs with certain sequence and structural properties. However, the efficiency of RNA interference has also been shown to depend on the structure of the target mRNA, primarily through studies using exogenous transcripts with well-defined secondary structures in the vicinity of the target sequence. While these studies provide a means for examining the impact of target sequence and structure independently, the predicted secondary structures for these transcripts are often not reflective of structures that form in full-length, native mRNAs where interactions can occur between relatively remote segments of the mRNAs. Here, using a combination of experimental results and analysis of a large dataset, we demonstrate that the accessibility of certain local target structures on the mRNA is an important determinant in the gene silencing ability of siRNAs. siRNAs targeting the enhanced green fluorescent protein were chosen using a minimal siRNA selection algorithm followed by classification based on the predicted minimum free energy structures of the target transcripts. Transfection into HeLa and HepG2 cells revealed that siRNAs targeting regions of the mRNA predicted to have unpaired 5'- and 3'-ends resulted in greater gene silencing than regions predicted to have other types of secondary structure. These results were confirmed by analysis of gene silencing data from previously published siRNAs, which showed that mRNA target regions unpaired at either the 5'-end or 3'-end were silenced, on average, ~10% more strongly than target regions unpaired in the center or primarily paired throughout. We found this effect to be independent of the structure of the siRNA guide strand. Taken together, these results suggest minimal requirements for nucleation of hybridization between the siRNA guide strand and mRNA and that both mRNA and guide strand structure should be considered when choosing candidate siRNAs.

Keywords

RNA interference; small interfering RNA; siRNA; target; mRNA structure; guide strand structure

Introduction

RNA interference (RNAi) is a natural phenomenon resulting in potent and primarily specific gene silencing that occurs in most eukaryotes (Fire et al., 1998; Hannon, 2002). RNAi is initiated in cells by the presence of either short interfering RNAs (siRNAs) or microRNAs (miRNAs), which are small non-coding RNA molecules ~21 nucleotides (nt) long (Elbashir et al., 2001a). The guide strand of these small RNAs is incorporated into the active RNA induced silencing complex (RISC), which then targets mRNAs possessing regions

Address correspondence to: Dr. S. Patrick Walton, Applied Biomolecular Engineering Laboratory of the Cellular and Biomolecular Laboratory, Department of Chemical Engineering and Materials Science, 3249 Engineering Building, East Lansing, MI 48824-1226, Tel. 517-432-8733, Fax. 517-432-1105, Email: spwalton@egr.msu.edu Web: <http://www.egr.msu.edu/abel/>.

complementary to the guide strand sequence. Upon hybridization to the target message, RISC prohibits its translation, either by cleavage of the target mRNA when guided by siRNAs, or non-degradative translational repression when guided by miRNAs. Utilization of exogenously delivered siRNAs to silence desired targets by RNAi has become a powerful tool for facilitating disease diagnosis and treatment and improving our general understanding of fundamental biological processes (Hannon, 2002; Mello and Conte, 2004).

Unfortunately, the proportion of siRNAs that are successful in repressing a target gene is low, not unlike what was found for antisense oligonucleotides (asODNs) (Stein, 1998). As with asODNs, early strategies for choosing siRNAs focused on the sequence of siRNAs, eliminating sequences based on GC content and stretches of greater than four consecutive identical bases (e.g., GGGG) (Elbashir et al., 2002). These guidelines made siRNA synthesis more convenient but were not based on mechanistic understanding. After identification of siRNAs that would be amenable to synthesis, genomic uniqueness would be verified by BLAST searching. The remaining candidate siRNAs would then be tested for activity in cell culture to identify the best silencers.

As data from siRNA experiments has accumulated, more elaborate selection algorithms have been developed that further discriminate the most important siRNA structural and sequence features. One important design rule is based on the relative end stabilities of the siRNA duplex, with the strand with the more weakly hybridized 5'-end incorporated preferentially into RISC (Khvorova et al., 2003; Schwarz et al., 2003; Tomari et al., 2004). Specific positional base preferences have been identified that tend to yield the desired differential stability between the two ends (Jagla et al., 2005; Reynolds et al., 2004; Ui-Tei et al., 2004). Additionally, formation of secondary structure in the siRNA guide strand can impair the ability of RISC to interact with its target mRNA (Patzel et al., 2005), analogous to what was shown with asODN (Mathews et al., 1999; Walton et al., 1999). More recently, attempts have been made to select against sequences resulting in nonspecific effects, including off-target silencing and immune responses (Fedorov et al., 2006; Jackson et al., 2003; Reynolds et al., 2006; Scacheri et al., 2004; Sledz et al., 2003).

However, current siRNA selection guidelines have not typically included possible impacts of the target mRNA structure on silencing efficiency. It has been shown that siRNAs can be equally effective when targeting inside the coding region of the mRNA or the 5'- and 3'- untranslated regions (e.g., (Yoshinari et al., 2004)). Several reports have used well-defined helices to demonstrate that local target structure has a direct impact on siRNA efficacy (Ameres et al., 2007; Bohula et al., 2003; Brown et al., 2005; Far and Sczakiel, 2003; Overhoff et al., 2005; Schubert et al., 2005; Shao et al., 2007; Vickers et al., 2003; Westerhout and Berkhout, 2007; Yoshinari et al., 2004). Gene silencing decreased when the orientation or the degree of partial base-pairing of a target construct was varied for a single siRNA (Schubert et al., 2005; Westerhout and Berkhout, 2007). This idea was extended to full-length transcripts (ICAM-1 and *survivin*) where siRNAs targeting inaccessible regions were ten-fold less active than accessible sites (Overhoff et al., 2005). Native mRNA structures likely inhibit the interaction of RISC with the target RNA, again echoing that which was found for target mRNA structure on asODN function (Vickers et al., 2000; Walton et al., 2002). One particularly intriguing indication of the importance of target mRNA structure on RNAi is illustrated by the ability of HIV-1 to overcome silencing by mutating the siRNA target sequence, doing so in a fashion that also alters its local RNA secondary structure (Westerhout et al., 2005).

In this report, we show that the influence of local mRNA target structure on the efficiency of siRNA-mediated RNAi is universal, applying to both endogenous and exogenous transcripts, across a variety of cell types. In addition, this influence can be reliably captured through prediction of the minimum free energy secondary structure of the full-length target mRNA.

