Genetic And Functional Analyses of Archaeal ATP-Dependent RNA Ligase
（古細菌 RNA リガーゼの機能解明研究）

2022

筑波大学グローバル教育院

School of the Integrative and Global Majors in University of Tsukuba

Ph.D. Program in Human Biology

Yancheng Liu

筑波大学

University of Tsukuba

博士（人間生物学）学位論文

PhD dissertation in Human biology

Genetic And Functional Analyses of Archaeal ATP-Dependent RNA Ligase
（古細菌 RNA リガーゼの機能解明研究）

2022

筑波大学グローバル教育院

School of the Integrative and Global Majors in University of Tsukuba

Ph.D. Program in Human Biology

Yancheng Liu

Acknowledgement

I am here sincerely to thank all the people who supported me to make my thesis possible.

First of all, I would like to thank my research supervisor, Dr. Kiong Ho for his excellent guidance on my experiments and paperwork during my graduate school life. At the same time, I also want to thank the Ph. D. Program in human biology from the University of Tsukuba for giving me sufficient financial support to focus on study and the opportunity to study in this laboratory.

I would like to acknowledge my core collaborators, Dr. Yuko Takagi, Dr. Shigeo Yoshinari and Dr. Katsuhiko Murakami, for providing the RNA ligase and TbDcp2 research materials and establishing experiment systems. I also want to acknowledge all of my colleagues especially Dr. Shuhei Ueshima, Dr. Jianhuang Lin and Dr. Yifan Hu, who were always discussing my research problems with me and taught me a lot of experimental techniques.

I would like to thank all my old or new friends who helped me to enjoy the life in Japan, balanced the study and rest and shared memorable time. Last but not least, I want to specially thank my family and my wife for their continuous encouragement and support of my Ph.D. study.

table of contents

Abstract

Background: Living cells are under constant attack by various environmental and cellular agents that induce breakage of both DNA and RNA. While elaborate DNA repair pathways are understood in detail, our knowledge of how damage to RNA molecules is recognized and repaired remains very limited. One of the few facts that have been established is that RNA ligase, an enzyme that recognizes breaks in an RNA chain and joins the two ends together, is a key component of this repair process. The archaea ATP-dependent RNA ligase adopts a unique homodimeric structure and catalyze the circularization of small RNA. The RNA immunoprecipitation analysis in *P. abyssi* suggest that RNA ligase can interact with circular non-coding RNAs.

Purpose: The purpose of this project is to understand the role of ATP-dependent RNA ligase in hyperthermophilic archaeon *Thermococcus kodakarensis* (TkoRnl). The aim of this project is to determine its physiological RNA targets and its mechanism of action.

Methods: Comparative transcriptomics analysis was performed on RNA isolated from *T. kodakarensis* Rnl (TkoRnl) deletion strain and its parental strain (WT). The high-throughput RNA-sequencing was performed using Illumina HighSeq2500 technology. A custom Perl script was used to identify RNA-Seq reads representing circular RNA (circRNA) with a following criteria: RNA-Seq reads containing two segments of at least 20-nts matching the reference genome and the two matched segments should be encoded in the same direction but in inverse order in the reference genome. The circRNA junction reads with less than 100 independent read counts were eliminated. A similar strategy was used to identify circRNAs from the small RNA-Seq datasets, except circRNA reads predicted to be longer than 10 kb and read counts of less than 20 were eliminated. RT-PCR was performed to detect circRNAs from the total RNA isolated from *T. kodakarensis*. Primers containing gene-specific sequences were used for reverse transcription reaction and subsequently amplified by PCR using the circular junction and gene-specific primers. RNA ligase (Rnl) was produced in E. coli and purified from soluble lysate using Ni-agarose affinity chromatography. The in vitro transcribed RNAs or chemically synthesized RNA oligonucleotides were labeled at the 5'-end with [γ- 32P] ATP. The ligase was incubated at 55°C with the

radiolabeled 5’-monophosphate RNA substrates, and the products were separated on a polyacrylamide gel.

Results: Comparative transcriptomics analysis revealed that TkoRnl could alter the abundance of small circRNA species. These included 18 C/D box sRNAs, 3 non-coding RNAs, and 3 genes annotated as protein-coding genes which were either absent or significantly reduced in the TkoRnl deletion. The RT-PCR analysis was performed to verify that the predicted circular RNA species are present in *T. kodakarensis*. There were no significant changes in circularization of tRNA intron and rRNA intermediates between the WT and TkoRnl deletion.

Two C/D box sRNAs, sR42 and sR31, which were found to be circularized in the WT, but not in TkoRnl deletion strain, were assayed for RNA circularization in vitro. *Methanobacterium* RNA ligase (MthRnl) a homolog of TkoRnl, could efficiently circularize both sR42 and sR31. The circularization activity was not significantly affected when the conserved C/D box sequence was substituted with alternative bases. However, a mutation that disrupts the hybridization between the terminus was a poor substrate for ligation. Consistent with these findings, the majority of the sRNA that was found to be circularized in WT and not in TkoRnl deletion, possess a stable terminal stem structure. I conclude that terminal stem structures are critical for the intramolecular ligation.

Recent studies in *Pyrococcus furiosus* suggest that the 3'-end of 23S rRNA could processed by excision of 40-nts helix 98 (H98) located ~100 nucleotides upstream of the predicted 3’-end. RNA-Seq reads that corresponds to H98 segment was also missing in *T. kodakarensis*. Small RNA fragments that correspond to the 3’-end 23S rRNA accumulated in the WT strain. In contrast, RNA fragments correspond to the upstream segment of H98 accumulated in the TkoRnl deletion, suggest that TkoRnl may act on processing/repairing the rRNA.

Discussion: In this thesis, I analyzed the effect of an allelic knock-out of ATP-dependent RNA ligase in *T. kodakarensis*. Deletion of TkoRnl abolishes circularization of a subset of small RNA species, the majority of which were C/D box sRNA. The archaeal Rnl could not circularize unstructured RNA of a similar length or C/D box sRNAs that have disrupted terminal stem structures. Thus, the proximity of the two ends could be critical for intramolecular ligation reaction. The archaeal Rnl may preferentially recognize the terminal stem structure to selectively initiate the ligation reaction.

The functional significance of circular C/D Box sRNA is unclear. C/D box sRNAs have been reported to function as a guide RNA for methylating tRNAs and rRNAs. Circular C/D Box sRNA may alter the specificity to guide RNA to regulate rRNA methylation. While significant differences in the expression of rRNA genes were not detected between the WT and TkoRnl deletion, the fact that RNA fragments upstream of H98 accumulated in TkoRnl deletion suggest its role in rRNA processing. Rnl may either directly involved in joining the breaks generated near H98 or indirectly through circular C/D box sRNA formation.

Abbreviations

aa: amino acid

AMP: Adenosine 5'-monophosphate

ATP: Adenosine 5'-triphosphate

AMPPNP: adenosine 5'-[β, γ-imido]triphosphate

AppRNA: RNA-adenylate

EpA : Enzyme-AMP

BHB: Bulge-helix-bulge

ENDA: archaeal RNA splicing endonuclease A

GTP: Guanosine 5'-triphosphate

NMP: Nucleotide 5'-monophosphate

GMP: Guanosine 5'-monophosphate

tRNA: transfer RNA

rRNA: Ribosomal RNA

5'-ETS: 5' external transcribed spacer

3'-ETS: 3' external transcribed spacer

ITS: internal transcribed spacer

mRNA: messenger RNA

gRNA: guide RNA

circRNA: circular RNA

Rnl: ATP dependent RNA ligase

Rnl1: ATP dependent RNA ligase 1

Rnl2: ATP dependent RNA RNA ligase 2

Rnl3: ATP dependent RNA RNA ligase 3

Trl1: yeast tRNA ligase

Pnkp: polynucleotide kinase

REL: Kinetoplastid RNA editing ligase

NTase: Nucleotidyltransferase

CPD: cyclic phosphodiesterase

TUTase: 3' terminal uridyl transferase

ExoUase: 3' uridine exonuclease

MthRnl: Methanobacterium thermoautotrophicum RNA ligase

PabRnl: Pyrococcus abyssi RNA ligase

TkoRnl: Thermococcus kodakarensis KOD1 RNA ligase

N domain: N-terminal domain

C domain: C-terminal domain

PNK: polynucleotide kinase

WT: wild type

KO: knock out

***Δrnl*: ATP dependent RNA ligase deletion**

RtcB: GTP-dependent RNA ligases

5'-PO4: 5'-phosphate

5’-OH: 5’-hydroxyl

3’-OH: 3’-hydroxyl

RPKM: Read Per Kilobase per Million mapped reads

Chapter 1 : Introduction

1-1 Ligation Mechanism of RNA ligases

There are two RNA ligation system, ATP-dependent RNA ligases and GTP-dependent RNA ligases, have been identified, which ligate different RNA termini[1].

ATP-dependent RNA ligases have a similar ligation mechanism with ATP-dependent DNA ligases and have relatively conserved catalytic domain with those DNA ligases[1, 2]. Some ATP-dependent RNA ligases could ligate both DNA and RNA. In the ligation reaction, the RNA ligase catalyzes the phosphodiester bonds formation between a 5’ phosphate and 3’ hydroxyl termini of RNAs by three reaction steps[3]. In step one, the RNA ligase consumes an ATP to form a covalent ligase-(lysyl-N)-AMP intermediate. In step two, the RNA ligase catalyzes the formation of 5’ adenylated RNA (AppRNA), an inverted pyrophosphate bridge structure by transferring the AMP from the ligase to the phosphate on the 5’ end of the target RNA. In step three, the RNA ligase catalyzes the 5’ - 3’ phosphodiester bonds formation and liberates AMP by inducing a 3’-hydroxyl group of an RNA to attack the 5’ adenylated RNA (AppRNA) (Figure 1). In vitro biochemical assay, RNA ligases can perform intramolecular ligation to form covalently closed RNA circles and intermolecular ligation to join two individual RNA molecules to a single longer RNA molecule with or without a help of complementary bridging polynucleotides pairing to the two RNA molecules[2, 4].

GTP-dependent RNA ligases (RtcB superfamilies) have a different catalytic domain which has no conserved homologs with ATP-dependent RNA ligases. In the RtcB ligation, RtcB could directly ligate the RNA with 5’ -OH end and the RNAs with 2’,3’ cyclic phosphate (or the RNA with 3’ phosphate end) to form a 5’ - 3’ phosphodiester bonds [5, 6]. RtcB ligation reaction also have three steps. In step one, RtcB consumes a GTP to form a covalent ligase-(histidinyl-N)-GMP intermediate. In step two, RtcB transfer the GMP to the 3’ end of RNA to generate a polynucleotide-(3’)pp(5’)G intermediate (RNA-pp-G). In this step, if the 3’ end of RNA is 2’,3’ cyclic

phosphate, RtcB will hydrolyze the termini to 3’-phosphate end and then transfer the GMP. In step three, RtcB join the two ends of RNA to form a 5’ - 3’ phosphodiester bonds by inducing the 5’-OH end to attack the RNA-pp-G end (Figure 1).

1-2 Phylogenetic distribution of ATP-dependent RNA ligases

ATP-dependent RNA ligases are widely identified in all three phylogenetic domains, bacteria, archaea and eukarya. There are six families of ATP-dependent RNA ligases. They are classified into six families because of the difference of their primary structure and RNA substrate specificity. Type 1 ATP-dependent RNA ligases (Rnl1) are identified in bacteriophage, baculoviruses, bacteria, and plant and prefers tRNA as its substrates[7-11]. Type 2 ATP-dependent RNA ligases (Rnl2) is found in all three phylogenetic domains and prefers double stranded RNAs as its substrates[12, 13]. Type 3 ATP-dependent RNA ligases (Rnl3) now is only found in various archaeal species. Via *in vitro* ligation assay, Torchia *et al.* showed that Rnl3 could ligate single stranded DNA and RNA [4]. Type 4 ATP-dependent RNA ligases (Rnl4) are identified in bacteria such as *Clostridium thermocellum* and is a key component for RNA repair system*[14]*. Type 5 ATP-dependent RNA ligases (Rnl5) are identified in bacteria, viruses, and unicellular eukarya such as *Naegleria gruberi* and prefers double stranded RNAs as its substrates[15, 16]. Type 6 ATP-dependent RNA ligases (Rnl6) are identified in fungi and are essential for tRNA repair system[17].

1-3 Biological function of Type 1, Type 4 and Type 6 ATP-dependent RNA ligases

There is a common biological function of Type 1, Type 4 and Type 6 ATP-dependent RNA ligases, called tRNA repair. Before the crystal structure analysis, all the ATP-dependent RNA ligases which ligate tRNA fragment *in vivo* were thought to be Type 1 ATP-dependent RNA ligase (Rnl1). However, nowadays, these ligases were classified into Type 1, Type 4 and Type 6 ATP-dependent RNA ligases because of the different protein structures.

Type 1 ATP-dependent RNA ligase was first identified in T4 bacteriophage (T4 Rnl1) [18]. T4Rnl1 is responsible for host tRNA repair in E.coli. After T4 infection, bacteria cells will express several host anticodon nucleases for breaking its own tRNA

to inhibit virus protein translation as an antiviral response. Lysine tRNA of the bacteria is cleaved at the anticodon loop. The tRNA becomes two RNA fragments. 5' end of tRNA fragment has a 2', 3'-cyclicphosphate on its 3' end while 3' end of tRNA fragment has a 5'-OH end[19]. However, T4 phage will repair the breakage of the host tRNA for restarting its proliferation. T4 polynucleotide kinase/phosphorase (Pnkp) converts the ends on the cutting site to 3'-OH and 5'-PO_4, and then T4 Rnl1 ligates the two Lys-tRNA fragments at the cutting site to reproduce the full length lysine tRNA[20]. This finding raised a hypothesis that RNAs may not just be degraded after using. There could be an RNA repair pathway and RNA ligases could be one key protein in the repair pathway.

In bacteria *Clostridium thermocellum*, the tRNA repair system require polynucleotide kinase/phosphorase (Pnkp) for RNA end-healing and end-sealing. The crystal structure analysis of Pnkp shows a unique ATP-dependent RNA ligase domain[14, 21]. Therefore, the Pnkp is classified as Type 4 ATP-dependent RNA ligases.

Type 6 ATP-dependent RNA ligase (Trl1) is involved in tRNA repair and Non-canonical RNA splicing which is independent with the RNA spliceosome. The crystal structure analysis of Trl1 shows a N-terminal ATP-dependent ligase domain (LIG), a central GTP-dependent polynucleotide kinase domain (KIN) and a C-terminal cyclic phosphodiesterase domain (CPD)[17, 22]. In yeast, the tRNA ligase Trl1, was first found to be involved in tRNA splicing [23]. In yeast, some tRNAs have intron such as Ser-tRNA. For tRNA maturation, the pre-tRNA were first cut out the intron with a site-specific endonuclease. The tRNA becomes two RNA fragments. 5' end of tRNA fragment has a 2', 3'-cyclicphosphate on its 3' end while 3' end of tRNA fragment has a 5'-OH end. Then the Trl1 works for the tRNA repair: the cyclic phosphodiesterase domain (CPD) is responsible for modify the 3' end to 3'-OH; the GTP-dependent polynucleotide kinase domain (KIN) is responsible for modify the 5' end to 5' - PO_4; N-terminal ATP-dependent ligase domain (LIG) is responsible for ligating the two tRNA fragments [24-29].

Besides tRNA repair, the Trl1 have been found to be involved in Non-canonical RNA splicing. In yeast, the mRNA of *HAC1* which is a key protein in the unfolded protein response (UPR), is found to has a Non-canonical RNA splicing [30]. In *HAC1* mRNA splicing, *HAC1* pre-mRNA is exported to cytoplasm after transcription without splicing. When cell stress comes, *HAC1* pre-mRNA is first cleaved at two

sites by an endonuclease and releases the intron. At last step, the yeast tRNA ligase Trl1 ligates the *HAC1* exons to create a new open reading frame which could be translated to active *HAC1* protein. Next the activated *HAC1* protein induces UPR proteins expression for helping cell survival[30].

1-4 Biological function of GTP-dependent RNA ligase

GTP-dependent RNA ligases (RtcB superfamilies) are widely identified in all three phylogenetic domain, bacteria, archaea and Eukarya. RtcB is found to be involved in tRNA splicing, RNA repair and nonspliceosomal mRNA splicing in various organisms[31-37]. Recent studies found RtcB implicating in RNA recombination and biogenesis of circular RNAs[38-40].

In 2011, The first GTP-dependent RNA ligase, RtcB, is identified in the tRNA splicing pathway in *Methanopyrus kandleri* [31]. In tRNA maturation, precursor tRNAs with intron were first cleaved by the tRNA splicing endonuclease and became to the tRNA fragments with 2', 3'-cyclic phosphate end and 5'-OH end. In yeast, the end of tRNA fragments need to be change to 5' phosphorate end and 3' hydroxyl end and then the tRNA fragments were ligated by ATP dependent RNA ligase (Trl1) to form maturated tRNA. However, RtcB could directly ligates the tRNA fragments with 2', 3'-cyclic phosphate end and 5'-OH end to form maturated tRNA. In human beings, RtcB homolog, HSPC117, has been identified to work directly in the tRNA maturation that depletion of HSPC117 inhibits the pre-tRNA maturation in Hela cells and *in vitro* assay[32].

In the URP pathway, yeast cells use Trl1 to conduct *HAC1* mRNA splicing while mammalian cells use RtcB to conduct *XBP1* mRNA splicing[35].

The biogenesis of circular tRNA intron were observed in various Archaea and Eukarya species including human beings. RtcB were found to circularize the tRNA intron after tRNA splicing[39].

It was also reported that RtcB could join long interspersed element-1 sequences (LINE-1) and U6 small nuclear RNA to form chimeric RNA in several human cell lines[40]. LINE-1 could generate various genomic insertions via the retrotransposition and LINE-1 now comprise about 17% of human genome. U6 small nuclear RNA plays an essential role in nuclear intron splicing. The RtcB generated chimeric RNA

could increase genetic variation in the human genome, but the exact biological functions still need to reveal.

Many Archaea and Eukarya species contain both ATP-dependent RNA ligation pathway and GTP-dependent RNA ligation pathway. The two kind of RNA ligases may share some biological function and have its unique biological function at the same time.

1-5 Biological function of Type 2 and Type 5 ATP-dependent RNA ligase

Type 2 ATP-dependent RNA ligase have been identified in all three phylogenetic domains. These RNA ligases prefer to ligate two RNA fragments with a guide RNA in a double-stranded RNA structure. In kinetoplastid, type 2 RNA ligase is named as mRNA Editing Ligase (REL). In kinetoplastid, most of the mitochondrial mRNAs need RNA editing for maturation. These pre-mRNAs are edited to forming new open reading frame with new start codons, stop codons and the coding regions in the transcripts via insertion or deletion of uridines on the pre-mRNAs [41-44]. In the mRNA editing, guide RNAs (gRNAs), small RNA molecules with around 70 nt, hybridize with the pre-mRNAs as template; endonucleases cleave the pre-mRNA at editing site. For insertion editing, 3’ terminal uridyl transferases (TUTase) add uridines onto the 5’ cleavage site and normally RNA ligase REL2 is responsible for sealing the edited mRNA but if REL2 has been downregulated, REL1 could also seal the U insertion edited mRNA [45]. For deletion editing, 3’-U-exonucleases (ExoUase) delete uridines on the cleavage site and only RNA ligase REL1 could seal the U deletion edited mRNA [46, 47]. Via RNAi downregulation, it was found that the disruption of REL1 is lethal to the cells but down regulation of REL2 does not affect cell growth and U deletion RNA editing [48-50]. Via homolog searching with the sequence of REL1 and REL2, T4 RNA ligase 2 (T4 Rnl2) is identified in gene gp24.1 from bacteriophage T4 [51]. Via *in vitro* RNA ligation assay, it was found that T4 Rnl2 prefers to ligate RNAs with an RNA or DNA template but still could ligate single strand RNAs to circular form and linear multimer form [51] [52]. Until now, the biological targets and functions of T4 Rnl2 are still unknown. The potential function of T4 Rnl2 may be like T4 Rnl1 to perform ligation reaction on different tRNA restriction systems or like trypanosome RELs to be involved in RNA editing

pathways.

Type 5 ATP-dependent RNA ligase also prefer to ligate double strand RNA. The first Type 5 ATP-dependent RNA ligase (DraRnl) was found in bacterium *Deinococcus radiodurans*. *Deinococcus radiodurans* is famous for its extreme resistance to ionizing radiation(IR) which could cause lethal effects such as DNA damage, oxidative bursts and protein damage[15]. The deletion of DraRnl could not affect the cell growth under normal condition. However, deletion of DraRnl could cause the cell become sensitive to the ionizing radiation[53].

1-6 Identification and *in vitro* characterization of Type 3 ATP-dependent RNA ligase

Type 3 ATP-dependent RNA ligases (Rnl3) were identified in various archaea species via homology analyses with the nucleotidyltrasferase domain of T4 Rnl2[4, 54]. Type 3 RNA ligases (PabRnl) from the archaeon *Pyrococcus abyssi* and MthRnl from Archaeon *Methanobacterium thermoautotrophicum* are purified and analyzed via *in vitro* biochemical assay individually. *In vitro* assays show that both PabRnl and MthRnl have standard ligation mechanism of RNA ligases which react with ATP to form ligase-adenylates first, then generate RNA adenylate intermediates and last seal the RNAs.

Unlike Type 1 and 2 RNA ligases, Type 3 RNA ligases prefer to seal an RNA itself to form a circular RNA but not to join two RNA fragments into multimer linear RNA. This suggests that type 3 RNA ligases have specific mechanism to protect single strand RNA circularization and prevent linear multimer ligation with two RNAs. For the polynucleotide substrate, MthRnl shows a dual-specificity to ligate the single stranded DNA with similar efficiency to the single stranded RNA, while type 1 or type 2 RNA ligases can not circularize the single stranded DNA[4]. Furthermore, MthRnl could seal the chimeric 24-mer oligo which has a DNA 5’-end and an RNA 3’ end. MthRnl could also seal the chimeric 24-mer oligo which has an RNA 5’-end and a DNA 3’ end. These results suggests that MthRnl does not discriminate between RNA and DNA on either 5’ or 3’ end for phosphodiester bond formation. Moreover, glycerol gradient sedimentation analyses find that type 3 RNA ligase (Rnl3) is a homodimer while type 1 and 2 RNA ligases are monomer. Domain deletion analyses show that the C-terminal domain deletion could disrupt Rnl3 dimerization and

dimerization are essential for Rnl3 to conduct RNA ligation[4].

1-7 protein structure of ATP-dependent RNA ligases and their substrates specificity

All DNA ligases, ATP-dependent RNA ligases and GTP-dependent mRNA capping enzymes belong to the superfamily of covalent nucleotidyl-transferases which could capture NMP to form an enzyme-(lysyl-N)-NMP intermediates and transfer the NMP nucleotide to the polynucleotide 5’ ends[55]. This active domain is classified as nucleotidyl-transferase (NTase) domain. Among all the three types of ATP-dependent RNA ligases, the NTase domain is located in N-terminal domain. Deletion and mutational analysis suggest that the feature functions of each type of RNA ligase come from their C-terminal domain. The deletion of the C-terminal domain of T4Rnl1 causes loss of specificity on tRNA ligation. The surface electrostatics of T4Rnl1 shows that C-terminal domain of T4Rnl1 have accumulated positive charges accumulate on its surface while the rest of the protein surface is negatively charged, which suggests that tRNA binds at the C-terminal domain [56]. Mutational analyses find that Arg-318 and Lys-319, two positively charged residues, are the potential residues to interact with the tRNA [57].

In T4Rnl2, the truncated T4Rnl2 with N-terminal domain (NTase domain) only can catalyze steps 1 and 3 of the RNA ligation reaction but is not able to conduct an entire ligation reaction and not able to bind the nicked duplex substrate. Furthermore, mutational analysis identifies two C-terminal domain residues, Arg-266 and Arg-292, which are essential for conducting an entire ligation reaction with nicked double strand RNA but dispensable for sealing a pre-adenylated RNA substrate[3].

In archaea, the type 3 RNA ligases have a unique homodimeric quaternary structure while all identified type 1 and 2 RNA ligases are monomeric. There are two crystal structures of type 3 RNA ligases which have been reported, PabRnl and MthRnl. In the C-terminal domain, a dimer interface (aa 245-313) has been identified and it is an all-α-helical structure[54]. Dozens of mutational analyses with alanine substitution are performed on MthRnl to identify the key residual for each ligation step and RNA binding[58]. Mutated MthRnl with Alanine substitution of the amino acid Thr117 or Arg118 is not able to seal the substrate RNA (the step 3 reaction). Instead, this mutated MthRnl tends to deadenylate the 5’-AMP from the RNA-adenylate

(AppRNA), which reverse the step 2 ligation reaction. Tyr159, Phe281 and Glu285, which are conserved among archaeal ATP-dependent RNA ligases and are situated on the surface of the enzyme, are required for RNA binding. Deletion analysis with MthRnl suggests that NTase domain on its own could form ligase-AMP and RNA-adenylate formation, although the activities are much lower than full-length MthRnl. This suggests that C-terminal domain from Rnl3 family is required for the full activity of all steps along the ligation pathway.

