nature structural & molecular biology

Recognition of the pre-miRNA structure by *Drosophila* Dicer-1

Akihisa Tsutsumi[1,2], Tomoko Kawamata[1], Natsuko Izumi[1], Hervé Seitz[3,4] & Yukihide Tomari[1,2]

Drosophila melanogaster **has two Dicer proteins with specialized functions. Dicer-1 liberates miRNA-miRNA* duplexes from precursor miRNAs (pre-miRNAs), whereas Dicer-2 processes long double-stranded RNAs into small interfering RNA duplexes. It was recently demonstrated that Dicer-2 is rendered highly specific for long double-stranded RNA substrates by inorganic phosphate and a partner protein R2D2. However, it remains unclear how Dicer-1 exclusively recognize pre-miRNAs. Here we show that fly Dicer-1 recognizes the single-stranded terminal loop structure of pre-miRNAs through its N-terminal helicase domain, checks the loop size and measures the distance between the 3′ overhang and the terminal loop. This unique mechanism allows fly Dicer-1 to strictly inspect the authenticity of pre-miRNA structures.**

MicroRNAs (miRNAs) are a class of endogenous small RNAs found in plants, animals, some unicellular eukaryotes and their viruses[1]. miRNAs regulate a diverse array of biological processes by post-transcriptionally silencing their target genes[1]. The miRNA pathway begins with transcription of long primary miRNAs (pri-miRNAs), typically by polymerase II (ref. 2). Pri-miRNAs are first processed in the nucleus by the RNase III enzyme Drosha, together with its double-stranded RNA (dsRNA)-binding protein partner, called Pasha in flies and DGCR8 in mammals[2]. The resultant ~70-nt hairpin RNAs are called precursor miRNAs (pre-miRNAs). Pre-miRNAs are then exported to the cytoplasm by Exportin-5 and further processed there by another RNase III enzyme, Dicer, assisted by its dsRNA-binding protein partner, called Loquacious (Loqs) in flies and TAR RNA-binding protein (TRBP) and PKR activating protein (PACT) in mammals[2–9]. Dicer processing results in ~22-nt RNA duplexes with 2-nt overhangs at both 3′ ends and frequent internal mismatches, called miRNA-miRNA* duplexes. The miRNA-miRNA* duplexes are then loaded into Argonaute (Ago) proteins, the core component of the effector complex called RNA-induced silencing complex (RISC)[10–13], with the aid of a dynamic conformational change of Ago proteins by the Hsc70-Hsp90 chaperone machinery and by ATP consumption[14–16]. One of the two strands in miRNA-miRNA* duplexes, often the miRNA strand, is selectively retained in Ago proteins, forming functionally mature RISC. This strand-separation process, or unwinding, does not require ATP[17,18] or the action of the chaperone machinery[14].

A parallel small RNA pathway, the small interfering RNA (siRNA) pathway, starts with endogenous or exogenous long dsRNAs in the cytoplasm. Long dsRNAs are processed by Dicer into siRNA duplexes[2], which, similarly to miRNA-miRNA* duplexes, are loaded into Ago proteins by the chaperone machinery[14–16]. In flies, Dicer-2 (Dcr-2) and its partner protein R2D2 are essential for loading of siRNA duplexes into Ago2 (refs. 19–21), but the requirement of Dicer for human RISC loading is unclear. Following RISC loading, one strand—the passenger stand—is cleaved and discarded[22–25]. The resulting mature RISC, containing only the guide strand, usually cleaves target RNAs with perfectly complementary sites, a phenomenon called RNA interference (RNAi)[26].

In humans, a single Dicer processes both long dsRNAs and pre-miRNAs into siRNA duplexes and miRNA duplexes, respectively[27,28]. In contrast, *Drosophila melanogaster* and many other insects have two Dicer paralogs, Dicer-1 (Dcr-1) and Dcr-2, with mutually exclusive functions[29]: Dcr-1 liberates miRNA-miRNA* duplexes from pre-miRNAs, whereas Dcr-2 processes long dsRNAs into siRNA duplexes. What determines the substrate specificities of these two insect Dicer proteins? It was recently demonstrated that although Dcr-2 can potentially process pre-miRNAs, inorganic phosphate and R2D2 render Dcr-2 highly specific for long dsRNA substrates[30]. However, how Dcr-1 specifically recognizes pre-miRNAs remains unknown. There are no common sequences conserved among pre-miRNAs. Instead, pre-miRNAs share several structural features: a 2-nt 3′ overhang, a double-stranded stem region, mismatches in the stem region and a thermodynamically unstable, single-stranded terminal loop region (**Fig. 1a**). Therefore, fly Dcr-1 and/or associating cofactors must recognize some of these features to distinguish pre-miRNAs from long dsRNAs. To learn whether Dcr-1 itself recognizes pre-miRNAs specifically, and how it does this, we carried out a series of *in vitro* dicing assays and gel-shift analyses. Our data show that fly Dcr-1 recognizes the single-stranded terminal loop structure of pre-miRNAs through its N-terminal helicase domain, checks the terminal loop size and measures the length from the 3′ overhang to

[1]Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [2]Department of Medical Genome Sciences, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [3]Université de Toulouse, Université Paul Sabatier, Laboratoire de Biologie Moléculaire Eucaryote, Toulouse, France. [4]Centre National de la Recherche Scientifique, Laboratoire de Biologie Moléculaire Eucaryote, Toulouse, France. Correspondence should be addressed to Y.T. (tomari@iam.u-tokyo.ac.jp).

Received 19 January; accepted 13 July; published online 18 September 2011; doi:10.1038/nsmb.2125

the terminal loop. This unique mechanism allows fly Dcr-1 to strictly inspect the authenticity of pre-miRNA structures.

RESULTS

Dcr-1 does not distinguish the stem structure

We first tested if Dcr-1 recognizes internal mismatches in the stem region characteristic of pre-miRNAs. The wild-type fly pre-*let-7* has two mismatches and three G•U wobbles in its stem region. We constructed *let-7* shRNA, where all the mismatches and G•U wobbles are replaced with Watson-Crick base pairs, and *let-7* shRNA (mm9), where a single mismatch was reintroduced at guide position 9, without changing the *let-7* guide sequence (**Supplementary Table 1**). Fly Dcr-1, tagged with a streptavidin-binding peptide (SBP) at the N-terminal, was expressed in S2 cells and purified to apparent homogeneity (**Supplementary Fig. 1a**). The RNA substrates were radiolabeled at their 5′ phosphates and incubated with recombinant Dcr-1 at 25 °C. Dicing efficiency was nearly identical for all the substrates (**Fig. 1b**). As a control, we similarly purified the SBP-tagged catalytic mutant of Dcr-1, in which two glutamate residues in the RNase IIIa and IIIb domains are replaced with alanine[31] (E1908A E2139A, **Supplementary Fig. 1a**). No dicing activity was observed with the catalytic mutant (**Supplementary Fig. 1b**). We concluded that Dcr-1 does not distinguish the structure of the stem region.

