

Functional polarity is introduced by Dicer processing of short substrate RNAs

Scott D. Rose, Dong-Ho Kim¹, Mohammed Amarzguioui¹, Jeremy D. Heidel², Michael A. Collingwood, Mark E. Davis², John J. Rossi^{1,3} and Mark A. Behlke*

Integrated DNA Technologies, Inc. 1710 Commercial Park, Coralville, IA 52241, USA, ¹Division of Molecular Biology, Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010-3011, USA, ²Chemical Engineering, California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA and ³Graduate School of Biological Sciences, City of Hope and Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010-3011, USA

Received May 13, 2005; Revised and Accepted July 6, 2005

ABSTRACT

Synthetic RNA duplexes that are substrates for Dicer are potent triggers of RNA interference (RNAi). Blunt 27mer duplexes can be up to 100-fold more potent than traditional 21mer duplexes (1). Not all 27mer duplexes show increased potency. Evaluation of the products of *in vitro* dicing reactions using electrospray ionization mass spectrometry reveals that a variety of products can be produced by Dicer cleavage. Use of asymmetric duplexes having a single 2-base 3'-overhang restricts the heterogeneity that results from dicing. Inclusion of DNA residues at the ends of blunt duplexes also limits heterogeneity. Combination of asymmetric 2-base 3'-overhang with 3'-DNA residues on the blunt end result in a duplex form which directs dicing to predictably yield a single primary cleavage product. It is therefore possible to design a 27mer duplex which is processed by Dicer to yield a specific, desired 21mer species. Using this strategy, two different 27mers can be designed that result in the same 21mer after dicing, one where the 3'-overhang resides on the antisense (AS) strand and dicing proceeds to the 'right' ('R') and one where the 3'-overhang resides on the sense (S) strand and dicing proceeds to the 'left' ('L'). Interestingly, the 'R' version of the asymmetric 27mer is generally more potent in reducing target gene levels than the 'L' version 27mer. Strand targeting experiments show asymmetric strand utilization between the two different 27mer forms, with the 'R' form favoring S strand and the 'L' form favoring AS strand silencing. Thus, Dicer

processing confers functional polarity within the RNAi pathway.

INTRODUCTION

RNA interference (RNAi) is a conserved pathway present in most eukaryotes where double-stranded RNA (dsRNA) triggers a series of biochemical events that culminates in sequence-specific suppression of gene expression (2–4). Long dsRNAs have been employed for many years as a means to modulate gene expression in plants (5), yeast (6) and *Caenorhabditis elegans* (7). Similar attempts in mammalian cells failed due to interferon activation. In *Drosophila*, dsRNA of >150 bp length efficiently induces an RNAi response. This response becomes weaker as RNA length decreases and 25–38 bp duplexes are inactive; very short duplexes of 19–23 bp length regain activity (8). The biochemistry is slightly different in mammalian cells and 25–30 bp duplexes will strongly induce an RNAi response, but potency does decrease with length and 45 bp duplexes are mostly inactive (1). *In vivo*, long dsRNAs are cleaved by the RNase III class endoribonuclease Dicer into 21–23 base duplexes having 2-base 3'-overhangs (9,10). These species, called 'small interfering RNAs' (siRNAs), enter the RNA induced silencing complex (RISC) and serve as a sequence-specific guide to target degradation of complementary mRNA species (8).

The discovery of siRNAs permitted RNAi to be used as an experimental tool in higher eukaryotes. Typically, siRNAs are chemically synthesized as 21mers with a central 19 bp duplex region and symmetric 2-base 3'-overhangs on the termini. These duplexes are transfected into cells lines, directly mimicking the products made by Dicer *in vivo*. Most siRNA sequences can be administered to cultured cells or to animals without eliciting an interferon response (11–13). There are

*To whom correspondence should be addressed. Tel: +1 319 626 8432; Fax: +1 319 626 8444; Email: mbehlke@idtdna.com

some reports that particular motifs can induce such a response when delivered via lipids (13–15), although a cyclodextrin-containing polycation system has been shown to deliver siRNA containing one such putative immunostimulatory motif that achieves target gene down-regulation in mice without triggering an interferon response (16), even in a disseminated tumor model.

RNAi has rapidly become the favored method to knock down single genes for detailed study or hundreds to thousands of genes in high-throughput functional genomics surveys. The potential of 21mer siRNAs for use as therapeutic agents to reduce activity of specific gene products is also receiving considerable attention and successful knockdown of gene expression in mice has already been demonstrated by several groups (17–20).

We recently described that chemically synthesized RNA duplexes of 25–30 base length can have as much as a 100-fold increase in potency compared with 21mers at the same location. At the site most extensively examined in this study, EGFPS1, only minor differences in potency were seen between duplexes with blunt, 3'-overhang or 5'-overhang ends, and a blunt 27mer duplex was most potent (1). Increased potency has similarly been described for 29mer stem short hairpin RNAs (shRNAs) when compared with 19mer stem hairpins (21). While the primary function of Dicer is generally thought to be cleavage of long substrate dsRNAs into short siRNA products, Dicer also introduces the cleaved siRNA duplexes into nascent RISC in *Drosophila* (22–24). Dicer is involved in RISC assembly and is itself part of the pre-RISC complex (25). The observed increased potency obtained using longer RNAs in triggering RNAi is theorized to result from providing Dicer with a substrate (27mer) instead of a product (21mer) and that this improves the rate or efficiency of entry of the siRNA duplex into RISC.

Unfortunately, not all 27mers show this kind of increased potency. It is well known that shifting a 21mer siRNA by a few bases along the mRNA sequence can change its potency by 10-fold or more (26–28). Different products that result from dicing can have different functional potency, and control of the dicing reaction may be necessary to best utilize Dicer-substrate RNAs in RNAi. The EGFPS1 blunt 27mer studied in Kim *et al.* (1) is diced into two distinct 21mers. Vermeulen and colleagues reported studies where synthetic 61mer duplex RNAs were digested using recombinant human Dicer *in vitro* and examined for cut sites using a ³²P-end-labeled gel assay system. Heterogeneous cleavage patterns were observed and the presence of blunt versus 3'-overhang ends altered precise cleavage sites (29). Dicing patterns for short 25–30mer RNA substrates have not been published and processing rules that enable accurate prediction of these patterns do not exist. We therefore studied dicing patterns at a variety of sites using different duplex designs to see if cleavage products could be predicted. We find that a wide variety of dicing patterns can result from blunt 27mer duplexes. An asymmetric duplex having a single 2-base 3'-overhang generally has a more predictable and limited dicing pattern where a major cleavage site is located 21–22 bases from the overhang. Including DNA residues at the 3' end of the blunt side of an asymmetric duplex further limits heterogeneity in dicing patterns and makes it possible to design 27mer duplexes that result in predictable products after dicing.

We find that position of the 3'-overhang influences potency and asymmetric duplexes having a 3'-overhang on the antisense strand are generally more potent than those with the 3'-overhang on the sense strand. This can be attributed to asymmetrical strand loading into RISC, as the opposite efficacy patterns are observed when targeting the antisense transcript. Novel designs described here that incorporate a combination of asymmetric 3'-overhang with DNA residues in the blunt end offer a reliable approach to design Dicer-substrate RNA duplexes for use in RNAi applications.

MATERIALS AND METHODS

Chemically synthesized siRNAs

All RNA oligonucleotides described in this study were synthesized and purified using high-performance liquid chromatography (HPLC) at Integrated DNA Technologies (Coralville, IA). All oligonucleotides were examined by electrospray ionization mass spectrometry (ESI-MS) and were within $\pm 0.02\%$ predicted mass and were further examined by capillary electrophoresis and were $>90\%$ molar purity. Final duplexes were prepared in sodium salt form. Duplexes are named by site (EGFPS2 is Site 2 in enhanced green fluorescent protein) and by strand length. EGFPS2 27/27 is a blunt 27mer RNA duplex. EGFPS2 25/27 has a 25 base top (sense, 'S') strand and 27 base bottom (antisense, 'AS') strand with a 2-base 3'-overhang. EGFPS2 27/25 has a 27 base sense strand with a 2-base 3'-overhang and a 25 base antisense strand. DNA bases have been substituted at various locations and inclusion is indicated as 'D' in compound names and DNA residues are identified using bold lower case letters in sequence; 'p' represents 5'-phosphate. Starting with a 21mer sequence, an 'R' 27mer has added bases extending to the right side of this sequence (3' with respect to the target) and an 'L' 27mer has added bases extending to the left side of this sequence (5' with respect to the target). All RNA duplexes used in this study are listed in Table S1 (online Supplementary Material).

In vitro dicing assays

RNA duplexes (100 pmol) were incubated in 20 μ l of 20 mM Tris, pH 8.0, 200 mM NaCl, 2.5 mM MgCl₂ with or without 1 U of recombinant human Dicer (Stratagene, La Jolla, CA) at 37°C for 24 h. Samples were desalted using a Performa SR 96 well plate (Edge Biosystems, Gaithersburg, MD). Electrospray ionization liquid chromatography mass spectroscopy (ESI-LC-MS) of duplex RNAs pre- and post-treatment with Dicer were done using an Oligo HTCS system (Novatia, Princeton, NJ) (30), which consisted of ThermoFinnigan TSQ7000, Xcalibur data system, ProMass data processing software and Paradigm MS4™ HPLC (Michrom BioResources, Auburn, CA). The liquid chromatography step employed before injection into the mass spectrometer (LC-MS) removes most of the cations complexed with the nucleic acids; some sodium ion can remain bound to the RNA and are visualized as minor +22 or +44 species, which is the net mass gain seen with substitution of sodium for hydrogen. All dicing experiments were performed at least twice.

EGFP RNAi assays

HEK 293 cells were split in 24-well plates to 60% confluency in DMEM media one day prior to transfection. The indicated amounts of reporter and internal control DNAs and siRNAs were diluted in 50 μ l of Opti-MEM I (Invitrogen, Carlsbad, CA) and mixed with 50 μ l of Opti-MEM I diluted Lipofectamine 2000 (1 μ l per well) (Invitrogen). After incubation at room temperature for 20–30 min, the complexes were added to cells in 0.4 ml of DMEM media. To normalize for transfection efficiency, either firefly luciferase or red fluorescent protein (RFP) reporter plasmids were included as internal controls. For the luciferase assay, the Steady Glo Luciferase Assay Kit (Promega, Madison, WI) was used according to manufacturer's instruction. For RFP co-transfection, the indicated amount of EGFP reporter plasmid (pLEGFP-C1 vector, Clontech, Palo Alto, CA) was co-transfected with 20 ng of RFP reporter plasmid (pDsRed2-C1, BD Sciences, Franklin Lakes, NJ). After 24 h, RFP expression levels were monitored by fluorescence microscopy. Only experiments where transfection efficiencies were >90% (as assessed by RFP expression) were evaluated. Levels of EGFP expression were measured 24 h later. EGFP expression was determined either from the median number of EGFP-fluorescent cells determined by FACS (live cells) or by fluorometer readings (cell extracts). All transfections were minimally performed in duplicate and data averaged. Assays done in triplicate (or more) include error bars in data reporting.

EGFP reporter assays of hnRNPH RNAi

To facilitate throughput, efficacy of hnRNPH-specific duplexes were assayed using a synthetic reporter system where the complete coding region of the hnRNPH gene was cloned into the XhoI and BamHI sites of the EGFP gene in the expression vector pLEGFP-C1 (Clontech, Palo Alto, CA) to make an EGFP-hnRNPH fusion protein. A PCR product was made from a human hnRNPH cDNA clone using the XhoI containing forward primer 5'-ACGCAGAACTCGAGTGTCTA-3' and the BamHI containing reverse primer 5'-TCACTGCTCC-TAGGTTACCT-3'. The resulting PCR product was digested using XhoI and BamHI, and cloned into pLEGFP-C1 that had been similarly digested. Products were verified by DNA sequencing. The fusion protein reporter construct was used to directly measure activity of anti-hnRNPH reagents by change in EGFP fluorescence (31). Transfection and EGFP assays were performed as described above.

