Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs

Sébastien Muller, Fabrice Leclerc, Isabelle Behm-Ansmant, Jean-Baptiste Fourmann, Bruno Charpentier and Christiane Branlant*

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP, Nancy Université, Faculté des Sciences et Techniques, 54506 Vandoeuvre-lès-Nancy, France

Received January 15, 2008; Revised February 7, 2008; Accepted February 8, 2008

ABSTRACT
How far do H/ACA sRNPs contribute to rRNA pseudouridylation in Archaea was still an open question. Hence here, by computational search in three Pyrococcus genomes, we identified seven H/ACA sRNAs and predicted their target sites in rRNAs. In parallel, we experimentally identified 17 residues in P. abyssi rRNAs. By in vitro reconstitution of H/ACA sRNPs, we assigned 15 out of the 17 residues to the 7 identified H/ACA sRNAs: one H/ACA motif can guide up to three distinct pseudouridylation. Interestingly, by using a 23S rRNA fragment as the substrate, one of the two remaining residues could be formed in vitro by the aCBF5/aNOP10/aGAR1 complex without guide sRNA. Our results shed light on structural constraints in archaeal H/ACA sRNPs: the length of helix H2 is of 5 or 6bps, the distance between the ANA motif and the targeted U residue is of 14 or 15nts, and the stability of the interaction formed by the substrate rRNA and the 3’-guide sequence is more important than that formed with the 5’-guide sequence. Surprisingly, we showed that a sRNA–rRNA interaction with the targeted uridine in a single-stranded 5’-UNN-3’ trinucleotide instead of the canonical 5’-UN-3’ dinucleotide is functional.

INTRODUCTION
Conversion of uridine into pseudouridine (Ψ) residues is one of the most abundant post-transcriptional modifications of tRNAs, rRNAs and UsnRNAs (1). Compared to U residues, Ψ residues can form an additional hydrogen bond at the N1–H position. Furthermore, the carbon–carbon link between the ribose and the base limits the flexibility of the ribose backbone of Ψ residues, which favours and stabilizes base-pair interactions (2). A role of Ψ residues in stabilization of the tRNA 3D structure has also been well documented (3), and the functional importance of Ψ residues in U2 snRNA for the activity of splicing complexes was demonstrated (4–7). Eukaryal rRNAs contain a large number of Ψ residues compared to bacterial rRNAs and they are concentrated in rRNA regions expected to play important functional roles, in particular in domains IV and V, which are directly involved in the peptidyl transferase activity (8–10). Taken individually, pseudouridylation in rRNAs are not essential for cell growth. However, the complete abolition of Ψ formation in rRNAs is deleterious for ribosome assembly and activity (10–12). Recent data suggest their possible involvement in: (i) subunit interaction (12–14), (ii) the translocation step (14,15), (iii) translation termination (12,16) and (iv) folding of 23S rRNA in an active form at the peptidyl transferase centre (PTC) (11,12,17). The large number of pseudouridylation sites in eukaryal rRNAs compared to bacterial rRNAs is explained by the use of different catalysts: standalone enzymes carrying both the RNA recognition capability and the catalytic activity are used in bacteria (10,18), whereas U to Ψ conversions are catalyzed by H/ACA snoRNPs in eukarya (19,20). The H/ACA snoRNPs contain four proteins and an H/ACA snoRNA that defines the targeted U residue by base-pair interaction with the rRNA. Recent data revealed a similar RNA-guided system in archaea (21–24). Most of the eukaryal H/ACA snoRNAs contain two characteristic stem-loop structures, with an internal loop (pseudouridylation pocket), that is complementary to two nucleotide stretches bordering the targeted U residue. Each of the stem-loops is flanked by a conserved motif (H and

*To whom correspondence should be addressed. Tel: +33 3 83 68 43 03; Fax: +33 3 83 68 43 07; Email: christiane.branlant@maem.uhp-nancy.fr

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
ACA, respectively). In the following part of this manuscript, one stem-loop structure containing a pseudouridylation pocket flanked by one of the conserved 3’ sequence will be denoted an H/ACA motif. In the RNA duplex formed between the H/ACA motif and the substrate rRNA sequence, the targeted U residue and its 3’ nucleotide are both single-stranded. The archaeal sRNA counterparts of the snoRNAs have more diverse architectures. They are composed of one, two or three H/ACA motifs (21–26). In addition, each archaeal H/ACA stem-loop structure contains a K-turn or a K-loop which binds motif, and aGAR1 may facilitate the H/ACA sRNP turnover (22–24). Both L7Ae and aGAR1 strongly reinforce the sRNP activity (23). The crystal structures of H/ACA sRNP protein complexes and of an entire H/ACA sRNP were recently solved at high resolution (27–31). The ANA sequence is needed for binding of aCBF5 to the guide RNA (28,29,31). In the crystal structure, aCBF5 also interacts with the pseudouridylation pocket, helix H1 and helix H2 of the H/ACA motif (31). Recently a 3D structure obtained for an H/ACA sRNP devoid of the L7Ae protein and bound to its RNA substrate, revealed contacts between aCBF5 and the target rRNA–H/ACA sRNA duplex (30). However, little is known on the structural constraints required for formation of an active RNA target–H/ACA sRNA interaction. The two NMR structures established for a complex formed between an H/ACA stem-loop structure and a small target RNA revealed the capability of the two RNAs to interact together, in the absence of protein, at the high concentration used for NMR analysis (32,33). In these structures, the heterologous helix formed by interaction of the RNA target with the 3’-guide element of the sRNA is stacked on helix H1, while the heterologous helix formed by interaction of the RNA target with the 5’-guide element of the sRNA is stacked on helix H2.

