10554–10566 *Nucleic Acids Research, 2012, Vol. 40, No. 20* *Published online 31 August 2012*
doi:10.1093/nar/gks826

X-ray structure of the fourth type of archaeal tRNA splicing endonuclease: insights into the evolution of a novel three-unit composition and a unique loop involved in broad substrate specificity

Akira Hirata[1], Kosuke Fujishima[2,3], Ryota Yamagami[1], Takuya Kawamura[1], Jillian F. Banfield[4,5], Akio Kanai[2] and Hiroyuki Hori[1,6,*]

[1]Department of Materials Science and Biotechnology, Graduate School of Science and Engineering, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime 790-8577, [2]Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan, [3]NASA Ames Research Center, Moffett Field, CA 94035, [4]Department of Earth and Planetary Science, University of California, Berkeley, CA 94270, [5]Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA and [6]Venture Business Laboratory, Ehime University, Bunkyo 3, Matsuyama, Ehime 790-8577, Japan

Received July 23, 2012; Revised and Accepted August 8, 2012

ABSTRACT

Cleavage of introns from precursor transfer RNAs (tRNAs) by tRNA splicing endonuclease (EndA) is essential for tRNA maturation in Archaea and Eukarya. In the past, archaeal EndAs were classified into three types (α'_2, α_4 and $\alpha_2\beta_2$) according to subunit composition. Recently, we have identified a fourth type of archaeal EndA from an uncultivated archaeon *Candidatus Micrarchaeum acidiphilum*, referred to as ARMAN-2, which is deeply branched within Euryarchaea. The ARMAN-2 EndA forms an ε_2 homodimer and has broad substrate specificity like the $\alpha_2\beta_2$ type EndAs found in Crenarchaea and Nanoarchaea. However, the precise architecture of ARMAN-2 EndA was unknown. Here, we report the crystal structure of the ε_2 homodimer of ARMAN-2 EndA. The structure reveals that the ε protomer is separated into three novel units (α^N, α and β^C) fused by two distinct linkers, although the overall structure of ARMAN-2 EndA is similar to those of the other three types of archaeal EndAs. Structural comparison and mutational analyses reveal that an ARMAN-2 type-specific loop (ASL) is involved in the broad substrate specificity and that K161 in the ASL functions as the RNA recognition site. These findings suggest that the broad substrate specificities of ε_2 and $\alpha_2\beta_2$ EndAs were separately acquired through different evolutionary processes.

INTRODUCTION

Transfer RNA (tRNA) is an adapter molecule that acts as a translator of genetic information from nucleotide sequence of messenger RNA to amino acid sequence of protein. Because tRNA needs to go through a maturation process in order to synthesize proteins correctly and smoothly, tRNA maturation is essential for life. tRNA splicing, which removes introns and joins exons in precursor (pre)-tRNA, is an important process in tRNA maturation.

Many interruptions of pre-tRNA with introns have been found in all three domains of life. In Eukarya, most introns are predominantly located in the canonical position between nucleotide positions 37 and 38 in the anticodon loop of tRNA, while archaeal introns are located not only in the canonical position but also in various non-canonical positions including the D-and T-loops, the variable region and the aminoacyl stem (1). In some cases, single archaeal pre-tRNAs include two or three introns, called multiple-introns, in non-canonical positions (2,3). The introns in eukaryotic cytoplasmic and archaeal pre-tRNA are removed by a tRNA splicing endonuclease (EndA) (4–6). The eukaryotic EndA consists of four subunits (SEN2, SEN15, SEN34 and SEN54) (7,8). In contrast, archaeal EndAs are classified into three types by subunit composition, namely, homotetramer (α_4), homodimer (α'_2) and heterotetramer ($\alpha_2\beta_2$) [(9) and Figure 1]. Figure 1 shows the structures and characteristics of the three types of archaeal EndAs. The α subunit in the archaeal EndAs is a catalytic subunit and

*To whom correspondence should be addressed. Tel: +81 89 927 8548; Fax: +81 89 927 9941; Email: hori@eng.ehime-u.ac.jp

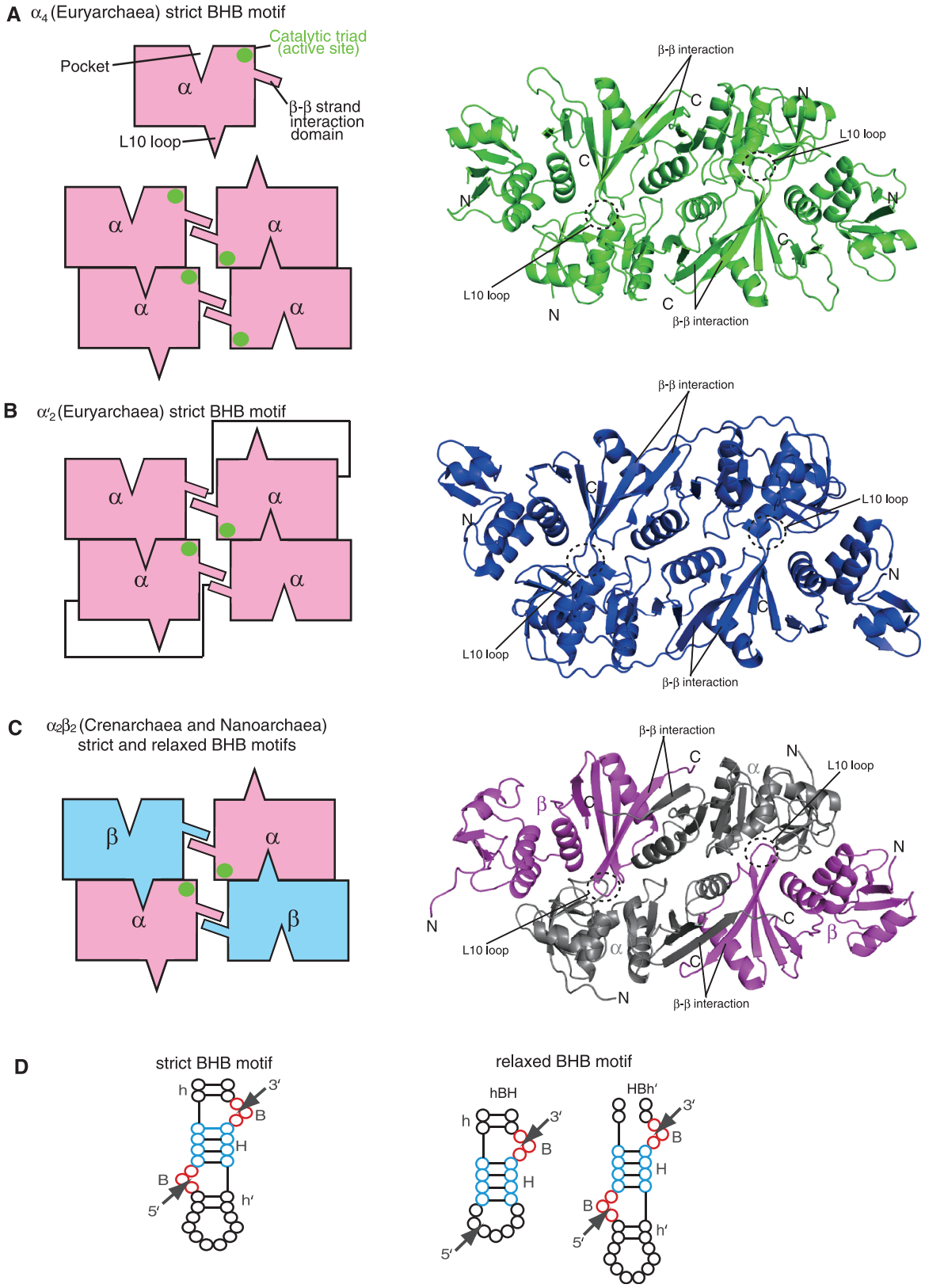

Figure 1. Structures and characteristics of three types of archaeal EndAs. The subunit interactions are represented by cartoon models on the left side. The β–β interaction responsible for inter/intraunit formation, the L10 loop and pocket responsible for dimer/tetramer formation are highlighted. The catalytic triads are marked by green circles. The right panels show the ribbon models of EndAs. The sources of EndAs are shown in parentheses and the substrate specificities of EndAs are shown next to the parentheses: (**A**) α_4 type MJ-EndA; (**B**) α'_2 type AFU-EndA; (**C**) $\alpha_2\beta_2$ type APE-EndA. Full names of the archaea species are as follows: MJ, *Methanocaldococcus jannaschii*; AFU, *Archaeoglobus fulgidus*; APE, *Aeropyrum pernix*. The secondary structure diagrams of substrate RNA motifs. The splicing sites are indicated using arrows. B and H represent bulge and helix, respectively. The h and h′ indicate the helices close to the 3-nt bulge on the exonic side and intronic side, respectively. (**D**) Left, strict BHB motif; Right, relaxed BHB motifs (hBH and HBh′).

