

Structural determinants of RNA recognition and cleavage by Dicer

Ian J MacRae^{1,2}, Kaihong Zhou^{1,2} & Jennifer A Doudna^{1–4}

A hallmark of RNA interference is the production of short double-stranded RNA (dsRNA) molecules 21–28 nucleotides in length by the specialized RNase III protein Dicer. Dicer enzymes uniquely generate RNA products of specific lengths by mechanisms that have not been fully elucidated. Here we show that the PAZ domain responsible for dsRNA end recognition confers this measuring ability through both its structural position and RNA-binding specificity. Point mutations define the dsRNA-binding surface and reveal a protein loop important for cleavage of substrates containing perfect or imperfect base pairing. On the basis of these results, we reengineered Dicer with a U1A RNA-binding domain in place of the PAZ domain to create an enzyme with altered end-recognition specificity and RNA product length. These results explain how Dicer functions as a molecular ruler and provide a structural basis for modifying its activity in cells.

Dicer enzymes are a specialized family of RNase III proteins that produce and help traffic small dsRNAs during RNA interference (RNAi)¹. Found in the cytoplasm of nearly all eukaryotic cells, Dicer recognizes the 5' and 3' helical ends of dsRNA and cleaves a specific distance away to produce 21- to 28-nucleotide (nt) short interfering RNAs (siRNAs)² or microRNAs³. In addition, Dicer helps load these RNA products into multiprotein RNA-induced silencing complexes^{4–7}, where they direct cognate gene silencing by targeted mRNA degradation⁸, translational repression⁹ and heterochromatin formation¹⁰.

Dicer contains two copies of the universally conserved catalytic domain of RNase III proteins. Dicer enzymes also typically include an N-terminal ATPase/DEXD helicase domain, a small domain of unknown function (DUF283) and a C-terminal dsRNA-binding domain¹¹. A sixth domain called PAZ is located a variable distance upstream of the two catalytic domains. Notably, a few Dicers seem to lack either the ATPase/helicase domain or dsRNA-binding domain, or both, suggesting that these regions modify dicing activity but do not contribute directly to RNA measurement or catalysis. Furthermore, several Dicer proteins, including *Tetrahymena thermophila* Dcr1 and Dcr2 (ref. 12) and *Schizosaccharomyces pombe* Dcr1, do not have identifiable PAZ domains, raising the possibility of alternative mechanisms of RNA recognition.

Structural insight into Dicer function came from the crystal structure of an intact and fully active Dicer from *Giardia intestinalis*¹³. This enzyme naturally contains only the PAZ and two catalytic domains found within the larger Dicers. The two catalytic domains associate with each other to form an 'internal dimer' that resembles the RNase homodimer of bacterial RNase III enzymes¹⁴. These catalytic domains are linked to the PAZ domain at the opposite end of the molecule by a long α -helix and a flat platform surface on the

face of the protein. The 65-Å distance between the PAZ domain and the catalytic domain active sites closely matches the length spanned by 25–27 base pairs (bp) of dsRNA, the length of the RNAs produced by *G. intestinalis* Dicer. This structure, coupled with available biochemical data, has led to a model for dsRNA binding and cleavage similar to that proposed for human Dicer¹⁵. In the model, Dicer binds dsRNA 3' ends using the PAZ domain and positions the dsRNA substrate along the flat face of the enzyme, in agreement with a recent crystal structure of bacterial RNase III bound to dsRNA¹⁶.

To test this model, we investigated the contributions of distinct parts of the Dicer structure to dsRNA recognition, product-length specificity and end-dependent dsRNA cleavage. We show here that the PAZ domain is a self-contained structure whose location within the Dicer architecture is sufficient to define the lengths of siRNAs. RNA binding and activity assays with a series of Dicer mutants show that electrostatic interactions contribute the bulk of the substrate-binding energy. A loop adjacent to one of the catalytic domain active sites is responsible for maintaining the correct size of diced RNAs produced from perfectly and imperfectly base-paired substrates. On the basis of these results, we created a new form of Dicer with altered dsRNA substrate specificity by replacing the PAZ domain with the U1A RNA loop-recognition domain, thereby changing RNA substrate specificity without affecting product length. These results explain how Dicer functions as a molecular ruler and suggest a mechanism for its regulation in cells.

RESULTS

The PAZ domain is required to specify siRNA length

Dicer's intrinsic measuring mechanism is postulated to rely on the PAZ domain. To test this idea directly, a mutant *G. intestinalis* Dicer

¹Howard Hughes Medical Institute, ²Department of Molecular and Cell Biology and ³Department of Chemistry, University of California, Berkeley, California 94720, USA. ⁴Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. Correspondence should be addressed to J.A.D. (doudna@berkeley.edu).

protein was designed in which the entire 113-residue PAZ domain was deleted (Δ PAZ) without disrupting the structure of the remaining protein (Fig. 1a). As predicted, Δ PAZ Dicer produced dsRNA products of variable length (Fig. 1b), suggesting that the truncated enzyme engages and cleaves at random positions in the dsRNA substrate. The Δ PAZ enzyme also had reduced affinity for dsRNA compared to that of the wild-type enzyme (data not shown). The finding that removal of the PAZ domain abolished cleavage site selectivity by *G. intestinalis* Dicer is consistent with the structural model and suggests that the dsRNA helical end is the primary determinant of cleavage site selection by Dicer.