Firefly luciferase RNAi assays

HeLa cells were seeded at 5e4 cells per well in 24-well plates 24 h prior to transfection in DMEM (Mediatech, Herndon, VA) containing 10% fetal bovine serum (Invitrogen). In each well, 100 μ l of OptiMEM I containing 2.5 μ l Lipofectamine was added to an equal volume of OptiMEM I containing 1 μ g pGL3-Control Vector (Promega), and the indicated duplex at either 20, 2 or 0.4 nM and incubated at room temperature for ~30 min prior to transfection. Plates were washed with PBS and incubated for 4 h at 37°C with lipoplex solution. Lipoplex solutions were replaced with 1 ml complete growth medium. At 48 h post-transfection, growth medium was replaced with 100 μ l 1 \times Cell Culture Lysis Reagent (Promega). Plates were incubated at room temperature with gentle shaking

for ~1 h to allow complete cell lysis. The luminescence of cell lysates was determined using a Monolight 3010 luminometer (BD Pharmingen, San Diego, CA) and the Luciferase assay system (Promega). Substrate solution (100 μ l) was added to 10 μ l cell lysate, and the resulting luminescent signal was integrated over 10 s. Results are presented in relative light units (RLU) as a percentage of the average signal for plasmid-only samples on the same plate. Mean and SD values for triplicate wells are presented for all treatments.

La RNAi assays

RNA duplexes specific for the human La antigen (NM_003142) were transfected into HEK293 cells in 6-well plates at 30% confluency as described above. Cellular extracts were prepared 72 h post-transfection. Western blots were performed as described previously (1). The anti-Enolase antibody was obtained from Biogenesis (Kingston, NH) and the anti-La antibody was kindly provided by Dr Ger Pruijn (Katholieke Universiteit Nijmegen, Netherlands).

HeLa cells were split in 24-well plates at 35% confluency and were transfected the next day with Oligofectamine (Invitrogen) using 1 μ l per 65 μ l OptiMEM I with RNA duplexes at the indicated concentrations. All transfections were performed in triplicate. RNA was harvested at 24 h post-transfection using SV96 Total RNA Isolation Kit (Promega). RNA was checked for quality using a Bioanalyzer 2100 (Agilent, Palo Alto, CA) and cDNA was prepared using 500 ng total RNA with SuperScript-II reverse transcriptase (Invitrogen) per manufacturer's instructions using both oligo-dT and random hexamer priming. Real-time PCRs were done using an estimated 33 ng cDNA per 25 μ l reaction using Immolase DNA Polymerase (Bioline, Randolph, MA) and 200 nM primers and probe. La-specific primers were La-For 5'-GACCAACA-AGAATCCCTAAACA, La-Rev 5'-CTTGCCCTGAACTGTACTT and probe La-P 5'-FAM-AAGGGTAATAAAGC-TGCCAGCCTGGGT-IowaBlackFQ. Cycling conditions employed were as follows: 50°C for 2 min and 95°C for 10 min followed by 40 cycles of 2-step PCR with 95°C for 15 s and 60°C for 1 min. PCR and fluorescence measurements were done using an ABI Prism™ 7000 Sequence Detector (Applied Biosystems Inc., Foster City, CA). All data points were performed in triplicate. Expression data was normalized to internal control human acidic ribosomal phosphoprotein P0 (RPLP0) (NM_001002) levels which were measured in separate wells in parallel using primers RPLP0-For 5'-GGCGACCTGGAAGTCCAACCT, RPLP0-Rev 5'-CCAT-CAGCACCACAGCCTTC, and probe RPLP0-P 5'-FAM-ATCTGCTGCATCTGCTTGGAGCCCA-IowaBlackFQ (32).

Luciferase reporter vectors with EGFP and hnRNPH, S versus AS targeting

A PCR generated fragment of the EGFP coding region spanning sites EGFPs1 and EGFPs2 was cloned into the XhoI site located in the 3'-untranslated region (3'-UTR) of the humanized Renilla luciferase gene of plasmid psiCHECK™-2 (Promega). Primers containing an XhoI restriction site and EGFP nucleotides 67–85 (For, 5'-TTTCTCGAGTAAACGG-CCACAAGTTCA-3') and 291–311 (Rev, 5'-TTTCTCGAG-TCGTCCTTGAAGAAGATGGTG-3') were used to generate a 245 bp EGFP PCR product. The PCR product was digested

with XhoI and the fragment was cloned into a unique XhoI site located in the psiCHECK™-2 vector. Clones with both orientations (S and AS) were obtained and verified by DNA sequencing.

A PCR generated fragment of the hnRNPH coding region spanning sites H1 and H3 was similarly cloned into the XhoI site located in the 3'-UTR of the humanized Renilla luciferase gene of plasmid psiCHECK™-2. The fragment was 343 bp long including the region 90–432 of reference sequence NM_005520. Clones with both orientations (S and AS) were obtained. Maps of the 'S' and 'AS' psiCHECK-2 reporter vectors are shown in Figure S2 in the Supplementary Material.

HEK293 cells were transfected with 150 ng of reporter vectors Luc-EGFP-'S' or Luc-EGFP-'AS' with the indicated amounts of EGFP2 or control duplex RNAs as described above. HCT116 cells were transfected with 100 ng of reporter vectors Luc-hnRNPH-'S' or Luc-hnRNPH-'AS' with the indicated amounts of H3 or control duplex RNAs. Luciferase assays were performed 24 h post-transfection. Changes in expression of Renilla luciferase (target) were calculated relative to firefly luciferase (internal control).

RESULTS

Dicing patterns for EGFP RNA duplexes

The products that result from *in vitro* digestion of various substrate RNA duplexes with recombinant human Dicer were visualized using ESI-MS (hereafter referred to as the 'ESI-dicing assay'). A blunt 27mer derived from EGFP Site-1 (EGFPS1) was previously shown to produce two primary 21mer cleavage products by ESI-dicing (1). RNA duplexes derived from EGFP Site 2 (EGFPS2) sequence are studied here in greater detail and are shown in Figure 1 and Figure S1 (online Supplementary Material). In general, *in vitro* dicing of 27mer duplexes using recombinant human Dicer results in a heterogeneous set of products, most of which are 21mer or 22mer species with 5'-phosphate. More rarely, 20mer and 23mer species are also seen; these species are usually inconsistent between repetitions. If a cleavage product includes the 5' end from the original chemically synthesized substrate duplex (i.e. no enzymatic cleavage event was needed to produce that end), the 5' end of the diced product remained identical to the substrate (5'-phosphate or 5'-OH). 5' ends resulting from internal cleavage events show mass values consistent with the presence of 5'-phosphate. This spectrum of diced products is consistent with the pattern seen by Elbashir in a collection of sequenced clones derived from diced longer dsRNAs (8). Peaks seen in a mass spectrometry trace can sometimes be identified to represent a single, unambiguous, unique sequence. In other cases, it is not possible to unambiguously identify a specific sequence based solely upon mass if more than one 21mer with same base composition (and therefore the same mass) could be produced from cleavage of the substrate 27mer. In deconvolution of the mass spectra data, we have assigned duplex identity to unique species wherever possible and otherwise have identified products which are consistent with 21mer or 22mer duplexes with 2-base 3'-overhangs (instead of other combinations that would yield blunt duplexes or 1-base overhangs, etc.).

ESI-dicing of the blunt 27mer substrate duplex EGFPS2 R 27/27 shows a variety of mass peaks which are consistent with five duplex products, shown in Figure 1A. Using other blunt 27mer duplex sequences (from other sites), we have seen 2, 3, 4 or more duplex products result from dicing (data not shown). Elbashir reported *in vitro* dicing of 39, 52 and 111 bp dsRNA substrates using *Drosophila* extracts. Products were cloned and a heterogeneous collection of 21–22 base fragments were identified by DNA sequencing. While dicing appeared to generally initiate from one or both ends, no single product or subset of products was dominant (8). We cannot identify any obvious pattern that allows prediction of what specific 21mer(s) will result from dicing a given blunt 27mer, making rational design of 27mer RNAi reagents difficult. We therefore tested a series of design variants to see if this heterogeneity could be reduced with the goal of finding a design for which cleavage patterns could be reliably predicted.

Modification of one end of a blunt 27mer duplex with fluorescein inhibits dicing. Modification of both 3' ends with fluorescein completely blocks dicing (1). We tested if substitution of three DNA residues at the 5' or 3' end of either strand of a blunt 27mer duplex affected dicing patterns. Results are shown in Figure S1 (online Supplementary Material). DNA residues cannot be cleaved by Dicer and therefore directly alter cleavage patterns. Further, the DNA residues seem to reduce the likelihood of Dicer binding that end (assuming that Dicer binds the RNA duplex via the PAZ domain and then cuts 21–22 bases away) (33), especially when positioned on the 3' end. Curiously, EGFPS2 R 27/27(3'D), the duplex with DNA residues positioned at the 3' end of the antisense strand, was not diced to produce any 21–22mer cleavage product (identical results were obtained in three attempts). While the use of terminal DNA residues can change and somewhat simplify dicing patterns, this approach alone is insufficient to truly direct dicing to predictable patterns.

Naturally occurring substrates for Dicer include microRNAs (miRNAs). These species typically originate in the nucleus as long primary transcripts where they are processed to ~70mer pre-miRNAs by the RNase III class endonuclease Drosha (34,35). Drosha products are shRNAs, which structurally look like an asymmetric duplex, having a single 2-base 3'-overhang on one side and a hairpin loop on the other side. These species are exported to the cytoplasm (36) where final processing by Dicer takes place (37). The 3'-overhang may be important for Dicer processing. The presence of a single 2-base 3'-overhang has been shown to help direct 'correct' dicing of synthetic shRNAs (21) and alters dicing patterns of synthetic linear 61mer duplexes (29).