The impact of mRNA target structure on silencing was also found to be independent of the predicted structure of the siRNA guide strand, arguing that multiple structural factors should be taken into account when designing siRNAs.

Materials and Methods

RNA secondary structure prediction and siRNA selection

The predicted secondary structure of the EGFP mRNA coding sequence was obtained using default settings on the mfold web server version 3.2 (<http://bioinfo.rpi.edu/applications/mfold>) or default settings on UNAFold version 3.4, an updated algorithm replacing mfold (Markham and Zuker, 2005; Zuker, 2003). siRNAs targeting the EGFP transcript were identified using siRNA Target Finder from Ambion Inc. (http://www.ambion.com/techlib/misc/siRNA_finder.html). A subset of these siRNAs was further selected according to the structure of the mRNA target as predicted in the minimum free energy (MFE) structure (Figure 1, Figure A1, and Results). Chemically synthesized siRNAs with 3'-UU overhangs were purchased from Dharmacon. siRNA sequences and GC content are available in Table I, and the predicted EGFP mRNA structures targeted by the siRNAs are shown in Figure A1. All siRNAs are referred to relative to the target region, where the 5'-end is position 1, which corresponds to position 19 of the guide siRNA strand.

Cell culture and transfection

Human cervical carcinoma (HeLa) and human hepatocellular carcinoma (HepG2) cells were grown in Dulbecco's modified Eagle medium (DMEM; Invitrogen) supplemented with 10% (v/v) fetal bovine serum (Biomed), 100 U/ml penicillin and 100 µg/ml streptomycin (Invitrogen) at 37°C in a 5% CO₂ humidified incubator. For transfection, HeLa cells were plated in 6-well plates at 200,000 cells/well and HepG2 cells were plated in 12-well plates at 400,000 cells/well 24 h before use in medium containing serum but lacking antibiotic. EGFP plasmid (4.0 µg for HeLa, 1.6 µg for HepG2, d2EGFP (EGFP variant with an ~2 hr half-life), Clontech) and 30 nM siRNA were transfected using Lipofectamine 2000 (3.5 and 6.0 µl, respectively; Invitrogen) in OptiMEM (Invitrogen) according to the manufacturer's protocol. After four hours, the transfection solution was replaced with complete culture medium.

Transfection efficiency was monitored by fluorescence microscopy 24 hrs post-transfection using fluorescently tagged siRNAs (targeting position 126 of the EGFP transcript or a non-targeting sequence (Block-iT Fluorescent Oligo; Invitrogen)). In all cases, transfection efficiency was >85%. Consistent cell viability, independent of siRNA and/or EGFP plasmid transfection, was confirmed by both microscopy and CellTiter-Blue Cell Viability Assay (Promega), performed according to the manufacturer's instructions. Efficiency of EGFP plasmid transfection was similarly monitored by microscopy. EGFP expression and cell viability were consistent throughout the experiments.

EGFP protein expression quantification

EGFP expression was quantified from live cells 24 hrs post-transfection. Culture medium was aspirated, cells were washed twice with phosphate buffered saline (PBS), and then a volume of PBS equal to the culture medium volume was added. Finally, fluorescence levels were measured using a SPECTRAMax GEMINI EM plate reader (Molecular Devices) with excitation at 480 nm, emission at 525 nm, and cutoff filter at 515 nm, as recommended by the manufacturer. Relative fluorescence units (RFU) were scaled to EGFP transfected cells with no siRNA (100%) and mock transfected cells with no siRNA or plasmid (0%). Fluorescence levels from mock transfected cells differed by less than 2% from wells containing only PBS. Experiments were repeated at least eight times ($n \geq 8$).

Data and analysis

RNA secondary structure predictions were performed on the genes targeted by siRNAs in the dataset compiled by Shabalina, et al. (2006) that contained 653 siRNAs. Prior to the analysis, we removed siRNAs targeting four non-human genes and deleted five genes that contained over 6,000 nts, as this was the maximum length that could be analyzed by mfold. Two other genes were removed due to difficulty in identifying the appropriate target mRNA sequences. The final dataset consisted of 548 siRNAs (533 from (Shabalina et al., 2006) and 15 from this work) targeting 42 different genes. As such, we expect that the pool of sequences used in our analyses is still sufficiently large and targets enough different mRNAs to be representative of the entire siRNA target landscape.

Structures targeted by siRNAs were identified from the MFE prediction of the full-length mRNA sequence, as listed by the National Center for Biotechnology Information (NCBI) on 12 April 2007, and analyzed as described in Results. The MFE secondary structure of each siRNA guide strand was also determined using default mfold settings and analyzed as described. Due to the variability in siRNA overhang composition utilized throughout the literature (i.e., 3'-UU vs. 3'-dTdT vs. matching the target site), only the effects of the 19 nucleotide core on target structure were considered. All other calculations including unpaired Student's t-test to compare two independent groups were performed using Microsoft Excel.

Results

siRNA selection and mRNA target structure determination

We sought to establish a relationship between the gene silencing activity of siRNAs and the structure of the target mRNA. siRNAs were designed using Ambion's siRNA Target Finder such that siRNA structures were only limited to having a 19 nt duplex with 3'-UU overhangs. No other sequence restrictions were selected. Of the 35 siRNAs returned by the algorithm, 15 were chosen for synthesis based on their GC content and the predicted structure of the target region of the mRNA (Table I and Figure A1). For our experiments, an EGFP reporter gene exhibiting an ~2 hr half-life was selected. This particular EGFP has been used frequently for RNAi assessment, with several effective siRNAs in existence (Kim and Rossi, 2003).