shares homology with SEN2 and SEN34 subunits of eukaryotic EndA, implying a common evolutionary origin between the eukaryotic and archaeal EndAs (7,10). Both the eukaryotic and archaeal EndAs are proposed to use a similar cleavage chemistry to that of ribonuclease A (6,11). The archaeal α subunit has a catalytic triad, L-10 loop and pocket (Figure 1). In the case of α'_2 EndAs, the α subunit contains two α units joined by a polypeptide linker (Figure 1B). In Figure 1, the locations of the catalytic triad in archaeal EndAs are marked by green circles. The negatively-charged L-10 loop and positively-charged pocket contribute to subunit interaction and are conserved in the three types of archaeal EndAs. However, the substrate recognition mechanisms of EndAs are different to some extent. The eukaryotic EndA requires the mature domain of pre-tRNA for the recognition of cleavage sites in the canonical position (12), although the three types of archaeal EndAs can remove introns with a bulge–helix–bulge (BHB) motif irrespective of the existence of the pre-tRNA mature domain. Furthermore, the $\alpha_2\beta_2$ type of archaeal EndA possesses a broad substrate specificity that recognizes relaxed BHB motifs of various lengths and disruption of either the 5′- or 3′-bulge in the BHB (so-called HBh′ and hBH) as well as the strict BHB motif (Figure 1C and D). In contrast, the α'_2 and α_4 type EndAs recognize only the strict BHB motif (13–19). The HBh′ and hBH motifs are often found at non-canonical positions in introns of pre-tRNAs from Crenarchaea and Nanoarchaea, consistent with the possession of the $\alpha_2\beta_2$ type EndA (3). Some of these pre-tRNAs are spliced into two or three gene fragments at different loci and are called split or tri-split tRNAs, respectively (20,21). Furthermore, permutated tRNA, in which the 5′ and 3′ halves of the coding sequences separated by intervening elements have their positions switched, has been discovered in some genera of Crenarchaea (22). Only the $\alpha_2\beta_2$ type of archaeal EndA with broad substrate specificity has the ability to excise non-canonical introns, suggesting the coevolution of disrupted tRNA gene diversity and EndA architecture (16,21). In Crenarchaeal EndAs, the Crenarchaea-specific loop (CSL) is conserved in the catalytic α subunit (19). In our previous study, it was revealed that the CSL is responsible for the broad substrate specificity and that a conserved Lys residue in the CSL functions as the substrate recognition site (23).

Recently, we found a fourth type of archaeal EndA from an uncultivated archaeon *Candidatus Micrarchaeum acidiphilum*, referred to as ARMAN (Archaeal Richmond Mine Acidphilic Nanoorganism)-2 (24), which was discovered in an acid mine drainage site at Iron Mountain in Northern California (25). Our biochemical and bioinformatic analyses have led us to propose that the ARMAN-2 EndA has a novel three-unit architecture that consists of two duplicated catalytic α units and one structural β-unit encoded on a single gene (24). Our cross-linking analysis showed that two three-unit protomers are assembled into the functional ε_2, where ε represents the union of three units (α^P–α–β) (24). The amino acid sequences of the α- and β-units (127 and 97 amino acids, respectively) are similar to those of the catalytic α and structural β subunits in the other three types (α'_2, α_4 and $\alpha_2\beta_2$) of archaeal EndAs. In contrast, the α^P unit (163 amino acids) is a pseudo-catalytic unit since three residues (His, Tyr and Lys) comprising the catalytic triad and positively-charged residues responsible for dimer formation are mutated. The question therefore arises as to how the α^P unit interacts with the α and β units in the ε_2 architecture? The precise architecture of three-unit interactions will provide new insights into the molecular evolution of archaeal EndA. Furthermore, remarkably, the ARMAN-2 EndA possesses a broad substrate specificity that cleaves introns with both strict and relaxed BHB motifs despite lacking the CSL region. What structural properties of ARMAN-2 EndA confer the broad substrate specificity? Structural determination of ARMAN-2 EndA is necessary to address these issues. We present herein an X-ray crystal structure of ARMAN-2 EndA, demonstrating a novel three-unit arrangement of the ε_2 homodimeric complex. Our structural comparison of ARMAN-2 (ε_2) and the other three types (α'_2, α_4 and $\alpha_2\beta_2$) of archaeal EndAs shows that the ARMAN-2 EndA possesses an ARMAN-2 type-specific loop (ASL). Our structure-guided mutagenesis study identified the catalytic residues and revealed that the ASL is responsible for the broad substrate specificity. Furthermore, our study suggests that the Lys residue in the ASL plays the same role as the Lys residue in the CSL for the broad substrate recognition and that the ASL has been acquired by a distinctly independent evolutionary pathway to the CSL.

MATERIALS AND METHODS

Protein expression and purification

A pET-23 b vector (Novagen) harboring an ARMAN-2 EndA gene attached to a 6× His tag at its C-terminus has been previously constructed (24). The plasmid was used for overexpressing the recombinant ARMAN-2 EndA in *Escherichia coli* Rosetta 2(DE3) strain (Novagen). *Escherichia coli* cells harboring the plasmid were grown in LB media supplemented with 100 μg/ml of ampicillin at 37°C, and then isopropylthio–β-galactoside (IPTG) was added to a final concentration of 0.5 mM when the cells density reached OD_{600} = ~0.8. After cultivation at 20°C for 24 h, the cells were harvested by centrifugation (6000 rpm at 4°C for 20 min). The cells were suspended in 15 ml buffer A [20 mM Tris–HCl (pH 7.6), 200 mM KCl, 20 mM imidazole 10 mM 2-mercaptoethanol and 5% glycerol] supplemented with protease inhibitor cocktail (Roche) and then disrupted with an ultrasonic disruptor (model VCX-500, Sonics & Materials., Inc., USA). A fraction of *E. coli* proteins was denatured by heat treatment at 50°C for 20 min and removed by centrifugation (18 000 rpm at 4°C for 20 min). The supernatant was loaded onto a Ni–NTA Superflow column (Qiagen) equilibrated with buffer A and then the enzyme was eluted by buffer A containing 500 mM imidazole. The eluted fractions were collected and then loaded onto a HiTrap Heparin–Sepharose column (GE Healthcare) equilibrated with buffer B [20 mM Tris–HCl (pH 7.6), 50 mM KCl, 10 mM

2-mercaptoethanol and 5% glycerol]. The bound protein was eluted by a linear gradient of buffer B from 50 mM to 1 M KCl. The eluted fractions were collected and then concentrated to ~3 ml volume using Amicon Ultra-15 centrifugal filter units. Finally, the concentrated protein was applied to a HiLoad 16/60 Superdex 75 pg column (GE Healthcare) equilibrated with buffer C [20 mM Tris–HCl (pH 7.6), 700 mM NaCl, 10 mM 2-mercaptoethanol and 5% glycerol]. The single peak fractions were collected. Mutant genes were generated using the QuickChange site-directed mutagenesis kit (Stratagene), and the mutations were verified by DNA sequencing. Mutant proteins were expressed and purified in the same manner as the wild-type protein. The recombinant *Archaeaoglobus fulgidus* (AFU)-EndA and its chimera mutants were prepared as reported previously (23). The protein purities were confirmed by SDS–PAGE (Supplementary Figure S1).

Crystallization

The single-peak fractions from the Superdex-75 gel-filtration column were pooled and then concentrated to ~10 mg/ml using Vivaspin 15R centrifugal filter units (Sartorius stedim biotech). Initial trials for crystallization of the ARMAN-2 EndA were performed by the hanging-drop vapor diffusion method using a Crystal Screening Kit (Hampton Research). The drop solution was equilibrated against 200 µl of reservoir solution at 22°C. A few crystals were obtained under some of the tested conditions which contained PEG 3350 as the precipitant. Based on the initial crystallization conditions, we then searched for optimum conditions. When the ARMAN-2 EndA protein solution was mixed with an equal volume of a crystallization solution that contained 18% PEG3350 and 0.2 M tri-ammonium citrate (pH 7.0), crystals grew within 5 days at 22°C producing full-sized rectangular-shaped ($200 \times 100 \times 100$ µm) crystals. For the experimental phase determination by single-wavelength anomalous dispersion (SAD) method, the crystal was soaked in mother liquor supplemented with 0.4 mM $KPtCl_4$ at 22°C for 16 h. Cryo-protection of the native and Pt-induced crystals was achieved by stepwise transfer to the respective artificial mother liquor containing 25% glycerol. The crystals were then flash-frozen in liquid nitrogen.

Data collection and structure determination

X-ray diffraction data sets from native crystals ($\lambda = 1.0000$) and SAD data sets from Pt-induced crystals ($\lambda = 1.0717$) were collected at 100 K on the BL38B1 beamline at SPring-8 (Hyogo, Japan). All data sets were processed, merged and scaled using the HKL2000 program (26). Using the deduced Pt-SAD data set, all 19 Pt positions were identified and refined in the orthorhombic space group $P2_12_12_1$, and the initial phase was calculated by using AutoSol in PHENIX (27), followed by automated model building using RESOLVE (28). The resulting map and partial model were used for manually building the model using COOT (29). The model was further refined by using PHENIX (27). Using the native data set and the refined model as a search coordinate, the structure of the ARMAN-2 EndA was determined by molecular replacement with the Phaser program (30). The model was further manually built with COOT (29) and refined with PHENIX (27). The structure of ARMAN-2 EndA was refined to R_{work}/R_{free} of 21.8%/25.7% at 2.25 Å resolution (Table 1). The space group of the crystal belonged to $P3_2$, where two ARMAN-2 EndA molecules are present in an asymmetric unit. The final model contained residues 2-387 (chain A and B) and 152 water molecules. The final model of the ARMAN-2 EndA structure was further checked using PROCHECK (31), showing the quality of the refined model. Ramachandran plots (%) of the ARMAN-2 EndA structure are tabulated in Table 1. The structure factor and coordinates have been deposited in the Protein Data Bank (PDB code 4FZ2). All structural figures were generated by PyMOL (DeLano Scientific, Palo Alto, CA).

