ARCHAEAL RIBOSOMES: BIOGENESIS, STRUCTURE AND FUNCTION

EDITED BY: Paola Londei, Anna La Teana and Sébastien Ferreira-Cerca
PUBLISHED IN: Frontiers in Microbiology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-141-0
DOI 10.3389/978-2-88974-141-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ARCHAEAL RIBOSOMES: BIOGENESIS, STRUCTURE AND FUNCTION

Topic Editors:
Paola Londei, Sapienza University of Rome, Italy
Anna La Teana, Polytechnic University of Marche, Italy
Sébastien Ferreira-Cerca, University of Regensburg, Germany

Citation: Londei, P., La Teana. A., Ferreira-Cerca, S., eds. (2022). Archaeal Ribosomes: Biogenesis, Structure and Function. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-141-0

Table of Contents

frontiers in Microbiology

EDITORIAL
published: 07 December 2021
doi: 10.3389/fmicb.2021.800052

Editorial: Archaeal Ribosomes: Biogenesis, Structure and Function

Sébastien Ferreira-Cerca[1]*, *Anna La Teana*[2]* and *Paola Londei*[3]*

[1] Biochemistry III-Regensburg Center for Biochemistry, Institute for Biochemistry, Genetics and Microbiology, University of Regensburg, Regensburg, Germany, [2] Department of Life and Environmental Science, New York-Marche Structural Biology Center (NY-MaSBiC), Polytechnic University of Marche, Ancona, Italy, [3] Department Molecular Medicine, Policlinico Umberto I° Viale Regina, University of Rome Sapienza, Rome, Italy

Keywords: ribosome, archaea, translation, ribosome biogenesis, RNA modification

Editorial on the Research Topic

Archaeal Ribosomes: Biogenesis, Structure and Function

At the end of the 1970s, Carl Woese identified microorganisms that belong to a separate domain of life, the archaea (Fox et al., 1977; Albers et al., 2013). In the subsequent years, in-depth studies of the molecular and cellular biology of archaea, as well as of their physiology and environmental distribution, have revealed that archaea possess some very unique and distinctive biological traits (Albers et al., 2013; Tahon et al., 2021).

Such unique archaeal traits are also observed in fundamental cellular mechanisms like translation and ribosome biogenesis. Despite the universality of the translation process performed by the ribosome, significant molecular and functional similarities, as well as differences, underlying the mechanism of translation and ribosome's biology have been established across the tree of life.

Due to the central position of the translation machinery for gene expression and regulation, unraveling details of ribosome biogenesis and structure, as well as understanding its role in protein homeostasis is an emerging topic. Notably, and in contrast to bacteria and eukaryotes where a wealth of information is available, little is known as yet about the biogenesis and function of archaeal ribosomes.

The present Research Topic "*Archaeal Ribosomes: Biogenesis, Structure, and Function*" covers fundamental aspects of archaeal ribosome biology, from biogenesis to function.

In the first part of this Research Topic, three reviews and three original articles provide key insights into ribosome biogenesis in Archaea.

Londei and Ferreira-Cerca give an overview of ribosome biogenesis in archaea and highlight the similarities and differences with respect to the paradigm ribosome biogenesis pathways determined in bacterial and eukaryotic model organisms. Furthermore, Czekay and Kothe and Breuer et al. provide an in-depth summary on our knowledge of RNA-guided ribosomal RNA modifications, which require molecular machineries uniquely shared between archaea and eukaryotes, and discuss possible functional implications for ribosome biogenesis and function.

Ribosome biogenesis requires the action of trans-acting factors, known as ribosome biogenesis or assembly factors. However, little is known about putative archaeal ribosome biogenesis factors and their conservation across archaea. Birikmen et al. provide an in-depth bioinformatic analysis, and a state-of-the-art overview of ribosome biogenesis factors conservation across archaea and eukaryotes. Their study describes a core set of putative ribosome biogenesis factors widely conserved among archaea that may contribute to ribosome synthesis, although this is still to be proven experimentally. Notably, this study also reveals that the expansion of ribosome biogenesis factors characteristic of the eukaryotic pathway is not present in the Asgard archaea phylum which has been proposed to be more closely related to eukaryotes (Tahon et al., 2021). Altogether,

Edited and reviewed by:
Simonetta Gribaldo, Institut Pasteur, France

***Correspondence:**
Sébastien Ferreira-Cerca
sebastien.ferreira-cerca@ur.de
orcid.org/0000-0002-0522-843X
Anna La Teana
a.lateana@univpm.it
Paola Londei
paola.londei@uniroma1.it

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

Received: *22 October 2021*
Accepted: *09 November 2021*
Published: *07 December 2021*

Citation:
Ferreira-Cerca S, La Teana A and Londei P (2021) Editorial: Archaeal Ribosomes: Biogenesis, Structure and Function. Front. Microbiol. 12:800052. doi: 10.3389/fmicb.2021.800052

this comprehensive study provides an important resource that will further help to functionally explore the archaeal ribosome biogenesis pathway and its evolution.

Detailed data on *in vivo* functional characterization of ribosome biogenesis in archaea remain relatively scarce. Using a recently developed gene repression system based on CRISPR technology (Stachler and Marchfelder, 2016), Schwarz et al. provide new insights into ribosomal RNA maturation and the *in vivo* role of the tRNA splicing endonuclease. The latter, in addition to its classical function in tRNA maturation, is also required for ribosomal RNA maturation *in vivo*, in agreement with recent *in vitro* data (Qi et al., 2020). This study further supports the functional requirement of precursor rRNA circularization, a specific characteristic of most archaeal ribosome biogenesis pathways analyzed thus far (Tang et al., 2002; Danan et al., 2012; Jüttner et al., 2020; Qi et al., 2020).

Finally, the study of Birkedal et al. also provides new insight into ribosome biogenesis diversity. Indeed, this work reveals a non-canonical 23S rRNA maturation pathway and describes the first known example of circularly permuted ribosomal RNA. This study also highlights the fact that ribosome biogenesis in archaea might be more diverse than so far anticipated.

Overall, there is still much to learn from the archaeal world, and much room for new surprising and exciting discoveries that will have key implications for our general understanding of ribosome biogenesis and evolution.

In the second part of this Research Topic, ribosome function is the focus of three reviews and two original research articles.

An emerging topic is the presence of transcription-translation coupling in archaea. Coupled transcription-translation has been proposed as a general mechanism of gene expression regulation in bacteria and archaea (Irastortza-Olaziregi and Amster-Choder, 2021); however, evidence for this coupling remains limited to a few archaeal species (French et al., 2007). Based on recent discoveries in bacteria and archaea, Weixlbaumer et al.'s perspective article explores the possible molecular basis of functional transcription-translation coupling in archaea. Furthermore, by using a systematic co-purification approach in *Haloferax volcanii*, Schramm et al. provide additional evidence for transcription-translation coupling in this cellular context, as well as additional insights into the interactions among translation initiation factors as well as among initiation factors and RNA polymerase components. Obviously, many future studies will be required to gain a better and wider understanding of the general prevalence of transcription-translation coupling in archaea, and on its molecular mechanism and functional advantages.

It has been previously established that translation initiation shares similar molecular features in archaea and eukaryotes. However, the structural and functional characteristics of assembly/disassembly of the pre-initiation and initiation complexes in archaea remain to be fully explored. Schmitt et al. provide an overview of the recent structural and functional advances on archaeal translation initiation. This summary is also complemented by additional insights from Lo Gullo et al. describing the release from the ribosome of the translation initiation factor aIF6 by the action of translation elongation factor EF-2.

Finally, De Lise et al. offer an overview on specific deviations from the standard ribosome decoding that have been identified in archaea.

In summary, the present Research Topic not only provides a state-of-the-art summary on our understanding of ribosome biogenesis and function in archaea, but also offers new perspectives and insights in this still emerging field, thus contributing to inspire and orient future research.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

The topic editors wish to thank all authors for providing significant contributions to the Research Topic. Moreover, all participating reviewers are deeply acknowledged for their generous time and insightful suggestions and comments. Finally, we like to thank all the team at Frontiers in Microbiology for their constant support and constructive suggestions.

REFERENCES

Albers, S.-V., Forterre, P., Prangishvili, D., and Schleper, C. (2013). The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. *Nat. Rev. Micro* 11, 713–719. doi: 10.1038/nrmicro3124

Danan, M., Schwartz, S., Edelheit, S., and Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* 40, 3131–3142. doi: 10.1093/nar/gkr1009

Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S., and Woese, C. R. (1977). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. U.S.A.* 74, 4537–4541. doi: 10.1073/pnas.74.10.4537

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

Irastortza-Olaziregi, M., and Amster-Choder, O. (2021). Coupled transcription-translation in prokaryotes: an old couple with new surprises. *Front. Microbiol.* 11:3532. doi: 10.3389/fmicb.2020.624830

Jüttner, M., Weiß, M., Ostheimer, N., Reglin, C., Kern, M., Knüppel, R., et al. (2020). A versatile cis-acting element reporter system to study the function, maturation and stability of ribosomal RNA mutants in archaea. *Nucleic Acids Res.* 48, 2073–2090. doi: 10.1093/nar/gkz1156

Qi, L., Li, J., Jia, J., Yue, L., and Dong, X. (2020). Comprehensive analysis of the pre-ribosomal RNA maturation pathway in a methanoarchaeon exposes the conserved circularization and linearization mode in archaea. *RNA Biol.* 17, 1427–1441. doi: 10.1080/15476286.2020.1771946

Stachler, A.-E., and Marchfelder, A. (2016). Gene repression in haloarchaea using the CRISPR (Clustered regularly interspaced short palindromic repeats)-Cas I-B system. *J. Biol. Chem.* 291, 15226–15242. doi: 10.1074/jbc.M116.724062

Tahon, G., Patricia, G., and Ettema, T. J. G. (2021). Expanding archaeal diversity and phylogeny: past, present, and future. *Annu. Rev. Microbiol.* 75, 359–381. doi: 10.1146/annurev-micro-040921-050212

Tang, T. H., Rozhdestvensky, T. S., d'Orval, B. C., Bortolin, M.-L., Huber, H., Charpentier, B., et al. (2002). RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res.* 30, 921–930. doi: 10.1093/nar/30.4.921

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

REVIEW
published: 18 September 2020
doi: 10.3389/fmicb.2020.584152

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

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

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

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

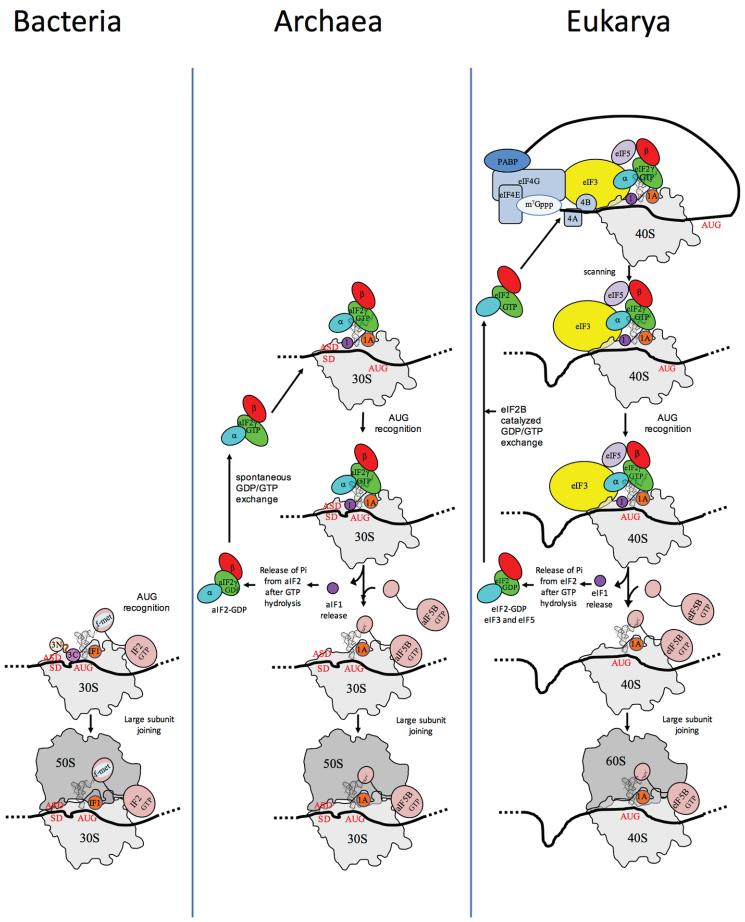


FIGURE 1 | Schematic views of the translation initiation (TI) steps in the three domains of life. The figure illustrates the main steps in bacteria (left), in archaea (middle), and in eukarya (right). Bacterial 30S subunit recruits the messenger RNA (mRNA), often due to the base pairing between a Shine-Dalgarno 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/aIF1A, e/aIF1, e/aIF2, and e/aIF5B (see also **Figure 2**). e/aIF2 heterotrimer is represented with a three-color code (α subunit in cyan, β subunit in red, and γ 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 aIF1, aIF1A, aIF2, and aIF5B 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/aIF1A (or IF1 in bacteria) and e/aIF5B (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 maturated with a m^7G-cap at the 5′ end and a polyadenylated tail at the 3′ end. The cap-dependent canonical translation initiation model involves a pre-initiation complex (43S PIC) containing the SSU, the ternary complex eIF2-GTP-Met-$tRNA_i^{Met}$, 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

Abbreviations: TI, Translation initiation; SSU, Small ribosomal subunit; LSU, Large ribosomal subunit; IC, Initiation complex; TC, Ternary complex e/aIF2:GTP:Met-$tRNA_i^{Met}$; LUCA, Last universal common ancestor.

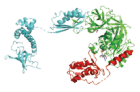
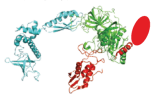
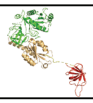
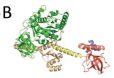
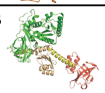
Bacteria	Archaea	Eukarya	Function
-	aIF2 (α,β,γ) 	eIF2 (α,β,γ) 	Binds tRNAi, GTP hydrolysis and releases Pi coupled to start codon recognition.
~IF3* 	aIF1 	eIF1 	Participates in accuracy of start codon selection.
IF1 	aIF1A 	eIF1A 	Participates in accuracy of start codon selection.
IF2 	aIF5B 	eIF5B 	Binds tRNAi, facilitates joining of the large ribosomal subunit
-	-	eIF5	Accelerates hydrolysis of GTP by eIF2 (GTPase activating protein, GAP).
-	aIF2B (α,β,δ)*	eIF2B (α,β,γ,δ,ε)	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 eIF2β 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 α–β 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 Ficner, 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 γ and ε subunits of eIF2B are missing in archaea. The function of the eIF2B α, β, δ 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$_i^{Met}$), 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 ARCHAEAL 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)	21 (15u, 6b)
		50S	23S (2891)	33 (18u, 15b)
			5S (117)	
Archaea	70S	30S	16S (1483)	25 (15u, 9e, 1a)
		50S	23S (2967)	39 (18u, 20e)
			5S(122)	
Eukarya	80S	40S	18S (1860)	33 (15u, 18e)
		60S	28S (4039)	46 (18u, 28e)
			5S (120)	
			5.8S (158; *S. cerevisiae*)	

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).

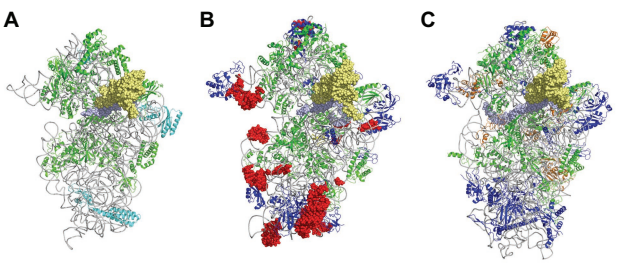


FIGURE 3 | The small ribosomal subunit in the three domains of life. **(A)** 30S from *Thermus thermophilus* (PDB 5LMV; Hussain et al., 2016). **(B)** 30S from *Pyrococcus abyssi* (PDB 6SWC; Coureux et al., 2020). **(C)** 40S from *Kluyveromyces lactis* (PDB 6FYX; Llacer et al., 2018). Ribosomal proteins are colored as follows; universal in green, bacterial in cyan, eukaryotic and archaeal in dark blue, and eukaryotic only (as compared to *P. abyssi*) in orange. The P site tRNA is in yellow spheres and the mRNA in light blue spheres. rRNA is in gray. Regions of the ribosomal RNA of the *P. abyssi* small subunit that are not in the common core as defined in Bernier et al. (2018) are shown with red spheres (middle view, **Table 2**).

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.rna.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; Torarinsson 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), *Methanolobus psychrophilus* (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 crenarchaeaota, 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 Lokiarchaeaote 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 (≤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

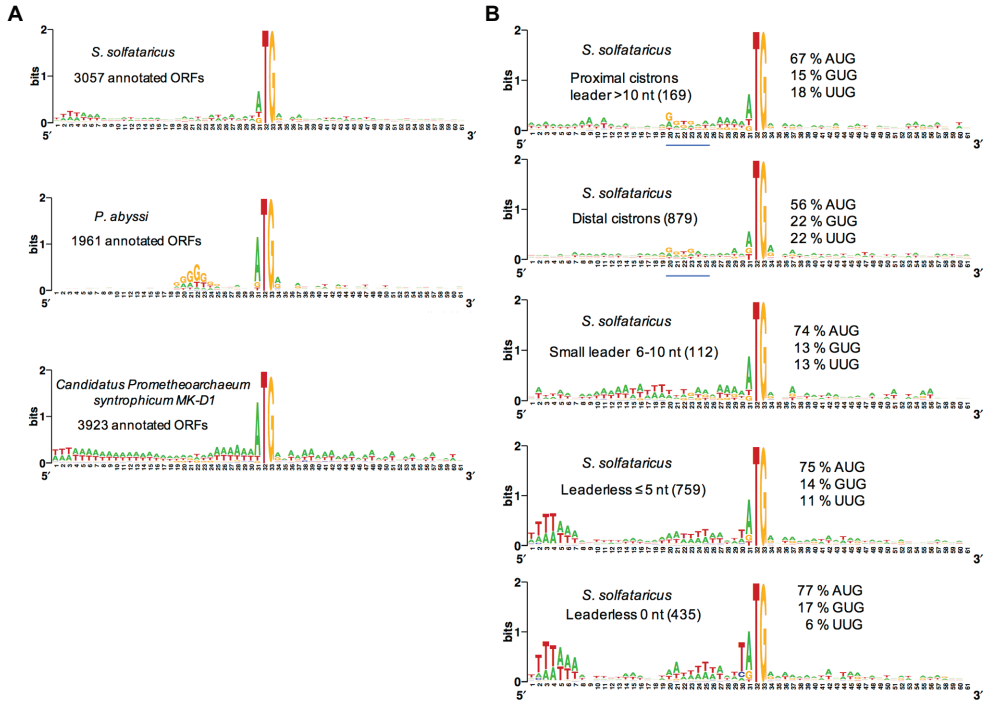


FIGURE 4 | Analysis of translation start regions. **(A)** Analysis of the translation start regions in *Sulfolobus solfataricus*, *P. abyssi*, and the Asgard *Candidatus Prometeoarchaeum syntrophicum*. DNA sequences (60 nt around the first base of the start codon) were extracted from the genomic sequences (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[1]). 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 aIF2:GTP:Met-tRNA$_i^{Met}$ and the two small initiation factors aIF1 and aIF1A 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 biscistronic 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).

[1]http://www.rna.icmb.utexas.edu/

The GGAGGUCA SD motif of the *gvp*H 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) UUUGGAGGUGAUUUAAA(+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 $^{5'}$AUCACCUCC$^{3'}$ 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

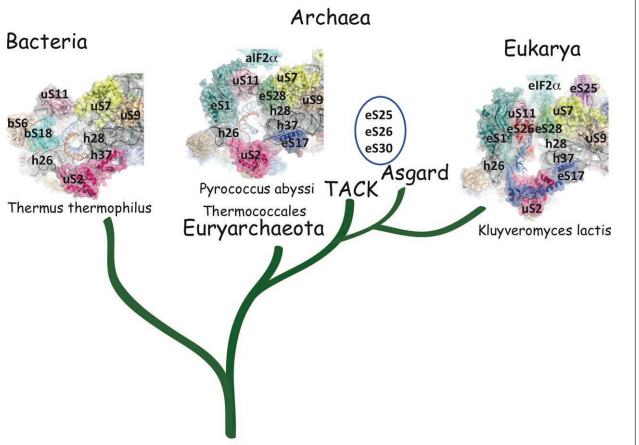


FIGURE 5 | Comparison of the mRNA exit channels in the three domains of life. Surface representations of the mRNA exit channels of representative structures in the three domains of life. The mRNA is shown in blue and the 3'extremity of the rRNA is shown in orange. R-proteins are labeled using the Ban 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 α subunit of eIF2. Such an interaction was not observed with the homologous protein aIF2α 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$_i^{Met}$ in a GTP dependent manner (Yatime et al., 2004, 2006; Pedulla et al., 2005; Sokabe et al., 2006; Stolboushkina et al., 2008). γ is the core subunit that binds GTP (Schmitt et al., 2002; Dubiez et al., 2015). α and β are bound to γ 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 α and the domains I and II of γ, while the aIF2β 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$_{REMOTE}$, the anticodon stem-loop of the Met-tRNA$_i^{Met}$ is out of the P site. aIF2γ 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 aIF2γ. In the second conformation, called IC1-P$_{IN}$, the anticodon stem-loop of the Met-tRNA$_i^{Met}$ is bound to the P site, while the position of aIF2γ 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$_{REMOTE}$ and IC1-P$_{IN}$ 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 aIF2γ are lost. Both these h44 movements and the release of contacts between aIF1 and aIF2γ 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 aIF2γ 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.

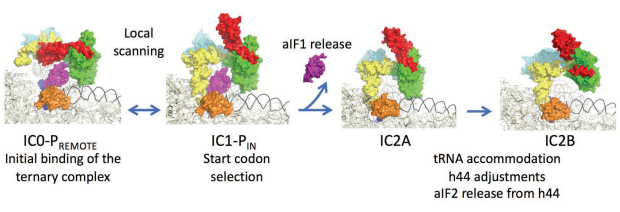


FIGURE 6 | Steps of translation initiation in *P. abyssi*. Surface representation of successive *P. abyssi* translation initiation complexes. aIF2, aIF1, and aIF1A are shown with the same color code as in **Figure 2**. mRNA is in dark blue, initiator tRNA is in yellow, and the h44 helix is in black. The figure shows start codon selection in the full archaeal TI complex where IC0-P$_{REMOTE}$ and IC1-P$_{IN}$ are in equilibrium until a start codon is found in the P site (Coureux et al., 2016). Codon:anticodon pairing stabilizes the IC1 state and triggers aIF1 release. In IC2A, the initiator tRNA fully accommodates. Release of aIF1 would cause both Pi release from aIF2γ and h44 adjustments leading to irreversible aIF2 release. The figure is adapted from Coureux et al. (2020).

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.

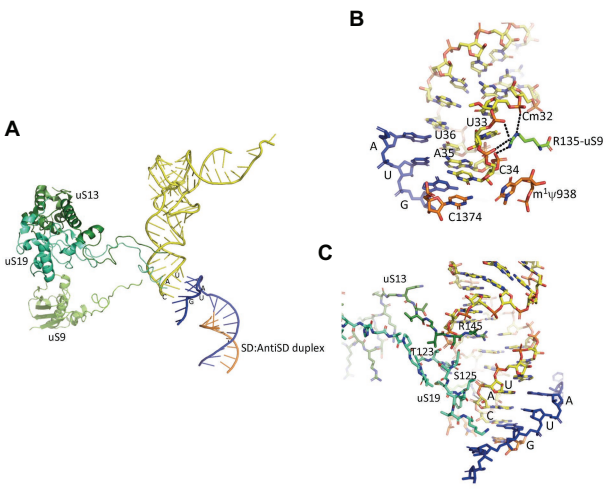


FIGURE 7 | Interaction of the initiator tRNA at the P site with the universal proteins uS9, uS13, and uS19. **(A)** Overall view of the accommodated tRNA as observed in the IC2 complex from *P. abyssi* (Coureux et al., 2016). The color code is the same as in **Figure 6**. **(B)** Close up showing the interaction of the strictly conserved terminal arginine of uS9 with the codon:anticodon duplex. **(C)** Close up showing the interaction of the uS13 and uS19 tails with the major groove of the anticodon stem-loop of the initiator tRNA.

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^3U1467 (m^3U1498, *E. coli* numbering) and the two dimethyladenosines m_2^6A1487,

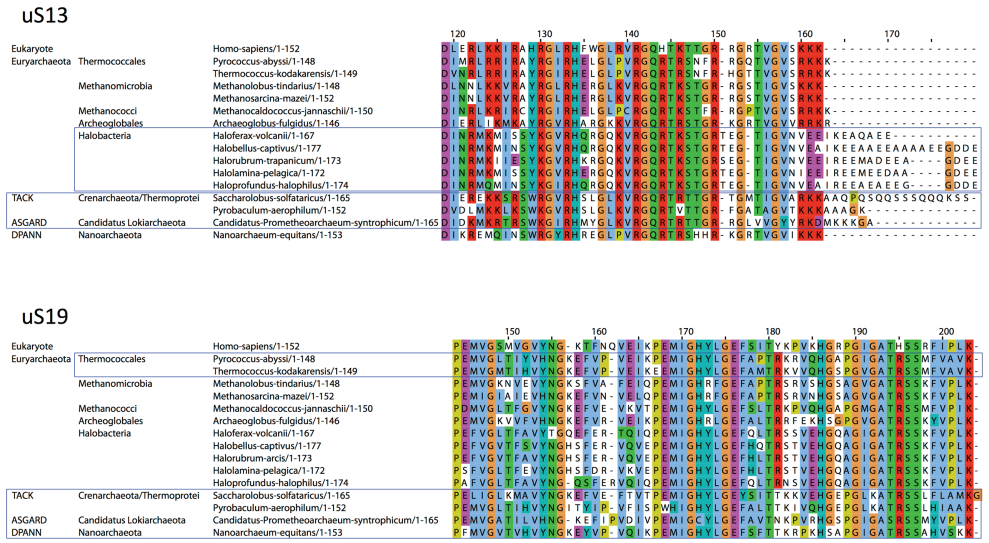


FIGURE 8 | Sequence alignments of the uS13 and uS19 C-terminal tails. uS13 and uS19 sequences were extracted and aligned using Pipealign (Plewniak et al., 2003). After visual inspection, several families regarding the C-terminal tail specificities in Archaea were identified. For uS13, ca. 250 archaeal 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.

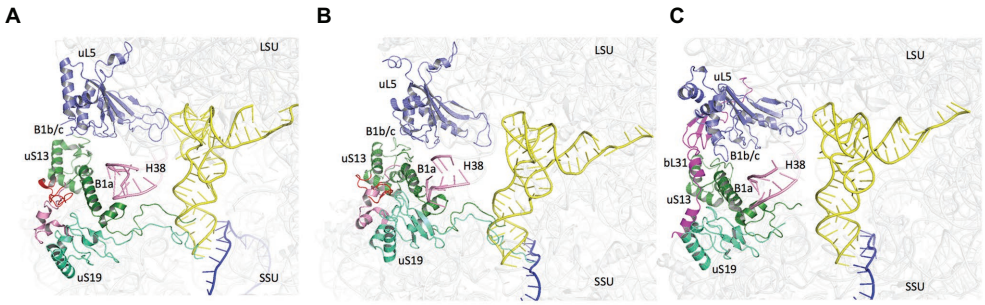


FIGURE 9 | Domain specificities of uS13 and uS19 cores. The three panels show the B1a-B1b/c bridge. **(A)** Archaeal case. The view is a composite using the SSU from PDB 6SWC (Coureux et al., 2020) and the LSU from PDB 4V6U (Armache et al., 2013). **(B)** Structure of human ribosome from PDB 6Y0G (Bhaskar 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 $m^6{}_2$A1487, $m^6{}_2$1488, 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 α-helix (helix h12) to domain IV (**Figure 2**; Roll-Mecak et al., 2004). Domain IV contains a β-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 learderless 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 dowstream 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 λ 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 RRF, i.e., under conditions where 70S ribosomes were abundant, translation of leaderless mRNAs was maintained whereas that of SD-leadered 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$_f^{Met}$, 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 (Kaberdina 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$_i^{Met}$ 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 γ 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 leardeless. 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., Vishveshwara, 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 A 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 H2-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-tRNAiMet 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

Coureux, 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

Coureux, 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 eIF2alpha and eIF2Balpha: 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 A 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 aIF2Bbeta and aIF2Bdelta. *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 Bläsi, 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 $tRNA^{Met}$ 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 e/aIF5B, 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 Blasi, U. (2008). Translation initiation factor a/eIF2(-γ) 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, archaeabacterial, 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

Kaberdina, 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., Spåhr, 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 *Methanolobus 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., Mattossovich, 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.fmrre.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 increases 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 Blasi, 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 Blasi, 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., Lazennec-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., Hasenöhrl, 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.141222098

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., Graschopf, 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., Zytnicki, 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 November 2020
doi: 10.3389/fmicb.2020.594838

Splicing Endonuclease Is an Important Player in rRNA and tRNA Maturation in Archaea

*Thandi S. Schwarz[1], Sarah J. Berkemer[2,3,4], Stephan H. Bernhart[2,5], Matthias Weiß[6†], Sébastien Ferreira-Cerca[6], Peter F. Stadler[2,3,5,7,8,9,10,11,12] and Anita Marchfelder[1]**

[1] Biologie II, Ulm University, Ulm, Germany, [2] Bioinformatics, Department of Computer Science, Leipzig University, Leipzig, Germany, [3] Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany, [4] Competence Center for Scalable Data Services and Solutions, Leipzig University, Leipzig, Germany, [5] Interdisciplinary Center for Bioinformatics, Leipzig University, Leipzig, Germany, [6] Regensburg Center for Biochemistry, Biochemistry III – Institute of Biochemistry, Genetics and Microbiology, University of Regensburg, Regensburg, Germany, [7] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, [8] Research Center for Civilization Diseases, Leipzig University, Leipzig, Germany, [9] Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia, [10] Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria, [11] Center for RNA in Technology and Health, University of Copenhagen, Frederiksberg, Denmark, [12] Santa Fe Institute, Santa Fe, NM, United States

OPEN ACCESS

Edited by:
Francisco Rodriguez-valera, Miguel Hernández University of Elche, Spain

Reviewed by:
Yulong Shen, Shandong University, Qingdao, China
Ruben Michael Ceballos, University of Arkansas, United States

***Correspondence:**
Anita Marchfelder
anita.marchfelder@uni-ulm.de

†Present address:
Matthias Weiß, Institute of Epigenetics and Stem Cells, Helmholtz Center Munich, Munich, Germany

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

Received: *14 August 2020*
Accepted: *21 October 2020*
Published: *20 November 2020*

Citation:
Schwarz TS, Berkemer SJ, Bernhart SH, Weiß M, Ferreira-Cerca S, Stadler PF and Marchfelder A (2020) Splicing Endonuclease Is an Important Player in rRNA and tRNA Maturation in Archaea. Front. Microbiol. 11:594838. doi: 10.3389/fmicb.2020.594838

In all three domains of life, tRNA genes contain introns that must be removed to yield functional tRNA. In archaea and eukarya, the first step of this process is catalyzed by a splicing endonuclease. The consensus structure recognized by the splicing endonuclease is a bulge-helix-bulge (BHB) motif which is also found in rRNA precursors. So far, a systematic analysis to identify all biological substrates of the splicing endonuclease has not been carried out. In this study, we employed CRISPRi to repress expression of the splicing endonuclease in the archaeon *Haloferax volcanii* to identify all substrates of this enzyme. Expression of the splicing endonuclease was reduced to 1% of its normal level, resulting in a significant extension of lag phase in *H. volcanii* growth. In the repression strain, 41 genes were down-regulated and 102 were up-regulated. As an additional approach in identifying new substrates of the splicing endonuclease, we isolated and sequenced circular RNAs, which identified excised introns removed from tRNA and rRNA precursors as well as from the 5′ UTR of the gene HVO_1309. *In vitro* processing assays showed that the BHB sites in the 5′ UTR of HVO_1309 and in a 16S rRNA-like precursor are processed by the recombinant splicing endonuclease. The splicing endonuclease is therefore an important player in RNA maturation in archaea.

Keywords: splicing endonuclease, tRNA intron, rRNA intron, tRNA processing, rRNA processing

INTRODUCTION

tRNA molecules are key players within cells since they translate genetic information into protein. Generation of functional tRNA molecules requires a plethora of processing steps starting with the removal of 5′ leader and 3′ trailer sequences from pre-tRNA [for a review see (Clouet-d'Orval et al., 2018; **Figure 1**)]. Some tRNA genes contain introns that must also be removed from precursor RNA to yield mature functional tRNAs.

While some bacterial tRNA genes contain self-splicing group I introns, archaeal and eukaryotic tRNA introns are removed by proteins. The initial step of intron removal in eukaryotes and archaea is catalyzed by an RNA splicing endonuclease. The resulting splice products are ligated by a tRNA ligase, thereby generating mature tRNA as well as a circular intron (**Figure 1**).

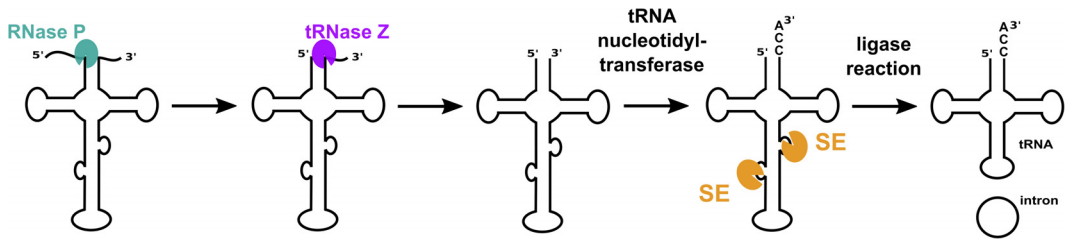


FIGURE 1 | tRNA maturation in archaea. Functional mature tRNAs are obtained after several processing steps. The 5′ leader and 3′ trailer sequences are removed by the endonucleases RNase P and tRNase Z. A terminal CCA triplet is added to the 3′ end by tRNA nucleotidyl transferase. Introns are removed by a splicing endonuclease (SE), exons are ligated, and the intron is circularized by an RNA ligase. The order of the processing steps can differ from organism to organism.

In some organisms, (e.g., *S. cerevisiae*) the enzyme 2′-phosphotransferase is required to remove the 2′ phosphate at the ligation junction to yield mature tRNA (Steiger et al., 2005). Structures for archaeal splicing endonucleases have been determined. Four different quaternary structures have emerged: an α_4 homotetramer from *Methanocaldococcus jannaschii* (Lykke-Andersen and Garrett, 1997; Li et al., 1998), an α_2 homodimer from *H. volcanii* (Kleman-Leyer et al., 1997), a $(\alpha\beta)_2$ dimer of heterodimers from *Nanoarchaeum equitans* (Mitchell et al., 2009) and, more recently, an irregular hexamer, ε_2 from Candidatus *Microarchaeum acidiphilum* (Fujishima et al., 2011). Each of the two ε substructures of the ε_2 complex consists of three proteins: a catalytic α subunit, a β subunit and a pseudocatalytic α subunit, denoted as α^P, forming an irregular homodimeric "six-unit" architecture [$\varepsilon = (\alpha^P\text{-}\alpha\text{-}\beta)$].

Despite structural differences, all archaeal splicing endonucleases recognize a conserved structure, referred to as a bulge-helix-bulge motif (BHB), which features a four-nucleotide helix flanked by two bulges, each comprised of three nucleotides (**Figure 2**; Marck and Grosjean, 2003; Fujishima et al., 2011). The heterodimeric $(\alpha\beta)_2$ form of the splicing endonuclease which is found mainly in the TACK[1] superphylum accepts a broader substrate spectrum with BHB motifs that contain various bulge lengths (Watanabe et al., 2002; Clouet-d'Orval et al., 2018). Notably, non-canonical BHB motifs are prevalent in the crenarchaea (Marck and Grosjean, 2003; Fujishima et al., 2011). Similarly, the splicing endonuclease ε_2 found in ARMAN[2] archaea has a broad substrate spectrum (Fujishima et al., 2011; Clouet-d'Orval et al., 2018). The BHB motif is usually found in the tRNA anticodon arm but can also be present in other locations in the tRNA (Marck and Grosjean, 2003). In addition, the BHB motif was found in rRNAs (Kjems and Garrett, 1991; Qi et al., 2020) and in one mRNA (Watanabe et al., 2002). In another study (Tang et al., 2002), it was suggested that circular 16S and 23S pre-rRNAs may be generated upon cleavage at the BHB site and subsequent ligation of exposed ends. Indeed, the presence of circular pre-rRNAs was confirmed in subsequent studies (Danan et al., 2012; Jüttner et al., 2020); however the enzyme catalyzing the reaction *in vivo* was unknown. More recently, using *in vitro* methods, it was shown that a splicing endonuclease catalyses this reaction in methanoarchaea (Qi et al., 2020).

Haloferax volcanii is a halophilic archaeon belonging to the kingdom euryarchaeal. *H. volcanii* contains a α_2 type splicing endonuclease, which strictly requires a hBHBh′[3] motif and will not cleave any other BHB (Thompson and Daniels, 1990). The *Haloferax* genome encodes three intron containing tRNA genes with intron lengths of 31 to 103 nucleotides: $tRNA^{Trp}_{CCA}$: 103, $tRNA^{Met}_{CAT}$: 75, $tRNA^{Gln}_{TTG}$: 31, respectively. Moreover, the precursor RNAs for 16S rRNA and 23S rRNAs both contain a BHB motif (**Figure 2**). Splicing endonuclease processing at these BHB motifs would release the rRNAs from the multicistronic pre-rRNA. The *Haloferax* splicing endonuclease has been investigated in detail and was shown to be active at low salt concentrations *in vitro* (Thompson and Daniels, 1990). This is surprising since *Haloferax* contains high intracellular salt concentrations (Mullakhanbhai and Larsen, 1975; Oren, 2008). Substrate specificities for the *Haloferax* endonuclease have also been reported (Thompson and Daniels, 1990; Armbruster and Daniels, 1997). A potential cellular function for the excised intron was investigated for $tRNA^{Trp}$ in *Haloferax* (Clouet d'Orval et al., 2001; Singh et al., 2004). The intron for this tRNA contains a box C/D sRNA that is required to direct 2′-*O*-methylation of nucleotides C34 and U39 in $tRNA^{Trp}$. Removal of the intron part of the tRNA gene has no effect on cell viability under standard conditions, suggesting that the intron is not maintained due to the sRNA (Joardar et al., 2008). To date, no other systematic analyses have been performed *in vivo* to identify all cellular substrates and functions of the archaeal splicing endonucleases. Moreover, very few ribonucleases have been functionally characterized in archaea. The archaeal tRNA 3′ end processing enzyme tRNase Z has been identified (Späth et al., 2008) and shown to catalyze not only tRNA 3′ processing but also 5′ end maturation of the 5S rRNA (Hölzle et al., 2008). To investigate whether the splicing endonuclease also processes substrates other than tRNAs, we set out to determine all biological substrates of the splicing endonuclease in *Haloferax*. Here, we use a systematic analysis of splicing endonuclease substrates *in vivo* to show that, in addition to the known tRNA substrates, one additional mRNA and the predicted rRNA substrates are cleaved

[1]The TACK superphylum encompasses Thaumarchaeota, Aigarchaeota, Crenarchaeota Bathyarchaeota and Korarchaeota.

[2]ARMAN: Archaeal Richmond Mine Acidophilic Nanoorganism.

[3]hBHBh′: a canonical BHB splicing motif with the two outer helices having at least two Watson-Crick base pairs; h, the helix closing the 3-nt bulge on the exonic side; h′, the helix closing the 3-nt bulge on the intronic side.

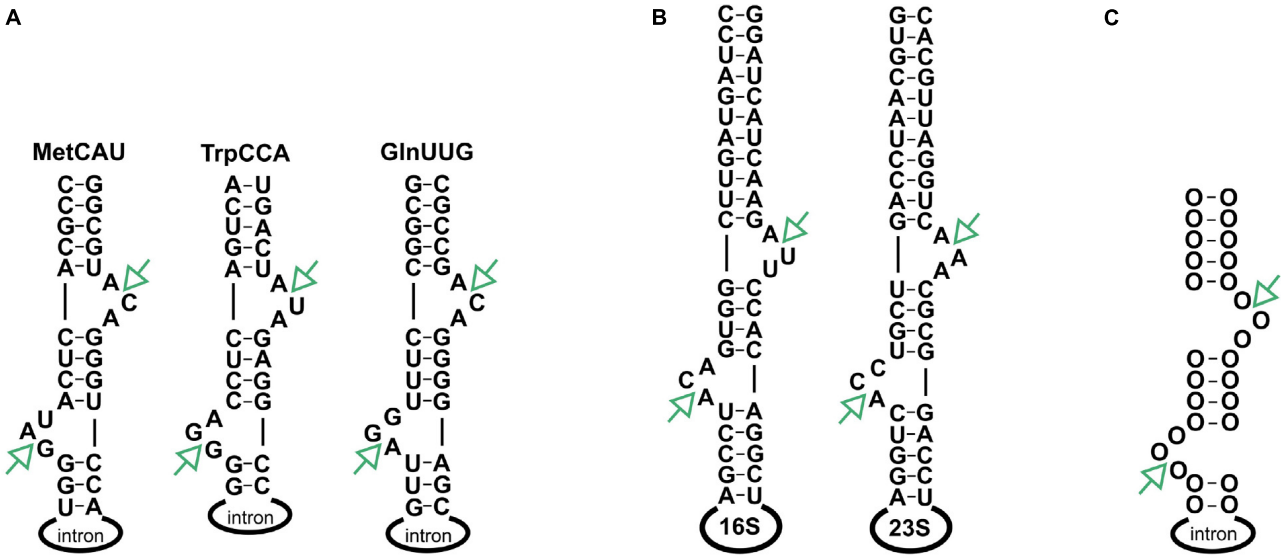


FIGURE 2 | Location of BHB motifs. **(A)** Three tRNAs contain an intron with the hBHBh′ motif forming at the exon-intron boundary. **(B)** hBHBh′ motifs are also flanking the 16S and 23S rRNAs. **(C)** General structure of the hBHBh′ motif. SE processing sites are indicated by green arrows.

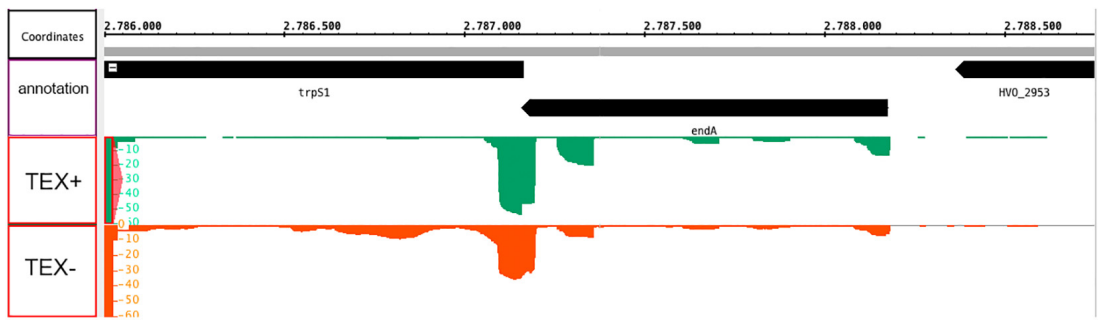


FIGURE 3 | Genomic location of the *endA* gene and transcription start site (TSS) analysis. The *endA* gene is encoded on the minus strand and overlaps the downstream *trpS1* gene by four nucleotides. dRNA-Seq analysis was performed to identify TSSs in *H. volcanii* (Babski et al., 2016). Green signals: the isolated RNA was treated with terminal exonuclease (+TEX), red signals: the RNA was not treated with TEX (–TEX). Comparison of reads from both fractions allowed us to determine the TSSs. The genome sequence is shown in the middle (with the genomic location in nucleotides, annotated genes are shown in black bars). Green (+TEX) and red (–TEX) regions represent dRNA-Seq reads, and the height corresponds to the coverage (shown at the left).

by the splicing endonuclease. These *in vivo* observations were confirmed by *in vitro* assays.

RESULTS

RNA Splicing Endonuclease From *Haloferax volcanii*

Haloferax volcanii encodes an RNA splicing endonuclease (HVO_2952, *endA*) in a bicistronic operon together with a tryptophanyl-tRNA synthetase (HVO_2951, *trpS1*) (**Figure 3**), the two genes overlap by four nucleotides. Both the prokaryotic operon database (Taboada et al., 2012) and our transcriptome (Berkemer et al., 2020), TSS (Babski et al., 2016) and transcription termination site (TTS) data (Berkemer et al., 2020) confirm this genomic organization. According to our TSS studies (Babski et al., 2016), the gene for the splicing endonuclease seems to contain two promoters for the downstream *trpS1* gene (**Figure 3**), therefore the *trpS1* gene could be transcribed independently from the *endA* gene. TSS data also show, that transcription from the *endA* promoter is comparatively low. In addition, the *endA* mRNA is not detectable on northern blots (data not shown). But proteome data clearly show that the splicing endonuclease is present in the proteome (Jevtić et al., 2019; Schulze et al., 2020).

To investigate the biological function of the splicing endonuclease, we aimed to repress its expression using CRISPR interference (CRISPRi) (Stachler and Marchfelder, 2016; Stachler et al., 2019). The CRISPRi approach uses the endogenous CRISPR-Cas system to repress transcription (Stachler and Marchfelder, 2016; Stachler et al., 2019). The Cas protein complex Cascade is guided by a crRNA to the

promoter region or transcription start site leading to hindering transcription initiation. If we repress transcription initiation at the *endA* promoter, the promoters for *trpS1* located in the *endA* gene are still present and able to initiate transcription of the *trpS1* gene.

Knockdown of the *endA* Gene Results in Growth Defects and Defects in tRNA Splicing

Transcription start sites of *Haloferax* were determined in a previous work, allowing the identification of upstream located promoter regions (Babski et al., 2016). We designed four crRNAs to target the promoter, the transcription start site and open reading frame of the *endA* gene (**Figure 4A**). The *Haloferax* strain HV30 was transformed with plasmids expressing these crRNAs (Stachler and Marchfelder, 2016; Stachler et al., 2019), and the resulting transformants were analyzed with respect to their growth behavior, showing that the expression of crRNA SEa2 had the strongest effect on growth (**Figure 4B** and **Supplementary Figure 2**). Cells expressing this crRNA showed a severe growth defect. RNA was extracted from these cells to determine how strong the *endA* mRNA was repressed. Since the amount of *endA* mRNA present in the cell was too low to use northern blots we used quantitative RT-PCR (qRT-PCR) to compare mRNA concentrations (**Figure 4C**).

Expression of crRNA SEa2 reduces the *endA* mRNA concentration down to 1.1% (±0.5%) of its wild type level in the exponential phase. To determine the effect of the lower expression of *endA* on the maturation of the intron containing tRNAs, northern blots were performed to detect the $tRNA^{Trp}$ precursors, processing intermediates and mature form. Northern blot analyses clearly showed that the expression of crRNA SEa2 results in an accumulation of unspliced tRNAs (**Figure 5**).

Repression of the Splicing Endonuclease Changes the Relative Abundances of (Pre-) Ribosomal RNA

Ribosomal precursor RNAs in *H. volcanii* contain two BHB motifs (**Figure 6A**). Recently, the structural integrity of these BHB motifs has been shown to be required for the synthesis of archaeal specific circular-pre-rRNA intermediates and mature rRNA (Jüttner et al., 2020). The presence and structure of this motif within the ribosomal processing stems is conserved in most archaea analyzed so far (Jüttner et al., 2020; Qi et al., 2020). The effect on the overall relative abundances of (pre-)rRNA in *endA* CRISPRi cells was investigated by qRT-PCR combining different sets of primers, which showed that pre-rRNAs that are not cleaved at the BHB motif slightly accumulate (**Figure 6B**). In addition, the amount of the circular pre-rRNAs normally cleaved at the BHB motif and circularized by the ligase is strongly decreased, showing that cleavage of the BHB motif by the splicing endonuclease (SE) is impaired. Furthermore, the consequence of *endA* repression is an overall reduced amount of total 16S and 23S rRNAs in CRISPRi cells (**Figure 6**). Taken together, these results established that the splicing endonuclease is essential for efficient rRNA maturation *in vivo*.

The *trpS1* Gene Is Individually Expressed

Transcriptome data show that the *trpS1* gene encoded downstream of *endA* is only slightly affected in the *endA* CRISPRi strain and is not down-regulated but slightly up-regulated (up-regulation 0.8, data not shown). Thus, down-regulation of transcription at the promoter upstream of *endA* does not repress expression of TrpS1. The promoters for *trpS1* detected in the *endA* gene (**Figure 3**) seem to be sufficient for adequate TrpS1 expression. To investigate whether *endA* and *trpS1* are transcribed into a bicistronic mRNA, we performed RT-PCR using primers targeting the *endA* gene and the *trpS1* gene (**Figure 7A**). A bicistronic mRNA was found in RNA from wild type cells; however, only low amounts of such an mRNA were present in RNA from *endA* CRISPRi cells (**Figure 7B**).

Reduction of Splicing Endonuclease Expression Results in Changes in the Transcriptome

To investigate the impact of splicing endonuclease repression on the cell, we compared the transcriptomes of wild type and *endA* CRISPRi cells. RNA was isolated from triplicates of both strains, and cDNA libraries were generated and subjected to RNA-Seq (for details see Material and Methods). The resulting sequencing data were investigated, and DESeq2 (Love et al., 2014) was used to test for differentially expressed genes in *endA* CRISPRi cells in comparison to the wild type cells. Differential expression was tested for all annotated genes of *Haloferax*. Altogether, 102 genes were found to be up-regulated and 41 were down-regulated (when a threshold of $\log_2$: 2/−2 was applied) (**Tables 1**, **2** and **Supplementary Tables 1**, **2**).

The transcriptome data clearly confirm down-regulation of the splicing endonuclease, since it is the most down-regulated gene. All three intron-containing tRNAs were also down-regulated (**Table 1** and **Supplementary Table 1**). In addition to the intron-containing tRNAs, seven other tRNAs were likewise down-regulated (**Supplementary Table 1**). The rRNAs were not included in the analysis since our RNA-Seq protocol included an rRNA removal step.

The 10 down-regulated tRNAs all are decoding preferentially used mRNA codons according to the codon bias determined for *Haloferax* (Nakamura and Sugiura, 2007)[4], which might lead together with the reduced rRNA production to a general decrease in protein synthesis. The other 31 down-regulated genes included 17 genes coding for proteins and RNAs with unknown function (**Supplementary Table 1**).

The most up-regulated gene encodes L-lactate permease. This gene is encoded in a bicistronic operon together with the gene for FAD-dependent oxidoreductase (HVO_1697), which is also up-regulated. Six of the most up-regulated genes are located in a cluster on pHV3 (HVO_B0042-HVO_B0047) and are related to iron metabolism. Altogether, 11 transport related proteins and 22

[4]http://www.kazusa.or.jp/codon/

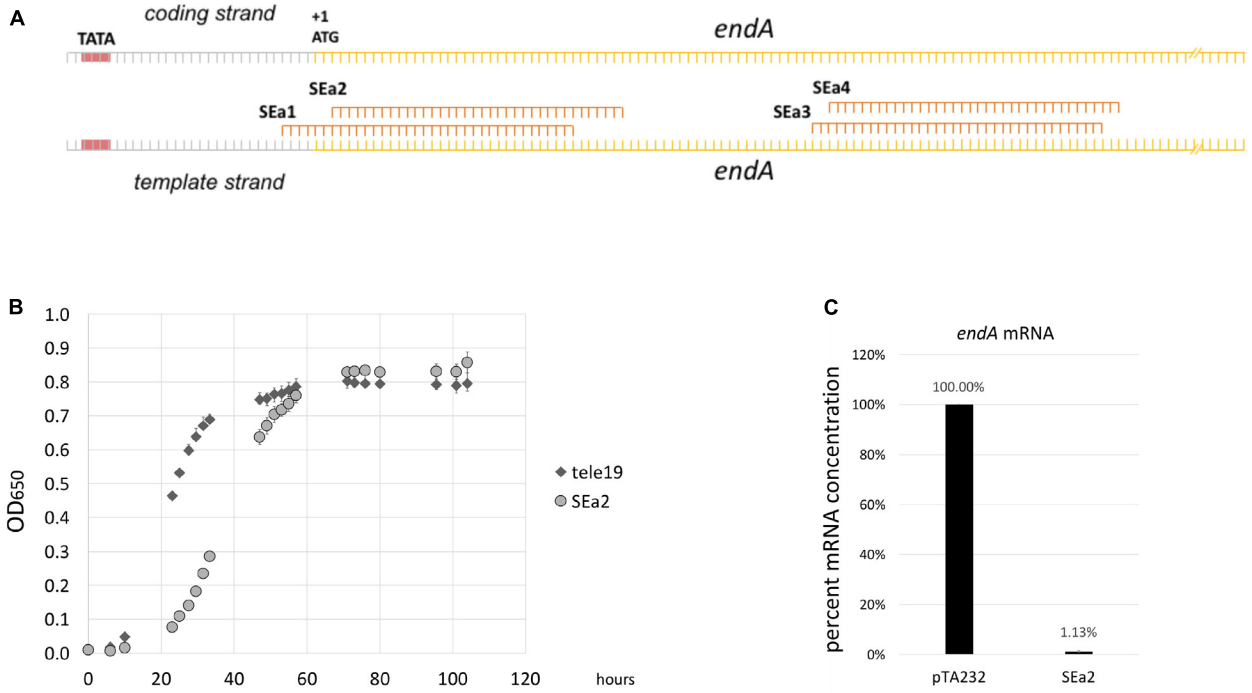


FIGURE 4 | CRISPRi against the *endA* gene results in reduced *endA* mRNA concentrations. **(A)** Location of crRNAs. Four different crRNAs (SEa1-SEa4) were designed against the *endA* transcription start site and beginning of the ORF. The promoter is indicated with a red box and TATA; the transcription start site is indicated with +1; the *endA* ORF is indicated in yellow and the ATG is also shown. TSS and the A of the ATG are identical, showing that the *endA* mRNA is leaderless. For a more detailed view see **Supplementary Figure 1**. **(B)** Expression of crRNA SEa2 influences growth. Expression of crRNA SEa2 results in a severe lag phase (light gray circles, SEa2). As a control, a crRNA that does not bind any target in the cell was expressed (dark gray diamonds, tele19). **(C)** Reduction of *endA* mRNA concentrations upon expression of crRNA SEa2. qRT-PCR was performed to measure the concentrations of *endA* mRNA in wild type (column pTA232, cells HV30 × pTA232) and CRISPRi cells (column SEa2, cells HV30 × pTA232SEa2). The mRNA concentration in wild type cells was set to 100%. Repression with CRISPRi reduces the mRNA concentration to 1.1% (±0.5) of that of the wild type.

genes coding for proteins and RNAs of unknown functions were up-regulated (**Supplementary Table 2**).

Circular RNAs That Are Products of the SE Reaction

As an additional approach to identify all substrates of the splicing endonuclease we enriched and sequenced the circular RNAs of *H. volcanii* (circRNA-seq). The SE cleaves its substrates at the BHB motifs, and the resulting intron is circularized by a ligase. Thus, circRNA-seq identifies among others, the circular splicing products of the SE. Altogether, seven circRNAs flanked by a BHB motif were identified (**Table 3**). Two belong to the known tRNA and four to the predicted rRNA SE substrates; one is newly identified. The circularized intron of the third tRNA, $tRNA^{Gln}_{TTG}$, was not found, which might be due to its short length (31 nucleotides). It has been reported previously that the circRNA-seq method has a bias against short RNAs (Danan et al., 2012). Analyses of the intron position in the newly identified substrate showed that it is located in the 5′ UTR of HVO_1309 (**Figure 8**). HVO_1309 codes for a peptidase from the M24 family protein and transcriptome data show that HVO_1309 is down-regulated in *endA* CRISPRi cells (**Supplementary Table 1**). This suggests that splicing might be required for efficient expression or RNA stability.

The Splicing Endonuclease Processes a Pre-16S and the Newly Identified mRNA Substrate *in vitro*

The *Haloferax* splicing endonuclease can be expressed in *E. coli* and processes intron-containing tRNAs effectively *in vitro* (Thompson and Daniels, 1990). To confirm that the ribosomal RNA precursors are substrates for SE, we processed a truncated 16S rRNA precursor *in vitro* with recombinant SE (**Figure 9**). To confirm the activity of the recombinant SE, a control reaction with the intron containing the $tRNA^{Trp}$ precursor was performed that showed efficient processing (data not shown). The full length 16S rRNA precursor is more than 1,600 nucleotides long and such long precursors are suboptimal substrates for *in vitro* processing assays. Therefore, we generated a 568 nt long rRNA substrate, that had most of the mature rRNA part deleted, leaving only the 5′ and 3′ ends of the 16S rRNA and the BHB motif (**Figure 9A**). Incubation with the recombinant SE showed that this substrate is indeed efficiently processed (**Figure 9B**). To test whether the newly identified BHB motif found in the 5′ UTR of HVO_1309 (**Table 3**) is cleaved by SE, we incubated the respective precursor

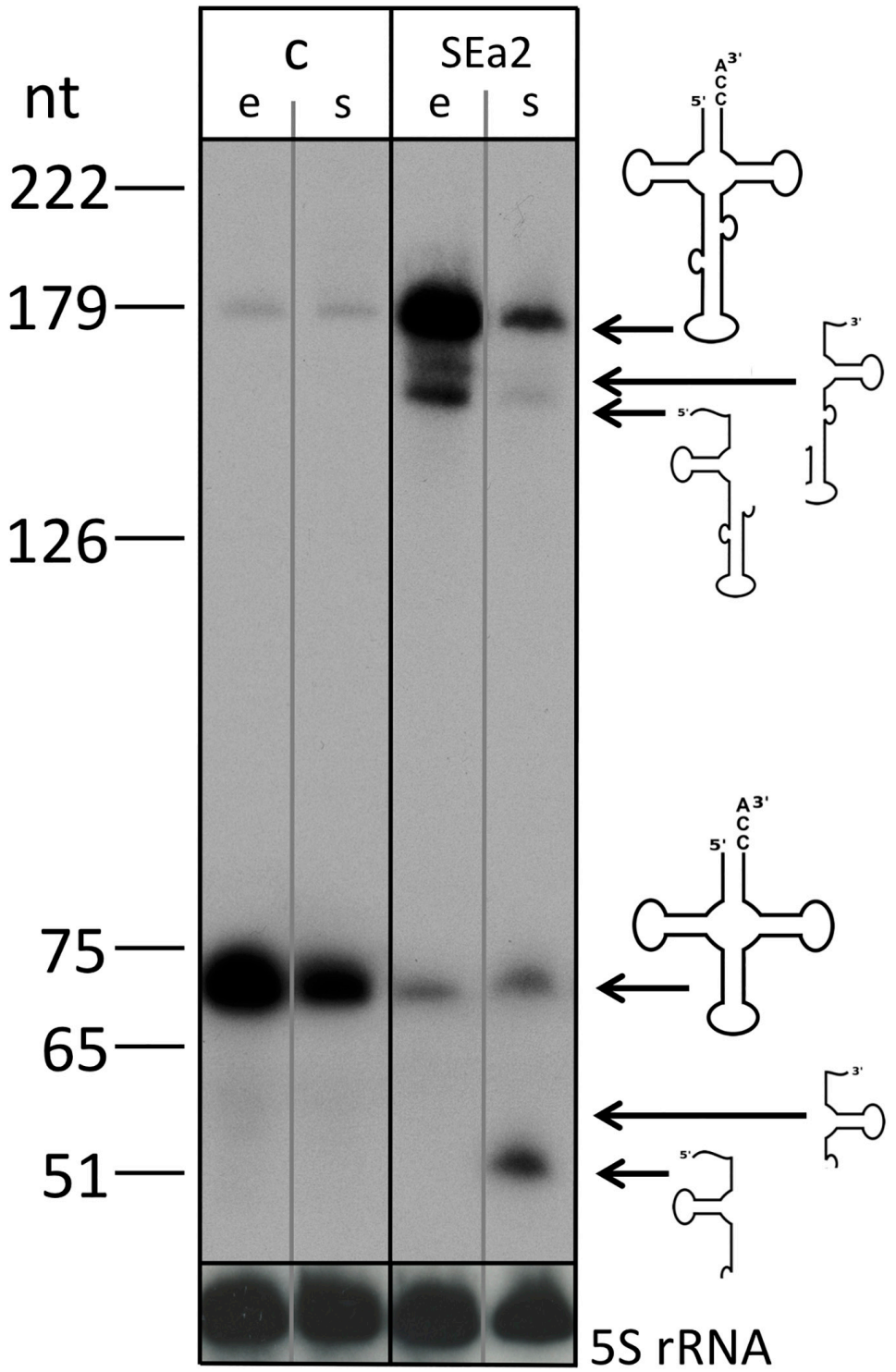


FIGURE 5 | tRNA maturation is severely impeded in CRISPRi cells. RNA was isolated from control cells (lane c, HV30 × pTA232) and cells expressing crRNA SEa2 (lane SEa2, HV30 × pTA232SEa2) (lanes e: cells were grown to exponential phase, lanes s: cells were grown to stationary phase). RNA was separated by 8% PAGE and transferred to nylon membranes. Blots were hybridized with a probe against the tRNA intron that also detects the mature tRNA (upper panel; lower panel: the same blot was hybridized with a probe against the 5S rRNA). Expression of crRNA SEa2 results in accumulation of unspliced tRNA and processing intermediates. Only very small amounts of mature tRNA can be detected. A DNA size marker is given at the left. Precursors, processing intermediates and the mature product are shown schematically on the right.

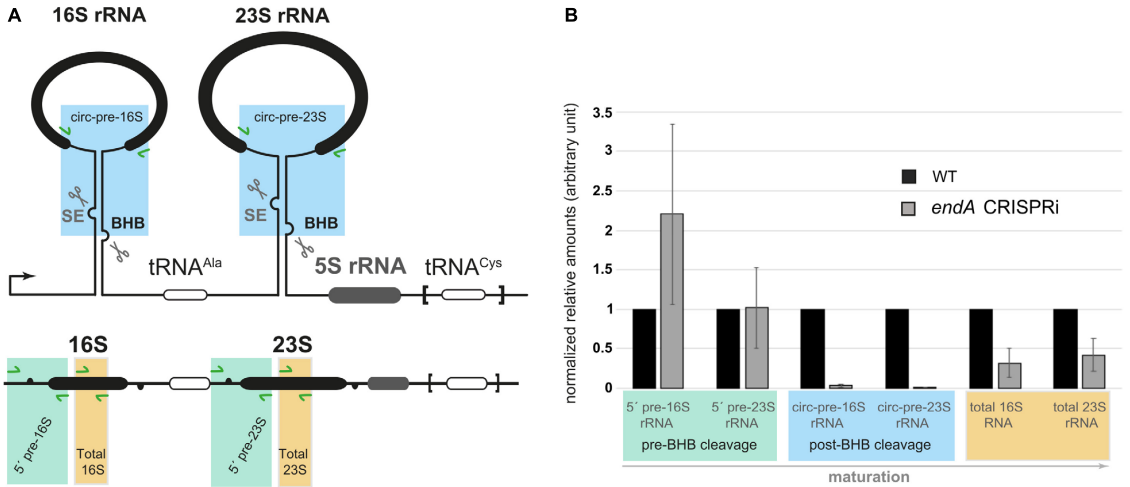


FIGURE 6 | Analysis of pre-rRNA maturation in cells depleted of the RNA splicing endonuclease. **(A)** Schematic representation of the *H. volcanii* rDNA operon. One of the two rDNA operons (operon A: HVO_3038-HVO_3042 and flanking regions) present in *H. volcanii* and characterized by an additional $tRNA^{Cys}$ gene at its 3′ end is schematically depicted. The inverted repeats flanking the mature 16S/23S rRNAs establish double stranded RNA regions containing bulge-helix-bulge (BHB) motifs, putative substrates for the splicing endonuclease (SE, scissors). Primers used for quantitative RT-PCR analysis of total 16/23S rRNAs (shown in orange), pre-circular 16S/23S rRNAs (post-BHB cleavage pre-rRNA species, blue) and 5′ pre-16S/23S rRNAs containing regions upstream of the respective 5′ bulge (pre-BHB cleavage pre-rRNA species, green) are indicated (green arrows). **(B)** Pre-rRNA maturation profile of cells depleted of SE analyzed by qRT-PCR.

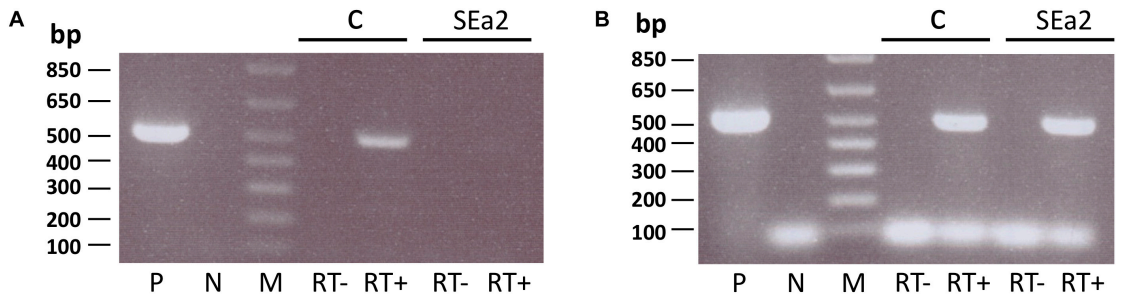


FIGURE 7 | *trpS1* can be transcribed independently from *endA*. RNA was isolated from control cells (HV30 × pTA232, lanes c) and cells expressing crRNA SEa2 (HV30 × pTA232SEa2, lanes SEa2). RNAs were reverse transcribed using random hexamer primers. PCR was performed using primers spanning the overlapping region of *endA* and *trpS1* **(A)** or a part of *trpS1* **(B)**. PCR products were resolved on a 0.8% agarose gel. P: Amplification from HV30 gDNA as a PCR control. N: control without the addition of template DNA. M: 1 kb plus DNA ladder, with sizes given on the left. RT–, isolated RNA was added to exclude DNA contamination; RT+, cDNA template was added (from control cells or CRISPRi cells). Signals below 100 bp in panel B are derived from primers.

in an *in vitro* assay with recombinant SE (**Figure 10**), which revealed that this substrate was processed. The two products flanking the BHB are clearly visible (26 nt and 29 nt). The central fragment (47 nt) is either running slower and visible at approximately 65 nucleotides (a fragment of this size is also visible in the control reaction) or degraded.

DISCUSSION

The Splicing Endonuclease Processes tRNAs, rRNAs, and a mRNA

Compared to bacteria and eukarya, only a few ribonucleases have been characterized in archaea. Furthermore, the ribonucleases which catalyze many key processing steps in cellular reactions remain unknown. It has been suggested that the few currently known archaeal ribonucleases may have a broader substrate spectrum and catalyze more processing steps than their bacterial and eukaryal homologs (Hölzle et al., 2012). For instance the *H. volcanii* tRNase Z – an endonuclease known for catalyzing tRNA 3′ processing (Schiffer et al., 2002; Späth et al., 2008)- has been shown to have an additional function besides tRNA processing; it is also involved in maturation of 5S ribosomal RNA (Hölzle et al., 2008).

Here, we aimed to identify all biological substrates of the tRNA splicing endonuclease in *H. volcanii* by applying the CRISPRi approach to repress *endA* expression to determine whether SE processes additional substrates beyond the known tRNA intron substrates. The growth of *endA* CRISPRi cells was severely affected, confirming the important cellular role of SE. We confirmed that all three intron containing tRNAs were down-regulated in the *endA* CRISPRi strain and that precursor

TABLE 1 | Genes down-regulated in the *endA* CRISPRi strain.

Gene	log_2	Annotation[a]
HVO_2952	−7,9	tRNA splicing endonuclease
HVO_0575	−4,7	Homolog to NAD-dependent epimerase/dehydratase
HVO_2519_T	−4,6	$tRNA^{Met}$ (contains an intron)
HVO_2566	−3,8	$tRNA^{Gly}$
HVO_0864_T	−3,5	$tRNA^{Gln}$ (contains an intron)
HVO_B0149	−3,3	Oleate hydratase
HVO_1800	−2,9	Hypothetical protein
HVO_1561	−2,9	Conserved hypothetical protein
HVO_2804_T	−2,9	$tRNA^{Thr}$
HVO_1276_T	−2,7	$tRNA^{Trp}$ (contains an intron)

The 10 most down-regulated genes are shown. The gene coding for the splicing endonuclease is the most down-regulated gene. All three tRNAs with introns are among the 10 most down-regulated genes and three genes encoding proteins of unknown function. Column "gene": HVO gene number; column log_2: log_2 fold change; column "annotation": gene annotation. [a]Annotation according to HaloLex (Pfeiffer et al., 2008).

TABLE 2 | Genes up-regulated in the *endA* CRISPRi strain.

Gene	log_2	Annotation[a]
HVO_1696	7.0	L-lactate permease
HVO_B0044	4.8	Siderophore biosynthesis protein, lucA
HVO_B0042	4.7	Probable 1,3-diaminopropane N-3-monooxygenase, lucD
HVO_B0045	4.5	Diaminobutyrate decarboxylase
HVO_1228	4.2	DUF5059 domain
HVO_B0043	4.1	Probable N4-hydroxy-1-aminopropane O-acetyltransferase, lucB
HVO_1697	4.1	FAD-dependent oxidoreductase (GlcD/DLD_GlcF/GlpC domain fusion protein)
HVO_B0047	4.0	ABC-type transport system periplasmic substrate-binding protein (probable substrate iron-III)
HVO_B0046	3.9	Diaminobutyrate decarboxylase diaminobutyrate-2-oxoglutarate aminotransferase
HVO_B0197	3.8	ABC-type transport system periplasmic substrate-binding protein (probable substrate iron-III)

The 10 most up-regulated genes are shown. The gene HVO_1696, coding for an L-lactate permease is the most up-regulated gene. Six of the most up-regulated genes are located in a cluster on pHV3 (HVO_B0042-HVO_B0047) and are related to iron metabolism. Column "gene": HVO gene number; column log_2: log_2 fold change; column "annotation": gene annotation. [a]Annotation according to HaloLex (Pfeiffer et al., 2008).

tRNAs accumulated in the CRISPRi strain, establishing that the splicing endonuclease catalyses intron removal *in vivo*. In addition, we showed that the BHB motifs flanking the two larger ribosomal RNAs are also cleaved by SE *in vivo*. The processing of the 16S ribosomal RNA precursor by the SE could be confirmed with an *in vitro* processing experiment. This is in line with the observed processing of ribosomal RNAs in *Sulfolobus solfataricus* and *Methanolobus psychrophilus*, where maturation is also performed by the SE (Ciammaruconi and Londei, 2001; Qi et al., 2020).

To identify additional SE substrates, we employed circRNA-seq that identifies among others, the circular introns that are products of the splicing reaction. The rRNA and two of the tRNA circular introns were detected, and only the 31 nucleotide intron from $tRNA^{Gln}$ was not found. It has been reported that circRNA-seq has a bias against shorter circRNAs (Danan et al., 2012) which might explain this observation. CircRNA-seq identified an additional circular RNA derived from the 5′ UTR of HVO_1309, and splicing removes 46 nucleotides from the 192 nucleotide long 5′ UTR. 5′ UTRs are known from bacteria/eukarya to be used as regulatory elements. 5′ UTRs can bind small regulatory RNAs or proteins to regulate the expression of downstream genes. They can also fold into specific structures binding distinct ligands, thereby also regulating downstream genes. Removal of the 46 nucleotides could eliminate a protein or sRNA binding site or inhibit/allow a specific structure to form. In archaea, only a few examples of interactions of proteins with 5′ UTRs have been reported (Daume et al., 2017; Li et al., 2019). The entire transcribed sequence of the HVO_1309 gene is conserved in haloarchaea, suggesting that the sequence of the 5′ UTR might be important. Since HVO_1309 is down-regulated in the CRISPRi strain, removal of the intron seems to be important for efficient expression of HVO_1309. Processing of the BHB sites in the 5′ UTR of HVO_1309 was confirmed by the *in vitro* assay. To date, only one example of an additional BHB intron has been reported in archaea situated in the *cbf5* mRNA in crenarchaeota (Watanabe et al., 2002; Yokobori et al., 2009). Whether the intron has a specific function is not yet known.

Repression of the *endA* Gene Expression Results in Regulation of a Plethora of Genes

Comparison of the *endA* CRISPRi transcriptome with the wild type transcriptome showed that a plethora of genes are regulated upon repression of *endA*. Many tRNAs are affected by repression, suggesting that there is feedback from the defect in processing intron-containing tRNAs and perhaps even pre-rRNAs. The down-regulated tRNAs decode preferentially used codons; thus, depletion of these tRNAs surely has an impact on translation in general. However, little is known about how tRNA and rRNA expression are co-regulated and whether tRNA levels are co-regulated in archaea. In addition, it is difficult to assess the direct effects of inefficient tRNA and rRNA intron splicing and secondary effects that are due to low concentrations of mature tRNAs and rRNAs.

endA/trpS1 Operon Structure

The *endA* gene is located upstream of the *trpS1* gene, and the genes overlap with four nucleotides. We showed that *trpS1* is transcribed together with *endA* but it is also separately transcribed from independent promoters into a monocistronic mRNA. This is confirmed by the transcriptome data that show that the *trpS1* mRNA concentrations are not down-regulated in the CRISPRi strain. The expression of *trpS1* can be regulated from two promoters located in the *endA* ORF which allows regulation of *trpS1* expression independent of *endA* expression. This setup is another example of the independent regulation of genes that are part of the same operon (Berkemer et al., 2020).

