NUCLEIC ACID THERAPEUTICS
Volume 26, Number 5, 2016

DOI: 10.1089/nat.2016.0612

Terminal Duplex Stability and Nucleotide Identity Differentially Control siRNA Loading and Activity in RNA Interference

Phillip A. Angart, Rebecca J. Carlson, Kwasi Adu-Berchie, and S. Patrick Walton

Efficient short interfering RNA (siRNA)-mediated gene silencing requires selection of a sequence that is complementary to the intended target and possesses sequence and structural features that encourage favorable functional interactions with the RNA interference (RNAi) pathway proteins. In this study, we investigated how terminal sequence and structural characteristics of siRNAs contribute to siRNA strand loading and silencing activity and how these characteristics ultimately result in a functionally asymmetric duplex in cultured HeLa cells. Our results reiterate that the most important characteristic in determining siRNA activity is the 5′ terminal nucleotide identity. Our findings further suggest that siRNA loading is controlled principally by the hybridization stability of the 5′ terminus (Nucleotides: 1–2) of each siRNA strand, independent of the opposing terminus. Postloading, RNA-induced silencing complex (RISC)–specific activity was found to be improved by lower hybridization stability in the 5′ terminus (Nucleotides: 3–4) of the loaded siRNA strand and greater hybridization stability toward the 3′ terminus (Nucleotides: 17–18). Concomitantly, specific recognition of the 5′ terminal nucleotide sequence by human Argonaute 2 (Ago2) improves RISC half-life. These findings indicate that careful selection of siRNA sequences can maximize both the loading and the specific activity of the intended guide strand.

Keywords: siRNA, RNAi, functional asymmetry, Ago2, terminal nucleotide, duplex thermodynamics

Introduction

The RNA interference (RNAi) pathway is a central regulatory mechanism in mammalian cells, where endogenous small RNAs, once incorporated into a ribonucleoprotein complex, target complementary sequences and cause post-transcriptional gene silencing [1,2]. Short interfering RNAs (siRNAs), a subset of small RNAs recognized by the RNAi pathway, are selectively loaded into the Argonaute 2 (Ago2) protein [3], an endoribonuclease [4], and enact gene silencing through binding and cleaving complementary mRNAs [5,6]. The ability to design and deliver exogenous siRNAs for targeted protein knockdown offers an attractive therapeutic strategy [7].

RNAi is initiated when the pathway proteins recognize RNA duplexes with the characteristic siRNA structure, ~19 internal base pairs and 2 nt 3′ overhangs [6,8]. Either strand from the duplex can be processed and loaded into Ago2 to form the mature RNA-induced silencing complex (RISC) [9]. The quantity of intended RISC formed and its activity determine the efficacy of a siRNA in silencing its target. While generally only one siRNA strand (the guide strand) can target a specific mRNA, in systems where targets exist for both siRNA strands, both strands can silence [3,10]. Even in the absence of a specific target, loading and function of the unintended strand (the passenger strand) can result in significant nonspecific off-target effects [11,12]. Thus, to achieve maximal activity and specificity of silencing, it is essential to design siRNAs for loading and activity of only the intended guide strand. To do so requires an understanding of how to control (1) selection and loading of the guide strand and (2) the half-life and activity of RISC. Concomitantly, these same design rules should be applied to minimizing the loading and activity of the passenger strand, thereby achieving maximal functional asymmetry, which we define as the ratio of the silencing activities of the siRNA strands.

The concept of asymmetry has been essential to designing siRNAs since their initial structure (double-stranded) and their functional form (single-stranded) were determined [13]. Factors that differentiate between the two siRNA strands have been investigated, with the objective of determining how the RNAi pathway proteins identify the guide and passenger strand of each siRNA. The first feature discovered for predicting siRNA asymmetry was relative terminal hybridization stability ($\Delta\Delta$G) (Fig. 1A) [13], which captures the relative stability of the duplex structures at the

Department of Chemical Engineering and Materials Science, Michigan State University, East Lansing, Michigan.

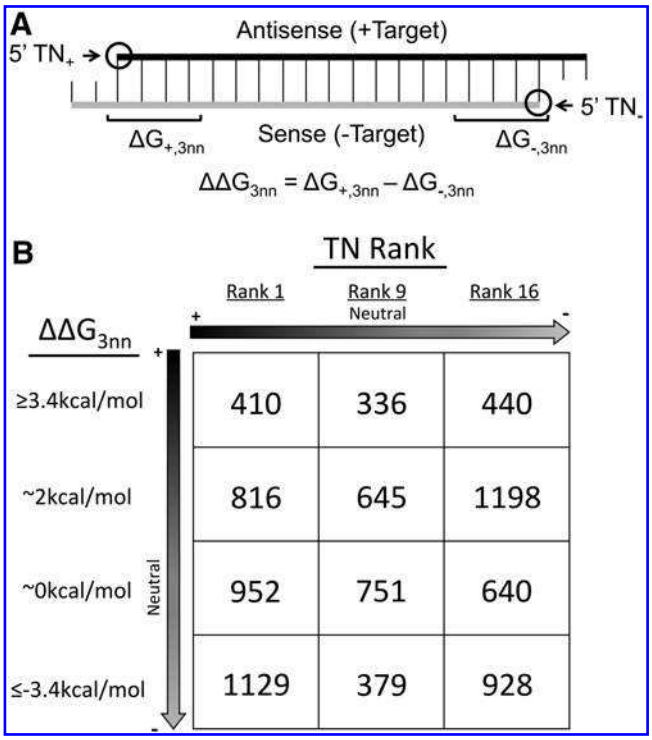


FIG. 1. Characteristics of *PKR*-targeting siRNAs. **(A)** Definition of siRNA relative terminal hybridization energy ($\Delta\Delta G_{3nn}$). 5′ TNs (5′ TN_+ and 5′ TN_-) used to classify sequences according to 5′ TN rankings (TN Rank) are listed in Table 1. **(B)** Feature predicted to favor greater antisense (+) strand activity (*black*) and feature predicted to favor greater sense (−) strand activity (*gray*). siRNAs are named by the 5′ position on the *PKR* mRNA targeted by the siRNA (+) strand. siRNA sequences are provided in Supplementary Table S1. PKR, Protein Kinase R; siRNA, short interfering RNA; TN, terminal nucleotide.

siRNA termini. $\Delta\Delta$G was shown to predict the functional asymmetry of siRNAs [13,14] and their directional binding by proteins of the RNAi pathway [15,16]. In our prior work, we classified siRNA asymmetries according to the 5′ terminal nucleotide (TN) on each strand (TN Rank) (Fig. 1A and Table 1) [17–19], showing this feature to be more predictive of siRNA activity than $\Delta\Delta$G.

