



# Recent Advances in Archaeal Translation Initiation

Emmanuelle Schmitt\*, Pierre-Damien Coureux, Ramy Kazan, Gabrielle Bourgeois, Christine Lazennec-Schurdevin and Yves Mechulam

Laboratoire de Biologie Structurale de la Cellule, BIOc, Ecole Polytechnique, CNRS-UMR7654, Institut Polytechnique de Paris, Palaiseau, France

## OPEN ACCESS

### Edited by:

Sébastien Ferreira-Cerca,  
University of Regensburg, Germany

### Reviewed by:

Paola Londei,  
Sapienza University of Rome, Italy  
Umesh Varshney,  
Indian Institute of Science (IISc), India

### \*Correspondence:

Emmanuelle Schmitt  
emmanuelle.schmitt@polytechnique.  
edu

### Specialty section:

This article was submitted to  
Biology of Archaea,  
a section of the journal  
Frontiers in Microbiology

**Received:** 16 July 2020

**Accepted:** 24 August 2020

**Published:** 18 September 2020

### Citation:

Schmitt E, Coureux P-D, Kazan R,  
Bourgeois G, Lazennec-Schurdevin C  
and Mechulam Y (2020) Recent  
Advances in Archaeal  
Translation Initiation.  
Front. Microbiol. 11:584152.  
doi: 10.3389/fmicb.2020.584152

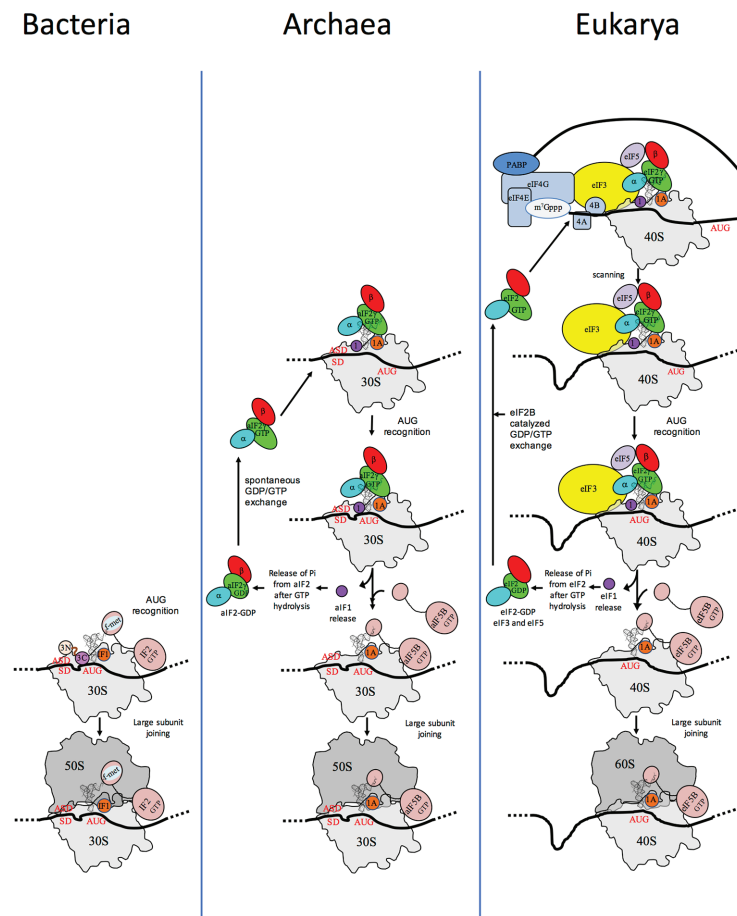
Translation initiation (TI) allows accurate selection of the initiation codon on a messenger RNA (mRNA) and defines the reading frame. In all domains of life, translation initiation generally occurs within a macromolecular complex made up of the small ribosomal subunit, the mRNA, a specialized methionylated initiator tRNA, and translation initiation factors (IFs). Once the start codon is selected at the P site of the ribosome and the large subunit is associated, the IFs are released and a ribosome competent for elongation is formed. However, even if the general principles are the same in the three domains of life, the molecular mechanisms are different in bacteria, eukaryotes, and archaea and may also vary depending on the mRNA. Because TI mechanisms have evolved lately, their studies bring important information about the evolutionary relationships between extant organisms. In this context, recent structural data on ribosomal complexes and genome-wide studies are particularly valuable. This review focuses on archaeal translation initiation highlighting its relationships with either the eukaryotic or the bacterial world. Eukaryotic features of the archaeal small ribosomal subunit are presented. Ribosome evolution and TI mechanisms diversity in archaeal branches are discussed. Next, the use of leaderless mRNAs and that of leadered mRNAs having Shine-Dalgarno sequences is analyzed. Finally, the current knowledge on TI mechanisms of SD-leadered and leaderless mRNAs is detailed.

**Keywords:** mRNA, Shine-Dalgarno, leaderless, ribosomal proteins, evolution

## INTRODUCTION

Translation initiation (TI) allows accurate selection of the initiation codon on a messenger RNA (mRNA), which then defines the reading frame of the protein to be synthesized. In all domains of life, translation initiation generally occurs within a macromolecular complex made up of the small ribosomal subunit (SSU), the mRNA, a specialized methionylated initiator tRNA, and translation initiation factors (IFs). Once the start codon is selected at the P site of the ribosome and the large subunit is associated, the IFs are released and a ribosome competent for elongation is formed. However, even if the general principles are the same in the three domains of life, the molecular mechanisms are different in bacteria, eukaryotes, and archaea and may also vary depending on the mRNA (**Figure 1**).

In bacteria, mRNAs are not further processed after transcription and the 5' untranslated region (5'-UTR) often carries a "Shine-Dalgarno" (SD) sequence containing a GGAGG consensus complementary to the 3' end of the 16S rRNA of the SSU (Shine and Dalgarno, 1974). They can also be devoid of SD sequence or even have no 5'-UTR at all. The methionylated initiator



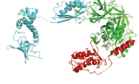
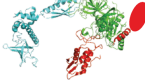









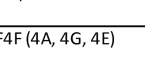
**FIGURE 1** | Schematic views of the translation initiation (TI) steps in the three domains of life. The figure illustrates the main steps in bacteria (left), in archaea (middle), and in eukarya (right). Bacterial 30S subunit recruits the messenger RNA (mRNA), often due to the base pairing between a Shine-Dalgarno sequence (SD) with an ASD sequence at the 3'-end of 16S rRNA. Three initiation factors, IF1, IF2, and IF3 favor the recruitment of the initiator tRNA and its pairing with the start codon. The formyl-methionyl moiety of the initiator tRNA is important for recognition by IF2. After start codon recognition, IF3 is released and the large ribosomal subunit is recruited with the help of IF2 (see Mechulam et al., 2011; Rodnina, 2018 for reviews). Archaea and eukarya share a common set of factors comprising e/alf1A, e/alf1, e/alf2, and e/alf5B (see also **Figure 2**). e/alf2 heterotrimer is represented with a three-color code ( $\alpha$  subunit in cyan,  $\beta$  subunit in red, and  $\gamma$  subunit in green). In canonical eukaryotic translation, a pre-initiation complex, containing the small ribosomal subunit, the methionylated initiator tRNA, and initiation factors, forms at the 5'-capped end of the mRNA. The complex then scans the mRNA until a start codon in a suitable environment is found. Base-pairing of the tRNA anticodon with the AUG start codon triggers eIF1 release followed by the release of Pi resulting from GTP hydrolysis by eIF2 (Algire et al., 2005). In turn, eIF2, eIF3, and eIF5 are released; eIF5B-GTP is recruited and favors joining with the large ribosomal subunit (see Hinnebusch, 2017 for a review). Archaea often use an SD sequence for mRNA recruitment. The 30S subunit is then definitely positioned with the start codon in the P site thanks to base-pairing with the tRNA anticodon. Overall, the four initiation factors alf1, alf1A, alf2, and alf5B play similar roles as their eukaryotic counterparts (see text and Schmitt et al., 2019 for a mechanism-oriented review). In the three cases, the translation competent IC is formed after the release of e/alf1A (or IF1 in bacteria) and e/alf5B (or IF2 in bacteria). In eukarya, the complex formed by eIF4E + eIF4G + eIF4A is known as eIF4F. eIF3, composed of 6 (yeast) to 13 (mammals) subunits is represented as a yellow oval. The figure is adapted from Schmitt et al. (2019).

tRNA is formylated and the formyl group is crucial for its accurate selection by the initiation complex (Guillon et al., 1993, 1996). Only three initiation factors are involved, IF1, IF2, and IF3 (for reviews see, for example, Mechulam et al., 2011; Rodnina, 2018).

In eukaryotes, translation initiation is more complicated with many IFs involved (**Figure 2**). mRNAs are matured with a m<sup>7</sup>G-cap at the 5' end and a polyadenylated tail at the 3' end.

**Abbreviations:** TI, Translation initiation; SSU, Small ribosomal subunit; LSU, Large ribosomal subunit; IC, Initiation complex; TC, Ternary complex e/alf2:GTP:Met-tRNA<sup>Met</sup>; LUCA, Last universal common ancestor.

The cap-dependent canonical translation initiation model involves a pre-initiation complex (43S PIC) containing the SSU, the ternary complex eIF2-GTP-Met-tRNA<sup>Met</sup>, the two small factors, eIF1 and eIF1A, and two proteins with regulatory functions, eIF5 and eIF3. eIF5 is the activating protein for the eIF2 GTPase, and eIF3 is a large multimeric architectural protein involved in mRNA binding. In the presence of factors of the eIF4 family and of the poly(A)-binding protein (PABP) associated with the poly(A) tail of the mRNA, the 43S PIC is recruited at the 5'-cap extremity of the mRNA, thereby forming the 48S PIC. In mammals, direct interaction of eIF3 with eIF4F favoring

Bacteria	Archaea	Eukarya	Function
-	aIF2 ( $\alpha, \beta, \gamma$ ) 	eIF2 ( $\alpha, \beta, \gamma$ ) 	Binds tRNA <sub>i</sub> , GTP hydrolysis and releases Pi coupled to start codon recognition.
$\sim$ IF3*	aIF1 	eIF1 	Participates in accuracy of start codon selection.
IF1 	aIF1A 	eIF1A 	Participates in accuracy of start codon selection.
IF2 	aIF5B 	eIF5B 	Binds tRNA <sub>i</sub> , facilitates joining of the large ribosomal subunit
-	-	eIF5 	Accelerates hydrolysis of GTP by eIF2 (GTPase activating protein, GAP).
-	aIF2B ( $\alpha, \beta, \delta$ )*	eIF2B ( $\alpha, \beta, \gamma, \delta, \epsilon$ ) 	eIF2B catalyzes GDP/GTP exchange on eIF2 (guanine nucleotide exchange factor, GEF).
-	-	eIF3 (6 to 13 subunits)	Participates in the formation of the pre-initiation complex.
-	aIF4A**	eIF4F (4A, 4G, 4E)	eIF4F tethers the pre-initiation complex to the 5' cap of the mRNA before scanning.
-	-	eIF4B	Assists eIF4F.

**FIGURE 2 |** Translation initiation factors in the three domains of life. The structures of the archaeal translation initiation factors and of their orthologues in eukaryotes and bacteria (when present) are shown. e/aIF2 is colored as in **Figure 1**. The unknown structure of the N-domain specific of eukaryotic eIF2 $\beta$  is shown as an oval. The structure of aIF2 is from PDB 3V11 (Schmitt et al., 2012), those of aIF1 and aIF1A are from Coureux et al. (2016). The structures of eIF2, eIF1, and eIF1A are from PDB 6FYX (Llacer et al., 2018). IF1 is from PDB 3I4O (Hatzopoulos and Mueller-Dieckmann, 2010). Bacterial IF3 is a two-domain protein. The correspondence between IF3 and e/aIF1 is based on a structural and functional resemblance of the IF3 C-terminal domain with e/aIF1. Despite this resemblance, the topologies of the two  $\alpha$ - $\beta$  folds are different. This suggests that they do not derive from a common ancestor. aIF5B is from PDB 1G7T (Roll-Mecak et al., 2000), eIF5B is from PDB 4N3N (Kuhle and Figner, 2014), and IF2 is from PDB 5LMV (Hussain et al., 2016). The color code for e/aIF5B/IF2 is as follows: G-domain and domain II in green, domain III in light orange, linker in yellow, and domain IV in red. The specific archaeal helix in domain IV is shown in blue. The catalytic  $\gamma$  and  $\epsilon$  subunits of eIF2B are missing in archaea. The function of the eIF2B  $\alpha, \beta, \delta$  homologues in archaea is not clear and may be unrelated to translation initiation (Dev et al., 2009; Gogoi et al., 2016). \*\*The aIF4A orthologue is present in many archaea. However, deletion of the corresponding gene in *Haloferax volcanii* showed only a small phenotype (Gäbel et al., 2013).

the formation of the 48S complex was shown (Korneeva et al., 2000). However, these interactions were not detected in *Saccharomyces cerevisiae* (Jivotovskaya et al., 2006). The 48S PIC then scans the mRNA until an AUG codon in a correct context (Kozak motif) is found (Kozak, 1986). Recognition of the AUG codon stops scanning, causes the release of factors and the assembly of an 80S complex competent for elongation *via* the junction with the large subunit, using eIF5B and eIF1A (for a review see, for example, Hinnebusch, 2017). Besides this canonical mechanism, a certain number of alternative starting routes have also been described (Shatsky et al., 2018).

Archaeal TI harbors bacterial and eukaryotic features. In archaea, mRNAs are not further processed after transcription. They have Shine-Dalgarno sequences or very short 5'-UTR. Hence, archaeal mRNA features are close to bacterial mRNA features. In contrast, genomic analyses showed that archaeal initiation factors correspond to a subset of eukaryotic translation initiation factors. Indeed, aIF1, aIF1A, aIF2, and aIF5B homologous to the corresponding eukaryotic factors are present (**Figures 1, 2;**

Dennis, 1997; Kyrpides and Woese, 1998; Benelli and Londei, 2011; Gäbel et al., 2013; Schmitt et al., 2019). Thus, even if there are obvious differences between archaea and eukaryotes, in particular for the recruitment of mRNA, *via* SD sequences vs. long-range scanning, the selection of the start codon is carried out within a same structural core composed of the small ribosomal subunit, mRNA, methionylated initiator tRNA (Met-tRNA<sub>i</sub><sup>Met</sup>), and the three initiation factors e/aIF1, e/aIF1A, and e/aIF2 (Schmitt et al., 2019). Finally, the late steps of TI preceding the formation of a ribosome competent for elongation are controlled by initiation factors that are conserved in the three domains of life, IF1-e/aIF1A, and IF2-e/aIF5B.

Because TI mechanisms have evolved lately, their studies bring important information about the evolutionary relationships between extant organisms. In this context, recent structural data on ribosomal complexes and genome-wide studies are particularly valuable.

This review will focus on archaeal translation initiation highlighting its relationships with either the eukaryotic or the

bacterial world. We first describe eukaryotic features of the archaeal small ribosomal subunit possibly related to TI mechanisms and discuss the diversity of the archaeal ribosome among archaeal phyla. Next, we discuss the occurrence of leaderless mRNAs and that of leadered mRNAs having Shine-Dalgarno sequences. The current knowledge on TI mechanisms of SD-leadered and leaderless mRNAs is then presented.