Dicer measures from the 3' end of dsRNA substrates

To test whether *G. intestinalis* Dicer has the same preference for dsRNA substrates with free helical ends as observed for human Dicer¹⁷, we generated a covalently closed, circular dsRNA substrate similar in sequence to the hairpin precursor of let-7 microRNA³. In contrast to the hairpin substrate, Dicer did not hydrolyze the ligated circular dsRNA substrate, even after extended incubation (Supplementary Fig. 1 online). Thus, Dicer requires an open helical end for proper substrate recognition and processing.

We further probed the structural requirements for RNA helix end recognition by testing cleavage site selection in three 37-nt dsRNA substrates. The first substrate contained 3' 2-bp overhangs on each end of the duplex, the second substrate had blunt ends and the third substrate had a 3' 2-bp recession on both ends of the duplex (Fig. 1c). In each case, incubation with *G. intestinalis* Dicer yielded products resulting from cleavage 25 or 26 nt from the 3' end of both RNA strands (Fig. 1d). Cleavage from the 3' end of the radiolabeled strand always yielded an 11-nt product, consistent with measuring from that 3' end. A mixture of product sizes from a single substrate is commonly observed in the Dicer family of enzymes¹⁵. The exact product size and distribution of sizes seems to depend on substrate sequence¹⁸. Notably, Dicer generated RNA fragments 29 or 30 nt long from the substrate with the recessed 3' ends, which is ~4 nt longer than canonical *G. intestinalis* Dicer products. Thus, substrate structure contributes to siRNA length, making it possible for Dicer to generate longer RNA fragments than previously thought.

Figure 1 Role of the PAZ domain and helical end structure in cleavage site selection by Dicer.

(a) Crystal structure of wild-type *G. intestinalis* Dicer (left) and a probable model for Δ PAZ Dicer (right). Key domains and loops are labeled. (b) Fixed-time activity assays (25 nM enzyme, ~50 nM substrate, 3 h) with wild-type and Δ PAZ Dicers. (c) Schematic representation of the three dsRNA substrates used. Arrows point to observed cleavage sites. Lines attached to arrows denote distances measured by Dicer. (d) Fixed-time assay (25 nM enzyme, ~75 nM substrate, 10 min) of three dsRNA substrates. -OH indicates partial alkaline hydrolysis ladder of the labeled RNA. **Figure 1a**, **Figure 2b** and **Figure 4c** were generated using PyMOL²⁸.

A dsRNA-binding surface between the PAZ and catalytic domains

The structure of *G. intestinalis* Dicer revealed several patches of positively charged residues on the surface of the protein, particularly along the surface connecting the PAZ and the catalytic domains. To test the contributions of these residues to dsRNA substrate recognition, nine Dicer protein mutants were produced in which two to four surface residues were mutated to alanine (Table 1). Upon incubation with a 153-bp dsRNA substrate, the wild-type Dicer processed nearly all of the starting material into 25- to 27-nt RNAs over the course of the assay (Fig. 2a). In contrast, many of the mutant Dicers were much slower, producing only a fraction of the amount of products produced by wild-type enzymes. All of the mutations that impaired dicing activity mapped to the one side of the molecule, the 'front' (Fig. 2b). Mutation of positively charged patches elsewhere on the protein surface had relatively little effect on dicing activity. These results indicate that the positively charged residues on the front of the platform and PAZ domains are crucial to Dicer function.

Electrostatic interactions confer dsRNA binding affinity

We further characterized the surface mutants by assaying their activity with various amounts of the 37-nt dsRNA substrate shown in Figure 1c (left). Similar to the substrate used in a previous study¹⁵, this substrate can be cleaved once by Dicer from either end of the duplex to produce dsRNA products. Assays were carried out with duplexes in which a trace amount of one of the two RNA strands was labeled with ³²P on its 5' end.

Wild-type *G. intestinalis* Dicer cleaved the labeled RNA into fragments 25 or 26 nt long and 11 or 12 nt long (Fig. 1d), consistent