However, using activity as the readout minimizes the degree to which identified features can be linked to specific steps in the RNAi mechanism. Our goals in the present study were to understand whether TN Rank and $\Delta\Delta$G predict siRNA strand selection in addition to activity and to quantify the degree to which strand selection determines the final activity of a siRNA strand. Our results showed that both TN Rank and $\Delta\Delta$G predict asymmetric strand loading, although less accurately than predicting siRNA functional asymmetry. While the influence of TN on strand activity is principally post-loading, presumably through interactions between the guide strand and Ago2, hybridization stability, particularly at the 5′ terminus, influences both strand selection and post-loading events (e.g., interactions with the mRNA target). The findings presented in this study increase our understanding of siRNA structure–function relationships and can be useful in the selection of active and specific siRNAs.

Table 1. 5′ Terminal Nucleotide Rankings (from [22])

TN Rank	*5′ TN_+: 5′ TN_-*
1	U:G
2	A:G
3	U:C
4	U:A
5	A:C
6	U:U
7	C:G
8	G:C
9	A:A
10	G:G
11	A:U
12	C:C
13	G:A
14	G:U
15	C:A
16	C:U

Arrow denotes TN Ranks predicted to favor greater antisense (+) strand activity (*black*) or favor greater sense (−) strand activity (*gray*). TN, terminal nucleotide.

Materials and Methods

Cell culture and transfection

HeLa cells were maintained in DMEM High Glucose (cat# 11965-092; Life Technologies) supplemented with 10% FBS (cat# 16000-044; Life Technologies) and 1% Pen-Strep (cat# 15240062; Life Technologies) and incubated at 37°C and 5% CO_2 in a humidified incubator. Before transfection, cells were plated at 15,000 cells/well in 96-well plates or 350,000 cells/well in 6-well plates in media without Pen-Strep for 24 h. Forward transfections were performed with Lipofectamine 2000 (cat# 11668019; Life Technologies) at a final concentration of 2.33 μg/mL, based on final well volume. Lipoplexes were prepared as per the manufacturer's instructions in Opti-MEM (cat# 31985-070; Life Technologies); siRNA and plasmid concentrations are specified for each experiment.

Nucleic acids

Protein Kinase R (*PKR*)-targeting siRNAs were ordered from Dharmacon as duplexes, designed with a 19-bp antisense strand targeting the *PKR* gene and 3′ uridine dinucleotide overhangs. siRNA sequences are listed in Supplementary Table S1; Supplementary materials are available online at www.liebertpub.com/nat. siRNA nearest-neighbor energy parameters were obtained from the DINAMelt Web Server [20,21] and used to calculate relative terminal hybridization stabilities (Fig. 1A and Supplementary Table S1). All DNA primers for this study were ordered from Integrated DNA Technologies.

For the luciferase experiments, the *PKR* gene was PCR amplified from pET28a-PKR [22] (kindly provided by Dr. Philip C. Bevilacqua) and subcloned into the psiCHECK2-AS34a vector [23] (Addgene No. 37099) in the forward (psiCHECK2-PKR+) or reverse (psiCHECK2-PKR−) direction

using Clontech In-Fusion cloning (Supplementary Fig. S1). Plasmids were verified by Sanger sequencing. Primers used for cloning and sequencing are listed in Supplementary Table S2.

Primers for stem-loop reverse transcription quantitative polymerase chain reaction (RT-qPCR) were designed and prepared according to published specifications [24], with a constant hairpin region and a 6 nt overlap with the amplification target. All sequences used are listed in Supplementary Table S3. qPCR was performed with a sequence-specific forward primer and a universal reverse primer. All primers were verified by spiking in known quantities of siRNA and verifying amplification efficiency. Standard curves were generated for each siRNA sequence to account for variation in the amplification efficiency of each siRNA-specific primer set.

Dual-Luciferase assay

HeLa cells were seeded in a 96-well plate at 15,000 cells/well in 100 μL of media 24 h before transfection. After 24 h, the media was changed, and both the *PKR* and nontargeting siRNAs were diluted in Opti-MEM and cotransfected with 40 ng/well of either the (+) or (−) psiCHECK2-AS34a plasmid, using 0.35 μL/well of Lipofectamine 2000 for a total transfection volume of 50 μL. Serial dilutions of *PKR* siRNAs ranging from 3.2 to 10,000 pM were transfected in duplicate, and transfections were performed twice for each siRNA. The *PKR* siRNA was diluted in a nontargeting siRNA to maintain total siRNA concentration at 10,000 pM per well. 24 h post-transfection, the media was aspirated, replaced with 79 μL of Dulbecco's PBS (cat# 14040133; Invitrogen), and lysed using 79 μL of Dual-Glo Luciferase Reagent (cat# E2940; Promega). After incubating at room temperature for 15 min on a rocker, 150 μL of the solution was transferred to a solid white 96-well plate, and firefly activity was measured using a Synergy H4 microplate reader (Biotek). Seventy five microliters of Dual-Glo Stop & Glo Reagent (cat# E2940; Promega) was added, and after a 10 min incubation period, *Renilla* activity was measured. Relative siRNA activity was determined by subtracting background, dividing *Renilla* signal by the firefly luciferase signal within a well, and then normalizing the ratio to the ratio from a well treated with a nontargeting siRNA. Values of 0 and 1 indicate complete silencing and no silencing, respectively. IC_{50} values were determined by fitting data to the following equation,

$$Y = \frac{1}{1 + 10^{X - Log(IC_{50})}}$$

using least-square regression, where X is the concentration of targeting siRNA and Y is the relative siRNA activity.

Stem-loop RT-qPCR

HeLa cells were plated in a 6-well plate and transfected after 24 h with 10 nM *PKR* siRNA and 10 nM of a nontargeting siRNA (internal standard). siRNA-Ago2 coimmunoprecipitation was similar to the procedure described [25], with several exceptions. Immunoprecipitations were scaled down to 400 μg of cell lysate; cell lysate protein concentrations were determined using Bradford assay (cat# 5000201; Bio-Rad) with a BSA standard. Immunoprecipitation of Ago2 was performed with 3.375 μg of Ago2 Antibody (Clone 11A9 cat# SAB4200085-200UL; Sigma-Aldrich) prebound to 25 μL of Protein G magnetic beads (cat# S1430S; NEB). Proteinase K Digestion Buffer was also supplemented with ~400 μg/mL tRNA (cat# 10109541001; Roche), to act as a carrier, and Ago2-bound RNAs purified using Direct-zol RNA Purification Kit according to the manufacturer's protocol (cat# R2051; Zymo Research).