TABLE 3 | Circular RNAs Identified by circRNA-seq.

Left site	Right site	Intron length	Folding energy	Gene
1,195,172	1,195,218	46	−14.9	5′ UTR HVO_1309
2,384,876	2,384,951	75	−11.6	$tRNA^{Met}_{CAT}$ (intron)
1,165,777	1,165,880	103	−9.6	$tRNA^{Trp}$ (intron)
2,770,085	2,771,742	1658	−11.8	16S rRNA
1,598,083	1,599,741	1658	−11.8	16S rRNA
1,600,017	1,602,996	2979	−14.2	23S rRNA
2,766,828	2,769,808	2980	−12.8	23S rRNA

Seven circRNAs were identified by in silico analysis to be flanked by a BHB motif with a folding energy lower than −9. Introns of tRNAs $tRNA^{Trp}$ and $tRNA^{Met}$ could be identified, as could circular 16S and 23S rRNA precursors. Columns: left site/right site: position of the BHB splice site; intron length: length of the respective intron; folding energy: folding energy of the BHB motif (in kcal/mol); gene: gene in which the intron is located.

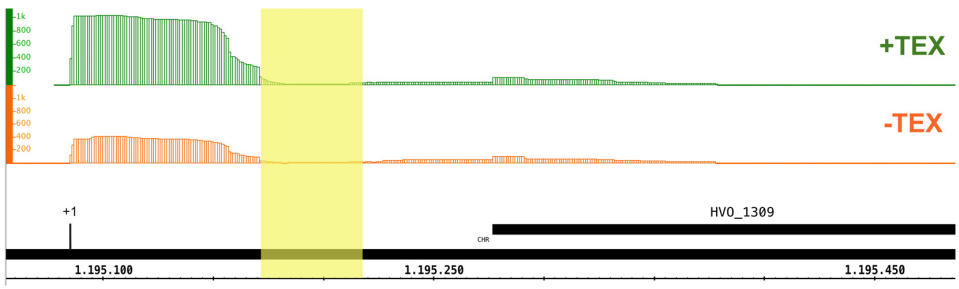


FIGURE 8 | An intron is located in the 5′ UTR of HVO_1309. CircRNA-seq identified a circular RNA that is flanked by a BHB motif derived from the 5′UTR of the HVO_1309 gene. The position of the intron is shaded in yellow. The TSS is indicated by a black line and +1. TSSs were identified by dRNA-Seq analysis (Babski et al., 2016). Green: RNA treated with terminal exonuclease (+TEX), red: RNA not treated with TEX (–TEX). The height of the regions corresponds to the coverage of dRNA-Seq reads (read numbers are given at the left). Genome coordinates are indicated in nucleotides at the bottom. The HVO_1309 gene is shown as a black bar.

trpS1 codes for tryptophanyl-tRNA synthetase and is therefore important for loading $tRNA^{Trp}$ with the aminoacyl group. Reduced levels of mature $tRNA^{Trp}$ and therefore less tRNA to be loaded with an aminoacyl group might lead to a slight up-regulation of *trpS1* expression as a feedback mechanism.

MATERIALS AND METHODS

The strains, plasmids and primers used are listed in **Supplementary Table 3**.

Growth of Strains

Haloferax volcanii strain HV30 was grown under aerobic conditions at 45°C in Hv-YPC medium or HV-Min medium containing the respective supplements (Allers et al., 2004). *E. coli* strain DH5α (Invitrogen™, Thermo Fisher Scientific) was grown aerobically at 37°C in 2YT medium (Miller, 1972).

Growth Experiment

Haloferax volcanii cells were grown in glass test tubes under shaking at 45°C in triplicates. The $OD_{650\ nm}$ was measured at different time points.

RT-PCR

TRIzol™ Reagent (Invitrogen™, Thermo Fisher Scientific) was used to isolate total RNA from *H. volcanii* cells in the early exponential growth phase (OD_{650}: ∼0.3). RNA was treated with RQ1 RNAfree DNase (Promega) to remove residual DNA. To investigate whether *endA* and *trpS1* are transcribed as bicistronic or two monocistronic transcripts, cDNA synthesis was carried out with 2 μg of DNA-free total RNA, RevertAid Reverse Transcriptase and random hexamer primers (Thermo Fisher Scientific). The subsequent PCR was performed with the gene-specific primers endAtrpS1_fw and endAtrpS1_rev to amplify the overlap of *endA* and *trpS1*. The primers trpS1intfw/rev were used to compare the signal of the *trpS1* transcript to the signals from the overlapping section. PCR products were separated on a 0.8% agarose gel in TAE buffer.

qRT-PCR

Determination of the *endA* mRNA Concentrations

To determine the relative amount of *endA* transcript, RNA was isolated from *H. volcanii* cells in exponential growth phase from strains carrying only a control plasmid (HV30 × pTA232) or a plasmid expressing a spacer against the transcription start site of the *endA* gene (HV30 × pTA232SEa1-a4). Ten micrograms of total RNA was treated with 30 μl of RQ1 DNase (Promega), and cDNA synthesis was performed using random hexamer primers and RevertAid Reverse Transcriptase (Thermo Fisher Scientific). The quantitative PCRs were performed with the KAPA™ SYBR fast Mastermix for Roche LightCycler® Kapa Biosystems and the LightCycler® 480 System (Roche). The primer sets used were q_endA_fw2 and q_endA_rev2 to amplify the *endA* transcript, q_tsgA3_fw2 and q_tsgA3_rev2 for tsgA3 and

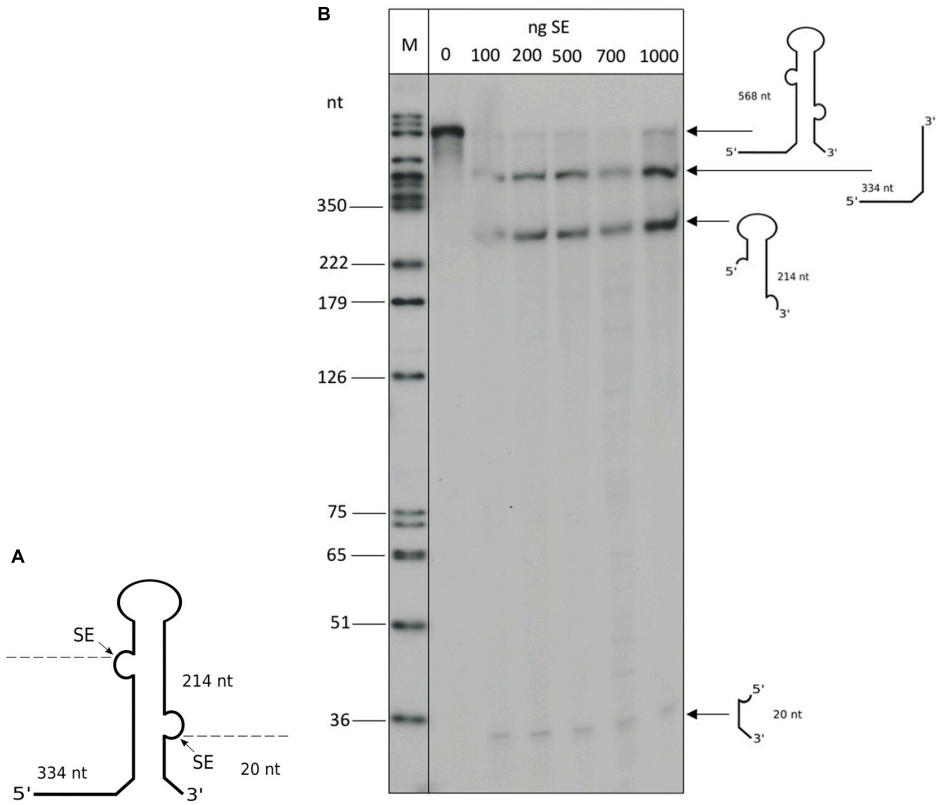


FIGURE 9 | *In vitro* processing of a truncated 16S precursor. **(A)** Schematic drawing of the truncated 16S rRNA precursor used for the *in vitro* reaction: 1,450 nucleotides of the internal part of the 16S rRNA were deleted. Arrows and dashed lines indicate the SE splice sites and processing products. The resulting fragment sizes are given in nucleotides. **(B)** *In vitro* processing was performed with recombinant SE and the radioactively labeled substrate. Different amounts of SE protein were used (between 100 and 1,000 ng), as indicated at the top of the lanes. A DNA size marker is given on the left. The 568 nt long precursor is processed by the SE, resulting in three fragments: a 334 nt leader fragment, a 214 nt 16S rRNA loop fragment and a 20 nt trailer fragment. SE processing products can be observed beginning at 100 ng. Processing fragments are shown schematically on the right.

q_trmB1_fw2 and q_trmB1_rev2 for trmB1. Cycling conditions were as follows: initial denaturation for 5 min at 95°C; 40 cycles of denaturation for 15 s at 95°C, annealing of primers for 10 s at 59°C and elongation for 10 s at 72°C; followed by a melting curve determination for 5 s at 95°C and 1 min at 65°C. The reaction was completed by cooling for 30 sec at 40°C. qPCRs were carried out in triplicate in three independent experiments. Normalization of *endA* levels was obtained by using *tsgA3* and *trmB1* as references. Relative levels were calculated with the Roche LightCycler® software according to the E-Method.

Determination of the Relative Abundance of Pre-rRNAs

Total RNA from logarithmically growing cells, cDNA synthesis and qRT-PCR analysis were essentially performed as described previously (Knüppel et al., 2017). For cDNA synthesis primers oHv040 (16S rRNA), oHv042 (23S rRNA) and oHv390 (Tfb) were used. The following primer pair combinations were used for qRT-PCR: oHv040/oHv039 (Hv_circ-pre-16S rRNA), oHv041/oHv042 (Hv_circ-pre-23S rRNA) oHv040/oHv200 (Hv_5' extended-pre-16S rRNA) oHv042/oHv201 (Hv_5' extended-pre-23S rRNA), oHv040/oHv205 (Hv_total 16S rRNA), oHv042/oHv206 (Hv_total 23S rRNA), and oHv391/oHv392 (Tfb) (Jüttner et al., 2020). Relative quantification analysis was performed using a comparative analysis software module (Rotor-gene 6 – Corbett Research/Qiagen). Relative levels were calculated according to the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) using the transcription initiation factor TFB (HVO_0226) mRNA level as reference. Relative RNA levels were normalized to WT. To ensure accuracy of the data, experiments were performed using biological triplicates (three CRISPRi transformants) and technical triplicates of serial dilutions (minimum two dilutions) of the cDNA samples were run.

CRISPRi

crRNAs against the region around the transcription start side of the *endA* gene were synthesized by Invitrogen GeneArt Gene Synthesis (Thermo Fisher Scientific). The plasmids carry a synthetic promoter and terminator for *H. volcanii* (Maier et al., 2015), tRNA-like-elements that flank the crRNA, and the crRNA gene that contains an 8 nt 5′ handle and the 36 nt spacer. The complete insert was cut from the plasmid with *Kpn*I and *Bam*HI and ligated with the *Haloferax* shuttle vector pTA232, yielding plasmids pTA232-SEa1 to 4. HV30 was transformed

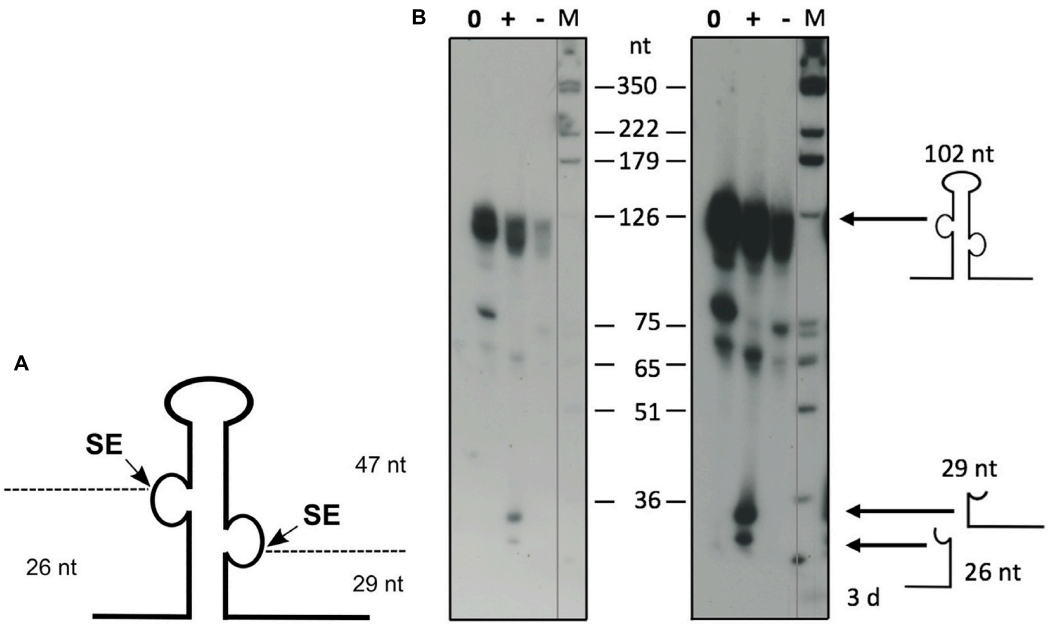


FIGURE 10 | *In vitro* processing of the newly identified BHB containing RNA. **(A)** Schematic drawing of the precursor substrate. As a substrate for the *in vitro* processing reaction, the 5′ UTR of HVO_1309 was used, encompassing the intron flanked by the BHB motif with additional nucleotides up- and downstream. **(B)** Processing reaction. The substrate was transcribed *in vitro* and labeled throughout with [α − −^{32}P]-UTP. *In vitro* processing was performed with recombinant SE, and RNAs were separated by 8% PAGE. Samples were incubated for 60 min. A DNA size marker is given at the sides, and sizes are shown in the middle (M). A control reaction without enzyme was performed at the beginning and end of the reaction time (indicated as NC0 and NC60 for 0 and 60 min of incubation, respectively). Samples with enzyme were incubated for 60 min (SE). Left panel: short exposure, right panel: long exposure. The precursor RNA has a length of 102 nucleotides, and the processing fragments are 26, 29, and 47 nucleotides long. Substrate and processing products are shown schematically at the side.

with one of the generated plasmids pTA232-SEa1 to 4 or the control plasmid pTA232. Cells were grown in Hv-Min selection medium containing tryptophan and uracil and harvested at selected $OD_{650\ nm}$ values. The repression effect was determined by northern blot analysis and qRT-PCR.

Analysis of Differentially Expressed Genes

RNA isolated from CRISPRi and control cultures was treated with TURBO^{TM} DNase (Invitrogen^{TM}, Thermo Fisher Scientific) to remove all genomic DNA that was carried over during RNA preparation. Ten micrograms of total RNA was treated with 20 µl (40 units) of TURBO^{TM} DNase in a volume of 200 µl. Ten micrograms of DNA-free RNA was sequenced with next generation sequencing by Vertis Biotechnologie AG. The quality of raw sequencing reads was checked using FastQC version 0.11.4 (Andrews, 2010), and adaptor sequences and low-quality reads were trimmed using Cutadapt version 1.10 (Martin, 2011). Read mapping was performed using segemehl version 0.2.0 (Hoffmann et al., 2009; Otto et al., 2014). This was done for reads of three replicates from the wild type and three replicates from the CRISPRi mutant. The numbers of mapped and unmapped reads are listed in **Table 4**.

After the mapping, htseq-count (Anders et al., 2015) was used to calculate read counts as an input to DESeq2 (Love et al., 2014). DESeq2 was then applied to the data with the wild type sample as the control and the CRISPRi samples as the condition.

The resulting down- and upregulated genes are listed in **Supplementary Tables 1**, **2**.

CircRNA-Seq and Detection of BHB Elements

RNA was isolated and treated as described above. To enrich circular RNAs the ribodepleted RNA samples were digested with Ribonuclease R (Lucigen, United States). RNA was then fragmented using ultrasound (4 pulses of 30 s each at 4°C). For cDNA synthesis oligonucleotide adapters were ligated to the 5′ and 3′ ends of the RNA fragments. First-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3′ adapter as the primer. The resulting cDNAs were PCR-amplified using a high-fidelity DNA polymerase. The cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analyzed by capillary electrophoresis. The cDNA pools were sequenced on an Illumina NextSeq 500 system using a 1 × 75 bp read length. The quality of raw sequencing reads was checked using FastQC version 0.11.4 (Andrews, 2010), and adaptor sequences and low-quality reads were trimmed using Cutadapt version 1.10 (Martin, 2011). Read mapping was performed using segemehl version 0.2.0 (Hoffmann et al., 2009; Otto et al., 2014) with the split-read option such that split reads and circularized sequences were additionally reported. This was done for reads of 3 replicates from the wild type and 3 replicates from the CRISPRi mutant. The numbers of mapped and unmapped reads are listed in the **Table 5**.

TABLE 4 | Reads mapped to genome for the analysis of differentially expressed genes.

Sample	Replica	Total reads	Mapped reads	Uniquely mapped	Un-mapped	% of mapped reads
Wild type	S1	21,418,618	20,445,547	14,927,255	973,071	95.46%
	S2	21,816,231	20,725,427	14,408,066	1,090,804	95.00%
	S3	22,918,081	22,051,753	16,240,084	866,328	96.22%
CRISPRi	S1	22,626,113	21,810,284	17,903,446	815,829	96.39%
	S2	20,067,042	19,250,373	15,557,489	816,669	95.93%
	S3	21,902,393	21,043,653	16,769,061	858,740	96.08%

The number of reads obtained (column total reads) and mapped (columns mapped reads, uniquely mapped reads, unmapped reads) from the different samples are shown.

TABLE 5 | Reads mapped to genome for CircRNA-seq.

Sample	Replica	Total reads	Mapped reads	Uniquely mapped	Un-mapped	% of mapped reads
Wild type	S1	13,356,856	13,086,209	11,602,090	270,647	97.97%
	S2	11,955,747	11,718,382	10,305,229	237,365	98.01%
	S3	14,669,440	14,457,586	11,515,934	211,854	98.56%
CRISPRi	S1	9,902,921	9,724,986	8,982,553	177,935	98.20%
	S2	11,089,425	10,868,626	10,090,045	220,799	98.01%
	S3	13,426,209	13,247,376	12,346,730	178,833	98.67%

The number of reads obtained (column total reads) and mapped (columns mapped reads, uniquely mapped reads, unmapped reads) from the different samples are shown.

TABLE 6 | Introns in wild type and CRISPRi strains.

	Wild type			CRISPRi		
	Linear	Circularized	Total	Linear	Circularized	Total
Number	85	12	97	80	9	89
Average coverage	956.722	2,437.394	1,139.898	1,995.882	4,351.267	2,234.067

Numbers and and average coverage of linear and circularized introns in wild type and CRISPRi strains is shown. The coverage is higher for putative introns in CRISPRi displaying the lower splicing activity.

Mapped reads were processed using samtools version 1.3 (Li et al., 2009). To calculate genome coverage and intersections of data sets, we used bedtools (bedtools v2.26.0) (Quinlan and Hall, 2010). Splice sites reported after the mapping were filtered such that only splice sites with a coverage of at least two that appear in all three replicates were kept. A pair of splice sites (left and right) describe the location of an intron. Various reported introns differ by only a few bases in their start or end coordinates. Such sites were merged if their distance was at most 10 nt. The coordinates of the site with the highest coverage were kept for further analyses. The numbers and varieties of linear and circularized introns in the wild type and CRISPRi mutant do not differ significantly; however, their coverage shows large differences regarding circularized introns (**Table 6**). The coverage is higher for putative introns in CRISPRi, which can be explained by lower splicing activity such that coverage refers to unspliced fragments, whereas the wild type introns undergo splicing and thus degradation more quickly.

The search for BHB motifs was conducted using covariance models of known BHB motifs in *Haloferax* and programs of the infernal suite version 1.1 (Nawrocki and Eddy, 2013) for the detection of new motifs. BHB motifs are secondary structures that are formed out of two remotely located parts around the intron and act as a signal for the splicing machinery. The covariance model consists of a multiple sequence alignment of known BHB motifs and a consensus secondary structure. As the intronic sequences are not part of the motif and differ for each of the elements, sequences around the splice sites are cut and glued together such that intronic sequences are modeled as inserts. The program cmbuild from the infernal suite was used to create the covariance model. To set the focus on the secondary structure, the options –eset 0 and –hand were used. The model is shown below, where square brackets show base pairs and "b" denotes the bulge positions. Splice sites occur inside the bulges. Thus, areas around the splice sites are extracted from the sequences, where "∼" stands for the intronic sequence, which is not part of the model and thus is modeled as an insertion.

```
[[[[[[[[[bbb[[[~~~~~~~~~]]]]]]]bbb]]]]]
----exon--|--------intron-------|-exon-
```

Reported splice sites are then used to extract sequences around the splice sites themselves to restrict the search space to a reasonable number of sequences and positions. As the BHB motif is a relatively small motif without sequence conservation, restriction of the search space is necessary for the current type of covariance models which only report hits with an overall good score. However, unmatched positions receive a bad score; thus, unrestricted searches would result in mostly badly scored hits. The program cmalign was used to apply the BHB covariance model to the set of target sequences. Here, a global alignment of the sequences to the model was used, to force the program to map sequences to the BHB motif instead of creating insertions or deletions. The results of cmalign were then additionally checked to only keep sequences where the splice site was located inside the bulge region. Another step was to use RNAcofold from ViennaRNA package [version 2.0, (Lorenz et al., 2011)] to check if sequences predicted to form a BHB motif truly fold into the

predicted structure. RNAcofold additionally outputs the folding energy of the secondary structure motif, which indicates stability and thus probable occurrences of the motif.

Northern Blot Analysis

Total RNA was isolated from *H. volcanii* cells in the early exponential growth phase ($OD_{650\ nm}$: ∼0,3) with TRIzol™ Reagent (Invitrogen™, Thermo Fisher Scientific). DNase treatment to remove residual DNA was performed with RQ1 RNase-Free DNase (Promega). Ten micrograms of total RNA was separated by 8% PAGE and transferred to a nylon membrane (Hybond-N+, GE Healthcare). The membrane was hybridized with radioactively labeled oligonucleotide probes. Probes were labeled at the 5′ end by T4-polynucleotide kinase (PNK; Thermo Fisher Scientific) and [γ-^{32}P]-ATP. To quantify the RNA, imaging plates (BAS-MS, Fujifilm) were exposed to radioactivity on the hybridized membranes and analyzed with the FLA-3000 scanner (GE Healthcare; software BASreader 3.14). Signal intensity was measured with ImageJ and set in relation to the signal of 5S rRNA that was detected and used as a loading control. The proportion of RNA is given as a percentage from cells expressing a targeting spacer and those carrying a control plasmid. The amount of RNA in the control strains was set to 100%, and the amounts in strains expressing a targeting spacer were calculated in relation to this value. Northern blot analyses were performed with biological triplicates of the control and CRISPRi strains.

In vitro Processing

The splicing endonuclease was expressed from the plasmid pTA28a(+)-*endA* in *E. coli* BL21Ai cells as fusion protein with an N-terminal His-tag. Expression of the protein was induced with 2 g/l arabinose and 1 mM IPTG at $OD_{600\ nm}$ 0.6 for 3 h at room temperature and under constant shaking. Cells were lysed by sonification (Sonofier 250, Branson Ultrasonics). After centrifugation, a His-tag purification was carried using the supernatant and N-terminal His-tag and Poly-Prep® Chromatography Columns (Bio-Rad) with Protino® Ni-NTA Agarose (Macherey-Nagel). The purity of the recombinant protein was confirmed by Coomassie stained SDS-PAGE gels and western blot. Templates for *in vitro* processing were cloned containing putative introns and exons of different lengths. The PCR product consisted of the T7 promoter and the respective genomic region. Primers for amplification were as follows: T7ptRNATrp and 3-tTrpTrailer, 1309-fw and 1309-rev and 5-T7prom-16SPrim, 3-*Xba*I- Leader-Trailer 16SPrim, 5-*Xba*I-16S3Trailer and 3-*Kpn*I-16S3Trailer. PCR products were generated by amplification with *Pfu* DNA Polymerase (Promega), treated with *Pfu* DNA Polymerase for 30 min at 72°C to generate blunt ends and separated on a 1% agarose gel. These DNA templates were used for *in vitro* transcription using T7 RNA Polymerase (New England Biolabs) in the presence of [α-^{32}P]-UTP. After transcription, substrates were separated on a 8% PAA-TBE-urea-gel. RNAs were recovered from the gel by overnight elution at 4°C in elution buffer containing 0.5 M ammonium acetate, 0.1 mM EDTA and 0.1% SDS at pH 5. For *in vitro* processing, RNA substrates were denatured in SE buffer (10× SE buffer: 400 mM Tris pH 7.5, 25 mM spermidine) at 80°C for 5 min and renatured for 30 min at 21°C. Reactions with and without enzyme were carried out at 37°C and samples were taken at different time points. RNA was extracted from the processing reactions by phenol-/chloroform and precipitated with ethanol. RNA was separated on an 8%-PAA-TBE-urea-gel and transferred to filter paper and a photosensitive film was exposed to the gel.

DATA AVAILABILITY STATEMENT

The data were deposited in ENA with study accessions numbers: PRJEB40302 (primary) and ERP123923 (secondary).

AUTHOR CONTRIBUTIONS

TS, SBk, SBn, and MW did the experiments and data curation. AM conceptualized the project. TS, SBk, SF-C, PS, and AM wrote the original draft. SF-C, PS, and AM reviewed the draft, edited the draft, and provided the resources and funding. All authors contributed to the article and approved the submitted version.

FUNDING

SF-C was supported by intramural funding from the Department of Biochemistry III – House of the Ribosome and by the DFG [FE1622/2-1, and Collaborative Research Center SFB/CRC 960 (SFB960-B13)]. Work in the AM laboratory was funded by the DFG (Ma1538/21-1).

ACKNOWLEDGMENTS

We thank Dr. Karina Haas for generating the 16S substrate plasmid pUC18-16S-leader-trailer (Haas, 2016) and Susanne Schmidt for expert technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.594838/full#supplementary-material

REFERENCES

Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the leuB and trpA genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/aem.70.2.943-953.2004

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data.*

Armbruster, D. W., and Daniels, C. J. (1997). Splicing of intron-containing tRNATrp by the archaeon *Haloferax volcanii* occurs independent of mature tRNA structure. *J. Biol. Chem.* 272, 19758–19762. doi: 10.1074/jbc.272.32.19758

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

Berkemer, S. J., Maier, L. K., Amman, F., Bernhart, S. H., Wortz, J., Markle, P., et al. (2020). Identification of RNA 3 ends and termination sites in *Haloferax volcanii*. *RNA Biol.* 17, 663–676. doi: 10.1080/15476286.2020.1723328

Ciammaruconi, A., and P. Londei (2001). In vitro processing of the 16S rRNA of the thermophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 183, 3866–3874. doi: 10.1128/JB.183.13.3866-3874.2001

Clouet d'Orval, B., Bortolin, M. L., Gaspin, C., and Bachellerie, J. P. (2001). Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res.* 29, 4518–4529. doi: 10.1093/nar/29.22.4518

Clouet-d'Orval, B., Batista, M., Bouvier, M., Quentin, Y., Fichant, G., Marchfelder, A., et al. (2018). Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiol. Rev.* 42, 579–613. doi: 10.1093/femsre/fuy016

Danan, M., Schwartz, S., Edelheit, S., and Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* 40, 3131–3142. doi: 10.1093/nar/gkr1009

Daume, M., Uhl, M., Backofen, R., and Randau, L. (2017). RIP-Seq Suggests Translational Regulation by L7Ae in Archaea. *mBio* 8:e00730-17. doi: 10.1128/mBio.00730-17

Fujishima, K., Sugahara, J., Miller, C. S., Baker, B. J., Di Giulio, M., Takesue, K., et al. (2011). A novel three-unit tRNA splicing endonuclease found in ultrasmall Archaea possesses broad substrate specificity. *Nucleic Acids Res.* 39, 9695–9704. doi: 10.1093/nar/gkr692

Haas, K. A. (2016). *Untersuchung des CRISPR-Cas-Systems und der RNase G/E in Archaeen*. Ph.D. thesis, Ulm University, Ulm.

Hölzle, A., Fischer, S., Heyer, R., Schütz, S., Zacharias, M., Walther, P., et al. (2008). Maturation of the 5S rRNA 5' end is catalyzed in vitro by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA* 14, 928–937. doi: 10.1261/rna.933208

Hoffmann, S., Otto, C., Kurtz, S., Sharma, C. M., Khaitovich, P., Vogel, J., et al. (2009). Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol.* 5:e1000502. doi: 10.1371/journal.pcbi.1000502

Hölzle, A., Stoll, B., Schnattinger, T., Schöning, U., Tjaden, B., and Marchfelder, A. (2012). tRNA-like elements in *Haloferax volcanii*. *Biochimie* 94, 940–946. doi: 10.1016/j.biochi.2011.12.002

Jevtić, Ž, Stoll, B., Pfeiffer, F., Sharma, K., Urlaub, H., Marchfelder, A., et al. (2019). The response of *Haloferax volcanii* to salt and temperature stress: a proteome study by label-free mass spectrometry. *Proteomics* 19:e1800491. doi: 10.1002/pmic.201800491

Joardar, A., Gurha, P., Skariah, G., and Gupta, R. (2008). Box C/D RNA-guided 2'-O methylations and the intron of tRNATrp are not essential for the viability of *Haloferax volcanii*. *J. Bacteriol.* 190, 7308–7313. doi: 10.1128/jb.00820-08

Jüttner, M., Weiß, M., Ostheimer, N., Reglin, C., Kern, M., Knüppel, R., et al. (2020). A versatile cis-acting element reporter system to study the function, maturation and stability of ribosomal RNA mutants in archaea. *Nucleic Acids Res.* 48, 2073–2090. doi: 10.1093/nar/gkz1156

Kjems, J., and Garrett, R. A. (1991). Ribosomal RNA introns in archaea and evidence for RNA conformational changes associated with splicing. *Proc. Natl. Acad. Sci. U.S.A.* 88, 439–443. doi: 10.1073/pnas.88.2.439

Kleman-Leyer, K., Armbruster, D. W., and Daniels, C. J. (1997). Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell* 89, 839–847. doi: 10.1016/s0092-8674(00)80269-x

Knüppel, R., Kuttenberger, C., and Ferreira-Cerca, S. (2017). Toward time-resolved analysis of RNA metabolism in archaea using 4-thiouracil. *Front. Microbiol.* 8:286. doi: 10.3389/fmicb.2017.00286

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, H., Trotta, C. R., and Abelson, J. (1998). Crystal structure and evolution of a transfer RNA splicing enzyme. *Science* 280, 279–284. doi: 10.1126/science.280.5361.279

Li, J., Zhang, B., Zhou, L., Qi, L., Yue, L., Zhang, W., et al. (2019). The archaeal RNA chaperone TRAM0076 shapes the transcriptome and optimizes the growth of *Methanococcus maripaludis*. *PLoS Genet.* 15:e1008328. doi: 10.1371/journal.pgen.1008328

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lorenz, R., Bernhart, S. H., Zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Lykke-Andersen, J., and Garrett, R. A. (1997). RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. *EMBO J.* 16, 6290–6300. doi: 10.1093/emboj/16.20.6290

Maier, L. K., Stachler, A.-E., Saunders, S. J., Backofen, J., and Marchfelder, A. (2015). An active immune defense with a minimal CRISPR (clustered regularly interspaced short palindromic repeats) RNA and without the Cas6 protein. *J. Biol. Chem.* 290, 4192–4201. doi: 10.1074/jbc.M114.617506

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Marck, C., and Grosjean, H. (2003). Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA* 9, 1516–1531. doi: 10.1261/rna.5132503

Miller, J. H. (1972). *Experiments in Molecular Genetics*. New York: Cold Spring Harbour Laboratory Press.

Mitchell, M., Xue, S., Erdman, R., Randau, L., Söll, D., and Li, H. (2009). Crystal structure and assembly of the functional Nanoarchaeum equitans tRNA splicing endonuclease. *Nucleic Acids Res.* 37, 5793–5802. doi: 10.1093/nar/gkp537

Mullakhanbhai, M. F., and Larsen, H. (1975). *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* 104, 207–214. doi: 10.1007/bf00447326

Nakamura, M., and Sugiura, M. (2007). Translation efficiencies of synonymous codons are not always correlated with codon usage in tobacco chloroplasts. *Plant J.* 49, 128–134. doi: 10.1111/j.1365-313X.2006.02945.x

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4:2. doi: 10.1186/1746-1448-4-2

Otto, C., Stadler, P. F., and Hoffmann, S. (2014). Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics* 30, 1837–1843. doi: 10.1093/bioinformatics/btu146

Pfeiffer, F., Broicher, A., Gillich, T., Klee, K., Mejia, J., Rampp, M., et al. (2008). Genome information management and integrated data analysis with HaloLex. *Arch. Microbiol.* 190, 281–299. doi: 10.1007/s00203-008-0389-z

Qi, L., Li, J., Jia, J., Yue, L., and Dong, X. (2020). Comprehensive analysis of the pre-ribosomal RNA maturation pathway in a methanoarchaeon exposes the conserved circularization and linearization mode in archaea. *RNA Biol.* 17, 1427–1441. doi: 10.1080/15476286.2020.1771946

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Schiffer, S., Rösch, S., and Marchfelder, A. (2002). Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. *EMBO J.* 21, 2769–2777. doi: 10.1093/emboj/21.11.2769

Schulze, S., Adams, Z., Cerletti, M., De Castro, R., Ferreira-Cerca, S., Fufezan, C., et al. (2020). The Archaeal Proteome Project advances knowledge about archaeal cell biology through comprehensive proteomics. *Nat. Commun.* 11:3145. doi: 10.1038/s41467-020-16784-7

Singh, S. K., Gurha, P., Tran, E. J., Maxwell, E. S., and Gupta, R. (2004). Sequential 2'-O-methylation of archaeal pre-tRNATrp nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA. *J. Biol. Chem.* 279, 47661–47671. doi: 10.1074/jbc.m408868200

Späth, B., Schubert, S., Lieberoth, A., Settele, F., Schütz, S., Fischer, S., et al. (2008). Two archaeal tRNase Z enzymes: similar but different. *Arch. Microbiol.* 190, 301–308. doi: 10.1007/s00203-008-0368-4

Stachler, A. E., and Marchfelder, A. (2016). Gene repression in Haloarchaea using the CRISPR (clustered regularly interspaced short palindromic repeats) - Cas I-B system. *J. Biol. Chem.* 291, 15226–15242. doi: 10.1074/jbc.m116.724062

Stachler, A. E., Schwarz, T. S., Schreiber, S., and Marchfelder, A. (2019). CRISPRi as an efficient tool for gene repression in archaea. *Methods* 172, 76–85. doi: 10.1016/j.ymeth.2019.05.023

Steiger, M. A., Jackman, J. E., and Phizicky, E. M. (2005). Analysis of 2'-phosphotransferase (Tpt1p) from Saccharomyces cerevisiae: evidence for a conserved two-step reaction mechanism. *RNA* 11, 99–106. doi: 10.1261/rna.7194605

Taboada, B., Ciria, R., Martinez-Guerrero, C. E., and Merino, E. (2012). ProOpDB: prokaryotic operon database. *Nucleic Acids Res.* 40, D627–D631. doi: 10.1093/nar/gkr1020

Tang, T. H., Rozhdestvensky, T. S., d'Orval, B. C., Bortolin, M. L., Huber, H., Charpentier, B., et al. (2002). RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res.* 30, 921–930. doi: 10.1093/nar/30.4.921

Thompson, L. D., and Daniels, C. J. (1990). Recognition of exon-intron boundaries by the *Halobacterium volcanii* tRNA intron endonuclease. *J. Biol. Chem.* 265, 18104–18111.

Watanabe, Y., Yokobori, S., Inaba, T., Yamagishi, A., Oshima, T., Kawarabayasi, Y., et al. (2002). Introns in protein-coding genes in Archaea. *FEBS Lett.* 510, 27–30. doi: 10.1016/s0014-5793(01)03219-7

Yokobori, S., Itoh, T., Yoshinari, S., Nomura, N., Sako, Y., Yamagishi, A., et al. (2009). Gain and loss of an intron in a protein-coding gene in Archaea: the case of an archaeal RNA pseudouridine synthase gene. *BMC Evol. Biol.* 9:198. doi: 10.1186/1471-2148-9-198

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 December 2020
doi: 10.3389/fmicb.2020.582022

The 23S Ribosomal RNA From *Pyrococcus furiosus* Is Circularly Permuted

*Ulf Birkedal[1], Bertrand Beckert[2], Daniel N. Wilson[2] and Henrik Nielsen[1,3]**

[1]*Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark,* [2]*Institut für Biochemie und Molekularbiologie, Universität Hamburg, Hamburg, Germany,* [3]*Genomics Group, Nord University, Bodø, Norway*

Synthesis and assembly of ribosomal components are fundamental cellular processes and generally well-conserved within the main groups of organisms. Yet, provocative variations to the general schemes exist. We have discovered an unusual processing pathway of pre-rRNA in extreme thermophilic archaea exemplified by *Pyrococcus furiosus*. The large subunit (LSU) rRNA is produced as a circularly permuted form through circularization followed by excision of Helix 98. As a consequence, the terminal domain VII that comprise the binding site for the signal recognition particle is appended to the 5´ end of the LSU rRNA that instead terminates in Domain VI carrying the Sarcin-Ricin Loop, the primary interaction site with the translational GTPases. To our knowledge, this is the first example of a true post-transcriptional circular permutation of a main functional molecule and the first example of rRNA fragmentation in archaea.

Keywords: archaea, *Pyrococcus*, ribosomal RNA processing, circular permutation, fragmented ribosomal RNA

OPEN ACCESS

Edited by:
Sébastien Ferreira-Cerca, University of Regensburg, Germany

Reviewed by:
Xiuzhu Dong, Chinese Academy of Sciences, China
Lennart Randau, University of Marburg, Germany
Akio Kanai, Keio University, Japan

****Correspondence:***
Henrik Nielsen
hamra@sund.ku.dk

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

Received: *10 July 2020*
Accepted: *16 November 2020*
Published: *10 December 2020*

Citation:
Birkedal U, Beckert B, Wilson DN and Nielsen H (2020) The 23S Ribosomal RNA From Pyrococcus furiosus Is Circularly Permuted. Front. Microbiol. 11:582022. doi: 10.3389/fmicb.2020.582022

INTRODUCTION

Synthesis and processing of rRNA is fundamental to all living organisms. A highly conserved feature at the transcriptional level is that the two main species, SSU and LSU rRNAs, are co-transcribed with the consequence that they are expressed directly into 1:1 stoichiometry. 5S rRNA on the other hand can be part of the same transcriptional unit or transcribed elsewhere in the genome. Another common principle applies to early pre-rRNA processing, namely that the 5´ and 3´ ends of each of SSU and LSU rRNA comes together as a prerequisite for their release from the ribosomal precursor. As a consequence, only fully transcribed precursors give rise to mature rRNA. In bacteria, the nucleotides corresponding to the mature rRNA ends as well as immediate flanking regions base pair to form stem structures that are recognized by RNase III, and the mature ends are subsequently formed by exonucleolytic trimming. In archaea, Bulge-Helix-Bulge (BHB) motifs are formed, cleaved by tRNA splicing endonuclease, and ligated by tRNA ligase to form circular intermediates. This is followed by further endonucleolytic cleavages and exonucleolytic trimming to form the mature rRNA ends. In the most studied models of eukarya (yeast and human), the primary transcript is not looped through base pairing and the initial cleavages are not coordinated in a single cleavage activity. Instead, multiple endonucleolytic cleavages and exonucleolytic trimming reactions result in the release of the mature rRNA species. The details of pre-rRNA processing and variations on the general schemes have been the subject

of numerous reviews in bacteria (Deutscher, 2009), archaea (Clouet-d´Orval et al., 2018), and eukarya (Mullineux and Lafontaine, 2012; Henras et al., 2015). Considering the deep evolutionary conservation and central role in cell metabolism, it is fascinating that many variations in the topology of the mature rRNA can be found, including fragmentation, scrambling, and circularization.

Circularization of an RNA molecule followed by linearization by cleavage outside the circularization junction creates a circularly permuted RNA in which the transcriptional order of two continuous sequence elements AB is reorganized into BA. Additionally, such molecules are characterized by having a new sequence junction and a discontinuity compared to the co-transcriptional, linear counterpart. These structural differences may confer different biological properties to the circularly permuted molecular species. Circular permutation has been used as a method to study various properties of RNA molecules, e.g., the order of folding (Pan, 2000; Lease et al., 2007), including the folding (Kitahara and Suzuki, 2009) or tethering of rRNA (Fried et al., 2015; Orelle et al., 2015) and to make molecular tools, e.g., permuted group I introns as vehicles for producing circularized exons (Puttaraju and Been, 1992; Ford and Ares, 1994) or permuted rRNA in a protocol for the incorporation of non-natural nucleoside analogs (Erlacher et al., 2011). In nature, circular permuted RNAs can be produced by RNA processing or by genomic rearrangement. Any of the mechanisms that generate circular RNA (reviewed in Petkovic and Muller, 2015; Chen, 2016; Ebbesen et al., 2016; Panda et al., 2017) can in principle give rise to circularly permuted molecules, but their existence remains to be documented. RNA splicing is particularly potent in generating circular RNA, and the spliceosome can generate circular RNA by back-splicing or internal splicing of lariat intermediates that result from exon skipping reactions. The point of this circularization, however, appears to be the stabilization of the RNA and when the circle is re-opened, e.g., in the case of the miRNA sponge ciRS-7, the circularly permuted product is immediately degraded to liberate bound miRNAs (Hansen et al., 2013). Genomic rearrangements resulting in circularly permuted tRNA genes have been described from six species of algae and one archaeon (Soma, 2014). Here, the 3´ half is transcribed upstream of the 5´ half in a single transcript, that is, circularized and subsequently linearized to restore a conventional tRNA with functional ends. The processing is likely mediated by the tRNA intron splicing enzymes and it has been speculated that the permuted gene organization eliminates viral integration sites, similar to what has been proposed for fragmentation of tRNA genes by introns (Randau and Soll, 2008). In two species of the archaeon *Thermoproteus*, the SRP RNA gene is genomically rearranged, such that the transcription is initiated from a position that is normally found within the gene (Plagens et al., 2015). The precursor has short leader and trailers that form a BHB motif. This is processed by the tRNA splicing machinery to form a circular species that restores a fully functional SRP RNA. Here, the formation of a covalently closed molecule is seen as a mechanism to prevent unfolding of the RNA at extreme growth temperatures. These two latter examples have in common that the tRNA splicing machinery has been recruited to serve functions different from removal of introns.

Here, we demonstrate that the LSU rRNA of the extreme thermophile *Pyrococcus furiosus* is circularly permuted. Unlike the tRNA and SRP RNAs that are rearranged at the DNA level, the LSU rRNA in *P. furiosus* is transcribed as a conventionally organized LSU rRNA within a precursor encoding the SSU and LSU rRNA separated by a tRNA. The LSU is removed by a tRNA splicing-like mechanism that circularizes the LSU exactly at extended 5´ and 3´ ends. However, due to precise excision of helix 98 (H98), the mature LSU RNA appears as a linear RNA species that has helices 99–101 appended to the 5´ end and terminates in helices H94–97 comprising the Sarcin-Ricin Loop (SRL).

MATERIALS AND METHODS

Cells Culture and RNA Extraction

Pyrococcus furiosus cells were grown in 1/2 SME medium (pH 6.5) supplemented with 0.1% yeast extract, 0.1% peptone, and 0.1% starch. The gas phase was N_2/CO_2 (80:20) and the incubation temperature was 95°C. The culture was grown for 30 h to a final cell density of 3×10^7 cells/ml. RNA was extracted by resuspending 0.5 g of pelleted cells in 2.5 ml TMN buffer (50 mM Tris-Cl, pH 7.5, 10 mM $MgCl_2$, and 100 mM NH_4Cl). Cells were run twice through a French press, followed by addition of 4 ml TRIzol (Life Technologies). After 5 min incubation at room temperature, 0.8 ml chloroform was added. The samples were centrifuged at 12,000 × *g* for 5 min and the aqueous phase transferred to a new tube for extraction with phenol/chloroform. Total RNA was then precipitated by ethanol precipitation.

RiboMeth-Seq and Transcriptomics Analyses

The RiboMeth-seq analysis was performed in duplicates according to previously described protocols (Birkedal et al., 2015; Krogh et al., 2017). Briefly, the RNA was degraded by alkaline into short fragments and the 20–40 nt fraction purified from gels. Then, adaptors were ligated to the library fragments using a modified *Arabidopsis* tRNA ligase. Finally, the library was sequenced on the Ion Proton sequencing platform. The reads were mapped to non-coding RNAs annotated in the *P. furiosus* genome (GenBank: CP003685.1) and scored for read-end counts. Analyses of transcriptomics data were based on datasets deposited at the European Nucleotide Archive: SRX501747 (*Thermococcus kodakarensis*), SRX2118858 (*Pyrobaculum aerophilum*), SRX3467357 (*Sulfolobus acidocaldarius*), and SRX5547671 (*Pyrococcus furiosus*).

Northern Blotting and Primer Extension

Northern Blot analysis was performed according to standard protocols as described in Josefsen and Nielsen (2010). The oligonucleotides used as probes were 5S rRNA: 5´-GGA TCG CTG GGG GGC TT, H98: 5´-GCC GGT CGC CCA GGC CCA, and H99–H101: 5´-GCA GGA CCT CGG GCG AT.

AMV (Promega) or SuperScript IV (Invitrogen) reverse transcriptase was used in primer extension experiments according to the information provided by the supplier. Two oligos were used to map the 5′ end corresponding to the H98 excision site (oligo 1: 5′-GCA GGA CCT CGG GCG AT and oligo 2: 5′-ATC CCC GCC CTA TCA ACC GGG TCT T) and two for the conventionally assigned 5′ end (oligo 3: 5′-TAG CGT CCT AGC CCC TCT A and oligo 4: 5′-GGC GGC TTA GCG TCC TA). The RT-PCR experiment was made by first-strand cDNA synthesis of whole cell RNA using AMV reverse transcriptase (Promega) and dN_6 primers, followed by standard PCR.

Figure Making

RNA base pairing schemes were assisted by RNAfold from the ViennaRNA Web Services (http://rna.tbi.univie.ac.at/) and supported by covariance analysis of sequences from species of Thermococcales. Structure figures were prepared using PyMol and UCSF-Chimera.

RESULTS

LSU rRNA in *Pyrococcus furiosus* Has Covalently Joined 5′ and 3′ Ends and Lacks H98

During a RiboMeth-seq analysis of ribose methylation in *P. furiosus*, we observed several anomalies in sequencing read patterns. RiboMeth-seq is a method designed for profiling of ribose methylations in RNA (Birkedal et al., 2015; Krogh and Nielsen, 2019). In brief, it consists of partial alkaline fragmentation of RNA followed by cloning and sequencing based on the 5′ OH and 2′, 3′ cyclic phosphates generated by alkaline cleavage. 2′-*O*-Me protects against alkaline cleavage and, thus, methylated sites can be deduced from counting read-ends in the sequencing of library fragments. The RiboMeth-seq data on methylations in *P. furiosus* are deposited at the NCBI Gene Expression Omnibus database and will be reported elsewhere in the present article collection on archaeal ribosomes. Here, we focus on additional information provided by the RiboMeth-seq analysis on RNA organization. Each internal position of the RNA molecule being analyzed is covered by a 5′ read-end and a 3′ read-end corresponding to sequencing from either end of the library fragments. However, toward the ends of the molecule, reads are only obtained from one end due to a gel purification step in the protocol that removes fragments <20 nt (**Figure 1A**). Thus, 5′ and 3′ ends are recognized by depletion of 3′ or 5′ read-ends, respectively (**Figure 1B**). As a consequence, hidden breaks, e.g., as found in *Tetrahymena* LSU rRNA (Eckert et al., 1978), leave a distinct feature. Furthermore, ends that carry 5′ phosphates result in chimeric reads during cloning (**Figure 1C**), whereas 5′ OH ends do not. In this way, it is possible to distinguish ends originating from the two main types of cleavages, cleavage by transesterification and hydrolytic cleavage. Finally, circularization junctions can be extracted from reads as with any other RNA-seq method. During the analysis of *P. furiosus* rRNA, we observed very low coverage of nucleotides 2,927-2,967 corresponding to LSU H98 (**Figure 2A**) as well as the characteristic 5′ and 3′ read-end features signifying free ends (**Figure 2B**). Moreover, the many apparent single-nucleotide polymorphisms (SNPs) at nucleotides preceding C2968 (**Figure 2B**) suggested a large fraction of chimeric reads that were subsequently confirmed by inspection of actual reads. The chimera had 5′ parts originating from random fragments in the sample and thus derived from ligation during the cloning of library fragments demonstrating that C2968 carried a 5′ phosphate unlike the fragments generated by alkaline cleavage. Thus, H98 (41 nt) appeared to be cleanly excised at its base by a hydrolytic cleavage at the A2967-C2968 junction and likely by a similar reaction at the U2926-C2927 junction although we cannot exclude other mechanisms, e.g., removal of H98 by a combination of endonucleases and exonucleases. Chimeric reads from the latter cleavage reaction were probably depleted because cleavage fragments carrying the 5′ phosphate would be too small to be recovered in the protocol′s gel purification step. Excision of H98 should leave a 3′ LSU rRNA fragment of 129 nt. Due to failure to recover reads corresponding to the annotated ends of LSU rRNA during RiboMeth-seq analysis, we suspected that the terminal 129 nt was appended to the 5′ end. We then remapped all reads to a reorganized cyclic reference sequence. This resulted in consistent mapping with thousands of reads spanning the C3096-G1 junction (**Figure 2C** and **Supplementary Figure S1A**). Thus, we conclude from RiboMeth-seq analysis that the predominant form of LSU rRNA in *P. furiosus* is circularly permuted.

LSU Circularization Involves a BHB Motif and Is Conserved Among Thermococcales

Next, we inspected the *P. furiosus* pre-rRNA sequence for features that could underlie the proposed processing reactions. LSU rRNA and flanking sequences can form two stretches of extensive base pairing separated by a canonical BHB motif that presents the mature LSU rRNA 5′ and 3′ ends for cleavage and subsequent ligation by the tRNA splicing machinery (**Figure 3A**). The BHB motif is composed of a 4-base pair stem flanked by two 3-nt bulges (thus, a 3-4-3 BHB motif). The BHB motif and flanking helices are conserved among Thermococcales, e.g., in *T. kodakarensis* (**Figures 3A,B**). Cleavage of a BHB motif is catalyzed by tRNA splicing endonuclease and occurs at symmetrical positions within the bulges, resulting in 5′ OH and 2′, 3′ cyclic phosphate ends that are subsequently joined by a ligase, presumably a tRNA ligase. This type of processing of LSU rRNA has been proposed in other archaea. However, in *P. furiosus*, the mature ends become joined in contrast to the situation in, e.g., *A. fulgidus* and *S. solfataricus* (Tang et al., 2002), where nucleotides in flanking regions are joined and the mature ends are subsequently formed by endonucleolytic cleavage and exonucleolytic trimming reactions to form free 5′ and 3′ ends. The organization in *P. furiosus* further suggested that the flanking sequences became joined to generate a linear ITS2-3′ ETS molecule and, indeed, a junction demonstrating the existence of such a molecule could be detected among the reads (**Supplementary Figure S1B**).

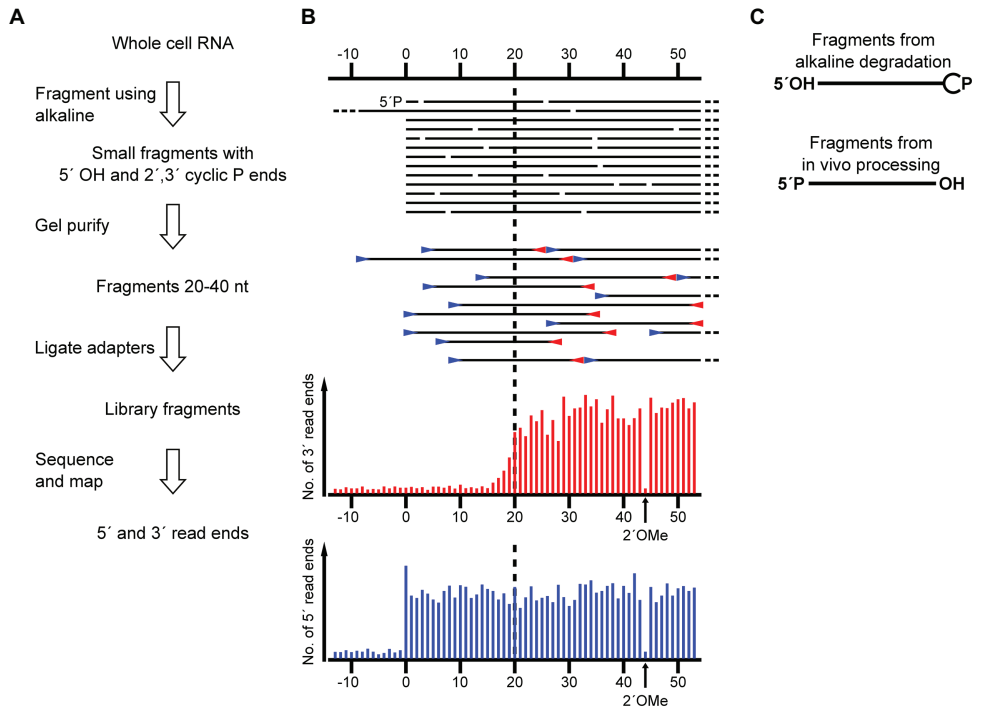
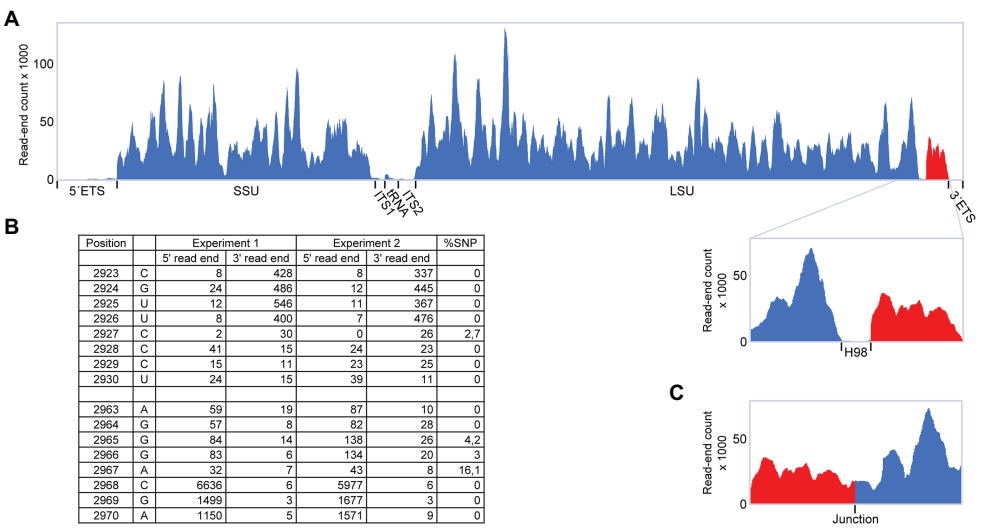


FIGURE 1 | Schematic showing the application of RiboMeth-seq for analysis of native RNA ends. **(A)** Flow diagram of the steps in RiboMeth-seq. **(B)** Drawing of the fragments generated during the RiboMeth-seq protocol and the resulting read end coverage obtained at a native end. Note that the high number of read ends at the position corresponding to the native end as well as the depletion of 3´ read ends corresponding to the first 20 nt of the fragment carrying the native 5´ end. The depletion is caused by the loss of fragments <20 nt in length in the gel purification step. The signature resulting from a single 2´OMe in the RNA is indicated in the read end diagram. **(C)** Drawing highlights the end groups of internal fragments from alkaline degradation contrasted with RNA processing fragments from hydrolytic cleavages.



Position		Experiment 1		Experiment 2		%SNP
		5' read end	3' read end	5' read end	3' read end	
2923	C	8	428	8	337	0
2924	G	24	486	12	445	0
2925	U	12	546	11	367	0
2926	U	8	400	7	476	0
2927	C	2	30	0	26	2,7
2928	C	41	15	24	23	0
2929	C	15	11	23	25	0
2930	U	24	15	39	11	0
2963	A	59	19	87	10	0
2964	G	57	8	82	28	0
2965	G	84	14	138	26	4,2
2966	G	83	6	134	20	3
2967	A	32	7	43	8	16,1
2968	C	6636	6	5977	6	0
2969	G	1499	3	1677	3	0
2970	A	1150	5	1571	9	0

FIGURE 2 | RiboMeth-seq analysis of *Pyrococcus furiosus* rRNA. **(A)** Read-coverage of the entire rRNA operon using a linear reference sequence. The 5´ and 3´ parts are in blue and red, respectively, to emphasize their rearrangement during circularization. **(B)** Table of actual read-end numbers at the borders of H98 showing the signature pattern for free ends at the H98 flanking regions. The fraction of single nucleotide polymorphisms (SNP) at read ends is shown in the rightmost column. **(C)** Read-coverage around the proposed circularization junction after re-mapping of the data in **(A)** using a circularized reference sequence.

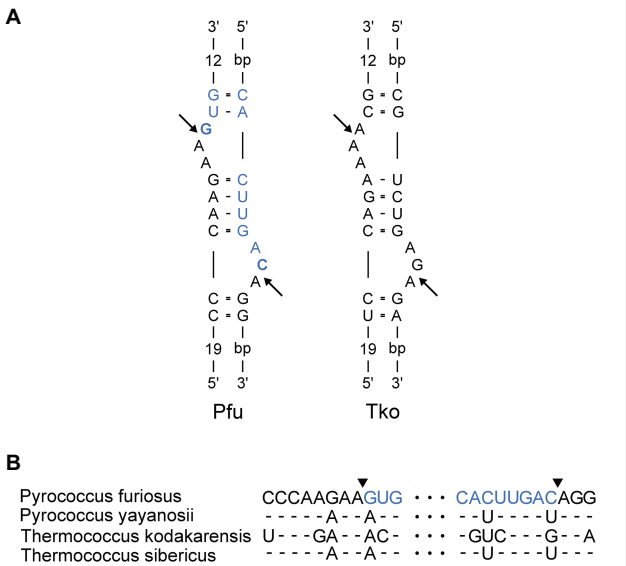


FIGURE 3 | (A) Bulge-Helix-Bulge (BHB) motif involved in processing of *Pyrococcus furiosus* LSU rRNA. The arrows indicate canonical cleavage sites that will result in circularization of the rRNA. The mature rRNA in *Pfu* is colored in blue. The nucleotides that are joined to form the circularization junction in *Pfu* are in bold. Pfu: *Pyrococcus furiosus*. Tko: *Thermococcus kodakarensis*. **(B)** Alignment of sequences in the LSU BHB motif in species of Thermococcales. Arrowheads and color coding in the *Pfu* sequence corresponds to **(A)**.

In contrast to the LSU rRNA, the SSU rRNA appeared to be processed according to the conventional scheme. The 5′-ETS and ITS1 can co-fold to form a BHB motif separating two extended helices of 13 and 19 bp, respectively. The predicted cleavage sites are located 62 nt upstream of the mature SSU rRNA 5′ end and 27 nt downstream of the mature SSU rRNA 3′ end. The circularization sites creating a circular SSU rRNA intermediate and a chimeric 5′-ETS-ITS1 molecule could both be identified as low abundance reads, as expected. Specifically, we counted 95 and 42 reads, respectively, for the two junctions derived from SSU rRNA processing compared to 2,319 reads for the LSU rRNA circularization junction and 11 reads for the ITS2-3′ ETS junction. Further processing to form the mature SSU rRNA ends supposedly involves the action of endonucleases and exonucleases (Tang et al., 2002).

Circularization and H98 Excision Are Supported by Northern Blotting and Primer Extension Experiments

In order to experimentally validate the proposed processing scheme, we first performed northern blotting and primer extension experiments. We were unable to detect the excised H98 by northern blotting analysis of small RNAs (**Figure 4A**) and primer extension experiments (data not shown). Thus, we exclude the possibility that the H98 exists as a prominent free RNA compared to 5S rRNA that was used as a control. Instead, we detected substoichiometric amounts of H98 co-migrating with LSU rRNA in agarose gels (**Figure 4B**). We estimated H98 to be present at <1% of the amount of the LSU rRNA 3′ part that was used as a control. This is consistent with the low number of sequencing reads from H98 in RiboMeth-seq analysis. We conclude that H98 is absent from mature *P. furiosus* ribosomes but can be detected at low levels in processing intermediates or incorrectly processed precursors. In parallel, we analyzed for the presence of the 129 nt 3′ end of LSU rRNA (H99–H101) that would exist as a separate fragment after H98 excision in the absence of ligation of the 5′ and 3′ end of LSU rRNA. We detected the free 129 nt fragment at a level of <5% compared to 5S rRNA based on the hybridization analysis as well as a gel stain (**Figure 4A**). Northern blot analysis of an agarose gel revealed that the bulk of the 3′ end is associated with a conventional LSU rRNA species, as expected from circularization of the precursor transcript (**Figure 4B**). This demonstrated that the ligation step is slightly less efficient than the preceding endonucleolytic step, at least under the present growth conditions. Primer extension analysis confirmed the absence of a 5′ end of LSU rRNA corresponding to pos. 1 on the gene map (primers #3 and #4) and instead demonstrated a strong signal at pos. C2968 (all four primers) in accordance with the proposed idea of a circularly permuted LSU rRNA (**Figure 4C**).

H98 Is Located at the Surface of the Large Subunit and Is Optional in Archaea

H98 is a highly variable helix that is nevertheless found in most organisms, including *Escherichia coli*, yeast, and humans. In bacteria, such as *E. coli*, H98 comprises 15 nucleotides that are located adjacent to the 5′ and 3′ ends of the 23S rRNA in the vicinity of ribosomal proteins L3 and L13 at the back of the large subunit (**Figures 5A,B**). By inspection of sequences deposited in the UCSC archaeal genome browser (Chan et al., 2012), the comparative RNA website (Cannone et al., 2002), and the full-length rRNA organismal alignment (FLORA) database (Doris et al., 2015), it appears that H98 is optional in archaea. It is generally found in species belonging to Crenarchaeota with occasional losses (e.g., in *Ignicoccus hospitalis* among the Desulfurococcales), but absent in the two representatives of Nanoarchaeota, e.g., *Nanoarchaeum equitans*. In Euryarchaeota, it is found in, e.g., Methanococcales (with occasional losses, e.g., in *Methanococcus infernus*) and Thermococcales, but appears absent in, e.g., Methanomicrobiales, Halobacteriales, and Archaeoglobales (**Supplementary Table S1**). This is also confirmed by the X-ray crystal structure of the large subunit from the Halophile *Haloarcula marismortui* (Gabdulkhakov et al., 2013) and the cryo-EM structure of the 50S subunit from the methanophile *Methanothermobacter thermautotrophicus* (Greber et al., 2012), where H98 is replaced by a small linker of four nucleotides in both cases (**Figures 5C,D**). Our observations add further complexity to this picture, because we show that H98 is encoded within the Thermococcales, but subsequently excised and degraded.

Our observations of the sequence structure of *P. furiosus* LSU rRNA appear inconsistent with a published cryo-EM model of

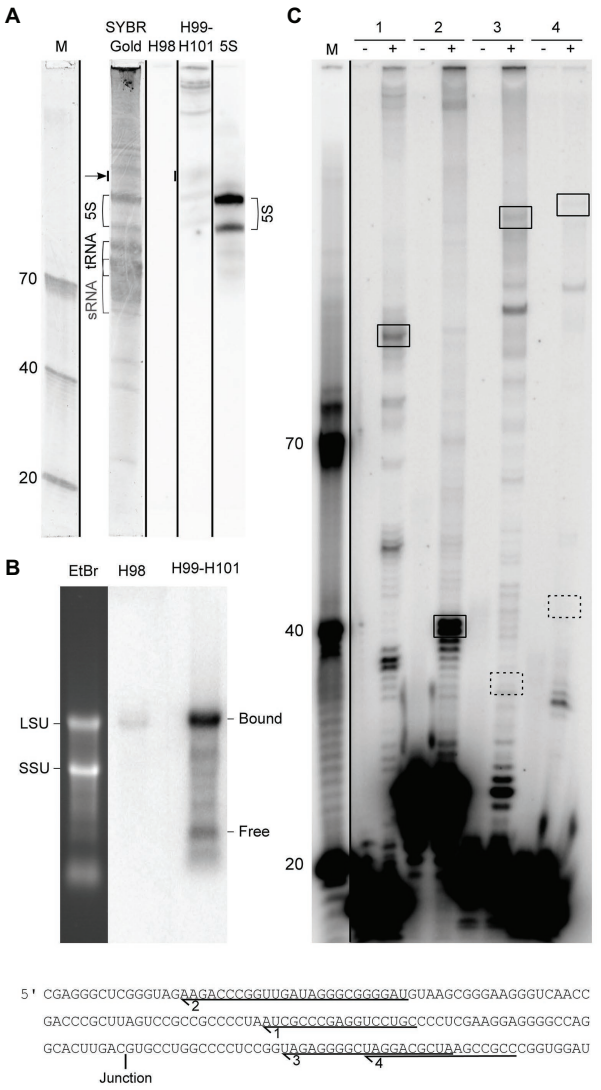


FIGURE 4 | Experimental analysis of LSU rRNA circularization and H98 removal. **(A)** Northern blot analysis of the small RNA fraction of whole cell RNA fractionated on a 10% UPAG. The DNA oligonucleotide size marker (M) is followed by a SYBR Gold stained lane with indications of sRNA, tRNA, and 5S rRNA, followed by parallel hybridizations using probes against H98, H99–H101, and 5S rRNA (control). The lanes are separated by black lines to indicate that they were parallel lanes from the same gel. The arrow indicates the signal that was used to estimate the substoichiometric (compared to 5S rRNA) amounts of free 3′ terminal LSU rRNA fragment (129 nt; comprising H99-H101). The two forms of 5S rRNA detected may represent the two copies of the gene annotated with slightly different lengths in the genome browser (Chan et al., 2012). **(B)** Northern blot analysis of whole cell RNA fractionated on a 1% FA-agarose gel analyzed in parallel with H98 and H99–H101 probes. The positions of SSU and LSU rRNA, respectively, are marked on the gel stain, and the positions of bound (circularized) and free 3′ terminal LSU rRNA fragment (corresponding to H99-H101) are marked on the northern blots. **(C)** Primer extension analysis of the 5′ end of mature LSU rRNA. Comparison of the expected signals deriving from circularization and H98 removal (boxed with full lines) and conventional processing as inferred from the literature (boxed with dashed lines). The sequence in the lower part shows the location of the primers used in the experiment in relation to the circularization junction. (–) and (+) refer to addition of reverse transcriptase. M: DNA oligonucleotide marker.

ribosomes at 6.6 Å resolution (Armache et al., 2013). Here, the cryo-EM data were modeled based on GenBank Acc. no. AE009950 and appears to lack 22 nt at the 5′ end and 25 nt at the 3′ end, respectively, compared to our sequence derived from rRNA sequencing with very high coverage. Moreover, re-examining the cryo-EM maps for the *P. furiosus* 70S ribosome reveals that the density in this region does not support the model proposed in the paper. Instead, the density would correspond with H98 being absent and the additional density being assigned to the extensions present at the 5′ and 3′ ends of the 23S rRNA (**Figure 5E**). Similarly, the examination of cryo-EM map of *T. kodakarensis* 70S ribosome at 16 Å resolution (Armache et al., 2013) also reveals the electron density that would be consistent with the absence of H98 and nucleotide extensions circularizing the 5′ and 3′ ends of the 23S rRNA (**Figure 5F**). Recently, high resolution (2.5-3.0 Å) cryo-EM structures of T. kodakarensis 70S ribosome were reported (Sas-Chen et al., 2020), where structural models for H98 are lacking due to the lack of density for H98. Re-analysis of the cryo-EM map, including low-pass filtering, revealed additional density extending from the 5′ and 3′ ends (**Supplementary Figures S2A**,**B**), analogous to that observed in the previous *T. kodakarensis* 70S ribosome structure (Armache et al., 2013). These observations are supported by re-analysis of existing transcriptomics data that confirm the circularization junction in mature LSU rRNA as well as the absence of H98 (**Supplementary Figure S2**). Thus, in conclusion, the cryo-EM density maps support the excision of H98 from the Thermococcales of the Euryarchaeota, consistent with the high-throughput sequencing data.

DISCUSSION

We have shown by high-throughput sequencing as well as by northern blotting and primer extension that the main form of LSU rRNA in *P. furiosus* is circularly permuted. Based on our experiments, we propose a model for pre-rRNA processing in *P. furiosus* and other species of Thermococcales that is conventional with respect to formation of SSU rRNA as well as the tRNA located between the SSU and LSU rRNAs but has two unconventional features with respect to processing of LSU rRNA (**Figure 6A**). First, the LSU rRNA is released by a tRNA splicing-like mechanism exactly at its 5′ and 3′ ends. As a consequence of the release mechanism, the ends are covalently joined. Second, H98 is excised, likely by two coordinated hydrolytic cleavage reactions and subsequently degraded. Thus, LSU rRNA in *P. furiosus* exists as a circularly permuted RNA.

Circularization of RNA is widespread in archaea (Danan et al., 2012) and *P. furiosus* appears to be particularly prone to RNA circularization, e.g., with abundant circularization of box C/D RNAs (Starostina et al., 2004). Recently, an RNA ligase that ligates 5′-P and 3′-OH ends was characterized from *Pyrococcus abyssi* (the *Pab* 1,020 RNA ligase) and proposed to circularize many different cellular RNAs (Becker et al., 2017). The excision of H98 in *P. furiosus* curiously resembles a step in rRNA processing in chloroplasts. Here, the 3′ terminal part of LSU rRNA corresponding to helices 99–101 is encoded as

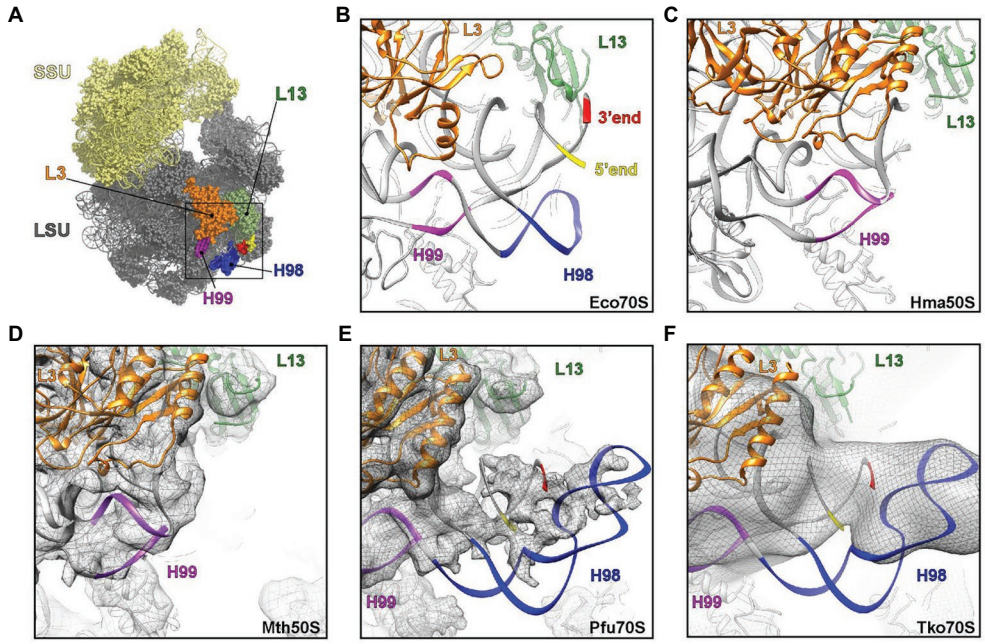


FIGURE 5 | Visualization of H98 in structures of ribosomal particles from bacteria and archaea. **(A)** Overview of the *E. coli* ribosome with the SSU in yellow and the LSU in gray. Ribosomal proteins L3 (orange) and L13 (green) are highlighted for reference as are 23S rRNA helices H98 (blue), H99 (purple), and the 5′ (red) and 3′ (yellow) ends. **(B–F)** View of the large ribosomal subunit highlighting the region, where H98 (blue), H99 (purple), and the 5′ and 3′ ends are located in **(B,C)** the crystal structures of the **(B)** *E. coli* 70S ribosome (PDB ID 5GMP) and the **(C)** *Haloarcula marismortui* 50S subunit (PDB ID 4V9F), as well as cryo-electron microscopy structures of the **(D–F)** *Methanothermobacter thermautotrophicus* 50S subunit (EMD-2012, PDB ID 4ADX), **(E)** *Pyrococcus furiosus* 70S ribosome (EMD-2009, PDB ID 4V6U) and the **(F)** *Thermococcus kodakarensis* 70S ribosome (EMD-2170 with PDB ID 4V6U fitted). In **(D–F)**, the cryo-EM map density is shown as a gray mesh.

a 4.5S rRNA that is separated from the bulk of the LSU rRNA by an internal transcribed spacer (Whitfeld et al., 1978), much like 5.8S rRNA that is encoded as the 5′ part of LSU rRNA in eukaryotes. The location of the internal transcribed spacer coincides with H98. In spinach chloroplasts, the spacer comprises 115 nt and can adopt a compact structure with a base-paired stem. Although the mechanism of removal of this spacer is unknown, it is formally very similar to H98 excision in *P. furiosus* ribosomes.

