Diversity of circular RNAs and RNA ligases in archaeal cells

Hubert F Becker, Caroline L'Hermitte-Stead, Hannu Myllykallio

Elsevier Editorial System(tm) for Biochimie + OA Mirror

Manuscript Draft

Manuscript Number: BIOCHI-D-19-00117R2

Title: Diversity of circular RNAs and RNA ligases in archaeal cells

Article Type: Mini Review

Section/Category: Special issue

Keywords: Circular RNA
RNA-seq
archaea
RNA ligase
Bioinformatics analyses

Corresponding Author: Dr. Hubert F. Becker,

Corresponding Author's Institution: Ecole Polytechnique

First Author: Hubert F. Becker

Order of Authors: Hubert F. Becker; Caroline L'Hermitte-Stead; Hannu Myllykallio

Manuscript Region of Origin: FRANCE

Abstract: Circular RNAs (circRNAs) differ structurally from other types of RNAs and are resistant against exoribonucleases. Although they have been detected in all domains of life, it remains unclear how circularization affects or changes functions of these ubiquitous nucleic acid circles. The biogenesis of circRNAs has been mostly described as a backsplicing event, but in archaea, where RNA splicing is a rare phenomenon, a second pathway for circRNA formation was described in the cases of rRNAs processing, tRNA intron excision, and Box C/D RNAs formation. At least in some archaeal species, circRNAs are formed by a ligation step catalyzed by an atypic homodimeric RNA ligase belonging to Rnl3 family. In this review, we describe archaeal circRNA transcriptomes obtained using high throughput sequencing technologies on Sulfolobus solfataricus, Pyrococcus abyssi and Nanoarchaeum equitans cells. We will discuss the distribution of circular RNAs among the different RNA categories and present the Rnl3 ligase family implicated in the circularization activity. Special focus is given for the description of phylogenetic distributions, protein structures, and substrate specificities of archaeal RNA ligases.

Opposed Reviewers:

Palaiseau, June 3th 2019

Hubert BECKER
hubert.becker@polytechnique.edu
Ecole Polytechnique
Laboratoire d'Optique et Biosciences
91128 Palaiseau

Dr Richard Buckingham
Editorial Board
Biochimie Journal

Object : Revision2 of the mini review BIOCHI-D-19-00117

Dear Dr. Buckingham,

Thank you for your message and the comments concerning our revised version of the mini-review "Diversity of circular RNAs and RNA ligases in archaeal cells"by Becker *et al*.

We have taken into account the reviewer's comments as described below :

1. We replace tRNA by Intron-tRNA when it was appropriate (Figure 1B, Table 1)

2. The number of circRNA against total RNA was indicated in the legend of Figure 1, for the most abundant circRNA classes.

3. In section 3, we simplified the classification of circRNA by combining intron-tRNA and Box C/D RNA in the same non coding RNA family.

Thank you for your attention to our response.

Cordially yours,

Hubert F. Becker

Conflicts of interest

Authors declare no conflict of interests.

Concerning the Mini-review submission:
Diversity of circular RNAs and RNA ligases in archaeal cells
Authors : Hubert F. Becker, C. L'Hermitte-Stead, H. Myllykallio

Author Agreement

On behalf of all the authors, I declare that all authors have seen and approved the final version of the manuscript being submitted. All authors are in agreement for publishing this article.

Hubert F. Becker

Review *Becker et al*, Diversity of circular RNAs and RNA ligases in archaeal cells

Abstract

Circular RNAs (circRNAs) differ structurally from other types of RNAs and are resistant against exoribonucleases. Although they have been detected in all domains of life, it remains unclear how circularization affects or changes functions of these ubiquitous nucleic acid circles. The biogenesis of circRNAs has been mostly described as a backsplicing event, but in archaea, where RNA splicing is a rare phenomenon, a second pathway for circRNA formation was described in the cases of rRNAs processing, tRNA intron excision, and Box C/D RNAs formation. At least in some archaeal species, circRNAs are formed by a ligation step catalyzed by an atypic homodimeric RNA ligase belonging to Rnl3 family. In this review, we describe archaeal circRNA transcriptomes obtained using high throughput sequencing technologies on *Sulfolobus solfataricus*, *Pyrococcus abyssi* and *Nanoarchaeum equitans* cells. We will discuss the distribution of circular RNAs among the different RNA categories and present the Rnl3 ligase family implicated in the circularization activity. Special focus is given for the description of phylogenetic distributions, protein structures, and substrate specificities of archaeal RNA ligases.

***Highlights (for review)**

Diversity of archaeal circular RNAs.

Enzymatic circularization of Box C/D RNAs.

Atypical dimeric RNA ligase (Rnl3) in hyperthermophiles

Identification of circular RNAs by RNA-seq and bioinformatics analyses

Diversity of circular RNAs and RNA ligases in archaeal cells

Hubert F. Becker[1,2,*], C. L'Hermitte-Stead[1], H. Myllykallio[1]

[1]LOB, Ecole Polytechnique, CNRS, INSERM, Institut Polytechnique de Paris, 91128 Palaiseau, France.

[2] Sorbonne Université, Faculté des Sciences et Ingénierie, 75005, Paris, France.

* Corresponding author

Abstract

Circular RNAs (circRNAs) differ structurally from other types of RNAs and are resistant against exoribonucleases. Although they have been detected in all domains of life, it remains unclear how circularization affects or changes functions of these ubiquitous nucleic acid circles. The biogenesis of circRNAs has been mostly described as a backsplicing event, but in archaea, where RNA splicing is a rare phenomenon, a second pathway for circRNA formation was described in the cases of rRNAs processing, tRNA intron excision, and Box C/D RNAs formation. At least in some archaeal species, circRNAs are formed by a ligation step catalyzed by an atypic homodimeric RNA ligase belonging to Rnl3 family. In this review, we describe archaeal circRNA transcriptomes obtained using high throughput sequencing technologies on *Sulfolobus solfataricus*, *Pyrococcus abyssi* and *Nanoarchaeum equitans* cells. We will discuss the distribution of circular RNAs among the different RNA categories and present the Rnl3 ligase family implicated in the circularization activity. Special focus is given for the description of phylogenetic distributions, protein structures, and substrate specificities of archaeal RNA ligases.

1. Introduction

Circular RNAs (circRNAs) differ structurally from other types of RNAs as their 3' and 5' ends are not free but covalently joined to form a circular structure. The lack of free ends confers to circular RNAs resistance toward exonucleases and therefore enhance their stability, but it remains unclear how the circularization affects or changes their function. CircRNAs are ubiquitous since they have been detected in all domains of life; even if in the past, they were observed as by-products of pre-RNA processing and were therefore interpreted as the result of erroneous splicing [1-3]. Following the discovery of the first circRNAs molecules, several types and categories of circRNAs have been detected in various organisms including *i)* the circular single-stranded RNA genomes of plant viroids and the hepatitis delta, *ii)* circular excised introns such as group I and II introns, and intron-tRNAs, *iii)* circular rRNAs intermediates in rRNA processing reactions, and *iv)* noncoding RNAs in archaea with snRNP functions (see [4, 5] for reviews). Recently, genome-wide RNA-seq experiments have revealed an abundant genome-wide presence of circRNAs in the mammalian transcriptome, with evolutionary conservation in terms of sequence and expression, suggesting specific cellular functions [6-9]. One of the most well-studied examples is the circular CDR1as RNA observed in human and mouse brain, with a direct implication in gene expression by miRNA sequestering. CDR1as was described as a sponge for miR-7 RNA since it massively bound mir-7 with more than 70 binding sites [10, 11]. Other known functions of circRNAs include roles as splicing or transcriptional regulators (see [12-14] for reviews). The use of circular RNA as biomarkers is also envisaged not only because the level of circRNAs between healthy and diseased subjects differ, but also because circRNAa are stable and highly abundant [15].