The protocol for RT-qPCR quantification was developed based on published methods [24,26,27]. Stem-loop primers were folded by heating to 95°C for 10 min, ramping to 75°C over 10 min, then holding at 75°C, 68°C, 65°C, 62°C, and 60°C for 30 min each, before ramping to 4°C over 2 h. RNA samples were heated to 65°C for 5 min and snap cooled on ice for 5 min immediately before reverse transcription. Pulsed reverse transcription reactions were assembled with 1× First-Strand Buffer, 10 mM DTT, 0.25 mM dNTP mix, 20 U Superscript III (cat# 18080044; Life Technologies), 0.2 U of SUPERaseIn (cat# AM2696; Ambion), and 1 nM stem-loop primer in 20 μL with 1 μL of purified RNA and cycled as follows: 16°C for 30 min followed by 60 cycles of 30°C for 30 s, 42°C for 30 s, and 50°C for 1 s, and reverse transcriptase was then heat inactivated at 85°C for 5 min. qPCRs were assembled with 300 nM of a siRNA-specific forward primer, 300 nM universal reverse primer, 1× IQ SYBR Green Supermix (cat# 170882; Bio-Rad), and 1.8 μL of cDNA in a 25 μL reaction. qPCR cycling was performed as follows: 95°C for 10 min followed by 40 cycles of 95°C for 10 s and 60°C for 10 s on a MyiQ Thermocycler (Bio-Rad). (+) siRNA strand, (−) siRNA strand, and internal standard siRNA strand were reverse transcribed and quantified separately (Supplementary Table S4). Amplification specificity was verified by performing stem-loop RT-qPCR on samples that were transfected with a siRNA that was not complementary to the stem-loop primers. Control PCRs were performed on all samples following reverse transcription in the absence of reverse transcriptase. No amplification that interfered with quantification was observed in any PCR control.

Statistical methods

Linear regression analysis was performed in python with the pandas and statsmodels packages, using an ordinary least square regression. Briefly, 1- and 2- factor models were built with either siRNA IC_{50} values or siRNA loading into Ago2 as the response variable and correlated with TN Rank and $\Delta\Delta G_{3nn}$ (double-stranded analyses) or with TN and ΔG (single-stranded analyses). P values for 1-factor model were used to test if the slope was nonzero. 2-factor models were built using either (1) TN Rank with $\Delta\Delta G_{3nn}$ or (2) TN and ΔG at each position along the duplex using P values to test if the slope of the $\Delta\Delta G_{3nn}$ or ΔG parameter was nonzero. A Bonferroni correction was applied to P values to correct for multiple comparisons. All other statistical analyses were performed using GraphPad Prism 6, with the exception of ΔAICc, which was calculated using RStudio with the AICcmodavg package. ΔAICc and Akaike weights were used to compare linear regression models. Akaike weights provide relative probabilities that one model better fits the data.

ΔAICc and Akaike weights contain a correction for models with different numbers of variables, allowing direct comparison among models. When an additional variable is added to a model and the Akaike weight favors the new model, the variable is providing new explanatory information.

Results

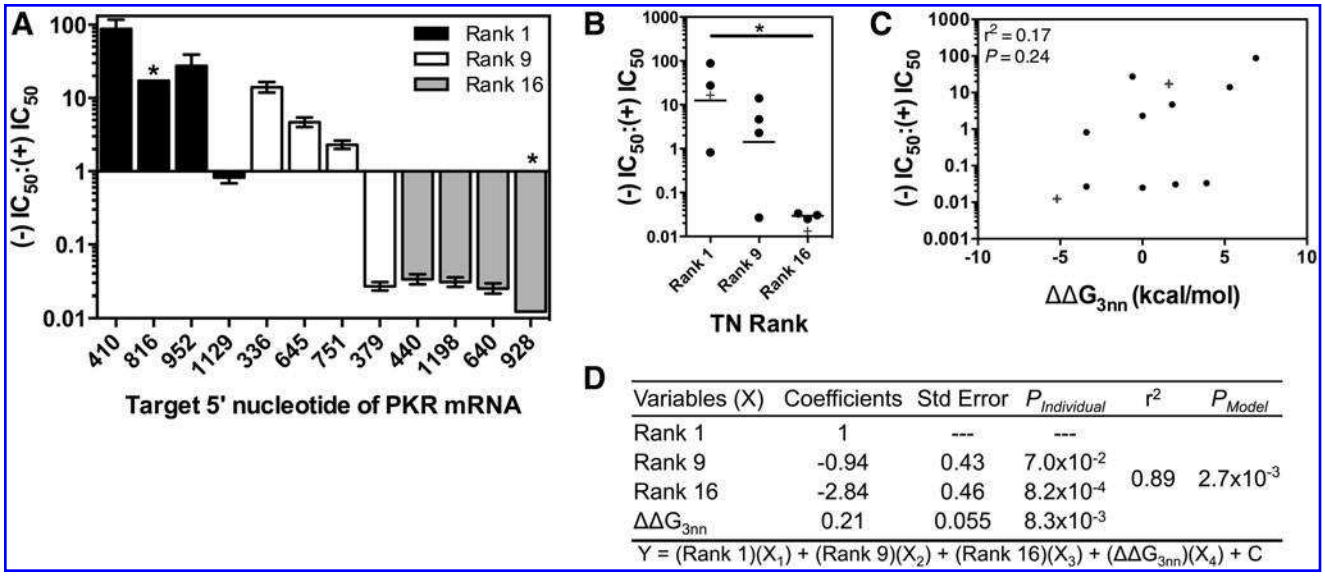
TN Rank and $\Delta\Delta G_{3nn}$ influence siRNA functional asymmetry in cultured HeLa cells

Our previous work described the importance of TN Rank (Table 1) and $\Delta\Delta G_{3nn}$ (Fig. 1A) in predicting siRNA activity [18]. The set of siRNAs chosen to investigate TN Rank and $\Delta\Delta G_{3nn}$ was selected to possess favorable, neutral, and unfavorable rankings of the two features, with favorable referring to a characteristic that would be associated with higher silencing of the siRNA against its target mRNA, in this case *PKR* (Fig. 1B) [18]. In this study, we wished to determine if these parameters could also predict siRNA functional asymmetry, as the most highly active and specific siRNAs will have functional asymmetries highly biased in favor of the intended guide strand.