The circularization reaction and the excision of H98 are unlikely to be mechanistically coupled because they occur at distant sites and by different mechanisms. However, the reactions may be linked through a conformational switching mechanism (Nagel and Pleij, 2002). Prior to circularization, the BHB motif is flanked by two extraordinarily stable helices of 21 and 14 bp. The cleavage and ligation reactions remove the 21 bp lower of the two helices, and we speculate that the upper helix can now switch into a conformation in which the two strands fold back on themselves to form two independent helices flanking the circularization junction (**Figure 6B**). The two proposed alternative structures have roughly the same number of base pairs. Importantly, the conformational switch changes the structure of the junction from which H98 emerge (**Figure 6C**), and we speculate that only one of the two structures presents the base of H98 for cleavage. The structure of H98 and its flanking sequences as well as the possibility to undergo conformational switching upon circularization is conserved in *T. kodakarensis*. If circularization in this scheme precedes excision of H98, it should be possible to capture full-length circular molecules by RT-PCR. Unfortunately, this proved technically challenging. The opposite order of reactions, i.e., H98 excision prior to circularization, would lead to transient existence of two large cleavage fragments of the ribosomal RNA precursor. We did not detect such molecular species at sensitivity of analysis that would reveal rare to medium abundant mRNAs. Thus, the experimental discrimination between the two possibilities remains unsettled.

From an evolutionary point of view, the reorganization of LSU rRNA may have resulted from loss of a processing endonuclease that processed the circular intermediate following processing at the BHB motif. H1 is known to interact with H98 in archaea and the extended H1 or other structures resulting from the extended sequence (**Figure 6B**) may have structurally clashed with H98 driving evolution toward H98 elimination. In a recent paper on expansion segments in Asgard archaea (Penev et al., 2020), the *P. furiosus* is highlighted because it is supposed to have the longest H98 among archaea. H98 is the basis of ES39 that is "supersized" in Asgard archaea and of considerable length in eukaryotes. *Pyrococcus furiosus* carries a "μ-ES39" and its predicted structure presents a problem to the accretion model of rRNA expansion because it does not overlap in 3D structure when compared to *E. coli* in which H98 interacts within the H1 minor groove through an A-minor interaction. We propose instead that the lack of H98 provides space for an extended structure replacing H1, which is opposite

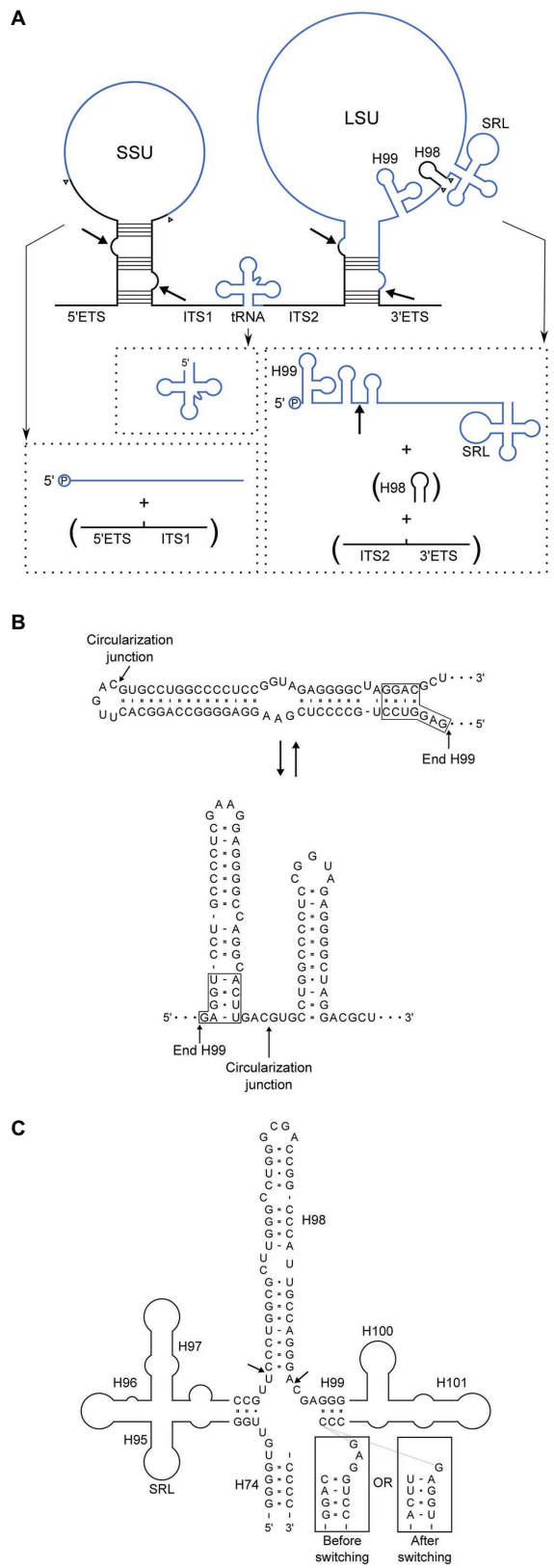


FIGURE 6 | (Continued)

FIGURE 6 | (A) Model of pre-rRNA processing in *Pyrococcus furiosus* including the rRNA precursor and final products. By-products that are turned-over are in parentheses. Intermediates, e.g., putative circular forms, were not included because the order of processing steps was not determined with certainty. Arrows indicate the proposed processing sites at BHB motifs and the circularization junction in the mature LSU rRNA, and arrowheads show the final processing sites for generation of the mature products. End-products are in blue and spacers in black. **(B)** Two alternatives base pairing schemes after circularization. The upper correspond to H1 in conventional annotation extended by 15 bp and closed by a loop. In the lower structure this helix is resolved into two helices spanning the circularization junction. **(C)** Structural context of H98 removal. H98 is located between H95 and H97 comprising the Sarcin-Ricin Loop (SRL) and H99–H101 involved in binding of the signal recognition particle (SRP). The arrows indicate the processing sites mapped by RiboMeth-seq analysis. The boxed sequences are similar to those in **(B)** and are appended to the 3′-strand of H99 to show how the alternative base pairing affects the structure of the junction at which H98 excision takes place.

to eukaryotes in which H1 is lacking giving way for a much expanded H98-ES39.

A recent, comprehensive review described the endoribonucleases and exoribonucleases known from archaea (Clouet-d´Orval et al., 2018). Whereas the enzymes responsible for the circularization step most likely are the tRNA splicing enzymes, the enzyme(s) responsible for H98 excision are unknown. The present data provide two clues to the activity. First, the observation of chimeric reads in RiboMeth-seq suggests that cleavages occur by hydrolysis leaving 5´-phosphate and 3´-OH. Second, the conformational switching model suggests that a particular structure at the cleavage junction may be induced. Furthermore, the excision removes a double stranded region and leaves unpaired ends, which argue for an initial endonucleolytic cleavage followed by exonucleolytic degradation of H98 and structural protection of the mature ends. Several endonucleases involved in rRNA maturation are known from bacteria (Taverniti et al., 2011; Clouet-d´Orval et al., 2015; Vercruysse et al., 2016). Of particular interest, PPR-SMR proteins are responsible for some endonucleolytic cleavages in chloroplast rRNA processing (Zhou et al., 2017) and related proteins exist in archaea, but the activity specifically responsible for processing of the spacer upstream of 4.5S RNA, that resembles H98 removal, remains to be characterized. RNase III is sporadic in archaea (Nicholson, 2014) but a protein of unknown function comprising an RNase III domain has been annotated in *Pyrococcus*. Recently, the RNase E-like FAU-1 endonuclease from *Pyrococcus* and *Thermococcus* was shown to be involved in pre-5S rRNA processing (Ikeda et al., 2017). It was also observed that LSU rRNA in ΔFAU-1 cells is approximately 50 nt longer than wt LSU rRNA, roughly corresponding to the length of H98. However, there is no obvious resemblance between pre-5S rRNA and the invoked structures of the sequences flanking H98. Future experiments should be directed toward experimental characterization of the activity based on cell extracts and model RNA substrates as well as recent developments in *in vivo* characterization of pre-rRNA processing (Juttner et al., 2020).

Processing of LSU rRNA in archaea is not understood in full detail despite recent progress (Grünberger et al., 2019; Juttner et al., 2020; Qi et al., 2020). However, clues to the diversity in processing can be obtained from both RiboMeth-seq

analyses and the existing transcriptomics data obtained for other purposes. From RiboMeth-seq studies, we conclude that species of Methanococcales as well as *I. hospitalis* and *N. equitans* have free 5′ and 3′ ends and are uninterrupted in the H97-(H98)-H99 region (data not shown). In addition to the circularly permuted forms in *P. furiosus* and *T. kodakarensis* (**Figure 2** and **Supplementary Figure S2**), we find, by analysis of transcriptomics data deposited in the European Nucleotide Archive, that LSU rRNA from *P. aerophilum* is circularized at a BHB motif located at the mature ends and retains H98 (**Supplementary Figure S3**). *S. acidocaldarius* is well-studied and has a canonical BHB motif at a distance from the proposed mature LSU rRNA ends and are thought to employ endonucleases and exonucleases to further process the circular intermediate formed at the BHB motif (Tang et al., 2002). However, processing at these sites have not been characterized, and although coverage is relatively low in this region, we find more that 10,000 reads spanning the circularization junction and too few read ends to support free LSU rRNA ends in this region (**Supplementary Figure S4**), suggesting that this species is the mature form rather than a processing intermediate. The circularized species has previously been demonstrated as an abundant species (Danan et al., 2012; Juttner et al., 2020). H98 is clearly present, suggesting that LSU rRNA in *S. acidocaldarius* remains circular. However, this must await direct characterization by independent methods. It is a characteristic of LSU rRNA processing that the two ends come together at an early stage of ribosome formation. The covalent joining of the ends through rRNA circularization demonstrated in this study presents the most radical of strategies toward this.

Fragmentation of rRNA has not been reported from archaea but is not uncommon and can result from genome organization or rRNA processing. In the mitochondria of *Tetrahymena pyriformis*, the 5′ end of LSU rRNA is transcribed downstream of the main gene body giving rise to a discontinuous rRNA synthesized in an unconventional order (Heinonen et al., 1987). Many bacteria transcribe pre-rRNA that becomes fragmented through the removal of intervening sequences by RNase III cleavage and trimming by other RNases (Evguenieva-Hackenberg, 2005). RNA fragmentation has been proposed to regulate the ribosome concentration by allowing faster turn-over (Hsu et al., 1994). This type of fragmentation is reminiscent of the "hidden gaps" found in protozoans and insects (Gray and Schnare, 1996). More dramatic fragmentation through removal of several intervening sequences is seen in, e.g., *Euglena gracilis* (Schnare and Gray, 2011) and in the LSU rRNA of several Trypanosomes (Hernandez and Cevallos, 2014; Shalev-Benami et al., 2016; Zhang et al., 2016; Liu et al., 2017).

As a consequence of the processing reactions in *P. furiosus*, LSU rRNA is circularly permuted and has a dramatic re-organization of two key elements in the ribosome. First, helices 99–101 are appended to the 5′ end of the molecule. H100 (or its structural equivalent) is involved in docking of the signal recognition particle (SRP) in bacteria and chloroplasts (Bieri et al., 2017). Second, helices H94–97 in Domain VI comprising H95 with the Sarcin-Ricin Loop (SRL) are placed at the very 3′ end. The SRL is the primary site of interaction between the ribosome and translational GTPases during protein synthesis. One intriguing possibility is that these elements of the ribosome are incorrectly positioned prior to the excision of H98, such that this step could serve as a late quality control for release of functional ribosomes. Quality control steps in ribosome biogenesis based on RNA cleavage or RNA modification are widespread (reviewed in Mullineux and Lafontaine, 2012; Sloan et al., 2016). The reorganization of the LSU rRNA may also serve as an adaptation for growth at elevated temperatures. The circularization may ensure the association of the 5′ and 3′ ends during ribosome biogenesis or protect the ends in the mature ribosome from accessibility due to thermal melting, as suggested for archaeal SRP RNA (16; Plagens et al., 2015). Alternative mechanisms may serve to protect the ends originating from H98 excision. Cryo-EM structures of spinach chloroplast ribosomes reveal an interaction of the 5′ end of 4.5S RNA and the 3′ end of the upstream LSU rRNA stabilized by an N-terminal extension of uL13 (Bieri et al., 2017; Graf et al., 2017). Such an extension is not present on *Pyrococcus* L13, where other strategies may apply.

In conclusion, we have demonstrated the existence of a naturally occurring, circularly permuted, and functional rRNA in the thermophilic archaeon, *P. furiosus*. This organism is widely studied and the organization of its rRNA was not anticipated in previous studies. However, re-analysis of existing data, e.g. from (Grünberger et al., 2019; **Supplementary Figure S5**) support our model. It will be of interest to study the structural and functional consequences on the ribosome of this re-organization of rRNA, and its possible role in adaptation of this organism to existence in a harsh environment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at https://www.ncbi.nlm.nih.gov/ and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153501. Coordinates for the mentioned cryo-EM structures can be found in the Protein Data bank under accession numbers 5GMP (*E. coli* 70S ribosome), 4V9F (*H. marismortui* 50S subunit), 4ADX, (*M. thermautotrophicus* 50S subunit), 4V6 (*P. furiosus* 70S ribosome), and 4V6U (*T. kodakarensis* 70S ribosome).

AUTHOR CONTRIBUTIONS

UB and HN planned the RiboMeth-seq and biochemical experiments that were conducted by UB. DW supervised the cryo-EM analysis. DW and BB prepared the structure Figures. HN wrote the first manuscript draft. All authors contributed to the article and approved the submitted version.

FUNDING

Work in the Nielsen lab was supported by the Danish Council for Independent Research (DFF-0602-02396 and DFF-4183-00486)

and the Lundbeck Foundation (R198–2015-174). Funding for open access charge: the Danish Council for Independent Research (DFF-0602-02396 and DFF-4183-00486).

ACKNOWLEDGMENTS

Whole cell RNA from *P. furiosus* was provided by Stephen Douthwaite (University of Southern Denmark). The RNA was extracted from cell pellets provided by Harald Huber (University of Regensburg).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.582022/full#supplementary-material

REFERENCES

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

Becker, H. F., Heliou, A., Djaout, K., Lestini, R., Regnier, M., and Myllykallio, H. (2017). High-throughput sequencing reveals circular substrates for an archaeal RNA ligase. *RNA Biol.* 14, 1075–1085. doi: 10.1080/15476286.2017.1302640

Bieri, P., Leibundgut, M., Saurer, M., Boehringer, D., and Ban, N. (2017). The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J.* 36, 475–486. doi: 10.15252/embj.201695959

Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J., and Nielsen, H. (2015). Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed. Eng.* 54, 451–455. doi: 10.1002/anie.201408362

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

Chan, P. P., Holmes, A. D., Smith, A. M., Tran, D., and Lowe, T. M. (2012). The UCSC Archaeal genome browser: 2012 update. *Nucleic Acids Res.* 40, D646–D652. doi: 10.1093/nar/gkr990

Chen, L. L. (2016). The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 17, 205–211. doi: 10.1038/nrm.2015.32

Clouet-d´Orval, B., Batista, M., Bouvier, M., Quentin, Y., Fichant, G., Marchfelder, A., et al. (2018). Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiol. Rev.* 42, 579–613. doi: 10.1093/femsre/fuy016

Clouet-d´Orval, B., Phung, D. K., Langendijk-Genevaux, P. S., and Quentin, Y. (2015). Universal RNA-degrading enzymes in Archaea: prevalence, activities and functions of beta-CASP ribonucleases. *Biochimie* 118, 278–285. doi: 10.1016/j.biochi.2015.05.021

Danan, M., Schwartz, S., Edelheit, S., and Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* 40, 3131–3142. doi: 10.1093/nar/gkr1009

Deutscher, M. P. (2009). Maturation and degradation of ribosomal RNA in bacteria. *Prog. Mol. Biol. Transl. Sci.* 85, 369–391. doi: 10.1016/S0079-6603(08)00809-X

Doris, S. M., Smith, D. R., Beamesderfer, J. N., Raphael, B. J., Nathanson, J. A., and Gerbi, S. A. (2015). Universal and domain-specific sequences in 23S-28S ribosomal RNA identified by computational phylogenetics. *RNA* 21, 1719–1730. doi: 10.1261/rna.051144.115

Ebbesen, K. K., Kjems, J., and Hansen, T. B. (2016). Circular RNAs: identification, biogenesis and function. *Biochim. Biophys. Acta* 1859, 163–168. doi: 10.1016/j.bbagrm.2015.07.007

Eckert, W. A., Kaffenberger, W., Krohne, G., and Franke, W. W. (1978). Introduction of hidden breaks during rRNA maturation and ageing in *Tetrahymena pyriformis*. *Eur. J. Biochem.* 87, 607–616. doi: 10.1111/j.1432-1033.1978.tb12413.x

Erlacher, M. D., Chirkova, A., Voegele, P., and Polacek, N. (2011). Generation of chemically engineered ribosomes for atomic mutagenesis studies on protein biosynthesis. *Nat. Protoc.* 6, 580–592. doi: 10.1038/nprot.2011.306

Evguenieva-Hackenberg, E. (2005). Bacterial ribosomal RNA in pieces. *Mol. Microbiol.* 57, 318–325. doi: 10.1111/j.1365-2958.2005.04662.x

Ford, E., and Ares, M. Jr. (1994). Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4. *Proc. Natl. Acad. Sci. U. S. A.* 91, 3117–3121. doi: 10.1073/pnas.91.8.3117

Fried, S. D., Schmied, W. H., Uttamapinant, C., and Chin, J. W. (2015). Ribosome subunit stapling for orthogonal translation in *E. coli*. *Angew. Chem. Weinheim Bergstr. Ger.* 127, 12982–12985. doi: 10.1002/anie.201506311

Gabdulkhakov, A., Nikonov, S., and Garber, M. (2013). Revisiting the *Haloarcula marismortui* 50S ribosomal subunit model. *Acta Crystallogr. D Biol. Crystallogr.* 69, 997–1004. doi: 10.1107/S0907444913004745

Graf, M., Arenz, S., Huter, P., Donhofer, A., Novacek, J., and Wilson, D. N. (2017). Cryo-EM structure of the spinach chloroplast ribosome reveals the location of plastid-specific ribosomal proteins and extensions. *Nucleic Acids Res.* 45, 2887–2896. doi: 10.1093/nar/gkw1272

Gray, M. W., and Schnare, M. N. (1996). "Evolution of rRNA gene organization" in *Ribosomal RNA*. eds. R. Zimmermann and A. Dahlberg (Boca Raton, Florida: CRC Press), 49–69.

Greber, B. J., Boehringer, D., Godinic-Mikulcic, V., Crnkovic, A., Ibba, M., Weygand-Durasevic, I., et al. (2012). Cryo-EM structure of the archaeal 50S ribosomal subunit in complex with initiation factor 6 and implications for ribosome evolution. *J. Mol. Biol.* 418, 145–160. doi: 10.1016/j.jmb.2012.01.018

Grünberger, F., Knüppel, R., Jüttner, M., Fenk, M., Borst, A., Reichelt, R., et al. (2019). Exploring prokaryotic transcription, operon structures, rRNA maturation and modifications using nanopore-based native RNA sequencing. bioRxiv. [Preprint]. doi: 10.1101/2019.12.18.880849

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993

Heinonen, T. Y., Schnare, M. N., Young, P. G., and Gray, M. W. (1987). Rearranged coding segments, separated by a transfer RNA gene, specify the two parts of a discontinuous large subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J. Biol. Chem.* 262, 2879–2887.

Henras, A. K., Plisson-Chastang, C., O´Donohue, M. F., Chakraborty, A., and Gleizes, P. E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* 6, 225–242. doi: 10.1002/wrna.1269

Hernandez, R., and Cevallos, A. M. (2014). Ribosomal RNA gene transcription in trypanosomes. *Parasitol. Res.* 113, 2415–2424. doi: 10.1007/s00436-014-3940-7

Hsu, D., Shih, L. M., and Zee, Y. C. (1994). Degradation of rRNA in *Salmonella* strains: a novel mechanism to regulate the concentrations of rRNA and ribosomes. *J. Bacteriol.* 176, 4761–4765. doi: 10.1128/jb.176.15.4761-4765.1994

Ikeda, Y., Okada, Y., Sato, A., Kanai, T., Tomita, M., Atomi, H., et al. (2017). An archaeal RNA binding protein, FAU-1, is a novel ribonuclease related to rRNA stability in *Pyrococcus* and *Thermococcus*. *Sci. Rep.* 7:12674. doi: 10.1038/s41598-017-13062-3

Josefsen, K., and Nielsen, H. (2010). "Northern blotting analysis" in *RNA: Methods and protocols.* ed. H. Nielsen (New York, Dordrecht, Heidelberg, London: Humana Press), 87–105.

Juttner, M., Weiss, M., Ostheimer, N., Reglin, C., Kern, M., Knuppel, R., et al. (2020). A versatile cis-acting element reporter system to study the function, maturation and stability of ribosomal RNA mutants in archaea. *Nucleic Acids Res.* 48, 2073–2090. doi: 10.1093/nar/gkz1156

Kitahara, K., and Suzuki, T. (2009). The ordered transcription of RNA domains is not essential for ribosome biogenesis in *Escherichia coli*. *Mol. Cell* 34, 760–766. doi: 10.1016/j.molcel.2009.05.014

Krogh, N., Birkedal, U., and Nielsen, H. (2017). RiboMeth-seq: profiling of 2´-O-me in RNA. *Methods Mol. Biol.* 1562, 189–209. doi: 10.1007/978-1-4939-6807-7_13

Krogh, N., and Nielsen, H. (2019). Sequencing-based methods for detection and quantitation of ribose methylations in RNA. *Methods* 156, 5–15. doi: 10.1016/j.ymeth.2018.11.017

Lease, R. A., Adilakshmi, T., Heilman-Miller, S., and Woodson, S. A. (2007). Communication between RNA folding domains revealed by folding of circularly permuted ribozymes. *J. Mol. Biol.* 373, 197–210. doi: 10.1016/j.jmb.2007.07.007

Liu, Z., Gutierrez-Vargas, C., Wei, J., Grassucci, R. A., Sun, M., Espina, N., et al. (2017). Determination of the ribosome structure to a resolution of 2.5 a by single-particle cryo-EM. *Protein Sci.* 26, 82–92. doi: 10.1002/pro.3068

Mullineux, S. T., and Lafontaine, D. L. (2012). Mapping the cleavage sites on mammalian pre-rRNAs: where do we stand? *Biochimie* 94, 1521–1532. doi: 10.1016/j.biochi.2012.02.001

Nagel, J. H., and Pleij, C. W. (2002). Self-induced structural switches in RNA. *Biochimie* 84, 913–923. doi: 10.1016/s0300-9084(02)01448-7

Nicholson, A. W. (2014). Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip. Rev. RNA* 5, 31–48. doi: 10.1002/wrna.1195

Orelle, C., Carlson, E. D., Szal, T., Florin, T., Jewett, M. C., and Mankin, A. S. (2015). Protein synthesis by ribosomes with tethered subunits. *Nature* 524, 119–124. doi: 10.1038/nature14862

Pan, T. (2000). Probing RNA structure and function, by circular permutation. *Methods Enzymol.* 317, 313–330. doi: 10.1016/s0076-6879(00)17022-3

Panda, A. C., Grammatikakis, I., Munk, R., Gorospe, M., and Abdelmohsen, K. (2017). Emerging roles and context of circular RNAs. *Wiley Interdiscip. Rev. RNA* 8:e1386. doi: 10.1002/wrna.1386

Penev, P. I., Fakhretaha-Aval, S., Patel, V. J., Cannone, J. J., Gutell, R. R., Petrov, A. S., et al. (2020). Supersized ribosomal RNA expansion segments in Asgard archaea. *Genome Biol. Evol.* 12, 1694–1710. doi: 10.1093/gbe/evaa170

Petkovic, S., and Muller, S. (2015). RNA circularization strategies in vivo and in vitro. *Nucleic Acids Res.* 43, 2454–2465. doi: 10.1093/nar/gkv045

Plagens, A., Daume, M., Wiegel, J., and Randau, L. (2015). Circularization restores signal recognition particle RNA functionality in *Thermoproteus*. *elife* 4:e11623. doi: 10.7554/eLife.11623

Puttaraju, M., and Been, M. D. (1992). Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons. *Nucleic Acids Res.* 20, 5357–5364. doi: 10.1093/nar/20.20.5357

Qi, L., Li, J., Jia, J., Yue, L., and Dong, X. (2020). Comprehensive analysis of the pre-ribosomal RNA maturation pathway in a methanoarchaeon exposes the conserved circularization and linearization mode in archaea. *RNA Biol.* 17, 1427–1441. doi: 10.1080/15476286.2020.1771946

Randau, L., and Soll, D. (2008). Transfer RNA genes in pieces. *EMBO Rep.* 9, 623–628. doi: 10.1038/embor.2008.101

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

Schnare, M. N., and Gray, M. W. (2011). Complete modification maps for the cytosolic small and large subunit rRNAs of *Euglena gracilis*: functional and evolutionary implications of contrasting patterns between the two rRNA components. *J. Mol. Biol.* 413, 66–83. doi: 10.1016/j.jmb.2011.08.037

Shalev-Benami, M., Zhang, Y., Matzov, D., Halfon, Y., Zackay, A., Rozenberg, H., et al. (2016). 2.8-Å Cryo-EM structure of the large ribosomal subunit from the eukaryotic parasite leishmania. *Cell Rep.* 16, 288–294. doi: 10.1016/j.celrep.2016.06.014

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J., and Bohnsack, M. T. (2016). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 5, 1138–1152. doi: 10.1080/15476286.2016.1259781

Soma, A. (2014). Circularly permuted tRNA genes: their expression and implications for their physiological relevance and development. *Front. Genet.* 5:63. doi: 10.3389/fgene.2014.00063

Starostina, N. G., Marshburn, S., Johnson, L. S., Eddy, S. R., Terns, R. M., and Terns, M. P. (2004). Circular box C/D RNAs in *Pyrococcus furiosus*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14097–14101. doi: 10.1073/pnas.0403520101

Tang, T. H., Rozhdestvensky, T. S., d´Orval, B. C., Bortolin, M. L., Huber, H., Charpentier, B., et al. (2002). RNomics in *Archaea* reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res.* 30, 921–930. doi: 10.1093/nar/30.4.921

Taverniti, V., Forti, F., Ghisotti, D., and Putzer, H. (2011). *Mycobacterium smegmatis* RNase J is a 5´-3´ exo−/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol. Microbiol.* 82, 1260–1276. doi: 10.1111/j.1365-2958.2011.07888.x

Vercruysse, M., Kohrer, C., Shen, Y., Proulx, S., Ghosal, A., Davies, B. W., et al. (2016). Identification of YbeY-protein interactions involved in 16S rRNA maturation and stress regulation in *Escherichia coli*. *mBio* 7, e01785–e01716. doi: 10.1128/mBio.01785-16

Whitfeld, P. R., Leaver, C. J., Bottomley, W., and Atchison, B. (1978). Low-molecular-weight (4.5S) ribonucleic acid in higher-plant chloroplast ribosomes. *Biochem. J.* 175, 1103–1112. doi: 10.1042/bj1751103

Zhang, X., Lai, M., Chang, W., Yu, I., Ding, K., Mrazek, J., et al. (2016). Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. *Nat. Commun.* 7:13223. doi: 10.1038/ncomms13223

Zhou, W., Lu, Q., Li, Q., Wang, L., Ding, S., Zhang, A., et al. (2017). PPR-SMR protein SOT1 has RNA endonuclease activity. *Proc. Natl. Acad. Sci. U. S. A.* 114, E1554–E1563. doi: 10.1073/pnas.1612460114

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.

frontiers in Microbiology

REVIEW
published: 11 March 2021
doi: 10.3389/fmicb.2021.654370

H/ACA Small Ribonucleoproteins: Structural and Functional Comparison Between Archaea and Eukaryotes

*Dominic P. Czekay and Ute Kothe**

Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute, University of Lethbridge, Lethbridge, AB, Canada

During ribosome synthesis, ribosomal RNA is modified through the formation of many pseudouridines and methylations which contribute to ribosome function across all domains of life. In archaea and eukaryotes, pseudouridylation of rRNA is catalyzed by H/ACA small ribonucleoproteins (sRNPs) utilizing different H/ACA guide RNAs to identify target uridines for modification. H/ACA sRNPs are conserved in archaea and eukaryotes, as they share a common general architecture and function, but there are also several notable differences between archaeal and eukaryotic H/ACA sRNPs. Due to the higher protein stability in archaea, we have more information on the structure of archaeal H/ACA sRNPs compared to eukaryotic counterparts. However, based on the long history of yeast genetic and other cellular studies, the biological role of H/ACA sRNPs during ribosome biogenesis is better understood in eukaryotes than archaea. Therefore, this review provides an overview of the current knowledge on H/ACA sRNPs from archaea, in particular their structure and function, and relates it to our understanding of the roles of eukaryotic H/ACA sRNP during eukaryotic ribosome synthesis and beyond. Based on this comparison of our current insights into archaeal and eukaryotic H/ACA sRNPs, we discuss what role archaeal H/ACA sRNPs may play in the formation of ribosomes.

Keywords: H/ACA RNA, pseudouridine, RNA modification, ribosome biogenesis, pre-rRNA processing, telomerase, Dyskeratosis congenita, dyskerin

OPEN ACCESS

Edited by:
Anna La Teana,
Polytechnic University of Marche, Italy

Reviewed by:
Yi-Tao Yu,
University of Rochester, United States
Sunny Sharma,
Rutgers, The State University of New Jersey, United States

***Correspondence:**
Ute Kothe
ute.kothe@uleth.ca

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

Received: *16 January 2021*
Accepted: *18 February 2021*
Published: *11 March 2021*

Citation:
Czekay DP and Kothe U (2021) H/ACA Small Ribonucleoproteins: Structural and Functional Comparison Between Archaea and Eukaryotes. Front. Microbiol. 12:654370. doi: 10.3389/fmicb.2021.654370

INTRODUCTION

Ribosomes are macromolecular components present in all living cells responsible for protein biosynthesis, one of the energetically most expensive processes in cells. Ribosome biogenesis begins with the transcription of ribosomal RNA (rRNA), which in both archaea and eukaryotes is mostly transcribed as a long precursor containing individual segments of rRNA although some archaea also have separate rRNA genes (Yip et al., 2013). During the early stages of ribosome biogenesis, the nascent pre-rRNA is subject to many site-specific RNA modifications, the most abundant of which are 2′-O-methylations and pseudouridines (Maden, 1990; Kos and Tollervey, 2010; Yip et al., 2013).

Pseudouridine is a structural isomer of uridine initially discovered using two-dimensional paper chromatography of yeast RNA extracts (Davis and Allen, 1957). This RNA modification

is characterized by its unique C–C glycosidic bond (**Figure 1**). The isomerization of uridine to pseudouridine results in an additional imino group acting as a hydrogen bond donor on the Hoogsteen edge of the base. Pseudouridine has been demonstrated to be more thermodynamically favorable than uridine when present in short duplexes of RNA (Davis, 1995; Kierzek et al., 2014). This can be partially explained by the fact that in crystal structures pseudouridine is observed to coordinate a water molecule between its nucleobase and nearby sugar-phosphate backbone, providing a rigidifying effect to the local RNA fold and increasing base stacking interactions (Arnez and Steitz, 1994).

The formation of pseudouridine is catalyzed by a conserved class of enzymes known as pseudouridine synthases. In bacteria, these enzymes exist solely as stand-alone proteins, which both recognize the modification site in rRNA or tRNA and catalyze their modification (reviewed in Hamma and Ferre-D'Amare, 2006). While archaea and eukaryotes also contain stand-alone enzymes for this purpose (reviewed in Rintala-Dempsey and Kothe, 2017), a more sophisticated system employing H/ACA sRNPs is responsible for nearly all rRNA modifications (Ganot et al., 1997a; Ni et al., 1997; Yu and Meier, 2014). H/ACA sRNPs are named after the H/ACA guide RNA component that determines their sequence specificity. In 1997, two groups discovered the hitherto unknown function of H/ACA RNAs and their associated proteins in directing the site-specific pseudouridylation in rRNA triggering a plethora of studies in yeast and other eukaryotes that provides the basis for our current understanding of H/ACA sRNP function (Ganot et al., 1997a; Ni et al., 1997). Eukaryotic H/ACA sRNPs are further distinguished as H/ACA small nucleolar RNPs (snoRNPs) or H/ACA small Cajal-body-specific RNPs (scaRNPs), which localize and function in nucleoli and Cajal bodies, respectively (Darzacq et al., 2002). Shortly after the description of H/ACA sRNAs in eukaryotes, their presence was also verified in archaea (Watanabe and Gray, 2000; Tang et al., 2002). Considering the lack of subnuclear compartments in archaea, the archaeal counterparts are simply designated as H/ACA snoRNP-like, or more commonly as H/ACA small ribonucleoproteins (sRNPs) (Omer et al., 2003). In this review, we will explore and compare the structures of archaeal and eukaryotic H/ACA sRNPs, the variety of functions of H/ACA sRNPs, and discuss what is known about their assembly and implications on ribosome biogenesis and beyond.

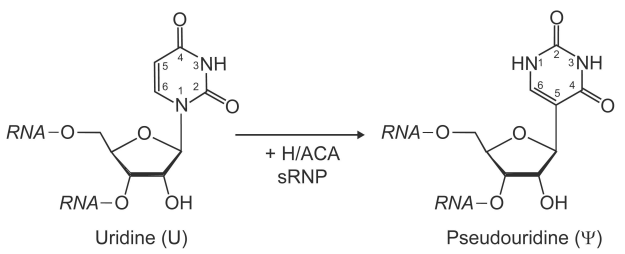


FIGURE 1 | Schematic representation of the isomerization of a uridine to pseudouridine by H/ACA sRNPs. Pseudouridine is characterized by a unique C–C glycosidic bond linking C1′ of the ribose with C5 of the base as well as an extra imino group (N1) with hydrogen bonding potential within the base. The Watson-Crick face in pseudouridine is unchanged and allows base-pairing with adenine.

H/ACA sRNPs SHARE A COMMON STRUCTURAL CORE ORGANIZATION

A mature H/ACA sRNP particle is composed of four different core proteins that assemble onto a H/ACA guide RNA scaffold (**Figure 2**). The archaeal proteins and their eukaryotic homologs that constitute H/ACA sRNPs are: L7ae (Nhp2 in eukaryotes), Nop10, Gar1, and the catalytic component, Cbf5 (dyskerin in humans) (Watanabe and Gray, 2000; Rozhdestvensky et al., 2003). Li and Ye (2006) reported the first structure revealing the organization of an archaeal H/ACA sRNP which was followed by a number of further structures of archaeal H/ACA sRNPs including structures showing the recognition of target RNA (Liang et al., 2007a; Duan et al., 2009; Liang B. et al., 2009; Zhou et al., 2010). The overall structural similarity of archaea and eukaryotic H/ACA sRNPs as well as some critical differences have subsequently been revealed by a structure of the *S. cerevisiae* Cbf5-Nop10-Gar1 complex, and more recently, by a cryo-electron microscopy structure of human telomerase, containing a H/ACA sRNP assembled on the 3′ end of human telomerase RNA (Li et al., 2011b; Nguyen et al., 2018). In the following, we will introduce the structural features of the H/ACA sRNP components and discuss their conservation and differences between archaea and eukaryotes.

H/ACA guide RNAs in archaea and eukaryotes have a few notable differences. In most studied eukaryotes (albeit with few exceptions like trypanosomes; Liang et al., 2004), all H/ACA snoRNAs conform to a hairpin-hinge-hairpin secondary structure where each hairpin is followed by one of two conserved consensus sequences, the H box (consensus ANANNA) and the ACA box, respectively. Within the ACA box, the adenines are most conserved, and alternative sequences (AUA, AAA, or AGA) can be found (Ganot et al., 1997b). In all cases, the ACA box is located strictly three nucleotides upstream of the 3′ end of the RNA (Balakin et al., 1996; Ganot et al., 1997b). However, some archaea display also highly atypical H/ACA RNA features (Bernick et al., 2012). Instead of the two-hairpin structure observed in almost all eukaryotes like yeast and humans, the vast majority of archaeal H/ACA sRNAs contain just one hairpin followed by an ACA box, but in rare instances archaeal H/ACA sRNAs have two or three hairpins (Rozhdestvensky et al., 2003). In both eukaryotes and archaea, the H and ACA box elements are necessary for H/ACA sRNP accumulation, localization, and pseudouridylation activities *in vivo* (Balakin et al., 1996; Ganot et al., 1997b; Bortolin et al., 1999; Narayanan et al., 1999; Caton et al., 2018). As obvious from the crystal structure, the single-hairpin H/ACA sRNA in archaea is bound by one set of core proteins (Li and Ye, 2006), and in analogy, it is assumed that each hairpin of H/ACA sRNAs in eukaryotes also binds a complete set of H/ACA proteins which is further supported by the set of two H/ACA proteins observed bound to human telomerase RNA (**Figure 2**; Nguyen et al., 2018).

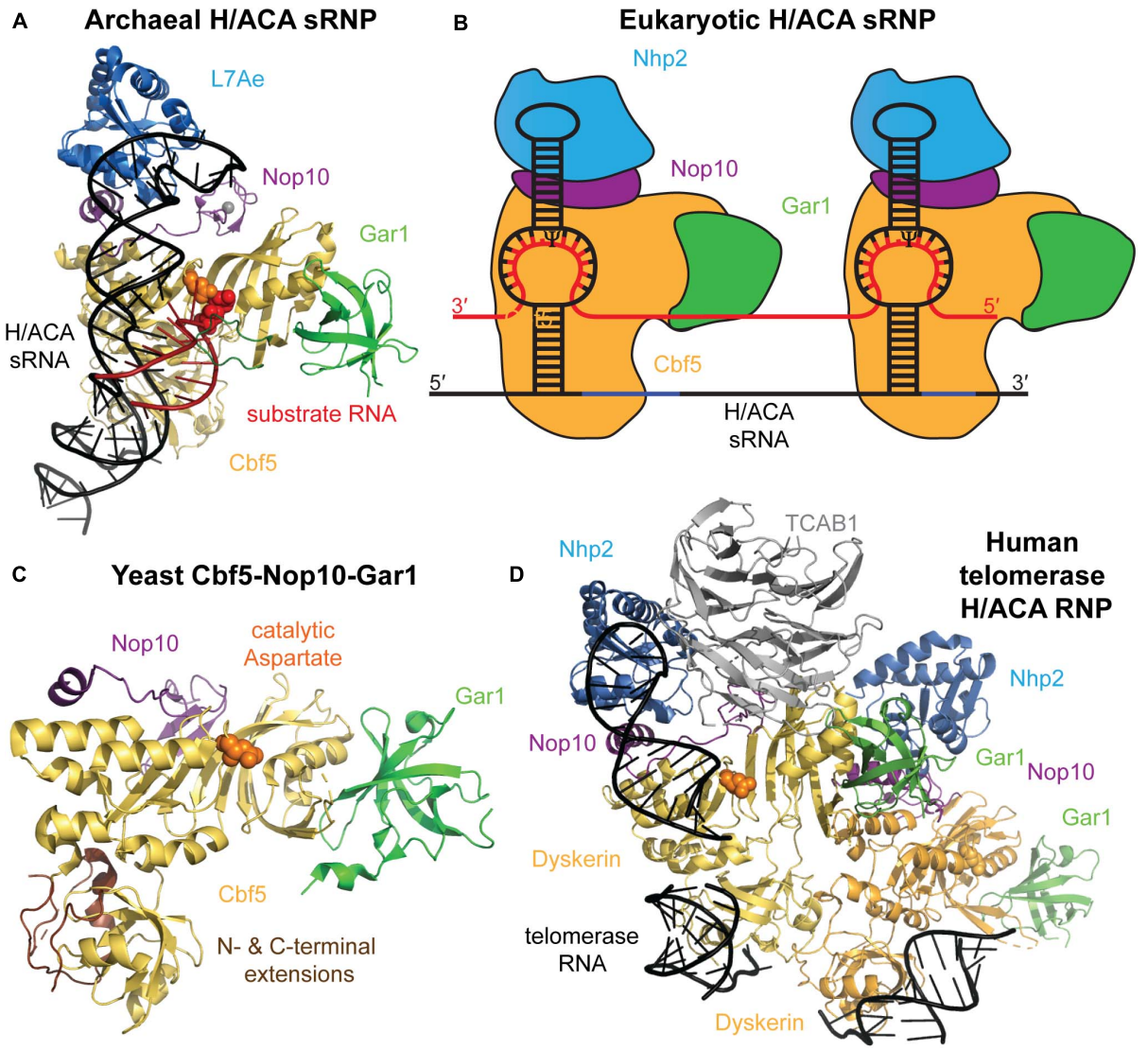


FIGURE 2 | Archaeal and eukaryotic H/ACA sRNP structure. **(A)** Crystal structure of an H/ACA sRNP bound to substrate RNA (red) from *P. furiosus* (PDB ID: 3HAY) (Duan et al., 2009). The single-hairpin H/ACA sRNA (black) binds to the four H/ACA proteins: Cbf5 (orange), Nop10 (purple), L7Ae (light blue), and Gar1 (green). In the active site of Cbf5, the catalytic aspartate residue is depicted in orange sphere adjacent to the target uridine (red sticks). **(B)** Schematic representation of the typical two-hairpin structure of a eukaryotic H/ACA sRNP bound to a target RNA (red). Each hairpin of the H/ACA snoRNA (black) is assumed to bind a complete set of H/ACA proteins. The proteins and RNAs are colored as in **(A)**. Note that Nhp2 is the eukaryotic homolog of archaeal L7Ae. **(C)** Crystal structure of the yeast Cbf5-Nop10-Gar1 complex (PDB ID: 3U28) shown in a similar conformation as the archaeal H/ACA sRNP complex in **(A)** (Li et al., 2011b). In eukaryotes, the PUA of Cbf5 (bottom) is larger than in archaea due to N- and C-terminal extensions shown in brown. **(D)** Structural model of the H/ACA sRNP complex assembled on the 3′ end of human telomerase RNA based on a cryo-electron microscopy structure of human telomerase (Nguyen et al., 2018). Two sets of all H/ACA proteins (including the human homolog of Cbf5 called dyskerin) are observed as well as the Telomerase Cajal body protein 1 (TCBA1, gray). All structure representations were prepared using Pymol software.

Within each hairpin, H/ACA RNAs contain single-stranded pockets, generally known as pseudouridylation pockets. The unpaired nucleotides in the pocket provide pseudouridylation specificity by selecting a uridine in the target RNA whose flanking nucleotides complement those in the H/ACA sRNA (Ganot et al., 1997a). Target RNA binding forms a three-way junction at which the targeted uridine as well as a nucleotide 3′ of the target uridine remain unpaired in the center of the guide RNA pocket, and the target uridine is inserted into the active site of Cbf5 for modification (Liang et al., 2007a; Wu and Feigon, 2007; Liang B. et al., 2009). In eukaryotes, at least 8 base-pairs are required between the H/ACA sRNA and its target RNA with no less than three base-pairs on either side of the pseudouridylation pocket in order to allow for efficient pseudouridine formation (De Zoysa et al., 2018; Kelly et al., 2019). Across all functional H/ACA sRNAs, there is a defined distance between the site of pseudouridylation and the nearest downstream consensus sequence element (H box or ACA box),

but this distance varies slightly between eukaryotes and archaea (Tang et al., 2002; Toffano-Nioche et al., 2015). In archaea this distance is typically 14–16 nt, whereas in eukaryotes it is generally 15–16 nt. Functionally, this distance acts as a molecular ruler that ensures proper positioning of the guide RNA relative to the active site in Cbf5 such that a bound substrate RNA target uridine can be positioned for catalysis (Caton et al., 2018).

A notable difference between eukaryotic and archaeal H/ACA sRNAs is the presence of a conserved K-turn or K-loop motif in the upper portion of the hairpin above the pseudouridylation pocket in archaeal H/ACA sRNA which is absent in eukaryotic H/ACA sRNAs (Rozhdestvensky et al., 2003). The K-turn or kink-turn motif is a common RNA motif that results in a characteristic kink in an RNA helix as first observed in ribosomal RNA (Klein et al., 2001). The kink is caused by a three-nucleotide internal bulge that is closed on one site by two canonical G-C base pairs and that is flanked on its other site by two sheared G-A base pairs. Whereas some archaeal H/ACA sRNAs have a longer upper hairpin stem harboring a K-turn motif, other archaeal H/ACA sRNAs with a shorter upper stem contain a variation of this motif called K-loop. Here, the G-A base-pairs are present, but instead of a 3-nucleotide bulge a 7-nucleotide loop is found. Notably, both the K-turn and the K-loop motif are always located 5–6 nucleotides above the pseudouridylation pocket (Rozhdestvensky et al., 2003).

The catalytic core component of H/ACA sRNPs is the protein Cbf5, a pseudouridine synthase of the TruB family. This family is defined by the essential PseudoUridine synthase and Archaeosine transglycosylase (PUA) domain, a common RNA binding domain that contributes to H/ACA sRNA binding in Cbf5 by interacting with the lower stem and the H or ACA box (Hamma et al., 2005; Hamma and Ferre-D'Amare, 2006). One notable difference between eukaryotic and archaeal Cbf5 is the presence of N- and C-terminal extensions in the eukaryotic protein that both contribute to a larger PUA domain but may also be partially unstructured based on the presence of many charged residues (**Figure 2**). The catalytic domain of Cbf5 harbors the core fold and conserved active site cleft residues that are shared across all pseudouridine synthase families (Hamma et al., 2005; Hamma and Ferre-D'Amare, 2006). The active site is characterized by the presence of a strictly conserved catalytic aspartate residue that is required for nucleophilic attack during isomerization (**Figure 2**; Huang et al., 1998; Veerareddygari et al., 2016). Additional active site residues include a conserved basic residue and a tyrosine residue that stacks with the target uracil base (Ferre-D'Amare, 2003). In TruB, the bacterial homolog of Cbf5, the conserved basic residues are shown to participate in an electrostatic network important for modification; meanwhile, the conserved tyrosine is suggested to maintain active site structure and may act as a general base during catalysis (Phannachet et al., 2005; Friedt et al., 2014). Interestingly, Cbf5 is an essential gene in eukaryotes, but it can be deleted in *Haloferax volcanii* suggesting a differential importance of H/ACA sRNPs in eukaryotes and archaea (Jiang et al., 1993; Blaby et al., 2011).

The pseudouridine synthase Cbf5 tightly interacts along its catalytic domain with the protein Nop10, a small (<10 kDa) protein that binds Cbf5 independent of other proteins or RNA (Hamma et al., 2005). Nop10 is organized into two domains that are separated by a linker. Although the linker and C-terminal domain of Nop10 are mostly unstructured in solution, Nop10 adopts structure upon binding to Cbf5 (Hamma et al., 2005; Khanna et al., 2006; Reichow and Varani, 2008). When bound, the central region of Nop10 supports the boundaries of the Cbf5 active site, and is speculated to potentially influence active site dynamics (Hamma et al., 2005). Unique to archaeal Nop10 is the presence of a highly stable N-terminal zinc-binding ribbon that is replaced by a smaller, only partially stable, β-hairpin in eukaryotic counterparts (Khanna et al., 2006). When in complex with Cbf5, a pair of solvent-exposed Nop10 aromatic residues moderately contribute to binding of the H/ACA RNA (Hamma et al., 2005). Moreover, Nop10 seems to stabilize the active site of Cbf5 thereby enhancing its catalytic activity (Kamalampeta and Kothe, 2012).

The third H/ACA sRNP protein is Gar1, an essential protein containing one large central domain flanked by two Glycine-Arginine Rich (GAR) regions, which are common amongst other yeast nucleolar proteins (Girard et al., 1992; Bagni and Lapeyre, 1998). Archaeal homologs of Gar1 are substantially smaller in size, as they lack both GAR regions found in their eukaryotic counterparts (Bridger et al., 2012). Consequently, only the central portion of the eukaryotic protein is conserved in archaea. Strikingly, a Gar1 central-domain only variant in yeast was demonstrated to be sufficient in performing all essential functions of full-length Gar1 *in vivo*, rescuing growth and pre-rRNA processing defects observed in Gar1-deficient strains (Girard et al., 1992, 1994). The central domain of Gar1 interacts with the catalytic domain of Cbf5, but is not in direct contact with the H/ACA sRNA (**Figure 2**; Li and Ye, 2006). Instead, Gar1 also enhances Cbf5's catalytic activity similar to Nop10 (Kamalampeta and Kothe, 2012), and it is critical for product release (Duan et al., 2009). The later function is achieved through an interaction of Gar1 with the so-called thumb loop of Cbf5: in the substrate-free, open state, Gar1 binds the thumb loop allowing Cbf5 to recruit substrate RNA. Subsequently, Cbf5's thumb loop is released from Gar1 and binds over the substrate RNA thereby stabilizing it in the active site of Cbf5. In order to allow for product release after pseudouridine formation, Gar1 has to once again bind the thumb loop of Cbf5 to allow for target RNA dissociation (Duan et al., 2009). Interestingly, yeast Gar1 has been reported to directly bind the essential H/ACA snoRNAs snR30 and snR10 (Bagni and Lapeyre, 1998). While this interaction is not observed in the H/ACA sRNP structures reported to-date, it could be that eukaryotic Gar1 fulfills additional functions by directly interacting with RNA.

The upper stem of the H/ACA sRNA is bound by the archaeal protein L7Ae or its respective eukaryotic homolog Nhp2 (Rozhdestvensky et al., 2003). L7Ae is a member a large family of RNA-binding proteins that specifically recognize K-turn and K-loop motifs (Rozhdestvensky et al., 2003; Hamma and Ferre-D'Amare, 2004; Gagnon et al., 2010). Notably, L7Ae is also a core component of archaeal C/D sRNPs where L7Ae also recognizes a K-loop motif. While Nhp2 is the eukaryotic homolog of L7Ae, it has lost the ability to specifically bind K turns or K loops in agreement with the absence of these motifs in eukaryotic H/ACA sRNPs (Henras et al., 2001). Moreover, Nhp2 is restricted

to H/ACA sRNPs only whereas eukaryotic C/D sRNPs contain the homolog Snu13p/14k which continues to recognize K turns and loops. Nevertheless, Nhp2 has retained the general ability to bind RNA (Henras et al., 2001). Unlike L7Ae, which shows little affinity for other H/ACA sRNP proteins, Nhp2 tightly binds to Nop10 in eukaryotic RNPs (Hamma et al., 2005). As a result, the recruitment of L7Ae and Nhp2 to the H/ACA sRNP differs: whereas Nhp2 is anchored to the H/ACA RNP through a protein-protein interaction with Nop10, L7Ae relies on binding to the K-turn of the H/ACA guide RNA and only forms a weak binding interface with Nop10 (Wang and Meier, 2004). Presumably, the conserved distance of 5–6 nucleotides between the pseudouridylation pocket and the K-turn or K-loop motif in archaeal H/ACA sRNA is required to allow for these week L7Ae-Nop10 interactions (Rozhdestvensky et al., 2003). Notably, both Nhp2 and L7Ae play an important role in anchoring the top of an H/ACA guide RNA hairpin and to position the pseudouridylation pocket in close proximity of the active site of Cbf5 which is important for pseudouridylation activity (Liang et al., 2007a, 2008; Caton et al., 2018). Thus, Nhp2 and L7Ae differ in their molecular interactions, but seem to fulfill the same function.

FUNCTIONAL ROLES OF H/ACA sRNPs IN RIBOSOME FORMATION AND BEYOND

H/ACA sRNPs play roles in several cellular pathways including ribosome biogenesis, but also in many other RNA-related processes (**Figure 3**). The most well defined and characteristic role of H/ACA RNPs is the site-specific introduction of pseudouridines in rRNA during ribosome synthesis (Bousquet-Antonelli et al., 1997; Ganot et al., 1997a; Ni et al., 1997). While the specific role of individual pseudouridines in rRNA remains unclear, collectively pseudouridines are critical for ribosome function, and the removal of select pseudouridines via the deletion of the respective H/ACA guide RNAs causes changes in ribosome structure and function (Penzo and Montanaro, 2018). Importantly, pseudouridines occur with the greatest frequency in functionally important regions of the ribosome such as the peptidyl transferase center, the decoding center, and intersubunit bridges (Bakin et al., 1994; Liang et al., 2007c). In yeast, the removal of H/ACA guide RNAs introducing pseudouridines in these regions influences ribosome structure, translation rate, translational fidelity, and biogenesis (King et al., 2003; Baxter-Roshek et al., 2007; Decatur et al., 2007; Baudin-Baillieu et al., 2009; Liang X.H. et al., 2009; Polikanov et al., 2015; Sloan et al., 2017). However, since archaea are less amenable to genetic manipulation, our understanding of the exact roles of rRNA pseudouridylation for the archaeal ribosome is lagging. Interestingly, for the organisms studied so far, it seems that archaeal ribosomes contain a much lower number of pseudouridines (e.g., 5 in *Sulfolobus acidocaldarius*) compared to their eukaryotic counterparts and even compared to some bacteria like *Escherichia coli* with 11 rRNA pseudouridines (Massenet et al., 1999; Hamma and Ferre-D'Amare, 2006). However, the pseudouridines detected in archaeal rRNA also reside in critical regions, namely the peptidyltransferase center and helix 69 of the 23S rRNA, and similar positions are also modified in bacteria (Ofengand and Bakin, 1997; Massenet et al., 1999; Blaby et al., 2011). Based on the conservation of rRNA pseudouridylation in all kingdoms of life, it seems therefore reasonable to assume that rRNA modification by archaeal H/ACA sRNPs plays in general similar roles in ribosome synthesis and translation as in eukaryotes.

The ribosomal A-site acts as the binding site for incoming aminoacyl-tRNAs during protein synthesis and contains several pseudouridines in eukaryotes. Removal of pseudouridines within the yeast ribosomal A-site alters the structure of the A-site, changing the positioning of critical bases involved in tRNA accommodation (Baxter-Roshek et al., 2007). The yeast A-site finger contains four pseudouridines that cause slight increases in frameshifting when removed individually; however, the removal of all four pseudouridines causes elevated UGA stop codon readthrough with increased + 1 frameshifting (Baudin-Baillieu et al., 2009). Deleting pseudouridines together with 2′-O-methylations in the ribosomal A-site further affects translation fidelity (Baudin-Baillieu et al., 2009). Given the sparsity of pseudouridines in archaeal rRNA, it is currently not clear whether some pseudouridines are fulfilling similar roles in the archaeal A site. Helix 69 of 25S rRNA is an important region of the ribosome and is part of the intersubunit bridge connecting the small and large ribosomal subunit. Depletion of rRNA modifications in this intersubunit bridge (Helix 69), which includes four pseudouridine residues, results in decreased growth rate, increased antibiotic sensitivity, and increased frameshifting during translation in yeast (Liang et al., 2007c). Most likely, the conserved pseudouridines in archaeal helix 69 have a similar function.

Aside from ribosomal RNA, pseudouridines have also been discovered in tRNA, small nuclear RNA (snRNA), long non-coding RNA (lncRNA) and mRNA in eukaryotes (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014; Li et al., 2015). Many (but not all) pseudouridines in snRNA are introduced by H/ACA RNAs called small Cajal body RNAs (scaRNAs), a subset of H/ACA sRNAs that do not exist in archaea. In addition to the H and ACA boxes common to all H/ACA RNAs, box H/ACA scaRNAs contain one additional sequence element named the CAB box (consensus UGAG) that is located at the terminal loops of each hairpin (Richard et al., 2003). Moreover, several pseudouridines in yeast and human mRNAs are dependent on Cbf5/dyskerin and are therefore most likely introduced by H/ACA sRNPs (Carlile et al., 2014; Schwartz et al., 2014; Li et al., 2015). In archaea, it is currently unknown whether H/ACA sRNPs can also target RNAs other than rRNA, but at least computational predictions suggest that this possibility should not be ruled out (Toffano-Nioche et al., 2015).

Transfer RNA (tRNA) is one of the most highly modified RNAs within all cells. Notably, the pseudouridylation of position 55 at the TΨC arm, is universally conserved across all domains of life in all elongator tRNAs. In eukaryotes, this pseudouridine is introduced by the standalone pseudouridine synthase Pus4, but interestingly in archaea, Ψ55 can be introduced by both the standalone enzyme Pus10 (which is not related to Pus4) as well as

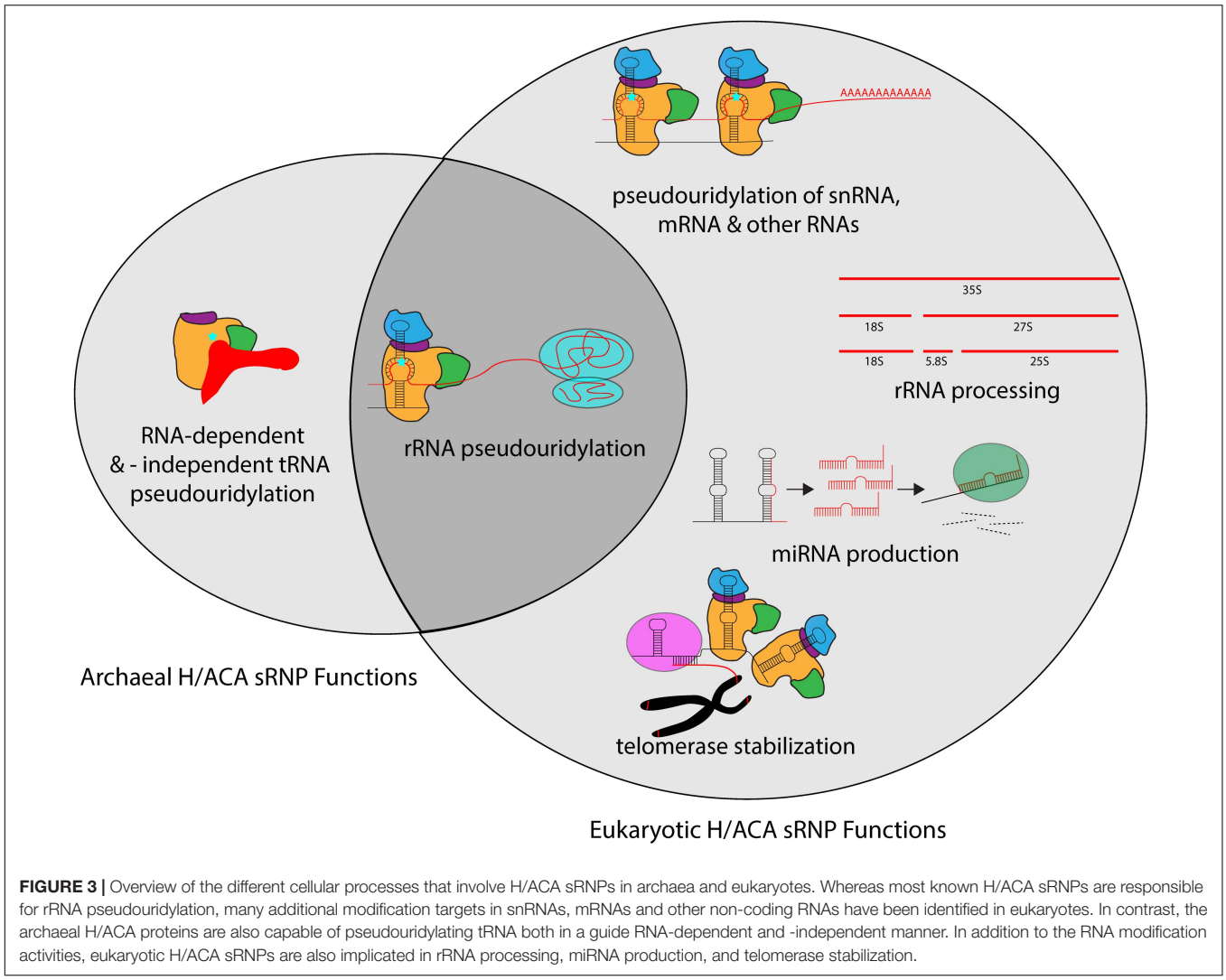


FIGURE 3 | Overview of the different cellular processes that involve H/ACA sRNPs in archaea and eukaryotes. Whereas most known H/ACA sRNPs are responsible for rRNA pseudouridylation, many additional modification targets in snRNAs, mRNAs and other non-coding RNAs have been identified in eukaryotes. In contrast, the archaeal H/ACA proteins are also capable of pseudouridylating tRNA both in a guide RNA-dependent and -independent manner. In addition to the RNA modification activities, eukaryotic H/ACA sRNPs are also implicated in rRNA processing, miRNA production, and telomerase stabilization.

by Cbf5 (Roovers et al., 2006; Gurha and Gupta, 2008). Strikingly, in the latter scenario, Cbf5 is capable of introducing the pseudouridine at position 55 in an RNA-independent manner, i.e., without H/ACA sRNA, and this activity is enhanced by the presence of the Nop10 and Gar1 proteins (Roovers et al., 2006; Kamalampeta and Kothe, 2012; Fujikane et al., 2018). To bind the tRNA in the absence of H/ACA sRNA, the archaeal Cbf5 PUA domain binds the CCA 3′ end of the tRNA tightly highlighting the versatility of the PUA domain in either binding the ACA motif of H/ACA sRNAs or the CCA motif of tRNAs (Roovers et al., 2006). However, Cbf5 is non-essential in *H. volcanii* in contrast to Pus10 indicating that *in vivo* Pus10 is the predominant tRNA Ψ55 modification enzyme (Blaby et al., 2011). In addition to this RNA-independent modification of tRNAs by archaeal Cbf5, it has also been reported that at least in some archaeal species such as *Sulfolobus solfataricus* the pseudouridine in position 35 of pre-tRNA^{Tyr} can be generated in an RNA-dependent mechanism by a complete H/ACA sRNP (Muller et al., 2009).

Whereas pseudouridylation by H/ACA sRNPs is their most studied activity, it is presumably not their most important cellular function. Notably, as mentioned, the catalytic protein of H/ACA sRNP, Cbf5, is not essential in archaea suggesting that ribosome biogenesis can occur in the absence of pseudouridylation in archaea (Blaby et al., 2011). Interestingly, the same is true in yeast. Whereas Cbf5 is essential (Jiang et al., 1993), yeast strains expressing only catalytically inactive Cbf5 show a significant growth defect, but are viable (Zebarjadian et al., 1999). These observations raise the intriguing question regarding additional functions of H/ACA sRNPs beyond RNA modification which have been identified in eukaryotes, but not (yet) in archaea (Mitchell et al., 1999a; Vos and Kothe, 2020).

Interestingly, modification H/ACA sRNAs are usually non-essential in eukaryotes, but this is not true for all H/ACA sRNAs providing insight into the most critical cellular function of H/ACA sRNPs. The one essential eukaryotic H/ACA sRNA is *S. cerevisiae* snR30/human U17 (Bally et al., 1988). Notably, there is no identified homologs of the snR30 RNA in archaea. Unlike typical H/ACA sRNAs, snR30/U17 has no known sites of pseudouridylation but instead has a crucial role for the processing of 35S pre-rRNA to generate 18S rRNA (Vos and

Kothe, 2020). Together with the core H/ACA proteins and several other ribosome biogenesis factors, snR30 facilitates the early endo-nucleolytic 35S pre-rRNA cleavage events (Zebarjadian et al., 1999; Atzorn et al., 2004). Therein, snR30 also base-pairs with rRNA in an unpaired pocket of its 3′ hairpin; however this interaction resides at the bottom rather than the top of the pocket and thus differs significantly from the rRNA interactions of modification H/ACA sRNPs (Fayet-Lebaron et al., 2009). The detailed molecular mechanism and the architecture of the snR30 H/ACA sRNP remain unknown, but evidently this complex is responsible for the most important function of eukaryotic H/ACA sRNPs during ribosome assembly. Although not essential, the yeast snR10 H/ACA sRNP is similarly implicated in 35S pre-rRNA processing, and consequently its deletion also increases cell doubling time, results in accumulation of 35S pre-rRNA, and causes a cold-sensitive phenotype (Tollervey, 1987; King et al., 2003). Given that processing of archaeal rRNA occurs entirely differently using an archaeal-specific splicing mechanism (summarized in Yip et al., 2013; Clouet-d'Orval et al., 2018), it seems unlikely that an archaeal H/ACA sRNP fulfills a similar role during rRNA processing as the eukaryotic snR30/U17 H/ACA sRNP, but it cannot be excluded that H/ACA sRNPs are differently involved in archaeal ribosome formation.

One interesting function of H/ACA sRNAs observed exclusively in vertebrates is the stabilization of telomerase RNA. The 3′ end of vertebrate telomerase RNA folds into a secondary structure that strongly resembles an H/ACA sRNA, and accordingly the 3′ end of telomerase RNA assembles with two complete sets of box H/ACA core proteins (**Figure 2D**; Mitchell et al., 1999a; Chen et al., 2000; Dragon et al., 2000). Similar to H/ACA sRNAs that direct pseudouridylation, telomerase RNA contains consensus H and ACA sequences that are also essential for its accumulation, 3′ end processing, and telomerase activity (Mitchell et al., 1999a). This function of vertebrate H/ACA sRNPs has been strongly implicated with a human premature aging syndrome called Dyskeratosis congenita characterized by leukoplakia, nail dystrophy, bone marrow failure, and increased susceptibility to some forms of cancer (Dokal, 2000). The disease has three forms: autosomal dominant, autosomal recessive, and X-linked (X-DC), which is the most severe of all forms. Many X-DC patients have mutations in dyskerin, the human homolog of *Cbf5* (Heiss et al., 1998) which cluster in the PUA domain as well as N- and C-terminal extensions of dyskerin which envelop the PUA domain (Hamma et al., 2005; Rashid et al., 2006). Notably, the mutated residues are generally conserved in eukaryotic Cbf5/dyskerin, but not in its archaeal homolog. In accordance with the role of the PUA domain for the binding to the ACA box in H/ACA sRNAs, many X-DC dyskerin variants do not bind telomerase RNA leading to its destabilization (Ashbridge et al., 2009). As a consequence, one key symptom of X-DC is the shortening of telomeres in cells derived from X-DC patients as well as reduced telomerase activity in primary cells (Mitchell et al., 1999b; Vulliamy T.J. et al., 2001). In addition, it was shown for certain X-DC mutations that they also impair rRNA pseudouridylation and reduce rRNA processing (Mochizuki et al., 2004), and in a mouse model with reduced dyskerin expression, which recapitulates Dyskeratosis congenita features, ribosomal defects appear before telomere shortening (Ruggero et al., 2003). The less severe autosomal dominant form of Dyskeratosis congenita is characterized by mutations that remove a portion of the H/ACA RNA-like structure of telomerase RNA (Vulliamy T. et al., 2001). Unlike dyskerin mutations in X-DC, autosomal dominant mutations do not reduce binding of dyskerin to telomerase RNA (Ashbridge et al., 2009). Another autosomal recessive form of Dyskeratosis congenita is also linked to H/ACA RNPs and is caused by mutations in the *nop10* or *nhp2* genes (Walne et al., 2007; Vulliamy et al., 2008). In summary, the importance of human H/ACA sRNPs for telomere maintenance and ribosome biogenesis is underlined by the molecular defects observed in the different forms of Dyskeratosis congenita.

Lastly, the functions of eukaryotic H/ACA sRNPs extend even further beyond RNA modifications, telomerase stabilization, and rRNA processing (McMahon et al., 2015). In at least one instance, a human H/ACA RNA has been shown to function as a micro RNA (miRNA) after processing by the Dicer enzyme *in vivo* (Ender et al., 2008). Many small RNAs (20–26 nt in length) created from ACA45, normally responsible for directing pseudouridylation of U37 in U2 spliceosomal RNA (snRNA), can stably associate with Argonaute (Ago) proteins and direct the degradation of transcriptional regulator CDC2L6 mRNA (Ender et al., 2008). Notably, other human miRNAs might also be derived from H/ACA sRNA-like precursors (Scott et al., 2009). Furthermore, some H/ACA sRNAs are associated with chromatin and may thus contribute to the regulation of transcription (Schubert et al., 2012). Lastly, H/ACA-like RNAs are critical for trans-splicing in trypanosomes through mediating pseudouridylation of the spliced leader RNA, the substrate for trans-splicing (Barth et al., 2005). Thus, H/ACA sRNAs and their complexes with proteins may have more functions than currently anticipated, and this may also hold true for archaeal H/ACA sRNAs.

THE ASSEMBLY PATHWAY OF H/ACA sRNPs

In eukaryotes, the formation of a functional H/ACA sRNP is a complex process that involves several factors working together to assemble and transport the premature H/ACA sRNP particles throughout different compartments of the cell and ultimately to their final location, i.e., the nucleolus or Cajal body (Kiss et al., 2006). In contrast, our current information suggests that archaeal H/ACA sRNPs can self-assemble as none of the additional assembly factors is conserved in archaea. Self-assembly of archaeal H/ACA sRNPs has been successful *in vitro* laying the ground for several biochemical and structural studies (Baker et al., 2005; Charpentier et al., 2005). In contrast, it was much more difficult to reconstitute a yeast H/ACA sRNP in the absence of assembly factors *in vitro* due to the instability of the isolated proteins (Li et al., 2011b; Caton et al., 2018). In the following sections, we will describe the process of H/ACA sRNP biogenesis beginning with the production of nascent Cbf5/dyskerin in the cytoplasm (**Figure 4**).

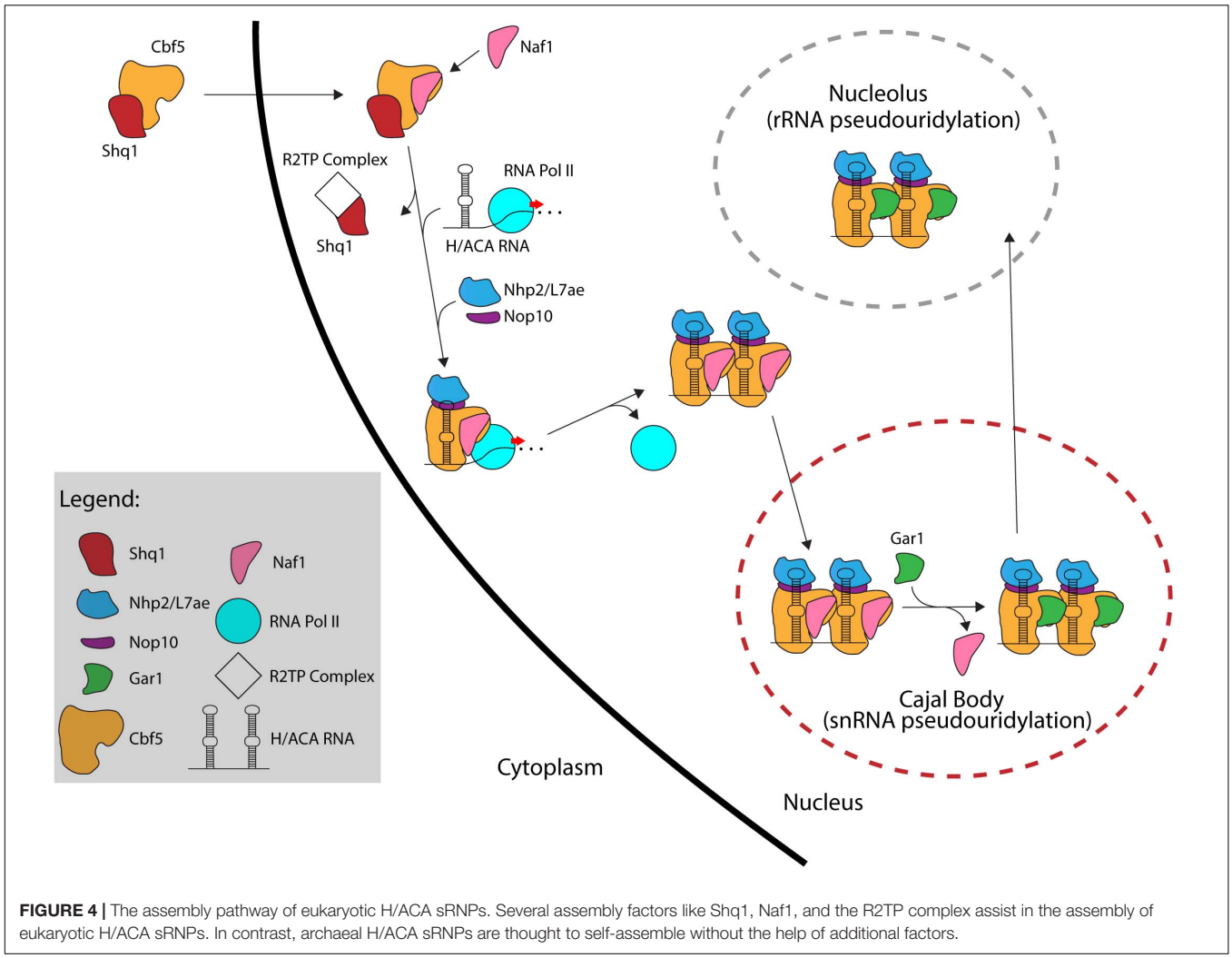


FIGURE 4 | The assembly pathway of eukaryotic H/ACA sRNPs. Several assembly factors like Shq1, Naf1, and the R2TP complex assist in the assembly of eukaryotic H/ACA sRNPs. In contrast, archaeal H/ACA sRNPs are thought to self-assemble without the help of additional factors.