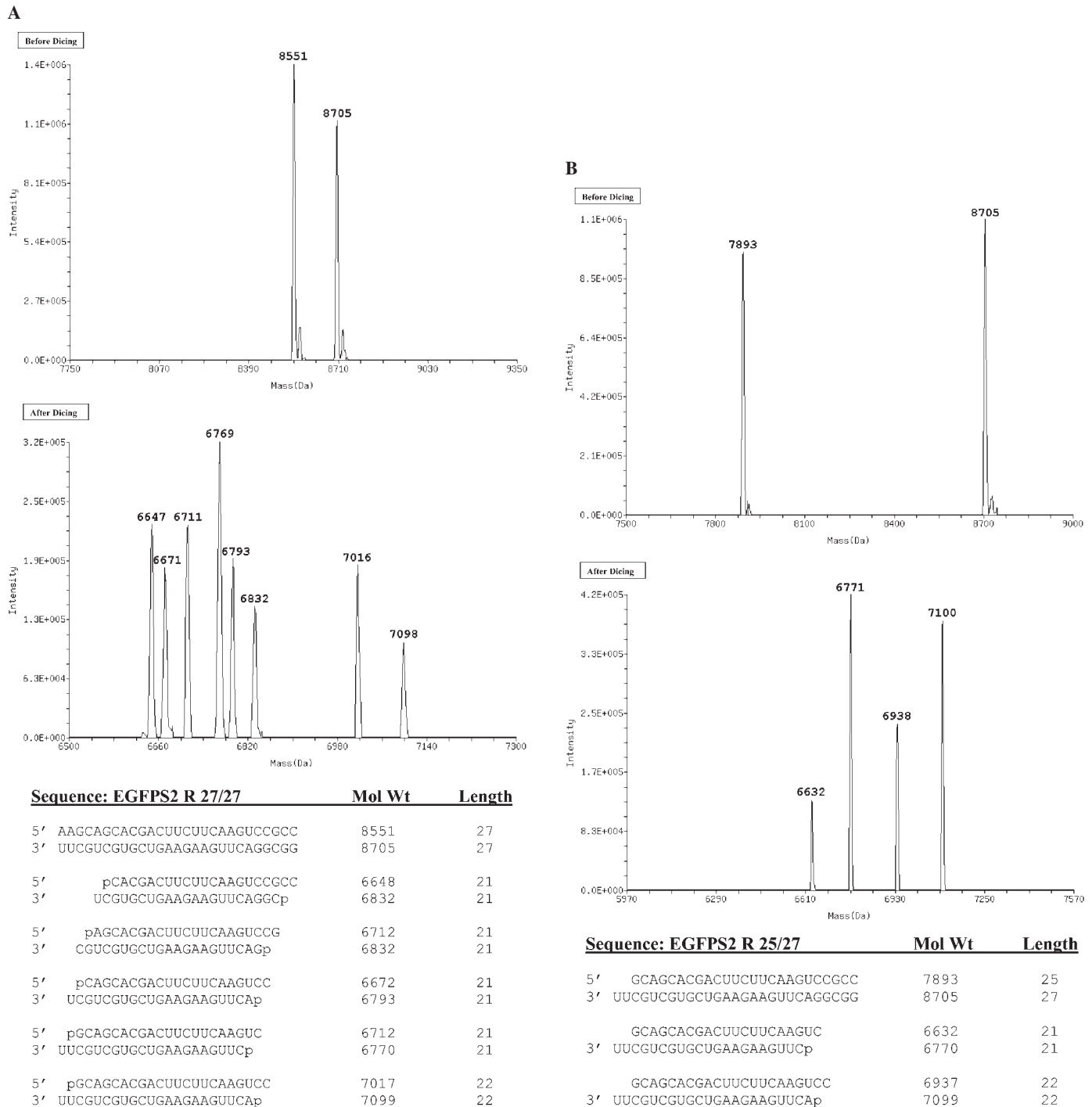
Asymmetric duplexes having one 2-base 3'-overhang and one blunt end were studied using the ESI-dicing assay. Duplex EGFPS2 R 25/27 has two bases removed from the 5' end of the sense strand compared with blunt duplex EGFPS2 R 27/27, providing a single 2-base 3'-overhang. This duplex showed a much simplified dicing pattern, Figure 1B. Only one 21mer and one 22mer duplex were produced. The observed pattern is consistent with a model where Dicer binds the 3'-overhang of the substrate RNA and cleavage takes place 21–22 bases distant from this site (33). In fact, this pattern could be considered a single event as dicing generates a single pair of related duplexes as 'the product' made

from digestion of the substrate duplex from a single unique binding or start site.

The presence of a 3'-overhang does not always restrict dicing to a simple pattern. Duplex EGFPS2 R 27/25 has two bases removed from the 5' end of the antisense strand compared with blunt duplex EGFPS2 R 27/27. This duplex showed a complex dicing pattern (Figure 1C) which is different from both the parent blunt duplex (Figure 1A) and the asymmetric duplex with the 3'-overhang on the opposite end (Figure 1B). Changing the bases present on the

3'-overhang from 'CC' to 'GG' did not reduce the complexity of dicing pattern seen for this duplex (data not shown).

The two approaches of modifying dicing patterns were combined and DNA residues were placed at the 3' end (blunt end) of the antisense strand of asymmetric duplex EGFPS2 R 27/25, resulting in duplex EGFPS2 R 27/25D. This duplex gave a simplified dicing pattern with only a single 21mer and a single 22mer species (Figure 1D). This new design strategy places an element on one end that is generally favorable for Dicer binding (2-base 3'-overhang) and an element on the other



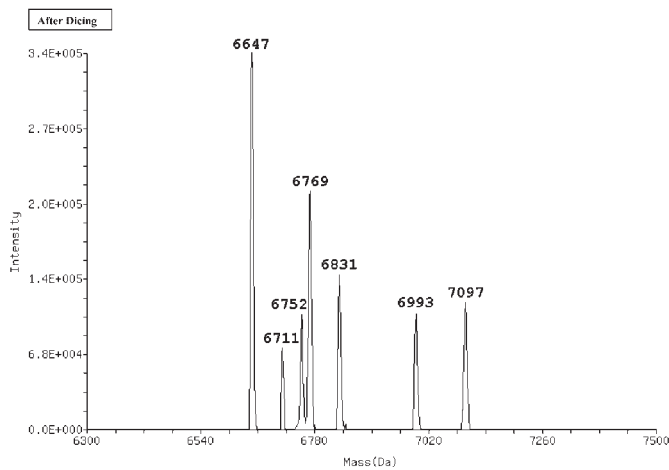
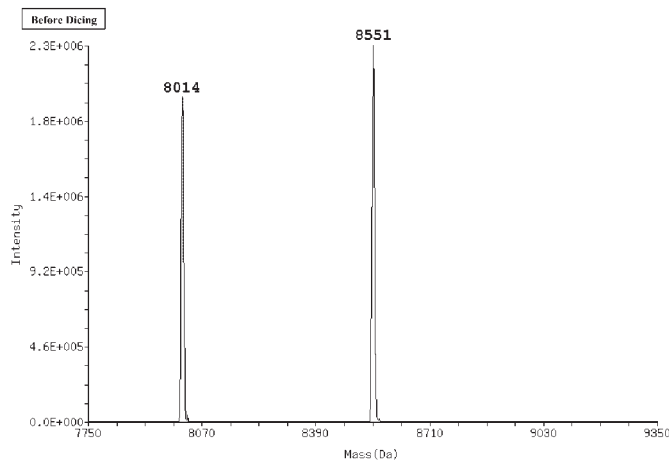
end that is generally unfavorable for Dicer binding (blunt end with 3'-DNA residues). This type of duplex appears to offer Dicer a single favorable binding site and would be predicted to result in a single 21mer product.

Dicing patterns from the new design having a 2-base 3'-overhang on one end and blunt with 3'-DNA residues on the other end were examined using additional duplexes. Duplex EGFPS2 R 25D/27 has a 3'-overhang on the antisense strand and two DNA bases at the 3' end (blunt end) of the sense strand. Duplex EGFPS2 L 27/25D has the opposite design with a 3'-overhang on the sense strand and two DNA bases at the 3' end (blunt end) of the antisense strand. If the design strategy works as predicted, these two different duplexes

should both be diced into the same 21mer product. EGFPS2 R 25D/27 diced into the expected 21mer/22mer pair with cleavage of 21–22 bases from the 3'-overhang (Figure 1E). Similarly, EGFPS2 L 27/25D diced into the expected 21mer/22mer pair with cleavage 21–22 bases from the 3'-overhang (Figure 1F). Thus, the same 21mer duplex was produced from two different substrate RNAs. Note, however, that the sister product (22mer) is necessarily different for each substrate since the single base addition occurs on opposite sides of the 21mer.

To confirm that this cleavage pattern holds at other sites and is truly predictable, a similar related pair of asymmetric 27mers were tested at EGFP Site-1 (Supplementary Table S1).

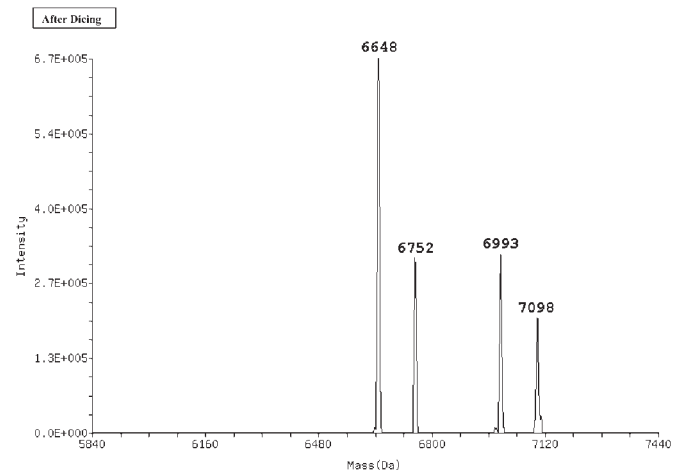
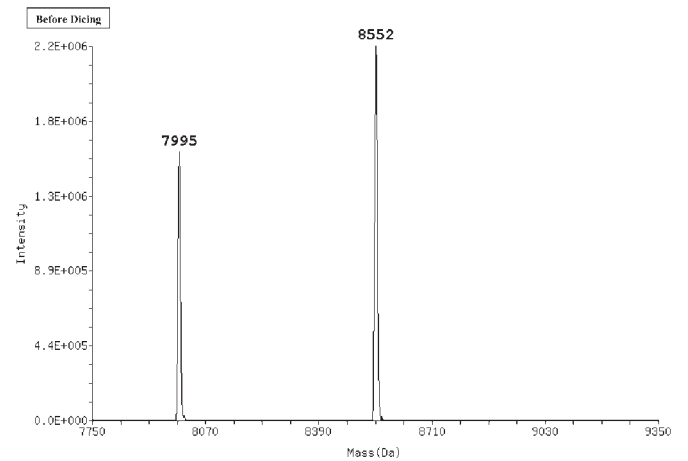
C



Sequence: EGFPS2 R 27/25 **Mol Wt** **Length**

5'	AAGCAGCAGACUUCUUAAGUCCGCC	8551	27
3'	UUCGUCGUGCUGAAGAAGUUCAGGC	8014	25
5'	pCAGCAGACUUCUUAAGUCCGCC	6648	21
3'	UCGUCGUGAAGAAGUUCAGGC	6752	21
5'	pAGCAGCAGACUUCUUAAGUCCG	6712	21
3'	CGUCGUGCUGAAGAAGUUCAGp	6832	21
5'	pGCAGCAGCAGACUUCUUAAGUC	6712	21
3'	UUCGUCGUGCUGAAGAAGUUCp	6770	21
5'	pGCAGCAGACUUCUUAAGUCCGCC	6993	22
3'	GUCGUCGUGAAGAAGUUCAGGC	7093	22

D



Sequence: EGFPS2 R 27/25(3'D) **Mol Wt** **Length**

5'	AAGCAGCAGACUUCUUAAGUCCGCC	8551	27
3'	ttc GUCGUGCUGAAGAAGUUCAGGC	7995	25
5'	pCAGCAGACUUCUUAAGUCCGCC	6648	21
3'	UCGUCGUGAAGAAGUUCAGGC	6752	21
5'	pGCAGCAGACUUCUUAAGUCCGCC	6993	22
3'	GUCGUCGUGAAGAAGUUCAGGC	7097	22