We initially measured the activity of each siRNA strand using a luciferase reporter assay where the full-length coding sequence of *PKR* [hereafter the (+) target, targeted by the (+) strand of each siRNA] or the full-length complement sequence of *PKR* [hereafter the (−) target, targeted by the (−) strand of each siRNA] was cloned as a fusion product downstream of the *Renilla* luciferase gene (Supplementary Fig. S1). siRNA strand IC_{50} values were determined by cotransfecting HeLa cells with one of the plasmids and one of the *PKR*-targeting siRNAs. Functional asymmetry was calculated as the ratio of IC_{50} values for each pair of complementary strands (Fig. 2A). Two siRNA strands, PKR816(−) and PKR928(+), did not display any measurable activity (Supplementary Fig. S2); for these siRNAs, functional asymmetry was estimated assuming an IC_{50} of 10 nM, the highest concentration tested, for the inactive strand. Examining our results for both TN Rank and $\Delta\Delta G_{3nn}$, we find that, as we found with predicting strand activity, TN Rank is a stronger predictor of functional asymmetry than $\Delta\Delta G_{3nn}$ (Fig. 2B, C), with the information provided by $\Delta\Delta G_{3nn}$ improving the correlation, but being insufficient to provide a significant correlation alone (Fig. 2C, D). The importance of the $\Delta\Delta G_{3nn}$ is most evident for sequences with intermediate TN Rank, where positive $\Delta\Delta G_{3nn}$ sequences favor (+) strand activities, while negative $\Delta\Delta G_{3nn}$ sequences favor (−) strand activities (Fig. 2A).

Asymmetric strand loading of siRNA strands

To better understand the steps of the mechanism that, on aggregate, result in functional asymmetry, we measured siRNA strand loading into Ago2 by transfecting HeLa cells with our siRNA of interest for 24 h, immunoprecipitating Ago2, and then measuring its associated RNA by stem-loop RT-qPCR. We observed that TN Rank correlated with siRNA asymmetric strand loading (Fig. 3A, B). $\Delta\Delta G_{3nn}$ was not correlated with siRNA asymmetric strand loading (similar to



Variables (X)	Coefficients	Std Error	$P_{Individual}$	r^2	P_{Model}
Rank 1	1	---	---		
Rank 9	-0.94	0.43	7.0×10^{-2}	0.89	2.7×10^{-3}
Rank 16	-2.84	0.46	8.2×10^{-4}		
$\Delta\Delta G_{3nn}$	0.21	0.055	8.3×10^{-3}		

$Y = (\text{Rank 1})(X_1) + (\text{Rank 9})(X_2) + (\text{Rank 16})(X_3) + (\Delta\Delta G_{3nn})(X_4) + C$

FIG. 2. siRNA Functional Asymmetry. **(A)** Relative siRNA functional asymmetry sorted left to right by TN Rank and then by decreasing $\Delta\Delta G_{3nn}$. Shown are the ratios of siRNA (−) to (+) strand IC_{50} values determined from curve fitting to four biological replicates. *Error bars* are ±1 SD. (*) represents values where one siRNA strand IC_{50} was not measurable and assumed to be 10 nM (the highest concentration tested) for calculating functional asymmetry. **(B)** Distribution of siRNA functional asymmetry with respect to TN Rank. *Horizontal lines* represent the mean activity within the TN Rank; overhead bar marked with (*) represents a statistically different pairwise comparison ($P < 0.05$) by one-way ANOVA with Tukey's *post hoc* analysis. **(C)** Correlation of functional asymmetry with $\Delta\Delta G_{3nn}$. **(B, C)** siRNAs PKR816 and PKR928 had one estimated IC_{50} and were marked with (+) and not included in any statistical analysis. **(D)** Multiple linear regression of siRNA functional asymmetry with TN Rank and $\Delta\Delta G_{3nn}$, where Y is the functional asymmetry [(−) IC_{50}:(+) IC_{50}] and X_i is the variable. Coefficients were normalized to Rank 1. $P_{individual}$ is calculated based on comparison of the individual coefficients with 0. P_{model} is calculated based on comparison of the model fit with a fit to random data, a measure of the predictive value of the model. C = 0.89. PKR, Protein Kinase R; siRNA, short interfering RNA; TN, terminal nucleotide.