Following its translation, Cbf5/dyskerin is quickly bound by Shq1, an essential protein related to Hsp90 cochaperones, that plays a crucial role in early H/ACA sRNP biogenesis by tightly binding the H/ACA sRNA binding interface of Cbf5 through RNA mimicry (Yang et al., 2002; Godin et al., 2009; Walbott et al., 2011). Shq1 binding to Cbf5 ensures that the RNA binding surface of Cbf5 is occupied thereby preventing non-specific RNA binding and aggregation prior to assembly on an H/ACA RNA (Grozdanov et al., 2009; Li et al., 2011a; Caton et al., 2018). Interestingly, mutations in Shq1 can also cause Dyskeratosis congenita (Bizarro and Meier, 2017).

The Cbf5/dyskerin complex with Shq1 is then imported into the nucleus to join the nascent H/ACA sRNA. In *S. cerevisiae*, H/ACA RNAs are typically encoded as single genes (Schattner et al., 2004), whereas H/ACA RNA genes are found within introns of protein coding genes in mammals (Schattner et al., 2006). Through computational and experimental screens, H/ACA snoRNA genes have been identified in several organisms showing a variety of different gene structures such as independent genes, intron-encoded genes, and polycistronic gene clusters (Liang et al., 2004, 2007b; Chen et al., 2008; Wang and Ruvinsky, 2010; Patra Bhattacharya et al., 2016). H/ACA sRNAs are transcribed by RNA Polymerase II followed by processing involving several factors. In yeast, processing of polycistronic H/ACA sRNA is mediated by the endonuclease RNase III (Rnt1) (Chanfreau et al., 1998). Intron-encoded H/ACA sRNAs are typically liberated through splicing and debranching. To mediate further processing in yeast, H/ACA sRNAs are polyadenylated by the poly(A) polymerase Pap1 or the alternative Tfr4, bound by polyA binding protein (Pab2 in fission yeast) and subsequently processed by the nuclear exosome (van Hoof et al., 2000; Grzechnik and Kufel, 2008; Lemay et al., 2010; Berndt et al., 2012). As there are only few pseudouridines in archaeal rRNA and thus only few H/ACA RNAs, the transcription and maturation of archaeal H/ACA RNAs has not been studied in detail, but many archaeal H/ACA RNAs have been identified (Rozhdestvensky et al., 2003; Muller et al., 2007, 2008; Randau, 2015; Toffano-Nioche et al., 2015; Clouet-d'Orval et al., 2018).

After the Cbf5-Shq1 complex enters the nucleus, Cbf5 is bound by the protein Naf1 which contains a Gar1 domain mediating its interaction with Cbf5 (Hoareau-Aveilla et al., 2006; Leulliot et al., 2007). Subsequently, Cbf5 is recruited to the site

of H/ACA RNA transcription. In eukaryotes, the recruitment of Cbf5 to the nascent H/ACA RNA is achieved through Naf1-mediated interactions with the C-terminal domain (CTD) of RNA polymerase II (Fatica et al., 2002; Richard et al., 2006). For snoRNAs transcribed from their own promoter in humans, an additional mode of recruitment is suggested that is mediated through TSG1 which is responsible for 5′ hypermethylation of snoRNAs and also interacts with dyskerin (Mouaikel et al., 2002). To enable Cbf5 binding to nascent H/ACA RNA, Shq1 is removed from Cbf5 by the R2TP complex, a multiprotein complex composed of two AAA + ATPases (Rvb1 and Rvb2 in yeast) and two Hsp90 interactors (Pih1 and Tah1 in yeast) that is involved in multiple cellular processes (King et al., 2001; Kakihara and Houry, 2012; Machado-Pinilla et al., 2012). The co-transcriptional assembly of Cbf5 on the H/ACA sRNA is likely protecting the nascent RNA from degradation by exonucleolytic proteins since Cbf5 is necessary for accumulation of all box H/ACA RNAs (Lafontaine et al., 1998; Berndt et al., 2012). Nop10 and Nhp2 are also recruited to the maturing H/ACA sRNP although the timing of their association is not entirely clear. However, the presence of Naf1 prevents Gar1 recruitment and renders the pre-sRNP complex inactive.

Currently, it is not entirely clear whether the Naf1-bound H/ACA pre-sRNPs localize to the Cajal bodies and are then shuttled to the nucleoli, or whether they migrate to the nucleoli directly. In any case, Naf1 is replaced by Gar1 forming the active RNP complexes. Although the process for exchanging these proteins is not fully known, the SMN complex, which like Gar1 is also highly concentrated in Cajal bodies, is implicated in this process supporting the hypothesis that H/ACA sRNPs migrate through the Cajal body (Pellizzoni et al., 2001; Whitehead et al., 2002). Finally, most H/ACA sRNPs are shuttled to the nucleolus to modify ribosomal RNAs while those required for snRNA modification (scaRNAs) remain in the Cajal bodies (Kiss, 2006).

DISCUSSION

H/ACA sRNPs are versatile ribonucleoprotein machines conserved across both archaea and eukaryotes that play critical roles during ribosome biogenesis through the site-directed formation of pseudouridine modifications in rRNA. In agreement with their conservation, the core structure and functionality of H/ACA sRNPs is the same in archaea and eukaryotes, but multiple adaptations have arisen to further expand the scope of cellular roles of these RNPs such as tRNA modification in archaea as well as modification of several RNAs, rRNA processing, telomerase stabilization, microRNA biogenesis and chromatin regulation in eukaryotes. Notably, some of these additional functions have only emerged recently, and we are still lacking a full understanding of the molecular mechanisms of H/ACA sRNPs in ribosome assembly and beyond. Moreover, H/ACA sRNPs can be utilized as bioengineering devices to site-specifically introduce novel pseudouridines, for example to enable stop codon read-through in yeast (Karijolich and Yu, 2011). As pseudouridines prevent the recognition of mRNA by the immune system and novel mRNA vaccines contain pseudouridines (Kariko et al., 2008; Pardi and Weissman, 2017), the engineering capability of H/ACA sRNPs holds future promising applications beyond the role of H/ACA sRNAs in ribosome formation. Given the current progress in understanding ribosome formation and H/ACA sRNP function, a number of interesting hypotheses are emerging regarding further roles of these ribonucleoproteins. These may hold true in archaea and/or eukaryotes and will likely shape the direction of future research.

Besides stabilizing rRNA through the introduction of pseudouridines, it has been a long-standing speculation that H/ACA sRNPs may also act as rRNA chaperones in both archaea and eukaryotes (Watkins and Bohnsack, 2012; Yip et al., 2013). By base-pairing with rRNA, H/ACA sRNPs may keep certain regions of the rRNA unfolded during the early stages of ribosome assembly thereby preventing premature folding or they may even be able to unfold wrong rRNA folding intermediates. As rRNA folding is a complex and poorly understood process due to the immense size of rRNA, this is an intriguing proposition that will require a concerted approach to be experimentally addressed. In this context, it is interesting to note that eukaryotic H/ACA sRNPs likely rely at least in part on RNA helicases such as Has1 and Rok1 to be removed from rRNA which may contribute to regulating the timing of rRNA folding (Liang and Fournier, 2006; Bohnsack et al., 2008). In contrast, we have no indication to date that helicases fulfill a similar role for H/ACA sRNPs during archaeal ribosome assembly. In addition to rRNA modification and possibly folding, it is noteworthy that one of the most critical functions of a eukaryotic H/ACA sRNA, namely snR30/U17, is to facilitate the processing of pre-rRNA which may also constitute a significant difference between the eukaryotic and archaeal kingdom of life. In all organisms, it will be interesting to understand the coordinated action of H/ACA sRNPs and the other ribosome assembly factors who will interact simultaneously with rRNA early during ribosome formation when the rRNA is still accessible and not yet folded into a compact form. Clearly, many molecular mechanisms and interactions remain to be unraveled regarding ribosome biogenesis in both archaea and eukaryotes.

AUTHOR CONTRIBUTIONS

DC and UK devised the conceptual structure of the manuscript. DC wrote the first draft including figures whereas UK refined the text and figures. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant RGPIN-2014-05954 and 2020-04965), Alberta Innovates (Strategic Research Chair 2015) which both provided salary support for DC and Canadian Institutes of Health Research (Project Grant 437623) which covers publication charges.

REFERENCES

Arnez, J. G., and Steitz, T. A. (1994). Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* 33, 7560–7567.

Ashbridge, B., Orte, A., Yeoman, J. A., Kirwan, M., Vulliamy, T., Dokal, I., et al. (2009). Single-molecule analysis of the human telomerase RNA.dyskerin interaction and the effect of dyskeratosis congenita mutations. *Biochemistry* 48, 10858–10865. doi: 10.1021/bi901373e

Atzorn, V., Fragapane, P., and Kiss, T. (2004). U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. *Mol. Cell Biol.* 24, 1769–1778.

Bagni, C., and Lapeyre, B. (1998). Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. *J. Biol. Chem.* 273, 10868–10873. doi: 10.1074/jbc.273.18.10868

Baker, D. L., Youssef, O. A., Chastkofsky, M. I., Dy, D. A., Terns, R. M., and Terns, M. P. (2005). RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev.* 19, 1238–1248. doi: 10.1101/gad.1309605

Bakin, A., Lane, B. G., and Ofengand, J. (1994). Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry* 33, 13475–13483.

Balakin, A. G., Smith, L., and Fournier, M. J. (1996). The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* 86, 823–834.

Bally, M., Hughes, J., and Cesareni, G. (1988). SnR30: a new, essential small nuclear RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 16, 5291–5303. doi: 10.1093/nar/16.12.5291

Barth, S., Hury, A., Liang, X. H., and Michaeli, S. (2005). Elucidating the role of H/ACA-like RNAs in trans-splicing and rRNA processing via RNA interference silencing of the *Trypanosoma brucei* CBF5 pseudouridine synthase. *J. Biol. Chem.* 280, 34558–34568. doi: 10.1074/jbc.M503465200

Baudin-Baillieu, A., Fabret, C., Liang, X. H., Piekna-Przybylska, D., Fournier, M. J., and Rousset, J. P. (2009). Nucleotide modifications in three functionally important regions of the *Saccharomyces cerevisiae* ribosome affect translation accuracy. *Nucleic Acids Res.* 37, 7665–7677. doi: 10.1093/nar/gkp816

Baxter-Roshek, J. L., Petrov, A. N., and Dinman, J. D. (2007). Optimization of ribosome structure and function by rRNA base modification. *PLoS One* 2:e174. doi: 10.1371/journal.pone.0000174

Berndt, H., Harnisch, C., Rammelt, C., Stohr, N., Zirkel, A., Dohm, J. C., et al. (2012). Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* 18, 958–972. doi: 10.1261/rna.032292.112

Bernick, D. L., Dennis, P. P., Hochsmann, M., and Lowe, T. M. (2012). Discovery of Pyrobaculum small RNA families with atypical pseudouridine guide RNA features. *RNA* 18, 402–411. doi: 10.1261/rna.031385.111

Bizarro, J., and Meier, U. T. (2017). Inherited SHQ1 mutations impair interaction with NAP57/dyskerin, a major target in dyskeratosis congenita. *Mol. Genet. Genomic Med.* 5, 805–808. doi: 10.1002/mgg3.314

Blaby, I. K., Majumder, M., Chatterjee, K., Jana, S., Grosjean, H., de Crecy-Lagard, V., et al. (2011). Pseudouridine formation in archaeal RNAs: the case of *Haloferax volcanii*. *RNA* 17, 1367–1380. doi: 10.1261/rna.2712811

Bohnsack, M. T., Kos, M., and Tollervey, D. (2008). Quantitative analysis of snoRNA association with pre-ribosomes and release of snR30 by Rok1 helicase. *EMBO Rep.* 9, 1230–1236. doi: 10.1038/embor.2008.184

Bortolin, M. L., Ganot, P., and Kiss, T. (1999). Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *EMBO J.* 18, 457–469. doi: 10.1093/emboj/18.2.457

Bousquet-Antonelli, C., Henry, Y., G'Elugne, J. P., Caizergues-Ferrer, M., and Kiss, T. (1997). A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. *Embo J.* 16, 4770–4776. doi: 10.1093/emboj/16.15.4770

Bridger, S. L., Lancaster, W. A., Poole, F. L. II, Schut, G. J., and Adams, M. W. (2012). Genome sequencing of a genetically tractable *Pyrococcus furiosus* strain reveals a highly dynamic genome. *J. Bacteriol.* 194, 4097–4106. doi: 10.1128/jb.00439-12

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., and Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143–146. doi: 10.1038/nature13802

Caton, E. A., Kelly, E. K., Kamalampeta, R., and Kothe, U. (2018). Efficient RNA pseudouridylation by eukaryotic H/ACA ribonucleoproteins requires high affinity binding and correct positioning of guide RNA. *Nucleic Acids Res.* 46, 905–916. doi: 10.1093/nar/gkx1167

Chanfreau, G., Legrain, P., and Jacquier, A. (1998). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J. Mol. Biol.* 284, 975–988. doi: 10.1006/jmbi.1998.2237

Charpentier, B., Muller, S., and Branlant, C. (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.* 33, 3133–3144. doi: 10.1093/nar/gki630

Chen, C. L., Chen, C. J., Vallon, O., Huang, Z. P., Zhou, H., and Qu, L. H. (2008). Genomewide analysis of box C/D and box H/ACA snoRNAs in *Chlamydomonas reinhardtii* reveals an extensive organization into intronic gene clusters. *Genetics* 179, 21–30. doi: 10.1534/genetics.107.086025

Chen, J. L., Blasco, M. A., and Greider, C. W. (2000). Secondary structure of vertebrate telomerase RNA. *Cell* 100, 503–514. doi: 10.1016/s0092-8674(00)80687-x

Clouet-d'Orval, B., Batista, M., Bouvier, M., Quentin, Y., Fichant, G., Marchfelder, A., et al. (2018). Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiol. Rev.* 42, 579–613. doi: 10.1093/femsre/fuy016

Darzacq, X., Jady, B. E., Verheggen, C., Kiss, A. M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J.* 21, 2746–2756. doi: 10.1093/emboj/21.11.2746

Davis, D. R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* 23, 5020–5026.

Davis, F. F., and Allen, F. W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* 227, 907–915.

De Zoysa, M. D., Wu, G., Katz, R., and Yu, Y. T. (2018). Guide-substrate base-pairing requirement for box H/ACA RNA-guided RNA pseudouridylation. *RNA* 24, 1106–1117. doi: 10.1261/rna.066837.118

Decatur, W. A., Liang, X. H., Piekna-Przybylska, D., and Fournier, M. J. (2007). Identifying effects of snoRNA-guided modifications on the synthesis and function of the yeast ribosome. *Methods Enzymol.* 425, 283–316. doi: 10.1016/s0076-6879(07)25013-x

Dokal, I. (2000). Dyskeratosis congenita in all its forms. *Br. J. Haematol.* 110, 768–779. doi: 10.1046/j.1365-2141.2000.02109.x

Dragon, F., Pogacic, V., and Filipowicz, W. (2000). In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. *Mol. Cell Biol.* 20, 3037–3048. doi: 10.1128/mcb.20.9.3037-3048.2000

Duan, J., Li, L., Lu, J., Wang, W., and Ye, K. (2009). Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol. Cell* 34, 427–439. doi: 10.1016/j.molcel.2009.05.005

Ender, C., Krek, A., Friedlander, M. R., Beitzinger, M., Weinmann, L., Chen, W., et al. (2008). A human snoRNA with microRNA-like functions. *Mol. Cell* 32, 519–528. doi: 10.1016/j.molcel.2008.10.017

Fatica, A., Dlakic, M., and Tollervey, D. (2002). Naf1 p is a box H/ACA snoRNP assembly factor. *RNA* 8, 1502–1514.

Fayet-Lebaron, E., Atzorn, V., Henry, Y., and Kiss, T. (2009). 18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences. *EMBO J.* 28, 1260–1270. doi: 10.1038/emboj.2009.79

Ferre-D'Amare, A. R. (2003). RNA-modifying enzymes. *Curr. Opin. Struct. Biol.* 13, 49–55. doi: 10.1016/s0959-440x(02)00002-7

Friedt, J., Leavens, F. M., Mercier, E., Wieden, H. J., and Kothe, U. (2014). An arginine-aspartate network in the active site of bacterial TruB is critical for catalyzing pseudouridine formation. *Nucleic Acids Res.* 42, 3857–3870. doi: 10.1093/nar/gkt1331

Fujikane, R., Behm-Ansmant, I., Tillault, A. S., Loegler, C., Igel-Bourguignon, V., Marguet, E., et al. (2018). Contribution of protein Gar1 to the RNA-guided and RNA-independent rRNA:Psi-synthase activities of the archaeal Cbf5 protein. *Sci. Rep.* 8:13815. doi: 10.1038/s41598-018-32164-0

Gagnon, K. T., Zhang, X., Qu, G., Biswas, S., Suryadi, J., Brown, B. A. II, et al. (2010). Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA* 16, 79–90. doi: 10.1261/rna.1692310

Ganot, P., Bortolin, M. L., and Kiss, T. (1997a). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* 89, 799–809.

Ganot, P., Caizergues-Ferrer, M., and Kiss, T. (1997b). The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.* 11, 941–956.

Girard, J. P., Bagni, C., Caizergues-Ferrer, M., Amalric, F., and Lapeyre, B. (1994). Identification of a segment of the small nucleolar ribonucleoprotein-associated protein GAR1 that is sufficient for nucleolar accumulation. *J. Biol. Chem.* 269, 18499–18506.

Girard, J. P., Lehtonen, H., Caizergues-Ferrer, M., Amalric, F., Tollervey, D., and Lapeyre, B. (1992). GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *EMBO J.* 11, 673–682.

Godin, K. S., Walbott, H., Leulliot, N., van Tilbeurgh, H., and Varani, G. (2009). The box H/ACA snoRNP assembly factor Shq1p is a chaperone protein homologous to Hsp90 cochaperones that binds to the Cbf5p enzyme. *J. Mol. Biol.* 390, 231–244. doi: 10.1016/j.jmb.2009.04.076

Grozdanov, P. N., Roy, S., Kittur, N., and Meier, U. T. (2009). SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs. *RNA* 15, 1188–1197. doi: 10.1261/rna.1532109

Grzechnik, P., and Kufel, J. (2008). Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol. Cell* 32, 247–258. doi: 10.1016/j.molcel.2008.10.003

Gurha, P., and Gupta, R. (2008). Archaeal Pus10 proteins can produce both pseudouridine 54 and 55 in tRNA. *RNA* 14, 2521–2527. doi: 10.1261/rna.1276508

Hamma, T., and Ferre-D'Amare, A. R. (2004). Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution. *Structure* 12, 893–903. doi: 10.1016/j.str.2004.03.015

Hamma, T., and Ferre-D'Amare, A. R. (2006). Pseudouridine synthases. *Chem. Biol.* 13, 1125–1135. doi: 10.1016/j.chembiol.2006.09.009

Hamma, T., Reichow, S. L., Varani, G., and Ferre-D'Amare, A. R. (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat. Struct. Mol. Biol.* 12, 1101–1107. doi: 10.1038/nsmb1036

Heiss, N. S., Knight, S. W., Vulliamy, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., et al. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat. Genet.* 19, 32–38. doi: 10.1038/ng0598-32

Henras, A., Dez, C., Noaillac-Depeyre, J., Henry, Y., and Caizergues-Ferrer, M. (2001). Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic Acids Res.* 29, 2733–2746. doi: 10.1093/nar/29.13.2733

Hoareau-Aveilla, C., Bonoli, M., Caizergues-Ferrer, M., and Henry, Y. (2006). hNaf1 is required for accumulation of human box H/ACA snoRNPs, scaRNPs, and telomerase. *RNA* 12, 832–840. doi: 10.1261/rna.2344106

Huang, L., Pookanjanatavip, M., Gu, X., and Santi, D. V. (1998). A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry* 37, 344–351. doi: 10.1021/bi971874+

Jiang, W., Middleton, K., Yoon, H. J., Fouquet, C., and Carbon, J. (1993). An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. *Mol. Cell Biol.* 13, 4884–4893. doi: 10.1128/mcb.13.8.4884

Kakihara, Y., and Houry, W. A. (2012). The R2TP complex: discovery and functions. *Biochim. Biophys. Acta* 1823, 101–107. doi: 10.1016/j.bbamcr.2011.08.016

Kamalampeta, R., and Kothe, U. (2012). Archaeal proteins Nop10 and Gar1 increase the catalytic activity of Cbf5 in pseudouridylating tRNA. *Sci. Rep.* 2:663. doi: 10.1038/srep00663

Karijolich, J., and Yu, Y. T. (2011). Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* 474, 395–398. doi: 10.1038/nature10165

Kariko, K., Muramatsu, H., Welsh, F. A., Ludwig, J., Kato, H., Akira, S., et al. (2008). Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol. Ther.* 16, 1833–1840. doi: 10.1038/mt.2008.200

Kelly, E. K., Czekay, D. P., and Kothe, U. (2019). Base-pairing interactions between substrate RNA and H/ACA guide RNA modulate the kinetics of pseudouridylation, but not the affinity of substrate binding by H/ACA small nucleolar ribonucleoproteins. *RNA* 25, 1393–1404. doi: 10.1261/rna.071043.119

Khanna, M., Wu, H., Johansson, C., Caizergues-Ferrer, M., and Feigon, J. (2006). Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA. *RNA* 12, 40–52. doi: 10.1261/rna.2221606

Kierzek, E., Malgowska, M., Lisowiec, J., Turner, D. H., Gdaniec, Z., and Kierzek, R. (2014). The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* 42, 3492–3501. doi: 10.1093/nar/gkt1330

King, T. H., Decatur, W. A., Bertrand, E., Maxwell, E. S., and Fournier, M. J. (2001). A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol. Cell Biol.* 21, 7731–7746. doi: 10.1128/mcb.21.22.7731-7746.2001

King, T. H., Liu, B., McCully, R. R., and Fournier, M. J. (2003). Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell* 11, 425–435.

Kiss, T. (2006). SnoRNP biogenesis meets Pre-mRNA splicing. *Mol. Cell* 23, 775–776. doi: 10.1016/j.molcel.2006.08.023

Kiss, T., Fayet, E., Jady, B. E., Richard, P., and Weber, M. (2006). Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb. Symp. Quant. Biol.* 71, 407–417. doi: 10.1101/sqb.2006.71.025

Klein, D. J., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. *EMBO J.* 20, 4214–4221. doi: 10.1093/emboj/20.15.4214

Kos, M., and Tollervey, D. (2010). Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol. Cell* 37, 809–820. doi: 10.1016/j.molcel.2010.02.024

Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* 12, 527–537.

Lemay, J. F., D'Amours, A., Lemieux, C., Lackner, D. H., St-Sauveur, V. G., Bahler, J., et al. (2010). The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. *Mol. Cell* 37, 34–45. doi: 10.1016/j.molcel.2009.12.019

Leulliot, N., Godin, K. S., Hoareau-Aveilla, C., Quevillon-Cheruel, S., Varani, G., Henry, Y., et al. (2007). The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *J. Mol. Biol.* 371, 1338–1353. doi: 10.1016/j.jmb.2007.06.031

Li, L., and Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* 443, 302–307.

Li, S., Duan, J., Li, D., Ma, S., and Ye, K. (2011a). Structure of the Shq1-Cbf5-Nop10-Gar1 complex and implications for H/ACA RNP biogenesis and dyskeratosis congenita. *EMBO J.* 30, 5010–5020. doi: 10.1038/emboj.2011.427

Li, S., Duan, J., Li, D., Yang, B., Dong, M., and Ye, K. (2011b). Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev.* 25, 2409–2421. doi: 10.1101/gad.175299.111

Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., et al. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 11, 592–597. doi: 10.1038/nchembio.1836

Liang, B., Kahen, E. J., Calvin, K., Zhou, J., Blanco, M., and Li, H. (2008). Long-distance placement of substrate RNA by H/ACA proteins. *RNA* 14, 2086–2094. doi: 10.1261/rna.1109808

Liang, B., Xue, S., Terns, R. M., Terns, M. P., and Li, H. (2007a). Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nat. Struct. Mol. Biol.* 14, 1189–1195.

Liang, B., Zhou, J., Kahen, E., Terns, R. M., Terns, M. P., and Li, H. (2009). Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nat. Struct. Mol. Biol.* 16, 740–746. doi: 10.1038/nsmb.1624

Liang, X. H., and Fournier, M. J. (2006). The helicase Has1p is required for snoRNA release from pre-rRNA. *Mol. Cell Biol.* 26, 7437–7450. doi: 10.1128/MCB.00664-06

Liang, X. H., Hury, A., Hoze, E., Uliel, S., Myslyuk, I., Apatoff, A., et al. (2007b). Genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in Leishmania major indicates conservation among trypanosomatids in the repertoire and in their rRNA targets. *Eukaryot. Cell* 6, 361–377. doi: 10.1128/EC.00296-06

Liang, X. H., Liu, Q., and Fournier, M. J. (2007c). rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol. Cell* 28, 965–977. doi: 10.1016/j.molcel.2007.10.012

Liang, X. H., Liu, Q., and Fournier, M. J. (2009). Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* 15, 1716–1728. doi: 10.1261/rna.1724409

Liang, X. H., Ochaion, A., Xu, Y. X., Liu, Q., and Michaeli, S. (2004). Small nucleolar RNA clusters in trypanosomatid Leptomonas collosoma. genome organization, expression studies, and the potential role of sequences present upstream from the first repeated cluster. *J. Biol. Chem.* 279, 5100–5109. doi: 10.1074/jbc.M308264200

Lovejoy, A. F., Riordan, D. P., and Brown, P. O. (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One* 9:e110799. doi: 10.1371/journal.pone.0110799

Machado-Pinilla, R., Liger, D., Leulliot, N., and Meier, U. T. (2012). Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs. *RNA* 18, 1833–1845. doi: 10.1261/rna.034942.112

Maden, B. E. (1990). The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* 39, 241–303. doi: 10.1016/s0079-6603(08)60629-7

Massenet, S., Ansmant, I., Motorin, Y., and Branlant, C. (1999). The first determination of pseudouridine residues in 23S ribosomal RNA from hyperthermophilic Archaea *Sulfolobus acidocaldarius*. *FEBS Lett.* 462, 94–100. doi: 10.1016/s0014-5793(99)01524-0

McMahon, M., Contreras, A., and Ruggero, D. (2015). Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdiscip Rev. RNA* 6, 173–189. doi: 10.1002/wrna.1266

Mitchell, J. R., Cheng, J., and Collins, K. (1999a). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol. Cell Biol.* 19, 567–576.

Mitchell, J. R., Wood, E., and Collins, K. (1999b). A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402, 551–555. doi: 10.1038/990141

Mochizuki, Y., He, J., Kulkarni, S., Bessler, M., and Mason, P. J. (2004). Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *Proc. Natl. Acad. Sci. U.S.A.* 101, 10756–10761. doi: 10.1073/pnas.0402560101

Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., and Bordonne, R. (2002). Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol. Cell* 9, 891–901. doi: 10.1016/s1097-2765(02)00484-7

Muller, S., Charpentier, B., Branlant, C., and Leclerc, F. (2007). A dedicated computational approach for the identification of archaeal H/ACA sRNAs. *Methods Enzymol.* 425, 355–387. doi: 10.1016/S0076-6879(07)25015-3

Muller, S., Leclerc, F., Behm-Ansmant, I., Fourmann, J. B., Charpentier, B., and Branlant, C. (2008). Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs. *Nucleic Acids Res.* 36, 2459–2475. doi: 10.1093/nar/gkn077

Muller, S., Urban, A., Hecker, A., Leclerc, F., Branlant, C., and Motorin, Y. (2009). Deficiency of the tRNATyr:Psi 35-synthase aPus7 in Archaea of the Sulfolobales order might be rescued by the H/ACA sRNA-guided machinery. *Nucleic Acids Res.* 37, 1308–1322. doi: 10.1093/nar/gkn1037

Narayanan, A., Lukowiak, A., Jady, B. E., Dragon, F., Kiss, T., Terns, R. M., et al. (1999). Nucleolar localization signals of box H/ACA small nucleolar RNAs. *EMBO J.* 18, 5120–5130. doi: 10.1093/emboj/18.18.5120

Nguyen, T. H. D., Tam, J., Wu, R. A., Greber, B. J., Toso, D., Nogales, E., et al. (2018). Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* 557, 190–195. doi: 10.1038/s41586-018-0062-x

Ni, J., Tien, A. L., and Fournier, M. J. (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* 89, 565–573.

Ofengand, J., and Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *J. Mol. Biol.* 266, 246–268. doi: 10.1006/jmbi.1996.0737

Omer, A. D., Ziesche, S., Decatur, W. A., Fournier, M. J., and Dennis, P. P. (2003). RNA-modifying machines in archaea. *Mol. Microbiol.* 48, 617–629. doi: 10.1046/j.1365-2958.2003.03483.x

Pardi, N., and Weissman, D. (2017). Nucleoside modified mRNA vaccines for infectious diseases. *Methods Mol. Biol.* 1499, 109–121. doi: 10.1007/978-1-4939-6481-9_6

Patra Bhattacharya, D., Canzler, S., Kehr, S., Hertel, J., Grosse, I., and Stadler, P. F. (2016). Phylogenetic distribution of plant snoRNA families. *BMC Genomics* 17:969. doi: 10.1186/s12864-016-3301-2

Pellizzoni, L., Baccon, J., Charroux, B., and Dreyfuss, G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Curr. Biol.* 11, 1079–1088. doi: 10.1016/s0960-9822(01)00316-5

Penzo, M., and Montanaro, L. (2018). Turning uridines around: role of rRNA pseudouridylation in ribosome biogenesis and ribosomal function. *Biomolecules* 8:38. doi: 10.3390/biom8020038

Phannachet, K., Elias, Y., and Huang, R. H. (2005). Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry* 44, 15488–15494. doi: 10.1021/bi050961w

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

Randau, L. (2015). Evolution of small guide RNA genes in hyperthermophilic archaea. *Ann. N. Y. Acad. Sci.* 1341, 188–193. doi: 10.1111/nyas.12643

Rashid, R., Liang, B., Baker, D. L., Youssef, O. A., He, Y., Phipps, K., et al. (2006). Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell* 21, 249–260. doi: 10.1016/j.molcel.2005.11.017

Reichow, S. L., and Varani, G. (2008). Nop10 is a conserved H/ACA snoRNP molecular adaptor. *Biochemistry* 47, 6148–6156. doi: 10.1021/bi800418p

Richard, P., Darzacq, X., Bertrand, E., Jady, B. E., Verheggen, C., and Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J.* 22, 4283–4293. doi: 10.1093/emboj/cdg394

Richard, P., Kiss, A. M., Darzacq, X., and Kiss, T. (2006). Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Mol. Cell Biol.* 26, 2540–2549. doi: 10.1128/mcb.26.7.2540-2549.2006

Rintala-Dempsey, A. C., and Kothe, U. (2017). Eukaryotic stand-alone pseudouridine synthases – RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol.* 14, 1185–1196. doi: 10.1080/15476286.2016.1276150

Roovers, M., Hale, C., Tricot, C., Terns, M. P., Terns, R. M., Grosjean, H., et al. (2006). Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res.* 34, 4293–4301. doi: 10.1093/nar/gkl530

Rozhdestvensky, T. S., Tang, T. H., Tchirkova, I. V., Brosius, J., Bachellerie, J. P., and Huttenhofer, A. (2003). Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.* 31, 869–877.

Ruggero, D., Grisendi, S., Piazza, F., Rego, E., Mari, F., Rao, P. H., et al. (2003). Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science* 299, 259–262. doi: 10.1126/science.1079447

Schattner, P., Barberan-Soler, S., and Lowe, T. M. (2006). A computational screen for mammalian pseudouridylation guide H/ACA RNAs. *RNA* 12, 15–25. doi: 10.1261/rna.2210406

Schattner, P., Decatur, W. A., Davis, C. A., Ares, M. Jr., Fournier, M. J., and Lowe, T. M. (2004). Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res.* 32, 4281–4296. doi: 10.1093/nar/gkh768

Schubert, T., Pusch, M. C., Diermeier, S., Benes, V., Kremmer, E., Imhof, A., et al. (2012). Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. *Mol. Cell* 48, 434–444. doi: 10.1016/j.molcel.2012.08.021

Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., Leon-Ricardo, B. X., et al. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159, 148–162. doi: 10.1016/j.cell.2014.08.028

Scott, M. S., Avolio, F., Ono, M., Lamond, A. I., and Barton, G. J. (2009). Human miRNA precursors with box H/ACA snoRNA features. *PLoS Comput. Biol.* 5:e1000507. doi: 10.1371/journal.pcbi.1000507

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J., and Bohnsack, M. T. (2017). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781

Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., et al. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7536–7541. doi: 10.1073/pnas.112047299

Toffano-Nioche, C., Gautheret, D., and Leclerc, F. (2015). Revisiting the structure/function relationships of H/ACA(-like) RNAs: a unified model for Euryarchaea and Crenarchaea. *Nucleic Acids Res.* 43, 7744–7761. doi: 10.1093/nar/gkv756

Tollervey, D. (1987). A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J.* 6, 4169–4175.

van Hoof, A., Lennertz, P., and Parker, R. (2000). Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol. Cell Biol.* 20, 441–452. doi: 10.1128/mcb.20.2.441-452.2000

Veerareddygari, G. R., Singh, S. K., and Mueller, E. G. (2016). The Pseudouridine Synthases Proceed through a Glycal Intermediate. *J. Am. Chem. Soc.* 138, 7852–7855. doi: 10.1021/jacs.6b04491

Vos, T. J., and Kothe, U. (2020). snR30/U17 small nucleolar ribonucleoprotein: a critical player during ribosome biogenesis. *Cells* 9:2195. doi: 10.3390/cells9102195

Vulliamy, T., Beswick, R., Kirwan, M., Marrone, A., Digweed, M., Walne, A., et al. (2008). Mutations in the telomerase component NHP2 cause the premature ageing syndrome Dyskeratosis congenita. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8073–8078. doi: 10.1073/pnas.0800042105

Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P. J., et al. (2001). The RNA component of telomerase is mutated in autosomal dominant Dyskeratosis congenita. *Nature* 413, 432–435. doi: 10.1038/35096585

Vulliamy, T. J., Knight, S. W., Mason, P. J., and Dokal, I. (2001). Very short telomeres in the peripheral blood of patients with X-linked and autosomal Dyskeratosis congenita. *Blood Cells Mol. Dis.* 27, 353–357. doi: 10.1006/bcmd.2001.0389

Walbott, H., Machado-Pinilla, R., Liger, D., Blaud, M., Rety, S., Grozdanov, P. N., et al. (2011). The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic. *Genes Dev.* 25, 2398–2408. doi: 10.1101/gad.176834.111

Walne, A. J., Vulliamy, T., Marrone, A., Beswick, R., Kirwan, M., Masunari, Y., et al. (2007). Genetic heterogeneity in autosomal recessive Dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Hum. Mol. Genet.* 16, 1619–1629. doi: 10.1093/hmg/ddm111

Wang, C., and Meier, U. T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.* 23, 1857–1867. doi: 10.1038/sj.emboj.7600181

Wang, P. P., and Ruvinsky, I. (2010). Computational prediction of Caenorhabditis box H/ACA snoRNAs using genomic properties of their host genes. *RNA* 16, 290–298. doi: 10.1261/rna.1876210

Watanabe, Y., and Gray, M. W. (2000). Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res.* 28, 2342–2352. doi: 10.1093/nar/28.12.2342

Watkins, N. J., and Bohnsack, M. T. (2012). The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev. RNA* 3, 397–414. doi: 10.1002/wrna.117

Whitehead, S. E., Jones, K. W., Zhang, X., Cheng, X., Terns, R. M., and Terns, M. P. (2002). Determinants of the interaction of the spinal muscular atrophy disease protein SMN with the dimethylarginine-modified box H/ACA small nucleolar ribonucleoprotein GAR1. *J. Biol. Chem.* 277, 48087–48093. doi: 10.1074/jbc.M204551200

Wu, H., and Feigon, J. (2007). H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6655–6660. doi: 10.1073/pnas.0701534104

Yang, P. K., Rotondo, G., Porras, T., Legrain, P., and Chanfreau, G. (2002). The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. *J. Biol. Chem.* 277, 45235–45242. doi: 10.1074/jbc.M207669200

Yip, W. S., Vincent, N. G., and Baserga, S. J. (2013). Ribonucleoproteins in archaeal pre-rRNA processing and modification. *Archaea* 2013:614735. doi: 10.1155/2013/614735

Yu, Y. T., and Meier, U. T. (2014). RNA-guided isomerization of uridine to pseudouridine–pseudouridylation. *RNA Biol.* 11, 1483–1494. doi: 10.4161/15476286.2014.972855

Zebarjadian, Y., King, T., Fournier, M. J., Clarke, L., and Carbon, J. (1999). Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol. Cell Biol.* 19, 7461–7472. doi: 10.1128/mcb.19.11.7461

Zhou, J., Liang, B., and Li, H. (2010). Functional and structural impact of target uridine substitutions on the H/ACA ribonucleoprotein particle pseudouridine synthase. *Biochemistry* 49, 6276–6281. doi: 10.1021/bi1006699

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.

frontiers in Microbiology

MINI REVIEW
published: 12 March 2021
doi: 10.3389/fmicb.2021.654029

Conservation of Archaeal C/D Box sRNA-Guided RNA Modifications

*Ruth Breuer, Jose-Vicente Gomes-Filho and Lennart Randau**

Prokaryotic RNA Biology, Philipps-Universität Marburg, Marburg, Germany

Post-transcriptional modifications fulfill many important roles during ribosomal RNA maturation in all three domains of life. Ribose 2'-*O*-methylations constitute the most abundant chemical rRNA modification and are, for example, involved in RNA folding and stabilization. In archaea, these modification sites are determined by variable sets of C/D box sRNAs that guide the activity of the rRNA 2'-*O*-methyltransferase fibrillarin. Each C/D box sRNA contains two guide sequences that can act in coordination to bridge rRNA sequences. Here, we will review the landscape of archaeal C/D box sRNA genes and their target sites. One focus is placed on the apparent accelerated evolution of guide sequences and the varied pairing of the two individual guides, which results in different rRNA modification patterns and RNA chaperone activities.

Keywords: C/D box, methylation, RNA modification, RNA folding, RNA structure

OPEN ACCESS

Edited by:
Sébastien Ferreira-Cerca, University of Regensburg, Germany

Reviewed by:
Teresa Carlomagno, Leibniz University Hannover, Germany
Béatrice Clouet-D'orval, Centre National de la Recherche Scientifique (CNRS), France

****Correspondence:***
Lennart Randau
lennart.randau@staff.uni-marburg.de

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

Received: *15 January 2021*
Accepted: *19 February 2021*
Published: *12 March 2021*

Citation:
Breuer R, Gomes-Filho J-V and Randau L (2021) Conservation of Archaeal C/D Box sRNA-Guided RNA Modifications. Front. Microbiol. 12:654029. doi: 10.3389/fmicb.2021.654029

INTRODUCTION

The chemical modification of RNA has long been known to play a role in a wide variety of cellular processes in all three domains of life. The manifold modifications can be introduced co- or post-transcriptionally and concern all classes of RNA molecules. The most abundant RNA modification is the ribose-2'-*O*-methylation, which is commonly found on ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) and also present on small nuclear RNAs (snRNAs) in archaea and eukaryotes (Maden et al., 1995; Kiss-László et al., 1996; Tycowski et al., 1998; Omer et al., 2000; Vitali and Kiss, 2019). This modification fulfills many different functions: It can protect RNA from ribonucleolytic cleavage, stabilize single base pairs, exhibit a chaperone function and influence folding at high temperatures (Kawai et al., 1992; Herschlag et al., 1993; Williams et al., 2001; Helm, 2006). The latter function is especially important in thermophilic organisms, therefore, it is no surprise that thermophilic archaea exhibit a significantly larger number of 2'-*O*-methylations than mesophilic archaea (Noon et al., 1998; Omer et al., 2000; Su et al., 2013).

In bacteria, 2'-*O*-methylations are comparatively rare and introduced by site- or region-specific protein-only enzymes (Decatur and Fournier, 2002). In contrast, methylation of the ribose moiety is more commonly observed in archaea and eukaryotes, which both utilize an RNA-dependent mechanism involving so-called C/D box s(no)RNAs. Here, the methylation reaction is performed by a ribonucleoprotein (RNP) complex carrying a small nucleolar RNA (snoRNA), or its archaeal homolog, the sno-like RNA (sRNA; **Figure 1A**; Lischwe et al., 1985; Ochs et al., 1985; Filipowicz and Kiss, 1993; Maxwell and Fournier, 1995; Gaspin et al., 2000; Omer et al., 2000).

C/D box sRNAs were found to be approximately 50–70 nt long in archaea and between 50 and 300 nt long in eukaryotes (Lui and Lowe, 2013). These RNA molecules are named for four conserved sequence elements: the box C and box C' motifs with the consensus sequence RUGAUGA and the box D and box D' motifs with the consensus sequence CUGA (Maxwell and Fournier, 1995; Kiss-Laszlo et al., 1998). During C/D box sRNA folding, the motifs C and D base-pair and form a helix-loop-helix structure termed kink-turn (k-turn). The k-turn is a short stem structure comprising non-canonical base pairs and carrying two sheared base pairs (AG and GA) at its top (Watkins et al., 2000; Klein et al., 2001). Base pairing between the motifs C' and D' results in a similar structure called the k-loop, which consists of a single stem closed by a terminal loop (Nolivos et al., 2005). The sequences located between the motifs C and D', and C' and D, are complementary to the target RNA sequences and therefore serve as guides for the identification of methylation sites (**Figure 1A**). The length of the archaeal guide sequences ranges from 10 to 12 nt. Target methylation occurs at the nucleotide complementary to the fifth nucleotide upstream of the box D/D' motif (Kiss-László et al., 1996; Tran et al., 2005). A recent study in *Drosophila* identifies the minimal functional eukaryotic C/D box snoRNA as a single-domain molecule with (i) a terminal stem with a consensus k-turn domain, (ii) one box C and one box D separated by a 14 nt long antisense element and (iii) a one-nucleotide spacer between box C and the antisense element (Deryusheva and Gall, 2019).

Interestingly, archaeal organisms harbor not only linear, but also circular C/D box sRNAs, though their role remains to be determined (Starostina et al., 2004; Danan et al., 2012; Randau, 2012; Su et al., 2013). The analysis of permuted RNA-seq reads allowed for the detection of circularization junctions of RNA molecules and revealed that C/D box sRNA termini can be fused. Inspection of these fusion sites indicated that the termini are not clearly defined, but can vary by few nucleotides for individual C/D box sRNA species. In addition, linear C/D box sRNAs are usually observed in parallel to circular variants. Notably, in *Sulfolobus solfataricus*, C/D box sRNAs occur predominantly in the linear form, whereas in *Pyrococcus furiosus* almost all C/D box sRNAs exist in both linear and circular forms with similar abundance (Starostina et al., 2004; Danan et al., 2012). Furthermore, archaeal circular RNA molecules exist among tRNA introns and rRNA processing intermediates (Danan et al., 2012; Jüttner et al., 2020). *Thermoproteus* species were found to require circularization of signal recognition particle (SRP) RNAs to yield functional molecules (Plagens et al., 2015). In these cases, the 5' and 3' ends of the RNA molecule fold into close contact and form a bulge-helix-bulge motif which is recognized and cleaved by the tRNA splicing endonuclease and subsequently ligated by the tRNA ligase RtcB (Trotta et al., 1997; Englert et al., 2011; Popow et al., 2011). However, C/D box sRNA termini usually do not form canonical BHB motifs and the exact method of circularization remains unclear (Starostina et al., 2004). Since circular sRNA molecules have been nearly exclusively found in thermophiles thus far, it is suggested that the circularization provides stability at elevated growth temperatures (Starostina et al., 2004; Danan et al., 2012). Here, it is plausible that the close proximity of C/D box sRNA termini upon protein binding facilitates RNA ligation, representing a statistic event that is positively selected for due to the increased stability of the circularized products.

The C/D box sRNA is part of the C/D box RNP complex which contains three highly conserved proteins in archaea and four proteins in eukaryotes (**Figure 1A**). Upon adopting its secondary structure, the k-turn and k-loop of the C/D box sRNA are bound and stabilized by the RNA-binding protein L7Ae (Snu13/15.5 K in yeast/human; Kuhn et al., 2002; Omer et al., 2002; Gagnon et al., 2010). Binding of the C/D box sRNA by L7Ae depends on three essential features: (i) the terminal stem at the 5' and 3' ends of the C/D box sRNA,

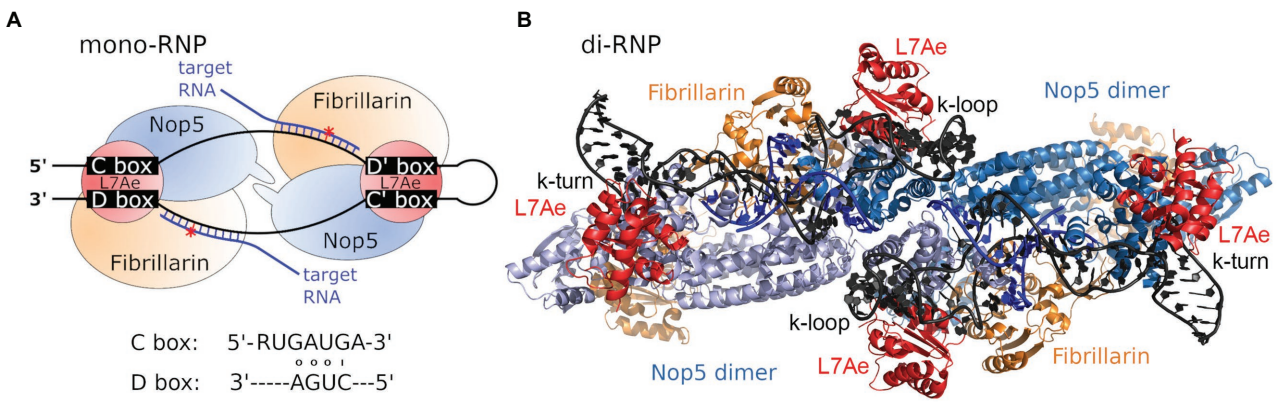


FIGURE 1 | C/D box sRNP architecture. **(A)** Schematic view of the archaeal C/D box mono-sRNP with bound target RNA (blue), consisting of the C/D box sRNA (black), L7Ae (red), Nop5 (blue), and fibrillarin (orange). The red asterisk denotes the position of methylation. Consensus C and D box sequences are indicated. **(B)** The C/D box sRNP complex has been observed to exist as a dimeric variant consisting of two C/D box sRNAs and four copies of each protein (pdb-id:4BY9; Lapinaite et al., 2013).

which juxtaposes the boxes C and D motifs, (ii) two sheared GA base pairs formed by pairing of the box C and box D motifs, and (iii) the box C uridine which is part of the k-turn's internal loop (Kuhn et al., 2002). After binding of the C/D box sRNA by L7Ae, the assembly of the RNP is completed by binding of the proteins Nop5 (Nop56/Nop58 heterodimer in yeast and humans) and fibrillarin (Nop1/fibrillarin in yeast/human; Omer et al., 2002; Bortolin et al., 2003). The N-terminal and C-terminal domains of Nop5 interact with fibrillarin and the C/D box sRNA, respectively. Furthermore, the coiled-coil domain of Nop5 mediates Nop5-dimerization for optimal interaction with the C/D box sRNA (Aittaleb et al., 2003).

Fibrillarin exhibits a conserved *S*-adenosyl-methionine (SAM) binding motif and possesses methyl transfer activity. It was found that this activity is dependent on C/D box sRNP formation and could not be observed independent of the complex (Wang et al., 2001; Omer et al., 2002). For the 2'-*O*-methylation reaction, fibrillarin uses *S*-adenosyl-L-methionine as a methyl group donor and after depositing the methyl group at the 2'-OH moiety, the ribose preferably adopts an endo-conformation, thereby blocking sugar-edge interactions (Kawai et al., 1992; Auffinger and Westhof, 1997; Hansen et al., 2002; Motorin and Helm, 2010). Conversely, a fibrillarin-Nop5 heterodimer of *Pyrococcus abyssi* was recently found to perform *in vitro* 2'-*O*-methylation of rRNA independently of L7Ae and C/D box sRNAs (Tomkuviene et al., 2017). In C/D box sRNPs containing fibrillarin, recent evidence shows that the guide RNA sequence determines the affinity of fibrillarin for the substrate and the extent of fibrillarin binding correlates with the efficiency of methylation (Graziadei et al., 2020).

First reports of the structure of the C/D box sRNP complex provided contradictory results for arrangement and number of associated proteins. However, it soon became clear that observed differences were caused by the type of C/D box sRNA that had been utilized in the *in vitro* experiments. While the usage of an artificial two-stranded RNA lacking the k-loop motif lead to the assembly of a monomeric complex consisting of one RNA and two copies of each protein, the usage of an *in vitro* transcribed natural C/D box sRNA sequence lead to the assembly of dimeric complex consisting of two RNAs and four copies of each protein (**Figure 1B**; Bleichert et al., 2009; Bleichert and Baserga, 2010; Xue et al., 2010; Lin et al., 2011; Bower-Phipps et al., 2012; Lapinaite et al., 2013). Accordingly, these results suggest that the nature of the RNA determines if mono- or diRNPs are assembled and influences these complexes' functional roles. These and other findings on the structural diversity of C/D box sRNPs are extensively reviewed by Yu et al. (2018).

C/D BOX sRNA TARGETS

Ribosomal Targets

A first study aiming to identify archaeal sRNAs employed co-immunoprecipitation with archaeal fibrillarin and Nop5 and identified 18 C/D box sRNAs in *Sulfolobus acidocaldarius*. Furthermore, methylations at the predicted target positions for six of these sRNAs were verified using deoxyribonucleotide triphosphate (dNTP) concentration-dependent primer extension assays (Omer et al., 2000). Subsequent experiments lead to the discovery of over 200 sRNAs across seven archaeal species, targeting mostly – though not exclusively – archaeal rRNAs. Here, it was also revealed that, in contrast to eukaryotes, most archaeal sRNAs possess two sequences able to guide methylation and that these double guides can target closely linked positions on the same RNA molecule (Omer et al., 2000). At the same time, another study reported the identification of a family of 46 archaeal sRNAs in the genomes of three species of the hyperthermophile *Pyrococcus* species Additionally, these sRNAs were experimentally verified in *P. abyssi* using Northern hybridization (Gaspin et al., 2000).

Shortly afterwards, another study used a combination of MALDI-MS and primer extension assays to locate conserved modification patterns in the A-loop region of the 23S rRNA in five archaeal and eubacterial species. The A-loop of the 23S rRNA (also known as helix 92), constitutes part of the peptidyl transferase loop in domain V of the 23S rRNA and its functional importance has been emphasized by several studies (Hansen et al., 2002). In fact, loss of the 2'-*O*-ribose methylation at position U2552 in the A-loop leads to decreased growth rate and reduced protein synthesis activity in *Escherichia coli* (Caldas et al., 2000). It was shown that despite variation in the exact positions of modifications in the helices 90–92, modifications in the A-loop are always present at positions equivalent to U2552 and/or G2553 in *E. coli* (Hansen et al., 2002). Projecting these previously identified modifications from *E. coli* onto their corresponding positions in the 2.4 Å X-ray crystal structure of the *Haloarcula marismortui* 50S ribosome subunit, all modifications were found to be clustered around the peptidyl transferase center (Ban et al., 2000; Hansen et al., 2002).

The advent of RNA-seq has enabled researchers to efficiently identify C/D box sRNAs among any organism's total RNA pool. Subsequently, their guide sequences can be used to computationally predict their potential RNA targets on the basis of their hybridization potential. Analyses of RNA-seq coverage revealed large numbers of abundantly transcribed small RNAs with readily identifiable C and D box sequences. These postulated C/D box sRNAs were, for example, described for model archaea of different archaeal phyla: *Nanoarchaeum equitans* (26 C/D box sRNAs), *Ignicoccus hospitalis* (128 C/D box sRNAs), *Methanococcus maripaludis* (7 C/D box sRNAs), *Methanopyrus kandleri* (127 C/D box sRNAs), *Pyrobaculum calidifontis* (88 C/D box sRNAs), *S. acidocaldarius* (61 C/D box sRNAs) and *Thermoproteus tenax* (52 C/D box sRNAs). Using the guide sequences of these C/D box sRNAs, 719 potential 2'-*O*-methylation sites in the archaeal 23S and 16S rRNA sequences were identified and hinted at common targets and rRNA regions (**Figure 2**). This dataset revealed some shared methylation targets but did not reveal a single position to be uniformly present in all seven species. Instead, it became clear that methylation targets cluster in hotspot regions of the rRNA molecules. Among all investigated species, these methylation hotspots have been detected in the functionally important and evolutionary conserved regions of the ribosome (Liang et al., 2009; Dennis et al., 2015; Lui et al., 2018).

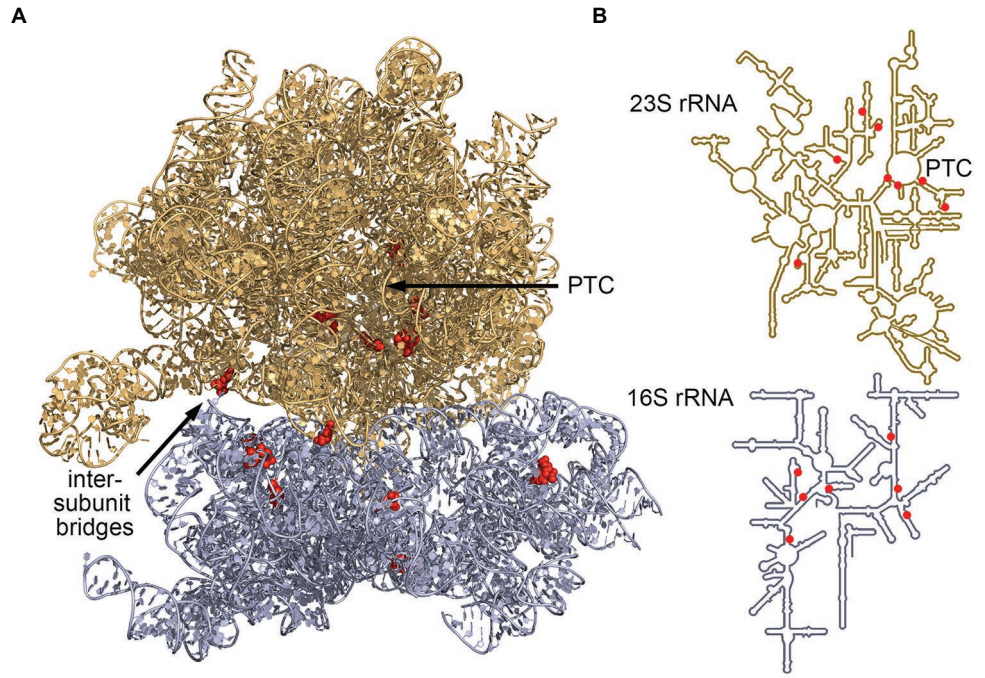


FIGURE 2 | Most conserved C/D box sRNA guide targets of archaeal 16S and 23S rRNA. Analysis of C/D box sRNA guides of seven archaeal species (Dennis et al., 2015) identified seven 16S rRNA sites and eight 23S rRNA sites that are targeted by a minimum of four guides. These sites (red) are clustered at the ribosome core [peptidyl-transferase center (PTC)] and at the intersubunit bridges. **(A)** Positions were mapped onto the ribosome structure of *Thermococcus kodakarensis* (pdb-id: 6SKF; Sas-Chen et al., 2020) and **(B)** onto the secondary structure representations of archaeal rRNAs (Petrov et al., 2014).

The archaeal consensus 23S rRNA structure exhibits six domains surrounding a central core. Conserved methylation hotspots are identified in domain II helices 35 and 35a, domain IV helices 61 and 68–71, and domain V helices 90–93 (Dennis et al., 2015). These regions correspond to the ancient core of the ribosome where domain V lies at the center of the large ribosomal subunit. One major cluster of hotspots surrounds the catalytic peptidyl transferase center located in domain V, where peptide bond formation and peptide release occurs (**Figure 2**; Petrov et al., 2013; Dennis et al., 2015; Lui et al., 2018). Another predicted cluster lies in domain IV where helices 68, 69, and 71 form part of the interface between the large and the small ribosomal subunits (Cate et al., 1999; Dennis et al., 2015; Lui et al., 2018).

The archaeal consensus 16S rRNA structure consists of four domains connected by a central core which is located close to the functional decoding center of the small ribosomal subunit (Wimberly et al., 2000). Here, conserved methylation hotspots are predicted in helices 3, 18, and 27 (Dennis et al., 2015). In fact, helix 18 is the core of the decoding center and responsible for monitoring the codon-anticodon pairings (Ogle et al., 2001). Due to their location, the methylated nucleotides likely contribute to stabilizing the decoding center as well as the association of the four domains. Furthermore, predicted methylation clusters are especially dense in regions which are not protected by RNA-binding proteins. It has therefore been proposed that the modifications help stabilize the structure of these exposed regions and in turn support subunit interactions (Dennis et al., 2015). These findings are corroborated by a study across six *Pyrobaculum* species, revealing that most rRNA-targeting C/D box sRNAs are dual guides targeting sites within 100 nt of each other (Lui et al., 2018).

Non-ribosomal Targets

The initial prediction and experimental verification of archaeal rRNA targets of C/D box sRNAs revealed additional antisense elements matching tRNAs (Gaspin et al., 2000; Omer et al., 2000). Further investigation revealed the presence of four C/D box sRNAs targeting the first position of the anticodon of the tRNAs tRNA-Leu (CAA), tRNA-Leu (UAA), tRNA-Met and tRNA-Trp in three *Pyrococcus* species. One of these sRNAs, termed sR50, corresponds to the intron of its predicted target, the pre-tRNA-Trp and was shown to guide the methylation of the pre-tRNA-Trp nucleotides Cm34 and Um39 in *in vitro* experiments in *Haloferax volcanii* (Omer et al., 2000; D'Orval et al., 2001). This proposed *cis*-acting mechanism was later shown to be a *trans*-acting or intramolecular mechanism by a study which also revealed the sequential pattern of C/D box sRNA-guided methylation (Bortolin et al., 2003; Singh et al., 2004). Recently, an in-depth analysis of C/D box sRNA families in *Pyrobaculum* species computationally predicted tRNA targets for 16% of the identified guide sequences (as opposed to 56% rRNA targets; Lui et al., 2018). Unsurprisingly, the tRNA methylation targets correspond to

structurally conserved regions, however, in contrast to rRNA methylation *via* double-guide sRNAs, tRNA methylation is mediated by sRNAs where one guide targets the tRNA while the other guide has no target. Interestingly, a tRNA-targeting sRNA can mediate methylation of either a single or many tRNAs, depending on whether its guide targets a unique sequence like the region surrounding the wobble base in the anticodon (position 34) or a conserved sequence shared across different tRNA families (Lui et al., 2018). Conversely, a few tRNAs in *S. acidocaldarius* and *P. furiosus* contain several methylated or predicted methylated nucleotides which have not yet been linked to corresponding sRNAs or stand-alone specific methyltransferases (Wolff et al., 2020).

Several computational studies also identified numerous "orphan guides," which are defined as snoRNA guide sequences that do not show complementary to any known RNA target (Hüttenhofer et al., 2001; Yang et al., 2006; Lui et al., 2018). Looking for target RNA sequences outside of rRNAs or tRNAs has relieved many guide sequences of their "orphan status," thereby contributing to the expanding functional diversity of snoRNAs (see also Discussion section and reviewed in Falaleeva et al., 2017; Bergeron et al., 2020). At least some orphan guides can be considered as a consequence of accumulated mutations in the guide sequence, eventually leading to the evolution of guides for novel methylation targets (Dennis et al., 2001; Lui et al., 2018).

C/D Box sRNAs and Their Role as RNA Chaperones

Shortly after their discovery, it was suggested that C/D box sRNAs might function as RNA chaperones (Steitz and Tycowski, 1995). This theory gained further support by a subsequent computer simulation of long-range rRNA interactions at elevated temperatures in the presence of double-guided C/D box sRNAs (Schoemaker and Gultyaev, 2006). Indeed, there exist several C/D box sRNAs whose guide sequences target positions at a considerable distance from each other. For example, *N. equitans* exhibits two instances where the predicted targets are distant in sequence but close in the secondary structure: the guides of sR17 target nucleotides situated on opposing strands of helix 28, the defining helix of domain III of the 16S rRNA and the guides of sR15 target nucleotides on opposing strands of helix 30 which defines a large subsection of domain III (Dennis et al., 2015). These findings lead to the conclusion that the C/D box sRNAs act as chaperones by bringing distant rRNA sequences together and facilitate their annealing, thereby assisting in ribosome subunit assembly (Gaspin et al., 2000; Dennis et al., 2015).

C/D BOX sRNA GENE AND GUIDE EVOLUTION

Genomic Context Variability

In yeast, most snoRNA genes are transcribed from independent RNA polymerase II or III promoters as mono- or polycistronic transcripts (Li et al., 2005; Dieci et al., 2009). In plants, the C/D box sRNA genes exist almost exclusively as polycistronic clusters, or, to a lesser extent, as dicistronic tRNA-C/D box snoRNAs (Leader et al., 1997; Kruszka et al., 2003; Barbezier et al., 2009). In vertebrates, independent promoters are rare and snoRNA genes are usually located in introns of protein-coding or non-protein-coding genes (Pelczar and Filipowicz, 1998; Weber, 2006) and only few of them exist as polycistrons (Leader et al., 1994; Tycowski et al., 2004). The polycistronic transcripts or intron-located C/D box snoRNAs are processed and matured by endo- and exoribonucleolytic activities (Caffarelli et al., 1994; Kiss and Filipowicz, 1995; Villa et al., 1998).

In archaea, analysis of the genomic context of C/D box sRNA genes revealed a variable organization of promoter and processing elements. A 2017 study on the genomic context of C/D box sRNAs in six archaeal model organisms concluded that only a minority of archaeal C/D box sRNAs are transcribed from independent promoters as only 20% of all investigated genes exhibited a conserved, TATA box-like motif in their 50 nt upstream region (Tripp et al., 2017). Instead, the majority of C/D box sRNA genes overlap with either the 5' or the 3' end of a neighboring open reading frame (ORF; Gaspin et al., 2000; Dennis et al., 2001; Randau, 2012; Tripp et al., 2017; Lui et al., 2018). Twenty-five percent of genes show overlap with a 3' end and carry the stop codon of the upstream gene within their sequence. Though the stop codon can be found in any of the four conserved motifs, in almost 50% of cases it was located in the C box motif (Tripp et al., 2017). Notably, the stop codon "UGA" is found in most of the C box and D box motifs and only a few nucleotide changes separate it from evolving into a k-turn element. Therefore, it was proposed that the start or stop codons of overlapping genes are responsible for the accelerated evolution of k-turn motifs in C/D box sRNA genes (Tripp et al., 2017). A 5' overlap was identified for 7% of the investigated C/D box sRNA genes. Similarly, the start codon of an overlapping downstream coding region was found to be located within different parts of C/D box sRNA sequences, most commonly however, in the guide sequence or downstream of the D box motif (Tripp et al., 2017). Analyses of the impact of these overlaps on the transcription rate of a reporter gene revealed neutral or only slightly negative effects for 3' overlapping C/D box sRNAs genes. However, a 5' overlap caused a significant reduction in transcription of the downstream gene, which is in agreement with the rare presence of this gene arrangement in nature (Tripp et al., 2017). In some cases, this scenario might result in the formation of pseudogene sequences downstream of C/D box sRNA genes.

A significant fraction of C/D box sRNA genes was found to occur in clusters of two or three genes indicating polycistronic transcription. Several dicistronic transcripts, including examples of tRNA- C/D box sRNA fusions, were identified (Tripp et al., 2017). In these cases, different C/D box sRNAs were, for example, found to be located directly downstream of genes coding for tRNASer in *I. hospitalis*, tRNAPro in *T. tenax*, and tRNAVal in *N. equitans*. Consequently, tRNA 3' maturation is suggested to generate the 5' terminus of the respective C/D box sRNAs. In some cases, C/D box

sRNAs were also found to be located within tRNA introns and shown to mediate tRNA methylation *in cis* (D'Orval et al., 2001). The majority of the remaining genes are located in intergenic regions and some of them exhibit the aforementioned, conserved motifs, indicating the presence of an independent promoter. Others are located up- or downstream of neighboring protein-coding genes at a distance of less than 25 nt. Consequently, most C/D box sRNA genes do not require independent promoters, as they are part of longer precursor transcripts that are subsequently processed into mature C/D box sRNAs (Tripp et al., 2017).

Mutational analyses of *S. acidocaldarius* upstream and downstream regions of a C/D box sRNA gene revealed that these surrounding sequences can be changed without affecting C/D box sRNA maturation. Instead, the presence of the conserved, internal box motifs responsible for forming the k-turn and k-loop structures was found to be essential (Tripp et al., 2017). These observations suggest that the insertion of a C/D box sRNA gene into a transcriptionally active genome context is sufficient to obtain mature C/D box sRNAs. In this model, C/D box sRNP formation would result in the protection of the C/D box sRNA body *via* protein-RNA contacts while the exposed RNA termini would gradually be processed by cellular nucleases and/or chemical RNA degradation at elevated temperatures. In addition, interactions with RNA ligases would then yield fractions of circularized C/D box sRNAs without accessible RNA termini.

Identification of Guide Sequences

The conserved box C and box D sequences of C/D box RNAs and their evenly spaced arrangement into two k-turns for L7Ae binding allow for the computational prediction of C/D box sRNA genes among archaeal sequences. One of the first programs that used these features to scan genomes for snoRNAs and their putative methylation targets in rRNA is snoScan (Lowe and Eddy, 1999). This program applies probabilistic models of snoRNAs and initially identified 22 novel guides in *Saccharomyces cerevisiae*. Although the initial development of snoScan focused on the prediction of rRNA methylation sites, different RNA sequences can also be used for target prediction. Another tool that utilizes probabilistic models is snoSeeker (Yang et al., 2006). This program searches for box C and box D elements, terminal stem pairing and, optionally, target sequences, enabling prediction of both guide and orphan sRNAs. The algorithm SnoReport utilizes support vector machines (SVM) and RNA secondary structure prediction to identify C/D box sRNA sequences (Hertel et al., 2008; de Araujo Oliveira et al., 2016). A more recent version, SnoReport 2.0, takes advantage of features of known C/D box sRNAs detected in invertebrates to improve its SVM during the training phase. In addition, a k-turn test, in which the predicted sRNAs must present G.A dinucleotides in box C and D, at least one uridine for the U-U pair and a Watson-Crick base pair between the sixth nt of the C box and the first nt of the D box significantly reduced the number of false positives (de Araujo Oliveira et al., 2016). Additionally, snoStrip (Bartschat et al., 2014) is a comprehensive pipeline that applies the following steps: first, a sequence-based homology search is performed using BLASTn (Altschul et al., 1990) and further complemented with the generation of covariance models. Next, the detection of characteristic C and D box motifs is performed through temporary alignments using MUSCLE (Edgar, 2004). If the location of a box motif agrees in all alignments, the position is annotated as a candidate box sequence. After defining conserved sequence elements, a secondary structure analysis is employed to ensure that only correctly folded C/D box RNAs are further analyzed. Finally, prediction of the putative targets is achieved using Plexy, a tool that calculates the optimal thermodynamic interactions of a C/D box sRNA with candidate targets (Kehr et al., 2011). As the repertoire of sequenced C/D box sRNAs increases, several databases have been created to categorize these molecules, including the Plant snoRNA database (Brown et al., 2001). For archaea, two databases can be highlighted: Rfam and snoRNAdb (Lowe and Eddy, 1999; Omer et al., 2000; Kalvari et al., 2021). The Rfam database uses a generalized search based on covariance models to annotate a wide diversity of non-coding RNAs, including C/D box sRNAs, that are conserved in three or more species. The database snoRNAdb compiles homologs of C/D box sRNAs that were predicted for crenarchaeal and euryarchaeal species, while also providing information about their putative targets.

Even though these tools and different strategies are available for C/D box sRNA prediction, this class of RNA is still underrepresented in archaeal annotations (Gardner et al., 2010) and only a combination of RNA-seq analyses, comparative genomics and computational methods allow for complete C/D box sRNA identification (Lui et al., 2018). Since most prediction algorithms were developed using eukaryotic C/D box sRNAs as training sets, it is hypothesized that features which are exclusive to archaeal C/D box sRNAs are absent, therefore impacting the overall efficiency and reliability of the predictions. Here, the lower degree of conservation of C' and D' boxes in eukaryotes in comparison to archaea represents one clear difference (Yang et al., 2020). The recent increase in the availability of archaeal transcriptome datasets is an asset to expand the repertoire of hand-curated C/D box sRNAs. Utilizing experimentally validated datasets, tools that are based on pre-generated covariance models (e.g. INFERNAL – cmsearch) can take advantage of the conserved C' and D' motifs to drastically increase the number of predicted C/D box sRNAs in Archaea (Lui et al., 2018). The reduction of the stringency of the search parameters for C/D box sRNA genes results in increasing amounts of false positive sequences resembling C/D box sRNA genes. These hits can be viewed as sequence space with increased probability of evolving novel C/D box sRNA elements and might impact the dynamics of guide sequence generation.

DISCUSSION

C/D box sRNAs were early found to possess other functions besides their established role in the 2'-*O*-methylation of rRNA and tRNA (Dennis et al., 2001). C/D box sRNAs of yeast and

eukaryotes (especially humans) have been shown to be involved in diverse functions including rRNA processing, RNA base acetylation, regulation of mRNA 3' processing, and alternative pre-mRNA splicing (Kass et al., 1990; Falaleeva et al., 2016; Huang et al., 2017; Sharma et al., 2017b). Recently, it was shown that C/D box snoRNA-guided methylation of mRNA regulates protein expression and enzyme activity (Elliott et al., 2019). Additionally, it was revealed that many snoRNAs are processed into shorter forms such as miRNA (called sno-miRNA) and efficiently exert gene regulatory functions (Brameier et al., 2011). It was also recently discovered that snoRNAs retained in longer RNAs can interact with non-canonical proteins and act as a decoy, thereby hindering their activity (Bergeron et al., 2020). In fact, the influence of C/D box snoRNAs in the human metabolism is very significant: the C/D box snoRNA *U60* is involved in intracellular cholesterol trafficking and regulation of cholesterol homeostasis (Brandis et al., 2013). Lack of expression of the C/D box snoRNA cluster *SNORD116* causes Prader-Willi-Syndrome, a neurobehavioral disorder manifesting itself in hyperphagia and leading to morbid obesity (Ding et al., 2008; Duker et al., 2010). Furthermore, the snoRNA *U50* was found to be deleted in several common cancers, with a particularly strong association in breast cancer and prostate cancer (Dong et al., 2008, 2009; Siprashvili et al., 2015). With an evident link between snoRNAs and human cancer and other systemic diseases being established, a strong resurgence of eukaryotic snoRNA research has been noted. New findings in this area continually expand our knowledge of diverse snoRNA functions and have most recently been reviewed by Deogharia and Majumder (2019), Liang et al. (2019) and Bratkovič et al. (2020).

Using RiboMethSeq to analyze the 2'-*O*-methylation patterns on eukaryotic rRNAs, it was shown that a knockdown of the methyltransferase fibrillarin (FBL) in HeLa-cells leads to a site-specific decrease of methylation levels. Affected sites were identified in conserved and/or functionally important regions of the ribosome, like its "core," close to the A- and P-sites, the intersubunit bridges and the peptide exit tunnel, while 2'-*O*-Me sites close to the peptidyl transferase center were not subject to variation in methylation levels upon FBL knockdown (Erales et al., 2017). Another study from the same year mapping 2'-*O*-methylation sites vulnerable to fibrillarin depletion on human rRNAs, also investigated the C/D box sRNAs whose guide sequences target these "vulnerable" methylation sites. However, these studies did not find a direct correlation between the sites with a variable methylation level and abundance of the sRNAs which target them (Sharma et al., 2017a). More recently, RiboMethSeq was adapted to map 2'-*O*-methylation sites on rRNAs in human breast cancer samples (Marcel et al., 2020). Here, the identified methylation sites were divided into two classes: one class encompassing a larger group of rRNA 2'-*O*-methylation sites with a low inter-patient variability, termed "stable" sites, and a second class encompassing a smaller group of rRNA 2'-*O*-methylation sites with a high inter-patient variability in methylation levels, termed "variable" sites. These stable sites were found to be located in the decoding center, the peptidyl transferase center and the polypeptide exit tunnel, while the variable sites were located in layers 1 or 2 nt away from these functional regions. Furthermore, it is suggested that the 2'-*O*-methylation levels at the variable rRNA sites are associated with breast cancer subtype and tumor grade, indicating that not only tumor size but also the pattern of rRNA 2'-*O*-methylation influences factors like tumor aggressiveness and patient survival (Marcel et al., 2020).

