Citation: Molecular Therapy–Nucleic Acids (2012) 1, e17; doi:10.1038/mtna.2012.6

www.nature.com/mtna

Deep Sequencing Analyses of DsiRNAs Reveal the Influence of 3′ Terminal Overhangs on Dicing Polarity, Strand Selectivity, and RNA Editing of siRNAs

Jiehua Zhou[1], Min-Sun Song[1], Ashley M Jacobi[2], Mark A Behlke[2], Xiwei Wu[3] and John J Rossi[1,4]

25/27 Base duplex RNAs that are substrates for Dicer have been demonstrated to enhance RNA interference (RNAi) potency and efficacy. Since the target sites are not always equally susceptible to suppression by small interfering RNA (siRNA), not all 27-mer duplexes that are processed into the corresponding conventional siRNAs show increased potency. Thus random designing of Dicer-substrate siRNAs (DsiRNAs) may generate siRNAs with poor RNAi due to unpredictable Dicer processing. Previous studies have demonstrated that the 3′-overhang affects dicing cleavage site and the orientation of Dicer entry. Moreover, an asymmetric 27-mer duplex having a 3′ two-nucleotide overhang and 3′-DNA residues on the blunt end has been rationally designed to obtain greater efficacy. This asymmetric structure directs dicing to predictably yield a single primary cleavage product. In the present study, we analyzed the *in vitro* and intracellular dicing patterns of chemically synthesized duplex RNAs with different 3′-overhangs. Consistent with previous studies, we observed that Dicer preferentially processes these RNAs at a site 21–22 nucleotide (nt) from the two-base 3′-overhangs. We also observed that the direction and ability of human Dicer to generate siRNAs can be partially or completely blocked by DNA residues at the 3′-termimi. To examine the effects of various 3′-end modifications on Dicer processing in cells, we employed Illumina Deep sequencing analyses to unravel the fates of the asymmetric 27-mer duplexes. To validate the strand selection process and knockdown capabilities we also conducted dual-luciferase psiCHECK reporter assays to monitor the RNAi potencies of both the "sense" (S) and "antisense" (AS) strands derived from these DsiRNAs. Consistent with our *in vitro* Dicer assays, the asymmetric duplexes were predictably processed into desired primary cleavage products of 21–22-mers in cells. We also observed the trimming of the 3′ end, especially when DNA residues were incorporated into the overhangs and this trimming ultimately influenced the Dicer-cleavage site and RNAi potency. Moreover, the observation that the most efficacious strand was the most abundant revealed that the relative frequencies of each "S" or "AS" strand are highly correlated with the silencing activity and strand selectivity. Collectively, our data demonstrate that even though the only differences between a family of DsiRNAs was the 3′ two-nuclotide overhang, dicing polarity and strand selectivity are distinct depending upon the sequence and chemical nature of this overhang. Thus, it is possible to predictably control dicing polarity and strand selectivity *via* simply changing the 3′-end overhangs without altering the original duplex sequence. These optimal design features of 3′-overhangs might provide a facile approach for rationally designing highly potent 25/27-mer DsiRNAs.
Molecular Therapy–Nucleic Acids (2012) **1**, e5; doi:10.1038/mtna.2012.6; published online 3 April 2012

Introduction

RNA interference (RNAi) is a sequence-specific post-transcriptional gene silencing process triggered by 21–25 nucleotide (nt) small interfering RNAs (siRNAs). In cells these siRNAs are generated by the ribonuclease III Dicer which processes these siRNAs from longer double-stranded RNAs.[1,2] In association with Dicer, the cleaved small RNA products possessing a 5′-phosphate and 2-base 3′ overhang are loaded into large multiprotein complexes termed RNA-induced silencing complexes (RISC) and one of the two strands is selected as a guide for the sequence-specific silencing of the complementary target RNA.[2–5] The PAZ domain, an RNA-binding domain of Dicer which is also found in Argonaute proteins, specifically recognizes the 3′ end of single-stranded RNA, suggesting it can function as a module for anchoring the 3′ end of the guide strand within the RISC.[5,6] For the Dicer-substrate siRNAs (DsiRNAs), the 3′ overhang therefore[7–9] affects dicing polarity (binding of Dicer) as well as subsequent strand selectivity in RISC,[10–15] consequently influencing the overall RNAi efficiency. It was previously reported that chemically synthesized 29 base duplex short hairpin RNAs that are substrates for Dicer are more potent RNAi triggers than 19 base duplex short hairpin RNAs.[16] Similarly, DsiRNAs of 25–30 nt can be up to 100-fold more potent than conventional 21-mer duplexes targeted to the same sequence location.[17] This

[1]Department of Molecular and Cellular Biology, Beckman Research Institute of City of Hope, Duarte, California, USA; [2]Integrated DNA Technologies, Inc. Coralville, Iowa, USA; [3]Department of Molecular Medicine, Beckman Research Institute of City of Hope, Duarte, California, USA; [4]Irell and Manella Graduate School of Biological Sciences, Beckman Research Institute of City of Hope, Duarte, California, USA
Correspondence: John J Rossi, Department of Molecular and Cellular Biology, Beckman Research Institute of City of Hope, 1500 East Duarte Road, Duarte, California 91010, USA.
E-mail address: jrossi@coh.org
Keywords: RNA interference; Dicer-substrate siRNA (DsiRNA); Deep sequencing; 3′-overhang; Dicing polarity
Received 23 January 2012; accepted 18 February 2012; advance online publication 3 April 2012. doi:10.1038/mtna.2012.6

increased potency might be attributed to the fact that Dicer-generated 21–23-mer siRNAs are more efficiently incorporated into RISC through the physical association of Dicer with the TAR RNA-binding protein and Argonaute proteins.

Dicer cleavage of a blunt ended 27-mer duplex generally can generate a variety of different 21–23-mers depending on its sequence parameters,[15,18] such as length/composition of the 3′-terminus,[19] GC content, inverted repeats, etc. Furthermore, it has been demonstrated that siRNA efficacy is highly dependent on the target position[20] and RNAi potency is susceptible to shifting a 21-mer siRNA even by a single base along the target mRNA sequence.[21] The overall RNAi efficacy of DsiRNAs critically depends on the composition and potency of the processing products. Thus acquiring a better understanding of DsiRNA designs is useful for enhancing RNAi efficacy. In this study, we show that it is possible to predictably control dicing polarity and strand selectivity *via* simply changing the sequences of the 3′-end overhangs without altering the original duplex sequence. Recently, an asymmetric 25/27-mer duplex having a 3′ two-nucleotide overhang and two 3′-DNA residues on the blunt end has been rationally designed to obtain greater efficacy.[22–24] This asymmetric structure directs dicing to predictably yield a single primary cleavage product.

In the present study, a simple 5′-end labeled gel assay was performed to analyze *in vitro* dicing patterns of chemically synthesized symmetric or asymmetric 25 base pair duplex *TNPO3* (*Homo sapiens transportin 3*) RNAs with different 3′ two-nucleotide overhangs. We observed that a series of heterogeneous products were generated by Dicer cleavage from these substrates and Dicer preferentially processes to a site 22 nt from the 3′-end of the substrates that have a two-base ribonucleotide overhang. The 3′ two-base ribonucleotide overhangs can help orient Dicer on the substrates and facilitate Dicer entry and cleavage. In contrast, the direction and

Table 1 The 21-mer siRNAs and 27-mer Dicer-substrate siRNAs targeting *TNPO3* and the target sequences are listed

	***TNPO3* 21-mer siRNA and 27-mer Dicer-substrate siRNA (DsiRNA)**		***TNPO3* target sequence (site I)(GC-natural match)**
			5′---CG CGACA TTGCAGCTCGTGTACCAG GC ---- 3′
Group I	Symmetric 21/21-mer	Original 21 (UU)/21 (UU)	5′ CGACAUUGCAGCUCGUGUA UU 3′ UU GCUGUAACGUCGAGCACAU
		Shifted 21 (UU)/21 (UU)	5′ UGCAGCUCGUGUACCAGGC UU 3′ UU ACGUCGAGCACAUGGUCCG
	Asymmetric 25/27-mer	25 (GC)/27 (UU)	5′ CGACAUUGCAGCUCGUGUACCAGGC 3′ UU GCUGUAACGUCGAGCACAUGGUCCG
		25 (GC)/27 (tt)	5′ CGACAUUGCAGCUCGUGUACCAGGC 3′ tt GCUGUAACGUCGAGCACAUGGUCCG
	Symmetric 27/27-mer	27 (UU)/27 (UU)	5′ CGACAUUGCAGCUCGUGUACCAGGC UU 3′ UU GCUGUAACGUCGAGCACAUGGUCCG
		27 (tt)/27 (tt)	5′ CGACAUUGCAGCUCGUGUACCAGGC tt 3′ tt GCUGUAACGUCGAGCACAUGGUCCG
		27 (UU)/27 (tt)	5′ CGACAUUGCAGCUCGUGUACCAGGC UU 3′ tt GCUGUAACGUCGAGCACAUGGUCCG
		27 (tt)/27 (UU)	5′ CGACAUUGCAGCUCGUGUACCAGGC tt 3′ UU GCUGUAACGUCGAGCACAUGGUCCG
Group II	Asymmetric 25/27-mer	25 (gc)/27 (GC)	5′ CGACAUUGCAGCUCGUGUACCAG *gc* 3′ CG GCUGUAACGUCGAGCACAUGGUCCG
		25 (gc)/27 (tt)	5′ CGACAUUGCAGCUCGUGUACCAG *gc* 3′ tt GCUGUAACGUCGAGCACAUGGUCCG
		25 (gc)/27 (UU)	5′ CGACAUUGCAGCUCGUGUACCAG *gc* 3′ UU GCUGUAACGUCGAGCACAUGGUCCG
		25 (gc)/27 (AA)	5′ CGACAUUGCAGCUCGUGUACCAG *gc* 3′ AA GCUGUAACGUCGAGCACAUGGUCCG
		25 (gc)/27 (GG)	5′ CGACAUUGCAGCUCGUGUACCAG *gc* 3′ GG GCUGUAACGUCGAGCACAUGGUCCG
		25 (gc)/27 (CC)	5′ CGACAUUGCAGCUCGUGUACCAG *gc* 3′ CC GCUGUAACGUCGAGCACAUGGUCCG

The sense strand is presented from 5′ to 3′ and is marked as black, while the antisense strand is presented 3′ to 5′ end is marked as gray. Two-nucleotide 3′-overhangs are underlined. Ribonucleotides are upper case and deoxyribonucleotides are in lower case. These RNAs are named by their strand length and overhangs: the number indicates the length of RNA strands; "(NN)" means the two-base 3′ ends. Starting with a 19-mer siRNA sequence, shown as the original 21 (UU)/21 (UU) mer, these 27-mer DsiRNAs have are extended by six bases upstream of the target sequence.
DsiRNA, Dicer-substrate small interfering RNA; *TNPO3*, *Homo sapiens* transportin 3; siRNA, small interfering RNA.

ability of human Dicer to generate siRNAs can be partially or completely blocked by DNA residues at the 3′-termimi. Consistent with previous studies, combining of such features into an asymmetric DsiRNA (25/27-mer DsiRNA) provides an optimal design for obtaining a single, primary cleavage product with the best RNAi potency.