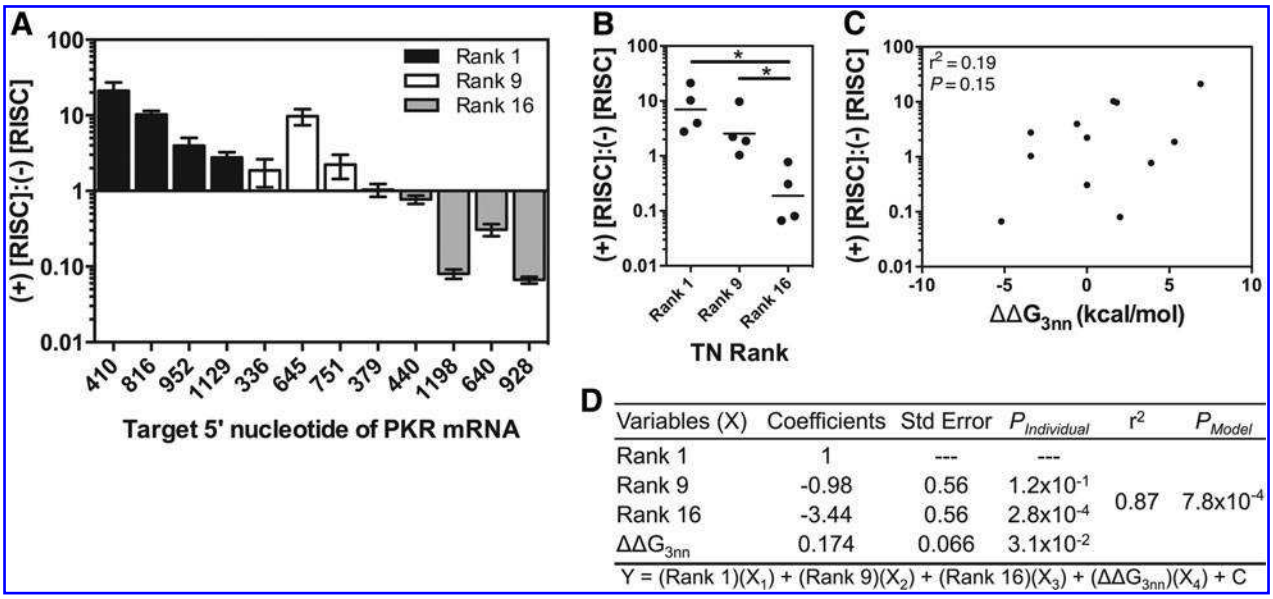


FIG. 3. siRNA Asymmetric Strand Loading. **(A)** Relative siRNA strand loading into Ago2 sorted left to right by TN Rank and then by decreasing $\Delta\Delta G_{3nn}$. Shown are the ratios of siRNA (+) strand to (–) strand loading. $n = 3$, *error bars* are ±1 SD. **(B)** Distribution of siRNA asymmetric loading by TN Rank. *Horizontal lines* represent the mean activity within the TN Rank; overhead bars marked with (*) represent statistically different pairwise comparisons ($P < 0.05$) by one-way ANOVA with Tukey's *post hoc* analysis. **(C)** Correlation of siRNA asymmetric strand loading with $\Delta\Delta G_{3nn}$. **(D)** Multiple linear regression of asymmetric strand loading with TN Rank and $\Delta\Delta G_{3nn}$, where Y is the relative strand loading ((+) [RISC]:(–) [RISC]) and X_i is the variable. Coefficients were normalized to Rank 1. $P_{individual}$ is calculated based on comparison of the individual coefficients with 0. P_{model} is calculated based on comparison of the model fit with a fit to random data, a measure of the predictive value of the model. C = 1.75. Ago2, Argonaute 2; RISC, RNA-induced silencing complex; siRNA, short interfering RNA; TN, terminal nucleotide.

its lack of correlation with functional asymmetry) (Fig. 3C). In this case, however, it provided considerably less predictive information when combined with TN Rank (Fig. 3D) than it did for relative strand activity, as shown by the relative differences in Akaike weights for the combined 2-factor models and their respective single factor models (i.e., a smaller difference in Akaike weights indicates a smaller gain in model information content; Table 2).

For a number of siRNAs, we observed differences in the relative loadings and relative activities. PKR1129, which exhibited symmetrical strand activity ($P = 0.24$, extra sum-of-squares F test), had asymmetric strand loading ($P = 0.02$, two-tailed t-test) in line with its TN Rank. Interestingly, it is the (–) strand of PKR1129 that has an unexpectedly high activity compared to the other sequences with a 5′ G and similar loading [PKR410(–) and PKR952(–); Supplementary Fig. S3A, B]. This is presumably because this sequence has a $\Delta\Delta G_{3nn}$ strongly in favor of the (–) siRNA strand, which may partially offset its unfavorable 5′ nucleotide. In addition, sequences PKR336, PKR379, and PKR440 displayed symmetrical strand loadings and asymmetric activities (Figs. 2A and 3A). As with PKR1129, these siRNAs have large $\Delta\Delta G_{3nn}$ values (either highly positive or negative), suggesting that $\Delta\Delta G_{3nn}$ may significantly influence siRNA function only after exceeding a threshold magnitude.

Table 2. ΔAICc and Akaike Weight Comparisons of 1- and 2-Factor Linear Regressions Using Terminal Nucleotide Rank and $\Delta\Delta G_{3nn}$ (Figs. 2D and 3D)

	(+) IC_{50}:(–) IC_{50}		(+) [RISC]:(–) [RISC]	
Model	*ΔAICc*	*Akaike weight*	*ΔAICc*	*Akaike weight*
TN Rank	14.9	0.001	1.1	0.365
$\Delta\Delta G_{3nn}$	16.8	0	10.4	0.003
Combined	0	0.999	0	0.632

RISC, RNA-induced silencing complex; TN, terminal nucleotide.

A direct comparison of siRNA strand loading with its activity demonstrates that while siRNA loading and activity are correlated, multiple strands diverge significantly from the trend (Supplementary Fig. S4). These findings suggest some modulation of activity post-siRNA loading. Furthermore, the change in Akaike weights between the individual factors (TN Rank and $\Delta\Delta G_{3nn}$) and the combined model indicates that TN Rank and $\Delta\Delta G_{3nn}$ are complementary predictors of functional asymmetry (Table 2). In comparison to activity, $\Delta\Delta G_{3nn}$ adds less complementary information to the prediction of asymmetric strand loading (Table 2), indicating that $\Delta\Delta G_{3nn}$ contains information more relevant to post-loading events. For clarity, as we discuss post-loading events, we will use *RISC-specific activity* to refer to the enzymatic characteristics of a mature RISC that are independent of the amount of RISC generated.

RISC-specific activity

Because RISC-specific activity is a feature of the siRNA after the separation of the two siRNA strands, we analyzed

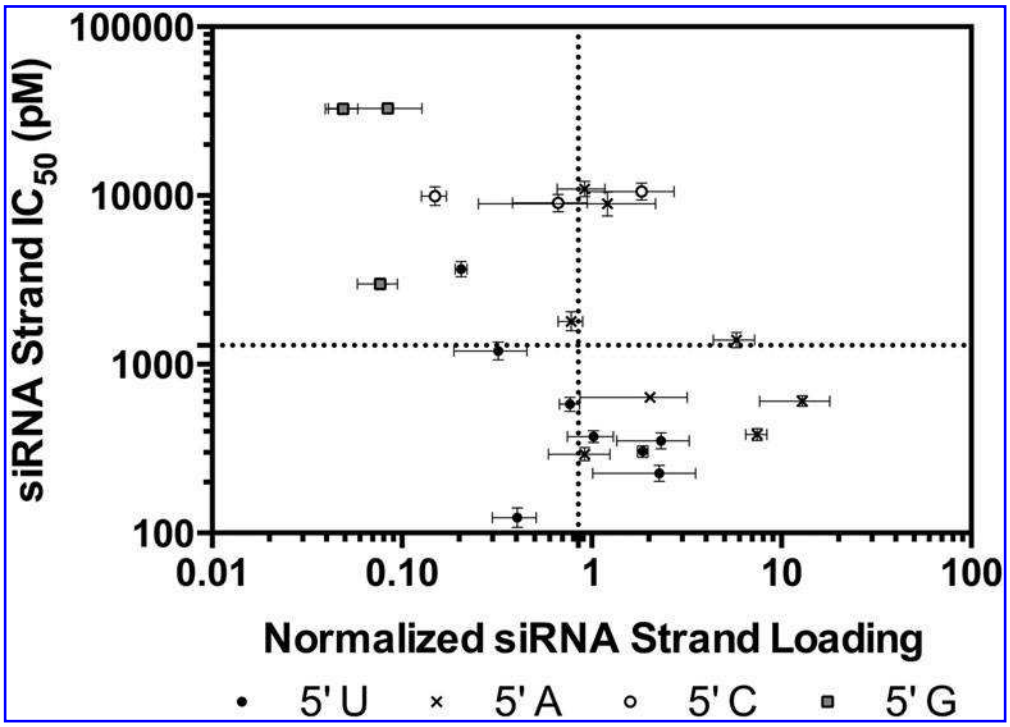


FIG. 4. Correlation of siRNA strand activity with Ago2 loading. *X* values are fitted IC_{50} values; $n=4$; error bars are ±1 SD. *Y* values are the normalized average siRNA loading; $n=3$; error bars are ±1 SD. *Dotted lines* represent the median of each data set. siRNA, short interfering RNA. Ago2, Argonaute 2.