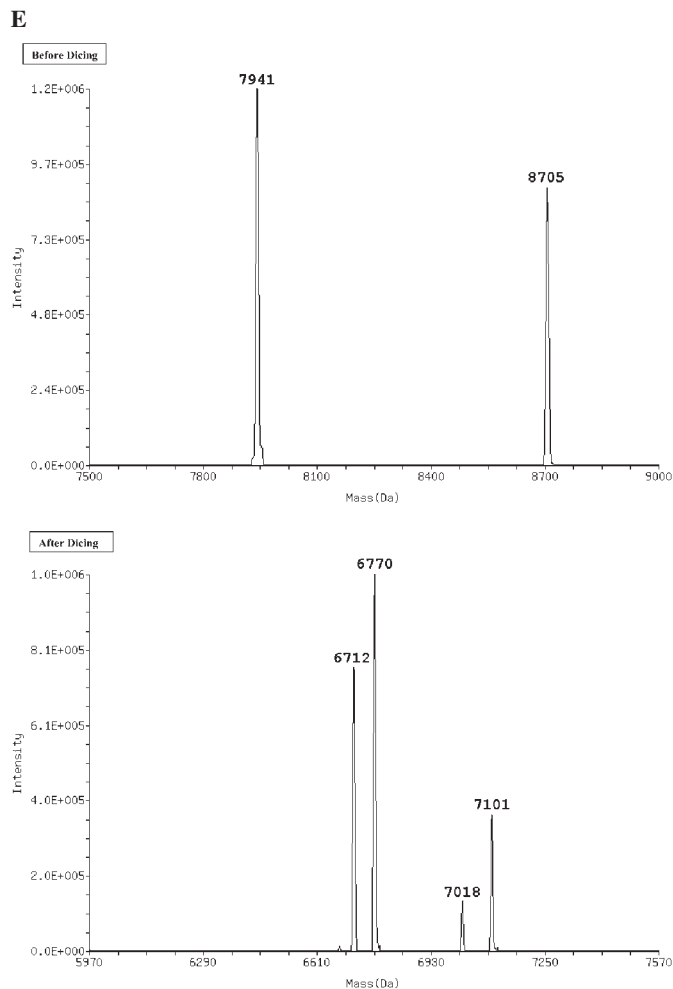
EGFPS1 R 25D/27 diced into the expected 21mer duplex with cleavage 21 bases from the 3'-overhang (Figure 1G). In this instance, no 22mer was seen. Similarly, EGFPS1 L 27/25D diced into the expected 21mer/22mer pair with cleavage of 21–22 bases from the 3'-overhang (Figure 1H). Again, the same 21mer duplex was produced from two different 27mer substrate RNAs.

Functional potency is different for 'R' versus 'L' form RNA duplexes

EGFPS2 duplexes were co-transfected into HEK293 cells with an EGFP expression plasmid. The blunt 27mer duplex (EGFPS2 R 27/27) was more potent in reducing EGFP expression levels than the 21mer duplex (EGFPS2 21/21) (1). The L 27/25D asymmetric duplex was slightly less potent than the blunt 27mer while the R 25D/27 asymmetric duplex

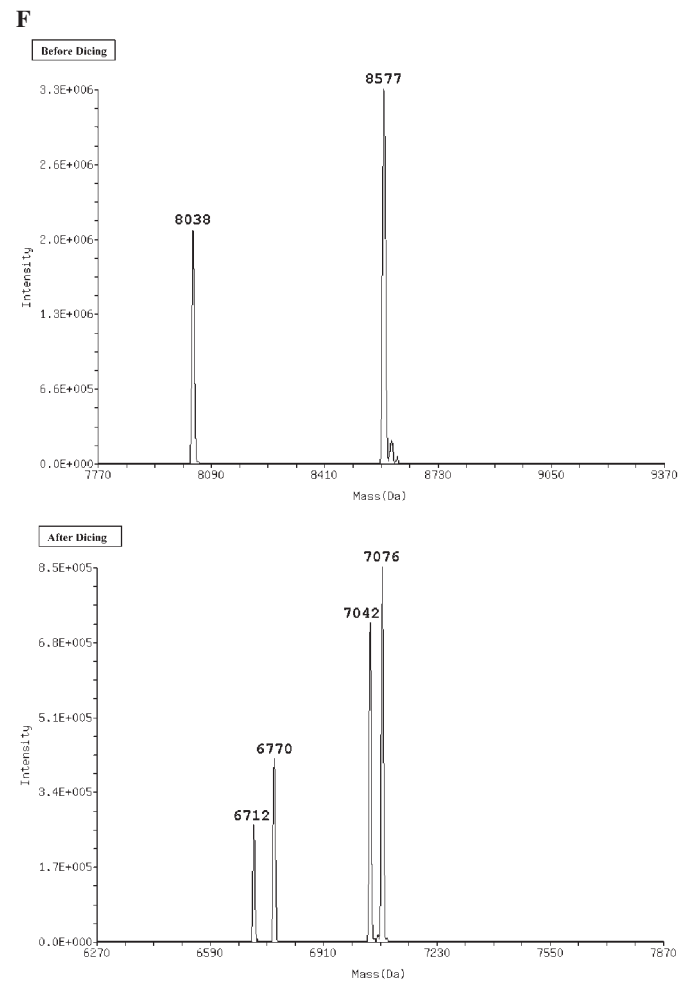
was significantly more potent than any other duplex tested at the EGFPS2 site (Figure 2A). A similarly matched pair of asymmetric 27mer duplexes was studied at EGFP Site-1 (Figure 2B). As was observed previously for the EGFPS2 site, the 'R' asymmetric duplex was significantly more potent than the 'L' asymmetric duplex.

We originally expected that the 'R'-form and 'L'-form duplexes would have similar functional potency since both result in the same antisense 21mer product after dicing (EGFPS2, Figure 1E and F; EGFPS1, Figure 1G and H). The 'L' and 'R' duplexes do produce different 22mer products after dicing. It is possible that the differential potency relates to these 22mer products. To test this possibility directly, RNA duplexes of 22mer length that correspond to the products identified from ESI-dicing of the 'L' and 'R'-form 27mers for both EGFPS1 and EGFPS2 were co-transfected into HEK293 cells with an EGFP expression plasmid. For site



Sequence: EGFPS2 R 25D/27

	Mol Wt	Length
5' pGCAGCAGCAGACUUCUUAAGUCCGcc	7940	25
3' UUCGUCGUGCGAAGAAGUUCAGCGG	8705	27
5' pGCAGCAGCAGACUUCUUAAGUC	6712	21
3' UUCGUCGUGCGAAGAAGUUCp	6770	21
5' pGCAGCAGCAGACUUCUUAAGUCC	7017	22
3' UUCGUCGUGCGAAGAAGUUCAp	7099	22



Sequence: EGFPS2 L 27/25D

	Mol Wt	Length
5' CAUGAAGCAGCAGCAGACUUCUUAAGUC	8576	27
3' gtACUUCGUCGUGCGAAGAAGUUCp	8037	25
5' pGCAGCAGCAGACUUCUUAAGUC	6712	21
3' UUCGUCGUGCGAAGAAGUUCp	6770	21
5' pAGCAGCAGCAGACUUCUUAAGUC	7041	22
3' CUUCGUCGUGCGAAGAAGUUCp	7075	22

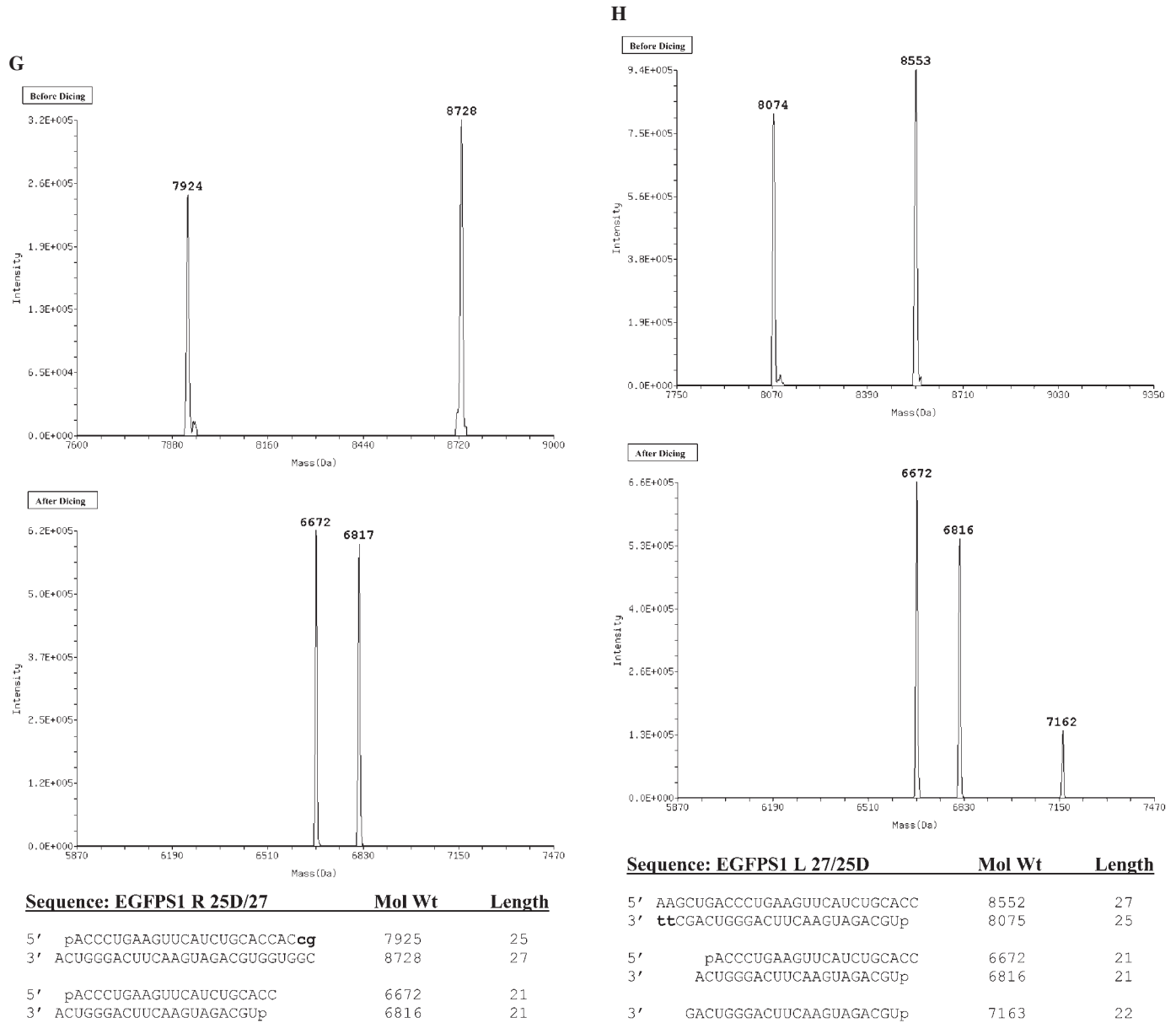
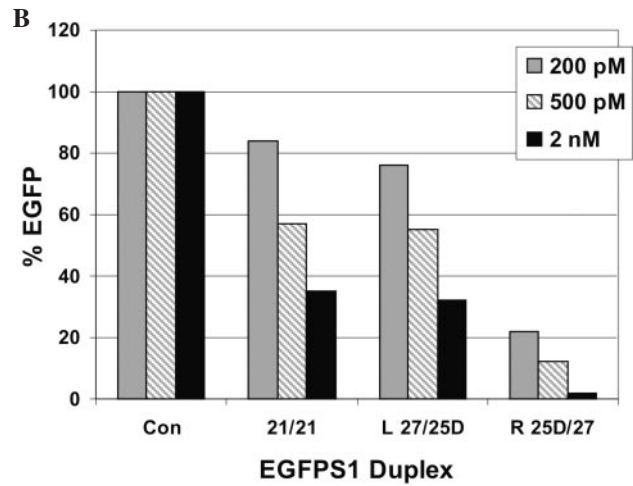
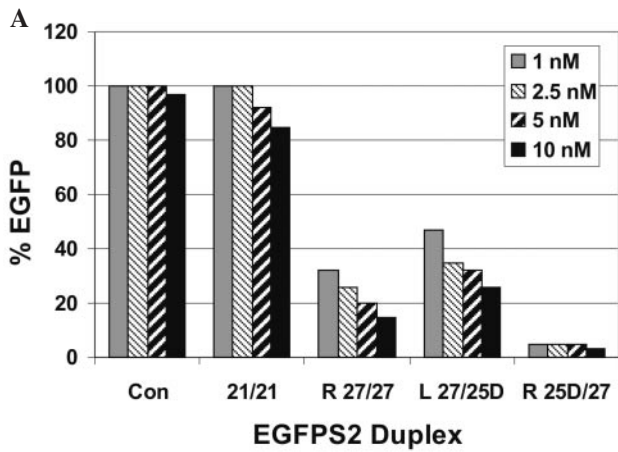


Figure 1. ESI-MS analysis of *in vitro* dicing reactions. Mass spectra of duplexes are shown before (top) and after (middle) digestion with recombinant human Dicer. Sequence of the substrate RNA duplex is provided with detected cleavage products along with calculated molecular weight and strand length (bottom). RNA bases are upper case, DNA bases are lower case bold and 'p' represents 5'-phosphate. (A) Blunt duplex EGFPS2 R 27/27; (B) asymmetric duplex EGFPS2 R 25/27; (C) asymmetric duplex EGFPS2 R 27/25; (D) asymmetric duplex EGFPS2 R 27/25(3'D); (E) asymmetric duplex EGFPS2 R 25D/27; (F) asymmetric duplex EGFPS2 L 27/25D; (G) asymmetric duplex EGFPS1 R 25D/27; (H) asymmetric duplex EGFPS1 L 27/25D.

