Molecular basis of bacterial protein Hen1 activating the ligase activity of bacterial protein Pnkp for RNA repair

Pei Wang, Chio Mui Chan1, David Christensen, Can Zhang, Kiruthika Selvadurai, and Raven H. Huang2

Department of Biochemistry, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Edited by Stewart Shuman, Sloan-Kettering Institute, New York, NY, and accepted by the Editorial Board July 3, 2012 (received for review June 8, 2012)

Ribotoxins cleave essential RNAs for cell killing in vivo, and the bacterial polynucleotide kinase-phosphatase (Pnkp)/hua enhancer 1 (Hen1) complex has been shown to repair ribotoxin-cleaved RNAs in vitro. Bacterial Pnkp/Hen1 is distinguished from other RNA repair systems by performing 3′-terminal 2′-O-methylation during RNA repair, which prevents the repaired RNA from repeated cleavage at the same site. To ensure the possibility of 2′-O-methylation by bacterial Hen1 during RNA repair and, therefore, maintain the quality of the repaired RNA, Pnkp/Hen1 has evolved to require the participation of Hen1 in RNA ligation, because Pnkp alone is unable to carry out the reaction despite possessing all signature motifs of an RNA ligase. However, the precise role of Hen1 in RNA ligation is unknown. Here, we present the crystal structure of an active RNA ligase consisting of the C-terminal half of Pnkp (Pnkp-C) and the N-terminal half of Hen1 (Hen1-N) from Clostridium thermocellum. The structure reveals that the N-terminal domain of Clostridium thermocellum (CtH) Hen1, shaped like a left hand, grabs the flexible insertion module of CtH Pnkp and locks its conformation via further interaction with the C-terminal addition module of CtH Pnkp. Formation of the CtH Pnkp/CtHen1-N heterodimer creates a ligation pocket with a width for two strands of RNA, depth for two nucleotides, and the adenosine monophosphate (AMP)-binding pocket at the bottom. The structure, combined with functional analyses, provides insight into the mechanism of how Hen1 activates the RNA ligase activity of Pnkp for RNA repair.

2′-O-methyltransferase | enzyme activation | RNA repair

Because of the presence of the 2′-OH group, RNA in living cells is much more susceptible to cleavage than DNA. The majority of the cleavage, carried out nonspecifically by protein ribonucleases, is part of normal RNA metabolism in living organisms. However, a variety of sequence- or site-specific RNA cleavage exists in living cells, and specific cleavage of some essential RNAs has been used for gene regulation and cell killing. In prokaryotes, a class of protein toxins named ribotoxins is mainly responsible for most site-specific RNA cleavage. Two well-characterized examples of ribotoxins for gene regulation are RelE (1, 2) and VapC (3). Several well-studied ribotoxins for cell killing are ricin (4), sarcin (5), PrtC (6), colicin ES (7), colicin D (8), and Kluyveromycetes lactis γ-toxin (9). All of them cleave essential RNAs involved in protein translation. With the exception of ricin, which uses a depurination mechanism, ribotoxins use a transesterification mechanism to cleave RNA.

To counter cell killing by ribotoxins, organisms have evolved protein enzymes that repair the cleaved RNAs. The first RNA repair system discovered consists of two proteins named Pnkp (polynucleotide kinase-phosphatase) and Rnl1 (RNA ligase 1) from bacteriophage T4 (6). Since the discovery of the T4 RNA repair system, a few multifunctional enzymes have been implicated in RNA repair (10–13). Of particular relevance to the present study is the bacterial Pnkp that was shown to possess kinase, phosphatase, and adenylyltransferase activities (11). Although bacterial Pnkp is able to process the two ends of a cleaved RNA, it is unable to ligate the two processed ends despite its possession of all signature motifs of an RNA or DNA ligase (14). To complete the RNA repair, we demonstrated that a second bacterial protein, termed Hen1, is required (15). Bacterial Hen1, like its eukaryotic counterpart involved in RNA interference (RNAi), is a unique methyltransferase that carries out 2′-O-methylation at the 3′-terminal nucleotide of RNA (16, 17). Our study further revealed that the 2′-OH group at the junction of repair is methylated by bacterial Hen1 during RNA repair by the bacterial Pnkp/Hen1 complex (15). Because the same 2′-OH group is responsible for the original RNA cut, its methylation results in the repaired RNA resisting future cleavage by the ribotoxin.

To maintain the quality of the repaired RNA (e.g., methylated repaired RNA), the bacterial Pnkp/Hen1 RNA repair system appears to have evolved to require participation of Hen1 in RNA ligation, ensuring the opportunity of Hen1 to carry out 2′-O-methylation during RNA repair. However, the involvement of Hen1 results in an active RNA ligase is unknown. Recent crystal structures of the ligase domain of Pnkp by Shuman and coworkers provide the first step toward our understanding of the mechanism that underlies the ability of Hen1 to activate Pnkp for RNA ligation (18). The study revealed that, in addition to the presence of the classical nucleotidyltransferase (NTase) module, the ligase domain of Pnkp has a flexible protein module inserted in the NTase module (insertion module) and a second protein module added at the C terminus of NTase (C-addition module). Here, we report the crystal structure of the C-terminal half of bacterial Pnkp (Pnkp-C) in complex with the N-terminal half of bacterial Hen1 (Hen1-N). The structure reveals that the formation of the Pnkp-C/Hen1-N heterodimer creates a ligation pocket to accommodate the two strands of RNA for ligation. In addition to the structure, we also carried out functional analyses of the complex, including protein adenylation reaction in a crystal, RNA adenylation reaction in solution, and structure-guided mutagenesis. The combined structural and functional studies of the Pnkp-C/Hen1-N complex shed light on how bacterial Hen1 activates the ligase activity of Pnkp for RNA repair.

