

Phosphate-binding pocket in Dicer-2 PAZ domain for high-fidelity siRNA production

Suresh K. Kandasamy^a and Ryuya Fukunaga^{a,1}

^aDepartment of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205

Edited by V. Narry Kim, Seoul National University, Seoul, Republic of Korea, and approved October 31, 2016 (received for review July 27, 2016)

The enzyme Dicer produces small silencing RNAs such as micro-RNAs (miRNAs) and small interfering RNAs (siRNAs). In *Drosophila*, Dicer-1 produces ~22–24-nt miRNAs from pre-miRNAs, whereas Dicer-2 makes 21-nt siRNAs from long double-stranded RNAs (dsRNAs). How Dicer-2 precisely makes 21-nt siRNAs with a remarkably high fidelity is unknown. Here we report that recognition of the 5'-monophosphate of a long dsRNA substrate by a phosphate-binding pocket in the Dicer-2 PAZ (Piwi, Argonaute, and Zwiille/Pinhead) domain is crucial for the length fidelity, but not the efficiency, in 21-nt siRNA production. Loss of the length fidelity, meaning increased length heterogeneity of siRNAs, caused by point mutations in the phosphate-binding pocket of the Dicer-2 PAZ domain decreased RNA silencing activity in vivo, showing the importance of the high fidelity to make 21-nt siRNAs. We propose that the 5'-monophosphate of a long dsRNA substrate is anchored by the phosphate-binding pocket in the Dicer-2 PAZ domain and the distance between the pocket and the RNA cleavage active site in the RNaseIII domain corresponds to the 21-nt pitch in the A-form duplex of a long dsRNA substrate, resulting in high-fidelity 21-nt siRNA production. This study sheds light on the molecular mechanism by which Dicer-2 produces 21-nt siRNAs with a remarkably high fidelity for efficient RNA silencing.

Dicer | siRNA | dsRNA | phosphate | RNA silencing

The enzyme Dicer produces small silencing RNAs such as micro-RNAs (miRNAs) and small interfering RNAs (siRNAs). The single human Dicer (also called Dicer1 based on its sequence similarity to *Drosophila* Dicer-1) produces both miRNAs and siRNAs. In contrast, *Drosophila* has two Dicers: Dicer-1, dedicated to the production of miRNAs, and Dicer-2, for siRNAs. *Drosophila* Dicer-1 makes ~22–24-nt miRNAs from pre-miRNAs, whereas Dicer-2 precisely produces 21-nt siRNAs from long dsRNAs (1). siRNA-dedicated Dicer-2 is found in arthropods including insects that transmit vector-borne diseases. Endogenous sources of long dsRNA for siRNA production include viral RNA genomes or replication intermediates, transposon transcripts, partially self-complementary hairpin RNAs, and convergent mRNA transcription (2–5). Long dsRNA can be introduced experimentally to trigger RNA interference (RNAi). After their production, miRNAs and siRNAs are loaded into Argonaute proteins, for which they act as sequence-specific guides. The difference in the length between miRNAs and siRNAs in arthropods is important because the miRNA effector protein Argonaute1 and the siRNA effector Argonaute2 preferentially bind ~22–24-nt RNAs and 21-nt RNAs, respectively (6). The miRNA–Argonaute1 complexes repress mRNA translation and promote mRNA turnover, whereas the siRNA–Argonaute2 complexes cleave their RNA targets.

Drosophila Dicer-1 and Dicer-2 and human Dicer share a common domain architecture (Fig. 1A): an N-terminal “helicase” domain; a central, atypical dsRNA-binding domain (dsRBD, previously known as DUF283); a PAZ (Piwi, Argonaute, and Zwiille/Pinhead) domain; two RNaseIII domains; and a C-terminal dsRBD. Only the helicase domain of *Drosophila* Dicer-2, but not that of *Drosophila* Dicer-1 or human Dicer, binds and hydrolyzes ATP. Dicer-2 is thought to use ATP to translocate along long dsRNA substrates, allowing it to produce siRNAs

processively from an end (7–9). The PAZ domain, which is named after three of the proteins containing it—Piwi, Argonaute, and Zwiille/Pinhead (10)—is found in both Dicer and Argonaute proteins and recognizes the characteristic 2-nt, 3'-overhang of RNA left by Drosha and Dicer cleavage (11, 12).

How Dicer-2 precisely produces siRNAs of exactly 21 nt length with a remarkably high fidelity is unknown. Previous studies suggested several possible mechanisms by which Dicer enzymes define small RNA length (13); these include the 3' counting rule (14, 15), the 5' counting rule (16), and the loop counting rule (17). The crystal structure of *Giardia* Dicer, which lacks the N-terminal helicase domain and the C-terminal dsRBD found in animal Dicers, suggested that it uses the 3' counting rule to define the product RNA length. In the 3' counting rule, the 3' 2-nt overhang of a dsRNA substrate is anchored to the PAZ domain and the specified distance between the anchored helical end of the dsRNA and the RNase active site in the RNaseIII domains serves as a molecular ruler (14, 15). In contrast, recent studies suggested the 5' counting rule, in which anchoring the 5' monophosphate, not the 3' overhang, of some pre-miRNAs in the PAZ domain of human Dicer determines the length of miRNAs (16, 18). Other studies suggested the loop-counting rule, in which human Dicer cleaves the stem of an RNA substrate at a position 2 nt away from the terminal loop (17). However, none of these studies can explain how *Drosophila* Dicer-2 precisely produces small RNAs from long dsRNA substrates, which lack a terminal loop compared with the miRNA-producing Dicers including *Drosophila* Dicer-1 and human Dicer that cleave pre-miRNAs containing a terminal loop.