## THE ARCHAEOAL RIBOSOME IS OF THE EUKARYOTIC TYPE

In the 1980s, Woese noted that rRNAs were excellent molecular chronometers that could be used to trace the molecular phylogenetic relationships between extant individuals. Indeed, rRNA are found in all organisms, are easily isolated and sequenced, and show positions that vary at different rates. Analysis of sequence/secondary structure variations in rRNAs allowed definition of Archaea as a third branch of the tree of life (Woese and Fox, 1977; Noller and Woese, 1981; Woese et al., 1983, 1990; Woese, 1987). Since these pioneering studies, many other works were dedicated to evolution of the ribosome (Cannone et al., 2002; Roberts et al., 2008; Fox, 2010; Petrov et al., 2015). Thanks to the increasing number of sequences and to the availability of high-resolution three-dimensional structures of ribosomes representative of each domain of life, evolutionary relationships between organisms appeared even clearer. A universal core making the structural and functional foundation of rRNAs of all cytoplasmic ribosomes was defined (Bernier et al., 2018). At the level of the SSU, this common core corresponds to 90% of bacterial rRNA and encompasses the decoding center of the small ribosomal subunit with in particular the 530 loop and the 1,490 region (*Escherichia coli* numbering) but not the 3' end corresponding to the mRNA exit region. Archaeal ribosomes have rRNA molecules very close in size to that of bacterial rRNAs explaining why the sedimentation coefficients of the archaeal ribosomal subunits are the same as that of bacterial ribosomes (Table 1). However, some regions are divergent (some of the divergent regions of the SSU colored in red in Figure 3), and rRNAs of archaeal ribosomes are closer to eukaryotic rRNAs than to bacterial rRNAs (Woese, 1987; Roberts et al., 2008; Bernier et al., 2018; Bowman et al., 2020; Penev et al., 2020).

Phylogenetic studies were further refined using ribosomal protein sequences. The archaeal ribosome contains ribosomal proteins (r-proteins) that are either universal (33 r-proteins) or specific to eukarya and archaea (29 r-proteins; Table 1). No r-proteins found only in the archaeal and bacterial domains are found. One protein that could be specific of the archaeal domain found in place of eukaryotic eS21 and, therefore, named aS21 was identified recently in the SSU of *Pyrococcus abyssi* (Coureux et al., 2020) and *Thermococcus celer* (Nurenberg-Goloub et al., 2020). However, further phylogenetic studies are required to firmly determine whether the protein is unique to archaea or distantly related to eS21. The 2014 system for naming ribosomal proteins is used throughout the manuscript. According to this naming, eukaryotic and archaeal specific proteins are

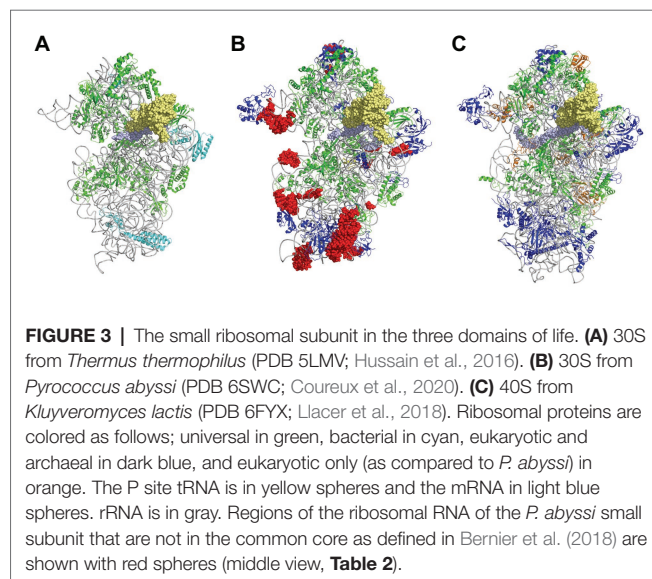
named eSX or eLX (Ban et al., 2014). Unfortunately, this naming does not directly distinguish r-proteins that are either present in eukaryotes and archaea from those present only in eukaryotes. Given the growing importance of studies of the archaeal ribosome, a naming including an ae prefix for specifying archaeal and eukaryotic proteins would now be desirable.

Structurally invariable cores are found in universal proteins. However, in addition to the core, protein segments or extensions show late evolution reflecting specialization in the three domains of life (Melnikov et al., 2018). Concerning the 29 r-proteins specific to eukarya and archaea, it is interesting to note that some of these proteins contact regions of the 16S rRNA outside of the common core (red patches in Figure 3 and Table 2).

**TABLE 1** | Ribosomes in the three domains of life.

Domain	Sedimentation coefficient	rRNA	Ribosomal proteins
Bacteria	70S	30S	16S (1493)
		50S	23S (2891) 5S (117)
Archaea	70S	30S	16S (1483)
		50S	23S (2967) 5S(122)
Eukarya	80S	40S	18S (1860)
		60S	28S (4039)
			5S (120)
			5.8S (158; <i>S. cerevisiae</i> )

Data concerning rRNA are from Bernier et al. (2018). Values in parentheses indicate the mean number of bases according to the SEREB (Sparse and Efficient Representation of Extant Biology) sampling except for the 5.8S rRNA. The number of species in the SEREB sample is 67 bacteria, 36 archaea, and 30 eukarya. Sequence alignments are accessible in Bernier et al. (2018). The ribosomal protein contents are indicated for *E. coli*, *S. cerevisiae*, and *P. abyssi*. These numbers slightly vary depending on the organism (Lecompte et al., 2002). The number of universal (u), eukaryotic (e), bacterial (b), archaeal (a) -type ribosomal protein is indicated. One protein possibly specific to archaea (a) has recently been identified in *P. abyssi* (Coureux et al., 2020; Nurenberg-Goloub et al., 2020).



**TABLE 2** | Regions of Pab-16S rRNA that are not in the common core and the r-proteins found nearby.

Regions of Pab-16S rRNA not in the common core	Archaeal and eukaryotic-specific r-proteins found nearby
(1405–1,437) h44	S8e; S6e
(176–178,207;209) h9	S8e
(181–205) h9	S4e; S8e
(214–227) h10	S4e
(420–428) h16	none
(440–456) h17	S24e
(561;614–615) h21	none
(841–842)	none
(963–975; 999–1,006) h33	S12e
(1102–1,104;1,109–1,111) h39; (1260) h41	S19e
(1141) h40	S17e
(1504–1,509) 3' extremity	mRNA exit tunnel

The regions indicated in the Table were from Bernier et al. (2018) and Woese (1987). R-proteins are named according to Ban et al. (2014). Numbering refers to the *P. abyssi* 16S rRNA sequence (GenBank AJ248283.1, [www.ma.icmb.utexas.edu](http://www.ma.icmb.utexas.edu)) also used in PDB 6SWC.

Phylogenetic analysis of archaeal/eukaryotic specific r-proteins show that content in r-proteins vary depending on the archaeal branch (Lecompte et al., 2002; Hartman et al., 2006; Yutin et al., 2012). In particular, Crenarchaeota, Thaumarchaeota, and Korarchaeota have more r-proteins than Euryarchaeota and Nanoarchaeota. More recently, an Asgard superphylum close to the TACK one has been identified (Eme et al., 2017). Examination of the first Asgard genomes (Akil and Robinson, 2018; Imachi et al., 2020) also revealed a higher content of r-proteins as observed in TACK. For instance, TACK and Asgard SSU contain S25e, S26e, and S30e, whereas these proteins are not found in euryarchaea. Altogether, these findings agree with the proposed origin of eukaryotes from within an archaeal superphylum close to Asgard and TACK (Hartman et al., 2006; Guy and Ettema, 2011; Yutin et al., 2012; Williams et al., 2013; Eme et al., 2017; Zaremba-Niedzwiedzka et al., 2017; Castelle and Banfield, 2018; Melnikov et al., 2020). Notably, a recent study in Lokiarchaea and Heimdallarchaea further decreased the gap between eukaryotes and archaea by identifying eukaryotic-like expansion segments in large subunit rRNA in these archaea (Penev et al., 2020). As discussed below, the availability of high-resolution structures of functional states of ribosomes in the three domains of life now provide data for functional and structural comparisons leading to validation of sequence-based models.

## ARCHAEAL mRNAs

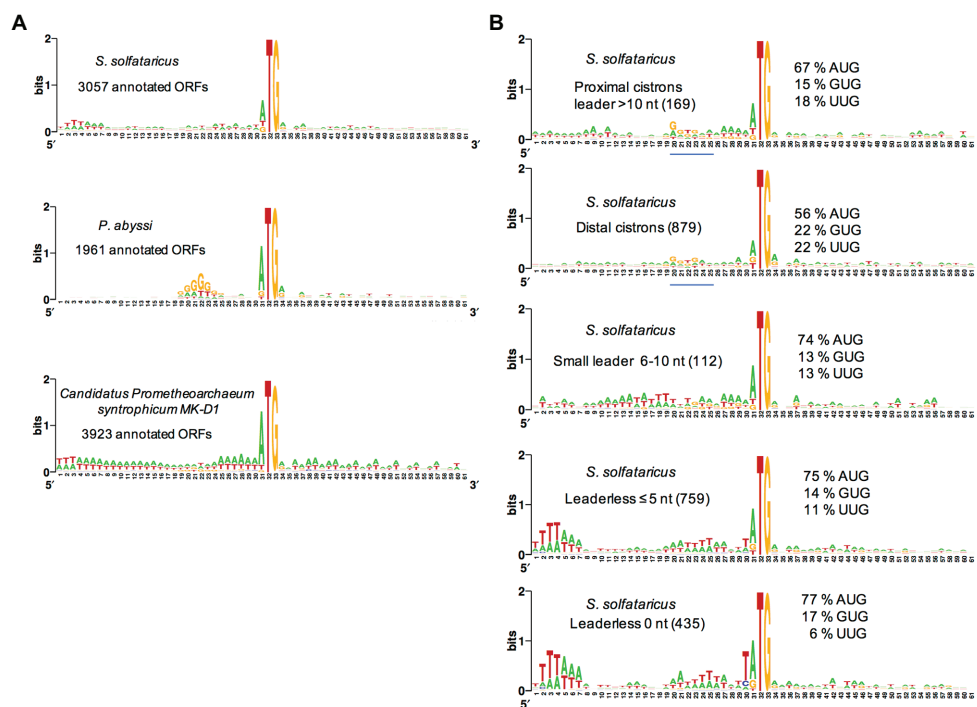
Organization of the archaeal mRNAs is of the bacterial type with many polycistronic genes organized into operons. mRNAs do not have a cap at the 5' end nor a 3' polyadenylated tail. Cryo-EM experiments performed on lysed *Thermococcus kodakaraensis* cells made it possible to observe that most of the polysomes were connected to strands of DNA, thus showing that the mRNA could begin to be translated before its synthesis is complete (French et al., 2007). Hence, from a functional

point of view, the prokaryotes archaea and bacteria differ from eukaryotes by the fact that, in the absence of nucleus, transcription and translation take place in the same compartment and that the two processes can, therefore, be coupled (Martin and Koonin, 2006).

Depending on the organism, archaeal mRNAs mainly have Shine-Dalgarno sequences or are mainly leaderless (Dennis, 1997; Ma et al., 2002; Benelli and Londei, 2011). mRNAs are generally considered leaderless if the number of nucleotides preceding the start codon is less or equal to 5 (Babski et al., 2016). Some authors have, however, chosen eight as the threshold, arguing that this is likely the minimal size of an UTR to allow efficient SD-aSD pairing (Jäger et al., 2014). The differences in 5'UTR of mRNAs reflect some diversity in translation initiation mechanisms (Tolstrup et al., 2000; Slupska et al., 2001; Torarinnsson et al., 2005; Brenneis et al., 2007; La Teana et al., 2013; Kramer et al., 2014; Schmitt et al., 2019). Recent genome-wide studies, most of them based on differential RNA-seq methods, highlighted mRNA organization in various archaeal branches (Jäger et al., 2009, 2014; Wurtzel et al., 2010; Toffano-Nioche et al., 2013; Li et al., 2015; Babski et al., 2016; Cho et al., 2017; Smollett et al., 2017; Grünberger et al., 2019; Gelsinger et al., 2020). Identification of transcription start points is particularly important in Archaea, where most gene annotations are generated from general computational pipelines that are not fully reliable. Hence, genome-wide transcriptomic studies have made it possible to correct automatic annotation of genomes and some theoretical models directly derived from these annotations.

Most euryarchaeal species studied to date mainly harbor Shine-Dalgarno sequences complementary to the 3' end of their 16S rRNA. Hence, in *Thermococcus onnurineus* (Cho et al., 2017), *Thermococcus kodakarensis* (Jäger et al., 2014), *Methanobrevibacter smithii* (Li et al., 2015), *Methanosarcina mazei* (Jäger et al., 2009), *P. abyssi* (Toffano-Nioche et al., 2013), and *Pyrococcus furiosus* (Grünberger et al., 2019), the abundance of leaderless mRNA is around 15% only. This is in contrast with the high percentage of leaderless mRNA observed in *Saccharolobus solfataricus* (69%; Wurtzel et al., 2010) and *Pyrobaculum aerophilum* (Slupska et al., 2001; Ma et al., 2002), both being crenarchaeota, and the euryarchaeota *Haloferax volcanii* (72%; Babski et al., 2016; Gelsinger et al., 2020). Interestingly, a quick analysis of the annotated translation initiation regions in the available Lokiarchaeote genome (Imachi et al., 2020) suggests that SD sequences are not prevalent (Figure 4A).

The availability of a genome-wide transcriptome from *H. volcanii* gave the opportunity to search for features of the many leaderless transcripts ( $\leq 5$  nt; Babski et al., 2016). First, it was noted that highly transcribed genes typically give leaderless mRNAs. Second, in leaderless mRNAs from abundantly transcribed genes, the AUG start codon was somewhat preferred over GUG. Third, the prevalence of A/G at the first position of the second codon was higher in leadered transcripts than in leaderless ones. However, this analysis did not highlight specific features of leaderless transcripts that may give clues on how they are recognized by the TI machinery. A genome-wide transcriptome is also available for *S. solfataricus* (Wurtzel et al., 2010). Again, comparative sequence logos from transcripts sorted by size of



**FIGURE 4 |** Analysis of translation start regions. **(A)** Analysis of the translation start regions in *Sulfolobus solfataricus*, *P. abyssi*, and the Asgard *Candidatus Prometheoarchaeum syntrophicum*. DNA sequences (60 nt around the first base of the start codon) were extracted from the genomic sequences (She et al., 2001; Cohen et al., 2003; Imachi et al., 2020). Annotations as corrected by Wurtzel et al. (2010) have been used for *S. solfataricus*. Sequence logos were created using Weblogo (Crooks et al., 2004). **(B)** Detailed analysis of the translation start regions in *S. solfataricus*. See also Tolstrup et al. (2000) for an earlier analysis. For each indicated category of transcript (number of ORFs in parentheses), the percentage of AUG, GUG, and UUG start codons are indicated. The position of potential 16S rRNA binding sites (Shine-Dalgarno sequences) in the upper two logos is shown by a blue line. Note that in fully leaderless genes (0 nt), the occurrence of T at the –1 position and the avoidance of UUG as start codon are likely linked to signals for RNA polymerase.

their leaders do not highlight obvious features of leaderless transcripts (Figure 4, see also Tolstrup et al., 2000). In contrast, ORFs that harbor a leader greater than 10 nucleotides, including distal cistrons in operons, show a G/T rich region 10 nucleotides upstream from the start codon, reflecting the occurrence of SD sequences in many cases (Figure 4). This agrees with a recent bioinformatics analysis (Huber et al., 2019), showing that many distal cistrons in overlapping gene pairs carry an SD sequence. Furthermore, the SD motif was found essential for translation of a distal cistron in *S. solfataricus* (Condo et al., 1999). Finally, it is interesting to note that at least in *S. solfataricus*, genes involved in protein translation are over-represented among leadered transcripts (Wurtzel et al., 2010).