1-8 Archaea Introduction

Carl Woese and his colleagues first proposed the domain of archaea in 1970 and the three-domain system was formally proposed in 1990 based on the difference of SSU rRNA sequences, sequences of ribosomal proteins and other core proteins [59]. Archaea species are organisms with single cell and usually live in environments with extremes in salinity, temperature or acidity. Archaea have similarities to both eukaryotes and bacteria. Like bacteria, archaea have no nucleus but own circular chromosomes and operons and share CRISPR-cas immunity system and horizontal gene transfer system. In transcriptional, post-transcriptional modification and translational systems, archaea are more similar to Eukarya and have various homologous genes which are working in these systems[60]. In the transcription system, Like eukaryotes, archaeal promoters also have TATA boxes and B recognition elements[61] In the post-transcriptional modification system, Archaea have no nucleolus but have snoRNA homologs, sno-like RNAs (C/D box sRNAs and H/ACA box sRNAs), and their core proteins to precisely target rRNA and tRNA and modify the target nucleotide of the RNA with 2’-O-methylation or pseudouridylation, like Eukarya. Therefore, evolutionary biologists have placed the Kingdom of Archaea closer to Eukarya than Bacteria in the phylogenetic tree.

Currently archaea domain has been classified into eight phyla. type 3 RNA ligases have been analyzed in Archaeon *Pyrococcus abyssi*, Archaeon *Methanobacter thermoautotrophicus str. Delta H* and Archaeon *Thermococcus kodakarensis KOD1*. Archaeon *Pyrococcus abyssi* and Archaeon *Thermococcus kodakarensis KOD1* belong to Thermococcaceae family in which all species in this family are isolated from hydrothermal vents in the ocean and are strictly anaerobes. Archaeon

Methanobacter thermoautotrophicus str. Delta H belongs to Methanobacteriales order in which all species in this order are able to produce methane in hypoxic conditions and most species are Gram-positive with rod-shaped bodies. All the three archaea belong to one of the eight phyla called ARMAN (Archaeal Richmond Mine Acidophilic Nanoorganisms).

1-9 tRNA and rRNA processing in Archaea

The tRNA needs various process to generate matured tRNA from long precursor tRNA. The processing include tRNA splicing, addition of a 3'-end CCA-end, extensive nucleotide modifications and, in some cases, removal of tRNA introns[62]. The archaeal tRNA genes were not all encoded in individual transcriptional units. Some of tRNA genes were encoded in operons with other tRNAs together. Some of tRNA genes were encoded in polycistronic units with ribosomal or messenger RNAs[63, 64]. Some of tRNA genes were even encoded separately and need a trans-splicing and ligation to join their 5'-halves and 3' -halves to a mature tRNA[65].

For tRNA intron splicing, Eukarya and Archaea species use similar two step processing. In the first step, An RNA endonuclease cleaves the pre-tRNA at specific site to generate a 2',3'-cyclic phosphate end and a 5'-hydroxyl end. In the second step, RNA ligases are responsible for join the two tRNA fragment. Yeasts use ATP dependent RNA ligase (Trl1) to ligate tRNAs and Human beings use GTP dependent RNA ligase (RtcB) to ligate tRNAs[23, 32]. In Archaea, *In vitro* assay suggest that many species use RtcB to ligate tRNAs for tRNA maturation[31].

In archaea, about 15% tRNAs contain introns with 11 to 129 nucleotides length[66]. The tRNA introns commonly locate between tRNA nucleotide 37 and 38 in the anticodon loop[67]. The archaeal splicing endonuclease (EndA) cleave the pre-tRNA by recognizing a bulge–helix–bulge (BHB) structure. The BHB motif have a 3-nt bulges, follow with a central 4-bp helix and end with a 3-nt bulges (Figure3)[68]. This structure helps the splicing endonuclease EndA to specifically cut the splice sites. The BHB processing pathway was widely used in archaea and were found in tRNA processing, rRNA processing, and mRNA processing[65, 69]. After cleavage, RtcB ligates the tRNA fragments directly to form the matured tRNA and in some case, the released tRNA introns were also circularized by RtcB.

rRNA genes are organized in ribosomal RNA operons(*rrn* operons) in all

species[70]. In Archaea, almost all archaeal genomes only have one to four *rrn* operons. In the Archaea species with only one *rrn* operons contain a 5' external transcribed spacer (5'-ETS) followed with a 16S rRNA gene, an internal transcribed spacer (ITS-1), a 23S rRNA gene, an ITS-2, a 5S rRNA gene and a 3' external transcribed spacer (3'-ETS)[71]. In other Archaea species, the 5S rRNA gene is scattered on the chromosome with an independent *rrn* operon while the 16S rRNA and 23S rRNA are contained together in one *rrn* operon. The ITS usually encodes an alanyl-tRNA gene (Figure 4) [72].

In Bacteria, the 5'-leader and 3'-trailer sequences of 16S pre-rRNA and 23S pre-rRNA form helix structures and are cleaved by the bacterial RNase III. In Archaea, the leader and trailer sequence of 16S/ 23S rRNA form BHB motifs and are cleaved by the endonuclease EndA as same as the tRNA splicing reaction[69, 73]. After BHB motif cutting, RNA ligases(possibly RtcB) ligate the 16S and 23S pre-rRNAs to circular 16S and 23S pre-rRNAs[74]. These circular 16S and 23S pre-rRNAs are only intermediates and receive multiple endonucleolytic cleavages and exonucleolytic trimming to form the mature ribosomal RNA[75]. The circular form of 5S rRNA intermediates have also been identified in *S. solfataricus*, but the clear 5S rRNA processing pathway is still waiting for revealing.

The 23S pre-rRNA have over hundred short helix structures which are named from H1 to H101. Helix 98 (H98) is generally located on the surface of the large ribosomal subunit and widely exists in most Archaea, bacteria and Eukarya species such as *T. kodakarensis*, E. coli and human beings[62, 76, 77]. In E. coli, H98 consists of 15 nucleotides and the crystal structures of ribosome suggests that H98 is located at the back of the large subunit and is near the ribosomal proteins L3 and L13[78]. However, in Archaea, H98 seems to be optional. In *Pyrococcus furiosus*, the H98 of 23S rRNA was found be precisely excised[78]. In the 23S rRNA maturation process, the 23S rRNA is circularized by the tRNA splicing-like mechanism and the helix 98 (H98) was removed in the later process which's mechanism is unknown. After H98 excision, the 3' end of downstream remaining 23S rRNA(H99-H101) were ligated to the 5' end of 23S rRNA and the new 3' end of 23S rRNA becomes to helices H94–97 comprising the Sarcin-Ricin Loop. As a consequence, the mature 23S rRNA of *Pyrococcus furiosus* is linear and have a trans-splicing where the H99-H101 fragment is appended to the 5' end of 23S rRNA. Still the biological importance of this rRNA fragmentation needs further study.

1-10 C/D box snoRNA/ sRNA biogenesis and functions

Small nucleolar RNAs(snoRNA) were first identified in Eukaryotic cells and were named because of their location, nucleolus. The first identified biological function of snoRNA was guiding methylase to modify ribosomal RNA, called 2'-O-methylation. Later, in Archaea, snoRNA's homolog, sno-like RNAs were identified. As Archaea has no nucleolus, these sno-like RNAs were named sRNA. The snoRNAs have two major families, C/D box snoRNA (C/D box sRNA in Archaea) and H/ACA box snoRNA (H/ACA box sRNA in Archaea), based on their different conserved sequence and structural elements[79-81].

The C/D box snoRNA/sRNA have a C box (RUGAUGA) motif at the 5' end of the snoRNA and a D box (CUGA) motif at the 3' end of the snoRNA. In the internal position, snoRNAs have less conserved sequence D' box and C' box, which allows more variance in the C/D box motif [82]. There is an obvious difference between C/D box snoRNA and C/D box sRNA. In Eukaryotic cells, C/D box snoRNA usually have a length between 50 to 300 nt. In Archaea, C/D box sRNAs usually have a length around 60nt.

For conducting 2'-O-methylation on rRNA or tRNA, C/D box snoRNA are associated with three core proteins to form a functional structure (sRNP). In the structure, the 5' end of a C/D box snoRNA adjoins the 3' end of the C/D box snoRNA via base-pairs between C box motif and D box motif. The first 'GA' nucleotides in C box motif (RU**GA**UGA) are paired with the 'GA' nucleotides in D box motif (CU**GA**), which caused a helix-loop-helix structure forming, called kink-turn (K-turn) structure. Similarly, D' box motif is paired with C' box motif and formed a K-loop structure. In the sRNP structure, two L7Ae protein first recognize K-turn and K-loop structure and bind them respectively for stabilizing[83]. Then another two core proteins, Nop5 and fibrillarin, bind L7Ae to assemble as active ribonucleoprotein (RNP) complex[84]. The fibrillarin protein is a methylase which are responsible for 2'-O-methylation on rRNA or tRNA [85-88]. For specifically modifying rRNA or tRNA, the C/D box sRNA sequence between C motif and D' motif or between C' motif and D motif work as guide sequence and form base-pairs with the target RNA (Figure 2). In archaea, usually, the guide sequence has 11 to 20 nucleotides. The nucleotide which undergoes 2'-O- methylation is complementary to the fifth nucleotide of guide sequence which is upstream of D motif or D' motif[89].

C/D box snoRNAs are found to be related with various human diseases, such as Prader-Willi syndrome, numerous breast and prostate cancers[90-92]. In human cells, 228 sites of modifications have been identified on the 80S rRNA[93]. Loss-of-function mutations in fibrillarin cause the global disappearance of 2'-O- methylation and cell death[94]. However only a few snoRNAs are indispensable in Saccharomyces cerevisiae. A group tried to genetically delete snoRNAs in yeast individually or multiply and found no significant adverse effect[95]. Thus, although snoRNAs induce 2'-O-methylation in same mechanism, the biological function of each 2'-O- methylation site could be different. In archaea, circular form of C/D box sRNA have been found universally[74, 96]. The ratio between circular form and linear form of C/D box sRNA is different between each archaea species. In *Sulfolobus solfataricus*, C/D box sRNAs have a higher amount of linear form C/D box sRNA than circular form while in *Pyrococcus furiosus*, the two form of C/D box sRNAs have similar abundance[74, 97]. The biological function for forming a circular form of C/D box sRNA is still not clear. As these archaea species are living in high temperature condition, the circular form of C/D box sRNA may help to stable the structure of C/D box sRNP.

Except nucleotide modification, C/D box snoRNAs have been found some other activities [98]. In mouse rRNA maturation, 5' external transcribed spacer (5' ETS) is cleaved in early rRNA processing stage via RNAse H. At least four C/D box snoRNAs facilitate the cleavage of the 5'ETS for rRNA maturation[99]. C/D box snoRNAs also show ability to modify other cellular RNAs, such as U6 spliceosomal RNA[100, 101].

1-11 Potential function of Rnl3

To understand the biological function of Rnl3, RNA samples from coimmunoprecipitation of PabRnl were sequenced via Ion Personal Genome Machine Sequencer. The sequencing data suggest that PabRnl could bind with C/D box sRNA, tRNA and rRNA [96]. Indeed, in archaea, various types of C/D box sRNA have circular forms; Trp tRNA gene has an intron which also has a circular form; 16S rRNA and 23S rRNA have circular forms as intermediates during rRNA maturation[74, 102-105]. Therefore, it is possible that Rnl3 is involved in those RNA

circularization. However, there are also reports that in Archaea, RtcB is involved in the tRNA splicing and rRNA processing after the ENDA cutting the BHB structure[31, 65, 106, 107]. MthRnl and PabRnl homologs are not found in all archaeal species. 29 out of 48 published archaeal genomes encode MthRnl homolog. From phylogenetic analysis, the coexistence of two RNA ligases, ATP dependent RNA ligase (MthRnl-like) and GTP dependent RNA ligase (Rtcb), are identified in some archaea, which may mean some kind of overlap functions between these two RNA ligases [108-110].

Our group member Shigeo Yoshinari identified a novel function of MthRnl that MthRnl could selectively cleave one adenosine on the 3' end of RNA substrates to generate a new 2',3'-cyclic phosphate end. This feature could give MthRnl novel biological functions.

1-12 Aim of this project

The main purpose of this project was to understand the biological function of Type 3 ATP dependent RNA ligase (Rnl3). In order to thoroughly define the biological functions of Rnl3 and identify its physiological targets, our team constructed T. kodakarensis Rnl3 Knock-out strain (TK1545 KO) and performed next generation sequencing to finding the alteration of transcriptome between WT and TK1545 KO strain.

The next question is how Rnl3 recognizes its biological targets. We used in vitro RNA ligation assay with different RNA substrates to elucidate this question.

Our group member Shigeo Yoshinari identified that MthRnl could selectively cleave one adenosine on the 3' end of RNA substrates to generate a new 2',3'-cyclic phosphate end. This feature could give MthRnl novel biological functions. In this project, we tried to further study this novel function.

Chapter 2 Method and Materials

2-1 TkoRnL knockout cell line construction

The homolog of Archaea RNA ligase 3 was found in T.kodakarensis annotated genome (NC_006624.1). The locus tag of this gene is *TK1545* and we named it as TkoRnl. The Katsuhiko Murakami laboratory at Penn State University contributed to the TkoRnl knockout cell line construction. The non-replicating targeting vertor carried the selectable marker of orotidine-5'-monpophosphate decarboxylase gene (*pyrF*) which was flanked by 700bp of upstream and downstream *TK1545* sequence. The targeting vector was transformed into the T.kodakarensis strain KW128 which has the deletion of *pyrF* gene. *TkoRNL* deleted transformants were selected via the culture with the absence of uracil. Dr. Takagi Youko confirmed the replacement of *TK1545* by the *pryF* marker via PCR genotyping and RT-PCR of total RNA.

2-2 *T.kodakarensis* total and small RNA isolation

T. kodakarensis total RNA was isolated from 100ml of wild type (WT) and *TK1545* knockout cell (*Δrnl*) cultures with TRIZOL (Invitrogen). The isolated RNA was treated with DNase I to prevent genomic DNA contamination. To collect small RNAs, total RNAs of WT and *Δrnl* were separated by 10% denaturing PAGE and the RNA population of 80-140nt was isolated via gel cutting and TE eluting. The eluted small RNA was ethanol precipitated with glycogen and resuspended with TE.

2-3 RNA sequencing via illumina HighSeq2500 system

All RNA samples were sent to the Genomic facility at SUNY at the State University of New York at Buffalo. They established the cDNA libraries and performed the RNA sequencing running with pair-ended running type via Illumina HighSeq2500 technology.

2-4 Gene expression profile analysis

Sequences in FASTQ format were first checked with FastQC and adaptor

sequences were trimmed via cutadapt. T.kodakarensis annotated genome (NC_006624.1) was indexed via bowtie 2.2.9. Post-trimmed reads were aligned to T.kodakarensis annotated genome via bowtie 2.2.9 with the custom setting. In alignment options, we selected the sensitive and local mode. In score options, we set the parameter as --score-min G,1.5,8. We also suppressed SAM records for reads that failed to align by setting SAM options as --no-unal. After alignment, T.kodakarensis annotated gene expression profiles were analyzed by featurecount[111] with the default setting.

2-5 Circular RNA analysis

RNA sequencing Reads in FASTQ format were analyzed via custom scripts which were written in PERL. To identify junction reads which indicated the circular RNA, I first aligned the Read with its 20nt of 5’ starting sequence and its 20nt of 3’ ending sequence separately. If all reads were first categorized into three classes, perfectly matched reads, double matched reads and discarded reads, the perfectly matched reads are the reads which could be mapped on the *T.kodakarensis* annotated genome(NC_006624.1) without any mismatch, the double matched reads are the reads which could be mapped on two distinct genome regions with their 5’ fragment sized 20bp and 3’ fragments sized 20bp but the entire read**s** cannot be perfectly mapped on the genome. The discarded reads represented all the remaining reads after the selection of perfectly matched reads and potentially processed reads.

The end matched reads were used for identifying cis-splicing sites (the 5’ fragment of the read is perfectly mapped near the 5’ end side of the genome while the 3’ fragment of the read is perfectly mapped near the 3’ end side of the genome) and trans-splicing sites (the 5’ fragment of the read is perfectly mapped near the 3’ end side of the genome while the 3’ fragment of the read is perfectly mapped near the 5’ end side of the genome). The identified trans-splicing sites could be circular RNA junction sites. For every end matched read, we repeatedly mapped its 5’ fragment from 21bp to 30bp on *T.kodakarensis* genome to find out the exact splicing site. Cis- or trans-splicing sites were decided by the relative genomic location between 5’ fragment and 3’ fragment of the end matched reads. The end matched reads, whose 5’ and 3’ fragment respectively have 5’-3’ sequential genome locations, were used to find cis-splicing sites. Further, the end matched reads, whose 5’ and 3’ fragments have

3’-5’ sequential genome locations, were used to find trans-splicing sites, the circular junction sites.

2-6 Circular RNA identification via Reverse Transcription PCR

Total RNAs from *T.kodakarensis* wild type or TkoRnL knockout (*Δrnl*) cells were used as templates for the Reverse Transcription PCR (RT-PCR). To identify the circular RNA predicted by NGS data, forward junction primer and reverse primer were designed for each predicted circular RNA. Circular RNAs were individually transcribed with their reverse primers. In the standard RT-PCR, 200ng of RNAs from WT or *Δrnl* were incubated at 55°C for 10mins in the 40 μl mixture containing 1X RT buffer, 0.5mM dNTP, 0.25μM reverse primer, and 50U ReverTra ace from Toyobo. The reverse transcriptase was inactivated by incubating at 95°C for 10mins. In the standard PCR, 1 μl of RT-PCR sample was incubated with the 50 μl mixture containing 1X Paq buffer, 0.2mM dNTP, 0.2μM reverse primer, 0.2μM forward junction primer, and 2.5U Paq5000 from Toyobo. The PCR program was 4mins incubation at 95°C and 25 cycles of 30s incubation at 95°C, 30s incubation at 60°C, and 10s incubation at 72°C. The PCR products were run with 3% low range ultra agarose gel and visualized by UV.

2-7 plasmid construction

Methanobacterium thermoautotrophicum strain "H genomic DNA was used as a PCR template to amplify the MTH1221 ORF, the predicted ATP dependent Archaea RNA ligase 3 gene (*MthRnl*). NdeI restriction site 5’ of the predicted translation start codon and a BamHI site at the 3’ of the predicted stop codon were introduced to the amplified fragments, 1.1-kb NdeI–BamHI fragments, via PCR. The fragment was inserted into pET16b (EMD Chemicals) with a N-terminal 10X his-tag (pET-MthRnl).

2-8 protein expression and purification

The pET-MthRnl plasmid was transformed into E. coli strain Rossetta2-(DE3). The

cells were cultured with LB media containing 880.1 mg/ml ampicillin until the A600 reached 0.3 at 37°C. Isopropyl-d-thiogalactopyranoside (IPTG) was then added to the culture to a final concentration of 0.3 mM for a further 3h incubation at 37°C. Cells were harvested by centrifugation and resuspended in 20 ml Buffer A (50 mM Tris–HCl (pH 7.5), 250 mM NaCl, 10% sucrose). Lysozyme was then added to a final concentration of 25μg/ml. The sample was then sonicated for 30s and cool down in ice for 2mins. Triton X-100 was added to 0.1 % final concentration and sonication was repeated twice. The cell lysate was centrifuged for 45 min at 17,000 rpm with a Beckman T14-50 rotor at 4°C. The supernatant was collected and applied to the column containing 1 ml Ni-NTA resin (Qiagen) which has been equilibrated with Buffer A containing 0.1% Triton X-100. The column was washed and eluted with 4ml Buffer B (50 mM Tris–HCl (pH 8.0), 250 mM NaCl, 10% glycerol) containing 0, 5, 50, 100, 200, 500 and 1000 mM imidazole orderly. The recombinant protein of MthRnl was eluted predominantly in the 200 mM fraction.

2-9 glycerol gradient sedimentation

2ml of MthRnl from the 200mM imidazole fraction was concentrated into 0.1ml via Amicon Ultra 10K device and layered onto 4.8ml 15-30% glycerol gradients containing buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2mM DTT, 0.05% Triton X-100). The gradients were centrifuged in the Beckman SW55i rotor at 45,000 rpm. for 18h at 4°C. Fractions (around 0.2 ml) were collected from the bottoms of the tubes. Besides, 30 μg of MthRnl from the 200mM imidazole fraction was mixed with marker proteins (catalase, BSA and cytochrome c) in 0.1 ml of buffer B and performed the same gradient sedimentation as described above. The MthRnl was mainly collected from the fraction of 17, 18, and 19.

2-10 *in vitro* RNA transcription

The templates and primers for *in vitro* RNA transcription were first amplified by PCR with the primer sets described in 2-16. The 50μl PCR reaction mixture for template amplification contained 1X KOD1 plus buffer, 0.2mM dNTP, 2mM $MgSO_4$, 1fmole DNA template, 1μM reverse primer, 1μM forward primer, and 1U KOD1 plus from Takara. The PCR program was 3mins incubation at 95°C and 25 cycles of 30s

incubation at 95°C, 30s incubation at 55°C, and 10s incubation at 72°C. The PCR product was purified by Favorgen gel/PCR purification kit with the manufacturer's instructions. Before *in vitro* RNA transcription, 1μg of DNA template and 1μg of 5' Forward primer were incubated with 100mM Tris HCl pH 7.5 and 100mM NaCl at 95°C for 3mins and 55°C for 5mins for hybridization in 15μl volume. All the 15μl hybridization products were added to 185μl reaction mixture to make the final reaction sample contain 1X T7 RNA transcription buffer, 50mM DTT, 2mM NTP, 0.01% TritonX-100, 8mM $MgCl_2$ and 100U T7 RNA polymerase from Takara. The reaction sample was incubated at 37°C for overnight. The *in vitro* transcribed RNAs were then gel purified with 8% native PAGE and treated with CIAP for preparing 5'-OH RNAs which will be labeled by PNK at 5' end in RNA ligase ligation assay.

2-11 T4 RNA Ligase 2 ligation Assay

The standard T4Rnl2 ligation reaction mixtures (10 μl) containing 50 mM Tris HCl (pH 6.5), 5 mM DTT, 1 mM $MgCl_2$, 0.5 pmol of 5' 32P-labeled ligase substrate, and with or without 1mM ATP and T4Rnl2 as specified were incubated at 22 °C with indicated times. The reactions were quenched by adding the same volume of 98% formamide, 20 mM EDTA. The samples were analyzed by electrophoresis through a 12% polyacrylamide gel containing 7 M urea in the running buffer with 50 mM Tris borate for **3**0mins at 30mA of constant current. The gel was then visualized by autoradiography and quantified via a Typhoon FLA7000 phosphoimager.

2-12 MthRnl ligation assay

The standard MthRnl ligation reaction mixtures (10 μl) containing 50 mM Tris HCl (pH 6.5), 1 mM $MgCl_2$, 0.5 pmol of 5' 32P-labeled ligase substrate, and with or without 1mM ATP and MthRnl as specified were incubated at 70 °C with indicated times. The reactions were quenched by adding the same volume of 98% formamide, 20 mM EDTA. The samples were analyzed by electrophoresis through a 12% polyacrylamide gel containing 7 M urea in the running buffer with 50 mM Tris-borate for 30mins at 30mA of constant current. The gel was then visualized by autoradiography and quantified via a Typhoon FLA7000 phosphoimager.

2-13 Electrophoretic mobility shift assay (EMSA)

The reaction mixtures of electrophoretic mobility shift assay (10 μl) containing 0.5 pmol of 5’ 32P-labeled ligase substrate, 22% glycerol, and wild-type MthRnl proteins with the indicated amount, were incubated for 5 mins at 27°C or 55°C. The protein-RNA complex formation was analyzed by electrophoresis for 30mis with 20mA constant current through a native 8% polyacrylamide gel. The gel was then visualized by autoradiography and quantified via a Typhoon FLA7000 phosphoimager.

2-14 MthRnl 3’-deadenylation assay

The standard MthRnl reaction mixtures (10 μl) contain 50 mM Tris HCl (pH 6.5), 0.5 mM MgCl2, 0.5 pmol of 5’ 32P-labeled ligase substrates, and MthRnl. The mixtures were incubated at 55 °C for 15mins. The reactions were quenched by adding same volume of 98% formamide, 20 mM EDTA. The samples were analyzed by electrophoresis through a 15% polyacrylamide gel containing 7 M urea in the running buffer with 50 mM Tris-borate for 30mins at 30mA of constant current. The gel was then visualized by autoradiography and quantified via a Typhoon FLA7000 phosphoimager.

2-15 MthRnl Proteolytic digestion assay

MthRnl (2.5 μg) was digested by chymotrypsin or trypsin with a titration. The reaction mixtures contain either 1mM MgCl2 or 2 mM EDTA. The reaction mixtures are incubated for 15 mins at 25°C. Digested products were separated by 15% SDS-PAGE and stained with Coomassie blue.