Results and Discussion

Reconstituting the CthPnkp/C/Hen1-N Heterodimer in Vitro As an Active RNA Ligase. The recombinant C-terminal half of CthPnkp (residues 444–870; 49 kDa) and N-terminal half of CthHen1...
(residues 1–230; 26 kDa) were overexpressed in *Escherichia coli* and purified to homogeneity. Size-exclusion chromatography of the purified proteins showed *Cth*Pnkp-C as a monomer and *Cth*Hen1-N as a homodimer (Fig. 1A, green and red curves). The formation of *Cth*Hen1-N homodimer was also observed previously by Jain and Shuman using a glycerol gradient sedimentation method (17). An equal molar mixture of *Cth*Pnkp-C and *Cth*Hen1-N, however, resulted in the formation of a *Cth*Pnkp-C/Hen1-N heterodimer instead of the expected heterotrimer (Fig. 1A, black). To carry out functional analyses, we chose cleaved tRNA<sup>Arg</sup>·ΔT (Fig. S1) as the RNA substrate for this study because it was the most efficient substrate among eight ribotoxin-cleaved RNAs in our recent study (19). RNA ligation assay using the processed RNAs showed that a combination of *Cth*Pnkp-C and *Cth*Hen1-N, but neither one alone, was able to carry out RNA ligation (Fig. 1B, lane 6), demonstrating that the *Cth*Pnkp-C/Hen1-N heterodimer is an active RNA ligase. We also tested a shorter version of *Cth*Pnkp-C (residues 479–870; 45 kDa) for RNA ligation, and we named it *Cth*Pnkp-C·ΔL because it essentially lacks the linker region as shown in Fig. 2A. *Cth*Pnkp-C·ΔL also required *Cth*Hen1-N for RNA ligation, and it showed higher activity than *Cth*Pnkp-C (Fig. 1B, lane 7).

**Architecture of the *Cth*Pnkp-C/Hen1-N Heterodimer.** We crystallized the *Cth*Pnkp-C/Hen1-N heterodimer and solved the structure at 2.6 Å (Table S1). We also obtained the structure of the *Cth*Pnkp-C/Hen1-N heterodimer covalently bound to an AMP by soaking a crystal of the *Cth*Pnkp-C/Hen1-N heterodimer with ATP and Mg<sup>2+</sup> before data collection (Table S1). Because the structure of the *Cth*Pnkp-C/Hen1-N/AMP ternary complex exhibits a bound AMP and has higher resolution, it was used for detailed structural analyses (Fig. 2). In addition, we also solved the structure of the *Cth*Hen1-N homodimer (Fig. 3A and Table S1).

The fold of *Cth*Hen1-N can be depicted as a seven-stranded antiparallel β-sheet stacked by two layers of helices (Figs. 2B and C and S4). We tentatively assigned the β-sheet together with the first layer of helices (Fig. 2, colored red) as the ligase-activating domain of *Cth*Hen1 and the second layer of helices (Fig. 2, colored yellow) as part of the linker region that connects the ligase-activating domain to the missing methyltransferase (MTase) domain (Fig. 2A). To facilitate structural and functional analyses, the ligase-activating domain of *Cth*Hen1-N is depicted as a left hand (Figs. 2C and S4 and Fig. S2).

**A** Dali (20) search using the structure of *Cth*Hen1-N from the *Cth*Pnkp-C/Hen1-N/AMP ternary complex identified structures of low-scoring hits that possess a fold of a β-sheet stacked by a layer of α-helices. A total of 355 hits met the criterion of Z-score above 2, demonstrating the prevalent fold of β-sheet/α-helices stacking in protein structures. The top-scoring hit is the crystal structure of a KH domain of zipcode binding protein (ZBP) 1 (21), having a Z-score of 5.9, r<sub>msd</sub> of 4.6 Å, and 13% sequence identities. *Cth*Hen1-N is significantly less similar to the remaining structures of the search (Z-scores of 4.2 or less), and none of the structures of the top ten-scoring hits possesses loops that constitute the thumb and fingers in *Cth*Hen1-N (Fig. S3). This analysis, combined with the information that the N-terminal half of bacterial Hen1 has no apparent sequence similarity to other proteins, suggests that the N-terminal half of bacterial Hen1 is distinct.