At present, in contrast to this extensive knowledge on H/ACA sRNPs, little is known on the number and location of Ψ residues in archaeal rRNAs. Their presence has only been investigated in *Halobacterium halobium* (34), *Halococcus marismortui* (35,36), *Sulfolobus acidocaldarius* (37) and *Archaeoglobus fulgidus* (21). Unfortunately, the search for putative H/ACA sRNA genes by computational analysis was made for other archaeal species: *Pyrococcus abyssi* (22,24,38), *Methanococcus jannaschii* (26) and *Thermococcus kodakarenis* (25). The utilization of RNomics approaches for the search of H/ACA sRNAs in archaea is even more scarce, it was only applied to the *A. fulgidus* (21) and *Sulfolobus solfataricus* (39) species: three H/ACA sRNAs and one single sRNA were identified, respectively. For a better definition of the rules that govern the H/ACA sRNP specificities and efficiencies, it was of high importance to identify all the putative H/ACA sRNAs of a given species as well as Ψ residues in its rRNAs and then, to try to assign the detected Ψ residues to the detected H/ACA sRNAs.

To this end, here, we used different experimental approaches in order to define the target sites of the putative H/ACA sRNAs that we identified by a computational analysis of the genomic sequences of three *Pyrococcus* species. These three species, *Pyrococcus abyssi*, *Pyrococcus furiosus* and *Pyrococcus horikoshii*, are hyperthermophiles (optimal growth temperature between 95 and 100°C). As a first step, we developed and used various computational approaches to identify all the putative H/ACA sRNAs of these species and their putative target sites in rRNAs. Then, to test for the presence of Ψ residues at the predicted sites in 16S and 23S rRNAs, we applied the RT-CMCT approach to large segments of the *P. abyssi* 16S and 23S rRNAs that were including the predicted pseudouridylation sites. Finally, to confirm the role of the identified H/ACA motifs in formation of the identified Ψ residues, we tested the activities of reconstituted H/ACA sRNPs on small rRNA fragments containing the expected target U residues. Finally, on the basis of the data obtained, we drew conclusions on structural constraints to which H/ACA sRNAs and the H/ACA sRNA–rRNA interactions are subjected.

**MATERIALS AND METHODS**

**Extraction of inter-coding-regions (ICR) from *Pyrococcus* genomes**

The sequences and annotations of the archaeal genomes *Pyrococcus abyssi* GE5, *Pyrococcus horikoshii* OT3 and *Pyrococcus furiosus* DSM3638 were downloaded in Fasta and Genbank formats from NCBI ftp site, ftp://ftp.ncbi.nih.gov/genomes/Bacteria. In each genome, the positions of DNA sequences corresponding to ORFs or template sequences of known stable RNAs (rRNA, tRNA, RNaseP and 7S RNA) were listed in a table denoted ‘position table’. Based on this table, the remaining segments of the genomes were extracted automatically from the genomic sequences, by using a script written in awk and perl languages (ExtractICR). Only sequences exceeding 15 nts were collected. ExtractICR flanks each of the linking sequences by their two 15-nt long bordering sequences. These exported sequences, denoted ICRs, were formatted and assembled in data bases, using Readseq (ftp://jubio.bio.indiana.edu/molbio/readseq) and Formatdb (NCBI toolkit).

**Selection of conserved ICRs and search for H/ACA sRNA genes**

A Blast version adapted for multi-alignment was used to compare the ICR sequences. First, repeated elements (more than 50-nt long segments repeated several times in the genome) were removed by comparison of all the
ICRs extracted from one given genome. Then, the ICRs from one given pyrococcal species were compared to those of the other two species. Based on a statistical analysis, the criteria retained to consider that an ICR shows a significant degree of conservation in the three species was the presence of at least a stretch of 18 nts with an identical sequence in these species or a 21-nt long sequence with only one base-pair substitution in one or two of the three species. The selected ICRs were then aligned using Clustal-W (40). GeneMarks (41) was used for elimination of the conserved ICRs corresponding to mis-annotated ORFs. An RNAMOT (42) descriptor, that was designed for C/D box sRNA gene detection, was used to identify these genes in the conserved ICRs. Finally, another RNAMOT descriptor was used to screen for the presence of H/ACA sRNA genes. It was based on some of the known structural features of the archaeal H/ACA motifs: the presence of two helices H1 (at least 7 bps) and H2 (at least 5 bps) flanking an internal loop (each strand including 5 to 11 nts). The length of the apical loop was allowed to vary between 8 and 35 nts. The descriptor included an ACA trinucleotide flanking helix H1. Only H/ACA motifs having at least one putative target site in rRNAs or tRNAs that fitted to the rules defined for H/ACA sRNA–rRNA interactions were retained. To this end, RNAMOT descriptors were built for each putative H/ACA motif as described in (25).

Search for H/ACA sRNA genes in entire genomic sequences

Then, based on the results obtained with RNAMOT, we used the ERPIN (43) software for searches with higher constraints in the entire Pyrococcus genomes. ERPIN builds helix and single-strand lod score profiles from sequence alignments and screens genome sequences for occurrences of these profiles. The H/ACA sRNAs that we identified by screening the conserved ICRs, together with the H/ACA sRNAs identified experimentally in A. fulgidus (21), were used to build the ERPIN profile as described in (25). As above, for each new candidate detected by this approach, RNAMOT was used to predict target sites in rRNAs (25).

The nucleotide sequences of all the H/ACA sRNAs detected in this study and the positions of their template sequences in the archaeal genomes are accessible at http://tagc.univ-mrs.fr/erpin/.

P. abyssi cultures

*P. abyssi* strain GE5 cells were grown as described (44) at 95°C in Vent Sulfothermophiles Medium (20 g/l NaCl, 0.25 g/l KCl, 0.05 g/l NaBr, 0.5 g/l SrCl₂-6H₂O, 0.08 g/l boric acid, 3 g/l PIPES, 1 g/l yeast extract, 4 g/l peptone, 1 g/l Resazurine, 200 g/l MgSO₄, 50 g/l CaCl₂, 50 g/l KH₂PO₄ pH 6.8). *P. abyssi* cell growth was stopped at the end of the exponential phase.

Total RNA isolation

The *P. abyssi* cells were collected by centrifugation. After washing in 1 M sorbitol, 25 mM Hepes pH 7.0, they were frozen and stored at −80°C. The method described by Chomczynski and Sacchi (1986) (45) was used for extraction of total RNA. About 10¹⁰ cells were dissolved in 4 ml of solution D (4 M guanidine thiocyanate, 25 mM sodium citrate pH 7.0, 0.5% sarcosyl, 0.1 M β-mercaptoethanol). The extracted RNAs were recovered by phenol/chloroform extraction, followed by ethanol precipitation using 0.3 M sodium acetate. They were dissolved in bi-distilled water and quantified.

