the FEBS Journal

Improved asymmetry prediction for short interfering RNAs

Amanda P. Malefyt[1,*], Ming Wu[2,†], Daniel B. Vocelle[1], Sean J. Kappes[1], Stephen D. Lindeman[1], Christina Chan[1,2,3] and S. Patrick Walton[1]

1 Department of Chemical Engineering and Materials Science, Michigan State University, East Lansing, MI, USA
2 Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA
3 Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, USA

Keywords
asymmetry; dsRNA-dependent protein kinase R; enhanced green fluorescent protein; short interfering RNA

Correspondence
S. Patrick Walton, Department of Chemical Engineering and Materials Science, Michigan State University, 428 S. Shaw Lane, East Lansing, MI 48824-1226, USA
Fax: +1 517 432 1105
Tel: +1 517 355 5135
E-mail: spwalton@egr.msu.edu

Present address:
*McKetta Department of Chemical and Bioprocess Engineering, Trine University, Angola, IN, USA
†Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA

(Received 3 June 2013, revised 28 August 2013, accepted 26 September 2013)

doi:10.1111/febs.12599

In the development of RNA interference therapeutics, merely selecting short interfering RNA (siRNA) sequences that are complementary to the mRNA target does not guarantee target silencing. Current algorithms for selecting siRNAs rely on many parameters, one of which is asymmetry, often predicted through calculation of the relative thermodynamic stabilities of the two ends of the siRNA. However, we have previously shown that highly active siRNA sequences are likely to have particular nucleotides at each 5′-end, independently of their thermodynamic asymmetry. Here, we describe an algorithm for predicting highly active siRNA sequences based only on these two asymmetry parameters. The algorithm uses end-sequence nucleotide preferences and predicted thermodynamic stabilities, each weighted on the basis of training data from the literature, to rank the probability that an siRNA sequence will have high or low activity. The algorithm successfully predicts weakly and highly active sequences for enhanced green fluorescent protein and protein kinase R. Use of these two parameters in combination improves the prediction of siRNA activity over current approaches for predicting asymmetry. Going forward, we anticipate that this approach to siRNA asymmetry prediction will be incorporated into the next generation of siRNA selection algorithms.

Introduction

Therapeutic applications of RNA interference (RNAi) make use of a conserved pathway for gene expression regulation that possesses the potential for exquisite sequence specificity through the complementarity of short interfering RNAs (siRNAs) for their targets [1–3]. Although the technology has yet to demonstrate its full potential in clinical applications [4,5], there remains major interest in developing siRNA-based therapeutics [6]. Because RNAi represents a therapeutic approach that can be applied to nearly any disease [7,8], improvements in the design and development of siRNA therapeutics have the potential to have a significant impact on clinical practice.

A number of intermolecular interactions are critical to the activity of siRNAs, including those with the delivery vehicle [9–11], the target mRNA [12–15], and the pathway proteins [16–19]. Whereas a single RNA guide strand and argonaute 2 are the minimal components required for active silencing *in vitro* [20], the proteins Dicer and TAR RNA-binding protein are important for RNA-induced silencing complex (RISC)

Abbreviations
EGFP, enhanced green fluorescent protein; H1299, human lung carcinoma; HepG2, hepatocellular carcinoma; LF2K, Lipofectamine 2000; PKR, dsRNA-dependent protein kinase R; RISC, RNA-induced silencing complex; RNAi, RNA interference; siRNA, short interfering RNA.

loading complex/RISC activity *in vivo* [21–23]. Other proteins, such as the protein activator of dsRNA-dependent protein kinase R (PKR) [24,25] and component 3 promoter of RISC [26], may also have important but as yet undefined functional roles in the RNAi process. One essential process executed by the pathway proteins is the identification and loading of the siRNA guide strand into RISC loading complex/RISC and the concomitant destruction of the passenger strand [2,27,28]. The likelihood of one siRNA strand becoming the guide strand relative to the other strand is termed asymmetry [27,29].

There are currently multiple proteins that are thought to participate in sensing the asymmetry of siRNA duplexes [18,30–32]. When the existence of siRNA asymmetry was first established, it was proposed that the relative hybridization stabilities of the two ends of the siRNA sequence constituted the principal means by which asymmetry was sensed by the pathway proteins [29]. Since that time, nearly all algorithms for selecting highly active siRNAs have used a thermodynamic calculation for asymmetry, among other parameters [29,33–37]. More recently, evidence has begun to accumulate that the terminal nucleotides on each 5′-end of the siRNA may be valuable for predicting the activity of an siRNA [18,30,38], in particular when classified according to the 16 possible combinations of nucleotides. When terminal nucleotide classification is combined with relative hybridization stability, the accuracy of predicting siRNA activity improves markedly [38].

In this study, we wanted to predict siRNA activities on the basis of only two asymmetry characteristics, terminal nucleotide classification and relative thermodynamic stability (Fig. 1), and establish experimentally their relative importance in determining the activity of an siRNA. Using a logistic regression model, we successfully predicted active and inactive sequences for the exogenous protein enhanced green fluorescent protein (EGFP) and the endogenous protein PKR. In addition, the combination of both end-sequence and thermodynamic stability features provided improved correlation with siRNA activity as compared with either feature individually. These results demonstrate that asymmetry may be determined by more factors than just relative stability, and algorithms for prediction of siRNA activity should also account for terminal nucleotide sequence classification in asymmetry calculations.

Results

Ranking and selection of EGFP-targeting siRNAs

Our ranking algorithm was initially tested on siRNAs to target the EGFP mRNA. From the cDNA sequence (Doc. S1), there were 824 possible siRNA sequences, which were ranked according to the difference between the algorithm's predicted likelihood of high and low activity. For comparison, commercial algorithms (Dharmacon and Ambion) were also used. These selection algorithms were chosen because their predictions are based solely on the characteristics of the siRNA and not on other factors used in some selection algorithms, such as target mRNA structure, which would make it difficult to directly compare the accuracy of our asymmetry-based predictions with predictions from more detailed selection approaches. The commercial rankings only included sequences predicted to have high activity, as opposed to the entire range of possible siRNA sequences. Although this is adequate for those needing effective siRNA sequences, it does not provide sufficient data to enable comparison of the characteristics of high-activity and low-activity siRNAs. The Dharmacon algorithm ranked the recommended siRNAs, whereas for the Ambion algorithm there were no distinctions among the top 35 candidate sequences. Interestingly, there was no overlap between the lists of recommended sequences provided by the commercial algorithms. Aggregating the commercial recommendations with our predictions, we chose 11 sequences to test experimentally that would allow us to preliminarily compare the relative utility of the three prediction approaches (Table 1).