The ligase domain of *Cth*Pnkp, as recently reported by Shuman and coworkers (18), is composed of an NTase module (Fig. 2, green), an insertion module (Fig. 2, blue), and a C-addition module (Fig. 2, cyan). In addition to these three modules, we also observed a structure containing an extra 35 aa at the N terminus of *Cth*Pnkp-C (Fig. 2, magenta) and tentatively assigned it as part of the linker region that connects the ligase domain to the missing phosphatase and kinase domains (Fig. 2A). In light of the new information from the structure of the *Cth*Pnkp-C/Hen1-N heterodimer, we made minor adjustments in defining the boundaries of the individual modules in *Cth*Pnkp-C (Fig. 2A and Fig. S4).

**Structural Changes upon Formation of the *Cth*Pnkp-C/Hen1-N Heterodimer.** Significant structural changes take place in both *Cth*Hen1-N and *Cth*Pnkp-C upon formation of the heterodimer. Comparison between *Cth*Hen1-N as a homodimer and *Cth*Hen1-N as a heterodimer reveals little change of the main body of the structure (β-strands plus two layers of α-helices) (Fig. 3A; r<sub>msd</sub> = 0.52 Å). On the other hand, the loops constituting the thumb and three of the four fingers are seen only with the formation of the heterodimer (Fig. 3A). The structures reveal that the palm of the ligase-activating domain is the common interface for both the *Cth*Hen1-N homodimer and the *Cth*Pnkp-C/Hen1-N heterodimer (Figs. 3A and 2C), and this structural insight provides a molecular explanation for the observed transition from the *Cth*Hen1-N homodimer to the *Cth*Pnkp-C/Hen1-N heterodimer in solution (Fig. 1A).
In the absence of CthHen1-N, the insertion module of CthPnkp-C is conformationally flexible, as demonstrated by three different folds and orientations in the structures of CthPnkp-C/ATP and CthPnkp-C/AMP complexes (18) (Fig. 3B). In the CthPnkp-C/Hen1-N/AMP complex, however, the conformation of the insertion module is fixed (Fig. 3B), which is made possible by the extensive interaction of CthHen1-N with the insertion module plus additional interaction of CthHen1-N with the C-addition module (Fig. 4). Although the insertion module in the structure of CthPnkp-C/AMP is oriented in a manner similar to the one in CthPnkp-C/Hen1-N/AMP, it is interesting to notice that the relative positions of the two helices (Fig. 3B, blue and light blue) that constitute the insertion module oppose one another. As a result of the formation of the CthPnkp-C/Hen1-N complex, the side chains of six conserved residues near the base of the insertion module, which are scattered in the structures of CthPnkp-C alone, are oriented in one direction (Fig. 3B). As discussed later, the identities of these residues and their correct orientations are important for the ligase activity of Pnkp. Finally, although most of the C-addition module structure is unchanged upon formation of the CthPnkp-C/Hen1-N complex, the region that interacts with the ligase-activating domain also has a small, but significant, structural change (Fig. S5).

Recognition of CthPnkp-C by CthHen1-N. The palm of the ligase-activating domain only interacts with the insertion module, and the interactions are predominantly hydrophobic (Fig. 4A). In addition to numerous hydrophobic interactions, the side chain of D55, whose position is stabilized by a hydrogen bond with the side chain of H37, forms a salt bridge with the side chain of R671 (Fig. 4A). The potential to form such a salt bridge is conserved among Pnkp/Hen1 complexes from different organisms (Figs. S2 and S4). Meanwhile, the thumb of the ligase-activating domain also interacts with the insertion module (Fig. 4B). The interactions mainly occur between the side chains of D153 and E154 from the ligase-activating domain and the side chains of R631, Y686, and R687 from the insertion module (Fig. 4B). In addition, the side chain of W159 from the ligase-activating domain stacks on the opposite side of the insertion module where the important conserved residues for RNA ligation are located (Figs. 4B and 3B).

Unlike the palm and thumb, finger2 and finger3 of the ligase-activating domain make contacts with both the insertion and the C-addition modules (Fig. 4C). The region is also where the insertion module and the C-addition module interact with each other. At the center of the three-way interaction is an extensive hydrogen bonding network formed by the side chains of Y83, R675, and D806, the main-chain carboxyl groups of L54 and P82, and a water molecule (Fig. 4C). Two additional interactions, a salt bridge between the side chains of D57 and R815 and a π-charge stacking between the side chains of Y77 and R869, further stabilize the interaction between the ligase-activating domain and the C-addition module (Fig. 4C). R869 is part of the strictly conserved DPRL motif at the C terminus of Pnkp (Fig. S4). The main chains of R869 and the terminal residue L870 form salt bridges with R794 and K792 (Fig. 4C), both of which are members of motif V of the NTase module (Fig. S4).

Ligation Pocket. Formation of the CthPnkp-C/Hen1-N heterodimer creates a deep pocket with the covalent AMP at the bottom (Fig. 5A and B). We call it the ligation pocket because it is most likely the locus of RNA ligation. The pocket is partially divided by the side chain of R565 (Fig. S4, marked with an asterisk), which is a member of motif Ia of the NTase module (Fig. S4). The NTase module essentially provides the entire floor of the ligation pocket, but it only contributes ~25% of the wall (Fig. S4, green). The C-addition module provides ~50% of the wall (Fig. S4, cyan), and the remaining ~25% of the wall is from a joint contribution of the insertion module and the ligase-activating domain (Fig. S4, blue and red). The ligation pocket is overwhelmingly positively charged (Fig. 5B), consistent with its proposed role in accommodating two strands of the negatively charged RNA.