2-16 Primer List

Primer name	primer sequence 5'-3'
T7_sR42template-F	TAATACGACTCACTATAGGGGGATGAAGAGCGTTTACCGGTCTGATTAGCTACCGGATGAAGAGACTGGCACTGGCCGACCCCTTCA
sR42PCRwithT7promoter+5ntprimer-F	GAAATTAATACGACTCACTATAGGGG
sR42_3endprimer-R	TGAAGGGGTCGGCCAGTGCC
T7_sR42templatemutate-F	TAATACGACTCACTATAGGGGGTTTTTGAGCGTTTACCGGTTTTTTTAGCTACCGGTTTTTG

	AGACTGGCACTGGTTTTCCCCTTCA
sR42mutPCRwithT7promoter+5ntprimer-F	GAAATTAATACGACTCACTATAGGGG
sR42mute_3endprimer-R	TGAAGGGGAAAACCAGTGCC
T7_sR31template-R	CGAAGGGCTCGGAGGGTGTTCACGTCATCATCTCCGGGTTCAGAATGCTTGCTACTCGTC ATTGCCCTATAGTGAGTCGTATTAATTTC
T7_sR31CDmutant-A-template-R	CGAAGGGCTGTGAGGGTGTTCACGTCAAAGTCTCCGGGTTGTGAATGCTTGCTACTCAAA GTTGCCCTATAGTGAGTCGTATTAATTTC
T7_sR31endmutant-B-template-R	CCACATACTGTGAGGGTGTTCACGTCAAAGTCTCCGGGTTGTGAATGCTTGCTACTCAAA GTTCCCCTATAGTGAGTCGTATTAATTTC
T7_sR31noterminalstem-C-template-R	CGAATCTATCGGAGGGTGTTCACGTCATCATCTCCGGGTTCAGAATGCTTGCTACTCGTCA TTGCCCTATAGTGAGTCGTATTAATTTC
D non-sR-R	CCACATACTGTGAGGGTGGCAAGCATTCTGAACCCGGAGATGATGACGTGAACACGATC CGATCCCCTATAGTGAGTCGTATTAATTTC
5endstartPrimer	GAAATTAATACGACTCACTATAGGG
sR31-Primer3ends	CGAAGGGCTCGGAGG
sR31-A-Primer3ends	CGAAGGGCTGTGAGGG
sR31-BD-Primer3ends	CCACATACTGTGAGGGTG
F-primer-Trp-intron	ACCCGGAGAAAGGTATGAGG
F-primer-C/D box sR01	GACCCCGTATGGGGGATGAA
F-primer-TK0058	TGACCCCCTAAGGGGAATGA
F-primer-C/D box sR05	CGACCCTCAGTTGGGGATGA
F-primer-C/D box sR15	AGGGACACCTGTCCGATGAT
F-primer-C/D box sR20	GAGGCCTTCTGGCCCGTGAT
F-primer-TK0705	CTCCCCTACTTGCCGGAAAC
F-primer-C/D box sR28	GAGCCTAAGTGGGCGATGAC
F-primer-C/D box sR31	GAGCCCTTCGGGGCAATGAC
F-primer-ncRNA01	AGCCTCCGTCGGCAATGATG
F-primer-C/D box sR34	ACCTTTTCCCAAGGCGTGAT
F-primer-C/D box sR37	GAGCTTTGAAGGGCTATGAT
F-primer-C/D box sR41	GAGGCCACGAGGCCGATGAC
F-primer-C/D box sR42	GACCCCTTCAGGGGGATGAA
F-primer-ncRNA02	CATCGGGCAAATGCCCTCAG
F-primer-C/D box sR46	ACCGGCATGAACGGGATGAT
F-primer-C/D box sR49	CTGACCCACATCGGGGAATG
F-primer-TK2034	GAGGTCTCATGACCGATGAT
F-primer-TK2109	TCATCGCCCTGTAGGGCTCA
F-primer-C/D box sR61	GAGGTGGTTTCACCGATGAT
R-primer-Trp-intron	CTCCAAAGGCTCATCCCCAA
R-primer-C/D box sR01	AGGGCGTGCTCATCACAGTC
R-primer-TK0058	GGCAGGGCTCTTCATCGTTT
R-primer-C/D box sR05	GCCCTCCGCTCATCACTTTT
R-primer-C/D box sR15	AAGACCGATGTCGTCACCGT
R-primer-C/D box sR20	TAGCTGGTGTCCTCACCGCT

R-primer-TK0705	GGATCCTGGGAAGATGGCTT
R-primer-C/D box sR28	CGAGGGTAGTCATCACCGGT
R-primer-C/D box sR31	GTGTTCACGTCATCATCTCCGG
R-primer-ncRNA01	CGTCGATCCCGGGCATCCAT
R-primer-C/D box sR34	AAGCGAAGAGTCATCACCGC
R-primer-C/D box sR37	CGGCCAAGCTCATCATCCAA
R-primer-C/D box sR41	ACCCTCTTTTCCTCATCAGT
R-primer-C/D box sR42	GTGCCAGTCTCTTCATCCGG
R-primer-ncRNA02	AAAGAGAGTGAGCTGAGAGA
R-primer-C/D box sR46	CCGCGAAACCTCTTCACCTC
R-primer-C/D box sR49	TCCTCGTCACCGCTCGGAAC
R-primer-TK2034	GGAACGGCTTCCACATCAAC
R-primer-TK2109	GTTCTCCTTTCTGATTTGTG
R-primer-C/D box sR61	AGTCCCTTCTCATCATCATT

Chapter 3 RNA ligase 3 biological targets identification and the substrates specificity characterization

3-1 Generation of T. kodakarensis ATP-dependent RNA ligase Knock-out strain and Performing Next generation RNA-sequencing

In order to thoroughly define the biological functions of Rnl3 and identify its physiological targets, our team constructed *T. kodakarensis* Rnl3 Knock-out strain (Δrnl). The gene of *TK1545* in KOD1 encodes for *T. kodakarensis* ATP-dependent RNA ligase which was conserved with Rnl3 family. We knock-outed *TK1545* in KOD1 via homologous recombination to replace *TK1545* gene with a selectable marker, orotidine-5'-monophosphate decarboxylase gene (*pyrF, TK2276*). The deletion of *TK1545* was confirmed by PCR[112].

We examined the RNA population between WT and *TK1545* KO via Polyacrylamide gel electrophoresis analysis. The results suggested that the abundance of short RNA species, ranging from 80 to 140 nt, have significantly difference between WT and *TK1545* KO strains. Compared to the WT strain, in the *TK1545* KO strain, the RNAs with around 140nt length was reduced while the RNAs with around 80nt length was increased. Therefore, we performed next generation sequencing with both whole transcriptome RNA (Total RNA sample) and short RNA species (small RNA sample) via Illumnia HiSeq system and Life technologies SOLiD system.

3-2 Illumnia NGS Data Analysis with Total RNA samples

Via Illumina HighSeq2500 technology (51-base; pair-ended read), The RNA-sequencing were performed with the WT whole transcriptome RNA (WT-Total) sample and the *TK1545* KO whole transcriptome RNA (*TK1545* KO-Total) sample. I have examined the Reads Quality via FastQC. For both WT-Total and *TK1545* KO-Total datasets, over 80 percent of the reads had Phred quality scores >35. The WT-

Total dataset had 101.4 million of pair-ended reads in which 94.86% of reads (aligned reads) were mapped to the genome concordantly exactly 1 time. The *TK1545* KO-Total had 58.22 million of pair-ended reads in which 94.54% of reads mapped to the genome concordantly exactly 1 time.

After gene alignment using FeatureCounts[111], I calculated the RPKM (Read Per Kilobase per Million mapped reads) of all *T. Kodakarensis* annotated genes with WT-total dataset and *TK1545* KO-total dataset. There are 2328 annotated genes in *T. Kodakarensis* genome including 46 types of tRNA genes and 4 types of ribosome RNA genes. In the WT-total dataset, the median of RPKM of genes was 9.8 and 1976/2328 genes have RPKM more than 2. The RPKM of TkoRnl gene (*TK1545*) was 9.1, which had a similar expression level to the median. In *TK1545* KO-total dataset, the median of RPKM of genes was 10.9 and 1995/2328 genes had RPKM more than 2. The RPKM of *TK1545* (TkoRnl) was 0.04 while the RPKM of *TK2276* (the selectable marker gene *pyrF*) was 2.96. This suggested that most of KOD1 genes had similar expression level between WT and Δrnl, and the TkoRnl gene (*TK1545*) was indeed knock-outed in the Δrnl sample.

Next, I compared the gene expression level, whose RPKM were more than 2, between WT-total and *TK1545* KO-total dataset (Figure 5A, Supplementary Data 1). I found that 13 different genes were down-regulated more than 2-fold in the KO-total sample and 10 different genes had been up-regulated more than 2 fold (Figure 5B). Many of these genes that exhibit differential gene expression are encoded on the same polycistronic transcription unit; iron transport proteins (TK0714, TK0715, and TK0716), phosphate transporter proteins (TK2060 and TK2061), sulfur reductase subunits (TK2071 and TK2072), and Acetyl-CoA acetyltransferase pathway (TK0179, TK0180, and TK0181). The genes encoded within each operon were either enhanced or reduced to a similar extent in *TK1545* KO-total dataset. In addition, all 4 types of rRNA genes and all 46 types of tRNA genes had similar expression level between WT and *TK1545* KO in which the expression change was less than 2 fold. We need further study to identify how TkoRnl affect the mRNA abundance of those gene.

3-3 Circular RNA identification with Total RNA samples

In Archaea, split tRNA, C/D box sRNA and rRNA intron circularization have

been identified. I wondered whether Rnl3 could possibly act on these non-coding RNAs to regulate their biological function. In 2017, Hubert et al reported that via co-immunoprecipitation, they pulled down 133 different circular RNAs which might interact with the Rnl3 (PabRnl) in *Pyrococcus abyssi*. They further listed 42 highly significant circular RNA molecules where 38/42 circular RNAs were C/D box sRNAs, 1 tRNA intron(tRNA-Trp) and 3 non-annotated circular RNAs. Therefore, I examined the circular RNA alternation between WT and *TK1545* KO mutants.

I thoroughly detected potential circular RNAs in our RNA sequencing datasets using the method described by Mini Danan et[74]. The key point to identify potential circular RNAs was to find the junction read which could span to the circularization junction (Figure 6).

I found 410073 copies of junction reads (0.20% of the data) in WT-total dataset and 281828 copies (0.24% of the data) in *TK1545* KO-total dataset (Supplementary Data 2). With these junction reads, I predicted the potential circular RNAs which have unique start and end sites on the *T. Kodakarensis* genome. I observed that some circular RNAs have adjacent circularization points within one to five nucleotide shifts on the KOD1 genome. This could be because of the slight heterogeneity in choosing the transcription initiation site or the heterogeneity in post-transcription processing. This also could be because of the difficulty to sequence circular RNAs with the illumine sequencing system. Therefore, I clustered such circular RNAs as one distinct circular RNA. In WT-total dataset, I found 12632 types of circular RNAs. In *TK1545* KO-total dataset, I found 9744 types of circular RNAs. As most of the predicted circular RNAs only had one junction read to support its existence, which could be an amplification artifact, I selected the highly supported circular RNAs which had more than 100 copies of junction reads. In the WT-total dataset, 113 types of highly supported circular RNAs were identified. In the *TK1545* KO-total dataset, 63 types of highly supported circular RNAs were identified.

I have compared these highly supported circular RNAs between WT-total dataset and *TK1545* KO-total dataset (Supplementary Data 2: Comparison of WT and tk1545KO sheet). In these circular RNAs, 100/114 types of circular RNAs were located in 16S rRNA – 23S rRNA region. Meanwhile, I also found 10 types of circular RNAs which were significantly reduced after Rnl3 deletion (Supplementary Data 2: Comparison of WT and tk1545KO sheet). After further analysis, I found most of them were C/D box sRNA.

3-4 ribosome RNA and tRNA analysis with Total RNA samples

It has been widely reported that tRNA introns and rRNA processing intermediates are circularized in various archaea species[74, 96]. My both WT-Total and *TK1545* KO-Total datasets had over 90% of reads which were mapped on 16S rRNA-23S rRNA region. Therefore, I performed ribosome RNA analysis with total RNA samples. In Archaea, 16S rRNA and 23S rRNA maturation required BHB motif dependent cleavage and circularization. I identified the circular form of 16S rRNA and 23S rRNA. Seq data reveal that tRNA-Trp intron (1,415 reads), 16S circRNA-1 (6,835 reads) and 23S circRNA-1 (282,955 reads) are likely cleaved at the Bulge-Helix-Bulge (BHB) motifs by tRNA splicing endonuclease and joined by tRNA ligase RtcB to form a circular rRNA processing intermediate, as reported in previous studies. I also identified a novel circular form of 23S rRNA (23S circRNA-2) which was 37nt shorter than the predicted 5'-and 3'-ends of 23S rRNA. Recently *Birkedal et al* reported that in *Pyrococcus furiosus*, 3'-end of 23S rRNA could be fused to the 5'-end by an RNA rearrangement. The 40-nts helix 98 (H98) which was located about 100 nucleotides upstream of the 3'-end of 23S rRNA was excised as a consequence[78]. In my data, I also noticed a low coverage of RNA-Seq reads within the H98 segment, which is similar to *P. furiosus* rRNA (Figure 7B). This suggest that the H98 could also be excised in *T. kodakarensis* during the 23S rRNA maturation.

Our team noted earlier that the abundance of short RNA species was difference between WT and *tk1545* KO (Δrnl). Takagi Yuko therefore performed small RNA cloning with the small RNAs fragments collected from WT or Δrnl via adaptor-mediate RNA ligation, cDNA Reverse transcription and TA cloning. The cDNA of isolated RNA species has the adaptor sequence with a degenerate sequence at the ends, which allowed us to verify that each obtained clone was derived from an independent RNA species and that no single clone was over-represented during PCR amplification.

After alignment analysis, we found that all isolated RNA species were derived from 16S to 23S rRNA operon. In WT, the isolated RNA species were all from 23S rRNA, which were aligned in the range from the excision site of H98 to the 3′-end of 23S rRNA (Figure 7A; Figure 19). However, In the Δrnl, only half of the isolated RNA specie were from the 23S rRNA, which were aligned upstream of the H98

(Figure 7A; Figure 19). This indicated that TkoRnl deletion may cause rRNA processing alternation. Consistent with this finding, from small RNA-Seq datasets, I found that the reads-coverage on the upstream of H98 were more abundant in the Δrnl than WT (Figure 7C). Taken together, from our finding, Rnl3 may participate in rRNA processing, especially in H98 processing.

In *T. Kodakarensis*, there are 46 tRNAs and 2 out of 46 tRNA ($tRNA^{Met}$-CAT and $tRNA^{Trp}$-CCA) have intron. It was reported that in *Sulfolobus tokodaii*, the maturation of these two tRNAs required BHB motif dependent cleavage and the intron of the tRNA will be cut out to form circular RNAs. In my four samples, I identified the circular RNA of the $tRNA^{Trp}$-CCA intron but did not identify the circular RNA of $tRNA^{Met}$-CAT intron. In Hubert's study, via co-immunoprecipitation, they pulled down circular RNAs of the $tRNA^{Trp}$-CCA intron with PabRnl antibody. They suggested that Rnl3 may be responsible for $tRNA^{Trp}$-CCA intron circularization. However, my RNA-seq results suggested that deletion of TkoRnl in *T. Kodakarensis* will not affect $tRNA^{Trp}$-CCA intron circularization. In addition, 16S rRNA and 23S rRNA also used BHB motif for rRNA circularization and maturation. All of them were not affected by TkoRnl deletion. This suggested that Rnl3 dependent circularization is independent of the BHB motif system.

3-5 Illumnia NGS Data Analysis with small RNA samples

Via Illumina HighSeq2500 technology (51-base; pair-ended read), The RNA-sequencing were performed with the WT short RNA species (WT-Small) sample and the KO short RNA species RNA (*TK1545* KO-Small) sample. I have examined the Reads Quality via FastQC. For both WT-Total and KO-Total datasets, over 80 percent of the reads had Phred quality scores >35. The WT-Small dataset had 83.5 million of pair-ended reads and the *TK1545* KO-Small dataset had 72.7 million of pair-ended reads. However, millions of adaptor sequence and RNA PCR primer sequence were identified by FastQC and filtered by Cutadapt program. Eventually the WT-small dataset had 22.21 million of pair-ended reads in which 79.94% reads mapped to the genome concordantly exactly 1 time. *TK1545* KO-small sample were collected 16.37 million of reads in which 13.33 million (81.43%) reads mapped to the genome concordantly exactly 1 time. I also analyzed and compared the gene expression level between WT-small and TK1545 KO-small dataset (Supplementary Data 1).

3-6 Circular RNA analysis with Small RNA samples

I then thoroughly detected potential circular RNAs in WT-small and KO-small samples with the same method. In small RNA samples, only short RNAs species were collected for RNA sequencing. I found 147251 copies of junction reads in the WT-small sample and 69848 copies in the *TK1545* KO-small sample (Supplementary Data 3). In the WT-small sample, after clustering similar circular RNAs, I predicted 2237 types of circular RNAs. In the *TK1545* KO-small sample, I found 1402 types of circular RNAs. In the small RNA datasets, only short RNAs were collected for RNA sequencing, therefore, predicted circular RNAs were further filtered to be shorter than 10 kb. I identified 280 distinct circular RNAs in the WT-small sample and 154 distinct circular RNAs in *TK1545* KO-small sample. As most of the predicted circular RNAs only had one junction read to support its existence, which could be an amplification artifact, I selected the highly supported circular RNAs which had more than 20 copies of junction reads. I identified 31 highly supported circular RNAs in the WT-small sample and only 8 highly supported circular RNAs in the *TK1545* KO-small sample (Table 1). The 8 highly supported circular RNAs were both highly supported in the WT-small sample and the *TK1545* KO-small sample.

3-7 Circular RNA alternation in TkoRnl deletion mutant of T. Kodakarensis

From the four RNA-seq samples, I identified many potential circular RNAs in both the WT samples and *TK1545* KO samples while I also identified many potential circular RNAs which only existed in the WT samples. This suggested that Rnl3 may be responsible for this circular RNA formation. Therefore, I defined the potentially Rnl3 related circular RNAs as the circular RNAs which were highly supported by junction reads and had more than 10 times copies of junction reads in the WT sample than in the *TK1545* KO sample.

After comparing WT-small and *TK1545* KO-small samples, I also identified 24 distinct TkoRnl related circular RNAs (Table 1). 18 out of 24 circular RNAs were C/D box sRNAs. The remaining 6 circular RNAs were likely to be C/D box sRNAs. They have similar lengths, around 60nt but less conserved C/D box elements.

To further understand the relationship between Rnl3 and C/D box sRNAs, I tried

to examine all C/D box sRNAs between WT and TkoRnl deletion samples. As C/D box sRNAs have not been annotated in *T. Kodakarensis* by any previous study, I predicted 61 different C/D box sRNAs in KOD1 annotated genomes by searching conserved C/D box motif sequences. I first examined the C/D box sRNA expression level in the WT-small and *TK1545* KO-small samples by counting the aligned reads in each C/D box sRNA region. Comparing the WT-small and *TK1545* KO-small samples, 55/61 C/D box sRNAs had similar expressions, 4/61 C/D box sRNAs had 2-fold down regulation in the *TK1545* KO-small dataset and 2/61 C/D box sRNAs had 2-fold up regulation in the *TK1545* KO-small dataset (Table 2). I then checked whether potential circular RNAs existed around those C/D box sRNAs locations in the WT-small and *TK1545* KO-small samples. I found that 26/61 different C/D box sRNAs had circular forms. 18 out 26 C/D box sRNAs were highly supported by junction reads as mentioned above. The remaining 6 out of 26 different C/D box sRNAs had circular forms but the junction reads for supporting these sRNAs were less than 21 copies. In these 26 different C/D box sRNAs, 24/26 different C/D box sRNAs were TkoRnl related, which meant that these circular C/D box sRNAs had more than 10 times copies of junction reads in the WT sample than in the *TK1545* KO sample.

We also performed RNA-sequencing with SOLid system. The both WT-SOLid dataset and *TK1545* KO-SOLid dataset have low reads quality and not passed FastQC checking. I performed circular RNA analysis (Supplementary Data 4, GSE207552) and only used these datasets for confirming the circular RNAs identified in Table1. As showed in Table 3, Most of the highly supported circular RNAs found in WT-Small dataset were also identified in SOLid datasets.

Hubert et al reported that Rnl3 interacted with circularized C/D box sRNAs in *P. abyssi*. Therefore, I analyzed the circular RNAs in *P. abyssi* using the RNA sequencing data (GSM1401488) from Toffano-Nioche's group's study[113]. I found that 39 different C/D box sRNAs had circular forms among the 59 annotated *P. abyssi* C/D box sRNAs (Table 2). After homolog searching with 61 predicted KOD1 C/D box sRNAs and 59 annotated *P. abyssi* C/D box sRNAs, I found 30 different conserved C/D box sRNAs. 17/30 different conserved C/D box sRNAs had circular forms in both the KOD1 WT-small samples and *P. abyssi* samples. 10/17 conserved circular C/D box sRNAs were highly supported, which had more than 10 copies of junction reads in both the *T. Kodakarensis* WT-small samples and *P. abyssi* samples. Those results suggested that Rnl3 pathway is widely used for C/D box sRNA

circularization in Archaea.

3-8 Circular RNA terminal structure analysis

I examined whether a common sequence or structure exists in the 24 highly supported circular RNAs regulated by TkoRnl. 18 out of 24 circular RNAs were C/D box sRNAs. The remaining 6 circular RNAs were likely to be C/D box sRNAs. They have similar lengths, around 60nt but less conserved C/D box elements. I predicted the terminal structure of these circular RNAs by first predicting the C/D box decided kink-turn structure of each circRNA and then predicting the potential nucleotide hybridization. I found that the terminal sequences of these circRNAs are GC-rich and the most of these circRNAs (21 out of 24) could have over three base pairings to form a terminal stem (Figure 8A). In contrast, I analyzed the predicted C/D box sRNAs which have no circRNA reads and have more than 1000 aligned reads in Table 2. I found that 16 out of 17 these non-circular C/D box sRNAs have less than two base pairings, which suggest that these non-circular C/D box sRNAs may not form a terminal stem (Figure 8B). From the secondary structure analysis, I suggest that hybridization between the two terminal ends could be critical for RNA circularization.

3-9 Circular RNA detection via RT-PCR

To independently detect the identified circular RNAs in RNA-seq, I selected 18 candidates for Reverse Transcription-PCR (Figure 9). I selected tRNA-Trp intron as my positive control because tRNA-Trp introns is circularized by RtcB and I detected the circRNA reads of tRNA-Trp introns in both WT-datasets and *TK1545* KO datasets. Forward junction primers and reverse primers were designed for each predicted circular RNA to perform RT-PCR (see 2-16 primer list). If there are circular RNAs, the Primer sets will generate a ladder of DNA fragments by RT-PCR as showed in positive control (Figure 9 (2)). If there is no circular RNAs, PCR results will show nothing or only primer dimer (Figure 9 (1) or Figure 6 (5)-*TK1545* KO). As a result, I detected 16 out of 18 candidate circRNAs in WT, but not in *TK1545* KO. This suggested that the predicted circRNAs from RNA-seq are existing in the *T. Kodakarensis* cells. I failed to detect the circular form of C/D box sR01 and TK2109 in both WT and TK1545 KO (Figure 9 (3) and (26), respectively). This is possibly

due to a heterogeneous mixture of circular junction sequences in these RNAs, which may have affected the PCR amplification step.

3-10 Rnl3 Ligation specificity on C/D box sRNAs

The next question was whether Rnl3 could recognize C/D box sRNAs specifically. I hypothesized that Rnl3 could recognize C/D box sRNAs specifically via terminal stems and C/D box sequences. Via *in vitro* transcription, I synthesized two C/D box sRNAs, sR31 and sR42, which had top 2 junction read counts, 4 bp terminal stems, and had homologs in *P. abyssi*.

We previously showed that TkoRnl is capable of circularizing 24-mer single-stranded RNA, but the circularization activity was weak compared to the M. thermoautotrophicus enzyme (MthRnl)[114]; thus, MthRnl was used for the ligation assay. MthRnl is a homolog of TkoRnl (NCBI BLAST E-value of 6 × 10−60). Previously, RNA ligase studies usually used 24 mer RNAs for characterizing RNA ligase biochemical features via *in vitro* ligation assay. The synthesized RNA substrates were labeled by T4 polynucleotide kinase (PNK) with radioactive ^{32}p-ATP at 5' end of RNA to form 5'-*pRNAs. I then performed *in vitro* RNA ligation assay with MthRnl and the labeled RNA substrates. After ligation assay, the ligation products were separated by UREA polyacrylamide gel electrophoresis (UREA-PAGE) and visualized by autoradiography. Here I compared the ligation efficiency of MthRnl with 24mer, sR31 and sR42. In standard *in vitro* Rnl ligation assay, MthRnl could ligate sR31 and sR42 with the same result as 24mer RNAs (Figure 10 and Figure 11) at 70 °C. I then performed ligation assay with 1mM ATP. There are three steps in ATP dependent RNA ligation (Figure 1). In step 1, RNA ligase reacted with ATP to form Enzyme-AMP (EpA). In step 2, RNA ligase will transfer the AMP to 5' phosphorylated RNA to form AppRNA. In step 3, RNA ligase will induce 3' hydroxy of RNA to attack the 5' AppRNA to complete RNA ligation and release AMP. With non-optimal substrates and 1mM ATP, RNA ligation assay will be suspended in step2 and AppRNA will be accumulated[4]. We suggested that with ATP, the second EpA formation will occur before step 3 and then EpA form of Rnl3 will not be able to conduct ligation steps, so AppRNA will be accumulated. However, under 1mM ATP condition, sR31 and sR42 were still ligated by MthRnl, instead of AppRNA accumulation as 24 mer RNA did (Figure 10). This suggests that when the substrates

were C/D box sRNAs, Rnl3 conducts ligation steps before the second EpA formation, which indicates that C/D box sRNAs are better substrates for Rnl3 than 24mer RNAs.