The crystal structures of the human Dicer fragment containing the PAZ domain showed that it has a phosphate-binding pocket

Significance

In *Drosophila*, Dicer-1 produces ~22–24-nt micro-RNAs (miRNAs) from pre-miRNAs, whereas Dicer-2 makes 21-nt siRNAs from long dsRNAs. siRNAs function in antiviral and antitransposon immunity. The molecular mechanism by which the length of small RNAs produced by Dicer enzymes is defined is not fully understood. We found that high fidelity in 21-nt siRNA production by Dicer-2 is important for efficient RNA silencing. We also found that recognition of the terminal monophosphate on RNA substrates by the Dicer-2 phosphate-binding pocket is crucial for high fidelity in siRNA production and mutation in the pocket increased length heterogeneity of siRNAs. Our study sheds light on the importance and the molecular mechanism of high-fidelity siRNA production.

Author contributions: R.F. designed research; S.K.K. and R.F. performed research; R.F. analyzed data; and R.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE84532).

¹To whom correspondence should be addressed. Email: fukunaga@jhmi.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1612393113/-DCSupplemental.

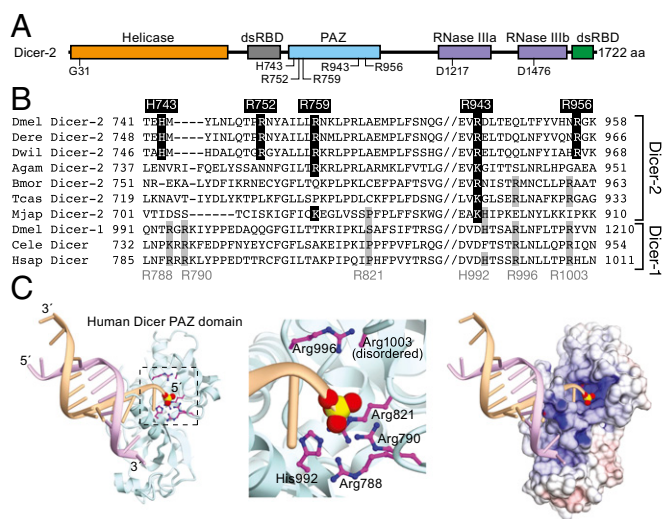


Fig. 1. A phosphate-binding pocket in Dicer PAZ domains. (A) Domain structures of *Drosophila melanogaster* Dicer-2. The residues mutated in this study are indicated. (B) Alignment of the PAZ domain sequences from Dicers from arthropods, *Caenorhabditis elegans*, and human. Agam, *Anopheles gambiae*; Bmor, *Bombyx mori*; Cele, *C. elegans*; Dere, *Drosophila erecta*; Dmel, *Drosophila melanogaster*; Dwil, *Drosophila willistoni*; Hsap, *Homo sapiens*; Mjap, *Marsupenaeus japonicus*; Tcas, *Tribolium castaneum*. (C) A crystal structure of a human Dicer fragment containing the PAZ domain bound with 10-bp dsRNA (light pink and orange) with 5'-monophosphorylated frayed ends [Protein Data Bank (PDB) ID code 4NH5]. The dsRNA 5'-monophosphate (shown as a ball and stick model in yellow and red) is bound in the phosphate-binding pocket, where it is recognized by the side chains of Arg788, Arg790, Arg821, and His992 (shown as a ball and stick model in magenta). (Right) The electrostatic potential map.

that recognizes the 5' monophosphate of pre-miRNAs (16, 18). The recognition of the 5' monophosphate of pre-miRNAs by the phosphate-binding pocket in the human Dicer PAZ domain is important for both the efficiency and fidelity in miRNA production (16). These studies suggested that basic residues that comprise the phosphate-binding pocket in human Dicer are conserved in miRNA-producing Dicers (Dicer-1) but not in siRNA-producing Dicers (Dicer-2). Therefore, they proposed that only miRNA-producing Dicers, but not siRNA-producing Dicers, have a phosphate-binding pocket. However, by multiple sequence alignment, we identified several basic amino acid residues that are conserved among Dicer-2 proteins and are located near the positions where human Dicer has the conserved basic residues comprising the phosphate-binding pocket (Fig. 1 B and C). We previously found that both the 5' monophosphate of RNA substrates and two arginine residues (Arg943 and Arg956) in the *Drosophila* Dicer-2 PAZ domain are required for cleavage of short dsRNAs, but not long dsRNAs, in vitro (7, 19). Based on these findings, we proposed that Dicer-2 has a unique phosphate-binding pocket in its PAZ domain, which is composed of the arginine residues, and that it is required to bind the 5' monophosphate of short dsRNA substrates but not long dsRNAs (7, 19). However, the role of the phosphate-binding pocket in the Dicer-2 PAZ domain, which is not required for the long dsRNA processing but causes the risk to process short dsRNA substrates such as pre-miRNAs, remained unknown.

In this study, we report that recognition of the 5' monophosphate on long dsRNA substrates by the unique phosphate-binding pocket of *Drosophila* Dicer-2 is crucial for the length fidelity, but not the efficiency, in 21-nt siRNA production. We also report that high-fidelity production of 21-nt siRNAs is important for efficient RNA silencing activity.