EGFPS2, the 'R-derived' 22mer was significantly more potent than the 'L-derived' 22mer. This species could contribute to the higher potency observed for the 'R' form EGFPS2 27mer. However, for site EGFPS1 the 'L-derived' 22mer was 3-fold more potent than the 'R-derived' 22mer (Supplementary Figure S3). In this case, 22mer products cannot contribute to the observed higher potency of the 'R' form EGFPS1 27mer. Therefore, although the different 22mer species that result from 27mer dicing can contribute to the ultimate potency of a given duplex, we believe that the generalized difference in potency seen between 'R' and 'L'-form 27mers results from other aspects of Dicer/RISC biochemistry. Vermeulen reported that asymmetric 21mer duplexes with a

single 3'-overhang on the antisense strand were generally more potent than duplexes with a single 3'-overhang on the sense strand (for 3 out of 4 duplexes studied in one gene target) (29). In this case, the duplexes tested were not Dicer substrates.

Dicer-substrate duplexes were studied in other target systems to see if the seeming functional asymmetry between duplexes that are processed to the same 21mer product was reproducible at additional sites. RNA duplexes targeting two adjacent sites in the hnRNPH gene were tested for potency in triggering RNAi. Site H3 and neighboring site H1 (two base 5'-shift) were studied for RNAi-mediated suppression of EGFP fluorescence using a modified EGFP expression



5' ccatgaagcagcagcagacttcttcaagtccgcatgc EGFP Site-2

5' GCAGCAGCACUUCUUAAGUC EGFPs2 21/21
3' UUCGUCGUGCUGAAGAAGUUC

5' AAGCAGCAGCACUUCUUAAGUCCGC EGFPs2 R 27/27
3' UUCGUCGUGCUGAAGAAGUUCAGGCG

5' CAUGAAGCAGCAGCACUUCUUAAGUC EGFPs2 L 27/25D
3' **gt**ACUUCGUGCUGAAGAAGUUCp

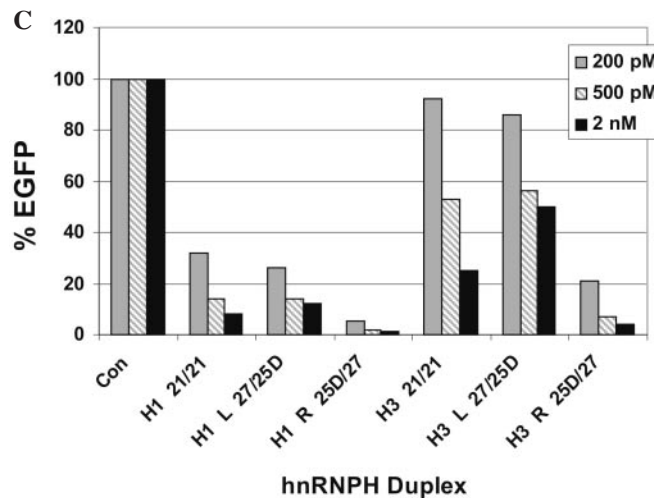
5' pGCAGCAGCACUUCUUAAGUCCGC EGFPs2 R 25D/27
3' UUCGUCGUGCUGAAGAAGUUCAGGCG

5' caagctgaccctgaagttcatctgcaccaccggcaa EGFP Site-1

5' ACCUGAAGUUAUCUGCACC EGFPs1 21/21
3' ACUGGGACUUAAGUAGACGU

5' pACCUGAAGUUAUCUGCACCACcg EGFPs1 R 25D/27
3' ACUGGGACUUAAGUAGACGUGGUGC

5' AAGCUGACCCUGAAGUUAUCUGCACC EGFPs1 L 27/25D
3' **tt**CGACUGGGACUUAAGUAGACGUp



5' ttgttgaacttgaatcagaagatgaagtcaaatggc hnRNPH

5' AACUUGAAUCAGAAGAUGAAG H1 21/21
3' ACUUGAACUUAGUCUUCUACU

5' UUGUUGAACUUGAAUCAGAAGAUGAAG H1 L 27/25D
5' **aa**CAACUUGAACUUAGUCUUCUACUp

5' pAACUUGAAUCAGAAGAUGAAGUCaa H1 R 25D/27
3' ACUUGAACUUAGUCUUCUACUUCAGUU

5' CUUGAAUCAGAAGAUGAAGUC H3 21/21
3' UUGAACUUAGUCUUCUACUUC

5' GUUGAACUUGAAUCAGAAGAUGAAGUC H3 L 27/25D
3' **ca**ACUUGAACUUAGUCUUCUACUUCp

5' pCUUGAAUCAGAAGAUGAAGUCAaat H3 R 25D/27
3' UUGAACUUAGUCUUCUACUUCAGUUUA

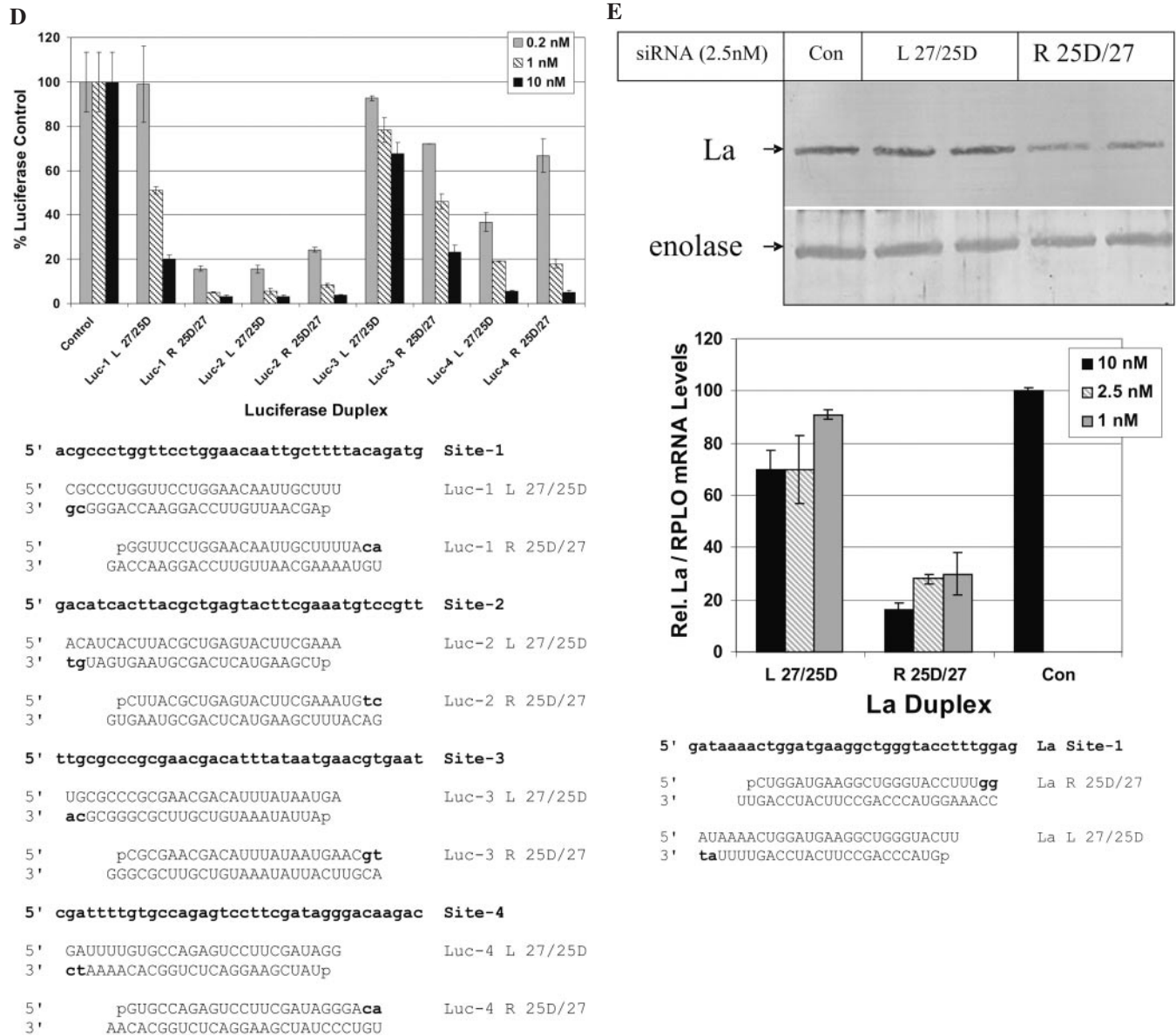


Figure 2. ‘R’ form duplexes are more potent than ‘L’ form duplexes. Top: relative expression data are shown. Bottom: target sequences (S strand) are shown and transfected RNAi duplexes are aligned beneath in duplex form with sense strand top (5’→3’) and antisense strand bottom (3’→5’). RNA bases are upper case, DNA bases are lower case bold, and ‘p’ represents 5’-phosphate. (A) Relative EGFP fluorescence was measured following co-transfection of an EGFP expression plasmid and EGFP2 RNA duplexes; (B) relative EGFP fluorescence was measured following co-transfection of an EGFP expression plasmid and EGFP1 RNA duplexes; (C) relative EGFP fluorescence was measured following co-transfection of an EGFP-hnRNPH fusion vector and hnRNPH RNA duplexes; (D) relative light emission from firefly luciferase was measured following co-transfection of a luciferase expression vector and luciferase RNA duplexes; (E) RNA duplexes specific for La antigen mRNA were transfected into cells at 2.5 nM concentration and protein extracts were prepared 72 h post-transfection. Western blots were done using anti-La antibodies and anti-enolase (control) antibodies. RNA duplexes were transfected into cells at 1, 2.5 and 10 nM concentrations, and RNA was prepared 24 h post-transfection. Quantitative real-time RT-PCR was used to measure relative La mRNA levels, using RPLP0 as an internal normalization standard.

construct that includes the complete hnRNPH coding sequence fused in frame with the C-terminal end of EGFP. RNA duplexes were co-transfected into HEK293 cells with the EGFP-hnRNPH reporter vector and EGFP expression was measured at 48 h (Figure 2C). The ‘L’ duplex H3 L 27/25D showed similar potency to the H3 21mer duplex while the ‘R’ duplex H3 R 25D/27 was markedly more potent. The same trend was observed at site H1. The ‘L’ duplex H1 L 27/25D had similar potency to the H1 21mer duplex while the ‘R’ duplex H1 R 25D/27 was markedly more potent. Thus, at both

these new sites, the ‘R’ version of the asymmetric 27mer duplex was found to be significantly more potent than the ‘L’ version duplex. It is also interesting to note the large difference in potency seen between the H1 and H3 21mers even though these duplexes are shifted by only two bases along the hnRNPH sequence.