To investigate the influence of the two-base 3′ overhang on dicing of substrates in cells we employed Illumina Deep sequencing analyses to unravel the fates of the asymmetric 25/27-mer *TNPO3* duplexes in HEK293 cells. To validate the strand selection process and target knockdown capabilities we also conducted dual-luciferase psiCHECK reporter assays to monitor the RNAi potencies of both the "S" and "antisense" (AS) strands derived from these DsiRNAs. In similarity to our *in vitro* Dicer assays, the asymmetric duplexes were predictably processed into the desired primary cleavage products of 21–22 nts in cells. We also observed trimming of the 3′ ends and some additional Uracils added as well. Our observation that the most efficacious strand was the most abundant revealed that the relative frequencies of each "S" or "AS" strand are highly correlated with the silencing activity and strand selectivity. A similar observation was also found in a separate pair of asymmetric 25/27-mer duplexes targeting heterogeneous nuclear ribonucleoprotein H (*hnRNP H1*), further validating the role of the sequence composition of 3′ double-nucleotide overhang and the proposed optimal design features.

Taken together, our data demonstrate that even though the only differences between a family of DsiRNAs was the 3′ two-nucleotide overhang, dicing polarity and strand selectivity are distinct depending upon the sequence and chemical nature of this overhang. Thus, it is possible to predictably control dicing polarity and strand selectivity *via* simply changing the 3′-end overhangs without altering the original duplex sequence. These optimal design features of 3′-overhangs provide a facile approach for rationally designing highly potent 27-mer DsiRNAs.

Results

Design of DsiRNAs against *TNPO3* and *in vitro* Dicing approaches

We designed and synthesized a series of 25 base pair, two-base overhang containing RNAs targeting the mRNA produced from the *TNPO3* gene (*Homo sapiens transportin 3*) which is one of the many HIV-1 dependency factors[25] (**Table 1**). The original 21 (UU)/21 (UU) siRNA was previously reported to efficiently knockdown target gene expression.[25] The shifted 21-mer siRNAs were also designed to target sites shifted downstream of the original 21-mer siRNA in increments of 6 nt. Based upon the original 21-mer sequence, we designed the 27-mers with an added six bases of the *TNPO* sequence The various arrangements of 3′ two-base overhangs are labeled as groups I and II, which include four symmetric and six asymmetric DsiRNAs as listed in **Table 1**.

To evaluate the effects of the 3′-overhang in determining the position and pattern of Dicer cleavage, P^{32}-end labeled duplexes were processed *in vitro* with recombinant human Dicer, and the dicing products were electrophoresed in denaturing polyacrylamide gels and visualized by autoradiography. **Figure 1** depicts the two types of experimental approaches used (method A and B). The dicing patterns were designated by the direction of Dicer entry into the substrates. "L-R" indicates Dicer enters

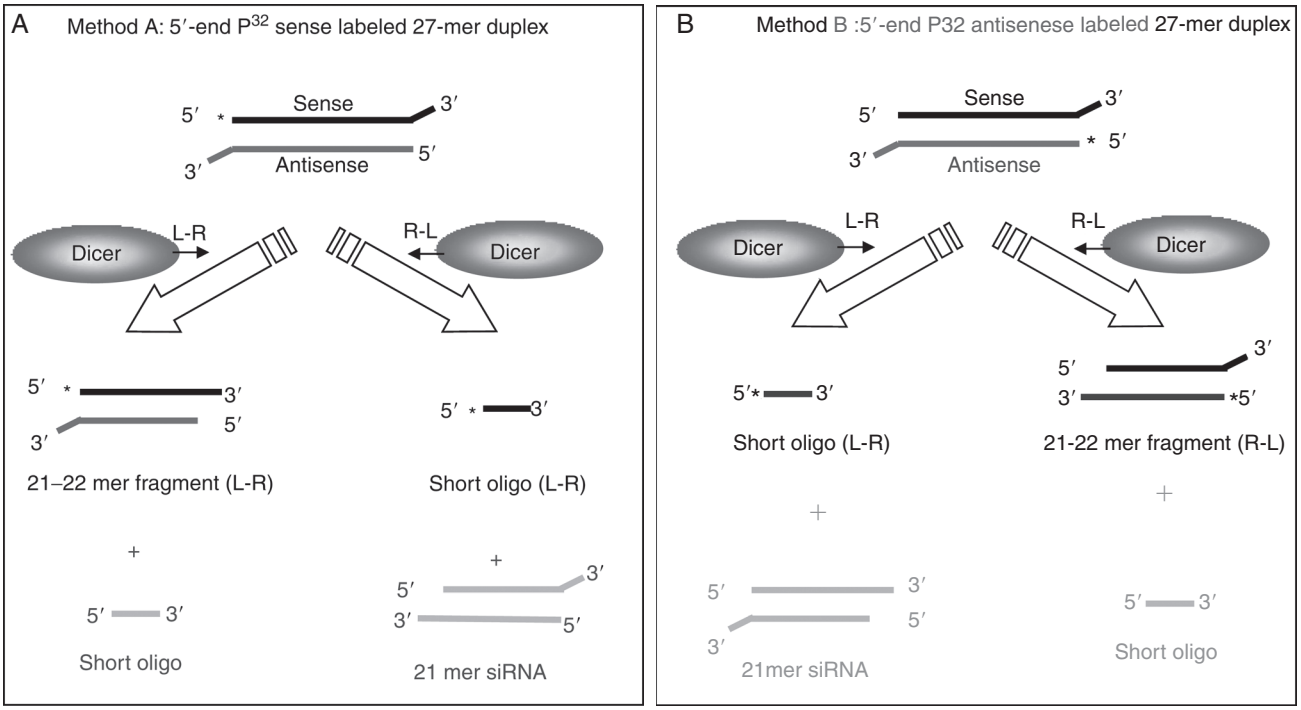


Figure 1 ***In vitro*** **Dicer processing of (a) 5′ P^{32}-end labeled sense or antisense (b) duplexes.** Two types of experimental approaches are displayed as method A and B, respectively. The dicing patterns are named by the direction of Dicer entering the duplex as left to right (L-R) and right to left (R-L).

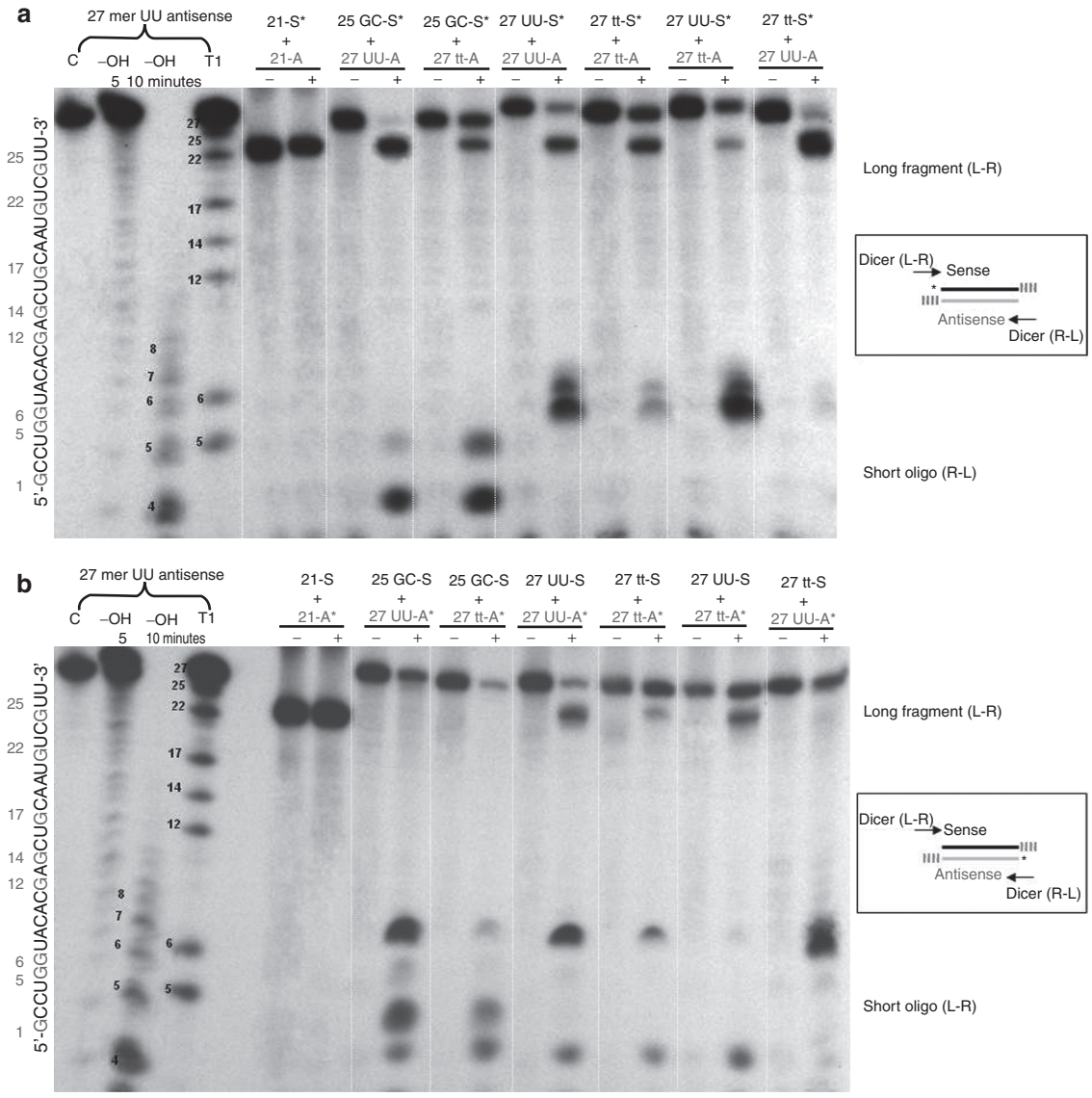


Figure 2 ***In vitro*** **Dicer cleavage of 5′ P^{32}-end labeled sense or antisense strands.** Group I duplexes. The cleavage fragments that result from dicing L-R and R-L were visualized autoradiographically following denaturing gel electrophoresis of the Dicer-cleavage products: (a) sense and (b) antisense. The two experimental approaches (method A and B) are defined in the legend to Figure 1.

from left to right, while the "R-L" model depicts Dicer processing from right to left. The proportions of the cleaved/uncleaved fragments that result from "L-R" and "R-L" dicing directions reflect the dicing efficiency for the 27-mer duplex.

The 3′-overhangs influence *in vitro* dicing patterns

Initially, the products that result from *in vitro* digestion of group I duplexes were analyzed using 5′-end P^{32} labeled RNAs and a denaturing gel electrophoresis assay. In method A, Dicer cleavage of the 5′-end P^{32} S strand-labeled duplexes generated two different sized S strand fragments (**Figure 2a**). For "L-R" the longer species 21–22 mer contained the P^{32}, whereas for "R-L" short 4–5 mers from the 25-mer S or 6–7 mers from the 27-mer AS strand, were produced, respectively. These results demonstrate that Dicer enters these 27-mer duplex RNAs by either "L-R" or "R-L." A comparable proportion of long and short species (the ratio of "L-R" versus "R-L" = 1) was observed in the 27 (UU)/27 (UU) duplex. However, the 27 (tt)/27 (UU) duplex with a two-base deoxyribonucleotide (tt) 3′ overhang on the S strand yielded greater amounts of long "L-R" fragments and a very small amount of "R-L" products (the ratio of "L-R" versus "R-L" >>1). In the reverse overhang setting 27 (UU)/27 (tt) duplex, the ratio of "L-R" to "R-L" cleavage products was <1, confirming that Dicer does not readily enter a duplex with dTdT 2-base overhangs. Furthermore, in comparison with the 27 (tt)/27 (tt) duplex, the 27 (UU)/27 (tt) duplex has a dominant "R-L" dicing polarity and the 27 (tt)/27 (UU) duplex has a major "L-R" polarity, demonstrating that the 3′-UU overhang facilitates Dicer binding and entry. A blunt end in the asymmetric duplexes (the 25 (GC)/27 (UU) and the 25 (GC) /27 (tt) duplexes) seemed to show less dicing polarity.