Dcr-1 recognizes the terminal loop region

We next focused on the terminal loop region. Wild-type pre-*let-7* bears a 14-nt single-stranded terminal loop region (pre-*let-7* (L14)). We generated small-loop variants pre-*let-7* (L8) with an 8-nt loop region and pre-*let-7* (L4) with a 4-nt loop region (**Supplementary Table 1**). Dcr-1's dicing activity was markedly reduced with these small-loop mutant substrates (**Fig. 1c** and **Supplementary Fig. 1c**), suggesting that the terminal loop region has an important role in substrate recognition by Dcr-1.

To examine whether the terminal loop structure per se or its single-stranded property is important, we constructed an RNA duplex with the 14-nt terminal loop of pre-*let-7* cut in half, leaving two 7-nt single-stranded fragments (*let-7* duplex (ss7+7), **Supplementary Table 1**). Dcr-1 diced *let-7* duplex (ss7+7) as efficiently as it diced wild-type pre-*let-7* (L14), indicating that Dcr-1 requires a single-stranded region rather than the characteristic terminal loop structure for substrate recognition (**Fig. 1d**).

To further explore the mechanism by which Dcr-1 recognizes the single-stranded region, we systematically prepared a series of *let-7* duplexes with different lengths of single-stranded RNAs at the terminal region: ss7+7 (two halves of the natural 14-nt terminal loop), ss7+3, ss3+7, ss3+3, ss15+3 and ss3+15 (**Supplementary Table 1**). As expected from the data for the small-loop pre-*let-7* derivatives (**Fig. 1c**), Dcr-1 diced the ss7+7 duplex efficiently but not the ss3+3 duplex (**Fig. 1e**). Both the ss7+3 and ss3+7 duplexes showed moderate dicing efficiency (**Fig. 1e**), suggesting that the presence of a 7-nt single-stranded RNA on either one side of the two stem strands alone is not sufficient for recognition by Dcr-1. By contrast, having a 15-nt single-stranded RNA on either side (ss15+3 and ss3+15 duplexes) greatly rescued the dicing activity (**Fig. 1f**). We concluded that Dcr-1 does not necessarily require single-stranded regions on both sides of the two stem strands but rather simply requires single-stranded RNA(s) of sufficient length.

Dcr-1 measures the distance from the 3′ overhang to the loop

It is known that the PAZ domain of the Dicer proteins binds to the 3′-overhang structure[32,33]. Indeed, in our study, removing the 3′ 2-nt overhang from the pre-*let-7* completely blocked the dicing reaction, regardless of whether the 5′ end was base-paired or frayed (**Fig. 2a**). Therefore, the fly Dcr-1's PAZ domain specifically recognizes the 3′-overhang structure.

We then investigated the importance of stem length for dicing. Wild-type pre-*let-7* has a 22-nt stem region (pre-*let-7* (S22)). We constructed a long-stem series of pre-*let-7* variants with 25-, 28-, 31- and 34-nt stem regions (pre-*let-7* (S25) to pre-*let-7* (S34), **Supplementary Table 1**). Whereas the length of the dicing products remained ~21 nt, irrespective of the stem length, the dicing activity decreased markedly as the stem became longer (**Fig. 2b** and **Supplementary Fig. 1d**). A terminally single-stranded duplex substrate with a long stem region (*let-7* duplex (S34, ss7+7)) also showed reduced dicing compared to the wild-type 22-nt stem region (*let-7* duplex (ss7+7)) (**Fig. 2c**). When the terminal 7-nt + 7-nt single-stranded region of the *let-7* duplex (S34, ss7+7) was repositioned at the twenty-second nucleotide from the 3′ overhang as an internal bulge (*let-7* duplex (S22, ss7+7, S12); **Supplementary Table 1**), without changing the total length, the dicing efficiency was partially rescued (**Fig. 2c**). Therefore, Dcr-1 recognizes a single-stranded region, not necessarily at the opposite terminus but at an appropriate distance from the 3′ overhang.

Together, the above observations suggest that Dcr-1 recognizes the 3′-overhang structure and a single-stranded region of an appropriate size and measures the distance between these two structural features. In this way, Dcr-1 inspects whether the substrate is 'authentic' or not.

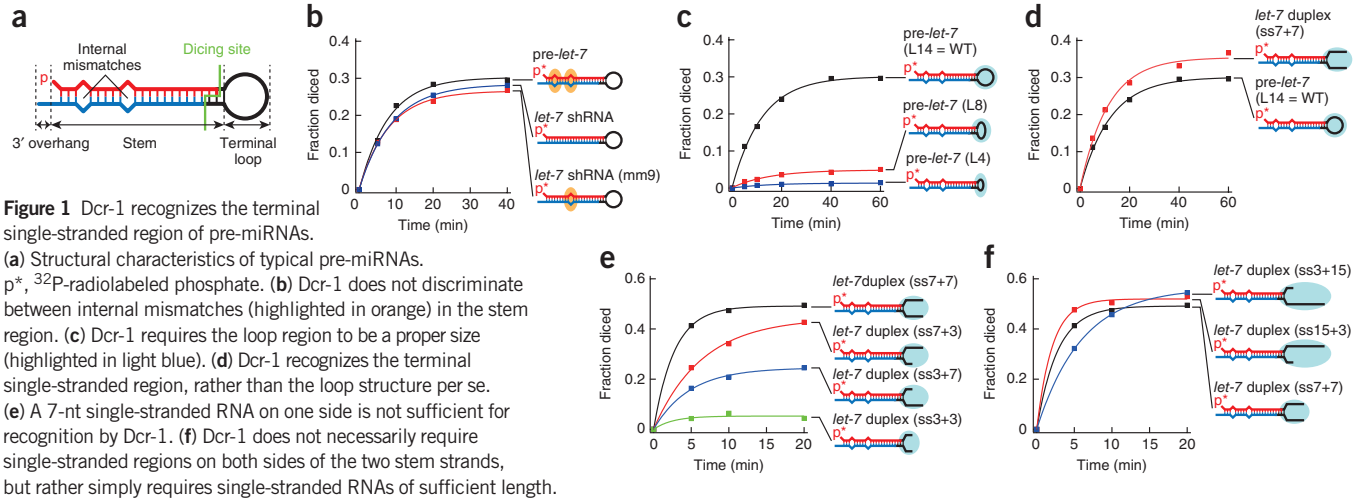

Figure 1 Dcr-1 recognizes the terminal single-stranded region of pre-miRNAs. (**a**) Structural characteristics of typical pre-miRNAs. p*, ^{32}P-radiolabeled phosphate. (**b**) Dcr-1 does not discriminate between internal mismatches (highlighted in orange) in the stem region. (**c**) Dcr-1 requires the loop region to be a proper size (highlighted in light blue). (**d**) Dcr-1 recognizes the terminal single-stranded region, rather than the loop structure per se. (**e**) A 7-nt single-stranded RNA on one side is not sufficient for recognition by Dcr-1. (**f**) Dcr-1 does not necessarily require single-stranded regions on both sides of the two stem strands, but rather simply requires single-stranded RNAs of sufficient length.