Results and Discussion

5' Monophosphate of Long dsRNA Substrates Is Important for Length Fidelity in siRNA Production in Vitro. To probe the role of a terminal phosphate of long dsRNA in siRNA production, we first asked whether the length fidelity in siRNA production is affected if a long dsRNA substrate lacks a 5' monophosphate. We analyzed the length of the first, second, and third siRNAs (counting from the dsRNA end) produced by recombinant wild-type Dicer-2 protein in vitro from 5' monophosphate (Fig. 2 A–C) or 5' hydroxyl (Fig. 2 D–F) 73-bp dsRNAs ³²P-radiolabeled at a specific internal position. The other end of the dsRNAs was blocked by two deoxynucleotides (7). The length fidelity in the first siRNA production from the 73-bp dsRNA with a 5' monophosphate end was high; the ratios of 20-, 21-, and 22-nt siRNAs were 6 ± 1%, 85 ± 1%, and 9 ± 1%, respectively (Fig. 2). In contrast, that from the 73-bp dsRNA with a 5' hydroxyl was significantly lower; the ratios of 20-, 21-, and 22-nt siRNAs were 11 ± 0%, 52 ± 1%, and 37 ± 2%, respectively (*P* value < 0.000001). Dicer-2 cleavage leaves a 5' monophosphate on both the produced siRNA duplex and the remaining long dsRNA. Therefore, for the second and third siRNAs produced during processive siRNA generation, the intermediate long dsRNA substrates have a 5' monophosphate even if the original unprocessed long dsRNA had a 5' hydroxyl in the first siRNA production. We found that the fidelity in the second and third siRNA production was high in both the initial 5' monophosphate 73-bp dsRNA substrate and the initial 5' hydroxyl 73-bp dsRNA substrate. The ratios of 21-nt siRNA in the second and third siRNA production were 96 ± 2% and 91 ± 1%, respectively, from the 5' monophosphate 73-bp dsRNA and 96 ± 1% and 91 ± 2%, respectively, from the 5' hydroxyl 73-bp dsRNA. Together, we concluded that if a long dsRNA substrate lacks a 5' monophosphate, the length fidelity in only the first siRNA production is lowered, meaning that the length

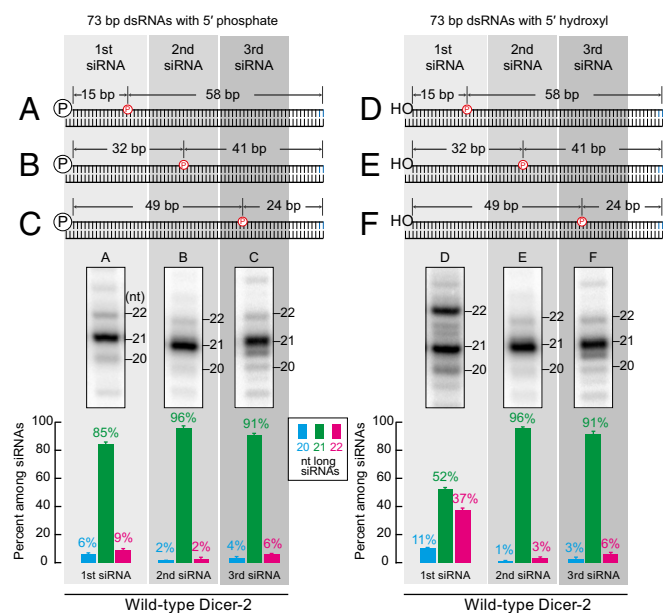


Fig. 2. 5' monophosphate of long dsRNA is crucial for fidelity in the first siRNA production. Shown are the lengths of the first, second, and third siRNAs produced from 100 nM of the site-specifically ³²P-radiolabeled (indicated in orange) 73-bp dsRNAs with 2-nt 3' overhang and either 5' monophosphate (dsRNAs, A–C) or 5' hydroxyl (dsRNAs, D–F) by 8 nM of the recombinant wild-type Dicer-2. The other end of the 73-bp dsRNAs was blocked by two deoxynucleotides (blue). Representative gel images and quantification are shown. Data are mean ± SD for three or four independent trials.

heterogeneity of the first siRNA is increased, while the fidelity is recovered in the subsequent processive siRNA production.

Design of Phosphate-Binding Pocket Mutant Dicer-2. To probe the role of the unique phosphate-binding pocket in the Dicer-2 PAZ domain in vivo, we created transgenic flies expressing point mutants of Dicer-2 in which the phosphate-binding pocket is disrupted. We replaced Arg943 and Arg956 with alanine (Phos-mut1: R943A, R956A). The amino acid residue corresponding to Arg943 is highly conserved as a basic residue (arginine or lysine) among arthropod Dicer-2, whereas Arg956 is conserved among *Drosophila* species (Fig. 1B). We previously showed in vitro that the recombinant protein of Phos-mut1 Dicer-2 processes long dsRNAs as efficiently as wild-type Dicer-2 but cannot process short dsRNAs, unlike wild-type Dicer-2 (19). Because we were unsure whether the point mutation introduced in Phos-mut1 completely disrupted the phosphate-binding ability of the pocket, we designed two more point mutants. In Phos-mut2 (H743A, R752A, R759A, R943A, R956A), three additional basic residues conserved in *Drosophila* near the basic residues comprising the phosphate-binding pocket in the human Dicer PAZ domain were mutated to alanine residues (Fig. 1B and C). In Phos-mut3 (R943E, R956E), the basic Arg943 and Arg956 residues were replaced with the acidic glutamic acid residues. In addition to Phos-mut Dicer-2, we also created the transgenic fly strains expressing wild-type Dicer-2 as a positive control and ATP-mut (G31R) and RNaseIII-mut (D1217A, D1476A) as negative controls. ATP-mut Dicer-2 cannot bind ATP in the helicase domain, and RNaseIII-mut Dicer-2 cannot catalyze RNA cleavage in the RNaseIII active sites (1, 7, 19).

Phosphate-Binding Pocket in Dicer-2 PAZ Domain Is Important for Efficient RNA Silencing. We expressed the transgenic Dicer-2 constructs in flies under a UAST promoter using the ubiquitous Act5C-Gal4 driver in the background of the endogenous *dicer-2* null [*dicer-2^{LS11fsx}* = *dicer-2^{null}* (1)] to test their rescuing ability. We used the GMR-wIR (*white* inverted repeat) transgene reporter system (20) to test the RNA silencing activity and siRNA production in vivo. GMR-wIR produces an inverted repeat hairpin RNA corresponding to *white* exon 3 during eye development. Dicer-2 processes the wIR hairpin into siRNAs, which in turn cleaves and thereby reduces the levels of endogenous *white* mRNAs.

