Structural Basis for Double-Stranded RNA Processing by Dicer

Ian J. MacRae,[1,3] Kaihong Zhou,[1,3] Fei Li,[1] Adrian Repic,[1] Angela N. Brooks,[1] W. Zacheus Cande,[1] Paul D. Adams,[4] Jennifer A. Doudna[1,2,3,4]*

The specialized ribonuclease Dicer initiates RNA interference by cleaving double-stranded RNA (dsRNA) substrates into small fragments about 25 nucleotides in length. In the crystal structure of an intact Dicer enzyme, the PAZ domain, a module that binds the end of dsRNA, is separated from the two catalytic ribonuclease III (RNase III) domains by a flat, positively charged surface. The 65 angstrom distance between the PAZ and RNase III domains matches the length spanned by 25 base pairs of RNA. Thus, Dicer itself is a molecular ruler that recognizes dsRNA and cleaves a specified distance from the helical end.

RNA interference (RNAi) is an ancient gene-silencing process that plays a fundamental role in diverse eukaryotic functions including viral defense (*1*), chromatin remodeling (*2*), genome rearrangement (*3*), developmental timing (*4*), brain morphogenesis (*5*), and stem cell maintenance (*6*). All RNAi pathways require the multidomain ribonuclease Dicer (*7*). Dicer first processes input dsRNA into small fragments called short interfering RNAs (siRNAs) (*8*), or microRNAs (miRNA) (*9*), which are the hallmark of RNAi. Dicer then helps load its small RNA products into large multiprotein complexes termed RNA-induced silencing complexes (RISC) (*10*). RISC and RISC-like complexes use the small RNAs as guides for the sequence-specific silencing of cognate genes through mRNA degradation (*11*), translational inhibition (*12*), and heterochromatin formation (*13*).

Dicer products are typically 21 to 25 nucleotides long, which is the ideal size for a gene silencing guide, because it is long enough to provide the sequence complexity required to uniquely specify a single gene in a eukaryotic genome. Several models have been proposed for how Dicer generates RNA fragments of this specific size (*14*–*16*), but structural information is lacking. In an effort to deepen our understanding of the initiation step of RNAi, we determined the crystal structure of an intact and fully active Dicer enzyme.

Conservation of a highly active Dicer in *Giardia intestinalis*. We identified an open reading frame in *Giardia intestinalis* that encodes the PAZ and tandem RNase III domains characteristic of Dicer (*7*), but lacks the N-terminal DExD/H helicase, C-terminal double-stranded RNA binding domain (dsRBD), and extended interdomain regions associated with Dicer in higher eukaryotes (Fig. 1A). A recombinant form of this protein possesses robust dicing activity in vitro (Fig. 1B). The RNA fragments produced by *Giardia* Dicer are 25 to 27 nucleotides long, which is similar to a class of small RNAs associated with RNAi-mediated DNA elimination in *Tetrahymena* (*17*) and RNA-directed DNA methylation in plants (*18*). dsRNA cleavage by *Giardia* Dicer is magnesium-dependent, although several other divalent cations including Mn^{2+}, Ni^{2+}, and Co^{2+} also support catalytic activity (*19*). The presence of discrete dicing intermediates separated by intervals of ~25 nucleotides indicates that *Giardia* Dicer processes dsRNA from the helical end in a fashion similar to human Dicer (*20*). However, in contrast to human Dicer, *Giardia* Dicer has a low affinity for its small RNA product (~1 μM) (*19*) and displays multiple turnover kinetics (*20*).

Structural overview. We determined the crystal structure of the full-length *Giardia* Dicer at 3.3 Å resolution (table S1). The structure reveals an elongated molecule that, when viewed from the front, takes on a shape resembling a hatchet; the RNase III domains form the blade and the PAZ domain makes up the base of the handle (Fig. 2A). The PAZ domain is directly connected to the RNase IIIa domain by a long α helix that runs through the handle of the molecule. This "connector" helix is encircled by the N-terminal residues of the protein, which form a platform domain composed of an antiparallel β sheet and three α helices. A large helical domain bridges the two RNase III domains and forms the back end of the blade. Viewing Dicer from the side reveals a contiguous flat surface that extends along one face of the molecule.