The available data suggest that leaderless mRNAs and leadered, SD containing, mRNAs co-exist in almost all archaea, including those for which leaderless mRNAs are prevalent (Tolstrup et al., 2000; Ma et al., 2002; Karlin et al., 2005; Wurtzel et al., 2010; Huber et al., 2019). Moreover, AUG, GUG, and UUG can serve as start codons in all types of mRNAs, whatever the size of the leader. Thus, it is likely that most, if not all, archaea have a TI machinery capable of translating both leaderless and leadered mRNAs (Benelli et al., 2003). This raises the question of the mechanism of mRNAs recruitment by the ribosome.

## SD-MEDIATED mRNA RECRUITMENT

Sequence analyses show that the 3' extremity of archaeal 16S rRNAs is highly conserved and corresponds to a  $5'$ AUCACCUCCU $3'$  consensus (note that crenarchaeota often lack the last CU nucleotides<sup>1</sup>). This sequence is complementary to the SD motif comprising GGAGG. By analogy with bacteria, it can be proposed that in Archaea formation of the SD:antiSD duplex facilitates the recruitment of the small ribosomal subunit. Then, the assembly of the archaeal initiation complex (IC) containing the ternary complex  $\alpha$ IF2:GTP:Met-tRNA<sup>Met</sup> and the two small initiation factors  $\alpha$ IF1 and  $\alpha$ IF1A is favored. Within this complex, the three initiation factors ensure accurate selection of the start codon (Pedulla et al., 2005; Hasenöhrl et al., 2006, 2009; Gäbel et al., 2013; Coureux et al., 2016, 2020; Monestier et al., 2018).

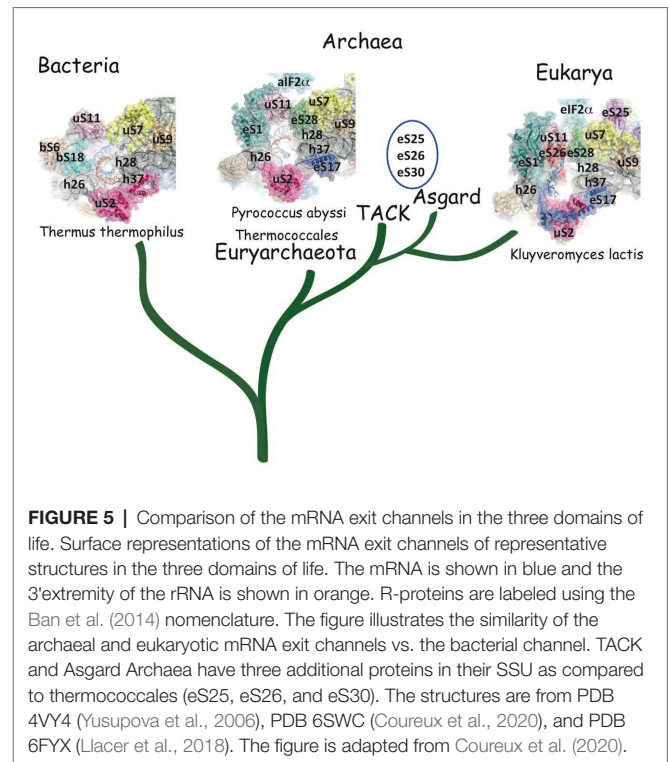
The role of the SD sequences in translation was experimentally studied in only few archaeal species. Using a cell-free system, the SD motifs were shown essential for translation of a bicistronic mRNA from *S. solfataricus* (Condo et al., 1999). In recent developments of this *S. solfataricus* cell-free system, translation is stimulated by a strong SD motif placed ahead of the start codon of the chosen gene (Lo Gullo et al., 2019).

<sup>1</sup><http://www.rna.icmb.utexas.edu/>

The GGAGGUCA SD motif of the *gvpH* gene from *Halobacterium salinarium*, involved in the gas vesicle formation, was shown to enhance translation efficiency using an *in vivo* assay in the related halophilic archaeon *H. volcanii* (Sartorius-Neef and Pfeifer, 2004). However, in *H. volcanii* (72% leaderless mRNAs), the SD motif was shown to be non-functional in translation initiation of the monocistronic *sod* mRNA (Kramer et al., 2014). Moreover, in this halophilic archaea, translational coupling was demonstrated for overlapping gene pairs. In this case, the SD motif in the distal cistrons appeared more important for reinitiation than for *de novo* initiation (Huber et al., 2019). This raises the possibility that distal cistrons in overlapping gene pairs are translated by a mechanism where 70S ribosomes terminate and then reinitiate without dissociation. Alternatively, the terminating ribosome may dissociate but the SSU would remain bound to the mRNA thanks to the SD sequence (Huber et al., 2019). It should, however, be noted that translation of an SD-leadered distal cistron in a *S. solfataricus* cell-free system was found to be independent of the translation of the first cistron (Condo et al., 1999). Notably, *S. solfataricus* and *H. volcanii* both are organisms that widely use leaderless mRNAs. It cannot be excluded that in these organisms, TI mechanism evolved in such a way that interaction of the SD motif with the antiSD sequence of the 16S rRNA became less important for the stability of the TI complex. Unfortunately, to our knowledge, the role of the SD sequence in TI efficiency was not studied *in vivo* in euryarchaea, where a SD sequence is found in the major parts of the genes. Nevertheless, recent structural studies showing the formation of an SD:antiSD duplex in the mRNA exit chamber of the SSU of *P. abyssi* (71% SD-leadered genes; Ma et al., 2002; Toffano-Nioche et al., 2013) strongly suggest that as in bacteria, the SD motif stabilizes the TI complex.

## THE SD DUPLEX IS BOUND IN AN mRNA EXIT CHANNEL THAT DIFFERS FROM THAT OF BACTERIA

The cryo-EM structure of a TI complex from *P. abyssi* (Pa) using an mRNA derived from that of the gene coding for Pa-aEF1A containing a strong SD sequence [A(-17)UUUGGAGGUGAUUAAA(+1)UGCCAAAG(+9)] is known at 3.4 Å resolution (Coureux et al., 2020). In the mRNA exit chamber, the SD duplex is extended to nine nucleotides and involves the <sup>5</sup>AUCACCUCC<sup>3</sup> sequence of the 3'-end of the 16S rRNA. The SD helix is positioned in the mRNA exit chamber delineated by uS11, eS1, and h26 on the one side and by uS7, eS28, h28, and h37, on the other side (Figure 5). Interactions of uS11 with eS28 and uS7 connect the platform to the head and form the SD duplex channel. uS2 and eS17 are located at the end of the mRNA exit chamber. The archaeal mRNA exit chamber was compared to the bacterial one. As shown in Figure 5, bS6 and bS18 are found in place of eS1 in the bacterial ribosome. In bacteria, eS28 is absent and uS2 possesses a supplementary inserted helical domain occupying the position of eS17. Interestingly, comparison of the bacterial



**FIGURE 5** | Comparison of the mRNA exit channels in the three domains of life. Surface representations of the mRNA exit channels of representative structures in the three domains of life. The mRNA is shown in blue and the 3' extremity of the rRNA is shown in orange. R-proteins are labeled using the Ban et al. (2014) nomenclature. The figure illustrates the similarity of the archaeal and eukaryotic mRNA exit channels vs. the bacterial channel. TACK and Asgard Archaea have three additional proteins in their SSU as compared to thermococcales (eS25, eS26, and eS30). The structures are from PDB 4VY4 (Yusupova et al., 2006), PDB 6SWC (Coureux et al., 2020), and PDB 6FYX (Llacer et al., 2018). The figure is adapted from Coureux et al. (2020).

structures with the archaeal one showed that the spacing between the AUG codon and the SD sequence changed the position of the duplex in the chamber, probably explaining how it influences translation initiation efficiency (Coureux et al., 2020). Overall, archaeal and bacterial exit channels appear as two structural solutions for binding the SD duplex. These two solutions reflect an early divergence of the ribosomes from these two domains (Figure 5).

The mRNA exit tunnel of the euryarchaeota *P. abyssi* is of the eukaryotic type (Figure 5). Notable differences are, however, observed. First, in the yeast ribosome, eS17 has a long C-terminal extension contacting the mRNA and second, eS26 stabilizes the 5' end of the mRNA (Llacer et al., 2018; Simonetti et al., 2020). Interestingly, eS26 was proposed to be involved in recognition of Kozak sequence elements (Ferretti et al., 2017). Importantly, in eukaryotes, initiation factors were shown to be involved in stabilization of the mRNA in the exit channel. Indeed, Kozak consensus nucleotides are recognized in the E site by domain 1 of the  $\alpha$  subunit of eIF2. Such an interaction was not observed with the homologous protein aIF2 $\alpha$  in the *P. abyssi* complex (Coureux et al., 2020). In addition, the eukaryotic-specific eIF3a subunit would also stabilize the mRNA at the exit channel pore (Llacer et al., 2018). These differences illustrate how eukaryotic and thermococcal ribosomes evolved the mRNA binding modes in the exit pocket, in relation with the canonical eukaryotic scanning mode vs. the SD-assisted AUG recognition mode occurring in many genes in the archaeal domain. In this view, it is notable that eS26 is absent in euryarchaeotes but present in TACK/Asgard genomes (Lecompte et al., 2002; Schutz et al., 2018). Because the archaeal version of the exit chamber is a simplified version of the eukaryotic one, this

argues in favor of the controversial hypothesis that eukaryotic ribosomes have evolved from within the archaeal version (Zaremba-Niedzwiedzka et al., 2017; Eme and Ettema, 2018).

## START CODON SELECTION MECHANISM

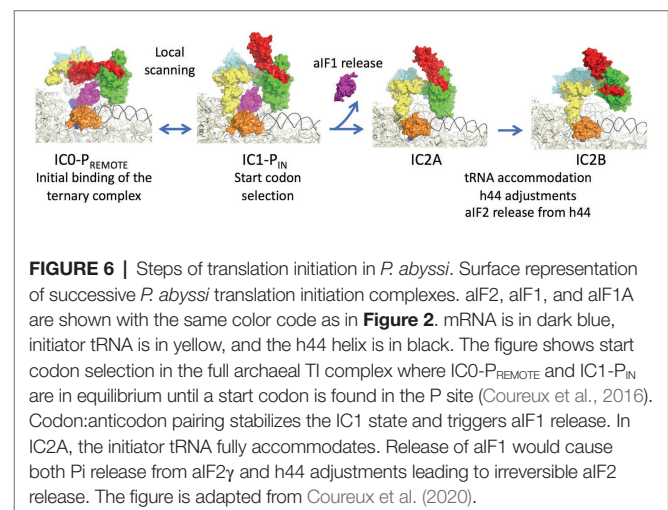
Three archaeal initiation factors, aIF1, aIF1A, and aIF2, participate in start codon selection on the SSU. The biochemical and structural data concerning these factors have been recently reviewed (Schmitt et al., 2019). Briefly, aIF1 is a small protein of ca. 100 residues that binds to the 30S in front of the P site (Coureux et al., 2016). Biochemical data using *P. abyssi* and *S. solfataricus* aIF1 have shown that the factor favored mRNA binding and formation of the initiation complex (Hasenöhrl et al., 2006, 2009; Monestier et al., 2018). The factor was also shown to induce a dynamic character of the IC favoring proofreading of erroneous initiation complexes (Hasenöhrl et al., 2009; Monestier et al., 2018). The role of aIF1 in translation fidelity is consistent with that observed for eIF1 in eukaryotes (Algire et al., 2005; Nanda et al., 2009). This function is also reminiscent of that of the bacterial translation initiation factor IF3 whose C-terminal domain has some structural resemblance with e/aIF1 (Figure 2; Rodnina, 2018). Notably, IF3 C-terminal domain alone is sufficient to sustain the growth of an IF3-deficient *E. coli* strain (Ayyub et al., 2017).

aIF1A is a small protein of ca. 100 residues that contains an OB-fold. Like its eukaryotic homologue, the factor occupies the A site on the SSU. Importantly, the eukaryotic version of the factor contains N and C-terminal extensions necessary for the scanning of the PIC along the mRNA (Figure 2; Pestova et al., 1998). aIF2 is a heterotrimeric protein that binds Met-tRNA<sup>Met</sup> in a GTP dependent manner (Yatime et al., 2004, 2006; Pedulla et al., 2005; Sokabe et al., 2006; Stolboushkina et al., 2008).  $\gamma$  is the core subunit that binds GTP (Schmitt et al., 2002; Dubiez et al., 2015).  $\alpha$  and  $\beta$  are bound to  $\gamma$  but do not interact together (Yatime et al., 2004; Schmitt et al., 2010). A 5 Å crystal structure of the TC (aIF2:GDPNP:Met-tRNA) showed that the initiator tRNA is bound to aIF2 *via* the C-terminal domain of  $\alpha$  and the domains I and II of  $\gamma$ , while the aIF2 $\beta$  subunit did not strongly contribute to the tRNA binding (Schmitt et al., 2012). A cryo-EM study of an archaeal initiation complex from *P. abyssi* containing the three initiation factors showed two conformations (Coureux et al., 2016). Analysis of these two conformations led to a model of start codon selection. In the major conformation, called IC0-P<sub>REMOTE</sub>, the anticodon stem-loop of the Met-tRNA<sup>Met</sup> is out of the P site. aIF2 $\gamma$  is bound to helix h44 and interacts with aIF1. The N-terminal domain of aIF1 would contact the two switch regions that control the nucleotide state of aIF2 $\gamma$ . In the second conformation, called IC1-P<sub>IN</sub>, the anticodon stem-loop of the Met-tRNA<sup>Met</sup> is bound to the P site, while the position of aIF2 $\gamma$  on h44 has not changed. aIF1A is still bound within the A site and aIF1 still located in front of the P site. The IC0-P<sub>REMOTE</sub> and IC1-P<sub>IN</sub> positions are in equilibrium and the transition from one position to the other, accompanied by a 30S head motion, reflects the dynamics of

the PIC during start codon selection in the P site (Coureux et al., 2016; Monestier et al., 2018). As observed for eukaryotic PIC (Lomakin et al., 2003; Weisser et al., 2013; Llacer et al., 2015), definitive stabilization of the tRNA in the P site is impaired by aIF1. This is consistent with the function of aIF1 in destabilization of erroneous TI complexes (Hasenöhrl et al., 2009; Hussain et al., 2014; Llacer et al., 2015; Monestier et al., 2018). Moreover, interaction of aIF2 with h44 of the 30S also counteracts final accommodation of the tRNA in the P site unless the start codon is base-paired with the tRNA anticodon. Indeed, codon-anticodon pairing compensates for the restoring force exerted on the tRNA by aIF2 because of its interaction with h44. Such a compensation would allow a longer stay of the initiator tRNA in the P site and trigger further events, including aIF1 departure because of steric hindrance with the tRNA, and release of aIF2 in its GDP bound form (Figure 6).