Northern blot analysis

About 10 μg of *P. abyssi* total RNA and 5′-end labelled DNA size markers (100 bp DNA ladder, MBI Fermentas) were fractionated in parallel on 6% denaturing polyacrylamide gel (8 M urea, 0.5× TBE buffer). After electrophoresis on a Hybond-N+ membrane (Amersham) and by UV-crosslinking, a pre-hybridization was carried out for 1 h at 58°C in SSPE buffer (0.9 M NaCl, 47 mM Na₂HPO₄-2H₂O, 6 mM EDTA pH 7.4, containing 1 g/l Ficoll, 1 g/l polyvinylpyrrolidone, 1 g/l BSA, 0.5% SDS). Oligonucleotide probes complementary to the predicted H/ACA sRNAs (Table S1 in Supplementary data) were 5′-end labelled with [γ-³²P]ATP and T4 polynucleotide kinase for 1 h at 37°C. Hybridization was carried out at 58°C for 16 h in the presence of 100 ng of the labelled oligonucleotide. The membranes were washed four times in SSPE buffer at 42°C for 5 min and the hybridization bands were visualized on a Typhoon 9410 (Amersham Biosciences).

Mapping of pseudouridine (Ψ) residues in *P. abyssi* sRNAs

The N-cyclohexyl-N′-(2-morpholinoethyl)-carboximide metho-p-toluolsulfonate (CMCT) RNA treatment was adapted from Bakan and Olfengard (1993) (46), as previously described (47). CMCT modifications were performed on 5 μg of *P. abyssi* total RNA. Sequences of CMCT modifications were identified by primer extension analysis, using the AMV RT (QBiogene, USA). The sequences of the 5′-end labelled primers that we used are given in Table S1 (Supplementary data). RNA sequencing was done on 20 μg of *P. abyssi* total RNA.

Recombinant protein production

The recombinant GST-L7Ae, GST-aNOP10, GST-agaR1, GST-aCBF5 proteins were produced in *E. coli* BL21 CodonPlus cells (Novagen). The GST-tag was cleaved by the protease pre tease in the course of purification performed on Glutathione-Sepharose 4B (Pharmacia), as previously described (23).

In vitro transcription of H/ACA sRNAs and their rRNA substrates

The Pab19, Pab21, Pab35, Pab40, Pab91, Pab105 and Pab160 sRNA sequences were PCR amplified from the GE5 genomic DNA using a forward primer containing the T7 RNA polymerase promoter and a second primer complementary to the expected 3′ end of the sRNA (Table S1 in Supplementary data). The amplified DNA fragments were cloned in the pTAdv vector (Clontech) and sequenced. In *vitro* transcriptions with the T7 RNA polymerase were performed as
The H/ACA sRNA transcripts were purified by gel electrophoresis. 

Template DNA encoding the RNA targets derived from *P. abyssi* rRNAs were obtained by annealing two complementary oligonucleotides. In *vitro* transcription was performed as previously described (23), in the presence of 20μCi [α-32P] NTP (800 Ci/mmol), 0.13 mM of the same NTP and 4 mM of each of the three other NTPs. When using RNase P1 for the digestion, [α-32P]UTP was used for labelling. When RNase T2 was used, the identity of the [α-32P]NTP used was defined by the residue located 3′ to the targeted U residue. The sequences of the in *vitro* transcribed substrates are given in Table S2 in Supplementary data.

**H/ACA sRNP reconstitution and test of their activity**

As previously described (48), the unlabelled sRNA (4 pmol) and [α-32P] NTP-labelled targets (150 fmol) were mixed with the four L7Ac, aCBF5, aNOP10 and aGAR1 recombinant proteins at a 200 nM concentration, at room temperature in buffer D (150 mM KCl; 1.5 mM MgCl2; 0.2 mM EDTA; 20 mM HEPES, pH 7.9). The mixture was incubated at 65°C for 80 min. Then, formation of /C9 residue was detected after either T2 or P1 RNase digestion. The 3′-monophosphate nucleotides produced by RNase T2 digestion or the 5′-monophosphate residues obtained after P1 RNase digestion were fractionated by thin-layer chromatography (TLC) (mono or 2D) as previously described, using the N1 buffer for 1D TLC and N1-N2 or N1-R2 buffers for 2D TLC (48). The radioactivity in the spots or the bands was quantified with a phosphorimager, by using the ImageQuant software. When digestion was achieved with RNase T2, the yield of /C9 formation (expressed in mol of /C9 residue per mol of target RNA) was estimated taking into account the total number of residues located 5′ to the incorporated.

**Figure 1.** The computational process used for the search of H/ACA genes in ICRs. Panel A: ICR definition. ICRs include 15 bps upstream and downstream of the segments linking the ORFs or template sequences for known RNAs. Panel B: the helices H1 and H2, and loop size criteria used to design the RNAMOT descriptor devoted to the search for H/ACA sRNA genes are shown. Panel C: Example of base-pair criteria used to search for possible rRNA targets of the candidate H/ACA motifs. Panel D: the 5 computer-based steps used for the search of H/ACA genes: (1) sequences and annotations were downloaded, (2) sequences of ICRs were extracted from the genomic sequences and assembled in databases, (3) after elimination of ICRs containing repeated sequences, the ICR sequences from one species were compared to those of other species by using Blast, 119 conserved ICRs were selected, (4) protein-coding sequences and known C/D box sRNA genes were filtered out of the conserved sequences and (5) RNAMOT descriptors were used for the search of H/ACA genes in ICRs, five putative genes were detected.
labelled nucleotide. After P1 RNase digestion, we only took into consideration the total number of U residues in the target RNA.