To measure RNA silencing activity quantitatively, we measured the levels of *white* mRNA in hand-dissected fly heads by quantitative (q)RT-PCR using a primer pair to amplify a region in its exon 6 that is not present in the wIR transgene. As expected, the levels of *white* mRNA in *dicer-2^{null}* flies were significantly higher compared with the control flies that have endogenous wild-type *dicer-2*, showing that the RNA silencing activity is lost in *dicer-2^{null}* flies (Fig. 3A). The levels of *white* mRNA were low in the *dicer-2^{null}* flies rescued by the wild-type Dicer-2 transgene, similar to the control flies, showing that the wild-type Dicer-2 transgene fully rescued the RNA silencing activity. In contrast, as expected, the levels of *white* mRNA were high in the negative control ATP-mut and RNaseIII-mut Dicer-2 flies, showing that these transgenes failed to rescue the RNA silencing activity. The levels of *white* mRNA in all three Phos-mut flies were significantly higher compared with the control and wild-type Dicer-2 rescue flies but lower compared with *dicer-2^{null}*, ATP-mut, and RNaseIII-mut flies, showing that the RNA silencing activity is only partially rescued by these Phos-mut transgenes. We concluded that the phosphate-binding pocket in the Dicer-2 PAZ domain is important for efficient RNA silencing in vivo.

Besides siRNA production from long dsRNAs, Dicer-2 also transfers siRNAs to Argonaute2 (21, 22). ATP-mut Dicer (*Dicer-2^{G31R}*) retains the activity to transfer siRNAs to

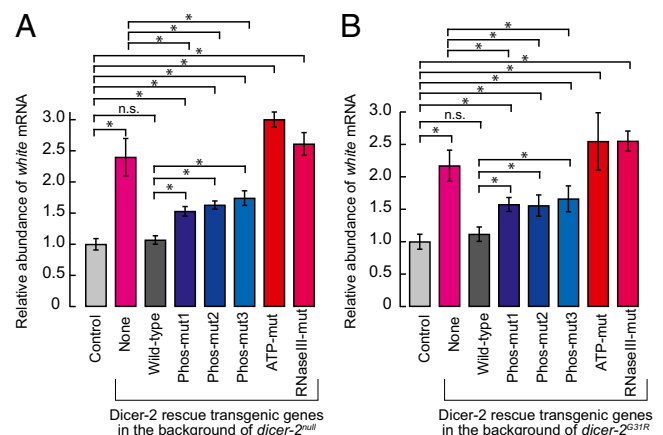


Fig. 3. A phosphate-binding pocket of Dicer-2 is important for efficient RNA silencing. The levels of *white* mRNA in the heads of the female flies in the background of endogenous (A) *dicer-2^{null}* and (B) *dicer-2^{G31R}* in the presence of wIR, normalized by *rp49* relative to the mean of the control flies, determined by qRT-PCR. Data are mean \pm SD ($n = 3$). * $P < 0.05$; n.s., non-significant. *dicer-2^{null}* background: wIR; *dicer-2^{null}*, Act5C-Gal4/*dicer-2^{null}*, UAST-HA-Dicer-2 (wild-type or mutant)/+. *dicer-2^{G31R}* background: wIR; *dicer-2^{G31R}*, Act5C-Gal4/*dicer-2^{G31R}*, UAST-HA-Dicer-2 (wild-type or mutant)/+. Control flies are wIR; *dicer-2^{null}*, Act5C-Gal4/CyO/+; Phos-mut1, -2, and -3 are Dicer-2^{R943A,R956A}, Dicer-2^{H743A,R752A, R759A,R943A,R956A}, and Dicer-2^{R943E,R956E}, respectively. ATP-mut and RNaseIII-mut are Dicer-2^{G31R} and Dicer-2^{D1217A,D1476A}, respectively.

Argonaute2, whereas it cannot generate siRNAs from long dsRNAs (1, 7, 19, 23). To test whether the observed partial loss of RNA silencing activity in Phos-mut flies was due to a loss of the activity to transfer siRNAs rather than a deficiency to produce siRNAs, this time we quantitated the levels of *white* mRNA in the background of endogenous ATP-mut *dicer-2* (*dicer-2^{G31R}*) (1), instead of *dicer-2^{null}*. If the partial loss of the RNA silencing activity was due to deficiency to transfer siRNAs, then coexpression of endogenous ATP-mut Dicer-2 and transgenic Phos-mut Dicer-2 would rescue the RNA silencing activity. We found that this was not the case. The levels of *white* mRNA in the Phos-mut flies were still higher compared with the control and wild-type Dicer-2 rescue flies but lower compared with the nonrescue, ATP-mut, and RNaseIII-mut flies (Fig. 3B). These results suggest that the partial loss of the RNA silencing activity in Phos-mut Dicer-2 transgenic flies is likely due to a reduction in quantity and/or quality of the siRNAs produced by Phos-mut Dicer-2.

Phosphate-Binding Pocket in Dicer-2 PAZ Domain Is Important for Length Fidelity but Not Efficiency in siRNA Production in Vivo. To test the hypothesis that the quantity and/or quality of the siRNAs produced in the Phos-mut Dicer-2-rescued flies are reduced, we high-throughput-sequenced small RNA libraries prepared from heads hand-dissected from the female flies with or without the Dicer-2 transgenes in the background of GMR-wIR and endogenous *dicer-2^{null}* (Table S1). First, we analyzed the quantity of the wIR-derived siRNAs. The wIR-derived 19–23-nt siRNAs were observed in the control flies, and their abundance (normalized by the total non-rRNA mapping reads, a majority of which are miRNA reads) was reduced in *dicer-2^{null}* flies, as expected (Fig. 4A). The abundance was rescued in the flies containing the wild-type Dicer-2 transgene but not in ATP-mut or RNaseIII-mut Dicer-2 transgenic flies. The abundance was rescued in the flies with Phos-mut Dicer-2 transgenes, revealing that the quantity of wIR-derived siRNAs produced by Phos-mut Dicer-2 was not largely changed.