I then hypothesized that Rnl3 could recognize C/D box sRNAs specifically via terminal stems and C/D box motif sequences. I synthesized sR31-a and sR42-a in which C/D box motif sequences were mutated to other nucleotides. I synthesized sR31-b in which RNA sequences near 3' end was mutated to impair terminal stem formation. I synthesized sR31-c in which both the C/D box motif sequences and the terminal stems were mutated. In addition, as sR31 has 67nt, the difference of RNA length may cause ligation efficiency change between sR31 and 24mer RNAs. Therefore, I synthesized a 67mer linear RNAs without any predicted secondary structure.

The RNA ligation assay showed that sR31-a and sR42-a have the same ligation efficiency with the wild type sR31 and sR42(Figure 10 and Figure 11). This suggested that C/D box motif sequences were not essential for Rnl3 recognition. The RNA ligation assay with sR31-b showed that Rnl3 had less than half of catalytical activity with sR31-b, compared with Wild Type sR31 (Figure 11). This suggested that terminal stem is important for Rnl3 to ligate C/D box sRNAs. The RNA ligation assay with sR31-c and 67mer linear RNA showed that MthRnl was not able to ligate 67mer linear RNAs and mutated sR31 in which both C/D box motif and terminal stem sequences were mutated. To some degree, sR31-c could also be considered as random linear RNA because sR31-c had not the C/D box sequence feature and had no RNA structure feature of the terminal stem. To sum up, we suggested that both the C/D box and terminal stems could help Rnl3 to recognize C/D box sRNAs. Terminal stem mutation will reduce half Rnl3 ligation efficiency on C/D box sRNAs. Although to disrupt C/D box will not affect Rnl3 ligation efficiency, yet double mutation sR31-c can not be ligated by MthRnl, which suggested C/D box sequences had some positive effects on Rnl3 related circularization.

3-11 T4Rnl2 Ligates 67mer linear RNAs and C/D box sRNAs without specificity

It is possible that the 5' and 3' ends of linear RNAs are far from each other while the terminal stem and C/D box sequence of C/D box sRNAs bring 5' end near to 3' end, which may facilitate ligation efficiency of RNA ligases. To figure out this, I

checked whether T4Rnl2 has lower ligation efficiency on 67mer linear RNAs than C/D box sRNAs. In standard Rnl ligation assay, T4Rnl2 could ligate 67mer linear RNA, sR31 and sR31-c efficiently (Figure 12) at 22 ℃. In the previous studies, RNAs with about 20mer were widely used for determining ligation efficiency of various RNA ligases. Our MthRnl could ligate 24mer RNAs but not 67mer linear RNAs while T4Rn2 could ligate both 24mer RNAs and 67mer linear RNAs. It is possible that Rnl3 could recognize the length of the substrates. The ligation reaction of MthRnl was performed at 70 ℃ which was much higher than T4Rnl2's. It is also possible that higher reaction temperature caused the 5' end of RNAs to have more difficulty to meet 3' end of RNAs in 67mer linear RNAs than 24mer RNAs. Therefore, I performed RNA ligation assay with T4Rnl2 and MthRnl at the same temperature, 37 ℃. At 37 ℃, T4Rnl2 still could ligate sR31 and sR31-c and MthRnl could ligate sR31 only but not sR31-c (Figure 13). At the same time, MthRnl could only adenylate 24mer RNA at 37 ℃. It suggested that Rnl3 had more strict substrate specificity at RNA length around 60nt than 24mer RNA.

3-12 Rnl3 Ligation products verification via RNAse R

The RNAs could be separated differently by UREA-PAGE with its linear form or circular form. The circular form of 24mer RNAs ran faster than the linear form of 24mer RNAs in UREA-PAGE (Figure 10). The circular form of 67nt sR31 or 70nt sR42 ran slower than their linear form in UREA-FAGE (Figure 11). To verify the Rnl3 ligation products, I used the exo-ribonucleolytic enzyme RNase R[115]. This exonuclease degrades linear RNA molecules in a 3' to 5' direction but leaves circular RNA molecules and some highly structured RNAs which do not expose their 3' end as a single strand. I verified the MthRnl ligation products of sR31, sR31-a, sR31-b and sR42 (Figure 14). The RNAse R could digest sR31, sR31-a and sR31-b but could not digest their ligation products. The RNAse R had difficulty to digest sR42. The RNAse R have slow digestion speed with sR42 while RNAse R can not digest circular sR42 at all. This suggested that the ligation products were of circular forms but not of intermolecular ligation which combines two RNAs into one longer RNA.

3-13 A Novel function of MthRnl : 3'A-deletion function

Toward identifying the optimal substrate for MthRnl ligation activity, our laboratory prepared various 21mer RNA substrates with different sequences and structures. Shigeo Yoshinari found that incubation of MthRnl with the 21mer RNAs with an adenine at 3' end generated a novel RNA species that migrated between the input linear pRNA and the circular RNA. He further demonstrated that the novel RNA species was 3'-deadenylated RNA[114]. MthRnl cleaved the adenine on the 3' end of RNA substrates. I also verified this novel function of MthRnl. I used unstructured 24mer RNAs, hairpin-structured RNAs with dual overhang (named 4-3AG) and hairpin-structured RNAs with 3' overhang (named 1-6AG) (Figure 15). After the MthRnl ligation, about half of 4-3AG RNAs were circularized, and another half of 4-3 AG RNAs were de-adenylated (Figure 15). However, after the MthRnl ligation, almost all 1-6AG RNAs were de-adenylated (Figure 15). This suggested that MthRnl could efficiently remove 3'-adenosine from RNAs that have more than 5nt flexible and overhanged 3' end. However, it was less efficient in the cases of RNAs that bear a recessed 3'-end.

I next analyzed the optimal pH and temperature for MthRnl to remove 3'-adenosine. I performed MthRnl 3'-deadenylation assay with 4-3AG RNA. I found that with an increased pH, the yield of 3'-deadenylated RNAs increased and circular RNA levels declined. The 3'-deadenylated 4-3AG RNA was not detectable under pH 5.5(Figure 16A). To find the optimal reaction temperature, I performed MthRnl 3'-deadenylation assay with 4-3AG RNA and set the temperature from 35℃ to 90℃. I found that MthRnl could deadenylate the RNA substrates over a wide range of temperatures and the optimal temperature was 55 ℃ while optimal RNA ligation temperature was 70 ℃ (Figure 16B). The two activities, ligation and 3'-deadenylation, have two distinct optimal temperatures and pH values, implying that the ligation and 3'-deadenylation reactions were independent of one another.

The next question is whether 3'-deadenylation needed the ATP binding site of MthRnl. I have performed MthRnl 3'-deadenylation assay with 24mer RNA and 4-3AG RNA. I found that adding ATP or AMP-PNP could inhibit the MthRnl 3'-deadenylation activity with these two RNAs (Figure17). AMP-PNP is an ATP analog which can occupy the ATP binding site but can not react with the ligase and form EpA in the ligation reaction step two. This suggested that the ATP binding site of

MthRnl were essential for 3’-deadenylation. Our group showed that Divalent cation such as Magnesium, was essential to MthRnl for RNA ligation and RNA 3’deadenylation. I analyzed whether divalent cation affected RNA substrate binding or protein folding via EMSA and Limited Proteolytic Digestion. The results suggested that divalent cation did not affect RNA substrate binding or protein folding (Figure18).

Chapter 4 Discussion

4-1 TkoRnl deletion alter the gene expression profile

In my project, we generated an allelic knock-out of Type 3 ATP-dependent RNA ligase (Rnl3) in T. kodakarensis to determine the biological targets of the Rnl3. Although TkoRnl was not essential for the growth of T. kodakarensis under standard laboratory conditions, the Polyacrylamide gel electrophoresis analysis showed the transcriptome between Wild Type (WT) strain and Knock-out (KO) strain have significant difference. The gene expression profile suggested that 13 different genes were down-regulated more than 2 fold in the KO sample and 10 different genes had been up-regulated more than 2 fold. Many of these genes that exhibit differential gene expression are encoded on the same polycistronic transcription unit. We need further study to identify how TkoRnl affect the mRNA abundance of those gene. From the NGS data, I did not found mRNA splicing on these genes, therefore, it is more possible that TkoRnl acts in an indirect way to affect these genes.

4-2 Rnl3 is responsible for the circularization of C/D box sRNAs

In archaea, various types of C/D box sRNA have circular forms; Trp tRNA gene has an intron which also has a circular form; 16S rRNA and 23S rRNA have circular forms as intermediates during rRNA maturation [74, 102-105]. Therefore, I tried to figure out whether Rnl3 is responsible for generating circRNAs by comparing the circRNA reads obtained from the small RNA-Seq datasets of WT and TkoRnl KO. Under my criteria, I have identified 31 putative circRNAs in WT sample and 8 putative circRNAs in KO sample. From these putative circRNA, I found 7 circRNA including the Trp-tRNA intron and ribosomal RNA, have no abundance change after TkoRnl deletion. It has been reported that GTP-dependent RNA ligases (RtcB) are responsible for tRNA halves ligation during tRNA maturation via BHB splicing pathway (31). In *T. Kodakarensis*, 16S and 23S ribosomal RNA also form BHB motif and use same BHB splicing pathway to generate circular intermediates of 16S and

23S rRNA, which suggests that RtcB are responsible for 16S and 23S rRNA circularization in *T. Kodakarensis*. In Hubert's study, via co-immunoprecipitation, they pulled down circular RNAs of the Trp-tRNA intron with pab-Rnl3 antibody. They suggested that Rnl3 may be responsible for Trp-tRNA intron circularization. However, my RNA-seq results suggested that deletion of Rnl3 in *T. Kodakarensis* will not affect Trp-tRNA intron circularization. In addition, the circularized Trp-tRNA intron was also detected in my RT-PCR analysis in abundance, both in WT and TkoRnl KO samples. This suggested that Rnl3 dependent circularization is independent of the BHB splicing pathway.

From the 31 putative circRNAs, I also identified 24 putative circRNAs which are significantly reduced after TkoRnl deletion. It is highly possible that type 3 RNA ligase is involved in this RNA circularization. 18 out of 24 circular RNAs were C/D box sRNAs. The remaining 6 circular RNAs were likely to be C/D box sRNAs. They have similar lengths, around 60nt but less conserved C/D box elements. As C/D box sRNAs have not been annotated in *T. Kodakarensis* by any previous study, I predicted 61 different C/D box sRNAs in *T. Kodakarensis* annotated genomes by searching conserved C/D box motif sequences. I found that 26/61 different C/D box sRNAs had circular forms and 24/26 different C/D box sRNAs have significantly reduced their circular form after TkoRnl deletion. This suggests that TkoRnl are responsible for circularization of C/D box sRNAs. Hubert et al reported that Rnl3 had interaction with circularized C/D box sRNAs in *P. abyssi*. Those results suggested that Rnl3 pathway is widely used for C/D box sRNA circularization in Archaea.

The functional significance of circular C/D Box sRNA is unclear. C/D box sRNAs have been reported to function as a guide RNA for methylating tRNAs and rRNAs. Circular C/D Box sRNA may alter the specificity to guide RNA to regulate rRNA methylation. However, we did not detect significant differences in the expression, or the overall read coverage, of rRNA genes from the whole transcriptome RNA-seq analysis, between the WT and TkoRnl deletion strains. Researcher suggested that Archaea species usually lived in high temperature condition, so the forming circular RNA may help C/D box sRNAs stable. My RNA-seq data showed that most of the C/D box sRNAs have similar expression level between WT and TkoRnl KO samples. However, there are still 4 different C/D Box sRNAs which have reduced RNA abundance after TkoRnl deletion. While it is clear that TkoRnl is responsible for C/D box sRNA circularization, further analysis is necessary to

evaluate the biological function of Rnl in archaea.

The BHB processing pathway was widely used in archaea and were found in tRNA processing, rRNA processing, and mRNA processing[65, 69]. After cleavage, RtcB ligates the tRNA fragments directly to form the matured tRNA and in some case, the released tRNA introns were also circularized by RtcB.

In Hubert's study, via co-immunoprecipitation, they pulled down circular RNAs of the $tRNA^{Trp}$-CCA intron with pab-Rnl3 antibody. They suggested that Rnl3 may be responsible for $tRNA^{Trp}$-CCA intron circularization. However, my RNA-seq results suggested that deletion of Rnl3 in *T. Kodakarensis* will not affect $tRNA^{Trp}$-CCA intron circularization. In addition, 16S rRNA and 23S rRNA also used BHB motif for rRNA circularization and maturation. All of them were not affected by Rnl3 deletion. This suggested that Rnl3 dependent circularization is independent of the BHB motif system.

4-3 The terminal stem structure of C/D box sRNA helps Rnl3 to efficiently ligate its substrates.

The next question was whether Rnl3 could recognize C/D box sRNAs specifically. I hypothesized that Rnl3 could recognize C/D box sRNAs specifically via terminal stems and C/D box sequences. I synthesized mutated C/D box sRNAs (sR31-a) in which C/D box motif sequences were mutated to other nucleotides. I synthesized sR31-b in which RNA sequences near 3' end was mutated to impair terminal stem formation. I synthesized sR31-c in which both the C/D box motif sequences and the terminal stems were mutated. In addition, as sR31 has 67nt, the difference of RNA length may cause ligation efficiency change between sR31 and 24mer RNAs. Therefore, I synthesized 67nt linear RNAs.

The RNA ligation assay showed that sR31-a have the same ligation efficiency with the wild type sR31(Figure 11). This suggested that C/D box motif sequences were not essential for Rnl3 recognition. The RNA ligation assay with sR31-b showed that Rnl3 had less than half of catalytical activity with sR31-b, compared with Wild Type sR31 (Figure 11). This suggested that terminal stem is important for Rnl3 to ligate C/D box sRNAs. The RNA ligation assay with sR31-c and 67nt linear RNA showed that MthRnl was not able to ligate 67nt linear RNAs and mutated sR31 in which both C/D box motif and terminal stem sequences were mutated. To some

degree, sR31-c could also be considered as random linear RNA because sR31-c had no the C/D box sequence feature and had no RNA structure feature of the terminal stem.

From RNA-seq analysis, I found not all C/D box sRNAs were circularized in *T. kodakarensis*. From the ligation assay, I suggested that the conserved C/D box sequence element was not sufficient for circularization. There are other missing regulating factors for the circularization of C/D box sRNAs. The Rnl3 could not circularize unstructured RNA of a similar length or C/D box sRNAs that have disrupted both terminal stem structures and C/D box motif. I concluded that Rnl3 may preferentially recognize the terminal stem, and the proximity of the two ends could be critical for intramolecular ligation.

4-4 MthRnl has a novel 3'-deadenylation function.

Dr. Shigeo Yoshinari has identified a novel function of MthRnl that MthRnl could selectively cleave one adenosine on the 3' end of RNA substrates to generate a new 2',3'-cyclic phosphate end.

I prepared several RNA substrates with different secondary structures. I found that MthRnl could efficiently remove 3'-adenosine from RNAs that have more than 5nt flexible and overhanged 3' end. However, it was less efficient in the cases of RNAs that bear a recessed 3'-end. These results suggest that MthRnl prefers a substrate with an unpaired /unstacked single-stranded RNA end. Whether MthRnl adenylates the 5'-PO4 ends for ligation or cleaves the 3'-adenosine likely depends on the accessibility of each RNA terminus. MthRnl could deadenylate the RNA substrates over a wide range of temperatures. The optimal temperature was 55 °C while optimal RNA ligation temperature was 70 °C. The two activities, ligation and deadenylation, have two distinct optimal temperatures and pH values, implying that the ligation and 3'-deadenylation reactions were independent of one another.

To figure out whether 3'-deadenylation needed the ATP binding site of MthRnl, Dr. Shigeo and I analyzed the effects of ATP, AMP-PNP, AMP and other rNTPs on MthRnl 3'-deadenylation activity. We noticed that adding ATP and AMP-PNP inhibited the MthRnl 3'-deadenylation activity while AMP or other rNTPs did not. This suggested that the ATP binding site of MthRnl were essential for 3'-deadenylation. Our group also performed MthRnl 3'-deadenylation assay with

different MthRnl mutants[114]. The MthRnl mutants which are not able to generate EpA, have impaired MthRnl 3'-deadenylation activity. However, the MthRnl mutants which can not continue the step 2 and step 3 ligation reaction, still have MthRnl 3'-deadenylation activity. Therefore, it is possible that ATP binding site could recognize the adenosine at the 3'-end of the RNA in step 1 of the ligation reaction.

The different biochemical characteristics of RNA ligation and 3'- deadenylation suggested that MthRnl have specified biological function with its 3'- deadenylation function on the 3'-adenylated RNAs to regulate their processing and downstream biological events. There are numerous RNAs that are terminated by 3'-adenosine such as tRNAs and some mRNAs. Some archaea species such as methanogens, have RNAs with heteropolymeric poly(A)-rich tails and regulate the RNA degradation by removing the poly(A)-rich tails. MthRnl may be responsible for the RNA degradation regulation by converting 3'-OH ends to 2', 3'-cyclic phosphate ends. MthRnl may also function in forming RtcB substrates which have a 2', 3'-cyclic phosphate ends.

4-5 Rnl3 may affect 23S rRNA processing

In *Pyrococcus furiosus*, the H98 of 23S rRNA was found be precisely excised[78]. In the 23S rRNA maturation process, the 23S rRNA is circularized by the tRNA splicing-like mechanism and the helix 98(H98) was removed in the later process which's mechanism is unknown. From the RNA-seq analysis, I found that T. KOD1 also excise its H98 from 23S rRNA, consistent with the finding that the H98 is not present in the cryo-EM structure of *T. kodakarensis* 70S rRNA[78, 116]. Also, from the RNA-seq analysis, I found that the reads mapping pattern of H98 is different between WT and TkoRnl deletion strain. My group member Dr. Takagi found that *T. kodakarensis* accumulates ~90 nts fragments consisting of a sequence that matches the H98 3'-cleavage site to the predicted 3'-end of 23S rRNA. We speculate that the excision of H98 releases the 3'-end fragment and may have accumulated in *T. kodakarensis*. Intriguingly, we did not retrieve the same fragments from the TkoRnl deletion strain. Instead, we recovered fragments that mapped upstream of H98 near the SRL. SRL interacts with the translational elongation factors that hydrolyze GTP during translocation and the cleavage or modification by ribotoxins could block ribosome translocation[117-119]. Therefore, Rnl3 may participates in the H98 processing. It is possible that Rnl3 joins the breakage upon excision of H98. However,

I fail to detect any RNA-Seq reads suggesting such “cis-splicing” events near the 3’-end of 23S rRNA. Therefore, Rnl3 ligation activity may not act on rRNA directly. It is plausible that Rnl3 may act indirectly through the formation of circular C/D box sRNA, which in turn could regulate rRNA processing.

Chapter 5 Figures and legends

Figure 1
ATP dependent RNA ligase
Step 1: **E+Appp→EpA + pp**
Step 2: **EpA+pRNA→E + AppRNA**
Step 3: **E+AppRNA+RNA-OH→RNA-p-RNA +Ap + E**

GTP dependent RNA ligase
Step 1: **E+Gppp→EpG + pp**
Step 2: **EpG+RNA>p→E + RNAppG**
Step 3: **E+OH-RNA+RNAppG→RNA-p-RNA +Gp + E**

Figure 1 Ligation mechanism of ATP dependent RNA ligase and GTP dependent RNA ligase. In the ATP dependent ligation reaction, the RNA ligase catalyzes the phosphodiester bonds formation between a 5’ phosphate and 3’ hydroxyl termini of RNAs by three reaction steps[3]. In step one, the RNA ligase consumes a ATP to form a covalent ligase-(lysyl-N)-AMP intermediate. In step two, the RNA ligase catalyzes the formaiton of 5’ adenylylated RNA (AppRNA), an inverted pyrophosphate bridge structure by transferring the AMP from the ligase to the phosphate on the 5’ end of the target RNA. In step three, the RNA ligase catalyzes the 5’ - 3’ phosphodiester bonds formation and liberates AMP by inducing a 3’-hydroxyl group of a RNA to attack the 5’ adenylylated RNA (AppRNA). In the RtcB ligation, RtcB could directly ligate the RNA with 5’ -OH end and the RNAs with 2’,3’ cyclic phosphate (RNA>p)(or the RNA with 3’ phosphate end) to form a 5’ - 3’ phosphodiester bonds. RtcB ligation reaction also have three steps.[1]. In step one, RtcB consumes a GTP to form a covalent ligase-(histidinyl-N)-GMP intermediate. In step two, RtcB transfer the GMP to the 3’ end of RNA to generate a polynucleotide-(3’)pp(5’)G intermediate (RNAppG). In this step, if the 3’ end of RNA is 2’,3’ cyclic phosphate, RtcB will hydrolyze the termini to 3’-phosphate end and then transfer the GMP. In step three, RtcB join the two end of RNA to form a 5’ - 3’ phosphodiester bonds by inducing the 5’-OH end to attack the RNA-pp-G end.

Figure 2

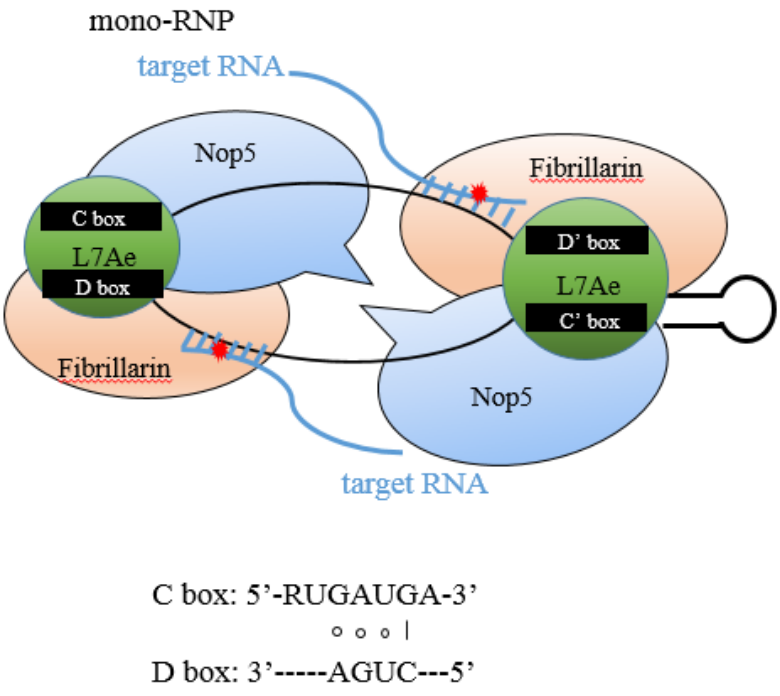


Figure2. The C/D box sRNP structure illustration. In the sRNP structure, two L7Ae protein first recognize K-turn and K-loop structure of C/D box sRNA (black line) and bind them respectively for stabilizing. Then another two core proteins, Nop5 and fibrillarin, bind L7Ae to assemble as active ribonucleoprotein (RNP) complex. The red asterisk indicates the methylation site on the target RNA[88]. This figure is adopted from Breuer et al[88].

Figure 3

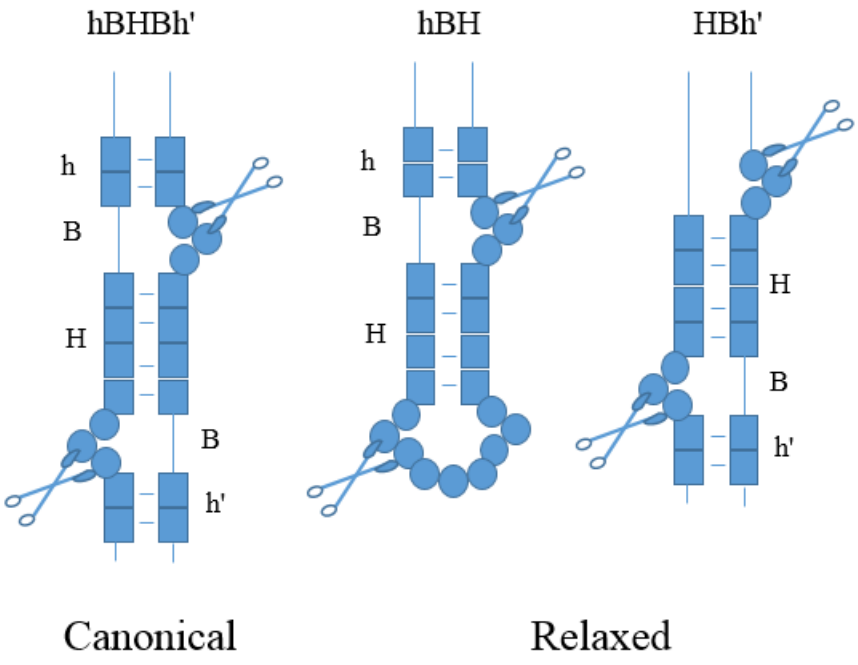


Figure3. Bulge-Helix-Bulge RNA (BHB motifs). The canonical BHB motif have a 3-nt bulges, follow with a central 4-bp helix and end with a 3-nt bulges showed in left panel. Right panel shows There are also relaxed versions of the BHB motif showed in right panel. This structure helps the splicing endonuclease EndA to specifically cut the splice sites. The BHB processing pathway was widely used in archaea and were found in tRNA processing, rRNA processing, and mRNA processing[62, 67]. The figure is adopted from Clouet-d'Orval, B., et al.[62]

Figure 4

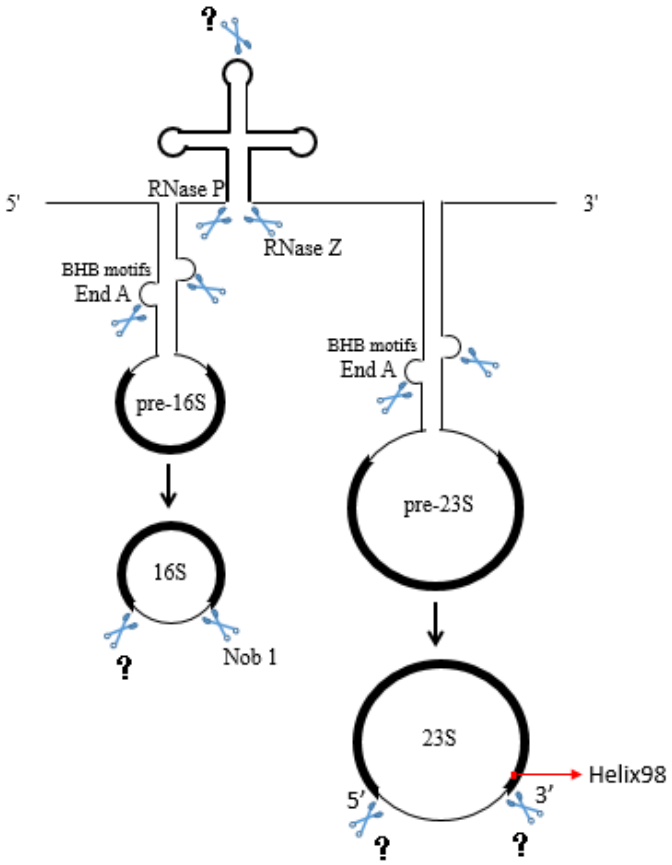


Figure 4. Archaea rrn operons. In the some Archaea species such as *T.kodakarensis*, the 5S rRNA gene is scattered on the chromosome with an independent rrn operon while the 16S rRNA and 23S rRNA are contained together in one rrn operon. The ITS usually encodes a alanyl-tRNA gene. The pre-16S rRNA and pre-23S rRNA will form BHB structures and are cleaved by the endonuclease EndA [69, 73]. After BHB motif cutting, RNA ligases (possibly RtcB) ligate the 16S and 23S pre-rRNAs to circular 16S and 23S pre-rRNAs. The circular 16S and 23S pre-rRNAs are only intermediates and receive multiple endonucleolytic cleavages and exonucleolytic trimming to form the mature ribosomal RNA. This process still needs further study to reveal all the rRNA maturation steps. Helix 98 (indicated in red box) is located at the 3'end of 23S rRNA. The figure is adopted from Clouet-d'Orval, B., et al[62].