RESULTS

Analysis of conserved ICRs in Pyrococcus identifies five conserved putative H/ACA sRNAs

In order to make a link between H/ACA sRNAs and pseudouridylation in rRNAs, in the P. abyssi species, we first made a complete identification of H/ACA–sRNA motifs in P. abyssi. This was done in two steps because, when we started this study, only a limited number of archaeal H/ACA sRNAs had been identified. As a first step, we applied a phylogenetic approach to the Pyrococcus genus. The idea was that DNA segments, which link ORFs and/or template sequences for stable known RNAs (rRNAs, tRNAs, RNase P, 7S RNA) (Inter-Coding-Regions, ICRs) and that bear long stretches of conserved sequences in three Pyrococcus species P. abyssi, P. furiosus and P. horikoshii, may correspond to genes for functional non-coding RNAs. Based on the few H/ACA sRNAs known at that time (three from A. fulgidus) (21) some characteristic features of H/ACA sRNA motifs could be delineated. They were used to build an RNAMOT descriptor for the search of H/ACA sRNA genes. Then, RNAMOT profiles were built for the search of the possible rRNA target sites of each candidate H/ACA motif. By using this approach (see the details in Materials and Methods), we detected five putative H/ACA sRNA genes, that were common to the three species. Four of them have been recently characterized in P. furiosus by the use of a computational approach based on G/C content analysis, namely, the Pf1, Pf3, Pf6 and Pf7 sRNAs (22,49). The counterparts that we identified in P. abyssi and P. horikoshii are designated as Pab21, Pab105, Pab35 and Pab40 sRNAs (P. abyssi) and Pho21, Pho105, Pho35 and Pho40 sRNAs (P. horikoshii), respectively (Figure 2). The fifth common H/ACA sRNA, which had not been characterized by other teams, is denoted Pab19, Phu19 and Pho19 in P. abyssi, P. furiosus and P. horikoshii, respectively. We previously used it to settle conditions for in vitro reconstitution of active H/ACA sRNPs (23). Taking into account the numerous compensatory base-pair mutations in the three species studied and in T. kodakarenisis, we could propose relevant secondary structures for each of the five sRNAs (Figure 2). Only for sRNA Pab21 and for motif 2 in sRNA Pab40, it was difficult to make a choice between two possible 2D structures that were both containing a K-loop motif (Figure 2). Production of the H/ACA sRNAs Pab21, Pab35, Pab40, Pab91 and Pab105 in P. abyssi was verified by Northern blot analysis (Figure 3), and two forms of Pab21 sRNA with or without the C/D box motif were found to be present in vivo (Figure 3).

Screening of complete genomes with ERPIN identifies two additional common putative H/ACA sRNAs in Pyrococci

Some H/ACA motifs may have escaped detection when using the above comparative approach. Therefore, we took advantage of the increased knowledge on H/ACA sRNAs, which was brought by the 15 pyrococcal H/ACA sRNAs that we identified, to develop another computational approach based on the ERPIN software (25,43). To this end, the sequences of the 3 known H/ACA sRNAs from A. fulgidus and the 15 pyrococcal H/ACA sRNAs were aligned by ERPIN, on the basis of their proposed 2D structures and conserved sequence elements. As recently described (25), by comparison of such a profile with complete archaeal genomic sequences, ERPIN can predict H/ACA sRNA genes with a high degree of efficiency. By using different sets of constraints for the lengths of the 3'- and 5'-guide strands of the pseudouridylation pocket, we detected two other putative H/ACA sRNAs common to the three Pyrococcus species. One of them was also recently detected in P. furiosus (sRNA Pf9) by another approach (24). Its counterparts in P. abyssi and P. horikoshii were denoted Pab160 and Pho160, respectively (Figure 2F). The second H/ACA sRNA identified (Pab19, Phu19 and Pho19 in P. abyssi, P. furiosus and P. horikoshii, respectively) had not been detected previously, probably because of its extended 5'-guide sequence (Figure 2). Nucleotide sequence conservation in the three species and secondary structure conservation by compensatory base changes strongly suggest the presence of a small 5' stem-loop structure upstream from the H/ACA motif in sRNA Pab19. Note that the Pab19 pseudouridylation pocket is larger than usual due to the length of its 5'-guide sequence (15 nts) (Figure 2F). By applying the ERPIN search approach with other kinds of relaxed constraints to each of the 3 Pyrococcus genomes, we did not find any other putative H/ACA sRNA gene. Therefore, we concluded that Pyrococcus species probably contain only seven H/ACA sRNAs, which are highly conserved in this genus. In addition, they are also conserved in Thermococcus kodakarenisis that belongs to the same order as the Pyrococcus genus (25).

The seven pyrococcal H/ACA sRNAs are predicted to target 21 sites in rRNAs

We used the RNAMOT software (25) to predict possible target sites in rRNAs for each of the seven identified putative H/ACA sRNAs, and this in the three species studied. One up to four distinct target sites were predicted for each putative pseudouridylation pocket (Figure 4). Taken together, 14 and 7 target sites were predicted in 23S and 16S rRNAs, respectively. Note that motif 1 in sRNA Pab35 and sRNA Pab160 are both predicted to guide modifications at position 922 in 16S rRNA and at position 2672 in 23S rRNA (Table 1). Only 8 of the 17 sites that we predicted for the five previously identified P. furiosus H/ACA sRNAs had been proposed (22). Strong conservation of the possibility to form the rRNA–sRNA interactions are observed in the three species (Figure 4). Note that in the non-canonical interaction proposed for sRNA Pab19, two single-stranded residues are found 3' to the targeted U residue (5'-UGC-3') (Figure 4G). We verified that none of the seven identified H/ACA sRNAs can act on tRNAs. In addition, no other putative H/ACA...
Figure 2. Sequences and proposed secondary structures of the seven pyrococcal H/ACA sRNA candidates. The sequences and proposed secondary structures for the seven candidate *P. abyssi* sRNAs, Pab21 (A), Pab35 (B), Pab40 (C), Pab91 (D), Pab105 (E), Pab160 (F) and Pab19 (G) sRNAs, are shown. The ANA sequence at the 3' end of the RNA and the K-turn or K-loop motif in the apical part of the H/ACA motif are boxed. Base substitutions in the *P. furiosus* (*P.f.*), *P. horikoshii* (*P.h.*), and *T. kodakarensis* (*T.k.*)(25) sRNAs are shown. The two putative foldings, which can be proposed for sRNA Pab21 and for motif 2 in sRNA Pab40, are shown (insets in panels A and C). Only the conformation shown in the entire molecule turned to be functional.
The experimental analysis confirmed Ψ formation at 12 of the 21 predicted sites (positions 27, 891, 922 in 16S rRNA, and positions 1932, 2549, 2554, 2588, 2672, 2685, 2697, 2794 and 2930 in 23S rRNA) (Table 1 and Figure 5). In addition, the absence of Ψ formation was clearly demonstrated at six of the predicted positions (positions 892, 995 and 1122 in 16S rRNA and 278, 2250, and 2552) (Table 1 and Figure 5). Note that three of these unmodified positions were previously proposed to be modified in *P. furiosus* (22). Based on the present data, they are unlikely to be modified in this species. For three of the positions predicted to be modified in *P. abyssi*, the experimental analysis was obscured by the presence of an RT pause at the level of the stop expected after CMCT modification (positions 1017 in 16S rRNA, 1377 and 2016 in 23S rRNA). Therefore, we could not determine whether U to Ψ conversion occurred at these positions. However, for one of them (2016 in 23S rRNA), formation of a Ψ residue was detected at the corresponding position in two other archaeal species: *H. halobium* (34) and *H. marismortui* (35,36), and in nearly all bacterial and eukaryal organisms which were studied up to now (10).