Two–metal-ion mechanism of dsRNA cleavage. The two RNase III domains of Dicer sit adjacent to each other in the blade region and form an internal heterodimer that is similar to the homodimeric structure of bacterial RNase III (fig. S1). Although previous bacterial RNase III crystal structures revealed a single catalytic metal ion in each RNase III domain (*21*), subsequent studies implicated two metal ions in the hydrolysis of each strand of the dsRNA (*22*).

During our biochemical characterization of *Giardia* Dicer, we noticed that the enzyme is potently inhibited by trivalent lanthanide cations such as Er^{3+} (*19*). Lanthanides often bind more tightly to cation binding sites than divalent cations do, a property previously used to identify transient Mn^{2+} binding sites in proteins (*23*). Inspection of the anomalous difference electron density map from a crystal derivatized with $ErCl_3$ revealed a pair of Er^{3+} cations in the active site of each RNase III domain of *Giardia* Dicer (Fig. 2B). The prominent Er^{3+} metal (M1) in each domain resides between four strictly conserved acidic residues, which make up the previously identified Mn^{2+} binding site of bacterial RNase III (*21*). The second Er^{3+} binding site (M2) lies adjacent to the first, outside of the acidic residue cluster. The distances between the two Er^{3+} metals in the RNase IIIa and IIIb domains are ~4.2 Å and ~5.5 Å, respectively. These distances are similar to those previously observed in the active site of RNase H (4.1 Å) (*24*), avian sarcoma virus (ASV) integrase (3.6 Å) (*25*), the restriction enzyme *EcoR*V (4.2 Å) (*26*), and the group I intron (3.9 Å) (*27*), all of which are thought to use a two–metal-ion mechanism of catalysis. The 17.5 Å distance between the

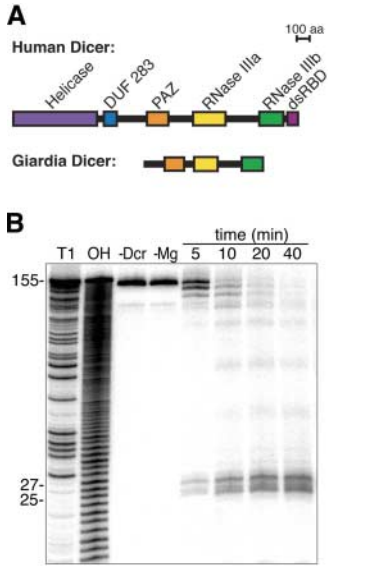

Fig. 1. *Giardia* encodes an active Dicer enzyme. (**A**) Schematic representation of the primary sequence of human and *Giardia* Dicers. (**B**) Time course of in vitro *Giardia* Dicer dsRNA cleavage assay. RNA product sizes were determined by comparison with RNase T1 and alkaline hydrolysis (OH) sequencing ladders (lanes 1 and 2). Dicing requires the protein (Dcr) and Mg^{2+} (lanes 3 and 4).

[1]Department of Molecular and Cell Biology, [2]Department of Chemistry, [3]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA. [4]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed: E-mail: doudna@berkeley.edu

