

2B, lower panel). An increase in the expression of arginine fermentation genes (*arcA*, *arcI*, and *arcC*) (Fig. 2B) in stationary phase could be a mechanism to cope with acidification (23). We found formal evidence for a total of 447 transcriptional units (336 monocistronic and 111 polycistronic), implying a high rate of alternative transcripts (42%) in this bacterium under the conditions studied, similar to that in eukaryotes (40%, although still under debate) (24) and archaea (40% in *H. salinarum*) (25). We found that genes that are split into different suboperons tend to belong to different functional categories (9). Thus, although genome reduction leads to longer operons accommodating genes with different functions (26), the latter can still retain internal transcription and termination sites under certain conditions.

The high frequency of alternative transcripts of *M. pneumoniae* genes hints at a situation similar to that in eukaryotes, where many factors contribute to the regulation of gene expression. To further support this hypothesis, we used gene expression clustering under the 62 distinct conditions (table S7) to identify groups of coexpressed genes and their possible common regulatory motifs. Using a correlation cutoff of 0.65, we identified 94 coexpression groups (table S6 and fig. S11), encompassing 416 genes. Thirty of the clusters contained genes from more than two operons. Of these, 14 share a specific sequence motif in their upstream region and another 8 have a specific combination of motifs (fig. S12), which might drive the coexpression (for example, 4 of the 14 motifs are found at splitting sites inside operons). The rest of the genes did not group together, implying complex and multiple levels of regulation orchestrated by the various environmental conditions. This is exemplified by the five

heat shock-induced genes containing a regulatory CIRCE (controlling inverted repeat of chaperone expression) element (27) (Fig. 2C). Not all of them clustered together, indicating at least one other regulatory element. Similarly, overexpression of a transcription factor (Fur, ferric uptake regulator) reveals a common motif in all genes significantly changing expression, although they belong to different coexpression clusters (fig. S13 and table S6).

Our work revealed an unanticipated complexity in the transcriptome of a genome-reduced bacterium. This complexity cannot be explained by the presence of eight predicted transcription factors (26). Furthermore, the fact that the proteome organization is not explainable by the genome organization (28) indicates the existence of other regulatory processes. The surprisingly frequent expression heterogeneity within operons, the change of operon structures leading to alternative transcripts in response to environmental perturbations, and the frequency of antisense RNA, which might explain some of these expression changes, suggest that transcriptional regulation in bacteria resemble that of eukaryotes more than previously thought.

References and Notes

1. B. Tjaden *et al.*, *Nucleic Acids Res.* **30**, 3732 (2002).
2. D. W. Selinger *et al.*, *Nat. Biotechnol.* **18**, 1262 (2000).
3. N. B. Reppas, J. T. Wade, G. M. Church, K. Struhl, *Mol. Cell* **24**, 747 (2006).
4. C. M. Nelson *et al.*, *BMC Genomics* **9**, 364 (2008).
5. T. Akama *et al.*, *J. Bacteriol.* **191**, 3321 (2009).
6. P. T. McGrath *et al.*, *Nat. Biotechnol.* **25**, 584 (2007).
7. A. Toledo-Arana *et al.*, *Nature* **459**, 950 (2009).
8. J. Vogel, E. G. Wagner, *Curr. Opin. Microbiol.* **10**, 262 (2007).
9. Materials and methods are available as supporting material on Science Online.

10. W. Huber, J. Toedling, L. M. Steinmetz, *Bioinformatics* **22**, 1963 (2006).
11. Z. Xu *et al.*, *Nature* **457**, 1033 (2009).
12. X. J. Wang, T. Gaasterland, N. H. Chua, *Genome Biol.* **6**, R30 (2005).
13. S. R. Henz *et al.*, *Plant Physiol.* **144**, 1247 (2007).
14. X. Ge, Q. Wu, Y. C. Jung, J. Chen, S. M. Wang, *Bioinformatics* **22**, 2475 (2006).
15. M. Lapidot, Y. Pilpel, *EMBO Rep.* **7**, 1216 (2006).
16. S. Brantl, E. G. Wagner, *EMBO J.* **13**, 3599 (1994).
17. G. Andre *et al.*, *Nucleic Acids Res.* **36**, 5955 (2008).
18. S. Brantl, *Curr. Opin. Microbiol.* **10**, 102 (2007).
19. S. Katayama *et al.*, *Science* **309**, 1564 (2005).
20. S. D. Hooper *et al.*, *Mol. Syst. Biol.* **3**, 72 (2007).
21. T. Washio, J. Sasayama, M. Tomita, *Nucleic Acids Res.* **26**, 5456 (1998).
22. M. J. de Hoon, Y. Makita, K. Nakai, S. Miyano, *PLOS Comput. Biol.* **1**, e25 (2005).
23. A. Budin-Verneuil, E. Maguin, Y. Auffray, S. Dusko Ehrlich, V. Pichereau, *Lait* **84**, 8 (2004).
24. S. Boue, I. Letunic, P. Bork, *Bioessays* **25**, 1031 (2003).
25. T. Koide *et al.*, *Mol. Syst. Biol.* **5**, 285 (2009).
26. E. Yus *et al.*, *Science* **326**, 1263 (2009).
27. L. J. Chang, W. H. Chen, F. C. Minion, D. Shiuian, *Biochem. Biophys. Res. Commun.* **367**, 213 (2008).
28. S. Kühner *et al.*, *Science* **326**, 1235 (2009).
29. J. Weiner III, R. Herrmann, G. F. Browning, *Nucleic Acids Res.* **28**, 4488 (2000).
30. We thank the Genomics core facility at EMBL (Heidelberg, Germany), J. Lozano for help with statistical analysis, and the Ultrasequencing Unit at CRG. This work was funded by the Foundation Marcelino Botín, the Ministry of Education of Spain (MEC)—Consolider, and the European Research Council. M.G. is funded by the Spanish MEC—Formación Profesorado Universitario. V.N. is funded by the Netherlands Organization for Scientific Research.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5957/1268/DC1