Interestingly, in the course of this analysis, Ψ residues were found at two positions that were not predicted to be targeted by any of the identified H/ACA sRNAs (positions 2585 and 2603 in 23S rRNA). Even by relaxing the constraints in the ERPIN search, we could not detect putative H/ACA sRNAs for these positions. Therefore, their formation may be sRNA independent.

### The seven H/ACA sRNPs are active in vitro

To get experimental supports for the relationships that we established between Ψ residues in *P. abyssi* rRNAs and the identified H/ACA sRNAs, we used the H/ACA sRNP *in vitro* reconstitution method which was developed in the laboratory (23). To this end, each of the seven identified H/ACA sRNAs was transcribed, and the four recombinant aCBF5, aNOP10, aGAR1 and L7Ae proteins were produced. Assembly of individual proteins and different combinations of them on the H/ACA sRNAs was tested by electrophoresis-mobility shift assay (EMSA). Each of the identified H/ACA sRNA could be assembled into a sRNP (data not shown), which was a clear demonstration that they were true H/ACA sRNAs. Then, we measured the activities of the reconstituted particles on all their predicted target sites. To this end, we used small *P. abyssi* rRNA fragments containing from 18 up to 31 nts, most of them had a single target U residue located in the middle of the molecule, except 3 substrates that were containing 2 or 3 target U residues, because of the close vicinity of these U residues in 23S rRNA (see Table S2 in Supplementary data). Therefore, the activity of one up to three distinct sRNPs was tested on a given RNA substrate. This activity was measured by the nearest-neighbour method as previously described (23,48). T2 RNase digestion was performed on RNA substrates labelled with [*γ-32P]*ATP, CTP, GTP or UTP. P1 RNase hydrolysis was done on [*γ-32P]*UTP labelled RNA substrates. Each RNA substrate was incubated with the four core proteins (LCNG) at 65°C as
Figure 4. Predicted rRNAs target sites of the 11 H/ACA motifs. Interactions between each H/ACA motif and its putative rRNA target sequences are shown for the seven candidates sRNAs Pab21 (A), Pab35 (B), Pab40 (C), Pab91 (D), Pab105 (E), Pab160 (F) and Pab19 (G). The expected targeted U residues are indicated by arrows and numbers giving their positions in the 16S or 23S rRNAs. The distance between the K-turn motif and the pseudouridylation pocket is given in bps, the length between the ANA sequence and the pseudouridylation pocket is given in nts. The two possible intermolecular interactions proposed for sRNA Pab21 and motif 2 in sRNA Pab40 result from the two possible alternative conformations of these H/ACA motifs (Figure 2 Panels A and C). The interactions found to be functional in in vitro assays are indicated by (+), the inactive ones are shown by (−).
Previously, described, in the presence or the absence of the H/ACA sRNA (23). Then, the RNAs were extracted, digested, and fractionated by thin layer chromatography (as described in Materials and Methods) (Figure 6).

Only 15 of the 21 predicted sites were found to be modified by the reconstituted sRNPs and the rates of U to Ψ conversion were ranging from 30 to 100% (Figure 6 and Table 1). The rRNA segments predicted to be modified by two distinct H/ACA motifs were in fact modified by only one of them (motif 1 in sRNA Pab35 modifies position 2672 in 23S rRNA and sRNA Pab160 acts position 922 in 16S rRNA). The data obtained confirmed that each of the 11 H/ACA motifs is capable to guide pseudouridylation. Even the Pab19 H/ACA motif, which has a large internal loop and forms a non-canonical interaction with its substrate, was active (Figure 6D). Deletion of the additional single-stranded C residue, which is located 3' to the 5'-UN-3'dinucleotide (mutant S-C1019Δ) (Figures 4G and 6D), showed that the Pab19 sRNP modifies the WT and mutated RNA to Ψ conversion at positions 1377 and 2794 in 23S rRNA. Motif 1 in sRNA Pab105 only guides modification at position 1377 and 2794 in 23S rRNA. Therefore, both sRNAs Pab35 and Pab105 can guide modifications at four positions in rRNAs. Interestingly, we found that all the H/ACA sRNAs that contain a single H/ACA motif (Pab21, Pab91, Pab105 and Pab19) sRNAs can guide modifications at more than one position in rRNAs.