We docked two strands of RNA (three nucleotides each) into the ligation pocket (Fig. 5C). The docking was guided by the position of the covalently bound AMP and also by the aligned structure of human DNA ligase I in complex with DNA (22). In our docking model, the two strands of RNA fit nicely into the
ligation pocket, with most phosphate groups near the positively charged wall (Fig. 5C). The ligation pocket is approximately two nucleotides deep, because the third nucleotide of both strands faces a charged wall (Fig. 5C). The ligation pocket is approximately two nucleotides deep, because the third nucleotide of both strands.

**Functional Analyses.** A classic DNA/RNA ligation reaction involves three enzymatic steps (23, 24): (i) a DNA/RNA ligase reacts with ATP to form a ligase-AMP covalent intermediate; (ii) AMP is transferred from the ligase to the 5′-phosphate to form a DNA- or RNA-adenylate intermediate; and (iii) ligase-catalyzed nucleophilic attack of the 3′-OH on the adenylated 5′-phosphate produces a phosphodiester bond that seals the two ends, accompanied by the release of AMP. Here, we performed step 1 in a crystal of the CthPnkp-C/5′-Pnkp-C/AMP/CthHen1-N heterodimer by soaking the crystal with ATP and Mg2+. Here, we performed step 2 in a crystal of the CthPnkp-C/5′-Pnkp-C/AMP/CthHen1-N heterodimer by soaking the crystal with ATP and Mg2+. Here, we performed step 2 in a crystal of the CthPnkp-C/5′-Pnkp-C/AMP/CthHen1-N heterodimer by soaking the crystal with ATP and Mg2+ before data collection (Fig. S6). The presence of an AMP covalently linked to the side chain of K531 is clearly observed (Fig. S6B). Also present in the active site is an Mg2+ ion, with the phosphate group of AMP and five water molecules as its ligands (Fig. S6B).

Pnkp alone is able to carry out step 1 (11, 18), and both Pnkp and Hen1 are required for step 3 (15). Therefore, what remains unknown is the prerequisite for step 2. To address this issue, we performed a step 2-specific enzymatic assay (Fig. S4). For this experiment and the experiments to follow, we used CthPnkp-C-ΔL, which we found to be more active as noted above (Fig. 1B, lane 7). We speculate that, in the absence of the phosphatase and kinase domains, the presence of partial linker region in CthPnkp-C might make the ligation pocket less accessible to a RNA substrate, causing CthPnkp-C to be less active than CthPnkp-C-ΔL. As expected, CthHen1-N alone has no activity (Fig. 6A, lanes 1 and 2). We were able to detect a trace amount of activity with CthPnkp-C-ΔL alone (Fig. 6A, lanes 3 and 4; 0.05% of activity), indicating that the essential elements to perform step 2 are already present in bacterial Pnkp. However, an efficient step 2 reaction requires CthHen1-N (Fig. 6A, lanes 5 and 6). Both one-stranded RNA and two-stranded ligation substrate mimic are efficient substrates, although one-stranded RNA is better (Fig. 6A, compare lanes 5 vs. 6).

To provide a detailed view of the roles of the conserved residues in RNA ligation, we carried out mutation and deletion studies (Fig. 6B–E). Here, we focused on residues beyond the

**Fig. 3.** Structural changes from the CthHen1-N homodimer and the CthPnkp-C monomer to the CthPnkp-C/CthHen1-N heterodimer. (A) Structure of the CthHen1-N homodimer and structural comparison of CthHen1-N from the homodimer to the heterodimer. (Left) Ribbon representation of the structure of the CthHen1-N homodimer. The ligase-activating domain is colored blue and orange, and the linker region is colored green and lemon, respectively. (Right) Superposition of the structure of a CthHen1-N in the homodimer (blue and green) with the structure of CthHen1-N in the CthPnkp-C/CthHen1-N heterodimer (red and yellow). The rmsd of the two structures is 0.52 Å. The structure of the ligase-activating domain is depicted as a left hand, with the β-sheet as the palm and five loops as the thumb and fingers. (B) Ribbon representation of the structures of the insertion module from the CthPnkp-C/ATP (PDB ID code 3TY9), the CthPnkp-C/AMP (PDB ID code 3TY9), and the CthPnkp-C/CthHen1-N/AMP complexes. The structure is in the same orientation as the one shown in Fig. 2B. Two helices of the structure are colored with different shades of blue to highlight their relative positions. The side chains of six conserved residues near the base of the insertion module are in stick and colored orange, with heteroatoms colored individually (nitrogen in blue and oxygen in red).