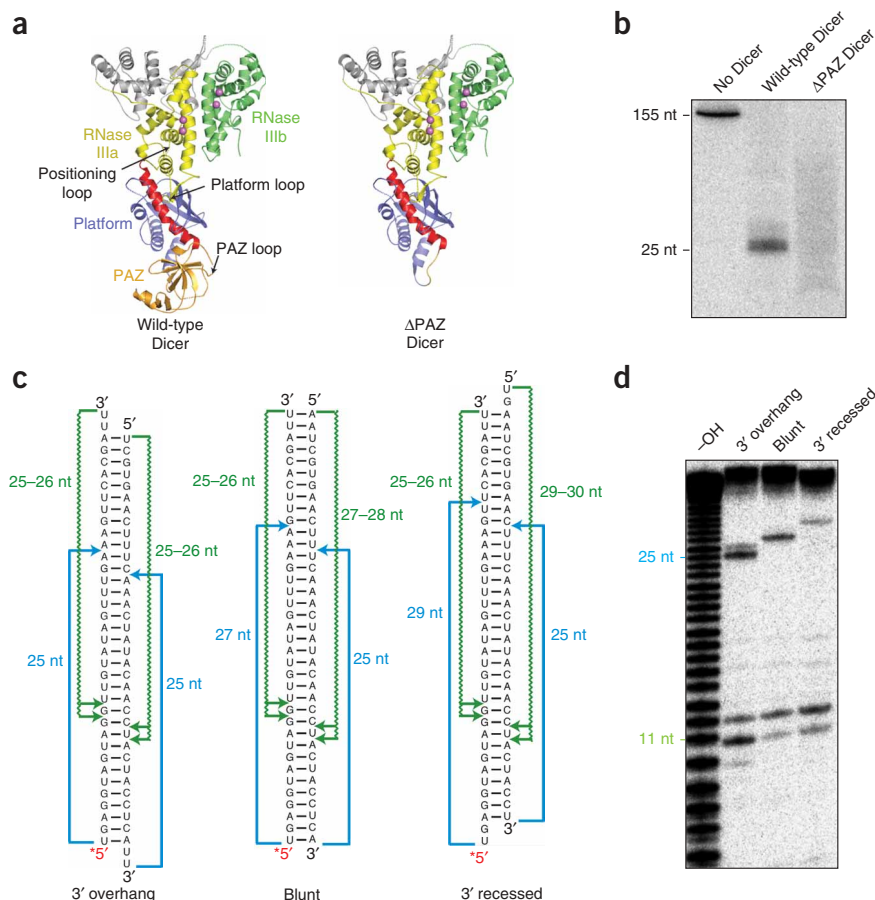


Table 1 Dicer mutants examined in this study

Mutant	Introduced mutations
Positioning loop 1	K400A K398A
Positioning loop 2	K400A K398A N394A N393A
Platform loop	R310A R312A
Platform	R99A H92A
PAZ loop 1	R217A K222A H224A
PAZ loop 2	K238A R240A
Backside bridge	R514A H513A
Underside PAZ	R206A H204A
Backside platform	K450A R454A K453A

with processing from both ends of the dsRNA substrate. Plotting the rate of product formation as a function of substrate concentration resulted in similar curves for both product sizes (Fig. 3a). Because a large molar excess of substrate over enzyme was used in these experiments, and because each end of a single dsRNA substrate can be recognized by Dicer, the ends act essentially as independent competitors of each other. Thus, the data can be fit to a modified version of the Henri-Michaelis-Menten equation, which accounts for two inseparable alternative substrates:

$$v_1 = [S_1]/(K_{m1} + [S_1] + K_{m1}[S_2]/K_{m2}) \quad (1)$$

where $[S_1]$ and $[S_2]$ are the concentrations of the two substrates, K_{m1} and K_{m2} are the Michaelis constants for each substrate and v_1 is the velocity at which substrate 1 is converted into product¹⁹. Taking into account that the concentrations of the two substrates must be equal ($[S_1] = [S_2]$) and that the velocity plots for the two products have roughly the same apparent K_m ($K_{m1} = K_{m2}$), equation (1) can be reduced to

$$v_1 = [S_1]/(K_{m1} + 2[S_1]) \quad (2)$$

Fitting the data to equation (2) yields a K_m value of 2.8 μM for both products and a theoretical maximal velocity of 24 s^{-1} and 22 s^{-1} for the 11- or 12-nt and the 25- or 25-nt products, respectively. The maximum velocity (V_{max}) of *G. intestinalis* Dicer is about ten-fold slower than that reported for *Escherichia coli* RNase III (228 s^{-1}), and the K_m of Dicer is about ten-fold greater than the reported K_m of bacterial RNase III for a viral dsRNA substrate (0.34 μM)²⁰.

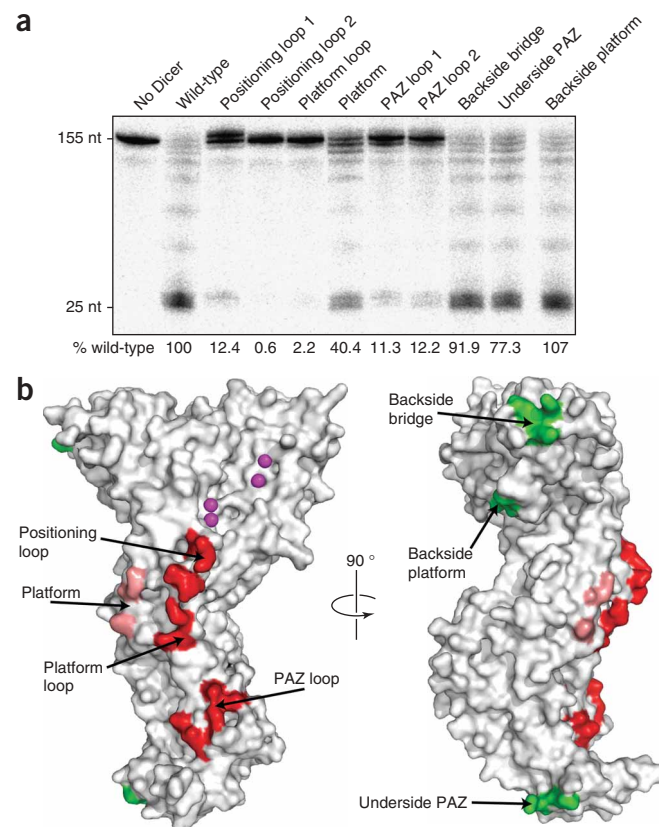
We next measured the initial velocity of the surface-mutant Dicers at various dsRNA concentrations. In most cases, the plot of velocity against substrate concentration was linear, indicating that even the highest dsRNA concentration used (10 μM) was well below the K_m values of most of the mutant enzymes (Fig. 3b). Only the plot of the platform mutant took on a hyperbolic shape, with a K_m about five times greater than that of the wild-type enzyme. The increased K_m values for the surface mutants suggest that the entire front face