In the MFE structure, 506 out of 846 (59.8%) of the nucleotides were predicted to exist in intramolecular base-pairs. We selected our 15 designed siRNAs to sample a number of uniquely structured regions. We considered the scenario where the loop portion of the stem-loop was at the 5'-end, the center, or at the 3'-end of the siRNA target site within the mRNA (Figure 1). These regions were chosen for our classification as each has been shown to contribute uniquely to the binding and activity of RISC (Haley and Zamore, 2004), and the structures of the ends of guide strands are also known to impact silencing efficiency (Patzel et al., 2005). For central loops, we expanded the region of availability to encompass 4 nt to either side of the cleavage site because hybridization stability in this region is known to impact silencing efficiency dramatically (Elbashir et al., 2001b; Martinez and Tuschl, 2004). A loop length of 4 consecutive unpaired nucleotides was chosen based on what has been shown to be required for nucleation in bimolecular hybridization (e.g., (Hargittai et al., 2004)). For each of these three target structure types, we selected siRNAs having low (<50%), medium (50% to 60%), and high (>60%) GC content. No siRNAs with very low GC content (<30%) were returned by the algorithm, but this fact was ignored as they have been shown to be less active (Reynolds et al., 2004). Search results were similarly limited for very high GC content (>70%) as only two sequences (positions 159 and 558) were identified. A lack of siRNAs with these GC contents is attributed to the overall GC content of the mRNA transcript and not necessarily to the search algorithm. To maximize the unique mRNA structures targeted, we also synthesized an siRNA

predicted to target a fully looped region (pos. 597). Coincidentally, the siRNA at position 159 targeted a fully stemmed region.

Silencing activity of siRNAs targeting the EGFP mRNA

The activity of each siRNA was assessed in HeLa and HepG2 cells by cotransfection with the reporter EGFP plasmid. These cell lines were chosen because of their 1) popularity of use and 2) difference in transfection efficiency. No significant toxicity due to transfection was observed (data not shown). EGFP protein expression was quantified directly from live cells (Figure 2). Of the 15 siRNAs selected, 10 were effective in silencing the EGFP >50% and 8 siRNAs were highly effective, resulting in silencing >75%. These highly-effective siRNAs appeared to favor 5'- and 3'-target structures, or structures predicted to be fully looped (i.e., single-stranded). Conversely, siRNAs targeting central loops, fully hybridized regions, or regions with very high GC content showed considerably lower silencing activity on average (Figure 2, siRNAs 159 and 558).

Consideration of larger siRNA dataset – data distribution

To support our observation that the structure of the siRNA target region plays a role in gene silencing activity, we performed our target structure analysis on the siRNA silencing database from Shabalina, et al. (2006). We analyzed the distribution of data within this dataset and found that the distribution of siRNAs was heavily skewed to siRNAs with high activity (low gene expression level) (Figure 3). Further investigation revealed that this was partially a result of incorporating siRNAs from Reynolds, et al. (2004), where 180 siRNAs targeting every other position of the cyclophilin B and firefly luciferase were used. Removal of these siRNAs from the dataset rendered the data relatively uniformly distributed, as would be expected for a well-sampled dataset. Our subsequent analyses were performed with and without the Reynolds data so as to characterize any uniqueness associated with this particular set of siRNAs.

Distribution of base pairs within the mRNA target and the effect on silencing activity

MFE mRNA secondary structures were predicted for the complete sequence of each gene, as published by NCBI, in the dataset using UNAFold v. 3.4, with default settings (Markham and Zuker, 2005; Zuker, 2003). These structures were used to determine the influence of the total number of base-pairs in the mRNA region targeted by each siRNA. We observed that siRNAs targeted structures ranging from those completely unpaired (zero base-pairs in the mRNA target region) to those fully paired (19 base-pairs in the target; Figure 4A), with the majority of target regions containing between 10 and 16 base-pairs. The average silencing efficiency tended to decrease as the total number of base-pairs increased up to 16, beyond which the number of siRNAs in each group is too limited to assess any significant trend (Figure 4B). The trend also appears consistent when comparing the two datasets, with only a shift in the absolute magnitude of the expected gene expression level.

mRNA target loop size and location effects

Since the total number of base-pairs within an mRNA target appeared to influence the degree of silencing, we hypothesized that the number and location of unpaired nucleotides would be equally influential in gene silencing. Using the large dataset, we determined the location of unpaired loops in the mRNA target sites. We considered scenarios where loops contained 1, 2, 3, 4, or 5 consecutive unpaired nucleotides and called this the “window size”. The window was walked along the target site (5'-end of mRNA target is position 1), and the average silencing activity for those siRNAs that were unpaired in that window was calculated (Figure 5). Even at a window size of 1, the silencing activity tended to be >5% better for sequences predicted to be unpaired at the 5'- and 3'-ends relative to those predicted to be available in the center, supporting the results we obtained in our experimental system (Figure 2). This trend

became more pronounced as the window size increased from 1 to 4 where the difference in silencing activity from ends to center was 8–10%. The change in trend at this window size supports the classifications we chose for our EGFP experiments, and, again, is in accord with the literature (Hargittai et al., 2004). Little change was observed increasing from a window size of 4 to 5 (Figure 5). We therefore defined our loop size for structure classification as four consecutive unpaired nucleotides ($W=4$) and refined our target structure matrix (Table II). For targets that contained both a central loop and either a 5' - or 3' -loop, we classified these as one of the latter.

The distribution of the data was similar in each group, with noticeably less skew when ignoring the Reynolds data (data not shown). The 5' - and 3' -loop structures showed improved silencing activity, while siRNAs targeting central loops demonstrated reduced functionality. Furthermore, stem structures gave ~12% lower silencing efficiency as compared to 3' -loops. The functionality of siRNAs within these groups (Table II) was then compared to each remaining group using an unpaired Student's t-test. The silencing activity of siRNAs in either the 5' - or 3' -groups was significantly higher than for siRNAs in the center, stem, and other groups (Table A1); the p-value for the combined 5' - and 3' -groups versus all others was 0.0038 and 0.0005 for the Complete Data and No Reynolds Data, respectively, providing statistical support to our group classifications.

siRNA guide strand structure effects

To demonstrate that siRNA guide strand structures were not confounding our analyses of mRNA local target structures, we determined the MFE secondary structure for each siRNA guide strand and performed the same analysis as described above for the mRNA. Guide structures of the 5' - and 3' -loop types (defined using $W=4$, as with our mRNA target structures) were twice as common as central loops, with < 2% forming stems; no guide strands formed "other" structures (Table II). Consistent with a previous report (Patzel et al., 2005), guide structures with 5' -loops resulted in improved gene silencing relative to central and 3' -loops (Table II), as confirmed by Student's t-test (Table A2).