In order to better understand the role of aIF1 in the mechanism, an IC2 complex made in its absence was studied by cryo-EM at 3.4 Å resolution. Consistent with the above ideas, the IC2 complex shows stable accommodation of the initiator tRNA in the P site (Monestier et al., 2018; Coureux et al., 2020). Comparison of all states (Figure 6) suggests that a first set of conformational adjustments of h44 accompanies aIF1 departure causing in turn the release of its contacts with aIF2. These events would lead to the release of aIF2-GDP. Re-adjustments of the position of h44 in the bulge region could explain how the contacts between h44 and aIF2 $\gamma$  are lost. Both these h44 movements and the release of contacts between aIF1 and aIF2 $\gamma$  could explain how aIF2 is detached from the ribosome after start codon recognition and aIF1 release. In eukaryotes, it was shown that full release of eIF2 is linked to the release of Pi coming from GTP hydrolysis on eIF2 (Algire et al., 2005). In archaea, the breakdown of contacts between the N-terminal domain of aIF1 and the switch regions of aIF2 $\gamma$  could explain the coupling between aIF1 release and Pi release.

Overall, the available data suggest similarity in the mechanisms involved in start codon selection by e/aIF1, e/aIF1A, and e/aIF2 on the SSU in eukaryotes and archaea.





### THREE UNIVERSAL PROTEINS PARTICIPATE IN THE SELECTION OF THE INITIATOR tRNA AT THE P SITE

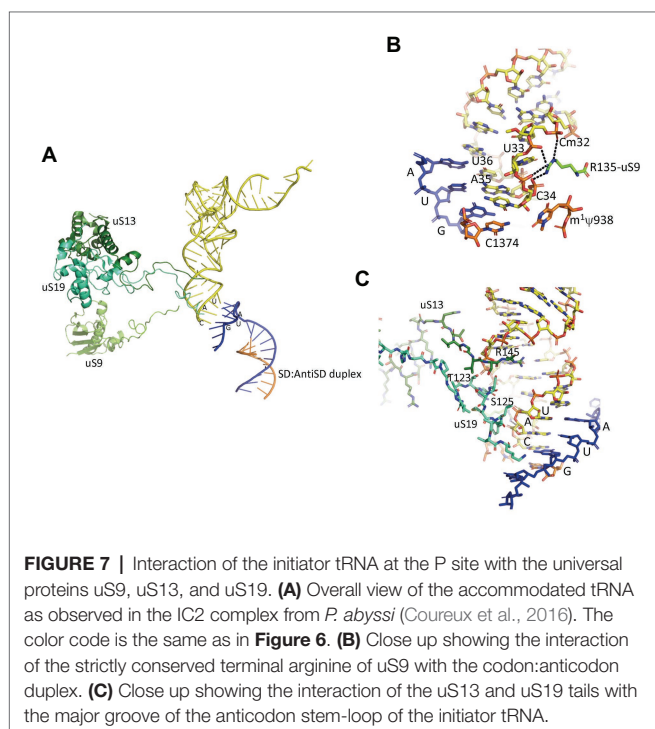
In the IC2 complex, the anticodon stem of the tRNA is tightly bound to the P site. A network of interactions involving the C-terminal tails of the three universal proteins uS9, uS13, and uS19 is observed (Figure 7). The C-terminal arginine R135 of uS9 is hydrogen bonded to the phosphate groups of Cm32, U33, A35, of the initiator tRNA. This interaction is also observed in eukaryotes (Llacer et al., 2018) and bacteria (Figure 7; Selmer et al., 2006; Fischer et al., 2015; Polikanov et al., 2015; Hussain et al., 2016). In the *P. abyssi* IC, the position of the C-terminal arginine uS9-R135 is further stabilized by interaction with uS19-R124. uS9 is highly conserved in the three domains of life and the protein systematically ends with an arginine residue (see alignments in Melnikov et al., 2018). The role of uS9 C-tail in fidelity was previously shown by studies with bacterial (Hoang et al., 2004; Noller et al., 2005; Arora et al., 2013a,b) and yeast systems (Ghosh et al., 2014; Jindal et al., 2019). Moreover, in eukaryotes, uS9 favors the recruitment of the TC on the ribosome (Jindal et al., 2019). Thus, the IC2 structure indicates a universal involvement of uS9 tail in the fidelity of TI. The constant role of the C-terminal arginine of uS9 would, therefore, have been acquired very early in evolution.

Concerning uS13 and uS19, the C-tails of these two proteins are oriented toward the major groove of the anticodon stem of the initiator tRNA (Figure 7). In particular, they interact with G30 of the second base pair of the almost universally conserved three GC base pairs of the anticodon stem of the initiator tRNA, that play a crucial role in translation fidelity (Samhita et al., 2012;

Hussain et al., 2016; Shetty et al., 2017; Ayyub et al., 2018). Sequence alignments of uS13 showed that the basic character of the C-tail is conserved in the three domains of life. Sequence conservation of the C-tail, though strong, is less strict than that of uS9, with some organisms, in all domains, having variable tail lengths (Melnikov et al., 2018). We refined this analysis by focusing on the uS13 tails in archaeal branches. It is striking that the tails have evolved in a branch-dependent manner (Figure 8). Halobacteria strongly differentiate by an acidic tail. On another hand, TACK and Asgard frequently display tails of variable lengths reminiscent of low complexity regions identified in various prokaryotic proteins (Ntountoumi et al., 2019). The C-tail of uS19 is very basic in bacteria. In archaea and in eukaryotes, the tails are eight residues longer with a less pronounced basic character (see alignments in Melnikov et al., 2018). Again, sequence alignments of archaeal representatives (Figure 8) highlight some branch specificities. Overall, the variations of the tails of uS13 and uS19 in archaea likely reflect tuning of translation mechanisms to peculiar environmental conditions.

Consistent with the archaeal case, the tails of uS13 and uS19 were recently observed in contact with the anticodon stem of the initiator tRNA in a mammalian late-stage initiation complex (Simonetti et al., 2020). On another hand, uS13 and uS19 are also involved in translation elongation, as observed recently in a mammalian elongation complex (Bhaskar et al., 2020) and during the translocation step in bacteria (Zhou et al., 2013).

Notably, the core domains of uS13 and uS19, located on the SSU head, are involved in the B1a and B1b/c bridges with the large ribosomal subunit (LSU). Several studies in yeast identified allosteric information pathways connecting functional centers in the LSU to the decoding center in the SSU through these bridges (Ben-Shem et al., 2011; Rhodin and Dinman, 2011; Bowen et al., 2015; Bhaskar et al., 2020). Interestingly, archaeal and eukaryotic uS13 and uS19 have sequence insertions in their core domains as compared to the bacterial proteins. These insertions expand the contact area between the two proteins (Figure 9). In bacteria, where the two insertions are missing, only few contacts between uS13 and uS19 are visible. However, a bacterial specific protein, bL31, interacting with uS19 was recently shown bridging the two subunits of the ribosome (Figure 9; Fischer et al., 2015). This bL31 protein could ensure a similar function as the two specific eukaryotic and archaeal extensions of uS13 and uS19. Overall, these observations likely reflect late evolutions of the mechanisms in the three domains of life. The eukaryotic and archaeal C-tails and insertions further argue in favor of the archaeal ribosome being of the eukaryotic-type.



### INVOLVEMENT OF BASE MODIFICATIONS IN START CODON SELECTION

Like in bacteria and eukaryotes, a series of rRNA modifications is observed around the P site (Coureux et al., 2020). Some of these rRNA modifications are classified as universally conserved. They correspond to m<sup>3</sup>U1467 (m<sup>3</sup>U1498, *E. coli* numbering) and the two dimethyladenosines m<sub>2</sub><sup>6</sup>A1487,

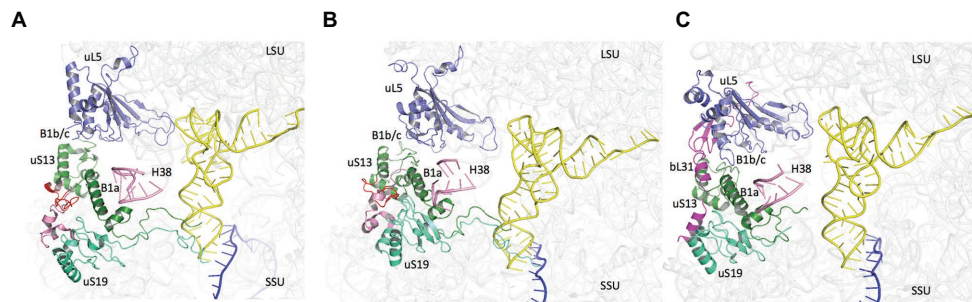
## uS13

			120	130	140	150	160	170
Eukaryote		Homo-sapiens/1-152	D L E R L K K I R A H R G L R F W G L R V R G G H T K T G R - R G R T V G V S R K K K					
Euryarchaeota	Thermococcales	Pyrococcus-abyssi/1-148	D I M R L R R I R A Y R G I R H E L G L P V R G Q T R S N F R - R G Q T V G V S R K K K					
		Thermococcus-kodakarensis/1-149	D V N L R R I R A Y R G I R H E L G L P V R G Q T R S N F R - H G T T V G V S R R K K					
	Methanomicrobia	Methanobolbus-tindarius/1-148	D L N L K R V A F R G I R H E R G L V R G Q T K S T G R - R S S T I G V S K K K					
		Methanosarcina-mazei/1-152	D I N L K V R A Y R G I R H E R G L V R G Q T K S T G R - R G S T V G V S R K K K					
	Methanococci	Methanocaldococcus-jannaschii/1-150	D I N L R K R I R C Y R G I R H E L G L P C R G Q T K S T F R - R G P T V G V S R R K K					
	Archaeoglobales	Archaeoglobus-fulgidus/1-146	D I E R L I K M A Y R G V R H A R G K V R G Q T R S T G R - K G R T V G V V R R K R					
	Halobacteria	Haloflexa-volcani/1-167	D I R R M K M I S Y K G V R R H R G Q V R G Q T K S T G T E G - T I G N V V E I K E A Q A E E					
		Halobellus-captivus/1-177	D I R R M K M I N S Y K G V R R Q R G Q V R G Q T K S T G T E G - T I G N V V A I K E E A A E A A A E E					
		Haloferax-thermautotrophicus/1-173	D I R R M K M I E S Y K G V R R K R G Q V R G Q T K S T G S E G - T I G N V V E I R E E M E D E A					
		Halolamina-pelagica/1-172	D I R R M K M I S Y K G V R R H R G Q V R G Q T K S T G T E G - T I G N V I E I R E E M E D E A					
		Haloprofundus-halophilus/1-174	D I R R M Q M I N S Y K G I R R R G Q V R G Q T K S T G T E G - T I G N V V A I R E E A E E G					
TACK	Crenarchaeota/Thermoprotei	Saccharolobus-solfataricus/1-165	D I E E K R S R S W G V R H S L G L V R G Q T M T I G R - T G M I T G V A R A A Q Q S Q S S S Q Q K S S					
ASGARD	Candidatus Lokiarchaeota	Pyrobaculum-aerophilum/1-152	D V G L M K K L R S W G V R H S L G L V R G Q T V T T G R - F G A L A G V T K K A A A G K					
DPANN	Nanoarchaeota	Candidatus-Prometheoarchaeum-syntrophicum/1-165	D I D K M K R T R S W K G I R H M Y G L V R G Q T R T T G R - R G L V V G V Y R K D M K K K A					
		Nanoarchaeum-equitans/1-153	D I K R E M Q I N S W R G V R R H R E G L P V R G Q T R S H H R - K G R T V G V I K K K					

## uS19

			150	160	170	180	190	200
Eukaryote		Homo-sapiens/1-152	P E M V G S M V G V Y N G - K I F N Q V E I K P E M I G H Y L G E F S I T Y K P V K H G R P G I G A T H S S R F I P L R					
Euryarchaeota	Thermococcales	Pyrococcus-abyssi/1-148	P E M V G L T I V V H N G K E F V P - V E I K P E M I G H Y L G E F A P T K R V D H G A P G I G A T R S S M F V A V K					
		Thermococcus-kodakarensis/1-149	P E M V G L T I V V H N G K E F V P - V E I K P E M I G H Y L G E F A M T K V V D H G A P G I G A T R S S M F V A V K					
	Methanomicrobia	Methanobolbus-tindarius/1-148	P E M V G K N V E Y N G S K F V A - F E I Q P E M I G H R F G E F A P T R S R V H S G S A G V G A T R S S K F V P L K					
		Methanosarcina-mazei/1-152	P E M I G I A I E V H N G K E F V N - V E I Q P E M I G H R F G E F A P T R S R V N H S G S A G V G A T R S S K F V P L K					
	Methanococci	Methanocaldococcus-jannaschii/1-150	P D M V G L T F G V Y N G K E F V E - V K V T P E M I G H Y L G E F S L T R K P V D H G A P G M G A T R S S M F V P I R					
	Archaeoglobales	Archaeoglobus-fulgidus/1-146	P E M V G K V V F V H N G K E F V R - V E I K P E M I G H R G E F A L T R F E K H S G S A G V G A T R S S K F V P L K					
	Halobacteria	Haloflexa-volcani/1-167	P E F V G L T F A V V T Q E F E R - I Q I Q P E M I G H Y L G E F Q L T R S S V E H Q A G I G A T R S S K F V P L K					
		Halobellus-captivus/1-177	P E F V G V T F V V Y N G H S F E R - V Q V E P E M I G H Y L G E F H Q T R S T V E H G Q A G I G A T R S S K F V P L K					
		Haloferax-thermautotrophicus/1-173	P E F V G V T F A V Y N G H S F E R - V Q V E P E M I G H Y L G E F H L T R S T V E H G Q A G I G A T R S S K F V P L K					
		Halolamina-pelagica/1-172	P S F V G L T F E V Y N G H S F D R - V K V E P E M I G H Y L G E F L T R S T V E H G Q A G I G A T R S S K F V P L K					
		Haloprofundus-halophilus/1-174	P A F V G L F E A V Y N G - D S F E R V I Q P E M I G H Y L G E F Q L T R S S V E H Q A G I G A T R S S K F V P L K					
TACK	Crenarchaeota/Thermoprotei	Saccharolobus-solfataricus/1-165	P E L I G L K M A V Y N G K E F V E - F T V T P E M I G H Y L G E Y S I T T K K V E H G E P L K A R S S L F L A M K G					
ASGARD	Candidatus Lokiarchaeota	Pyrobaculum-aerophilum/1-152	P E M V G A T I L V H N G - K E F I V D I V P E M I G C Y L G E F A V N K P V R H S G P G I G A T R S S L H I A A K					
DPANN	Nanoarchaeota	Nanoarchaeum-equitans/1-153	P F M V G V T V H V Y N G K E V V R - V D I K P E M I G H Y L G E F F E T K R K H S A P G I G A T R S S A H V S K K					

**FIGURE 8 |** Sequence alignments of the uS13 and uS19 C-terminal tails. uS13 and uS19 sequences were extracted and aligned using Pipealign (Plewniak et al., 2003). After visual inspection, several families regarding the C-terminal tail specificities in Archaea were identified. For uS13, ca. 250 archaeal sequences were used to which we added manually 100 sequences from halobacteria. For uS19, ca. 600 archaeal sequences were used. Typical representatives of each family are shown. The *Homo sapiens* sequence is used as a eukaryotic reference for comparison.



**FIGURE 9 |** Domain specificities of uS13 and uS19 cores. The three panels show the B1a-B1b/c bridge. **(A)** Archaeal case. The view is a composite using the SSU from PDB 6SWC (Coureux et al., 2020) and the LSU from PDB 4V6U (Armache et al., 2013). **(B)** Structure of human ribosome from PDB 6Y0G (Bhaskar et al., 2020). **(C)** Structure of *Escherichia coli* ribosome from PDB 5AFI (Fischer et al., 2015). The views show that archaeal and eukaryotic uS13 and uS19 have specific extensions (in red and pink, respectively) that contribute to the intersubunit bridge. Bacteria have instead a specific ribosomal protein bL31.