Some of the H/ACA motifs can guide modification at more than one position in rRNAs

Some of the H/ACA motifs can guide modifications at two or even three distinct positions. For instance, altogether, motifs 1 and 2 in sRNA Pab35 can guide U to Ψ conversion at positions 2672, 2930, 2549 and 2697 in 23S rRNA. Motif 1 in sRNA Pab105 only guides modification at one position (2554 in 23S rRNA), while motif 2 in this sRNA can act at positions 1377 and 2794 in 23S rRNA and 27 in 16S rRNA. Therefore, both sRNAs Pab35 and Pab105 can guide modifications at four positions in rRNAs. Interestingly, we found that all the H/ACA sRNAs that contain a single H/ACA motif (Pab21, Pab91, Pab105 and Pab19) guide modification at a unique position in rRNAs. Altogether, the 11 H/ACA motifs of the P. abyssi sRNAs can guide U to Ψ conversion at 15 sites in the P. abyssi rRNAs (Table 1). Presence of Ψ residue was detected at 12 of these sites (Figure 5 and Table 1). As mentioned above, for technical reasons, modification could not be tested experimentally at the three other positions. However, the strong activity measured in vitro at these three rRNA positions is a strong argument for the occurrence of these modifications in vivo.

Constraints on the H/ACA sRNA structure and H/ACA sRNA–target RNA interaction

Based on the above data, we tried to define rules for H/ACA sRNA structure and interactions. In several of
Figure 5. Identification of Ψ residues in the *P. abyssi* 16S (Panel A) and 23S (Panel B) rRNAs by the RT-CMCT method. The *P. abyssi* total RNA was treated in the absence (−) or the presence of CMCT (+), for 2, 10, or 20 min as indicated on the top of the lanes. The CMCT treatment was (+) or was not (−) followed by an alkaline treatment at pH 10.4 as indicated above the lanes. The appearance of an RT stop of increasing intensity after this treatment reveals the presence of a Ψ residue. They are indicated by horizontal arrows. Lanes U, G, C and A correspond to the RNA sequencing ladder. Positions of residues in 16S or 23S rRNAs are given on the right side of the autoradiograms.
Figure 6. In vitro activity tests of the seven reconstituted H/ACA sRNP s. H/ACA sRNP particles were assembled by using 4 pmol of in vitro transcribed H/ACA sRNA, the four recombinant proteins L7Ae (L), αCBF5 (C), αNOP10 (N) and αGAR1 (G) (LCNG) (200 nM each) and 150 fmol of radio-labelled in vitro transcribed RNA substrate as previously described (23). The putative target position in the RNA substrate is given on the left side of the chromatograms. After a 80-min incubation at 65 °C, the RNAs were extracted and digested with T2 (Panels A and B) or P1 nuclease (Panels C and D). The nucleotide 3′ or 5′ monophosphate, that are respectively released, were fractionated by 2D (Panel A, B and C) or 1D (Panel D) thin layer chromatography. When RNase T2 was used for the digestion, the RNA template was labelled by incorporation of [α-32P] ATP, CTP, GTP or UTP depending on the identity of the residue located 3′ to the targeted U residue. When P1 nuclease was used for digestion, labelling was achieved by [α-32P] UTP incorporation. 2D TLC on cellulose plates were performed using either the N1-N2 (Panels A and C), or the N1-R2 (Panel B) buffers and the N1 buffer was used for 1D TLC. Positions of Ap, Cp, Gp, Up and ψp or pU and pψ in the chromatograms are indicated as well as their numbers in the substrate RNAs. The ψ spots are circled. The indicated numbers of ψ moles formed per mole of RNA substrates were calculated as explained in Materials and Methods.
the predicted sRNA–target RNA interactions that turned to be non-functional, helix H2 was too short (4 bps) or too long (7 bps). The most frequent length in functional interactions (15 or 14 nts). Interestingly also, the base-pair distance between the ACA trinucleotide and the targeted residue U2585 or U2603 (Figure 7). The aCBF5/aNOP10/aGAR1 complex can modify position 55 in tRNAs in the absence of guide sRNA (50–52), we tested the in vitro activity of this complex at these two 23S rRNA positions. The assays were performed using two 20-nt long rRNA fragments (Table S1) containing, respectively, residue U2585 or U2603 (Figure 7). The aCBF5/aNOP10/aGAR1 complex was active on the rRNA fragment containing residue U2603 (46% yield) (Figure 7), but not on that containing residue U2585. Interestingly, we observed that the sequence containing the residue 2603 can be folded into a stem-loop structure showing some similarity with the TΨC stem-loop of tRNA (Figure 7). We concluded that formation of residue Ψ2603 may be catalyzed by the free aCBF5-aNOP10-aGAR1 complex, while an unidentified catalyst may act at position U2585.

**DISCUSSION**

Application of the various computational approaches that we developed for the search of H/ACA sRNAs in archaeal genomes turned to be highly efficient, since two of the seven sRNAs detected in *P. furiosus* were not found by other approaches previously applied to one of this species (22,24). Our blind detection of C/D box
sRNAs in conserved ICRs allowed the detection of 45 out of the 49 known C/D box sRNAs which are conserved in the three pyrococcal species. This finding illustrates the efficiency of our ICR-based approach for the search of non-coding RNAs of unknown structural characteristics. Up to now, computational predictions of the H/ACA sRNA–rRNA targets were not reliable. This is illustrated by the numerous corrections that we made in previous predictions. Our experimental search for Ψ residues in \textit{P. abyssi} rRNAs, together with the use of the H/ACA sRNPs reconstitution and activity assays, turned to be essential to identify the sites which are really targeted by the identified sRNAs. However, by using the structural rules that we established for functional sRNA–rRNA interactions, computational predictions will be more reliable.