The biogenesis of circRNAs has been mostly described as a backsplicing event wherein a down-stream 5' donor splice site is joined to an upstream acceptor splice site, generating a circular RNA product [16, 17]. It was also demonstrated that circularization and splicing processes compete against each other, leading to the coexistence of linear and circular RNA forms resulting from the splicing of the same RNA precursor [18]. In addition, a second pathway for circRNA formation was described in archaea in the cases of rRNAs processing, tRNA intron excision and Box C/D RNAs formation, for which the circularization result from the intron ligation step of the splicing reaction. In archaea, the 16S and 23S pre-rRNAs are circularized at sequences adjacent to a BHB (Bulge-Helix-Bulge) motif after endonucleases

cleavage [19]. In *H. volcanii*, the Box C/D RNAs are released from pre-tRNA-Trp by a possible similar mechanism, involving endonuclease and ligase activities [20]. However, in the majority of cases, Box C/D RNAs are not intron-encoded, and it remains to be determined how the circularization occurs on a mature RNA with already processed mature extremities.

In many cases, circRNA formation results from the enzymatic activity of ATP-dependent RNA ligases. They are classified in six different families (Rnl1-Rnl6) based of their structure and substrate specificity and catalyze the joining of RNA molecules *via* three nucleotidyltransferase steps similar to those described for ATP-dependant DNA ligases (see section 4. *Rnl3: a new RNA ligase family with circularization activity* [21]). Bacteriophage T4 RNA ligase 1 is the founding member of the Rnl1 family which recognize specifically tRNA substrates with a break in the anticodon loop [22-24]. For T4 RNA ligase 2 (Rnl2 family), it was observed that the C-terminal domain is essential for the RNA-adenylation step and the binding to a nicked duplex RNA substrate [25]. Members of the Rnl3 family are found in several archaea and present an atypical homodimeric quaternary structure and circularize single-stranded RNA [26, 27]. Pnkp defines a new RNA ligase family (Rnl4) with signature structural and functional properties. Rnl4 family was identified, in several bacterial species, and it was reported that a stable heterocomplex of two bacterial proteins, Pnkp and Hen1, was able to repair transfer RNAs cleaved by ribotoxins *in vitro* [28, 29]. For Rnl5 ligases, the recognition of a nicked RNA duplex is supported by a specific amino-terminal domain, since Rnl5 lacks a C-terminal domain [30, 31]. Recently, Banerjee *et al* proposed a fungal tRNA ligase (Trl1) as the founder of an Rnl6 clade of ATP-dependent RNA ligases, due to its unique C-terminal domain. Trl1 is involved in the repair of tRNA and mRNA breaks that are generated during tRNA splicing and non-canonical mRNA splicing [32].

In this review, we will describe archaeal circRNA transcriptomes obtained using next-generation sequencing technologies. These experiments have been performed on three archaeal species (*Sulfolobus solfataricus*, *Pyrococcus abyssi* and *Nanoarchaeum equitans).* In particular, we will discuss *i)* the computational approach for circular RNAs identification, *ii)* the distribution of circRNAs among the different RNA categories and *iii)* the Rnl3 ligase family implicated in the circularization activity, in terms of phylogenetic distribution, protein structure, and substrate specificity.

2. Methodology for Circular RNA identification by RNA-seq deep sequencing

The detection of circRNAS from RNA-seq data can be achieved by more than ten widely used algorithms, developed mostly for applications in mammalian cells (see [33-35] for reviews). These methods are all based on the identification of back-spliced junctions, with reads that do not align continuously, but rather discontinuously, and in the different orientation with the respect to the reference genome. These results, in so-called inverted reads, are detected computationally. The detection of circRNAs with all of these existing algorithms on the same set of data drew contradictory conclusions with regard to their performance. In the particular cases of low circRNAs expression, all methods showed some biased performance on detecting true positives and filtering false positives. One effective improvement in the strategy was to take advantage of the morphology of circRNAs, and their high stability against exoribonucleases. That is a reason why RNA samples were treated by RNases R, an exonuclease degrading linear RNAs, to enhance the representation of circRNAs in the libraries. One perpetual challenge to improve the existing algorithms is to balance between sensitivity and precision of detecting and filtering the true positive circRNAs candidates.

In archaea, RNA splicing is less common and mostly observed in pre-tRNA and pre-rRNA with a uniform splicing mechanism [36]. Introns were spliced by an apparently archaeal-specific mechanism in which an endoribonuclease cuts a “bulge-helix-bulge” (BHB) motif that forms at exon-intron junctions, before recruiting RNA ligases [37]. The different RNA splicing scenario in archaea required the development of an adapted computational program for the identification of circRNAs. To date, successful identification of a plethora of circular RNAs has been performed in several archaeal organisms by a combination of biochemical, RNA-seq and computational approaches.

The first archaeal circular transcriptome was determined for the archaeon *Sulfolobus solfataricus* (acidophile and hyperthermophile) by sequencing the total RNA fraction by the Illumina approach [38]. Sequencing reads were mapped to the reference genome (NC_002754) using BLASTn, which allows accounting for local matches of all reads. Circularization junctions were predicted by first identifying uniquely aligned permuted reads, i.e. with a downstream stretch of the read sequence aligning to an upstream position in the genome (Figure 1A). Then, more restrictive criteria were applied; such as suppression of redundancy (junction within 1-3 nucleotides), a specific selection of junctions supported by 2

or more reads, deletion of template switching and finally the addition of an exo-ribonucleolytic enzyme RNase R treatment. The RNA-seq approach yielded 27.5 million reads sized in average 40 bp, and 97.82% of them mapped to the genome of *Sulfolobus solfataricus* with at least 33bp, indicating a relatively high quality of sequencing data. After application of all computational criteria, the set of circRNAs identified in *S. solfataricus* contained 37 genes, divided into several functional groups.

For *Pyrococcus abyssi* (a hyperthermophile) [39]), RNA-seq was performed using the Ion Torrent PGM Sequencer on two different types of samples; a total RNA sample and an RNA fraction obtained by co-immunoprecipitation with RNA ligase Pab1020 after formaldehyde crosslinking. All the reads were mapped to the *P. abyssi* reference genome (NC_000868, 1.765.118 base pairs) using BLASTn and the read inversion criteria were used to identify circular RNAs (Figure 1A). Typical sequencing runs yielded ~ 400.000 reads with a read size of 80 to 90 base pairs, and all of them mapped to the genome. For the enrichment in circular RNAs, we used RNase R exonuclease treatment that specifically degrades linear RNAs molecules in a 5'-3' direction. After analysis of the resulting 285.000 reads, 11-15% of them, obtained using an RNase R enrichment, were classified as circular. Then multiple strict constraints were used to identify circRNAs. These included *(i)* localization of the circular junction within three nucleotides, *(ii)* the presence of the circRNA in two independent experiments and *(iii)* the junction must be supported by more than three individual sequencing reads. This pipe-line finally led to the identification of 133 individual circRNA loci in *P. abyssi.*

RNA-seq methodology was also used to obtain a global overview of the small RNA diversity present in the *Nanoarchaeum equitans* cell, a small archaeon placed in the phylum termed "Nanoarchaeota" with a minimal, compact genome of only 490 kilobases and an extremely high gene density with very little noncoding DNA [40]. In this approach, small RNAs were enriched from a total RNA fraction and sequenced using Illumina HiSeq2000 technology, generating ~16 million reads with an average length of 62 nucleotides. Reads longer than 15 nucleotides (~12 million reads) were mapped to the *N. equitans* reference genome (NC_005213) with CLC Genomics Workbench 5.0 (CLC Bio, Denmark), applying specific parameters (mismatch, insertion, deletion costs and length fraction). They revealed two distinct circular junctions in the Box C/D RNA family.