Transfection experiments were performed with human lung carcinoma (H1299)-EGFP cells at various siRNA concentrations (Fig. 2). Surprisingly, 81% of the sequences had some silencing effect as compared with control treatments, with 73% of sequences showing a > 75% reduction in protein levels at

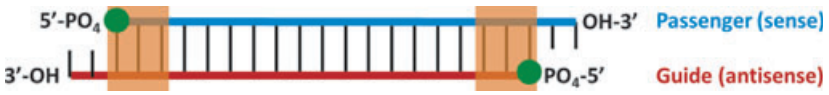

Fig. 1. Algorithm features for designing highly active siRNAs. From the selected mRNA target, all potential siRNAs were classified according to their terminal sequence and thermodynamic stability. For the terminal sequence (circles), the 5′-terminal nucleotides, antisense: sense strand, were used; thermodynamic stability (shaded rectangles) was calculated on the basis of $\Delta\Delta G$ values, using the three terminal nearest neighbors. All sequences were synthesized to contain UU overhangs on the 3′-ends.

Table 1. EGFP-targeting siRNA sequences selected for this study, sorted by algorithm rank.

5′ Target position	siRNA sequence	5′-End nucleotides Antisense: Sense	Rank	Thermodynamic $\Delta\Delta G$ (kcal·mol^{-1})	Algorithm rank	Testing rationale
415	UGUACUCCAGCUUGUGCCC	U:G	1	4.2	1	Highest algorithm ranking
416	UUGUACUCCAGCUUGUGCC	U:G	1	2.1	17	Highest Dharmacon rank (91)
274	UGCGCUCCUGGACGUAGCC	U:G	1	0.8	38	On Ambion list
757	UGGGCAGCGUGCCAUCAUC	U:G	1	−2.6	177	High end, low $\Delta\Delta G$
306	CUUGUAGUUGCCGUCGUCC	C:G	7	2.9	239	On Ambion list
396	CAGGAUGUUGCCGUCCUCC	C:G	7	0.5	389	On Ambion list
126	GAUGAACUUCAGGGUCAGC	G:G	10	1.9	418	On Ambion list
711	AUGGCUAAGCUUCUUGUAC	A:G	2	−3.7	551	Low rank, high Dharmacon (73)
783	CAUCCCGCUCUCCUGGGCA	C:U	16	3.1	695	Low end, high $\Delta\Delta G$
159	GGGCCAGGGCACGGGCAGC	G:G	10	−2.3	700	On Ambion list
253	CGGGCAUGGCGGACUUGAA	C:U	16	−5.2	824	Lowest algorithm rank

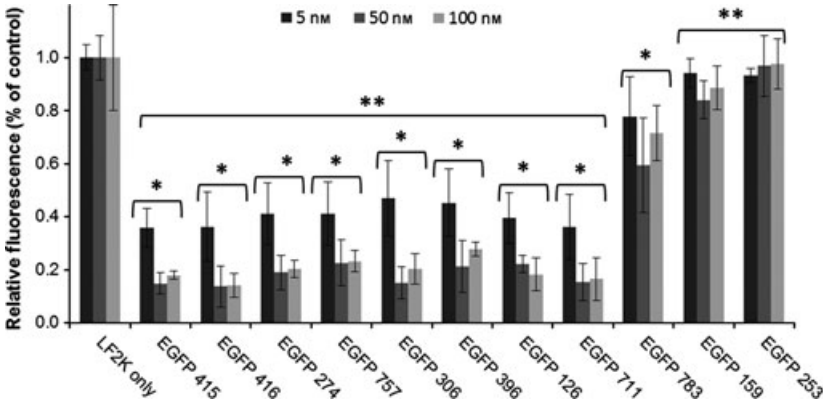

Fig. 2. Silencing of EGFP by selected siRNAs. H1299 cells expressing EGFP were treated with 5, 50 and 100 nM siRNA for 24 h with LF2K. Results are normalized to the EGFP fluorescence of control cells treated only with LF2K. siRNA treatments are ordered on the basis of algorithm predictions (415 = highest predicted activity; 253 = lowest predicted activity). Error bars represent ± 1 standard deviation; $n = 4$ (5 nM), $n = 5$ (50 nM), and $n = 3$ (100 nM). *Significant difference ($P < 0.05$) as compared with LF2K-only treatments. **Significant difference ($P < 0.05$) as compared with EGFP 783.

siRNA concentrations of 50 nM and 100 nM. One sequence (EGFP 783) showed intermediate silencing (the difference from other sequences is indicated by double asterisks), suggesting a gradient, rather than a step change, in silencing ability between active and inactive sequences. In general, sequences predicted by our algorithm to have higher activity showed increased inhibition of EGFP fluorescence. The rank order of activities was maintained at lower (5 nM) and saturating (50 nM and 100 nM) siRNA concentrations. The two sequences chosen on the basis of their opposing rankings between the two features in our approach, EGFP 757 (favorable terminal sequence; unfavorable $\Delta\Delta G$) and EGFP 783 (unfavorable terminal sequence; favorable $\Delta\Delta G$), ultimately showed activities that correlated with their terminal nucleotide classification rather than their thermodynamic stability.

To further investigate the hypothesis that terminal sequence classification was more important than thermodynamic stability in predicting siRNA activity, silencing efficiencies were compared against algorithm rank, terminal nucleotide rank and $\Delta\Delta G$ values individually (Fig. 3). The correlation between activity and terminal nucleotide rank was better than for thermodynamic stability alone, with the correlation with full algorithm rank being better than either alone. This agrees with our prior work showing that terminal nucleotide classification is generally a more informative predictor of siRNA activity, but that inclusion of the thermodynamic calculation provides some additional complementary information [38].