Figure 5

A **B**

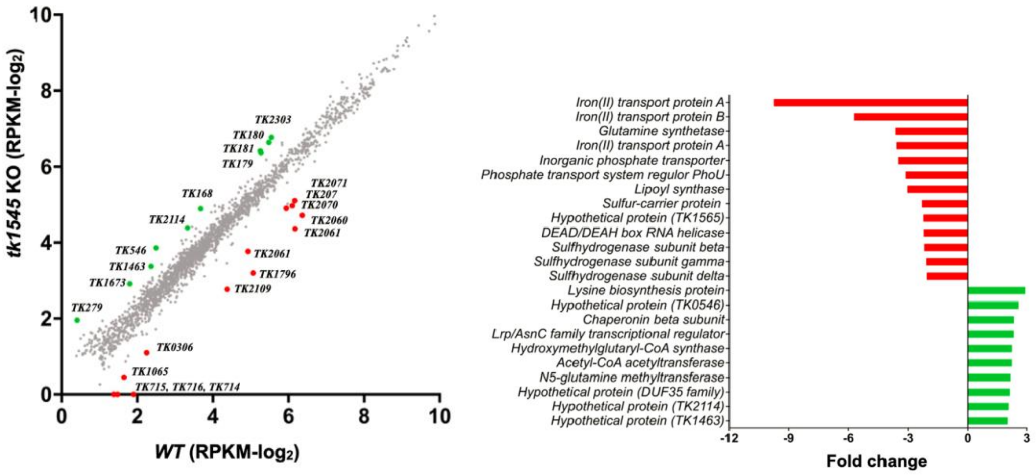


Figure 5 Gene abundance alternation after TkoRnl deletion. (A). T.Kodakarensis transcriptosome analysis between WT and Δrnl. Next generation sequencing was performed with Total RNA isolated from WT and Δrnl cells. After Gene profile analysis, the comparison between WT and Δrnl were drawn in Figure C according to the gene's read per million mapped reads (RPKM) values from WT and Δrnl cells. Any gene with less than one RPKM value was set to 0. Differentially expressed genes (over 2 fold RPKM value change) were drawn with colored dots(Green dots : up-regulated after TkoRnl deletion, Red dots : down-regulated). The RPKM for *tk1545* from the WT and Δrnl were 9.05 and 0.04, respectively. (B) The genes list which have greater than 2 fold change after TkoRnl deletion.

Figure 6

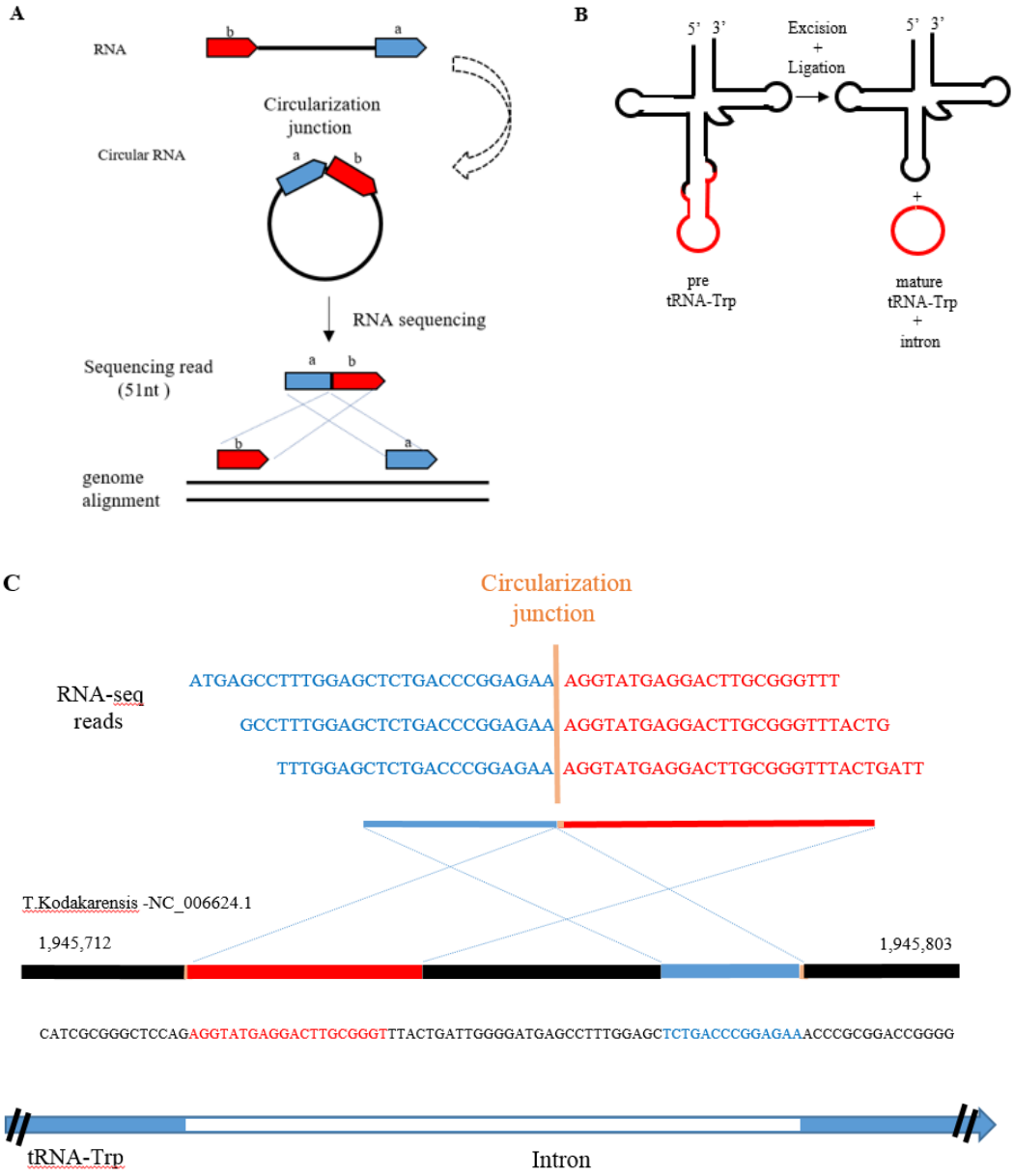


Figure 6. Illustration for circular RNA analysis from RNA-seq data. (A) The circular RNA junction sites were identified by collecting the RNA-seq reads which maps to the reference genome with an inverse order[74]. (B) Schematic representation of the T.Kodakarensis tRNA-Trp, which contains a 62nt intron that is cleaved in the process of tRNA maturation and becomes a stable RNA circle. (C) Some example Reads which mapped to circular junction site of the tRNA-Trp intron were shown here. This figure is adopted from Mini Danan et[74].

Figure 7

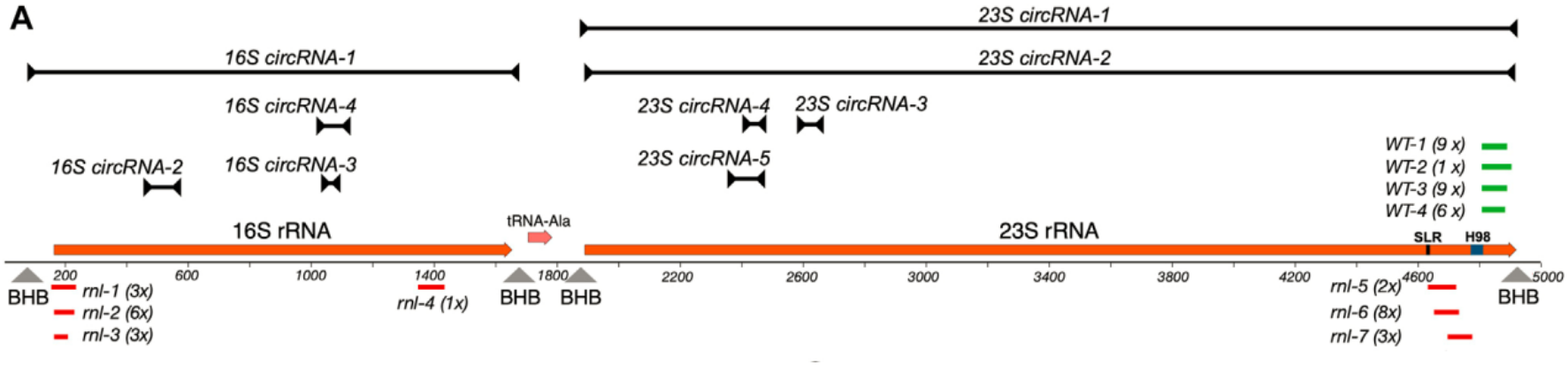


Figure 7 TkoRnl deletion affects rRNA circularization and processing. (A) Circular RNAs derived from T. kodakarensis 16S-tRNA-Ala-23S rRNA operon. A diagram on the bottom represents a 5 kb segment of T. kodakarensis (2022701–2027700 of NC_006624.1 reference genome) which contain the 16S-tRNA-Ala-23S rRNA operon. Each gene was colored in orange. Region of sarcin-ricin loop (SRL) and H98 were drawn with blue. The predicted BHB motifs position was pointed out by the gray arrowheads. The black lines above the diagram represented the expected size and position of highly represented circRNAs identified by my circular RNA analysis with Total RNA datasets. Takagi Yuko performed small RNA cloning with the small RNAs fragments collected from WT or Δrnl via adaptor-mediate RNA ligation. These cloned fragments (WT : green or Δrnl : red) were aligned to the reference genome. The number on the right represents number of identical RNA fragments recovered in the RNA cloning experiment (i.e., 2 x indicates twice). All cloned fragments from WT (*WT-1* to *WT-4*) were aligned after H98 while Δrnl fragments (*rnl-5* to *rnl-7*) were aligned just before H98.

Figure 7

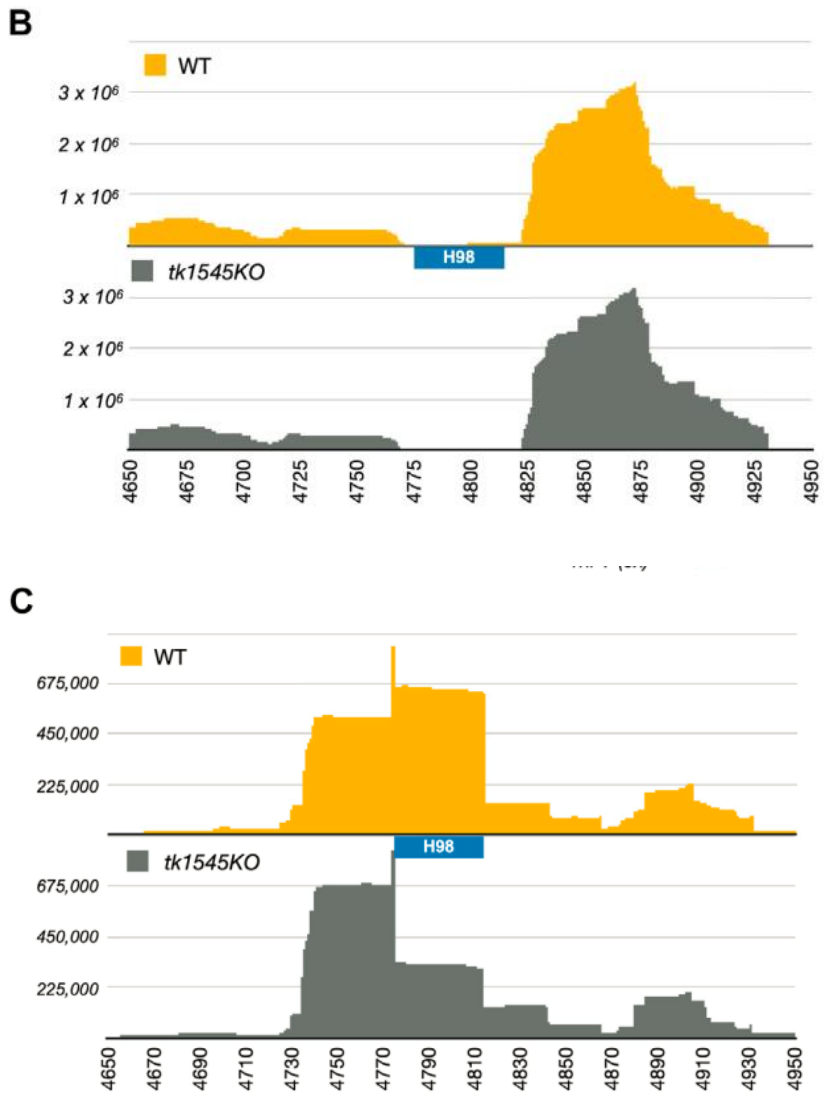


Figure 7 TkoRnl deletion affects rRNA circularization and processing. (B) H98 reads coverage changed after TkoRnl deletion. The Read-coverage pattern from Total RNA-Seq datasets were showed in the range 4650 to 4950 from the Figure 4(A) diagram which represent the 3'end of 23S rRNA (WT in top panel and *TK1545* KO in bottom panel). (C) was as same as (B) except that the datasets were from small RNA samples. The y-axis indicated the read counts. The read counts for the *TK1545* KO were normalized to the WT read counts. The position of H98 was indicated by the blue box.

Figure 8

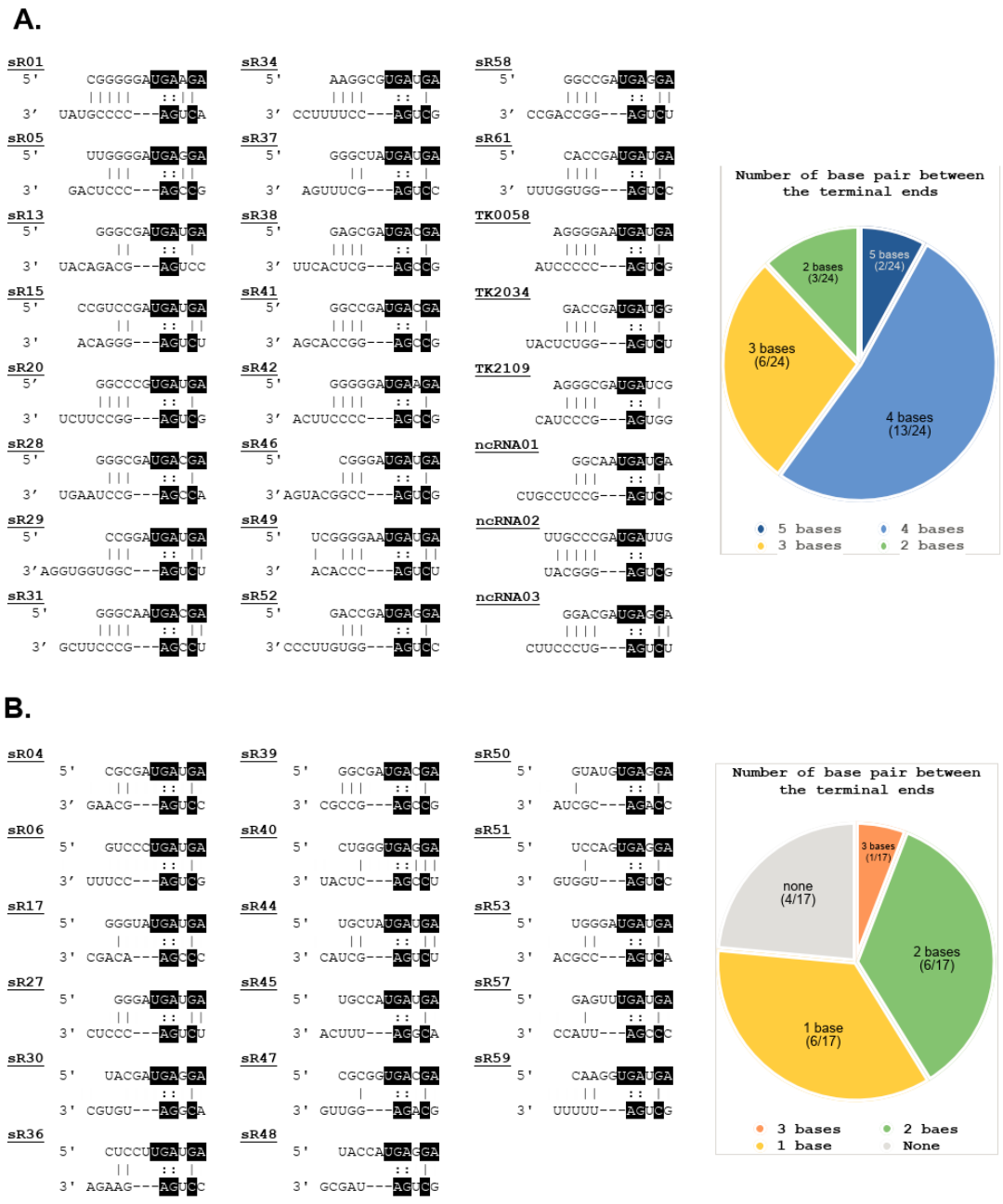


Figure 8 Majority of TkoRnl dependent circular RNAs contain terminal stem in Secondary structure. (A) TkoRnl dependent circular RNAs with more than twenty circRNA reads were performed secondary structure analysis. According to the C/D box sRNA structure, the C box and D box elements were highlighted with black background and the potential number of base pairings in the terminal stem region were calculated. The pie graph showed the RNA distribution. (B) the predicted C/D box sRNAs which have no circRNA reads and have more than 1000 aligned reads in Table 2, were analyzed.

Figure 9

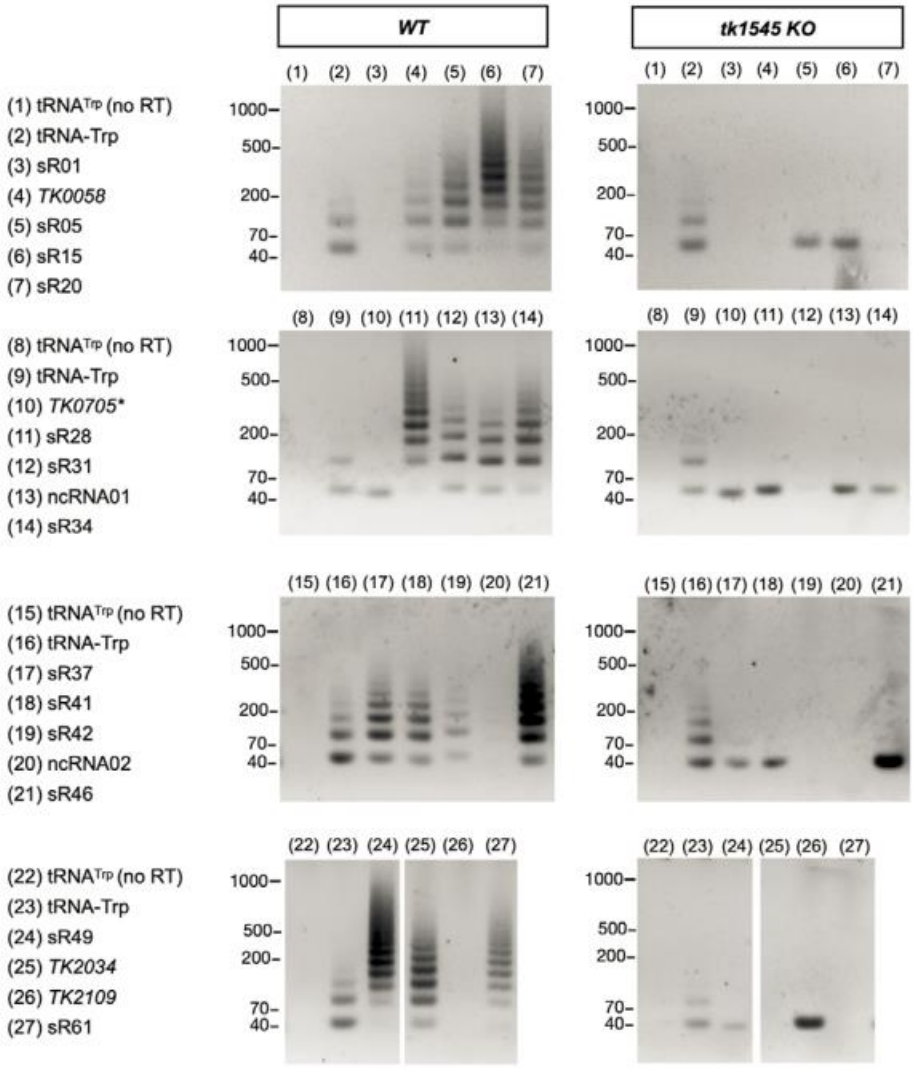
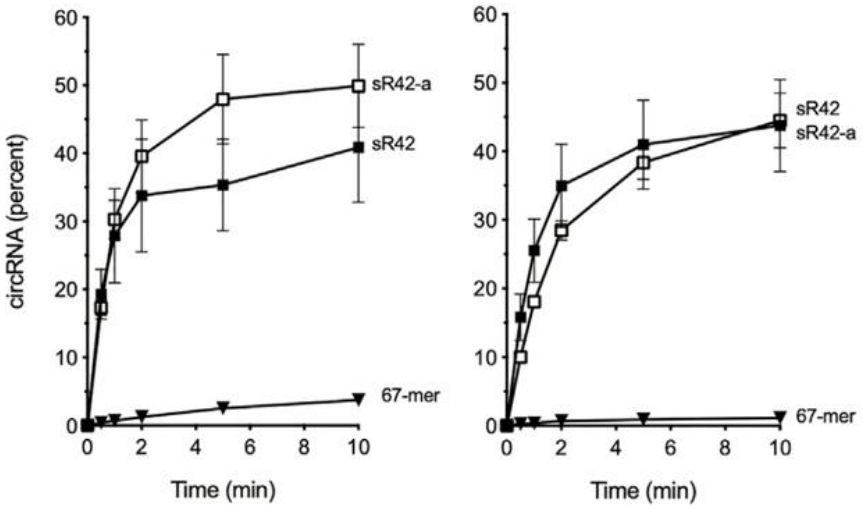


Figure 9 TkoRnl dependent circRNA were detected via RT-PCR. Total RNAs from T.kodakarensis WT or *⊿rnl* cells were used as templates for the Reverse Transcription PCR (RT-PCR). To identify the circular RNA predicted by NGS data, forward junction primer and reverse primer were designed for each predicted circRNAs (detailed in primer list). The circular RNA (tRNA-Trp intron) was used as a positive control (tRNA-Trp). As a negative control, the sample named $tRNA^{Trp}$ (no RT) was performed RT-PCR without reverse transcriptase. Another negative control, (10) TK0705 (Gene ID 3234407: NADP-dependent glyceraldehyde-3-phosphate dehydrogenase) is not circularized in both WT and *⊿rnl*. PCR products which migrate to near 40nts were likely to be products derived from a primer dimer.