To further validate these observations, we also 5′ P^{32} end-labeled the AS strand of the same 27-mer duplexes (method B) and preformed polyacrylamide denaturing gel

assays and autoradiography (**Figure 2b**). These 27-mer duplexes were also cleaved bidirectionally from both termini. The 27 (UU)/27 (UU) duplex generated a comparable proportion of long and short species. In contrast, there was very little processing of the 27 (tt)/27 (tt) duplex due to its two 3′-dTdT overhangs. The 25 (GC)/27 (UU) and the 27 (tt)/27 (UU) duplexes containing a 3′-UU overhang on the AS strand were readily processed from "L-R," primarily generating short fragments (**Figure 2b**), consistent with the results observed when these duplexes were labeled on the S strand in which the long fragments were the primary products (**Figure 2a**). Only a trace amount of the "R-L" products were observed following dicing of the above two duplexes due to the hindrance of the blunt end or 3′-dTdT overhang on the S strand. As observed in **Figure 2a**, the dicing pattern was completely reversed in the 27 (tt)/27 (UU) duplex and the 27 (UU)/27 (tt) duplexes. These results definitively demonstrate that the 3′-overhang orients Dicer entry and influences dicing preference, and DNA residues reduce the binding of Dicer. Interestingly, a heterogeneous collection of 4–7-nt fragments was produced by dicing of the asymmetric 27-mer duplexes (the 25 (GC)/27 UU and the 25 (GC)/27 (tt)). These fragments may be derived from a second round of Dicer cleavage of these siRNAs.

The 3′-overhangs affect RNAi potency

To investigate the influence of 3′-overhangs on RNAi activity, we measured the target knockdown efficacy of the group I duplexes using a quantitative real-time reverse transcription-PCR system (qRT-PCR) (**Figure 3**). The original 21-mer siRNA was more potent than the shifted-21-mer (62 versus 40% knockdown at 50 nmol/l and 40 versus 15% knockdown at 10 nmol/l). Compared with the original 21-mer siRNA, the most potent knockdown was mediated by the 27 (tt)/27 (UU) duplex, providing 78 and 60% knockdown at 50 and 10 nmol/l, respectively. The 25 (GC) /27 (UU) and 27 (UU)/27 (UU) duplexes enhanced (~10%) the RNAi potency, whereas other 27-mer duplexes were slightly less potent than the original 21-mer siRNA. The *in vitro* Dicer-cleavage reactions for the 27 (tt)/27 (UU) duplex predominantly generated "L-R" cleavage products identical in size to the original 21-mer siRNA. In contrast, major "R-L" dicing products were produced from the 27 (UU)/27 (tt) duplex. The "L-R′ cleavage species generate the more effective siRNAs.

The enhanced potency of the "L-R" cleavage might be attributed to Dicer-generated products which result in preferential handoff of the AS strand to RISC. Correspondingly, the unfavorable cleavage species deriving from "R-L" dicing displayed reduced RNAi. These results suggest that the 27-mer duplex can yield a specific, desired 21-mer species which is able to enhance RNAi activity, even though the only differences among the group of 27-mer duplexes are the 3′-overhangs.

Asymmetric DsiRNAs containing a single 3′-DNA blunt terminus improve RNAi potency

The "L-R" and "R-L" dicing patterns could be in equilibrium and therefore compete with each other generating a mixture of siRNAs. By preventing one of the two dicing directionalities *via* an unfavorable 3′-overhang it is possible to promote a single Dicer entry pattern. Since the 3′-overhang plays a

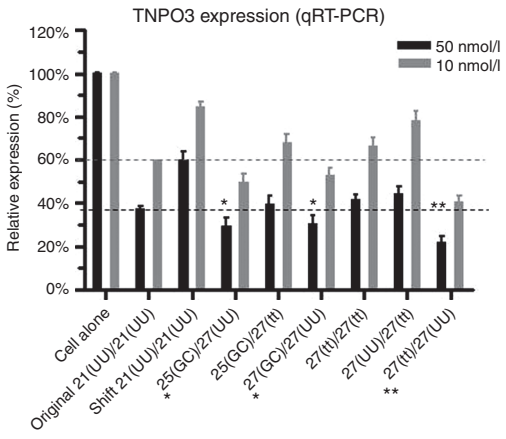


Figure 3 Silencing of *Homo sapiens* transportin 3 (*TNPO3*) by group I duplex Dicer-substrate small interfering RNAs (DsiRNAs). HEK 293 cells were transfected with 50 or 10 nmol/l of the experimental anti-*TNPO3* duplex DsiRNAs. *TNPO3* mRNA levels were detected by quantitative real-time reverse transcription (qRT-PCR). The data are normalized with GAPDH mRNA levels and represent the average of three replicate assays.

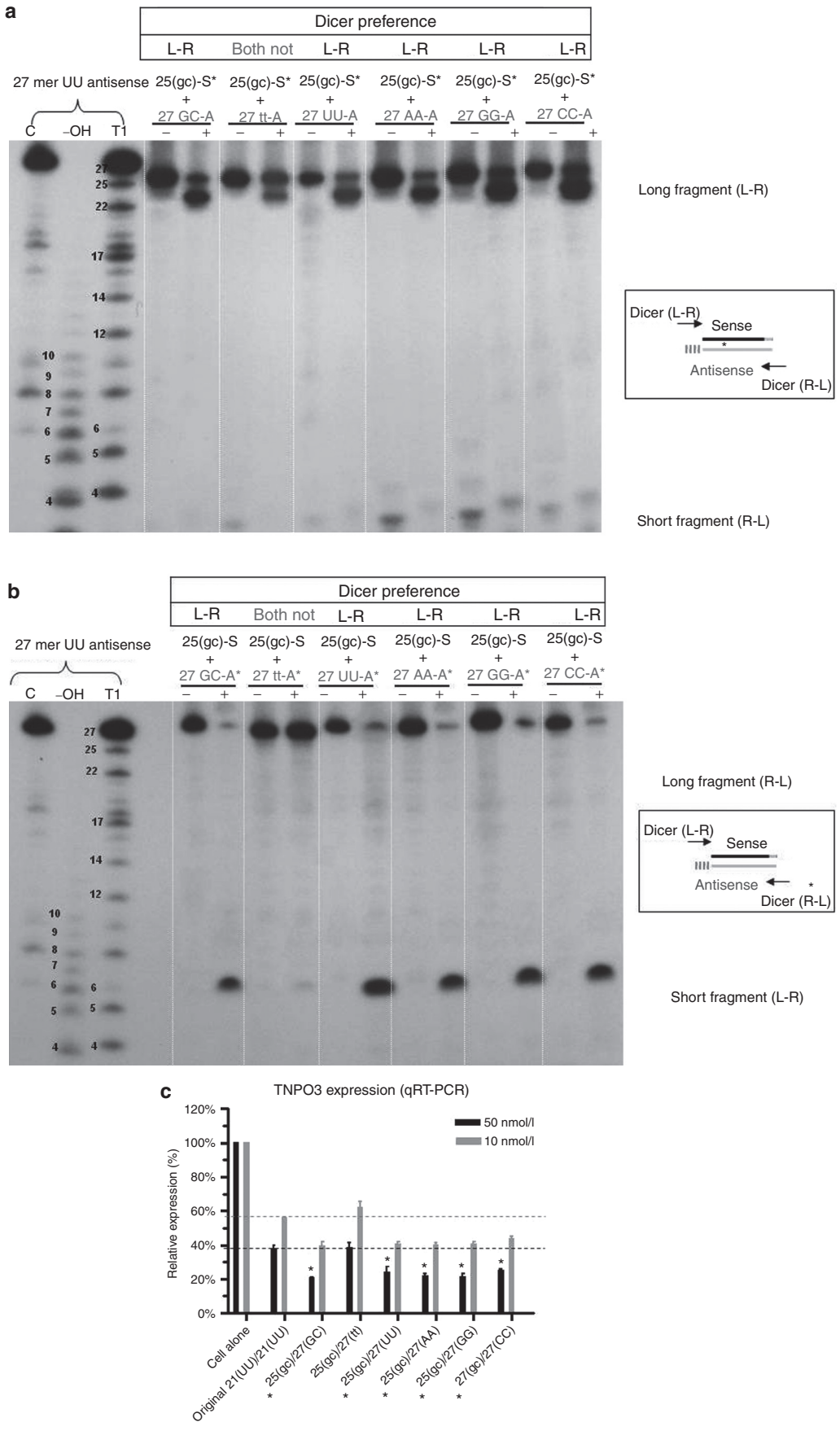
determinant role in orienting Dicer entry and RNAi efficiency, it is possible to generate the desired cleavage species *via* simply incorporating a favorable 3′ overhang and an unfavorable 3′ blunt end without altering the original 27-mer duplex sequence. For group I, 27-mer duplexes, the presence of a 3′-dTdT overhang or an RNA containing blunt end did not completely abolish Dicer bidirectional entry and different dicing patterns. To further restrict the dicing preference and entry orientation, we designed a series of asymmetric duplexes (group II in **Table 1**) with a single two-nucleotide 3′-overhang on the AS strand and two-nucleotide DNA residues at the 3′-blunt end of S strand, to provide a single favorable Dicer entry and a restricted dicing pattern.

In vitro Dicer processing of the group II duplexes was examined as previously described (**Figure 4b,c**). As expected, the desired "L-R" products derived from all these asymmetric duplexes were the overwhelming majority indicating that the two-base deoxyribonucleotide 3′ blunt end strongly impeded Dicer entry. When compared with the duplexes of group I, the asymmetric group II duplexes greatly simplified the *in vitro* dicing pattern, supporting the fact that terminal DNA residues impeded Dicer entry and can be used to direct Dicer entry onto a DsiRNA to obtain a single desired cleavage product.

To further evaluate the asymmetric design, we evaluated the target knockdown efficacy of these RNA duplexes (**Figure 4c**) *via* a qRT-PCR assay. As we observed in the dicing reactions, the asymmetric duplexes with various 3′ two-ribonucleotide overhangs had enhanced RNAi potency, whereas the duplex having a 3′-dTdT overhang had less efficient target knockdown efficiency. Moreover, the duplexes with a 3′ two-ribonucleotide overhang from group II showed comparable knockdown efficacies (**Figure 4c**) as well as dicing activities (**Figure 4b**).

The 3′-overhang sequences influence the guide strand selection of the asymmetric DsiRNAs

To identify whether the sequence composition of the 3′ two-base overhang is important for Dicer recognition and specificity, we used Illumina Deep sequencing analyses to


a
Dicer preference
L-R
Both not
L-R
L-R
L-R
L-R
27 mer UU antisense
25(gc)-S*
+
27 GC-A
27 tt-A
27 UU-A
27 AA-A
27 GG-A
27 CC-A
C
–OH
T1
Long fragment (L-R)
Dicer (L-R)
Sense
Antisense
Dicer (R-L)
Short fragment (R-L)
b
Dicer preference
25(gc)-S
27 GC-A*
27 tt-A*
27 UU-A*
27 AA-A*
27 GG-A*
27 CC-A*
Long fragment (R-L)
Short fragment (L-R)
c
TNPO3 expression (qRT-PCR)
50 nmol/l
10 nmol/l
Relative expression (%)
120%
100%
80%
60%
40%
20%
0%
Cell alone
Original 21(UU)/21(UU)
25(gc)/27(GC)
25(gc)/27(tt)
25(gc)/27(UU)
25(gc)/27(AA)
25(gc)/27(GG)
27(gc)/27(CC)

investigate the intracellular dicing products of the asymmetric 25/27-mer duplexes transfected in HEK293 cells. At 48 hours post-transfection with 10 nmol/l of these DsiRNAs, total RNAs were isolated and prepared for Illumina sequencing.