the behavior of the individual siRNA strands independent of their complements (Fig. 4). Likewise, in lieu of TN Rank and $\Delta\Delta G_{3nn}$, we examined TN (U, A, C, G) and the energies of all nearest-neighbor pairs ($\Delta G_{(x)-(y)nt}$) along the length of the duplex (Fig. 5). While similar, the nucleotide preferences for loading (A>U>C>G) (Supplementary Fig. S3D) are different from those for activity (U>A>C>G) (Supplementary Fig. S3C), indicating that TN impacts RISC-specific activity, as others have previously observed [28]. However, TN alone does not account for all of the observed variation in activities; for instance, siRNAs with a 5′ A displayed similar loading but varied substantially in their activities (Fig. 4).

To identify regions of the siRNA duplex important for controlling RISC-specific activity, we correlated loading and activity with the TN and ΔGs along the length of the duplex, looking for regions in the duplex important for siRNA activity, but not important for RISC loading (Fig. 5). We found that only the nearest-neighbor parameters surrounding the TN, $\Delta G_{(-2)-1nt}$, and ΔG_{1-2nt} correlated with activity (Fig. 5). To decouple the effects of the TN sequence from ΔG, we used multiple linear regression and reevaluated the correlation between ΔG and both siRNA loading and activity. We found that ΔG_{3-4nt} and $\Delta G_{17-18nt}$ were only predictive of siRNA activity not loading (Fig. 5). This indicates that RISC-specific

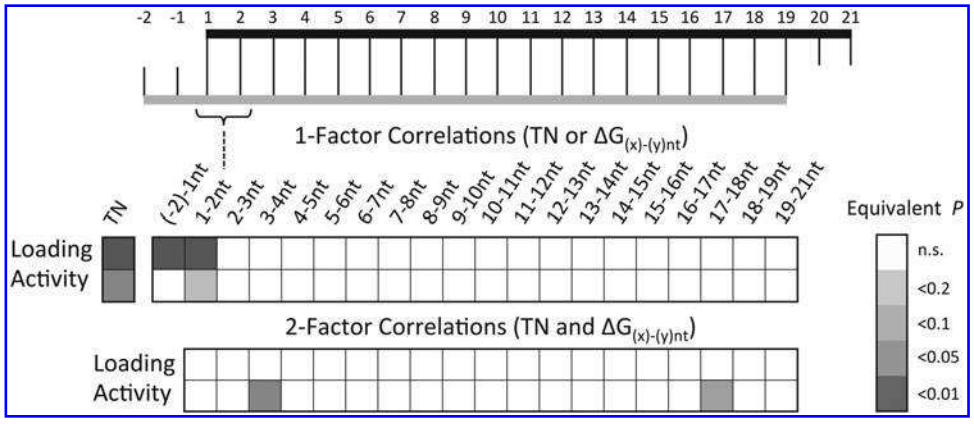


FIG. 5. Evaluation of siRNA loading and activity predictors. 1-factor and 2-factor correlations of siRNA loading and activity with TN and $\Delta G_{(x)-(y)nt}$ along the length of the duplex. 2-factor correlations use TN as one variable and a $\Delta G_{(x)-(y)nt}$ value as the other. Significance in these correlations indicates that the variable predicts loading or activity independent of TN. A Bonferroni correction was applied to correct for multiple comparisons. Regression statistics for 1-factor and 2-factor models are listed in Supplementary Tables S5 and S6, respectively. siRNA, short interfering RNA; TN, terminal nucleotide.

activity is improved for sequences with weaker (less negative) ΔG_{3-4nt} and stronger (more negative) $\Delta G_{17-18nt}$. These ΔG values are included in the calculation of $\Delta\Delta G_{3nn}$ and likely contribute to its utility in predicting siRNA activity.

Discussion

Designing siRNAs for maximal function requires the double-stranded structure to efficiently enter the RNAi pathway such that it can be easily processed to RISC loaded with the correct single strand. It is therefore critical to understand how the RNAi pathway proteins sense specific features of siRNA duplexes that differentiate the two siRNA strands. siRNA functional asymmetry was first characterized in *Drosophila* lysates and attributed to the difference in $\Delta\Delta G_{4nn}$ [13]. Many Argonaute proteins contain an additional structure that preferentially binds 5′ TNs [29]. Our previous work sought to explain how these two terminal features contribute to siRNA activity [18,19]. In this study, we showed how TN Rank and $\Delta\Delta G_{3nn}$ contribute to siRNA functional asymmetry and asymmetric strand loading. Furthermore, we have shown that, in cultured HeLa cells, functional asymmetry is modulated both before and after siRNA loading, with TN Rank and $\Delta\Delta G_{3nn}$ exerting influence at different stages of the mechanism.

Functional asymmetry and asymmetric strand loading were not perfectly correlated, indicating additional modulation of siRNA activity post-siRNA loading (Supplementary Fig. S4). Our top-down analysis identified free energies at two positions, ΔG_{3-4nt} and $\Delta G_{17-18nt}$, to be predictive of siRNA activity, but not siRNA loading (Fig. 5). Mature RISC splits siRNA-target mRNA interactions into five domains, the 5′ anchor (Nucleotide: 1), the seed region (Nucleotides: 2–8), the central region (Nucleotides: 9–12), the 3′ supplemental region (Nucleotides: 13–16), and the 3′ tail (Nucleotides: 17–21) [30]. The preference for weaker thermodynamic stability through nucleotides 1–4 of the siRNAs is well established [13,31–33]; our findings suggest that weaker hybridization in this region leads to greater RISC-specific activity (Fig. 5), in agreement with findings that high stability in the seed region decreases RISC turnover [34]. RISC-mRNA interactions beyond the 3′ supplemental region have been shown to be dispensable for target cleavage or in some cases even deleterious [28,30]. We contend that the preference for a stronger $\Delta G_{17-18nt}$ is also important in maintaining a stable interaction between RISC and its target, similar to base pairs in the 3′ supplemental region [30].