Next, to examine the quality of the wIR-derived siRNAs, we analyzed their length heterogeneity. In both the control and the wild-type Dicer-2 transgene-rescued flies, most of the

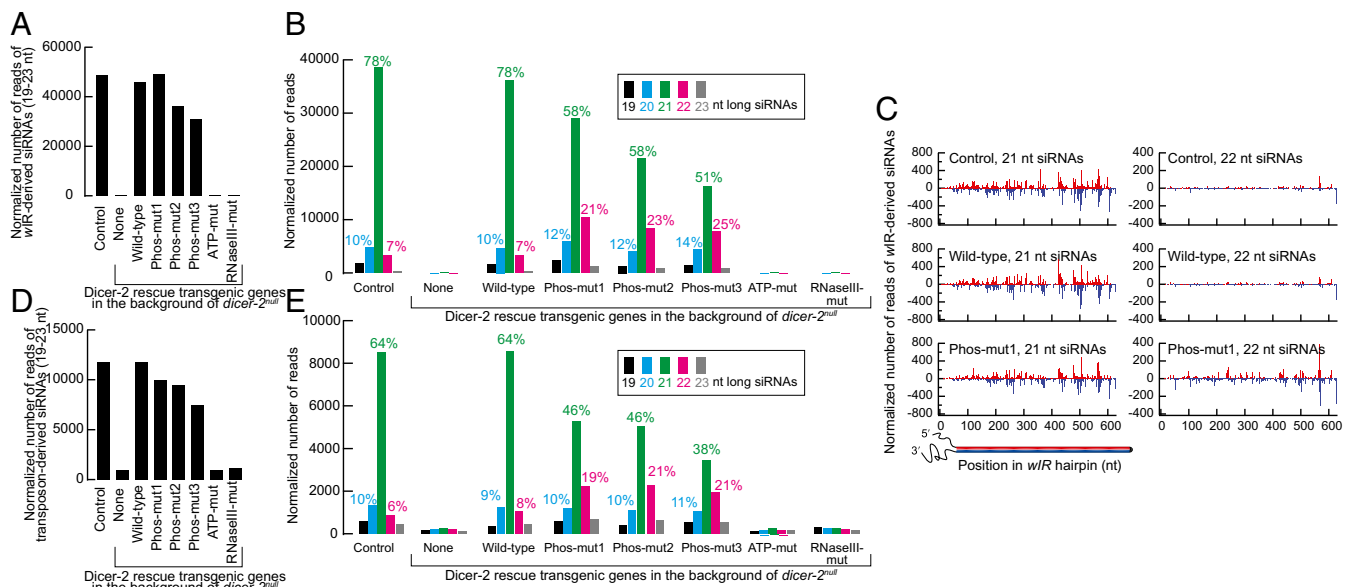


Fig. 4. A phosphate-binding pocket of Dicer-2 is important for high-fidelity 21-nt siRNA production in vivo. Sequencing results of small RNAs prepared from female fly heads are shown. The reads were normalized by the sequencing depth. (A) Normalized number of reads of 19–23-nt wIR-derived siRNAs. (B) Length distribution of wIR-derived siRNAs. (C) Abundance and 5' position of 21- and 22-nt wIR-derived siRNAs. Antisense siRNAs are shown in red, sense in blue. (D) Normalized number of reads of 19–23-nt transposon-derived siRNAs. (E) Length distribution of transposon-derived siRNAs.

wIR-derived siRNAs were 21 nt, composing 78% of all of the wIR-derived reads, whereas only 10% and 7% were 20 nt and 22 nt, respectively (Fig. 4B), showing high fidelity and precision in 21-nt siRNA production by wild-type Dicer-2. Strikingly, in all three Phos-mut Dicer-2 transgene-rescued flies, the ratio of the 21-nt wIR-derived siRNAs was reduced (58%, 58%, and 51% for Phos-mut1, -2, and -3, respectively), whereas those of 20 nt and 22 nt were increased (20 nt was 12%, 12%, and 14% and 22 nt was 21%, 23%, and 25% for Phos-mut1, -2, and -3, respectively), increasing the length heterogeneity of siRNAs. Also, the abundance of the wIR-derived siRNAs with each specific length was altered in all three Phos-mut Dicer-2 transgene-rescued flies. The abundance of the 21-nt wIR-derived siRNAs was decreased in Phos-mut flies, whereas those of the 20- and 22-nt siRNAs were increased (Fig. 4B). In Phos-mut flies, the siRNAs with altered length were produced across the whole length of the wIR hairpin rather than from a specific position or strand (Fig. 4C). The reduction in the length fidelity and the resultant increase in the length heterogeneity of siRNAs likely caused the partial loss of the RNA silencing activity in Phos-mut flies observed in vivo (Fig. 3), demonstrating the importance of high length fidelity in 21-nt siRNA production for efficient RNA silencing.

We also observed an increased length heterogeneity in transposon-derived siRNAs (Fig. 4D and E) and endogenous structural loci-derived siRNAs (Fig. S1) in Phos-mut flies. We concluded that the phosphate-binding pocket in the Dicer-2 PAZ domain is important for the length fidelity, but not the efficiency, in siRNA production in vivo.

Phosphate-Binding Pocket in Dicer-2 PAZ Domain Is Important for Length Fidelity in siRNA Production in Vitro. To test directly whether Phos-mut Dicer-2 lost the high fidelity required to produce 21-nt siRNAs, we performed an in vitro RNA processing assay using recombinant Dicer-2 protein. Wild-type Dicer-2 efficiently processed both the 5' monophosphate- and 5' hydroxyl-long (104-bp) dsRNAs (Fig. 5A) (19). In contrast, it efficiently cleaved the short (30-bp) dsRNA with a 5' monophosphate end but not the short (30-bp) dsRNA with a 5' hydroxyl end (Fig. 5A) (19). Thus, the 5' monophosphate is required for efficient cleavage

of short dsRNAs, but it is dispensable for cleavage of long dsRNAs. ATP-mut Dicer-2 exhibited reduced cleavage activities for the long dsRNAs but retained the wild-type activities to