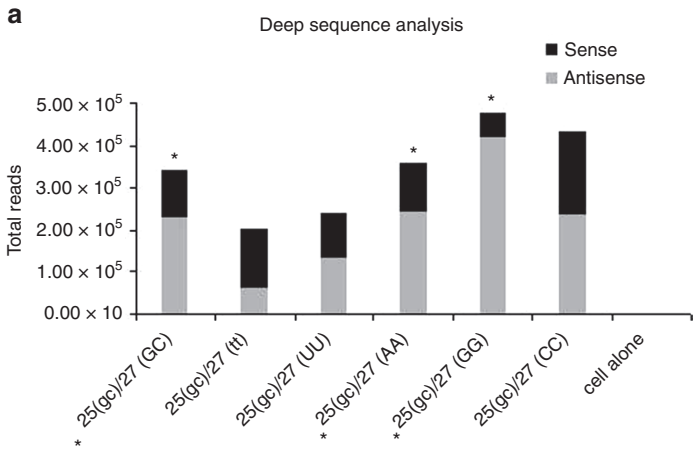
As shown in **Table 2a**, all the samples had similar total reads. From the total reads specific sequences were collected and aligned using the referenced sequences of the asymmetric DsiRNAs (AS or S strands). Surprisingly, the data demonstrated that the total reads of these RNA duplexes and the proportion of the AS strand to S strands were clearly distinct among the different duplexes (**Figure 5a,b**). Unequal transfection efficiencies or sample processing might cause such differences; however, the relative strand distributions suggest that the observed differences can largely be attributed to the 3′ overhang compositions. The relative abundance of AS to S sequences declined relative to the two-base overhangs as follows: GG > GC, AA > CC, UU >> tt. For example, despite the close number of total reads for the 25 (gc)/27 (GG) duplex (477,396) and the 25 (gc)/27 (CC) duplex (431,423) these two duplexes have distinctly different strand distributions with

Table 2 Illumina Deep sequence analyses and IC_{50} values of asymmetric 27-mer *TNPO3* Dicer-substrate siRNAs (group II)

All sequences	***25(gc)/27 (GC)***	***25(gc)/27 (tt)***	***25(gc)/27 (UU)***	***25(gc)/27 (AA)***	***25(gc)/27 (GG)***	***25(gc)/27 (CC)***	***Cell alone***
(a) Total reads of deep sequences, fraction and abundance of sense and antisense strands from each duplex are listed. To calculate the fractions, reads for each strand were divided by the total number of sense or antisense reads. The strand distribution was calculated as the ratio of the abundance of antisense to sense.							
Total reads from deep sequence	28,068,694	29,921,531	30,384,765	30,686,615	29,867,099	29,814,503	24,949,154
Total reads of antisense and sense	340,051	201,128	239,788	358,466	477,396	431,423	137
Reads of antisense	229,347	64,970	135,285	242,439	418,187	236,366	63
Reads of sense	110,704	136,158	104,503	116,027	59,209	195,057	74
Percentage of antisense	67.445%	32.303%	56.419%	67.632%	87.598%	54.788%	45.985%
Percentage of sense	32.555%	67.697%	43.581%	32.368%	12.402%	45.212%	54.015%
as/s ratio	2.0717	0.4772	1.2946	2.0895	7.0629	1.2118	0.8514
Abundance of antisense GG > GC, CC, AA > UU >> tt. Abundance of sense GG < GC, AA, UU < tt < CC. Ratio of antisense to sense GG > GC, AA > CC, UU >> tt.							
Top 10 sequences	***25(gc)/27 (GC)***	***25(gc)/27 (tt)***	***25(gc)/27 (UU)***	***25(gc)/27 (AA)***	***25(gc)/27 (GG)***	***25(gc)/27 (CC)***	
(b) Total reads of the top 10 sense and antisense strands from each duplex and the fraction of the desired "L-R" cleavage products from each strand are listed. L-R product reads from the sense or antisense strands were divided by the number of reads of each strand, respectively.							
Total reads of antisense and sense	237,338	134,334	183,025	228,955	318,968	317,477	
Reads of antisense	147,544	41,748	94,741	136,615	274,202	153,203	
"L-R" products from antisense	136,054	4,862	81,461	115,259	274,202	145,567	
Percentage of "L-R" products	92.21%	11.65%	85.98%	84.37%	100.00%	95.02%	
Reads of Sense	89,794	92,586	88,284	92,340	44,766	164,274	
"L-R" products from sense	88,380	86,928	87,604	87,824	42,409	161,621	
Percentage of "L-R" products	98.43%	93.89%	99.23%	95.11%	94.73%	98.39%	
Abundance of "L-R" products from Antisene: GG > GC, CC, AA > UU >> tt. Abundance of "L-R" products from Sene: GG < GC, AA, UU, tt < CC.							
psiCHECK assay (RNAi potency)	***25(gc)/27 (GC)***	***25(gc)/27 (tt)***	***25(gc)/27 (UU)***	***25(gc)/27 (AA)***	***25(gc)/27 (GG)***	***25(gc)/27 (CC)***	
(c) IC_{50} values of asymmetric 27-mer duplexes were determined using the psiCHECK assay as described in Materials and Methods. When the antisense or sense strands were used as the "guide" strand, their target knockdown efficiency is listed as the IC_{50}. The strand selectivity was calculated as the ratio of IC_{50} values of sense versus antisense targeting.							
IC_{50} of antisense (pmol/l)	61.34 ± 1.18	987.3 ± 1.31	119.6 ± 1.30	30.57 ± 1.18	21.06 ± 1.12	43.95 ± 1.21	
IC_{50} of sense (pmol/l)	1356 ± 1.55	2257 ± 1.46	307 ± 1.39	601.3 ± 1.24	8252 ± 1.60	186.2 ± 1.30	
Selectivity of AS to S	22.11	2.29	2.57	19.67	391.83	4.24	
RNAi potency of Antisense GG > AA, CC > GC > UU >> tt. RNAi potency of Sense GG << tt < GC < AA < UU < CC. Selectivity of Antisense to Sense GG >> GC, AA > CC > UU, tt.							

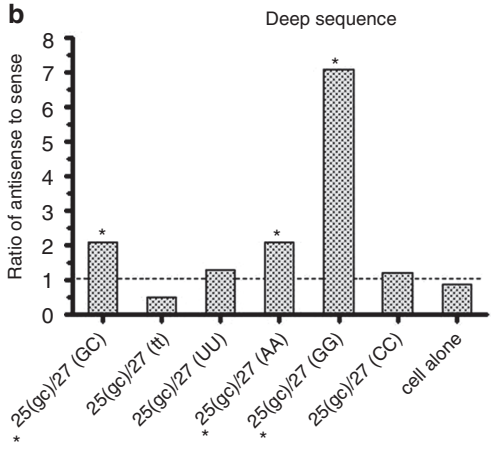
DsiRNA, Dicer-substrate small interfering RNA; *TNPO3*, *Homo sapiens* transportin 3; siRNA, small interfering RNA.

Figure 4 ***In vitro*** **Dicer cleavage of 5′ P^{32}-end labeled sense or antisense strands.** Group II asymmetric duplex Dicer-substrate small interfering RNAs (DsiRNAs) and RNA interference (RNAi) potency. The cleavage fragments that result from dicing L-R and R-L were visualized by autoradiography of the P^{32} 5′ end-labeled strands: (**a**) sense and (**b**) antisense. Two experimental approaches (method A and B) and the dicing pattern are defined the legend to Figure 1. (**c**) Silencing of *Homo sapiens* transportin 3 (*TNPO3*) triggered by group II duplex RNAs. HEK 293 cells were transfected with 50 or 10 nmol/l of the experimental anti-*TNPO3* duplex DsiRNAs. TNPO3 mRNA level were detected by quantitative real-time reverse transcription (qRT-PCR). The data were normalized with GAPDH mRNA and represent the average of three replicate assays.



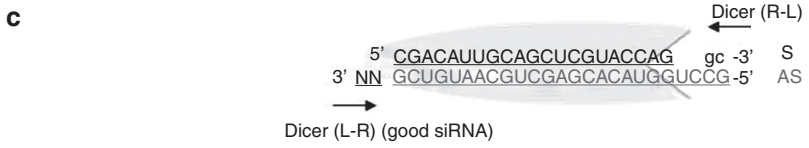
Asymmetric 27-mer duplexes with various 3'-overhangs

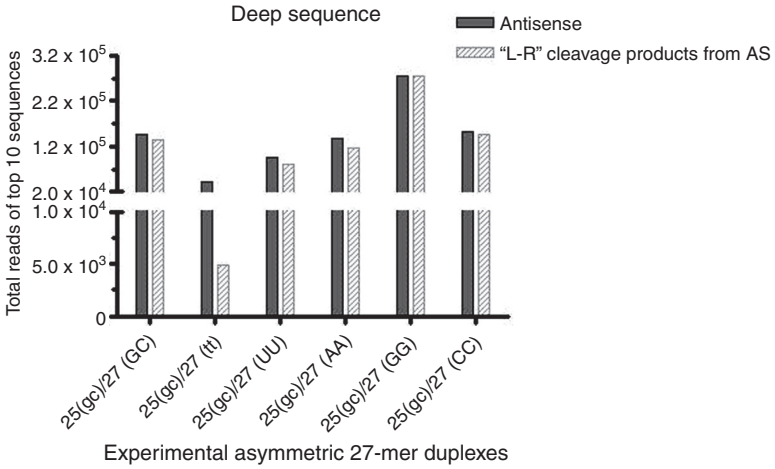
Abundance of antisense: GG > GC, CC, AA > UU >> tt
Abundance of sense: GG > GC, AA, UU < tt < CC



Experimental asymmetric 27-mer duplexes

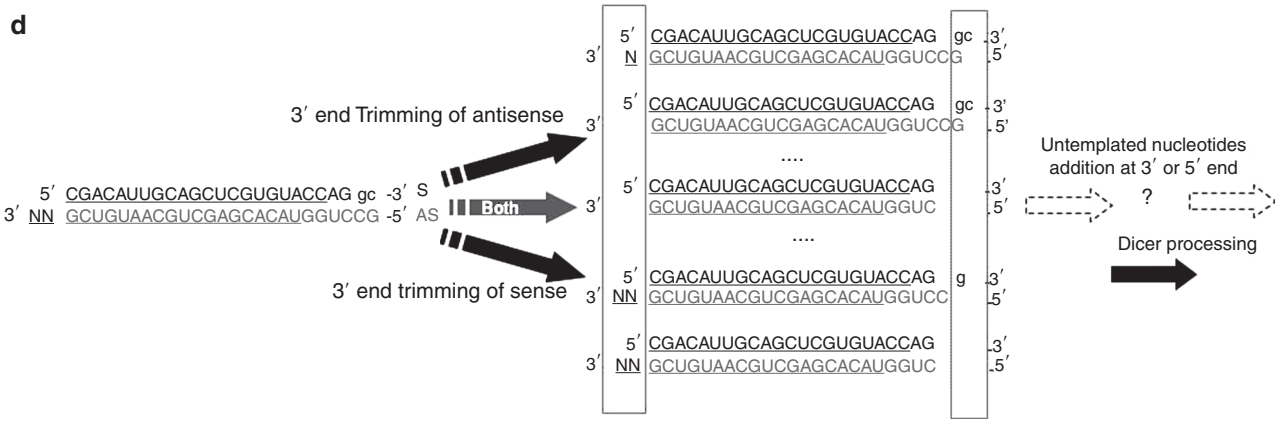
Ratio of AS/S: GG > GC, AA > CC, UU >> tt





Experimental asymmetric 27-mer duplexes

Abundance of "L-R" cleavage products from AS: GG > GC, CC, AA > UU >>tt



the relative ratios of AS to S being 7.06 and 1.21, respectively. Similarly, the 25 (gc)/27 (UU) duplex (239,788) and the 25 (gc)/27 (tt) duplex (201,128) have a similar number of total reads, whereas their relative strand distributions are very different. Therefore, these data demonstrate that the sequence composition of the 3′ overhang plays an important role in determining the fates of these DsiRNAs with respect to the dicing pattern, RISC loading and strand stability.