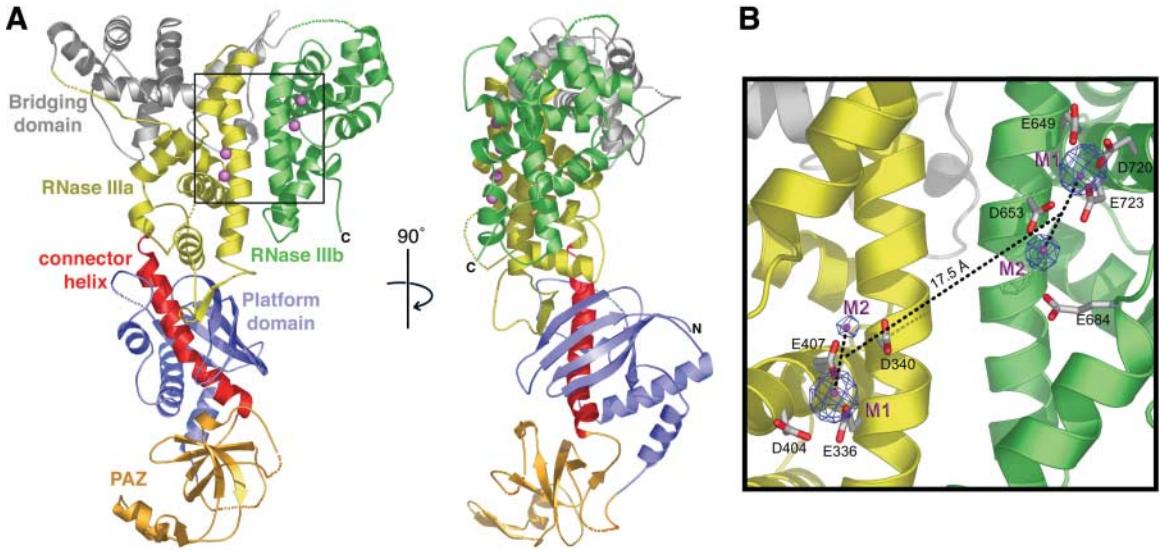

Fig. 2. Crystal structure of *Giardia* Dicer. (**A**) Front and side view ribbon representations of Dicer showing the N-terminal platform domain (blue), the PAZ domain (orange), the connector helix (red), the RNase IIIa domain (yellow), the RNase IIIb domain (green) and the RNase-bridging domain (gray). Disordered loops are drawn as dotted lines. (**B**) Close-up view of the Dicer catalytic sites; conserved acidic residues (sticks); erbium metal ions (purple); and erbium anomalous difference electron density map, contoured at 20σ (blue wire mesh). Dashed lines indicate distances described in the text.