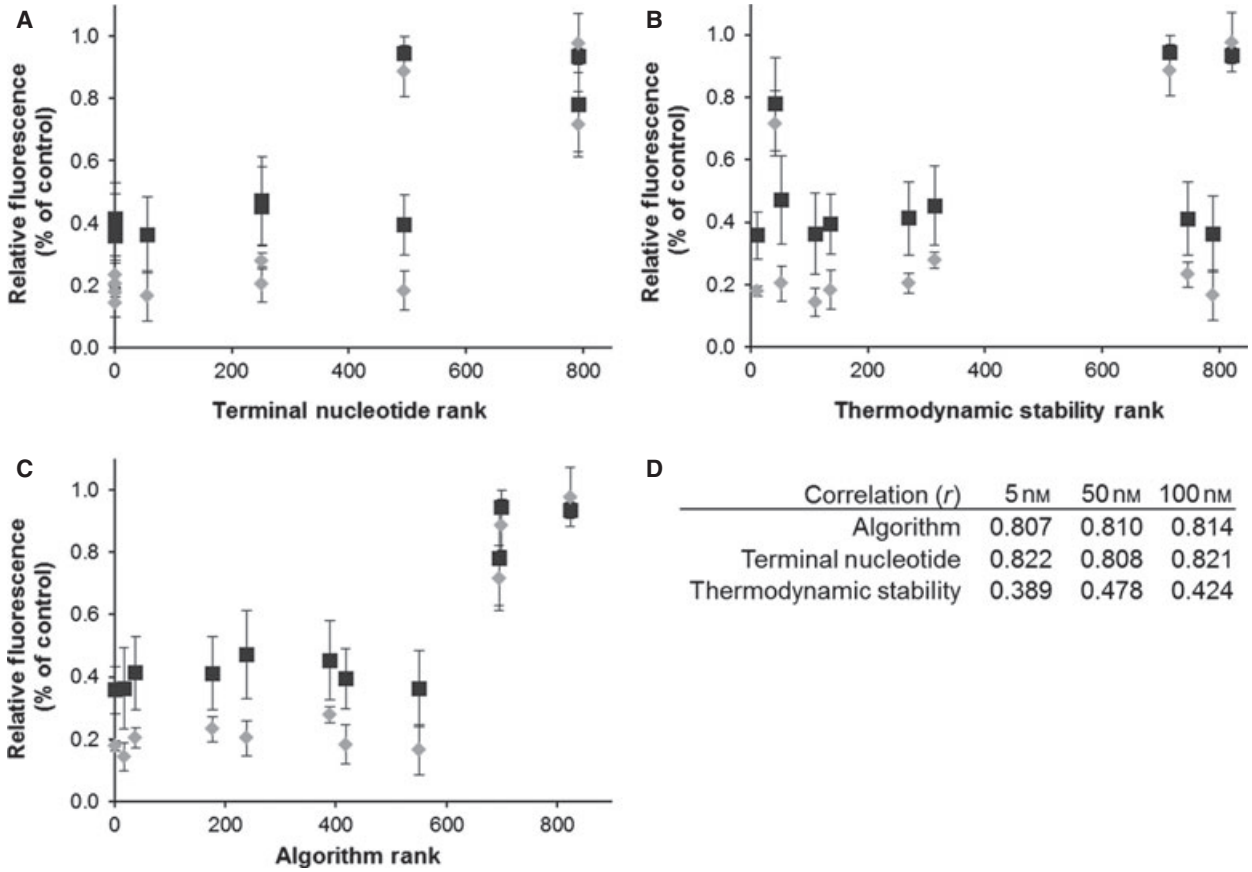

Fig. 3. Correlation of exogenous gene silencing with each feature of the algorithm. EGFP silencing results are plotted against terminal nucleotide rank only (A), $\Delta\Delta G$ values only (B), or algorithm rank (C). The squares represent 5 nM siRNA, and the diamonds represent 100 nM siRNA. For visual clarity, the 50 nM data points were not included in the plots, but they showed the same trend for silencing activity. The correlation coefficients for each dataset are tabulated (D).

Ranking and selection of PKR-targeting siRNAs

Although our algorithm successfully predicted active and inactive sequences for EGFP, we wanted to confirm similar results for an endogenous protein, the signaling pathway mediator PKR. In addition, through systematic selection of siRNAs of high, medium and low nucleotide ranking and siRNAs having high, medium and low relative thermodynamic stabilities (Table 2), we aimed to further explore the relative importance of each of these characteristics in predicting silencing activity. We selected PKR as a model endogenous protein on the basis of our prior work and expertise in silencing this protein [39]. Although PKR is a double-stranded RNA-responsive protein and is known to be functionally connected to proteins in the RNAi pathway [40], the lack of any cytotoxicity across all of our experiments suggested that it was not initiating any generalized immune response to the siRNAs that would confound our specific silencing results.

Transfection experiments were performed with hepatocellular carcinoma (HepG2) cells at 100 nM siRNA (Fig. 4). In this case, 55% of sequences showed a > 50% reduction in PKR protein levels as compared with control cells. Again, sequences predicted to have higher activity showed increased reduction in PKR protein levels. When sorted by end-nucleotide classification, sequences in the UG class showed the best silencing activity, on average, regardless of their thermodynamic stability. Conversely, sequences in the low-ranking terminal nucleotide class, CU, did not show significant silencing, even when they had highly positive $\Delta\Delta G$ values. In the intermediate category of end sequence (AA), silencing activity correlated strongly with the calculated thermodynamic stability, with a favorable value resulting in significant silencing and an unfavorable value not. Taken together with our results from the EGFP experiments, these results support the argument that terminal sequence classification is a stron-

Table 2. PKR-targeting siRNA sequences selected for this study, sorted by end-sequence ranking, followed by relative thermodynamic stability.