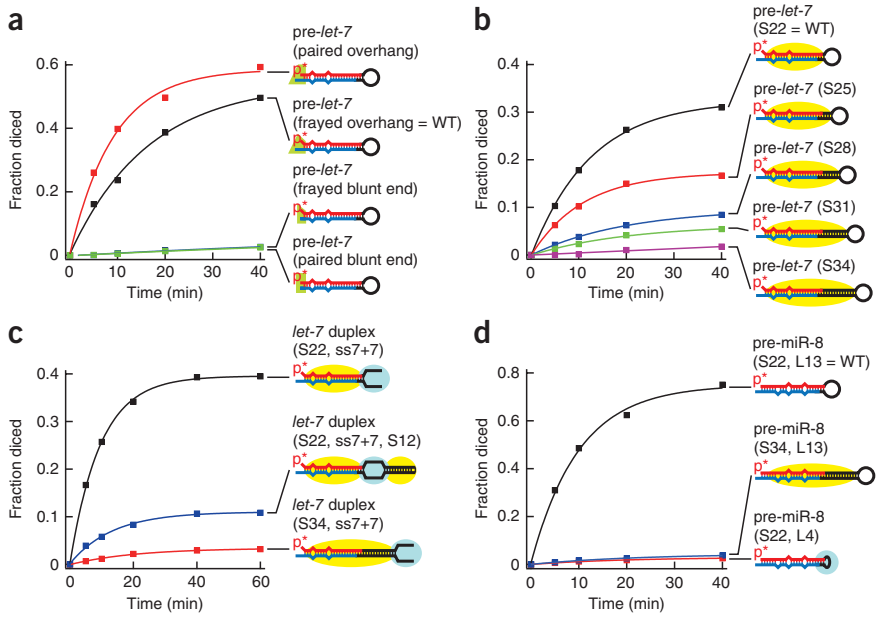

Figure 2 Dcr-1 measures the distance from the 3′ overhang to the terminal single-stranded region of pre-miRNAs. (**a**) Dcr-1 requires the 3′ overhang (highlighted in green) regardless of the base-pairing status at the 5′ end of pre-miRNAs. (**b**) Dcr-1 requires a stem region of the proper length (highlighted in yellow). (**c**) Dcr-1 recognizes the single-stranded RNA (highlighted in light blue) located at a proper distance from the 3′ overhang, but not necessarily at the opposite terminus from the 3′ overhang. (**d**) Dcr-1 efficiently dices wild-type pre-miR-8 but not its small-loop and long-stem variants.

To confirm the generality of this model, we prepared a series of fly pre-miR-8 derivatives: pre-miR-8 (S22, L13) (wild type), pre-miR-8 (S34, L13) (a long-stem variant) and pre-miR-8 (S22, L4) (a small-loop variant) (**Supplementary Table 1**). Much as it did the pre-*let-7* series, Dcr-1 robustly diced the wild-type pre-miR-8 but could barely dice the long-stem and small-loop variants of pre-miR-8 (**Fig. 2d**).

Binding affinity determines substrate recognition by Dcr-1

To examine whether the substrate binding step or subsequent steps that include catalysis have a crucial role in Dcr-1 specifically recognizing the pre-miRNA structure, we determined the binding affinity of Dcr-1 for various substrates. To this end, we used the E1908A E2139A catalytic mutant of Dcr-1 to avoid substrate cleavage (**Supplementary Fig. 1b**) and measured the K_d values for a series of pre-*let-7* variants by gel-shift analysis. We focused on three typical substrates: pre-*let-7* (S22, L14) (wild type), pre-*let-7* (S34, L14) (a long-stem variant) and pre-*let-7* (S22, L4) (a small-loop variant) (**Supplementary Table 1**). **Supplementary Figure 2a**,**b** shows representative binding curves, and **Table 1** summarizes the K_d values determined from three independent trials. Dcr-1 bound the long-stem variant about two times less tightly and the small-loop variant about six times less tightly than it bound the wild-type pre-*let-7* (**Table 1**), which agrees with their decreased dicing efficiency (**Figs. 1c** and **2b**). As expected, Dcr-1 barely bound to blunt-end substrates, regardless of whether the 5′ end was base-paired or frayed (**Supplementary Fig. 2e**). Thus, Dcr-1's binding affinity primarily determines its ability to discriminate between substrates.

Dcr-1 binds the terminal loop through its helicase domain

Dicer family proteins contain the DEAD-like helicase superfamily domain (DEXDc, also referred to as 'domain 1') and the helicase superfamily C-terminal domain (HELICc, also referred to as 'domain 2') at their N termini[34,35]. In general, these two tandemly repeated RecA-like domains cooperatively function as an ATP-dependent helicase[34–37]. Indeed, recent studies show that fly Dcr-2 is a dsRNA-stimulated ATPase; the helicase domain of Dcr-2 hydrolyzes ATP and fuels translocation of Dcr-2 along long dsRNAs to processively generate siRNAs[30,38]. By contrast, previous reports show that the dicing reaction by fly Dcr-1 does not require ATP[5]. We confirmed that the presence or absence of ATP or nonhydrolyzable ATP analogs did not affect the dicing efficiency (**Supplementary Fig. 2f**). Binding of the catalytic mutant Dcr-1 to pre-*let-7* was also unaffected by ATP (**Supplementary Fig. 2g**). Compared with fly Dcr-2 and human DICER1, fly Dcr-1's helicase domain is quite unique, in that its DEXDc domain, which includes motifs that are important for ATP binding, is poorly conserved, and two large linkers are inserted upstream and downstream of the HELICc domain (**Fig. 3a** and **Supplementary Figs. 3** and **4**). We hypothesized that this unique helicase domain of fly Dcr-1 confers its substrate specificity toward pre-miRNAs. To explore this possibility, we generated four deletion constructs of Dcr-1: ΔLinker, ΔHelicase, ΔDEXDc and ΔHELICc (**Fig. 3b** and **Supplementary Fig. 1a**), and we monitored the dicing activity of each deletion mutant for the wild-type pre-*let-7* and the long-stem and small-loop variants. Our data showed that deletion of the two large linkers (ΔLinker) did not affect the overall dicing efficiency or substrate specificity (**Fig. 3c**,**d**), suggesting that these linkers are dispensable for substrate recognition by Dcr-1. On the other hand, the deletion mutant that lacks the entire helicase domain (ΔHelicase) could not distinguish between the substrate structures (**Fig. 3e**); although the size of the dicing products remained unchanged (**Supplementary Fig. 5**), the dicing activity for wild-type pre-*let-7* was markedly reduced at a level similar to that of the long-stem and small-loop variants. ΔDEXDc and ΔHELICc mutants also lost their substrate specificity (**Fig. 3f**,**g**), suggesting that these two domains act together to recognize cognate substrates. The requirement of the helicase domain for accurate substrate recognition by Dcr-1 was also supported by a set of dicing experiments using lysate from S2 cells, in which endogenous Dcr-1 was knocked down and then untagged intact Dcr-1 or its ΔHelicase mutant was complementarily expressed (**Supplementary Fig. 6a–c**).