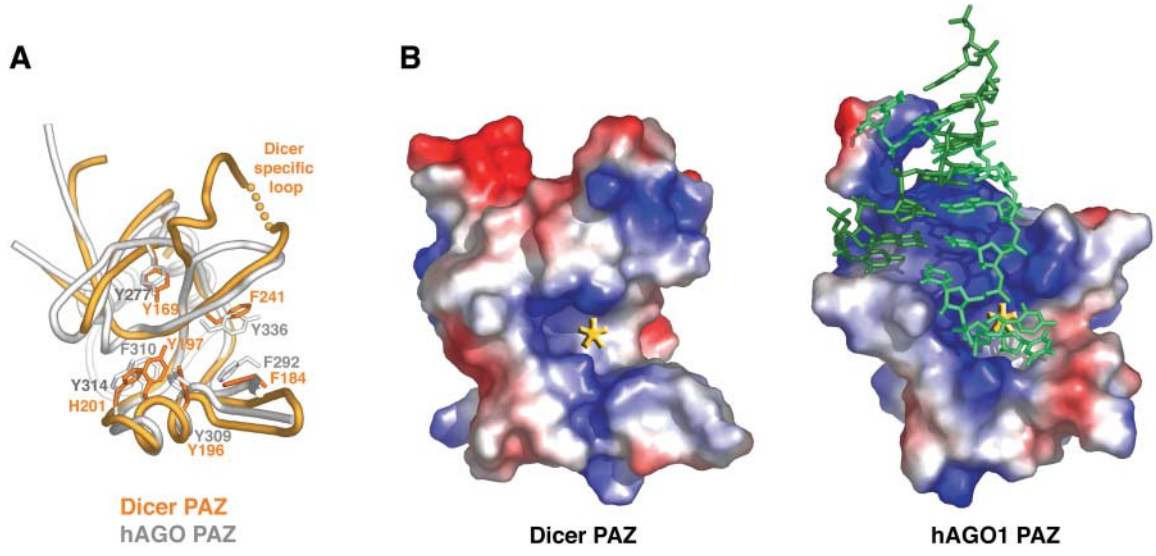

Fig. 3. Structural features of the Dicer PAZ domain. (**A**) Superposition of the C_α atoms of PAZ domains from *Giardia* Dicer (orange) and human Argonaute1 (white). Amino acids forming the 3′ overhang-binding pocket are shown as sticks. (**B**) Electrostatic surface representation of the PAZ domains of *Giardia* Dicer and Argonaute1 (hAGO1). Asterisks denote 3′ overhang-binding pockets. The RNA in Argonaute1 PAZ structure is drawn as green sticks.

metal-ion pairs closely matches the width of the dsRNA major groove. We also observed Mn^{2+} in all M1 and some M2 sites in crystals grown in high concentrations of $MnCl_2$. Therefore, we propose that the Er^{3+} metals seen in *Giardia* Dicer denote true catalytic metal-ion binding sites and that *Giardia* Dicer uses a two–metal-ion mechanism of catalysis. Given the high level of sequence conservation throughout the RNase III family, it is likely that all RNase III enzymes, including bacterial RNase III and Drosha, contain similar catalytic metal-ion binding sites.

Structural features of the Dicer PAZ domain. The PAZ domain is an RNA binding module found in Dicers and in the Argonaute family of proteins that are core components of RISC and other siRNA- and miRNA-containing complexes. Previous studies of PAZ domains from several Argonaute proteins revealed a degenerate oligonucleotide/oligosaccharide-binding (OB) fold that specifically recognizes dsRNA ends containing a 3′ two-base overhang (*28*–*31*). Superposition of the PAZ domains of *Giardia* Dicer and human Argonaute1 reveals that the two domains share the same overall fold and 3′ two-nucleotide RNA binding pocket (Fig. 3A).

The Dicer PAZ domain contains a large extended loop that is conserved among Dicer sequences and absent in Argonaute (fig. S1). The Dicer-specific loop dramatically changes the electrostatic potential and molecular surface surrounding the 3′ overhang-binding pocket relative to the Argonaute PAZ domain (Fig. 3B). The presence of many basic amino acid residues in the extra loop could substantially affect the way the RNA is recognized and perhaps handed off to other complexes by each family of proteins.

A model for siRNA formation. The structure of *Giardia* Dicer immediately suggests how Dicer enzymes specify siRNA length. Measuring from the active site of the RNase IIIa domain to the 3′ overhang-binding pocket in the PAZ domain gives a distance of ~65 Å (Fig. 4), which matches the length of 25 dsRNA base pairs. To produce a likely model of a Dicer-dsRNA complex, the positions of the metal-ion pairs bound in each RNase III domain were used to anchor the two scissile phosphates of an ideal A-form dsRNA helix into the RNase III active sites. This placement positions the twofold symmetry element of the dsRNA coincident with the pseudo twofold symmetry axis relating the two RNase III domains, which is analogous to how restriction enzymes typically bind dsDNA substrates (*32*). Bacterial RNase III has been proposed to bind dsRNA in a similar fashion (*16*, *33*). Outside of the RNase III region, the modeled dsRNA extends along a flat surface formed by the platform domain. This surface contains a large positively charged region that could interact directly with the nega-