Figure 2 Contributions of positively charged surface residues to dsRNA substrate recognition. (a) Fixed-time assay (25 nM enzyme, ~50 nM substrate, 1 h) of wild-type and surface-mutated *G. intestinalis* Dicers using a long dsRNA substrate. '% wild-type' indicates the amount of 25- to 27-nt RNA produced by each mutant over the course of the assay, as a percentage of the amount produced by the wild-type Dicer. (b) Surface representation of *G. intestinalis* Dicer showing residues important for enzyme activity clustered on the front of the enzyme. Mutants were characterized as unimpaired (>75% wild-type activity, green), impaired (>15% wild-type activity, pink) or severely impaired (<15% wild-type activity, red).

of Dicer contributes to substrate recognition and that upon binding there are many points of contact between the protein and dsRNA.

The positioning loop fine-tunes siRNA product length

Two of the surface mutant proteins we produced contained mutations in a 10-residue loop that we call the positioning loop (residues 391–401). This loop, part of the first catalytic domain (the RNase IIIa domain), resides directly below the active site at the junction between the platform and RNase IIIa domains (Fig. 1a). We were especially interested in this position of the protein because in the proposed model of dsRNA bound to *G. intestinalis* Dicer, the dsRNA had to be bent at the RNase IIIa–platform junction to allow a reasonable fit onto the Dicer crystal structure¹³. Notably, mutations in the loop markedly affected the activity of *G. intestinalis* Dicer, with the positioning loop 2 mutant being one of the most severely impaired surface mutants produced in this study (Table 1 and Fig. 2a). We also observed a subtler change in the positioning loop mutants: whereas the wild-type enzyme produced predominantly 25-nt RNAs from one end of the 37-nt substrate, positioning loop mutants produced a mixture of 25- and 26-nt RNAs (Fig. 4a). This effect was strongest in the positioning loop 2 mutant, which produced a higher ratio of 25- and 26-nt RNAs compared to the wild-type enzyme (Fig. 4b). The ratio of 25- and 26-nt products was constant throughout the course of the reaction for each enzyme in time-course experiments and was independent of substrate concentration (data not shown). The observed change in the size distribution of product RNAs suggests that the positioning loop is involved selecting the cleavage site and may have a role in placing the scissile phosphate correctly into the catalytic site. We propose that the positioning loop helps align or position dsRNA substrates into the enzyme active site as they extend across the RNase IIIa–platform domain junction.



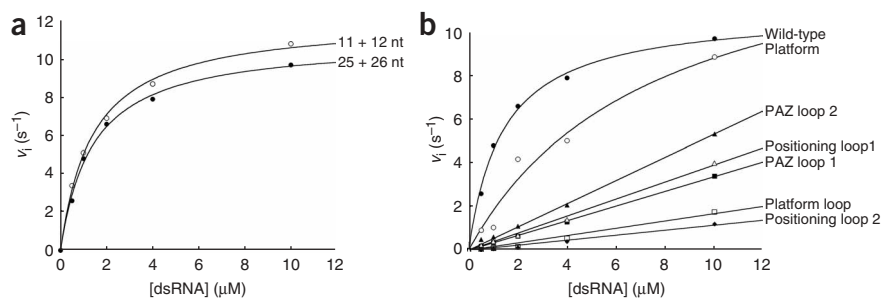
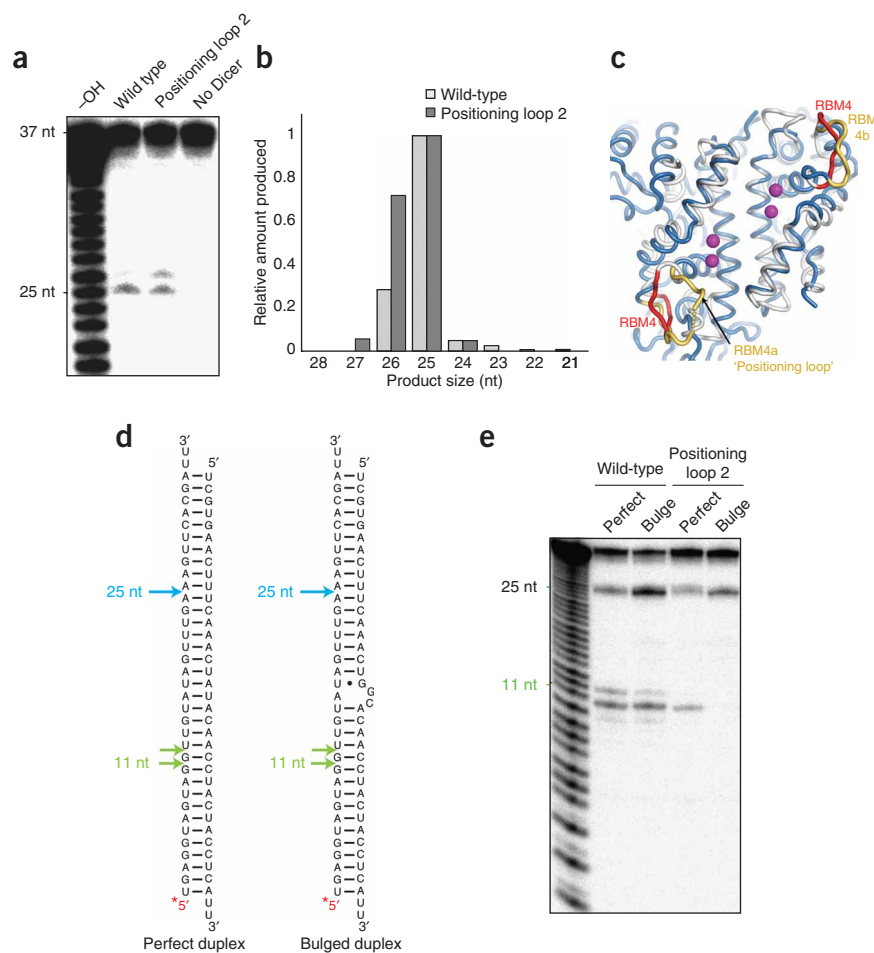