$m_2^6A1488$  ( $m_2^6A1518$ ,  $m_2^6A1519$  in *E. coli*). In contrast, for other positions, the pattern of rRNA modification is specific to eukaryotes and archaea (Coureux et al., 2020). A first layer of rRNA modification stabilizes the codon:anticodon duplex. A second layer made up  $m_2^6A1487$ ,  $m_2^6A1488$ ,  $hm^5C1378$ , stabilizes the first layer. Notably, the N-terminal part of eL41 contacts several modified bases linked to the P site. In particular, the conserved R15 (*P. abyssi* numbering) interacts with an acetylcytidine residue ( $Ac^4C1476$ , *P. abyssi* numbering). The case of eL41 is rather intriguing. Indeed, this protein was first identified as a protein belonging to the large ribosomal subunit explaining its naming. However, recent structures of eukaryotic and archaeal ribosomes revealed that the protein mainly interacts with the SSU. Because the protein interacts with the network of modified bases involved in the control of start codon selection, eL41 was proposed to be involved in the regulation of the

process. Up to now, eL41 has been found in eukaryotes and in most archaea (Lecompte et al., 2002; Coureux et al., 2020). Nevertheless, its identification is rendered difficult because of the small size of the protein (25–37 residues). Finally, it is notable that *P. abyssi* and *T. kodakarensis* ribosomes contain a large amount of  $ac^4C$ , likely involved in thermostability (Coureux et al., 2020; Sas-Chen et al., 2020). Indeed, the amount of  $ac^4C$  varies with growth temperature (Sas-Chen et al., 2020).

## LATE STEPS OF TRANSLATION INITIATION

In eukaryotes and in archaea, the late stage of TI occurring after e/aIF2 departure involves the two factors e/aIF1A and

e/aIF5B. These two factors ensure final check of the presence of the initiator tRNA and assembly with the LSU (**Figure 1**; Choi et al., 1998; Pestova et al., 2000; Maone et al., 2007). As other translational GTP-ases, e/aIF5B activity is related to the transition between an active GTP state and an inactive GDP state controlled by the movement of two switch regions (switch 1 and switch 2) that interact with the nucleotide. aIF5B is composed of four domains (**Figure 2**). The structure of aIF5B from *Methanobacterium thermoautotrophicum* showed that domains I (GTP binding domain), II, and III are packed together and linked by a long  $\alpha$ -helix (helix h12) to domain IV (**Figure 2**; Roll-Mecak et al., 2004). Domain IV contains a  $\beta$ -barrel and is responsible for the binding of the methionylated-CCA end of the initiator tRNA (Guillon et al., 2005). In eukaryotes, the integrity of the h12 helix and the multidomain nature of the factor were shown important for its function (Shin et al., 2011; Kuhle and Ficner, 2014; Huang and Fernández, 2020). Eukaryotic eIF5B contains an additional N-domain which displays little sequence conservation and was shown to be dispensable in yeast (Shin et al., 2002). eIF5B was shown to directly interact with the eukaryotic-specific C-tail of eIF1A (Choi et al., 2000; Marintchev et al., 2003; Zheng et al., 2014). This interaction is required for efficient subunit joining (Acker et al., 2006). Such an interaction between aIF1A and aIF5B has not been evidenced in archaea. Possibly, the two proteins do not directly interact because aIF1A does not possess the C-terminal extension and because aIF5B presents a supplementary helix at the position expected for aIF1A binding site (Murakami et al., 2016).

e/aIF5B-GTP binds the SSU (Maone et al., 2007) and accelerates the recruitment of the large ribosomal subunit (Pestova et al., 2000; Acker et al., 2009). It is possible that interaction of e/aIF5B with proteins of the P stalk contribute to favor the binding of the SSU-IC to the LSU (Murakami et al., 2018). The position of eIF5B on the 80S has been observed in several Cryo-EM studies (Fernandez et al., 2013; Yamamoto et al., 2014; Huang and Fernández, 2020) with domain IV holding the Met-CCA extremity of the initiator tRNA. In addition, the dynamics of its binding has been studied in real-time single-molecule experiments (Wang et al., 2019). A rearrangement of the 80-IC complex containing eIF5B would trigger GTP hydrolysis and the release of the factor leading to the formation of a ribosome competent for elongation.

Importantly, e/aIF5B and e/aIF1A are orthologues of the bacterial proteins IF2 and IF1, respectively. The assembly step has, therefore, a universal character. In bacteria, evolution might have selected formylation of the initiator tRNA to enhance specificity, whereas this improvement would have been gained in eukaryotes and archaea thanks to the emergence of e/aIF2. Interestingly, several studies have shown that in some non-canonical cases, eukaryotic translation initiation used eIF5B instead of eIF2 for the recruitment of the initiator tRNA (Terenin et al., 2008; Thakor and Holcik, 2012; Ho et al., 2018; Ross et al., 2018). This argues in favor of an ancestral translation initiation mechanism involving e/aIF5B/IF2 and e/aIF1A/IF1 that could have been used in the last common universal ancestor (LUCA) and that should also be discussed

in the light of what is known for translation initiation of leaderless mRNAs (Londei, 2005; Beck and Moll, 2018).

## TRANSLATION INITIATION OF LEADERLESS mRNAs

### Insights From Bacteria

The mechanisms for initiation of leaderless mRNAs translation in archaea have been addressed in a limited number of reports. Because leaderless mRNAs are found in the three domains of life (Janssen, 1993), some data concerning bacteria, and to a lesser extent mitochondria may be relevant to the archaeal case. Leaderless mRNAs can indeed be translated in *E. coli* (Balakin et al., 1992; Wu and Janssen, 1996; Grill et al., 2000; Moll et al., 2002b; Udagawa et al., 2004; Vesper et al., 2011; Yamamoto et al., 2016; Beck and Moll, 2018) and are even widespread in some bacterial species such as *Deinococcus* species (de Groot et al., 2014; Bouthier de la Tour et al., 2015) and mycobacterial species (Cortes et al., 2013; Shell et al., 2015; Li et al., 2017).

In bacteria and in Archaea, leaderless mRNAs are featured by a 5' tri-phosphate and an AUG (or GUG or UUG) start codon near the 5' extremity. Notably, in mammalian mitochondria, post-transcriptional processing of long transcript produces a free phosphate group at the 5' end of the mRNA. Accordingly, whereas in bacteria, a free phosphate group at the 5' end of a leaderless mRNA was shown important for TI (Giliberti et al., 2012), this is not the case in mitochondria (Christian and Spremulli, 2010). Finally, in *E. coli*, sequences located downstream to the start codon, called "downstream box" were proposed to contribute to TI efficiency of leaderless mRNAs (Sprengart et al., 1996; Martin-Farmer and Janssen, 1999) although this point is debated (Resch et al., 1996; O'Connor et al., 1999).

Studies in *E. coli* mainly used the mRNA encoding the *ci* repressor of the  $\lambda$  bacteriophage as a model leaderless mRNA and its translation from assembled 70S ribosomes was early proposed (Balakin et al., 1992). Numerous studies have demonstrated that 70S monomers were able to initiate translation of leaderless mRNAs *in vitro* (e.g., Moll et al., 2004; Udagawa et al., 2004; Yamamoto et al., 2016). Moreover, it was observed that in a strain deficient for the ribosome recycling factor RRE, i.e., under conditions where 70S ribosomes were abundant, translation of leaderless mRNAs was maintained whereas that of SD-leaderless mRNAs was inhibited (Moll et al., 2004). The possibility for the 70S monomers to initiate translation of leaderless mRNAs is connected to the accessibility of the ribosome for the mRNA (Yamamoto et al., 2016). Indeed, the channel in the 30S subunit can readily bind any mRNA at an internal site whereas a 70S ribosome must thread the mRNA from an extremity. Notably, initiation with 30S subunits has also been observed (Balakin et al., 1992; Grill et al., 2000).

In bacteria, the initiator tRNA and initiation factor 2 (IF2) have a crucial role in the recruitment of the leaderless mRNA, whether on the 30S subunit (Grill et al., 2000, 2001) or on the 70S ribosome (Udagawa et al., 2004; Yamamoto et al., 2016).

This strongly indicates that codon:anticodon base pairing has an important contribution to the affinity of the leaderless mRNA for the ribosome. Notably, the 5'-triphosphate group in the vicinity of the start codon may also be useful for leaderless mRNA binding (Giliberti et al., 2012). In this context, it is notable that the addition of 5 or 10 nucleotides without an SD sequence before the AUG codon abolished 70S initiation *in vitro* (Udagawa et al., 2004). IF3 was also shown important for TI of leaderless mRNAs though its role is less clear. In a purified translation system, leaderless mRNA translation was found strictly dependent on IF3 (Yamamoto et al., 2016). However, IF3 was also reported to be inhibitory, in particular at high concentrations (Tedin et al., 1999; Grill et al., 2001; Udagawa et al., 2004). The inhibitory effect of IF3 may be linked to its ribosomal subunits anti-associative activity (Dallas and Noller, 2001) and to its possible role in ribosome recycling (Peske et al., 2005; Zavialov et al., 2005). Hence, prolonged incubation with high levels of IF3 may decrease the availability of 70S ribosomes (Peske et al., 2005). Further, IF3 closely participates in start codon selection by destabilizing codon-anticodon interaction (Hartz et al., 1989; Gualerzi and Pon, 1990). Thus, because of the key importance of codon-anticodon pairing in the binding of a leaderless mRNA by the initiating ribosome, translation of leaderless mRNAs may be disfavored by high IF3 levels.

In summary, the most recent data rather favor translation initiation of leaderless mRNA in bacteria with 70S ribosomes, assisted by IF2, f-Met-tRNA<sup>Met</sup>, and IF3. However, the co-existence of a mechanism using 30S subunits cannot be fully excluded at this stage.

Specialized ribosomes were observed to translate leadered and leaderless mRNAs with different efficiencies. Notably, the absence of bS1 and uS2 favors leaderless mRNA translation by 70S ribosome (Moll et al., 2002a). Further, exposure of *E. coli* to the antibiotic kasugamycin induced 61S particles, devoid of six proteins in the small subunit (bS1, uS2, bS6, uS12, bS18, and bS21), that selectively translate leaderless mRNAs (Kaberina et al., 2009). Finally, MazEF, a toxin-antitoxin system induced by stress in *E. coli* was shown to function by generating specific leaderless mRNAs together with specialized ribosomes lacking the 16S rRNA region containing the anti-SD motif (Vesper et al., 2011).

Another example of leaderless translation is found in mammalian mitochondria. In these organelles, IF1 is absent whereas mt-IF2 and mt-IF3 have acquired specific structural extensions. Notably, mammalian mt-IF2 can replace both IF1 and IF2 for supporting *E. coli* growth (Gaur et al., 2008). Recent structural studies proposed that mt-IF3 would be necessary to stabilize the mt-SSU head for the accommodation of mt-IF2. After mt-IF3 release the recruitments of the initiator tRNA, of the mRNA and of the LSU would complete the initiation complex. Release of mt-IF3 would be an obligatory step for tRNA binding and mt-IF2 would be necessary for leaderless mRNA recruitment by the ribosome (Kummer et al., 2018; Koripella et al., 2019; Khawaja et al., 2020).

## The Archaeal Case

In archaea, recruitment of leaderless mRNAs by the ribosome has been studied *in vitro* using *S. solfataricus* components

(Condo et al., 1999; Benelli et al., 2003). It was observed that 30S subunits were unable to stably bind leaderless mRNA. However, addition of methionylated initiator tRNA was sufficient to form a complex, where the 30S subunit is positioned with the start codon of the leaderless mRNA in the P site, as assessed by toeprinting methods (Benelli et al., 2003). Further addition of aIF5B, the homologue of bacterial IF2 (Figure 2) did not significantly enhance the intensity of the toeprint (Benelli et al., 2003). Whether initiation of archaeal leaderless mRNAs translation occurs with 30S subunits or 70S ribosomes remains an open question. Considering the properties of archaeal initiation factors, several possibilities may be envisaged. In the first model, initiation would occur with 70S subunits. In this case, a direct involvement of the heterotrimeric aIF2 in tRNA recruitment is unlikely because the structure of the complex containing the 30S subunit, aIF2, and the initiator tRNA is not compatible with 50S assembly (Armache et al., 2013; Coureux et al., 2016, 2020). However, aIF5B, the archaeal homologue of bacterial IF2, can indeed bind both the 70S ribosome (Maone et al., 2007) and the methionylated initiator tRNA (Guillon et al., 2005). Thus, a 70S:aIF5B:Met-tRNA<sup>Met</sup> complex would be able to recruit the leaderless mRNA thanks to start codon-tRNA anticodon pairing and possibly binding of the 5'-triphosphate on the mRNA. A second possible mechanism would be mediated by the SSU, similarly to the SD-leadered mRNAs (see above). In such a model, the mRNA would be mainly tethered to the 30S subunit thanks to base pairing of the start codon with the anticodon of the initiator tRNA bound to aIF2-GTP and to the P site (Benelli et al., 2003). Whatever the mechanism, recruitment of the leaderless mRNA by either 30S or 70S ribosomes, it may be imagined that the leaderless mRNA can interact with the ribosome apart from the codon-anticodon pairing. In this view, it should be reminded that the  $\gamma$  subunit of *S. solfataricus* aIF2 is able to bind the 5'-triphosphorylated end of mRNAs, thereby protecting them against degradation (Hasenöhrl et al., 2008). It cannot be excluded that this activity also facilitates leaderless mRNA recruitment by the ribosome. Furthermore, as discussed in the present review (see Figure 5), the mRNA exit channel on the 30S subunit has archaeal-specific features, as well as features distinguishing archaeal branches. For instance, *S. solfataricus* and *P. aerophilum*, two archaea widely using leaderless mRNAs, have a larger set of 30S r-proteins. In contrast, Halobacteria and Thermoplasma, also widely use leaderless mRNAs but have a reduced set of r-proteins in the 30S (Lecompte et al., 2002). This opens the possibility that mechanisms of leaderless mRNA recruitment may somewhat vary within the archaeal world. These considerations deserve further investigation.

Finally, it has been reported that mRNAs with leaders not including an SD sequence were not translated in a cell-free *S. solfataricus* system, where leaderless mRNAs were efficiently used (Benelli et al., 2003). Leadered mRNAs with no obvious SD sequence apparently occur (Wurtzel et al., 2010). It may be hypothesized that translation of such mRNAs requires scanning of a 70S ribosome from the 5'-end of the mRNA, as proposed in the bacterial case (Yamamoto et al., 2016).

However, the mechanistic analogy between the bacterial and archaeal cases is limited by the observation that 70S-bound IF3 is mandatory for scanning in bacteria (Yamamoto et al., 2016). Whether aIF1 might play a similar role remains an open question. Hence, another possibility would be a mechanism involving an IC on the archaeal 30S resembling and foreshadowing eukaryotic scanning. Further studies are clearly needed to assess these various hypotheses or to decipher alternative mechanisms.

## CONCLUDING REMARKS

Study of TI mechanisms in Archaea has gained a new momentum in recent years thanks to the fast development of phylogenetic analyses, genome-wide studies, and 3D structure determinations. The archaeal ribosomes are of the eukaryotic type but they have specificities linked to the mode of mRNA recruitment. In many archaea, mRNAs carry SD sequences complementary to the 3' end of the ribosomal RNA. The formation of the SD:antiSD duplex favors the initiation complex and positioning of the initiator tRNA in the vicinity of the P site on the SSU. The SD:antiSD duplex is stabilized in an exit chamber on the SSU. The protein organization of this chamber is specific to the archaeal domain. It is very close to that in eukaryotes but very different from that in bacteria. Thus, bacteria and archaea have evolved two different structural solutions for the binding of the SD:antiSD duplex.