Intron-cleavage assay by the splicing endonuclease

The transcripts of ARMAN-2 pre-$tRNA^{Ile}$ (UAU) and pre-$tRNA^{Cys}$ (GCA) were prepared using T7 RNA polymerase as described in our previous report (24). Splicing reactions were performed as follows. 1.0 µg EndA was mixed with 0.2 nmol transcripts in 50 µl buffer D [50 mM

Table 1. Data collection and refinement statistics

	ARMAN-2 EndA	ARMAN-2 EndA Pt-derivative
Data collection		
Space group	$P3_2$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	112.02, 112.02, 81.08	75.62, 85.17, 140,19
α,β,γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	50 to 2.25 (2.33–2.25)	50 to 2.05 (2.07–2.00)
R_{merge}[a]	6.5 (49.2)	5.6 (35.1)
$I / \sigma I$	39.2 (4.7)	12.8 (10.6)
Completeness (%)	99.4 (95.4)	99.7 (98.6)
Redundancy	5.9 (5.5)	12.9 (11.7)
Refinement		
Resolution (Å)	37.4–2.25	
No. reflections	49 748	
R_{work}[b]/R_{free}[c]	21.8/ 25.7	
No. atoms	6302	
Protein	6454	
Water	152	
Avg. *B*-factors (Å^2)	51.5	
R.m.s.d.		
Bond lengths (Å)	0.006	
Bond angles (°)	1.0	
Ramachandran plot (%)		
Most favored	90.2	
Additional allowed	9.1	
Generously allowed	0.7	
Disallowed	0.0	

The value in the parentheses is for the highest resolution shell.
[a]$R_{merge} = \Sigma\Sigma_j| < I(h) > - I(h)_j|/\Sigma\Sigma_j| < I(h) > |$, where $< I(h) >$ is the mean intensity of symmetry-equivalent reflections.
[b]$R_{work} = \Sigma$ (IIF_p(obs) − F_p(calc)II)/ΣIF_p(obs)I.
[c]R_{free} = *R*-factor for a selected subset (10%) of reflections that was not included in earlier refinement calculations.

Tris–HCl (pH 7.6), 5 mM $MgCl_2$, 6 mM 2-mercaptoethanoland 50 mM KCl] and incubated at 50°C. Aliquots (10 µl) were removed at 0, 10, 30 and 60 min and were analyzed by 15% PAGE/7 M urea. The gel was stained with 0.05% toluidine blue.

RESULTS AND DISCUSSION

Overall structure

We initially crystallized an ARMAN-2 EndA to obtain structural information. Two different space groups were found in different ARMAN-2 EndA crystals under the same crystallization conditions. One crystal belonged to the orthorhombic space group $P2_12_12_1$, whereas the other belonged to the trigonal space group $P3_2$. In this study, we determined the structure of ARMAN-2 EndA from the latter crystal at 2.25 Å resolution (Figure 2A). Although the structure from the former crystal could be solved at 2.00 Å resolution, it exhibited many disordered regions due to the effects of crystal packing (data not shown). The final model of ARMAN-2 EndA contains two molecules per asymmetric unit. The two molecules are structurally almost identical (R-factor = 21.8 and R_{free}-factor = 25.7 in Table 1).

The ARMAN-2 EndA is composed of two ε protomers producing a homodimeric subunit structure, ε_2 (Figure 2A). The overall shape of the ε_2 homodimer structure is like a rectangular parallelepiped. The ε protomer consists of 10 α helices and 23 β strands (Figure 2B and C). Furthermore, the structure can be separated into three units, the α^N unit (2–97 residues; orange), the α unit (126–288 residues; pink) and the β^C unit (301–387 residues; cyan). The three units are connected by two linkers, linker 1 (98–125 residues; black) and linker 2 (289–300 residues; black): linker 1 connects the α^N and α units, whereas linker 2 connects the α and β^C units. The α^N unit (orange) is composed of a mixed anti-parallel and parallel β sheet (β1–β5), four α helices (α1–α4) and one β strand (β6). The β6 strand of the α^N unit, the β7 strand of linker 1 and five β strands (β18, β19, β21, β22 and β23) of the β^C unit participate in forming one mixed anti-parallel and parallel β sheet. This β7−β23 interaction probably prevents structural fluctuation of linker 1. The β^C unit (cyan) consists of one β sheet (β18, β19, β21, β22 and β23), one α helix (α10) and one β strand (β20). The β sheet is structurally sandwiched by two α helices (α4 and α10), thereby stabilizing the unit interaction between the α^N and β^C units. Furthermore, the β20 and β23 strands interact with the β16 and β17 strands in the α unit, respectively. These two anti-parallel β sheets connect the β^C and α units. The α unit (pink) can be separated into two subdomains, the N-terminus and C-terminus. The N-terminal subdomain is composed of a mixed anti-parallel and parallel β sheet (β8–β12 and three α helices (α5−α7), and the C-terminal subdomain is composed of a mixed anti-parallel and parallel β sheet (β13–β15 and β17), two α helices (α8−α9) and one β strand (β16). The five α helices (α5−α9) are placed around the two β sheets. Thus, these intra-unit interactions probably contribute to maintenance of the structural integrity of ARMAN-2 EndA. The configuration of secondary structures in the α^N and β^C unit overlaps with that of the α unit (Figure 2D). Thus, this configuration is commonly observed in the α and β subunits of the $\alpha_2\beta_2$ EndAs (19,23,32). Furthermore, our structure-based sequence alignment analysis has shown that the overlap region of the α^N, α and β^C units is found in the N-terminal subdomain of the α subunit in the α'_2 type EndAs , the entire domain of the α subunit in the $\alpha_2\beta_2$ type EndAs and the C-terminal subdomain of the β subunit in the $\alpha_2\beta_2$ type EndAs (Supplementary Figure S2), although the β^C unit includes an amino acid sequence (373–387 in ARMAN-2 EndA), which is found in the α subunit instead of the β subunit in the case of Crenarchaeal $\alpha_2\beta_2$ type EndAs (Supplementary Figure S2B). Based on these structural observations, we have redefined the fourth type of archaeal ε_2 EndA, where ε is three-units (α^N–α–β^C).

Structural comparison with three types of archaeal EndAs

Our current structural study clarified the ε_2 subunit structure of ARMAN-2 EndA (Figure 3). The architecture of the three units and subunit interactions were far beyond our previous expectations because two long linkers connect the three units in ARMAN-2 EndA. This architecture is not observed in the other three types of EndAs (Figure 1). Nevertheless, the overall shape and size of the ε_2 structure of ARMAN-2 EndA is very similar to those of the other three types of archaeal EndAs (Figure 1 and Supplementary Figure S3). In addition to the three types of archaeal EndAs, a structural homology search by the Dali sever (33) confirms that the structure of β^C and α units in ARMAN-2 EndA is homologous to that of a subunit (SEN15) of human EndA and that of prokaryotic DNA restriction enzymes.

As shown in Figure 1 and 3, two β−β strand interactions at the domain interface are conserved in all four types of archaeal EndAs (24). The interactions are shown to be responsible for intra/interunit interactions such as the α–α subunit assembly in the α_4 type EndA, the α–α domain assembly in the α'_2 type EndA and the α–β subunit assembly in the $\alpha_2\beta_2$ type EndA (11,19,32). However, in the case of ε_2 ARMAN-2 EndA, the β−β strand (β22–β23) interaction does not directly contribute to the interaction between the α and β^C units since the β22 and β23 strands are parts of the C-terminal β^C unit (Figures 2C and 3A). Instead, two anti-parallel β strand interactions (β20–β16 and β23–β17) connect the α and β^C units (Figure 2C). Furthermore, the β6–β18 strand interaction appears to contribute to the assembly of α^N and β^C units together with formation of a sandwich by the β sheet of the β^C unit and two α helices (α4 and α10). Because the three-unit architecture is connected by two linkers, the linkers enable it to easily form a complete ε protomer as compared to the α–α subunit assembly in α_4 type EndA and the α–β subunit assembly in $\alpha_2\beta_2$ type EndAs. Therefore, the linkers play an important role in the three-unit architecture. In contrast, as previously expected from our bioinformatics study (24), the three-unit molecule assembles with another through the interaction of a negatively-charged L10 loop with a