Additional functional roles of archaeal C/D box sRNAs are likely also to be discovered, which is supported by the existence of many "orphan" C/D box sRNA guides in archaea without easily detectable complementary methylation targets. Guide sequences of C/D box sRNAs define the methylation landscape of their hybridization targets. As these targets are mostly highly conserved rRNA molecules, it is initially surprising to see that C/D box sRNA do not exhibit a similar degree of conservation. As described in Ribosomal Targets section, ubiquitous methylation of functionally and structurally important rRNA regions can be achieved by different sets of C/D box sRNAs with varied guide sequences and guide sequence pairs. This dynamic evolution of guides has been analyzed in detail in six *Pyrobaculum* species containing 526 different C/D box sRNAs that were organized into 110 homologous families (Lui et al., 2018). At the genus level, less than two-thirds of the predicted targets were found to be conserved among the six *Pyrobaculum* species and guide sequences exhibited short insertions, deletions or substitutions. In the *Pyrobaculum* species dataset, 28% of guides showed no significant complementarity to potential RNA targets (Lui et al., 2018). As C/D box sRNA genes often overlap with adjacent genes that provide promoter elements and processing signals, it is also possible that the overlapping sequence results in the creation of an orphan guide sequence that is paired with a second guide that provides methylation benefits for the cell. Therefore, the presence of orphan guide sequences can partly be considered to be a consequence of the plasticity of the genomic context of C/D box sRNA genes. Here, it remains to be understood why C/D box sRNAs exhibit dynamic scenarios of polycistronic transcriptional units with different mRNA and tRNA partners. In mammalian cells, snoRNAs have been found in retroposable elements and it was proposed that retroposition followed by genetic drift would be able to increase snoRNA diversity and change their modification landscape (Weber, 2006). Mobile features of archaeal C/D box sRNA genes remain to be discovered.

AUTHOR CONTRIBUTIONS

RB, J-VG-F, and LR conceptualized and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided by the Deutsche Forschungsgemeinschaft (DFG), project RA 2169/4-1.

REFERENCES

Aittaleb, M., Rashid, R., Chen, Q., Palmer, J. R., Daniels, C. J., and Li, H. (2003). Structure and function of archaeal box C/D sRNP core proteins. *Nat. Struct. Biol.* 10, 256–263. doi: 10.1038/nsb905

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Auffinger, P., and Westhof, E. (1997). Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.* 274, 54–63. doi: 10.1006/jmbi.1997.1370

Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905–920. doi: 10.1126/science.289.5481.905

Barbezier, N., Canino, G., Rodor, J., Jobet, E., Saez-Vasquez, J., Marchfelder, A., et al. (2009). Processing of a dicistronic tRNA-snoRNA precursor: combined analysis in vitro and in vivo reveals alternate pathways and coupling to assembly of snoRNP. *Plant Physiol.* 150, 1598–1610. doi: 10.1104/pp.109.137968

Bartschat, S., Kehr, S., Tafer, H., Stadler, P. F., and Hertel, J. (2014). SnoStrip: a snorna annotation pipeline. *Bioinformatics* 30, 115–116. doi: 10.1093/bioinformatics/btt604

Bergeron, D., Fafard-Couture, É., and Scott, M. S. (2020). Small nucleolar RNAs: continuing identification of novel members and increasing diversity of their molecular mechanisms of action. *Biochem. Soc. Trans.* 48, 645–656. doi: 10.1042/BST20191046

Bleichert, F., and Baserga, S. J. (2010). Dissecting the role of conserved box C/D sRNA sequences in di-sRNP assembly and function. *Nucleic Acids Res.* 38, 8295–8305. doi: 10.1093/nar/gkq690

Bleichert, F., Gagnon, K. T., Brown, B. A. II, Maxwell, E. S., Leschziner, A. E., Unger, V. M., et al. (2009). A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science* 325, 1384–1388. doi: 10.1126/science.1176099

Bortolin, M. L., Bachellerie, J. P., and Clouet-d'Orval, B. (2003). In vitro RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNATrp. *Nucleic Acids Res.* 31, 6524–6535. doi: 10.1093/nar/gkg860

Bower-Phipps, K. R., Taylor, D. W., Wang, H. W., and Baserga, S. J. (2012). The box C/D sRNP dimeric architecture is conserved across domain archaea. *RNA* 18, 1527–1540. doi: 10.1261/rna.033134.112

Brameier, M., Herwig, A., Reinhardt, R., Walter, L., and Gruber, J. (2011). Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic Acids Res.* 39, 675–686. doi: 10.1093/nar/gkq776

Brandis, K. A., Gale, S., Jinn, S., Langmade, S. J., Dudley-Rucker, N., Jiang, H., et al. (2013). Box C/D small nucleolar RNA (snoRNA) U60 regulates intracellular cholesterol trafficking. *J. Biol. Chem.* 288, 35703–35713. doi: 10.1074/jbc.M113.488577

Bratkovič, T., Božič, J., and Rogelj, B. (2020). Functional diversity of small nucleolar RNAs. *Nucleic Acids Res.* 48, 1627–1651. doi: 10.1093/nar/gkz1140

Brown, J. W. S., Clark, G. P., Leader, D. J., Simpson, C. G., and Lowe, T. (2001). Multiple snoRNA gene clusters from *Arabidopsis*. *RNA* 7, 1817–1832.

Caffarelli, E., Arese, M., Santoro, B., Fragapane, P., and Bozzoni, I. (1994). In vitro study of processing of the intron-encoded U16 small nucleolar RNA in *Xenopus laevis*. *Mol. Cell. Biol.* 14, 2966–2974. doi: 10.1128/mcb.14.5.2966

Caldas, T., Binet, E., Bouloc, P., and Richarme, G. (2000). Translational defects of *Escherichia coli* mutants deficient in the Um2552 23S ribosomal RNA methyltransferase RrmJ/FTSJ. *Biochem. Biophys. Res. Commun.* 271, 714–718. doi: 10.1006/bbrc.2000.2702

Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science* 285, 2095–2104. doi: 10.1126/science.285.5436.2095

D'Orval, B. C., Bortolin, M. L., Gaspin, C., and Bachellerie, J. P. (2001). Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res.* 29, 4518–4529. doi: 10.1093/nar/29.22.4518

Danan, M., Schwartz, S., Edelheit, S., and Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* 40, 3131–3142. doi: 10.1093/nar/gkr1009

de Araujo Oliveira, J. V., Costa, F., Backofen, R., Stadler, P. F., Machado Telles Walter, M. E., and Hertel, J. (2016). SnoReport 2.0: new features and a refined support vector machine to improve snoRNA identification. *BMC Bioinformatics* 17:464. doi: 10.1186/s12859-016-1345-6

Decatur, W. A., and Fournier, M. J. (2002). rRNA modifications and ribosome function. *Trends Biochem. Sci.* 27, 344–351. doi: 10.1016/S0968-0004(02)02109-6

Dennis, P. P., Omer, A., and Lowe, T. (2001). A guided tour: small RNA function in *Aarchaea*. *Mol. Microbiol.* 40, 509–519. doi: 10.1046/j.1365-2958.2001.02381.x

Dennis, P. P., Tripp, V., Lui, L., Lowe, T., and Randau, L. (2015). C/D box sRNA-guided 2'-*O*-methylation patterns of archaeal rRNA molecules. *BMC Genomics* 16:632. doi: 10.1186/s12864-015-1839-z

Deogharia, M., and Majumder, M. (2019). Guide snoRNAs: Drivers or passengers in human disease? *Biology* 8, 1–16. doi: 10.3390/biology8010001

Deryusheva, S., and Gall, J. G. (2019). Small, smaller, smallest: minimal structural requirements for a fully functional box C/D modification guide RNA. *Biomolecules* 9:457. doi: 10.3390/biom9090457

Dieci, G., Preti, M., and Montanini, B. (2009). Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* 94, 83–88. doi: 10.1016/j.ygeno.2009.05.002

Ding, F., Li, H. H., Zhang, S., Solomon, N. M., Camper, S. A., Cohen, P., et al. (2008). SnoRNA Snord116 (Pwcr1/MBll-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS One* 3:e1709. doi: 10.1371/journal.pone.0001709

Dong, X. Y., Guo, P., Boyd, J., Sun, X., Li, Q., Zhou, W., et al. (2009). Implication of snoRNA U50 in human breast cancer. *J. Genet. Genomics* 36, 447–454. doi: 10.1016/S1673-8527(08)60134-4

Dong, X. Y., Rodriguez, C., Guo, P., Sun, X., Talbot, J. T., Zhou, W., et al. (2008). SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum. Mol. Genet.* 17, 1031–1042. doi: 10.1093/hmg/ddm375

Duker, A. L., Ballif, B. C., Bawle, E. V., Person, R. E., Mahadevan, S., Alliman, S., et al. (2010). Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur. J. Hum. Genet.* 18, 1196–1201. doi: 10.1038/ejhg.2010.102

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Elliott, B. A., Ho, H. T., Ranganathan, S. V., Vangaveti, S., Ilkayeva, O., Abou Assi, H., et al. (2019). Modification of messenger RNA by 2'-*O*-methylation regulates gene expression in vivo. *Nat. Commun.* 10:3401. doi: 10.1038/s41467-019-11375-7

Englert, M., Sheppard, K., Aslanian, A., Yates, J. R., and Söll, D. (2011). Archaeal 3'-phosphate RNA splicing ligase characterization identifies the missing component in tRNA maturation. *Proc. Natl. Acad. Sci. U. S. A.* 108, 1290–1295. doi: 10.1073/pnas.1018307108

Erales, J., Marchand, V., Panthu, B., Gillot, S., Belin, S., Ghayad, S. E., et al. (2017). Evidence for rRNA 2'-*O*-methylation plasticity: control of intrinsic translational capabilities of human ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* 114, 12934–12939. doi: 10.1073/pnas.1707674114

Falaleeva, M., Pages, A., Matuszek, Z., Hidmi, S., Agranat-Tamir, L., Korotkov, K., et al. (2016). Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proc. Natl. Acad. Sci. U. S. A.* 113, E1625–E1634. doi: 10.1073/pnas.1519292113

Falaleeva, M., Welden, J. R., Duncan, M. J., and Stamm, S. (2017). C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: old dogs show new tricks. *BioEssays* 39:201600264. doi: 10.1002/bies.201600264

Filipowicz, W., and Kiss, T. (1993). Structure and function of nucleolar snRNPs. *Mol. Biol. Rep.* 18, 149–156. doi: 10.1007/BF00986770

Gagnon, K. T., Zhang, X., Qu, G., Biswas, S., Suryadi, J., Brown, B. A., et al. (2010). Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA* 16, 79–90. doi: 10.1261/rna.1692310

Gardner, P. P., Bateman, A., and Poole, A. M. (2010). SnoPatrol: how many snoRNA genes are there? *J. Biol.* 9:4. doi: 10.1186/jbiol211

Gaspin, C., Cavaillé, J., Erauso, G., and Bachellerie, J. -P. (2000). Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.* 297, 895–906. doi: 10.1006/JMBI.2000.3593

Graziadei, A., Gabel, F., Kirkpatrick, J., and Carlomagno, T. (2020). The guide sRNA sequence determines the activity level of BOX C/D RNPs. *elife* 9, 1–27. doi: 10.7554/eLife.50027

Hansen, M. A., Kirpekar, F., Ritterbusch, W., and Vester, B. (2002). Posttranscriptional modifications in the A-loop of 23S rRNAs from selected archaea and eubacteria. *RNA* 8, 202–213. doi: 10.1017/S1355838202013365

Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* 34, 721–733. doi: 10.1093/nar/gkj471

Herschlag, D., Eckstein, F., and Cech, T. R. (1993). The importance of being ribose at the cleavage site in the Tetrahymena ribozyme reaction. *Biochemistry* 32, 8312–8321. doi: 10.1021/bi00083a035

Hertel, J., Hofacker, I. I., and Stadler, P. F. (2008). SnoReport: computational identification of snoRNAs with unknown targets. *Bioinformatics* 24, 158–164. doi: 10.1093/bioinformatics/btm464

Huang, C., Shi, J., Guo, Y., Huang, W., Huang, S., Ming, S., et al. (2017). A snoRNA modulates mRNA 3' end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res.* 45, 8647–8660. doi: 10.1093/nar/gkx651

Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J. P., et al. (2001). Rnomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.* 20, 2943–2953. doi: 10.1093/emboj/20.11.2943

Jüttner, M., Weiß, M., Ostheimer, N., Reglin, C., Kern, M., Knüppel, R., et al. (2020). A versatile cis-acting element reporter system to study the function, maturation and stability of ribosomal RNA mutants in archaea. *Nucleic Acids Res.* 48, 2073–2090. doi: 10.1093/nar/gkz1156

Kalvari, I., Nawrocki, E. P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* 49, D192–D200. doi: 10.1093/nar/gkaa1047

Kass, S., Tyc, K., Steitz, J. A., and Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell* 60, 897–908. doi: 10.1016/0092-8674(90)90338-F

Kawai, G., Yamamoto, Y., Watanabe, T., Yokoyama, S., Kamimura, T., Masegi, T., et al. (1992). Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* 31, 1040–1046. doi: 10.1021/bi00119a012

Kehr, S., Bartschat, S., Stadler, P. F., and Tafer, H. (2011). PLEXY: efficient target prediction for box C/D snoRNAs. *Bioinformatics* 27, 279–280. doi: 10.1093/bioinformatics/btq642

Kiss, T., and Filipowicz, W. (1995). Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev.* 9, 1411–1424. doi: 10.1101/gad.9.11.1411

Kiss-László, Z., Henry, Y., Bachellerie, J. -P., Caizergues-Ferrer, M., and Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* 85, 1077–1088. doi: 10.1016/S0092-8674(00)81308-2

Kiss-Laszlo, Z., Henry, Y. H., and Kiss, T. (1998). Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.* 17, 797–807. doi: 10.1093/emboj/17.3.797

Klein, D. J., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. *EMBO J.* 20, 4214–4221. doi: 10.1093/emboj/20.15.4214

Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S., et al. (2003). Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J.* 22, 621–632. doi: 10.1093/emboj/cdg040

Kuhn, J. F., Tran, E. J., and Maxwell, E. S. (2002). Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res.* 30, 931–941. doi: 10.1093/nar/30.4.931

Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F., and Carlomagno, T. (2013). The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* 502, 519–523. doi: 10.1038/nature12581

Leader, D. J., Clark, G. P., Watters, J., Beven, A. F., Shaw, P. J., and Brown, J. W. S. (1997). Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs. *EMBO J.* 16, 5742–5751. doi: 10.1093/emboj/16.18.5742

Leader, D. J., Sanders, J. F., Waugh, R., Shaw, P., and Brown, J. W. S. (1994). Molecular characterisation of plant U14 small nucleolar RNA genes: closely linked genes are transcribed as polycistronic U14 transcripts. *Nucleic Acids Res.* 22, 5196–5203. doi: 10.1093/nar/22.24.5196

Li, S. G., Zhou, H., Luo, Y. P., Zhang, P., and Qu, L. H. (2005). Identification and functional analysis of 20 box H/ACA small nucleolar RNAs (snoRNAs) from *Schizosaccharomyces pombe*. *J. Biol. Chem.* 280, 16446–16455. doi: 10.1074/jbc.M500326200

Liang, X. -H., Liu, Q., and Fournier, M. J. (2009). Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* 15, 1716–1728. doi: 10.1261/rna.1724409

Liang, J., Wen, J., Huang, Z., Chen, X., Zhang, B., and Chu, L. (2019). Small Nucleolar RNAs: insight into their function in cancer. *Front. Oncol.* 9:587. doi: 10.3389/fonc.2019.00587

Lin, J., Lai, S., Jia, R., Xu, A., Zhang, L., Lu, J., et al. (2011). Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* 469, 559–564. doi: 10.1038/nature09688

Lischwe, M. A., Ochs, R. L., Reddy, R., Cook, G., Yeoman, L. C., Tang, E. M., et al. (1985). Purification and partial characterization of a nucleolar scleroderma antigen (Mr = 34,000; pI, 8.5) rich in NG, NG-dimethylarginine. *J. Biol. Chem.* 260, 14304–14310. doi: 10.1016/S0021-9258(17)38718-5

Lowe, T. M., and Eddy, S. R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science* 283, 1168–1171. doi: 10.1126/science.283.5405.1168

Lui, L., and Lowe, T. (2013). Small nucleolar RNAs and RNAguided post-transcriptional modification. *Essays Biochem.* 54, 53–77. doi: 10.1042/BSE0540053

Lui, L. M., Uzilov, A. V., Bernick, D. L., Corredor, A., Lowe, T. M., and Dennis, P. P. (2018). Methylation guide RNA evolution in archaea: structure, function and genomic organization of 110 C/D box sRNA families across six *Pyrobaculum* species. *Nucleic Acids Res.* 46, 5678–5691. doi: 10.1093/nar/gky284

Maden, B. E. H., Corbett, M. E., Heeney, P. A., Pugh, K., and Ajuh, P. M. (1995). Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie* 77, 22–29. doi: 10.1016/0300-9084(96)88100-4

Marcel, V., Kielbassa, J., Marchand, V., Natchiar, K. S., Paraqindes, H., Nguyen Van Long, F., et al. (2020). Ribosomal RNA 2′*O*-methylation as a novel layer of inter-tumour heterogeneity in breast cancer. *NAR Cancer* 2, 1–12. doi: 10.1093/narcan/zcaa036

Maxwell, E. S., and Fournier, M. J. (1995). The small nucleolar RNAs. *Annu. Rev. Biochem.* 35, 897–934. doi: 10.1146/annurev.bi.64.070195.004341

Motorin, Y., and Helm, M. (2010). TRNA stabilization by modified nucleotides. *Biochemistry* 49, 4934–4944. doi: 10.1021/bi100408z

Nolivos, S., Carpousis, A. J., and Clouet-d'Orval, B. (2005). The K-loop, a general feature of the *Pyrococcus* C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Res.* 33, 6507–6514. doi: 10.1093/nar/gki962

Noon, K. R., Bruenger, E., and McCloskey, J. A. (1998). Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *J. Bacteriol.* 180, 2883–2888. doi: 10.1128/jb.180.11.2883-2888.1998

Ochs, R. L., Lischwe, M. A., Spohn, W. H., and Busch, H. (1985). Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. *Biol. Cell.* 54, 123–133. doi: 10.1111/j.1768-322X.1985.tb00387.x

Ogle, J. M., Brodersen, D. E., Clemons, J., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292, 897–902. doi: 10.1126/science.1060612

Omer, A. D., Lowe, T. M., Russell, A. C., Ebhardt, H., Eddy, S. R., and Dennis, P. P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science* 288, 517–522. doi: 10.1126/science.288.5465.517

Omer, A. D., Ziesche, S., Ebhardt, H. A., and Dennis, P. P. (2002). In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5289–5294. doi: 10.1073/pnas.082101999

Pelczar, P., and Filipowicz, W. (1998). The host gene for Intronic U17 small Nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol. Cell. Biol.* 18, 4509–4518. doi: 10.1128/mcb.18.8.4509

Petrov, A. S., Bernier, C. R., Gulen, B., Waterbury, C. C., Hershkovits, E., Hsiao, C., et al. (2014). Secondary structures of rRNAs from all three domains of life. *PLoS One* 9:e88222. doi: 10.1371/journal.pone.0088222

Petrov, A. S., Bernier, C. R., Hershkovits, E., Xue, Y., Waterbury, C. C., Hsiao, C., et al. (2013). Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Res.* 41, 7522–7535. doi: 10.1093/nar/gkt513

Plagens, A., Daume, M., Wiegel, J., and Randau, L. (2015). Circularization restores signal recognition particle RNA functionality in *Thermoproteus*. *elife* 4:e11623. doi: 10.7554/eLife.11623.001

Popow, J., Englert, M., Weitzer, S., Schleiffer, A., Mierzwa, B., Mechtler, K., et al. (2011). HSPC117 is the essential subunit of a human tRNA splicing ligase complex. *Science* 331, 760–764. doi: 10.1126/science.1197847

Randau, L. (2012). RNA processing in the minimal organism *Nanoarchaeum equitans*. *Genome Biol.* 13:R63. doi: 10.1186/gb-2012-13-7-r63

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

Schoemaker, R. J. W., and Gultyaev, A. P. (2006). Computer simulation of chaperone effects of Archaeal C/D box sRNA binding on rRNA folding. *Nucleic Acids Res.* 34, 2015–2026. doi: 10.1093/nar/gkl154.

Sharma, S., Marchand, V., Motorin, Y., and Lafontaine, D. L. J. (2017a). Identification of sites of 2'-*O*-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci. Rep.* 7, 1–15. doi: 10.1038/s41598-017-09734-9.

Sharma, S., Yang, J., van Nues, R., Watzinger, P., Kötter, P., Lafontaine, D. L. J., et al. (2017b). Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS Genet.* 13:e1006804. doi: 10.1371/journal.pgen.1006804

Singh, S. K., Gurha, P., Tran, E. J., Maxwell, E. S., and Gupta, R. (2004). Sequential 2'-*O*-methylation of archaeal pre-tRNA Trp nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA. *J. Biol. Chem.* 279, 47661–47671. doi: 10.1074/jbc.M408868200

Siprashvili, Z., Webster, D. E., Johnston, D., Shenoy, R. M., Ungewickell, A. J., Bhaduri, A., et al. (2015). The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer. *Nat. Genet.* 48, 53–58. doi: 10.1038/ng.3452

Starostina, N. G., Marshburn, S., Johnson, L. S., Eddy, S. R., Terns, R. M., and Terns, M. P. (2004). Circular box C/D RNAs in *Pyrococcus furiosus*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14097–14101. doi: 10.1073/pnas.0403520101

Steitz, J. A., and Tycowski, K. T. (1995). Small RNA chaperones for ribosome biogenesis. *Science* 270, 1626–1627. doi: 10.1016/s1937-6448(10)84002-x

Su, A. A. H., Tripp, V., and Randau, L. (2013). RNA-Seq analyses reveal the order of tRNA processing events and the maturation of C/D box and CRISPR RNAs in the hyperthermophile *Methanopyrus kandleri*. *Nucleic Acids Res.* 41, 6250–6258. doi: 10.1093/nar/gkt317

Tomkuviene, M., Ličyte, J., Olendraite, I., Liutkevičiute, Z., Clouet-D'Orval, B., and Klimašauskas, S. (2017). Archaeal fibrillarin-Nop5 heterodimer 2'-*O*-methylates RNA independently of the C/D guide RNP particle. *RNA* 23, 1329–1337. doi: 10.1261/rna.059832.116

Tran, E., Zhang, X., Lackey, L., and Maxwell, E. S. (2005). Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-*O*-methylation of target RNAs. *RNA* 11, 285–293. doi: 10.1261/rna.7223405.

Tripp, V., Martin, R., Orell, A., Alkhnbashi, O. S., Backofen, R., and Randau, L. (2017). Plasticity of archaeal C/D box sRNA biogenesis. *Mol. Microbiol.* 103, 151–164. doi: 10.1111/mmi.13549.

Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., et al. (1997). The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* 89, 849–858. doi: 10.1016/S0092-8674(00)80270-6.

Tycowski, K. T., Aab, A., and Steitz, J. A. (2004). Guide RNAs with 5† caps and novel box C/D snoRNA-like domains for modification of snRNAs in Metazoa. *Curr. Biol.* 14, 1985–1995. doi: 10.1016/j.cub.2004.11.003

Tycowski, K. T., You, Z. H., Graham, P. J., and Steitz, J. A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell* 2, 629–638. doi: 10.1016/S1097-2765(00)80161-6.

Villa, T., Ceradini, F., Presutti, C., and Bozzoni, I. (1998). Processing of the intron-encoded U18 small Nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exo- and endonucleolytic activities. *Mol. Cell. Biol.* 18, 3376–3383. doi: 10.1128/mcb.18.6.3376

Vitali, P., and Kiss, T. (2019). Cooperative 2'-*o*-methylation of the wobble cytidine of human elongator tRNAmet(cat) by a nucleolar and a cajal bodyspecific box C/D RNP. *Genes Dev.* 33, 741–746. doi: 10.1101/gad.326363.119.

Wang, W., Kim, R., Jancarik, J., Yokota, H., and Kim, S. H. (2001). Crystal structure of phosphoserine phosphatase from *Methanococcus jannaschii*, a hyperthermophile, at 1.8 Å resolution. *Structure* 9, 65–71. doi: 10.1016/S0969-2126(00)00558-X

Watkins, N. J., Ségault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., et al. (2000). A common Core RNP structure shared between the small Nucleoar box C/D RNPs and the Spliceosomal U4 snRNP. *Cell* 103, 457–466.

Weber, M. J. (2006). Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet.* 2:e205. doi: 10.1371/journal.pgen.0020205

Williams, D. J., Boots, J. L., and Hall, K. B. (2001). Thermodynamics of 2'-ribose substitutions in UUCG tetraloops. *RNA* 7, 44–53. doi: 10.1017/S1355838201001558

Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrheln, C., et al. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327–339. doi: 10.1038/35030006.

Wolff, P., Villette, C., Zumsteg, J., Heintz, D., Antoine, L., Chane-Woon-Ming, B., et al. (2020). Comparative patterns of modified nucleotides in individual tRNA species from a mesophilic and two thermophilic archaea. *RNA* 26, 1957–1975. doi: 10.1261/RNA.077537.120

Xue, S., Wang, R., Yang, F., Terns, R. M., Terns, M. P., Zhang, X., et al. (2010). Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol. Cell* 39, 939–949. doi: 10.1016/j.molcel.2010.08.022.

Yang, Z., Wang, J., Huang, L., Lilley, D. M. J., and Ye, K. (2020). Functional organization of box C/D RNA-guided RNA methyltransferase. *Nucleic Acids Res.* 48, 5094–5105. doi: 10.1093/nar/gkaa247

Yang, J. H., Zhang, X. C., Huang, Z. P., Zhou, H., Huang, M. B., Zhang, S., et al. (2006). snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. *Nucleic Acids Res.* 34, 5112–5123. doi: 10.1093/nar/gkl672.

Yu, G., Zhao, Y., and Li, H. (2018). The multistructural forms of box C/D ribonucleoprotein particles. *RNA* 24, 1625–1633. doi: 10.1261/rna.068312.118.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 24 March 2021
doi: 10.3389/fmicb.2021.631297

The Archaeal Elongation Factor EF-2 Induces the Release of aIF6 From 50S Ribosomal Subunit

Giada Lo Gullo[1], Maria Luisa De Santis[2], Alessandro Paiardini[3], Serena Rosignoli[3], Alice Romagnoli[4], Anna La Teana[4], Paola Londei[2] and Dario Benelli[2]**

[1] Department of Cellular Biotechnologies and Haematology, Sapienza University of Rome, Rome, Italy, [2] Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy, [3] Department of Biochemical Sciences, Sapienza University of Rome, Rome, Italy, [4] Department of Life and Environmental Science, New York-Marche Structural Biology Center (NY-MaSBiC), Polytechnic University of Marche, Ancona, Italy

The translation factor IF6 is a protein of about 25 kDa shared by the Archaea and the Eukarya but absent in Bacteria. It acts as a ribosome anti-association factor that binds to the large subunit preventing the joining to the small subunit. It must be released from the large ribosomal subunit to permit its entry to the translation cycle. In Eukarya, this process occurs by the coordinated action of the GTPase Efl1 and the docking protein SBDS. Archaea do not possess a homolog of the former factor while they have a homolog of SBDS. In the past, we have determined the function and ribosomal localization of the archaeal (*Sulfolobus solfataricus*) IF6 homolog (aIF6) highlighting its similarity to the eukaryotic counterpart. Here, we analyzed the mechanism of aIF6 release from the large ribosomal subunit. We found that, similarly to the Eukarya, the detachment of aIF6 from the 50S subunit requires a GTPase activity which involves the archaeal elongation factor 2 (aEF-2). However, the release of aIF6 from the 50S subunits does not require the archaeal homolog of SBDS, being on the contrary inhibited by its presence. Molecular modeling, using published structural data of closely related homologous proteins, elucidated the mechanistic interplay between the aIF6, aSBDS, and aEF2 on the ribosome surface. The results suggest that a conformational rearrangement of aEF2, upon GTP hydrolysis, promotes aIF6 ejection. On the other hand, aSBDS and aEF2 share the same binding site, whose occupation by SBDS prevents aEF2 binding, thereby inhibiting aIF6 release.

Keywords: IF6, EF2, ribosome, *Sulfolobus solfataricus*, protein synthesis, SBDS

OPEN ACCESS

Edited by:
Simonetta Gribaldo, Institut Pasteur, France

Reviewed by:
C. Martin Lawrence, Montana State University, United States
Béatrice Clouet-D'orval, Centre National de la Recherche Scientifique (CNRS), France

****Correspondence:***
Paola Londei
paola.londei@uniroma1.it
Dario Benelli
dario.benelli@uniroma1.it

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

Received: *19 November 2020*
Accepted: *11 February 2021*
Published: *24 March 2021*

Citation:
Lo Gullo G, De Santis ML, Paiardini A, Rosignoli S, Romagnoli A, La Teana A, Londei P and Benelli D (2021) The Archaeal Elongation Factor EF-2 Induces the Release of aIF6 From 50S Ribosomal Subunit. Front. Microbiol. 12:631297. doi: 10.3389/fmicb.2021.631297

INTRODUCTION

The process of protein synthesis is conserved in all living organisms and involves ribosomes, mRNA, and different translation factors. Although the overall size of archaeal ribosomes is similar to that of bacterial ones, their components have a closer homology to those of eukaryotic ribosomes. Indeed, as regards the ribosomal proteins (r-proteins), 33 are common to Archaea and Eukarya (A/E), while of the 34 r-proteins conserved in all three domains, the archaeal and eukaryotic homologs are more similar to each other than to the corresponding bacterial r-proteins (Lecompte et al., 2002; Yutin et al., 2012). Besides, the complexity of archaeal translation is also supported

by the larger-than-bacterial number of translation factors, notably translation initiation factors (Dennis, 1997; Benelli and Londei, 2011; Gäbel et al., 2013). The protein known as a/eIF6, a small monomeric polypeptide of about 25 kDa, is one of the translation factors shared by the Archaea and the Eukarya to the exclusion of Bacteria.

In eukaryotes, eIF6 was classified as a translation initiation factor for its ribosome anti-association activity. Indeed, early *in vitro* studies demonstrated the capacity of the protein to bind to the 60S subunit inhibiting its association with the 40S particle (Russell and Spremulli, 1979; Valenzuela et al., 1982). Subsequent structural data showed that eIF6 binds the sarcin-ricin loop (SRL), uL14, and eL24 on the intersubunit face of the large ribosomal subunit preventing ribosomal subunit joining (Gartmann et al., 2010; Klinge et al., 2011; Weis et al., 2015). Genetic studies in *Saccharomyces cerevisiae* showed that eIF6 has a function in the biogenesis and nuclear export of pre-60S subunits (Basu et al., 2001). Later studies confirmed that the removal of eIF6 from the 60S subunit is a late event of ribosome biogenesis and that this step requires the combined action of the GTPase Efl1 and SBDS (Bécam et al., 2001; Menne et al., 2007; Finch et al., 2011; Wong et al., 2011). Specifically, these two factors collaborate to a final quality control assessment for the integrity of the P-site and the GTPase center of the 60S subunit. In mammalian cells, the dislodgement of human eIF6 from the 60S subunit is also described by another model that requires the phosphorylation of the protein on residue S235 by PKCβII kinase recruited on the ribosomes by the receptor for activated C kinase 1 (RACK1) (Ceci et al., 2003). In Archaea, the eIF6 homolog shows a high degree of tertiary structure similarity. Indeed, the A/E factors display a conserved pentein fold (Groft et al., 2000) and this trait suggests that the proteins share a core function conserved in the eukaryal/archaeal line. Indeed, we demonstrated that, similarly to eukaryotes, aIF6 binds to the 30S interacting surface of the large ribosomal subunit, impairing the association between the two subunits (Benelli et al., 2009). Moreover, structural studies confirmed that the ribosome binding site of IF6 is the same as that of its eukaryotic counterpart (Greber et al., 2012).

To date, the molecular mechanism inducing the release of aIF6 from 50S subunits has not yet been determined in Archaea. Phylogenetic analysis of archaeal genomes showed that the ortholog of Efl1 is absent. However, Efl1 is highly homologous to the eukaryotic elongation factor 2 (eEF-2) since it displays the basic organization of a translocation factor composed of a five-domain architecture including the G domain. Moreover, Efl1 can compete with eEF-2 for ribosome binding resulting in the inhibition of the eEF-2 ribosome-dependent GTPase activity (Graindorge et al., 2005). Conversely, SBDS protein is highly conserved in Archaea and Eukaryotes. In humans, mutations of the SBDS gene are associated with the Schwachman–Diamond syndrome (SDS, OMIM 260400), an autosomal recessive disorder. Genetic and biochemical data from different organisms and SDS patient-derived cells support the hypothesis that SBDS is a human ribosomopathy caused by the impaired release and recycling of eIF6 from late cytoplasmic pre-60S ribosomal subunits (Finch et al., 2011; Burwick et al., 2012). In Archaea, the SBDS orthologs are located in a super-operon that encodes proteins constituting the exosome complex and *in vitro* studies have suggested that archaeal SBDS might be involved in RNA metabolism (Koonin et al., 2001; Luz et al., 2010).

In this work, we analyzed the role of both aEF2 and aSBDS in the release of archaeal IF6 from the large ribosomal subunit. Our results suggest that, similarly to eukaryotes, the release of aIF6 from the 50S subunit is a GTPase-dependent mechanism. The involved GTPase is the elongation factor 2 (aEF-2) which is necessary and sufficient to promote aIF6 detachment from the 50S subunit. However, the system does not appear to depend on aSBDS which instead has an inhibitory effect on the detachment of aIF6. To structurally interpret our data, we performed a molecular modeling of the complex aEF2-aSBDS-50S. The results suggest that the binding sites of aEF-2 and aSBDS on 50S subunit overlap. This model would justify the inhibitory effect of aSBDS on aEF2 GTPase activity through a competitive binding mechanism.

MATERIALS AND METHODS

Cloning of the *S. solfataricus* aSBDS and aEF2 Genes and Isolation of the Recombinant Proteins Under Native Conditions

The aEF-2 gene was PCR-amplified from *S.so.* genomic DNA using two synthetic DNA primers constructed on the sequence of the corresponding gene (*SSO0728*). Primer sequences used for aEF2 cloning were as follows: forward primer aEF2-*Nco*I (5′-TTTTTCCATGGCTTGCCTAGATATAAGACAGTAGAGC-3′) and reverse primer aEF2- *Bam*HI (5′- TTTTTGGATCC TCACGACAAGAAATCTTCCACTTTTGG-3′). The amplification product was then digested with *Nco*I/*Bam*HI enzymes and inserted into the corresponding sites of the pETM11(+) expression plasmid to yield the recombinant pETM-aEF2 (6His) plasmid. The construct adds a tag of six histidine residues to the N-terminus of the recombinant protein. It was sequenced and used to transform *E. coli* strain BL21 (DE3), transformants were grown at 37°C in LB medium containing kanamycin (30 μg/ml). aEF2 expression was induced with 1 mM IPTG at a growth curve of OD_{600} = 0.5 for a further 4 h before harvesting. The cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and sonicated. After centrifugation, the cleared lysate underwent a first step of purification for aEF2 by incubating the whole cell lysate at 70°C for 15 min to precipitate mesophilic *E. coli* proteins. Recombinant aEF-2 was purified by affinity chromatography on Ni–NTA agarose (Qiagen) and eluted under native conditions. The elution fraction was precipitated adding $(NH_4)_2SO_4$ at 70% of saturation, dialyzed against storage buffer (30 mM NH_4Cl, 20 mM Tris/HCl, pH 8.0) and stored at −80°C in aliquots. The open reading frame of *SSO0737* gene coding aSBDS protein was amplified using forward (5′-TTTTTTTAT GCTAGCATGACGAAGGAGCGTGATTATG-3′) and reverse primer (5′-CATGGTATGCTCGAGTCATCTCACTTGCAATAC

TTTAAC-3′) containing *Nhe*I and *Xho*I restriction site, respectively. The amplification product was then digested with *Nhe*I/*Xho*I enzymes and inserted into the corresponding sites of the pRSETB expression plasmid (Novagen) to yield the recombinant pRSETB-aSBDS (6His) plasmid. The construct adds a tag of six histidine residues to the N-terminus of the recombinant protein. It was sequenced and used to transform *E. coli* strain BL21 (DE3). The procedure for its expression and purification was the same described above for aEF2 excepted for the use of ampicillin instead of kanamycin as selector of cells containing the plasmid with the PCR insert. The purified recombinant protein aSBDS was dialyzed against the storage buffer containing 20 mM TEA pH 7.4, 10 mM KCl, 5% glycerol, and preserved in aliquots at −80°C.

Preparation of *S. solfataricus* Cellular Extracts and Cellular Fractions

Whole cell extracts were prepared starting from frozen *Sulfolobus solfataricus* cell pellets following the procedure previously described (Benelli and Londei, 2007). Crude cellular lysates (S30) were centrifuged in a Beckman Ti 50 rotor at 100,000×*g* and 4°C for 3 h to separate ribosomes from a supernatant (S-100) containing total cellular tRNAs and ribosome free cytoplasmatic proteins. The pellet of ribosomes (termed "crude" ribosomes, CRs) was resuspended in the extraction buffer (20 mM Tris/HCl pH 7.4, 10 mM $Mg(OAc)_2$, 40 mM NH_4Cl, 1 mM DTT). The proteins of S-100 cell fraction were concentrated, adding ammonium sulfate to 70% saturation. The precipitate was collected by centrifuging 10 min at 15,000 rpm; the pellet was dissolved in the resuspending buffer (20 mM Tris/HCl pH 7.4, 2 mM $Mg(OAc)_2$, and 2 mM DTT) and dialyzed against the same buffer. Ribosomes devoid of extrinsic proteins and some translation factors were obtained, resuspending crude ribosome pellet in salt-buffer (20 mM Tris/HCl pH 7.4, 500 mM NH_4Cl, 10 mM $Mg(OAc)_2$, 1 mM DTT), and then loaded on 18% (w/v) sucrose cushion in the same buffer. Then, they were centrifuged in a Beckman Ti 50 rotor at 100,000×*g* for 4 h at 4°C. The final ribosome pellet (termed "high-salt washed" ribosomes, 70S HSW) was resuspended in the extraction buffer containing 3% glycerol. The concentration of the ribosomes was determined by measuring the A_{260} and considering 1 OD_{260} 70S = 40 pmol. The supernatant recovered after the sedimentation of HSW was supplemented with ammonium sulfate at a final concentration of 70% and stirred on ice for about 1 h. The precipitate was collected by centrifuging for 10 min at 15,000 rpm; the pellet was dissolved in the resuspending buffer (20 mM Tris/HCl pH 7.4, 2 mM DTT, 5% glycerol) and dialyzed against the same buffer. This preparation was the high salt wash (HSW).

Isolation of Ribosomal Subunits

Aliquots of the salt-wash ribosomes (40 A_{260} units) were layered onto preparative 38 ml linear 10−30% (w/v) sucrose density gradients made in the ribosome-suspending buffer (20 mM Tris/HCl, pH 7.0, 40 mM NH_4Cl, 10 mM $Mg(CH_3COO)_2$, 2.0 mM dithiothreitol). The gradients were centrifuged in a Beckman SW 27 rotor operated at 18,000 rev/min and 4°C for 18 h. Fractions corresponding to the 30S and 50S peaks of A_{260} were separately pooled and the particles therein were precipitated by the addition of two volumes of ethanol. After low-speed centrifugation, the subunit pellets were resuspended in the ribosome extraction buffer containing 10% (v/v) glycerol and stored at −20°C.

GTP Hydrolysis Assay

The amount of inorganic phosphate released after GTP hydrolysis was monitored by the use of ammonium molybdate in sulfuric acid solution. In these experimental conditions, phosphate reacts with ammonium molybdate to form a yellow phosphorous molybdate complex showing an absorption peak at 660 nm. Measurement of aEF2 GTPase activity was carried out at 65°C for 20 min in a final volume of 0.05 ml containing 20 mM Tris/HCl, pH 7.4, 10 mM KCl, and 10 mM $MgCl_2$. The amount of protein used in each reaction is described in the legend of the corresponding figure. After terminating the reaction, the volume was brought up to 0.3 ml with the reaction buffer. This was followed by the addition of 0.7 ml of a reagent containing one part of 10% ascorbic acid and six parts of 0.42% ammonium molybdate $\cdot 4H_2O$ (prepared in 1 N H_2SO_4). After thoroughly mixing, the content was incubated at 45°C for 20 min permitting the color development that was read at 660 nm.

In vitro Translation

In vitro translation was performed by programming a whole cell lysate prepared as described before (Benelli and Londei, 2007). The samples contained in a final volume of 100 μl: 10 mM KCl, 20 mM TEA/HCl (pH 7.4), 20 mM $MgCl_2$, 3 mM ATP, 1 mM GTP, 4 μg of *S. solfataricus* total tRNA, 0.55 mg of S30 extract, and 4 μg of *in vitro* transcribed 104 mRNA. The samples were incubated for 45 min at 70°C. At the end of the reaction, fixation on ice with 1% formaldehyde for 30 min was performed to stabilize 70S ribosomes which are easily dissociated in *S. solfataricus* and the samples were layered on linear, 10–30% sucrose gradients containing 10 mM KCl, 20 mM TEA/HCl pH 7.4, and 20 mM $MgCl_2$. The gradients were centrifuged at 36,000 rpm for 4 h and 30 min in a Beckman SW41 rotor at 4°C and 36,000 rpm for 4 h and unloaded while monitoring absorbance at 260 nm.

Sucrose Gradient Analysis

The association of recombinant and/or endogenous proteins to ribosomal subunits was investigated by fractionating different samples on sucrose density gradient and then probing each fraction for the presence of the proteins by western blot with specific antibodies. Specifically, at the end of each reaction, the samples were layered on linear 10–30% sucrose gradients containing 10 mM KCl, 20 mM TEA-HCl (pH 7.5), and 20 mM $MgCl_2$; these were centrifuged in a Beckman SW41 rotor at 4°C and 36,000 rpm for 4 h or at 18,000 rpm for 17–18 h. After centrifugation, the gradients were unloaded while monitoring absorbance at 254 nm with the EM-1 Econo UV absorbance instrument (Bio-Rad). The individual fractions (0.5 ml) were collected in single tubes and precipitated adding 1/100 volume of 2% Na-deoxycholate and 1/10 of trichloroacetic acid 100%,

vortexed, and let sit over-night at 4°C. Then, the samples were centrifuged 15′ at 13,000×*g*, the protein pellets were resuspended in 20–40 μl of 1X Laemmli Sample Buffer, separated by 15% SDS–PAGE, and electroblotted to nitrocellulose membrane. On the basis of the protein analyzed, we probed the membrane with house made rabbit polyclonal antibodies (antibody against aSBDS and aIF6) or a 6x-His Tag monoclonal antibody (Thermo Fisher Scientific).

Western Blot Analysis

The protein concentration of different cell fractions was quantified using the Bradford assay. Equal amounts of protein samples were subjected to SDS-PAGE and transferred to nitrocellulose membrane (Amersham Protran-GE Healthcare, Little Chalfont, Buckinghamshire, United Kingdom). After blocking non-specific binding of antibody with 5% non-fat milk, blots were probed with one of the following antibodies: anti-aIF6 polyclonal rabbit antibodies (1:5,000), anti-aSBDS polyclonal rabbit antibodies (1:10,000), 6×-His Tag Monoclonal Antibody (4E3D10H2/E3; Thermo Fisher Scientific). Primary antibodies were detected by binding horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG-HRP (sc-2004; Santa Cruz Biotechnology), goat anti-mouse IgG-HRP (sc-2005; Santa Cruz Biotechnology), and using an enhanced chemiluminescent visualization system (ECL Western Blotting Substrate, Thermo Fisher Scientific-Pierce Biotechnology, Rockford, IL, United States). 6×-His Tag Monoclonal Antibody and secondary antibodies were diluted according to the manufacturer instructions. The images were captured by a BioRad ChemiDoc. MP Imaging system (Bio-Rad, Hercules, California, United States).

Protein Structure Analysis, Modeling, and Docking

The Combinatorial Extension (Shindyalov and Bourne, 1998) and PyMOL (Schrodinger, 2013) tools were used for structure superposition and visualization, respectively. Modeler v.9.9 (Sali and Blundell, 1993) and its graphical interface PyMod (Bramucci et al., 2012; Janson et al., 2017; Janson and Paiardini, 2020) were used for homology modeling purposes; models were validated using standard tools (Laskowski et al., 1996; Wiederstein and Sippl, 2007). The Phyre v2.0 server was used for finding candidate templates for homology modeling (Bennett-Lovsey et al., 2008). Prediction of the potential presence of protein-protein interaction sites was carried out with the consensus method implemented in meta-PPISP at the web site http://pipe.scs.fsu.edu/meta-ppisp (Qin and Zhou, 2007). Protein–protein docking was carried out starting from the original position of the homologous protein complexes and refined using the ClusPro method available at the server http://cluspro.bu.edu (Kozakov et al., 2010).

Size-Exclusion Chromatography (SEC)

Allyl dextran-based size-exclusion gel (Sephacryl S-300, GE Healthcare) was used as stationary phase. The gel column was prepared by filling a 15 cm long column with an appropriate amount of allyl dextran-based size-exclusion gel dilute 1:1 with the following solution buffer: 10 mM KCl, 20 mM TEA-HCl (pH 7.5), and 20 mM $MgCl_2$. The flow rate of the running buffer was 1 ml/min and the presence of molecules along the flow was monitored by reading the absorbance at 254 nm with the EM-1 Econo UV absorbance instrument (Bio-Rad). The speed of the recording pare was set to 1 cm/min.

Statistical Analysis

All data shown represent at least three independent experiments. Western blot bands intensities were captured and analyzed by a ChemiDoc MP Imaging system (Bio-Rad, Hercules, California, United States). Values represent the mean ± SEM. *P*-values listed represent a two-tailed Student's *t*-test *P*-value. $P < 0.05$ was considered statistically significant.

RESULTS

aIF6 Is Released From the 50S Subunits Through a Ribosome-Dependent GTPase Activity

The well-known role of the a/eIF6 protein as a ribosome anti-association factor leads to the assumption that the factor has to be released from the large ribosomal subunits to permit their access to the elongation cycle. Indeed, in our previous work we showed that lysates programmed for protein synthesis triggered the dissociation of aIF6 from the 50S subunits (Benelli et al., 2009). To elucidate the mechanism inducing aIF6 release we focused our attention on a simplified system consisting of just whole ribosomes. Specifically, we used one of the following fractions: (1) crude 70S, i.e., ribosomes obtained by high-speed centrifugation of whole cell lysates; (2) high salt purified ribosomes (70S HSW), i.e., purified ribosomes washed with a high salt buffer and devoid of most translation factors; (3) purified 50S subunits.

Initially, we performed *in vitro* studies incubating crude ribosomes in presence of GTP at 65°C for 15 min. We observed that under these conditions a substantial fraction of bound aIF6 was released (**Figure 1A**, 1st panel). This showed that ongoing translation is not required for aIF6 detachment. However, when the experiment was repeated using HSW 70S instead of crude ribosomes, aIF6 was not released, suggesting that the high-salt washing of ribosomes removed some factor essential for aIF6 detachment. Indeed, when the proteins removed by washing (HSW) were added back to the reaction mix, aIF6 release was again observed (**Figure 1A**, 2nd panel). Significantly, in all of the previous experiments, substituting GTP with GMP-PNP (a non-hydrolyzable analog of GTP) blocked aIF6 release, demonstrating that it was dependent on the hydrolysis of GTP. Hence, these preliminary results suggested that, similarly to the eukaryotes, some GTPase was implicated in removing aIF6 from the 50S subunits. Indeed, the GTPase assays shown in **Figure 1B** indicate that the crude ribosome fraction has a high GTPase activity which is lost upon high salt washing. Addition of HSW proteins to the washed 70S restored their GTPase activity to levels comparable to those of crude 70S. Overall, these experiments further support

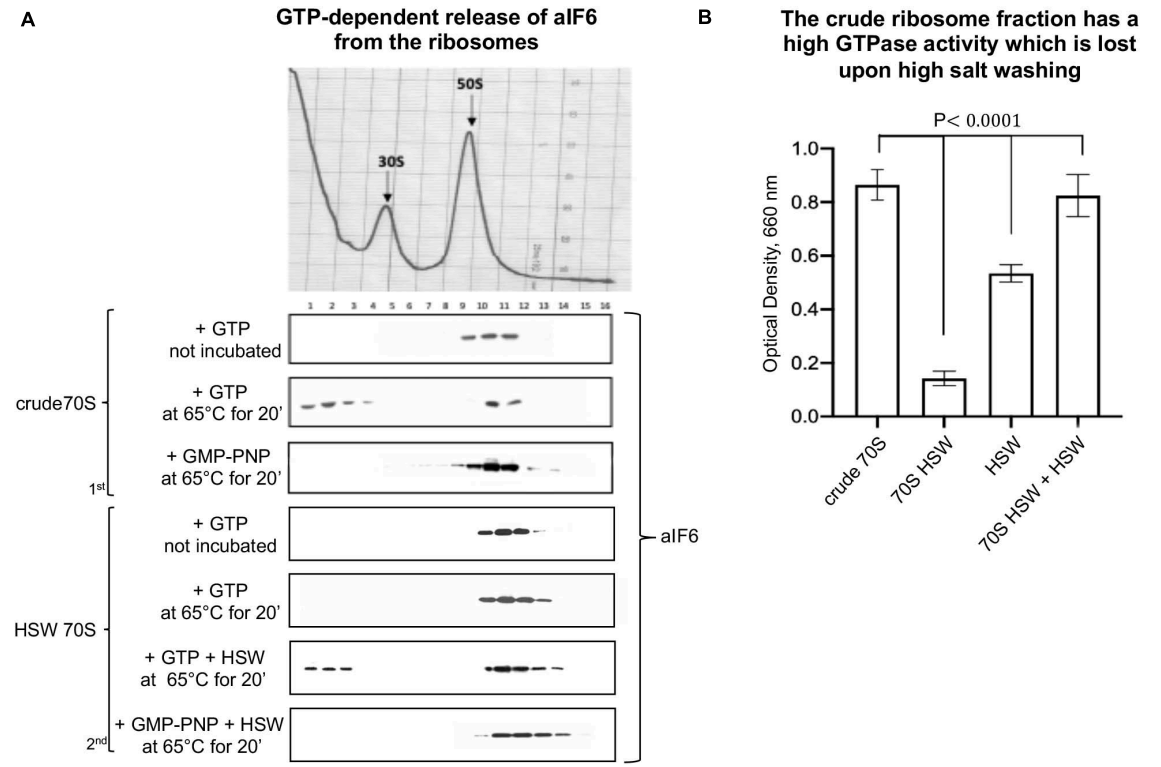


FIGURE 1 | The GTPase activity of extra-ribosomal 70S fraction induces the release of aIF6 from the 50S subunits. **(A)** Density gradient fractionation of crude 70S or HSW 70S in a mixture containing GTP or GMP-PNP and incubated at 65°C for 20 min. The curly brackets group experiments made with the same ribosome preparation. The distribution of aIF6, shown at the bottom of each gradient profile, was revealed by western blotting of the individual fractions with the anti-aIF6 antibodies. The distribution of ribosomal subunits was identified by the optical scans at OD254 nm of the gradients. **(B)** GTPase activity of different cell fractions was determined reading at 660 nm the amount of the phosphate/molybdate complex formed after GTP hydrolysis as described in the "Materials and Methods" section. Data are presented as mean value ± *SD* (*n* = 4). A representative image of at least three independent sucrose density experiments is shown for each analysis.

the idea that the detachment of aIF6 from 50S subunits requires the action of some critical GTPases loosely associated with the crude 70S ribosomes.

Ribosome-Dependent GTPase Activity of aEF-2 Induces the Release of aIF6

As said before, archaea do not possess homologs of the specialized GTPase Efl1. However, Efl1 is a close homolog of elongation factor 2 (EF-2), which raised the possibility that, in archaea, EF-2 itself could be the GTPase protein implicated in aIF6 detachment.

To verify this surmise, we decided to clone the *Sulfolobus solfataricus* gene *SSO0728* encoding the aEF-2 protein into an expression vector (pETM11+) adding a 6(His)-tag to the N-terminus of the recombinant protein (**Supplementary Figure S1A**). Upon expression in *E. coli,* the construct produced a recombinant aEF-2 protein devoid of the unique post translational modification specific of eukaryotic and most archaeal translational elongation factor 2 and known as diphthamide (Schaffrath et al., 2014; Narrowe et al., 2018). Therefore, we preliminarily verified whether our recombinant construct possessed a ribosome-dependent GTPase activity. The experiments in **Figure 2A** show that this was indeed the case, in accordance with previous evidence (de Vendittis et al., 1997). Successively, we analyzed the involvement of aEF-2 in aIF6 detachment from the 50S subunit incubating the HSW 70S in the presence of the recombinant protein at 65°C for 20 min. As shown in **Figure 2B**, under these conditions, aEF-2 was able to promote the release of aIF6; this ability was dependent on the hydrolysis of GTP, since the presence of GMP-PNP inhibited the reaction. These results were also reproduced using size-exclusion chromatography instead of density-gradient centrifugation (**Supplementary Figure S2**). Finally, to determine whether the presence of the 30S subunit was required for the aEF-2-induced aIF6 release, we performed the same experiments also using gradient-purified 50S subunits. As shown in the last lane of **Figure 2B**, aEF-2 was able to induce the release of aIF6 also in this case, suggesting that aIF6 detachment takes place on individual 50S ribosomal subunits that have not yet entered the translation cycle.

Localization of Archaeal SBDS in *S. solfataricus* Cell Extracts

The experiments described above establish the importance of aEF-2 in removing aIF6 from the 50S ribosomal subunit, thereby enabling the particles to enter the elongation cycle. However, they do not elucidate whether the aSBDS protein retains a

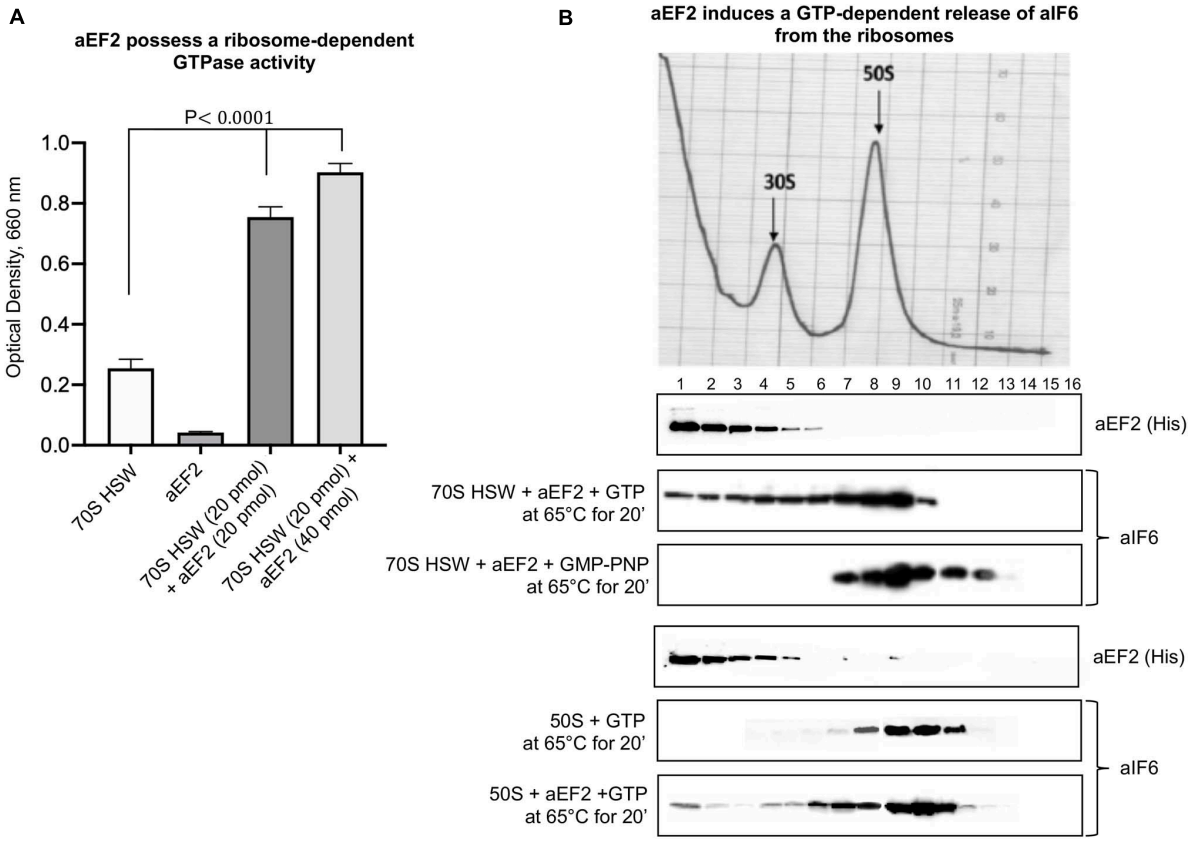


FIGURE 2 | aEF-2-induced release of aIF6 from ribosomes. **(A)** The GTPase activity of recombinant aEF-2 protein was analyzed by incubating 20 or 40 pmol of the protein with 20 pmol of 70S HSW and 1 mM GTP at 65°C for 20 min. At the end of the reaction, the inorganic phosphate released after GTP hydrolysis was revealed as described in the "Materials and Methods" section. Data are presented as mean value ± *SD* (*n* = 4). **(B)** Density gradient fractionation of 70S HSW (70 pmol) or 50S (50 pmol) incubated at 65°C for 20 min in the presence of aEF-2 (70 or 50 pmol) and 1 mM GTP or GMP-PNP. The distribution of aIF6 and recombinant aEF2 was revealed by western blotting of the individual fractions with anti-aIF6 and 6(His) antibodies, respectively. The distribution of the ribosomal subunits was identified by the optical scans at OD254 nm of the gradients. A representative image of at least three independent sucrose density experiments is shown for each analysis.

conserved evolutionary function, namely if it cooperates with aEF-2 in promoting the release of aIF6 from the 50S subunit. To investigate this point, we cloned the *S. solfataricus* gene *SSO0737* by PCR amplification on genomic DNA, inserted the amplified fragment in the expression plasmid pRSETB, expressed the plasmid in *E. coli* BL21 (DE3), and purified the recombinant protein from cell extracts by differential thermal denaturation and affinity chromatography. This procedure yielded a recombinant aSBDS protein (aSBDS$_r$) containing a 6xHis tag to its N-terminus that migrated as a single sharp band free of detectable contaminants (**Supplementary Figure S3A**). The purified protein was used to produce polyclonal antibodies to monitor the cellular distribution of the endogenous protein. When tested on both whole cell lysates and ribosome preparations, the aSBDS antiserum recognized a single polypeptide, which was abundant in the crude 70S but reduced in the HSW ribosomes (**Figure 3A**).

Translational Behavior of aSBDS

To investigate the behavior and localization of aSBDS during translation, sucrose density gradient analysis was performed on lysates programmed for protein synthesis as described earlier (Benelli and Londei, 2007). The programmed lysates were incubated at 70°C for 45 min to activate translation and were then fixed with formaldehyde to stabilize 70S ribosomes which are easily dissociated in *S. solfataricus*. As shown in **Figure 3B**, aSBDS was very abundant and widespread along the gradient, with stronger signals in the low-molecular weight fractions and in the fractions corresponding to the 50S peak. Some signal was also present in high-molecular weight fractions, similar to what was observed in yeast by other authors (Menne et al., 2007). A similar pattern was obtained upon gradient fractionation of crude 70S ribosomes (**Figure 3C**, 1st panel), while HSW 70S, which contain reduced amounts of aSBDS, yielded a more discrete localization of SBDS at the level of 50S subunits and higher fractions (**Figure 3C**, 2nd panel). In particular, the peak of SBDS observed in post-50S fractions may be due to the presence of the protein in high-mol-wt complexes formed with some other component present in the ribosome preparations. Artifacts due to precipitation and aggregation of SBDS were ruled out since the same reaction mixture devoid of ribosomes produced a signal of

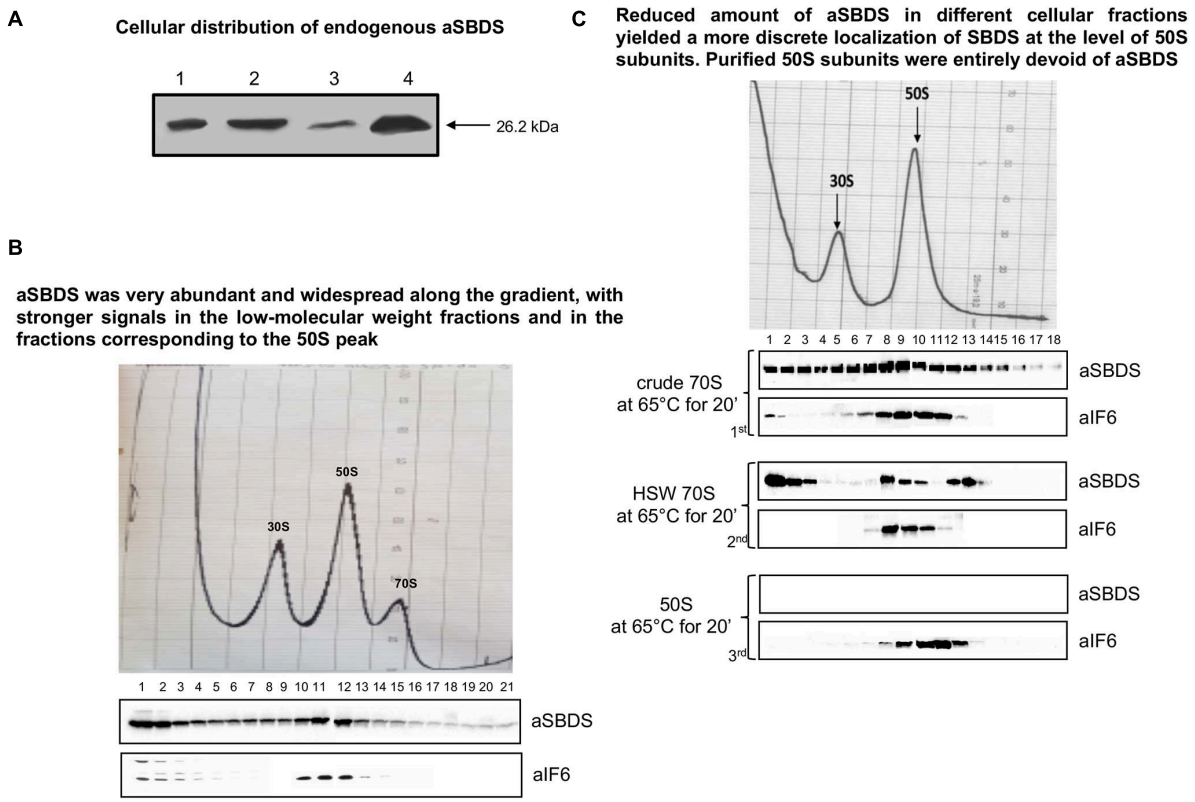


FIGURE 3 | Localization of endogenous aSBDS in a cell lysate. **(A)** Identification of endogenous aSBDS in different cell fractions by western blot: (1) cell extract; (2) crude ribosomes; (3) high-salt washed ribosomes; and (4) post-ribosomal supernatant (HSW). **(B)** Density gradient fractionation, after fixation with HCHO, of cell lysates programmed for translation and incubated at 70°C for 45 min. **(C)** Density-gradient fractionation of: crude 70S (70 pmol) in the 1st panel, HSW 70S (70 pmol) in the 2nd panel, and 50S subunits (50 pmol) in the 3rd panel. Each sample was incubated at 65°C for 20 min in the presence of 1 mM GTP. Braces group experiments made with the same ribosome preparation. The distribution of endogenous aIF6 and aSBDS shown at the bottom of each gradient profile was revealed by western blotting of the individual fractions with the anti-aIF6 and anti-aSBDS antibodies, respectively. In **(B,C)** the distribution of ribosomal subunits was identified by the optical scans at OD254 nm of the gradients. A representative image of at least three independent sucrose density experiments is shown for each analysis.

the recombinant aSBDS protein just in the first fractions (**Supplementary Figure S3B**). Gradient-purified 50S subunits were entirely devoid of aSBDS (**Figure 3C**, 3rd panel), demonstrating that the protein is not strongly associated with the ribosomes.

aSBDS Inhibits the GTPase Activity of aEF-2 and the Release of aIF6 From the Ribosomes

The role, if any, of aSBDS in the release of aIF6 from the large ribosomal subunit was directly investigated by adding the purified protein to a reaction mixture containing 70S HSW and aEF-2. Surprisingly, the presence of aSBDS effectively inhibited the aIF6 release from the ribosomes (**Figure 4A**, 1st lane). Similar results were also obtained when purified 50S subunits were used (**Figure 4A**, 4th lane). To get a better insight into this result, we repeated the experiments by adding aSBDS and aEF-2 at different times to the reaction mixture containing 70S HSW. As shown in **Figure 4A**, addition of SBDS 10 min after the start of the reaction with aEF-2 allowed a limited release of aIF6, while when SBDS was added at the outset and aEF-2 10 min later, aIF6 detachment was completely blocked. Furthermore, GTPase assays showed that aSBDS substantially inhibited the ribosome-dependent GTPase activity of aEF2 (**Figure 4B**). Upon the whole, the results suggested that aEF-2 and SBDS competed for a same ribosome-binding site, and that only ribosomes devoid of aSBDS were competent for aEF-2-induced aIF6 release.

DISCUSSION

In this work, the mechanism of release of the translation factor aIF6 from the large ribosomal subunit has been experimentally studied for the first time. Although a final mechanism has not been defined and will require further work, the results obtained have unveiled interesting homologies and differences with the corresponding eukaryotic process. Firstly, we could conclude that aIF6 release from archaeal large ribosomal subunit, similar to eukaryotes, is a GTPase-dependent event. The involved GTPase is the elongation factor 2 (aEF-2) which by itself is necessary and sufficient to induce aIF6 detachment from the ribosomes, even in the absence of ongoing translation. Indeed,

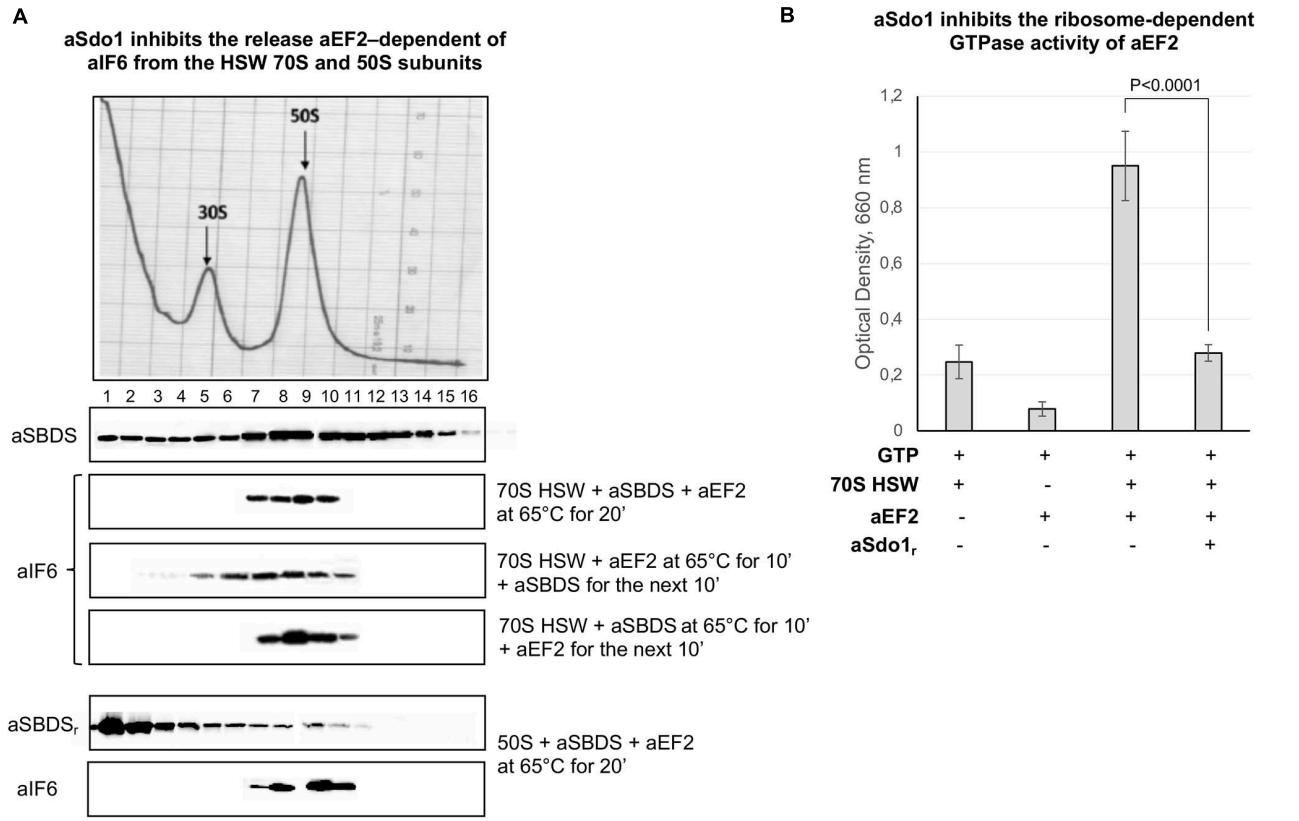


FIGURE 4 | aSBDS inhibits the release of aIF6 induced by aEF2. **(A)** Density gradient fractionation of HSW ribosomes (70 pmol) or 50S subunits (50 pmol) incubated with aSBDS (140 and 100 pmol, respectively), and aEF-2 (140 and 100 pmol, respectively), at 65°C for 20 min. Each sample was incubated in presence of 1 mM GTP. The distribution of aIF6 was revealed by western blotting of the individual fractions with the anti-aIF6 antibodies. The distribution of endogenous and recombinant aSBDS was revealed by western blotting of the individual fractions with the anti-aSBDS antibodies. The distribution of ribosomal subunits was identified by the optical scans at OD254 nm of the gradients. A representative image of at least three independent sucrose density experiments is shown for each analysis. **(B)** GTPase activity of aEF-2 (40 pmol) was analyzed by incubating the recombinant protein in presence/absence of 70S HSW (20 pmol), aSBDS (40 pmol), and 1 mM GTP at 65°C for 20 min. At the end of the reaction, the inorganic phosphate released after GTP hydrolysis was detected, as described in the "Materials and Methods" section. Data are presented as mean value ± *SD* (*n* = 4).

we observed the release of aIF6 from the 50S subunits in a reaction mixture containing just high-salt washed 70S, aEF-2, and GTP, without the other components necessary for translation such as tRNAs, mRNA, and translation factors. Since Archaea do not possess a homolog of the GTPase Efl1 involved in the eIF6 release in eukaryotes, a role of aEF-2 in the process had already been suggested both on the basis of the fact that Efl1 is a close homolog of aEF2, and because in eukaryotes Efl1 inhibits the GTPase activity of EF-2, probably because they compete for the same ribosome-binding site (Graindorge et al., 2005; Tanzawa et al., 2018). Indeed, we found that the reaction relied on the GTPase activity of the factor since the presence of GMP-PNP instead of GTP in the reaction inhibited the detachment of aIF6 from the ribosomes. However, release of aIF6 in Archaea does not appear to require the eukaryotic SBDS homolog. Instead, aSBDS seems to have an inhibitory effect on aIF6 detachment, probably because its ribosomal binding site overlaps with that of aEF-2 and the two factors compete for binding.