Figure 3 Kinetic measurements of Dicer surface mutants. (a) Steady-state velocity of wild-type *G. intestinalis* Dicer as a function of substrate concentration. Initial rates of 25- and 26-nt product formation and of 11- and 12-nt product formation are plotted. (b) Initial velocity of 25- and 26-nt product formation plotted against substrate concentration for surface-mutant Dicers.

The positioning loop in Dicer is structurally analogous to RNA-binding motif 4 (RBM4) in bacterial RNase III, which has also been implicated in substrate binding and catalytic efficiency²¹. However, the loop in the RNase IIIa domain of Dicer is considerably longer and more flexible²² (Fig. 4c), suggesting that the positioning loop has diverse modes of dsRNA binding. Notably, the positioning loop in human Dicer has an even larger extension that mediates interactions with Argonaute, the core subunit of the RNA-induced silencing complex²³. Thus, in mammals this loop has functions even beyond binding dsRNA substrates and assisting their hydrolysis. In contrast, the analogous loop in the RNase IIIb domain of Dicer is compact and structurally very similar to the RBM4 of bacterial RNase III, suggesting that it functions simply in substrate binding and catalytic efficiency.

We further examined the role of the positioning loop by challenging Dicer with an imperfect dsRNA substrate. The dsRNA

Figure 4 Role of the positioning loop in dsRNA recognition and cleavage. (a) Fixed-time assay of the wild-type Dicer (25 nM, ~75 nM substrate, 1 min) and positioning loop 2 mutant (125 nM, ~75 nM substrate, 10 min) acting on a 37-nt dsRNA substrate. Higher enzyme concentration and longer incubation time were required for the mutant to generate product levels comparable to the wild-type enzyme. -OH indicates a partial alkaline hydrolysis ladder of the labeled RNA. (b) Quantification of product size distribution from wild-type and positioning loop 2-mutant Dicer proteins. Amounts are shown relative to the most abundant product (25 nt). (c) Superposition of C α carbon atoms of the RNase domains of *G. intestinalis* Dicer (blue; PDB 2FFL) and the RBM4 region of bacterial RNase III (white; PDB 2EZ6). RBM4 of RNase III is shown in red, corresponding regions of Dicer (RBM4a and RBM4b) are in gold and catalytic metal ions are shown as purple spheres. (d) Schematic of perfect and bulged substrates. Arrows indicate Dicer cleavage sites. (e) Fixed-time assay of the wild-type Dicer (25 nM, ~75 nM substrate, 8 min) and positioning loop 2 mutant (125 nM, ~75 nM substrate, 120 min) acting on the perfect and bulged dsRNA substrates.



the bulged substrate were slightly enriched in 25-nt fragments relative to 11- and 12-nt fragments, suggesting that the bulge causes Dicer to favor processing from one end of the duplex over the other. This effect was exacerbated in the positioning loop 2 mutant, which did not produce any 11-nt products from the bulged substrate. The inability of the positioning loop 2 mutant to produce 11-nt products from the bulged substrate is consistent with a role for the loop in positioning the scissile phosphate for hydrolysis and suggests that, in some cases, the loop also helps Dicer recognize and process diverse dsRNA substrates.

substrate used was identical to the original except that it contained an internal bulge composed of a 2-bp mismatch and a single-base insertion (Fig. 4d). The bulge was designed to be in the middle of the duplex, which was expected to be near the RNase IIIa–platform junction when the duplex was engaged by Dicer from either end. The wild-type *G. intestinalis* Dicer processed the bulged and perfect substrates from both ends, producing a mixture of 25- or 26-nt and 11- or 12-nt products from each substrate (Fig. 4e). In contrast, the positioning loop 2 mutant produced 25- or 26-nt and 11-nt products but almost no 12-nt products, showing again that mutations in this loop subtly affect cleavage site selection. The products from