Figure 10

A



B

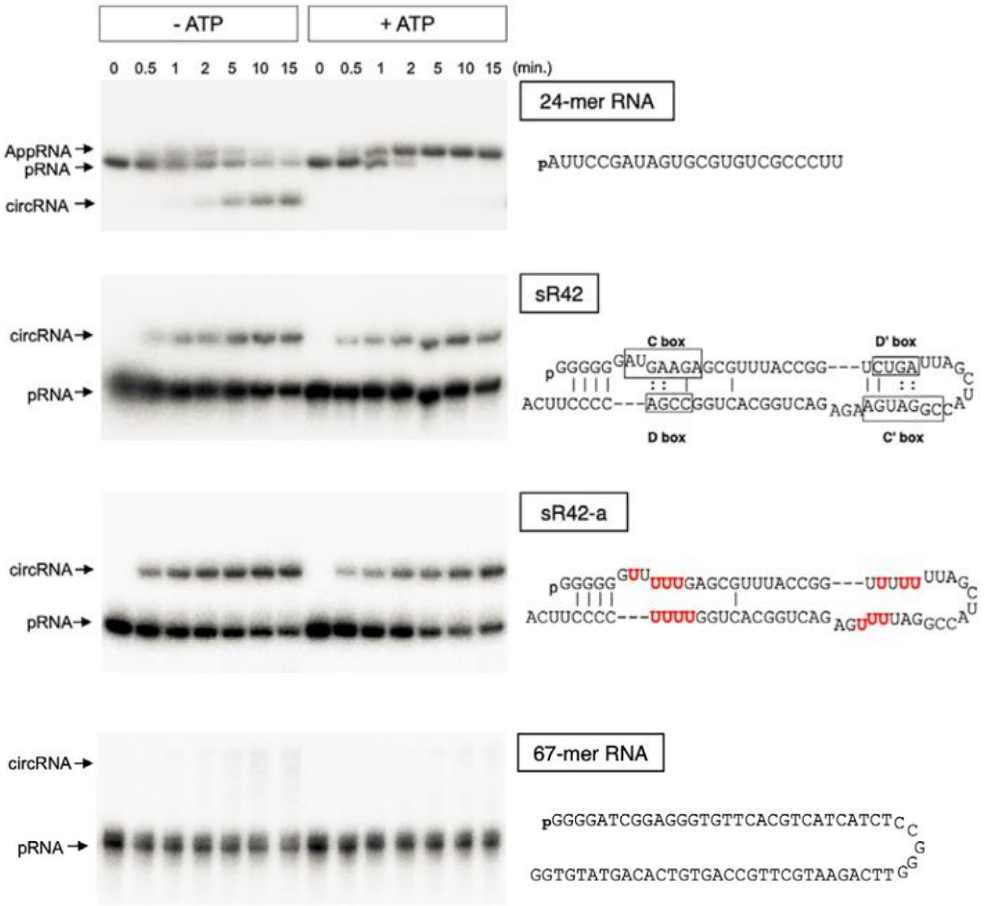
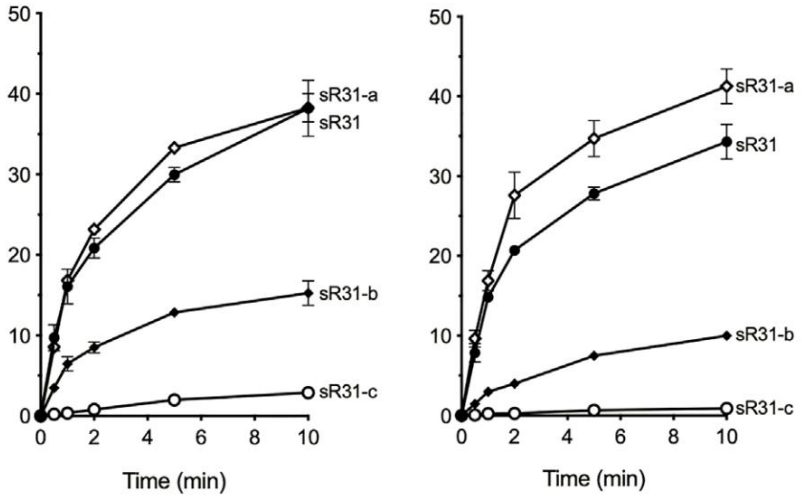


Figure 10. Characterization of RNA circularization activity on C/D box sRNA-42. (A) The time course of MthRnl ligation with C/D box sRNA-42 and indicated RNAs. MthRnl was incubated with the indicated pRNAs at 70°C in a

reaction mixture that contained 50 mM Tris-HCl (pH 6.5), 1 mM MgCl2, in the presence or absence of 1 mM ATP. Aliquots were withdrawn at the times indicated and the reaction products were separated by UREA-PAGE. The left panel represent the time course reaction without ATP and the right panel represent the time course reaction with ATP. The data plotted is the average of three separate ligation experiments with SE bars. (B) The photograph for the reaction products separated by UREA-PAGE. Positions of the substrate pRNAs and the ligation products, AppRNA, and circRNA are indicated by the arrow on the left. The sequences of substrates are listed on the right. The C/C’ box and D/D’ box elements in sR42 and the variant nucleotides in sR42 mutant (sR42-a) are highlighted.

Figure 11

A



B

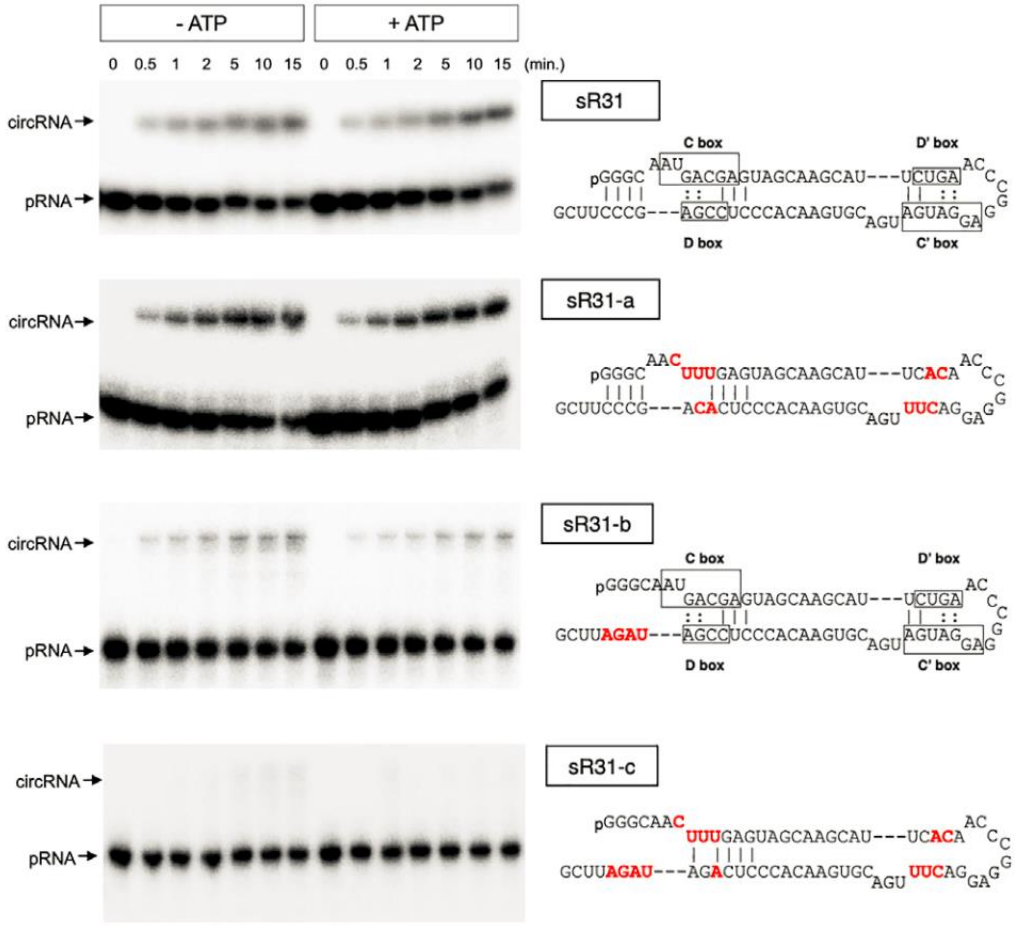


Figure 11. Characterization of RNA circularization activity on C/D box sRNA-31. (A) The time course of MthRnl ligation with C/D box sRNA-31 and indicated RNAs. MthRnl was incubated with the indicated pRNAs at 70°C in a

reaction mixture that contained 50 mM Tris-HCl (pH 6.5), 1 mM MgCl2, in the presence or absence of 1 mM ATP. Aliquots were withdrawn at the times indicated and the reaction products were separated by UREA-PAGE. The left panel represent the time course reaction without ATP and the right panel represent the time course reaction with ATP. The data plotted is the average of three separate ligation experiments with SE bars. (B) The photograph for the reaction products separated by UREA-PAGE. Positions of the substrate pRNAs and the ligation products, AppRNA, and circRNA are indicated by the arrow on the left. The sequences of substrates are listed on the right. The C/C' box and D/D' box elements in sR31 and the variant nucleotides in sR31 mutants are highlighted.

Figure 12

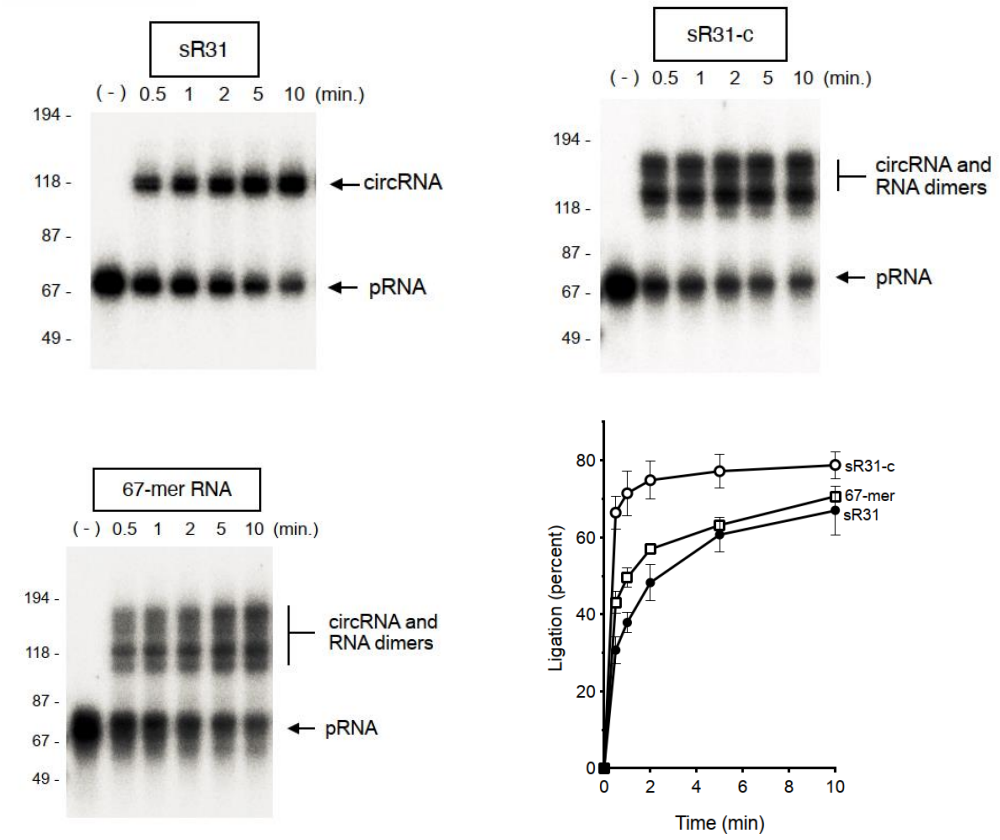


Figure 12. T4Rnl2 ligate C/D box sRNA and 67nt RNAs with similar ligation efficiency. The time course of T4Rnl2 ligation with C/D box sRNA-31 and indicated RNAs. T4Rnl2 was incubated with the indicated pRNAs at 22°C in a reaction mixture that contained 50 mM Tris-HCl (pH 6.5), 1 mM MgCl2, 5mM DTT. Aliquots were withdrawn at the times indicated and the reaction products were separated by UREA-PAGE. In control reaction (-), T4Rnl2 is not added. Positions of the substrate pRNAs and the ligation products, circRNA, and possible linear and circular RNA dimer are indicated by the arrow on the right. The data plotted is the average of three separate ligation experiments with SE bars.

Figure 13

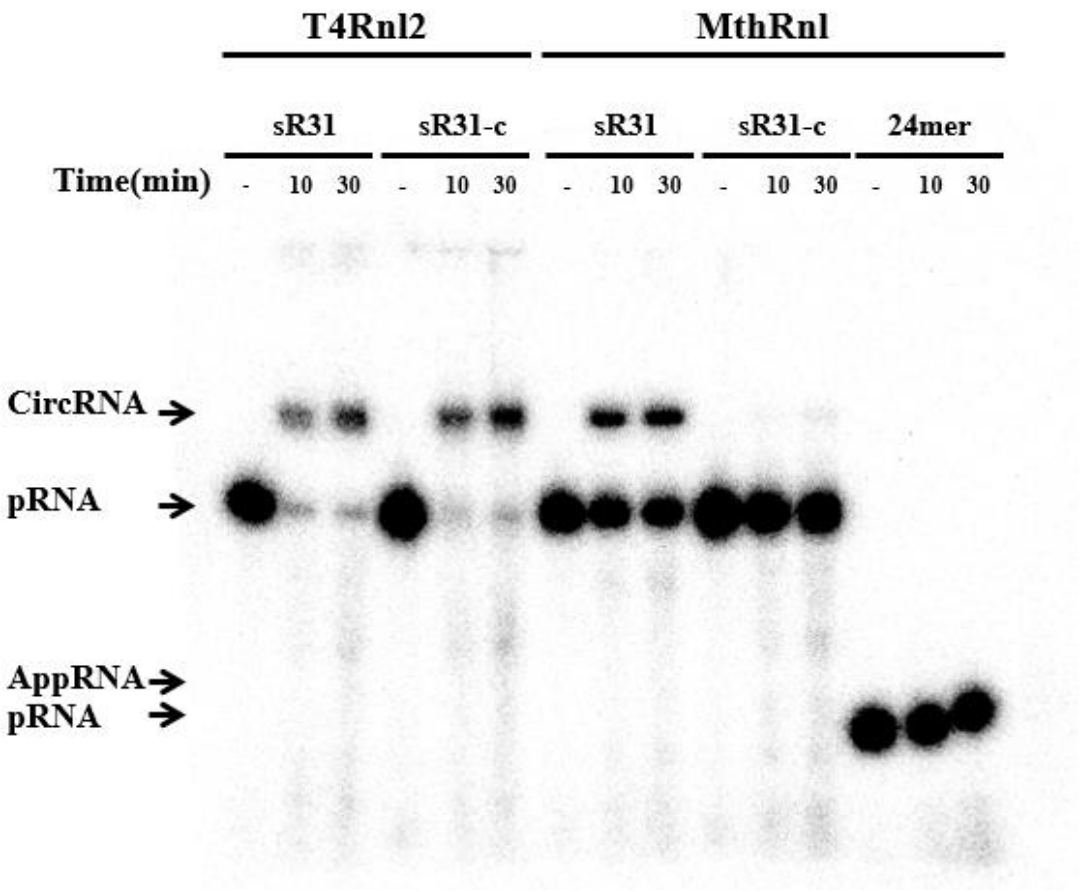


Figure 13. T4Rnl2 and MthRnl have different ligation preference. The time course reactions with T4Rnl2 and MthRnl were performed with each standard assay except that the incubation temperature of both reactions was set to 37°C. The RNA substrates are 24mer RNA, C/D box sRNA (sR31), and sR31-c with terminal stem mutant and C/D box element mutant. Positions of the substrate pRNAs and the ligation products, AppRNA, and circRNA are indicated by the arrow on the left.

Figure 14

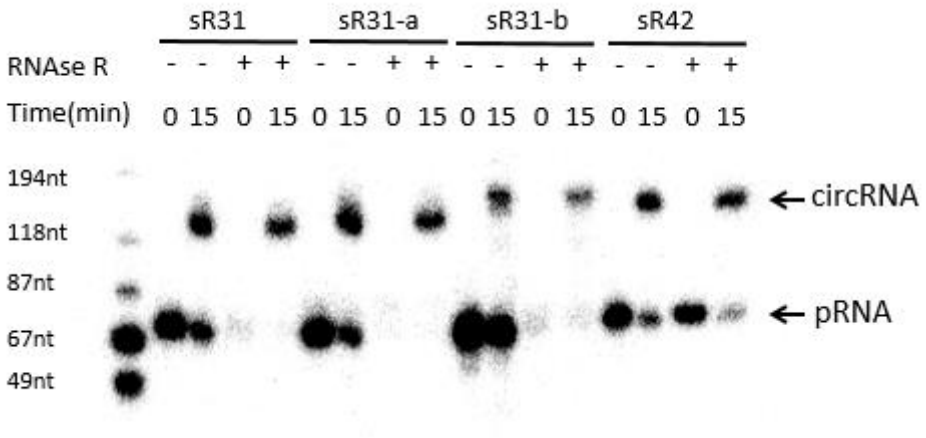
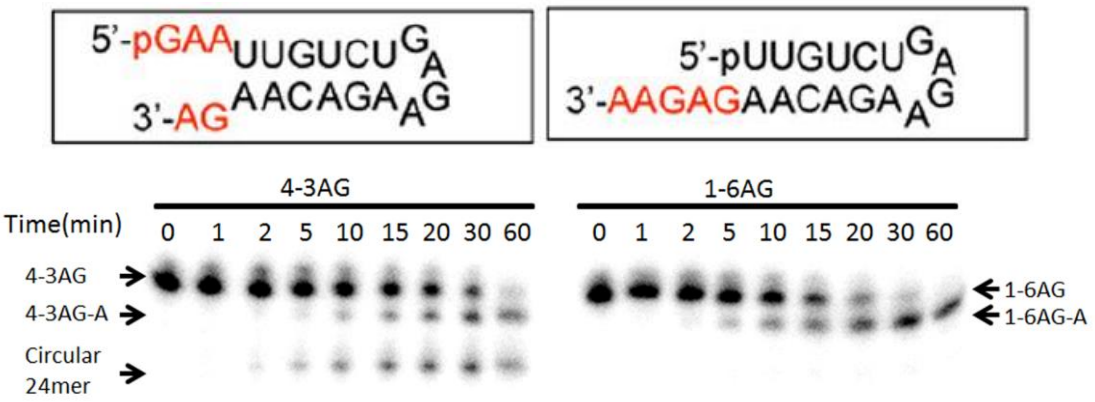


Figure 14 MthRnl ligation products checking via RNAse R. RNAse R digestion with MthRnl ligation products. 10μl Standard MthRnl ligation assay were performed with pRNAs as indicated. The reaction was quenched by 2 μl 50mM EDTA and half of each sample were digested with 7U RNAse R from Lucigen and 1X RNAse R buffer in 10μl volume with 60min incubation at 37°C. All aliquots were separated by 12% UREA-PAGE. Positions of the substrate pRNAs and the ligation products, circRNA are indicated by the arrow on the right.

Figure 15

A



B

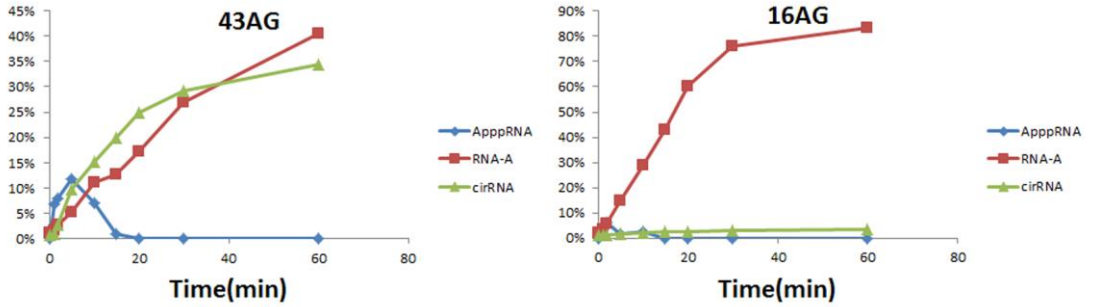
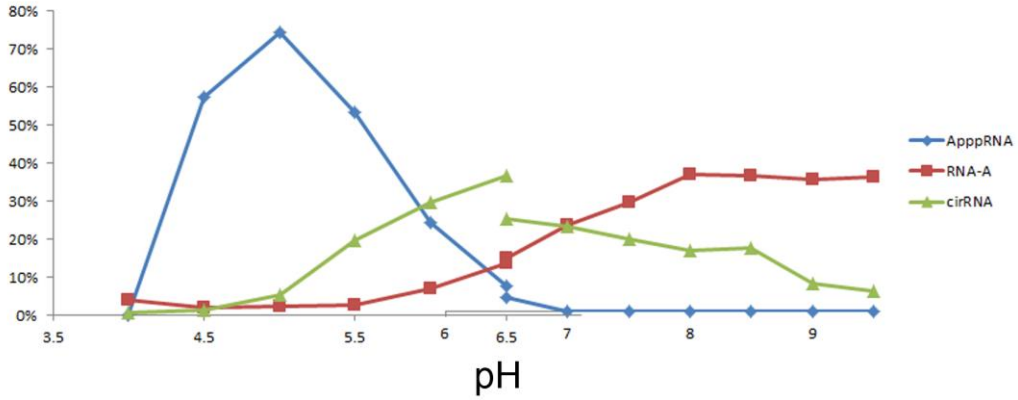


Figure 15 MthRnl could cleave the adenine on the 3' end of RNA substrates.

The time course of MthRnl 3'A deletion assay were performed with dual-over-hang RNA and 3'-over-hang RNAs. MthRnl was incubated with the indicated pRNAs at 55°C in a reaction mixture that contained 50 mM Tris-HCl (pH 6.5) and 0.5 mM MgCl2. Aliquots were withdrawn at the times indicated and the reaction products were separated by UREA-PAGE. (A) The photograph for the reaction products separated by UREA-PAGE. Positions of the substrate pRNAs(4-3AG and 1-6AG) and the ligation products, circRNA, and deadenylated RNA(4-3AG-A and 1-6AG-A) are indicated by the arrow. The sequence and secondary structure of 4-3AG and 1-6AG were shown on the top panel of figure 13 A. (B) The percentage of deadenylated RNA [RNA-A)], AppRNA, and cirRNA product is plotted as a function of incubation time.

Figure 16

A



B

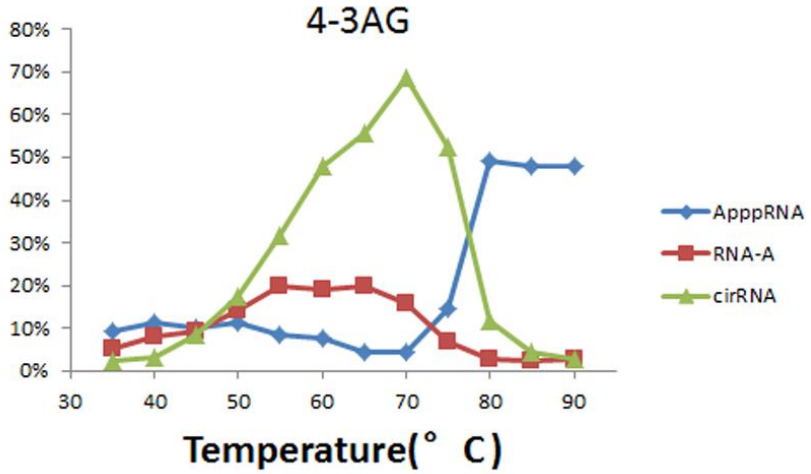


Figure 16 The optimal condition for MthRnl 3'-deadenylation is different with MthRnl ligation function. (A) Effects of pH on deadenlyation activity. The Standard MthRnl ligation assays were performed with the reaction mixtures containing 50 mM buffer (either Tris-Acetate between pH 4.0–6.5; or Tris-HCl between pH 6.5–9.5) and 1 pmol of dual overhang pRNA substrate (4-3AG) and 0.45 µg of MthRnl. The functions of pH on MthRnl ligation and 3'-deadenylation were plotted. The percentage of cirRNA (triangle), AppRNA (open diamond) and RNA-A (closed square) is plotted as a function of pH. (B) Temperature dependence. The Standard MthRnl ligation assays were performed with the reaction mixtures containing dual overhang pRNA substrate (4-3AG) and MthRnl were incubated for 15 min at indicated temperature. The percentage of cirRNA (triangle), AppRNA (open diamond) and RNA-A (closed square) is plotted as a function of temperature.