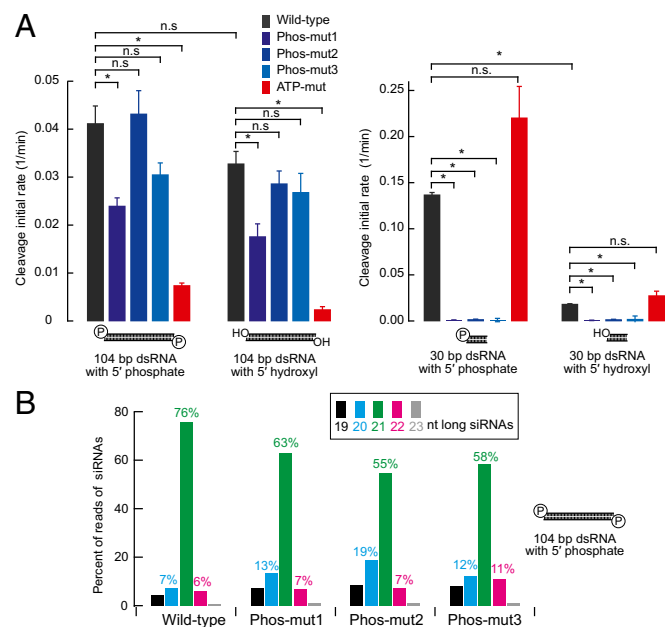


Fig. 5. A phosphate-binding pocket of Dicer-2 is important for high-fidelity 21-nt siRNA production in vitro. (A) Initial rates of substrate RNA cleavage by the wild-type and mutant recombinant Dicer-2 proteins in test tube. Data are mean \pm SD for three independent trials. * $P < 0.01$; n.s., nonsignificant. As long dsRNA, 104-bp dsRNAs with 2-nt 3' overhang and either 5' monophosphate or 5' hydroxyl were tested. As short dsRNA, 30-bp dsRNAs with 2-nt 3' overhang and either 5' monophosphate or 5' hydroxyl were tested. The other end of the 30-bp dsRNAs was blocked by two deoxynucleotides (7). (B) Length distribution of siRNAs produced from 104-bp dsRNA with 2-nt 3' overhang and 5' monophosphate by recombinant Dicer-2 proteins in test tube revealed by high-throughput sequencing.

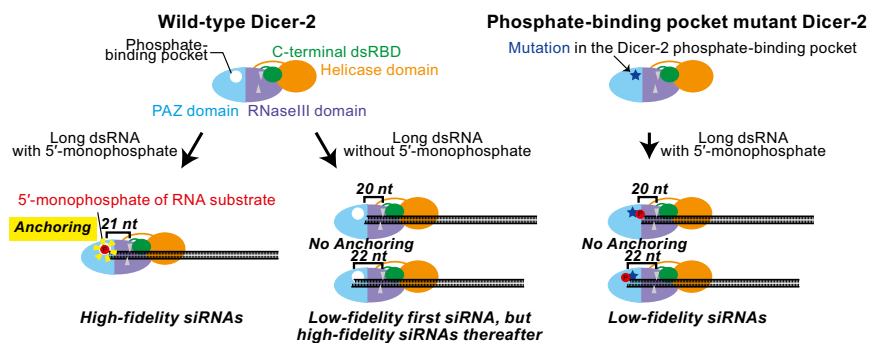


Fig. 6. Model for high-fidelity siRNA production by Dicer-2. The phosphate-binding pocket in the Dicer-2 PAZ domain anchors the 5' monophosphate of long dsRNA for precise measurement of 21-nt length between the 5' monophosphate and the RNaseIII active sites. In the absence of the recognition of the 5' monophosphate by the phosphate-binding pocket, siRNAs produced exhibit increased length heterogeneity. This model aligns well with the previous structural model of Dicer-2 (25).

cleave the short dsRNAs. These results support a previously proposed model that ATP binding and hydrolysis by the Dicer-2 helicase domain is necessary for Dicer-2 to translocate on a long dsRNA substrate to processively produce siRNAs (7–9).

Mutations in the Dicer-2 phosphate-binding pocket did not largely affect the long dsRNA processing rates by Dicer-2. Consistent with our *in vivo* results (Fig. 4*A* and *D*), Phos-mut2 and 3 Dicer-2 exhibited similar long dsRNA cleavage rates to wild-type Dicer-2 (Fig. 5*A*), revealing that the phosphate-binding pocket is not required for the efficiency to process long dsRNAs. Phos-mut1 exhibited slightly but statistically significant lowered activities to cleave the long dsRNAs compared with wild-type Dicer-2 (Fig. 5*A*), contradicting with our *in vivo* results that did not show changes (Fig. 4*A* and *D*), even though the same R943A and R956A mutations introduced in Phos-mut1 are included in the Phos-mut2 (H743A, R752A, R759A, R943A, and R956A). Currently, we do not know the cause for the lowered activity of Phos-mut1 *in vitro*. In contrast, all three phosphate-binding pocket mutants lost the activity to cleave the short dsRNAs, revealing that the phosphate-binding pocket is required to cleave short dsRNAs (Fig. 5*A*) (19). Together, these results suggest that a longer duplex associates with Dicer-2 in a manner that is less dependent on the terminal phosphate recognition by the Dicer-2 phosphate-binding pocket, probably relying more on stabilizing interactions between the proximal duplex region of the long dsRNA and the helicase domain and/or dsRBD of Dicer-2. Because short dsRNAs lack the proximal region that can interact with Dicer-2, their cleavage depends on the recognition of the terminal phosphate by the Dicer-2 phosphate-binding pocket.