In order to get a structural rationale of the results, we decided to model aIF6 (Uniprot ID: Q980G0) from *S. solfataricus*, based on the very high sequence identity with the homologous structure from *Methanocaldococcus jannaschii* (PDB: 1G61; Sequence identity: 47%), and to model also aSBDS (Uniprot ID: D0KTE1), based on the homologous from *Archaeoglobus fulgidus* (PDB: 1P9Q, sequence identity: 44%) (Savchenko et al., 2005; **Figure 5**). Moreover, the positions of aIF6 and aSBDS relative to the ribosomal subunit were obtained by superposing the predicted models with the homologous structures of the 60S ribosomal subunit from *Dictyostelium discoideum* (PDB 5ANB), and the 50S ribosomal subunits of *T. kodakarensis* (PDB 6SKG) (Sas-Chen et al., 2020) and *T. thermautotrophicus* (PDB 4ADX). Two loops of aSBDS (residues 170–175; 193–198) are mainly contacting in the model two regions of aIF6 (186–190; 206–210) suggesting that aSBDS could stabilize aIF6 in its interaction with the ribosome (**Figure 5**, upper right panel). On the other hand, modeling of aEF-2 (Uniprot ID: P30925) using as structural template the crystal structure of

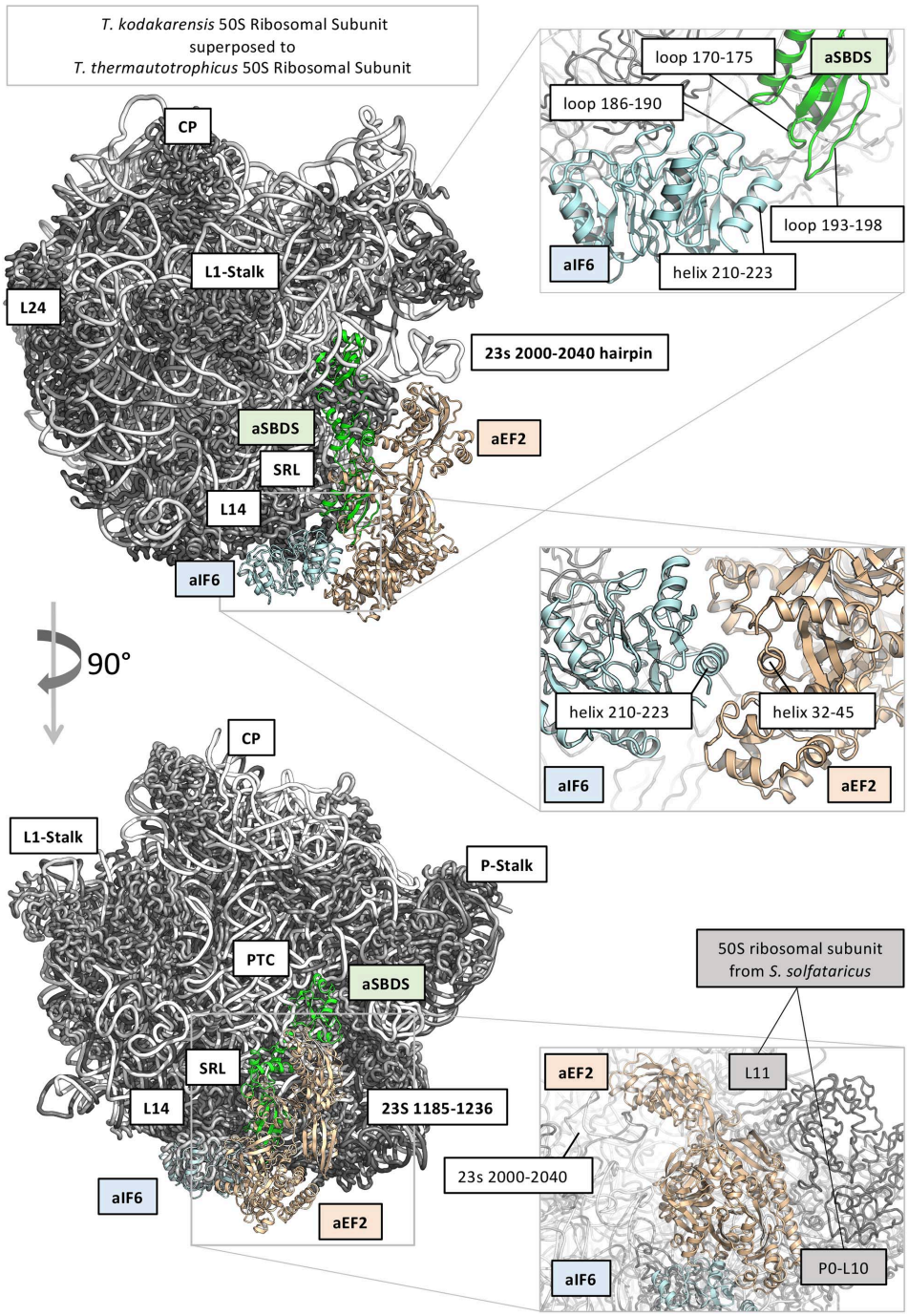


FIGURE 5 | Structural models of translation initiation factor 6 (aIF6), elongation factor 2 (aEF2), and ribosome maturation protein SBDS-like (aSBDS) from *Sulfolobus solfataricus* (cyan, gold, and green cartoons, respectively). The 50S ribosomal subunits from *T. kodakarensis* (PDB 6SKG) (Sas-Chen et al., 2020) and *T. thermautotrophicus* (PDB 4ADX) are shown as the reference in gray (protein) and white (rRNA) ribbons. The approximate position of aIF6 and aEF2 related to the 60S ribosomal subunit is based on the homologous structures from *Dictyostelium discoideum* (PDB 5ANB). The relative position of aEF2 and aSBDS suggests that the two proteins partially overlap and compete for the same binding site on the 50S subunit.

the homologous protein from *Pyrococcus horikoshii* (PDB: 5H7J, sequence identity: 50%) (Tanzawa et al., 2018), and its relative position on the ribosomal subunits as previously described, evidenced that aEF-2 is substantially smaller than eEF-2, and lacks an important domain region of eEF2 (PDB: 5ANB), namely 541–821, which is involved in binding and stabilizing SBDS in the eukaryotic complex. aEF-2 is instead stabilized by interactions with the archaeal proteins L10 and L11 (**Figure 5**, lower right

panel), and the archaeal 23s 2,000–2,040 hairpin. A partial overlap and competition are observed between aEF-2 and aSBDS in binding to the ribosome (**Figure 5**). Our modeling suggests that a tight interaction takes place between aIF6 and aEF-2, as previously observed (**Figure 5**, central panel). Therefore, it is conceivable that this interaction could be retained also after the conformational transition of aEF-2, upon GTP binding and hydrolysis. The overall effect of such conformational transition of aEF-2 would therefore be the displacement of aIF6 from its bound position on the ribosome.

To verify this, aEF-2 was modeled in its open conformation (based on the crystal structure of *Methanoperedens nitroreducens* EF2, PDB code: 6U45) (Fenwick and Ealick, 2020) and morphed between its open/closed states. Indeed, the model predicts that upon conformational transition of aEF2, aIF6 is displaced from its previous position.

SBDS has also been shown to share in part the same binding site with the GTPase Efl1 in eukaryotes: however, in the eukaryotic system, the arrival of Efl1 causes a conformational change of SBDS that is in turn required for the ejection of eIF6. In this view, eukaryotic SBDS functions as a cofactor for elongation factor-like GTPase 1 (Efl1). This does not seem to be the case in Archaea where, probably, the aEF-2-dependent detachment of aIF6 has to be preceded by the release of aSBDS from the ribosomes. In **Figure 6**, we present a model based on the previous results, which proposes a plausible explanation of the interplay among the translation factors in question.

In synthesis, aSBDS and aEF-2 would be two proteins that orchestrate, and participate in a distinct temporal manner to, the formation of a functional 50S. Specifically, aSBDS could be a protein belonging to the class of trans-acting factors known as "placeholders" which temporarily bind selected ribosomal sites until these have achieved a structure appropriate for exchanging the placeholder with another site-specific binding factor (Fenwick and Ealick, 2020). In the present case, the other factor would be aEF-2, whose action as a remover of aIF6 would be hindered by aSBDS until the biogenesis of the particle is completed.

However, the role of archaeal SBDS in the context of ribosome biogenesis or of translation is far from being clear and will require further experimentation to be fully elucidated. A certain amount of evidence would lead to speculate that aSBDS could be a part of the exosome system involved in the maturation of rRNA during ribosome biogenesis. First, in archaea, the aSBDS gene is located in a super-operon that encodes proteins constituting the exosome complex (**Supplementary Figure S4**). Second, *in vitro* studies have suggested that archaeal SBDS might be involved in RNA metabolism affecting RNA-exosome activity (Luz et al., 2010). Third, our present results show that aSBDS is very abundant in *Sulfolobus* cells and that it is widely distributed on density-gradient fractions, apparently being also included in high-mol-wt complexes of unknown composition. Indeed, there are data in literature showing that some of *Sulfolobus solfataricus* exosome components, in particular Rrp41, show a sedimentation pattern not unlike what we observed for aSBDS (Warren, 2018).

Upon the whole, the previous considerations could lead to conceive a tentative scenario, where the ancestral function of SBDS (retained in present-day Archaea) was to participate in the maturation of pre-rRNA on the large ribosomal subunit. SBDS would remain on the ribosome until this task was completed, also obstructing the binding site for ribosome-dependent GTPases such as aEF-2, thus preventing the premature release of the anti-association factor aIF6. In a later evolutionary stage, in the eukaryotic lineage, both SBDS and the RNA-exosome complex retained a pivotal role for the maturation of the pre-ribosomes but in two distinct temporal steps. A specific GTPase dedicated to eIF6 detachment also emerged (Menne et al., 2007; Witharana et al., 2012; Espinar-Marchena et al., 2017). The mechanism triggering the detachment of aSBDS from the archaeal ribosome remains to be understood; conceivably, it could be a conformational change induced by an unknown GTPase that accompanies the final maturation of the large subunit. Future research will hopefully shed light on this very interesting process.

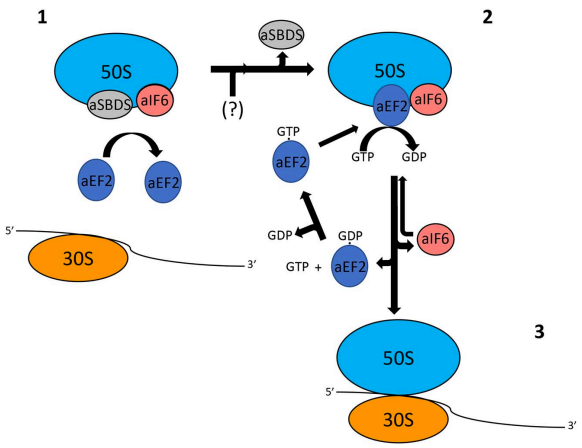


FIGURE 6 | A plausible model depicting the mode of action of aEF-2 on the 50S subunits for the release of aIF6 in *Sulfolobus solfataricus*. The picture represents 50S subunits with the aIF6 and aSBDS proteins bound on it. In the first step, the presence of aSBDS on the ribosomes does not permit the binding of aEF-2 (1). The activity of an unknown factor (here represented with the symbol "?") induces the release of aSBDS from the ribosomes facilitating the binding of aEF-2 to the 50S subunits (2). The hydrolysis of GTP bound to aEF-2 induces a conformational change in the ribosome and/or in the structure of aEF-2 itself with the consequent release of aIF6. aEF2 bound to GDP dissociates from the ribosomes and the next exchange of GDP/GTP on aEF-2 allows the recruitment of the protein to a new cycle of aIF6 release. Similarly, free aIF6 is ready to bind newly to the large ribosomal subunits while 50S subunits, free of aIF6, can instead complete the translation initiation phase (3).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

DB, PL, and AL conceived and designed the experiments. DB, GL, MD, and AR performed the experiments. DB, PL, AL, and AP analyzed the data. DB, PL, and AP wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants to PL from the Istituto Pasteur-Fondazione Cenci Bolognetti project "Detecting and characterizing specialized ribosomes translating specific classes of mRNAs in Archaea" and by grants to PL from "La Sapienza" University of Rome for the 2017 project "Role of the translation factor eIF6 in ribosome biogenesis and differential translation of specific mRNAs" (RM11715C7CE47750).

ACKNOWLEDGMENTS

The pETM-11 plasmid was kindly provided by Dr. Roberto Spurio (University of Camerino, Italy).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.631297/full#supplementary-material

REFERENCES

Basu, U., Si, K., Warner, J. R., and Maitra, U. (2001). The *Saccharomyces cerevisiae* TIF6 gene encoding translation initiation factor 6 is required for 60S ribosomal subunit biogenesis. *Mol. Cell. Biol.* 21, 1453–1462. doi: 10.1128/MCB.21.5.1453-1462.2001

Bécam, A. M., Nasr, F., Racki, W. J., Zagulski, M., and Herbert, C. J. (2001). Ria1p (Ynl163c), a protein similar to elongation factors 2, is involved in the biogenesis of the 60S subunit of the ribosome in Saccharomyces cerevisiae. *Mol. Genet. Genomics* 266, 454–462. doi: 10.1007/s004380100548

Benelli, D., and Londei, P. (2007). In vitro studies of archaeal translational initiation. *Methods Enzymol.* 430, 79–109. doi: 10.1016/S0076-6879(07)30005-0

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., Marzi, S., Mancone, C., Alonzi, T., la Teana, A., and Londei, P. (2009). Function and ribosomal localization of aIF6, a translational regulator shared by archaea and eukarya. *Nucleic Acids Res.* 37, 256–267. doi: 10.1093/nar/gkn959

Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J., and Kelley, L. A. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* 70, 611–625.

Bramucci, E., Paiardini, A., Bossa, F., and Pascarella, S. (2012). PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. *BMC Bioinformatics* 13:S2. doi: 10.1186/1471-2105-13-S4-S2

Burwick, N., Coats, S. A., Nakamura, T., and Shimamura, A. (2012). Impaired ribosomal subunit association in Shwachman-Diamond syndrome. *Blood* 120, 5143–5152. doi: 10.1182/blood-2012-04-420166

Ceci, M., Gaviraghi, C., Gorrini, C., Sala, L. A., Offenhäuser, N., Marchisio, P. C., et al. (2003). Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature* 426, 579–584. doi: 10.1038/nature02160

de Vendittis, E., Amatruda, M. R., Raimo, G., and Bocchini, V. (1997). Heterologous expression in *Escherichia coli* of the gene encoding an archaeal thermoacidophilic elongation factor 2. Properties of the recombinant protein. *Biochimie* 79, 303–308. doi: 10.1016/s0300-9084(97)83518-3

Dennis, P. P. (1997). Ancient ciphers: translation in Archaea. *Cell* 89, 1007–1010. doi: 10.1016/s0092-8674(00)80288-3

Espinar-Marchena, F. J., Babiano, R., and Cruz, J. (2017). Placeholder factors in ribosome biogenesis: please, pave my way. *Microb. Cell* 4, 144–168. doi: 10.15698/mic2017.05.572

Fenwick, M. K., and Ealick, S. E. (2020). Structural basis of elongation factor 2 switching. *Curr. Res. Struct. Biol.* 2, 25–34.

Finch, A. J., Hilcenko, C., Basse, N., Drynan, L. F., Goyenechea, B., Menne, T. F., et al. (2011). Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes Shwachman-Diamond syndrome. *Genes Dev.* 25, 917–929. doi: 10.1101/gad.623011

Gäbel, K., Schmitt, J., Schulz, S., Näther, 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

Gartmann, M., Blau, M., Armache, J. P., Mielke, T., Topf, M., and Beckmann, R. (2010). Mechanism of eIF6-mediated inhibition of ribosomal subunit joining. *J. Biol. Chem.* 285, 14848–14851. doi: 10.1074/jbc.C109.096057

Graindorge, J. S., Rousselle, J. C., Senger, B., Lenormand, P., Namane, A., Lacroute, F., et al. (2005). Deletion of EFL1 results in heterogeneity of the 60 S GTPase-associated rRNA conformation. *J. Mol. Biol.* 352, 355–369. doi: 10.1016/j.jmb.2005.07.037

Greber, B. J., Boehringer, D., Godinic-Mikulcic, V., Crnkovic, A., Ibba, M., Weygand-Durasevic, I., et al. (2012). Cryo-EM structure of the archaeal 50S ribosomal subunit in complex with initiation factor 6 and implications for ribosome evolution. *J. Mol. Biol.* 418, 145–160. doi: 10.1016/j.jmb.2012.01.018

Groft, C. M., Beckmann, R., Sali, A., and Burley, S. K. (2000). Crystal structures of ribosome anti-association factor IF6. *Nat. Struct. Biol.* 7, 1156–1164. doi: 10.1038/82017

Janson, G., and Paiardini, A. (2020). PyMod 3: a complete suite for structural bioinformatics in PyMOL. *Bioinformatics* 3:btaa849. doi: 10.1093/bioinformatics/btaa849

Janson, G., Zhang, C., Prado, M. G., and Paiardini, A. (2017). PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics* 33, 444–446. doi: 10.1093/bioinformatics/btw638

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011). Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* 334, 941–948. doi: 10.1126/science.1211204

Koonin, E. V., Wolf, Y. I., and Aravind, L. (2001). Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* 11, 240–252. doi: 10.1101/gr.162001

Kozakov, D., Hall, D. R., Beglov, D., Brenke, R., Comeau, S. R., Shen, Y., et al. (2010). Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. *Proteins* 78, 3124–3130.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486.

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

Luz, J. S., Ramos, C. R., Santos, M. C., Coltri, P. P., Palhano, F. L., Foguel, D., et al. (2010). Identification of archaeal proteins that affect the exosome function in vitro. *BMC Biochem.* 11:22. doi: 10.1186/1471-2091-11-22

Menne, T. F., Goyenechea, B., Sánchez-Puig, N., Wong, C. C., Tonkin, L. M., Ancliff, P. J., et al. (2007). The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nat. Genet.* 39, 486–495. doi: 10.1038/ng1994

Narrowe, A. B., Spang, A., Stairs, C. W., Caceres, E. F., Baker, B. J., Miller, C. S., et al. (2018). Complex evolutionary history of translation elongation factor 2 and diphthamide biosynthesis in archaea and parabasalids. *Genome Biol. Evol.* 10, 2380–2393. doi: 10.1093/gbe/evy154

Qin, S., and Zhou, H. X. (2007). meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics* 23, 3386–3387.

Russell, D. W., and Spremulli, L. L. (1979). Purification and characterization of a ribosome dissociation factor (eukaryotic initiation factor 6) from wheat germ. *J. Biol. Chem.* 254, 8796–8800.

Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.

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

Savchenko, A., Krogan, N., Cort, J. R., Evdokimova, E., Lew, J. M., Yee, A. A., et al. (2005). The Shwachman-Bodian-Diamond syndrome protein family is involved in RNA metabolism. *J. Biol. Chem.* 280, 19213–19220. doi: 10.1074/jbc.M414421200

Schaffrath, R., Abdel-Fattah, W., Klassen, R., and Stark, M. J. (2014). The diphthamide modification pathway from Saccharomyces cerevisiae–revisited. *Mol. Microbiol.* 94, 1213–1226. doi: 10.1111/mmi.12845

Schrodinger, L. L. C. (2013). *The PyMOL Molecular Graphics System, Version 1.6.0.0*. Portland, OR.

Shindyalov, I. N., and Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 11, 739–747. doi: 10.1093/protein/11.9.739

Tanzawa, T., Kato, K., Girodat, D., Ose, T., Kumakura, Y., Wieden, H. J., et al. (2018). The C-terminal helix of ribosomal P stalk recognizes a hydrophobic groove of elongation factor 2 in a novel fashion. *Nucleic Acids Res.* 46, 3232–3244. doi: 10.1093/nar/gky115

Valenzuela, D. M., Chaudhuri, A., and Maitra, U. (1982). Eukaryotic ribosomal subunit anti-association activity of calf liver is contained in a single polypeptide chain protein of Mr = 25,500 (eukaryotic initiation factor 6). *J. Biol. Chem.* 257, 7712–7719.

Warren, A. J. (2018). Molecular basis of the human ribosomopathy Shwachman-Diamond syndrome. *Adv. Biol. Regul.* 67, 109–127. doi: 10.1016/j.jbior.2017.09.002

Weis, F., Giudice, E., Churcher, M., Jin, L., Hilcenko, C., Wong, C. C., et al. (2015). Mechanism of eIF6 release from the nascent 60S ribosomal subunit. *Nat. Struct. Mol. Biol.* 22, 914–919. doi: 10.1038/nsmb.3112

Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410.

Witharana, C., Roppelt, V., Lochnit, G., Klug, G., and Evguenieva-Hackenberg, E. (2012). Heterogeneous complexes of the RNA exosome in *Sulfolobus solfataricus*. *Biochimie* 94, 1578–1587. doi: 10.1016/j.biochi.2012.03.026

Wong, C. C., Traynor, D., Basse, N., Kay, R. R., and Warren, A. J. (2011). Defective ribosome assembly in Shwachman-Diamond syndrome. *Blood* 118, 4305–4312. doi: 10.1182/blood-2011-06-353938

Yutin, N., Puigbò, P., Koonin, E. V., and Wolf, Y. I. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PLoS One* 7:e36972. doi: 10.1371/journal.pone.0036972

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.

frontiers in Microbiology

PERSPECTIVE
published: 29 April 2021
doi: 10.3389/fmicb.2021.661827

Coupling of Transcription and Translation in Archaea: Cues From the Bacterial World

*Albert Weixlbaumer[1,2,3,4], Felix Grünberger[5], Finn Werner[6] and Dina Grohmann[5,7]**

[1]Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Illkirch, France, [2]Université de Strasbourg, Strasbourg, France, [3]CNRS UMR7104, Illkirch, France, [4]INSERM U1258, Illkirch, France, [5]Institute of Microbiology and Archaea Centre, University of Regensburg, Regensburg, Germany, [6]RNAP Lab, Division of Biosciences, Institute for Structural and Molecular Biology, London, United Kingdom, [7]Regensburg Center for Biochemistry, University of Regensburg, Regensburg, Germany

The lack of a nucleus is the defining cellular feature of bacteria and archaea. Consequently, transcription and translation are occurring in the same compartment, proceed simultaneously and likely in a coupled fashion. Recent cryo-electron microscopy (cryo-EM) and tomography data, also combined with crosslinking-mass spectrometry experiments, have uncovered detailed structural features of the coupling between a transcribing bacterial RNA polymerase (RNAP) and the trailing translating ribosome in *Escherichia coli* and *Mycoplasma pneumoniae*. Formation of this supercomplex, called expressome, is mediated by physical interactions between the RNAP-bound transcription elongation factors NusG and/or NusA and the ribosomal proteins including uS10. Based on the structural conservation of the RNAP core enzyme, the ribosome, and the universally conserved elongation factors Spt5 (NusG) and NusA, we discuss requirements and functional implications of transcription-translation coupling in archaea. We furthermore consider additional RNA-mediated and co-transcriptional processes that potentially influence expressome formation in archaea.

Keywords: RNA polymerase, ribosome, archaea, expressome, Spt4/5, NusG, Nus

OPEN ACCESS

Edited by:
Paola Londei,
Sapienza University of Rome, Italy

Reviewed by:
Béatrice Clouet-d'Orval,
Centre National de la Recherche Scientifique (CNRS), France
Bertram Daum,
University of Exeter, United Kingdom

***Correspondence:**
Dina Grohmann
dina.grohmann@ur.de

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

Received: *31 January 2021*
Accepted: *30 March 2021*
Published: *29 April 2021*

Citation:
Weixlbaumer A, Grünberger F, Werner F and Grohmann D (2021) Coupling of Transcription and Translation in Archaea: Cues From the Bacterial World. Front. Microbiol. 12:661827. doi: 10.3389/fmicb.2021.661827

INTRODUCTION

The controlled and coordinated expression of genes plays a fundamental role in all cellular life forms and occurs in two steps: transcription of DNA to RNA by RNA polymerase (RNAP) and translation of RNA to protein by the ribosome. Cellular RNAPs share a conserved core architecture (Hirata et al., 2008; Korkhin et al., 2009; Werner and Grohmann, 2011; Jun et al., 2014; Griesenbeck et al., 2017). However, the archaeal RNAP structure, subunit composition, and use of basal transcription factors (TF) are more closely related to eukaryotic RNAP II than the bacterial counterpart. Ribosomes are large ribonucleoprotein particles that consist of two subunits that entail ribosomal proteins (r-proteins) and rRNAs. While the general organization and function of the ribosome is universally conserved, the complexity and protein content of ribosomes increases from bacteria to archaea to eukaryotes (Armache et al., 2013; Yusupova and Yusupov, 2014; Ferreira-Cerca, 2017). In fact, differences in the transcriptional and translational apparatus reflect the increase in complexity during evolution (Armache et al., 2013).

For example, major differences in ribosome subunit composition are already apparent in the four phylogenetically distinct (super-) phyla: Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota (TACK); Euryachaeota; Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaeota (DPANN); and Asgard archaea.

Prokaryotes lack a nucleus, so transcription and translation occur in the same cellular compartment, the cytoplasm. Biochemical evidence and electron micrographs of lysed bacteria led to the early proposal and realization that translation occurs co-transcriptionally (Byrne et al., 1964; Miller et al., 1970). This prompted the question whether coordination or coupling of elongating RNAP with the pioneering ribosome mutually influences transcription and translation. Data from bacteria provided direct evidence that rates of transcription and translation are interdependent, at least in some species and for some transcription units (Landick et al., 1985; Proshkin et al., 2010; Castro-Roa and Zenkin, 2012; Zhu et al., 2019; Johnson et al., 2020; Stevenson-Jones et al., 2020). However, recent work in *Bacillus subtilis* showed that coupling of transcription and translation is not conserved across all bacteria (Johnson et al., 2020). Recently, single-particle cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) was used to elucidate structural details of the coupled bacterial RNAP and ribosome, a macromolecular assembly termed "expressome." It highlighted roles of transcription elongation factors NusG and/or NusA that physically connect RNAP with the ribosome (Demo et al., 2017; Kohler et al., 2017; O'Reilly et al., 2020; Wang et al., 2020; Webster et al., 2020).

In contrast, little is known about the coupling of transcription and translation in archaea. It is unclear if direct interactions between RNAP and ribosomes occur or if their association is solely mediated by the shared mRNA. Likewise, the contribution and regulatory role of accessory transcription factors is unknown (McGary and Nudler, 2013; Artsimovitch, 2018). Based on the structural information of the bacterial expressome(s), we discuss whether a coupling between the archaeal RNAP and ribosome can take place in a comparable manner. While molecular structures often guide hypotheses about underlying molecular mechanisms, they rarely suffice to provide the complete picture. We discuss how additional functional evidence obtained *in vivo*, including reporter gene assays and systems biology data such as transcriptome analyses and ribosome profiling, can shed light on the coupled gene expression in archaea. Finally, gene expression takes place in the context of other essential physiological processes. Hence, events like RNA processing and degradation, and their impact on transcription, translation, and their coupling are important to consider.

STRUCTURAL INSIGHTS INTO THE BACTERIAL EXPRESSOME

Attempts to gain structural insights into bacterial expressomes were based on two approaches: (i) cryo-EM of samples formed by direct reconstitution of purified *Escherichia coli* components on mRNA substrates, which direct a precise spacing between RNAP and the 70S ribosome (Wang et al., 2020; Webster et al., 2020), or (ii) direct visualization using cryo-ET in combination with in-cell cross-linking mass spectrometry in *Mycoplasma pneumoniae* (O'Reilly et al., 2020). With sufficient mRNA separating the two machineries, RNAP adopts a wide range of orientations, the assembly is highly flexible, and the mRNA is the only consistent connection (**Figure 1A**). In *E. coli*, adding NusG restrains RNAP and aligns the mRNA with the ribosomal helicase (**Figure 1B**), proposed to prevent secondary structure formation in the transcript (Webster et al., 2020). Addition of the TF NusA stabilizes the NusG-coupled expressome (Wang et al., 2020; **Figures 1D,E**). In contrast, in *M. pneumoniae*, NusA alone appears to couple the two machineries without a role for NusG, albeit in a different relative orientation (O'Reilly et al., 2020; **Figure 1F**). This is consistent with the weak sequence conservation in the NusG KOW domain of *E. coli* and *M. pneumoniae* and suggests that a different mechanism for coupling evolved in this minimal genome species.

All three studies concluded that short spacings between RNAP and the ribosome, either directed by the mRNA or by adding a drug to halt RNAP, form expressomes that resemble an earlier lower-resolution reconstruction formed by collision of a translating ribosome with a stalled RNAP (**Figure 1C**; Kohler et al., 2017). Importantly, while RNAP is still mobile in this collided conformation, NusG cannot simultaneously bind RNAP and the ribosome and therefore cannot form a physical link.

While it is tempting to suggest uncoupled, NusG coupled, and collided expressome structures represent a ribosome approaching RNAP (in agreement with a reduction in RNA separating the two machineries), there is no other experimental evidence to support this chronological order of events, and this remains subject for further research.

IS TRANSCRIPTION COUPLED TO TRANSLATION IN ARCHAEA?

For the archaeon *Thermococcus kodakarensis*, DNA-attached polysomes have been visualized by electron microscopy (French et al., 2007) suggesting that transcription translation coupling (TTC) occurs in archaea. Given the bacterial expressome, the question arises whether the archaeal machineries are compatible with this architecture. To answer this question, the bacterial transcription and translation apparatus has to be compared concerning (i) the overall RNAP architecture, (ii) the RNA length bridging the RNAP active site with the ribosomal P-site (carrying the peptidyl-tRNA), (iii) the presence of NusG or NusA-like factors, and (iv) the conservation of interaction surfaces.

In contrast to bacterial RNAPs, archaeal-eukaryotic RNAPs contain subunits Rpo4/7 (the stalk domain), which binds nascent RNA (Todone et al., 2001; Meka et al., 2005) and stimulate RNAP processivity (Hirtreiter et al., 2009) suggesting the stalk guides the RNA away from RNAP once it emerges from the RNA exit channel. Complexes between bacterial RNAP and 70S ribosomes could be observed for RNA spacers as short

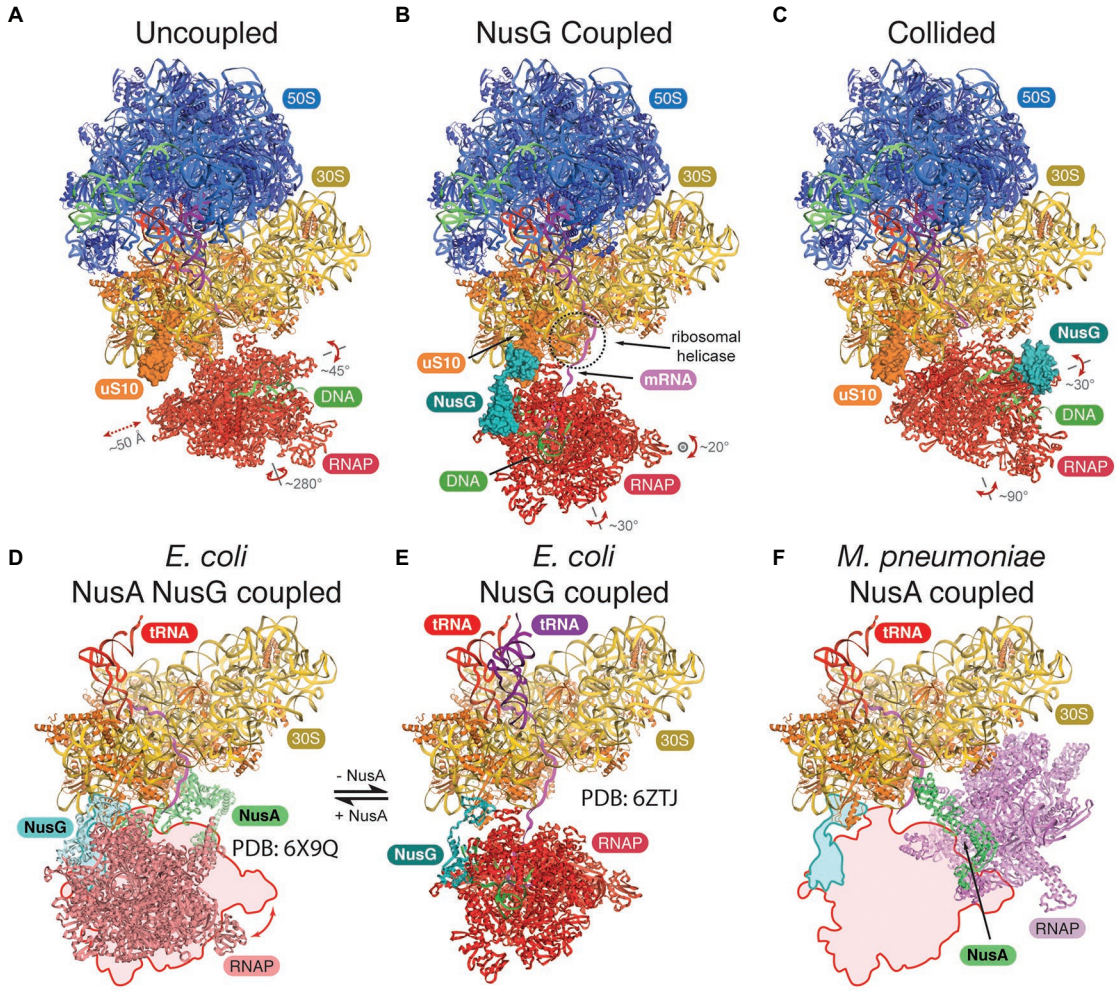


FIGURE 1 | Structures of the bacterial expressomes. **(A)** At mRNA spacings separating the RNA polymerase (RNAP) active site by more than ~35 nucleotides from the ribosomal P-site, and in absence of any coupling factor, RNAP adopts a wide range of orientations relative to the ribosome (uncoupled state; compare extent of translations and rotations). **(B)** Addition of the coupling factor NusG (turquoise) results in the formation of a physical link between RNAP and the ribosome (through uS10, orange) and restrains the rotational freedom. The emerging mRNA transcript (magenta) aligns with the ribosomal helicase (r-proteins uS3 and uS4, dashed circle; NusG coupled state). **(C)** Once the ribosome approaches RNAP further (mRNA spacings less than ~34 nucleotides), the expressome adopts a conformation similar to a previously observed collided state, where NusG can no longer form a physical bridge (collided state). **(D–F)** Comparison of transcription factor coupled states in *Escherichia coli* and *Mycoplasma pneumoniae*. The NusG-coupled state **(E)** gets stabilized by NusA in a similar relative orientation [**D**, compare outline with model, differences are within rotational freedom indicated in panel **(A)**]. In contrast, the NusA-coupled state observed in *M. pneumoniae* requires a major change in position and orientation of RNAP **(F)**, compare outline and model, *E. coli* RNAP and ribosome were docked into deposited cryo-electron tomography (cryo-ET) map (O'Reilly et al., 2020).

as 29 nt separating the RNAP active site from the ribosomal P-site (Wang et al., 2020). However, NusG-mediated coupling appears to be compatible only with spacer lengths greater than at least ~34 nt (Webster et al., 2020). Cryo-EM reconstructions (Bernecky et al., 2016; Ehara et al., 2017) and single-molecule FRET studies (Andrecka et al., 2008) of eukaryotic elongation complexes showed that transcripts of 14–29 nt reach the stalk base. Longer RNAs could not be mapped and appeared to be flexible. This suggests the attachment of longer RNAs to the stalk is transient or they are no longer associated with the stalk. In the context of the archaeal RNAP and assuming that the nascent RNA binds the stalk, for TTC to occur, a longer mRNA segment is required that can traverse the stalk before being fed into the ribosome in contrast to the bacterial situation. Alternatively, the mRNA might be detached from Rpo4/7 and directly enter the ribosome.

In the *E. coli* expressome, RNA-dependent TTC is further mediated by NusG, which is the only universally conserved TF (Werner, 2012). In archaea and eukaryotes, the NusG homolog is called Spt5 and forms a heterodimer with Spt4. NusG/Spt5 has an N-terminal NGN domain and a C-terminal KOW domain, which bind the RNAP clamp domain and the r-protein uS10, respectively (**Figure 2A**). At the majority of genes, archaeal Spt4/5 associates with the elongation complex proximal to the promoter and reflects the RNAP association pattern (Smollett et al., 2017). This suggests early Spt4/5

recruitment to RNAP even for short transcripts and thus a coupling function may also occur early in transcription. NusG/Spt5 are structurally conserved (**Figures 2B,C**; Hirtreiter et al., 2010; Martinez-Rucobo et al., 2011; Liu and Steitz, 2017). Hence, the interaction interfaces between NusG/Spt5 and RNAP and/or the ribosome might also be conserved.

First, we focus on the NusG-mediated contact between RNAP and ribosome because biochemical data suggest this to be the prevalent arrangement of the expressome *in vivo* (Saxena et al., 2018; Washburn et al., 2020). The binding site of NusG/Spt5 on RNAP is conserved according to structural data in all three kingdoms of life (**Figures 2B,C**; Klein et al., 2011; Martinez-Rucobo et al., 2011; Ehara et al., 2017; Kang et al., 2018). While structural data on archaeal Spt5 (aSpt5) interacting with the archaeal ribosome are missing, the length and mobility of the linker connecting the NGN and KOW domain in aSpt5 resembles NusG. Thus, a similar interaction as observed for the bacterial NusG-coupled expressome is feasible. Furthermore, bacterial RNAP exhibits substantial rotational and translational freedom with respect to the ribosome even in the NusG-coupled expressome. Modeling of an archaeal expressome based on bacterial RNAP orientations (Webster et al., 2020) shows that most orientations would require a different stalk orientation to avoid steric overlap with the 30S subunit (Coureux et al., 2020). Archaeal RNAP might either be more restricted in its orientation relative to the ribosome or adopt different orientations compatible with the stalk that have been modeled to be possible without steric clashes between RNAP and the ribosome (Kohler et al., 2017).

Bacterial uS10 provides a hydrophobic pocket for the KOW domain of NusG to insert several hydrophobic residues (Webster et al., 2020). Residues V84 and M88 in uS10 form one edge of the hydrophobic pocket in close proximity to F141, F144, and I164 in the NusG-KOW domain (Burmann et al., 2010; Webster et al., 2020). V84 and M88 in uS10 and F141 and I164 in the KOW domain are conserved among bacterial and archaeal proteins (**Figures 2B,C**; Melnikov et al., 2018) suggesting the hydrophobic interaction between NusG/Spt5-KOW and uS10 might be conserved. Moreover, the structure of bacterial and archaeal uS10 is conserved (**Figure 2C**) and residues in the putative interaction surface (β-strand 1 and 4, α-helix 2) of archaeal uS10 with Spt5 are highly conserved among archaeal uS10 proteins suggesting that the amino acid identity might play a role for the function and interaction of archaeal uS10 (Coureux et al., 2020; **Figure 2B**). Despite the conserved phenylalanine residues and overall sequence conservation of aSpt5, organisms of the euryarchaeal and crenarchaeal phylum do not share a high sequence conservation with bacterial KOW sequences.

The interactions in the expressome are not conserved across all bacteria and alternative coupling mechanisms have evolved. In *M. pneumoniae*, the bacterial elongation and termination factor NusA couples RNAP and the ribosome (O'Reilly et al., 2020). Commonly, bacterial NusA proteins contain an N-terminal domain (binds RNAP), and a S1 and two KH domains (bind RNA). *Mycoplasma pneumoniae* NusA contains an additional flexible C-terminal extension not found in *E. coli* or *B. subtilis*, which contacts multiple r-proteins on the ribosome (**Figure 1D**). Consequently, the relative orientation of the ribosome to RNAP differs significantly from the *E. coli* expressome architecture (**Figure 1D**). NusA is also able to stabilize NusG-coupled expressomes in *E. coli* mediated by one of the KH domains (Wang et al., 2020). NusA-like homologs can be found in all archaeal phyla indicating a widespread distribution of this transcription factor but its function is unclear (Shibata et al., 2007). The domain organization differs significantly from bacterial NusA because archaeal NusA (aNusA) only contains KH domains but lacks the NTD, S1 domain and C-terminal extension that interacts with the ribosome in *M. pneumoniae*. Nevertheless, the structure of the bacterial and archaeal KH domains in NusA are highly conserved (**Figures 2D,E**) and aNusA also binds RNA (Shibata et al., 2007). It has been suggested that the RNAP interaction platform and S1 domain of Rpo7 in conjunction with the two KH domains of aNusA form the domain complement of bacterial NusA (**Figure 2F**; Belogurov and Artsimovitch, 2015; Fouqueau et al., 2018).

It is noteworthy that the archaeal domain of life encompasses highly diverse organisms, of which only a few model organisms have been studied so far. As documented for the bacterial world (Irastortza-Olaziregi and Amster-Choder, 2020; Johnson et al., 2020), expressome formation might occur in some archaeal species but not in others.

CO-TRANSCRIPTIONAL PROCESSES AND TRANSCRIPTOMIC FEATURES AFFECTING TRANSCRIPTION-TRANSLATION COUPLING IN ARCHAEA

The expressome structures illustrate the highly coordinated interplay of two molecular machineries. However, the expressome is not an isolated complex but operates with high specificity in a crowded cytoplasm where myriads of molecular processes occur simultaneously. In archaea, a number of transcriptional and co-transcriptional steps have been identified that might prevent the immediate loading of the ribosome onto the mRNA. Among others, processes like co-transcriptional RNA processing, binding of non-coding RNAs (ncRNAs) to and association of RNA chaperones and transcription termination factors with the RNA may influence expressome formation and will be shortly discussed in this section (compare **Figures 2G–I**).

Coupling of the ribosome to RNAP requires the mRNA to span the distance between the RNAP active site and the ribosomal P-site to provide enough space for both machineries (**Figure 2G**). Typically, regulatory sequences that confer translation initiation are encoded in the 5' untranslated region (5'-UTR). Some archaeal mRNAs have a short 5' UTR or none at all (analyzed for *Haloferax volcanii*, *Thermococcus onnurineus*, *Pyrococcus abyssi*, *Saccharolobus solfataricus*, Heyer et al., 2012; Xu et al., 2012; Beck and Moll, 2018). The relative number of leaderless mRNAs ranges between 1.4 and 72%. The mechanism of mRNA recognition and ribosome association appears to be highly diverse in prokaryotes, and we do not know whether the initiation mechanism influences and correlates with

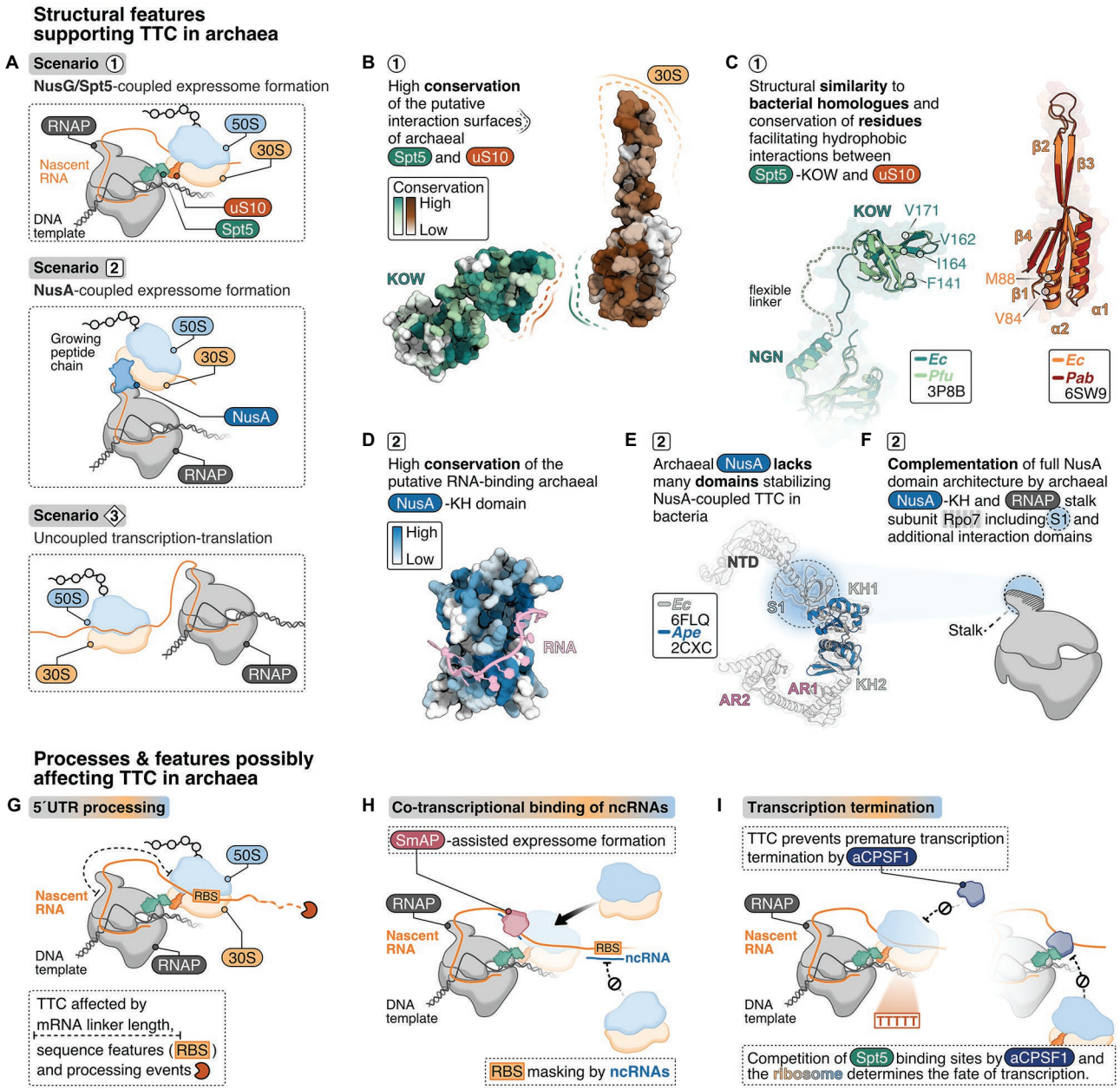


FIGURE 2 | Structural criteria and cellular processes that might mediate and influence transcription-translation coupling in archaea. **(A)** Possible scenarios for transcription translation coupling (TTC) in archaea derived from structural insights into expressome formation in bacteria: NusG/Spt5-coupled expressome formation is a possible scenario 1 representing the archaeal equivalent to the NusG-coupled state in *E. coli* (Webster et al., 2020). In addition, one could also imagine NusA-coupling (scenario 2), similar to what has been observed in *M. pneumoniae* (O'Reilly et al., 2020) or no coupling at all (scenario 3). **(B)** Analysis of conserved regions in archaeal Spt5 and uS10 using ConSurf (Ashkenazy et al., 2016). About 100 archaeal Spt5 and uS10 sequences were aligned, and their conservation score projected color-coded from white (0, not conserved) to dark green or dark-red (9, highly conserved), respectively, on the surface of *Pyrococcus furiosus* Spt5 and *Pyrococcus abyssi* uS10 structure. **(C)** Superimposition of archaeal and bacterial Spt5/NusG and (bacteria: dark-green, PDB: 6ZTJ, *E. coli*; archaea: light-green, PDB: 3P8B, *P. furiosus*) and uS10 (bacteria: orange, PDB: 6ZTJ, *E. coli*; archaea: red, PDB: 6SW9, *P. abyssi*). Residues important for the interaction with NusG are highlighted. **(D)** Conservation analysis of archaeal NusA proteins. Conservation scores from white (0, not conserved) to dark-blue (0, highly conserved) were calculated based on the comparison of archaeal NusA proteins and projected on the surface of *Aeropyrum pernix* NusA. A superimposed model of bacterial RNA (Beuth et al., 2005) is shown in pink. **(E)** The *E. coli* NusA structure (gray, PDB: 6FLQ) overlayed with *A. pernix* NusA (Shibata et al., 2007) (blue, PDB: 2CXC). **(F)** Cartoon depiction of the archaeal RNAP highlighting the hypothesis that the S1 domain of the stalk-forming subunit Rpo7 and the archaeal NusA form a homologue of bacterial NusA. TTC in archaea may be affected by co-transcriptional processes and features depicted in **(G–I)**, including 5'UTR length and processing **(G)**, co-transcriptional binding of non-coding RNAs (ncRNAs) **(H)** and the transcription termination pathway **(I)**.

expressome formation (Wen et al., 2020). mRNAs that lack a ribosomal binding site (RBS) can also emerge from RNA processing events at the 5'-end that lead to cleavage of the 5'-UTR (Qi et al., 2017; **Figure 2G**). As shown for several bacterial (Mäder et al., 2004; Ramirez-Peña et al., 2010; Lioliou et al., 2012) and for the archaeal organisms *Methanocaldococcus jannaschii* and *Marinobacter psychrophilus* (Zhang and Olsen, 2009; Qi et al., 2017), processing of the mRNAs can stabilize transcripts and regulate translation of r-proteins (Qi et al., 2017) and mRNAs from multicistronic operons. In this case, the timing of mRNA processing and translation seems important to avoid conflicts between these two processes.

Co-transcriptional binding of a small regulatory ncRNA to an mRNA is a common posttranscriptional regulation mechanism in prokaryotes that influences RNA stability and translational efficiency of mRNAs in response to changing environmental conditions (Babski et al., 2014; Hör et al., 2018). For *H. volcanii* and *Methanosarcina mazei* small ncRNAs have been detected that can potentially bind to the 5' UTR thereby potentially masking the RBS (Jäger et al., 2009; Soppa et al., 2009; Heyer et al., 2012; Gelsinger and DiRuggiero, 2018; **Figure 2H**). For example, the small RNA_{41} in *M. mazei* binds multiple RBS in a polycistronic mRNA and decouples transcription and translation (Buddeweg et al., 2018).

In bacteria, ncRNA-mRNA hybridization is often mediated by the RNA chaperone Hfq, which belongs to the Sm protein family (Vogel and Luisi, 2011). Hfq can bind RNA co-transcriptionally (Kambara et al., 2018) and plays a role in transcription termination/antitermination (Rabhi et al., 2011; Sedlyarova et al., 2016), ribosome biogenesis (Andrade et al., 2018) and ribosome association with the mRNA in bacteria (Chen et al., 2019). In archaea, a *bona fide* Hfq protein is rarely encoded. More often, single or multiple genes encode an archaeal Sm-like protein (SmAP; Reichelt et al., 2018). Similar to bacterial Hfq, archaeal SmAPs were shown to bind RNAs (Nielsen et al., 2007; Fischer et al., 2010; Märtens et al., 2015). Hence, co-transcriptional association of a ncRNA at the 5' UTR (potentially supported by a SmAP) would prevent ribosome association with the 5' UTR (**Figure 2H**). Co-immunoprecipitation experiments showed that SmAPs not only bind RNAs but also r-proteins (Fischer et al., 2010). It is conceivable that SmAPs participate in posttranscriptional regulation, translation, or act as a bridging factor to recruit ribosomes to the mRNA (**Figure 2H**).

Lastly, the transcription termination pathway might be decisive whether TTC can occur, or vice versa (**Figure 2I**). In archaea, transcription terminates *via* two mechanisms that are not necessarily mutually exclusive: (i) intrinsic termination at poly(U) stretches (Santangelo and Reeve, 2006; Hirtreiter et al., 2009, 2010; Santangelo et al., 2009; Dar et al., 2016; Berkemer et al., 2020) or (ii) factor-dependent termination assisted by the archaeal termination factor aCPSF1/FttA that binds the nascent RNA (Sanders et al., 2020; Yue et al., 2020). Importantly, aCPSF1 also enhances termination at poly(U) stretches. Termination *via* aCSPF1 involves cleavage of the transcript at the 3'-end. In *Methanococcus maripaludis* deletion of aCPSF1 resulted in altered expression levels for the majority of genes (Yue et al., 2020). Furthermore, aCPSF1-dependent termination gets stimulated by the presence of the stalk domain and Spt4/5 (Sanders et al., 2020). Even though a direct interaction between aCPSF1 and the stalk or Spt4/5 has not yet been experimentally verified, a physical interaction is likely and would be consistent with the observed increased termination efficiency. It is tempting to speculate that aCPSF1 and the ribosome interact with RNAP-bound Spt4/5 in a mutually exclusive fashion similar to Rho and the ribosome with RNAP-bound NusG in bacteria. As a consequence, transcription termination and ribosome coupling might be mutually exclusive. Ribosomes coupled to RNAP *via* Spt4/5 would prevent aCPSF1 interactions with the nascent RNA and prevent premature termination (**Figure 2H**). Alternatively, once aCPSF1 gains access to Spt4/5 it may interfere with TTC (**Figure 2H**). This would be reminiscent of the recruitment of Rho by NusG-KOW to RNAP leading to transcription termination of non-coding/untranslated RNA transcripts (Washburn et al., 2020). Whether TTC or termination prevails could be gene- or operon-specific, could be a target for regulation, and may vary from species to species. Direct, mRNA-independent interactions between the bacterial RNAP and ribosome have been shown. It is possible that in some instances, e.g., during transcription of short mRNAs, the archaeal ribosome might bind the mRNA close to the RNAP exit channel and direct contacts between the elongating RNAP and the ribosome (Wang et al., 2020; Webster et al., 2020).

FUTURE PERSPECTIVES

Are transcription and translation coupled in archaea similar to bacteria? We propose that this is likely, but definitive proof is still lacking. This problem can only be solved by a multidisciplinary effort that reaches beyond a molecular-structural analysis *in vitro*. In order to rationalize the underlying molecular mechanisms, we do need to understand the structural determinants of the RNAP-ribosome interactions and the potential role of general regulatory factors including NusG-Spt4/5 and NusA, as well as ribosomal proteins. A crucial question to be solved in the future is how co-transcriptional processes like SmAP binding, transcription termination, or RNA processing are coordinated with ribosome coupling in space and time. This also includes coordination of translation initiation and TTC. In bacteria, translation initiation was delineated in great detail showing that the 30S subunit is recruited to the mRNA with the help of the initiator tRNA and initiation factors before the 50S subunit joins to form the translation-competent ribosome (see for example, Milon et al., 2010; Tsai et al., 2012). In archaea, the situation is more complex as additional (eukaryotic-like) initiation factors are involved (Benelli et al., 2016). Nonetheless, even in the archaeal initiation complex, uS10 remains exposed and might be available for coupling to Spt5 (Coureux et al., 2020). Consequently, the 30S subunit is involved in translation initiation and coupling to RNAP, and it has to be seen whether these processes are compatible or mutually exclusive.

To elaborate on the finer points of biologically relevant interaction networks a combination of cross-linking/mass spectrometry experiments like the recent elegant study of Rappsilber and colleagues are necessary (O'Reilly et al., 2020). Complementary to these efforts are structural biology, functional genomics, and systems biology approaches that hold great promise to ascertain (i) to which extent the coupling applies to all transcription units or whether it is limited to specific subset or classes of operons, and (ii) whether the coupling-uncoupling is a dynamic process and dependent on environmental cues and stresses, i.e., whether it is subject to regulation. Key to this approach are experiments that monitor changes in the global characteristics of transcription, such as genome-wide RNAP occupancy profiles and transcriptome analyses, in response to perturbations of translation by using ribosome inhibitors/antibiotics or ribosome variants. We have to develop high-resolution methods that combine ribo-seq/proteomics and RNAP NET-seq or ChIP-exo/transcriptomics and integrate the data to obtain a complete view of the interdependence of transcription and translation. Finally, it is important to note that archaea are evolutionary diverse and tractable archaeal model organisms are scarce. Despite the conservation of NusG, the molecular mechanisms of transcription that were revealed for Crenarchaea and Euryarchaea are distinct in many ways including the RNAP subunit composition and chromatin structure. Likewise, we only know little about the mechanisms of translation across the archaeal phyla. The properties of their ribosomes are distinct including the molecular mechanisms of translation initiation, which might have an impact on the coupling of the leading ribosome to the RNAP.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AW prepared the **Figure 1**. FG prepared the **Figure 2**. All authors contributed to the article and approved the submitted version.

FUNDING

Research in the RNAP laboratory at UCL was funded by a Wellcome Investigator Award in Science "Mechanisms and Regulation of RNAP transcription" to FW (WT 207446/Z/17/Z). Publications costs were covered by the German Research Foundation (DFG) with the funding program Open Access Publishing. We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (SFB960-TP7 to DG), by the European Research Council (ERC starting grant TRANSREG 679734 to AW).

ACKNOWLEDGMENTS

FW, AW, and DG thank the members of the Werner, Weixlbaumer, and Grohmann labs for fruitful discussions and comments on this manuscript.

REFERENCES

Andrade, J. M., Dos Santos, R. F., Chelysheva, I., Ignatova, Z., and Arraiano, C. M. (2018). The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity. *EMBO J.* 37:e97631. doi: 10.15252/embj.201797631

Andrecka, J., Lewis, R., Brückner, F., Lehmann, E., Cramer, P., and Michaelis, J. (2008). Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* 105, 135–140. doi: 10.1073/pnas.0703815105

Armache, J.-P., Anger, A. M., Márquez, V., Franckenberg, S., Fröhlich, 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

Artsimovitch, I. (2018). Rebuilding the bridge between transcription and translation. *Mol. Microbiol.* 108, 467–472. doi: 10.1111/mmi.13964

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., et al. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350. doi: 10.1093/nar/gkw408

Babski, J., Maier, L.-K., Heyer, R., Jaschinski, K., Prasse, D., Jäger, D., et al. (2014). Small regulatory RNAs in archaea. *RNA Biol.* 11, 484–493. doi: 10.4161/rna.28452

Beck, H. J., and Moll, I. (2018). Leaderless mRNAs in the spotlight: ancient but not outdated! *Microbiol. Spectr.* 6, 155–170. doi: 10.1128/microbiolspec.RWR-0016-2017

Belogurov, G. A., and Artsimovitch, I. (2015). Regulation of transcript elongation. *Annu. Rev. Microbiol.* 69, 49–69. doi: 10.1146/annurev-micro-091014-104047

Benelli, D., La Teana, A., and Londei, P. (2016). "Evolution of translational initiation: from archaea to eukarya" in *Evolution of the Protein Synthesis Machinery and Its Regulation*. eds. G. Hernández and R. Jagus (Cham: Springer), 61–79.

Berkemer, S. J., Maier, L.-K., Amman, F., Bernhart, S. H., Wörtz, J., Märkle, P., et al. (2020). Identification of RNA 3' ends and termination sites in Haloferax volcanii. *RNA Biol.* 17, 663–676. doi: 10.1080/15476286.2020.1723328

Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. *Nature* 529, 551–554. doi: 10.1038/nature16482

Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R., and Taylor, I. A. (2005). Structure of a *Mycobacterium tuberculosis* NusA-RNA complex. *EMBO J.* 24, 3576–3587. doi: 10.1038/sj.emboj.7600829

Buddeweg, A., Sharma, K., Urlaub, H., and Schmitz, R. A. (2018). sRNA41 affects ribosome binding sites within polycistronic mRNAs in Methanosarcina mazei Gö1. *Mol. Microbiol.* 107, 595–609. doi: 10.1111/mmi.13900

Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E., et al. (2010). A NusE:NusG complex links transcription and translation. *Science* 328, 501–504. doi: 10.1126/science.1184953

Byrne, R., Levin, J. G., Bladen, H. A., and Nirenberg, M. W. (1964). The in vitro formation of a DNA-ribosome complex. *Proc. Natl. Acad. Sci. U. S. A.* 52, 140–148. doi: 10.1073/pnas.52.1.140

Castro-Roa, D., and Zenkin, N. (2012). In vitro experimental system for analysis of transcription-translation coupling. *Nucleic Acids Res.* 40:e45. doi: 10.1093/nar/gkr1262

Chen, J., Morita, T., and Gottesman, S. (2019). Regulation of transcription termination of small RNAs and by small RNAs: molecular mechanisms and biological functions. *Front. Cell. Infect. Microbiol.* 9:201. doi: 10.3389/fcimb.2019.00201

Coureux, 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

Dar, D., Prasse, D., Schmitz, R. A., and Sorek, R. (2016). Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nat. Microbiol.* 1:16143. doi: 10.1038/nmicrobiol.2016.143

Demo, G., Rasouly, A., Vasilyev, N., Svetlov, V., Loveland, A. B., Diaz-Avalos, R., et al. (2017). Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife* 6:e28560. doi: 10.7554/eLife.28560

Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., and Sekine, S.-I. (2017). Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* 357, 921–924. doi: 10.1126/science.aan8552

Ferreira-Cerca, S. (2017). "Life and death of ribosomes in archaea" in *RNA Metabolism and Gene Expression in Archaea*. ed. B. Clouet-d'Orval (Cham: Springer), 129–158.

Fischer, S., Benz, J., Späth, B., Maier, L.-K., Straub, J., Granzow, M., et al. (2010). The archaeal Lsm protein binds to small RNAs. *J. Biol. Chem.* 285, 34429–34438. doi: 10.1074/jbc.M110.118950

Fouqueau, T., Blombach, F., Cackett, G., Carty, A. E., Matelska, D. M., Ofer, S., et al. (2018). The cutting edge of archaeal transcription. *Emerg. Top. Life Sci.* 2, 517–533. doi: 10.1042/ETLS20180014

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

Gelsinger, D. R., and DiRuggiero, J. (2018). The non-coding regulatory RNA revolution in archaea. *Genes* 9:141. doi: 10.3390/genes9030141

Griesenbeck, J., Tschochner, H., and Grohmann, D. (2017). Structure and function of RNA polymerases and the transcription machineries. *Subcell. Biochem.* 83, 225–270. doi: 10.1007/978-3-319-46503-6_9

Heyer, R., Dörr, M., Jellen-Ritter, A., Späth, B., Babski, J., Jaschinski, K., et al. (2012). High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in Haloferax volcanii. *RNA Biol.* 9, 1011–1018. doi: 10.4161/rna.20826

Hirata, A., Klein, B. J., and Murakami, K. S. (2008). The X-ray crystal structure of RNA polymerase from archaea. *Nature* 451, 851–854. doi: 10.1038/nature06530

Hirtreiter, A., Damsma, G. E., Cheung, A. C. M., Klose, D., Grohmann, D., Vojnic, E., et al. (2010). Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* 38, 4040–4051. doi: 10.1093/nar/gkq135

Hirtreiter, A., Grohmann, D., and Werner, F. (2009). Molecular mechanisms of RNA polymerase—the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic Acids Res.* 38, 585–596. doi: 10.1093/nar/gkp928

Hör, J., Gorski, S. A., and Vogel, J. (2018). Bacterial RNA biology on a genome scale. *Mol. Cell* 70, 785–799. doi: 10.1016/j.molcel.2017.12.023

Irastortza-Olaziregi, M., and Amster-Choder, O. (2020). Coupled transcription-translation in prokaryotes: an old couple with new surprises. *Front. Microbiol.* 11:624830. doi: 10.3389/fmicb.2020.624830

Jäger, D., Sharma, C. M., Thomsen, J., Ehlers, C., Vogel, J., and Schmitz, R. A. (2009). Deep sequencing analysis of the Methanosarcina mazei Gö1 transcriptome in response to nitrogen availability. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21878–21882. doi: 10.1073/pnas.0909051106

Johnson, G. E., Lalanne, J.-B., Peters, M. L., and Li, G.-W. (2020). Functionally uncoupled transcription-translation in Bacillus subtilis. *Nature* 585, 124–128. doi: 10.1038/s41586-020-2638-5

Jun, S.-H., Hirata, A., Kanai, T., Santangelo, T. J., Imanaka, T., and Murakami, K. S. (2014). The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat. Commun.* 5:5132. doi: 10.1038/ncomms6132

Kambara, T. K., Ramsey, K. M., and Dove, S. L. (2018). Pervasive targeting of nascent transcripts by Hfq. *Cell Rep.* 23, 1543–1552. doi: 10.1016/j.celrep.2018.03.134

Kang, J. Y., Mooney, R. A., Nedialkov, Y., Saba, J., Mishanina, T. V., Artsimovitch, I., et al. (2018). Structural basis for transcript elongation control by NusG family universal regulators. *Cell* 173, 1650.e14–1662.e14. doi: 10.1016/j.cell.2018.05.017

Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X., and Murakami, K. S. (2011). RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. U. S. A.* 108, 546–550. doi: 10.1073/pnas.1013828108

Kohler, R., Mooney, R. A., Mills, D. J., Landick, R., and Cramer, P. (2017). Architecture of a transcribing-translating expressome. *Science* 356, 194–197. doi: 10.1126/science.aal3059

Korkhin, Y., Unligil, U. M., Littlefield, O., Nelson, P. J., Stuart, D. I., Sigler, P. B., et al. (2009). Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* 7:e1000102. doi: 10.1371/journal.pbio.1000102

Landick, R., Carey, J., and Yanofsky, C. (1985). Translation activates the paused transcription complex and restores transcription of the trp operon leader region. *Proc. Natl. Acad. Sci. U. S. A.* 82, 4663–4667. doi: 10.1073/pnas.82.14.4663

Lioliou, E., Sharma, C. M., Caldelari, I., Helfer, A.-C., Fechter, P., Vandenesch, F., et al. (2012). Global regulatory functions of the *Staphylococcus aureus* endoribonuclease III in gene expression. *PLoS Genet.* 8:e1002782. doi: 10.1371/journal.pgen.1002782

Liu, B., and Steitz, T. A. (2017). Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res.* 45, 968–974. doi: 10.1093/nar/gkw1159

Mäder, U., Hennig, S., Hecker, M., and Homuth, G. (2004). Transcriptional organization and posttranscriptional regulation of the Bacillus subtilis branched-chain amino acid biosynthesis genes. *J. Bacteriol.* 186, 2240–2252. doi: 10.1128/JB.186.8.2240-2252.2004

Märtens, B., Bezerra, G. A., Kreuter, M. J., Grishkovskaya, I., Manica, A., Arkhipova, V., et al. (2015). The heptameric SmAP1 and SmAP2 proteins of the crenarchaeon sulfolobus solfataricus bind to common and distinct RNA targets. *Life* 5, 1264–1281. doi: 10.3390/life5021264

Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M., and Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* 30, 1302–1310. doi: 10.1038/emboj.2011.64

McGary, K., and Nudler, E. (2013). RNA polymerase and the ribosome: the close relationship. *Curr. Opin. Microbiol.* 16, 112–117. doi: 10.1016/j.mib.2013.01.010

Meka, H., Werner, F., Cordell, S. C., Onesti, S., and Brick, P. (2005). Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res.* 33, 6435–6444. doi: 10.1093/nar/gki945

Melnikov, S., Manakongtreecheep, K., and Söll, 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

Miller, O. L., Hamkalo, B. A., and Thomas, C. A. (1970). Visualization of bacterial genes in action. *Science* 169, 392–395. doi: 10.1126/science.169.3943.392

Milon, P., Carotti, M., Konevega, A. L., Wintermeyer, W., Rodnina, M. V., and Gualerzi, C. O. (2010). The ribosome-bound initiation factor 2 recruits initiator tRNA to the 30S initiation complex. *EMBO Rep.* 11, 312–316. doi: 10.1038/embor.2010.12

Nielsen, J. S., Bøggild, A., Andersen, C. B. F., Nielsen, G., Boysen, A., Brodersen, D. E., et al. (2007). An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from Methanococcus jannaschii. *RNA* 13, 2213–2223. doi: 10.1261/rna.689007

O'Reilly, F. J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., et al. (2020). In-cell architecture of an actively transcribing-translating expressome. *Science* 369, 554–557. doi: 10.1126/science.abb3758

Proshkin, S., Rahmouni, A. R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328, 504–508. doi: 10.1126/science.1184939

Qi, L., Yue, L., Feng, D., Qi, F., Li, J., and Dong, X. (2017). Genome-wide mRNA processing in methanogenic archaea reveals post-transcriptional regulation of ribosomal protein synthesis. *Nucleic Acids Res.* 45, 7285–7298. doi: 10.1093/nar/gkx454

Rabhi, M., Espéli, O., Schwartz, A., Cayrol, B., Rahmouni, A. R., Arluison, V., et al. (2011). The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J.* 30, 2805–2816. doi: 10.1038/emboj.2011.192

Ramirez-Peña, E., Treviño, J., Liu, Z., Perez, N., and Sumby, P. (2010). The group A Streptococcus small regulatory RNA FasX enhances streptokinase activity by increasing the stability of the ska mRNA transcript. *Mol. Microbiol.* 78, 1332–1347. doi: 10.1111/j.1365-2958.2010.07427.x

Reichelt, R., Grohmann, D., and Willkomm, S. (2018). A journey through the evolutionary diversification of archaeal Lsm and Hfq proteins. *Emerg. Top. Life Sci.* 2, 647–657. doi: 10.1042/ETLS20180034

Sanders, T. J., Wenck, B. R., Selan, J. N., Barker, M. P., Trimmer, S. A., Walker, J. E., et al. (2020). FttA is a CPSF73 homologue that terminates transcription in archaea. *Nat. Microbiol.* 5, 545–553. doi: 10.1038/s41564-020-0667-3

Santangelo, T. J., Cubonová, L. U., Skinner, K. M., and Reeve, J. N. (2009). Archaeal intrinsic transcription termination in vivo. *J. Bacteriol.* 191, 7102–7108. doi: 10.1128/JB.00982-09

Santangelo, T. J., and Reeve, J. N. (2006). Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J. Mol. Biol.* 355, 196–210. doi: 10.1016/j.jmb.2005.10.062

Saxena, S., Myka, K. K., Washburn, R., Costantino, N., Court, D. L., and Gottesman, M. E. (2018). *Escherichia coli* transcription factor NusG binds to 70S ribosomes. *Mol. Microbiol.* 108, 495–504. doi: 10.1111/mmi.13953

Sedlyarova, N., Shamovsky, I., Bharati, B. K., Epshtein, V., Chen, J., Gottesman, S., et al. (2016). sRNA-mediated control of transcription termination in *E. coli*. *Cell* 167, 111.e13–121.e13. doi: 10.1016/j.cell.2016.09.004

Shibata, R., Bessho, Y., Shinkai, A., Nishimoto, M., Fusatomi, E., Terada, T., et al. (2007). Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochem. Biophys. Res. Commun.* 355, 122–128. doi: 10.1016/j.bbrc.2007.01.119

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

Soppa, J., Straub, J., Brenneis, M., Jellen-Ritter, A., Heyer, R., Fischer, S., et al. (2009). Small RNAs of the halophilic archaeon Haloferax volcanii. *Biochem. Soc. Trans.* 37, 133–136. doi: 10.1042/BST0370133

Stevenson-Jones, F., Woodgate, J., Castro-Roa, D., and Zenkin, N. (2020). Ribosome reactivates transcription by physically pushing RNA polymerase out of transcription arrest. *Proc. Natl. Acad. Sci. U. S. A.* 117, 8462–8467. doi: 10.1073/pnas.1919985117

Todone, F., Brick, P., Werner, F., Weinzierl, R. O. J., and Onesti, S. (2001). Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol. Cell* 8, 1137–1143. doi: 10.1016/S1097-2765(01)00379-3

Tsai, A., Petrov, A., Marshall, R. A., Korlach, J., Uemura, S., and Puglisi, J. D. (2012). Heterogeneous pathways and timing of factor departure during translation initiation. *Nature* 487, 390–393. doi: 10.1038/nature11172

Vogel, J., and Luisi, B. F. (2011). Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* 9, 578–589. doi: 10.1038/nrmicro2615

Wang, C., Molodtsov, V., Firlar, E., Kaelber, J., Blaha, G., Su, M., et al. (2020). Structural basis of transcription-translation coupling. *Science* 369, 1359–1365. doi: 10.1126/science.abb5317

Washburn, R. S., Zuber, P. K., Sun, M., Hashem, Y., Shen, B., Li, W., et al. (2020). *Escherichia coli* NusG links the lead ribosome with the transcription elongation complex. *iScience* 23:101352. doi: 10.1016/j.isci.2020.101352

Webster, M. W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Abdelkareem, M., et al. (2020). Structural basis of transcription-translation coupling and collision in bacteria. *Science* 369, 1355–1359. doi: 10.1126/science.abb5036

Wen, J.-D., Kuo, S.-T., and Chou, H.-H. D. (2020). The diversity of Shine-Dalgarno sequences sheds light on the evolution of translation initiation. *RNA Biol.* 1–12. doi: 10.1080/15476286.2020.1861406

Werner, F. (2012). A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* 417, 13–27. doi: 10.1016/j.jmb.2012.01.031

Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* 9, 85–98. doi: 10.1038/nrmicro2507

Xu, N., Li, Y., Zhao, Y.-T., Guo, L., Fang, Y.-Y., Zhao, J.-H., et al. (2012). Identification and characterization of small RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus. *PLoS One* 7:e35306. doi: 10.1371/journal.pone.0053387

Yue, L., Li, J., Zhang, B., Qi, L., Li, Z., Zhao, F., et al. (2020). The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of archaea via a 3'-end cleavage mode. *Nucleic Acids Res.* 48, 9589–9605. doi: 10.1093/nar/gkaa702

Yusupova, G., and Yusupov, M. (2014). High-resolution structure of the eukaryotic 80S ribosome. *Annu. Rev. Biochem.* 83, 467–486. doi: 10.1146/annurev-biochem-060713-035445

Zhang, J., and Olsen, G. J. (2009). Messenger RNA processing in Methanocaldococcus (Methanococcus) jannaschii. *RNA* 15, 1909–1916. doi: 10.1261/rna.1715209

Zhu, M., Mori, M., Hwa, T., and Dai, X. (2019). Disruption of transcription-translation coordination in *Escherichia coli* leads to premature transcriptional termination. *Nat. Microbiol.* 4, 2347–2356. doi: 10.1038/s41564-019-0543-1

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.

frontiers in Microbiology

REVIEW
published: 04 June 2021
doi: 10.3389/fmicb.2021.688061

Programmed Deviations of Ribosomes From Standard Decoding in Archaea

*Federica De Lise[1], Andrea Strazzulli[2,3], Roberta Iacono[1,2], Nicola Curci[1,2], Mauro Di Fenza[1], Luisa Maurelli[1], Marco Moracci[1,2,3] and Beatrice Cobucci-Ponzano[1]**

[1] Institute of Biosciences and BioResources – National Research Council of Italy, Naples, Italy, [2] Department of Biology, University of Naples Federico II, Complesso Universitario di Monte S. Angelo, Naples, Italy, [3] Task Force on Microbiome Studies, University of Naples Federico II, Naples, Italy

Genetic code decoding, initially considered to be universal and immutable, is now known to be flexible. In fact, in specific genes, ribosomes deviate from the standard translational rules in a programmed way, a phenomenon globally termed recoding. Translational recoding, which has been found in all domains of life, includes a group of events occurring during gene translation, namely stop codon readthrough, programmed ± 1 frameshifting, and ribosome bypassing. These events regulate protein expression at translational level and their mechanisms are well known and characterized in viruses, bacteria and eukaryotes. In this review we summarize the current state-of-the-art of recoding in the third domain of life. In Archaea, it was demonstrated and extensively studied that translational recoding regulates the decoding of the 21st and the 22nd amino acids selenocysteine and pyrrolysine, respectively, and only one case of programmed –1 frameshifting has been reported so far in *Saccharolobus solfataricus* P2. However, further putative events of translational recoding have been hypothesized in other archaeal species, but not extensively studied and confirmed yet. Although this phenomenon could have some implication for the physiology and adaptation of life in extreme environments, this field is still underexplored and genes whose expression could be regulated by recoding are still poorly characterized. The study of these recoding episodes in Archaea is urgently needed.