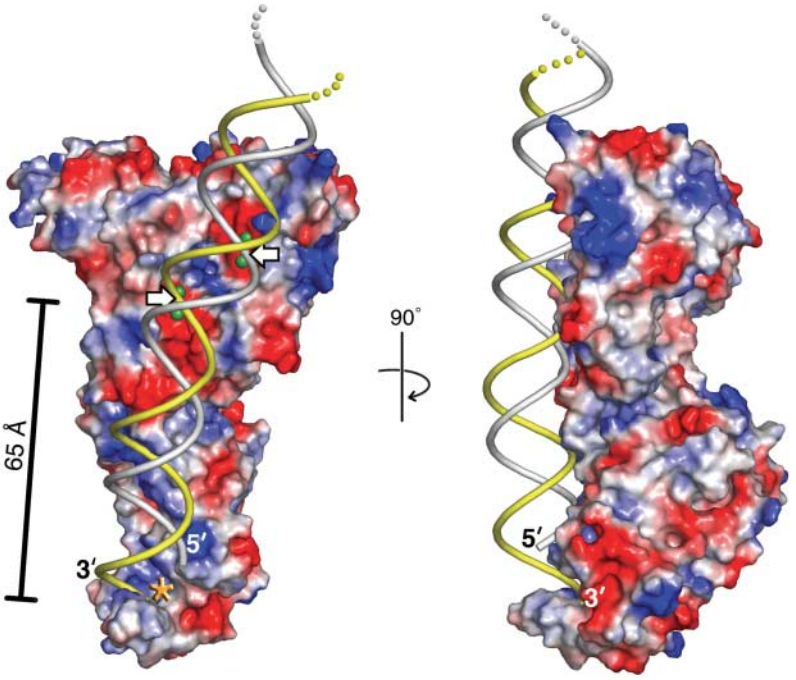

Fig. 4. A model for dsRNA processing by Dicer. Front and side views of a surface representation of *Giardia* Dicer with modeled dsRNA. Red and blue represent acidic and basic protein surface charge, respectively. Electrostatic surface potentials do not include contributions from bound metal ions. Putative catalytic metal ions are shown as green spheres. White arrows point to scissile phosphates. Asterisk denotes PAZ domain 3′ overhang-binding pocket.

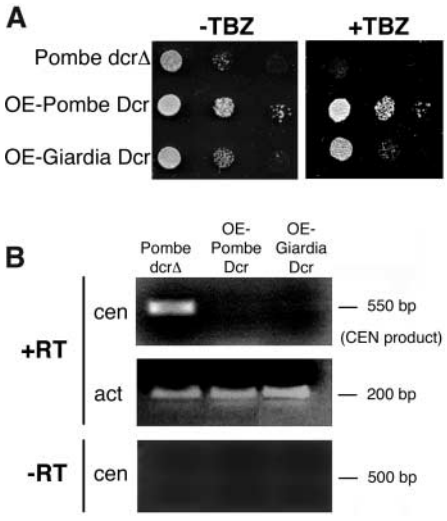

Fig. 5. *Giardia* Dicer supports RNAi in vivo. (**A**) Overexpression (OE) of *Giardia* Dicer rescues the TBZ sensitivity of the *S. pombe Dicer delete (dcr*Δ). Growth was assayed by spotting 10-fold serial dilutions of cultures indicated. (**B**) Overexpression of *Giardia* Dicer restores transcriptional silencing at centromeres (cen). Transcript levels were determined by semiquantitative reverse-transcriptase polymerase chain reaction. Actin (act) served as an internal control. bp, base pair.

tively charged phosphodiester backbone of the modeled dsRNA helix. The 3′ end of the RNA duplex falls directly into the 3′ overhang-binding pocket of the PAZ domain, and the 5′ end lies adjacent to the Dicer-specific PAZ domain loop. There are exactly 25 nucleotides between the 3′ end of the helix bound to the PAZ domain and the scissile phosphate in the RNase IIIa domain.

Thus, Dicer is a molecular ruler that measures and cleaves ~25 nucleotides from the end of a dsRNA. The length of the small RNAs produced by Dicer is set by the distance between the PAZ and RNase III domains, which is largely a function of the length of the connector helix. This model of dsRNA processing is consistent with the proposed architecture of human Dicer based on biochemical studies in which the RNase IIIa and IIIb domains were shown to produce the siRNA 5′ and 3′ ends, respectively (*16*). Furthermore, closing the ends of a dsRNA substrate by hybridization or ligation greatly diminished dicing activity (*20*, *34*), which may explain why circular viral dsRNA is resistant to RNAi (*35*).