Asymmetric ‘R’ and ‘L’ version RNA duplexes targeting four sites in firefly luciferase were compared for potency in triggering RNAi. Duplexes were co-transfected with a luciferase reporter vector into HeLa cells and cell extracts were

examined for suppression of luciferase light emission 48 h post-transfection (Figure 2D). At sites Luc-1 and Luc-3, the 'R' 27mer duplex was more potent than the 'L' 27mer duplex at all doses. At sites Luc-2 and Luc-4, the two forms were of similar potency at 10 and 1 nM, while the 'L' form duplexes were slightly more effective than the 'R' form at 200 pM.

To ensure that the functional polarity observed between 'R' versus 'L' Dicer-substrate RNAs was not an artifact of targeting synthetic reporter constructs, we targeted an endogenous transcript, the La protein encoding mRNA (NM_003142). We previously described that a blunt 27mer Dicer-substrate duplex was more potent at reducing expression of La mRNA than a 21mer duplex at the same site (1). 'R' and 'L' form asymmetric 27mer Dicer-substrate duplexes at this same site and were transfected into HEK293 cells at various concentrations. Protein extracts were prepared 72 h post-transfection and examined for La expression levels by western blot analysis (Figure 2E). The 'L'-sided duplex showed little if any suppression at 2.5 nM dose whereas the 'R'-sided duplex showed appreciable inhibition. The same La duplexes were transfected into HeLa cells and RNA was prepared at 24 h post-transfection. La mRNA levels were assayed using a quantitative real-time RT-PCR assay (Figure 2E). The 'R'-sided duplex was more potent than the 'L'-sided duplex at reducing La mRNA levels at all doses examined. Therefore, the same functional polarity is observed when targeting an endogenous transcript as previously observed using plasmid reporter constructs.

In summary, 'R' duplexes (3'-overhang on the antisense strand) were more active in triggering RNAi than 'L' duplexes (3'-overhang on the sense strand) at 7 out of 9 sites studied here using 27mer Dicer-substrate duplexes and for 3 out of 4 21mer duplexes reported by Vermeulen (29). The impact that end stability can have on 21mer duplex entry into RISC has been well described (28,38–40) and is a cornerstone for current siRNA design rules. New asymmetric terminal overhang rules may offer an additional basis for rational design of siRNAs.

'L' versus 'R' dicing polarity influences strand targeting

It is possible that functional polarity is introduced by the direction that dicing proceeds relative to the sense and antisense strands of the substrate RNA, and that this somehow influences strand entry into RISC. We directly tested this possibility using a synthetic reporter vector system where suppression of both sense and antisense strands target mRNA could be independently assayed. A fragment of the EGFP coding region that spans EGFP Site 2 was cloned into the 3'-UTR of the Renilla luciferase gene in the psiCHECKTM-2 vector in both orientations, producing a sense reporter vector, Luc-EGFP-'S', and an antisense reporter vector, Luc-EGFP-'AS'. Maps of 'S' versus 'AS' reporter constructs are shown in Figure S2 in the Supplementary Material. Inhibition of Luc-EGFP-'S' is an assay for functional loading of the antisense strand of the siRNA into RISC. Conversely, inhibition of Luc-EGFP-'AS' is an assay for functional loading of the sense strand of the siRNA into RISC.

The sense and antisense reporter vectors were co-transfected into HEK293 cells with various concentrations

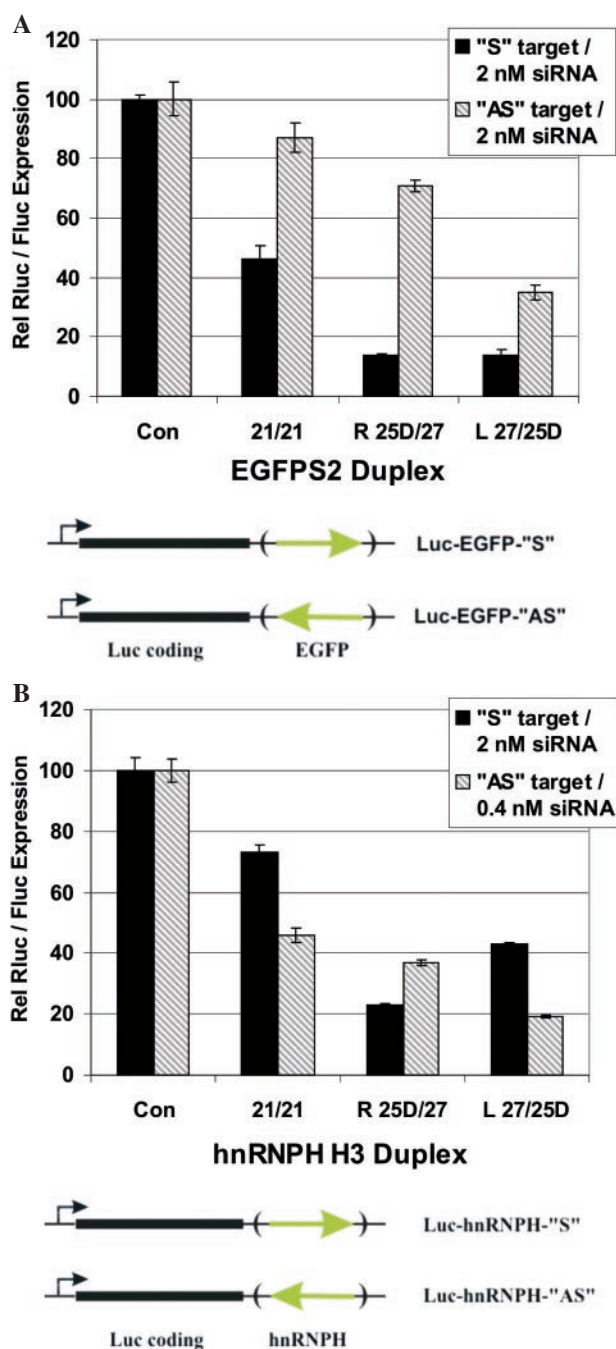


Figure 3. Strand bias is introduced by the direction of Dicer processing. Reporter constructs were made in the psiCheck-2TM vector that contained a fragment of EGFP or hnRNPH coding sequence cloned into the 3'-UTR of the Renilla luciferase gene in both 'S' and 'AS' orientations. Reporter plasmids were co-transfected into cells without (control) or with the indicated RNA duplexes and luciferase activity was measured 24 h post-transfection. Results are reported as the ratio of Renilla luciferase (target) to firefly luciferase (internal control); (A) EGFP strand targeting; (B) hnRNPH strand targeting. Schematic diagrams of the four synthetic reporter constructs employed with orientation of target sequences relative to Renilla luciferase are shown.

of EGFPS2 RNA duplexes and luciferase activity was assayed 24 h post-transfection (Figure 3A). All duplexes showed better inhibition of luciferase expression using the 'S' than the 'AS' orientation EGFP target, as would be expected for a 'good'

functional siRNA site, and both 27mers were more potent than the 21mer duplex. Preferential targeting of the sense reporter was observed for both 'R' and 'L' form duplexes, however relative inhibition of the antisense reporter was substantially better for the 'L' form duplex, indicating relative increased loading of the sense (top) strand of the 'L' form siRNA into RISC. Even though the same 21mer results from dicing (Figure 1E and F), the 'L' and 'R' form duplexes clearly show different patterns of 'S' versus 'AS' strand targeting.

A fragment of the hnRNPH genes that spans Site H3 was similarly cloned into the 3'-UTR of the Renilla luciferase gene in the psiCHECKTM-2 vector in both orientations (Supplementary Figure S2). Reporter plasmids were co-transfected into HCT116 cells with various concentrations of H3 RNA duplexes and luciferase activity was assayed 24 h post-transfection (Figure 3B). In this case, the 21mer duplex demonstrated preferential targeting of the antisense target, indicative of a 'poor' siRNA site. Both 27mers again showed improved potency relative to the 21mer duplex. The antisense strand was more efficiently suppressed relative to sense strand for both 'R' and 'L' form 27mers, however the relative ratio of strand targeting was markedly different between forms, with the 'L' form duplex being significantly more potent for antisense strand suppression.

Asymmetric Dicer-substrate duplexes having a 2-base 3'-overhang on the antisense strand (with Dicer processing proceeding to the right) generally show greater potency than duplexes having the 3'-overhang on the sense strand (with Dicer processing proceeding to the left). This asymmetry appears to result from a relative advantage of the antisense strand for incorporation into RISC from the 'R' form while the 'L' form favors sense strand incorporation. In addition to conferring greater functional potency, these findings also suggest that off-target effects arising from the sense strand may be reduced by using 'R' form duplexes.

Structural features that influence potency of Dicer-substrate RNAs

The EGFP-hnRNPH assay system was used to study additional aspects of duplex design. DNA residues were found to be less favorable than RNA for Dicer processing of blunt 3' ends and altered dicing patterns (Supplementary Figure S1). DNA residues were substituted for RNA in the 2-base 3'-overhang of the asymmetric 'R' version duplexes at both the H1 and H3 sites. These duplexes were co-transfected with the EGFP-hnRNPH reporter vector and EGFP expression was measured at 48 h (Figure 4A). DNA bases in the 3'-overhang reduced potency 2- to 3-fold compared with the original RNA-overhang versions. Thus DNA bases can be incorporated in the single-stranded 3'-overhang (with some loss of activity) if increased nuclease stability is needed, such as for *in vivo* applications. The effect of varying the 3'-overhang length was also tested

(Figure 4B). Asymmetric 'R' duplexes at site H3 having 1-base, 2-base, 3-base and 4-base RNA 3'-overhangs all had similar potency. Extremely long 3'-overhangs of 17–20 bases were shown to block dicing from that end and reduce potency in triggering RNAi (8).

It has become common practice to make 21mer siRNAs with DNA 'TT' 3'-overhangs without regard to complementarity with the target sequence. We observed that substitution of DNA for RNA in the 3'-overhang can reduce potency by 2- to 3-fold (Figure 3B). Elbashir varied sequence in both sense and antisense strand 3'-overhangs for a luciferase 21mer siRNA and found that certain base substitutions resulted in as much as a 5-fold reduction in potency. Perfect match 'UG' or related 'UdG', 'UU', or 'TT' sequences also worked well, and 'TT' performed best (41). Vermeulen compared the relative efficiency of 16 duplexes for cleavage in an *in vitro* dicing assay that had every possible 2-base combination in a single asymmetric 3'-overhang in a 61mer substrate (29). A 3'-terminal rA base was found to be favorable for dicing, however no correlation with functional potency was made. Dinucleotide overhang sequences that are favorable for cleavage (Dicer substrates) may be different from sequences that are favorable for functional performance in 21mers (Dicer products). The effect that base sequence of the 3'-overhang may have on functional potency of 27mer Dicer-substrate duplexes was tested.