**Fig. 4.** Details of the heterodimer interface. (A) Details of the interface between the palm of the ligase-activating domain and the insertion module. Cα-chains of the structure are represented and colored the same as in Fig. 2B and C. The side chains are in stick and colored orange except heteroatoms, which are colored individually (nitrogen in blue, oxygen in red, and sulfur in yellow). Hydrogen bonds and salt bridges are depicted with black dashed lines. Many hydrophobic residues involved in interaction between the palm and the insertion module are not labeled for clarity. (B) Details of the interface between the thumb of the ligase-activating domain and the insertion module. (C) Details of the interface of the three-way interaction among finger2 and finger3 of the ligase-activating domain, the insertion module, and the C-addition module. A water molecule is depicted in sphere and colored red. The covalent AMP is partially shown to provide the location of the active site.
two thirds). Alanine mutations of some conserved aromatic residues (Fig. 6D, lane 2) have the most effect, with W611A being the most severe mutant (Fig. 6D, lane 2; <1% activity). Deletion of the first residue of CthHen1-N results in a sevenfold reduction of activity (Fig. 6C, lane 2), which is consistent with the conservation of the precise N terminus in bacterial Hen1 (Fig. S2). The remaining three residues in this group are all positively charged (R63, K614, and K824). We suggest that these three residues are involved in guiding RNA substrate for step 2 or/and step 3. Among the mutants with severe defects, there are representatives from each of the three modules proposed to make up the ligation pocket. This fact supports the observed structural arrangement in the Pnkp-C/Hen1-N heterodimer. The substantial reduction of the enzymatic activity of these eight mutants most likely resulted from local structural alterations caused by the alanine mutations. However, we cannot rule out the possibility of the global structural changes of these mutants, particularly for alanine mutations of some large hydrophobic residues such as W611, Y623, Y689, and Y798.

The second group includes four mutants (L61A, Y77A, H194A, and Q622A) that have only modest effect on enzymatic activity (defined by a reduction in activity of less than one-half). They are all relatively distant from the active site (Fig. 6B).

Unexpected results were obtained from six mutants (L74A, L61A, E862A, D867A, R869A, and DL870), which displayed increased enzymatic activity (defined by an increase in activity of more than one-half). From a molecular point of view, the increased activity of these mutants can be explained by alteration of the ligation pocket both in size and charge, making the ligation pocket more readily accessible to the RNA substrate. However, our data cannot explain the biological significance of the restricted conservation of these six residues in bacterial Pnkp and Hen1 for RNA repair, and further studies are required to address this issue. Suggested future studies, which are beyond the scope of the current study, include employment of other RNA substrates, use of a Pnkp-C/Hen1-N heterodimer that is non-thermophilic, and mutation and deletion studies on the full-length CthPnkp/Hen1.

Concluding Remarks. In this study, we present the crystal structures of the CthPnkp-C/Hen1-N/AMP ternary complex, the CthPnkp-C/Hen1-N heterodimer, and the CthHen1-N homodimer. Our structural studies reveal that specific interactions between CthHen1-N and CthPnkp-C lock the conformation of an otherwise flexible insertion module in CthPnkp-C, creating a ligation pocket to accommodate two strands of RNA for ligation. Our structures, together with the crystal structures of the ligase domain of CthPnkp recently determined by Shuman and coworkers (18), provide a complete picture of the structural and conformational changes accompanying the transition from the CthPnkp-C monomer and the CthHen1-N homodimer to the CthPnkp-C/Hen1-N heterodimer. Furthermore, the structure of the CthPnkp-C/Hen1-N heterodimer and the locations of the two linker regions, in particular, provide a glimpse of the likely architecture of the full-length Pnkp/Hen1 RNA repair complex.

Despite the unexpected increase in activity of some mutants, the functional analyses described in this study provide additional insight into the mechanism underlying activation of the ligase activity of Pnkp by Hen1. The step 2-specific assay clearly demonstrates that Hen1 activates the ligase activity of Pnkp starting with step 2. Mutation studies reveal that efficient RNA ligation requires direct involvement of several conserved residues from the insertion module. Finally, in addition to playing an essential structural role in the formation of a ligation pocket and in aligning the conserved residues from the insertion module for RNA ligation, the ligase-activating domain of Hen1 may provide its own residue(s) that might directly participate in RNA substrate
binding for the ligation, as exemplified by a fivefold reduction in activity with the R63A mutation.

Materials and Methods

Recombinant Proteins. Cloning, expression, and purification of various recombinant proteins are described in SI Materials and Methods.

Crystallography. Crystallization, data collection, and structural determination of the CthPnkp-C/Hen1-N/VAMP ternary complex, the CthPnkp-C/Hen1-N heterodimer, and the CthHen1-N homodimer are described in SI Materials and Methods and Table S1.

Enzymatic Assays. RNA ligation, step 2, and steps 2 + 3 reactions are described in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank P. Smith and S. Shuman for providing the coordinates of the ligase domain of CthPnkp before publication. We thank J. Brunzelle and Z. Li for help with data collection. This work was supported by National Science Foundation Grant MCB-0920966.