Figure 17

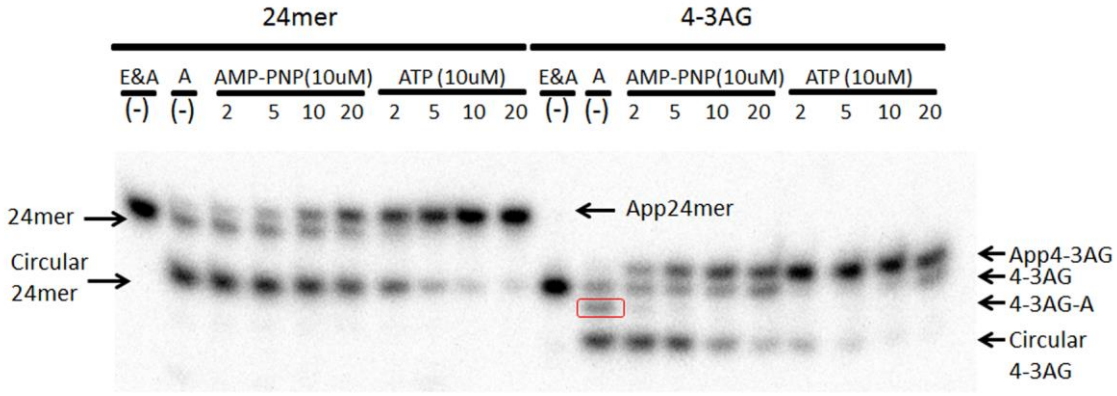
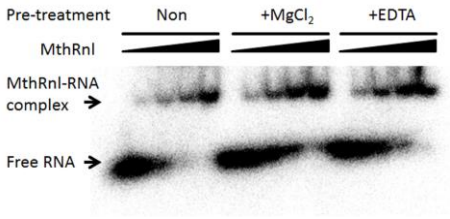


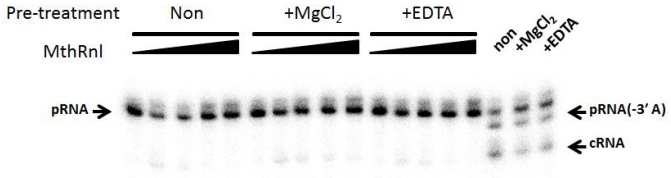
Figure 17 ATP binding site of MthRnl were essential for 3'-deadenylation. The Standard MthRnl ligation assays were performed with the reaction mixtures containing the indicated pRNA substrates and the indicated concentration of ATP or AMP-PNP. Positions of the substrate pRNAs (24mer and 4-3AG) and the ligation products, AppRNA (App24mer and App4-3AG), circRNA (circular 24mer and circular 4-3AG), and deadenylated RNA(4-3AG-A) are indicated by the arrow. The control assay (E&A (-)) have no MthRnl and ATP. The control assay (A (-)) have no ATP.

Figure 18

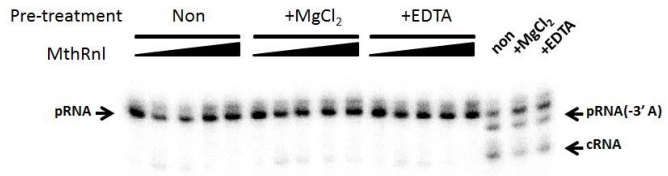
A



B



C



D

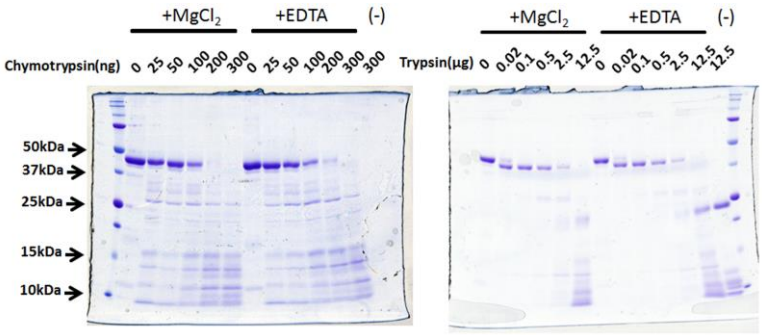
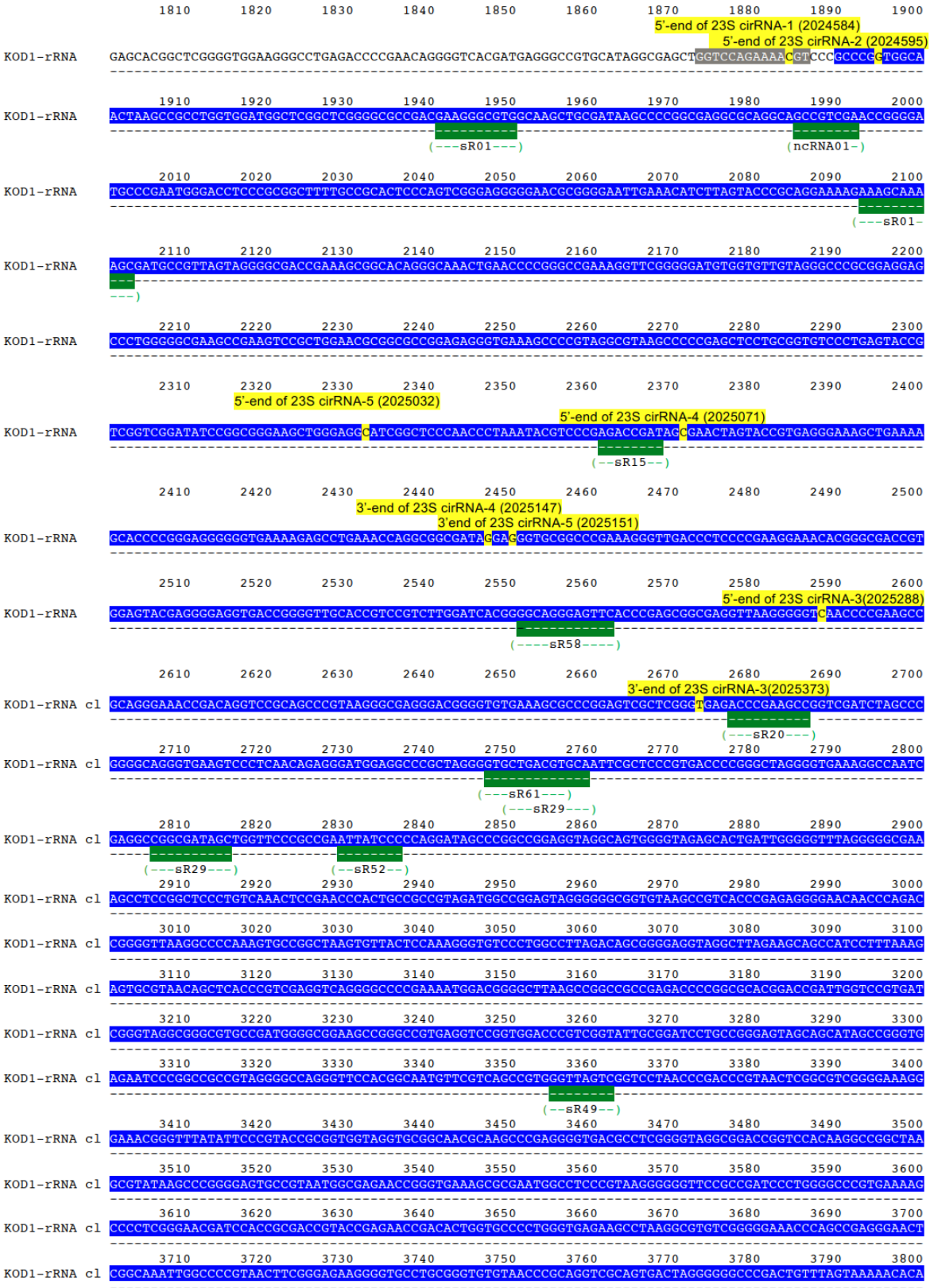


Figure 18 Divalent cation does not affect substrate binding or protein folding. (A) Electrophoretic mobility shift assay (EMSA) with MthRnl and the pRNA substrates. The MthRnl for assay performed pre-treatment to incubate with 10 mM MgCl2 (+MgCl2) or remove the metal ions by incubating with 2 mM EDTA (+EDTA). The MthRnl were incubated for 5 mins at room temperature. The free MgCl2 and EDTA were removed from the MthRnl by applying the incubated enzyme onto G-50 NICK Column (GE Healthcare). The protein titration were performed with the reaction mixture containing 50mM Tris-HCl (pH 6.5), 5% glycerol, 50 nM dual-overhang pRNA(4-3 AG). The MthRnl used in the titration assay were 0, 0.12, 0.23, 0.45, or 0.9 μg for each line. The control assay used

MthRnl without pre-treatment, named Non. MthRnl-RNA complex was visualized by EMSA assay. Positions of the protein-RNA complex and the free RNA were indicated by the arrow. (B) RNA Ligation. To verify the integrity MthRnl purified by gel filtration, RNA ligation assay was performed in the presence (lanes 16, 17, 18; 0.9 µg of enzyme) or absence of divalent cation (Lanes 2 -15; with the same enzyme titration series as panel A) using 50 nM dual-overhang pRNA. MthRnl (+MgCl2) or MthRnl (+EDTA) was active for RNA circularization and 3'-deadenylation only when divalent cation is present in the reaction, which implies that MgCl2 was successfully removed by gel filtration, and EDTA pre-treatment does not inactivate the enzyme. (C) Limited Proteolytic Digestion. MthRnl (2.5 µg) was incubated with indicated amount of either chymotrypsin (top panel) or trypsin (bottom panel) in a reaction buffer containing either 1mM MgCl2 or 2 mM EDTA. Incubation was for 15 mins at 25°C. Digested products were resolved by 15% SDS-PAGE and stained with Coomassie blue. Control reaction without MthRnl is shown in lane (-). Positions of marker proteins are indicated on the right.

Figure 19


5'-end of 16S cirRNA-1 (2022801)
(---sR61--)
(---sR15---)
5'-end of 16S cirRNA-2(2023166)
(----sR13----)
3'-end of 16S cirRNA-2(2023285)
(----sR46----)
(----sR05---)
5'-end of 16S cirRNA-4(2023722)
5'-end of 16S cirRNA-3(2023743)
3'-end of 16S cirRNA-3(2023804)
3'-end of 16S cirRNA-4(2023831)
(----sR58----)
3'-end of 16S cirRNA-1 (2024382)
(--sR58-)


5'-end of 23S cirRNA-1 (2024584)
5'-end of 23S cirRNA-2 (2024595)
KOD1-rRNA
(---sR01---)
(ncRNA01-)
(---sR01-
---)
5'-end of 23S cirRNA-5 (2025032)
5'-end of 23S cirRNA-4 (2025071)
(--sR15--)
3'-end of 23S cirRNA-4 (2025147)
3'end of 23S cirRNA-5 (2025151)
5'-end of 23S cirRNA-3(2025288)
(----sR58----)
3'-end of 23S cirRNA-3(2025373)
KOD1-rRNA cl
(---sR20---)
(---sR61---)
(---sR29---)
(---sR29---)
(--sR52--)
(--sR49--)

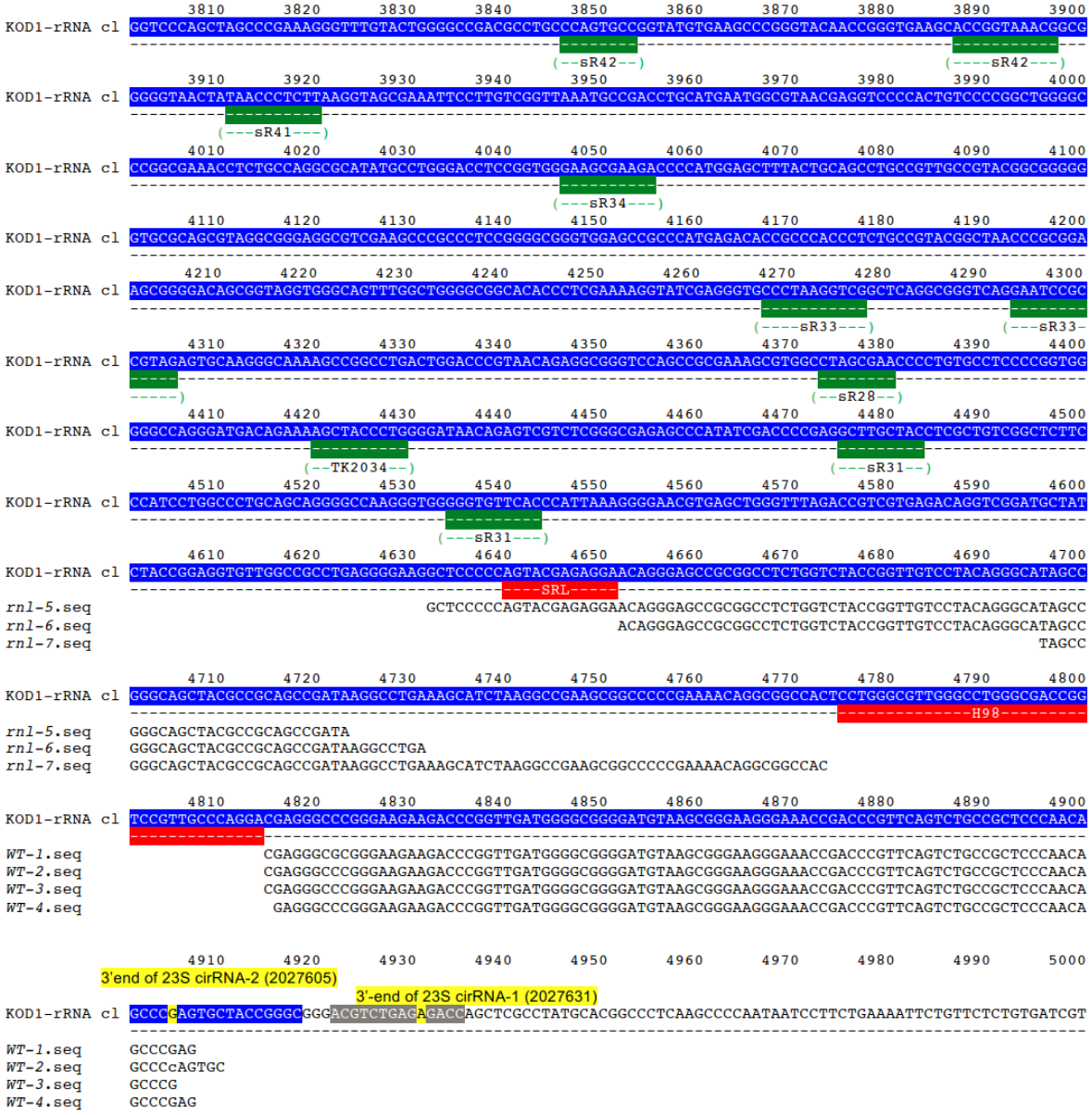


Figure 19 Analysis of *T. kodakarensis* rRNAs fragments obtained from WT and TkoRnl deletion cells. Small The full sequence of the diagram shown in Figure 4A were drawn here. The 5kb segment of *T. kodakarensis* (2022701–2027700 of NC_006624.1 reference genome) contains the 16S-tRNA-Ala-23S rRNA operon. The region that corresponds to 16S RNA (red), tRNA-Ala (light green), 23S rRNA (blue) are highlighted. The bulge-helix-bulge motifs are highlighted in gray. Nucleotides highlighted in yellow are the start and end position of the predicted circRNAs which shown in Figure 7. Predicted helix 98 (H98) and a conserved nucleotides in SRL are highlighted in red. Potential target sequences of the circularized C/D Box guide RNAs in Table 1 are indicated and colored in dark green. The full sequence of small RNA clonings (rnl-1 to rnl-7 and WT-1 to WT-4) were aligned to the 5kb segment.

Chapter 6 Tables and legends

Table 1 List of circRNAs in T. kodakarensis

Locus/ CircRNA name	Circular junction (+/– 5 nucleotides)	Predicted length[a]	*T. kodakarensis* (WT)		*T. kodakarensis* (*tk1545* KO)		Alias and predicted transcription start site[b]	
			Number of aligned read	Number of circular reads	Number of aligned read	Number of circular reads		
sR01	47786 to 47847	62	2,167	395	1,776	7	TKOcandSno19	47910
sR05	116401 to 116466	66	445,142	354	333,035	0	Tko-sR07	116468
sR13	279797 to 279863	67	59,144	47	58,048	0	Tko-sR14	279865
sR15	316130 to 316191	62	1,629,229	698	1,523,609	8	Tko-sR16	316202
sR20	558818 to 558879	62	1,622	21	1,400	0	Tko-sR20	558914
sR28	832364 to 832425	62	65,322	72	54,801	1	Tko-sR26	832424
sR29	940146 to 940209	64	475,600	36	283,215	0	Tko-sR29	940134
sR31	963853 to 963919	67	562,648	4,655	721,884	34	Tko-sR31	963842
sR34	1103565 to 1103626	62	20,161	119	19,347	1	Tko-sR35	1103625
sR37	1159583 to 1159644	62	356,961	44	195,173	1	Tko-sR37	1159732
sR38	1167276 to 1167338	63	11,287	106	4,940	0		
sR41	1226838 to 1226899	62	1,967	31	610	0	Tko-sR41	1226903
sR42	1226948 to 1227017	70	12,085	7,194	17,658	10	Tko-sR42	1226948
sR46	1371729 to 1371790	62	176,232	188	79,569	3	Tko-sR50	1371720
sR49	1446209 to 1446268	60	62,328	358	36,039	1	Tko-sR52	1446266
sR52	1476851 to 1476917	67	28,701	467	42,613	6	Tko-sR54	1476850
sR58	1947796 to 1947856	61	926	95	845	0		
sR61	2070055 to 2070116	62	100,212	92	31,284	0	Tko-sR67	2070114
ncRNA01	1053772 to 1053834	63	1,267,983	68	587,087	3	Tko-sR33	1053770
ncRNA02	1257020 to 1257081	62	8,320	139	6,342	0	TKOcandSno66	1257078
ncRNA03	1593064 to 1593134	71	7,565	403	9,633	1		
TK0058	51411 to 51477	67	1,816	589	780	1	Tko-sR01	51477
TK2109	1894519 to 1894579	61	14,240	28	11,130	0		
TK2034	1826865 to 1826930	66	771	66	477	2	Tko-sR22[c]	1826865
TK0894	779653 to 779796	144	317	28	215	28		
TK1980	1784662 to 1785579	918	1,322	42	739	53		
TK0135	108461 to 116468	8,008	457,060	90	576,950	110		
tRNA-Trp	1945728 to 1945789	62	28,907	2,023	23,156	1,292		
16S RNA-c1	2022801 to 2024382	1,582	404,505	96,319	268,720	53,186		
23S RNA-c1	2024584 to 2027631	3,048	867,077	27,733	610,944	11,628		
23S RNA-c2	2024595 to 2027605	3,011	867,051	83	610,939	35		

Table 1 List of circRNAs in T. kodakarensis. The 31 predicted circRNAs which were detected in WT-small or *TK1545* KO small RNA-Seq datasets were listed.
[a]Position of circular junction and predicted length of each circRNAs, were determined by mapping to the *T. kodakarensis* reference genome
[b]The alias and predicted start site of transcription are from Jäger et al.[120].
[c]Annotated as sR22 in Jäger et al[120].

Table 2 C/D box-like sRNAs encoded by T. kodakarensis

T. kodakarensis C/D box-like sRNA	Start and end (+/- 5 nucleotides)			Predicted length (nts)	*T. kodakarensis* (WT)		*T. kodakarensis* (*tk1545* KO)		Alias and predicted transcription start site+		*P. abyssi* C/D Box sRNA homologs	*P. abyssi*	
					Number of aligned read	Number of circular reads	Number of aligned read	Number of circular reads				Number of aligned read	Number of circular reads
sR01	47786	to	47847	62	2,167	395	1,776	7	TKOcandSno19	47910	Pab-sR09	0	9
sR02	54779	to	54838	60	253	0	140	0					
sR03	58300	to	58357	58	117,134	1	98,318	0	Tko-sR03	58356	Pab-sR29*	0	300
sR04	87729	to	87784	56	146,490	0	82,694	0	Tko-sR04	87730	Pab-sR46*	2,590	236
sR05	116401	to	116466	66	445,142	354	333,035	0	Tko-sR07	116468	Pab-sR18	191	0
sR06	159451	to	159515	65	51,789	0	28,668	0	Tko-sR11	159453	Pab-sR15	0	0
sR07	162929	to	162999	71	314	0	201	0					
sR08	178215	to	178280	66	171	0	124	0					
sR09	206252	to	206314	63	18	0	23	0					
sR10	254603	to	254660	58	13,578	2	7,752	0	Tko-sR12	254604	Pab-sR04*	0	10
sR11	268795	to	268861	67	185	0	131	0					
sR12	274087	to	274147	61	7,484	12	6,249	0	Tko-sR13	274205	Pab-sR10	0	0
sR13	279797	to	279863	67	59,144	47	58,048	0	Tko-sR14	279865			
sR14	287655	to	287727	73	16,443	5	4,240	0	Tko-sR15	287755			
sR15	316130	to	316191	62	1,629,229	698	1,523,609	8	Tko-sR16	316202	Pab-sR14*	0	95
sR16	368491	to	368563	73	777	0	459	0					
sR17	493594	to	493681	88	5,810	0	3,877	0					
sR18	534297	to	534365	69	588	0	317	0					
sR19	545080	to	545151	72	173	0	154	0					
sR20	558818	to	558879	62	1,622	21	1,400	0	Tko-sR20	558914	Pab-sR28	1,024	0
sR21	599335	to	599401	67	79	0	55	0					
sR22	605281	to	605338	58	201	0	172	0					
sR23	617316	to	617387	72	165	0	130	0					
sR24	621932	to	621995	64	588	0	389	0					
sR25	654868	to	654934	67	387	0	215	0					
sR26	669805	to	669874	70	42	0	53	0					
sR27	795506	to	795566	61	107,342	0	53,467	0	Tko-sR24	795566	Pab-sR02*	1,306	66
sR28	832364	to	832425	62	65,322	72	54,801	1	Tko-sR26	832424	Pab-sR39*	0	11
sR29	940146	to	940209	64	475,600	36	283,215	0	Tko-sR29	940134	Pab-sR07*	0	1
sR30	942666	to	942747	82	3,044	0	2,747	0					
sR31	963853	to	963919	67	562,648	4,655	721,884	34	Tko-sR31	963842	Pab-sR03*	0	28
sR32	1001641	to	1001710	70	292	0	264	0					
sR33	1100949	to	1101011	63	170,549	20	97,779	0	Tko-sR34	1101012	Pab-sR13*	3,727	120
sR34	1103565	to	1103626	62	20,161	119	19,347	0	Tko-sR35	1103625	Pab-sR45*	584	87
sR35	1108431	to	1108492	62	550,108	8	209,674	1	Tko-sR36	1108387	Pab-sR31*	483	287
sR36	1133458	to	1133521	64	2,487	0	1,500	0					
sR37	1159583	to	1159644	62	356,961	44	195,173	1	Tko-sR37	1159732	Pab-sR32*	543	35
sR38	1167276	to	1167338	63	11,287	106	4,940	0					
sR39	1167372	to	1167429	58	26,514	0	26,153	0	Tko-sR39	1167431	Pab-sR51	0	3
sR40	1183980	to	1184044	65	2,773	0	2,623	0					
sR41	1226838	to	1226899	62	1,967	31	610	0	Tko-sR41	1226903	Pab-sR34*	0	2,621
sR42	1226948	to	1227017	70	12,085	7,194	17,658	10	Tko-sR42	1226948	Pab-sR12*	1,840	122
sR43	1365876	to	1365945	70	148	0	92	0					
sR44	1368418	to	1368475	58	32,410	0	34,627	0	Tko-sR49	1368474	Pab-sR21*	664	49
sR45	1369006	to	1369070	65	1,385	0	1,076	0					
sR46	1371729	to	1371790	62	176,232	188	79,569	3	Tko-sR50	1371720	Pab-sR23	0	0
sR47	1400539	to	1400606	68	1,603	0	1,156	0					
sR48	1445711	to	1445786	76	1,874	0	2,488	0					
sR49	1446209	to	1446268	60	62,328	358	36,039	1	Tko-sR52	1446266	Pab-sR11*	555	46
sR50	1452262	to	1452324	63	2,020	0	1,736	0					
sR51	1465355	to	1465424	70	1,560	0	1,280	0					
sR52	1476851	to	1476917	67	28,701	467	42,613	6	Tko-sR54	1476850	Pab-sR55*	1,121	44
sR53	1626302	to	1626359	58	208,776	0	150,510	0	Tko-sR57	1626358	Pab-sR60*	18,423	6,365
sR54	1626388	to	1626447	60	157,804	4	43,537	0	Tko-sR58	1626445	Pab-sR26*	10,252	3,087
sR55	1726020	to	1726083	64	1,429,614	6	1,478,232	3	Tko-sR61	1726020	Pab-sR44*	595	41
sR56	1766626	to	1766698	73	615	0	407	0					
sR57	1787986	to	1788070	85	2,302	0	1,285	0	TKOcandSno104	1780963			
sR58	1947796	to	1947856	61	926	95	845	0			Pab-sR25*	2,153	447
sR59	2002453	to	2002515	63	198,717	0	113,994	0	Tko-sR66	2002525	Pab-sR05	95	1
sR60	2062992	to	2063053	62	712	0	419	0					
sR61	2070055	to	2070116	62	100,212	92	31,284	0	Tko-sR67	2070114	Pab-sR52	176	0

Table 2 C/D box-like sRNAs encoded by T. kodakarensis. A total of sixty-one putative C/D box sRNAs were identified in the *T. kodakarensis* genome (NC_006624.1) via searching C/D box motifs. Circular form of C/D box sRNAs were analyzed with WT and *tk1545* KO small RNA-Seq datasets. Jäger et al have reported several C/D box sRNAs in *T. kodakarensis[120]*. I listed their data in the The alias and predicted start site of transcription column for comparing. I performed circular RNA analysis with *P.abyssi* RNA-Seq dataset (GSM1401488)[113]. The *P. abyssi* C/D box sRNAs[96] which are homologs of *T. kodakarensis* C/D box sRNAs were listed with aligned reads and circRNA reads. The *P. abyssi* C/D box sRNAs which were identified with circular Reads under my analysis, were considered to have circular form and were indicated by asterisks.