After delivery of the siRNA to the cell cytoplasm, efficient guide strand loading requires nucleotide-independent phosphorylation of the siRNA 5′ TNs [35], outcompeting native miRNAs for binding to Ago2 [36,37], conformational changes within Ago2 facilitated by HSP90 and potentially other proteins [10,38–41], and cleavage and removal of the passenger strand [42,43]. We observed a large variation in the total amount of Ago2 loading even among siRNA strands with the same TN (Supplementary Fig. S3B). In analyzing other characteristics for their ability to predict loading, we found $\Delta G_{(-2)-1nt}$ and ΔG_{1-2nt} of the 5′ end of the guide strand to be the most predictive features of siRNA strand loading (Fig. 5). Interestingly, there was no correlation between siRNA strand loading and the ΔG values near the 3′ terminus (i.e., Nucleotides 17–21), suggesting that the interactions of Ago2 are primarily with the 5′ ends of siRNAs. Thus, parameters based on relative strand properties (e.g., TN Rank and $\Delta\Delta G$) are less useful in predicting absolute siRNA strand loading and activity than in predicting relative strand activities and loadings.

The established binding preferences for the Ago2 nucleotide specificity loop (U>A>C>G) match the order of siRNA activities in our data (Supplementary Fig. S3C), but differ from the loading preferences we observed (A>U>C>G) (Supplementary Fig. S3D). This suggests that 5′ TN affects RISC-specific activity, presumably through differences in RISC half-life [28], similar to what was observed in *Drosophila* Ago1 [44]. The correlation of siRNA loading with the 5′ terminal $\Delta G_{(-2)-1nt}$ and ΔG_{1-2nt} indicates that siRNA loading is controlled, in part, by the 5′ terminal duplex energy (Fig. 5), although TN remains a significant predictor of siRNA strand loading because of its collinearity with the $\Delta G_{(-2)-1nt}$ and ΔG_{1-2nt} terms. Recently, sensing of siRNA strand thermodynamics was proposed to occur through Ago2 MID domain interactions with the phosphate backbone of the first four nucleotides of the siRNA guide strand [45], sensing the accessibility and single-strand character of weakly base paired termini [45]. Collectively, our findings agree with this model, but suggest that the first and second nucleotides of the siRNA guide strand are the most significant in determining total strand loading (Fig. 5). siRNA design algorithms should thus account for relative siRNA features (TN Rank and $\Delta\Delta G$) to ensure correct strand selection and absolute strand features (TN and ΔG) to ensure high strand loading and RISC-specific activity.

While our experiments and analysis could not cover the entire siRNA feature space, the number of siRNAs tested was large enough to identify regions of the siRNA that explain some of the differences between activity and loading. 5′ C and 5′ G sequences are underrepresented in our individual siRNA strands; thus, a more comprehensive set of siRNAs could potentially have provided more insight into how features beyond the 5′ TN influence siRNA behavior both pre- and postloading. Nonetheless, the parameters identified indicate that it is possible to design siRNAs with the goal of high loading, high activity, or both.

Conclusion

Taken together, our findings suggest that siRNA loading and activity are partially independent, allowing for an additional degree of freedom in siRNA selection and design. The behavior of siRNA strands is strongly dependent on the 5′ nucleotide, but the duplex hybridization energies have a second order effect on both siRNA loading and RISC-specific activity. Specifically, $\Delta G_{(-2)-1nt}$ and ΔG_{1-2nt} are predictive of siRNA loading, and ΔG_{3-4nt} and $\Delta G_{17-18nt}$ are predictive of RISC-specific activity. Finally, we have shown that siRNA functional asymmetry is a result of many competing factors that ultimately control the activity of each siRNA strand. A more accurate understanding of the interplay among these factors will lead to better siRNAs.

Acknowledgments

The authors thank Professor Philip C. Bevilacqua for providing the *PKR* plasmid, Professor Judy Lieberman for the psiCHECK2-AS34a plasmid provided through Addgene

(Plasmid #37099). We would also like to thank the members of the Cellular and Biomolecular Engineering Laboratory for their technical and intellectual support, in particular Sarah Thorwall. Financial support for this work was provided, in part, by the Michigan State University, the National Science Foundation (CBET 1510895 and 1547518), and the National Institutes of Health (GM079688, GM089866, and CA176854).

Author Disclosure Statement

No competing financial interests exist.