As has been reported for RNA-seq studies in eukaryotic cells, it is also essential in archaea to apply multiple and restrictive constraints to allow attaining highly selective circRNAs identification. In addition to the developed computational criteria, we can highlight methodological improvements that have highly simplified circRNAs identification in *P. abyssi*, such as the circRNAs enrichment in the RNA sample by exonuclease treatment and the intersection of the circRNAs candidates obtained from independent and complementary experiments.

3. Functional classification of archaeal circRNAs

Few years before high-throughput sequencing approaches, circular RNAs belonging to several different functional RNA groups had already been identified in archaea. The presence of circular RNA molecules was revealed within the excised tRNA-introns in the mesophilic *Haloferax volcanii* [41, 42]. Once cleaved from its pre-tRNA-Trp and circularized, the Box C/D RNA included in the circular intron can guide methylation of nucleotides of the mature tRNA-Trp [20]. Several large circular introns derived from 23S and 16S rRNAs were described for hyperthermophilic archaea [43, 44], and proposed as rRNAs processing intermediates [19]. Moreover, the unexpected discovery of significant amounts of circular Box C/D RNAs, in addition to linear forms, in native complex with proteins required for Box C/D function was very intriguing [45]. In this case, the circular form of *P. furiosus* Box C/D RNAs could function as a guide for 2'-O-methylation on specific nucleotides in RNAs and tRNAs, by forming an RNP complex with three specific proteins (L7ae, Nop5, and fibrillarin) [45, 46].

Circular RNA transcriptome analysis by deep sequencing (RNA-Seq) has allowed a more exhaustive and non-biased identification of cirRNAs in the "third" domain of life. Interestingly, the results on circular transcriptomes for the three archaea (*P. abyssi*, *S. solfataricus,* and *N. equitans*) converged in terms of the functional categorization of circRNAs. Three non-coding RNA classes are frequently circular in all of these archaeal species (Box C/D RNAs, tRNA-intron and rRNA), but circRNAs were also identified in coding RNAs, and non-annotated RNAs (Figure 1B). The size of the sequencing reads is generally less than 200 nucleotides, in accordance with the size of most identified RNA categories, but for longer RNA genes (rRNA and coding RNAs), circles do not cover these loci completely. They were frequently found to present only a part of the given gene.

In all the studies described above, particularly high amount of circRNAs were found for Box C/D RNAs. For instance, 17 and 11 circular Box C/D RNAs loci were identified for *P. abyssi* and *S. solfataricus*, respectively (Table 1). More Box C/D RNA sequences had been identified for *P. abyssi* with a relaxation of selection criteria (only inversion of the reads during the mapping to the genome). This suggests that the majority of the Box C/D RNAs are circular in archaea (38 out of 56 in *P. abyssi* cells and 11/13 in *S. solfataricus* cells). RNA circularization in thermophiles could provide increased thermostability by limiting thermal denaturation of the stem structures formed between the 5' and 3' termini of Box C/D RNAs.

The excised intron from tRNA-Trp was also described as an abundant circRNA found in *S. solfataricus and P. abyssi* cells, with precise identification of the junction point. Intron from tRNA-Trp does not form a new circRNA category since the intron sequence encodes also for a Box C/D RNAs. It seems that the circularization occurs simultaneously during the splicing process and linear intermediates are not expected to occur. This has been suggested by the absence of enrichment in circular reads after exoribonuclease treatment during the library preparation [39] and in accordance with previous results that described a highly stable intron-tRNA-Trp in *Haloferax volcanii* [41]. Among the five circular tRNA identified in *S. solfataricus and P. abyssi* cells, tRNA-Trp is the only example overlapping between both studies. Both pre-tRNA-Trp contain a "BHB" motif, which will allow the circularization of the cleaved-out intron during the archaeal splicing mechanisms, as mentioned in section 2. Furthermore, the cleaved intron sequence carries a C/D Box motif that guides the 2'-O-methylation on nucleotides in the anticodon loop of mature tRNA-Trp [20], classifying the archaeal intron of tRNA-Trp in the Box C/D RNA family. The same functional conservation and circular-intron localization in the tRNA sequence was not observed for the other identified circular tRNA candidates.

Evidence for circular intermediates of rRNA (5S, 16S and 23S) was also observed but, in many cases, localization of the precise position of the junction point from permutated reads was far from evident, possibly reflecting the relatively high length and highly structured nature of these RNAs. Formation of the circular 5S rRNA could be compatible with the proposed maturation mechanism of the archaeal pre-5S-rRNA [19], but it is unclear whether this potential circular form corresponds to an additional processing intermediate or the mature 5S RNA.

Several additional circRNAs were observed within protein-coding genes, but the number of circular reads was generally low compared with linear reads for the same locus complicating their analyses. Furthermore, no spliced reads supporting a mature form of mRNA after intron excision were detected, suggesting that these circRNAs are probably not intron excision products [38]. Nevertheless, we cannot exclude an undescribed regulatory function for circRNAS detected within ORFs or the possibility that these "coding" circRNAs represent degradation products, based on their variability, small size and supposed random distribution between ORFs.

Finally, it is very likely that the number of identified circRNAs was an underestimate of the real number of circRNAs in archaeal cells, as it is hard to distinguish whether reads falling within an RNA belong to a circular or linear transcript. The threshold selection has been set arbitrarily to discriminate between background noise and significant results, which is particularly complicated when circRNAs mapped to repeat regions. In addition, very short circRNAs, compared to the read length or circRNAs with low expression level, might have escaped detection. Therefore, in quantitative and qualitative terms archaeal circRNAs might be even more abundant than reported in these RNA-Seq experiments.

As described above, for many examples in archaeal cells, circRNAs formation likely results from the enzymatic activity of ATP-dependent RNA ligases, which we are going to describe below.

4. Rnl3: a new RNA ligase family with circularization activity

RNA ligases, from all families, catalyze phosphodiester bond formation in a conserved three-step mechanism with different co-factors (ATP, GTP or NADP+) [21, 47]. In a first step, the side chain of lysine in the active site attacks the AMP group of ATP or NAD+, creating an enzyme-nucleotide ligase–AMP reaction intermediate. Secondly, the nucleotide is transferred from the enzyme to the 5'-phosphate of the RNA molecule, creating an activated 5' adenylated intermediate product (AppRNAs). In the third step, the phosphodiester bond is formed between the 3'-hydroxyl of the acceptor substrate and the 5'-adenylated phosphate of the donor, releasing AMP. The special feature of the RNA circularization reaction is that

donor and acceptor groups belong to the same RNA molecule, resulting in the intramolecular-ligation reaction of free termini.