Keywords: alpha-fucosidase, recoding, frameshifting, pyrrolysine, selenocysteine, archaea

OPEN ACCESS

Edited by:
Paola Londei,
Sapienza University of Rome, Italy

Reviewed by:
Emmanuelle Schmitt,
UMR 7654 Bases moléculaires et régulation de la biosynthèse protéique, France
Dario Benelli,
Sapienza University of Rome, Italy

***Correspondence:**
Beatrice Cobucci-Ponzano
beatrice.cobucciponzano@ibbr.cnr.it

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

Received: *30 March 2021*
Accepted: *04 May 2021*
Published: *04 June 2021*

Citation:
De Lise F, Strazzulli A, Iacono R, Curci N, Di Fenza M, Maurelli L, Moracci M and Cobucci-Ponzano B (2021) Programmed Deviations of Ribosomes From Standard Decoding in Archaea. Front. Microbiol. 12:688061. doi: 10.3389/fmicb.2021.688061

INTRODUCTION

Translation, in its basic mechanism, is universally conserved and is performed by one of the most complex and sophisticated cell machineries, the ribosomes, in which the majority of protein components are highly conserved in all of the domains of life. However, both the genetic code and its decoding are neither universal nor immutable due to the complex nature of translation. The genetic code is not quite universal; in fact, it is well established that the meaning of some

codons in certain organelles and organisms has been reassigned (codon reassignment) for all the mRNAs belonging to that organelle or organism. Unlike codon reassignment, in non-canonical translation mechanisms the alteration of the translation rules does not occur for the whole organism but is limited only to specific genes and often correlated to particular physiological conditions that regulate their translation. The discovery of these gene expression regulatory mechanisms has completely changed our view of the disrupted genes that are often found during genome sequencing. In fact, sequenced genomes often reveal interrupted coding sequences and they are generally considered sequencing errors or pseudogenes. It is now well known that the majority of these interrupted genes are functional and encode proteins whose expression is regulated. Non-canonical translation mechanisms have been identified in all steps of the translation: initiation, elongation and termination. Well known strategies related to the initiation of translation are internal ribosome entry, leaky scanning, non-AUG initiation, ribosome shunting and reinitiation. These strategies are used extensively by viruses, presumably providing alternative ways to express different proteins from a single mRNA, facilitating the access to overlapping ORFs and overcoming the structural differences present in viral transcripts in comparison with cellular mRNAs. Furthermore, it has been shown that cancer cells exploit these alternative modes of translation initiation for their survival and proliferation under stressful conditions (for comprehensive reviews see Firth and Brierley, 2012; Pooggin and Ryabova, 2018; Sriram et al., 2018; Yang and Wang, 2019; Cao and Slavoff, 2020).

Programmed deviations from the standard translational rules occuring during translational elongation or termination steps are termed *recoding* (Gesteland and Atkins, 1996; Firth and Brierley, 2012) and, often in competition with standard decoding, have crucial roles in the regulation of gene expression (Baranov et al., 2002). These universal mechanisms are *+1* or *–1 programmed frameshifting* (PRF) and *ribosome hopping,* which occur during the elongation step, and *stop codon readthrough*/*redefinition* occurring during the termination step (Farabaugh, 1996; Gesteland and Atkins, 1996; Baranov et al., 2002; Namy et al., 2004; Ling et al., 2015; Atkins et al., 2016; Rodnina et al., 2020).

In *stop codon readthrough* (**Figure 1**), the termination codon is decoded by a tRNA rather than a release factor, allowing ribosomes to synthesize an extended polypeptide. In specific genes, this tRNA carries the unusual amino acids selenocysteine (Hatfield and Gladyshev, 2002) or pyrrolysine (Namy et al., 2004), and specific stimulatory elements downstream to the stop codon regulate this process (Bertram et al., 2001). In PRF (**Figure 1**), ribosomes are induced to switch, upward or backward for +1 and –1 PRF, respectively, to an alternative, overlapping reading frame at a specific shift site (Farabaugh, 1996; Atkins et al., 2016). This is a regulated process and its frequency depends by genes and on the presence of stimulatory signals in the mRNA. PRF has been detected in organisms from all three domains of life, but it is very common in viruses (Baranov et al., 2006; Firth and Brierley, 2012), in which several recoding events have been described and characterized. R*ibosome hopping* (**Figure 1**) is a rarer recoding event in which the ribosome stops in a precise site of the mRNA and re-starts translation downstream bypassing few nucleotides. This mechanism has been discovered and studied in detail in the gene 60 of bacteriophage T4 (Herr et al., 2004). Ribosomal bypass occurs at *hop* elements where the ribosome block at the "take-off codon," immediately upstream of a stop codon followed by a hairpin, determining the dissociation of the peptidyl-tRNA which re-associates at the "landing triplet," 50 nt downstream, where the translation resumes. More recently, several bypassing elements (*byps*) have been reported in *Magnusiomyces capitatus* mitochondria, suggesting that hopping is more frequent than previously thought (Lang et al., 2014; Nosek et al., 2015). An updated list of genes regulated by recoding can be found in the Recode2 database ([1]Bekaert et al., 2010).

In recent years, bioinformatic analyses of sequenced genomes available in databases have allowed the identification of numerous interrupted genes that could be potential candidates for genes whose expression is regulated by recoding (van Passel et al., 2007; Cobucci-Ponzano et al., 2010; Sharma et al., 2011). However, to date, most of these have been identified serendipitously. A huge boost to the study of non-canonical translation mechanisms came from the development of ribosome profiling, or ribo-seq, a technique that provides genome-wide information on protein synthesis (GWIPS) *in vivo* (Ingolia et al., 2009). Ribosome profiling is based on the deep sequencing of ribosome-protected mRNA fragments and the high resolution of this technique allows the determination of ribosome density along individual cellular mRNA molecules. The real power of ribosome profiling is in its ability to obtain position-specific information regarding ribosome locations on mRNAs, allowing the identification of unpredictable non-canonical translation events. Since its invention, the ribosome profiling technique has been applied in a range of studies in both prokaryotic and eukaryotic organisms, but only one analysis is reported in Archaea (Michel and Baranov, 2013; Brar and Weissman, 2015; Gelsinger et al., 2020).

In Archaea, non-canonical translation events have been demonstrated only during the elongation and the termination steps. In particular, termination codon readthrough events regulating the incorporation of the amino acids selenocysteine and pyrrolysine (Nicholas et al., 2018; Rother and Quitzke, 2018), and –1 PRF allowing the expression of a functional α-L-fucosidase (Cobucci-Ponzano et al., 2003a,b, 2005a,b, 2006, 2012). More recently, –1 PRF was also reported in siphoviruses tailed virus 1 (HVTV-1) and three viruses (HCTV-1,2 and 5) that infect halophilic archaea (Pietila et al., 2013; Sencilo et al., 2013). Increasing evidence suggests that the flexibility of genetic code decoding is a trait selected during evolution to benefit microorganisms under certain physiological conditions, increasing their fitness (Ling et al., 2015). This could be particularly relevant for Archaea, often inhabiting extreme environments in which changes in nutrients, pH, temperatures, etc. are rather common and occur rapidly and reversibly, and may expose microbes to the necessity to modify reversibly gene expression through quick mechanisms (Iacono et al., 2020; Onofri et al., 2020; Strazzulli et al., 2020). Here, we summarize

[1]http://recode.ucc.ie

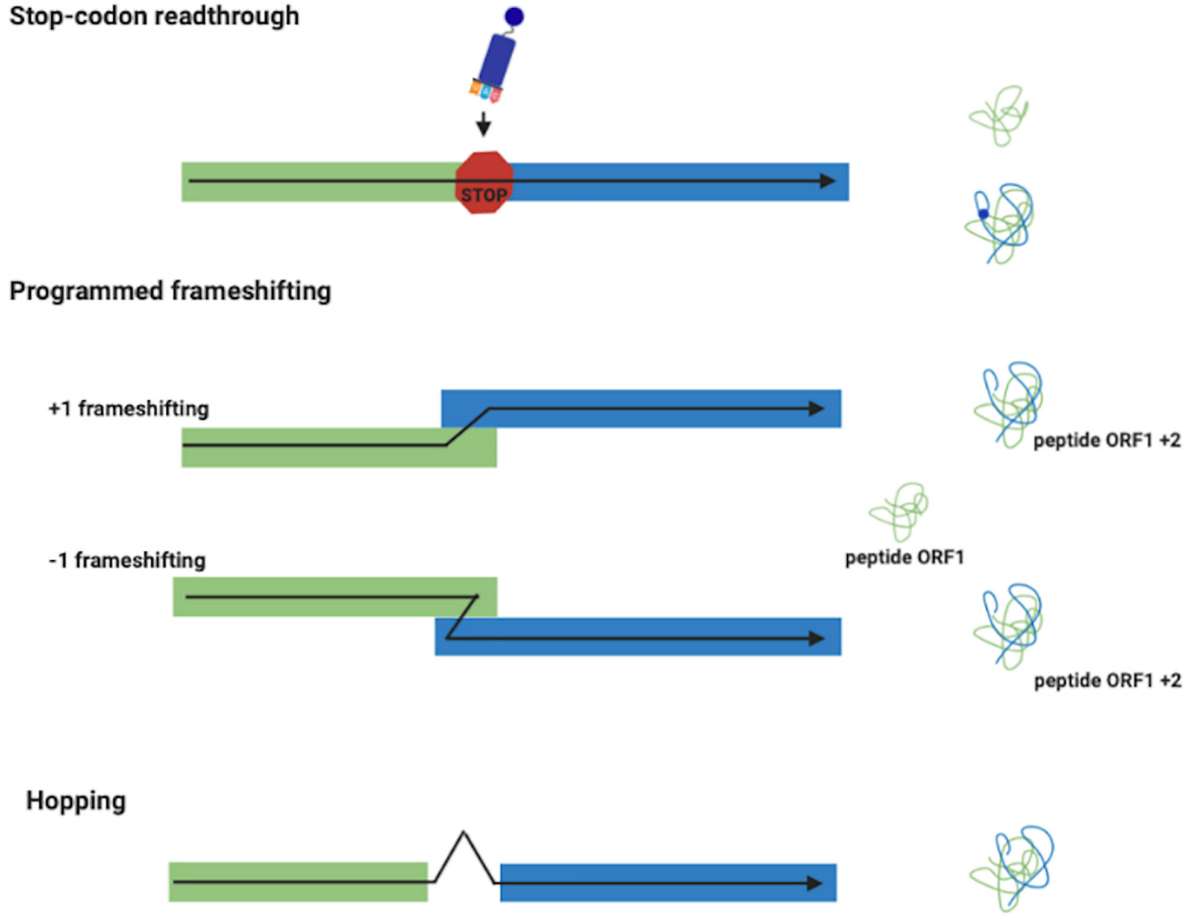


FIGURE 1 | Recoding events. Stop-codon readthrough: a different meaning is assigned to a stop codon with the insertion of the unusual amino acids selenocysteine and pyrrolysine. Frameshifting (+1 and −1): produces two polypeptides from different reading frames of the same mRNA. Ribosome Hopping: synthesizes one protein from two open discontinuous reading frames.

the current state of the art on the studies on the mechanisms of translational recoding found in Archaea, often living in extreme conditions, to provide an update of this interesting and relatively unknown mechanism of regulation of gene expression in the third domain of life.

STOP CODON READTHROUGH

In stop codon readthrough it is important to distinguish between two different mechanisms: 'reassignment' and 'recoding' (Atkins and Baranov, 2010). In codon reassignment, occurring for example in certain mitochondria (Barrell et al., 1979; Osawa et al., 1992), the meaning of particular codons is always reassigned. That codon has only the new meaning and this redefinition is context-independent. These reassignments mainly involve UAG or UGA codons encoding an amino acid instead of a termination signal. Instead, in context-dependent codon redefinition, such event only applies to particular stop codons. Stop codon readthrough is dynamic, with the new definition competing with the standard one, so only a part of the product reflects the new meaning. When it occurs, this redefinition mechanism is a recoding event (**Figure 1**) in which UAG or UGA specify for the amino acids selenocysteine and pyrrolysine, respectively.

The 21st Amino Acid: Selenocysteine

The twenty-first amino acid selenocysteine (Sec) contains selenium, an essential micronutrient for many organisms, and is translationally incorporated into proteins in Bacteria, Eukarya and Archaea (for a comprehensive review see Ambrogelly et al., 2007). Sec does not have a fully dedicated codon, but it is inserted in response to the UGA stop codons that are recoded in the presence of specific regulation signals *in cis*. When translating ribosomes encounter an UGA stop codon in the presence of regulative signals, they are loaded with a specific Sec-tRNA, promoting the insertion of a Sec residue in this location. In fact, in response to those signals, a Sec-specific elongation factor (SelB) replaces the standard EF-Tu uniquely for the translation of Sec UGA codons and recruits the specific Sec-tRNA (see below for the description of the mechanism of insertion). In bacteria, bSelB is homologous in the N-terminal part to the standard elongation

factor ET-Tu, while it has a C-terminal extension responsible for binding to SECIS elements. In contrast to that, the C-terminal extension of the archaeal aSelB is shorter and unrelated to that of bacteria and these structural features are conserved in the eukaryotic homolog eSelB (Kromayer et al., 1996; Fagegaltier et al., 2000; Tujebajeva et al., 2000; Yoshizawa et al., 2005).

This structural difference is most likely the cause of the lack of binding of aSelB to a cognate SECIS element *in vitro* (Rother et al., 2000).

The presence of Sec as selenium carrier in natural proteins, called selenoproteins, was first demonstrated in clostridial glycine reductase (Cone et al., 1976). Sec was then found in enzymes maintaining cell redox balance defending the cell against reactive oxygen species. In humans, the selenoproteome comprises 25 members, whose biological functions have been implicated in diverse human diseases ranging from cardiovascular and endocrine disorders to abnormalities in immune responses and cancer (Bellinger et al., 2009).

Selenoproteins are often enzymes with oxidoreductase function in which Sec is the catalytic redox active site. Homologs proteins in which Sec is replaced with cysteine (Cys) exist for the great majority of selenoproteins, although they perform the same reaction less efficiently (Fomenko and Gladyshev, 2012). It is generally accepted that Sec is used in place of Cys due to its higher reactivity, which leads to improved catalytic efficiency, although the exchangeability of Sec and Cys is debated (Gromer et al., 2003; Castellano, 2009; Hondal and Ruggles, 2011; Hondal et al., 2013). The fact that the Sec-containing proteins are more active if compared to the Cys-containing versions was elegantly demonstrated by inactivating the Sec-specific elongation factor SelB in *M. maripaludis* JJ and observing that this led to overexpression of Cys-containing versions of selenoproteins (Rother et al., 2003). Selenoproteins are not present in all organisms but their distribution is scattered among all the three domains of life in which, however, they perform different functions (Mariotti et al., 2015). In Bacteria, selenoproteins are involved in redox homeostasis, electron transport/energy metabolism, compound detoxification, and oxidative protein folding. In contrast, in Archaea they are involved in methanogenesis, with the only exception of selenophosphate synthetase (SPS), involved in Sec biosynthesis (Stock and Rother, 2009; Rother and Krzycki, 2010). In Eukarya, selenoproteins are mainly involved in redox regulation, antioxidant defense, protein repair, and oxidative protein folding, with very few examples involved in compound detoxification, electron transport, and energy metabolism (Labunskyy et al., 2014). However, Bacteria and Archaea share a larger number of selenoprotein families if compared to Eukarya (Mariotti et al., 2016).

In Archaea, Bacteria and Eukarya, Sec is synthesized in a tRNA-bound fashion, although the mechanisms of Sec synthesis and insertion show differences in the three domains of life (**Figure 2**). While archaea and eukaryotes first catalyze the synthesis of phospho-Ser with the protein phosphoseryl-tRNASec kinase (PSTK), and then convert it to Sec, bacteria directly synthesize Sec from Ser (For a review see Rother and Quitzke, 2018).

The insertion of Sec is driven by specific signals found in the selenoprotein gene transcripts in *cis*. These signals are RNA structures, named SECIS (SElenoCysteine Insertion Sequence) elements (Berry et al., 1991; **Figure 3**). In response to those signals, the specific elongation factor SelB replaces the standard EF-Tu and recruits the Sec-tRNA, promoting the insertion of Sec residues in a specific

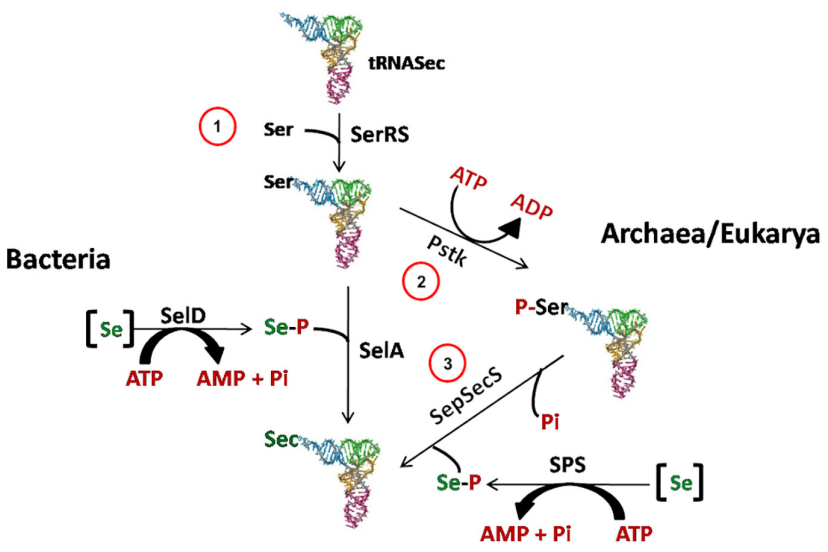


FIGURE 2 | Sec biosynthesis in the three domains of life. In Archaea, as well as in Eukarya, Sec is synthesized in three steps. First (1), SerRS acylates tRNASec with serine to generate Ser-tRNASec. Then (2) PSTK forms Sep-tRNASec, which is converted to Sec-tRNASec by SepSecS in the presence of selenophosphate produced by selenophosphate synthetase (SPS) (3). -[Se]: reduced Se species; -SerRS: seryl-tRNA synthetase; -SelD/SPS: selenophosphate synthetase; -SelA: bacterial Sec synthase; -PstK: seryl-tRNASec kinase; -SepSecS: O-phosphoseryl-tRNA:selenocysteyl-tRNA synthase.

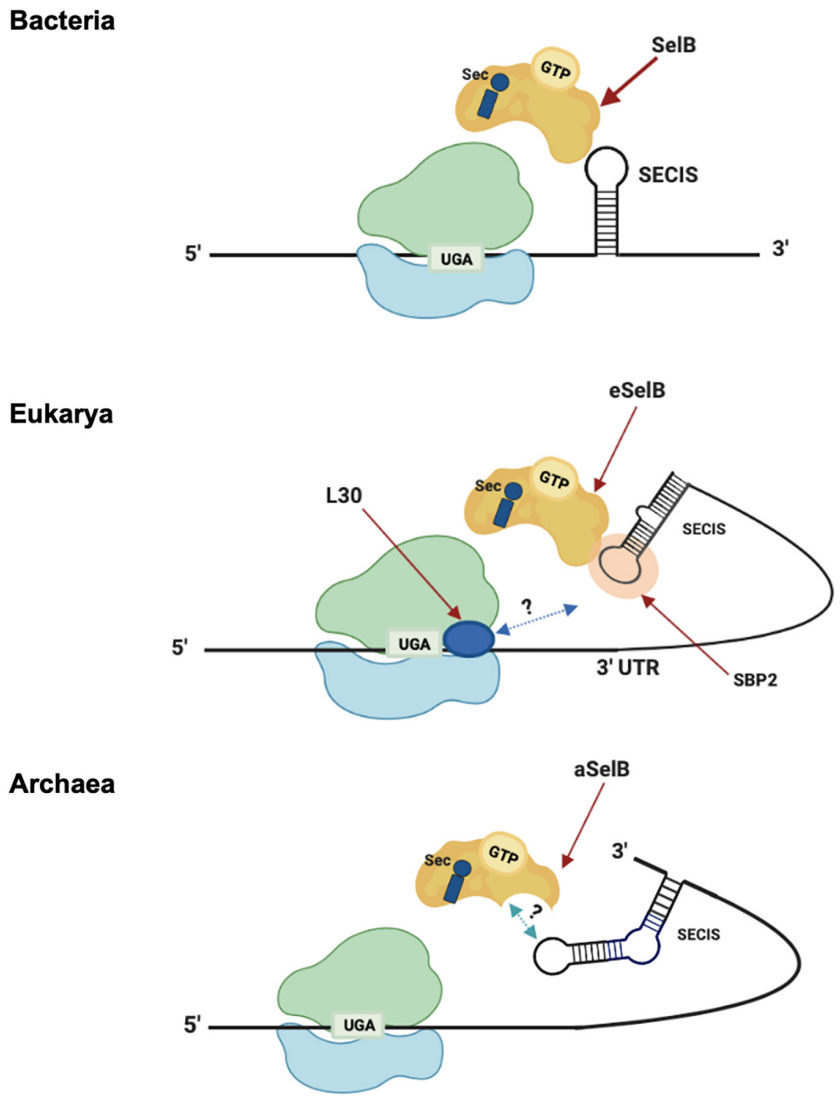


FIGURE 3 | Sec translation in the three domains. Model of Sec incorporation in Bacteria (top), Eukarya (middle), and Archaea (bottom). -3′-UTR: 3′-untranslated region; -L30: ribosomal protein L30; -SBP2: SECIS-binding protein 2; -SECIS: Sec insertion sequence; -SelB/aSelB/eSelB: Sec-specific elongation factor.

UGA (Hatfield and Gladyshev, 2002; Mariotti et al., 2016). Interestingly, SECIS elements do not share similarity in sequence or structure between the three domains of life (Krol, 2002). In bacteria, the SECIS element (bSECIS) is a stem–loop structure located within the coding sequence, immediately downstream of the recoded UGA. The bSECIS is bound directly by the elongation factor bSelB through its C-terminal extension (see above) (**Figure 3**). The eukaryotic SECIS elements are, instead, located in the 3′ UTR of selenoprotein transcripts and they do not interact directly with eSelB, but though the SECIS Binding Protein 2 SBP2 (Tujebajeva et al., 2000; Fletcher et al., 2001). In addition, it has been found that other factors are involved in eukaryal Sec insertion, as the ribosomal protein L30 (Chavatte et al., 2005).

The archaeal versions of SECIS (aSECIS) are characterized by two stems separated by an invariant asymmetric bulge (Krol, 2002; Kryukov and Gladyshev, 2004; Stock and Rother, 2009) and are normally located in the 3′ UTR of selenoprotein coding mRNA, with a single documented exception (Wilting et al., 1997). To date, no aSECIS binding factors have been identified. The SBP2 homolog has never been observed in archaea, and it has been shown that the archaeal SelB does not bind aSECIS elements (Mariotti et al., 2016). Thus, it has been proposed that the eukaryal Sec decoding mechanism, in which SBP2 is a key factor, evolved after the transition from archaeal to eukaryotic-like SECIS elements (Stock and Rother, 2009). From an evolutionary point of view, the distribution of selenoproteins in living organisms is consistent with the phylogenetic relationship between the organisms in the three domains of life (Mariotti et al., 2015). In addition, considering the clear homology between the key factors involved in the Sec pathway (tRNAsec, SelB, and the selenophosphate synthetase

SPS/SelD) (Santesmasses et al., 2017), it was highlighted that it is very likely that this pathway originated only once in the history of life and was already present in the Last Universal Common Ancestor (LUCA) (Mariotti et al., 2016). However, the presence of this pathway in different living organisms appears to be very dynamic, showing both clear events of horizontal gene transfer and independent loss in many lineages (Zhang et al., 2006; Lobanov et al., 2008; Mariotti et al., 2015).

Selenoproteins are a quite rare feature among the Archaea. Sec was found in formate dehydrogenase, formylmethanofuran dehydrogenase, F_{420} reducing and non-reducing hydrogenases, HesB-like protein and heterodisulfide reductases (Kryukov and Gladyshev, 2004; Stock et al., 2010). For a detailed list of putative and known archaeal selenoproteins and their properties see Rother and Quitzke (2018). Interestingly, genes encoding selenoproteins, belonging to different families, and the full set of genes encoding for the key factors involved in the Sec pathway, have been found in *Lokiarchaeota* (Spang et al., 2015), considered the closest cultured archaeal relative of eukaryotes (Mariotti et al., 2016). The selenoprotein families identified in *Lokiarchaeota* were previously reported in other archaeal lineages (Stock and Rother, 2009), with the exception of the thioredoxin-like superfamily, found by bioinformatic analysis, both in bacteria (Zhang and Gladyshev, 2008) and eukaryotes (Jiang et al., 2012; Mariotti et al., 2013). Moreover, although the selenoprotein genes in Lokiarchaeota are typical of archaea, they possess conserved RNA structures similar to eukaryotic SECIS elements. This finding is the basis of a new theory proposing that eukaryotes have not reinvented the mechanism of insertion of the Sec as previously proposed, but rather that the Sec pathway has passed vertically from Archaea to Eukarya (Rother and Quitzke, 2018).

The 22nd Amino Acid: Pyrrolysine

Pyrrolysine (Pyl) was identified in 2002 as the 22nd proteinogenic amino acid (Hao et al., 2002; Srinivasan et al., 2002). From a biochemical perspective, Pyl is a typical L-lysine amino acid to which a pyrrole ring is branched on the lateral chain through an amide bond. This chemical modification is different from those present in other L-lysine derivatives found in some proteins from archaea like hypusine or methyllysine (Eichler and Adams, 2005). In fact, while in hypusine and methyllysine the modifications originate from post-translational events, Pyl is translationally incorporated (for a review see Brugère et al., 2018). This unusual and highly specialized amino acid is found in a small number of archaea able to metabolize methylamine as well as a few bacteria. The first hint of the presence of pyrrolysine (Pyl) has been reported in several Methanosarcina species with a total of 21 genes of mono, di-, and trimethylamine methyltransferases (MtmB, MtbB, and MttB, respectively) showing an in-frame amber UAG codon (James et al., 2001). Initially, the amino acid inserted into the UAG codon was identified as a lysine. Later, the three-dimensional structure resolution of the enzyme MtmB allowed to demonstrate that the amino acid was a Pyl. Furthermore, the identification of a specific tRNA for Pyl confirmed the hypothesis that Pyl is inserted into proteins during translation by a mechanism of recoding (Hao et al., 2002; Srinivasan et al., 2002). From these preliminary discoveries, several new pieces of information have been collected that have allowed to define the key factors involved in the biosynthesis and insertion of Pyl, the molecular mechanism underlying this recoding mechanism, its distribution and evolution, and the catalytic role of this amino acid.

The five Pyl genes involved in the biosynthesis and genetic encoding of Pyl are pylTSBCD (Srinivasan et al., 2002; Krzycki, 2004; Zhang et al., 2005; Longstaff et al., 2007) and, in most cases, they are organized in an operon-like structure as shown in **Figure 4A**. The Pyl genes have been found in bacterial and archaeal genomes and are usually clustered near the genes encoding the methylamine methyltransferases and other genes involved in methylamine metabolism (for a detailed description of the genomic contexts of Pyl-related genes see Gaston et al., 2011).

Initially, it had been proposed that the synthesis of Pyl took place from the lysyl-tRNA-Pyl (Srinivasan et al., 2002; Polycarpo et al., 2003), similarly to how it occurs for the synthesis of the Sec starting from the seryl-tRNA (**Figure 2**; Yoshizawa and Bock, 2009; Rother and Krzycki, 2010). However, it is now well documented that Pyl is synthesized by the enzymes PylB, PylC, and PylD from two equivalents of lysine. The two other genes of the Pyl system, *pylT* and *pylS*, encode, respectively, for the tRNAPyl, whose anticodon is complementary to the UAG codon, and the subunit of the tRNAPyl synthetase which directly esterified Pyl to the 3′-hydroxyl of tRNAPyl, clearly demonstrating that Pyl is made independently of tRNAPyl (**Figure 4A**; Blight et al., 2004; Polycarpo et al., 2004; Nozawa et al., 2009; Gaston et al., 2011; Tharp et al., 2018). The complete pathway of biosynthesis of Pyl is reported in **Figure 4B** (for a review see Brugère et al., 2018).

Although possible sequences that regulate Pyl (named PYLIS by analogy to the SECIS sequences, see above) were initially postulated (Namy et al., 2007), bioinformatic (Zhang et al., 2005) and biochemical studies have shown that no *cis* element is found or required in *E. coli* for the recoding of the UAG stop codon into Pyl (Longstaff et al., 2007; Namy et al., 2007). It follows that there is no specific context in the mRNA driving the recoding event, therefore it was proposed that Pyl insertion relies only on the competition between release factors and the tRNA-Pyl during translation. However, how the cell prevents all stop codons from being recoded is still to be elucidated, especially considering that *cis* signals have not been found. Interestingly, it has been reported that in the clostridial *Acetohalobium arabaticum*, UAG specifies Pyl only when the cells are grown in trimethylamine, while, when the cells are grown on pyruvate as a carbon source, UAG only specifies termination (Prat et al., 2012). Thus, this result indicates that Pyl insertion is regulated in specific physiological conditions and could suggest the presence of a *trans*-acting regulation factor expressed only in particular conditions which must still be identified.

Pyl is found in all methanogen methylamine methyltransferase genes and in some cases the readthrough efficiency of the UAG codon is as high as 97%. In these enzymes Pyl is always present in the active site, capturing methylamines before transferring one methyl group to a Co(I)-corrinoid cofactor of an associated protein (MtmC/MtbC/MttC) (Hao et al., 2002), suggesting that

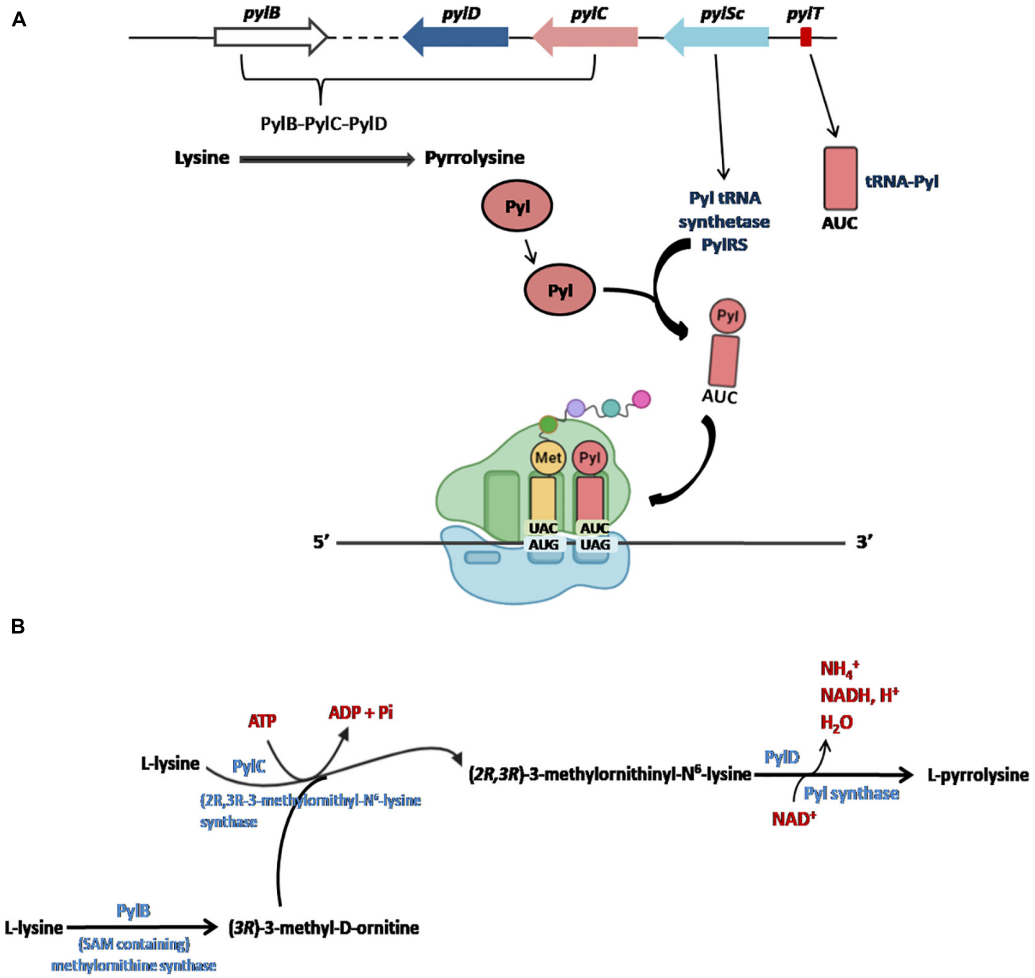


FIGURE 4 | The Pyl insertion system **(A)**. Pyl, synthesized by pylB, pylD, pylC, is charged on a specific tRNA (encoded by pylT) whose anticodon AUC recognizes UAG codons in a specific reaction catalyzed by PylRS (encoded PylSc). See text for details. Figures arranged from Brugère et al. (2018). Biosynthesis of Pyl **(B)**. The complete biosynthesis pathway of L-pyrrolysine from two lysines catalyzed by PylB, PylC and PylD.

its role is fundamental for methylamine metabolism. More recently, it has been reported that natural MttB analogs without Pyl found in *Desulfitobacterium hafniense* has a glycine betaine methyltransferase activity (Ticak et al., 2014), confirming that methyltransferases containing Pyl are related to methylamines metabolism. The only other known Pyl-containing proteins are some transposases (Zhang et al., 2005), and a tRNAHis-guanylyltransferase Thg1 (Heinemann et al., 2009) both present in a subset of Methanosarcinales.

In archaea, pyl genes were initially identified in anaerobic methanogens living in environments where methylamines are available, namely, several Methanosarcinales (Deppenmeier et al., 2002; Galagan et al., 2002; Maeder et al., 2006), in *Methanococcus burtonii* (psychrophile) (Goodchild et al., 2004), and in *Methanoalophilus mahii and Methanohalobium evestigatum* (halophiles) (Rother and Krzycki, 2010; Gaston et al., 2011). More recently, the genes for the synthesis and encoding of Pyl were identified in several new lineages of methanogens, discovered by metagenomic approaches and distantly related to those mentioned above, in which the methanogenesis is dependent on methyl-compounds (Borrel et al., 2013; Evans et al., 2015; Petitjean et al., 2015; Nobu et al., 2016; Vanwonterghem et al., 2016; Sorokin et al., 2017). Pyl-containing methyltransferases needed for methylamine utilization are always present in these new lineages of methanogens that contain the Pyl system, strengthening the hypothesis that the Pyl system is dedicated to the incorporation of Pyl in these methyltransferases, and thus associated to methylamine utilization. Methanohalophilus, in which the Pyl-containing methyltransferases are absent (Fricke et al., 2006) are also lacking the Pyl system, suggesting that this recoding mechanism is linked to methylamine methyltransferases rather than to archaea performing methanogenesis based on methyl-compounds. In addition, it has been found that uncultured sugar-fermenters of the candidate division of Persephonarchaea, thriving in a hypersaline environment, harbor a complete set of genes for Pyl synthesis and *mtmB*, *mtbB*, and *mttB* genes (Guan et al., 2017). The components of the Pyl system in these archaea

are phylogenetically related to those found in the bacteria *Acetohalobium arabaticum* who lives in the same environment, suggesting an event of horizontal gene transfer between these organisms (Guan et al., 2017). From its first discovery great advances have been made in understanding the role of this recoding event in archaea and allowing us to reveal that Pyl-system has a wide distribution and is not necessarily associated with methanogenesis in this domain of life (Brugère et al., 2018).

There are several hypotheses for the emergence of the Pyl system in living organisms. Among the others, one of the most recent, and strongly supported by current data, postulated that the Pyl trait is very ancient and probably only emerged once after LUCA and was linked to methanogenesis. The trait could have then evolved and preserved in organisms for which methylamine metabolism was fundamental to survive and could have been further spread across the bacterial and archaeal domains by horizontal gene transfer (Brugère et al., 2018).

PROGRAMMED RIBOSOMAL FRAMESHIFTING

During standard mRNA translation the ribosome initiates protein synthesis at a start codon and moves by decoding three nucleotides at a time until it reaches a stop codon where translation is terminated. However, in some cases the ribosomes switch to an alternative reading frame on the mRNA by determining a translational slippage in the +1 or −1 direction (Farabaugh, 1996; Gesteland and Atkins, 1996; **Figure 1**). In contrast to spontaneous frameshifting, which produces non-functional polypeptides, PRF is generally in competition with standard decoding and typically leads to the synthesis of a functional polypeptide from an alternative frame with efficiencies varying from very low to as high as 80% (Tsuchihashi and Brown, 1992; Atkins et al., 2009). At the functional level there are two more common classes of regulation of PRF. In a first class, often termed 'set ratio' frameshifting, the proportion of ribosomes that shift frame is constant, thereby generating an extra N-terminally coincident product. In a second class, frameshift efficiency is dependent by the level of translation initiation or responsive to a trans-acting factor. Here, frameshifting, acting as a sensor and/or effector has a regulatory function, allows the synthesis of a functional trans-frame encoded product or alters mRNA half-life (Atkins et al., 2016). It has been well demonstrated that in eukaryotes, PRF can regulate the stability of an mRNA. In fact, it has been seen that following a PRF event, the ribosomes encounter a stop codon in the new reading frame that activates the nonsense-mediated decay pathway (Belew et al., 2014).

The PRF has been studied extensively in viruses, where −1 PRF plays an important role in viral propagation by modulating synthesis of viral proteins in specific stoichiometric ratios (Jacks and Varmus. 1985; Plant et al., 2010). The use of a −1 PRF mechanism for the expression of a viral gene was first identified in the Rous sarcoma virus (Jacks and Varmus, 1985). To date, it is well known that, for example, all coronaviruses utilize −1 PRF to control the relative expression of their proteins. In general, the early translated viral proteins are involved in neutralizing the host cellular immune response (ORF1a) and in genome replication and RNA synthesis (ORF1b). ORF1b is in the −1 reading frame with respect to ORF1a, and all coronaviruses, as well as SARS-CoV-2, utilize −1 PRF as a mean to synthesize the ORF2 encoded proteins (Kelly et al., 2020). PRF is well documented in retrotransposons and insertion elements too, while it is less common in cellular genes. Among the chromosomal genes, the best studied examples are the Antizyme (Matsufuji et al., 1995) in which + 1 PRF frameshifting functions both as a sensor of the polyamine levels and as an effector of a self-regulating circuit from yeasts to mammals. In the bacterial DNA polymerase, γ and τ subunits are produced in 1:1 molar ratio by −1 PRF from *dnaX* gene (Tsuchihashi and Kornberg, 1990; Mangold, 2005; Chen et al., 2014). For a comprehensive review on the genes expressed by PRF in Bacteria, Eukarya and viruses see Atkins et al. (2016); Rodnina et al. (2020). Among PRF, −1 frameshifting is more widespread with examples in all three domains of life (Luthi et al., 1990; Tsuchihashi and Kornberg, 1990; Cobucci-Ponzano et al., 2006; Wills et al., 2006; Belew et al., 2014), many of which are phylogenetically conserved.

As stated above, −1 PRF is generally in competition with standard decoding but it is facilitated by two regulatory elements in the mRNA sequence, a slippery site, where the transition to the −1 frame takes place, and a secondary structure element (a pseudoknot, a steam and loop or a kissing loop) at a defined distance of 5 to 9 nucleotides from the slippery site (Brierley et al., 1992, 2010; Atkinson et al., 1997; Lin et al., 2012; Choi et al., 2020). The slippery site, usually in the form of a heptanucleotide sequence X-XXY-YYZ, in which X can be any base, Y is usually A or U, and Z is any base but G (codons are shown in the 0 reading frame), allows for base pairing between the tRNA anticodon and the mRNA codon after shifting into the −1 reading frame. Prokaryotic frameshifting sites may contain additional stimulatory elements, such as an internal Shine-Dalgarno (SD)-like sequence upstream of the slippery site (Larsen et al., 1997; Choi et al., 2020) or tandem rare codons (Caliskan et al., 2017) both with the function of slowing down the translating ribosome and increasing the frameshifting efficiency. −1 PRF can be also facilitated by miRNAs binding as reported in the human mRNA encoding the HIV-1 co-receptor CCR5 (Belew et al., 2014), or proteins, as reported in some viruses (Kobayashi et al., 2010; Napthine et al., 2017; Wang et al., 2019) to the sequence following the slippery site.

Detailed studies on the molecular mechanism by which −1 PRF occurs have only recently been reported. These studies suggest that the molecular mechanisms are mainly two and depend on the availability of the aa-tRNAs of the codons in the slippery sequence (Namy et al., 2006; Chen et al., 2013, 2014; Caliskan et al., 2014, 2017; Kim et al., 2014; Yan et al., 2015; Korniy et al., 2019a,b). When the tRNAs reading the slippery sequence codons are abundant, −1 PRF occurs at the late stage of translocation, with two tRNAs moving through the ribosome, and requires the presence of the stimulatory element within the mRNA sequence. By contrast, in conditions in which aa-tRNAs are limited, the −1 PRF occurs via one-tRNA slippage of the P-site tRNA, when the A site is vacant, and its efficiency is independent of the stimulatory element within the mRNA sequence. This latter

mechanism is often called "hungry" frameshifting, because it can be triggered by aa-tRNA limitation due to starvation (Gallant and Lindsley, 1992; Olubajo and Taylor, 2005; Temperley et al., 2010) (see below).

In Archaea only one case of –1 PRF has been reported (Cobucci-Ponzano et al., 2006). In the thermoacidophilic archaeon *Saccharolobus solfataricus* (formerly *Sulfolobus solfataricus*) (Sakai and Kurosawa, 2018) strain P2 the *fucA1* gene was found to be organized in two open reading frames (ORFs) SSO11867 and SSO3060 of 81 and 426 amino acids, respectively, which are separated by a –1 frameshifting in a 40 bases overlap. These ORFs encode, respectively, for the N- and C-terminal part of a α-*L*-fucosidase. The overlap region between the two ORFs had the characteristic features of the genes expressed by –1 PRF, including a heptanucleotide A-AAA-AAT (codons are shown in the zero frame), flanked by a putative stem and loop and the tandem rare codons CAC (**Figure 5**). To test if these gene fragments could lead to a functional enzyme, a full-length gene, named *framefucA*, was produced by inserting specific site-directed mutations in the *fucA1* gene, exactly in the position predicted by –1 PRF. In this way, the poly-A sequence of the *slippery* site was disrupted and a T nucleotide was inserted to restore a single reading frame between the two ORFs (Cobucci-Ponzano et al., 2003a). The *framefucA* mutant encoded for a polypeptide of 495 amino acids, that, remarkably, in recombinant form produced a fully functional α-*L*-fucosidase, named Ssα-fuc, which was thermophilic, thermostable and had an unusual non-americ structure (Cobucci-Ponzano et al., 2003b, 2005a; Rosano et al., 2004). The full-length protein FucA was expressed by –1 PRF in both *E. coli* and *S. solfataricus* showing for the first time that this kind of recoding is present in Archaea (Cobucci-Ponzano et al., 2006). The observation that the *fucA1* interrupted gene directed the expression of low α-*L*-fucosidase activity in *E. coli* led to the isolation and characterization of the polypeptides expressed in the recombinant form demonstrating that the *fucA1* gene produced in *E. coli* a mixture of two full-length polypeptides, both functional, with a total efficiency of about 5% (Xu et al., 2004; Cobucci-Ponzano et al., 2006). The identification of these polypeptides indicated that the translational recoding of *fucA1* might occur in two ways, at least in *E. coli* (**Figure 6**): a simultaneous backward slippage of the ribosome when both the P- and the A-site tRNAs are

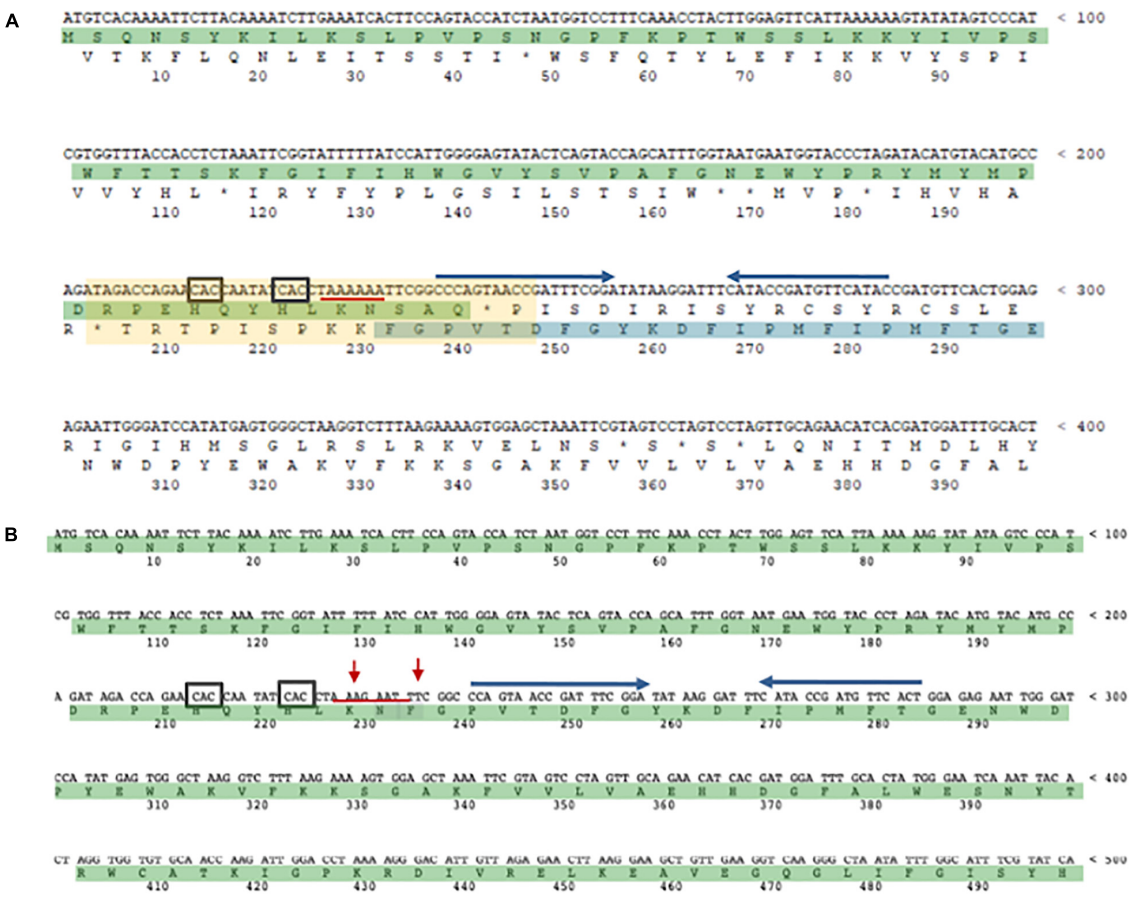


FIGURE 5 | The α-L-fucosidase gene. **(A)** The N-terminal SSO11867 ORF (highlighted in green) is in the zero frame, the C-terminal SSO3060 ORF (highlighted in blue), for which only a fragment is shown, is in the –1 frame. The 40 bp region of overlap bertween the two ORFs is indicated with a light yellow rectangle. The slippery heptameric sequence is underlined with a red line. The rare codons CAC are indicated with a black square. The putative stem and loop region is indicated with blu arrows. **(B)** *framefucA* mutant gene (only a fragment is shown). The red arrows indicate the mutated nucleotides in the slippery sequence.

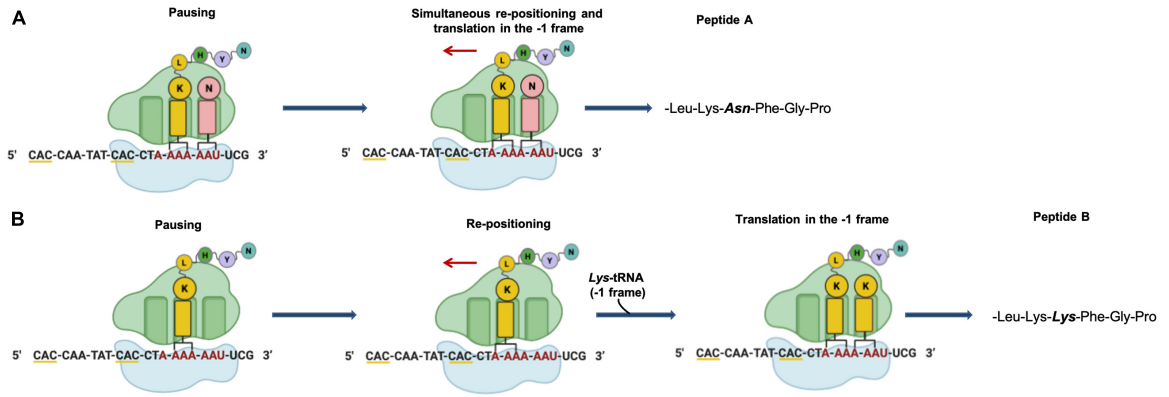


FIGURE 6 | Putative mechanism of programmed –1 frameshifting. **(A)** Simultaneous P- and A-site slippage; **(B)** P-site slippage. The slippery heptameric sequence is indicated in red. Rare codons are underlined with yellow line.

occupied (**Figure 6A**) and/or the repositioning of the ribosome in the –1 frame when only the P-site tRNA is bound (**Figure 6B**) (Cobucci-Ponzano et al., 2006). The analysis of *fucA1* –1 PRF in *S. solfataricus* by *in vitro* translation revealed that only the *wild type* slippery sequence led to the translation of a full-length product with good efficiency (about 10%), demonstrating that this process occurred in archaea (Cobucci-Ponzano et al., 2006). *In vivo*, full-length polypeptides from *fucA1* were identified in *S. solfataricus* extracts, and reverse real-time PCR experiments and specific enzymatic assays confirmed that this enzyme was functionally expressed though at very low levels.

Although these studies produced evidence that –1 PRF is present in archaea, several questions remain unanswered: to date, it is still not known why the translational of *fucA1* in *S. solfataricus* is regulated by recoding, and if other genes are expressed by this mechanism in this or other archaea. However, since there are no α-*L*-fucosidase genes regulated by PRF in Bacteria and Eukarya, it has been suggested that this sophisticated mechanism of translational regulation preexisted in *S. solfataricus* and it was applied to the α-*L*-fucosidase gene for physiological reasons. Very recently, it has been found that *fucA1* mRNA increases by 10 *fold* after *S. solfataricus* undergoes cold shock and in *S. solfataricus* cells grown in minimal medium containing the oligosaccharides of the hemicellulose xyloglucan (De Lise et al., 2021). Furthermore, this α-*L*-fucosidase has been shown to cooperate with other glycoside hydrolases from *S. solfataricus* for the hydrolysis of fucosylated xyloglucan oligosaccharides by removing the fucose moieties from this substrate with high efficiency *in vitro* (Curci et al., 2021). These new results will certainly need to be explored and could be of great help in understanding what the function of this enzyme is *in vivo*, and why its expression is regulated by 1- PRF.

Genomic sequencing showed that the *fucA1* gene was also present in other archaea, all belonging to Crenarchaeota (for the compilation of these genes, see the Carbohydrate Active enZyme database[2]). The α-L-fucosidases from Sulfolobales showed 96% amino acid sequence identity and are all full length with the exception of the *S. solfataricus*, strain 98/2 which presented the frameshifting in the same position as the gene from strain P2. However, all Sulfolobales genes showed 100% DNA sequence identity in the region of the frameshifting, maintaining the rare codon, the slippery sequence, in which the stretch of A is shortened by one nucleotide in full-length genes, and the putative stem loop. On the contrary, the slippery sequence is not conserved in full-length α-*L*-fucosidase homologs from *I. aggregans* and *C. maquilingensis*. Remarkably, full-length α-*L*-fucosidases, in the region of the slippery sequence, have the same Lys or Asn amino acids observed in the full- length product of the wild-type interrupted *fucA* (Cobucci-Ponzano et al., 2006).

More recently, PRF events have been reported in some archaeal viruses. In particular, –1 PRF seems to be used by the siphoviruses tailed virus 1 (HVTV-1) and three viruses (HCTV-1,2 and 5) that infect halophilic archaea, while an event of +1 PRF appears to be present in the haloarchaeal myovirus tailed virus 2 (HSTV-2) (Pietila et al., 2013; Sencilo et al., 2013). In addition, it has been suggested that a frameshifting is presumably involved in the synthesis of magnesium chelatase from the archaea *Methanocaldococcus* and *Methanococcus* (Antonov et al., 2013b). Unfortunately, genes with frameshifts could be difficult to annotate by standard procedures and often might be annotated as two separate adjacent hypothetical genes (Antonov et al., 2013b). In recent years some bioinformatic tools have been developed with the aim of identifying possible genes regulated by frameshifting (Antonov and Borodovsky, 2010; Antonov et al., 2013a). However, none of these have been systematically tested on Archaea and it would be very useful to know whether the parameters used allow to identify possible genes regulated by frameshifting in this domain of life.

CONCLUSION

The identification of novel genes whose expression could be regulated translational recoding is not easy, either because

[2]http://www.cazy.org/

disrupted genes are commonly considered non-functional pseudogenes or because technical limitations, and this is particularly true for Archaea for which molecular biology tools are still to be completely developed. Non-functional pseudogenes are present in organisms from all the living domains, though in some cases they have been demonstrated to be useful for an organism's survival and adaptation to particular environmental changes (Harrison and Gerstein, 2002; Balakirev and Ayala, 2003; Hirotsune et al., 2003). In Archaea, 15 different species have been bioinformatically analyzed revealing a high number of predicted pseudogenes, the highest of which (8.6% of the annotated protein coding sequences) being in *S. solfataricus*. The expression of these genes has not been tested but, remarkably, all the frameshifts occurred in A/T rich DNA tracts resembling the slippery sequences regulating –1 PRF in *cis* (van Passel et al., 2007). In addition, a different bioinformatic analysis of other 16 Archaea genomes, allowed to identify a large number of disrupted genes, some of which resulted to be functional, as demonstrated by a high throughput proteomic analysis and functional characterization of some of them from *S. solfataricus* strain P2 (Cobucci-Ponzano et al., 2010). Interestingly one of the interrupted gene whose expression could be regulated by –1 PRF is the putative universal translation initiation factor SUI-1/aIF1 (Cobucci-Ponzano et al., 2010). This protein is essential in yeast forming the translation initiation complex and monitoring the maintenance of the correct translational reading frame in eukaryotes, such as it was suggested that it might govern programmed –1 frameshifting as a trans-acting factor (Cui et al., 1998; Kyrpides and Woese, 1998). Similarly, *in vitro* experiments performed with *S. solfataricus* cell fractions showed that aIF1 promotes translation complex binding to the ribosome, promoting discrimination against non-canonical start codons and enhancing translation efficiency (Hasenöhrl et al., 2006 RNA.; 12: 674–682.; Hasenöhrl et al., 2009 RNA.; 15: 2288–2298.) In the genome annotation of *S. solfataricus,* P2 strain, this gene is reported as interrupted by –1 frameshfting, but, once re-sequenced, it was found to be full-length, suggesting a possible sequencing error (Hasenöhrl et al., 2006; Cobucci-Ponzano et al., 2010). However, a high-throughput proteomic analysis revealed the presence of two peptides, one deriving from the full-length gene and the other one deriving from the translation of the annotated interrupted gene by –1 PRF (Cobucci-Ponzano et al., 2010). These data merit further investigation and could be of some help to elucidate the possible mechanism of expression of this gene in *S. solfataricus* and to shed some light of its role *in vivo*.

It has been suggested that the flexibility of the genetic code decoding, typical of recoding mechanisms, is a trait selected during evolution that may increase microbial fitness under certain conditions (Ling et al., 2015). The majority of Archaea populate extreme environments, which are often spots (e.g., hydrothermal vents, solfataras, etc.) surrounded by environments with milder conditions and frequently subjected to sudden changes that greatly, and temporarily, modify the chemical-physical parameters to which microorganisms must adapt. It is tempting to speculate that in these extreme environments translational recoding could be a way to maintain in a latent state the expression of certain genes, and up- or down-regulate them under specific conditions. Another important aspect to be considered is related to the understanding of the molecular mechanisms that lead to the improved fitness as a result of genetic code variation (Ling et al., 2015). This fostered a new research area in engineering synthetic organisms with new genetic codes and non-canonical amino acids (for a review see Hoffman et al., 2018). These engineered synthetic organisms will be very important to study the physiological effect of genetic code evolution (Ling et al., 2015). Thus, the study of translational recoding in Archaea is particularly important for its possible implications in the evolution of the genetic code and the correlation between the flexibility of the genetic code decoding and improved fitness in extreme environments.

AUTHOR CONTRIBUTIONS

FDL and BC-P wrote the manuscript. AS, RI, NC, LM, MDF, and MM contributed to the article, edited English style, and approved the submitted version. MDF, MM, FDL, and BC-P edited the manuscript into its final format. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Italian Space Agency: "Life In Space (OPPS)" project (ASI N. 2019-3-U.O).

ACKNOWLEDGMENTS

We are grateful to Giovanni del Monaco, Chiara Nobile, and Marco Petruzziello (IBBR-CNR) for administrative assistance.

REFERENCES

Ambrogelly, A., Palioura, S., and Söll, D. (2007). Natural expansion of the genetic code. *Nat. Chem. Biol.* 3, 29–35. doi: 10.1038/nchembio847

Antonov, I., and Borodovsky, M. (2010). Genetack: frameshift identification in protein-coding sequences by the Viterbi algorithm. *J. Bioinform Comput. Biol.* 8, 535–551. doi: 10.1142/s0219720010004847

Antonov, I., Coakley, A., Atkins, J. F., Baranov, P. V., and Borodovsky, M. (2013b). Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic Acids Res.* 41, 6514–6530. doi: 10.1093/nar/gkt274

Antonov, I., Baranov, P., and Borodovsky, M. (2013a). GeneTack database: genes with frameshifts in prokaryotic genomes and eukaryotic mRNA sequences. *Nucleic Acids Res.* 41, D152–D156. doi: 10.1093/nar/gks1062

Atkins, J. F., and Baranov, P. V. (2010). The distinction between recoding and codon reassignment. *Genetics* 185, 1535–1536. doi: 10.1534/genetics.110.11901

Atkins, J. F., Gesteland, R. F., and Atkins, R. F. G. J. F. (2009). *Recoding.* Berlin: Springer.

Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E., and Baranov, P. V. (2016). Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* 44, 7007–7078. doi: 10.1093/nar/gkw530

Atkinson, J., Dodge, M., and Gallant, J. (1997). Secondary structures and starvation-induced frameshifting. *Mol. Microbiol.* 26, 747–753. doi: 10.1046/j.1365-2958.1997.6101959.x

Balakirev, E. S., and Ayala, F. J. (2003). Pseudogenes: are they "Junk" or functional DNA? *Annu. Rev. Genet.* 37, 123–151. doi: 10.1146/annurev.genet.37.040103.103949

Baranov, P. V., Fayet, O., Hendrix, R. W., and Atkins, J. F. (2006). Recoding in bacteriophages and bacterial IS elements. *Trends Genet.* 22, 174–181. doi: 10.1016/j.tig.2006.01.005

Baranov, P. V., Gesteland, R. F., and Atkins, J. F. (2002). Recoding: translational bifurcations in gene expression. *Gene* 286, 187–201. doi: 10.1016/s0378-1119(02)00423-7

Barrell, B. G., Bankier, A. T., and Drouin, J. (1979). A different genetic code in human mitochondria. *Nature* 282, 189–194. doi: 10.1038/282189a0

Bekaert, M., Firth, A. E., Zhang, Y., Gladyshev, V. N., Atkins, J. F., and Baranov, P. V. (2010). Recode-2: new design, new search tools, and many more genes. *Nucleic Acids Res.* 38, 69–74. doi: 10.1093/nar/gkp788

Belew, A., Meskauskas, A., Musalgaonkar, S., Advani, V., Sulima, S., Kasprzak, W., et al. (2014). Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nature* 512, 265–269. doi: 10.1038/nature13429

Bellinger, F. P., Raman, A. V., Reeves, M. A., and Berry, M. J. (2009). Regulation and function of selenoproteins in human disease. *Biochem. J.* 422, 11–22. doi: 10.1042/BJ20090219

Berry, M. J., Banu, L., Chen, Y. Y., Mandel, S. J., Kieffer, J. D., Harney, J. W., et al. (1991). Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3′ untranslated region. *Nature* 353, 273–276. doi: 10.1038/353273a0

Bertram, G., Innes, S., Minella, O., Richardson, J., and Stansfield, I. (2001). Endless possibilities: translation termination and stop codon recognition. *Microbiology* 147, 255–269. doi: 10.1099/00221287-147-2-255

Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., et al. (2004). Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* 431, 333–335. doi: 10.1038/nature02895

Borrel, G., O'Toole, P. W., Harris, H. M., Peyret, P., Brugere, J. F., and Gribaldo, S. (2013). Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis. *Genome Biol. Evol.* 5, 1769–1780. doi: 10.1093/gbe/evt128

Brar, G., and Weissman, J. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* 16, 651–664. doi: 10.1038/nrm4069

Brierley, I., Gilbert, R., Pennell, S., Atkins, J. F., and Gesteland, R. F. (2010). Recoding: expansion of decoding rules enriches gene expression. *Nucleic Acids and Mol. Biol.* 24, 149–174.

Brierley, I., Jenner, A., and Inglis, S. (1992). Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* 227, 463–479. doi: 10.1016/0022-2836(92)90901-U

Brugère, J. F., Atkins, J. F., and Paul, W. O. (2018). Toole, guillaume borrel; pyrrolysine in archaea: a 22nd amino acid encoded through a genetic code expansion. *Emerg. Top. Life Sci.* 2, 607–618. doi: 10.1021/ja2054034

Caliskan, N., Katunin, V. I., Belardinelli, R., Peske, F., and Rodnina, M. V. (2014). Programmed –1 frameshifting by kinetic partitioning during impeded translocation. *Cell* 157, 1619–1631. doi: 10.1016/j.cell.2014.04.041

Caliskan, N., Wohlgemuth, I., Korniy, N., Pearson, M., Peske, F., and Rodnina, M. V. (2017). Conditional switch between frameshifting regimes upon translation of dnaX mRNA. *Mol. Cell.* 66, 558–567. doi: 10.1016/j.molcel.2017.04.023

Cao, X., and Slavoff, S. A. (2020). Non-AUG start codons: expanding and regulating the small and alternative ORFeome. *Exp. Cell. Res.* 391:111973. doi: 10.1016/j.yexcr.2020.111973

Castellano, S. (2009). On the unique function of selenocysteine—insights from the evolution of selenoproteins. *Biochim. Biophys. Acta* 1790, 1463–1470. doi: 10.1016/j.bbagen.2009.03.027

Chavatte, L., Brown, B. A., and Driscoll, D. M. (2005). Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat. Struct. Mol. Biol.* 12, 408–416. doi: 10.1038/nsmb922

Chen, C., Zhang, H., Broitman, S., Reiche, M., Farrell, I., Cooperman, B., et al. (2013). Dynamics of translation by single ribosomes through mRNA secondary structures. *Nat. Struct. Mol. Biol.* 20, 582–588. doi: 10.1038/nsmb.2544

Chen, J., Petrov, A., Johansson, M., Tsai, A., O'Leary, S., and Puglisi, J. (2014). Dynamic pathways of -1 translational frameshifting. *Nature* 512, 328–332. doi: 10.1038/nature13428

Choi, J., O'Loughlin, S., Atkins, J. F., and Puglisi, J. D. (2020). The energy landscape of –1 ribosomal frameshifting. *Sci. Advan.* 6:eaax6969. doi: 10.1126/sciadv.aax6969

Cobucci-Ponzano, B., Conte, F., Benelli, D., Londei, P., Flagiello, A., Monti, M., et al. (2006). The gene of an archaeal alpha-L-fucosidase is expressed by translational frameshifting. *Nucleic Acids Res.* 34, 4258–4268. doi: 10.1093/nar/gkl574

Cobucci-Ponzano, B., Guzzini, L., Benelli, D., Londei, P., Perrodou, E., Lecompte, O., et al. (2010). Functional characterization and high- throughput proteomic analysis of interrupted genes in the archaeon *Sulfolobus solfataricus*. *J. Prot. Res.* 9, 2496–2507. doi: 10.1021/pr901166q

Cobucci-Ponzano, B., Mazzone, M., Rossi, M., and Moracci, M. (2005a). Probing the catalytically essential residues of the α-L-fucosidase from the hyperthermophilic archaeon Sulfolobus solfataricus. *Biochemistry* 44, 6331–6342. doi: 10.1021/bi047495f

Cobucci-Ponzano, B., Rossi, M., and Moracci, M. (2005b). Recoding in archaea. *Mol. Microbiol.* 55, 339–348. doi: 10.1111/j.1365-2958.2004.04400.x

Cobucci-Ponzano, B., Rossi, M., and Moracci, M. (2012). Translational recoding in archaea. *Extremophiles* 16, 793–803. doi: 10.1007/s00792-012-0482-8

Cobucci-Ponzano, B., Trincone, A., Giordano, A., Rossi, M., and Moracci, M. (2003a). Identification of an archaeal α-L-fucosidase encoded by an interrupted gene. Production of a functional enzyme by mutations mimicking programmed -1 frameshifting. *J. Biol. Chem.* 278, 14622–14631. doi: 10.1074/jbc.M211834200

Cobucci-Ponzano, B., Trincone, A., Giordano, A., Rossi, M., and Moracci, M. (2003b). Identification of the catalytic nucleophile of the family 29 α-L-fucosidase from *Sulfolobus solfataricus* via chemical rescue of an inactive mutant. *Biochemistry* 42, 9525–9531. doi: 10.1021/bi035036t

Cone, J. E., Del Río, R. M., Davis, J. N., and Stadtman, T. C. (1976). Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. *Proc. Natl. Acad. Sci. U.S.A.* 73, 2659–2663. doi: 10.1073/pnas.73.8.2659

Cui, Y., Dinman, J. D., Kinzy, T. G., and Peltz, S. W. (1998). The Mof2/Sui1 protein is a general monitor of translational accuracy. *Mol. Cell. Biol.* 18, 1506–1516. doi: 10.1128/mcb.18.3.1506

Curci, N., Strazzulli, A., Iacono, R., De Lise, F., Maurelli, L., Di Fenza, M., et al. (2021). Xyloglucan oligosaccharides hydrolysis by exo-acting glycoside hydrolases from hyperthermophilic microorganism *Saccharolobus solfataricus*. *Int. J. Mol. Sci.* 22:3325. doi: 10.3390/ijms22073325

De Lise, F., Iacono, R., Strazzulli, A., Giglio, R., Curci, N., Maurelli, L., et al. (2021). Transcript regulation of the recoded archaeal α-L-fucosidase in vivo. *Molecules* 26:1861. doi: 10.3390/molecules26071861

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., et al. (2002). The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microb. Biotech.* 4, 453–461.

Eichler, J., and Adams, M. W. (2005). Posttranslational protein modification in Archaea. *Microbiol. Mol. Biol. Rev.* 69, 393–425. doi: 10.1128/MMBR.69.3.393-425.2005

Evans, P. N., Parks, D. H., Chadwick, G. L., Robbins, S. J., Orphan, V. J., Golding, S. D., et al. (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 350, 434–438. doi: 10.1126/science.aac7745

Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P., and Krol, A. (2000). Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J.* 19, 4796–4805. doi: 10.1093/emboj/19.17.4796

Farabaugh, P. J. (1996). Programmed translational frameshifting. *Microbiol. Rev.* 60, 103–134.

Firth, A. E., and Brierley, I. (2012). Non-canonical translation in RNA viruses. *J. Gen. Virol.* 93, 1385–1409. doi: 10.1099/vir.0.042499-0

Fletcher, J. E., Copeland, P. R., Driscoll, D. M., and Krol, A. (2001). The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA* 7, 1442–1453.