Giardia **Dicer can support RNAi in fission yeast.** Given that *Giardia* Dicer lacks some of the domains commonly associated with Dicer enzymes, most notably the N-terminal helicase, we wondered if the structure represents an intact Dicer or merely the catalytic subunit of a larger complex required for complete Dicer function in vivo. To address this question, we introduced the *Giardia* Dicer gene into a strain of the fission yeast *Schizosaccharomyces pombe* that contains a deletion of its endogenous Dicer (*dcr*Δ).

Like most Dicer proteins, the *S. pombe* Dicer contains an N-terminal helicase domain and a C-terminal dsRBD. The *S. pombe dcr*Δ strain is defective in RNAi and is hypersensitive to the microtubule-destabilizing drug thiobendazole (TBZ) because of chromosome missegregation (*36*). Plasmid expression of *S. pombe dcr1*$^{+}$ fully rescued TBZ sensitivity of the *dcr*Δ cells. A partial functional rescue of TBZ sensitivity was also achieved by episomal expression of *Giardia* Dicer, indicating that *Giardia* Dicer can suppress the chromosome segregation defect (Fig. 5A). Furthermore, *Giardia* Dicer restores silencing of centromeric regions that are aberrantly transcribed in the *dcr*Δ mutant (Fig. 5B). These results demonstrate that *Giardia* Dicer is sufficient to function as an intact Dicer in vivo.

A conserved architecture in Dicer enzymes. Considering the structural role played by the connector helix that links the PAZ and RNase III domains (Fig. 2), we wondered whether larger Dicer proteins found in higher eukaryotes contain an analogous helix. Sequence alignment of the region directly following the PAZ domain of several evolutionarily diverse Dicer enzymes reveals a conserved pattern of hydrophobic and hydrophilic amino acids that is predicted to form a long α helix by secondary structural analysis (fig. S2). All Dicers contain a conserved proline about 11 amino acid residues from the predicted N terminus of the helix. In the crystal structure of *Giardia* Dicer, this proline induces a distinct kink that aids in directing the helix toward the RNase IIIa domain.

Most Dicer proteins contain a conserved region of ~100 amino acids termed ''domain of unknown function 283'' (DUF283), which lies between the helicase and PAZ domains in the primary sequence. Low but consistent sequence homology between the N-terminal domain of *Giardia* Dicer and DUF283 (fig. S3) suggests that in the Dicers of higher eukaryotes, DUF283 forms a platform structure similar to that of *Giardia* Dicer.

The conserved Dicer architecture, together with the demonstration that *Giardia* Dicer can substitute for *S. pombe* Dicer in vivo, argues that the mechanism of Dicer-catalyzed dsRNA processing is conserved. Moreover, these results indicate that all Dicers evolved from a common ancestral enzyme. Because *Giardia* is one of the most anciently diverged members of the eukaryotic kingdom, we may consider that the earliest eukaryotic organisms had a similar Dicer enzyme and therefore were capable of RNAi-like processes. It will be of evolutionary interest to determine the cellular function of Dicer and RNAi in *Giardia*.

The structure of *Giardia* Dicer also provides new insight into eukaryotic RNase III enzymes in general. This family of enzymes performs a range of specific cellular functions involving the cleavage of dsRNA [reviewed in (*37*)]. The structure of Dicer illustrates

how the presence of RNA binding modules, like the PAZ and platform domains, can impart a specific function to the otherwise nonspecific double-stranded RNase activity of the RNase III dimer (*38*). This is likely to be the structural paradigm for all eukaryotic RNase III enzymes that have specific activities and cellular functions.