On another hand, transcriptomic data show that in many archaea, mRNAs are leaderless. Thus, the mode of recruitment of these leaderless mRNAs would be different. Leaderless mRNAs are considered to be ancient and may reflect the original TI

mechanisms that existed in the LUCA. However, translation initiation mechanisms of these mRNAs are still unclear and the current data do not exclude an initiation mode involving the 30S or a pre-assembled 70S. In both cases, the role of the initiator tRNA would be essential. Interestingly, the presence of the three ribosomal proteins, eS26, eS25, and eS30, systematically found in eukaryotes, varies between the different branches of archaea. These three proteins have a direct link with translation initiation. In eukaryotes, eS26 is located in the mRNA exit channel, eS30 is located in the mRNA entry channel, and eS25 is observed in contact with the initiator tRNA. Additional studies are necessary to establish their function in Archaea. However, it is tempting to imagine that the presence of these proteins is related to an evolution of the ribosomes coupled to that of the organization of mRNAs. To answer these questions, the accumulation of new data is necessary. In this context, high-resolution cryo-EM structures will contribute to bring important information.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This work was supported by grants from the Centre National de la Recherche Scientifique and Ecole Polytechnique to Unité Mixte de Recherche n°7654 and by a grant from the Agence Nationale de la Recherche (ANR-17-CE11-0037).

## REFERENCES

- Acker, M. G., Shin, B. S., Dever, T. E., and Lorsch, J. R. (2006). Interaction between eukaryotic initiation factors 1A and 5B is required for efficient ribosomal subunit joining. *J. Biol. Chem.* 281, 8469–8475. doi: 10.1074/jbc.M600210200
- Acker, M. G., Shin, B. S., Nanda, J. S., Saini, A. K., Dever, T. E., and Lorsch, J. R. (2009). Kinetic analysis of late steps of eukaryotic translation initiation. *J. Mol. Biol.* 385, 491–506. doi: 10.1016/j.jmb.2008.10.029
- Akil, C., and Robinson, R. C. (2018). Genomes of Asgard archaea encode profilins that regulate actin. *Nature* 562, 439–443. doi: 10.1038/s41586-018-0548-6
- Algire, M. A., Maag, D., and Lorsch, J. R. (2005). Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell* 20, 251–262. doi: 10.1016/j.molcel.2005.09.008
- Armache, J. P., Anger, A. M., Marquez, V., Franckenberg, S., Frohlich, T., Villa, E., et al. (2013). Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. *Nucleic Acids Res.* 41, 1284–1293. doi: 10.1093/nar/gks1259
- Arora, S., Bhamidimarri, S. P., Bhattacharyya, M., Govindan, A., Weber, M. H., Vishveshwar, S., et al. (2013a). Distinctive contributions of the ribosomal P-site elements m2G966, m5C967 and the C-terminal tail of the S9 protein in the fidelity of initiation of translation in *Escherichia coli*. *Nucleic Acids Res.* 41, 4963–4975. doi: 10.1093/nar/gkt175
- Arora, S., Bhamidimarri, S. P., Weber, M. H., and Varshney, U. (2013b). Role of the ribosomal P-site elements of m(2)G966, m(5)C967, and the S9 C-terminal tail in maintenance of the reading frame during translational elongation in *Escherichia coli*. *J. Bacteriol.* 195, 3524–3530. doi: 10.1128/JB.00455-13
- Ayyub, S. A., Dobriyal, D., Shah, R. A., Lahry, K., Bhattacharyya, M., Bhattacharyya, S., et al. (2018). Coevolution of the translational machinery optimizes initiation with unusual initiator tRNAs and initiation codons in mycoplasmas. *RNA Biol.* 15, 70–80. doi: 10.1080/15476286.2017.1377879
- Ayyub, S. A., Dobriyal, D., and Varshney, U. (2017). Contributions of the N- and C-terminal domains of initiation factor 3 to its functions in the fidelity of initiation and antiassociation of the ribosomal subunits. *J. Bacteriol.* 199, e00051–e00017. doi: 10.1128/JB.00051-17
- Babski, J., Haas, K. A., Nather-Schindler, D., Pfeiffer, F., Forstner, K. U., Hammelmann, M., et al. (2016). Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* 17:629. doi: 10.1186/s12864-016-2920-y
- Balakin, A. G., Skripkin, E. A., Shatsky, I. N., and Bogdanov, A. A. (1992). Unusual ribosome binding properties of mRNA encoding bacteriophage lambda repressor. *Nucleic Acids Res.* 20, 563–571. doi: 10.1093/nar/20.3.563
- Ban, N., Beckmann, R., Cate, J. H., Dinman, J. D., Dragon, F., Ellis, S. R., et al. (2014). A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* 24, 165–169. doi: 10.1016/j.sbi.2014.01.002
- Beck, H. J., and Moll, I. (2018). Leaderless mRNAs in the spotlight: ancient but not outdated! *Microbiol. Spectr.* 6, 1–15. doi: 10.1128/microbiolspec.RWR-0016-2017
- Benelli, D., and Londei, P. (2011). Translation initiation in Archaea: conserved and domain-specific features. *Biochem. Soc. Trans.* 39, 89–93. doi: 10.1042/BST0390089
- Benelli, D., Maone, E., and Londei, P. (2003). Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation. *Mol. Microbiol.* 50, 635–643. doi: 10.1046/j.1365-2958.2003.03721.x
- Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* 334, 1524–1529. doi: 10.1126/science.1212642

- Bernier, C. R., Petrov, A. S., Kovacs, N. A., Penev, P. I., and Williams, L. D. (2018). Translation: the universal structural core of life. *Mol. Biol. Evol.* 35, 2065–2076. doi: 10.1093/molbev/msy101
- Bhaskar, V., Graff-Meyer, A., Schenk, A. D., Cavadini, S., Von Loeffelholz, O., Natchiar, S. K., et al. (2020). Dynamics of uS19 C-terminal tail during the translation elongation cycle in human ribosomes. *Cell Rep.* 31:107473. doi: 10.1016/j.celrep.2020.03.037
- Bouthier de la Tour, C., Blanchard, L., Dulermo, R., Ludanyi, M., Devigne, A., Armengaud, J., et al. (2015). The abundant and essential HU proteins in *Deinococcus deserti* and *Deinococcus radiodurans* are translated from leaderless mRNA. *Microbiology* 161, 2410–2422. doi: 10.1099/mic.0.000186
- Bowen, A. M., Musalgaonkar, S., Moomau, C. A., Gulay, S. P., Mirvis, M., and Dinman, J. D. (2015). Ribosomal protein uS19 mutants reveal its role in coordinating ribosome structure and function. *Translation* 3:e1117703. doi: 10.1080/21690731.2015.1117703
- Bowman, J. C., Petrov, A. S., Frenkel-Pinter, M., Penev, P. I., and Williams, L. D. (2020). Root of the tree: the significance, evolution, and origins of the ribosome. *Chem. Rev.* 120, 4848–4878. doi: 10.1021/acs.chemrev.9b00742
- Brenneis, M., Hering, O., Lange, C., and Soppa, J. (2007). Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea. *PLoS Genet.* 3:e229. doi: 10.1371/journal.pgen.0030229
- Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'souza, L. M., Du, Y., et al. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2. doi: 10.1186/1471-2105-3-2
- Castelle, C. J., and Banfield, J. F. (2018). Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* 172, 1181–1197. doi: 10.1016/j.cell.2018.02.016
- Cho, S., Kim, M. S., Jeong, Y., Lee, B. R., Lee, J. H., Kang, S. G., et al. (2017). Genome-wide primary transcriptome analysis of H<sub>2</sub>-producing archaeon *Thermococcus onnurineus* NA1. *Sci. Rep.* 7:43044. doi: 10.1038/srep43044
- Choi, S. K., Lee, J. H., Zoll, W. L., Merrick, W. C., and Dever, T. E. (1998). Promotion of met-tRNA<sup>iMet</sup> binding to ribosomes by yIF2, a bacterial IF2 homolog in yeast. *Science* 280, 1757–1760. doi: 10.1126/science.280.5370.1757
- Choi, S. K., Olsen, D. S., Roll-Mecak, A., Martung, A., Remo, K. L., Burley, S. K., et al. (2000). Physical and functional interaction between the eukaryotic orthologs of prokaryotic translation initiation factors IF1 and IF2. *Mol. Cell Biol.* 20, 7183–7191. doi: 10.1128/MCB.20.19.7183-7191.2000
- Christian, B. E., and Spremulli, L. L. (2010). Preferential selection of the 5'-terminal start codon on leaderless mRNAs by mammalian mitochondrial ribosomes. *J. Biol. Chem.* 285, 28379–28386. doi: 10.1074/jbc.M110.149054
- Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., et al. (2003). An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol. Microbiol.* 47, 1495–1512. doi: 10.1046/j.1365-2958.2003.03381.x
- Condo, I., Ciammaruconi, A., Benelli, D., Ruggero, D., and Londei, P. (1999). Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* 34, 377–384. doi: 10.1046/j.1365-2958.1999.01615.x
- Cortes, T., Schubert, O. T., Rose, G., Arnvig, K. B., Comas, I., Aebersold, R., et al. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* 5, 1121–1131. doi: 10.1016/j.celrep.2013.10.031
- Coureur, P. D., Lazennec-Schurdevin, C., Bourcier, S., Mechulam, Y., and Schmitt, E. (2020). Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation. *Commun. Biol.* 3:58. doi: 10.1038/s42003-020-0780-0
- Coureur, P. D., Lazennec-Schurdevin, C., Monestier, A., Larquet, E., Cladiere, L., Klaholz, B. P., et al. (2016). Cryo-EM study of start codon selection during archaeal translation initiation. *Nat. Commun.* 7:13366. doi: 10.1038/ncomms13366
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Dallas, A., and Noller, H. F. (2001). Interaction of translation initiation factor 3 with the 30S ribosomal subunit. *Mol. Cell* 8, 855–864. doi: 10.1016/S1097-2765(01)00356-2
- de Groot, A., Roche, D., Fernandez, B., Ludanyi, M., Cruveiller, S., Pignol, D., et al. (2014). RNA sequencing and proteogenomics reveal the importance of leaderless mRNAs in the radiation-tolerant bacterium *Deinococcus deserti*. *Genome Biol. Evol.* 6, 932–948. doi: 10.1093/gbe/evu069
- Dennis, P. P. (1997). Ancient ciphers: translation in Archaea. *Cell* 89, 1007–1010. doi: 10.1016/S0092-8674(00)80288-3
- Dev, K., Santangelo, T. J., Rothenburg, S., Neculai, D., Dey, M., Sicheri, F., et al. (2009). Archaeal aIF2B interacts with eukaryotic translation initiation factors eIF2 $\alpha$  and eIF2 $\beta$ : implications for aIF2B function and eIF2B regulation. *J. Mol. Biol.* 392, 701–722. doi: 10.1016/j.jmb.2009.07.030
- Dubiez, E., Aleksandrov, A., Lazennec-Schurdevin, C., Mechulam, Y., and Schmitt, E. (2015). Identification of a second GTP-bound magnesium ion in archaeal initiation factor 2. *Nucleic Acids Res.* 43, 2946–2957. doi: 10.1093/nar/gkv053
- Eme, L., and Ettema, T. J. G. (2018). The eukaryotic ancestor shapes up. *Nature* 562, 352–353. doi: 10.1038/d41586-018-06868-2
- Eme, L., Spang, A., Lombard, J., Stairs, C. W., and Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* 15, 711–723. doi: 10.1038/nrmicro.2017.133
- Fernandez, I. S., Bai, X. C., Hussain, T., Kelley, A. C., Lorsch, J. R., Ramakrishnan, V., et al. (2013). Molecular architecture of a eukaryotic translational initiation complex. *Science* 342:1240585. doi: 10.1126/science.1240585
- Ferretti, M. B., Ghalei, H., Ward, E. A., Potts, E. L., and Karbstein, K. (2017). Rps26 directs mRNA-specific translation by recognition of Kozak sequence elements. *Nat. Struct. Mol. Biol.* 24, 700–707. doi: 10.1038/nsmb.3442
- Fischer, N., Neumann, P., Konevega, A. L., Bock, L. V., Ficner, R., Rodnina, M. V., et al. (2015). Structure of the *E. coli* ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* 520, 567–570. doi: 10.1038/nature14275
- Fox, G. E. (2010). Origin and evolution of the ribosome. *Cold Spring Harb. Perspect. Biol.* 2:a003483. doi: 10.1101/cshperspect.a003483
- French, S. L., Santangelo, T. J., Beyer, A. L., and Reeve, J. N. (2007). Transcription and translation are coupled in Archaea. *Mol. Biol. Evol.* 24, 893–895. doi: 10.1093/molbev/msm007
- Gäbel, K., Schmitt, J., Schulz, S., Nather, D. J., and Soppa, J. (2013). A comprehensive analysis of the importance of translation initiation factors for *Haloferax volcanii* applying deletion and conditional depletion mutants. *PLoS One* 8:e77188. doi: 10.1371/journal.pone.0077188
- Gaur, R., Grasso, D., Datta, P. P., Krishna, P. D. V., Das, G., Spencer, A., et al. (2008). *Mol. Cell* 29, 180–190. doi: 10.1016/j.molcel.2007.11.021
- Gelsinger, D. R., Dallon, E., Reddy, R., Mohammad, F., Buskirk, A. R., and Diruggiero, J. (2020). Ribosome profiling in archaea reveals leaderless translation, novel translational initiation sites, and ribosome pausing at single codon resolution. *Nucleic Acids Res.* 48, 5201–5216. doi: 10.1093/nar/gkaa304
- Ghosh, A., Jindal, S., Bentley, A. A., Hinnebusch, A. G., and Komar, A. A. (2014). Rps5-Rps16 communication is essential for efficient translation initiation in yeast *S. cerevisiae*. *Nucleic Acids Res.* 42, 8537–8555. doi: 10.1093/nar/gku550
- Giliberti, J., O'donnell, S., Etten, W. J., and Janssen, G. R. (2012). A 5'-terminal phosphate is required for stable ternary complex formation and translation of leaderless mRNA in *Escherichia coli*. *RNA* 18, 508–518. doi: 10.1261/rna.027698.111
- Gogoi, P., Srivastava, A., Jayaprakash, P., Jeyakanthan, J., and Kanaujia, S. P. (2016). In silico analysis suggests that PH0702 and PH0208 encode for methylthioribose-1-phosphate isomerase and ribose-1,5-bisphosphate isomerase, respectively, rather than aIF2 $\beta$  and aIF2 $\delta$ . *Gene* 575, 118–126. doi: 10.1016/j.gene.2015.08.048
- Grill, S., Gualerzi, C. O., Londei, P., and Blasi, U. (2000). Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *EMBO J.* 19, 4101–4110. doi: 10.1093/emboj/19.15.4101
- Grill, S., Moll, I., Hasenöhrl, D., Gualerzi, C. O., and Blasi, U. (2001). Modulation of ribosomal recruitment to 5'-terminal start codons by translation initiation factors IF2 and IF3. *FEBS Lett.* 495, 167–171. doi: 10.1016/S0014-5793(01)02378-X
- Grünberger, F., Reichelt, R., Bunk, B., Spröer, C., Overmann, J., Rachel, R., et al. (2019). Next generation DNA-seq and differential RNA-seq allow re-annotation of the *Pyrococcus furiosus* DSM 3638 genome and provide insights into archaeal antisense transcription. *Front. Microbiol.* 10:1603. doi: 10.3389/fmicb.2019.01603