The Rnl3 RNA ligase family was revealed by structural (X-ray crystallography) characterization of two archaeal RNA ligase from *Pyrococcus abyssi* (PDB ID 2VUG) [26] and *Methanobacterium thermoautotrophicum* (PDB ID 5D1P) [27]. These structures revealed a homodimeric structure, which is highly atypical for nucleic acid ligases. A structure of the third homolog was resolved and deposited on PDB website for an RNA ligase from *Aquifex aeolicus* (PDB ID 3QWU) (Osipiuk J. *et al*, not published), which is a thermophilic bacteria. A homodimeric structure of Rnl3 proteins is highly conserved, as revealed by the superimposition of the three Rnl3 high-resolution structures. A perfect superimposition was obtained for all domains of these proteins using MolSoft, an improved superimposition algorithm (Figure 2A). Each monomer consists of four domains: an N-terminal domain, a catalytic domain (nucleotidyltransferase domain), a dimerization domain and a C-terminal domain. The N-terminal domain contains a generally common topology with T4 RNA ligase 1 (T4 Rnl1) and has only been observed in these two enzymes to date. The catalytic domain showed, unsurprisingly, structural similarities with DNA and RNA ligases, and in a more general manner, with other members of the nucleotidyltransferase superfamily. Rnl3 proteins possess two additional C-terminal domains to the nucleotidyltransferase domain, which has not been observed in previous RNA ligase structures. The dimerization domain belongs to three-dimensional swapping domains described during the intertwined dimer formation process between two identical proteins [48, 49]. In this phenomenon, proteins often swap only a single secondary structure element such as a β-strand or α-helix rather than a whole domain of structure, which is also true for Rnl3 proteins. Swapping domains are believed to play an important role in the mechanism of dimerization or oligomerization in the protein evolution since some proteins remain functional only in their oligomeric state. A similar result was observed for Pab1020 RNA ligase since key mutations in this oligomerization domain generate aggregation of the protein and loss of activity. Dimerization of Rnl3 proteins could also be a common adaptation associated with thermophily in archaeal proteins [50]. The C-terminal domain is specific to this protein family and has no known structural homologs. It was shown that this C-terminal domain is essential for sealing of 3’-hydroxyl and 5’-phosphate ends, but not for ligase adenylation or for the

phosphodiester bond formation at the 5'-pre-adenylated RNA ([51], Becker unpublished results). Both C-terminal domains, flanking the nucleotidyltransferase domain, seem *i)* to be essential for the correct folding and solubility of the Rnl3 RNA ligase and ii) assist the catalytic domain for efficient ligation of the ends of RNA substrates.

5. Phylogenetic distribution of RNA ligases belonging to the Rnl3 family

We note that the RNA ligases Rnl3 display two unique features: a homodimeric structure and a single-stranded RNA circularization activity. Two complementary bioinformatic approaches, Pfam and UniProt, were used to identify new potential members of Rnl3 family in archaeal and bacterial kingdoms, using Pab1020 sequence and structure as entries for the search. For the computational analysis of UniProt BLAST results, only proteins with a level of similarity higher than 90% and conservation of all four protein domains, were selected. More than a typical BLAST search, Pfam uses alignments and a profile of hidden Markov models (HMM), which allow better remote homology detection [52, 53]. Results confirmed the distribution of Rnl3 members mainly within archaea (97 sequences) but also within bacteria [31 sequences (Figure 3)]. The observation of all examined phyla revealed that Rnl3 members were very frequently found in thermophiles or hyperthermophiles that typically grow between 50°C and 95°C. As mentioned in section 4, the dimeric organization of the Pab1020 RNA ligase is essential for its stability, solubility, and activity, i.e. the introduction of a mutation at essential residues in the dimerization domain would lead to total disorganization of the protein. On the other hand, deletion of the dimerization and the C-terminal domains restore the protein solubility, but the truncated Pab1020 RNA ligase exhibit now a monomeric structure, devoid of circular activity; confirming a direct correlation between the atypical dimeric structure and the circularization activity of Pab1020 RNA ligase. These results could be in accordance with observations that under drastic conditions, proteins require several adaptations that confer their ability to retain structure and function at high temperature. Without these necessary adaptations, proteins undergo irreversible unfolding, resulting in their aggregation [54]. Thermophilic proteins tend to have a prominent hydrophobic core, compensated by an increased number of salt-bridging, disulfide bridges and electrostatic interactions to maintain their activity at high temperatures [50]. Recent characterization of quaternary organization for thermophilic

proteins described an organization in dimer or higher oligomeric states to increase the rigidity of the individual subunits, promote tighter packing of the hydrophobic core, and reduce exposure of hydrophobic residues to solvent, e.g. for *Sulfolobus solfataricus* phosphotriesterase [55], acetyl-CoA synthetases from *Ignicoccus hospitalis [56]* and *Pyrobaculum aerophilum* [57]. Therefore, the higher-order quaternary structure can be advantageous for the stability of thermophilic proteins and may be a common feature of thermoadaptation.

6. Circularization activity of archaeal RNA ligases

To understand the mechanistic similarities between Rnl3 family members, we have compared enzymatic kinetic data obtained for RNA ligases Pab1020 from *P. abyssi* [26, 39] and MTH 1221 from *M. methanobacterium* [27, 58].

Both enzymes require the presence of a divalent cation, preferentially magnesium or manganese, and are ATP dependent. The mutants K95G (Pab1020) and K97A (MTH1221) were completely inert in forming circular RNAs or 5'-adenylated-RNAs, illustrating that in both active sites an invariant lysine is adenylated and represent the crucial catalytic residue (highlighted in yellow in Figure 2B). Adenyltransferase or ligation assays were performed at 50°C *in vitro*, and it was observed that only the last step of the reaction, the intramolecular phosphodiester bond formation, was affected and declined as the temperature decreased from optimal temperatures to 30°C. These results suggest conformational changes in the dimer architecture of Rnl3 proteins during the last step of the reaction. However, no information is available to consider if a sharing of both nucleotidyltransferase domains of the dimer is possible, or if they are fully independent. At first sight, both nucleotidyltransferase domains are distant from each other with an approximate distance of 3.5 nm between the catalytic residues of each monomer (Figure 2B). Nevertheless, the surface-charge distribution of Pab1020 RNA ligase shows clearly the presence of two positive grooves on each monomer until joining each catalytic site, and extending to the dimerization domain (Figures 2 C, D). The dimer interface may serve as a platform for RNA binding and can be a bridging point between both active sites, allowing RNA substrate to cross the dimer symmetry axis. This hypothesis is already supported by experimental data since mutation or

deletion of the dimerization domain of Pab1020 and MTH1221 abolish ligation reaction and trigger a partial aggregation of the proteins [27].

Studies addressing RNA specificity compared to DNA were also undertaken for both enzymes without leading to the same conclusion. In absence of ATP, MTH1221 converted a synthetic 24-mer pDNA into a DNA circle [27] whereas no ligation reaction was observed for Pab1020 RNA ligase, with a synthetic 18-mer pDNA substrate [39]. Recently, we confirmed the absence of DNA ligation for Pab1020 enzyme, with a physiological-like substrate. The DNA homolog sequence of SR4 Box C/D RNA, formed a stable DNA/protein complex with Pab1020, on native gel mobility assay, but no trace of circular DNA was detected. However, it has been shown that Pab1020 and MTH1221 allow scalable high yield adenylation of 5'-phosphorylated ssDNA [59] (Becker *et al*, results not published).

For the joining of RNA molecules, it was shown that synthetic RNA substrates, with the corresponding RNA sequence as tested previously for synthetic DNA, were efficiently circularized by both enzymes [27, 39]. By a combination of RNA immunoprecipitation assays and RNA-seq approaches, our genome-wide studies revealed 133 individual circRNA loci in *P. abyssi* cells [39]. To test whether RNAs interacting with Pab1020 RNA ligase correspond to physiologically significant substrates of this enzyme, we assayed the ligation activity of Pab1020 RNA ligase using the linear fluorescent Cy5-RNA transcripts for three circRNAs identified during the circRNAs high-throughput sequencing project on *P. abyssi* cells. Becker et al choose two Box C/D RNAs (SR4 and SR29) and the 5S rRNA for this *in vitro* ligation tests on respective transcripts. EMSA assays indicated that the three analyzed transcripts formed specific RNA-protein complexes at near stoichiometric conditions with increased affinity compared to synthetic RNA. To identify circular RNA molecules on 5S rRNA and Box C/D RNAs SR4 or SR29, we used RNase R exoribonuclease treatment to discriminate between circular products and linear substrate RNAs, since only circular RNAs are resistant to exoribonuclease treatment. The circular form was observed for all three selected transcripts (Box C/D RNAs SR4, SR29 and the 5S rRNA,) after incubation with Pab1020 RNA ligase and resistance to RNase R treatment. The RNA ligation activity of Pab1020 was confirmed by using inverse PCR, with divergent primers, which is expected to amplify only circular templates. For these three selected RNA molecules, we observed a full-length RT-PCR product confirming RNA circularization by Pab1020 [39].