Materials and Methods

SOM Text

Figs. S1 to S14

Tables S1 to S8

References

28 May 2009; accepted 2 October 2009

10.1126/science.1176951

Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme

David M. Shechner,^{1,2} Robert A. Grant,² Sarah C. Bagby,^{1,2} Yelena Koldobskaya,³ Joseph A. Piccirilli,³ David P. Bartel^{1,2*}

Primordial organisms of the putative RNA world would have required polymerase ribozymes able to replicate RNA. Known ribozymes with polymerase activity best approximating that needed for RNA replication contain at their catalytic core the class I RNA ligase, an artificial ribozyme with a catalytic rate among the fastest of known ribozymes. Here we present the 3.0 angstrom crystal structure of this ligase. The architecture resembles a tripod, its three legs converging near the ligation junction. Interacting with this tripod scaffold through a series of 10 minor-groove interactions (including two A-minor triads) is the unpaired segment that contributes to and organizes the active site. A cytosine nucleobase and two backbone phosphates about the ligation junction; their location suggests a model for catalysis resembling that of proteinaceous polymerases.

The RNA world hypothesis proposes that early life forms lacked DNA and coded proteins, depending instead on RNA for both chemical catalysis and information storage

(1). Central to this RNA world would have been polymerase ribozymes able to replicate RNA. Among the efforts to generate ribozymes with such ability, the most productive have started

with the class I RNA ligase ribozyme (2–4). This ribozyme was originally isolated from a large pool of random sequences (5, 6). It has since been improved by mutation and selection and has served as a platform for modeling ribozyme evolution in vitro (6–8). Because it rapidly promotes a reaction with chemistry identical to that catalyzed by proteinaceous enzymes that replicate RNA (Fig. 1A) (6), the ligase has provided the catalytic engine for more sophisticated RNA enzymes that use nucleoside triphosphates and the information from an external RNA template to synthesize short strands of RNA (2, 3, 9). Although more efficient with some templates than with others, this primer-extension reaction is gen-

¹Whitehead Institute for Biomedical Research and Howard Hughes Medical Institute, 9 Cambridge Center, Cambridge, MA 02142, USA. ²Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ³Department of Chemistry, Department of Biochemistry and Molecular Biology, and Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu

eral in that all templates tested support detectable extension (2–4). To understand the structural basis behind RNA-catalyzed RNA polymerization, we have solved the crystal structure of the class I ligase ribozyme.

The ligase sequence variant we crystallized was the product of three successive *in vitro* se-

lection experiments, the last of which mutagenized segments not participating in known base pairs (termed “joining regions”) and selected variants that folded and reacted within milliseconds (5, 6, 10). This experiment produced an improved variant that, unlike its predecessor, yielded useful crystals (data to 3.0 Å, tables S1 to S3, figs.

S1 and S2) (11). This variant is more tolerant of low Mg^{2+} concentrations; it reacts 15 times faster than the predecessor in 1 mM Mg^{2+} (10) but only slightly faster than the predecessor in high Mg^{2+} [predecessor reaction rate, 800/min in 60 mM Mg^{2+} , pH 9 (12)]. As with the predecessor, its reaction is pH-dependent, slowing to