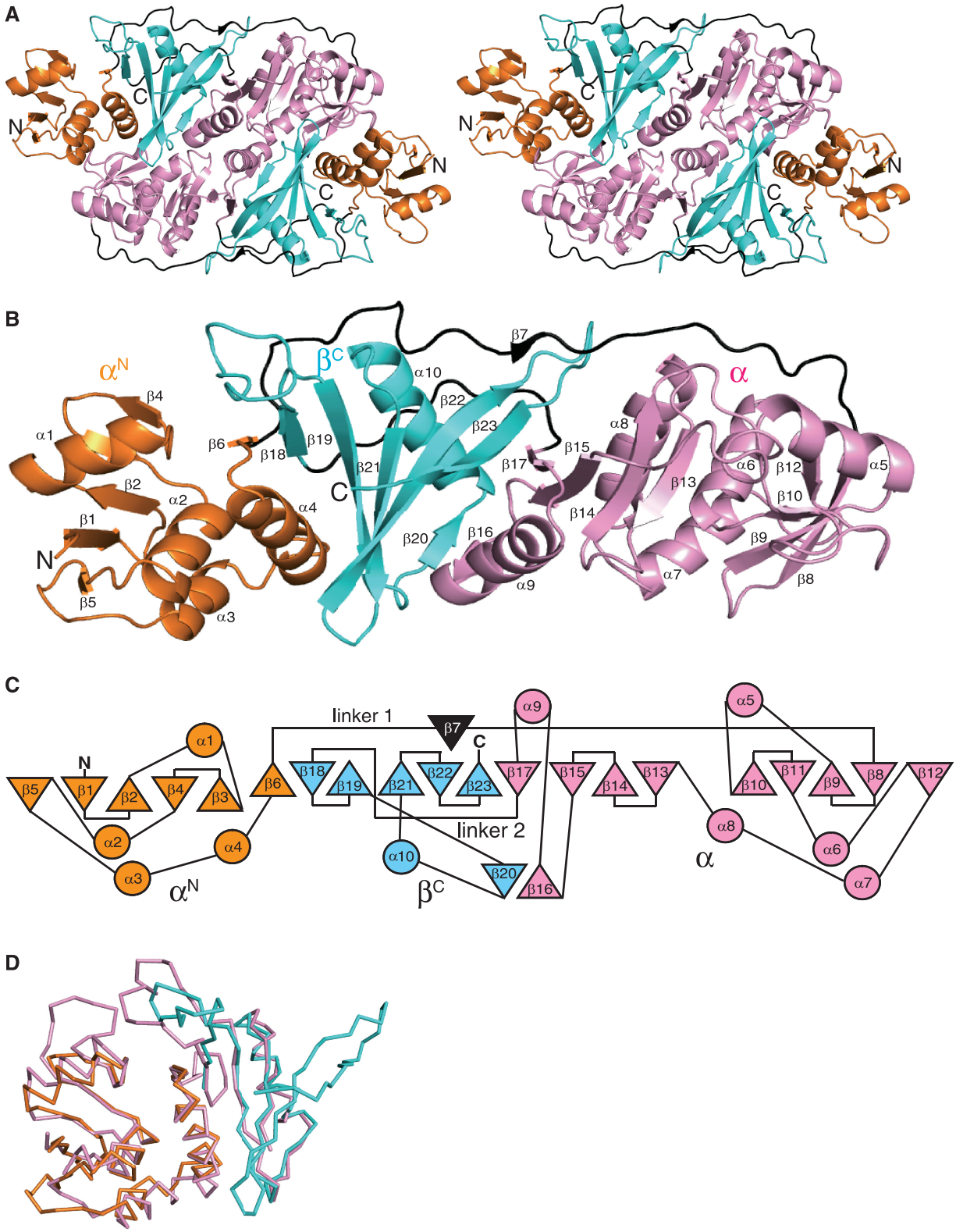

Figure 2. Crystal structure of ARMAN-2 EndA. (**A**) Ribbon stereo diagram of the overall structure of the functional ε_2 homodimeric complex, where ε stands for the union of three units (α^N–α–β^C). The α^N unit, α unit, β^C unit and two linker regions are colored orange, pink, cyan and black, respectively. The N- and C-terminal ends are labeled as N and C, respectively. (**B**) Ribbon diagram of the ε protomer. The secondary structures of the α helix and β strand are labeled (in order) as the α and β, respectively. The α^N unit, α unit, β^C unit and two linker regions are colored as described above, respectively. (**C**) A secondary structure topology diagram of the ε protomer. The α helices and β strands are represented by circles and triangles, respectively. The α^N, α and β^C units are colored as in Figure 2A and B. (**D**) Superimposition diagram of Cα atoms of the α^N (orange) and β^C (cyan) units onto that of the α unit (pink) of ARMAN-2 EndA.

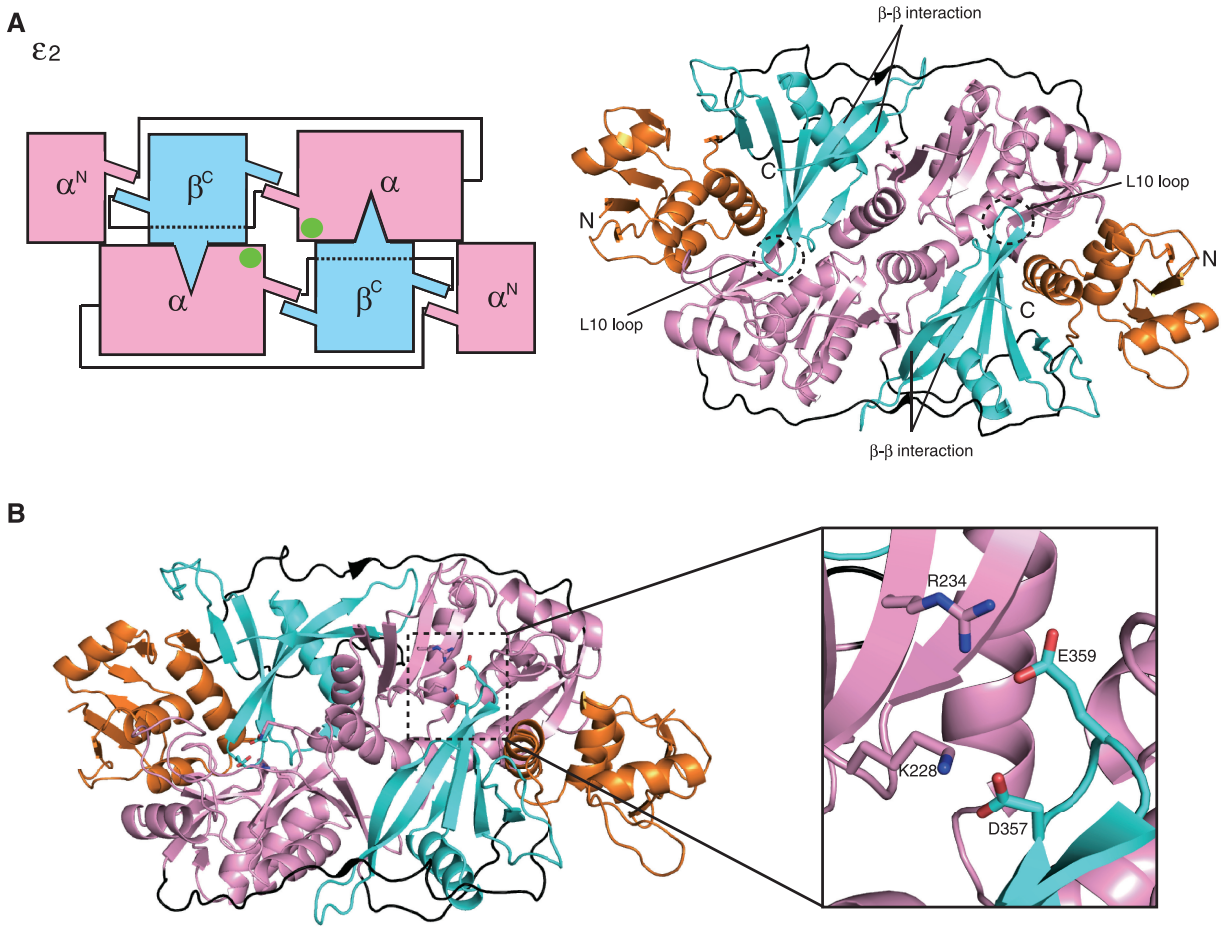

Figure 3. Structural properties of ARMAN-2 EndA (**A**) The structure of ARMAN-2 EndA is shown in the same way as in Figure 1. Left, cartoon representation of the structural model of the ε_2 type ARMAN-2 EndA. Right, ribbon diagram of the ε_2 type ARMAN-2 EndA. The α^N unit, α unit, β^C unit and two linker regions are colored as in Figure 2. (**B**) Close-up view of electrostatic interaction between positively-charged pocket and negatively-charged L10 loop responsible for dimer formation in ARMAN-2 EndA. Two salt bridges (K228-D357 and K234-E359) are highlighted as stick models.

positively-charged pocket of the α unit in the opposing three-unit molecule (Figure 3B). These electrostatic interactions are observed in other EndA structures (Figure 1), suggesting the structural and/or functional importance of these interactions. Figure 3B shows the molecular interaction of the L10 loop and positively-charged pocket in the ARMAN-2 EndA. Two salt-bridge interactions (D357–K228 and E359–R234) are observed. The positively and negatively-charged amino acid residues are conserved in almost all EndAs as reported previously (24). In fact, our previous mutagenesis study has shown that the D357A mutant of ARMAN-2 EndA barely cleaves the introns from pre-tRNA^{Ile} and pre-tRNA^{Cys} (24), suggesting that the salt-bridge interaction (D357–K228) is required for the formation of functional ε_2 homodimer of ARMAN-2 EndA. Although we could not observe the dissociation of ε_2 homodimer into ε protomer in the D357A mutant under our cross-linking analysis (24), the breakage of the salt-bridge interaction (D357–K228) probably induce the conformational change of the side chain of K228. In the ARMAN-2 EndA structure, the side chain of K228 is located close to that of R275, which is the putative RNA recognition residue as described in the next section, at the 4.2 Å distance between the K228 Cγ and R275 Nη2. Thus, the conformation of the side chain of R275 can also be changed by the electrostatic repulsion between the K228 and R275 in the D357A mutant, resulting in the loss of intron-cleavage activity.