To examine the length fidelity in siRNA production *in vitro*, we prepared and sequenced libraries of the small RNAs produced from 104-bp dsRNA with 5' monophosphate ends by the recombinant Dicer-2 proteins (Table S2). Similar to our *in vivo* findings (Fig. 4), we found that Phos-mut Dicer-2 produced fewer 21-nt siRNAs and higher amounts of 20- and 22-nt siRNAs compared with wild-type Dicer-2 *in vitro* (Fig. 5*B*). Seventy-six percent of siRNAs produced by wild-type Dicer-2 were 21 nt long. However, only 63%, 55%, and 58% of siRNAs produced by Phos-mut1, -2, and -3, respectively, were 21 nt long. The ratios of 20 nt and 22 nt produced by Phos-mut1 (13% and 7%), Phos-mut2 (19% and 7%), and Phos-mut3 (12% and 11%) were higher compared with wild-type Dicer-2 (7% and 6%). We concluded that Phos-mut Dicer-2 lost the length fidelity to produce 21-nt siRNAs without a major loss in the efficiency to produce siRNAs *in vitro*. Thus, even with an RNA substrate with a terminal monophosphate, the mutations in the Dicer-2 phosphate-binding pocket mimic the fidelity-lowering effect caused by the lack of the terminal monophosphate on long dsRNA

observed in the first siRNA production by the wild-type Dicer-2 (Fig. 2).

Model for the Role of the Dicer-2 Phosphate-Binding Pocket in High-Fidelity 21-nt siRNA Production. We found that the 5' monophosphate of long dsRNA is crucial for high-fidelity 21-nt siRNA production from the terminus *in vitro*. Moreover, our *in vivo* and *in vitro* studies revealed that the phosphate-binding pocket in the Dicer-2 PAZ domain plays an important and direct role in high-fidelity 21-nt siRNA production. High-fidelity 21-nt siRNA production was crucial for efficient RNA silencing activity *in vivo*, showing the importance of high-fidelity production of 21-nt siRNAs. This *in vivo* finding is consistent with the previous *in vitro* studies that revealed that synthetic 21-nt siRNA duplexes are significantly more potent than synthetic 20-nt or 22-nt siRNA duplexes to silence a synthetic target mRNA in *Drosophila* embryo extract *in vitro* (24). Based on these findings, we propose the following model by which the Dicer-2 phosphate-binding pocket ensures high-length fidelity in 21-nt siRNA production (Fig. 6). The phosphate-binding pocket binds the 5' monophosphate of a long dsRNA substrate, thereby anchoring the end of the RNA substrate. The N-terminal helicase domain and/or the C-terminal dsRBD bind the body of the long dsRNA substrate and align it along the RNaseIII active sites in the RNaseIII domains. The distance between the phosphate-binding pocket and the RNaseIII active sites corresponds to that of the 21-nt pitch in the A-form dsRNA duplex and thus functions as a molecular ruler. Therefore, anchoring the 5' monophosphate of an RNA substrate at the phosphate-binding pocket and aligning the rest of the RNA toward the RNaseIII active site allows precise measurement of 21-nt length and thus ensures high-fidelity production of 21-nt siRNAs. During the processive production of multiple siRNAs from a single long dsRNA substrate, Dicer-2 repeatedly uses the 5' monophosphate-anchoring mechanism to precisely produce 21-nt siRNAs with high fidelity as it translocates along the length of the dsRNA. When a long dsRNA substrate lacks a 5' monophosphate, a siRNA with a wrong length is produced from the RNA terminus. Although the end of the substrate RNA cannot be properly anchored due to the lack of a 5' monophosphate, the N-terminal helicase domain and/or the C-terminal dsRBD can still bind and align the long dsRNA to the RNaseIII active sites, resulting in the production of siRNAs with altered length without a major loss in the production efficiency. Although the length heterogeneity is increased, the siRNAs produced are still close to the correct 21-nt length, suggesting other interactions between the dsRNA end and the Dicer-2 PAZ domain such as the 3' binding pocket. High fidelity is recovered in the subsequent processive siRNA production, as Dicer-2 can start to use the 5' monophosphate-anchoring mechanism after the first siRNA production. When the phosphate-binding pocket

of the Dicer-2 PAZ domain is mutated, the 5' monophosphate of neither the original dsRNA substrate nor its intermediate during the processive siRNA production can be anchored. However, the N-terminal helicase domain and/or the C-terminal dsRBD can still bind and align the long dsRNA to the RNaseIII active sites. In addition, the dsRNA end may still interact with the 3' binding pocket in the PAZ domain. Thus, the phosphate-binding pocket mutant Dicer-2 produces siRNAs with altered length across the whole length of the RNA substrates without a major loss in siRNA production efficiency. This model aligns well with the previous structural model of Dicer-2 (25). Our studies shed light on the mechanism by which Dicer-2 produces 21-nt siRNAs with a remarkably high fidelity.

The recognition of the 5' monophosphate of pre-miRNAs by the phosphate-binding pocket in the human Dicer PAZ domain was shown to be important for both the efficiency and length fidelity in miRNA production (16, 18). Almost all of the pre-miRNAs have a 5' monophosphate. Therefore, human Dicer may use the 5' monophosphate on pre-miRNAs to select against incorrect RNA substrates that lack a 5' monophosphate. In contrast, our studies demonstrated that recognition of the 5' monophosphate of a long dsRNA substrate by the unique phosphate-binding pocket of *Drosophila* Dicer-2 PAZ domain is important for the length fidelity, but not the efficiency, during siRNA production. Unlike human Dicer, why does *Drosophila* Dicer-2 efficiently process RNAs lacking a 5' monophosphate into siRNAs with altered length rather than not to produce any siRNAs? We speculate that with this strategy *Drosophila* Dicer-2 can cleave the terminal end of diverse long dsRNA substrates such as viral RNAs and transposon RNAs that lack a 5' monophosphate and instead have a 5' hydroxyl, 5' triphosphate, or 5' cap end (26, 27). Initial terminal cleavage by Dicer-2 of such long dsRNA substrates lacking a 5' monophosphate can produce a 5' monophosphate on the remaining intermediate long dsRNA substrates. Therefore, although the initial terminal siRNAs produced may have a lower length fidelity due to the lack of a 5' monophosphate, Dicer-2 can use the 5' monophosphate-anchoring mechanism to

produce high-fidelity 21-nt siRNAs during subsequent processive siRNA production (as in Fig. 2). Thus, the unique phosphate-binding pocket of *Drosophila* Dicer-2 can achieve both the high fidelity in siRNA production and the relaxed end structure requirement in the initial long dsRNA substrate cleavage, maximizing its overall antiviral and antitransposon activities. This antiviral and antitransposon defense system mediated by the unique Dicer-2 enzyme is crucial in arthropods, as they lack acquired immunity (28, 29).