A set of 16 asymmetric 'R' form duplexes were synthesized at site H-3 in hnRNPH comprising every possible dinucleotide pair in the single 2-base 3'-overhang. Duplexes were transfected into HEK293 cells with the Luc-hnRNPH-'S' reporter construct and assayed 48 h post-transfection for luciferase activity (Figure 4C). The overhang sequence had a significant impact on duplex potency, with the duplex having a perfect match 'UU' overhang being 2–5 times more potent than the other duplexes tested. Overhang sequences that improved the efficiency of *in vitro* dicing of a 61mer duplex RNA (29) performed poorly in our functional assay of Dicer-substrate 27mer duplexes in suppressing target gene expression. More sequence sets will need to be compared to determine if complementary overhangs in the antisense strand will generally perform better or if the 'UU' sequence itself confers some advantage.

We have developed improved design criteria for short Dicer-substrate RNA duplexes for use as triggers of RNAi. Optimized design features include (i) asymmetric duplex with a single 2-base 3'-overhang on the antisense strand ('R' form duplex, 25D/27 type), (ii) use of RNA bases in the 3'-overhang complementary to the target and (iii) substitution of two DNA bases at the 3' end of the sense strand (blunt end). By exploiting the substrate preferences for human Dicer, this design generally leads to a single, predictable cleavage product after dicing and invokes the functional polarity conferred by the dicing reaction to favor antisense strand entry into RISC with enhanced sense strand targeting.

Figure 4. Structure-activity features of Dicer-substrate RNA duplexes. (A) Comparison of activity of 'R' form 27mer duplexes with RNA versus DNA bases in the single 2-base 3'-overhang. (B) Comparison of activity of 'R' form 27mer duplexes with 1, 2, 3 or 4-base RNA 3'-overhangs. (C) Comparison of activity of 'R' form 27mer duplexes with every possible RNA 2-base 3'-overhang. Top: For A and B, an EGFP-hnRNPH fusion protein expression plasmid was transfected into cells without (control) or with the indicated RNA duplexes and EGFP fluorescence was measured at 48 h post-transfection. For C, a luciferase expression vector containing hnRNPH sequence in the 3'-UTR was transfected into cells without (control) or with the indicated RNA duplexes and luciferase activity was assayed 48 h post-transfection. Bottom: hnRNPH target sequence (S strand) is shown and duplexes employed in transfection are aligned beneath in duplex form with sense strand top (5'→3') and antisense strand bottom (3'→5'). RNA bases are upper case, DNA bases are lower case bold, and 'p' represents 5'-phosphate. Two neighboring sites were targeted, H-1 and H-3 shifted by two bases along the hnRNPH sequence.

DISCUSSION

Degradative RNAi occurs through a complex series of linked biochemical events involving a large number of protein components, many of which are still being identified. The overall process can be conceptualized as occurring in three stages, including (i) processing of long dsRNAs into short 21–23mer functional siRNAs, (ii) assembly of a mature RISC and (iii) sequence-specific cleavage of target ssRNA (i.e. mRNA) followed by degradation. In the initiator phase, long dsRNAs are processed by Dicer into short 21–23mer siRNAs which have 2-base 3'-overhangs and 5'-terminal phosphate groups (10). In cell extracts, this process is accelerated by ATP (9,10); however, recombinant human Dicer is ATP-independent (42). In *Drosophila*, the potency of dsRNA to trigger an RNAi response decreases as length is shortened from 130 to 29 bp (8). Recombinant human Dicer efficiently cleaves long dsRNAs but shows some loss of activity as the dsRNA substrate decreases from 100 to 50 bp length (43). Shorter sequences regain activity and, in fact, RNA duplexes as short as 23mer length are substrates for human Dicer (cleaved to 21–22mer products) (1). More than one Dicer molecule can bind a 130 bp dsRNA molecule (possibly one on each end). Processing preferentially begins by cleaving duplexes at their ends and eventually results in complete degradation into 21–23 base fragments (42). It is not clear if sequential cleavage events occur in a processive fashion along a single substrate RNA molecule or if Dicer cleaves, dissociates and re-binds the substrate before cleaving again.

Dicer is a complex 220 kDa protein comprising a dsRNA binding domain (dsRBD), a PAZ domain, a DExH RNA helicase/ATPase domain, and two RNase III class domains (42,43). The RNase III domains cooperatively function to cleave substrate dsRNA into smaller 21–23 base fragments and orientation for this cleavage is assisted by flanking dsRNA binding domains, dsRBD and PAZ (33). Crystal structure of the PAZ domain from human Argonaute eIF2c1 has been determined and suggests that this element specifically functions as a binding site for 2-base 3'-overhangs (44). Similar conclusions were made for the *Drosophila* Ago-2 protein based upon NMR solution structure (45). The PAZ/PIWI domains seem to serve as anchors that helps spatially oriented bound RNAs in the enzyme active site (46,47). Argonaute cleaves the substrate RNA 10 bases away from the PIWI anchor site whereas Dicer cleaves 21–22 bases away from the PAZ anchor site. This model fits well with our experimental observations of dicing patterns for 27mer dsRNA substrates. Blunt 27mer substrates do not provide an optimal structure for binding the Dicer PAZ domain so the 'anchor' step takes place with imprecision and results in heterogeneous cleavage products. Asymmetric duplexes with a single 2-base 3'-overhang provide a single favorable PAZ binding site, so these substrates usually have a single unique anchor site and cleavage occurs 21–22 bases away with limited heterogeneity. The presence of a 3'-overhang promotes 'correct' dicing of hairpin RNA substrates (21). Interestingly, the actual base sequence of the 3'-overhang can influence dicing (29) and not all asymmetric duplexes with a single 3'-overhang show a simple dicing pattern (EGFPS2 R 27/25, Figure 1C). Adding DNA residues to the blunt end of an asymmetric duplex seems to help direct binding to the 3'-overhang even for

'unfavorable' sequences, possibly because the presence of DNA makes the blunt end an even worse structure for PAZ binding (EGFPS2 R 27/25D, Figure 1D). Thus, asymmetric duplexes with one 2-base 3'-overhang in combination with 3'-DNA residues on the blunt end present Dicer with a substrate that is cleaved into predictable products.

Functional interpretation of the ESI-MS dicing data assumes that the cleavage properties of endogenous human Dicer will parallel the patterns observed using purified recombinant human Dicer *in vitro*. In *Drosophila*, Dicer functions as a heterodimer with the RNA binding protein R2D2, which forms a complex with Dicer and the siRNA duplex (48). A possible human ortholog for R2D2 has recently been identified as TRBP (49). SiRNA-mediated depletion of TRBP resulted in loss of recruitment of Ago-2 and destabilization of Dicer. Behavior of human Dicer *in vitro* in the absence of other RNAi pathway proteins such as TRBP may be different from its properties *in vivo*. Since the actual functional potency of RNA duplexes when transfected into cells follows predictions made based upon dicing patterns obtained for the same duplexes *in vitro*, it seems unlikely that *in vivo* dicing patterns will be significantly different from those observed *in vitro*.

Dicer plays more than one role in the RNAi pathway. In addition to endonuclease cleavage of long dsRNA into siRNAs, Dicer is involved with entry of the siRNA into RISC and participates in RISC assembly (25). *Drosophila* has two Dicer proteins, a Dicer-1 nuclease that is involved in miRNA processing and a Dicer-2 nuclease that is involved in siRNA processing (22). Mutants lacking Dicer-2 activity have defective RISC assembly, even when provided with 21mer siRNAs that do not require cleavage (23). We theorize that the RISC assembly function of Dicer is involved with the increased potency seen for Dicer-substrate RNA duplexes compared with short 21mer siRNAs; providing Dicer with a 'substrate', rather than a 'product', may improve efficiency of the RISC entry step.

While association with Dicer is required for entry of siRNAs into RISC in *Drosophila*, the pathways available for RISC loading may be slightly different in mammals. In one study, single-stranded RNA was shown to be capable of directing sequence-specific target cleavage via RNAi pathways in HeLa cell extracts while only duplex siRNAs could direct target cleavage in *Drosophila* extracts. Further, immunodepletion of Dicer from the HeLa cell extracts did not block target cleavage triggered by siRNAs (50). In another study, a mouse ES cell line was established that was homozygous for disruption of the *dcr-1* gene and had no functional Dicer activity. As expected, these cells were deficient for miRNA production and could not initiate an RNAi response from shRNA compounds, both of which require Dicer processing. Unlike the *Drosophila* Dicer-2 mutants, however, the Dicer-deficient mouse cells could support RNAi if provided exogenous 21mer duplex siRNAs (51). Thus two independent studies suggest that Dicer is not required for siRNA entry into RISC in mammalian cells. However, Dicer may still play some role in siRNA loading into RISC. In studies performed in both human and mouse cells, Doi and colleagues demonstrated that knockdown of Dicer using RNAi triggered by 21mer siRNAs significantly reduced the efficiency of siRNA-mediated silencing of a luciferase reporter target (52). In a study that reconciles some of the apparent differences

between the human and *Drosophila* systems, Chendrimada *et al.* (49) reported that RNAi-mediated knockdown of either Dicer or TRBP, the proposed human R2D2 ortholog, reduced efficiency of siRNA-mediated silencing of a luciferase reporter target in a human cell line. Taken together, these studies suggest that RISC loading and functional triggering of an RNAi response is more efficient when Dicer and TRBP are present, even though Dicer is clearly not required for siRNA loading into RISC in mammals.

The roles played by various protein and their interactions with the siRNA during RISC assembly are currently better defined in *Drosophila* than in mammals. In *Drosophila*, cooperative interaction of the R2D2 protein and Dicer-2 is required for entry of a siRNA into RISC (48). R2D2 preferentially binds the end of the siRNA duplex having greater thermodynamic stability and specifically associates with the 5' end of whichever strand is present at this end. Conversely, Dicer binds the 5' end of the strand on the opposing end (24). This complex is joined by a variety of other protein components as RISC assembly proceeds. Eventually, the siRNA duplex is unwound and R2D2 exits RISC with its associated single-strand (the 'passenger strand'), leaving Dicer and its associated strand in RISC. The retained strand later serves to direct sequence-specific targeting (the 'guide strand'). Ago-2 replaces R2D2 at the 3' end of the retained guide strand (24) and is the Argonaute protein family member that functions as the actual 'slicer' endonuclease activity in RISC (53).

Thermodynamic bias leads to preferential association of R2D2 on one end of an siRNA duplex and Dicer on the other end, and thereby directs as to which strand remains in RISC ('guide strand') and which strand is ejected ('passenger strand'). The terminal base sequence therefore plays a significant role in determining sense versus antisense strand targeting. Other factors may also influence which strand enters RISC. Elbashir observed that the orientation of Dicer processing of a model 52mer RNA duplex defined which strand directed target cleavage. Long single-stranded overhangs block dicing from starting at that end. A blunt 52mer duplex directed cleavage of both sense and antisense strand targets. A 52mer duplex with 20 base overhangs on both ends was inactive and did not direct cleavage of either strand targets. Asymmetric 52mer dsRNA duplexes showed strand bias. If the 20-base 3'-overhang extended from the sense strand, Dicer processing started from the blunt end and antisense strand sequence preferentially directed cleavage of a sense target. Conversely, if the 20-base 3'-overhang extended from the antisense strand, Dicer processing started from the blunt end and sense-strand sequence preferentially directed cleavage of an antisense target (8).