We next considered a pair-wise comparison of each mRNA structure group versus each guide structure group (Table III). siRNAs with any combination of 5' - or 3' -loops in the mRNA or guide structures tended to give the best silencing, with >5% difference in activity compared to combinations including central, stem, or other groups. siRNAs with central loops in the guide strand or targeting stems in the mRNA had lower silencing activities, as was expected. Taken together, these results suggest that inclusion of basic mRNA secondary structural information, in concert with siRNA guide strand structure details, can improve the likelihood of identifying active siRNAs.

Discussion

Accurately accounting for significant controlling parameters in siRNA design continues to pose a considerable problem for RNAi applications. Ideally, all mRNAs targeted for cleavage by RISC would be entirely single-stranded and free from ribosomes or other bound molecules, allowing for uninhibited hybridization of any siRNA guide strand to its complementary target. Even ignoring the presence of ribosomes, it has been shown that the accessibility of the target mRNA through complex secondary structures can influence RNAi (Ameres et al., 2007; Bohula et al., 2003; Brown et al., 2005; Long et al., 2007; Overhoff et al., 2005; Schubert et al., 2005; Shao et al., 2007; Vickers et al., 2003; Westerhout and Berkhout, 2007; Yoshinari et al., 2004). In this work, our goal was to facilitate incorporation of target mRNA structure into siRNA design algorithms by investigating which mRNA structure types were most amenable to silencing. We experimentally showed that specific regions of the target mRNA were more susceptible to RNAi silencing than others. Furthermore, these results were in

agreement with computational analyses of data available in the literature. The majority of our results (41 of 42 genes) are based on the silencing of endogenous mRNAs, thus avoiding any artifacts that may have arisen by silencing only exogenous or engineered constructs. Despite the variety of systems and readouts used to obtain it, the data clearly showed that siRNAs targeting regions predicted to be unpaired at either the 5'- or 3'-end silenced, on average, 8% better than siRNAs targeting regions unpaired in the center or without any unpaired windows. Though this net effect could be argued to be small, our observation is consistent with a recent report showing an average difference in knockdown of ~14% for 101 shRNAs when accounting for an energy parameter in target accessibility calculations (Shao et al., 2007). It is noteworthy that in this same report a difference of ~35% was found when only considering shRNAs that have a favorable duplex asymmetry. While we did not take into account duplex asymmetry in our analyses, these results strongly suggest that our differences in average silencing would be improved by removing siRNAs that are not asymmetric. Regardless, provided that it comes at low computational and time costs, the additional information gained by using mRNA target structure predictions is warranted for incorporation into siRNA design algorithms to enhance the likelihood of selection of active siRNAs.

As the number of experimentally tested siRNAs increases, it becomes vitally important to note the conditions for both design and application of effective *and ineffective* sequences when reporting results. By doing so, one could greatly enhance the set of siRNAs available for bioinformatics studies and development of design algorithms (Matveeva et al., 2007). Algorithms would be similarly assisted by using silencing results from siRNAs randomly selected for their target, since more elaborate selection methods may bias sequences towards parameters already shown to be relevant, such as GC content and other particular positional base preferences (Reynolds et al., 2004). We therefore designed our siRNAs targeting EGFP using an algorithm developed to take advantage of a simple method for enzymatic siRNA synthesis (http://www.ambion.com/techlib/misc/siRNA_finder.html). This algorithm only required that the target site lie immediately downstream of 'AA' dinucleotides, and so all of our experimental sequences contained UU overhangs (making the entire 21 nt of the siRNA complementary to the target mRNA). This is certainly a potential source of bias and may be one reason why our initial experiments yielded such a high proportion of active siRNAs (8 out of 15 silenced the EGFP more than 75%). It is noteworthy that the siRNA targeting a fully unpaired region resulted in >75% silencing (Figure 2; target 597) and that the siRNAs targeting fully paired regions (targets 159 and 396) yielded only ~20–40% reduction, an expected trend for the most and least ideal target types, respectively. Moreover, the total number of base-pairs within the mRNA target site, another metric for target site stability, showed a normal distribution around 11 base-pairs, with or without the data from Reynolds, et al. (Figure 4A). This suggests that most siRNA target sites have at least half of the target region sequestered in native structure. As with the loop and stem targeting siRNAs, siRNAs targeting regions with fewer total base-pairs tended to reduce gene expression more effectively (Figure 4B).

When analyzing data for parametric information, it is important that the dataset is as unbiased as possible and contains enough data points to draw the appropriate conclusions. Beginning with the entire database of Shabalina, et al., we pared it down to only those siRNAs targeting human genes shorter than 6,000 nts (see Results). This dramatically increased our structural prediction speed while only reducing our dataset by 18%, from 653 to 533 siRNAs. After examining the activity profile of the remaining sequences, we observed that the data were skewed to siRNAs that reduced gene expression levels below 35% (Figure 3), due to the Reynolds data. Despite our concerns about the unequal weighting in the data, inclusion of these data in our analyses did not significantly alter the results for 5'-, central, or 3'-loops (Figure 4 and Figure 5, Table II and Table A1). Some effect, though, was seen on the sequences in the "other" grouping. "Other" refers to those sequences that have loops but not in regions to be classified in the 5'-loop, 3'-loop, or central class. The average gene expression remaining for

this group increased from 42% to 51% upon removal of the Reynolds data. Similarly, the analyses implemented in Table A1 (and data not shown) revealed no statistically significant effect from this group. These results indicate that structures that fall into the “other” class are not as critical for defining silencing efficiency, though this issue is still up for debate (Katoh and Suzuki, 2007; Schwarz et al., 2006).