Materials and Methods

Fly Strains. P[HA-Dicer-2] rescuing transgenes containing an N-terminal HA tag were generated by subcloning the coding sequence of *dicer-2* into a pUASTattB plasmid vector. Point mutations were introduced using PCR. The transgenes were integrated at position 68E1 on the third chromosome using the BDRC fly strain 24485. The mini-*white* gene (*w⁺trmC*) derived from the integrated plasmids and the RFP gene originally present in the fly strain to mark the landing site were removed by using Cre-Lox. qRT-PCR was performed as described previously (30). Detailed information is provided in *SI Materials and Methods*.

In Vitro Dicing Assay. In vitro Dicing assay was performed as previously described (19). Detailed information is provided in *SI Materials and Methods*.

Small RNA Sequencing. Small RNA libraries were prepared, sequenced on HiSeq4000 (Illumina), and analyzed as previously described (19, 30–32). The sequencing statistics of the small RNAs are summarized in *Tables S1* and *S2*. The Gene Expression Omnibus (GEO) accession number for the small RNA libraries reported in this paper is GSE84532.

ACKNOWLEDGMENTS. We are grateful to Dr. Richard Carthew (Northwestern University) for fly strains of *white*-inverted repeat, *dicer-2^{null}*, and *dicer-2^{G31R}*; Dr. Bo Han (Pacific Biosciences) and Dr. Junko Tsuji (University of Massachusetts Medical School) for their help in the bioinformatics analysis; and Susan Liao and Dr. Li Zhu, members of R.F. laboratory, for their comments on the manuscript. This work was supported by Grant 15SDG23220028 from the American Heart Association and funds provided by the Department of Biological Chemistry at the Johns Hopkins School of Medicine (to R.F.).

- Lee YS, et al. (2004) Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117(1):69–81.
- Ghildiyal M, et al. (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320(5879):1077–1081.
- Kawamura Y, et al. (2008) *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453(7196):793–797.
- Okamura K, Balla S, Martin R, Liu N, Lai EC (2008) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol* 15(6):581–590.
- Okamura K, et al. (2008) The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453(7196):803–806.
- Ameres SL, Hung JH, Xu J, Weng Z, Zamore PD (2011) Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two *Drosophila* Argonaute proteins. *RNA* 17(1):54–63.
- Cenik ES, et al. (2011) Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol Cell* 42(2):172–184.
- Welker NC, et al. (2011) Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol Cell* 41(5):589–599.
- Sinha NK, Trettin KD, Aruscavage PJ, Bass BL (2015) *Drosophila* dicer-2 cleavage is mediated by helicase- and dsRNA termini-dependent states that are modulated by Loquacious-PD. *Mol Cell* 58(3):406–417.
- Cerutti L, Mian N, Bateman A (2000) Domains in gene silencing and cell differentiation proteins: The novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem Sci* 25(10):481–482.
- Lingel A, Simon B, Izaurralde E, Sattler M (2003) Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* 426(6965):465–469.
- Lingel A, Simon B, Izaurralde E, Sattler M (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* 11(6):576–577.
- Zhang H, Kolb FA, Jaskiewicz W, Westhof E, Filipowicz W (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell* 118(1):57–68.
- Macrae IJ, et al. (2006) Structural basis for double-stranded RNA processing by Dicer. *Science* 311(5758):195–198.
- Macrae IJ, Zhou K, Doudna JA (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* 14(10):934–940.
- Park JE, et al. (2011) Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* 475(7355):201–205.
- Gu S, et al. (2012) The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell* 151(4):900–911.
- Tian Y, et al. (2014) A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer. *Mol Cell* 53(4):606–616.
- Fukunaga R, Colpan C, Han BW, Zamore PD (2014) Inorganic phosphate blocks binding of pre-miRNA to Dicer-2 via its PAZ domain. *EMBO J* 33(4):371–384.
- Lee YS, Carthew RW (2003) Making a better RNAi vector for *Drosophila*: use of intron spacers. *Methods* 30(4):322–329.
- Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD (2004) A protein sensor for siRNA asymmetry. *Science* 306(5700):1377–1380.
- Liu Q, et al. (2003) R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* 301(5641):1921–1925.
- Förstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD (2007) *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* 130(2):287–297.
- Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T (2001) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J* 20(23):6877–6888.
- Lau PW, et al. (2012) The molecular architecture of human Dicer. *Nat Struct Mol Biol* 19(4):436–440.
- Yang D, Leibowitz JL (2015) The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res* 206:120–133.
- Byzewska M, Śmietański M, Purta E, Bujnicki JM (2014) RNA methyltransferases involved in 5' cap biosynthesis. *RNA Biol* 11(12):1597–1607.
- Bronkhorst AV, van Rij RP (2014) The long and short of antiviral defense: Small RNA-based immunity in insects. *Curr Opin Virol* 7:19–28.
- Karlikow M, Goic B, Saleh MC (2014) RNAi and antiviral defense in *Drosophila*: Setting up a systemic immune response. *Dev Comp Immunol* 42(1):85–92.
- Fukunaga R, et al. (2012) Dicer partner proteins tune the length of mature miRNAs in flies and mammals. *Cell* 151(3):533–546.
- Han BW, Wang W, Li C, Weng Z, Zamore PD (2015) Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* 348(6236):817–821.
- Han BW, Wang W, Zamore PD, Weng Z (2015) piPipes: A set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics* 31(4):593–595.