A high number of Ψ residues in \textit{P. abyssi} rRNAs as compared to other archaea

Based on experimental analysis of Ψ residues in rRNAs and reconstitution of H/ACA sRNPs, the \textit{P. abyssi} rRNA regions that we studied (20 and 27% of the 16S and 23S rRNAs, respectively) probably contain 17 Ψ residues. These rRNA regions were selected because they contain the highest number of post-transcriptionally modified residues in all living organisms (10). However, we cannot exclude the possibility that some additional Ψ residues are formed without the use of guide RNA, in \textit{P. abyssi} rRNA segments located outside of these regions. Up to now, 17 Ψ residues is the highest number of Ψ residues found in archaeal rRNAs. Only three and four Ψ residues were found upon complete analysis of the \textit{H. marismortui}, \textit{H. halobium} 16S and 23S rRNAs, respectively (34–36). Six Ψ residues were detected in domains II, IV and V of the \textit{S. acidocaldarius} 23S rRNA and five Ψ residues were found in \textit{A. fulgidus} rRNAs when looking for H/ACA sRNA target sites (21,37). Interestingly, the two halophilic species, \textit{H. marismortui} and \textit{H. halobium}, which have the smallest number of Ψ residues, grow at 50 and 42°C, respectively. \textit{S. acidocaldarius} and \textit{A. fulgidus} grow at 80 and 83°C, respectively, whereas \textit{P. abyssi} optimally grows at 98°C. Therefore, as already proposed for archaeal 2'-O-methylations (53), there may be a correlation between the number of Ψ residues in archaeal rRNAs and the growth temperature of the organism. Confirmation of this statement requires further analysis on a larger number of archaeal species growing at different temperatures.

Ψ residues in \textit{P. abyssi} 23S rRNA are concentrated at the PTC

Interestingly, 2 and 8 of the 13 Ψ residues expected to be present in \textit{P. abyssi} 23S rRNA are located in the functional domains IV and V, respectively (Figure 8). Noticeably, residue 2016 in domain IV corresponds to one of two highly conserved Ψ residues in stem-loop structure 69 (SLS69). Its conservation is probably explained by the high functional importance of SLS69: (i) it contacts both the 16S rRNA and the tRNA bound at the acceptor site (A site) (12,14), (ii) it is located at the subunit interface and, (iii) it was proposed to play a role in tRNA translocation (9,54,55) and in translation termination (12,16). Formation of two Ψ residues in the terminal loop of SLS69, including the \textit{P. abyssi} Ψ2016 counterpart, was found to confer a growth advantage in \textit{S. cerevisiae} (12,17).

Eight Ψ residues are concentrated in domain V of the \textit{P. abyssi} 23S rRNA. This domain is expected to be directly involved in the catalytic activity of the PTC. Interestingly, these 8 Ψ residues are all located in helical segments (Figure 8). They may stabilize the conformation of this 23S rRNA region at the high growth temperature of \textit{P. abyssi}. A cooperative effect of Ψ residues in stabilizing RNA conformation, was previously proposed (10). A need for this cooperative effect may explain the presence of 4, 2 and 2 Ψ residues in close vicinity in helices 89, 90 and 93, respectively. Only three of these eight Ψ residues were detected in several eukarya, including \textit{S. cerevisiae} (Figure 8) (34,56).

Altogether, four pseudouridylation sites in \textit{P. abyssi} 23S rRNAs are highly conserved in the large eukaryal rRNAs (positions 2616 in domain IV, 2549, 2588 and 2603 in domain V) (Figure 8). In addition, residues Ψ2930 detected in domain VI is also present in mouse and human rRNAs. Position 2698 in \textit{P. abyssi} 23S rRNA corresponds to a frequently pseudouridylated position in eukarya. Instead of U2698, U2697 is converted into a Ψ residue in \textit{P. abyssi}. Altogether, this comparison of \textit{P. abyssi} and eukaryal pseudouridylation in rRNAs suggests that conserved pseudouridylation sites in the large rRNA may have appeared early in evolution. Interestingly, one of the conserved Ψ residues in domain V (Ψ2603) is not guided by an H/ACA sRNA in \textit{P. abyssi}.

A base substitution in sRNA Pab40 may generate a different specificity as compared to sRNAs Afu4, Pf7 and Pho40

Like the \textit{P. abyssi} Pab40 sRNA, the Afu4, Pf7, and Pho40 sRNAs contain three H/ACA motifs. The target sites proposed for motifs 1 and 3 in Afu4 and Pf7 sRNAs (21,22) are identical to the ones determined experimentally for motifs 1 and 3 in Pab40 sRNA, respectively. In contrast, whereas motif 2 in the Afu4 and Pf7 sRNAs were proposed to guide modification at positions 2601 and 302 in the \textit{A. fulgidus} and \textit{P. furiosus} 23S rRNAs, respectively (corresponding to positions 2647 and 278 in \textit{P. abyssi} 23S rRNA, respectively), we found that motif 2 in sRNA Pab40 is active at position 1932 in \textit{P. abyssi} 23S rRNA. A difference of specificity between motif 2 in Pab40 and motifs 2 in Afu4, Pf7, and Pho40 sRNAs can be explained by a point mutation in the K-turn motif (a G to U substitution at position 1932) (corresponding to positions 2647 and 278 in \textit{P. abyssi} 23S rRNA, respectively). The target sites previously proposed for motifs 2 in Afu4 and Pf7 sRNAs are not in agreement with the structural rules that we established for active H/ACA sRNA–rRNA interactions. According to these rules, positions 2632 in the \textit{A. fulgidus} 23S rRNA and 2701 in the \textit{P. furiosus} 23S rRNA, that both correspond to position 2678 in \textit{P. abyssi} 23S rRNA, are expected to be the true target...
sites (Figure S1). U 2678 is not converted into a \( \Psi \) residue in *P. abyssi*, however, this is the case in *H. marismortui* (35,36), *H. halobium* (34), *Drosophila melanogaster* and *Mouse musculus* (10) (Figure 8). The absence of modification at position 2678 in the *P. abyssi* 23S rRNA shows how a single point mutation can dramatically modify the specificity of an H/ACA motif.

Note that two other proposed target sites in *P. furiosus* rRNAs (corresponding to positions 892 in 16S rRNA and 2552 in *P. abyssi* 23S rRNA and which were expected to be guided by sRNAs Pf1 and Pf3 respectively, Table 1), are invalidated by our experimental rules. In addition, some *bona fide* target sites were not previously predicted in *P. furiosus* rRNAs (Table 1). Altogether, these data
strenthen the importance to verify experimentally the proposed H/ACA target sites.