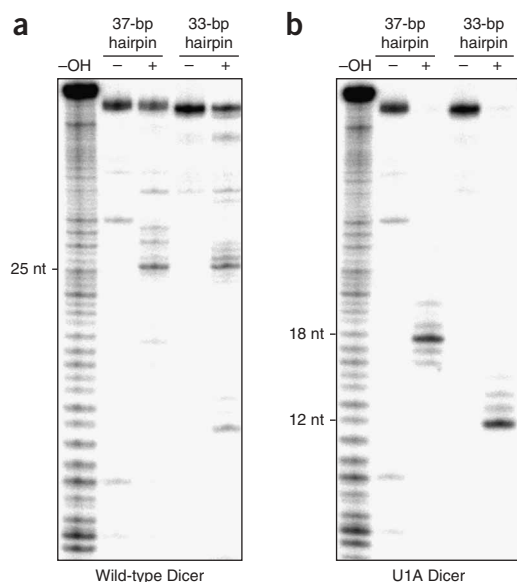


Figure 5 Activity of an engineered Dicer with altered RNA recognition specificity. (a,b) Fixed-time assay (25 nM enzyme, ~50 nM substrate, 15 min) of wild-type Dicer (a) and reengineered U1A Dicer (b) processing 37- and 33-bp RNA hairpins. Both hairpins are ^{32}P -labeled on the open 5' end and have the U1A recognition loop on the other end.

Dicer can be reprogrammed to recognize novel RNA substrates

The above results show that the PAZ domains in the Dicer protein function in substrate cleavage site selection, whereas the positively charged surface residues contribute to substrate binding affinity and subtle positioning. These findings imply that there is nothing inherently special about the PAZ domain; it should be possible to alter the substrate specificity of Dicer by replacing the PAZ domain with a completely different RNA-recognition motif. To test this idea, we designed, expressed and purified a mutant *G. intestinalis* Dicer protein in which the RNA-binding domain of the spliceosomal protein U1A was substituted for the PAZ domain. The U1A RNA-binding domain has a high affinity for RNA loops containing the 14-nt sequence 5'-CCAUUGCACUCCGG-3'. Two RNA hairpin substrates containing a U1A recognition loop at the end of a 37- or 33-bp stem were used to test the RNase activity and specificity of Dicer containing the wild-type or U1A RNA-binding domain.

The wild-type Dicer processed the two hairpin substrates in the same way: in each reaction, the predominant labeled product was 25 nt long (Fig. 5a), consistent with Dicer measuring from the open helical end. In contrast, the U1A Dicer processed the two RNA hairpins differently, producing 18- and 12-nt labeled RNAs from the 37- and 33-bp hairpins, respectively (Fig. 5b). Furthermore, substrates containing mutations in the U1A recognition sequence abolished hairpin recognition and cleavage by the U1A Dicer (data not shown). These results suggest that, as designed, the U1A Dicer measures and cleaves a set distance from the U1A recognition loop. The measuring distance of 19–21 nt by the U1A Dicer is slightly less than that of the wild-type *G. intestinalis*, which probably reflects differences between the depths of the RNA-binding pockets in the U1A and the PAZ domains. Notably, under the experimental conditions used (~50 nM hairpin RNA) the U1A Dicer was more efficient than the wild-type Dicer at cleaving the RNA substrates. This may reflect the fact that the U1A RNA-binding domain has a much higher

affinity for its cognate RNA-binding site ($K_d = 20$ pM) than the PAZ domain does for open helical ends (1 μM)^{24,25}.

DISCUSSION

Understanding how Dicer efficiently generates short RNAs for targeted gene silencing is essential to determining the molecular mechanisms and evolution of RNAi and enabling its manipulation for therapeutic purposes. On the basis of the crystal structure of *G. intestinalis* Dicer, mutants were designed to test the contributions of the PAZ domain and a variety of exposed loops and side chains to dsRNA recognition and processing. Our analysis shows that the Dicer PAZ domain is required for helical end-dependent measuring of dsRNA substrates. RNA measuring by the PAZ domain is strictly from the 3' end of dsRNA substrates, enabling production of RNAs longer than the canonical size from substrates containing a 5' extension. Although it is unknown whether similar substrates occur *in vivo*, it is notable that such activity by Dicer could yield longer RNA products not previously ascribed to Dicer. Furthermore, this activity by the *G. intestinalis* Dicer might explain the presence of small RNAs ~30 nt in length in *G. intestinalis*²⁶.

Extensive site-directed mutagenesis of Dicer surface residues has revealed that the primary interaction between Dicer and its substrates occurs along the flat, positively charged surface that connects the PAZ domain to the two catalytic domains. Amino acid changes in this region uniformly decreased dicing activity by increasing the apparent K_m of the enzyme for dsRNA substrates. Many of the residues in this region of the protein seem to contribute binding affinity through electrostatic interactions with the negatively charged RNA backbone.

A 10-residue segment termed the positioning loop, located at the junction between the platform and catalytic domains, has a subtler and potentially crucial role in correctly placing substrates in the enzyme active sites. Comparison of the four independent copies of Dicer in the crystallographic asymmetric unit has shown that this loop is highly flexible²². Mutation of the positioning loop markedly reduced dicing activity and changed the size distribution of RNA products, suggesting that the positioning loop helps align or direct dsRNA substrates into the enzyme active sites. This may be particularly important in enabling Dicer to process dsRNA substrates containing noncanonical base pairs such as precursor microRNAs³.