7. Circular RNA biogenesis in organisms without identified Rnl3 enzyme

It will be necessary to obtain the circular transcriptome for more archaeal species to understand if a single or multiple pathways are required for the biogenesis of circRNAs. For that purpose, one first interesting question will be to resolve the mechanism of formation of the circular intron derived from pre-tRNA-Trp in *Haloferax volcanii* while this organism does not contain an RNA ligase from Rnl3 family. First, the existence of this circular-intron was confirmed by different approaches and second, the implication of an RNA ligase seems expected since the circular intron derived from an endonuclease-ligase two-step reaction, as described above. The absence of an RNA ligase from Rnl3 family, in *Haloferax volcanii*, implies that another protein catalyzes the circularization. Potential candidates are DNA ligases because it was shown that Rnl3 proteins are closely related to the CDC9 domain found in ATP-dependent DNA ligases, generating misannotation of members of the Rnl3 family as DNA ligases [39]. An excellent example of this is the putative *Aquifex aeolicus* RNA ligase (PDB code 3qwu) annotated as DNA ligase in the PDB databank. As discussed previously, this bacterial protein is very likely an RNA ligase, as it contains a dimer interface structure that is only conserved among Rnl3 family members.

Another option is that a clearly identified DNA ligase can ligate both DNA and RNA substrates, as it was described for a viroid that subverts host DNA ligase 1 into an RNA ligase [60]. Moreover, in vitro assays on Arabidopsis thaliana DNA ligase 1, using oligonucleotides substrates, have shown that this enzyme can ligate acceptor RNA to donor DNA [61].

8. Concluding remarks

With the recent high-throughput sequencing technologies, it is now clear that circRNAs are a large group of RNAs that are gaining considerable interest. Their mechanisms of formation are not well described today, but it is clear that circRNAs could be formed by backsplicing or by the direct ligation of 5’ and 3’ ends of linear RNAs. In archaea, it seems that the direct

ligation pathway may be preferred and that Rnl3 RNA ligases are the key enzymes required for the circularization activity.

In archaea, circRNAs seem to be stable and widely expressed, whereas their biological functions need to be better understood. Remarkably, emerging evidence indicates that circRNAs display important functions with implications in physiological and pathological processes in human cells. Thus, understanding the biogenesis and the biological functions of circRNAs will be highly relevant for furthering our understanding of these novels and ubiquitous circular molecules.

Conflicts of interest

Authors declare no conflict of interests.

Acknowledgments

We acknowledge Roxane Lestini for critical reading of the manuscript.

Figures legends

Figure 1: Identification of circRNAs using high throughput sequencing. (A) Obtained reads were aligned to the reference genome using Blastn. Reads were considered circular if at least two permuted matches were detected, whereas linear reads were not retained. (B) Representation of the different functional RNA classes considered circular in sequencing experiments obtained for *P. abyssi* and *S. solfataricus*. The percentage of reads supporting circular RNA of the different functional groups were indicated for *P. abyssi* / *S. solfataricus*. Box C/D RNA are the most abundant circular RNAs in *P. abyssi* with 38 circular form out of 59 Box C/D RNA identified in the genome, and a ratio of 11/13 for *S. solfataricus.* For tRNA containing an intron, 1/2 was observed circular for *P. abyssi* and 5/19 for *S. solfataricus*. For rRNA, circular forms were observed in 5S, 16S and 23S rRNA as intermediates during the maturation process as described in section 3.

Figure 2: Architecture of the protein members of the Rnl3 family. (A) Superimposition of the structures of three RNA ligases, Pab1020 (PDB code 2VUG), MTH1221 (PDB code 5D1P) and AQ1106 (PDB code 3QWU). The four domains of the proteins are totally superimposed and

are colored as follows: N-terminal domain, orange; catalytic domain, blue; dimerization, green; C-terminal domain, grey. MolSoft was used to superimpose multiple protein structures. (B) Structure of the Pab1020 homodimer. For each monomer, the nucleotidyltransferase domain is colored in blue and the catalytic lysine residue in yellow. (C) and (D) The surface electrostatic potential of Pab1020 protein was displayed using the Chimera software, with the negative potentials (red) and positive potentials (blue). Two orientations of the protein are represented with a yellow highlight of the catalytic lysine residue in the extension of a positive groove in panel D.

Figure 3: Phylogenetic distribution of RNA ligases from Rnl3 family in archaea and bacteria. The phylogenetic distribution was compiled from the Pfam protein families database in 2019, https://pfam.xfam.org and from Uniprot website https://www.uniprot.org/blast/. Next to the name of kingdom, class and order, the number of sequences / number of species are shown.

Bibliography

[1] B. Capel, A. Swain, S. Nicolis, A. Hacker, M. Walter, P. Koopman, P. Goodfellow, R. Lovell-Badge, Circular transcripts of the testis-determining gene Sry in adult mouse testis, Cell, 73 (1993) 1019-1030.
[2] C. Cocquerelle, B. Mascrez, D. Hetuin, B. Bailleul, Mis-splicing yields circular RNA molecules, FASEB J, 7 (1993) 155-160.
[3] J.M. Nigro, K.R. Cho, E.R. Fearon, S.E. Kern, J.M. Ruppert, J.D. Oliner, K.W. Kinzler, B. Vogelstein, Scrambled exons, Cell, 64 (1991) 607-613.
[4] I. Chen, C.Y. Chen, T.J. Chuang, Biogenesis, identification, and function of exonic circular RNAs, Wiley Interdiscip Rev RNA, 6 (2015) 563-579.
[5] E. Lasda, R. Parker, Circular RNAs: diversity of form and function, RNA, 20 (2015) 1829-1842.
[6] W.R. Jeck, J.A. Sorrentino, K. Wang, M.K. Slevin, C.E. Burd, J. Liu, W.F. Marzluff, N.E. Sharpless, Circular RNAs are abundant, conserved, and associated with ALU repeats, RNA, 19 (2013) 141-157.
[7] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S.D. Mackowiak, L.H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. le Noble, N. Rajewsky, Circular RNAs are a large class of animal RNAs with regulatory potency, Nature, 495 (2013) 333-338.
[8] A. Rybak-Wolf, C. Stottmeister, P. Glazar, M. Jens, N. Pino, S. Giusti, M. Hanan, M. Behm, O. Bartok, R. Ashwal-Fluss, M. Herzog, L. Schreyer, P. Papavasileiou, A. Ivanov, M. Ohman, D. Refojo, S. Kadener, N. Rajewsky, Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed, Mol Cell, 58 (2015) 870-885.
[9] J. Salzman, C. Gawad, P.L. Wang, N. Lacayo, P.O. Brown, Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types, PLoS One, 7 (2012) e30733.
[10] T.B. Hansen, T.I. Jensen, B.H. Clausen, J.B. Bramsen, B. Finsen, C.K. Damgaard, J. Kjems, Natural RNA circles function as efficient microRNA sponges, Nature, 495 (2013) 384-388.
[11] M. Piwecka, P. Glazar, L.R. Hernandez-Miranda, S. Memczak, S.A. Wolf, A. Rybak-Wolf, A. Filipchyk, F. Klironomos, C.A. Cerda Jara, P. Fenske, T. Trimbuch, V. Zywitza, M. Plass, L. Schreyer, S.