References

1. Snead NM and JJ Rossi. (2010). Biogenesis and function of endogenous and exogenous siRNAs. Wiley Interdiscip Rev RNA 1:117–131.
2. Carthew RW and EJ Sontheimer. (2009). Origins and mechanisms of miRNAs and siRNAs. Cell 136:642–655.
3. Yoda M, T Kawamata, Z Paroo, X Ye, S Iwasaki, Q Liu and Y Tomari. (2010). ATP-dependent human RISC assembly pathways. Nat Struct Mol Biol 17:17–23.
4. Liu JD, MA Carmell, FV Rivas, CG Marsden, JM Thomson, JJ Song, SM Hammond, L Joshua-Tor and GJ Hannon. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science 305:1437–1441.
5. Martinez J, A Patkaniowska, H Urlaub, R Lührmann and T Tuschl. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110:563–574.
6. Elbashir SM, J Martinez, A Patkaniowska, W Lendeckel and T Tuschl. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J 20:6877–6888.
7. Angart P, D Vocelle, C Chan and SP Walton. (2013). Design of siRNA therapeutics from the molecular scale. Pharmaceuticals 6:440–468.
8. Zamore PD, T Tuschl, PA Sharp and DP Bartel. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101:25–33.
9. Rivas FV, NH Tolia, JJ Song, JP Aragon, JD Liu, GJ Hannon and L Joshua-Tor. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol 12:340–349.
10. Sakurai K, M Amarzguioui, D Kim, J Alluin, B Heale, M Song, A Gatignol, MA Behlke and JJ Rossi. (2011). A role for human Dicer in pre-RISC loading of siRNAs. Nucleic Acids Res 39:1510–1525.
11. Jackson AL, SR Bartz, J Schelter, SV Kobayashi, J Burchard, M Mao, B Li, G Cavet and PS Linsley. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol 21:635–637.
12. Jackson AL, J Burchard, J Schelter, BN Chau, M Cleary, L Lim and PS Linsley. (2006). Widespread siRNA ''off-target'' transcript silencing mediated by seed region sequence complementarity. RNA 12:1179–1187.
13. Schwarz D, G Hutvagner, T Du, Z Xu, N Aronin and P Zamore. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell 115:199–208.
14. Lu ZJ and DH Mathews. (2008). Efficient siRNA selection using hybridization thermodynamics. Nucleic Acids Res 36:640–647.
15. Tomari Y, C Matranga, B Haley, N Martinez and PD Zamore. (2004). A protein sensor for siRNA asymmetry. Science 306:1377–1380.
16. Noland CL, E Ma and JA Doudna. (2011). siRNA repositioning for guide strand selection by human dicer complexes. Mol Cell 43:110–121.
17. Gredell JA, MJ Dittmer, M Wu, C Chan and SP Walton. (2010). Recognition of siRNA asymmetry by TAR RNA binding protein. Biochemistry 49:3148–3155.
18. Malefyt AP, M Wu, DB Vocelle, SJ Kappes, SD Lindeman, C Chan and SP Walton. (2013). Improved asymmetry prediction for siRNAs. FEBS J 281:320–330.
19. Walton SP, M Wu, JA Gredell and C Chan. (2010). Designing highly active siRNAs for therapeutic applications. FEBS J 277:4806–4813.
20. Markham NR and M Zuker. (2005). DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Res 33: W577–W581.
21. Markham NR and M Zuker. (2008) UNAFold: software for nucleic acid folding and hybridization. In: *Bioinformatics, Volume II: Structure, Function, and Applications*. J.M. Keith, ed. Humana Press, Totowa, NJ, pp 3–31.
22. Bevilacqua PC and TR Cech. (1996). Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. Biochemistry 35:9983–9994.
23. Navarro F, D Gutman, E Meire, M Caceres, I Rigoutsos, Z Bentwich and J Lieberman. (2009). miR-34a contributes to megakaryocytic differentiation of K562 cells independently of p53. Blood 114:2181–2192.
24. Varkonyi-Gasic E and RP Hellens. (2011) Quantitative stem-loop RT-PCR for detection of microRNAs. In: *RNAi and Plant Gene Function Analysis: Methods and Protocols*. H. Kodama and A. Komamine, eds. Humana Press, New York, NY; pp. 145–157.
25. Beitzinger M and G Meister. (2011). Experimental identification of microRNA targets by immunoprecipitation of Argonaute protein complexes. Methods Mol Biol 732: 153–167.
26. Chen CF, DA Ridzon, AJ Broomer, ZH Zhou, DH Lee, JT Nguyen, M Barbisin, NL Xu, VR Mahuvakar, *et al.* (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res 33:e179.
27. Kramer MF. (2011) Stem-loop RT-qPCR for miRNAs. In: *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., Hoboken, NJ; pp. 15.10.11–15.10.15.
28. De N, L Young, PW Lau, NC Meisner, DV Morrissey and IJ MacRae. (2013). Highly complementary target RNAs promote release of guide RNAs from human Argonaute2. Mol Cell 50:344–355.
29. Frank F, N Sonenberg and B Nagar. (2010). Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. Nature 465:818–822.
30. Wee LM, CF Flores-Jasso, WE Salomon and PD Zamore. (2013). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell 152: 1055–1067.
31. Reynolds A, D Leake, Q Boese, S Scaringe, WS Marshall and A Khvorova. (2004). Rational siRNA design for RNA interference. Nat Biotechnol 22:326–330.
32. Amarzguioui M and H Prydz. (2004). An algorithm for selection of functional siRNA sequences. Biochem Biophys Res Commun 316:1050–1058.
33. Hibio N, K Hino, E Shimizu, Y Nagata and K Ui-Tei. (2012). Stability of miRNA 5′ terminal and seed regions is correlated with experimentally observed miRNA-mediated silencing efficacy. Sci Rep 2:996.

34. Salomon WE, SM Jolly, MJ Moore, PD Zamore and V Serebrov. (2015). Single-molecule imaging reveals that argonaute reshapes the binding properties of its nucleic acid guides. Cell 162:84–95.
35. Weitzer S and J Martinez. (2007). The human RNA kinase hClp1 is active on 3′ transfer RNA exons and short interfering RNAs. Nature 477:222–226.
36. Koller E, S Propp, H Murray, W Lima, B Bhat, TP Prakash, CR Allerson, EE Swayze, EG Marcusson and NM Dean. (2006). Competition for RISC binding predicts in vitro potency of siRNA. Nucleic Acids Res 34:4467–4476.
37. Tanudji M, D Machalek, GM Arndt and L Rivory. (2010). Competition between siRNA duplexes: impact of RNA-induced silencing complex loading efficiency and comparison between conventional-21 bp and Dicer-substrate siRNAs. Oligonucleotides 20:27–32.
38. Iwasaki S, M Kobayashi, M Yoda, Y Sakaguchi, S Katsuma, T Suzuki and Y Tomari. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell 39:292–299.
39. Bernard MA, L Wang and SD Tachado. (2015). DICER-ARGONAUTE2 complex in continuous fluorogenic assays of RNA interference enzymes. PLoS One 10:e0120614.
40. Willkomm S, A Deerberg, J Heidemann, F Flugge, J Meine, R Hu, R Kretschmer-Kazemi Far and T Restle. (2016). Recombinant hTRBP and hPACT modulate hAgo2-catalyzed siRNA-mediated target RNA cleavage in vitro. PLoS One 11:e0146814.
41. Ye X, N Huang, Y Liu, Z Paroo, C Huerta, P Li, S Chen, Q Liu and H Zhang. (2011). Structure of C3PO and mechanism of human RISC activation. Nat Struct Mol Biol 18: 650–657.
42. Matranga C, Y Tomari, C Shin, D Bartel and P Zamore. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell 123:607–620.
43. Leuschner PJF, SL Ameres, S Kueng and J Martinez. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. EMBO Rep 7:314–320.
44. Kawamata T, M Yoda and Y Tomari. (2011). Multilayer checkpoints for microRNA authenticity during RISC assembly. EMBO Rep 12:944–949.
45. Suzuki HI, A Katsura, T Yasuda, T Ueno, H Mano, K Sugimoto and K Miyazono. (2015). Small-RNA asymmetry is directly driven by mammalian Argonautes. Nat Struct Mol Biol 22:512–521.

Address correspondence to:
S. Patrick Walton, ScD
Department of Chemical Engineering and Materials Science
Michigan State University
428 South Shaw Lane, Room 3249
Engineering Building
East Lansing, MI 48824-1226

E-mail: spwalton@egr.msu.edu

Received for publication February 17, 2016; accepted after revision May 6, 2016.

This article has been cited by:

1. Rahul Maheshwari, Muktika Tekade, Piyush Gondaliya, Kiran Kalia, Antony D'Emanuele, Rakesh Kumar Tekade. 2017. Recent advances in exosome-based nanovehicles as RNA interference therapeutic carriers. *Nanomedicine* **12**:21, 2653-2675. [Crossref]