Fig. 6. Functional assays of step 2 and step 3 with the wild-type and mutant enzymes. (A) Denaturing polyacrylamide gel electrophoresis (DPAGE) analysis of step 2 carried out by CthHen1-N, CthPnkp-C, or a combination of both using either one-stranded RNA (3′-half-P) or two-stranded RNA (5′-half-P/3′-half-P) as the substrate. The presence of 2′,3′-cyclic phosphate at the 3′-end of 5′-half-P allows step 2 to occur but blocks step 3 by the lack of 3′-OH. Data in the bar graph are represented as means ± SD of three separate sets of experiments. (B) Surface of an expanded view of the ligation pocket showing the locations of 18 residues (except M1 of CthHen1-N, which is distant from the ligation pocket) selected for mutation and deletion analyses. Residues are colored as the same as those shown in Figs. 2 and S.A. (C–E) DPAGE analyses of step 2 (Top) and steps 2 + 3 (Middle) carried out by the wild-type and mutant enzymes. The RNA substrate for step 2 is 5′-half-P/3′-half-P, and the substrate for steps 2 + 3 is the 3′P-internally labeled cleaved RNA(3′-P/3′-P). The reactions were quantified based on the amount of radioactivity incorporated into 3′-half-P (step 2) or the percentage of the cleaved RNA(3′-P/3′-P) to be repaired (steps 2 + 3), followed by normalization to the activity of the wild-type enzyme, which is defined as 100%. The black bar represents the relative activity for step 2, and the white bar represents the relative activity for steps 2 + 3.


Supporting Information

Wang et al. 10.1073/pnas.1209805109

SI Materials and Methods

Protein Expression and Purification. DNA fragments corresponding to the C-terminal half of CthPnkp (residues 444–870) and the N-terminal half of CthHen1 (residues 1–230) were PCR-amplified from the genomic DNA of Clostridium thermocellum and inserted into the pETDuet-1 vector to create expression vectors for CthPnkp-C and CthHen1-N. After structural determination of the CthPnkpc/CthHen1-N heterodimer, the expression vectors for CthPnkpc-C∆L (residues 479–870) and CthHen1-N-ΔL (residues 1–199) were created with an attempt to define the minimum domains of the ligase complex. Whereas CthPnkpc-C∆L was expressed as a soluble protein in E. coli, CthHen1-N-ΔL was not. Because the CthPnkpc-C∆L/CthHen1-N heterodimer was more active, the expression vectors for single-alanine mutants were created based on the expression vectors of CthPnkpc-C∆L and CthHen1-N using the QuikChange method. The expression vectors for deleting mutants were created analogous to the original CthPnkpc-C∆L and CthHen1-N, except that the last and the first residues of CthPnkpc-C∆L and CthHen1-N were deleted, respectively. All proteins were expressed in E. coli strain BL21(DE3) and were purified to homogeneity using DEAE ion exchange, heparin affinity, Mono-Q ion exchange, and Superdex 200 size-exclusion chromatography. To prepare the CthPnkpc/CthHen1-N heterodimer for crystallization, the two proteins were mixed in equal molar, and the mixture was incubated at 4 °C for 1 h before the heterodimer was purified with Superdex 200 size-exclusion chromatography.

Crystallization, Data Collection, and Structure Determination. Crystals of the CthPnkpc-CthHen1-N heterodimer were grown by the hanging-drop vapor diffusion method at 30 °C. The purified CthPnkpc-CthHen1-N heterodimer (~6 mg/mL) was mixed with a reservoir solution containing 16% (wt/vol) PEG4000, 0.5 M NaCl, 25 mM CaCl2, 5% (vol/vol) methanol, 50 mM Hepes (pH 7.5), and 50 mM Tris-HCl (pH 8.0). Thin plate crystals appeared after 2–3 d. Crystals of the CthHen1-N homodimer were grown in a similar manner, except at 18 °C, and the reservoir solution contained 15% (wt/vol) PEG4000, 0.1 M NaCl, and 0.1 M Mes (pH 6.5). Crystals were soaked briefly in a cryoprotecting solution containing all of the components of the reservoir solution supplemented with 25% (vol/vol) glycerol, and the cryoextracted crystals were then mounted in a nylon loop and flash-frozen in liquid nitrogen. To obtain the crystal structure of the CthPnkpc/CthHen1-N/AMP ternary complex, a crystal of the CthPnkpc/CthHen1-N heterodimer was soaked with 5 mM ATP and 5 mM Mg2+ in the cryoprotecting solution overnight before crystal mounting and data collection. Data were collected at 21-ID beam lines at the Advanced Photon Source (APS) and processed using the HKL2000 program (1). Phase for the structure of the CthHen1-N homodimer was determined based on a SeMet SAD data using the Phenix program (2). A partial model was automatically built using Phenix. The remaining model was manually built using the Coot program (3). Refinement was carried out using Phenix. The structure of one of the monomers of CthHen1-N homodimer, together with the structure of the ligase domain of CthPnkpc (minus the insert module), was used as the initial search model for the phase of the CthPnkpc/CthHen1-N heterodimer. Two copies of each were found using the molecular replacement method in Phenix. The build of the model and the structural refinement of the CthPnkpc/CthHen1-N heterodimer and the CthPnkpc-CthHen1-N/AMP ternary complex were analogous to the procedures for the structure of the CthHen1-N homodimer. Statistics of the data collection and structural refinement are summarized in Table S1.

RNA Preparation. Unlabeled and 32P-internally radiolabeled cleaved tRNA\(^{\text{Npg}}\)ΔT were prepared as described previously (4).