The active site

It has been reported that three catalytic resides (tyrosine, histidine and lysine) as well as two substrate recognition residues (two argentines) are conserved in the α subunit of the EndA from Euryarchaea (6,11). Our structural study and amino acid sequence alignment strongly suggest that the Y236, H251 and K282 residues are the catalytic residues and that the R275 and W384 are the possible substrate recognition residues in the case of the ARMAN-2 EndA (Supplementary Figure S2B). These five residues are located on the enzyme surface around the expected catalytic pocket (Figure 4A) and can be arranged at similar locations to that of their counterpart residues in the three types of archaeal EndAs (Supplementary Figure S3). To clarify whether the catalytic triad of ARMAN-2 EndA is indeed formed

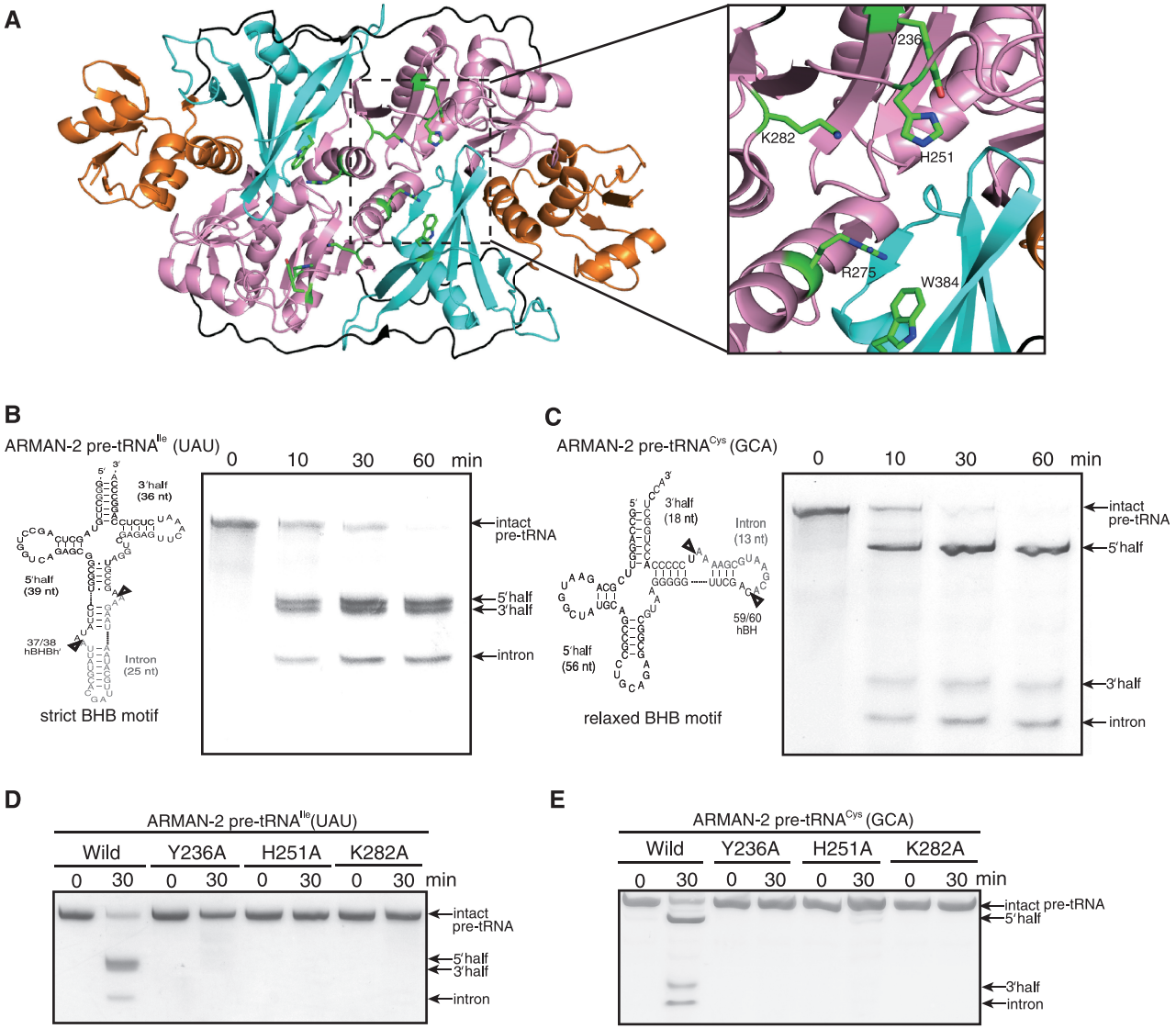

Figure 4. The active site of ARMAN-2 EndA. (**A**) Close-up view of the active site. The catalytic triad comprised of three catalytic residues (Y236, H251 and K282) and two putative RNA recognition residues (R275 and W384) are shown by stick model (green). (**B**) Time-dependent cleavage of ARMAN-2 pre-tRNA^{Ile}(UGU) by the wild-type ARMAN-2 EndA. Predicted secondary structure of ARMAN-2 pre-tRNA^{Ile}(UGU) labeled with two arrows indicating the splicing sites is shown at the left side of the gel. (**C**) Time-dependent cleavage of ARMAN-2 pre-tRNA^{Cys}(GCA). Predicted secondary structure of ARMAN-2 pre-tRNA^{Cys}(GCA) indicating the splicing sites is shown at the left side of the gel. (**D**) Cleavage activities of the wild-type and three mutants (Y236A, H251A and K282A) using ARMAN-2 pre-tRNA^{Ile}(UGU). (**E**) Cleavage activities of wild-type and three mutants (Y236A, H251A and K282A) using ARMAN-2 pre-tRNA^{Cys}(GCA). Reaction mixtures were separated on 15% polyacrylamide/7 M urea gels. The cleaved products are shown using arrows at the right side of the gel.

by the three predicted residues (Y236, H251 and K282), we constructed three alanine mutants (Y236A, H251A and K282A) and then performed an intron-cleavage assay with the mutants. Prior to the mutant study, we optimized the assay conditions by using the wild-type ARMAN-2 EndA and two pre-tRNA transcripts (pre-tRNA^{Ile} and pre-tRNA^{Cys}) previously used as substrates (24). The pre-tRNA^{Ile} and pre-tRNA^{Cys} contain a strict BHB motif at the canonical position and relaxed BHB motif at a non-canonical position, respectively (Figure 4B and C). The ARMAN-2 EndA completely removed the introns from both the pre-tRNA^{Ile} and pre-tRNA^{Cys} within 60 min. Next, we assayed for the removal of intron by the three mutants, Y236A, H251A and K282A. All three mutants failed to remove the introns from both the pre-tRNA^{Ile} and pre-tRNA^{Cys} (Figure 4D and E), suggesting that the Y236, H251 and K282 residues play an important role as the catalytic triad of ARMAN-2 EndA. Of the RNA recognition residues (R275 and W384) of ARMAN-2 EndA, the typtophan is only conserved in the $\alpha_2\beta_2$ type EndAs from Crenarchaea and Nanoarchaea (Supplementary Figure S2B). Instead of the tryptophan residue, an arginine residue is conserved in the α'_2 and α_4 EndAs from Euryarchaea. Two arginine residues in the α'_2 EndA capture the adenine base in the first bulge of the BHB motif by cation–π interactions (6,11).

A tryptophan residue can act as an alternative for the arginine because its indole ring can form a hydrophobic interaction with the nucleotide instead of the cation–π interaction.

Broad substrate specificity

Because the catalytic and substrate recognition residues of ARMAN-2 EndA are conserved in all types of archaeal EndAs as described above, these residues are probably not involved in the broad substrate specificity of the ARMAN-2 EndA. We searched for the specific regions responsible for the specificity of ARMAN-2 EndA based on the structure-based sequence alignment. As a result, two specific regions were found (Supplementary Figure S2B and S2C highlighted in cyan and green). To confirm the findings, we performed a structural comparison of the specific regions of ARMAN-2 EndA and counterparts, namely, the $\alpha_2\beta_2$ type *Aeropyrum pernix* (APE)-EndA and *Nanoarchaeum equitans* (NEQ)-EndAs (Figure 5A and B). Shown in Figure 5A is the specific region (158–168 residues) of ARMAN-2 EndA which forms a loop structure on the enzyme surface close to the catalytic triad. The ARMAN-2 type-specific loop (ASL) is positioned at a similar location to the CSL in APE-EndA, which plays a significant role in the broad substrate specificity. The conformation of ASL resembles that of CSL, although amino acid similarity and identity are not found with the exception of one positively-charged residue (K161 in ARMAN-2 EndA and K44 in APE-EndA in Figure 5A right panel). Notably, the configurations of K161 in ASL and K44 in CSL are such that they are positioned in the same direction. The K44 residue in APE-EndA has been shown to be essential for the enzymatic activity and broad substrate specificity (23). Accordingly, we hypothesized that the K161 residue in the ASL is probably involved in the splicing activity and broad substrate specificity of ARMAN-2 EndA. When we superimposed the structure of ARMAN-2 EndA onto that of NEQ-EndA (Supplementary Figure S3D and S3E), another specific loop (240–247 residues) of ARMAN-2 EndA was found to be positioned in the same way as the specific loop (90–98 residues) of NEQ-EndA (Figure 5B). These loops are close to the catalytic triad, and the catalytic His residue is located in both of these specific loops. Furthermore, it was expected that some positively-charged residues (K93, K94, K96 and R97) in the specific loop of NEQ-EndA would be important for the broad-specificity (32,34). We hypothesized that the K241 and K244 residues of ARMAN-2 EndA may correspond to the positively-charged residues of NEQ-EndA, although there is no sequence similarity or identity between the ARMAN-2 and NEQ-EndAs in this specific region.