Predicted secondary structures for RNA have been determined using algorithms such as mfold (Mathews et al., 1999; Zuker, 2003) or the Vienna RNA Package (Hofacker, 2003) that utilize nearest-neighbor energies to obtain an MFE structure and a set of suboptimal structures. However, others have suggested that this unnecessarily assumes that the MFE structure is the most prevalent structure of the mRNA in the cell (Ding et al., 2004). Instead, an ensemble of foldings encompassing a statistically significant sampling (~1000 structures) is considered more appropriate (Ding et al., 2004). This lends itself to a stochastic approach where the probabilities of interactions can be assessed. Both methods have proven useful in attempting to account for mRNA target structure in RNAi (Ameres et al., 2007; Heale et al., 2005; Long et al., 2007; Overhoff et al., 2005; Patzel et al., 2005; Schubert et al., 2005; Shao et al., 2007; Westerhout and Berkhout, 2007). While we did not consider ensembles in this work, for design purposes, our results (Figure 2 and Table II and Table III) and those of several others (Far and Sczakiel, 2003; Luo and Chang, 2004; Patzel et al., 2005; Schubert et al., 2005; Westerhout and Berkhout, 2007) show that mRNA MFE structures determined by mfold alone provide information that is useful in refining the pool of candidates for selection of active siRNAs.

It is well-established that RNA secondary structure predictions worsen as the length of the sequence to be folded increases. It is therefore common to predict folded structures for sequences of a given length, e.g., < 700 nt (Doshi et al., 2004; Mathews et al., 2004). For predictions of local structures, the folded sequence is generally specified to encompass a given length (~100 nt) of sequence upstream and downstream of the target (Heale et al., 2005; Long et al., 2007). This approach, though possibly valuable for gleaning additional information about the predicted target region structures, would not be of general interest to individuals designing siRNAs due to the computational expense, the number of folds to analyze, and the lack of a rigorous definition as to the appropriate sequence length to fold. As such, we did not investigate the use of a segmented approach for prediction of our target structures.

The extent to which certain regions of the siRNA guide strands, and consequently the mRNA regions targeted by those strands, influence RNAi has been somewhat controversial (Haley and Zamore, 2004; Long et al., 2007; Patzel et al., 2005; Shao et al., 2007; Westerhout and Berkhout, 2007). One report found that the 5'-end of the siRNA guide strand contributes more to the binding of the target mRNA than the center and 3'-end, which affect the helical geometry of the siRNA/mRNA hybrid (Haley and Zamore, 2004). This conclusion is consistent with our analysis showing a preference for unpaired 5'-ends in the guide strand over both central and 3'-loops (Table III and Table A2). Our guide strand results support those of Patzel and colleagues who found that for guide strands forming stem-loop structures the 5'-end was more influential than the 3'-end, with either being more critical than an unpaired central region (Koberle et al., 2006; Patzel et al., 2005). These results imply that 3'- and 5'-ends of the mRNA regions targeted by the guide strands are also important. This is further supported by a recent report that describes accessibility of the 3'-end of the target site as an important determinant of RISC activity (Ameres et al., 2007). It has also been shown using a reporter construct that a free 3'-end in the target region improved RNAi silencing (Westerhout and Berkhout, 2007), though only limited structures were considered which perhaps masked the contributions from the 5'-end. Destabilizing native mRNA structure both upstream and downstream of the siRNA target sequence can enhance the association rate of RISC as well (Ameres et al., 2007; Brown et al., 2005), which would suggest an improved likelihood of silencing. Another recent study found that improved silencing occurs at target sites bordered on both the 5'- and 3'-ends by

regions of high AU content (Nielsen et al., 2007), as these would presumably have relatively weaker native mRNA structure. However, two other reports reached no such conclusions, instead noting that siRNA/mRNA nucleation can occur anywhere along the target site (Long et al., 2007; Shao et al., 2007). These papers used a threshold energy term in the analysis of miRNA (Long et al., 2007) and siRNA (Shao et al., 2007) binding to structured targets. After nucleation, a second energy parameter regarding helix elongation was required to fully describe the miRNA behavior. This report (Long et al., 2007) also showed that nucleation of four consecutive unpaired nucleotides gave better correlations to gene inhibition. This latter result is in agreement with our observations of siRNA initiated gene silencing (Figure 5). Unfortunately, our attempts to describe the energetics of the siRNA/mRNA interaction using the RNAup algorithm from the Vienna RNA Package (Muckstein et al., 2006) or the algorithm of Heale, et al. (2005) were unsuccessful (Figure A2 and Figure A3). The fact that an effect was observed at both the 5'- and 3'-ends of the target suggest that the seed site (i.e., the 5'-end) and the relative differential stability between the 5'- and 3'-ends are not the only important parameters for si/miRNA functionality. Most likely, it is easier for mRNA targets with four consecutive unpaired nucleotides at one or both ends to form a stable hybrid with a guide strand also lacking structure at one or both ends. Once initiated, the dsRNA helix can elongate, possibly facilitated by the helicase component associated with active RISC (Robb and Rana, 2007).

Conclusions

Using a broad analysis of siRNA-mediated RNAi from the literature, we have shown that siRNAs targeting mRNAs with predicted regions of four consecutive unpaired nucleotides at either the 5'- or 3'-ends of the target site are more potent for inducing RNAi-based gene silencing than when the center is unpaired or when the target site has no unpaired regions. Additionally, these observations are consistent with the mechanistic understanding of guide strand loading in the RNAi pathway. As these results were based on predicted MFE structures, they demonstrate the utility of mRNA secondary structure predictions in enhancing the likelihood of identifying active siRNAs.

Appendix

The predicted EGFP mRNA structures targeted by the 15 siRNAs designed in this work are shown in Figure A1.

In order to assess the statistical significance of our structure groupings, we performed an unpaired Student's t-test on independent groups of siRNAs (Table A1 and Table A2). The silencing activities for siRNAs in the group in the left-hand column were compared to the activities for siRNAs in the group along the top row and the resulting p-value is shown. Note that the p-value for the 5' versus 3' comparison did not take into account the fact that ~5% of the siRNAs in the mRNA target structure analysis and ~30% of the siRNAs in the guide strand structure analysis can be classified as both 5'- and 3'-loops; therefore the data are not completely independent and thus those p-values are larger than if the groupings were independent. However, in the comparison of siRNA activities for groups in the left-hand column versus activities from all other siRNAs not in that group ("All" group in the top row), only independent siRNAs were considered.

We also compared siRNAs in either the 5' or 3' groups together versus "All". For the analysis of mRNA target structures (Table A1) this p-value was lower than either the 5' versus "All" or the 3' versus "All" p-values because the 5' and 3' groups were not significantly different from each other, yet they were included in the "All" grouping. However, for the analysis of guide strand structures (Table A2) the 5' group, and not the 3' group, was significantly different

than the “All” group. Therefore the significance in the comparison of 5’ and 3’ versus “All” is due to the inclusion of the 5’ group results and not the 3’ group.