Manual inspection of the AS and S strands focusing only on the top 10 sequences, narrowed the list to 60~70% of the total reads (**Table 2b**), but this analysis still maintained the same propensity for relative abundance of AS/S (such as, GG > GC, AA > CC, UU >> tt). We further identified the Dicer-cleaved RNA species for the top 10 sequences. **Figure 5c** and **Supplementary Figure S1** display two major dicing patterns as previously determined: "L-R" cleavage generating the desired siRNAs and "R-L" cleavage producing the undesired siRNAs. Consistent with the *in vitro* dicing assay, the asymmetric duplexes having a 3′ two-ribonucleotide overhang were predictably processed into the desired "L-R" cleavage products of 21-22 mers in cells (**Supplementary Figure S1**). The population of "L-R" cleavage products is summarized in **Table 2b**. Approximately 85% of the AS strands were preferentially generated from "L-R" cleavage products. In contrast, the AS strand of the 25 (gc)/27 (tt) duplex only produced 11.65% of the desired "L-R" cleavage species, which corresponds with the *in vitro* dicing results.

Interestingly, we also observed RNA editing of both strands, such as trimming of a single or double-nucleotide on the 3′ end or addition of untemplated nucleotides to the 3′ or 5′ termini. As shown in **Figure 5d** and **Supplementary Figure S1**, trimming of the 3′ end occurred on both strands, especially when DNA residues were incorporated into the overhangs or the blunt end. Untemplated nucleotides were added to the 3′ or 5′-termini, most likely following the trimming or dicing reactions. For example, the 3′-GC overhang on the AS strand of the 25 (gc)/27 (GC) duplex was trimmed by a single "C" and subsequently extended by a U or A (**Supplementary Figure S1a**, entry 7 and 8). For the 25 (gc)/27 (tt) duplex, after its 3′ DNA (gc) blunt end on the S strand was trimmed, the resulting RNA was subjected to uridylation of the 3′ end (**Supplementary Figure S1b** S, entry 6, 7, and 10). Moreover, we also found that nucleotides most frequently added to these duplexes were U and A. The extension of a U or A was much more prevalent on the 3′ end than on the 5′ end. Although the RNA editing appeared to be widespread among all the duplexes, the frequency is generally lower than 10% of the total reads. The highest frequency of editing exceeded 50% in the 25 (gc)/27 (GC) duplex, suggesting the 3′ overhang composition also affects RNA extension and ultimately influences the Dicer-cleavage site and RNAi potency.

Figure 5 Illumina Deep sequence analyses of asymmetric 27-mer *Homo sapiens* transportin 3 (*TNPO3*) Dicer-substrate siRNAs (group II). HEK 293 cells were transfected with 10 nmol/l of the asymmetric group II RNA duplexes. Forty hours post-transfection the total RNAs were isolated and prepared for Illumina Deep sequencing. The data collection and alignment are described in Materials and Methods section. (**a**) The total reads and abundance of sense and antisense strands from each duplex. (**b**) The strand distribution was calculated as the ratio of the abundance of antisense to sense. The ratio of antisense (AS) to sense (S) is ranked by the 3′ overhang GG > GC, AA > CC, UU >> tt. (**c**) The dicing pattern L-R and R-L are as previously described. The L-R pattern generates the desired siRNA species for target knockdown. Total reads from the top 10 antisense strands and the abundance of L-R cleavage products from the antisense strand. (**d**) Two types of RNA editing: trimming of the 3′ end and post-transcriptional addition of nucleotides at the 3′ or 5′ ends.

The 3′-overhang sequence compositions influence strand selectivity and RNAi activity of asymmetric DsiRNAs

It is known that selection of the guide strand depends upon the differences in the thermodynamic stability of the two ends of the 21/22-mer duplexes.[5,26] The 3′ two-nucleotide overhangs can alter the stability of the duplex ends, eventually affecting strand selection by the Argonaute proteins. Argonaute 2 binds the guide strand in an oriented manner which facilitates cleavage of the passenger strand of the siRNA during RISC loading.[27] The potential for competition of the passenger strand for RISC entry necessitates proper design strategies to optimize the desired guide strand selection and to minimize off target activity by the undesired passenger strand.

To evaluate the RNAi activities of the asymmetric duplexes, we used the psiCHECK reporter system in which the target sequence for the AS or S strands are inserted in the 3′ UTR of the *Renilla luciferase* gene. The siRNA-mediated inhibition of luciferase activity for both the S and AS target orientations were independently tested (**Figure 6a** and **Table 2c**). The strand selectivity was calculated as a measure of the relative target inhibition efficiencies for each target orientation (**Figure 6b** and **Table 2c**). When compared with the duplex harboring a 3′ dTdT overhang, all the duplexes with 3′ two base-ribonucleotide overhangs tested in this study showed superior inhibition of luciferase expression (IC_{50} 20–100 pmol/l) when using the AS guide strand against the S target. However, when the S strands were used as guides for the AS orientation of the target, relatively lower efficiencies of target inhibition were observed (IC_{50} 200–8,000 pmol/l). The strand selectivity (**Figure 6b**) favoring the AS strand relative to the S strand as guides for these asymmetric duplexes demonstrated that the asymmetric design favors handoff to RISC of the AS strand as guide following dicing. Moreover, there was a profound bias in AS versus S strand function for the 3′ GG overhang. The data revealed that the most efficacious strand was the most abundant and the relative frequencies of the "S" or "AS" strands are highly correlated with the silencing activity and strand selectivity. In brief, the rankings for the RNAi efficiency for of the various 3′ overhangs is GG > GC, CC, AA > UU >> tt. Correspondingly, the abundance of the "L-R" products for the various 3′ two-base overhangs is also GG > GC, CC, AA > UU >> tt.

Validation of the role of the two-base 3′-overhang for a different pair of asymmetric DsiRNAs

To verify the role of the 3′-overhang in guide strand selection and function, we analyzed a separate set of asymmetric DsiRNAs that target *hnRNP H1*. These two asymmetric DsiRNAs differ by a single nucleotide, but have strikingly different

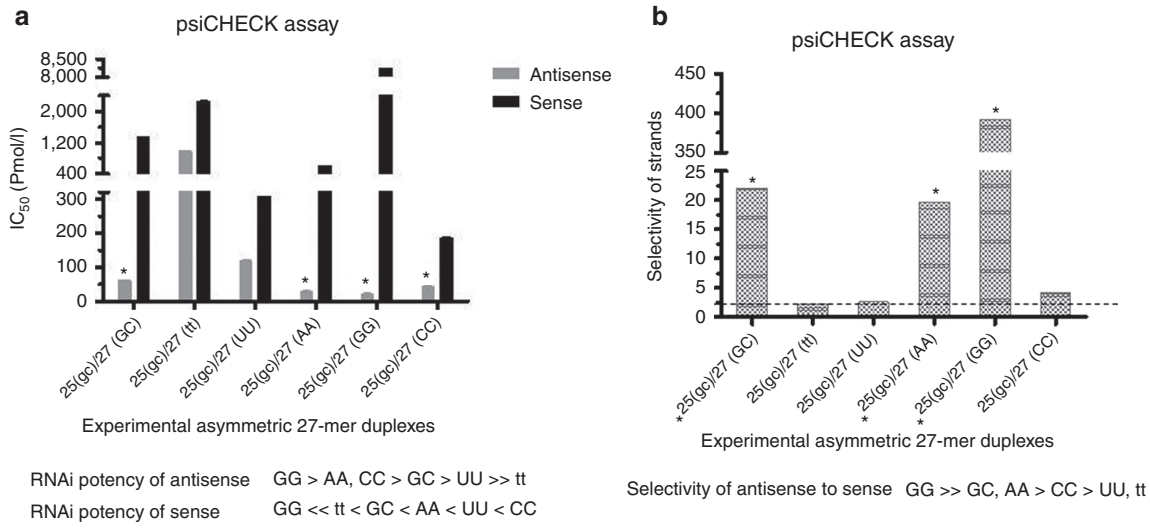


Figure 6 IC_{50} values and strand selectivity of asymmetric 27-mer *Homo sapiens* transportin 3 (*TNPO3*) Dicer-substrate siRNAs (group II). (**a**) IC_{50} values of the asymmetric 27-mer duplexes were determined using the psiCHECK assays as described Materials and Methods section. When the antisense or sense was used as "guide" strand, the target knockdown efficiency is listed as the IC_{50}. **(b)** The strand selectivity was calculated as the ratio of IC_{50} values of the sense to antisense targets.

functional activities, and thus represent an interesting test for the role of the two-base 3′ overhangs (**Table 3a**). The RNAi potencies of these RNAs were also evaluated using the psiCHECK system. As depicted in **Table 3b**, the knockdown efficiencies of both the S and AS strands of these asymmetric *hnRNP H1* DsiRNAs were independently assessed and the strand selectivity was calculated.

There is only a one-base shift along the target mRNA sequence between the site 325 DsiRNAs and the site 324 DsiRNAs. The site 324 DsiRNAs showed better overall RNAi efficiency (combined knockdown of both S and AS targets) when compared with the site 325 DsiRNAs containing the same 3′-overhangs. The site 324 DsiRNAs showed a comparable knockdown efficiency for either the S or AS strands as guides against their corresponding targets (**Table 3b**). Despite the weak strand selectivity for the 324 DsiRNAs, there was a relative tendency for biased strand selectivity for the 25 (gg)/27 (GG) duplexes compared to the 3′-CC, UU and AA overhangs. Substantial strand selectivity was observed for the site 325 DsiRNAs. For example, the selectivity of AS to S of the 25 (tg)/27 (GG) duplex is 6.73 and the 25 (tg)/27 (AA) duplex is 8.86, respectively. For the site 325 DsiRNAs it was interesting to see that the 3′-AA overhang provided the best RNAi activity and strongest strand selectivity. In this case, the 3′-AA overhang makes the siRNA completely complementary to the *hnRNP H1* target site, implying a perfectly matched 3′-overhang probably contributes to enhanced RNAi activity. Additionally, we also chose two representative DsiRNAs from each site to evaluate the target knockdown efficacy *via* a qRT-PCR assay (**Supplementary Figure S3**). Consistent with our observation in the *TNPO3* case, the asymmetric duplexes with perfectly matched 3′-overhangs have enhanced RNAi potency, whereas the duplex having a 3′-dTdT overhang had less efficient target knockdown efficiency. For example, the site 325 DsiRNA 25 (tg)/27 (AA) showed a lower IC_{50} value (44.37 pmol/l) compared with the 25 (tg)/27 (tt) duplex (IC_{50} = 284 pmol/l).

We also used Illumina Deep sequencing analyses to interrogate the relative strand abundance of the *hnRNP H1* asymmetric 25/27-mer duplexes. Total RNAs were isolated and prepared for Illumina Deep sequencing at 24 hours post-transfection of these DsiRNAs in HCT116 cells. According to their RNAi potency and strand selectivity, we chose two representative DsiRNAs for each target site. For example, the site 324 DsiRNAs 25 (tg)/27 (GG) and 25 (tg)/27 (AA) that showed a big difference in RNAi activity and selectivity were selected for deep sequence analyses. Similarly, deep sequencing analyses were performed with the DsiRNA 325 25 (gg)/27 (GG) and 25 (gg)/27 (CC) duplexes. To simplify the comparisons we only focused on the top 10 sequences. As shown in **Table 3b** and **Figure 7a,b**, the total reads of these RNA duplexes and the proportions of the AS to S strands were distinct among the experimental duplexes. In similarity to our previous observation with the *TNPO3* DsiRNAs, the relative abundances of the AS to S strands for both site 324 and 325 DsiRNAs were consistent with the strand selectivity in the RNAi assays. For example, the relative distribution of AS to S (**Figure 7b**) ratios ranked the 25 (tg)/27 (GG) >> 25 (tg)/ 27 (AA) for site 324, and GG > CC for site 325 which is the same trend for strand selectivity shown in **Table 3b**. Thus these observations validated that the 3′-overhang composition contributes to the fates of the DsiRNAs and ultimately RNAi activity.