Next, to test if the helicase domain contributes to substrate discrimination by its direct binding, we examined the binding affinity of the catalytic mutant of ΔHelicase Dcr-1 for pre-*let-7* and the structural variants (**Supplementary Fig. 2c**,**d**). Overall, the determined K_d values of ΔHelicase Dcr-1 were much greater (that is, affinity was much weaker) than the full-length Dcr-1, as summarized in **Table 2**. Notably, the difference between the relative affinities ($K_{relative}$) of

Table 1 The measured and relative affinities (±s.d.) of full-length Dcr-1 for pre-*let-7* and two structural variants

Substrate	K_d (nM)	$K_{relative}$
pre-*let-7* (S22, L14) (wild type)	25.4 ± 3.2	1.0 ± 0.4
pre-*let-7* (S34, L14) (long stem)	56.5 ± 4.4	2.2 ± 0.7
pre-*let-7* (S22, L4) (small loop)	147.7 ± 60.6	5.8 ± 2.5

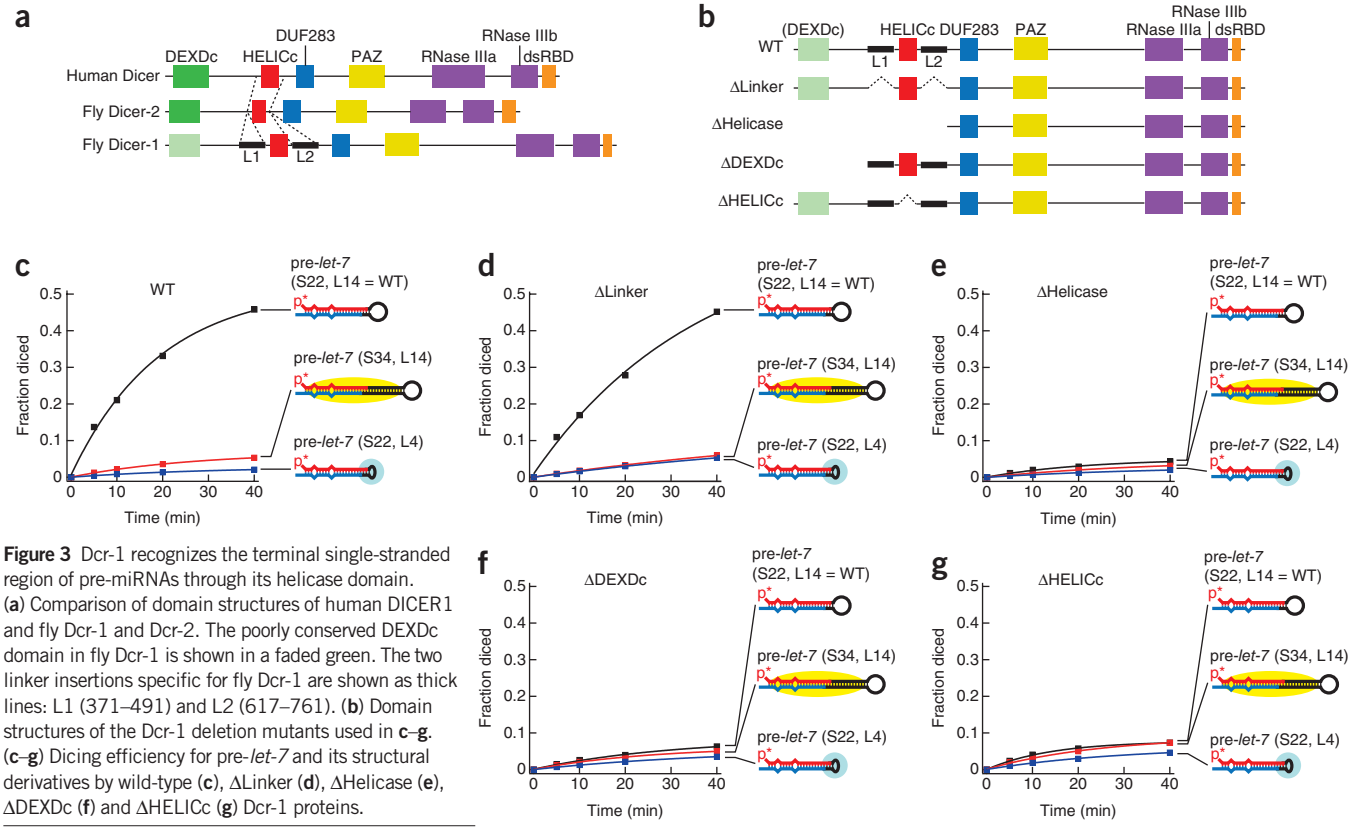

Figure 3 Dcr-1 recognizes the terminal single-stranded region of pre-miRNAs through its helicase domain. (**a**) Comparison of domain structures of human DICER1 and fly Dcr-1 and Dcr-2. The poorly conserved DEXDc domain in fly Dcr-1 is shown in a faded green. The two linker insertions specific for fly Dcr-1 are shown as thick lines: L1 (371–491) and L2 (617–761). (**b**) Domain structures of the Dcr-1 deletion mutants used in **c**–**g**. (**c**–**g**) Dicing efficiency for pre-*let-7* and its structural derivatives by wild-type (**c**), ΔLinker (**d**), ΔHelicase (**e**), ΔDEXDc (**f**) and ΔHELICc (**g**) Dcr-1 proteins.