The difference of functional potency that we see between 'R' versus 'L' version asymmetric Dicer-substrate duplexes probably relates to interactions between the same protein components within the RNAi pathway. In the case of asymmetric 27mers, the blunt end is unfavorable for Dicer binding and the dicing process starts at the end with a 2-base 3'-overhang. For 'L' duplexes, the 3'-overhang resides on the sense (top) strand. For 'R' duplexes, the 3'-overhang resides on the antisense (bottom) strand. 'R' duplexes generally show increased potency in directing silencing of sense-strand targets, a pattern consistent with the observations of Tuschl and co-workers (8). In the model of RISC formation proposed by Zamore and

co-workers (24), Dicer associates with the 5' end of the strand that is retained in RISC and R2D2 associates with the 5' end of the discarded passenger strand. The recent discovery of the required role for TRBP in mammalian RNAi opens the possibility that this protein serves a function similar to the *Drosophila* R2D2 (49). For 21mer siRNAs, where no dicing occurs, thermodynamic end stability might be the dominant factor directing which protein binds which end. For longer dsRNAs, where dicing occurs, polarity of the dicing reaction may also affect final protein binding patterns. Binding of a 3'-overhang may spatially orient the siRNA cleavage product in the correct position for association of the 5' end of that strand with Dicer (and therefore remain in RISC) and allow access of the other end to associate with a human R2D2 ortholog. This model assumes that some fraction of Dicer/siRNA complexes remain intact following cleavage and directly enter a developing RISC complex. If Dicer freely dissociated from the siRNA product after cleavage and later rebound on the basis of thermodynamic end stability, then polarity of the dicing reaction should not influence potency or confer strand bias. The nascent siRNA produced by Dicer cleavage has symmetric 2-base 3'-overhangs. Any effects introduced by structure of the asymmetric substrate should be lost if product siRNA and Dicer separate. This model is also consistent with the pattern of increased potency observed for asymmetric short siRNAs by Khvorova and co-workers (29). Here, binding of an antisense-strand 3'-overhang by the PAZ domain may similarly serve to favorably orient the 5' end of the antisense strand within Dicer for subsequent retention in RISC, in this case without the need for a cleavage event.

Strand bias introduced by Dicer processing of asymmetric duplexes confers a relative, not absolute, advantage to retention of the 3'-overhang strand in RISC. For example, the strand targeting experiments shown in Figure 3 demonstrate that use of the 'R' versus 'L' form duplexes significantly alter the ratio of 'S' versus 'AS' strand targeting; however, in all cases tested both 'S' and 'AS' strand targeting still occurs. In a more biologically relevant example, the active strand in miRNAs can be derived from either the top or bottom strand of the precursor miRNA and strand selection seems to be primarily determined by thermodynamic asymmetry rules without regard for the direction of Dicer processing (54). The relative contribution of various factors that contribute to preferential strand loading into RISC is complex.

Use of thermodynamic end stability rules (38,40,55) and empiric design parameters (28,56) has helped speed widespread use of RNAi in mammalian biology by making more potent reagents easier to obtain. The asymmetric duplexes described here similarly improve design of Dicer-substrate RNAs by exploiting the functional polarity introduced by Dicer processing.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Brian Elliott for assistance with ESI-MS analysis and Stephanie McConahay for assistance with preparation of the

figures. D. Kim is a Beckman Fellow. M. Amarzguioui is a postdoctoral fellow of the Norwegian Research Council. This work was supported in part by a grant from the Arnold and Mabel Beckman Foundation and the National Institutes of Health (AI29329, AI42552 and HL074704 to J.J.R.). Funding to pay the Open Access publication charges for this article was provided by Integrated DNA Technologies, Inc.

Conflict of interest statement. Scott D. Rose, Michael A. Collingwood and Mark A. Behlke are employed by Integrated technologies, Inc. (IDT), which has filed at least one patent application on the inventions described in this manuscript, and which offers oligonucleotides for sale similar to the oligonucleotides described in this manuscript. IDT is, however, not a publicly traded company, and Scott D. Rose, Michael A. Collingwood and Mark A. Behlke personally do not own any shares or equity in IDT. None of the authors have any conflicts to declare.

REFERENCES

- Kim, D.H., Behlke, M.A., Rose, S.D., Chang, M.S., Choi, S. and Rossi, J.J. (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat. Biotechnol.*, **23**, 222–226.
- Mello, C.C. and Conte, D., Jr (2004) Revealing the world of RNA interference. *Nature*, **431**, 338–342.
- Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature*, **431**, 343–349.
- Hannon, G.J. and Rossi, J.J. (2004) Unlocking the potential of the human genome with RNA interference. *Nature*, **431**, 371–378.
- Napoli, C., Lemieux, C. and Jorgensen, R. (1990) Introduction of a chimeric chalcone synthase gene into *Petunia* results in reversible co-suppression of homologous genes in trans. *Plant Cell*, **2**, 279–289.
- Romano, N. and Macino, G. (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.*, **6**, 3343–3353.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
- Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, **15**, 188–200.
- Zamore, P.D., Tuschl, T., Sharp, P.A. and Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, **101**, 25–33.
- Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.
- Heidel, J.D., Hu, S., Liu, X.F., Triche, T.J. and Davis, M.E. (2004) Lack of interferon response in animals to naked siRNAs. *Nat. Biotechnol.*, **22**, 1579–1582.
- Ma, Z., Li, J., He, F., Wilson, A., Pitt, B. and Li, S. (2005) Cationic lipids enhance siRNA-mediated interferon response in mice. *Biochem. Biophys. Res. Commun.*, **330**, 755–759.
- Judge, A.D., Sood, V., Shaw, J.R., Fang, D., McClintock, K. and Maclachlan, I. (2005) Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat. Biotechnol.*, **23**, 457–462.
- Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H. and Williams, B.R. (2003) Activation of the interferon system by short-interfering RNAs. *Nature Cell Biol.*, **5**, 834–839.
- Hornung, V., Guenther-Biller, M., Bourquin, C., Ablasser, A., Schlee, M., Uematsu, S., Noronha, A., Manoharan, M., Akira, S., de Fougerolles, A. et al. (2005) Sequence-specific potent induction of IFN- α by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nature Med.*, **11**, 263–270.
- Hu-Lieskovan, S., Heidel, J.D., Bartlett, D.W., Davis, M.E. and Triche, T.J. (2005) Sequence-specific knockdown of EWS-FLI1 by targeted, non-viral delivery of siRNA inhibits tumor growth in a murine model of metastatic Ewing's sarcoma. *Cancer Res.*, in press.
- Sorensen, D.R., Leirdal, M. and Sioud, M. (2003) Gene silencing by systemic delivery of synthetic siRNAs in adult mice. *J. Mol. Biol.*, **327**, 761–766.
- Hamar, P., Song, E., Kokeny, G., Chen, A., Ouyang, N. and Lieberman, J. (2004) Small interfering RNA targeting Fas protects mice against renal ischemia-reperfusion injury. *Proc. Natl Acad. Sci. USA*, **101**, 14883–14888.
- Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., Elbashir, S., Geick, A., Hadwiger, P., Harborth, J. et al. (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature*, **432**, 173–178.
- Ge, Q., Filip, L., Bai, A., Nguyen, T., Eisen, H.N. and Chen, J. (2004) Inhibition of influenza virus production in virus-infected mice by RNA interference. *Proc. Natl Acad. Sci. USA*, **101**, 8676–8681.
- Siolas, D., Lerner, C., Burchard, J., Ge, W., Linsley, P.S., Paddison, P.J., Hannon, G.J. and Cleary, M.A. (2005) Synthetic shRNAs as potent RNAi triggers. *Nat. Biotechnol.*, **23**, 227–231.
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J. and Carthew, R.W. (2004) Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, **117**, 69–81.
- Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W. and Sontheimer, E.J. (2004) A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell*, **117**, 83–94.
- Tomari, Y., Matranga, C., Haley, B., Martinez, N. and Zamore, P.D. (2004) A protein sensor for siRNA asymmetry. *Science*, **306**, 1377–1380.
- Sontheimer, E.J. (2005) Assembly and function of RNA silencing complexes. *Nature Rev. Mol. Cell Biol.*, **6**, 127–138.
- Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E. and Prydz, H. (2002) Positional effects of short interfering RNAs targeting the human coagulation trigger tissue factor. *Nucleic Acids Res.*, **30**, 1757–1766.
- Harborth, J., Elbashir, S.M., Vandenberg, K., Manning, H., Scaringe, S.A., Weber, K. and Tuschl, T. (2003) Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Antisense Nucleic Acid Drug Dev.*, **13**, 83–105.
- Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S. and Khvorova, A. (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.*, **22**, 326–330.
- Vermeylen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W.S., Karpilow, J. and Khvorova, A. (2005) The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA*, **11**, 674–682.
- Hail, M.E., Elliott, B. and Anderson, K. (2004) High-throughput analysis of oligonucleotides using automated electrospray ionization mass spectrometry. *Am. Biotechnol. Lab.*, **12**, 12–14.
- Kumar, R., Conklin, D.S. and Mittal, V. (2003) High-throughput selection of effective RNAi probes for gene silencing. *Genome Res.*, **13**, 2333–2340.
- Bieche, I., Nogues, C., Paradis, V., Olivi, M., Bedossa, P., Lidereau, R. and Vidaud, M. (2000) Quantitation of hTERT gene expression in sporadic breast tumors with a real-time reverse transcription-polymerase chain reaction assay. *Clin. Cancer Res.*, **6**, 452–459.
- Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E. and Filipowicz, W. (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell*, **118**, 57–68.
- Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. and Hannon, G.J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231–235.
- Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–240.
- Lund, E., Guttlinger, S., Calado, A., Dahlberg, J.E. and Kutay, U. (2004) Nuclear export of microRNA precursors. *Science*, **303**, 95–98.
- He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.*, **5**, 522–531.
- Aza-Blanc, P., Cooper, C.L., Wagner, K., Batalov, S., Deveraux, Q.L. and Cooke, M.P. (2003) Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol. Cell*, **12**, 627–637.
- Schwarz, D.S., Hutvagner, G., Haley, B. and Zamore, P.D. (2002) Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. *Mol. Cell*, **10**, 537–548.
- Khvorova, A., Reynolds, A. and Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.
- Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. and Tuschl, T. (2001) Functional anatomy of siRNAs for mediating efficient

- RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.*, **20**, 6877–6888.
42. Zhang, H., Kolb, F.A., Brondani, V., Billy, E. and Filipowicz, W. (2002) Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.*, **21**, 5875–5885.
 43. Provost, P., Dishart, D., Doucet, J., Frenthewey, D., Samuelsson, B. and Radmark, O. (2002) Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.*, **21**, 5864–5874.
 44. Ma, J.B., Ye, K. and Patel, D.J. (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, **429**, 318–322.
 45. Lingel, A., Simon, B., Izaurralde, E. and Sattler, M. (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature Struct. Mol. Biol.*, **11**, 576–577.
 46. Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T. and Patel, D.J. (2005) Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*, **434**, 666–670.
 47. Parker, J.S., Roe, S.M. and Barford, D. (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, **434**, 663–666.
 48. Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P. and Wang, X. (2003) R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science*, **301**, 1921–1925.
 49. Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K. and Shiekhattar, R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, doi:10.1038/nature03868.
 50. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. and Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, **110**, 563–574.
 51. Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M. and Rajewsky, K. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.*, **19**, 489–501.
 52. Doi, N., Zenno, S., Ueda, R., Ohki-Hamazaki, H., Ui-Tei, K. and Saigo, K. (2003) Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. *Curr. Biol.*, **13**, 41–46.
 53. Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. and Hannon, G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441.
 54. Krol, J., Sobczak, K., Wilczynska, U., Drath, M., Jasinska, A., Kaczynska, D. and Krzyzosiak, W.J. (2004) Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J. Biol. Chem.*, **279**, 42230–42239.
 55. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199–208.
 56. Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R. and Saigo, K. (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.*, **32**, 936–948.