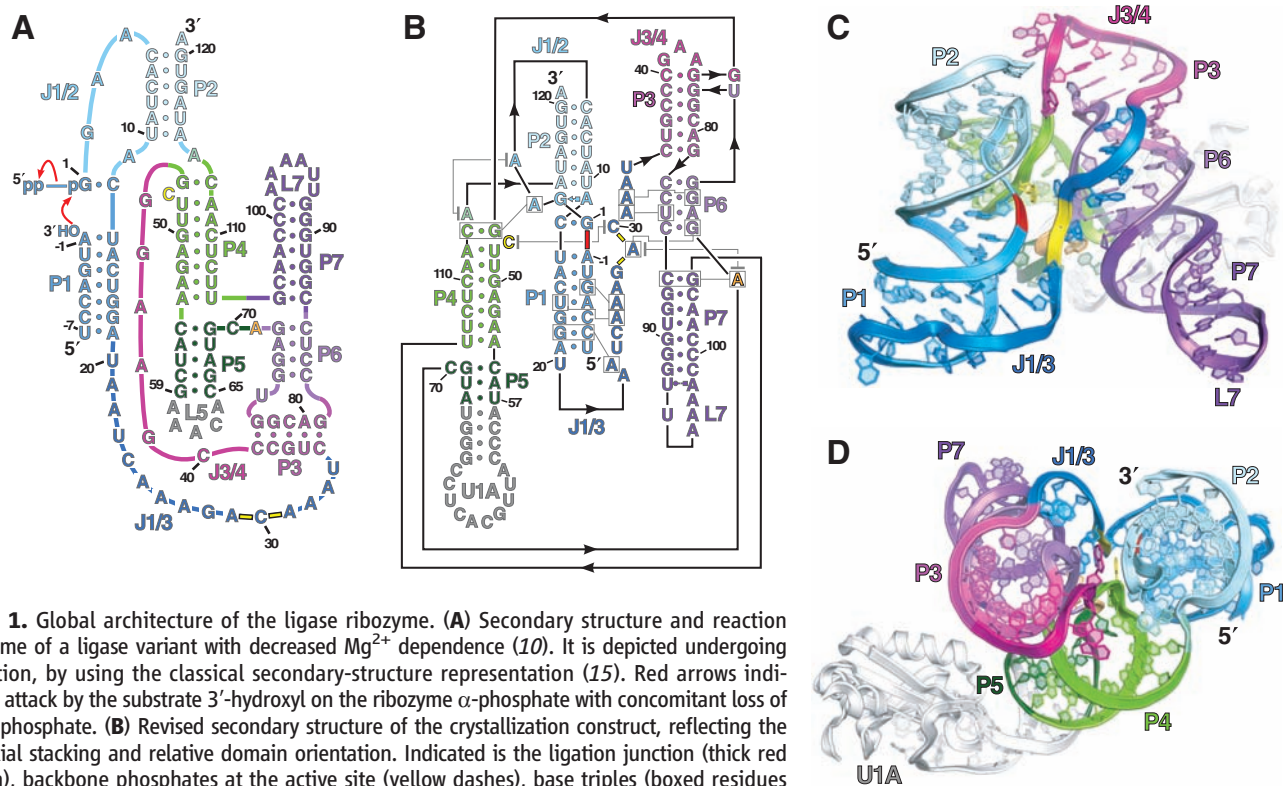
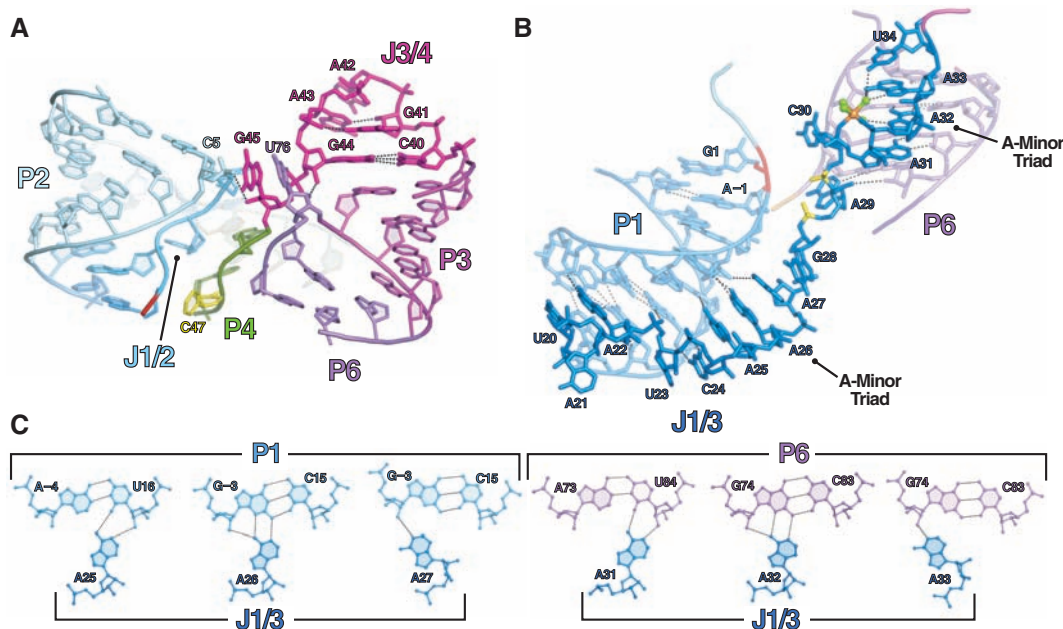


Fig. 1. Global architecture of the ligase ribozyme. (A) Secondary structure and reaction scheme of a ligase variant with decreased Mg^{2+} dependence (10). It is depicted undergoing ligation, by using the classical secondary-structure representation (15). Red arrows indicate attack by the substrate 3'-hydroxyl on the ribozyme α -phosphate with concomitant loss of pyrophosphate. (B) Revised secondary structure of the crystallization construct, reflecting the coaxial stacking and relative domain orientation. Indicated is the ligation junction (thick red dash), backbone phosphates at the active site (yellow dashes), base triples (boxed residues connected with gray lines), and stacking interactions (residues vertically aligned or connected with gray lines terminating in gray bars). Nucleotides numbered as in (A); those in gray were added to facilitate crystallization. Base-pair geometries indicated using nomenclature of (27). (C) Ribbon representation of ligase structure, as if peering into the active site (yellow) and ligation junction (red). (D) Top-down view, relative to (C).