ΔHelicase Dcr-1 for wild-type pre-*let-7* and for the structural variants (~1.4–1.9 fold, **Table 1**) was much smaller than that of full-length Dcr-1 (2–6 fold, **Table 2**), supporting the idea that the ΔHelicase mutant has largely lost its ability to distinguish the substrate structures. Given that the helicase domain in general shows RNA-binding activity[34–37], the simplest explanation is that Dcr-1 recognizes the terminal single-stranded region by direct binding through its helicase domain. Such recognition by Dcr-1 requires the availability of flexible single-stranded RNAs of appropriate length (**Fig. 1e**,**f**), which should ensure a sufficient association rate between the single-stranded region and Dcr-1's helicase domain.

DISCUSSION

Here we show that fly Dcr-1 specifically inspects the pre-miRNA structure. It is known that the interaction between Dcr-1 and its partner protein Loqs, especially a specific isoform of Loqs (Loqs-PB), enhances Dcr-1's dicing activity[3–5]. Indeed, we confirmed overall enhancement by recombinant Loqs-PB in our *in vitro* dicing assay. However, Loqs-PB did not affect the relative substrate specificity by Dcr-1 (**Supplementary Fig. 6d**,**e**), indicating that Dcr-1 alone can distinguish the substrate structures. Based on our data, we postulate that the helicase domain of Dcr-1 binds to the terminal loop and checks the loop size. Dcr-1 also measures the distance from the 3′ overhang to the terminal single-stranded loop through the PAZ domain and the helicase domain, respectively. Therefore, the PAZ and helicase domains of Dcr-1 together inspect the canonical pre-miRNA structure, before the actual cleavage occurs (**Fig. 4**).

Because the size of the Dcr-1 products was independent of the length of the stem region of the substrate pre-miRNAs (**Supplementary Fig. 1d**), the catalytic core of the two RNase III domains must be positioned at a fixed distance from the PAZ domain (**Fig. 4**), as previously demonstrated by the crystal structure of the minimum *Giardia intestinalis* Dicer that consists simply of the PAZ and RNase III domains and no helicase domain[33]. Indeed, ΔHelicase Dcr-1 diced pre-miRNAs at exactly the same position as the wild-type Dcr-1 did, albeit with lower efficiency (**Fig. 3e** and **Supplementary Fig. 5**). The dicing site of typical pre-miRNAs, including pre-*let-7*, resides only a few nucleotides away from the terminal loop. Thus, our model postulates that the helicase domain is located adjacent to the two RNase III domains on the opposite side of the PAZ domain (**Fig. 4**). Recent single-particle EM analyses have revealed that human DICER1 has an L-shaped overall structure, with the helicase domain located at the base arm of the L[39]. Experiments in which the crystal structure of *G. intestinalis* Dicer was docked into the EM density map of human DICER1 have resulted in two alternative models regarding the orientation of the PAZ domain and the RNase III domains[39,40]. In Model A, the PAZ domain resides near the top of the L and the RNase III domains are located at the base of the L[40], whereas in Model B, the PAZ domain is oriented toward the base of the L and the RNase III domains reside at the top of the L[39]. Our biochemical data for fly Dcr-1 strongly support Model A, in which the helicase domain and the RNase III domains are closely positioned at the base of the L (**Fig. 4**). Model A is also consistent with a recently proposed model for fly Dcr-2, in which its helicase domain acts to hold onto and translocate along

Table 2 The measured and relative affinities (±s.d.) of ΔHelicase Dcr-1 for pre-*let-7* and two structural variants

Substrate	K_d (nM)	$K_{relative}$
pre-*let-7* (S22, L14) (wild type)	121.0 ± 7.8	1.0 ± 0.1
pre-*let-7* (S34, L14) (long stem)	170.9 ± 14.6	1.4 ± 0.2
pre-*let-7* (S22, L4) (small loop)	229.4 ± 9.4	1.9 ± 0.1

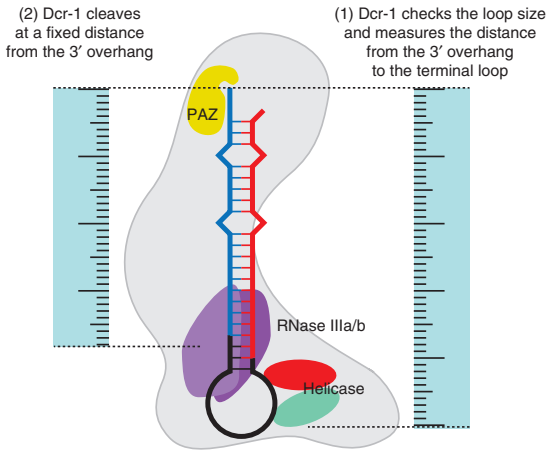

Figure 4 A proposed model for recognition of authentic pre-miRNA structures by Dcr-1. Dcr-1 checks the loop size by means of its helicase domain and measures the distance from the 3′ overhang to the single-stranded terminal loop region through its PAZ and helicase domains. Only when this distance is 'correct', can Dcr-1 dice the substrate through RNase IIIa and RNase IIIb domains located at a fixed distance from the 3′ overhang.

the substrates of long dsRNA[30,38]. However, given the functional diversity of the helicase domain of Dicer proteins (see below), it is currently unknown whether all the Dicer proteins have the same molecular architecture.