To examine whether these residues (K161, K241 and K244) are implicated in the enzymatic activity and broad substrate specificity of ARMAN-2 EndA, we constructed three alanine mutants (K161A, K241A and K244A) and subsequently conducted an intron cleavage assay of the mutants (Figure 6A and B). As shown in Figure 6A and B, the wild-type ARMAN-2 EndA, K241A and K244A mutants could remove introns from both the pre-tRNA^{Ile} and pre-tRNA^{Cys}. In contrast, the K161A mutant did not cleave the introns. These results demonstrate that the K161 residue is essential for the cleavage of introns with strict and relaxed BHB motifs. To understand the importance of the K161 residue structurally, we constructed a docking model of ARMAN-2 EndA complex with RNA based on the reported α'_2 type AFU-EndA and RNA complex structure (Figure 7A) (6). The RNA in the reported complex contains a BHB motif. The docking model shows that the K161 residue is situated near the 3′ phosphate group adjacent to the bulge structure of the RNA, suggesting that the K161 residue captures this 3′-phosphate group (or 3′-phosphate of the third nucleotide in the loop structure), fixes the substrate, and thereby is essential for cleavage activity. To clarify whether the K161 residue in the ASL plays a key role in determining the substrate specificity, we initially created an AFU EndA mutant protein (AFU-ASL) in which Lys175 was replaced by the ASL sequence (GTYKVSEH) of ARMAN-2 EndA (Figure 7B and C). We also made one additional mutant (ASL-K178A mutant), in which the K178 residue of the AFU-ASL mutant, corresponding to the K161 residue of AMRNA-2 EndA, was replaced with alanine. We then analyzed the substrate specificity of these two mutants. As shown in Figure 7D, the AFU-ASL and ASL-K178A cleaved the intron with a strict BHB motif from the anticodon loop in a similar manner as wild-type ARMAN-2 and AFU-EndAs. The wild-type AFU-EndA and ASL-K178A mutant, however, barely cleaved the intron with a relaxed BHB motif from the T-loop of the pre-tRNA^{Cys} (Figure 7E). In contrast, the AFU-ASL mutant effectively cleaved the intron from the pre-tRNA^{Cys} just as well as the wild-type ARMAN-2 EndA did although the cleavage fragment of 3′-half with intron was shown. Thus, these results clearly demonstrate that the insertion of the ASL conferred ARMAN-2 EndA-like broad substrate specificity to AFU-EndA, which otherwise has narrow substrate specificity. Furthermore, it was demonstrated that the K161 residue in the ARMAN-2 EndA plays a key role in the broad substrate specificity acting as the RNA recognition site, in a similar way to the function of the K44 residue in the CSL of APE-EndA (23).

Evolution of the fourth type of archaeal EndA

With respect to the evolution of three archaeal EndA families (α'_2, α_4 and $\alpha_2\beta_2$), it has been proposed that the α subunit gene of α_4 type EndA was first duplicated and then one was subfunctionalized to encode the β subunit, thereby suggesting that the α and β subunits are evolutionarily derived from the same origin (9). Because of the striking structural and sequential similarities between the ARMAN-2 EndA and the three other types of archaeal EndAs (Figure 2, Supplementary Figure S2 and S3), the ε protomer of the ARMAN-2 EndA is likely to also share the common evolutionary origin of the α and β subunits.

The uncultured acidophilic archaeon ARMAN-2 and its lineages were predominantly found in a chemoautotrophic biofilm and grown in acidic and metal-rich

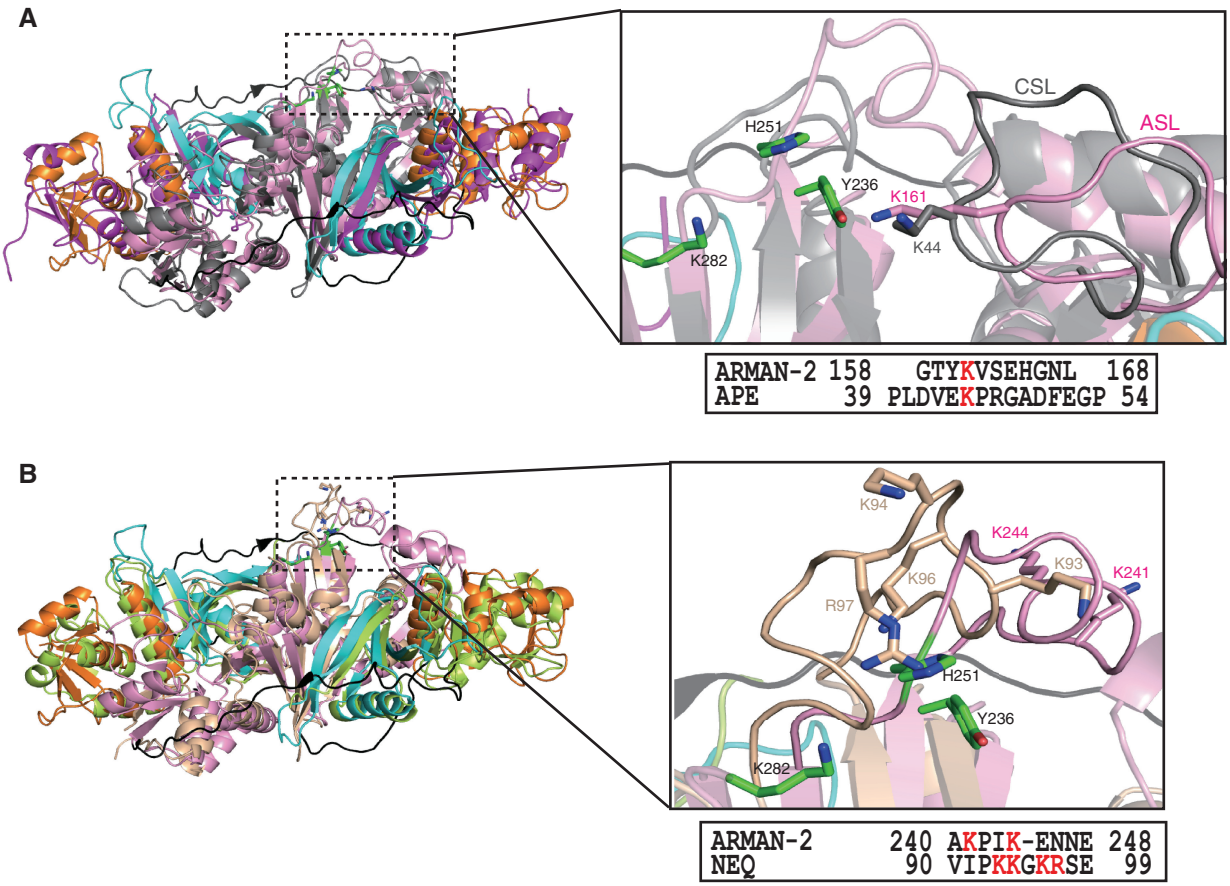

Figure 5. Comparison of ASL, Crenarchaea-specific loop (CSL) and NEQ-specific loop. (**A**) Left: superimposed structures of ARMAN-2 EndA and APE-EndA. Ribbon diagram of the ARMAN-2 EndA and APE-EndA are colored as in Figure 2A and D. Right: close-up view of the structure of ASL region (pink) of ARMAN-2 EndA superimposed on the structure of the CSL region (grey) of APE-EndA. The catalytic triad comprised of three catalytic residues (Y236, H251 and K282) are shown by stick model (green). The structure-based sequence alignment is shown at the bottom of superimposed structures. The conserved K161 in ASL and K44 in CSL are highlighted in red. (**B**) Left: superimposed structures of ARMAN-2 EndA and NEQ-EndA. Ribbon diagram of the ARMAN-2 EndA and NEQ-EndA are colored as in Figure 2A and Supplementary Figure 3D, respectively. The α and β subunits in NEQ-EndA are depicted by wheat and lime-green colors, respectively. Right: close-up view of the structure of insertion loop (pink) of ARMAN-2 EndA superimposed on the structure of the corresponding loop (wheat) of NEQ-EndA. The positively-charged Lys and Arg residues are shown as stick models. The structure-based sequence alignment is shown at the bottom of the superimposed structures. The positively-charged residues in the insertion loops of ARMAN-2 EndA and NEQ-EndA are highlighted in red. The catalytic triad comprised of three catalytic residues (Y236, H251 and K282) are shown as stick models (green). Full names of the archaea species are as follows; APE, *Aeropyrum pernix*; NEQ, *Nanoarchaeum equitans*.

solutions (25). In the biofilm, several eubacteria and archaea including the order of *Thermoplasmatales* are found. Intriguingly, the ARMAN lineages were shown to physically connect to the *Thermoplasmatales* using a 3D cryo-electron tomographic reconstruction (35). Furthermore, a virus was found to be on the cell wall of the ARMAN lineages, indicating an infection of the ARMAN lineages with the virus. Therefore, genetic diversity may occur in the biofilm community via horizontal and/or lateral gene transfer. In fact, ARMAN-2 has many genes homologous to those of Crenarchaea and eubacteria despite the phylogenetic affiliation to the deeply branched Euryarchaea. It is noteworthy that our previous (24) and current studies demonstrate the recombination of the EndA gene in ARMAN-2 cells. Our structure-based sequence alignment shows that the α^N unit is homologous to the N-terminal subdomain of the α subunit from Euryarchaeal EndAs, and that the α and β^C units share homology with the α subunit and C-terminal subdomain of the β subunit from Crenarchaeal EndAs, respectively (Supplementary Figure S2). Accordingly, ARMAN-2 EndA appears to have undergone a genetic recombination of the three subunits. As a result, the α^N, α and β^C units are currently found as the structural and functional element of ARMAN-2 EndA. At the end of the β^C unit (Figure 2 and Supplementary Figure S2), the amino acid sequence (373–387 residues), which folds into the β23 strand, is similar to that of the C-terminal subdomain of the α subunit from Crenarchaeal EndAs. At position 384, a tryptophan residue responsible for RNA recognition site is only found at the end of the α subunit of the EndAs from Crenarchaea and Nanoarchaea. Given these findings, it is likely that the C-terminal subdomain of the Crenarchaeal β subunit is

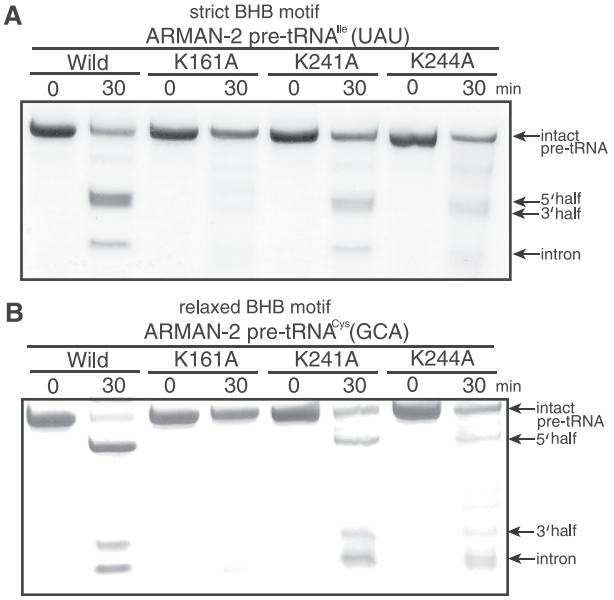

Figure 6. Intron cleavage activities and specificities of the wild-type ARMAN-2 EndA and its mutants (K161A, K241A and K244A): (**A**) ARMAN-2 pre-tRNA^{Ile}(UGU); (**B**) ARMAN-2 pre-tRNA^{Cys}(GCA). The cleaved products are shown using arrows at the right side of the gel.

incorporated into the end of the Crenarchaeal α subunit, resulting in the formation of the $\alpha-\beta^C$ unit of ARMAN-2 EndA. Thus, this could be understood as an example of so-called 'domain shuffling' occurring naturally.