Perhaps the most exciting finding of this study is that Dicer can be reengineered to recognize and cleave specific dsRNA substrates by replacement of the PAZ domain with a different RNA-recognition domain. Here, a Dicer mutant in which the PAZ domain was replaced by the U1A RNA-binding domain recognized and cleaved dsRNA substrates containing a U1A recognition sequence. The resulting dsRNA products were consistently 19–21 nt in length, indicating the accuracy of the structural model for dsRNA recognition and cleavage. This result also shows that the PAZ domain is not uniquely capable of providing the measuring function that characterizes Dicer endonucleolytic activity. By replacing the PAZ domain with different types of RNA-binding modules, Dicer could be programmed so that only a particular RNA substrate would be recognized and processed *in vivo*.

The discovery that Dicer can be reprogrammed provides a possible explanation for the apparent lack of identifiable PAZ domains in several Dicer proteins, including *T. thermophila* Dcr1 and Dcr2 and *S. pombe* Dcr1. Although the PAZ domain sequence may simply be too divergent to be recognized by common algorithms, it is also conceivable that a different domain has replaced PAZ in these proteins, or that the PAZ domain has mutated substantially to enable recognition of different families of RNA substrates. It is also possible that domains providing an RNA end- or sequence-recognition

function interact with these Dicers *in trans*, a mechanism analogous to the activity of nuclear Drosha RNase III-family enzymes with DGCR8-Pasha RNA-binding proteins²⁷.

METHODS

Preparation of mutant Dicer proteins. The QuikChange method for site-directed mutagenesis (Stratagene) was used to generate DNA plasmids encoding surface-mutated and ΔPAZ Dicer proteins. The template in each reaction was a wild-type copy of the *G. intestinalis* Dicer gene (NCBI Protein EAA41574) inserted between the SfoI and SalI sites of pFastBac HTA (Invitrogen). For the surface mutations, basic amino acid residues were changed to alanine (see Table 1). For ΔPAZ, codons 138–250 were replaced with the sequence 5'-GGATCCTCCGGTGGGA-3', which encodes a short Gly-Ser-Ser-Gly-Gly linker sequence between Ala137 and Pro251 of the wild-type sequence. To generate the U1A Dicer, the linker codons of the ΔPAZ plasmid were first changed to the sequence 5'-GGAGCTAGCGGTGGGA-3' to introduce a unique NdeI site. DNA encoding the U1A RNA binding domain was then cloned into the NdeI site as an XbaI fragment. The U1A insert was generated by PCR using the DNA oligonucleotides 5'-TGCAGGATCTAGAG GAGGGGGTGGCGGAGGGGAGACCCGCCCTAACCACTATTTAT-3' and 5'-AAATTCATCTGTCTAGACCCTCCGCCACTTTCATCTTGCAATGATA TCTGAGTCG-3' with the human U1A RNA-binding domain (NCBI Protein NP_004587) as a template. The resulting plasmid encodes a protein with the following amino acid sequence inserted between Ala137 and Pro215 of wild-type *G. intestinalis* Dicer: Gly-Ala-Arg-Gly₆, followed by amino acid residues Glu5 through Lys98 of U1A, followed by Gly₄-Ser-Ser.

All mutant DNA constructs were confirmed by DNA sequencing and used in the Bac-to-Bac system (Invitrogen) for the production of recombinant baculovirus. Mutant Dicer proteins were expressed in Sf9 cells and purified by a combination of nickel-nitrilotriacetic acid (Qiagen) and size exclusion chromatography as described previously for the wild-type enzyme¹³.

Preparation of RNA substrates. The 153-bp dsRNA substrate was prepared using *in vitro* transcription reactions spiked with [α -³²P]ATP as described¹³. Short RNA oligonucleotides were purchased from Integrated DNA Technologies, purified by denaturing PAGE and ³²P-labeled at the 5' end using T4 polynucleotide kinase (New England Biolabs).

The RNA hairpin used in **Supplementary Figure 1** was transcribed *in vitro* from a precursor let-7-like RNA³ with the following sequence: 5'-GCC CUUUGGGGUGAGGUAGUAGGUAGUAGUUUGGGGUCUCUGCCUGC UAUGGGGAUACUAUACAUCUACUGUCUUUCUGAAGUGGC-3'. Transcribed RNA was treated with calf intestinal phosphatase (New England Biolabs), ³²P-labeled at the 5' end with T4 polynucleotide kinase and purified by denaturing PAGE. The hairpin was closed by ligation using T4 RNA ligase, and the ligated product was gel purified. Ligation was confirmed by limited hydrolysis under alkaline conditions, which nicks the RNA and shifts its electrophoretic mobility to that of the unligated hairpin (**Supplementary Fig. 1**).

Overlap PCR was used to generate DNA templates for the U1A hairpin RNAs. PCR products were cloned into pUC19 and then transcribed *in vitro* with T7 RNA polymerase. U1A hairpin RNAs were designed to have the following nucleotide sequences: U1A-1 (37 bp), 5'-GGAGGUAGUAGGUGC UAGGCCUACCGACACACAUC^{CAU}UGCACUCCGGGAUGUGUGCGGU AGGCCUAGCACCUCUACCUCCUG-3'; and U1A-2 (33 bp), 5'-GGAGUAG GUGCUAGGCCUACCGACACACAUC^{CAU}UGCACUCCGGGAUGUGUGUGU CGGUAGGCCUAGCACCUCUACCUCCUG-3' (U1A recognition elements underlined). Transcription products were treated with calf intestinal phosphatase, ³²P-labeled at the 5' end and PAGE-purified before use.