The helicase domain in Dicer proteins belongs to the RIG-I–like family in the helicase superfamily 2 (SF2)[34,35], yet functions of this domain appear extremely diverse. Unlike fly Dcr-1, fly Dcr-2 requires ATP and the functional helicase domain for processive dicing of long dsRNAs[29,30,38,41]. *Caenorhabditis elegans* DCR-1 also requires ATP and the helicase domain to process long dsRNAs with blunt or 5′-overhanging termini; helicase-domain mutants of DCR-1 are defective for production of a subset of endogenous siRNAs but retain the dicing activity of pre-miRNAs[38,42]. *Arabidopsis thaliana* DICER-LIKE 1 (DCL1), which possesses unique functionality to process pri-miRNAs into miRNA-miRNA* duplexes from the base to the terminal loop or the other way around[43–46], also requires ATP[47]. On the other hand, human DICER1 does not require ATP[27,28], and mutations or deletion of its helicase domain show no apparent defect[27] or even an enhanced dicing activity for pre-miRNAs and long dsRNAs[48,49]. What creates such diverse functionality? The alignment of the amino acid sequences of the helicase domains of animal Dicer proteins suggests that insect Dcr-2 proteins have strongly conserved canonical motifs that are similar to those in RIG-I. The helicase domain of RIG-I is linked to rigorous ATP-dependent dsRNA translocase activity, which is thought to act as a sensor for replicating viruses[50]. A translocase-like action is presumably required for processive dicing of long dsRNAs by insect Dcr-2 proteins[30,38], which can be explained by their well-conserved RIG-I–like helicase domain (**Supplementary Figs. 3** and **4**). By contrast, insect Dcr-1 proteins have the least conserved motifs, especially motif I, which is associated with ATP binding and hydrolysis, and motif III, which is important for intradomain interactions, both of which are largely disrupted in the DEXDc domain (**Supplementary Fig. 3**). Such a unique helicase domain of insect Dcr-1 proteins may be optimized for the stable single-stranded RNA binding required for the recognition of the terminally single-stranded region of pre-miRNAs. On the other hand, human DICER1 and *C. elegans* DCR-1, which are single Dicers that possess the dual functionality of processing both pre-miRNAs and long dsRNAs, appear to lie between insect Dcr-1 and Dcr-2 proteins in the phylogenetic tree, according to the sequence alignment of their helicase domains (**Supplementary Figs. 3** and **4**). Indeed, recombinant human DICER1 did not discriminate between wild-type pre-*let-7* and the structural derivatives (**Supplementary Fig. 7b**), reinforcing the idea that it recognizes substrates promiscuously. *C. elegans* DCR-1, whose helicase domain is more similar to insect Dcr-2 proteins, retains ATP-dependent processive dicing of long dsRNAs[38], but human DICER1, whose helicase domain is more homologous to insect Dcr-1 proteins, does not apparently require ATP[27,28,48,49]. Therefore, these two dual-functioning Dicer proteins seem to use different strategies to recognize both long dsRNAs and pre-miRNAs. Together, these results lead us to propose that the sequences and functions of the Dicer helicase domains have evolved to accommodate their specific substrates.

Given that fly Dcr-1 (and presumably other insect Dcr-1 proteins) exclusively recognizes the pre-miRNA structure, natural miRNA genes in insects may, in theory, have evolved to adapt their structures to Dcr-1's stringent substrate specificity. If so, pre-miRNAs in insects may conserve more homogeneous structures than those in other organisms. To address this question, we compared the structural heterogeneity of pre-miRNAs in *D. melanogaster*, *C. elegans*, *Homo sapiens* and *Mus musculus*. We used publicly available deep-sequencing libraries, which accurately describe the real pre-miRNA sequences, and predicted the secondary structures (see Methods). The average size of the loop and the average distance from the 3′ overhang to the terminal loop of fly pre-miRNAs agreed well with our biochemical characterization of substrate recognition by fly Dcr-1 (**Supplementary Fig. 7c–e**). However, our data indicate that the loop size and the overhang-to-loop distance are highly similar across species (**Supplementary Fig. 7c–e**). Therefore, it is more likely that the miRNA gene structure has been shaped by other constraints (for example, a loop that is too long will make the miRNA transcripts vulnerable to promiscuous RNases and a stem that is too long will make the miRNA gene unstable in the genome) . It is also more likely that insect Dcr-1 proteins have evolved their amino acid sequences and functions so that they can specifically recognize pre-miRNAs but not long dsRNAs. At the same time, insect Dcr-2 may have evolved to be optimized for efficient dicing of long dsRNAs while excluding pre-miRNAs with the aid of inorganic phosphate and R2D2 (ref. 30). Such functional compartmentalizing of Dicer proteins[29,30] as well as Ago proteins[11,51–53] would minimize the competition between the siRNA pathway, which is actively used by insects to fight against viruses and transposons, and the miRNA pathway, which is vital for gene regulation.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We are grateful to Q. Liu (University of Texas Southwestern Medical Center) for the pFastBac-His-Dcr-1 wild type and E1908A E2139A plasmids, to M. Siomi and H. Siomi (Keio University) for the plasmid for expressing Loqs-PB, and to S. Katsuma for his advice on protein expression. We thank P. Bas Kwak for assistance with the amino acid sequence alignment, and we thank K. Förstemann and members of the Tomari laboratory for discussions, suggestions and critical comments on the manuscript. This work was supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas ('Functional machinery for noncoding RNAs') and a a Grant-in-Aid for Young Scientists from the Japan Ministry of Education, Culture,

Sports, Science and Technology to Y.T.; and a Carrier Development Award from The International Human Frontier Science Program Organization to Y.T. N.I. is a recipient of a Research Fellowship from the Japan Society for the Promotion of Science.