Ayoub, C. Kocks, R. Kuhn, C. Rosenmund, C. Birchmeier, N. Rajewsky, Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function, Science, 357 (2017).
[12] L.L. Chen, The biogenesis and emerging roles of circular RNAs, Nat Rev Mol Cell Biol, 17 (2016) 205-211.
[13] M. Cortes-Lopez, P. Miura, Emerging Functions of Circular RNAs, Yale J Biol Med, 89 (2016) 527-537.
[14] S. Qu, Y. Zhong, R. Shang, X. Zhang, W. Song, J. Kjems, H. Li, The emerging landscape of circular RNA in life processes, RNA Biol, 14 (2017) 992-999.
[15] W.L. Ng, T.B. Mohd Mohidin, K. Shukla, Functional role of circular RNAs in cancer development and progression, RNA Biol, 15 (2018) 995-1005.
[16] R. Ashwal-Fluss, M. Meyer, N.R. Pamudurti, A. Ivanov, O. Bartok, M. Hanan, N. Evantal, S. Memczak, N. Rajewsky, S. Kadener, circRNA biogenesis competes with pre-mRNA splicing, Mol Cell, 56 (2014) 55-66.
[17] Y. Wang, Z. Wang, Efficient backsplicing produces translatable circular mRNAs, RNA, 21 (2015) 172-179.
[18] S. Starke, I. Jost, O. Rossbach, T. Schneider, S. Schreiner, L.H. Hung, A. Bindereif, Exon circularization requires canonical splice signals, Cell Rep, 10 (2015) 103-111.
[19] T.H. Tang, T.S. Rozhdestvensky, B.C. d'Orval, M.L. Bortolin, H. Huber, B. Charpentier, C. Branlant, J.P. Bachellerie, J. Brosius, A. Huttenhofer, RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing, Nucleic Acids Res, 30 (2002) 921-930.
[20] B. Clouet d'Orval, M.L. Bortolin, C. Gaspin, J.P. Bachellerie, Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp, Nucleic Acids Res, 29 (2001) 4518-4529.
[21] J.M. Pascal, DNA and RNA ligases: structural variations and shared mechanisms, Curr Opin Struct Biol, 18 (2008) 96-105.
[22] K. El Omari, J. Ren, L.E. Bird, M.K. Bona, G. Klarmann, S.F. LeGrice, D.K. Stammers, Molecular architecture and ligand recognition determinants for T4 RNA ligase, J Biol Chem, 281 (2006) 1573-1579.
[23] R. Silber, V.G. Malathi, J. Hurwitz, Purification and properties of bacteriophage T4-induced RNA ligase, Proc Natl Acad Sci U S A, 69 (1972) 3009-3013.
[24] L.K. Wang, J. Nandakumar, B. Schwer, S. Shuman, The C-terminal domain of T4 RNA ligase 1 confers specificity for tRNA repair, RNA, 13 (2007) 1235-1244.
[25] J. Nandakumar, C.K. Ho, C.D. Lima, S. Shuman, RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2, J Biol Chem, 279 (2004) 31337-31347.
[26] M.A. Brooks, L. Meslet-Cladiere, M. Graille, J. Kuhn, K. Blondeau, H. Myllykallio, H. van Tilbeurgh, The structure of an archaeal homodimeric ligase which has RNA circularization activity, Protein Sci, 17 (2008) 1336-1345.
[27] C. Torchia, Y. Takagi, C.K. Ho, Archaeal RNA ligase is a homodimeric protein that catalyzes intramolecular ligation of single-stranded RNA and DNA, Nucleic Acids Res, 36 (2008) 6218-6227.
[28] C.M. Chan, C. Zhou, R.H. Huang, Reconstituting bacterial RNA repair and modification in vitro, Science, 326 (2009) 247.
[29] P. Smith, L.K. Wang, P.A. Nair, S. Shuman, The adenylyltransferase domain of bacterial Pnkp defines a unique RNA ligase family, Proc Natl Acad Sci U S A, 109 (2012) 2296-2301.
[30] A. Raymond, S. Shuman, Deinococcus radiodurans RNA ligase exemplifies a novel ligase clade with a distinctive N-terminal module that is important for 5'-PO4 nick sealing and ligase adenylylation but dispensable for phosphodiester formation at an adenylylated nick, Nucleic Acids Res, 35 (2007) 839-849.
[31] M.C. Unciuleac, S. Shuman, Characterization of a novel eukaryal nick-sealing RNA ligase from Naegleria gruberi, RNA, 21 (2015) 824-832.
[32] A. Banerjee, S. Ghosh, Y. Goldgur, S. Shuman, Structure and two-metal mechanism of fungal tRNA ligase, Nucleic Acids Res, (2018).

[33] T.B. Hansen, Improved circRNA Identification by Combining Prediction Algorithms, Front Cell Dev Biol, 6 (2018) 20.
[34] T. Jakobi, C. Dieterich, Computational approaches for circular RNA analysis, Wiley Interdiscip Rev RNA, 10 (2019) e1528.
[35] X. Zeng, W. Lin, M. Guo, Q. Zou, A comprehensive overview and evaluation of circular RNA detection tools, PLoS Comput Biol, 13 (2017) e1005420.
[36] J. Lykke-Andersen, C. Aagaard, M. Semionenkov, R.A. Garrett, Archaeal introns: splicing, intercellular mobility and evolution, Trends Biochem Sci, 22 (1997) 326-331.
[37] M. Englert, K. Sheppard, A. Aslanian, J.R. Yates, 3rd, D. Soll, Archaeal 3'-phosphate RNA splicing ligase characterization identifies the missing component in tRNA maturation, Proc Natl Acad Sci U S A, 108 (2011) 1290-1295.
[38] M. Danan, S. Schwartz, S. Edelheit, R. Sorek, Transcriptome-wide discovery of circular RNAs in Archaea, Nucleic Acids Res, 40 (2011) 3131-3142.
[39] H.F. Becker, A. Heliou, K. Djaout, R. Lestini, M. Regnier, H. Myllykallio, High-throughput sequencing reveals circular substrates for an archaeal RNA ligase, RNA Biol, 14 (2017) 1075-1085.
[40] L. Randau, RNA processing in the minimal organism Nanoarchaeum equitans, Genome Biol, 13 (2012) R63.
[41] S.R. Salgia, S.K. Singh, P. Gurha, R. Gupta, Two reactions of Haloferax volcanii RNA splicing enzymes: joining of exons and circularization of introns, RNA, 9 (2003) 319-330.
[42] S.K. Singh, P. Gurha, E.J. Tran, E.S. Maxwell, R. Gupta, Sequential 2'-O-methylation of archaeal pre-tRNATrp nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA, J Biol Chem, 279 (2004) 47661-47671.
[43] J.Z. Dalgaard, R.A. Garrett, Protein-coding introns from the 23S rRNA-encoding gene form stable circles in the hyperthermophilic archaeon Pyrobaculum organotrophum, Gene, 121 (1992) 103-110.
[44] J. Lykke-Andersen, R.A. Garrett, Structural characteristics of the stable RNA introns of archaeal hyperthermophiles and their splicing junctions, J Mol Biol, 243 (1994) 846-855.
[45] N.G. Starostina, S. Marshburn, L.S. Johnson, S.R. Eddy, R.M. Terns, M.P. Terns, Circular box C/D RNAs in Pyrococcus furiosus, Proc Natl Acad Sci U S A, 101 (2004) 14097-14101.
[46] B. Clouet-d'Orval, C. Gaspin, A. Mougin, Two different mechanisms for tRNA ribose methylation in Archaea: a short survey, Biochimie, 87 (2005) 889-895.
[47] S. Shuman, C.D. Lima, The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases, Curr Opin Struct Biol, 14 (2004) 757-764.
[48] S. Kundu, R.L. Jernigan, Molecular mechanism of domain swapping in proteins: an analysis of slower motions, Biophys J, 86 (2004) 3846-3854.
[49] F. Rousseau, J. Schymkowitz, L.S. Itzhaki, Implications of 3D domain swapping for protein folding, misfolding and function, Adv Exp Med Biol, 747 (2012) 137-152.
[50] C.J. Reed, H. Lewis, E. Trejo, V. Winston, C. Evilia, Protein adaptations in archaeal extremophiles, Archaea, 2013 (2013) 373275.
[51] C.K. Ho, L.K. Wang, C.D. Lima, S. Shuman, Structure and mechanism of RNA ligase, Structure, 12 (2004) 327-339.
[52] S. El-Gebali, J. Mistry, A. Bateman, S.R. Eddy, A. Luciani, S.C. Potter, M. Qureshi, L.J. Richardson, G.A. Salazar, A. Smart, E.L.L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S.C.E. Tosatto, R.D. Finn, The Pfam protein families database in 2019, Nucleic Acids Res, 47 (2019) D427-D432.
[53] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, Nucleic Acids Res, 39 (2011) W29-37.
[54] S.J. Tomazic, A.M. Klibanov, Mechanisms of irreversible thermal inactivation of Bacillus alpha-amylases, J Biol Chem, 263 (1988) 3086-3091.
[55] P. Del Vecchio, M. Elias, L. Merone, G. Graziano, J. Dupuy, L. Mandrich, P. Carullo, B. Fournier, D. Rochu, M. Rossi, P. Masson, E. Chabriere, G. Manco, Structural determinants of the high thermal stability of SsoPox from the hyperthermophilic archaeon Sulfolobus solfataricus, Extremophiles, 13 (2009) 461-470.