Fig. 2. Tertiary contacts involving the three longest joining regions. (A) Interactions bridging the three domains. (B) The path of J1/3. (C) Hydrogen bonds of the two A-minor triads (fig. S8).



2.2/min in our crystallization conditions (10 mM Mg^{2+} , pH 6.0).

To promote crystallization, we replaced loop 5 (L5) with the U1A-binding loop, and grew crystals of the ligated product complexed with U1A (Fig. 1B) (11, 13). A parallel effort used a phage-display system to generate antibodies for cocrystallization (14), which yielded crystals with data to 3.1 Å. The ligase structure in this second crystal form, solved independently, confirmed the structure presented here (all-atom root mean square deviation = 1.48 Å) (fig. S3).

The global structure features three coaxially stacked domains: P1-P2, P3-P6-P7, and P4-P5 (Fig. 1, B and C), consistent with the previously predicted topology (15), but with the three domains placed at relative angles of 58° to 71°, converging near the ligation junction so as to resemble a tripod (Fig. 1, C and D). Because the tripod legs protrude into solvent, the fraction of surface area occluded from solvent is less than that of similarly sized RNAs (fig. S4).

Positioning these domains are tertiary interactions at the top of the tripod (Fig. 2A). G45

stacks on U76, the joining residue of the P6-P3 pseudoknot. This interaction pulls the 5' strands of P4 and P6 close to—and nearly parallel with—the J1/2 joining region, facilitating a contact between a C5 nonbridging oxygen and the 2'-hydroxyl of G45, a group with confirmed function (10). Preceding G45 is an unexpected Watson-Crick pair, G44:C40, which we confirmed biochemically (fig. S5) (11). This pair extends the P3 helical stack and closes a 3-nucleotide (nt) loop resembling a GNRA tetraloop (in which N is A, C, G, or U; R is A or G). In addition to this loop, two other regions (L7 and part of J1/3) resemble previously defined substructures (figs. S6 and S7).

Interacting with the tripod scaffold is J1/3, which docks into the P1 and P6 minor grooves, passing from one to the other near the ligation junction (Figs. 1C and 2B). Of the 10 minor-groove interactions, eight involve adenosines of J1/3. Residues A25–A26–A27 dock into the fourth and fifth base pairs of P1 (Fig. 2C and fig. S9), which corresponds to the primer-template duplex used by the polymerase. Each interaction could form irrespective of the P1 base-pair identity. Of particular note are the hydrogen bonds involving the 2'-hydroxyls of U16 and G-3. In the polymerase, 2'-deoxy substitution is more detrimental at the position analogous to U16 than at any other template residue, and 2'-deoxy substitution at the position analogous to G-3 is among the most detrimental primer substitutions (16). Hence A25–A26–A27 make defined, yet sequence-independent, contacts that help explain the ability of the polymerase to utilize primer-template helices of any sequence (2–4).

At the other end of J1/3, A31–A32–A33 dock into the P6 minor groove, passing from one helical strand to the other through a succession of hydrogen bonds identical to that of A25–A26–A27 (Fig. 2, B and C). We call this recurring motif the A-minor triad and note another instance in the small subunit of the bacterial ribosome (fig. S8). The P6 A-minor triad helps form a Mg^{2+} -binding site (Figs. 2B and 3A). Direct metal coordination by the A31 and A32 nonbridging (pro-*R*_p) phosphate oxygens brings these oxygens ~3.1 Å from one another, inducing a 90° kink that positions C30 out of the helical docking register of A31–A33. Outer-sphere contacts involving N7 of A32, N7 and N2 of A33, and O4 of U34 further stabilize this interaction—roles that, in concert with their packing into P6, explain both the absolute conservation of these nucleotides in active ribozyme isolates (2, 4, 8, 10, 17) and the deleterious effects of chemically modifying them (10).

Between the two A-minor triads lies the active site (Figs. 2B and 3, A to C). Forming the “floor” of the active site is A71, an absolutely conserved residue at the center of the four-way junction linking the P4-P5 and P3-P6-P7 domains (Fig. 1, A and B). A71 forms an imperfect type I A-minor interaction (18) with C86:G105, the first base pair of P7 (Fig. 3A and fig. S9). Chemical modification of A71 or loss of the C86 2'-hydroxyl impairs