5′ Target position	siRNA sequence	5′-End nucleotides Antisense: Sense	Rank	Thermodynamic $\Delta\Delta G$ (kcal·mol^{-1})	Algorithm rank	Testing rationale
410	UAAUGAAAUCCUUCUGGCC	U:G	1	6.3	1	High end, high $\Delta\Delta G$
193	UCUUUGAUCUACCUUCACC	U:G	1	1.9	61	Previously validated
816	UUUAAAAUCCAUGCCAAAC	U:G	1	0	230	High end, medium $\Delta\Delta G$
952	UUGCCAAUGCUUUUACUUC	U:G	1	−4.2	948	High end, low $\Delta\Delta G$
336	AAUUCUAUUGAUAAGGCCU	A:A	9	5.3	205	Medium end, high $\Delta\Delta G$
645	AUCUGCUGAGAAGUCACCU	A:A	9	0	1154	Medium end, medium $\Delta\Delta$G
379	AUGCACACUGUUCAUAAUU	A:A	9	−5.2	1926	Medium end, low $\Delta\Delta G$
1198	CAAAGAGUUCCAAAGCCAA	C:U	16	5.5	1606	Low end, high $\Delta\Delta G$
640	CUGAGAAGUCACCUUCAGA	C:U	16	0	2206	Low end, medium $\Delta\Delta G$
928	CCGCCUUCUCGUUAUUAUA	C:U	16	−5.2	2352	Low end, low $\Delta\Delta G$

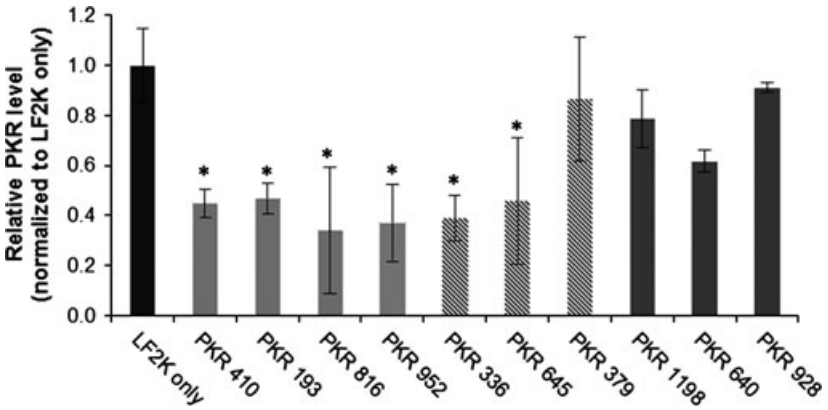

Fig. 4. Silencing of PKR by selected siRNAs. HepG2 cells were reverse-transfected with 100 nM siRNA for 48 h with LF2K. Results are normalized to both total protein level and PKR levels of control cells treated only with LF2K (black). siRNAs were grouped by end-sequence rank: UG (light gray), AA (striped), and CU (medium gray). Error bars represent ± 1 standard deviation; $n = 3$. *Significant difference ($P < 0.05$) as compared with LF2K-only treatment.

ger predictor of siRNA activity that relative hybridization stability.

Our algorithm achieved a better correlation between rank and siRNA activity than achieved by either terminal nucleotide or thermodynamic stability independently (Fig. 5). Although the PKR terminal nucleotide correlation is less than that achieved from the EGFP data, the thermodynamic stability and overall algorithm rank correlation coefficients remain similar, even with a larger range of possible siRNA sequences (2352 for PKR versus 824 for EGFP), showing the algorithm's fidelity for targets of various sizes. It is noteworthy that, for both sequences, the top-ranked sequence, i.e. the one that would have been chosen for predicting the activity of an siRNA against a new target, was highly active in silencing the target, further supporting our hypothesis that using only two parameters was sufficient for identifying active siRNAs against novel targets.

Discussion

The use of asymmetry is well established as being useful and important for selecting active siRNA sequences. Multistep workflow protocols for selecting effective siRNAs have been developed [41,42]. However, the selection algorithms themselves are based in part on using relative thermodynamic stability as the sole factor in determining sequence asymmetry. Other reported algorithms lack a consensus on the best way to calculate thermodynamic asymmetry for siRNA activity prediction [36,43–47]. When the commercially