**Poor conservation of pseudouridylation sites between archaeal orders**

Except for motif 2 in sRNA Pab40, H/ACA sRNAs and their target sites are conserved in the three *Pyrococcus* species studied as well as in the phylogenetically related *T. kodakarenis* species (25). In contrast, conservation of P positions is poor between species of different orders. Only one of the three P residues found in the *H. halobium* and *H. marismortui* 23S rRNAs and only one of the six P residues found in the *S. acidocaldarius* 23S rRNA are conserved in *P. abyssi* (Figure 8). *A. fulgidus* is more closely related to *Pyrococcus* species than *Sulfolobus*, *Halobacterium*, and *Halocarcula* species. Accordingly, our experimental data show that three of the six P residues detected in the *A. fulgidus* 23S rRNA are conserved in *P. abyssi* as well as one modified position in 16S rRNA.

The four P residues detected in 16S rRNA are located in functional areas of the 30S subunit

The location of P residues in the bacterial and eukaryal SSU rRNAs is rather variable from one species to the other. Interestingly, two of the four P residues that we detected in the *P. abyssi* 16S rRNA (positions 27 and 891), are located within or very close to the essential central pseudoknot of the 16S rRNA (57). They may play a role in its formation or stability. This central pseudoknot is in close vicinity of the P site (58). Residue P22 also belongs to a 16S rRNA segment located at the P site, and residue P1017 belongs to a segment involved in A site formation. Hence, the four P residues present in the *P. abyssi* 16S rRNA are located at or very near the A and P sites. Noticeably, no P residue was detected in the 16S rRNA from *H. volcanii* while *S. solfataricus* 16S rRNA was estimated to contain five P residues based on mass spectrometry analysis (59).

The aCBF5/aNOP10/aGAR1 complex may act on rRNA without guide sRNA

Among the small *P. abyssi* rRNA substrates that we used to test the activity of reconstituted sRNPs (Table S2), only that containing residue U2603 was modified in the absence of H/ACA sRNA (Figure 7). Therefore, the aCBF5/aNOP10/aGAR1 complex may be a specific catalyst for this position in 23S rRNA. The stem-loop structure formed by the small RNA substrate used in the assay is not the one expected to be formed in the 50S subunit (Figures 7 and 8). However, we cannot exclude the possibility that it is formed at some stage during 23S rRNA synthesis or 50S subunit assembly. Although a stable stem-loop structure could also be formed by the small substrate containing residue U2585 (Figure 7), the aCBF5/aNOP10/aGAR1 complex did not modify it. No activity was detected with the recombinant Pus10 enzyme (data not shown). Two other RNA: P synthases are expected to be present in archaea, the E. coli TruA and TruD (Pus7) homologues (60).

The implication of Pus7 is unlikely, since residue 2585 is not located in a sequence that fits the consensus sequence recognized by this enzyme (RSU/N/PAR (R = purine, S = G/C, N = any nucleotide) (61 and our unpublished data). Therefore, it would be interesting to test the activity of the TruA homologue at position 2585 in *P. abyssi* 23S rRNA.

In agreement with the absence of *E. coli* RluE, RluB and RluC homologues in *P. abyssi*, the pseudouridylation at the three positions modified by these enzymes in *E. coli* are catalyzed by one H/ACA sRNA, Pab35.

**Structural determinants for H/ACA sRNA specificity**

Our data shed light on structural determinants of the sRNA-rRNA interaction that will be useful for the identification of new H/ACA sRNAs and their target sites. They can be explained taking structural data into account: (i) NMR analysis of the H/ACA snoRNA–rRNA interaction showed that the P1S and P2S intermolecular helices formed with the 3’- and 5’-guide sequences, respectively, can be coaxially stacked on helices H1 and H2 of the sRNA (32,33), and in this structure, the substrate is folded into a U-shape, with the targeted U residue protruding in the middle (32), (ii) based on the sRNP crystal structure, the H2-P2S pseudo-helix likely interacts with protein L7Ae bound to the K-turn or K-loop (22–24) and to a lesser extent with aNOP10 and aCBF5 through helix H2 (31), (iii) the various X-ray structures which have been established show that several interactions are formed between the aCBF5 and aNOP10 proteins (27–31) and the L7Ae and aNOP10 proteins (31) and (iv) finally, the H1-P1S pseudo-helix is expected to interact with protein aCBF5 bound to the ACA triplet (23,24,30,31). Therefore, it is highly conceivable that variations of the distances that separate the ACA box and the targeted uridine residue on the one hand (14 or 15-nt long in archaea, versus 16-nt long in eukarya) (62), and the K-turn and targeted U residue on the other hand (5 or 6 bps), disturb the positioning of aCBF5 relative to the targeted U residue and/or prevent some protein–protein interactions in the core protein structure (23,51). Importantly, we noticed that helix H2 never contains more than one bulged nucleotide.

Noticeably, for the first time, we show a greater importance of the rRNA interaction formed with the 3’-guide sequence (P1S), compared to that formed with the 5’-guide sequence (PS2). This difference can be explained by the close contact of aCBF5 with the 3’-guide sequence, compare to its limited contact with the 5’-guide sequence in the sRNP 3D structure (30). Remarkably, also the minimal length of the overall rRNA–sRNA base-pair interaction seems to be shorter in archaea compared to eukaryal system. This decreased stringency in archaea may explain why, in archaea but not in eukarya, a given H/ACA motif can guide P formation at up to three different positions in rRNAs.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.
ACKNOWLEDGEMENTS

S. Muller and J-B Fourmann were doctoral fellows supported by the French Ministère de la Recherche et des Nouvelles Technologies (MRNT). The work was supported by the Centre National de la Recherche Scientifique (CNRS), the MRNT and the Pôle de Recherche Scientifique et Technologique (PRST) « Bioingénierie » of Région Lorraine. Fujihiko Matsunaga and Patrick Forgette (UMR CNRS 8621, Orsay, France) are warmly thanked for their generous gift of P. abyssi cells. J. Ugolini is acknowledged for her efficient technical assistance. Funding to pay the Open Access publication charge was provided by CNRS.

Conflict of interest statement. None declared.

REFERENCES

Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Halococcus marismortui* and *Deinococcus radiodurans*. *RNA*, 11, 210–219.