- Gualerzi, C. O., and Pon, C. L. (1990). Initiation of mRNA translation in prokaryotes. *Biochemistry* 29, 5881–5889. doi: 10.1021/bi00477a001
- Guillon, J. M., Heiss, S., Suturina, J., Mechulam, Y., Laalami, S., Grunberg-Manago, M., et al. (1996). Interplay of methionine tRNAs with translation elongation factor Tu and translation initiation factor 2 in *Escherichia coli*. *J. Biol. Chem.* 271, 22321–22325. doi: 10.1074/jbc.271.37.22321
- Guillon, J. M., Mechulam, Y., Blanquet, S., and Fayat, G. (1993). Importance of formylability and anticodon stem sequence to give tRNA<sup>Met</sup> an initiator identity in *Escherichia coli*. *J. Bacteriol.* 175, 4507–4514. doi: 10.1128/JB.175.14.4507-4514.1993
- Guillon, L., Schmitt, E., Blanquet, S., and Mechulam, Y. (2005). Initiator tRNA binding by eIF5B, the eukaryotic/archaeal homologue of bacterial initiation factor IF2. *Biochemistry* 44, 15594–15601. doi: 10.1021/bi051514j
- Guy, L., and Ettema, T. J. (2011). The archaeal “TACK” superphylum and the origin of eukaryotes. *Trends Microbiol.* 19, 580–587. doi: 10.1016/j.tim.2011.09.002
- Hartman, H., Favaretto, P., and Smith, T. F. (2006). The archaeal origins of the eukaryotic translational system. *Archaea* 2, 1–9. doi: 10.1155/2006/431618
- Hartz, D., Mcpheeters, D. S., and Gold, L. (1989). Selection of the initiator tRNA by *Escherichia coli* initiation factors. *Genes Dev.* 3, 1899–1912. doi: 10.1101/gad.3.12a.1899
- Hasenöhrl, D., Benelli, D., Barbazza, A., Londei, P., and Bläsi, U. (2006). *Sulfolobus solfataricus* translation initiation factor 1 stimulates translation initiation complex formation. *RNA* 12, 674–682. doi: 10.1261/rna.2289306
- Hasenöhrl, D., Fabbretti, A., Londei, P., Gualerzi, C. O., and Bläsi, U. (2009). Translation initiation complex formation in the crenarchaeon *Sulfolobus solfataricus*. *RNA* 15, 2288–2298. doi: 10.1261/rna.1662609
- Hasenöhrl, D., Lombo, T., Kaberdin, V., Londei, P., and Bläsi, U. (2008). Translation initiation factor a/eIF2(- $\gamma$ ) counteracts 5' to 3' mRNA decay in the archaeon *Sulfolobus solfataricus*. *Proc. Natl. Acad. Sci. U. S. A.* 105, 2146–2150. doi: 10.1073/pnas.0708894105
- Hatzopoulos, G. N., and Mueller-Dieckmann, J. (2010). Structure of translation initiation factor 1 from *Mycobacterium tuberculosis* and inferred binding to the 30S ribosomal subunit. *FEBS Lett.* 584, 1011–1015. doi: 10.1016/j.febslet.2010.01.051
- Hinnebusch, A. G. (2017). Structural insights into the mechanism of scanning and start codon recognition in eukaryotic translation initiation. *Trends Biochem. Sci.* 42, 589–611. doi: 10.1016/j.tibs.2017.03.004
- Ho, J. J. D., Balukoff, N. C., Cervantes, G., Malcolm, P. D., Krieger, J. R., and Lee, S. (2018). Oxygen-sensitive remodeling of central carbon metabolism by archaic eIF5B. *Cell Rep.* 22, 17–26. doi: 10.1016/j.celrep.2017.12.031
- Hoang, L., Fredrick, K., and Noller, H. F. (2004). Creating ribosomes with an all-RNA 30S subunit P site. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12439–12443. doi: 10.1073/pnas.0405227101
- Huang, B. Y., and Fernández, I. S. (2020). Long-range interdomain communications in eIF5B regulate GTP hydrolysis and translation initiation. *Proc. Natl. Acad. Sci. U. S. A.* 117, 1429–1437. doi: 10.1073/pnas.1916436117
- Huber, M., Faure, G., Laass, S., Kolbe, E., Seitz, K., Wehrheim, C., et al. (2019). Translational coupling via termination-reinitiation in archaea and bacteria. *Nat. Commun.* 10:4006. doi: 10.1038/s41467-019-11999-9
- Hussain, T., Llacer, J. L., Fernandez, I. S., Munoz, A., Martin-Marcos, P., Savva, C. G., et al. (2014). Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* 159, 597–607. doi: 10.1016/j.cell.2014.10.001
- Hussain, T., Llacer, J. L., Wimberly, B. T., Kieft, J. S., and Ramakrishnan, V. (2016). Large-scale movements of IF3 and tRNA during bacterial translation initiation. *Cell* 167, 133–144. doi: 10.1016/j.cell.2016.08.074
- Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., et al. (2020). Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577, 519–525. doi: 10.1038/s41586-019-1916-6
- Jäger, D., Forstner, K. U., Sharma, C. M., Santangelo, T. J., and Reeve, J. N. (2014). Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* 15:684. doi: 10.1186/1471-2164-15-684
- Jäger, D., Sharma, C. M., Thomsen, J., Ehlers, C., Vogel, J., and Schmitz, R. A. (2009). Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21878–21882. doi: 10.1073/pnas.0909051106
- Janssen, G. R. (1993). “Eubacterial, archaeobacterial, and eukaryotic genes that encode leaderless mRNA” in *Industrial microorganisms: Basic and applied molecular genetics*. ed. E. A. Baltz (Washington, DC: ASM Press), 59–67.
- Jindal, S., Ghosh, A., Ismail, A., Singh, N., and Komar, A. A. (2019). Role of the uS9/yS16 C-terminal tail in translation initiation and elongation in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 47, 806–823. doi: 10.1093/nar/gky1180
- Jivotovskaya, A. V., Valásek, L., Hinnebusch, A. G., and Nielsen, K. H. (2006). Eukaryotic translation initiation factor 3 (eIF3) and eIF2 can promote mRNA binding to 40S subunits independently of eIF4G in yeast. *Mol. Cell Biol.* 26, 1355–1372. doi: 10.1128/MCB.26.4.1355-1372.2006
- Kaberina, A. C., Szaflarski, W., Nierhaus, K. H., and Moll, I. (2009). An unexpected type of ribosomes induced by kasugamycin: a look into ancestral times of protein synthesis? *Mol. Cell* 33, 227–236. doi: 10.1016/j.molcel.2008.12.014
- Karlin, S., Mrázek, J., Ma, J., and Brocchieri, L. (2005). Predicted highly expressed genes in archaeal genomes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7303–7308. doi: 10.1073/pnas.0502313102
- Khawaja, A., Itoh, Y., Remes, C., Spähr, H., Yukhnovets, O., Höfig, H., et al. (2020). Distinct pre-initiation steps in human mitochondrial translation. *Nat. Commun.* 11:2932. doi: 10.1038/s41467-020-16503-2
- Koripella, R. K., Sharma, M. R., Risteff, P., Keshavan, P., and Agrawal, R. K. (2019). Structural insights into unique features of the human mitochondrial ribosome recycling. *Proc. Natl. Acad. Sci. U. S. A.* 116, 8283–8288. doi: 10.1073/pnas.1815675116
- Korneeva, N. L., Lamphear, B. J., Hennigan, F. L., and Rhoads, R. E. (2000). Mutually cooperative binding of eukaryotic translation initiation factor (eIF) 3 and eIF4A to human eIF4G-1. *J. Biol. Chem.* 275, 41369–41376. doi: 10.1074/jbc.M007525200
- Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292. doi: 10.1016/0092-8674(86)90762-2
- Kramer, P., Gäbel, K., Pfeiffer, F., and Soppa, J. (2014). *Haloferax volcanii*, a prokaryotic species that does not use the Shine Dalgarno mechanism for translation initiation at 5'-UTRs. *PLoS One* 9:e94979. doi: 10.1371/journal.pone.0094979
- Kuhle, B., and Ficner, R. (2014). eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining. *EMBO J.* 33, 1177–1191. doi: 10.1002/embj.201387344
- Kummer, E., Leibundgut, M., Rackham, O., Lee, R. G., Boehringer, D., Filipovska, A., et al. (2018). Unique features of mammalian mitochondrial translation initiation revealed by cryo-EM. *Nature* 560, 263–267. doi: 10.1038/s41586-018-0373-y
- Kyrpides, N. C., and Woese, C. R. (1998). Universally conserved translation initiation factors. *Proc. Natl. Acad. Sci. U. S. A.* 95, 224–228. doi: 10.1073/pnas.95.1.224
- La Teana, A., Benelli, D., Londei, P., and Bläsi, U. (2013). Translation initiation in the crenarchaeon *Sulfolobus solfataricus*: eukaryotic features but bacterial route. *Biochem. Soc. Trans.* 41, 350–355. doi: 10.1042/BST20120300
- Lecompte, O., Ripp, R., Thierry, J. C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* 30, 5382–5390. doi: 10.1093/nar/gkf693
- Li, X., Mei, H., Chen, F., Tang, Q., Yu, Z., Cao, X., et al. (2017). Transcriptome landscape of *Mycobacterium smegmatis*. *Front. Microbiol.* 8:2505. doi: 10.3389/fmicb.2017.02505
- Li, J., Qi, L., Guo, Y., Yue, L., Li, Y., Ge, W., et al. (2015). Global mapping transcriptional start sites revealed both transcriptional and post-transcriptional regulation of cold adaptation in the methanogenic archaeon *Methanobrevibacterium psychrophilus*. *Sci. Rep.* 5:9209. doi: 10.1038/srep09209
- Llacer, J. L., Hussain, T., Marler, L., Aitken, C. E., Thakur, A., Lorsch, J. R., et al. (2015). Conformational differences between open and closed states of the eukaryotic translation initiation complex. *Mol. Cell* 59, 399–412. doi: 10.1016/j.molcel.2015.06.033
- Llacer, J. L., Hussain, T., Saini, A. K., Nanda, J. S., Kaur, S., Gordiyenko, Y., et al. (2018). Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *elife* 7:e39273. doi: 10.7554/eLife.39273
- Lo Gullo, G., Mattosovich, R., Perugino, G., La Teana, A., Londei, P., and Benelli, D. (2019). Optimization of an *in vitro* transcription/translation system based on *Sulfolobus solfataricus* cell lysate. *Archaea* 2019:9848253. doi: 10.1155/2019/9848253