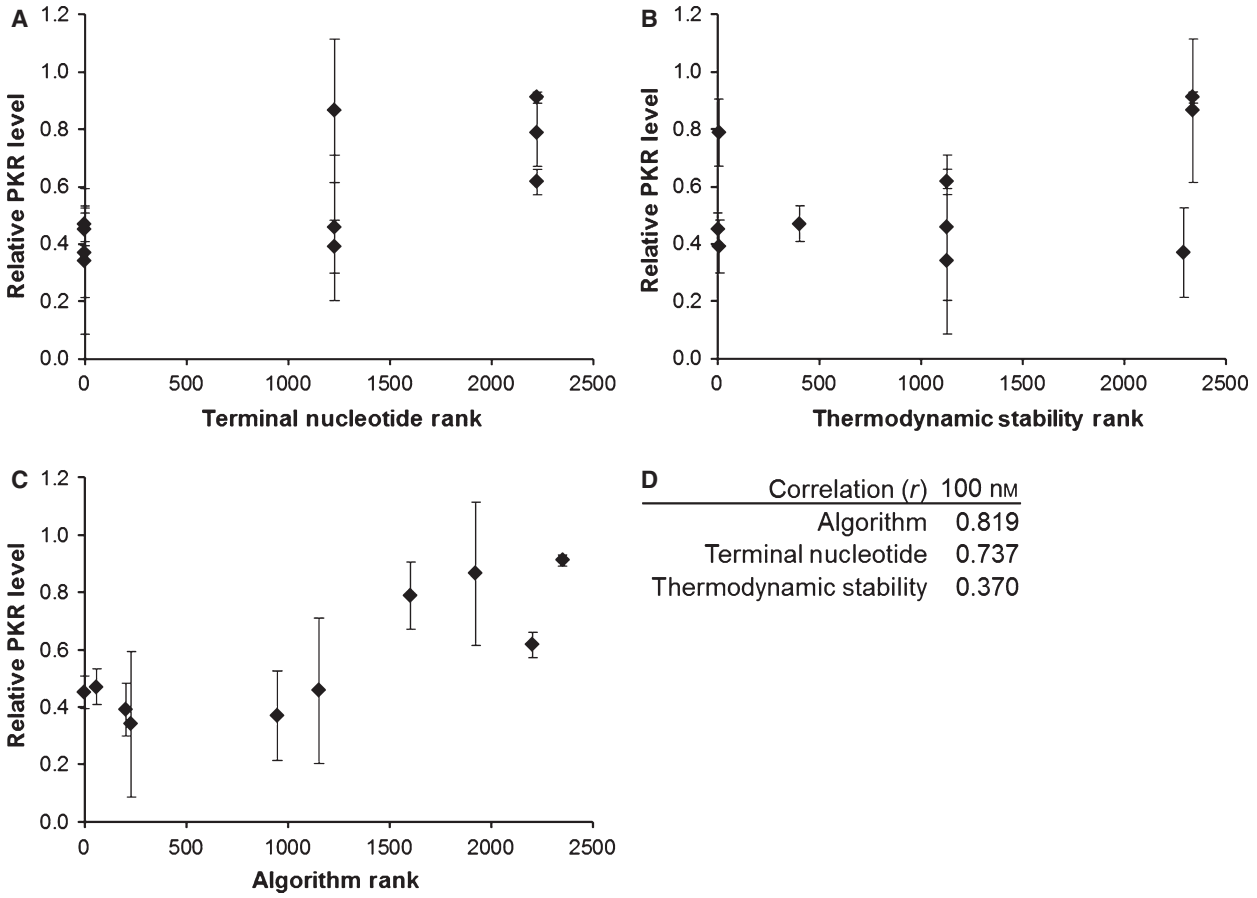

Fig. 5. Correlation of endogenous gene silencing with each feature of the algorithm. PKR silencing results are plotted against terminal nucleotide rank only (A), $\Delta\Delta G$ values only (B), or algorithm rank (C). The correlation coefficients for each dataset are tabulated (D).

available algorithms were utilized for comparison in this study, there were no overlaps between the sequences predicted to be highly active by Dharmacon and those predicted by Ambion, further illustrating the need for a consensus approach to selecting highly active siRNAs, including which parameters are most important/useful for such predictions.

The results described here further illustrate that accounting only for thermodynamic asymmetry ignores a more important feature in asymmetry, terminal nucleotide classification. Although others have identified terminal nucleotides as factors relevant to siRNA design (e.g. [35]), our approach of pairing the antisense and sense termini and weighting each pairing individually provides a unique and improved classification for sequences and their activities. Our selection technique achieves correlation coefficient values of > 0.8, whereas previously reported algorithms typically achieve correlation coefficients of 0.5–0.7 between algorithm predictions and experimental results [41,48]. As our approach is focused on the contribution of asymmetry in the siRNA to its ultimate activity, it was important to compare our approach with other ways of determining asymmetry. In calculating thermodynamic asymmetry, it is common to use one, three or four nearest neighbors at each end of the siRNA as the basis for calculation [29,36,48,49]. In our prior work [38], we showed that, in concert with terminal nucleotide classification, calculating thermodynamic asymmetry with three nearest neighbor parameters provided the most information, whereas one nearest neighbor provided the most information in the absence of terminal nucleotide classification. On the basis of this context, we compared the correlation of our data with thermodynamic calculations performed with one, three and four nearest neighbor parameters (Fig. 6). In all cases, the correlation of our experimental data was best with rankings including terminal nucleotide classification (Table 3). This strongly supports our contention that all siRNA selection algorithms would be improved by

Table 3. The correlation coefficients between EGFP silencing at 100 nM and the four different ranking methods in Fig. 6.

Correlation coefficient (r)	100 nM
Algorithm	0.814
$\Delta\Delta G$ (1 nearest neighbor)	0.539
$\Delta\Delta G$ (3 nearest neighbors)	0.424
$\Delta\Delta G$ (4 nearest neighbors)	0.327

the inclusion of our asymmetry approach in place of their current asymmetry calculation.

Our algorithm, as structured, only ranks sequences according to the likelihood of them silencing the intended target. It was our intention in this study to determine whether the factors that we tested were useful in predicting sequences of high activity. Our ranking approach does not account for potential off-target effects of the sequences. For the long-term design of siRNA therapeutics, it is essential that off-target effects be taken into account as well. That said, it is our belief that beginning therapeutic design with the most highly active sequence, which can then be modified, if needed, to improve its specificity, is a better approach to obtaining a useful therapeutic than beginning with the most highly specific sequence, which may then need to be modified to improve its silencing activity against the intended target.

Although our approach is useful for predicting active (and inactive) siRNAs, we have not yet established the causal relationship between the terminal nucleotide classification and siRNA processing and activity. Indeed, it is well established that the properties of the siRNA alone provide only partial information regarding the likely activity of the siRNA [14,15,50,51]. However, studies are increasingly reporting that more active siRNAs and microRNAs tend to contain specific nucleotides at the 5′-position of the guide strand [18,33,52], possibly a result of argonaute 2 binding specificity [30]. We expect that, going forward, our ongoing work and that of others will more firmly tie the presence of particular 5′-end nucleotides on both the guide and passenger strands with important siRNA–protein interactions that occur in the pathway to ensure proper siRNA processing.

Experimental procedures

Algorithm design and parameters

Using information from both terminal nucleotide classification and thermodynamic stability, we developed a 17-parameter logistic regression model based on both the 16 possible end-sequence combinations and the relative thermodynamic stability (Table 4; additional details in Doc. S1). The relative stability is calculated from the difference between the hybridization free energy from the 5′-end of the antisense strand and the 5′-end of the sense strand, termed ΔΔG, based on the three terminal nearest neighbor pairs [38,53] (Fig. 1). $\Delta\Delta G$ calculations were based on 21-nucleotide siRNAs with equivalent UU overhangs on the end of each strand. This calculation technique, when coupled with terminal sequence information, was shown to have the best predictive accuracy when tested on existing siRNA activity databases [38]. The weighting factors for each of the 17 parameters were based on fitting the model to the same siRNA databases [43,44]. From a cDNA sequence input, the algorithm predicts the probability that the given sequence will have high, medium or low activity. By use of the difference between the high and low probabilities, each siRNA sequence for a given target was ranked from the highest to the lowest difference. The cDNA sequences used for EGFP and PKR are included in Doc. S1, with siRNA target regions highlighted. For comparison with other asymmetry approaches (Fig. 6), asymmetry calculations were also performed with one and four nearest neighbor parameters. To ensure the most accurate comparisons across ranking approaches, sequences with equivalent values were all given the best possible ranking.

Table 4. Values of the coefficients used as weighting factors for predicting the probability of siRNAs having high (top third among all possible sequences) and low (bottom third among all possible sequences) activity.