[56] F. Mayer, U. Kuper, C. Meyer, S. Daxer, V. Muller, R. Rachel, H. Huber, AMP-forming acetyl coenzyme A synthetase in the outermost membrane of the hyperthermophilic crenarchaeon Ignicoccus hospitalis, J Bacteriol, 194 (2012) 1572-1581.
[57] C. Brasen, C. Urbanke, P. Schonheit, A novel octameric AMP-forming acetyl-CoA synthetase from the hyperthermophilic crenarchaeon Pyrobaculum aerophilum, FEBS Lett, 579 (2005) 477-482.
[58] H. Gu, S. Yoshinari, R. Ghosh, A.V. Ignatochkina, P.D. Gollnick, K.S. Murakami, C.K. Ho, Structural and mutational analysis of archaeal ATP-dependent RNA ligase identifies amino acids required for RNA binding and catalysis, Nucleic Acids Res, 44 (2016) 2337-2347.
[59] A.M. Zhelkovsky, L.A. McReynolds, Simple and efficient synthesis of 5' pre-adenylated DNA using thermostable RNA ligase, Nucleic Acids Res, 39 (2011) e117.
[60] M.A. Nohales, R. Flores, J.A. Daros, Viroid RNA redirects host DNA ligase 1 to act as an RNA ligase, Proc Natl Acad Sci U S A, 109 (2012) 13805-13810.
[61] Y.Q. Wu, B. Hohn, A. Ziemienowic, Characterization of an ATP-dependent type I DNA ligase from Arabidopsis thaliana, Plant Mol Biol, 46 (2001) 161-170.

Table 1

P. abyssi				*S. solfataricus*		*N. equitans*	
Class*	Size*	Class**	Size**	Class**	Size**	Class	Size
38 Box C/D	59-72	17 Box C/D	59-72	11 Box C/D	26-62	2 Box C/D	71-74
5 tRNA	26 -114	1 tRNA-intron	71	5 tRNA-intron	21-65		
6 rRNA	26-86			3 rRNA	123-3096		
13 NA	21-101	3 NA	61-65	9 non coding RNA	31-69		
71 inside ORFs	14-216			5 inside ORFs	26-163		
				1 H/ACA RNA	70		

Table 1 : CircRNAs distribution in functional RNA categories for *P. abyssi, S. solfataricus and N. equitans* RNA-seq experiments. NA= non-annotated, * circRNA identification by inverted reads alignment, ** application of more strict criteria including the circular junction identification. For *P. abyssi*, among the 5 tRNAs identified, the circularization has been only supported for the intron of tRNA-Trp (second column), with a high number of reads and the exact identification of the junction-nucleotides.

Figure 1

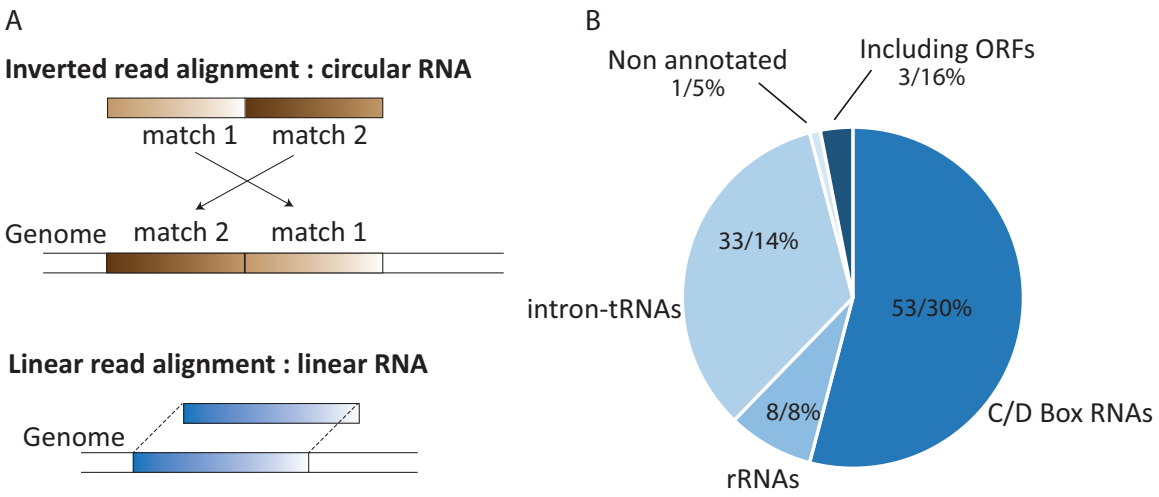

A
Inverted read alignment : circular RNA
match 1
match 2
Genome
match 2
match 1
Linear read alignment : linear RNA
Genome
B
Non annotated
1/5%
Including ORFs
3/16%
33/14%
intron-tRNAs
53/30%
8/8%
C/D Box RNAs
rRNAs