References and Notes

1. D. Baulcombe, *Trends Microbiol.* **10**, 306 (2002).
2. T. A. Volpe *et al.*, *Science* **297**, 1833 (2002).
3. K. Mochizuki, N. A. Fine, T. Fujisawa, M. A. Gorovsky, *Cell* **110**, 689 (2002).
4. A. Grishok *et al.*, *Cell* **106**, 23 (2001).
5. A. J. Giraldez *et al.*, *Science* **308**, 833 (2005).
6. S. D. Hatfield *et al.*, *Nature* **435**, 974 (2005).
7. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).
8. S. M. Elbashir, W. Lendeckel, T. Tuschl, *Genes Dev.* **15**, 188 (2001).
9. G. Hutvagner *et al.*, *Science* **293**, 834 (2001).
10. Q. Liu *et al.*, *Science* **301**, 1921 (2003).
11. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, *Nature* **404**, 293 (2000).
12. R. S. Pillai *et al.*, *Science* **309**, 1573 (2005).
13. A. Verdel *et al.*, *Science* **303**, 672 (2004).
14. M. A. Carmell, G. J. Hannon, *Nat. Struct. Mol. Biol.* **11**, 214 (2004).
15. P. D. Zamore, *Mol. Cell* **8**, 1158 (2001).
16. H. Zhang, F. A. Kolb, L. Jaskiewicz, E. Westhof, W. Filipowicz, *Cell* **118**, 57 (2004).
17. C. D. Malone, A. M. Anderson, J. A. Motl, C. H. Rexer, D. L. Chalker, *Mol. Cell. Biol.* **25**, 9151 (2005).
18. Z. Xie *et al.*, *PLoS Biol.* **2**, E104 (2004).
19. I. J. MacRae, K. Zhou, J. A. Doudna, data not shown.
20. H. Zhang, F. A. Kolb, V. Brondani, E. Billy, W. Filipowicz, *EMBO J.* **21**, 5875 (2002).
21. J. Blaszczyk *et al.*, *Structure* (*Camb*) **9**, 1225 (2001).
22. W. Sun, A. Pertzev, A. W. Nicholson, *Nucleic Acids Res.* **33**, 807 (2005).
23. M. Sundaramoorthy, H. L. Youngs, M. H. Gold, T. L. Poulos, *Biochemistry* **44**, 6463 (2005).
24. M. Nowotny, S. A. Gaidamakov, R. J. Crouch, W. Yang, *Cell* **121**, 1005 (2005).
25. G. Bujacz *et al.*, *J. Biol. Chem.* **272**, 18161 (1997).
26. I. B. Vipond, G. S. Baldwin, S. E. Halford, *Biochemistry* **34**, 697 (1995).
27. M. R. Stahley, S. A. Strobel, *Science* **309**, 1587 (2005).
28. J. B. Ma, K. Ye, D. J. Patel, *Nature* **429**, 318 (2004).
29. J. J. Song *et al.*, *Nat. Struct. Biol.* **10**, 1026 (2003).
30. A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nature* **426**, 465 (2003).
31. K. S. Yan *et al.*, *Nature* **426**, 468 (2003).
32. A. K. Aggarwal, *Curr. Opin. Struct. Biol.* **5**, 11 (1995).
33. D. L. Akey, J. M. Berger, *Protein Sci.* **14**, 2744 (2005).
34. A. Repic, J. A. Doudna, unpublished data.
35. J. Chang, P. Provost, J. M. Taylor, *J. Virol.* **77**, 11910 (2003).
36. I. M. Hall, K. Noma, S. I. Grewal, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 193 (2003).
37. D. Drider, C. Condon, *J. Mol. Microbiol. Biotechnol.* **8**, 195 (2004).
38. W. Sun, E. Jun, A. W. Nicholson, *Biochemistry* **40**, 14976 (2001).