Classifier	P (high)	P (low)
$\Delta\Delta G$	0.2973	−0.1661
U:G	1.8767	−0.9852
A:G	0.3737	−0.0193
U:C	1.8744	−1.3088
U:A	1.7387	−1.0301
A:C	0.4764	−0.406
U:U	1.5084	−1.0877
C:G	−0.2463	0.4025
G:C	−0.6404	0.3337
A:A	0.2381	0.0810
G:G	−0.7873	0.2945
A:U	−0.5707	0.3471
C:C	−0.3552	−0.0987
G:A	−1.1939	0.6077
G:U	−2.2329	1.3262
C:A	−1.1755	0.9055
C:U	−2.2846	1.2957
Intercept	−0.0498	−0.2145

Positive values indicate features that are positively correlated with high siRNA activity. By application of these weights in our algorithm, siRNAs were ranked by the magnitude of the difference between the probability of having high activity and the probability of having low activity. The 'Intercept' is applied to all sequences, and arises from the regression model (Doc. S1).

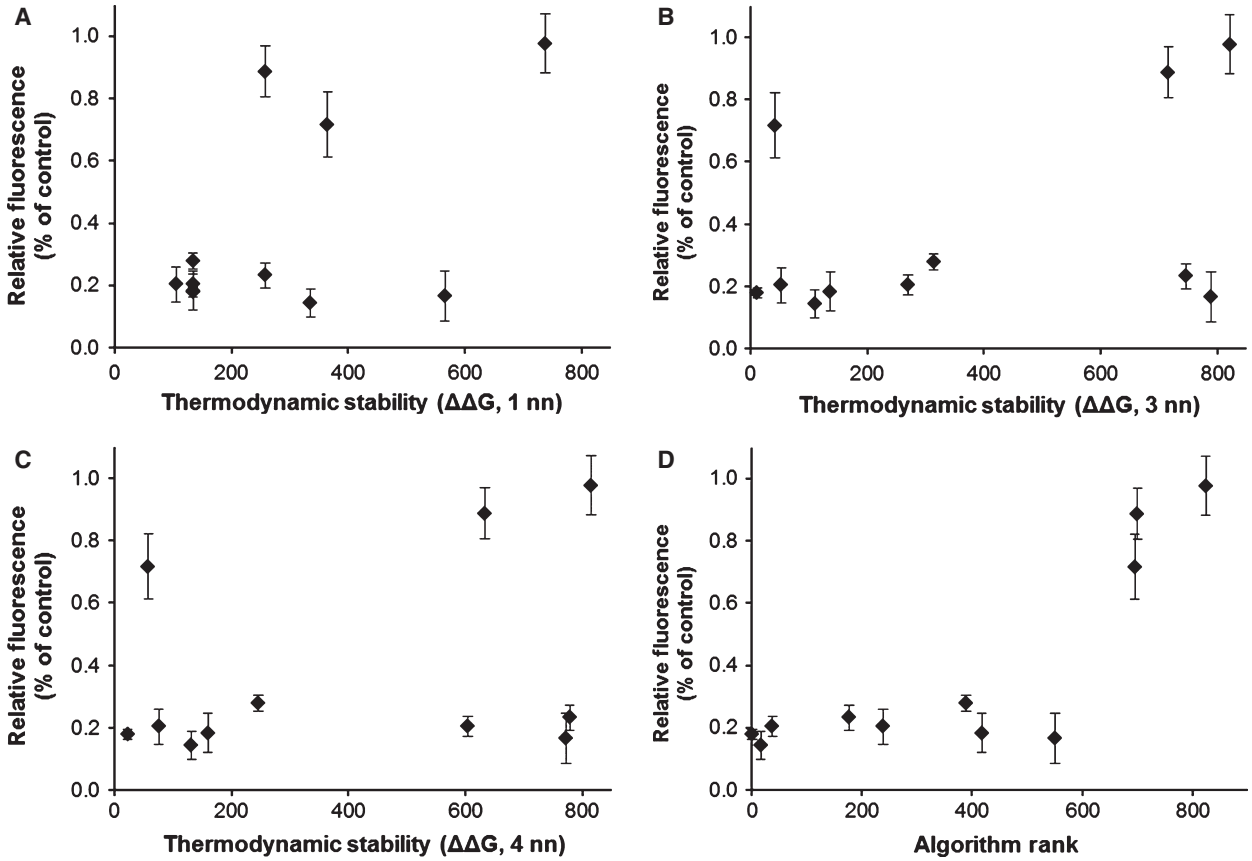

Fig. 6. Correlation of asymmetry calculations with experimental data. Results for 100 nM EGFP silencing plotted against rankings among all possible siRNA sequences, with $\Delta\Delta G$ calculated with one nearest neighbor (A), three nearest neighbors (B), four nearest neighbors (C), or our algorithm (three nearest neighbors and terminal nucleotide classification) (D).

Dharmacon ranking information was obtained from http://dharmacon.com/designcenter/DesignCenterPage (now available at http://www.thermoscientificbio.com/design-center/). Potential sequences were determined by entering the nucleotide sequence directly, removing any GC restriction requirements, and choosing the 'No BLAST' option to provide a direct comparison with the algorithm. Ambion ranking information was obtained from http://ambion.com/techlib/misc/siRNA_finder.html (no longer active after acquisition of Ambion by Life Technologies). A list of potential siRNAs was generated by entering the nucleotide sequence; no other selection parameters were required.

Materials

Lipofectamine 2000 (LF2K) was purchased from Invitrogen. All EGFP and PKR siRNA sequences were 21 nucleotides long (19 bp plus UU overhangs), and were purchased from Dharmacon. Opti-Mem (Gibco) was used for preparation of all transfection solutions. Monoclonal primary antibody against PKR (Y117) was purchased from Novus Biologicals. Monoclonal primary antibody against β-actin was purchased from Sigma. Secondary antibodies (anti-rabbit and anti-mouse) were purchased from ThermoScientific.

Cell culture

H1299 cells constitutively expressing a form of EGFP, modified to have a 2-h half-life, were generously provided by J. Kjems (University of Aarhus, Denmark). HepG2 cells were purchased from the American Type Culture Collection. Cell culture medium was prepared with DMEM High Glucose (Invitrogen) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin/streptomycin (Gibco). For the H1299 cells, 1% Geneticin (Gibco) was added to maintain EGFP expression. Cells were incubated at 37 °C in 5% CO_2, at 100% relative humidity, and subcultured every 4–5 days by trypsinization.