Figure2

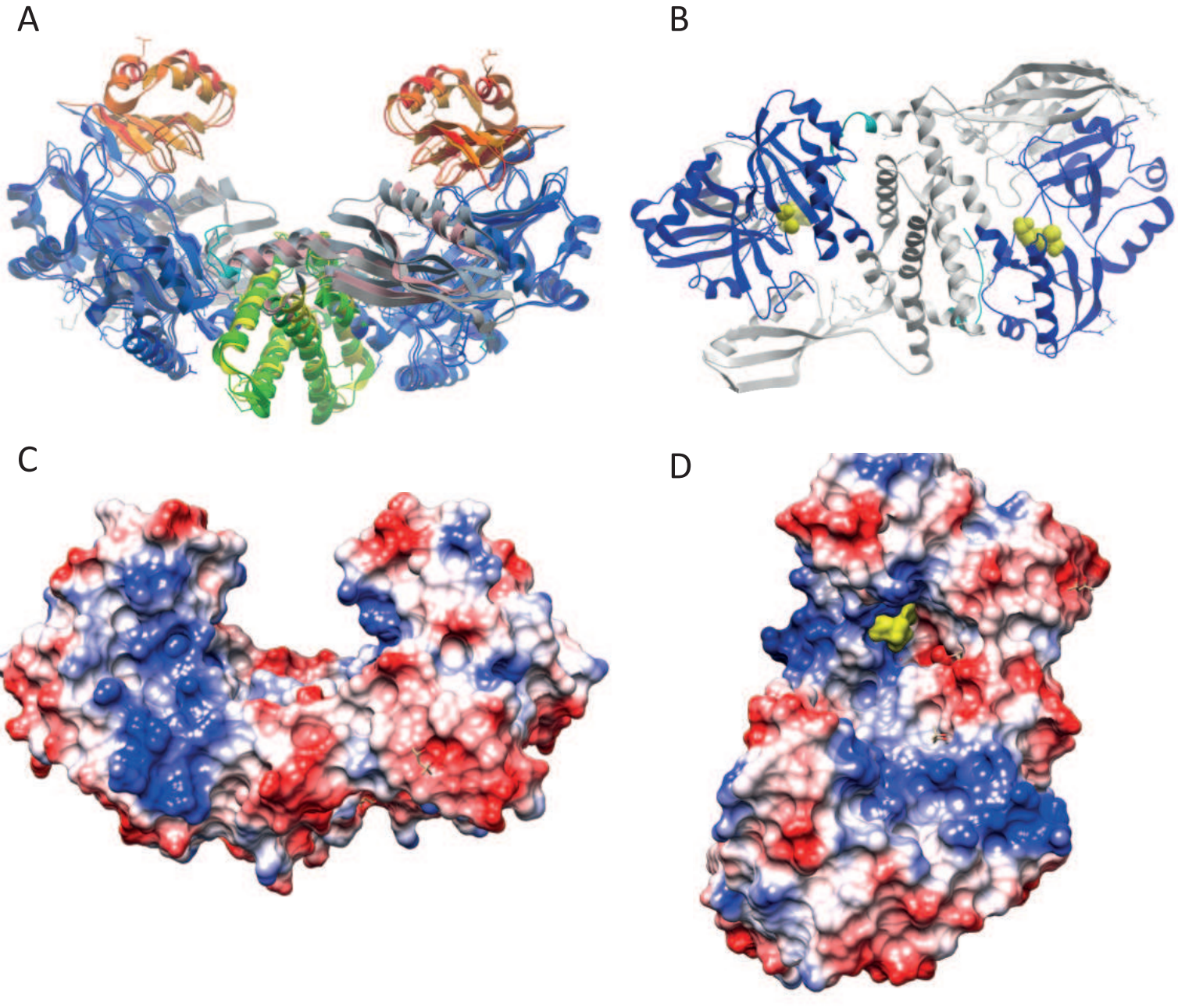
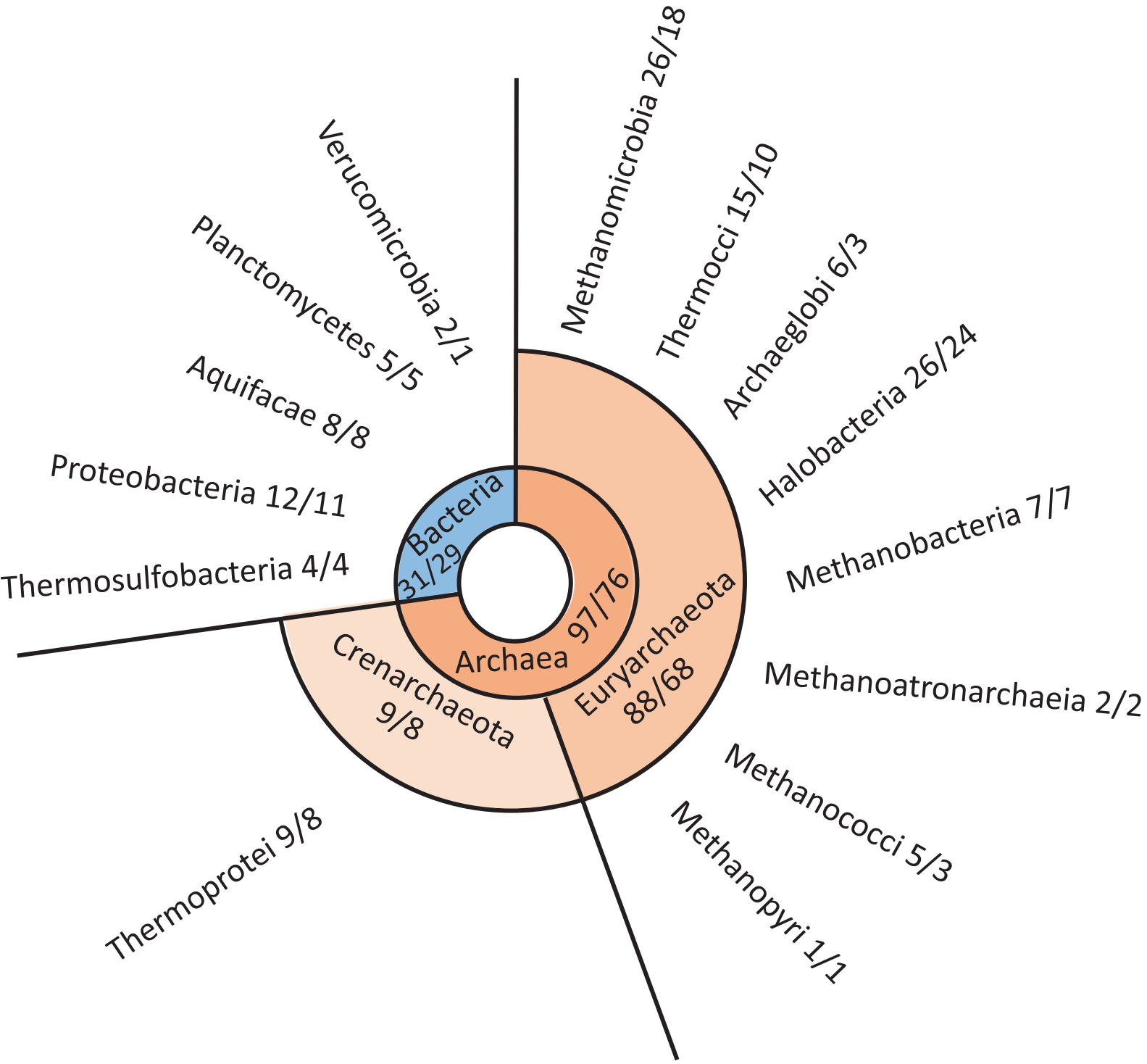


Figure3


Methanomicrobia 26/18
Thermocci 15/10
Archaeglobi 6/3
Halobacteria 26/24
Methanobacteria 7/7
Methanoatronarchaeia 2/2
Methanococci 5/3
Methanopyri 1/1
Euryarchaeota 88/68
Crenarchaeota 9/8
Thermoprotei 9/8
Archaea 97/76
Bacteria 31/29
Thermosulfobacteria 4/4
Proteobacteria 12/11
Aquifacae 8/8
Planctomycetes 5/5
Verucomicrobia 2/1