We attempted to describe the binding energies of the siRNAs to their mRNA targets using the RNAup algorithm (Muckstein et al., 2006) to determine if there was any correlation with their silencing activity. No correlation was identified (Figure A2). We also performed an analysis using the algorithm from Heale et al., (2005) that takes into account a similar energy of interaction, and noted no correlation (Figure A3).

Acknowledgements

The authors thank Hemant Kini, Sophie Carrell, and other members of the Cellular and Biomolecular Laboratory for helpful discussions. Financial support for this work was provided in part by Michigan State University, the National Science Foundation (#0425821), and the National Institutes of Health (#CA126136, #GM079688).

Nomenclature

bp, base-pair; dsRNA, double-stranded RNA; EGFP, enhanced green fluorescent protein; MFE, minimum free energy; nt, nucleotide; RNAi, RNA interference; siRNA, small interfering RNA.

References

- Ameres SL, Martinez J, Schroeder R. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* 2007;130(1):101–112. [PubMed: 17632058]
- Bohula EA, Salisbury AJ, Sohail M, Playford MP, Riedemann J, Southern EM, Macaulay VM. 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* 2003;278(18):15991–15997. [PubMed: 12604614]
- Brown KM, Chu CY, Rana TM. Target accessibility dictates the potency of human RISC. *Nat Struct Mol Biol* 2005;12(5):469–470. [PubMed: 15852021]
- Ding Y, Chan CY, Lawrence CE. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 2004;32(Web Server issue):W135–W141. [PubMed: 15215366]
- Doshi K, Cannone J, Cobaugh C, Gutell R. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* 2004;5(1):105. [PubMed: 15296519]
- Elbashir SM, Harborth J, Weber K, Tuschl T. Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* 2002;26(2):199–213. [PubMed: 12054897]
- Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* 2001a;15(2):188–200. [PubMed: 11157775]
- Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *Embo J* 2001b;20(23):6877–6888. [PubMed: 11726523]
- Far R, Sczakiel G. The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Res* 2003;31(15):4417–4424. [PubMed: 12888501]
- Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K, Leake D, Marshall WS, Khvorova A. Off-target effects by siRNA can induce toxic phenotype. *RNA* 2006;12(7):1188–1196. [PubMed: 16682561]
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998;391:806–811. [PubMed: 9486653]
- Haley B, Zamore PD. Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol* 2004;11(7):599–606. [PubMed: 15170178]
- Hannon GJ. RNA interference. *Nature* 2002;418:244–251. [PubMed: 12110901]

- Hargittai MR, Gorelick RJ, Rouzina I, Musier-Forsyth K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J Mol Biol* 2004;337(4):951–968. [PubMed: 15033363]
- Heale BS, Soifer HS, Bowers C, Rossi JJ. siRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res* 2005;33(3):e30. [PubMed: 15722476]
- Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res* 2003;31(13):3429–3431. [PubMed: 12824340]
- Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 2003;21(6):635–637. [PubMed: 12754523]
- Jagla B, Aulner N, Kelly PD, Song DA, Volchuk A, Zatorski A, Shum D, Mayer T, De Angelis DA, Ouerfelli O, et al. Sequence characteristics of functional siRNAs. *RNA* 2005;11(6):864–872. [PubMed: 15923373]
- Katoh T, Suzuki T. Specific residues at every third position of siRNA shape its efficient RNAi activity. *Nucleic Acids Res* 2007;35(4):e27. [PubMed: 17259216]
- Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. *Cell* 2003;115(2):209–216. [PubMed: 14567918]
- Kim DH, Rossi JJ. Coupling of RNAi-mediated target downregulation with gene replacement. *Antisense Nucleic Acid Drug Dev* 2003;13(3):151–155. [PubMed: 12954115]
- Koberle C, Kaufmann SH, Patzel V. Selecting effective siRNAs based on guide RNA structure. *Nat Protoc* 2006;1(4):1832–1839. [PubMed: 17487166]
- Long D, Lee R, Williams P, Chan CY, Ambros V, Ding Y. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol* 2007;14(4):287. [PubMed: 17401373]
- Luo KQ, Chang DC. The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region. *Biochem Biophys Res Commun* 2004;318(1):303–310. [PubMed: 15110788]
- Markham NR, Zuker M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* 2005;33(Web Server issue):W577–W581. [PubMed: 15980540]
- Martinez J, Tuschl T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev* 2004;18(9):975–980. [PubMed: 15105377]
- Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* 2004;101(19):7287–7292. [PubMed: 15123812]
- Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 1999;288(5):911. [PubMed: 10329189]
- Matveeva O, Nechipurenko Y, Rossi L, Moore B, Saetrom P, Ogurtsov AY, Atkins JF, Shabalina SA. Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. *Nucleic Acids Res* 2007;35(8):e63. [PubMed: 17426130]
- Mello CC, Conte D Jr. Revealing the world of RNA interference. *Nature* 2004;431(7006):338–342. [PubMed: 15372040]
- Muckstein U, Tafer H, Hackermuller J, Bernhart SH, Stadler PF, Hofacker IL. Thermodynamics of RNA-RNA binding. *Bioinformatics* 2006;22(10):1177–1182. [PubMed: 16446276]
- Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* 2007;13(11):1894–1910. [PubMed: 17872505]
- Overhoff M, Alken M, Far RK, Lemaitre M, Lebleu B, Sczakiel G, Robbins I. Local RNA target structure influences siRNA efficacy: a systematic global analysis. *J Mol Biol* 2005;348(4):871–881. [PubMed: 15843019]
- Patzel V, Rutz S, Dietrich I, Koberle C, Scheffold A, Kaufmann SH. Design of siRNAs producing unstructured guide-RNAs results in improved RNA interference efficiency. *Nat Biotechnol* 2005;23(11):1440–1444. [PubMed: 16258545]