EGFP silencing and fluorescence analysis

H1299-EGFP cells were seeded in 96-well, black-sided, clear-bottomed plates (Fisher Scientific) at a density of 20 000 cells per well in 0.1 mL of complete medium without

antibiotics. After 24 h, 50-μL solutions of various siRNAs and LF2K were prepared in Opti-Mem, and allowed to mix for 30 min prior to their addition to cells at final concentrations of 5–100 nM siRNA and 2.3 $\mu g \cdot mL^{-1}$ LF2K. Cells were incubated in the transfection solutions at 37 °C in 5% CO_2, at 100% humidity. After 24 h, cells were washed twice with Dulbecco's $NaCl/P_i$ (Gibco), and EGFP fluorescence was quantified with a Gemini EM fluorescent plate reader (Molecular Devices) at 480-nm excitation and 525-nm emission. Fluorescence intensity was normalized to control wells that were treated with transfection reagent but no siRNA. Cytotoxicity was assessed by microscopy, and was not seen in any of the treatments (Fig. S1).

PKR silencing and western blotting

For siRNA transfections targeting human PKR in HepG2 cells, reverse transfection was performed. Briefly, 250-μL solutions of various siRNAs and LF2K were prepared in Opti-Mem, and allowed to mix for 30 min prior to their addition to standard six-well tissue culture plates. Freshly trypsinized HepG2 cells suspended in antibiotic-free medium were added to the six-well culture plates at a density of 1.5×10^6 cells per well to achieve final concentrations of 100 nM siRNA and 2.3 $\mu g \cdot mL^{-1}$ LF2K. The cells were then incubated at 37 °C in 5% CO_2, at 100% humidity, for 48 h and collected. PKR levels were measured by western blot analysis (Fig. S2). The cells were washed twice with cold $NaCl/P_i$, and lysed in 200 μL per well of CelLytic M cell lysis buffer (Sigma) supplemented with protease inhibitor cocktail (Sigma). The cell lysate was clarified by centrifugation at 15 000 ***g*** for 10 min, and the supernatant was collected. Total protein levels were quantified with the Quick Start Bradford protein assay (BioRad). Approximately 20 μg of total protein was resolved, with 8% resolving, 5% stacking SDS/PAGE gels. Proteins were then transferred to nitrocellulose membranes, and probed with primary and secondary antibodies. Biotinylated protein ladders (Cell Signaling Technology) were loaded onto one well of each SDS/PAGE gel, and antibody against biotin was used to detect the protein ladders on the western blots. Antibody detection was performed with the SuperSignal West Femto Chemiluminescence substrate kit (ThermoScientific), and imaging was performed with the Molecular Imager ChemiDoc XRS System (Bio-Rad). Band intensity was first normalized to actin to control for protein isolation and loading, and the ratio was then normalized to the ratio for control cells that received transfection reagent but no siRNA. Cytotoxicity was assessed by microscopy and CellTiter-Blue assay (Promega) (Fig. S1), and was not seen in any of the treatments.

Statistical analyses

Multiple comparisons between protein levels across different siRNA treatments conditions were performed with one-way (PKR data) or two-way (EGFP data) ANOVA followed by Tukey's HSD *post hoc* analysis with the *P*-value cut-off set at 0.05. Analyses were performed with either Microsoft Excel or MINITAB.

Acknowledgements

Financial support for this work was provided in part by Michigan State University (MSU Foundation, Center for Systems Biology, and MSU Graduate School), the National Institutes of Health (GM079688, GM089866, RR024439, DK081768, and DK088251), and the National Science Foundation (CBET 0941055). We thank all members of the Cellular and Biomolecular Laboratory for their advice and support. We also thank J. Kjems (University of Aarhus, Denmark) for providing us with the EGFP cells.

References

1 Fire A, Xu S, Montgomery M, Kostas S, Driver S & Mello C (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811.
2 Matranga C, Tomari Y, Shin C, Bartel DP & Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620.
3 Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K & Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498.
4 Pollack A (2011) Drugmakers' fever for the power of RNA interference has cooled. New York Times.
5 Krieg AM (2011) Is RNAi dead? *Mol Ther* **19**, 1001–1002.
6 Gitig D (2012) Use of siRNA in therapeutic arena on the upswing. *Genet Eng Biotech News*, **12**, 24–26.
7 Davidson BL & McCray PB (2011) Current prospects for RNA interference-based therapies. *Nat Rev Genet* **12**, 329–340.
8 Angart P, Vocelle D, Chan C & Walton SP (2013) Design of siRNA therapeutics from the molecular scale. *Pharmaceuticals* **6**, 440–468.
9 Portis AM, Carballo G, Baker GL, Chan C & Walton SP (2010) Confocal microscopy for the analysis of siRNA delivery by polymeric nanoparticles. *Microsc Res Tech* **73**, 878–885.
10 Lu JJ, Langer R & Chen JZ (2009) A novel mechanism is involved in cationic lipid-mediated functional siRNA delivery. *Mol Pharmaceut* **6**, 763–771.
11 Siegwart DJ, Whitehead KA, Nuhn L, Sahay G, Cheng H, Jiang S, Ma M, Lytton-Jean A, Vegas A, Fenton P *et al.* (2011) Combinatorial synthesis of