The two specific loop regions of ARMAN-2 EndA were proposed as candidates responsible for the broad substrate specificity from our structural comparison (Figure 5). One of these, the ASL, has been shown to be involved in the broad substrate specificity of ARMAN-2 EndA (Figures 6 and 7). The ASL probably has the same function as the CSL from Crenarchaeal EndA. However, no significant sequence similarity is found except for the conserved Lys residue that functions as the substrate recognition site. Therefore, this suggests that the ASL was acquired by a distinctly independent evolutionary pathway to the CSL, so-called 'convergent evolution'. In contrast, another specific loop is positioned at a location similar to the specific loop of NEQ-EndA (Figure 5B), but has not been shown to be involved in the enzymatic activity and substrate specificity (Figure 6A and B). If the loop had continuous positively-charged residues like the NEQ-EndA, ARMAN-2 EndA may possess the broad substrate specificity even if the ASL is missing. In any case, there seems to be two different structural strategies to obtain the broader specificity as observed in the $\alpha_2\beta_2$ and ε_2 types of archaeal EndAs gained by convergent evolution. First, the conserved Lys residue on either the ASL or CSL adjacent to the active site functions as the substrate recognition site. Second, the continuous positively-charged residues on the specific loop including the catalytic His residue function as the substrate recognition site. In the case of the ARMAN-2 EndA, the second strategy may have been lost because of an earlier acquisition of the ASL.

A bona fide role for tRNA introns remains unclear except for the methylation on the 2′-*O*-ribose of G32 and G34 in tRNA^{Trp} from *Haloferax volcanii* (36). Randou and Söll have argued that the gain of tRNA introns provides protection against integration of mobile genetic elements, such as conjugative plasmids and viruses (37). A total of 56% of tRNA genes are interrupted with both the strict and relaxed BHB motif introns in the ARMAN-2 (23). In contrast, the lineages, ARMAN-4 and ARMAN-5 have only the strict BHB motif introns in 15% of tRNA genes, consistent with possession of the α_4 type EndA (24). Because the prototype of archaeal EndA was proposed to be an α_4 type (9), transition from the α_4 to the ε_2 type might allow an increase in the number and diversity of tRNA introns at non-canonical positions in ARMAN-2 cells. Furthermore, the CRISPR immune system that protects from virus (38) is absent from the genomes of all three ARMAN groups. Therefore, inclusion of the ASL in the ε_2 type of ARMAN-2 EndA may expand the disrupted tRNA genes for defense against the integration of mobile genetic elements as previously proposed in the case of inclusion of the CSL in the Crenarchaeal EndA (23). Moreover, given the report demonstrating that the gain of tRNA introns occurred relatively recently (39,40), incorporation of the ASL region into the ARMAN-2 EndA might have been a dominant advantage for the survival of ARMAN-2 cells in the biofilm community.

In conclusion, our structural study of the ARMAN-2 EndA, which is the fourth type of archaeal EndA, has shown the precise architecture of the ε protomer that consists of three units (α^N, α and β^C). The three units form the ε_2 homodimer. There is striking structural and functional similarity among all four types (α'_2, α_4, $\alpha_2\beta_2$ and ε_2) of archaeal EndAs, suggesting that the four types of archaeal EndAs are derived from a common ancestor. However, the two linker loops connecting the three-unit and the ASL are distinct in the ARMAN-2 EndA. The two linkers play an important role to facilitate the three-unit formation, and the ASL confers the broad substrate specificity on ARMAN-2 EndA. Furthermore, our structure based sequence alignment of the ARMAN-2 EndA exhibits a trace of gene recombination of the α and β subunits from Euryarchaeal and Crenarchaeal EndAs. These results broaden understanding of the mechanism underlying gain of function in protein architecture. In the ASL, the K161 residue functions as a RNA recognition site and thereby broadens the specificity of ARMAN-2 EndA. The ASL has arisen from convergent evolution to play a similar role to the CSL of Crenarchaeal EndA. Inclusion of the ASL in ARMAN-2 EndA may have allowed increasing number and diversity of tRNA introns for the protection from mobile genetic elements. Thus, our findings further enhance the possibility of coevolution of the archaeal EndA architecture and disrupted tRNA genes.

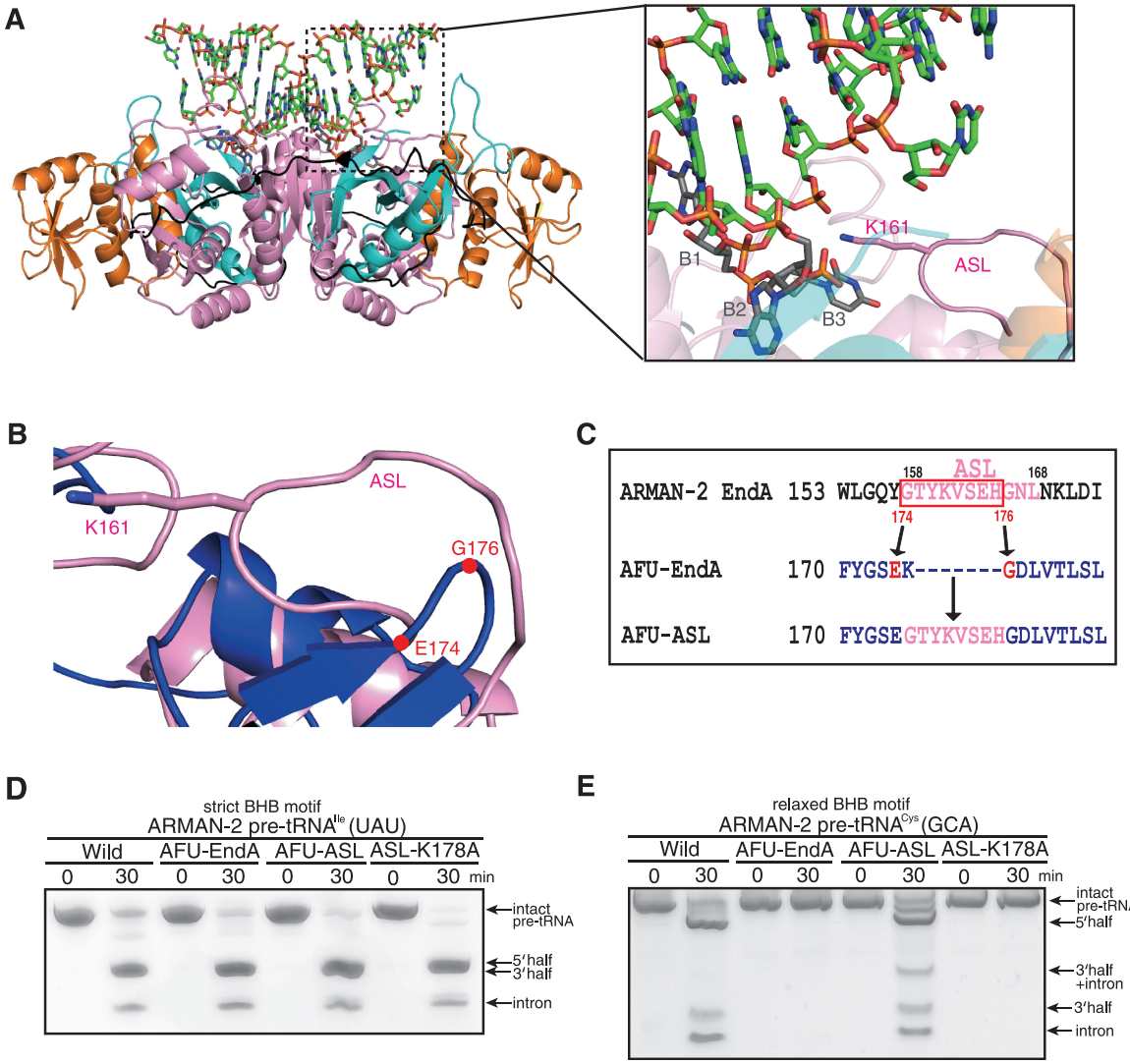

Figure 7. The conserved K161 residue in ASL is responsible for broad substrate specificity. (**A**) Model of the complex formed between the ARMAN-2 EndA and an RNA substrate (stick model, green) that contains a BHB motif (Left). The dotted square shows the active site. Close-up view of the active site of the enzyme-RNA complex (Right). Stick model in grey show the bulge structure (B1–B2–B3) in the BHB motif. The K161 in ASL is shown as a stick model (pink). (**B**) Ribbon diagram of the structure of the ASL region of ARMAN-2 EndA (pink) superimposed on the structure of the corresponding region of AFU-EndA (blue). The insertion positions (E174 and G176), where the ASL peptide (G158–H165) was inserted to create the AFU-ASL chimera, are indicated in red. **(C)** Schematic diagram of creation of the AFU-ASL chimera: the ASL peptide (G158–H165, red box) of ARMAN-2 EndA was inserted into the region between E174 and G176 (red) in AFU-EndA. Cleavage activities of wild-type ARMAN-2 EndA, AFU-EndA and its mutants (AFU-ASL and ASL-K178A): (**D**) ARMAN-2 pre-tRNA^{Ile}(UGU); (**E**) ARMAN-2 pre-tRNA^{Cys}(GCA). The cleaved products are shown using arrows at the right side of the gel.