- Reynolds A, Anderson EM, Vermeulen A, Fedorov Y, Robinson K, Leake D, Karpilow J, Marshall WS, Khvorova A. Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA* 2006;12(6):988–993. [PubMed: 16611941]
- Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. *Nat Biotechnol* 2004;22(3):326–330. [PubMed: 14758366]
- Robb GB, Rana TM. RNA Helicase A Interacts with RISC in Human Cells and Functions in RISC Loading. *Mol Cell* 2007;26(4):523–537. [PubMed: 17531811]
- Scacheri PC, Rozenblatt-Rosen O, Caplen NJ, Wolfsberg TG, Umayam L, Lee JC, Hughes CM, Shanmugam KS, Bhattacharjee A, Meyerson M, et al. Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc Natl Acad Sci U S A* 2004;101(7):1892–1897. [PubMed: 14769924]
- Schubert S, Grunweller A, Erdmann VA, Kurreck J. Local RNA Target Structure Influences siRNA Efficacy: Systematic Analysis of Intentionally Designed Binding Regions. *J Mol Biol* 2005;348(4):883–893. [PubMed: 15843020]
- Schwarz DS, Ding H, Kennington L, Moore JT, Schelter J, Burchard J, Linsley PS, Aronin N, Xu Z, Zamore PD. Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLoS Genet* 2006;2(9):e140. [PubMed: 16965178]
- Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the Assembly of the RNAi Enzyme Complex. *Cell* 2003;115:199–208. [PubMed: 14567917]
- Shabalina SA, Spiridonov AN, Ogurtsov AY. Computational models with thermodynamic and composition features improve siRNA design. *BMC Bioinformatics* 2006;7:65. [PubMed: 16472402]
- Shao Y, Chan CY, Maliyekkel A, Lawrence CE, Roninson IB, Ding Y. Effect of target secondary structure on RNAi efficiency. *RNA* 2007;13(10):1–10. [PubMed: 17123956]
- Sledz CA, Holko M, de Veer MJ, Silverman RH, Williams BR. Activation of the interferon system by short-interfering RNAs. *Nat Cell Biol* 2003;5(9):834–839. [PubMed: 12942087]
- Stein CA. How to design an antisense oligodeoxynucleotide experiment: a consensus approach. *Antisense Nucleic Acid Drug Dev* 1998;8(2):129–132. [PubMed: 9593051]
- Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD. A protein sensor for siRNA asymmetry. *Science* 2004;306(5700):1377–1380. [PubMed: 15550672]
- Ui-Tei K, Naito Y, Takahashi F, Haraguchi T, Ohki-Hamazaki H, Juni A, Ueda R, Saigo K. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res* 2004;32(3):936–948. [PubMed: 14769950]
- Vickers TA, Koo S, Bennett CF, Crooke ST, Dean NM, Baker BF. Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J Biol Chem* 2003;278(9):7108–7118. [PubMed: 12500975]
- Vickers TA, Wyatt JR, Freier SM. Effects of RNA secondary structure on cellular antisense activity. *Nucleic Acids Res* 2000;28(6):1340–1347. [PubMed: 10684928]
- Walton SP, Stephanopoulos GN, Yarmush ML, Roth CM. Prediction of antisense oligonucleotide binding affinity to a structured RNA target. *Biotechnol Bioeng* 1999;65(1):1–9. [PubMed: 10440665]
- Walton SP, Stephanopoulos GN, Yarmush ML, Roth CM. Thermodynamic and kinetic characterization of antisense oligodeoxynucleotide binding to a structured mRNA. *Biophys J* 2002;82(1 Pt 1):366–377. [PubMed: 11751323]
- Westerhout EM, Berkhout B. A systematic analysis of the effect of target RNA structure on RNA interference. *Nucleic Acids Res* 2007;35(13):4322–4330. [PubMed: 17576691]
- Westerhout EM, Ooms M, Vink M, Das AT, Berkhout B. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucleic Acids Res* 2005;33(2):796–804. [PubMed: 15687388]
- Yoshinari K, Miyagishi M, Taira K. Effects on RNAi of the tight structure, sequence and position of the targeted region. *Nucleic Acids Res* 2004;32(2):691–699. [PubMed: 14762201]
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31(13):3406–3415. [PubMed: 12824337]

Figure 1. Local mRNA target structure groupings

Four local mRNA structures were considered when grouping siRNA target regions.

Classification in any group required an mRNA target region (black) that contained four consecutive unpaired nucleotides (A) at the 5'-end, '5'-loop'; (B) at the 3'-end, '3'-loop'; (C) centered around the siRNA cleavage site, 'central-loop'; or (D) nowhere along the target, 'stem'.

Figure 2. mRNA target structure dependent gene silencing

siRNAs targeting the EGFP mRNA (see Materials and Methods) were cotransfected into HeLa and HepG2 cells and EGFP fluorescence levels were measured after 24 hrs. siRNAs are grouped according to the predicted structure (Figure 1; 5' – 5'-loop; C – central loop; 3' – 3'-loop; S – stem; L – completely unpaired, 'loop') of their target sites and by GC content ('-' – <50%; 'o' – 50–60%; '+' – 60–70%; '++' – >70%). Values represent mean $\pm$ standard deviation for at least eight independent experiments ($n \geq 8$). All siRNA treatments were significantly different from Control (Student's t-test; $p < 0.01$) except those denoted by * for HepG2 cells. Comparisons were also made for each siRNA relative to the sequence giving the smallest change in EGFP expression level for the corresponding cell line. HeLa: all siRNA treatments are significantly different from siRNA 159 except those denoted by # ($p < 0.01$). HepG2: all siRNA treatments are significantly different from siRNA 558 except those denoted by \$ ($p < 0.01$).

Figure 3. Distribution of siRNAs

The frequency of siRNAs giving an average gene expression remaining after siRNA treatment (grouped into bins of 5%; e.g., 0–5, 5–10, etc.) is shown for the siRNAs from the complete dataset and after the data from Reynolds, et al. was removed.

Figure 4. Influence of the number of base-pairs within the siRNA target site

Shown are (A) the frequency of siRNAs targeting mRNA sites with the predicted number of base-pairings and (B) the effect of the number of base-pairs in the target site on the average gene expression level remaining after siRNA treatment. For (B), only points where more than 15 siRNAs were tested are shown. The linear regressions show a positive correlation. As in Figure 3, data are plotted with and without the data from Reynolds, et al.