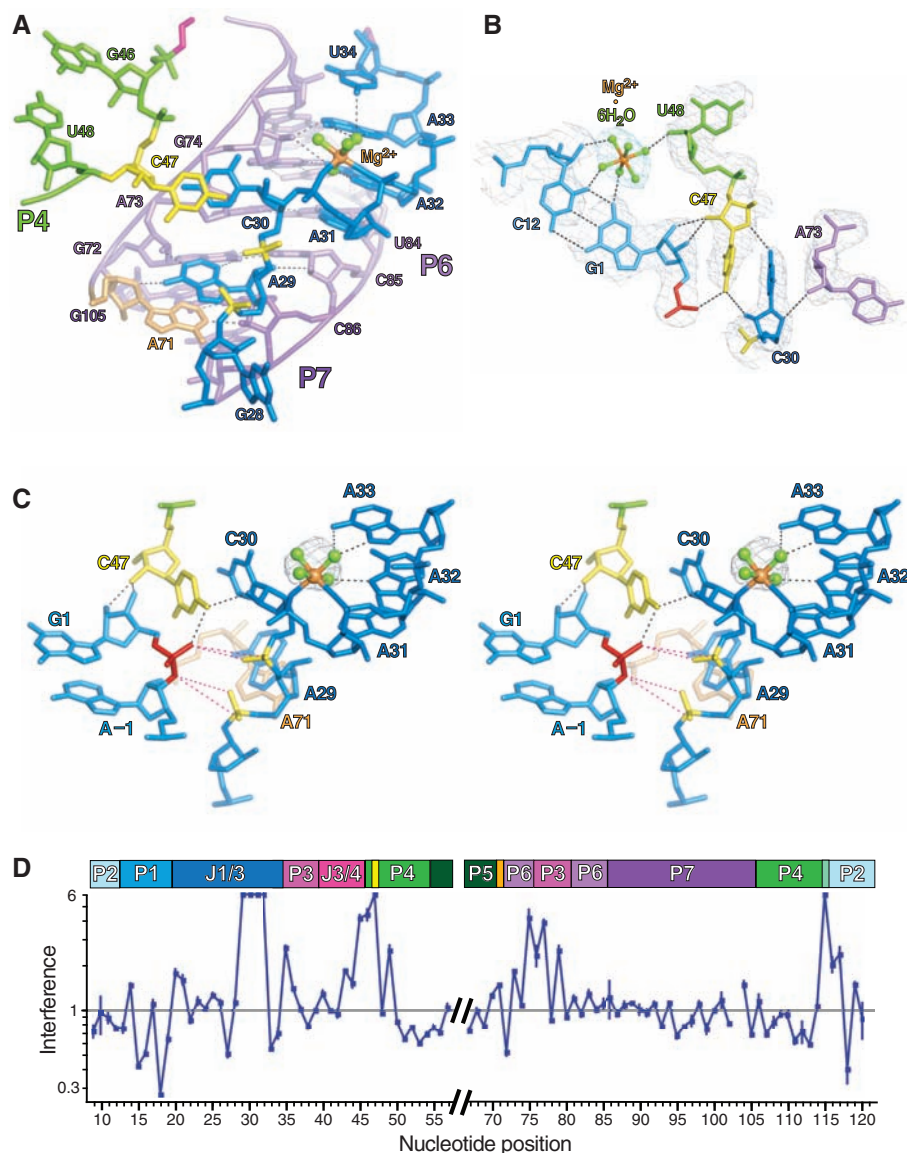


Fig. 3. Architecture of the active site. (A) The active site, as viewed from the ligation junction, with P1-P2 removed for clarity. (B) Interactions near G1:C12, which is analogous to the NTP-template pair during polymerization (2–4, 9). Meshes are simulated-annealing omit maps in which active-site nucleotides (gray, contoured at 2σ) or the hydrated metal cluster (aqua, 4σ) were excluded from map calculations. (C) Stereograph of the active site. Black dashes indicate hydrogen bonds; magenta dashes indicate proximity between A29 and C30 phosphate oxygens and the ligation junction (red). Mesh represents a simulated-annealing omit map (4.5σ) in which the hydrated metal was excluded from map calculations. (D) Mean interference values (\pm SD) from three α -phosphorothiolate NAIM experiments. The secondary structure is aligned above. Interference values were truncated at the detection limit, 6.0 (10, 22). Missing positions are those modified to facilitate crystallization (hashes) or too close to the termini to measure.

catalysis (10). A29 stacks on the A71 floor, itself forming an A-minor base triple (fig. S9). Residues A29 and C30 are hence enclosed between the A71 A-minor interaction and the metal-stabilized backbone kink (Fig. 3C). The consequences of this enclosure are twofold. First, C30 is extruded from the minor groove of P6, the rotation of its base constrained by A73 so as to form a cross-strand stack with C47 (Fig. 3, A to C). C47 is likewise extruded from helix P4, the plane of its base roughly perpendicular to the adjoining base pairs, with its N4 exocyclic amine positioned 3.1 Å away from the ligation junction. Second, the A71 interaction prevents G28 from stacking below A29. The consequent rotation of G28 places the phosphates of A29 and C30 in close proximity to one another (~5 Å between phosphorus centers) and facing P1. The nonbridging oxygens at these residues are 4.5 to 5.3 Å from the 3'-hydroxyl and phosphate oxygens of the ligation junction.

We therefore propose that C47, as positioned by C30, and the backbone phosphates of A29 and C30 compose the ligase active site. Both cytidines are conserved among active isolates (2, 4, 8, 10, 17), although their contributions to activity differ (fig. S10). The C30U substitution decreases activity by a factor of five, perhaps from disrupting the hydrogen bond between the C30 N4 and the C47 O4' (Fig. 3B), whereas the C47U substitution diminishes activity by a factor of $>10^4$, consistent with a more direct role in catalysis.