To prepare RNA substrate for ligation reaction, the cleaved 32P-internally radiolabeled tRNA\(^{\text{Npg}}\)ΔT (100 μM) was treated with CthPnkpc-N (10 μM) at 45 °C for 45 min to phosphorylate the 5′-OH and dephosphorylate the 2′,3′-cyclic phosphate in a reaction buffer [25 mM Tris-HCl (pH 8.0), 50 mM KCl, 2.5 mM MgCl2, 0.05 mM EDTA, and 5 mM DTT] plus 0.5 mM ATP and 0.25 mM Mn2+. The processed RNAs were then purified by DPAGE. The purified phosphorylated 3′-half (3′-half-P*) and dephosphorylated 5′-half (5′-half*) were combined in equal molar in the reaction buffer minus DTT. The RNA was annealed by heating at 90 °C for 3 min, followed by slow cooling to 25 °C.

To prepare RNA substrates for step 2, the phosphorylation reaction described above was carried out with the unlabeled 3′-half tRNA\(^{\text{Npg}}\)ΔT. The phosphorylated 3′-half RNA (3′-half-P*) was then annealed with either Tris-EDTA (TE) buffer or an equal molar of the 32P-internally radiolabeled 5′-half RNA (5′-half-P*) to prepare two RNA substrates for step 2. 5′-half-P* served two purposes: (i) it annealed to 3′-half-P to produce an RNA substrate that mimics the ligase substrate (the 2′,3′-cyclic phosphate at the 3′-end of 5′-half-P* prevents the annealed two strands of RNA to be ligated); and (ii) it was used as the internal reference during quantification of step 2 on 3′-half-P.

The cleaved 32P-internally radiolabeled tRNA\(^{\text{Npg}}\)ΔT was annealed without any pretreatment for steps 2 + 3.

Assays for RNA Ligation, Step 2, and Steps 2 + 3. For all enzymatic assays, the required proteins were mixed and preincubated at 25 °C for 1 h before their use to promote formation of the heterodimer (as needed).

For RNA ligation assays, the annealed 5′-half*3′-half-P (1 μM) was incubated with the enzyme (0.5 μM) and 0.2 mM ATP in the reaction buffer at 45 °C for 30 min. The reaction was stopped with phenol extraction, followed by ethanol precipitation to recover RNA. The recovered RNA was dissolved in 20 μL of an equal volume mixture of TE buffer and DPAGE loading buffer, and 6 μL of the dissolved sample was analyzed by a 15% DPAGE. The radioactivity of RNA was quantified using a PhosphorImage system (Molecular Dynamics).

For step 2–specific assays, enzyme (1 μM) was incubated with 1 μCi of 32P-α-ATP in the reaction buffer at 45 °C for 10 min. ATP (0.1 mM) was then added, and the sample was incubated at 45 °C for another 10 min. An equal volume of the RNA substrate (2 μM) in the reaction buffer was mixed with the pretreated enzyme, and the mixture was incubated at 45 °C for 10 min to carry out step 2. The reaction was stopped with phenol extraction, followed by ethanol precipitation, and the recovered RNA was analyzed analogous to the protocols described in the RNA ligation assays.

For steps 2 + 3 assays, the enzyme–AMP complex intermediate was formed the same as in the assay for step 2, except incubation with 32P-α-ATP was omitted. To phosphorylate the 5′-end and dephosphorylate the 3′-end of the RNA substrate, the annealed 5′-half-P*3′-half* (2 μM) was incubated with CthPnkpc-N (2 μM) in the reaction buffer at 45 °C for 15 min in the presence of 0.1 mM ATP and 0.25 mM Mn2+. The pretreated enzyme and RNA substrate were then combined in equal volume, and the combined sample was incubated at 45 °C for
20 min to carry out steps 2 + 3. The reaction was stopped with phenol extraction followed by ethanol precipitation, and the recovered RNA was analyzed in a manner similar to the protocols described in the RNA ligation assays.


**Fig. S1.** Cloverleaf view of *E. coli* tRNA^{Arg}\textsubscript{ΔT} with the TΨC stem-loop deleted (tRNA^{Arg-ΔT}). The site to be cleaved by colicin D is marked with an arrow.
Fig. S2. Conservation of the N-terminal half of Hen1 in bacteria. Amino acid sequences of the N-terminal half of 10 representative bacterial Hen1 were aligned. The sequences were selected based on the scores of BLASTp of CthHen1 against bacterial genomes at National Center for Biotechnology Information (NCBI) to represent a broad spectrum of bacterial Hen1. The conserved residues are boxed in color, with completely conserved residues in magenta, identical residues in yellow, and similar residues in cyan. Residue number over the alignment corresponds to CthHen1-N. The secondary structure of CthHen1-N is depicted above the primary sequence, with α-helices highlighted as cylinders, β-strands as arrows, loops as solid lines, and disordered residues as dotted lines. The structure of CthHen1-N is portrayed as a left hand (Fig. 3A), and the regions corresponding to the thumb and fingers are underlined and labeled. Ava, Anabaena variabilis; Bsp, Bradyrhizobium sp. BTAi1; Cmi, Clavibacter michiganensis; Cth, C. thermocellum; Dma, Deinococcus marinispolis; Fal, Frankia alni; Hau, Herpetosiphon aurantiacus; H/H, CthHen1-N homodimer; Nmu, Nakamurella multipartita; P/H, CthPnkp-C/Hen1-N heterodimer; Sco, Streptomyces coelicolor; Sve, Streptomyces venezuelae.