Table 3 Analysis of T. kodakarensis circRNAs on an ABI/SOLiD sequencing platform

	Illumina dataset (from Table 1)				SOLiD small RNA-Seq dataset 1				SOLiD small RNA-Seq dataset 2			
	T. kodakarensis (WT)		*T. kodakarensis* (*tk1545* KO)		*T. kodakarensis* (WT)		*T. kodakarensis* (*tk1545* KO)		*T. kodakarensis* (WT)		*T. kodakarensis* (*tk1545* KO)	
Locus / CircRNA name	RPKM aligned read	RPKM circular read	RPKM aligned read	RPKM circular read	RPKM aligned read	RPKM circular read	RPKM aligned read	RPKM circular read	RPKM aligned read	RPKM circular read	RPKM aligned read	RPKM circular read
sR01	212.8	39.4	200.3	0.8	13,051.6	65.6	9,265.3	0.0	13,461.4	85.2	10,608.5	1.5
sR05	40,401.3	35.3	34,717.0	0.0	3,654.2	50.2	4,740.6	0.0	4,124.4	46.8	3,440.2	1.5
sR13	5,287.8	4.7	5,960.9	0.0	1,476.2	24.1	1,757.9	0.0	1,950.1	25.1	2,283.8	0.0
sR15	157,409.4	69.7	169,074.4	0.9	99,416.6	834.0	245,412.3	9.0	135,113.7	841.7	268,802.4	4.5
sR20	156.7	2.1	155.4	0.0	818.6	1,023.2	1,588.2	0.0	2,024.9	1,204.1	2,026.0	3.0
sR28	6,311.1	7.2	6,081.3	0.1	3,909.8	112.9	3,636.8	0.0	4,483.9	145.3	3,766.8	1.5
sR29	46,703.8	3.6	31,943.5	0.0	5,354.8	41.5	7,684.5	7.8	6,407.5	53.4	8,489.8	9.1
sR31	50,304.0	464.7	74,129.1	3.9	15,632.3	419.9	31,288.8	5.2	17,365.0	502.7	31,803.1	9.1
sR34	1,947.9	11.9	101,282.5	0.1	536.0	1.9	485.8	0.0	743.2	30.1	567.8	0.0
sR37	34,488.1	4.4	21,658.3	0.1	309.7	3.9	308.3	0.0	208.8	5.0	388.5	0.0
sR38	1,073.2	10.6	539.5	0.0	3,911.0	312.8	3,734.9	0.0	3,512.3	297.3	3,513.4	0.0
sR41	190.0	3.1	67.7	0.0	2,141.5	205.6	1,514.9	0.0	2,350.5	215.4	1,535.3	1.5
sR42	1,034.2	718.2	1,735.6	1.1	14,684.6	840.8	22,989.6	2.6	16,401.8	945.2	21,267.2	0.0
sR46	16,756.5	18.8	8,689.6	0.3	5,365.3	422.8	4,827.2	2.6	6,090.5	496.0	4,638.3	0.0
sR49	6,222.6	35.7	4,132.5	0.1	3,301.7	91.7	2,829.8	2.6	2,015.7	111.9	1,055.1	0.0
sR52	2,773.0	46.6	4,728.8	0.7	115,289.3	262.6	165,785.0	7.8	158,209.2	405.8	182,679.9	15.1
sR58	90.9	9.5	95.3	0.0	86.4	7.7	90.8	0.0	75.4	11.7	102.5	0.0
sR61	9,682.1	9.2	3,471.6	0.0	2,028.7	790.6	866.4	1.3	1,698.0	1,095.6	1,019.8	3.0
ncRNA01	126,590.9	6.8	67,320.5	0.3	104,085.9	56.0	98,215.4	40.1	137,796.2	66.8	108,032.8	30.3
ncRNA02	803.8	13.9	703.8	0.0	28,622.4	0.0	43,833.6	0.0	35,074.6	8.4	38,178.2	0.0
ncRNA03	638.2	40.2	933.5	0.1	1,056.0	61.8	1,897.5	2.6	1,199.6	38.4	2,470.5	0.0
TK0058	162.4	58.8	80.1	0.1	3,495.3	20,502.8	1,853.4	51.7	4,087.0	18,856.5	1,934.7	45.4
TK2034	70.0	6.6	49.7	0.2	2,965.7	16.4	9,579.0	0.0	3,023.6	20.0	10,668.1	0.0
TK2109	1,398.4	2.8	1,255.8	0.0	2,110.4	330.1	3,048.1	1.3	1,163.1	643.0	2,619.8	0.0
tRNA-Trp	2,792.9	202.0	2,569.6	148.2	10,630.5	368.7	18,133.9	697.9	15,068.7	679.7	18,339.5	713.0

Table 3 Analysis of T. kodakarensis circRNAs on an ABI/SOLiD sequencing platform. Small RNA libraries were prepared from the WT and tk1545 KO cells using NEB Next Multiplex Small RNA Library Prep Set for SOLiD (New England Biolabs). The Next generation RNA-sequencings were performed on ABI/SOLiD platform twice (SOLiD small RNA-Seq dataset 1 and SOLiD small RNA-Seq dataset 2). The circular RNA analysis was performed with these datasets (Supplementary Data 4). The circRNAs in Table 1 were listed and were confirmed with SOLiD small RNA-Seq dataset 1 and SOLiD small RNA-Seq dataset 2.

References

1. Burroughs, A.M. and L. Aravind, *RNA damage in biological conflicts and the diversity of responding RNA repair systems.* Nucleic Acids Res, 2016. **44**(18): p. 8525-8555.
2. Ho, C.K., et al., *Structure and mechanism of RNA ligase.* Structure, 2004. **12**(2): p. 327-39.
3. Nandakumar, J., S. Shuman, and C.D. Lima, *RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward.* Cell, 2006. **127**(1): p. 71-84.
4. Torchia, C., Y. Takagi, and C.K. Ho, *Archaeal RNA ligase is a homodimeric protein that catalyzes intramolecular ligation of single-stranded RNA and DNA.* Nucleic Acids Res, 2008. **36**(19): p. 6218-27.
5. Chakravarty, A.K. and S. Shuman, *The sequential 2',3'-cyclic phosphodiesterase and 3'-phosphate/5'-OH ligation steps of the RtcB RNA splicing pathway are GTP-dependent.* Nucleic Acids Res, 2012. **40**(17): p. 8558-67.
6. Chakravarty, A.K., et al., *RNA ligase RtcB splices 3'-phosphate and 5'-OH ends via covalent RtcB-(histidinyl)-GMP and polynucleotide-(3')pp(5')G intermediates.* Proc Natl Acad Sci U S A, 2012. **109**(16): p. 6072-7.
7. Martins, A. and S. Shuman, *Characterization of a baculovirus enzyme with RNA ligase, polynucleotide 5'-kinase, and polynucleotide 3'-phosphatase activities.* J Biol Chem, 2004. **279**(18): p. 18220-31.
8. Blondal, T., et al., *Isolation and characterization of a thermostable RNA ligase 1 from a Thermus scotoductus bacteriophage TS2126 with good single-stranded DNA ligation properties.* Nucleic Acids Res, 2005. **33**(1): p. 135-42.
9. Englert, M. and H. Beier, *Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins.* Nucleic Acids Res, 2005. **33**(1): p. 388-99.
10. Wang, L.K., et al., *Structure-function analysis of the kinase-CPD domain of yeast tRNA ligase (Trl1) and requirements for complementation of tRNA splicing by a plant Trl1 homolog.* Nucleic Acids Res, 2006. **34**(2): p. 517-27.
11. Zhuang, F., et al., *Structural bias in T4 RNA ligase-mediated 3'-adapter ligation.* Nucleic Acids Res, 2012. **40**(7): p. e54.
12. McManus, M.T., et al., *Identification of candidate mitochondrial RNA editing ligases from Trypanosoma brucei.* RNA, 2001. **7**(2): p. 167-75.
13. Yin, S., et al., *Characterization of bacteriophage KVP40 and T4 RNA ligase 2.* Virology, 2004. **319**(1): p. 141-51.
14. Wang, P., et al., *Molecular basis of bacterial protein Hen1 activating the ligase activity of bacterial protein Pnkp for RNA repair.* Proc Natl Acad Sci U S A, 2012. **109**(33): p. 13248-53.
15. Martins, A. and S. Shuman, *An RNA ligase from Deinococcus radiodurans.* J Biol Chem, 2004. **279**(49): p. 50654-61.
16. Unciuleac, M.C. and S. Shuman, *Characterization of a novel eukaryal nick-sealing RNA ligase from Naegleria gruberi.* RNA, 2015. **21**(5): p. 824-32.

17. Banerjee, A., et al., *Structure and two-metal mechanism of fungal tRNA ligase.* Nucleic Acids Res, 2019. **47**(3): p. 1428-1439.
18. Silber, R., V.G. Malathi, and J. Hurwitz, *Purification and properties of bacteriophage T4-induced RNA ligase.* Proc Natl Acad Sci U S A, 1972. **69**(10): p. 3009-13.
19. Amitsur, M., I. Morad, and G. Kaufmann, *In vitro reconstitution of anticodon nuclease from components encoded by phage T4 and Escherichia coli CTr5X.* EMBO J, 1989. **8**(8): p. 2411-5.
20. Amitsur, M., R. Levitz, and G. Kaufmann, *Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA.* EMBO J, 1987. **6**(8): p. 2499-503.
21. Smith, P., et al., *The adenylyltransferase domain of bacterial Pnkp defines a unique RNA ligase family.* Proc Natl Acad Sci U S A, 2012. **109**(7): p. 2296-301.
22. Remus, B.S., Y. Goldgur, and S. Shuman, *Structural basis for the GTP specificity of the RNA kinase domain of fungal tRNA ligase.* Nucleic Acids Res, 2017. **45**(22): p. 12945-12953.
23. Peebles, C.L., et al., *Splicing of yeast tRNA precursors: a two-stage reaction.* Cell, 1979. **18**(1): p. 27-35.
24. Greer, C.L., et al., *Mechanism of action of a yeast RNA ligase in tRNA splicing.* Cell, 1983. **32**(2): p. 537-46.
25. Phizicky, E.M., R.C. Schwartz, and J. Abelson, *Saccharomyces cerevisiae tRNA ligase. Purification of the protein and isolation of the structural gene.* J Biol Chem, 1986. **261**(6): p. 2978-86.
26. Apostol, B.L., et al., *Deletion analysis of a multifunctional yeast tRNA ligase polypeptide. Identification of essential and dispensable functional domains.* J Biol Chem, 1991. **266**(12): p. 7445-55.
27. Westaway, S.K., et al., *Novel activity of a yeast ligase deletion polypeptide. Evidence for GTP-dependent tRNA splicing.* J Biol Chem, 1993. **268**(4): p. 2435-43.
28. Culver, G.M., et al., *A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae.* J Biol Chem, 1997. **272**(20): p. 13203-10.
29. Abelson, J., C.R. Trotta, and H. Li, *tRNA splicing.* J Biol Chem, 1998. **273**(21): p. 12685-8.
30. Sidrauski, C., J.S. Cox, and P. Walter, *tRNA ligase is required for regulated mRNA splicing in the unfolded protein response.* Cell, 1996. **87**(3): p. 405-13.
31. Englert, M., et al., *Archaeal 3'-phosphate RNA splicing ligase characterization identifies the missing component in tRNA maturation.* Proc Natl Acad Sci U S A, 2011. **108**(4): p. 1290-5.
32. Popow, J., et al., *HSPC117 is the essential subunit of a human tRNA splicing ligase complex.* Science, 2011. **331**(6018): p. 760-4.
33. Tanaka, N. and S. Shuman, *RtcB is the RNA ligase component of an Escherichia coli RNA repair operon.* J Biol Chem, 2011. **286**(10): p. 7727-31.
34. Tanaka, N., B. Meineke, and S. Shuman, *RtcB, a novel RNA ligase, can catalyze tRNA splicing and HAC1 mRNA splicing in vivo.* J Biol Chem, 2011. **286**(35): p. 30253-30257.
35. Jurkin, J., et al., *The mammalian tRNA ligase complex mediates splicing of XBP1 mRNA and controls antibody secretion in plasma cells.* EMBO J, 2014. **33**(24): p. 2922-36.
36. Kosmaczewski, S.G., et al., *The RtcB RNA ligase is an essential component of the metazoan unfolded protein response.* EMBO Rep, 2014. **15**(12): p. 1278-85.
37. Lu, Y., F.X. Liang, and X. Wang, *A synthetic biology approach identifies the mammalian UPR RNA ligase RtcB.* Mol Cell, 2014. **55**(5): p. 758-70.

38. Schmidt, C.A., et al., *Molecular determinants of metazoan tricRNA biogenesis.* Nucleic Acids Res, 2019. **47**(12): p. 6452-6465.
39. Litke, J.L. and S.R. Jaffrey, *Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts.* Nat Biotechnol, 2019. **37**(6): p. 667-675.
40. Moldovan, J.B., et al., *RNA ligation precedes the retrotransposition of U6/LINE-1 chimeric RNA.* Proc Natl Acad Sci U S A, 2019. **116**(41): p. 20612-20622.
41. Arts, G.J. and R. Benne, *Mechanism and evolution of RNA editing in kinetoplastida.* Biochim Biophys Acta, 1996. **1307**(1): p. 39-54.
42. Alfonzo, J.D., O. Thiemann, and L. Simpson, *The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria.* Nucleic Acids Res, 1997. **25**(19): p. 3751-9.
43. Stuart, K., et al., *RNA editing in kinetoplastid protozoa.* Microbiol Mol Biol Rev, 1997. **61**(1): p. 105-20.
44. Gott, J.M. and R.B. Emeson, *Functions and mechanisms of RNA editing.* Annu Rev Genet, 2000. **34**: p. 499-531.
45. Blum, B., N. Bakalara, and L. Simpson, *A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information.* Cell, 1990. **60**(2): p. 189-98.
46. Schnaufer, A., et al., *Separate insertion and deletion subcomplexes of the Trypanosoma brucei RNA editing complex.* Mol Cell, 2003. **12**(2): p. 307-19.
47. Schnaufer, A., et al., *An RNA ligase essential for RNA editing and survival of the bloodstream form of Trypanosoma brucei.* Science, 2001. **291**(5511): p. 2159-62.
48. Gao, G. and L. Simpson, *Is the Trypanosoma brucei REL1 RNA ligase specific for U-deletion RNA editing, and is the REL2 RNA ligase specific for U-insertion editing?* J Biol Chem, 2003. **278**(30): p. 27570-4.
49. McDermott, S.M., et al., *The Architecture of Trypanosoma brucei editosomes.* Proc Natl Acad Sci U S A, 2016. **113**(42): p. E6476-E6485.
50. Drozdz, M., et al., *TbMP81 is required for RNA editing in Trypanosoma brucei.* EMBO J, 2002. **21**(7): p. 1791-9.
51. Ho, C.K. and S. Shuman, *Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains.* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12709-14.
52. Nandakumar, J., et al., *RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2.* J Biol Chem, 2004. **279**(30): p. 31337-47.
53. Schmier, B.J., et al., *Deletion of the rnl gene encoding a nick-sealing RNA ligase sensitizes Deinococcus radiodurans to ionizing radiation.* Nucleic Acids Res, 2017. **45**(7): p. 3812-3821.
54. Brooks, M.A., et al., *The structure of an archaeal homodimeric ligase which has RNA circularization activity.* Protein Sci, 2008. **17**(8): p. 1336-45.
55. Shuman, S. and C.D. Lima, *The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases.* Curr Opin Struct Biol, 2004. **14**(6): p. 757-64.
56. El Omari, K., et al., *Molecular architecture and ligand recognition determinants for T4 RNA ligase.* J Biol Chem, 2006. **281**(3): p. 1573-9.
57. Wang, L.K., et al., *The C-terminal domain of T4 RNA ligase 1 confers specificity for tRNA repair.* RNA, 2007. **13**(8): p. 1235-44.
58. Gu, H., et al., *Structural and mutational analysis of archaeal ATP-dependent RNA ligase identifies amino acids required for RNA binding and catalysis.* Nucleic Acids Res, 2016. **44**(5):

p. 2337-47.
59. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4576-9.
60. Brochier-Armanet, C., P. Forterre, and S. Gribaldo, *Phylogeny and evolution of the Archaea: one hundred genomes later.* Curr Opin Microbiol, 2011. **14**(3): p. 274-81.
61. Reeve, J.N., *Archaeal chromatin and transcription.* Mol Microbiol, 2003. **48**(3): p. 587-98.
62. Clouet-d'Orval, B., et al., *Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea.* FEMS Microbiol Rev, 2018. **42**(5): p. 579-613.
63. Babski, J., et al., *Genome-wide identification of transcriptional start sites in the haloarchaeon Haloferax volcanii based on differential RNA-Seq (dRNA-Seq).* BMC Genomics, 2016. **17**(1): p. 629.
64. Smith, D.R., et al., *Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics.* J Bacteriol, 1997. **179**(22): p. 7135-55.
65. Randau, L., et al., *The heteromeric Nanoarchaeum equitans splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves.* Proc Natl Acad Sci U S A, 2005. **102**(50): p. 17934-9.
66. Yoshihisa, T., *Handling tRNA introns, archaeal way and eukaryotic way.* Front Genet, 2014. **5**: p. 213.
67. Marck, C. and H. Grosjean, *Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications.* RNA, 2003. **9**(12): p. 1516-31.
68. Diener, J.L. and P.B. Moore, *Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge-helix-bulge motif.* Mol Cell, 1998. **1**(6): p. 883-94.
69. Thompson, L.D. and C.J. Daniels, *Recognition of exon-intron boundaries by the Halobacterium volcanii tRNA intron endonuclease.* J Biol Chem, 1990. **265**(30): p. 18104-11.
70. Lafontaine, D.L. and D. Tollervey, *The function and synthesis of ribosomes.* Nat Rev Mol Cell Biol, 2001. **2**(7): p. 514-20.
71. Deppenmeier, U., et al., *The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea.* J Mol Microbiol Biotechnol, 2002. **4**(4): p. 453-61.
72. Hartman, A.L., et al., *The complete genome sequence of Haloferax volcanii DS2, a model archaeon.* PLoS One, 2010. **5**(3): p. e9605.
73. Chant, J. and P. Dennis, *Archaebacteria: transcription and processing of ribosomal RNA sequences in Halobacterium cutirubrum.* EMBO J, 1986. **5**(5): p. 1091-7.
74. Danan, M., et al., *Transcriptome-wide discovery of circular RNAs in Archaea.* Nucleic Acids Res, 2012. **40**(7): p. 3131-42.
75. Tang, T.H., et al., *RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing.* Nucleic Acids Res, 2002. **30**(4): p. 921-30.
76. Deutscher, M.P., *Maturation and degradation of ribosomal RNA in bacteria.* Prog Mol Biol Transl Sci, 2009. **85**: p. 369-91.
77. Henras, A.K., et al., *An overview of pre-ribosomal RNA processing in eukaryotes.* Wiley Interdiscip Rev RNA, 2015. **6**(2): p. 225-42.
78. Birkedal, U., et al., *The 23S Ribosomal RNA From.* Front Microbiol, 2020. **11**: p. 582022.
79. Balakin, A.G., L. Smith, and M.J. Fournier, *The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions.* Cell, 1996. **86**(5): p.

823-34.

80. Ganot, P., M. Caizergues-Ferrer, and T. Kiss, *The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation.* Genes Dev, 1997. **11**(7): p. 941-56.
81. Xia, L., N.J. Watkins, and E.S. Maxwell, *Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing.* RNA, 1997. **3**(1): p. 17-26.
82. Cavaillé, J., et al., *Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14311-6.
83. Nolivos, S., A.J. Carpousis, and B. Clouet-d'Orval, *The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn.* Nucleic Acids Res, 2005. **33**(20): p. 6507-14.
84. Omer, A.D., et al., *In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5289-94.
85. Baserga, S.J., X.D. Yang, and J.A. Steitz, *An intact Box C sequence in the U3 snRNA is required for binding of fibrillarin, the protein common to the major family of nucleolar snRNPs.* EMBO J, 1991. **10**(9): p. 2645-51.
86. Tyc, K. and J.A. Steitz, *U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus.* EMBO J, 1989. **8**(10): p. 3113-9.
87. Wang, H., et al., *Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution.* EMBO J, 2000. **19**(3): p. 317-23.
88. Breuer, R., J.V. Gomes-Filho, and L. Randau, *Conservation of Archaeal C/D Box sRNA-Guided RNA Modifications.* Front Microbiol, 2021. **12**: p. 654029.
89. Lui, L.M., et al., *Methylation guide RNA evolution in archaea: structure, function and genomic organization of 110 C/D box sRNA families across six Pyrobaculum species.* Nucleic Acids Res, 2018. **46**(11): p. 5678-5691.
90. Siprashvili, Z., et al., *The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer.* Nat Genet, 2016. **48**(1): p. 53-8.
91. Dong, X.Y., et al., *Implication of snoRNA U50 in human breast cancer.* J Genet Genomics, 2009. **36**(8): p. 447-54.
92. Cassidy, S.B., et al., *Prader-Willi syndrome.* Genet Med, 2012. **14**(1): p. 10-26.
93. Taoka, M., et al., *Landscape of the complete RNA chemical modifications in the human 80S ribosome.* Nucleic Acids Res, 2018. **46**(18): p. 9289-9298.
94. Tollervey, D., et al., *Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly.* Cell, 1993. **72**(3): p. 443-57.
95. Parker, R., et al., *Genetic analysis of small nuclear RNAs in Saccharomyces cerevisiae: viable sextuple mutant.* Mol Cell Biol, 1988. **8**(8): p. 3150-9.
96. Becker, H.F., et al., *High-throughput sequencing reveals circular substrates for an archaeal RNA ligase.* RNA Biol, 2017. **14**(8): p. 1075-1085.
97. Starostina, N.G., et al., *Circular box C/D RNAs in Pyrococcus furiosus.* Proc Natl Acad Sci U S A, 2004. **101**(39): p. 14097-101.
98. Falaleeva, M., et al., *C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks.* Bioessays, 2017. **39**(6).

99. Enright, C.A., et al., *5'ETS rRNA processing facilitated by four small RNAs: U14, E3, U17, and U3.* RNA, 1996. **2**(11): p. 1094-9.
100. Kiss, T., *Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions.* Cell, 2002. **109**(2): p. 145-8.
101. Tycowski, K.T., et al., *Modification of U6 spliceosomal RNA is guided by other small RNAs.* Mol Cell, 1998. **2**(5): p. 629-38.
102. Thompson, L.D., et al., *Transfer RNA intron processing in the halophilic archaebacteria.* Can J Microbiol, 1989. **35**(1): p. 36-42.
103. Kleman-Leyer, K., D.W. Armbruster, and C.J. Daniels, *Properties of H. volcanii tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems.* Cell, 1997. **89**(6): p. 839-47.
104. Lykke-Andersen, J., et al., *Archaeal introns: splicing, intercellular mobility and evolution.* Trends Biochem Sci, 1997. **22**(9): p. 326-31.
105. Klenk, H.P., et al., *The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.* Nature, 1997. **390**(6658): p. 364-70.
106. Gomes, I. and R. Gupta, *RNA splicing ligase activity in the archaeon Haloferax volcanii.* Biochem Biophys Res Commun, 1997. **237**(3): p. 588-94.
107. Zofallova, L., Y. Guo, and R. Gupta, *Junction phosphate is derived from the precursor in the tRNA spliced by the archaeon Haloferax volcanii cell extract.* RNA, 2000. **6**(7): p. 1019-30.
108. Filipowicz, W., et al., *RNA 3'-terminal phosphate cyclase activity and RNA ligation in HeLa cell extract.* Nucleic Acids Res, 1983. **11**(5): p. 1405-18.
109. Laski, F.A., et al., *Characterization of tRNA precursor splicing in mammalian extracts.* J Biol Chem, 1983. **258**(19): p. 11974-80.
110. Zillmann, M., M.A. Gorovsky, and E.M. Phizicky, *Conserved mechanism of tRNA splicing in eukaryotes.* Mol Cell Biol, 1991. **11**(11): p. 5410-6.
111. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.
112. Liu, Y., et al., *Genetic and Functional Analyses of Archaeal ATP-Dependent RNA Ligase in C/D Box sRNA Circularization and Ribosomal RNA Processing.* Front Mol Biosci, 2022. **9**: p. 811548.
113. Toffano-Nioche, C., et al., *RNA at 92 °C: the non-coding transcriptome of the hyperthermophilic archaeon Pyrococcus abyssi.* RNA Biol, 2013. **10**(7): p. 1211-20.
114. Yoshinari, S., et al., *Cleavage of 3'-terminal adenosine by archaeal ATP-dependent RNA ligase.* Sci Rep, 2017. **7**(1): p. 11662.
115. Suzuki, H., et al., *Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing.* Nucleic Acids Res, 2006. **34**(8): p. e63.
116. Sas-Chen, A., et al., *Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping.* Nature, 2020. **583**(7817): p. 638-643.
117. Wool, I.G., A. Glück, and Y. Endo, *Ribotoxin recognition of ribosomal RNA and a proposal for the mechanism of translocation.* Trends Biochem Sci, 1992. **17**(7): p. 266-9.
118. Szewczak, A.A., et al., *The conformation of the sarcin/ricin loop from 28S ribosomal RNA.* Proc Natl Acad Sci U S A, 1993. **90**(20): p. 9581-5.
119. Schmeing, T.M., et al., *The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA.* Science, 2009. **326**(5953): p. 688-694.
120. Jäger, D., et al., *Primary transcriptome map of the hyperthermophilic archaeon Thermococcus*

kodakarensis. BMC Genomics, 2014. **15**: p. 684.