ACCESSION NUMBERS

The structure factor and coordinates have been deposited in the Protein Data Bank (PDB code 4FZ2).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–3 and Supplementary References [41,42].

ACKNOWLEDGEMENTS

The authors thank the staff members of the beam-line facility at SPring-8 (Hyogo, Japan) for their technical support during data collection. The synchrotron radiation experiments were performed at the BL38B1 in the SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (Proposal No. 2012A1098).

FUNDING

Funding for open access charge: The Grant-in-Aid for Young Scientists (B) [No. 24770125 to A.H.]; Grant-in-Aid for Science Research (B) [No. 23350081 to H.H.]; Japan Society for the Promotion of Science.

Conflict of interest statement. None declared.

REFERENCES

1. Marck,C. and Grosjean,H. (2003) Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA*, **9**, 1516–1531.
2. Marck,C. and Grosjean,H. (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA*, **8**, 1189–1232.
3. Sugahara,J., Kikuta,K., Fujishima,K., Yachie,N., Tomita,M. and Kanai,A. (2008) Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order thermoproteales. *Mol. Biol. Evol.*, **25**, 2709–2716.
4. Abelson,J., Trotta,C. and Li,H. (1998) tRNA splicing. *J. Biol. Chem.*, **273**, 12685–12688.
5. Fabbri,S., Fruscoloni,P., Bufardeci,E., Di Nicola Negri,E., Baldi,M., Attardi,D., Mattoccia,E. and Tocchini-Valentini,G. (1998) Conservation of substrate recognition mechanisms by tRNA splicing endonucleases. *Science*, **280**, 284–286.
6. Xue,S., Calvin,K. and Li,H. (2006) RNA recognition and cleavage by a splicing endonuclease. *Science*, **312**, 906–910.
7. Trotta,C., Miao,F., Arn,E., Stevens,S., Ho,C., Rauhut,R. and Abelson,J. (1997) The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell*, **89**, 849–858.
8. Di Nicola Negri,E., Fabbri,S., Bufardeci,E., Baldi,M., Gandini Attardi,D., Mattoccia,E. and Tocchini-Valentini,G. (1997) The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements. *Cell*, **89**, 859–866.
9. Tocchini-Valentini,G., Fruscoloni,P. and Tocchini-Valentini,G. (2005) Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization. *Proc. Natl Acad. Sci. USA*, **102**, 8933–8938.
10. Kleman-Leyer,K., Armbruster,D. and Daniels,C. (1997) Properties of H. volcanii tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell*, **89**, 839–847.
11. Calvin,K. and Li,H. (2008) RNA-splicing endonuclease structure and function. *Cell. Mol. Life Sci.*, **65**, 1176–1185.
12. Reyes,V. and Abelson,J. (1988) Substrate recognition and splice site determination in yeast tRNA splicing. *Cell*, **55**, 719–730.
13. Calvin,K., Hall,M., Xu,F., Xue,S. and Li,H. (2005) Structural characterization of the catalytic subunit of a novel RNA splicing endonuclease. *J. Mol. Biol.*, **353**, 952–960.
14. Randau,L., Pearson,M. and Söll,D. (2005) The complete set of tRNA species in *Nanoarchaeum equitans*. *FEBS Lett.*, **579**, 2945–2947.
15. Randau,L., Calvin,K., Hall,M., Yuan,J., Podar,M., Li,H. and Söll,D. (2005) The heteromeric *Nanoarchaeum equitans* splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. *Proc. Natl Acad. Sci. USA*, **102**, 17934–17939.
16. Tocchini-Valentini,G., Fruscoloni,P. and Tocchini-Valentini,G. (2005) Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. *Proc. Natl Acad. Sci. USA*, **102**, 15418–15422.
17. Yoshinari,S., Itoh,T., Hallam,S., DeLong,E., Yokobori,S., Yamagishi,A., Oshima,T., Kita,K. and Watanabe,Y. (2006) Archaeal pre-mRNA splicing: a connection to hetero-oligomeric splicing endonuclease. *Biochem. Biophys. Res. Commun.*, **346**, 1024–1032.
18. Tocchini-Valentini,G., Fruscoloni,P. and Tocchini-Valentini,G. (2007) The dawn of dominance by the mature domain in tRNA splicing. *Proc. Natl Acad. Sci. USA*, **104**, 12300–12305.
19. Yoshinari,S., Shiba,T., Inaoka,D., Itoh,T., Kurisu,G., Harada,S., Kita,K. and Watanabe,Y. (2009) Functional importance of crenarchaea-specific extra-loop revealed by an X-ray structure of a heterotetrameric crenarchaeal splicing endonuclease. *Nucleic Acids Res.*, **37**, 4787–4798.
20. Randau,L., Münch,R., Hohn,M., Jahn,D. and Söll,D. (2005) *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature*, **433**, 537–541.
21. Sugahara,J., Fujishima,K., Morita,K., Tomita,M. and Kanai,A. (2009) Disrupted tRNA gene diversity and possible evolutionary scenarios. *J. Mol. Evol.*, **69**, 497–504.
22. Chan,P.P., Cozen,A.E. and Lowe,T.M. (2011) Discovery of permuted and recently split transfer RNAs in Archaea. *Genome Biol.*, **12**, R38.
23. Hirata,A., Kitajima,T. and Hori,H. (2011) Cleavage of intron from the standard or non-standard position of the precursor tRNA by the splicing endonuclease of *Aeropyrum pernix*, a hyper-thermophilic Crenarchaeon, involves a novel RNA recognition site in the Crenarchaea specific loop. *Nucleic Acids Res.*, **39**, 9376–9389.
24. Fujishima,K., Sugahara,J., Miller,C.S., Baker,B.J., Di Giulio,M., Takesue,K., Sato,A., Tomita,M., Banfield,J.F. and Kanai,A. (2011) A novel three-unit tRNA splicing endonuclease found in ultrasmall Archaea possesses broad substrate specificity. *Nucleic Acids Res.*, **39**, 9695–9704.
25. Baker,B.J., Tyson,G.W., Webb,R.I., Flanagan,J., Hugenholtz,P., Allen,E.E. and Banfield,J.F. (2006) Lineages of acidophilic archaea revealed by community genomic analysis. *Science*, **314**, 1933–1935.
26. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
27. Adams,P., Grosse-Kunstleve,R., Hung,L., Ioerger,T., McCoy,A., Moriarty,N., Read,R., Sacchettini,J., Sauter,N. and Terwilliger,T. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1948–1954.
28. Terwilliger,T.C. (2003) Automated side-chain model building and sequence assignment by template matching. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 45–49.
29. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
30. McCoy,A., Grosse-Kunstleve,R., Adams,P., Winn,M., Storoni,L. and Read,R. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
31. Laskowski,R., MacArthur,M., Moss,D. and Thoronton,J. (1992) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst*, **26**, 283–291.
32. Mitchell,M., Xue,S., Erdman,R., Randau,L., Söll,D. and Li,H. (2009) Crystal structure and assembly of the functional *Nanoarchaeum equitans* tRNA splicing endonuclease. *Nucleic Acids Res.*, **37**, 5793–5802.
33. Holm,L. and Rosenström,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
34. Okuda,M., Shiba,T., Inaoka,D.K., Kita,K., Kurisu,G., Mineki,S., Harada,S., Watanabe,Y. and Yoshinari,S. (2011) A conserved lysine residue in the Crenarchaea-specific loop is important for the Crenarchaeal splicing endonuclease activity. *J. Mol. Biol.*, **405**, 92–104.
35. Baker,B.J., Comolli,L.R., Dick,G.J., Hauser,L.J., Hyatt,D., Dill,B.D., Land,M.L., Verberkmoes,N.C., Hettich,R.L. and Banfield,J.F. (2010) Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl Acad. Sci. USA*, **107**, 8806–8811.
36. Singh,S.K., Gurha,P., Tran,E.J., Maxwell,E.S. and Gupta,R. (2004) Sequential 2'-O-methylation of archaeal pre-tRNATrp nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA. *J. Biol. Chem.*, **279**, 47661–47671.
37. Randau,L. and Söll,D. (2008) TRNA genes in pieces. *EMBO Rep.*, **9**, 623–628.
38. Karginov,F.V. and Hannon,G.J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell*, **37**, 7–19.
39. Fujishima,K., Sugahara,J., Tomita,M. and Kanai,A. (2010) Large-scale tRNA intron transposition in the archaeal order Thermoproteales represents a novel mechanism of intron gain. *Mol. Biol. Evol.*, **27**, 2233–2243.
40. Sugahara,J., Fujishima,K., Nunoura,T., Takaki,Y., Takami,H., Takai,K., Tomita,M. and Kanai,A. (2012) Genomic heterogeneity in a natural archaeal population suggests a model of tRNA gene disruption. *PLoS One*, **7**, e32504.
41. Chenna,R., Sugawara,H., Koike,T., Lopez,R., Gibson,T., Higgins,D. and Thompson,J. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.*, **31**, 3497–3500.
42. Gouet,P., Courcelle,E., Stuart,D. and Métoz,F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.