Figure 5. Effect of window size on gene expression levels

The profiles of the average gene expression level are shown for siRNAs (complete dataset) that were completely unpaired in the reading window ($W=1, 2, 3, 4$, or 5 consecutive nucleotides). Because the window size changes, the number of possible windows changes concomitantly; i.e., there are 19 windows of size 1 but only 16 windows of size 4 along a 19 nt long siRNA.

Figure A1. mRNA predicted structures for siRNAs targeting EGFP

The global mRNA structures were predicted (see Materials and Methods) and the local target site (black line, * denotes 5'-end of site) is shown for each siRNA. Target start position is indicated for each of the 15 siRNAs.

Figure A2. siRNA-mRNA interaction energies were calculated by the RNAup algorithm from the Vienna RNA Package

The linear regression (grey line) distribution coefficient (R) is 0.1190.

Figure A3. siRNA-mRNA relative accessibilities were calculated by the algorithm of Heale, et al. (2005)

The linear regression (grey line) distribution coefficient (R) is 0.0045.

Table I

Details for EGFP-targeting siRNAs experimentally validated in this study*

Start	End	Antisense	Sense	GC Content (%)
71	89	ACGCUGAACUUGUGGCCGU	ACGGCCACAAGUUCAGCGU ^I	52
81	99	CUCGCCGGACACGCUGAAC	GUUCAGCGUGUCCGGCGAG	62
126	144	GAUGAACUUCAGGGUCAGC	GCUGACCCUGAAGUUCAUC ^I	48
159	177	GGGCCAGGGCACGGGCAGC	GCUGCCCGUGCCUGGCCC	76
274	292	UGCGCUCCUGGACGUAGCC	GGCUACGUCCAGGAGCGCA	62
306	324	CUUGUAGUUGCCGUCGUCC	GGACGACGCAACUACAAG	52
318	336	CUCGGCGCGGGUCUUGUAG	CUACAAGACCCGCGCCGAG	62
396	414	CAGGAUGUUGCCGUCCUCC	GGAGGACGGCAACAUCUG	57
441	459	GAUAUAGACGUUGUGGCUG	CAGCCACAACGUCUAUAUC	43
471	489	CUUGAUGCCGUUCUUCUGC	GCAGAAGAACGGCAUCAAG	48
495	513	GUUGUGGCGGAUCUUGAAG	CUUCAAGAUCGCGCACAAAC	48
501	519	CUCGAUGUUGUGGCGGAUC	GAUCCGCCACAACAUCGAG	52
558	576	GCCGUCGCCGAUGGGGGUG	CACCCCAUCGGCGACGGC	71
597	615	CUGGGUGCUCAGGUAGUGG	CCACUACCUGAGCACCCAG	57
639	657	CAUGUGAUCGCGCUUCUCG	CGAGAAGCGCGAUCACAUG	52

* Start sites are relative to the start codon of the EGFP coding sequence. GC Content is per 21 nt.

^I siRNA sequences obtained from (Kim and Rossi, 2003).

Table II
Definition of RNA structures and siRNA activity within each group

Structure Type	Location of Unpaired Nucleotides	Structured mRNA			Structured Guide Strand		
		# of siRNAs	Average Gene Expression Level	# of siRNAs	Average Gene Expression Level	# of siRNAs	Average Gene Expression Level
		Complete Data	No Reynolds	Complete Data	No Reynolds	Complete Data	No Reynolds
5'	1–5	121	84	39.4	44.8	233	159
Central	7–13	98	60	47.2	51.7	119	70
3'	15–19	123	85	40.3	42.9	241	170
Stem	None	144	98	48.4	55.5	9	5
Other	3–9 or 11–17	88	53	41.6	50.8	N/A	N/A
						39.4	44.7
						48.1	52.8
						44.0	49.6
						52.5	68.5
						N/A	N/A

Biotechnol Bioeng. Author manuscript; available in PMC 2009 July 1.

Target structures were defined when 4 or more consecutive unpaired nucleotides were present in the specified regions. Nucleotide positions are relative to the 5'-end of the target mRNA. See Figure 1 for illustration. N/A, not applicable.

Table III
Distribution of siRNA functionality for each target structure classification

Guide Strand				
	5'	Central	3'	Stem
mRNA	5'	35.4	40.9	18.9
		0.4996	0.0876	
		43	64	4
	Central	41.6	50.7	56.0
		0.0149	0.1948	
		31	29	1
	3'	35.9	36.7	70.2
		0.3969	0.1405	
		64	50	4
	Stem	45.0	49.9	N/A
		0.0035	0.1011	
		66	71	0

Target structures were defined when 4 or more consecutive unpaired nucleotides were present in the specified regions of either the siRNA guide strand or the mRNA target. The entries in each box represent the average gene expression level (top), the p-value from a one-sided Student's t-test (middle), and the total number of siRNAs within that group (bottom) for siRNAs in the Complete Dataset.

Table A1
Statistical analysis of siRNA activity between groups for mRNA target structures.

Complete Data					
	Central	3'	Stem	Other	All
5'	0.0281	0.4101	0.0100	0.3007	0.0503
Central		0.0472	0.3812	0.1095	0.0853
3'			0.0192	0.3772	0.1039
Stem				0.0589	0.0150
Other					0.2834
5' and 3'					0.0038

No Reynolds Data					
	Central	3'	Stem	Other	All
5'					
Central	0.0867	0.3439	0.0094	0.1321	0.0822
3'		0.0446	0.2267	0.4410	0.2075
Stem			0.0034	0.0752	0.0232
Other				0.1940	0.0068
5' and 3'					0.3028
					0.0005

Table A2

Statistical analysis of siRNA activity between groups for guide strand structures.

	Complete Data			
	Central	3'	Stem	All
5'	0.0058	0.0491	0.1566	0.0029
Central		0.1255	0.3645	0.0363
3'			0.2539	0.3730
Stem				0.2372
5' and 3'				0.0227
	No Reynolds Data			
	Central	3'	Stem	All
5'	0.0331	0.0718	0.1264	0.0129
Central		0.2347	0.2168	0.1123
3'			0.1749	0.3105
Stem				0.1619
5' and 3'				0.0583