As previously described for the *TNPO3* DsiRNA system, the two major Dicer-cleavage products ("L-R" and "R-L") in the top 10 sequences were identified (**Figure 7c** and **Supplementary Figure S2**). **Table 3c** lists the population of "L-R" cleavage products producing the desired siRNA species. Consistent with the results obtained with the *TNPO3* system, >80% of the AS strands or S strands of the *hnRNP H1* DsiRNAs preferentially produced the desired primary "L-R" cleavage products. As shown in **Supplementary Figure S2**, RNA editing took place at the ends of both strands as previously observed. For example, trimming of a single or double-

Table 3 Deep sequence analyses and IC_{50} values of the *hnRNP H1* 25/27-mer Dicer-substrate siRNAs

hnRNP H1 25/27-mer Dicer-substrate siRNAs (DsiRNAs)		***hnRNP H1 target sequence***
		5′--- CT T TGAATCAGAAGATGAAGTCAAA TT GG ---3′
(**a**) The asymmetric 27-mer duplex RNAs against *hnRNP H1* (the site 324 and site 325 DsiRNAs) in this study and the target sequence are listed. The sense strand is presented from 5′ to 3′ and is marked as black and the antisense strand is presented from 3′ to 5′ and is marked as gray. Two-nucleotide 3′-overhangs are underlined. Ribonucleotides are in upper case and deoxyribonucleotides are in lower case. These RNAs are designated by their strand length and overhangs: the number indicates the length of RNA strands; "(NN)" means the two-base of 3′ ends		
Site 324 DsiRNAs	**25 (tg)/27 (CC)**	**5′ UGAAUCAGAAGAUGAAGUCAAAU tg 3′** **3′ CC ACUUAGUCUUCUACUUCAGUUUA AC**
	25 (tg)/27 (UU)	**5′ UGAAUCAGAAGAUGAAGUCAAAU tg 3′** **3′ UU ACUUAGUCUUCUACUUCAGUUUA AC**
	25 (tg)/27 (GG)	**5′ UGAAUCAGAAGAUGAAGUCAAAU tg 3′** **3′ GG ACUUAGUCUUCUACUUCAGUUUA AC**
	25 (tg)/27 (AA)	**5′ UGAAUCAGAAGAUGAAGUCAAAU tg 3′** **3′ AA ACUUAGUCUUCUACUUCAGUUUA AC**
Site 325 DsiRNAs	**25 (gg)/27 (CC)**	**5′ GAAUCAGAAGAUGAAGUCAAAUU gg 3′** **3′ CC CUUAGUCUUCUACUUCAGUUUAA CC**
	25 (gg)/27 (UU)	**5′ GAAUCAGAAGAUGAAGUCAAAUU gg 3′** **3′ UU CUUAGUCUUCUACUUCAGUUUAA CC**
	25 (gg)/27 (GG)	**5′ GAAUCAGAAGAUGAAGUCAAAUU gg 3′** **3′ GG CUUAGUCUUCUACUUCAGUUUAA CC**
	25 (gg)/27 (AA)	**5′ GAAUCAGAAGAUGAAGUCAAAUU gg 3′** **3′ AA CUUAGUCUUCUACUUCAGUUUAA CC**

psiCHECK assay (RNAi potency)	***Site 324 dsiRNAs (GA-natural match)***			
	25(tg)/27 (CC)	***25(tg)/27 (UU)***	***25(tg)/27 (GG)***	***25(tg)/27 (AA)***
(**b**) IC_{50} values of asymmetric 27-mer duplexes were determined using the psiCHECK assay. When the antisense or sense strand was the "guide" strand, their target knockdown efficiencies are listed as IC_{50} values. The strand selectivity was calculated as the ratio of IC_{50} values of sense to antisense strand target knockdown				
IC_{50} of antisense (pmol/l)	26.56 ± 1.08	34.24 ± 1.1	16.00 ± 1.06	35.04 ± 1.11
IC_{50} of sense (pmol/l)	12.72 ± 1.08	14.14 ± 1.08	16.03 ± 1.06	11.84 ± 1.08
Selectivity of AS to S	0.48	0.41	1.00	0.34
RNAi potency of antisense	GG > CC > UU, AA			
RNAi potency of sense	AA, CC > UU, GG			
Selectivity of antisense to sense	GG > CC, UU > AA			
psiCHECK assay (RNAi potency)	***Site 325 dsiRNAs (AA-natural match)***			
	25(gg)/27 (CC)	***25(gg)/27 (UU)***	***25(gg)/27 (GG)***	***25(gg)/27 (AA)***
IC_{50} of Antisense (pmol/l)	61.54 ± 1.09	91.84 ± 1.08	31.82 ± 1.1	5.73 ± 1.07
IC_{50} of Sense (pmol/l)	56.15 ± 1.11	296.1 ± 1.11	214.3 ± 1.11	50.77 ± 1.09
Selectivity of AS to S	0.91	3.22	6.73	8.86
RNAi potency of antisense	AA > GG > CC > UU			
RNAi potency of sense	AA, CC > GG, UU			
Selectivity of antisense to sense	AA > GG > CC > UU			
Top 10 sequences	***Site 324 25(tg)/27 (AA)***	***Site 324 25(tg)/27 (GG)***	***Site 325 25(gg)/27 (CC)***	***Site 325 25(gg)/27 (GG)***
(c) Illumina Deep sequence analyses of *hnRNP H1* DsiRNAs. HCT116 cells were transfected with 10 nmol/l of the asymmetric duplexes site 324 25 (tg)/27 (AA), site 324 25 (tg)/27 (GG), site 325 25 (gg)/27 (CC) and site 325 25 (gg)/27 (GG). The data collection and alignment protocols are described in Materials and Methods. The total reads of the top 10 sense and antisense strands from each duplex and the fraction of desired "L-R" cleavage products from each strand are listed. To calculate this fraction, "L-R" product reads from sense or antisense strands were divided by the separate reads of each strand				
Total reads of antisense and sense	158,404	1,322,436	446,025	744,339
Reads of antisense	91,124	1,281,110	431,859	732,830
"L-R" products from antisense	75,601	1,174,340	431,859	675,568
Percentage of "L-R" products	82.96%	91.67%	100.00%	92.19%
Reads of sense	67,280	41,326	14,166	11,509
"L-R" products from sense	67,280	41,326	14,166	11,509
Percentage of "L-R" products	100.00%	100.00%	100.00%	100.00%

DsiRNA, Dicer-substrate small interfering RNA; *hnRNP H1*, heterogeneous nuclear ribonucleoprotein H; siRNA, small interfering RNA.

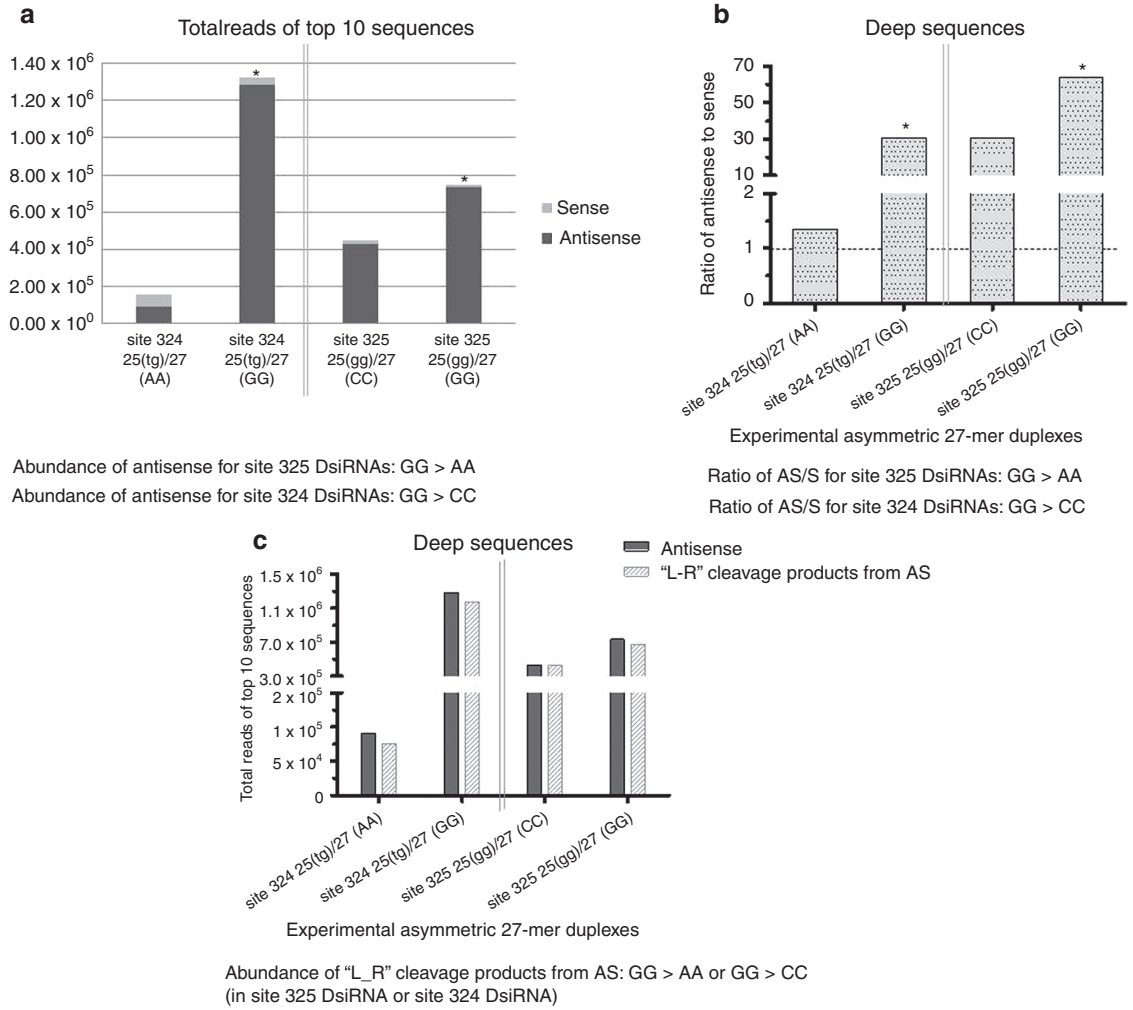


Figure 7 Illumina Deep sequence analyses of asymmetric 27-mer heterogeneous nuclear ribonucleoprotein H (*hnRNP H1*) Dicer-substrate small interfering RNA (siRNAs). The data collection and alignment were as described above. (**a**) Total reads of the top 10 sense and antisense strands from each duplex. (**b**) The strand distribution was calculated as the ratio of the abundance of antisense to sense. Ratio of antisense (AS) to sense (S) is ranked by the 3′ overhang GG > AA in site 324 Dicer-substrate small interfering RNAs (DsiRNAs) and GG > CC in site 325 DsiRNAs. (**c**) The dicing pattern L-R and R-L are as previously described. The L-R model generates the desired siRNA species for target knockdown. The total reads of the top 10 antisense strands and the abundance of L-R cleavage products are presented.

nucleotide at the 3′ end occurred often on the AS strand. Also 1 or 2 untemplated nucleotides (such as "U," "C," "CC," "UU," or "CU") were most frequently added to the 3′ end of both strands. Although RNA editing appears to be a widespread phenomenon among all the tested duplexes, the mechanism is unknown.