- Lomakin, I. B., Kolupaeva, V. G., Marintchev, A., Wagner, G., and Pestova, T. V. (2003). Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev.* 17, 2786–2797. doi: 10.1101/gad.1141803
- Londei, P. (2005). Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol. Rev.* 29, 185–200. doi: 10.1016/j.fmmre.2004.10.002
- Ma, J., Campbell, A., and Karlin, S. (2002). Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* 184, 5733–5745. doi: 10.1128/JB.184.20.5733-5745.2002
- Maone, E., Di Stefano, M., Berardi, A., Benelli, D., Marzi, S., La Teana, A., et al. (2007). Functional analysis of the translation factor aIF2/5B in the thermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* 65, 700–713. doi: 10.1111/j.1365-2958.2007.05820.x
- Marintchev, A., Kolupaeva, V. G., Pestova, T. V., and Wagner, G. (2003). Mapping the binding interface between human eukaryotic initiation factors 1A and 5B: a new interaction between old partners. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1535–1540. doi: 10.1073/pnas.0437845100
- Martin, W., and Koonin, E. V. (2006). A positive definition of prokaryotes. *Nature* 442:868. doi: 10.1038/442868c
- Martin-Farmer, J., and Janssen, G. R. (1999). A downstream CA repeat sequence causes translation from leadered and unleadered mRNA in *Escherichia coli*. *Mol. Microbiol.* 31, 1025–1038. doi: 10.1046/j.1365-2958.1999.01228.x
- Mechulam, Y., Blanquet, S., and Schmitt, E. (2011). Translation initiation. *EcoSal Plus* 4, 1–43. doi: 10.1128/ecosalplus.4.2.2
- Melnikov, S., Kwok, H. S., Manakongtreecheep, K., van Den Elzen, A., Thoreen, C. C., and Söll, D. (2020). Archaeal ribosomal proteins possess nuclear localization signal-type motifs: implications for the origin of the cell nucleus. *Mol. Biol. Evol.* 37, 124–133. doi: 10.1093/molbev/msz207
- Melnikov, S., Manakongtreecheep, K., and Soll, D. (2018). Revising the structural diversity of ribosomal proteins across the three domains of life. *Mol. Biol. Evol.* 35, 1588–1598. doi: 10.1093/molbev/msy021
- Moll, I., Grill, S., Gründling, A., and Bläsi, U. (2002a). Effects of ribosomal proteins S1, S2 and the DeaD/CsdA DEAD-box helicase on translation of leaderless and canonical mRNAs in *Escherichia coli*. *Mol. Microbiol.* 44, 1387–1396. doi: 10.1046/j.1365-2958.2002.02971.x
- Moll, I., Grill, S., Gualerzi, C. O., and Bläsi, U. (2002b). Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Microbiol.* 43, 239–246. doi: 10.1046/j.1365-2958.2002.02739.x
- Moll, I., Hirokawa, G., Kiel, M. C., Kaji, A., and Bläsi, U. (2004). Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res.* 32, 3354–3363. doi: 10.1093/nar/gkh663
- Monestier, A., Lazenec-Schurdevin, C., Coureux, P. D., Mechulam, Y., and Schmitt, E. (2018). Role of aIF1 in *Pyrococcus abyssi* translation initiation. *Nucleic Acids Res.* 46, 11061–11074. doi: 10.1093/nar/gky850
- Murakami, R., Miyoshi, T., Uchiumi, T., and Ito, K. (2016). Crystal structure of translation initiation factor 5B from the crenarchaeon *Aeropyrum pernix*. *Proteins* 84, 712–717. doi: 10.1002/prot.25009
- Murakami, R., Singh, C. R., Morris, J., Tang, L., Harmon, I., Takasu, A., et al. (2018). The interaction between the ribosomal stalk proteins and translation initiation factor 5B promotes translation initiation. *Mol. Cell. Biol.* 38, e00067–e00018. doi: 10.1128/MCB.00067-18
- Nanda, J. S., Cheung, Y. N., Takacs, J. E., Martin-Marcos, P., Saini, A. K., Hinnebusch, A. G., et al. (2009). eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation. *J. Mol. Biol.* 394, 268–285. doi: 10.1016/j.jmb.2009.09.017
- Noller, H. F., Hoang, L., and Fredrick, K. (2005). The 30S ribosomal P site: a function of 16S rRNA. *FEBS Lett.* 579, 855–858. doi: 10.1016/j.febslet.2004.11.026
- Noller, H. F., and Woese, C. R. (1981). Secondary structure of 16S ribosomal RNA. *Science* 212, 403–411. doi: 10.1126/science.6163215
- Ntountoumi, C., Vlastaridis, P., Mossialos, D., Stathopoulos, C., Iliopoulos, I., Promponas, V., et al. (2019). Low complexity regions in the proteins of prokaryotes perform important functional roles and are highly conserved. *Nucleic Acids Res.* 47, 9998–10009. doi: 10.1093/nar/gkz730
- Nurenberg-Goloub, E., Kratzat, H., Heinemann, H., Heuer, A., Kotter, P., Berninghausen, O., et al. (2020). Molecular analysis of the ribosome recycling factor ABCE1 bound to the 30S post-splitting complex. *EMBO J.* 39:e103788. doi: 10.15252/embj.2019103788
- O'Connor, M., Asai, T., Squires, C. L., and Dahlberg, A. E. (1999). Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA. *Proc. Natl. Acad. Sci. U. S. A.* 96, 8973–8978. doi: 10.1073/pnas.96.16.8973
- Pedulla, N., Palermo, R., Hasenöhr, D., Blasi, U., Cammarano, P., and Londei, P. (2005). The archaeal eIF2 homologue: functional properties of an ancient translation initiation factor. *Nucleic Acids Res.* 33, 1804–1812. doi: 10.1093/nar/gki321
- Penev, P. I., Fakhretaha-Aval, S., Patel, V. J., Cannone, J. J., Gutell, R. R., Petrov, A. S., et al. (2020). Eukaryotic-like ribosomal RNA in Asgard archaea. *bioRxiv* [Preprint]. doi: 10.1101/2019.12.25.888164
- Peske, F., Rodnina, M. V., and Wintermeyer, W. (2005). Sequence of steps in ribosome recycling as defined by kinetic analysis. *Mol. Cell* 18, 403–412. doi: 10.1016/j.molcel.2005.04.009
- Pestova, T. V., Borukhov, S. I., and Hellen, C. U. (1998). Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* 394, 854–859. doi: 10.1038/29703
- Pestova, T. V., Lomakin, I. B., Lee, J. H., Choi, S. K., Dever, T. E., and Hellen, C. U. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* 403, 332–335. doi: 10.1038/35002118
- Petrov, A. S., Gulen, B., Norris, A. M., Kovacs, N. A., Bernier, C. R., Lanier, K. A., et al. (2015). History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci. U. S. A.* 112, 15396–15401. doi: 10.1073/pnas.1509761112
- Plewniak, F., Bianchetti, L., Brelivet, Y., Carles, A., Chalmel, F., Lecompte, O., et al. (2003). PipeAlign: a new toolkit for protein family analysis. *Nucleic Acids Res.* 31, 3829–3832. doi: 10.1093/nar/gkg518
- Polikanov, Y. S., Melnikov, S. V., Soll, D., and Steitz, T. A. (2015). Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.* 22, 342–344. doi: 10.1038/nsmb.2992
- Resch, A., Tedin, K., Gründling, A., Mündlein, A., and Bläsi, U. (1996). Downstream box-anti-downstream box interactions are dispensable for translation initiation of leaderless mRNAs. *EMBO J.* 15, 4740–4748. doi: 10.1002/j.1460-2075.1996.tb00851.x
- Rhodin, M. H., and Dinman, J. D. (2011). An extensive network of information flow through the B1b/c intersubunit bridge of the yeast ribosome. *PLoS One* 6:e20048. doi: 10.1371/journal.pone.0020048
- Roberts, E., Sethi, A., Montoya, J., Woese, C. R., and Luthey-Schulten, Z. (2008). Molecular signatures of ribosomal evolution. *Proc. Natl. Acad. Sci. U. S. A.* 105, 13953–13958. doi: 10.1073/pnas.0804861105
- Rodnina, M. V. (2018). Translation in prokaryotes. *Cold Spring Harb. Perspect. Biol.* 10:a032664. doi: 10.1101/cshperspect.a032664
- Roll-Mecak, A., Alone, P., Cao, C., Dever, T. E., and Burley, S. K. (2004). X-ray structure of translation initiation factor eIF2gamma: implications for tRNA and eIF2alpha binding. *J. Biol. Chem.* 279, 10634–10642. doi: 10.1074/jbc.M310418200
- Roll-Mecak, A., Cao, C., Dever, T. E., and Burley, S. K. (2000). X-ray structures of the universal translation initiation factor IF2/eIF5B: conformational changes on GDP and GTP binding. *Cell* 103, 781–792. doi: 10.1016/S0092-8674(00)00181-1
- Ross, J., Bressler, K., and Thakor, N. (2018). Eukaryotic initiation factor 5B (eIF5B) cooperates with eIF1A and eIF5 to facilitate uORF2-mediated repression of ATF4 translation. *Int. J. Mol. Sci.* 19:4032. doi: 10.3390/ijms19124032
- Samhita, L., Shetty, S., and Varshney, U. (2012). Unconventional initiator tRNAs sustain *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 109, 13058–13063. doi: 10.1073/pnas.1207868109
- Sartorius-Neef, S., and Pfeifer, F. (2004). *In vivo* studies on putative Shine-Dalgarno sequences of the halophilic archaeon *Halobacterium salinarum*. *Mol. Microbiol.* 51, 579–588. doi: 10.1046/j.1365-2958.2003.03858.x
- Sas-Chen, A., Thomas, J. M., Matzov, D., Taoka, M., Nance, K. D., Nir, R., et al. (2020). Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature* 583, 638–643. doi: 10.1038/s41586-020-2418-2
- Schmitt, E., Blanquet, S., and Mechulam, Y. (2002). The large subunit of initiation factor aIF2 is a close structural homologue of elongation factors. *EMBO J.* 21, 1821–1832. doi: 10.1093/emboj/21.7.1821
- Schmitt, E., Coureux, P. D., Monestier, A., Dubiez, E., and Mechulam, Y. (2019). Start codon recognition in eukaryotic and archaeal translation initiation: a common structural core. *Int. J. Mol. Sci.* 20:939. doi: 10.3390/ijms20040939



- Schmitt, E., Naveau, M., and Mechulam, Y. (2010). Eukaryotic and archaeal translation initiation factor 2: a heterotrimeric tRNA carrier. *FEBS Lett.* 584, 405–412. doi: 10.1016/j.febslet.2009.11.002
- Schmitt, E., Panvert, M., Lazennec-Schurdevin, C., Coureux, P. D., Perez, J., Thompson, A., et al. (2012). Structure of the ternary initiation complex aIF2-GDPNP-methionylated initiator tRNA. *Nat. Struct. Mol. Biol.* 19, 450–454. doi: 10.1038/nsmb.2259
- Schutz, S., Michel, E., Damberger, F. F., Oplova, M., Pena, C., Leitner, A., et al. (2018). Molecular basis for disassembly of an importin:ribosomal protein complex by the escortin Tsr2. *Nat. Commun.* 9:3669. doi: 10.1038/s41467-018-06160-x
- Selmer, M., Dunham, C. M., Murphy, F. V. T., Weixlbaumer, A., Petry, S., Kelley, A. C., et al. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942. doi: 10.1126/science.1131127
- Shatsky, I. N., Terenin, I. M., Smirnova, V. V., and Andreev, D. E. (2018). Cap-independent translation: what's in a name? *Trends Biochem. Sci.* 43, 882–895. doi: 10.1016/j.tibs.2018.04.011
- She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., et al. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U. S. A.* 98, 7835–7840. doi: 10.1073/pnas.14122098
- Shell, S. S., Wang, J., Lapierre, P., Mir, M., Chase, M. R., Pyle, M. M., et al. (2015). Leaderless transcripts and small proteins are common features of the mycobacterial translational landscape. *PLoS Genet.* 11:e1005641. doi: 10.1371/journal.pgen.1005641
- Shetty, S., Shah, R. A., Chembazhi, U. V., Sah, S., and Varshney, U. (2017). Two highly conserved features of bacterial initiator tRNAs license them to pass through distinct checkpoints in translation initiation. *Nucleic Acids Res.* 45, 2040–2050. doi: 10.1093/nar/gkw854
- Shin, B. S., Acker, M. G., Kim, J. R., Maher, K. N., Arefin, S. M., Lorsch, J. R., et al. (2011). Structural integrity of {alpha}-helix H12 in translation initiation factor eIF5B is critical for 80S complex stability. *RNA* 17, 687–696. doi: 10.1261/rna.2412511
- Shin, B. S., Maag, D., Roll-Mecak, A., Arefin, M. S., Burley, S. K., Lorsch, J. R., et al. (2002). Uncoupling of initiation factor eIF5B/IF2 GTPase and translational activities by mutations that lower ribosome affinity. *Cell* 111, 1015–1025. doi: 10.1016/S0092-8674(02)01171-6
- Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U. S. A.* 71, 1342–1346. doi: 10.1073/pnas.71.4.1342
- Simonetti, A., Guca, E., Bochler, A., Kuhn, L., and Hashem, Y. (2020). Structural insights into the mammalian late-stage initiation complexes. *Cell Rep.* 31:107497. doi: 10.1016/j.celrep.2020.03.061
- Slupska, M. M., King, A. G., Fitz-Gibbon, S., Besemer, J., Borodovsky, M., and Miller, J. H. (2001). Leaderless transcripts of the crenarchaeal hyperthermophile *Pyrobaculum aerophilum*. *J. Mol. Biol.* 309, 347–360. doi: 10.1006/jmbi.2001.4669
- Smollett, K., Blombach, F., Reichelt, R., Thomm, M., and Werner, F. (2017). A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat. Microbiol.* 2:17021. doi: 10.1038/nmicrobiol.2017.21
- Sokabe, M., Yao, M., Sakai, N., Toya, S., and Tanaka, I. (2006). Structure of archaeal translational initiation factor 2 betagamma-GDP reveals significant conformational change of the beta-subunit and switch 1 region. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13016–13021. doi: 10.1073/pnas.0604165103
- Sprengart, M. L., Fuchs, E., and Porter, A. G. (1996). The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. *EMBO J.* 15, 665–674. doi: 10.1002/j.1460-2075.1996.tb00399.x
- Stolboushkina, E., Nikonov, S., Nikulin, A., Blasi, U., Manstein, D. J., Fedorov, R., et al. (2008). Crystal structure of the intact archaeal translation initiation factor 2 demonstrates very high conformational flexibility in the alpha- and beta-subunits. *J. Mol. Biol.* 382, 680–691. doi: 10.1016/j.jmb.2008.07.039
- Tedin, K., Moll, I., Grill, S., Resch, A., Graspöf, A., Gualerzi, C. O., et al. (1999). Translation initiation factor 3 antagonizes authentic start codon selection on leaderless mRNAs. *Mol. Microbiol.* 31, 67–77. doi: 10.1046/j.1365-2958.1999.01147.x
- Terenin, I. M., Dmitriev, S. E., Andreev, D. E., and Shatsky, I. N. (2008). Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat. Struct. Mol. Biol.* 15, 836–841. doi: 10.1038/nsmb.1445
- Thakor, N., and Holcik, M. (2012). IRES-mediated translation of cellular messenger RNA operates in eIF2alpha-independent manner during stress. *Nucleic Acids Res.* 40, 541–552. doi: 10.1093/nar/gkr701
- Toffano-Nioche, C., Ott, A., Crozat, E., Nguyen, A. N., Zytnecki, M., Leclerc, F., et al. (2013). RNA at 92 degrees C: the non-coding transcriptome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *RNA Biol.* 10, 1211–1220. doi: 10.4161/rna.25567
- Tolstrup, N., Sensen, C. W., Garrett, R. A., and Clausen, I. G. (2000). Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. *Extremophiles* 4, 175–179. doi: 10.1007/s007920070032
- Torarinsson, E., Klenk, H. P., and Garrett, R. A. (2005). Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.* 7, 47–54. doi: 10.1111/j.1462-2920.2004.00674.x
- Udagawa, T., Shimizu, Y., and Ueda, T. (2004). Evidence for the translation initiation of leaderless mRNAs by the intact 70 S ribosome without its dissociation into subunits in eubacteria. *J. Biol. Chem.* 279, 8539–8546. doi: 10.1074/jbc.M308784200
- Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A. C., Engelberg-Kulka, H., et al. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell* 147, 147–157. doi: 10.1016/j.cell.2011.07.047
- Wang, J., Johnson, A. G., Lapointe, C. P., Choi, J., Prabhakar, A., Chen, D. H., et al. (2019). eIF5B gates the transition from translation initiation to elongation. *Nature* 573, 605–608. doi: 10.1038/s41586-019-1561-0
- Weisser, M., Voigts-Hoffmann, F., Rabl, J., Leibundgut, M., and Ban, N. (2013). The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF1A. *Nat. Struct. Mol. Biol.* 20, 1015–1017. doi: 10.1038/nsmb.2622
- Williams, T. A., Foster, P. G., Cox, C. J., and Embley, T. M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504, 231–236. doi: 10.1038/nature12779
- Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271. doi: 10.1128/MMBR.51.2.221-271.1987
- Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5088–5090. doi: 10.1073/pnas.74.11.5088
- Woese, C. R., Gutell, R., Gupta, R., and Noller, H. F. (1983). Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.* 47, 621–669. doi: 10.1128/MMBR.47.4.621-669.1983
- Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576
- Wu, C. J., and Janssen, G. R. (1996). Translation of vph mRNA in *Streptomyces lividans* and *Escherichia coli* after removal of the 5' untranslated leader. *Mol. Microbiol.* 22, 339–355. doi: 10.1046/j.1365-2958.1996.00119.x
- Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B. A., and Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. *Genome Res.* 20, 133–141. doi: 10.1101/gr.100396.109
- Yamamoto, H., Unbehaun, A., Loerke, J., Behrmann, E., Collier, M., Burger, J., et al. (2014). Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. *Nat. Struct. Mol. Biol.* 21, 721–727. doi: 10.1038/nsmb.2859
- Yamamoto, H., Wittek, D., Gupta, R., Qin, B., Ueda, T., Krause, R., et al. (2016). 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 113, E1180–E1189. doi: 10.1073/pnas.1524554113
- Yatime, L., Mechulam, Y., Blanquet, S., and Schmitt, E. (2006). Structural switch of the gamma subunit in an archaeal aIF2 alpha gamma heterodimer. *Structure* 14, 119–128. doi: 10.1016/j.str.2005.09.020
- Yatime, L., Schmitt, E., Blanquet, S., and Mechulam, Y. (2004). Functional molecular mapping of archaeal translation initiation factor 2. *J. Biol. Chem.* 279, 15984–15993. doi: 10.1074/jbc.M311561200
- Yusupova, G., Jenner, L., Rees, B., Moras, D., and Yusupov, M. (2006). Structural basis for messenger RNA movement on the ribosome. *Nature* 444, 391–394. doi: 10.1038/nature05281
- Yutin, N., Puigbo, P., Koonin, E. V., and Wolf, Y. I. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PLoS One* 7:e36972. doi: 10.1371/journal.pone.0036972
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Backstrom, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358. doi: 10.1038/nature21031

- Zavialov, A. V., Hauryliuk, V. V., and Ehrenberg, M. (2005). Splitting of the posttermination ribosome into subunits by the concerted action of RRF and EF-G. *Mol. Cell* 18, 675–686. doi: 10.1016/j.molcel.2005.05.016
- Zheng, A., Yu, J., Yamamoto, R., Ose, T., Tanaka, I., and Yao, M. (2014). X-ray structures of eIF5B and the eIF5B-eIF1A complex: the conformational flexibility of eIF5B is restricted on the ribosome by interaction with eIF1A. *Acta Crystallogr. D Biol. Crystallogr.* 70, 3090–3098. doi: 10.1107/S1399004714021476
- Zhou, J., Lancaster, L., Donohue, J. P., and Noller, H. F. (2013). Crystal structures of EF-G-ribosome complexes trapped in intermediate states of translocation. *Science* 340:1236086. doi: 10.1126/science.1236086

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Schmitt, Coureux, Kazan, Bourgeois, Lazenec-Schurdevin and Mechulam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.