Fig. S3. Structural comparison of CthHen1-N with the top-scoring hit of a Dali search. Stereoview of Cα superposition of the aligned structures of CthHen1-N and a KH domain of ZBP1 (PDB ID code 3KRM). The structure of CthHen1-N is colored the same as shown in Fig. 2, and the structure of the KH domain is colored blue.
Fig. S4. Conservation of the C-terminal half of Pnkp in bacteria. Amino acid sequences of the C-terminal half of 10 representative bacterial Pnkp were aligned. The organisms and their orders are the same as those for Hen1-N shown in Fig. S1. Residue number over the alignment corresponds to CthPnkp-C. The NTase motifs conserved in DNA ligases, RNA ligases, and RNA-capping enzymes are underlined and labeled.
Fig. S5. Structures of the C-addition module of CthPnkp-C. Cα superposition of the aligned structures of the C-addition module of CthPnkp-C in the same orientation as the one shown in Fig. 2B. The structure from the CthPnkp-C/Hen1-N heterodimer is colored cyan, whereas the structures from CthPnkp-C·ATP and CthPnkp-C/AMP are in black. The side chains of eight conserved residues either near or in the ligation pocket (Y798, K824, E862, D867, and L870), interacting with CthHen1-N (D806, R815), or both (R869), are highlighted in stick. The side chains of residues from the CthPnkp-C/Hen1-N heterodimer are colored orange, with heteroatoms colored individually (nitrogen in blue and oxygen in red). The side chains of residues from CthPnkp-C·ATP and CthPnkp-C/AMP are in black. Six residues near or in the ligation pocket were selected for mutation and deletion studies as shown in Fig. 6E.

Fig. S6. Structural changes accompanying the adenylation reaction carried out in a crystal of the CthPnkp-C/Hen1-N heterodimer. (A and B) Structural details of the active site without (A) and with (B) a crystal of the CthPnkp-C/Hen1 heterodimer soaked with ATP and Mg^{2+} before data collection. The structures are colored and depicted in a manner similar to those shown in Fig. 4. Only the AMP and the side chains of residues of significant conformational changes are shown in stick. An Mg^{2+} ion and five coordinated water molecules are depicted in spheres and colored gray and red, respectively. Mg^{2+} and two of the five water molecules are not labeled for clarity. The simulated annealing omit map covering AMP, Mg^{2+}, and five water molecules is contoured at 1.5 σ.
Table S1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>CthPnkp-C/Hen1-N/AMP</th>
<th>CthPnkp-C/Hen1-N</th>
<th>CthHen1-N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space Group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>59.6, 59.8, 101.5</td>
<td>59.7, 60.0, 101.7</td>
<td>68.5, 68.5, 165.7</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>81.4, 87.8, 88.8</td>
<td>81.2, 87.6, 88.3</td>
<td>90.0, 90.0, 120.0</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0–2.4</td>
<td>50.0–2.6</td>
<td>50.0–2.3</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9787</td>
<td>0.9787</td>
<td>0.9787</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.7</td>
<td>3.8</td>
<td>7.5</td>
</tr>
<tr>
<td>Completeness (%)*</td>
<td>89.7 (79.7)</td>
<td>98.0 (94.2)</td>
<td>98.7 (100)</td>
</tr>
<tr>
<td>Average I/σ</td>
<td>14.8 (2.9)</td>
<td>11.8 (2.5)</td>
<td>14.9 (2.8)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>5.2 (19.7)</td>
<td>5.6 (23.0)</td>
<td>6.1 (20.4)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>34.6–2.4</td>
<td>42.4–2.6</td>
<td>40.4–2.3</td>
</tr>
<tr>
<td>No. reflections</td>
<td>49,244</td>
<td>41,067</td>
<td>19,215</td>
</tr>
<tr>
<td>Rwork/Rfree (%)</td>
<td>16.8/22.2</td>
<td>18.5/24.5</td>
<td>20.4/25.7</td>
</tr>
<tr>
<td>No. atoms</td>
<td>10,530</td>
<td>10,046</td>
<td>2,796</td>
</tr>
<tr>
<td>Protein</td>
<td>10,054</td>
<td>9,886</td>
<td>2,734</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>46</td>
<td>6</td>
<td>N/A</td>
</tr>
<tr>
<td>Water</td>
<td>430</td>
<td>154</td>
<td>62</td>
</tr>
<tr>
<td>Average B factor, Å²</td>
<td>36.9</td>
<td>48.6</td>
<td>56.3</td>
</tr>
<tr>
<td>rmsd values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.08</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.08</td>
<td>1.28</td>
<td>1.38</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>97.3</td>
<td>96.5</td>
<td>97.2</td>
</tr>
<tr>
<td>Outliners (%)</td>
<td>0.3</td>
<td>0.7</td>
<td>0.3</td>
</tr>
</tbody>
</table>

*Highest resolution shell is shown in parenthesis.