Immediately before each assay, RNA substrates were incubated at 65 °C in the standard dicing reaction buffer¹³ for 10 min. Open duplex substrates were cooled slowly to room temperature over the course of 30 min. Hairpin substrates were snap-cooled by plunging into an ice bath and then brought back to room temperature. Open duplex substrates were typically annealed at a concentration ten times the working concentration in the subsequent dicing assay. Hairpin substrates were annealed at working assay concentrations (typically 50–75 nM). Analytical native PAGE was used to determine the exact

ratio of the two RNAs used to form duplexes and to validate that the hairpin RNAs had correctly folded.

Dicer activity assays. All dicing assays were carried out at 37 °C and contained 50 mM NaCl, 3 mM MgCl₂, 100 mM HEPES (pH 7.5), ³²P-labeled RNA and purified *G. intestinalis* Dicer in a final volume of 20 μl. For the fixed-time assays, ~75 nM labeled RNA was incubated with 25 nM Dicer unless stated otherwise.

Time-course assays, in which substrate concentration was varied (**Fig. 3**), contained 50 nM unlabeled dsRNA spiked with a trace quantity of substrate labeled with ³²P at the 5' end. Unlabeled RNA oligonucleotides also contained 5'-phosphate groups. For these assays, 5–125 nM Dicer was used, with Dicer concentration always at least ten-fold below the substrate concentration. For each substrate concentration, data were collected at five time points, at 2-min intervals. Enzyme concentrations were adjusted in each reaction so that the rate of product formation was linear over the measured time course and reaction rates were directly proportional to the amount of enzyme used. In all assays, at least 90% of the input substrate RNA remained intact at the final time point.

All reactions were stopped by the addition of an equal volume of formamide, and RNA products were resolved by denaturing 16% PAGE. RNA products were visualized by phosphorimaging and quantified using ImageQuant (GE Healthcare Life Sciences).

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

ACKNOWLEDGMENTS

We thank the members of the Doudna lab for comments and suggestions. This work was supported by the Howard Hughes Medical Institute and US National Institutes of Health grant 5R01GM073794-02.

AUTHOR CONTRIBUTIONS

I.J.M. designed and performed experiments and wrote the manuscript, K.Z. performed experiments, and J.A.D. discussed results and wrote the manuscript.

Published online at <http://www.nature.com/nsmb>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- Hammond, S.M. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Lett.* **579**, 5822–5829 (2005).
- Elbashir, S.M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
- Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838 (2001).
- Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P.D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380 (2004).
- Maniaki, E. & Mourelatos, Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* **19**, 2979–2990 (2005).
- Gregory, R.I., Chendrimada, T.P., Cooch, N. & Shiekhattar, R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**, 631–640 (2005).
- Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W. & Sontheimer, E.J.A. QJ:Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* **117**, 83–94 (2004).
- Hammond, S.M., Caudy, A.A. & Hannon, G.J. Post-transcriptional gene silencing by double-stranded RNA. *Nat. Rev. Genet.* **2**, 110–119 (2001).
- Olsen, P.H. & Ambros, V. The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–680 (1999).
- Volpe, T.A. *et al.* Regulation of heterochromatin silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
- Mochizuki, K. & Gorovsky, M.A.A. Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev.* **19**, 77–89 (2005).
- Macrae, I.J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006).
- Błaszczak, J. *et al.* Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure* **9**, 1225–1236 (2001).
- Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004).
- Gan, J. *et al.* Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* **124**, 355–366 (2006).

17. 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).
18. Rose, S.D. *et al.* Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res.* **33**, 4140–4156 (2005).
19. Segel, I.H. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems* 957 (John Wiley & Sons, New York, 1975).
20. Sun, W., Jun, E. & Nicholson, A.W. Intrinsic double-stranded-RNA processing activity of *Escherichia coli* ribonuclease III lacking the dsRNA-binding domain. *Biochemistry* **40**, 14976–14984 (2001).
21. Pertzev, A.V. & Nicholson, A.W. Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* ribonuclease III. *Nucleic Acids Res.* **34**, 3708–3721 (2006).
22. Macrae, I.J., Li, F., Zhou, K., Cande, W.Z. & Doudna, J.A. Structure of Dicer and mechanistic implications for RNAi. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 73–80 (2006).
23. Sasaki, T. & Shimizu, N. Evolutionary conservation of a unique amino acid sequence in human DICER protein essential for binding to Argonaute family proteins. *Gene* **396**, 312–320 (2007).
24. Hall, K.B. & Stump, W.T. Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res.* **20**, 4283–4290 (1992).
25. 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).
26. Ullu, E., Lujan, H.D. & Tschudi, C. Small sense and antisense RNAs derived from a telomeric retroposon family in *Giardia intestinalis*. *Eukaryot. Cell* **4**, 1155–1157 (2005).
27. Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).
28. DeLano, W.L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, California, USA, 2002).