AUTHOR CONTRIBUTIONS

A.T. conducted biochemical experiments with the assistance of T.K. and N.I., H.S. conducted bioinformatic analyses, Y.T. supervised the study; A.T. and Y.T. wrote the manuscript, and all authors discussed the results and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
2. Kim, V.N., Han, J. & Siomi, M.C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
3. Förstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
4. Saito, K., Ishizuka, A., Siomi, H. & Siomi, M.C. Processing of pre-microRNAs by the Dicer-1–Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3**, e235 (2005).
5. Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* **19**, 1674–1679 (2005).
6. Chendrimada, T.P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744 (2005).
7. Haase, A.D. *et al.* TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.* **6**, 961–967 (2005).
8. Lee, Y. *et al.* The role of PACT in the RNA silencing pathway. *EMBO J.* **25**, 522–532 (2006).
9. Kok, K.H., Ng, M.H., Ching, Y.P. & Jin, D.Y. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J. Biol. Chem.* **282**, 17649–17657 (2007).
10. Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R. & Hannon, G.J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
11. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M.C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
12. Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
13. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
14. Iwasaki, S. *et al.* Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* **39**, 292–299 (2010).
15. Iki, T. *et al. In vitro* assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol. Cell* **39**, 282–291 (2010).
16. Miyoshi, T., Takeuchi, A., Siomi, H. & Siomi, M.C. A direct role for Hsp90 in pre-RISC formation in *Drosophila*. *Nat. Struct. Mol. Biol.* **17**, 1024–1026 (2010).
17. Kawamata, T., Seitz, H. & Tomari, Y. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat. Struct. Mol. Biol.* **16**, 953–960 (2009).
18. Yoda, M. *et al.* ATP-dependent human RISC assembly pathways. *Nat. Struct. Mol. Biol.* **17**, 17–23 (2010).
19. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
20. Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W. & Sontheimer, E.J.A. Dicer-2-dependent 80S complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* **117**, 83–94 (2004).
21. Tomari, Y. *et al.* RISC assembly defects in the *Drosophila* RNAi mutant *armitage*. *Cell* **116**, 831–841 (2004).
22. Matranga, C., Tomari, Y., Shin, C., Bartel, D.P. & Zamore, P.D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).
23. Rand, T.A., Petersen, S., Du, F. & Wang, X. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**, 621–629 (2005).
24. Leuschner, P.J., Ameres, S.L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320 (2006).
25. Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M.C. Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev.* **19**, 2837–2848 (2005).
26. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
27. Zhang, H., Kolb, F.A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002).
28. Provost, P. *et al.* Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.* **21**, 5864–5874 (2002).
29. Lee, Y.S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
30. Cenik, E.S. *et al.* Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease. *Mol. Cell* **42**, 172–184 (2011).
31. Ye, X., Paroo, Z. & Liu, Q. Functional anatomy of the *Drosophila* microRNA-generating enzyme. *J. Biol. Chem.* **282**, 28373–28378 (2007).
32. Ma, J.B., Ye, K. & Patel, D.J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).
33. Macrae, I.J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006).
34. Fairman-Williams, M.E., Guenther, U.P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010).
35. Jankowsky, E. RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* **36**, 19–29 (2011).
36. Linder, P. Dead-box proteins: a family affair—active and passive players in RNP-remodeling. *Nucleic Acids Res.* **34**, 4168–4180 (2006).
37. Hilbert, M., Karow, A.R. & Klostermeier, D. The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *Biol. Chem.* **390**, 1237–1250 (2009).
38. Welker, N.C. *et al.* Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol. Cell* **41**, 589–599 (2011).
39. Wang, H.W. *et al.* Structural insights into RNA processing by the human RISC-loading complex. *Nat. Struct. Mol. Biol.* **16**, 1148–1153 (2009).
40. Lau, P.W., Potter, C.S., Carragher, B. & MacRae, I.J. Structure of the human Dicer-TRBP complex by electron microscopy. *Structure* **17**, 1326–1332 (2009).
41. Nykänen, A., Haley, B. & Zamore, P.D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
42. Welker, N.C. *et al.* Dicer's helicase domain is required for accumulation of some, but not all, *C. elegans* endogenous siRNAs. *RNA* **16**, 893–903 (2010).
43. Mateos, J.L., Bologna, N.G., Chorostecki, U. & Palatnik, J.F. Identification of microRNA processing determinants by random mutagenesis of *Arabidopsis MIR172a* precursor. *Curr. Biol.* **20**, 49–54 (2010).
44. Song, L., Axtell, M.J. & Fedoroff, N.V. RNA secondary structural determinants of miRNA precursor processing in *Arabidopsis*. *Curr. Biol.* **20**, 37–41 (2010).
45. Werner, S., Wollmann, H., Schneeberger, K. & Weigel, D. Structure determinants for accurate processing of miR172a in *Arabidopsis thaliana*. *Curr. Biol.* **20**, 42–48 (2010).
46. Bologna, N.G., Mateos, J.L., Bresso, E.G. & Palatnik, J.F. A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J.* **28**, 3646–3656 (2009).
47. Dong, Z., Han, M.H. & Fedoroff, N. The RNA-binding proteins HYL1 and SE promote accurate *in vitro* processing of pri-miRNA by DCL1. *Proc. Natl. Acad. Sci. USA* **105**, 9970–9975 (2008).
48. Ma, E., MacRae, I.J., Kirsch, J.F. & Doudna, J.A. Autoinhibition of human dicer by its internal helicase domain. *J. Mol. Biol.* **380**, 237–243 (2008).
49. Soifer, H.S. *et al.* A role for the Dicer helicase domain in the processing of thermodynamically unstable hairpin RNAs. *Nucleic Acids Res.* **36**, 6511–6522 (2008).
50. Myong, S. *et al.* Cytosolic viral sensor RIG-I is a 5′-triphosphate-dependent translocase on double-stranded RNA. *Science* **323**, 1070–1074 (2009).
51. Tomari, Y., Du, T. & Zamore, P.D. Sorting of *Drosophila* small silencing RNAs. *Cell* **130**, 299–308 (2007).
52. Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y. & Zamore, P.D. *Drosophila* microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell* **130**, 287–297 (2007).
53. Iwasaki, S., Kawamata, T. & Tomari, Y. *Drosophila* Argonaute1 and Argonaute2 use distinct mechanisms for translational repression. *Mol. Cell* **34**, 58–67 (2009).

ONLINE METHODS

Substrate RNA preparation. The sequences of pre-*let-7* and its derivatives used in this study are shown in **Supplementary Table 1**. All the RNA oligonucleotides were chemically synthesized (Dharmacon, Hokkaido System Science and Gene Design). To construct the long-stem series of pre-*let-7* derivatives, the 5′-half- and 3′-half fragments were synthesized, annealed and ligated with T4 RNA ligase (Takara Bio), then 5′ radiolabeled by T4 polynucleotide kinase (Takara Bio) and [γ-^{32}P] ATP. The substrates were gel-purified, dissolved in lysis buffer (30 mM HEPES-KOH, pH 7.4, 100 mM KOAc and 2 mM $Mg(OAc)_2$), heated at 95 °C for 2 min and annealed at room temperature for 1 h.

Plasmid construction. The gene sequences of Dcr-1 and its catalytic mutant (E1908A E2139A) were amplified from pFastBac-His-Dcr-1 wild-type and E1908A E2139A mutant vectors[31] by PCR using oligonucleotide primers (5′-CACCATGGCGTTCCACTGGTGCG-3′) and (5′-TTAGTCTTTTTTGGCTATCAAGC-3′), and cloned into pENTR/D-TOPO (Invitrogen). The Flag-tag sequence in pAFW (*Drosophila* Gateway vector collection, Actin5C promoter and N-terminal 3 × Flag-tag) was replaced with the SBP-tag sequence to construct the pASW destination vector[14]. pENTR-Dcr-1 wild type and E1908A E2139A mutant were recombined with pASW by an LR Clonase enzyme (Invitrogen) reaction. The expression vectors of Dcr-1 deletion mutants were constructed from pASW-Dcr-1 wild type or E1908A E2139A mutant using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the following oligonucleotide primers: 5′-CCCCTTCACCAGCACCACAAAGGACTTGG-3′ and 5′-TTGTGGTGCTGGTGAAGGGGGCGG-3′ for ΔHelicase; 5′-CCCCTTCACCGTGCACACCCAAGCGGACGG-3′ and 5′-GGGTGTGCACGGTGAAGGGGGCGGCCGCGG-3′ for ΔDexDc; and 5′-TGGCAGCGACAGTTATAAAAGCCCAACTG-3′ and 5′-GCTTTTATAACTGTCGCTGCCATCATTGTGATC-3′ for ΔHELICc. For ΔLinker, site-directed mutagenesis was first carried out from pASW-Dcr-1 wild type using oligonucleotide primers 5′-CGTCGCTCCGAGCACCACAAAGGACTTGGTG-3′ and 5′-TTGTGGTGCTCGGAGCGACGAGAATGACATG-3′. Using the resulting vector as the template, subsequent site-directed mutagenesis was carried out using oligonucleotide primers 5′-GCCGGAAGAGACGCTCTGCGCACTGATTTAC-3′ and 5′-CGCAGAGCGTCTCTTCCGGCTTGAAGCACC-3′. To construct the expression vector for untagged Dcr-1, pENTR-Dcr-1 wild type was recombined with pAWH (*Drosophila* Gateway vector collection, Actin5C promoter and C-terminal 3× hemagglutinin (HA)-tag, Invitrogen). Note that pENTR-Dcr-1 wild type has an in-frame stop codon, so the C-terminal 3× HA-tag of pAWH is not translated. For pAWH-Dcr-1 ΔHelicase, site-directed mutagenesis was carried out with pAWH-Dcr-1 wild type using the same set of oligonucleotide primers used to construct pASW-Dcr-1 ΔHelicase.