chemically diverse core-shell nanoparticles for intracellular delivery. *Proc Natl Acad Sci USA* **108**, 12996–13001.
12 Ameres SL, Martinez J & Schroeder R (2007) Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* **130**, 101–112.
13 Brown KM, Chu C-Y & Rana TM (2005) Target accessibility dictates the potency of human RISC. *Nat Struct Mol Biol* **12**, 469–470.
14 Gredell JA, Berger AK & Walton SP (2008) Impact of target mRNA structure on siRNA silencing efficiency: A large-scale study. *Biotechnol Bioeng* **100**, 744–755.
15 Kiryu H, Terai G, Imamura O, Yoneyama H, Suzuki K & Asai K (2011) A detailed investigation of accessibilities around target sites of siRNAs and miRNAs. *Bioinformatics* **27**, 1788–1797.
16 MacRae IJ, Ma E, Zhou M, Robinson CV & Doudna JA (2007) In vitro reconstitution of the human RISC-loading complex. *Proc Natl Acad Sci USA* **105**, 512–517.
17 Maniataki E & Mourelatos Z (2005) A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev* **19**, 2979–2990.
18 Gredell JA, Dittmer MJ, Wu M, Chan C & Walton SP (2010) Recognition of siRNA asymmetry by TAR RNA binding protein. *Biochemistry* **49**, 3148–3155.
19 Kini HK & Walton SP (2009) Effect of siRNA terminal mismatches on TRBP and Dicer binding and silencing efficacy. *FEBS J* **276**, 6576–6585.
20 Rivas FV, Tolia NH, Song JJ, Aragon JP, Liu JD, Hannon GJ & Joshua-Tor L (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* **12**, 340–349.
21 Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K & Shiekhattar R (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744.
22 Haase AD, Jaskiewicz L, Zhang H, Laine S, Sack R, Gatignol A & Filipowicz W (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep* **6**, 961–967.
23 Lima WF, Murray H, Nichols JG, Wu H, Sun H, Prakash TP, Berdeja AR, Gaus HJ & Crooke ST (2009) Human Dicer binds short single-strand and double-strand RNA with high affinity and interacts with different regions of the nucleic acids. *J Biol Chem* **284**, 2535–2548.
24 Kok KH, Ng MHJ, Ching YP & Jin DY (2007) Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J Biol Chem* **282**, 17649–17657.
25 Lee Y, Hur I, Park SY, Kim YK, Suh MR & Kim VN (2006) The role of PACT in the RNA silencing pathway. *EMBO J* **25**, 522–532.
26 Ye X, Huang N, Liu Y, Paroo Z, Huerta C, Li P, Chen S, Liu Q & Zhang H (2011) Structure of C3PO and mechanism of human RISC activation. *Nat Struct Mol Biol* **18**, 650–657.
27 Tomari Y, Matranga C, Haley B, Martinez N & Zamore PD (2004) A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380.
28 Leuschner PJF, Ameres SL, Kueng S & Martinez J (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep* **7**, 314–320.
29 Schwarz D, Hutvagner G, Du T, Xu Z, Aronin N & Zamore P (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208.
30 Frank F, Sonenberg N & Nagar B (2010) Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822.
31 Noland CL, Ma E & Doudna JA (2011) siRNA repositioning for guide strand selection by human dicer complexes. *Mol Cell* **43**, 110–121.
32 Betancur JG & Tomari Y (2012) Dicer is dispensable for asymmetric RISC loading in mammals. *RNA* **18**, 24–30.
33 Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS & Khvorova A (2004) Rational siRNA design for RNA interference. *Nat Biotechnol* **22**, 326–330.
34 Amarzguioui M & Prydz H (2004) An algorithm for selection of functional siRNA sequences. *Biochem Biophys Res Commun* **316**, 1050–1058.
35 Ui-Tei K, Naito Y, Takahashi F, Haraguchi T, Ohki-Hamazaki H, Juni A, Ueda R & Saigo K (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res* **32**, 936–948.
36 Lu ZJ & Mathews DH (2008) Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res* **36**, 640–647.
37 Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, Shinmi J, Yatsuya H, Qiao S, Takahashi M *et al.* (2007) Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activites. *Nucleic Acids Res* **35**, e123.
38 Walton SP, Wu M, Gredell JA & Chan C (2010) Designing highly active siRNAs for therapeutic applications. *FEBS J* **277**, 4806–4813.
39 Yang X & Chan C (2009) Repression of PKR mediates palmitate-induced apoptosis in HepG2 cells through regulation of Bcl-2. *Cell Res* **19**, 469–486.
40 Laraki G, Clerzius G, Daher A, Melendez-Pena C, Daniels S & Gatignol A (2008) Interactions between the double-stranded RNA-binding proteins TRBP and

PACT define the Medipal domain that mediates protein–protein interactions. *RNA Biol* **5**, 92–103.
41 Birmingham A, Anderson E, Sullivan K, Reynolds A, Boese Q, Leake D, Karpilow J & Khvorova A (2007) A protocol for designing siRNAs with high functionality and specificity. *Nat Protoc* **2**, 2068–2078.
42 Mysara M, Garibaldi JM & El Hefnawi M (2011) MysiRNA-Designer: a workflow for efficient siRNA design. *PLoS ONE* **6**, e25642.
43 Huesken D, Lange J, Mickanin C, Weiler J, Asselbergs F, Warner J, Meloon B, Engel S, Rosenberg A, Cohen D *et al*. (2005) Design of a genome-wide siRNA library using an artificial neural network. *Nat Biotechnol* **23**, 995–1001.
44 Shabalina SA, Spiridonov AN & Ogurtsov AY (2006) Computational models with thermodynamic and composition features improve siRNA design. *BMC Bioinformatics* **7**, 65.
45 Yuan B, Latek R, Hossbach M, Tuschl T & Lewitter F (2004) siRNA selection server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Res* **32**, W130–W134.
46 Takasaki S (2009) Selecting effective siRNA target sequences by using Bayes' theorem. *Comput Biol Chem* **33**, 368–372.
47 Ladunga I (2006) More complete gene silencing by fewer siRNAs: transparent optimized design and biophysical signature. *Nucleic Acids Res* **35**, 433–440.
48 Matveeva OV, Nechipurenko YD, Rossi L, Moore B, Saetrom P, Ogurtsov AY, Atkins JF & Shabalina SA (2007) Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. *Nucleic Acids Res* **35**, e63.
49 Khvorova A, Reynolds A & Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216.
50 Overhoff M, Alken M, Far RK-K, Lemaitre M, Lebleu B, Sczakiel G & Robbins I (2005) Local RNA target structure influences siRNA efficacy: a systematic global analysis. *J Mol Biol* **348**, 871–881.
51 Shao Y, Chan CY, Maliyekkel A, Lawrence CE, Roninson IB & Ding Y (2007) Effect of target secondary structure on RNAi efficiency. *RNA* **13**, 1631–1640.
52 Seitz H, Tushir J & Zamore P (2011) A 5′-uridine amplifies miRNA/miRNA* asymmetry in Drosophila by promoting RNA-induced silencing complex formation. *Silence* **2011**, 2:4.
53 Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C & Turner DH (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* **37**, 14719–14735.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Doc. S1. The supporting information contains the following: details on the logistic regression model algorithm, EGFP and PKR target sequence information, Dharmacon and Ambion siRNA selection outputs, EGFP toxicity data, and PKR western blot images.

Fig. S1. Images and cytotoxicity data for siRNA transfection experiments.

Fig. S2. Western blots for siRNA silencing.