39. We thank members of the Doudna and Berger labs for helpful discussions, A. Fischer for work with tissue culture, and D. King for mass spectrometry analysis. We are grateful to C. Ralston and J. Dickert for technical support on beam lines 8.2.1 and 8.2.2 at the Advanced Light Source at the Lawrence Berkeley National Lab. I.J.M. is a Howard Hughes Medical Institute fellow of the Life Sciences Research Foundation. This work was supported in part by a grant from NIH (to J.A.D.). Dicer coordinates and structure factors have been deposited in the Protein Data Bank with accession code 2FFL.

Supporting Online Material

www.sciencemag.org/cgi/content/full/311/5758/195/DC1
Materials and Methods
Figs. S1 to S5
Table S1
References

21 October 2005; accepted 7 December 2005
10.1126/science.1121638

REPORTS

The Nature of the 660-Kilometer Discontinuity in Earth's Mantle from Global Seismic Observations of *PP* Precursors

Arwen Deuss,[1]* Simon A. T. Redfern,[1] Kit Chambers,[2] John H. Woodhouse[2]

The 660-kilometer discontinuity, which separates Earth's upper and lower mantle, has been detected routinely on a global scale in underside reflections of precursors to *SS* shear waves. Here, we report observations of this discontinuity in many different regions, using precursors to compressional *PP* waves. The apparent absence of such precursors in previous studies had posed major problems for models of mantle composition. We find a complicated structure, showing single and double reflections ranging in depth from 640 to 720 kilometers, that requires the existence of multiple phase transitions at the base of the transition zone. The results are consistent with a pyrolite mantle composition.

The characteristics of the 660-km discontinuity determine the convective style of the mantle, distinguishing potentially between whole-mantle and layered-mantle convection (*1*, *2*). The key question is whether this discontinuity is caused by a phase transition (*3*, *4*) or by a change in chemical composition (*5*). The latter would favor two-layer mantle convection, whereas the former would allow whole-mantle convection. Answers have been sought in seismology by studying the detailed seismic characteristics of this discontinuity and comparing them with predictions from mineral physical models of the mantle. Most seismic studies (*6*–*8*) find evidence in favor of the post-spinel phase transition in olivine from ringwoodite to perovskite and magnesiowüstite (*9*).

Precursors to the *SS* and *PP* waves have been used extensively to study the global seismic characteristics of the transition-zone discontinuities, in particular the 660-km discontinuity. These precursors arrive before the major *PP* (or *SS*) wave because they are not reflected from Earth's surface but from a discontinuity below the surface bounce point of the *PP* (or *SS*) wave (fig. S1). Precursors are named *PdP* (or *SdS*), where *d* is the depth of the discontinuity, thus *P*660*P* (or *S*660*S*) are the underside reflections of the 660-km discontinuity. These data are unusual in probing the mantle discontinuities with good global coverage, beneath both continents and oceans, and they can be used to draw inferences in many different regions.

Studies of the 660-km discontinuity have been hampered because, although it is easily observed in long-period *SS* precursors on a global scale, it has not been seen before using long-period *PP* precursors (*10*). This was interpreted to mean that the bulk modulus is continuous across the boundary (*11*), placing major constraints on the mineral physics of the post-spinel phase transition and on average mantle composition. In particular, these observations cannot be reconciled with a pyrolite mantle composition and are not consistent with the widely used Preliminary Reference Earth Model (PREM) (*12*); both models predict large reflection amplitudes in both *PP* and *SS* for the 660-km discontinuity (Fig. 1). It also implied a rather small density jump at the base of the transition zone, which would favor whole-mantle convection. Matters are further complicated by short-period observations of P′P′ precursors (*13*), which show that the 660-km discontinuity is sharp and prominent in a few isolated places in the Southern Hemisphere (*14*). Although P′P′ has a different ray path, it has been difficult to reconcile sharp short-period reflections with the absence of long-period reflections for the same discontinuity.

[1]Department of Earth Sciences, University of Cambridge, Cambridge CB3 0EZ, UK. [2]Department of Earth Sciences, University of Oxford, Oxford OX1 3PR, UK.

*To whom correspondence should be addressed. E-mail: deuss@esc.cam.ac.uk.