Expression and purification of recombinant proteins. To express SBP-tagged Dcr-1 or mutant proteins, S2 cells were transfected with pASW-Dcr-1 wild type or mutant vectors by using the Fugene HD Transfection Reagent (Roche). Cells were plated onto 10-cm dishes at a density of 1×10^6 cells per ml in antibiotic-free medium, transfected with 10 μg plasmid DNA with 30 μl Fugene HD per dish and harvested after 48 h. Typically, 200 μl lysate from S2 cells expressing SBP-tagged Dcr-1 proteins was incubated in 40 μl Streptavidin Sepharose High Performance (GE healthcare) for 3 h at 4 °C. The beads were washed three times with lysis buffer containing 0.5 M NaCl and 1 mM DTT, and rinsed twice with lysis buffer containing 1 mM DTT. SBP-tagged Dcr-1 proteins were eluted with 50 μl lysis buffer containing 5 mM biotin, pH 8.0, 50% (v/v) glycerol, 0.01% (w/v) BSA and 1 mM DTT (**Supplementary Fig. 1a**). The concentration of each recombinant protein was determined by quantitative amino acid analysis of the protein in a slice from an SDS-PAGE assay (Keck Biotechnology Resource Laboratory). Recombinant Loqs-PB was expressed in *E. coli* and purified as described previously[4]. Recombinant human DICER1 was purchased from Invitrogen.

S2 cell RNAi and lysate preparation. Dcr-1 3′ UTR sequence was PCR amplified using primers 5′-GAAGTCATCGCTGCAATCGTGTTTCATTAGATTTAATGATTATTTTTTTAGACATAAGAAC-3′ and 5′-TTATTCGACCATAGACAATCTATTTAAGATTTGTACAATTAAAGTTCTTATGTCTAAAAAAAT-3′ and subsequently cloned into pMT20-T vector (Takara Bio). The dsRNA transcription template was amplified using oligonucleotide primers 5′-TAATACGACTCACTATAGGGAGAGAAGTCATCGCTGCAATCG-3′ and 5′-TAATACGACTCACTATAGGGAGATTATTCGACCATAGACAATC-3′, both containing the T7 promoter sequence. The dsRNA was synthesized from the template using the AmpliScribe T7 High Yield Transcription Kit (Epicentre) and used to knock down endogenous Dcr-1 by the soaking method, as previously described[3]. The expression vectors of untagged Dcr-1 wild type and ΔHelicase were transfected into S2 cells 4 d after initial dsRNA treatment. Cells were harvested 6 d after first dsRNA treatment. S2 lysate was prepared as previously descried[3].

Dicing assay. *In vitro* dicing assays were typically carried out in 10 μl lysis buffer, containing 5% (v/v) glycerol, 1 mM DTT, 0.1 unit μl^{-1} RNasin Plus RNase Inhibitor (Promega), 1 nM 5′-radiolabeled substrate RNAs and 5 nM recombinant Dicer proteins. The reaction products were resolved by electrophoresis on 10% denaturing Page, detected by FLA-7000 image analyzer (Fujifilm) and quantified by MultiGauge software (Fujifilm). ATP and its analogs were added at a final concentration of 1 mM (**Supplementary Fig. 2f**).

Gel shift assay. Two μL of 1 nM 5′-radiolabeled substrate RNAs were incubated with 2 μl of recombinant wild type or mutant Dcr-1 proteins in lysis buffer containing 30% (v/v) glycerol, 1 mM DTT, 0.1 unit μl^{-1} RNasin Plus RNase Inhibitor (Promega) at 25 °C for 20 min. Free RNA and Dcr-1–RNA complexes were resolved by electrophoresis on 6% nondenaturing Page, detected by a FLA-7000 image analyzer (Fujifilm), quantified by MultiGuage software (Fujifilm) and fit to the Hill equation with IGOR Pro software (WaveMetrics).

Analysis of pre-miRNA stem length and loop size. Pre-miRNA sequences were annotated using miRBase (version 16: http://www.mirbase.org/) and available deep-sequencing libraries (*D. melanogaster*: GSM246084, GSM239041, GSM239050, GSM239051, GSM239052, GSM239053, GSM239054, GSM239055, GSM239056, GSM272651, GSM275691, GSM272652, GSM272653, GSM180332, GSM180329, GSM180335, GSM180333, GSM180330 and GSM180336; *C. elegans*: GSM336052, GSM336055, GSM336056, GSM336058, GSM336059, GSM336060, GSM336086, GSM336572, GSM139137, GSM297751, GSM297757, GSM297743, GSM297747, GSM297746, GSM297750, GSM297744, GSM297742, GSM297745 and GSM297748; *H. sapiens*: GSM569185, GSM569186, GSM569187, GSM569188, GSM569189, GSM569190, GSM565978, GSM565979, GSE20417, GSM494809, GSM494810, GSM494811, GSM494812, GSM541796, GSM541797, GSM458535, GSM458536, GSM458537, GSM458538, GSM458539, GSM458540, GSM458541, GSM458542, GSM458543, GSM458544, GSM458545, GSM458546, GSM522374, GSM337570 and GSM337571; *M. musculus*: GSM314552, GSM314558, GSM304914, GSM237107, GSM237110, GSM261959 and GSM261957). Only pre-miRNAs yielding at least 100 reads in the pooled deep-sequencing libraries were considered. For each pre-miRNA, alternative Drosha sites were weighted according to the abundance of the corresponding reads in the libraries. The pre-miRNA secondary structure was predicted by defining the base of the loop as the last base pair of the stem with a pairing probability above 0.5 (pairing probability was evaluated using RNAfold -p, from the Vienna RNA package 1.8.3 (http://www.tbi.univie.ac.at/~ivo/RNA/).