With only minor perturbation, the A29 and C30 phosphates could provide a binding site for a catalytic metal ion, as observed at active sites for some natural ribozymes (19, 20) and the L1 ligase, an artificial ribozyme that promotes a reaction resembling that of our ligase (21). Although we observed no electron density for such a metal ion in the crystal structure of the product, a metal might be bound more tightly before catalysis. To

test for a functional role for these and other backbone phosphate oxygens, we performed nucleotide-analog interference mapping (NAIM) (10, 22), randomly incorporating R_p -phosphorothioate substitutions and identifying those that interfered with activity (Fig. 3D). Substitution at residues A29 through A32 resulted in maximal interference. Substitution at A31 and A32 would disrupt the metal-stabilized kink needed to form the C30 to C47 stack at the active site (Fig. 3C). That substitution at A29 or C30 similarly abrogated function supports the hypothesis that these other phosphates coordinate at least one magnesium ion that is catalytically critical in the transition state but not bound tightly in the crystallized product.

We propose a preliminary model for catalysis by the class I ligase and its polymerase derivatives that resembles the mechanism of proteinaceous enzymes. Proteinaceous nucleic acid polymerases require a pair of aspartic acid-bound divalent metal ions supplemented by a general acid that stabilizes and protonates the pyrophosphate leaving group (Fig. 4A) (23, 24). In our model, the substrate α -phosphate and backbone phosphates of A29 and C30 jointly bind a catalytic magnesium ion (Fig. 4B). This metal activates the primer 3'-hydroxyl for nucleophilic attack and stabilizes the transition-state geometry, akin to Metal A of proteinaceous polymerases (23). In addition, because free nucleotide triphosphates (NTPs) bind divalent cations, we suggest that the G1 triphosphate (or the incoming NTP) enters the active site complexed with a second metal, which, after binding, would remain coordinated by oxygens on the β - and γ -phosphates. At the transition state, this second metal helps stabilize the developing negative charge on the pyrophosphate leaving group. This stabilization is aided by the exocyclic amine of C47, which hydrogen bonds to the (α,β) bridging oxygen.

Our model, which postulates a hydrogen bond to the leaving oxygen in the transition state, differs

from that of proteinaceous polymerases, which involves proton transfer to this oxygen in the transition state (24). Although nucleobases can act as general acids at ribozyme active sites (25), we disfavor ascribing such a function to C47. With increasing pH, the ribozyme ligation rate increases log-linearly with a slope of 1.0 (pH 5.7 to 8.5), consistent with the net loss of one proton, presumably that of the nucleophile, when proceeding from the ground state to the transition state (12). If general-acid catalysis by C47 were dominant at the transition state, the pH-rate profile would likely deviate from linearity over this pH range. Moreover, if C47 were a general acid, the functional group donating the proton would differ from that of the active-site cytidine of the hepatitis delta virus (HDV) ribozyme, wherein the N3 imine is thought to act as the proton donor (25). For the ligase, methylating N3 has little effect on catalysis, which rules out direct participation of N3 but not N4 (10).

By identifying the residues at the active site, the ligase crystal structure will facilitate directed examination of the catalytic mechanism of RNA-catalyzed RNA polymerization. Our model also provides insights into how known polymerase ribozymes recognize primer-template duplexes, the feature most in need of improvement for developing a self-replicating polymerase ribozyme (4, 26), and one that now can be targeted more explicitly in design and selection experiments.

References and Notes

- G. F. Joyce, L. E. Orgel, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999), pp. 49–77.
- W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, *Science* **292**, 1319 (2001).
- M. S. Lawrence, D. P. Bartel, *RNA* **11**, 1173 (2005).
- H. S. Zaher, P. J. Unrau, *RNA* **13**, 1017 (2007).
- D. P. Bartel, J. W. Szostak, *Science* **261**, 1411 (1993).
- E. H. Eklund, J. W. Szostak, D. P. Bartel, *Science* **269**, 364 (1995).
- M. C. Wright, G. F. Joyce, *Science* **276**, 614 (1997).
- G. F. Joyce, *Annu. Rev. Biochem.* **73**, 791 (2004).
- E. H. Eklund, D. P. Bartel, *Nature* **382**, 373 (1996).
- S. C. Bagby, N. H. Bergman, D. M. Shechner, C. C. Yen, D. P. Bartel, *RNA*, published online 27 November 2009 (10.1261/rna.191250).
- Materials and methods and supporting text are available on *Science* Online.
- M. E. Glasner, N. H. Bergman, D. P. Bartel, *Biochemistry* **41**, 8103 (2002).
- A. R. Ferré-D'Amaré, J. A. Doudna, *J. Mol. Biol.* **295**, 541 (2000).
- J. D. Ye *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 82 (2008).
- N. H. Bergman, N. C. Lau, V. Lehnert, E. Westhof, D. P. Bartel, *RNA* **10**, 176 (2004).
- U. F. Müller, D. P. Bartel, *Chem. Biol.* **10**, 799 (2003).
- E. H. Eklund, D. P. Bartel, *Nucleic Acids Res.* **23**, 3231 (1995).
- P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4899 (2001).
- M. R. Stahley, S. A. Strobel, *Science* **309**, 1587 (2005).
- N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle, *Science* **320**, 77 (2008).
- M. P. Robertson, W. G. Scott, *Science* **315**, 1549 (2007).
- S. A. Strobel, *Curr. Opin. Struct. Biol.* **9**, 346 (1999).
- N. Sträter, W. N. Lipscomb, T. Klambunde, B. Krebs, *Angew. Chem. Int. Ed. Engl.* **35**, 2024 (1996).
- C. Castro *et al.*, *Nat. Struct. Mol. Biol.* **16**, 212 (2009).