Discussion

The RNAse III family member Dicer initiates RNAi by processing double-stranded RNAs into 21–23 nt double-stranded RNAs (either miRNAs or siRNAs) generating a two-base 3′-overhang.[7,28] In association with Dicer, the siRNA products are loaded into RISC such that only one of the original strands is incorporated and used as a guide for the sequence-specific post-transcriptional silencing of cognate genes. The remaining strand, known as the antiguide or passenger strand, is degraded. Previous studies have demonstrated in mammals that the PAZ domain of Dicer is a single-stranded RNA-binding module that has preference for two-base single strand overhangs produced by another RNAse III family member Drosha, which processes primary miRNAs into pre-miRNAs.[6,11,29] The PAZ/PIWI domains of the Ago2 protein serve as anchors to spatially orient the bound RNA substrates in the enzyme active site.[30,31] Therefore, DsiRNAs with two-nucleotide 3′-overhangs that are favorable for Dicer binding/cleavage and subsequent Ago anchoring are believed to enhance RNAi potency.[9] Addition of RNA tetraloops, unfavorable DNA residues or fluorescent groups at the ends of double-stranded RNAs have been demonstrated to partially or completely block cleavage by human Dicer.[6,16]

The single-stranded two-base 3′ overhangs present in traditional 21-mer siRNAs are required for optimal siRNA function. As a regular practice in designing siRNAs, the two-base 3′-deoxynucleotide overhangs (such as "tt") are often added to

19-mer siRNAs, often without regard to complementarity with the target sequence. However, here we find that when this feature is applied to 25/27-mer DsiRNAs, it adversely affects dicing and subsequent RNAi activity. Moreover, we also find that the sequence composition of ribose 3′ two-base overhangs significantly affect dicing polarity and strand selectivity.

We first carried out investigations of *in vitro* dicing products using symmetric 25 base pair duplexes with various overhang configurations (group I, **Table 1**). Our results showed that a 3′ tt overhang attenuates Dicer entry onto the substrate while 3′ UU overhangs are favorable for Dicing entry and subsequent cleavage. When Dicer enters the substrate from either end of a duplex ("L-R" or "R-L" models), resulting in heterogeneous cleavage products, these can consequently impact on RNAi potency. Even though the only minor difference among all the symmetric 27-mer DsiRNAs tested is the sequence composition of the two-base 3′ overhang, the 27 (tt)/27 (UU) DsiRNA preferentially generated more "L-R" dicing products generating the desired siRNAs with the highest target knockdown efficiencies, whereas other duplexes with a 3′ tt overhang on the AS strand were primarily processed by "R-L" Dicer entry generating siRNAs with poor efficacy. Because a 3′=tt overhang or a ribose blunt end does not completely abolish Dicer entry/cleavage in the group I design, bidirectional dicing products were still observed. We further restricted the dicing preference in order to obtain the desired "L-R" product through rational design of 3′-termini. These optimized asymmetric DsiRNAs (group II, **Table 1**) have a single, favorable 3′ two-ribonucleotide overhang on the AS strand and an unfavorable two-nucleotide blunt DNA residue on the 3′ end of the S strand[22] which simplifies the Dicing pattern, predictably generating a single, desired "L-R" product, thereby enhancing RNAi potency. It is noteworthy that these design features can provide an optimal structure for binding by the Dicer PAZ domain along with an anchor site for Ago2 in RISC. The increased RNAi efficacy mediated by the 27-mer DsiRNAs can be attributed to two features (i) the major "L-R" products generate siRNAs with desired sequences; and (ii) the "L-R" products that have a 3′ two-base overhang on the AS strand result in preferential utilization of the AS strand as a guide in the RNAi machinery.

Since "L-R" and "R-L" Dicer entry patterns can coexist in a dicing reaction, they will compete each other. By blocking "R-L" dicing *via* the 3′-DNA blunt end on the S strand and simultaneously facilitating "L-R" entry *via* optimized 3′ two-base ribonuleotide overhangs on the desired AS/guide strand, it is possible to significantly promote the desired "L-R" dicing products and strand selectivity as previously demonstrated.[22] Furthermore, the RNA slicing-based silencing pathway is involved in multiple cycles of target binding, cleavage and product release mediated by the Argonaute 2 protein. In this scenario, the Argonaute 2 protein remains bound to the guide strand promoting guide strand selection by slicing the passenger strand, thereby establishing and protecting the guide strand from degradation.[4,5]

We have been interested in determining the influence of the sequence composition of the two-base 3′ overhangs on Dicer processing and strand selectivity. We have examined the fates of asymmetric *TNPO3* duplexes in cells by Illumina Deep sequencing analyses. Additionally the RNAi potencies of both the "S" and "AS" strands derived from these duplexes were also evaluated by psiCHECK reporter gene assays. In similarity to our *in vitro* Dicer assays, the asymmetric duplexes with two-base 3′ ribonucleotide overhangs were predictably and primarily processed into the desired "L-R" cleavage products of 21–22 mers in cells (**Table 2** and **Supplementary Figure S1**).

The effect of the sequence composition of the two-base 3′ overhang is an important point of discussion. In several previous studies, a preference in Dicer binding and Dicer-cleavage efficiency for DsiRNAs containing purine/purine nucleotide 3′ overhangs over pyrimidine/pyrimidine 3′-overhangs was observed. Dicer binding and activity was ordered by the 3′ overhangs CC > GC > GG > AA > UU,[15] GG > AA > UU > CC > tt,[9] and AA > GC >> GG > CC > UU.[19] However, other studies showed different tendencies. For example, a molecular dynamic simulation study[32] indicated that the 3′-UU overhang made a relatively more stable complex with the PAZ domain compared to 3′-GG, AA, and CC overhangs (UU > GG > AA > CC). In addition, an asymmetric 27 mer DsiRNA with a 3′-UU overhang on the AS strand was shown to be the most potent inhibitor of gene expression compared to DsiRNAs having different sequence compositions of the 3′ overhang (UU > GG, GC, CC > AA).[17,22] It was noted that in this case the 3′-UU overhang was completely complementary to the target, implying a perfectly matched 3′-overhang may contribute to RNAi activity. Our data for both the *TNPO3* and *hnRNP H1* targets also demonstrated that asymmetric DsiRNAs with a 3′-overhang complementary to the target mRNA have enhanced RNAi potency. Although the optimal sequence composition of the 3′ termini is still controversial overall for RNAi, the 3′ overhang composition undoubtedly contributes to Dicer binding and RNAi potency. Our data further demonstrate that the composition of the 3′ two-base overhangs significantly influences the relative strand abundance and selectivity into RISC. Our observation that the most efficacious strands were also the most abundant revealed that the relative frequencies of "S" or "AS" strands are highly correlated with overall strand selectivity and hence silencing activity. For example, in the *TNPO3* DsiRNA system, there is a strong bias in the abundance of "L-R" products and selection of the AS strand with the order of preference for two-base 3′ overhang being GG > GC, CC, AA > UU >> tt. Similarly, the strand distribution (Ratio of AS to S) follows the same 3′ overhang order of GG > GC, AA > CC, UU >> tt. Similar observations on biases dictated by the two-base overhang were also found for another set of asymmetric 25/27-mer duplexes targeting the *hnRNP H1* mRNA further validating the significance of the role of the sequence composition of the 3′ two-base overhang. With both the *TNPO3* and *hnRNP H1* mRNAs, our results support the observation that DsiRNAs with purine/purine (GG, AA) nucleotide overhangs are generally preferred over pyrimidine/pyrimidine overhangs (CC, UU).

In addition to the role of the two-base overhang in DsiRNA selection and RNAi, we have observed two types of RNA editing on both strands of the processed DsiRNAs. These are trimming of a single or double-nucleotide on the 3′ end and addition of untemplated nucleotides to the 3′ or 5′ termini subsequent to the trimming (**Figure 5d** and **Supplementary Figures S1** and **S2**). The deep sequences reveal that the

trimming of the 3′ end occurs on both strands following Dicer processing. There is more trimming when DNA residues were incorporated into the overhangs or blunt end. Moreover, the nucleotides most frequently added to these duplexes are U and A. The addition of a U or A was much more prevalent on the 3′ end than on the 5′ end. Interestingly, the highest frequency of RNA editing, which exceeded 50% of the sequences analyzed was observed for the *TNPO3* 25 (gc)/27 (GC) duplex, in which the 3′ end of the AS was extended by a U or A. Considering the direct relationship between sequence abundance and RNAi activity of the AS strand, we conjecture that the untemplated nucleotide additions to the 3′ end of the AS strand somehow facilitate RISC loading. It remains to be investigated what enzymes are participating in the RNA editing. We do not know if the addition of untemplated nucleotides takes place before or after Dicer cleavage.

In conclusion, consistent with previous reports,[22,24] our data demonstrate that 3′ RNA residues are more favorable than DNA residues for Dicer processing of DsiRNAs. Dicer entry onto DsiRNAs is oriented by the nature of the 3′ ends. In addition, the entry of Dicer also influences guide strand selection and ultimately RNAi potency. Since the PAZ domain is sensitive to the type of 3′ overhang sequence composition,[6] Dicing patterns could be predictably controlled through the rational design of the 3′ end. Furthermore, the sequence composition of the 3′ two-base overhang appears to influence RNA editing of the siRNAs, as well as the Dicing pattern and recruitment of the siRNA guide strand into the RISC. Further research will be needed to understand the mechanism of RNA recognition and editing events in the RNAi pathway.

Materials and methods

Materials. Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich (St Louis, MO), T4 PNK enzymes and buffer were obtained from New England BioLabs (Ipswich, MA) and all cell culture products were purchased from GIBOC (Gibco BRL/Life Technologies (Carlsbad, CA), a division of Invitrogen, Carlsbad, CA). The cell lines HEK 293 and HCT116 are from the ATCC (Manassas, VA). Random primers (Invitrogen); Lipofectamine 2000 (Invitrogen).

siRNAs. All the siRNAs were synthesized and purified using high-performance liquid chromatography at Integrated DNA Technologies (Coralville, IA). All RNA duplexes against *TNPO3* (group I and group II) used in this study are listed in the **Table 1**. All RNA duplexes against *hnRNP H1* used in this study are listed in the **Table 3a**.

In vitro Dicer assays. All the S strands and AS strands were end-labeled with T4 polynucleotide kinase and γ-^{32}P-ATP. Unlabeled S or AS RNAs were annealed with equal molar amounts of 5′-end-labeled corresponding AS or S strands in HBS buffer in order to form siRNA duplexes. siRNA duplexes (1 pmol) were incubated at 37 °C for 40 minutes in the presence or in the absence of 1 U of human recombinant Dicer enzyme following the manufacturer's recommendations (Ambion, Austin, TX). Reactions were stopped by phenol/chloroform extraction and the resulting solutions were electrophoresed in a 20% polyacrylamide denaturing gel. The gels were subsequently exposed to X-ray film.

Cell culture. HEK 293 cells and HCT116 cells were purchased from ATCC and cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum according to their respective data sheets. Cells were cultured in a humidified 5% CO_2 incubator at 37 °C.

Determination of TNPO3 gene silencing (qRT-PCR analysis). HEK 293 cells were split in 24-well plates to 60–70% confluency in Dulbecco's modified Eagle's medium media 1 day before transfection. The cells were transfected with 10 or 50 nmol/l of experimental anti-*TNPO3* duplex RNAs using Lipofectamine 2000 following the manufacturer's recommendations (Invitrogen). Forty eight hours post-transfection total RNAs were isolated with TriZol reagent (Invitrogen). Expression of the *TNPO3* gene was analyzed by quantitative real time-PCR using a 2× iQ SyberGreen Mastermix (Bio-Rad) and specific primer sets at a final concentration of 400 nmol/l. Primers were as follows: *TNPO3* forward primer: 5′-CCT GGA AGG GAT GTG TGC-3′; *TNPO3* reverse primer: 5′-AAA AAG GCA AAG AAG TCA CAT CA-3′; GAPDH forward primer: 5′ -CAT TGA CCT CAA CTA CAT G-3′; GAPDH reverse primer: 5′-TCT CCA TGG TGG TGA AGA C-3′.

TriZol reagent was used to extract total RNA according to the manufacturer's instruction (Invitrogen). Residual DNA was digested using the DNA-free kit per the manufacturer's instructions (Ambion). cDNA was produced using 2 µg of total RNA, Moloney murine leukemia virus reverse transcriptase and random primers in a 15 µl reaction according to the manufacturer's instructions (Invitrogen).GAPDH expression was used for normalization of the qPCR data.

Illumina Deep sequence and data analysis. HEK 293 cells were split into 24-well plates at 60–70% confluency in Dulbecco's modified Eagle's medium media one day prior to transfection. The cells were transfected with 10 nmol/l of the asymmetric group II RNA duplexes using Lipofectamine 2000 following the manufacturer's recommendations (Invitrogen). Forty eight hours post-transfection the total RNAs were isolated with TriZol reagent (Invitrogen) and prepared for Illumina Deep sequencing.

To identify the most frequent S and antisense products from each of the DsiRNA molecules, the sequences generated from Illumina Pipeline v1.6 were aligned with the S and AS strands of each siRNA molecule using Novoalign v2.05 (http://www.novocraft.com). All subsequent analyses were carried out using the R statistical environment and Bioconductor packages "Biostrings" and "ShortRead."[33] Only sequences that could be aligned to the siRNA sequences without mismatches were retained. The relative starting and ending positions of the siRNA sequences were determined based on their aligned positions and lengths, and the frequency of each product was counted. Only the 10 most frequent products are reported.

To examine whether there are nucleotide additions or deletions at either end of the Dicer processed products, the raw sequences were matched to the siRNA AS sequence with a seed size of 16 after removing the 3′-adapter using the Bioconductor package "ShortRead." For example, for a siRNA

sequence length of 23, the Illumina sequences were aligned totally to eight seeds which are the subsequences from bases 1–16, 2–17, and so on, of the original siRNA sequence. The matched sequences were then reduced to a set of unique sequences along with their number of occurrences. This set of sequences was then aligned with the siRNA reference sequence using the ClustalX2 multiple alignment tool[34] not allowing gaps. The multiple aligned sequences were visualized and exported using JalView.[35] The extra bases at either end of the product were highlighted manually.

Dual luciferase assay (detection of IC_{50} value). The 45 base pair oligomer of *TNPO3* cDNA was inserted into the *Spe*I and *Xho*I restriction endonuclease sites downstream of the humanized *Renilla luciferase* gene in the psiCHECK-2 vector (Promega, Fitchburg, WI) to generate plasmids psi-CHECK-*TNPO3*-AS (passenger strand reporter) and psi-CHECK-*TNPO3*-S (guide strand reporter). HCT116 cells were cotransfected in a 96-well format (25,000 cells/well) with 10 ng of the respective psiCHECK-*TNPO3*-AS or psi-CHECK-*TNPO3*-S vector, 100 fmol/l–50 nmol/l DsiRNAs and 0.1 µl Lipofectamine2000 (Invitrogen) per well. Cells were lysed in 1× Passive Lysis Buffer (Promega) 24 hours after transfection and analyzed using the Dual-Luciferase Reporter System (Promega) on a Veritas microplate luminometer (Turner Biosystems, Sunnyvale, CA). The average values were calculated from three replicates to set *Renilla/Firefly* luciferase expression to 100%. An IC_{50} curve was generated using Prism 5.01 software (GraphPad, La Jolla, CA). Sigmoidal dose responses were calculated according to $Y = \text{Bottom} + (\text{Top} - \text{Bottom})/(1 + 10)\,Y\,((\text{LogEC}_{50} - X))$; where X is the logarithm of concentration and Y is the response.

The sequences of Oligomers.

TNPO3 AS

TNPO3 AS_S: 5′-CCg CTCGAG ggagcaaagc cgacattgca gctcgtgtac caggcagtgc aggcg ACTAGT CC-3′;

TNPO3 AS_AS: 5′-GG ACTAGT cgcct gcactgcctg gtacacgagc tgCaatgtcg gctttgctcc CTCGAG CGG-3′

TNPO3 sense(S)

TNPO3 S_S: 5′-CCG CTCGAG cgcct gcactgcctg gtacacgagc tgCaatgtcg gctttgctcc ACTAGT CC-3′;

TNPO3 S_AS: 5′-GG ACTAGT ggagcaaagc cgacattgca gctcgtgtac caggcagtgc aggcg CTCGAG CGG-3′

The fragment of 343 base pair in *hnRNP H1* cDNA includes the region of bases 90–432 in the reference sequence NM_005520. The reporter plasmids psi-*hnRNP H*-S (sense reporter) and psi-*hnRNP H*-AS (AS reporter) were derived by cloning the hnRNPH sequences in the 3′-UTR of the humanized *Renilla luciferase* gene in the psiCHECK-2 (Promega).[22]

Supplementary material

Figure S1. Illumina Deep sequence analysis of asymmetric *TNPO3* 27-mer Dicer-substrate siRNAs (group II).

Figure S2. Illumina Deep sequence analyses of asymmetric 27-mer *hnRNP H1* Dicer-substrate siRNAs.

Figure S3. Silencing of *hnRNP H1* by the site 324 and 325 duplex DsiRNAs.

Acknowledgments. We thank Kumi Sakurai, Britta Hoehn, and Soifer Harris for helpful discussions. We thank City of Hope DNA sequencing core (Harry Gao and Jinhui Wang) for Illumina Deep Sequencing and City of Hope Bioinformatics Core facility (Xiwei Wu and Haiqing Li) for data analyses. This work was supported by grants from the National Institutes of Health awarded to J.J.R. AI29329, AI42552, and HL07470. J.J.R. and M.A.B. are cofounders of Dicerna Pharmaceuticals, a pharmaceutical company based on Dicer-substrate technology.

1. Fire, A, Xu, S, Montgomery, MK, Kostas, SA, Driver, SE and Mello, CC (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**: 806–811.
2. Zamore, PD, Tuschl, T, Sharp, PA and Bartel, DP (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25–33.
3. Hammond, SM, Bernstein, E, Beach, D and Hannon, GJ (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**: 293–296.
4. Parker, JS (2010). How to slice: snapshots of Argonaute in action. *Silence* **1**: 3.
5. Jinek, M and Doudna, JA (2009). A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**: 405–412.
6. Ma, JB, Ye, K and Patel, DJ (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**: 318–322.
7. Zhang, H, Kolb, FA, Jaskiewicz, L, Westhof, E and Filipowicz, W (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**: 57–68.
8. Zhang, H, Kolb, FA, 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.
9. Lima, WF, Murray, H, Nichols, JG, Wu, H, Sun, H, Prakash, TP *et al.* (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.
10. Hammond, SM, Boettcher, S, Caudy, AA, Kobayashi, R and Hannon, GJ (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**: 1146–1150.
11. Song, JJ, Liu, J, Tolia, NH, Schneiderman, J, Smith, SK, Martienssen, RA *et al.* (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10**: 1026–1032.
12. Liu, J, Carmell, MA, Rivas, FV, Marsden, CG, Thomson, JM, Song, JJ *et al.* (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**: 1437–1441.
13. O'Toole, AS, Miller, S, Haines, N, Zink, MC and Serra, MJ (2006). Comprehensive thermodynamic analysis of 3' double-nucleotide overhangs neighboring Watson-Crick terminal base pairs. *Nucleic Acids Res* **34**: 3338–3344.
14. O'Toole, AS, Miller, S and Serra, MJ (2005). Stability of 3' double nucleotide overhangs that model the 3' ends of siRNA. *RNA* **11**: 512–516.
15. Vermeulen, A, Behlen, L, Reynolds, A, Wolfson, A, Marshall, WS, Karpilow, J *et al.* (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA* **11**: 674–682.
16. Siolas, D, Lerner, C, Burchard, J, Ge, W, Linsley, PS, Paddison, PJ *et al.* (2005). Synthetic shRNAs as potent RNAi triggers. *Nat Biotechnol* **23**: 227–231.
17. Kim, DH, Behlke, MA, Rose, SD, Chang, MS, Choi, S and Rossi, JJ (2005). Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat Biotechnol* **23**: 222–226.
18. Reynolds, A, Leake, D, Boese, Q, Scaringe, S, Marshall, WS and Khvorova, A (2004). Rational siRNA design for RNA interference. *Nat Biotechnol* **22**: 326–330.
19. DiNitto, JP, Wang, L and Wu, JC (2010). Continuous fluorescence-based method for assessing dicer cleavage efficiency reveals 3' overhang nucleotide preference. *BioTechniques* **48**: 303–311.
20. Holen, T, Amarzguioui, M, Wiiger, MT, 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.
21. Nowotny, M and Yang, W (2009). Structural and functional modules in RNA interference. *Curr Opin Struct Biol* **19**: 286–293.
22. Rose, SD, Kim, DH, Amarzguioui, M, Heidel, JD, Collingwood, MA, Davis, ME *et al.* (2005). Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res* **33**: 4140–4156.
23. Amarzguioui, M and Rossi, JJ (2008). Principles of Dicer substrate (D-siRNA) design and function. *Methods Mol Biol* **442**: 3–10.
24. Amarzguioui, M, Lundberg, P, Cantin, E, Hagstrom, J, Behlke, MA and Rossi, JJ (2006). Rational design and *in vitro* and *in vivo* delivery of Dicer substrate siRNA. *Nat Protoc* **1**: 508–517.
25. Brass, AL, Dykxhoorn, DM, Benita, Y, Yan, N, Engelman, A, Xavier, RJ *et al.* (2008). Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319**: 921–926.
26. Wang, B, Li, S, Qi, HH, Chowdhury, D, Shi, Y and Novina, CD (2009). Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. *Nat Struct Mol Biol* **16**: 1259–1266.

27. Wang, Y, Juranek, S, Li, H, Sheng, G, Wardle, GS, Tuschl, T *et al.* (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**: 754–761.
28. Gan, J, Shaw, G, Tropea, JE, Waugh, DS, Court, DL and Ji, X (2008). A stepwise model for double-stranded RNA processing by ribonuclease III. *Mol Microbiol* **67**: 143–154.
29. Lingel, A, Simon, B, Izaurralde, E and Sattler, M (2004). Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* **11**: 576–577.
30. Ma, JB, Yuan, YR, Meister, G, Pei, Y, Tuschl, T and Patel, DJ (2005). Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**: 666–670.
31. Parker, JS, Roe, SM and Barford, D (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**: 663–666.
32. Lee, HS, Lee, SN, Joo, CH, Lee, H, Lee, HS, Yoon, SY *et al.* (2007). Contributions of 3′-overhang to the dissociation of small interfering RNAs from the PAZ domain: molecular dynamics simulation study. *J Mol Graph Model* **25**: 784–793.
33. Morgan, M, Anders, S, Lawrence, M, Aboyoun, P, Pagès, H and Gentleman, R (2009). ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **25**: 2607–2608.
34. Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
35. Waterhouse, AM, Procter, JB, Martin, DM, Clamp, M and Barton, GJ (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

Supplementary Information accompanies this paper on the Molecular Therapy–Nucleic Acids website (http://www.nature.com/mtna)