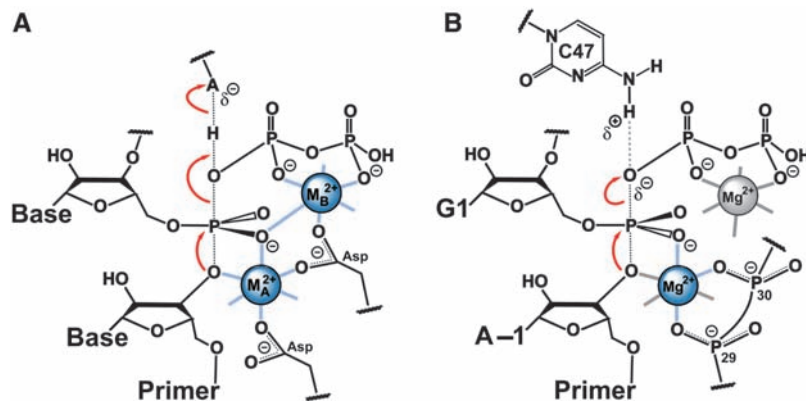


Fig. 4. Transition-state stabilization by polymerases built from either protein or RNA. **(A)** Catalysis by proteinaceous polymerases (23, 24). Indicated are bonds formed or broken during the transition state (red arrows), coordination of catalytic metal ions, M_A and M_B (blue solid lines), and an active-site acid ($A \cdots H$). **(B)** Model for catalysis by the ligase ribozyme. Notation as in (A), with the addition of a hydrogen bond between C47 N4 and the leaving group (dashed gray line). Some magnesium ligands are not specified; for those that are, relative orientations are unknown. A proposed contact to the reactive phosphate pro- R_p oxygen (28) and two speculative contacts implied by NAIM are in blue. Metal ion and coordinations not supported (or refuted) by structural or biochemical evidence are in gray.

25. P. C. Bevilacqua, R. Yajima, *Curr. Opin. Chem. Biol.* **10**, 455 (2006).
26. M. S. Lawrence, D. P. Bartel, *Biochemistry* **42**, 8748 (2003).
27. N. B. Leontis, E. Westhof, *RNA* **7**, 499 (2001).
28. M. E. Glasner, C. C. Yen, E. H. Eklund, D. P. Bartel, *Biochemistry* **39**, 15556 (2000).
29. We thank A. Ferré-D'Amaré, J. Doudna, R. Batey, and K. Zhou for the generous donation of reagents and protocols; K. Rajashankar for assistance in phasing; E. Duguid, V. Malashkevich, C. Drennan, T. Schwartz, U. RajBhandary,

J. Cochrane, M. Stahley, S. Strobel, M. Robertson, and W. Scott for advice, encouragement, and helpful discussions; and members of the Bartel laboratory for support and comments. Supported by NIH grant GM61835. X-ray data were collected with the Northeastern Collaborative Access Team at the Advanced Photo Source, Argonne National Laboratory (11). D.M.S. was an NSF predoctoral fellow; S.C.B. was a Howard Hughes Medical Institute (HHMI) predoctoral fellow. D.B. and J.A.P. are HHMI investigators. Coordinates for the U1A-bound and Fab-bound ribozymes are deposited in the Protein Data Bank (3HHN and 3IVK, respectively).

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5957/1271/DC1
Materials and Methods
SOM Text
Figs. S1 to S12
Tables S1 to S3
References

8 April 2009; accepted 24 July 2009
10.1126/science.1174676

A High-Resolution Structure of the Pre-microRNA Nuclear Export Machinery

Chimari Okada,^{1,6*} Eiki Yamashita,^{1,*} Soo Jae Lee,^{2,*†} Satoshi Shibata,³ Jun Katahira,^{3,4} Atsushi Nakagawa,¹ Yoshihiro Yoneda,^{3,4,5†} Tomitake Tsukihara^{1,6†}

Nuclear export of microRNAs (miRNAs) by exportin-5 (Exp-5) is an essential step in miRNA biogenesis. Here, we present the 2.9 angstrom structure of the pre-miRNA nuclear export machinery formed by pre-miRNA complexed with Exp-5 and a guanine triphosphate (GTP)-bound form of the small nuclear guanine triphosphatase (GTPase) Ran (RanGTP). The x-ray structure shows that Exp-5:RanGTP recognizes the 2-nucleotide 3' overhang structure and the double-stranded stem of the pre-miRNA. Exp-5:RanGTP shields the pre-miRNA stem from degradation in a baseball mitt-like structure where it is held by broadly distributed weak interactions, whereas a tunnel-like structure of Exp-5 interacts strongly with the 2-nucleotide 3' overhang through hydrogen bonds and ionic interactions. RNA recognition by Exp-5:RanGTP does not depend on RNA sequence, implying that Exp-5:RanGTP can recognize a variety of pre-miRNAs.

Mature microRNAs (miRNAs), short non-coding RNAs present in a wide range of eukaryotes (1, 2), play important roles in the regulation of biological processes including development, cell proliferation, cell differentiation, apoptosis, transposon silencing, and antiviral defense (3–6). miRNA biogenesis (7) begins in the nucleus, where capped and polyadenylated primary miRNAs, several kilobases in length, are transcribed. These are processed by the nuclear ribonuclease (RNase) III enzyme Drosha to generate ~65-nucleotide (nt) pre-miRNAs that have stem-loop structures containing 2-nt 3' overhangs. Exp-5 translocates pre-miRNAs from the nucleus to the cytoplasm through the nuclear pore complex (8–12). In the cytoplasm, the pre-

miRNAs are further processed by the cytoplasmic RNase III enzyme Dicer, which excises a ~22-base pair (bp) RNA duplex. One strand of the duplex binds to its target mRNA with imperfect complementarity, usually within the target's 3' untranslated region, assisted by the RNA-induced silencing complex (7).

Exp-5 facilitates miRNA biogenesis not only by acting as the nuclear export factor for pre-miRNAs but also by protecting pre-miRNAs from digestion by nucleases. Loss of Exp-5 results in the loss of cytoplasmic miRNA expression without nuclear accumulation of pre-miRNAs (10). Pre-miRNA binding to Exp-5 requires the guanine triphosphatase (GTPase) Ran (RanGTP). The Exp-5:RanGTP:pre-miRNA heteroternary complex formed in the nucleus is exported to the cytoplasm. Ran GTPase-activating protein, which promotes guanine triphosphate (GTP) hydrolysis in conjunction with RanBP1 and/or RanBP2, is exclusively localized in the cytoplasm and triggers the conformation change of Ran to induce release of the pre-miRNA cargo from Exp-5 (13, 14).

Here, we report the structure of the Exp-5:RanGTP:pre-miRNA complex at 2.9-Å resolution (Fig. 1A and fig. S1). This complex contains full-length human Exp-5, canine RanGTP residues 1 to 176 (removal of residues 177 to 216 stabilizes the GTP-bound conformation), and the 48-nt human pre-miRNA-30a stem domain, which includes the 2-nt 3' overhang (nucleotide

numbers 1 to 24 and 40 to 63 of human pre-miRNA-30a). Phase information used for the crystal structure analysis was derived from crystals containing Se-methionine-substituted Exp-5, and the RNA sequence was assigned from the Br anomalous signal information in crystals containing pre-miRNA 5-bromo-oxyuracil derivatives. The structure was refined to an *R* factor of 0.247 and free *R* factor of 0.312, and phasing statistics are provided in table S1. We modeled 1082 of 1204 residues of Exp-5. Several loop regions in the 20 HEAT repeats and 55 residues at the C terminus could not be modeled (details in fig. S1), and 13 residues at the C terminus were modeled as a polyalanine α helix. The residues 1 to 6 of Ran were not modeled because of their disordered structure. Electron density for the pre-miRNA was detected for nucleotides 1 to 11, 14 to 24, and 40 to 63 (fig. S2). The pre-miRNA-30a adopted a typical A-form RNA helical structure, 60 Å in length and 20 Å in diameter.

The Exp-5:RanGTP:pre-miRNA complex is an ellipsoid with dimensions of 65 Å by 80 Å by 110 Å. The crystal structure contains two ternary complexes, labeled A and B, in the asymmetric unit, which are essentially similar [root mean square (RMS) of 1.84 Å, where B is slightly more open than A] and present the same recognition modes for the pre-miRNA. Detailed structural comparison of ternary complexes A and B is described in (15). The structure of Exp-5 resembles a tightly wound spring, as seen in other members of the importin- β family. Such conformations are expected to be intrinsically flexible, so small changes in the relative orientation of successive HEAT repeats could cumulatively generate substantial changes in the helicoidal pitch (16). Ternary complex A yielded more contrast in its electron density map than did complex B; thus, all structural descriptions of the ternary complex in the following discussion will be restricted to ternary complex A. The Exp-5:RanGTP complex forms a baseball mitt-like structure in which the pre-miRNA is packed (Fig. 1B). A tunnel-like structure at the bottom of the mitt connects the inner space of the mitt with the outer space (Fig. 1B).

The pre-miRNA stem is caught in the mitt formed by the Exp-5:RanGTP complex (Fig. 1), whereas the 15-Å 2-nt 3' overhang is inserted into a tunnel formed from elements of HEAT repeats 12 to 15 (Figs. 2 and 3 and fig. S3). The inner surface of the tunnel is positively charged

¹Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita, Osaka 565-0871, Japan. ²College of Pharmacy, Chungbuk National University, Gaeshin-dong, Heungduk-gu, Cheongju City, Chungbuk, Korea. ³Department of Frontier Biosciences, Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁴Department of Biochemistry, Graduate School of Medicine, Osaka University, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁵Japan Science and Technology Agency (JST), Core Research for Evolutional Science and Technology (CREST), Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁶Department of Life Science, University of Hyogo, 3-2-1 Koto, Kamigori, Akoh, Hyogo 678-1297, Japan.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: sjlee@chungbuk.ac.kr (S.J.L.); yyoneda@anat3.med.osaka-u.ac.jp (Y.Y.); tsuki@protein.osaka-u.ac.jp (T.T.)