Fomenko, D. E., and Gladyshev, V. N. (2012). Comparative genomics of thiol oxidoreductases reveals widespread and essential functions of thiol based redox control of cellular processes. *Antiox Redox Sign.* 16, 193–201. doi: 10.1089/ars.2011.3980

Fricke, W. F., Seedorf, H., Henne, A., Krüer, M., Liesegang, H., Hedderich, R., et al. (2006). The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H2 for methane formation and ATP synthesis. *J. Bacteriol.* 188, 642–658. doi: 10.1128/JB.188.2.642-658.2006

Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., et al. (2002). The genome of M. acetivorans reveals extensive metabolic and physiological diversity. *Genome Res.* 12, 532–542. doi: 10.1101/gr.223902

Gallant, J., and Lindsley, D. (1992). Leftward ribosome frameshifting at a hungry codon. *J Mo. Biol.* 223, 31–40. doi: 10.1016/0022-2836(92)90713-t

Gaston, M. A., Zhang, L., Green-Church, K. B., and Krzycki, J. A. (2011). The complete biosynthesis of the genetically encoded amino acid pyrrolysine fromlysine. *Nature* 471, 647–650. doi: 10.1038/nature09918

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

Gesteland, R. F., and Atkins, J. F. (1996). Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem.* 65, 741–768. doi: 10.1146/annurev.bi.65.070196.003521

Goodchild, A., Saunders, N. F., Ertan, H., Raftery, M., Guilhaus, M., Curmi, P. M. G., et al. (2004). A proteomic determination of cold adaptation in the Antarctic archaeon, *Methanococcoides burtonii*. *Mol. Microbiol.* 53, 309–321. doi: 10.1111/j.1365-2958.2004.04130.x

Gromer, S., Johansson, L., Bauer, H., Arscott, L. D., Rauch, S., Ballou, D. P., et al. (2003). Active sites of thioredoxin reductases: why selenoproteins? *Proc. Natl. Acad. Sci. U.S.A.* 100, 12618–12623. doi: 10.1073/pnas.2134510100

Guan, Y., Haroon, M. F., Alam, I., Ferry, J. G., and Stingl, U. (2017). Single-cell genomics reveals pyrrolysine-encoding potential in members of uncultivated archaeal candidate division MSBL-1. *Environ. Microbiol. Rep.* 9, 404–410. doi: 10.1111/1758-2229.12545

Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A., and Chan, M. K. (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 296, 1462–1466. doi: 10.1126/science.1069556

Harrison, P. M., and Gerstein, M. (2002). Studying Genomes through the aeons: protein families, pseudogenes and proteome evolution. *J. Mol. Biol.* 318, 1155–1174. doi: 10.1016/s0022-2836(02)00109-2

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

Hatfield, D. L., and Gladyshev, V. N. (2002). How selenium has altered our understanding of the genetic code. *Mol. Cell. Biol.* 22, 3565–3576. doi: 10.1128/mcb.22.11.3565-3576.2002

Heinemann, I. U., O'Donoghue, P., Madinger, C., Benner, J., Randau, L., Noren, C. J., et al. (2009). The appearance of pyrrolysine in tRNAHisguanylyltransferase by neutral evolution. *Proc. Natl Acad. Sci. U.S.A.* 106, 21103–21108. doi: 10.1073/pnas.0912072106

Herr, A. J., Wills, N. M., Nelson, C. C., Gesteland, R. F., and Atkins, J. F. (2004). Factors that influence selection of coding resumption sites in translational bypassing: minimal conventional peptidyl-tRNA: mRNA pairing can suffice. *J. Biol. Chem.* 279, 11081–11087. doi: 10.1074/jbc.M311491200

Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., et al. (2003). An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423, 91–96. doi: 10.1038/nature01535

Hoffman, K. S., Crnković, A., and Söll, D. (2018). Versatility of synthetic tRNAs in genetic code expansion. *Genes* 9:537. doi: 10.3390/genes9110537

Hondal, R. J., Marino, S. M., and Gladyshev, V. N. (2013). Selenocysteine in thiol/ disulfide-like exchange reactions. *Antioxid Redox Signal.* 18, 1675–1689. doi: 10.1089/ars.2012.5013

Hondal, R. J., and Ruggles, E. L. (2011). Differing views of the role of selenium in thioredoxin reductase. *Amino Acids* 41, 73–89. doi: 10.1007/s00726-010-0494-6

Iacono, R., Cobucci-Ponzano, B., De Lise, F., Curci, N., Maurelli, L., Moracci, M., et al. (2020). Spatial metagenomics of three geothermal sites in Pisciarelli hot Spring focusing on the biochemical resources of the microbial consortia. *Molecules* 25:4023. doi: 10.3390/molecules25174023

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009). Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223. doi: 10.1126/science.1168978

Jacks, T., and Varmus, H. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* 230, 1237–1242. doi: 10.1126/science.2416054

James, C. M., Ferguson, T. K., Leykam, J. F., and Krzycki, J. A. (2001). The amber codon in the gene encoding the monomethylamine methyltransferase isolated from Methanosarcina barkeri is translated as a sense codon. *J. Biol. Chem.* 276, 34252–34258. doi: 10.1074/jbc.M102929200

Jiang, L., Ni, J., and Liu, Q. (2012). Evolution of selenoproteins in the metazoan. *BMC Genomics* 13:446. doi: 10.1186/1471-2164-13-446

Kelly, J. A., Olson, A. N., Neupane, K., Munshi, S., Emeterio, J. S., Pollack, L., et al. (2020). Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS-CoV-2. *bioRxiv [preprint]* doi: 10.1101/2020.03.13.991083

Kim, H., Liu, F., Fei, J., Bustamante, C., Gonzalez, R. Jr., and Tinoco, I. Jr. (2014). A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5538–5543. doi: 10.1073/pnas.1403457111

Kobayashi, Y., Zhuang, J., Peltz, S., and Dougherty, J. (2010). Identification of a cellular factor that modulates HIV-1 programmed ribosomal frameshifting. *J. Biol. Chem.* 285, 19776–19784. doi: 10.1074/jbc.M109.085621

Korniy, N., Goyal, A., Hoffmann, M., Samatova, E., Peske, F., Pohlmann, S., et al. (2019a). Modulation of HIV-1 Gag/Gag-Pol frameshifting by tRNA abundance. *Nucleic Acids Res.* 47, 5210–5222. doi: 10.1093/nar/gkz202

Korniy, N., Samatova, E., Anokhina, M., Peske, F., and Rodnina, M. (2019b). Mechanisms and biomedical implications of –1 programmed ribosome frameshifting on viral and bacterial mRNAs. *FEBS Lett.* 593, 1468–1482. doi: 10.1002/1873-3468.13478

Krol, A. (2002). Evolutionarily different RNA motifs and RNA-protein complexes to achieve selenoprotein synthesis. *Biochimie* 84, 765–774. doi: 10.1016/s0300-9084(02)01405-0

Kromayer, M., Wilting, R., Tormay, P., and Böck, A. (1996). Domain structure of the prokaryotic selenocysteine-specific elongation factor SelB. *J. Mol. Biol.* 262, 413–420. doi: 10.1006/jmbi.1996.0525

Kryukov, G. V., and Gladyshev, V. N. (2004). The prokaryotic selenoproteome. *EMBO Rep.* 5, 538–543. doi: 10.1038/sj.embor.7400126

Krzycki, J. A. (2004). Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr. Opin. Chem. Biol.* 8, 484–491. doi: 10.1016/j.cbpa.2004.08.012

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

Labunskyy, V. M., Hatfield, D. L., and Gladyshev, V. N. (2014). Selenoproteins: molecular pathways and physiological roles. *Physiol. Rev.* 94, 739–777. doi: 10.1152/physrev.00039.2013

Lang, B. F., Jakubkova, M., Hegedusova, E., Daoud, R., Forget, L., Brejova, B., et al. (2014). Massive programmed translational jumping in mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5926–5931. doi: 10.1073/pnas.1322190111

Larsen, B., Gesteland, R. F., and Atkins, J. F. (1997). Structural probing and mutagenic analysis of the stem-loop required for *Escherichia coli* dnaX ribosomal frameshifting: programmed efficiency of 50%. *J. Mol. Biol.* 271, 47–60. doi: 10.1006/jmbi.1997.1162

Lin, Z., Gilbert, R., and Brierley, I. (2012). Spacer-length dependence of programmed -1 or -2 ribosomal frameshifting on a U_6A heptamer supports a

role for messenger RNA (mRNA) tension in frameshifting. *Nucleic Acids Res.* 40, 8674–8689. doi: 10.1093/nar/gks629

Ling, J., O'Donoghue, P., and Söll, D. (2015). Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. *Nat. Rev. Microbiol.* 13, 707–772. doi: 10.1038/nrmicro3568

Lobanov, A. V., Hatfield, D. L., and Gladyshev, V. N. (2008). Reduced reliance on the trace element selenium during evolution of mammals. *Genome Biol.* 9:R62. doi: 10.1186/gb-2008-9-3-r62

Longstaff, D. G., Larue, R. C., Faust, J. E., Mahapatra, A., Zhang, L., Green-Church, K. B., et al. (2007). A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1021–1026. doi: 10.1073/pnas.0610294104

Luthi, K., Moser, M., Ryser, J., and Weber, H. (1990). Evidence for a role of translational frameshifting in the expression of transposition activity of the bacterial insertion element IS1. *Gene* 88, 15–20. doi: 10.1016/0378-1119(90)90054-u

Maeder, D. L., Anderson, I., Brettin, T. S., Bruce, D. C., Gilna, P., Han, C. S., et al. (2006). The Methanosarcina barkeri genome: comparative analysis with Methanosarcina acetivorans and Methanosarcina mazei reveals extensive rearrangement within methanosarcinal genomes. *J. Bacteriol.* 188, 7922–7931. doi: 10.1128/JB.00810-06

Mangold, U. (2005). The antizyme family: polyamines and beyond. *IUBMB Life* 57, 671–676. doi: 10.1080/15216540500307031

Mariotti, M., Lobanov, A. V., Guigo, R., and Gladyshev, V. N. (2013). SECISearch3 and Seblastian: new tools for prediction of SECIS elements and selenoproteins. *Nucleic Acids Res.* 41:e149. doi: 10.1093/nar/gkt550

Mariotti, M., Lobanov, A. V., Manta, B., Santesmasses, D., Bofill, A., Guigó, R., et al. (2016). *Lokiarchaeota* marks the transition between the archaeal and eukaryotic selenocysteine encoding systems. *Mol. Biol. Evol.* 33, 2441–2453. doi: 10.1093/molbev/msw122

Mariotti, M., Santesmasses, D., Capella-Gutierrez, S., Mateo, A., Arnan, C., Johnson, R., et al. (2015). Evolution of selenophosphate synthetases: emergence and relocation of function through independent duplications and recurrent subfunctionalization. *Genome Res.* 25, 1256–1267. doi: 10.1101/gr.190538.115

Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J. F., Gesteland, R. F., et al. (1995). Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* 80, 51–60. doi: 10.1016/0092-8674(95)90450-6

Michel, A. M., and Baranov, P. V. (2013). Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *WIREs RNA* 4, 473–490. doi: 10.1002/wrna.1172

Namy, O., Moran, S., Stuart, D., Gilbert, R., and Brierley, I. (2006). A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* 441, 244–247. doi: 10.1038/nature04735

Namy, O., Rousset, J. P., Sawsan, N., and Brierley, I. (2004). Reprogrammed genetic decoding review in cellular gene expression. *Mol. Cell.* 13, 157–168. doi: 10.1016/s1097-2765(04)00031-0

Namy, O., Zhou, Y., Gundllapalli, S., Polycarpo, C. R., Denise, A., Rousset, J. P., et al. (2007). Adding pyrrolysine to the *Escherichia coli* genetic code. *FEBS Lett.* 581, 5282–5288. doi: 10.1016/j.febslet.2007.10.022

Napthine, S., Ling, R., Finch, L., Jones, J., Bell, S., Brierley, I., et al. (2017). Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat. Commun.* 8:15582. doi: 10.1038/ncomms15582

Nicholas, P. R., Jean-François Brugère, J. F., Atkins, P. W., and O'Toole, G. B. (2018). Pyrrolysine in archaea: a 22nd amino acid encoded through a genetic code expansion. *Emerg. Top. Life Sci.* 2, 607–618. doi: 10.1042/ETLS20180094

Nobu, M. K., Narihiro, T., Kuroda, K., Mei, R., and Liu, W. T. (2016). Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. *ISME J.* 10, 2478–2487. doi: 10.1038/ismej.2016.33

Nosek, J., Tomaska, L., Burger, G., and Lang, B. F. (2015). Programmed translational bypassing elements in mitochondria: structure, mobility, and evolutionary origin. *Trends Genet.* 31, 187–194. doi: 10.1016/j.tig.2015.02.010

Nozawa, K., O'Donoghue, P., Gundllapalli, S., Araiso, Y., Ishitani, R., Umehara, T., et al. (2009). Pyrrolysyl-tRNA synthetase- tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* 457, 1163–1167. doi: 10.1038/nature07611

Olubajo, B., and Taylor, E. (2005). A -1 frameshift in the HIV-1 env gene is enhanced by arginine deficiency via a hungry codon mechanism. *Mutat. Res.* 579, 125–132. doi: 10.1016/j.mrfmmm.2005.02.018

Onofri, S., Balucani, N., and Barone, V. (2020). The Italian National project of astrobiology-life in space-origin, presence, persistence of life in space, from molecules to extremophiles. *Astrobiology* 20, 580–582. doi: 10.1089/ast.2020.2247

Osawa, S., Jukes, T. H., Watanabe, K., and Muto, A. (1992). Recent evidence for evolution of the genetic code. *Microbiol. Rev.* 56, 229–264.

Petitjean, C., Deschamps, P., López-García, P., Moreira, D., and Brochier-Armanet, C. (2015). Extending the conserved phylogenetic core of archaea disentangles the evolution of the third domain of life. *Mol. Biol. Evol.* 32, 1242–1254. doi: 10.1093/molbev/msv015

Pietila, M. K., Laurinmaki, P., Russell, D. A., Ko, C. C., Jacobs-Sera, D., Butcher, S. J., et al. (2013). Insights into head-tailed viruses infecting extremely halophilic archaea. *J. Virol.* 87, 3248–3260. doi: 10.1128/JVI.03397-12

Plant, E., Rakauskaite, R., Taylor, D., and Dinman, J. (2010). Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *J. Virol.* 84, 4330–4340. doi: 10.1128/JVI.02480-09

Polycarpo, C., Ambrogelly, A., Bérubé, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., et al. (2004). An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12450–12454. doi: 10.1073/pnas.0405362101

Polycarpo, C., Ambrogelly, A., Ruan, B., Tumbula-Hansen, D., Ataide, S. F., Ishitani, R., et al. (2003). Activation of the pyrrolysine suppressor tRNA requires formation of a ternary complex with class I and class II lysyl- tRNA synthetases. *Mol. Cell.* 12, 287–294. doi: 10.1016/S1097-2765(03)00280-6

Pooggin, M. M., and Ryabova, L. A. (2018). Ribosome shunting, polycistronic translation, and evasion of antiviral defenses in plant pararetroviruses and beyond. *Front. Microbiol.* 9:644. doi: 10.3389/fmicb.2018.00644

Prat, L., Heinemann, I. U., Aerni, H. R., Rinehart, J., O'Donoghue, P., and Söll, D. (2012). Carbon source-dependent expansion of the genetic code in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21070–21075. doi: 10.1073/pnas.121861311

Rodnina, M. V., Korniy, N., and Klimova, M. (2020). Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Res.* 48, 1056–1067. doi: 10.1093/nar/gkz783

Rosano, C., Zuccotti, S., Cobucci-Ponzano, B., Mazzone, M., Rossi, M., Moracci, M., et al. (2004). Structural characterization of the nonameric assembly of an Archaeal alpha-L-fucosidase by synchrotron small angle X-ray scattering. *Biochem. Biophys. Res. Commun.* 320, 176–182. doi: 10.1016/j.bbrc.2004.05.149

Rother, M., and Krzycki, J. A. (2010). Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea* 2010:453642. doi: 10.1155/2010/453642

Rother, M., Mathes, I., Lottspeich, F., and Böck, A. (2003). Inactivation of the *selB* gene in *Methanococcus maripaludis*: effect on synthesis of selenoproteins and their sulfur-containing homologs. *J. Bacteriol.* 185, 107–114. doi: 10.1128/JB.185.1.107-114.2003

Rother, M., and Quitzke, V. (2018). Selenoprotein synthesis and regulation in Archaea. *Biochimt. Biophys. Acta (BBA) - Gen Subj.* 1862, 2451–2462. doi: 10.1016/j.bbagen.2018.04.008

Rother, M., Wilting, R., Commans, S., and Böck, A. (2000). Identification and characterisation of the selenocysteine-specific translation factor SelB from the archaeon *Methanococcus jannaschii*. *J. Mol. Biol.* 299, 351–358. doi: 10.1006/jmbi.2000.3756

Sakai, H. D., and Kurosawa, N. (2018). *Saccharolobus caldissimus* gen. nov., sp. nov., a facultatively anaerobic iron-reducing hyperthermophilic archaeon isolated from an acidic terrestrial hot spring, and reclassification of *Sulfolobus solfataricus* as *Saccharolobus solfataricus* comb. nov. and *Sulfolobus shibatae* as *Saccharolobus shibatae* comb. nov. *Int. J. Syst. Evol. Microbiol.* 68, 1271–1278. doi: 10.1099/ijsem.0.002665

Santesmasses, D., Mariotti, M., and Guigo, R. (2017). Computational identification of the selenocysteine tRNA (tRNASec) in genomes. *PLoS Comput. Biol.* 13:e1005383. doi: 10.1371/journal.pcbi.1005383

Sencilo, A., Jacobs-Sera, D., Russell, D. A., Ko, C. C., Bowman, C. A., Atanasova, N. S., et al. (2013). Snapshot of haloarchaeal tailed virus genomes. *RNA Biol.* 10, 803–816. doi: 10.4161/rna.24045

Sharma, V., Firth, A. E., Antonov, I., Fayet, O., Atkins, J. F., Borodovsky, M., et al. (2011). A pilot study of bacterial genes with disrupted ORFs reveals a surprising profusion of protein sequence recoding mediated by ribosomal frameshifting and transcriptional realignment. *Mol. Biol. Evol.* 28, 3195–3211. doi: 10.1093/molbev/msr155

Sorokin, D. Y., Makarova, K. S., Abbas, B., Ferrer, M., Golyshin, P. N., Galinski, E. A., et al. (2017). Discovery of extremely halophilic, methyl-reducing Euryarchaea provides insights into the evolutionary origin of methanogenesis. *Nat. Microbiol.* 2:17081. doi: 10.1038/nmicrobiol.2017.81

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., et al. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179. doi: 10.1038/nature14447

Srinivasan, G., James, C. M., and Krzycki, J. A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG- decoding specialized tRNA. *Science* 296, 1459–1462. doi: 10.1126/science.1069588

Sriram, A., Bohlen, J., and Teleman, A. A. (2018). Translation acrobatics: how cancer cells exploit alternate modes of translational initiation. *EMBO Rep.* 19:e45947. doi: 10.15252/embr.201845947

Stock, T., and Rother, M. (2009). Selenoproteins in archaea and Gram-positive bacteria. *Biochim. Biophys. Acta* 1790, 1520–1532. doi: 10.1016/j.bbagen.2009.03.022

Stock, T., Selzer, M., and Rother, M. (2010). In vivo requirement of selenophosphate for selenoprotein synthesis in archaea. *Mol. Microbiol.* 75, 149–160. doi: 10.1111/j.1365-2958.2009.06970.x

Strazzulli, A., Cobucci-Ponzano, B., Iacono, R., Giglio, R., Maurelli, L., Curci, N., et al. (2020). Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *FEBS J.* 287, 1116–1137. doi: 10.1111/febs.15080

Temperley, R., Richter, R., Dennerlein, S., Lightowlers, R., and Chrzanowska-Lightowlers, Z. (2010). Hungry codons promote frameshifting in human mitochondrial ribosomes. *Science* 327:301. doi: 10.1126/science.1180674

Tharp, J. M., Ehnbom, A., and Liu, W. R. (2018). tRNAPyl: structure, function, and applications. *RNA Biol.* 15, 441–452. doi: 10.1080/15476286.2017.1356561

Ticak, T., Kountz, D. J., Girosky, K. E., Krzycki, J. A., and Ferguson, D. J. (2014). A nonpyrrolysine member of the widely distributed trimethylamine methyltransferase family is a glycine betaine methyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4668–E4676. doi: 10.1073/pnas

Tsuchihashi, Z., and Brown, P. O. (1992). Sequence requirements for efficient translational frameshifting in the *Escherichia coli* dnaX gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. *Genes Dev.* 6, 511–519. doi: 10.1101/gad.6.3.511

Tsuchihashi, Z., and Kornberg, A. (1990). Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci. U.S.A.* 87, 2516–2520. doi: 10.1073/pnas.87.7.2516

Tujebajeva, R. M., Copeland, P. R., Xu, X. M., Carlson, B. A., Harney, J. W., Driscoll, D. M., et al. (2000). Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep.* 1, 158–163. doi: 10.1093/embo-reports/kvd033

van Passel, M. W., Smillie, C. S., and Ochman, H. (2007). Gene decay in archaea. *Archaea* 2, 137–143. doi: 10.1155/2007/165723

Vanwonterghem, I., Evans, P. N., Parks, D. H., Jensen, P. D., Woodcroft, B. J., Hugenholtz, P., et al. (2016). Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* 1:16170. doi: 10.1038/nmicrobiol.2016.170

Wang, X., Xuan, Y., Han, Y., Ding, X., Ye, K., Yang, F., et al. (2019). Regulation of HIV-1 Gag-Pol Expression by Shiftless, an Inhibitor of Programmed -1 Ribosomal Frameshifting. *Cell* 176, 625–635. doi: 10.1016/j.cell.2018.12.030

Wills, N., Moore, B., Hammer, A., Gesteland, R., and Atkins, J. (2006). A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. *J/ Biol. Chem.* 281, 7082–7088. doi: 10.1074/jbc.M511629200

Wilting, R., Schorling, S., Persson, B. C., and Bock, A. (1997). Selenoprotein synthesis in archaea: identification of an mRNA element of *Methanococcus jannaschii* probably directing selenocysteine insertion. *J Mol Biol.* 266, 637–641. doi: 10.1006/jmbi.1996.0812

Xu, J., Hendrix, R. W., and Duda, R. L. (2004). Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol. Cell.* 16, 11–21. doi: 10.1016/j.molcel.2004.09.006

Yan, S., Wen, J., Bustamante, C., and Tinoco, I. Jr. (2015). Ribosome excursions during mRNA translocation mediate broad branching of frameshift pathways. *Cell* 160, 870–881. doi: 10.1016/j.cell.2015.02.003

Yang, Y., and Wang, Z. (2019). IRES-mediated cap-independent translation, a path leading to hidden proteome. *J. Mol. Cell Biol.* 11, 911–919. doi: 10.1093/jmcb/mjz091

Yoshizawa, S., and Bock, A. (2009). The many levels of control on bacterial selenoprotein synthesis. *Biochim. Biophys. Acta* 1790, 1404–1414. doi: 10.1016/j.bbagen.2009.03.010

Yoshizawa, S., Rasubala, L., Ose, T., Kohda, D., Fourmy, D., and Maenaka, K. (2005). Structural basis for mRNA recognition by elongation factor SelB. *Nat. Struct. Mol. Biol.* 12, 198–203. doi: 10.1038/nsmb890

Zhang, Y., and Gladyshev, V. N. (2008). Trends in selenium utilization in marine microbial world revealed through the analysis of the global ocean sampling (GOS) project. *PLoS Genet.* 13,4:e1000095. doi: 10.1371/journal.pgen.1000095

Zhang, Y., Baranov, P. V., Atkins, J. F., and Gladyshev, V. N. (2005). Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* 280, 20740–20751. doi: 10.1074/jbc.M501458

Zhang, Y., Romero, H., Salinas, G., and Gladyshev, V. N. (2006). Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome Biol.* 7:R94. doi: 10.1186/gb-2006-7-10-r94

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.

frontiers in Microbiology

REVIEW
published: 22 July 2021
doi: 10.3389/fmicb.2021.686977

Ribosome Biogenesis in Archaea

Paola Londei[1]* *and Sébastien Ferreira-Cerca*[2]*

[1] Department of Molecular Medicine, University of Rome Sapienza, Rome, Italy, [2] Biochemistry III – Regensburg Center for Biochemistry, Institute for Biochemistry, Genetics and Microbiology, University of Regensburg, Regensburg, Germany

Making ribosomes is a major cellular process essential for the maintenance of functional ribosome homeostasis and to ensure appropriate gene expression. Strikingly, although ribosomes are universally conserved ribonucleoprotein complexes decoding the genetic information contained in messenger RNAs into proteins, their biogenesis shows an intriguing degree of variability across the tree of life. In this review, we summarize our knowledge on the least understood ribosome biogenesis pathway: the archaeal one. Furthermore, we highlight some evolutionary conserved and divergent molecular features of making ribosomes across the tree of life.

Keywords: archaea, ribosome, ribosome biogenesis, ribosomal RNA, ribosomal proteins, RNA modifications

OPEN ACCESS

Edited by:
Eveline Peeters,
Vrije University Brussel, Belgium

Reviewed by:
Ute Kothe,
University of Lethbridge, Canada
Fabian Blombach,
University College London, United Kingdom
Henrik Nielsen,
University of Copenhagen, Denmark

***Correspondence:**
Sébastien Ferreira-Cerca
sebastien.ferreira-cerca@ur.de
orcid.org/0000-0002-0522-843X
Paola Londei
paola.londei@uniroma1.it

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

Received: *28 March 2021*
Accepted: *14 May 2021*
Published: *22 July 2021*

Citation:
Londei P and Ferreira-Cerca S (2021) Ribosome Biogenesis in Archaea. Front. Microbiol. 12:686977. doi: 10.3389/fmicb.2021.686977

RIBOSOMAL SUBUNIT COMPOSITION: ARCHAEAL SPECIFICITY AND COMMON FEATURES

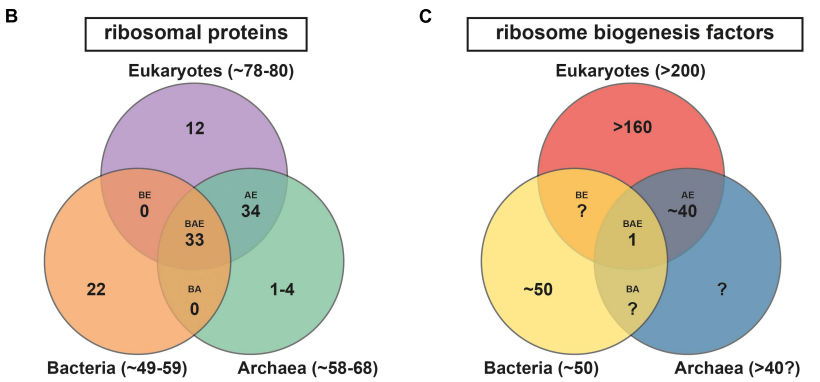
The ribosome is a universally conserved ribonucleoprotein (RNP) complex required for the synthesis of polypeptides from the intermediate molecule carrying the genetic information, the messenger RNA (Melnikov et al., 2012; Bowman et al., 2020). The birth of a ribosome itself is a highly energy-consuming and complicated orchestrated molecular dance that culminates in the formation of translation-competent mature ribosomal subunits (Nomura, 1999; Warner, 1999). The mature ribosome is composed of two ribosomal subunits, the small and large ribosomal subunits (hereafter SSU and LSU, respectively). These ribosomal subunits can be further divided into two main classes of structural components, the ribosomal RNA (rRNA) and the ribosomal proteins (r-protein). Despite its universality, the sequence and composition of the ribosomal subunits' structural components diverge across and within the different domains of life (Melnikov et al., 2012; Ban et al., 2014; Bowman et al., 2020). Notably, the sequence variabilities seen among the universally conserved ribosome structural components were recognized and harnessed at the end of the 1970s by the pioneering studies of Carl Woese and his collaborators and are still the cornerstone of modern molecular phylogenetic analysis and microbial taxonomy (Fox et al., 1977; Woese and Fox, 1977; Woese et al., 1990; Albers et al., 2013; Bahram et al., 2019).

Similar to their bacterial counterparts, archaeal ribosomes are composed of three types of rRNAs: the SSU 16S rRNA and the LSU 23S and 5S rRNAs, which interact with 60–70 r-proteins, establishing an intricate macromolecular network (Melnikov et al., 2012; Ban et al., 2014; Bowman et al., 2020; **Figure 1**).

Up to now and due to the size and sequence similarities among organisms lacking a cell nucleus, the archaeal rRNA molecules have been largely seen as being of a prokaryotic nature (**Figure 1**). Particularly and in contrast to canonical prokaryotic rRNAs, most eukaryotic rRNAs

A

	Bacteria	Archaea	Eukarya
rDNA repeats numbers [1]	1-16	1-4	150-200 (> 1,000 in plants)
rRNAs composition	16S, 23S and 5S	16S, 23S and 5S	18S, 25/28S, 5.8S and 5S
rRNA size	Ec: 16S ≈ 1,500 nt/ 23S ≈ 2,900 nt	Hv: 16S ≈ 1,500 nt/ 23S ≈2,900 nt	Sc: 18S ≈ 1,800 nt / 25S ≈ 3,400 nt Hs: 18S ≈ 1,900 nt / 28S ≈ 5,000 nt
rRNA expansion segments [2]	yes (uncommon)	yes (variable - only few)	yes (variable amounts and length)
ribosomal proteins [3]		33 universal r-proteins	
	≈ 49-59 44 ubiquitous	≈ 58-68 54 ubiquitous / 71 described	≈ 78-80 78 ubiquitous
***in vitro* reconstitution** [4]	yes	yes	no*
rRNA modifications [5]	protein-based (≈36 modif)	protein- and sRNP-based	protein- and sRNP-based
2´O-Methylation / Pseudourydilation	4/11	Sso (67/9); Hv (4/2) 7-127 C/D box sRNA	Sc (55/49); Hs (≈110/95)
additional "stand-alone" base modifications	≈ 20	Hv (7)/Tko (>100)	12
Ribosome biogenesis factors [6]	≈ 50	>40?	>200



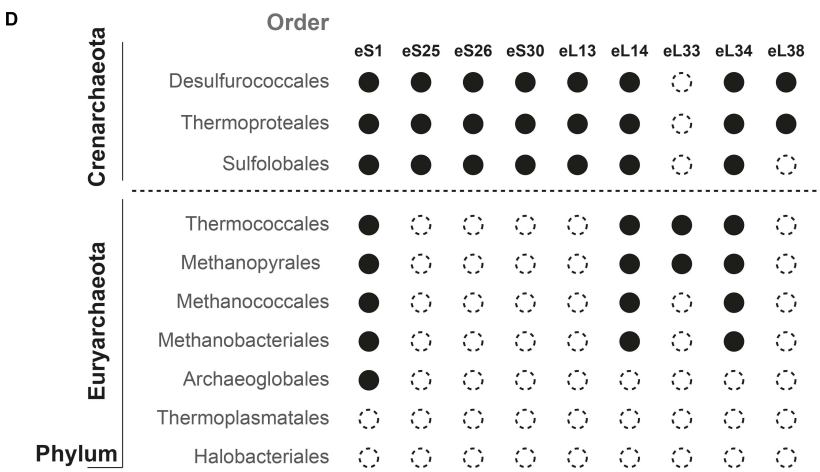


FIGURE 1 | Ribosome and ribosome biogenesis key features overview across the tree of life. **(A)** Summary of ribosome and ribosome biogenesis key features. Modified from Ferreira-Cerca (2017) according to [1](Hadjiolov, 1985; Warner, 1999; Klappenbach et al., 2001; Stoddard et al., 2015); [2](Gerbi, 1986, 1996; Armache et al., 2013; Parker et al., 2015; Petrov et al., 2015); [3](Lecompte et al., 2002; Nakao et al., 2004; Yutin et al., 2012); [4](Londei et al., 1986; Sanchez et al., 1990, 1996; Nierhaus, 1991; Mangiarotti and Chiaberge, 1997; Culver, 2003; Nierhaus and Lafontaine, 2004); [5](Lafontaine and Tollervey, 1998; Grosjean et al., 2008; Dennis et al., 2015; Sharma and Lafontaine, 2015; Krogh et al., 2016; Sloan et al., 2016; Taoka et al., 2018; Coureux et al., 2020; Grünberger et al., 2020; Sas-Chen et al., 2020); and [6](Hage and Tollervey, 2004; Thomson et al., 2013; Woolford and Baserga, 2013; Ebersberger et al., 2014; Grosjean et al., 2014; Henras et al., 2015). The detailed list of putative eukaryotic ribosome biogenesis factors conserved in archaea is depicted in Ebersberger et al. (2014). Abbreviations used: Sso, *Saccharolobus solfataricus*; Hv, *Haloferax volcanii;* Tko*, Thermococcus kodakarensis;* Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*. **(B,C)** Summary of shared ribosomal proteins **(B)** and ribosome biogenesis factors **(C)** across the three domains of life. Numbers of r-proteins and putative ribosome biogenesis factors sequence homologs shared between bacteria, archaea, and eukarya (BAE); bacteria, archaea (BA), archaea and eukarya (AE), bacteria and eukarya (BE), or unique to bacteria (B), or archaea (A), or eukarya (E), are indicated [based on (Lecompte et al., 2002; Hage and Tollervey, 2004; Nakao et al., 2004; Márquez et al., 2011; Yutin et al., 2012; Thomson et al., 2013; Woolford and Baserga, 2013; Ban et al., 2014; Ebersberger et al., 2014; Grosjean et al., 2014; Henras et al., 2015; Coureux et al., 2020; Nürenberg-Goloub et al., 2020)]. **(D)** Exemplary gene distribution of selected archaeal ribosomal proteins shared between archaea and eukaryotes across two major archaeal Phyla. Black circle denotes the presence and open circle denotes the absence of sequence homolog for the indicated ribosomal protein of the small (S) or large (L) ribosomal subunits, respectively. Adapted from Lecompte et al. (2002); Yutin et al. (2012) using the nomenclature proposed in Ban et al. (2014).

are characterized by the presence of so-called expansion segments (ES), which are additional RNA elements of various sizes incorporated into the universal prokaryotic rRNA core (Gerbi, 1996; Bowman et al., 2020; **Figure 1**). These ES increase the size and complexity of the respective rRNAs; however, recent analyses have provided evidence for the presence of such ES in both bacteria and archaea (Armache et al., 2013; Penev et al., 2020; Tirumalai et al., 2020; Stepanov and Fox, 2021). Although most of these sequence additions are limited in size and number (Armache et al., 2013; Penev et al., 2020; Tirumalai et al., 2020; Stepanov and Fox, 2021), larger ES, similar in size to those commonly observed in eukaryotes, have been recently described in the Asgard archaeal phylum (Penev et al., 2020), which is proposed to be the cradle of the eukaryotic lineage (Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017; Liu Y. et al., 2021). However, a common evolutionary relationship—based on sequence and/or structure homology—of the larger archaeal and eukaryotic ES could not be established (Penev et al., 2020). Recently, a role of some of these ES in ribosomal biogenesis and/or function has been established in eukaryotes (Ramesh and Woolford, 2016; Fujii et al., 2018; Díaz-López et al., 2019; Shankar et al., 2020). Accordingly, determining both the respective function(s) and evolutionary origin(s) of these additional rRNA segments in archaea is of general interest for the field and will be crucial to distinguish between the archaeal origin of eukaryotic features from the independent but convergent evolution trajectories of structural elements present in both archaea and eukaryotes.

The archaeal ribosomal proteins can be divided into three different groups: (1) the universally conserved r-proteins that form, with the rRNAs, the universal ribosomal core (Melnikov et al., 2012), (2) the r-proteins exclusively shared between archaea and eukaryotes, and (3) the archaeal-specific r-proteins (Lecompte et al., 2002; Márquez et al., 2011; Yutin et al., 2012; Ban et al., 2014; Coureux et al., 2020; Nürenberg-Goloub et al., 2020; **Figure 1**). The absence of exclusively shared r-proteins between bacteria and archaea remains an intriguing observation.

Among the 70 different r-proteins described in archaea, only 54 are known to be ubiquitous across archaea; among them, 33 are universally conserved (Lecompte et al., 2002; Yutin et al., 2012; Ban et al., 2014; **Figure 1**). The composition variability of the r-protein complement also correlates with a general decrease in complexity of the r-proteins composition at the domain scale (Lecompte et al., 2002; Yutin et al., 2012; **Figure 1**). In other words, the r-protein counterpart of the last archaeal common ancestor was likely more complex than that of most of its descendent lineages (Lecompte et al., 2002; Yutin et al., 2012). The functional consequences and additional adaptations underlying such r-protein reductive evolution for archaeal ribosome biogenesis and function is currently unknown. Furthermore, recent studies also indicate the presence of archaeal-specific ribosomal proteins (Márquez et al., 2011; Coureux et al., 2020; Nürenberg-Goloub et al., 2020), suggesting that the discovery of new additional archaeal-specific r-proteins is still incomplete. Last, organism-specific insertion, extension, deletion, or sequence variations within conserved r-proteins are not unusual, and may play an important role for the cellular adaptation of ribosome biogenesis and function (Ferreira-Cerca et al., 2007; Melnikov et al., 2018; Dao Duc et al., 2019). However, the functional contributions of the additional archaeal-specific r-protein features for ribosome assembly and function remain to be explored.

Another particularity of the r-protein composition of some archaeal ribosomal subunits is the presence of intra- and inter-subunit promiscuous r-proteins, which leads to an increase of the respective r-protein stoichiometry and to the presence of shared structural components of both the SSU and LSU (Armache et al., 2013). This peculiarity is in stark contrast to what is typically observed in the bacterial and eukaryotic systems, in which r-proteins are thought to be exclusive structural components of one or the other ribosomal subunit present in one copy per ribosomal subunit, with the exception of the LSU stalk r-proteins (Armache et al., 2013). The functional implications of these molecular peculiarities remain to be analyzed.

In conclusion, the core structural components of the archaeal ribosomal subunits are of prokaryotic origin, to which archaeal-specific and shared archaeal-eukaryotic features have been added. Together, the structural and functional constraints and/or advantages of these structural and compositional idiosyncrasies for ribosome biogenesis and function remain to be explored.

rRNA ORGANIZATION, SYNTHESIS, AND PROCESSING IN ARCHAEA

The organization of the rRNA genes and the maturation of the transcripts thereof to yield mature rRNA molecules is the most widely studied and best understood aspect of ribosome biogenesis in archaea (Yip et al., 2013; Ferreira-Cerca, 2017; Clouet-d'Orval et al., 2018). Because a large literature, including a number of excellent reviews, exist on this topic, here only the features most relevant from an evolutionary point of view are described.

As described, archaeal ribosomes are composed of one 30S and one 50S ribosomal subunit, the former containing a 16S rRNA and the latter 23S and 5S rRNAs. The genomic organization of the rRNA genes, however, presents marked differences in the different archaeal groups. Most euryarchaeota have a typically bacterial operon organization with the 16S-23S-5S rRNA genes linked in this order, separated by spacer sequences, and transcribed all together. In most cases the spacer separating the 16S and the 23S rRNA genes contains an Ala-tRNA gene; some euryarchaea also have a second tRNA gene, Cys-tRNA, in the 3′ETS downstream of the 5S rRNA gene (**Figure 2**). By contrast, in the crenarchaeota and probably in most members of the TACK superphylum, the 5S rRNA genes are physically separated from the other two larger rRNAs and transcribed independently (**Figure 2**). There are also a few special situations, such as that of the euryarchaeon *Themoplasma acidophilum*, where the three 16S, 23S, and 5S rRNA genes are unlinked and separately transcribed (Yip et al., 2013; Brewer et al., 2020; **Figure 2**).

The primary rRNA transcripts are maturated following pathways that follow neither the bacterial nor the eukaryal paradigm, albeit having features reminiscent of both.

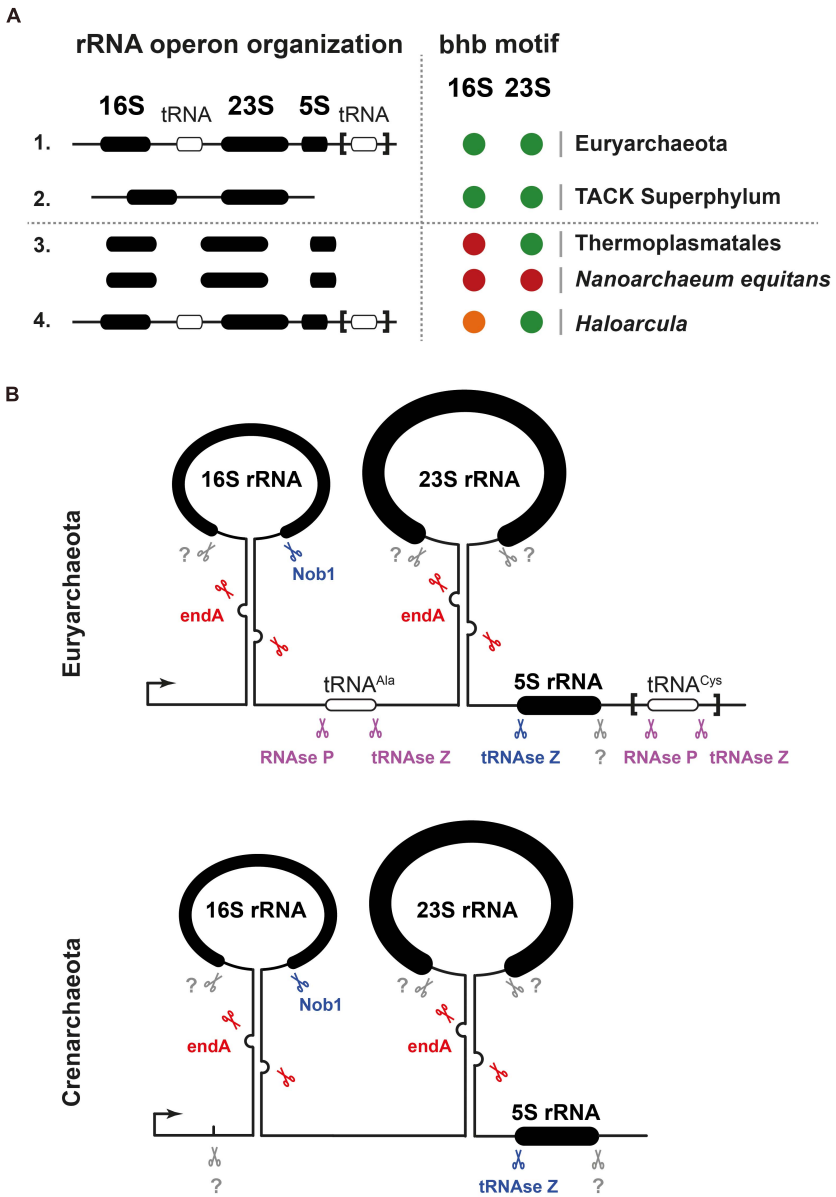


FIGURE 2 | rDNA gene organization and processing of pre-rRNA in archaea. **(A)** Ribosomal DNA gene organization and rRNA BHB motif conservation across archaea. A selected survey of archaeal rRNA operon organizations suggests two predominant classes of linked rRNA organization found in representative organisms of the Euryarchaeota and TACK Superphylum (Thaumarchaeota–Aigarchaeota–Crenarcheota–Korarchaeota) and one minor class of unlinked organization (e.g., *Thermoplasmata* class/*Nanoarchaeum equitans*). 16S and 23S rRNAs processing stem secondary structures were predicted using the ViennaRNA Web servers. Presence of predicted BHB is indicated in black. Presence of heterogeneous rRNA operons with heterogeneous presence of BHB motif within the processing stem is depicted in orange (*Haloarcula* genus). Absence of predictable BHB motifs is depicted by a red circle (e.g., *Thermoplasmata* class/*Nanoarchaeum equitans*). Modified from Jüttner et al. (2020) under CC-BY License. **(B)** Schematic representation of exemplary rRNA processing sites and the known respective ribonuclease activities required for the maturation or the pre-rRNA are indicated. Unknown activities are indicated in gray, putative activities in lilac, activities base on *in vitro* analysis in blue, and activities based on *in vivo* analysis in red. Upper panel represents common organization found in Euryarchaeota and lower panel in Crenarchaeota. Modified from Ferreira-Cerca (2017); Jüttner et al. (2020).

As in bacteria, the sequences flanking the rRNA genes have extended complementarity and pair, forming double-helical stems that are the target of certain endonucleases starting rRNA maturation. However, although, in bacteria, these stems are cleaved by RNAse III, in most archaea, they typically contain Bulge-helix-Bulge (BHB) motifs that are recognized and cleaved by the archaeal-specific *endA* splicing endonuclease (Tang et al., 2002; Ferreira-Cerca, 2017; Clouet-d'Orval et al., 2018; Qi et al., 2020; Schwarz et al., 2020; **Figure 2**). Consequently, the pre-16S and pre-23S rRNAs are ligated and first released

in a circular pre-rRNA form, which is subsequently opened and matured by other enzymes that have not yet been characterized (Tang et al., 2002; Ferreira-Cerca, 2017; Clouet-d'Orval et al., 2018; Jüttner et al., 2020; Qi et al., 2020; Schwarz et al., 2020). For a comprehensive review of the rRNA maturating/modifying enzymes, see Clouet-d'Orval et al. (2018) and Ferreira-Cerca (2017).

In certain members of the crenarchaeota, the processing of 16S rRNA has features that present some homology with the eukaryotic process; specifically, there are endonucleases that introduce 1-2 cuts within the 5′ETS (Durovic and Dennis, 1994; **Figure 2**). The most distal of these processing sites, termed site 0, lies some 70 nucleotides ahead of the 16S mature 5′ end, is probably conserved in most crenarchaeota, and has similarity to the processing site termed A0 in eukaryotes. Site A0 is generally present in eukaryotic pre-rRNAs and is one of the earliest processing sites starting its maturation (Mullineux and Lafontaine, 2012). In archaea, endonucleolytic cleavage at site 0 is independent of the formation of the processing stems containing the BHB motifs. Instead, its recognition is guided by a specific sequence containing a conserved CUU motif also found in the eukaryotic counterpart. This CUU motif is shown to be essential for cleavage in *S. solfataricus* (Ciammaruconi and Londei, 2001). Notably, in the eukarya, cleavage at site A0 requires a RNP particle containing the small nucleolar RNA U3, but in the archaea this does not seem to be the case. The archaeal endonuclease cutting at site 0 has not yet been identified; interestingly, it seems to be closely associated with the 60 kDa chaperonin, at least in *S. solfataricus* (Ruggero et al., 1998).

Although homologs of eukaryotic small nucleolar RNAs do not seem to be involved in rRNA processing in archaea, they do participate massively in another prominent feature of archaeal rRNA maturation, that is, guiding chemical modifications of specific nucleotides, which is described in the next paragraph.

RIBOSOMAL RNA MODIFICATIONS

RNA modifications were discovered in the early 1950s, and since then, more than 100 different types of chemical modifications have been described (Littlefield and Dunn, 1958; Boccaletto et al., 2018). These modifications are expanding the chemical and structural properties of the classical RNA alphabet (Li and Mason, 2014; Kadumuri and Janga, 2018).

Ribosomal RNA modifications are found in all rRNAs studied thus far (Piekna-Przybylska et al., 2008; Boccaletto et al., 2018); however, their diversity (respective chemical nature, number, and position) can be diverging across archaea (Grosjean et al., 2008; Dennis et al., 2015; Boccaletto et al., 2018; Coureux et al., 2020; Sas-Chen et al., 2020). rRNA modifications can be grouped into two main types: (1) base and (2) ribose modifications. Furthermore, the machineries involved in the rRNA modification process can be also subdivided into two major groups: (1) stand-alone enzymes, which are found across all domains of life, and (2) RNA-guided modifications, which utilize RNP complexes to guide and modify the target rRNA in an RNA sequence-dependent manner (Lafontaine and Tollervey, 1998; Omer et al., 2003; Yip et al., 2013). Notably, these RNP complexes are ubiquitous in both archaea and eukaryotes but are absent from bacteria and are responsible for the two major types of rRNA modifications, i.e., 2′O-methylation of the ribose moiety by the C/D box sRNPs and isomerization of the uridine base into pseudouridine by the H/ACA box sRNPs (Lafontaine and Tollervey, 1998; Omer et al., 2003; Yip et al., 2013). Moreover, in eukaryotes, few snoRNPs do not have any known rRNA modification function but are instead required for pre-rRNA processing (Lafontaine and Tollervey, 1998; Sharma and Lafontaine, 2015; Sloan et al., 2016). Among these, the snoRNA U3 is required for early processing steps of the SSU and to avoid premature folding of the SSU central pseudoknot structure (Baßler and Hurt, 2019; Klinge and Woolford, 2019). In archaea, U3 and snoRNPs facilitating rRNA processing and folding independently of rRNA modification activity are not known. More details about these two classes of RNPs and their rRNA modifications in archaea can be found in the two accompanying reviews in this special issue by Randau and collaborators (C/D box sRNPs; Breuer et al., 2021) and Kothe and collaborators (H/ACA box sRNPs; Czekay and Kothe, 2021).

In addition to the two main types of rRNA modifications mentioned, additional base modifications are also found. Commonly, base methylations (m1, m3, m5, m6A, ...) and also acetylation or larger types of modifications (e.g., acp3) are decorating the rRNAs (Piekna-Przybylska et al., 2008; Boccaletto et al., 2018). Generally, most of these modifications cluster within the ribosomal subunit functional centers (A-, P-, E-sites, and subunit bridges) and are believed to stabilize and/or support the activity of the translation machinery (Piekna-Przybylska et al., 2008; Sharma and Lafontaine, 2015; Sloan et al., 2016). Interestingly, the position and/or chemical nature of these modifications is apparently flexible across the tree of life, suggesting that the functional contribution of the respective rRNA modification(s) in their respective structural environments prevails over their exact chemical nature and/or relative position (Piekna-Przybylska et al., 2008; Sharma and Lafontaine, 2015; Sloan et al., 2016; Ferreira-Cerca, 2017).

The total amounts and types of rRNA modifications strongly vary across archaea. For instance, halophilic archaea possess a lower total amount of rRNA modifications (e.g., *H. volcanii* ∼10 known modifications; Grosjean et al., 2008). For example, the archaeal homologs of the eukaryotic methyltransferase Nep1 are not found in the phylogenetically related Methanogen class II and Haloarchaea (see also **Figure 3**). This decrease in the number of RNA modifications also correlates with a generally reduced amount of r-proteins and ribosome biogenesis factors in these organisms (Lecompte et al., 2002; Yutin et al., 2012; Ebersberger et al., 2014; see above and below). In contrast, the total amount of rRNA modifications in thermophiles and hyperthermophiles is particularly increased (Dennis et al., 2015). For example, representative organisms of the Thermococcales order, which can grow at remarkably high temperatures (near the boiling point of water), contain a large amount of base acetylations, presumably introduced by the archaeal homolog of the eukaryotic RNA cytidine acetyltransferase Kre33/Nat10 (Sleiman and Dragon, 2019; Coureux et al., 2020;

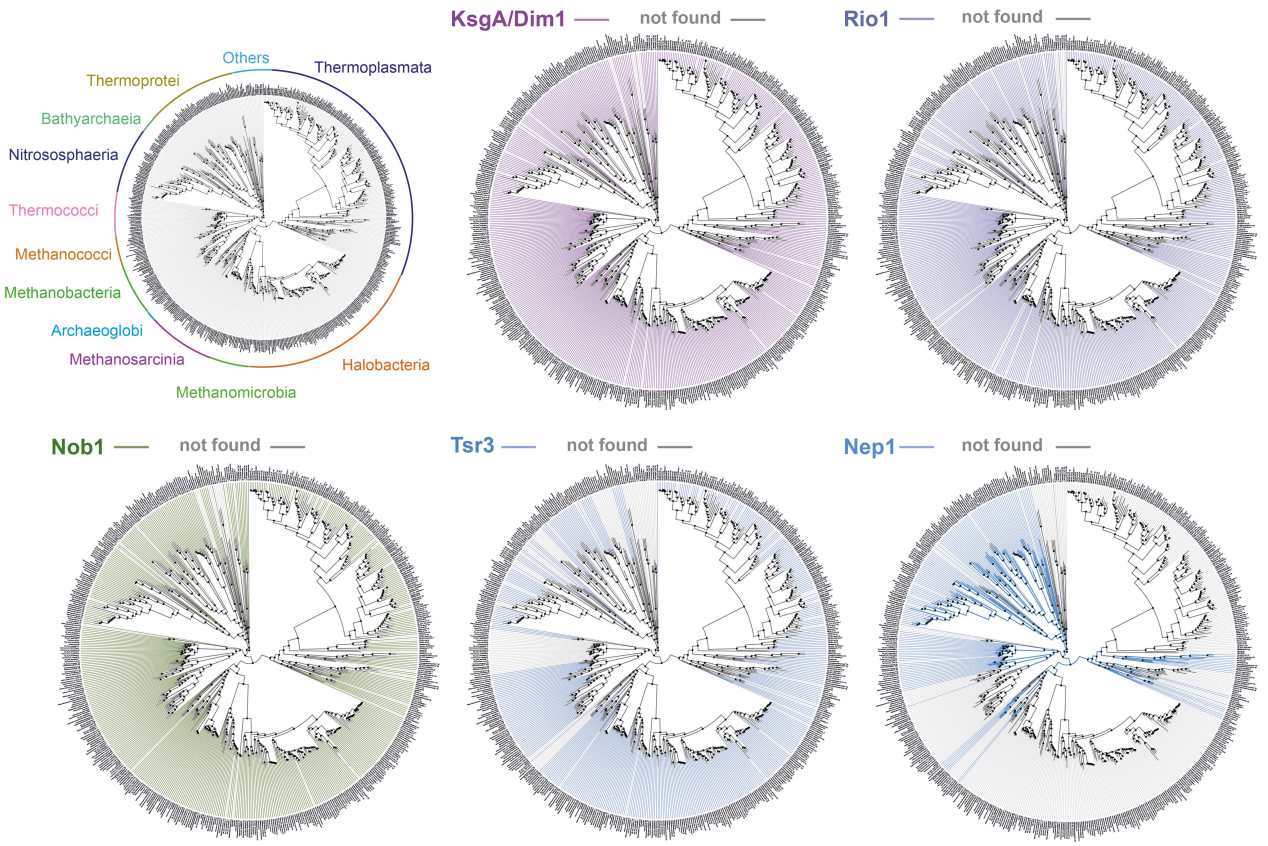


FIGURE 3 | Exemplary conservation of selected putative ribosome biogenesis factors involved in small ribosomal subunit biogenesis in archaea. Phylogenetic conservation profile of the indicated known or putative small ribosomal subunit ribosome biogenesis factors across 1,500 archaeal genomes were generated using AnnoTree (http://annotree.uwaterloo.ca; Mendler et al., 2019). Archaeal classes are annotated in a phylogenetic tree (upper left) as provided by AnnoTree. Absence of sequence homolog in a define organism is indicated by a gray line, whereas its presence is indicated by a colored line. Note the absence of significant homology for Nep1 (e.g., Thermoplasmata, Halobacteria, and more) or Tsr3 (e.g., Thermococcales) in a large group of organisms, in contrast to the more widespread distribution of KsgA/Dim1, Rio1, and Nob1 archaeal homologs.

Grünberger et al., 2020; Sas-Chen et al., 2020). Moreover, and in contrast to the clustered distribution of rRNA modifications normally observed, these acetylations are scattered throughout the rRNA sequences (Coureux et al., 2020; Grünberger et al., 2020; Sas-Chen et al., 2020). Furthermore, the total amount of these acetylations seems to vary according to the growth temperature (Sas-Chen et al., 2020).

Remarkably, among all the known stand-alone enzymes, the SSU dimethyltransferase KsgA/Dim1 carrying the dimethylation of two universally conserved adenosines at the 3′end of the SSU rRNA is the only almost universally conserved factor involved in ribosome biogenesis (Lafontaine et al., 1994; Connolly et al., 2008; Seistrup et al., 2017; Knüppel et al., 2021). Despite its widespread distribution, several functional aspects of the KsgA/Dim1 biology, such as assembly/release mechanisms and the modification process itself (e.g., completion) strikingly diverge between different organisms and across the different domain of life (Van Buul et al., 1984; Formenoy et al., 1994; Lafontaine et al., 1994; Connolly et al., 2008; Zorbas et al., 2015; Ghalei et al., 2017; Seistrup et al., 2017; Knüppel et al., 2021; Liu K. et al., 2021).

Overall, these observations suggest that the relative amount of rRNA modifications and their diversity may reflect organism-specific adaptation to their respective environmental conditions and/or organism-specific evolutionary trajectories (Dennis et al., 2015; Seistrup et al., 2017; Sas-Chen et al., 2020; Knüppel et al., 2021). The functional significance of the variability in rRNA modifications, of the presence of different modification machineries on the ribosome biogenesis pathway in archaea, and how these machineries have contributed to (re-)shape the ribosome assembly pathway remains to be determined.

ASSEMBLY OF ARCHAEAL RIBOSOMES: *IN VITRO* STUDIES

The capability of bacterial ribosomes to assemble spontaneously *in vitro* from the separate RNA and protein components was first demonstrated in the late 60's by the Nomura laboratory with *Escherichia coli* 30S subunit (Traub and Nomura, 1968) and later by the Nierhaus laboratory with the 50S subunit from the same organism (Nierhaus and Dohme, 1974). Ribosomes from

other bacterial species were also successfully reconstituted *in vitro* (Green and Noller, 1999; Agalarov et al., 2016).

These experiments are important in showing that, even in a huge macromolecular complex such as the ribosome, the components contain in themselves all the necessary information to interact in an orderly way so as to form a functional particle. Even more importantly, they highlight a definite assembly hierarchy, in which a subset of ribosomal proteins starts the ribosome biogenesis process by binding directly to specific sites on the rRNA. These "early assembly" proteins, together with the rRNA, create a "core particle" that has to undergo certain conformational changes before binding the missing proteins and being converted into the final functional particle.

Experiments of *in vitro* assembly with purified components, could define an "assembly map," i.e., the stepwise binding of ribosomal proteins to the rRNAs leading to the formation of intermediate particles that are finally converted into a complete functional ribosomal subunit (Roth and Nierhaus, 1980). However, the necessary experimental conditions (e.g., time, temperature, and ionic strength, etc.) to enable these *in vitro* reconstitution experiments are commonly incompatible with the physiological conditions of the respective organisms, thereby suggesting the presence of facilitating molecular mechanisms *in vivo*.

Among these mechanisms, the "assembly gradient" originally proposed by Knud Nierhaus suggests that cotranscriptional and directional assembly of r-proteins (5′ to 3′ direction), facilitate the initial steps of ribosomal assembly *in vivo* (Nierhaus, 1991). Similar principles of ribosomal assembly seem to apply in some eukaryotes [see, e.g., Cheng et al. (2017); de la Cruz et al. (2015); Ferreira-Cerca et al. (2007), but see also Cheng et al. (2019) and references therein] and may, therefore, likely operate in the archaeal context. For example, our recent work suggests a 5′ to 3′ coordination of the initial pre-rRNA maturation steps in *H. volcanii* (Jüttner et al., 2020). Moreover, recent studies in Sulfolobales suggest local clustering of the rRNA and r-protein operon genes, which may potentially have implication for early steps of ribosome assembly in some archaea (Takemata and Bell, 2021). However, the conservation of the topology and organization of the ribosome synthesis machinery remains to be explored (Cockram et al., 2021; Sobolev et al., 2021).

Furthermore, additional ribosome biogenesis factors facilitating or speeding up ribosome assembly were also identified later (Bunner et al., 2010; Nikolay et al., 2018; see below). Even if the pathways for *in vitro* ribosome assembly are likely to be at least in part different from those adopted *in vivo*, the results from *in vitro* studies reveal that ribosome biogenesis is a highly coordinated process that requires a number of specific sequential steps to be completed successfully.

In vitro reconstitution experiments were also employed to explore the degree of conservation of ribosomal components among different bacterial species. It was demonstrated that hybrid, active ribosomes could be successfully reconstituted from proteins and rRNA from different sources, thus further highlighting the high degree of functional and structural conservation of bacterial ribosomes (Higo et al., 1973; Vogel et al., 1984).

That archaeal ribosomes were also capable of spontaneous self-assembly *in vitro* was demonstrated some years later with the particles of two different extremophilic archaea: the 50S subunits of *Saccharolobus* (formerly *Sulfolobus*) *solfataricus* (Londei et al., 1986), an extreme thermophile, and both 30S and 50S subunits of *Haloferax mediterranei*, a halophilic organism (Sanchez et al., 1990, 1996). The challenge here was not only to obtain spontaneous reassembly of the ribosomal particles from a different domain of life, but also to explore how living in extreme environments affected ribosome biogenesis.

The thermophilic archaeon *S. solfataricus* is a particularly interesting case because it thrives optimally at a temperature of 80–85°C and because it is known to have more protein-rich ribosomes than its bacterial counterparts (Schmid and Böck, 1982; Londei et al., 1983). *S. solfataricus* 50S subunits could be functionally reassembled from the separate RNA and protein components only at high temperatures (80°C) and using high polyamine (thermine) concentrations. Interestingly, the best conditions for *Sulfolobus* 50S subunits *in vitro* assembly entailed a two-step procedure such as for the case of the corresponding *E. coli* particles. As in *E. coli*, the first step is performed at a relatively low temperature (60°C) and yields complete but functionally inactive particles. Activation is only achieved upon incubation at temperatures close to the one optimal for *Sulfolobus* growth (85°C), suggesting the requirement for a temperature-driven conformational change. The presence of a high concentration of the polyamine thermine, which is physiologically present in *S. solfataricus*, is most probably required to stabilize and promote the RNA/protein interactions (Londei et al., 1986).

Notably, however, it was never possible to achieve *in vitro* reconstitution of functional *S. solfataricus* 30S subunits despite the lower complexity of these particles with respect to the 50S ones. More precisely, *in vitro* assembly of 30S particles containing the 16S rRNA and the whole complement of 30S ribosomal proteins was easily obtained, but they were not active in translation (Londei, unpublished). The reason for this unexpected result is unclear. It may be due to the substantially higher protein content of *S. solfataricus* 30S subunits with respect to bacterial particles (28 r-proteins vs 20–21), and/or to the requirement for some additional assembly-promoting factor (see below). If so, biogenesis of *S. solfataricus* 30S subunits may present interesting homologies with the eukaryotic process that would be worth exploring in better detail.

As to halophilic ribosomes, *Haloferax mediterranei* 30S and 50S subunits could be reassembled successfully only at very high concentrations of salt, close to the physiological concentration within the cell. Two types of monovalent cations were the most effective in promoting reconstitution, K^+ and NH_4^+. Unlike what happens for both *E. coli* and *S. solfataricus, H. mediterranei* ribosomes could be reconstituted using a single-step incubation at 42°C., i.e., within the optimal temperature range for physiological growth of this organism. The procedure was similar for 30S and 50S subunits except that reconstitution of 30S subunits had a higher tolerance to ionic strength than that of 50S subunits and was independent of the Mg^{2+} concentration present in the assay (Sanchez et al., 1990, 1996).

One important outcome of the *in vitro* reconstitution experiments with archaeal ribosomes was the possibility of studying the assembly pathways and to identify the assembly-initiating r-proteins. Indeed, using purified rRNA and r-proteins from *S. solfataricus* large ribosomal subunits, it was shown that the initial RNA–protein interactions leading to the formation of a definite but still incomplete assembly intermediate did not require high temperatures, but took place optimally at about 20°C (Altamura et al., 1991). High temperatures, plus the missing proteins, were instead mandatory to convert the low-temperature assembly intermediate into active complete subunits. The assembly intermediate contains 16 of the 34 total 50S subunit r-proteins; among these, the actual primary RNA-binding proteins were identified by experiments of rRNA binding to membrane-immobilized *S. solfataricus* large subunit proteins. These turned out to be 8–9 r-proteins, well in accordance with the number of primary RNA-binding proteins in bacterial 50S ribosomes. It is probable that some, or even all, of these proteins belong to the universally conserved set of r-proteins, but because their identity was not assessed in the study in question, this cannot be stated with certainty. In any event, that the r-proteins present in the low-temperature-assembly intermediate are the innermost in the body of the 50S subunit was also confirmed by preparing ribosomal "cores," i.e., stripping the outer r-proteins with high concentrations of LiCl, a salt known to disrupt weak RNA/protein interactions (Altamura et al., 1991).

Finally, the availability of methods for *in vitro* reconstitution of archaeal ribosomes allows exploring the degree of evolutionary conservation of the assembly pathways and of rRNA/r-protein interactions. In one study, it was found that incubation of *S. solfataricus* LSU proteins with the 23S rRNAs from a distantly related archaeon (*H. mediterranei*) or from *E. coli* led to the formation of a definite and compact 40S particle, containing most of the proteins previously identified as early assembly proteins in *S. solfataricus*, including all of the primary RNA-binding ones (Altamura et al., 1991). These results suggest that the basic architecture of the ribosome and the primary rRNA/r-protein interactions are conserved to a large extent in the two prokaryotic domains of life.

Other data in agreement with this surmise is the complete functional exchangeability of 5S rRNA between *S. solfataricus* and *E. coli* LSUs (Teixidò et al., 1989).

In contrast, incubation of the *S. solfataricus* whole complement of 50S ribosomal proteins with LSU rRNAs from yeast produced no particle, but only an heterogeneous array of RNP complexes, further indicating that both ribosome structure and assembly pathways have undergone a marked divergence from the prokaryotic model in the course of eukaryotic evolution (Altamura et al., 1991).

In summary, probably the most important lesson to be learned from the *in vitro* assembly experiment is that strong similarities exist in the basic architecture and assembly pathways of archaeal and bacterial ribosomes in spite of the presence of unique features in both and of certain "eukaryotic" features in archaea, especially as regards rRNA structure and maturation. The greater complexity of ribosome assembly in eukaryotes is best documented by the fact that, despite many efforts, *in vitro* reconstitution of functional eukaryotic ribosomes from the separated components was largely unsuccessful. The one study claiming success in this task was performed with the ribosomes of *Dictyostelium discoideum* (Mangiarotti and Chiaberge, 1997). Interestingly, *in vitro* assembly of functional *D. discoideum* ribosomes could not be achieved using 18S and 28S rRNA species from mature cytoplasmic ribosomes but required still immature rRNA extracted from nuclear ribosomes. Furthermore, a small RNA species—presumably nucleolar—is apparently required for successful reconstitution. Although this study was never replicated, it agrees with the fact that ribosome assembly is inherently more complex in eukaryotes, developing along a pathway that makes use of many additional extra-ribosomal nuclear/nucleolar factors. Also, the similarity in operon organization and in processing pathways of archaeal and bacterial rRNAs with respect to the eukaryotic ones speaks in favor of a greater evolutionary conservation between the two prokaryotic domains. The presence of a single cellular compartment in which everything happens, from transcription of rRNAs, to maturation of rRNA transcripts, to ribosome assembly and activation, must have dictated the need for a simpler and more streamlined process of ribosome biogenesis than it is the case for eukaryotes. However, more work is required to assess these points, especially *in vivo* experiments, which, at present, are almost completely lacking in archaea.

RIBOSOME BIOGENESIS FACTORS: ARCHAEAL SPECIFICITY AND SHARED FEATURES

Ribosome biogenesis also requires the participation of additional ribosome biogenesis factors, also known as assembly factors or *trans*-acting factors. These factors have been analyzed in great detail in bacteria and eukaryotes. Generally, these factors transiently interact with the nascent ribosomal subunits and are believed to facilitate the ribosome biogenesis process. Among these factors, a significant fraction homes various enzymatic activity, mostly NTPase activity (ATPase, GTPase, and RNA helicases. . .), which may contribute to promote energy-dependent steps of the ribosomal subunit biogenesis process. Interestingly, whereas GTP-dependent processes are predominant in bacteria, ATP-dependent processes are strikingly more frequent in Eukaryotes (Shajani et al., 2011; Thomson et al., 2013; Davis and Williamson, 2017; Baßler and Hurt, 2019; Klinge and Woolford, 2019). Paradoxically, and despite the universal conservation of the ribosomal subunits, most of the ribosome biogenesis factors are (1) not conserved across evolution, and (2) their numbers are dramatically increasing in eukaryotes (Hage and Tollervey, 2004; Ebersberger et al., 2014; Ferreira-Cerca, 2017; **Figures 1**, **3**). This observation suggests that the ribosome biogenesis pathway has been reengineered multiple times during evolution and may reflect early adaptation to molecular constraints present within the respective cellular lineage ancestors. Still, there are remarkable similarities and/or analogies between the different ribosome biogenesis pathways

that may exist and are worth being highlighted. First, the presence of ribosome biogenesis factor sequence homologs between archaea and eukaryotes suggests a common origin of the archaeal–eukaryotic ribosome biogenesis pathway (Ebersberger et al., 2014). Intriguingly, these sequence homologs are known to predominantly act during the latest steps of eukaryotic SSU and LSU maturation. Second, the presence of structural and/or functional mimicry conserved across the tree of life suggests that, despite an apparent sequence/structure divergence between most ribosome biogenesis factors, some steps have similar molecular constraints across the tree of life that are overcome by functionally equivalent molecular inventions [discussed in Ferreira-Cerca (2017); Jüttner et al. (2020)]. This seems to be particularly true in the context of the late steps of the small ribosomal subunit biogenesis (Ferreira-Cerca, 2017; Knüppel et al., 2018). Notably, despite the absence of apparent sequence and structural conservation between bacterial and eukaryotic ribosome biogenesis factors, those, for example, involved in the late steps of SSU maturation remarkably cluster within an analogous structural region on the nascent SSU, i.e., regions that form the future functional centers. This suggests that binding of these ribosome biogenesis factors may ensure functional testing and avoid premature release of the nascent ribosomal subunits into the translational pool (Strunk et al., 2011, 2012; Ferreira-Cerca, 2017; Ghalei et al., 2017; Parker et al., 2019).

Furthermore, the ribosome biogenesis factors sequence homologs are not evenly distributed across all archaeal genomes, but follow the reductive evolution trend observed for the r-proteins, thereby suggesting a simplification of the ribosome biogenesis pathway in these organisms, e.g., euryarchaeota or nanoarchaeota, whereas ribosome synthesis in the TACK superphylum may generally be more complex due to the presence of additional r- proteins or ribosome biogenesis factors (Lecompte et al., 2002; Yutin et al., 2012; Ebersberger et al., 2014; Zaremba-Niedzwiedzka et al., 2017; **Figures 1**, **3**). However, in organisms showing an apparent reduced ribosome biogenesis complexity, the addition or molecular exchange by unknown archaeal specific r-proteins and/or ribosome biogenesis factors cannot be fully excluded.

So far, the functional analysis of archaeal ribosome biogenesis factors is rather limited, and only a few have been established *in vivo*. Most of these characterized factors are sequence homologs of genuine eukaryotic ribosome biogenesis factors [see Ebersberger et al. (2014) for a complete list of candidates]. Among them, the dimethyltransferase KsgA/Dim1 [see above and Grünberger et al. (2020); Knüppel et al. (2021)], or the Rio ATPase/Kinase family members are implicated in the late steps of SSU maturation, where they probably play a role similar to their eukaryotic counterparts (Knüppel et al., 2018). Similarly, the endonuclease Nob1 is implicated in the maturation of the 16S rRNA 3′end *in vitro* (Veith et al., 2012; Qi et al., 2020; **Figure 2**). Collectively, these analyses suggest that the late steps of archaeal SSU biogenesis is a simplified version of the late steps of eukaryotes SSU maturation (Ferreira-Cerca, 2017; Knüppel et al., 2018). However, the degree of functional conservation and interactions of ribosome biogenesis factors such as the archaeal homologs of Rio1, Fap7, Dim1, Pno1, or Nob1, which form an important functional network involved in the late steps of eukaryotic SSU maturation, remains to be explored. Gaining information on these points will surely offer important insights on the molecular evolution and adaptation of the ribosome biogenesis pathway.

Last, the endonuclease *endA* known to be involved in the maturation of intron-containing tRNAs (Clouet-d'Orval et al., 2018) has been recently implicated in rRNA processing, thereby indicating a functional coordination of tRNA and rRNA maturation in archaea (Qi et al., 2020; Schwarz et al., 2020; **Figure 2**).

PERSPECTIVES AND OUTLOOK

Among the numerous challenges and outstanding questions ahead, the comprehensive identification and functional characterization of factors implicated in archaeal ribosome biogenesis are a key step to further understanding the common and specific features of archaeal ribosome biogenesis. In addition, recent improvement of cryo-electron microscopy analysis has been instrumental to better characterize bacterial and eukaryotic ribosome biogenesis pathways (Davis and Williamson, 2017; Baßler and Hurt, 2019; Klinge and Woolford, 2019). A similar revolution is still to come in the archaeal ribosome biogenesis field and will be important to decipher functional and structural analogies conserved across the tree of life and further improve our view on the evolutionary history of the ribosome biogenesis pathway and how molecular and environmental constraints may have (re-)shaped the ribosome biogenesis molecular dance.

Furthermore, and as discussed, the ribosome biogenesis sequence homologs and r-proteins are not ubiquitously distributed across archaea. Therefore, it is of interest to define the extent of archaeal ribosome biogenesis diversity and functional adaptation (Seistrup et al., 2017; Birkedal et al., 2020; Sas-Chen et al., 2020; Knüppel et al., 2021). Additionally, future metagenomics analyses will certainly increase the numbers of newly identified archaea. Accordingly, learning from archaeal biodiversity, changes and adaptation of the ribosome biogenesis pathway are expected to be discovered; however, the formal analysis of this biodiversity is only possible with the advance of culturomics (Bilen et al., 2018; Lewis et al., 2021) and the fast implementation of genetic manipulation in multiple archaeal organisms.

AUTHOR CONTRIBUTIONS

PL and SF-C wrote the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

Work in the SF-C laboratory is supported by intramural funding from the department of Biochemistry III "House of the Ribosome," by the German Research Foundation

(DFG): individual research grant (FE1622/2-1), and collaborative research center SFB/CRC 960 (SFB960-AP1 and SFB960-B13) "RNP biogenesis: assembly of ribosomes and non-ribosomal RNPs and control of their function." Open access publishing of this work was partly supported by the German Research Foundation (DFG) within the funding program Open Access Publishing.

ACKNOWLEDGMENTS

The authors would like to apologize to our colleagues whose work could not be highlighted and/or discussed in this review. SF-C like to acknowledge Michael Jüttner and Robert Knüppel (University of Regensburg) for fruitful and inspiring discussions on archaeal ribosome biogenesis evolution.

REFERENCES

Agalarov, S., Yusupov, M., and Yusupova, G. (2016). "Reconstitution of Functionally Active Thermus thermophilus 30S Ribosomal Subunit from Ribosomal 16S RNA and Ribosomal Proteins," in *Nucleic Acid Crystallography: Methods and Protocols*, ed. E. Ennifar (New York, NY: Springer New York), 303–314. doi: 10.1007/978-1-4939-2763-0_19

Albers, S.-V., Forterre, P., Prangishvili, D., and Schleper, C. (2013). The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. *Nat. Rev. Micro.* 11, 713–719. doi: 10.1038/nrmicro3124

Altamura, S., Caprini, E., Sanchez, M., and Londei, P. (1991). Early assembly proteins of the large ribosomal subunit of the thermophilic archaebacterium Sulfolobus. Identification and binding to heterologous rRNA species. *J. Biol. Chem.* 266, 6195–6200. doi: 10.1016/s0021-9258(18)38103-1

Armache, J.-P., Anger, A. M., Márquez, V., Franckenberg, S., Fröhlich, 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

Bahram, M., Anslan, S., Hildebrand, F., Bork, P., and Tedersoo, L. (2019). Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. *Environ. Microbiol. Rep.* 11, 487–494. doi: 10.1111/1758-2229.12684

Ban, N., Beckmann, R., Cate, J. H. D., 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.

Baßler, J., and Hurt, E. (2019). Eukaryotic Ribosome Assembly. *Annu. Rev. Biochem.* 88, 281–306. doi: 10.1146/annurev-biochem-013118-110817

Bilen, M., Dufour, J.-C., Lagier, J.-C., Cadoret, F., Daoud, Z., Dubourg, G., et al. (2018). The contribution of culturomics to the repertoire of isolated human bacterial and archaeal species. *Microbiome* 6:94.

Birkedal, U., Beckert, B., Wilson, D. N., and Nielsen, H. (2020). The 23S Ribosomal RNA From Pyrococcus furiosus Is Circularly Permuted. *Front. Microbiol.* 11:582022. doi: 10.3389/fmicb.2020.582022

Boccaletto, P., Machnicka, M. A., Purta, E., Piątkowski, P., Bagiński, B., Wirecki, T. K., et al. (2018). MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* 46, D303–D307.

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

Breuer, R., Gomes-Filho, J.-V., and Randau, L. (2021). Conservation of Archaeal C/D Box sRNA-Guided RNA Modifications. *Front. Microbiol.* 12:654029. doi: 10.3389/fmicb.2021.654029

Brewer, T. E., Albertsen, M., Edwards, A., Kirkegaard, R. H., Rocha, E. P. C., and Fierer, N. (2020). Unlinked rRNA genes are widespread among bacteria and archaea. *ISME J.* 14, 597–608. doi: 10.1038/s41396-019-0552-3

Bunner, A. E., Nord, S., Wikström, P. M., and Williamson, J. R. (2010). The Effect of Ribosome Assembly Cofactors on In Vitro 30S Subunit Reconstitution. *J. Mol. Biol.* 398, 1–7. doi: 10.1016/j.jmb.2010.02.036

Cheng, J., Baßler, J., Fischer, P., Lau, B., Kellner, N., Kunze, R., et al. (2019). Thermophile 90S Pre-ribosome Structures Reveal the Reverse Order of Co-transcriptional 18S rRNA Subdomain Integration. *Mol Cell* 75, 1256–1269.e7.

Cheng, J., Kellner, N., Berninghausen, O., Hurt, E., and Beckmann, R. (2017). 3.2-Å-resolution structure of the 90S preribosome before A1 pre-rRNA cleavage. *Nat. Struct. Mol. Biol.* 24, 954–964. doi: 10.1038/nsmb.3476

Ciammaruconi, A., and Londei, P. (2001). In Vitro Processing of the 16S rRNA of the Thermophilic Archaeon Sulfolobus solfataricus. *J. Bacteriol.* 183, 3866–3874. doi: 10.1128/jb.183.13.3866-3874.2001

Clouet-d'Orval, B., Batista, M., Bouvier, M., Quentin, Y., Fichant, G., Marchfelder, A., et al. (2018). Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiol. Rev.* 42, 579–613. doi: 10.1093/femsre/fuy016

Cockram, C., Thierry, A., Gorlas, A., Lestini, R., and Koszul, R. (2021). Euryarchaeal genomes are folded into SMC-dependent loops and domains, but lack transcription-mediated compartmentalization. *Mol. Cell* 81, 459–472.e10.

Connolly, K., Rife, J. P., and Culver, G. (2008). Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. *Mol. Microbiol.* 70, 1062–1075. doi: 10.1111/j.1365-2958.2008.06485.x

Coureux, 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.

Culver, G. M. (2003). Assembly of the 30S ribosomal subunit. *Biopolymers* 68, 234–249.

Czekay, D. P., and Kothe, U. (2021). H/ACA Small Ribonucleoproteins: structural and Functional Comparison Between Archaea and Eukaryotes. *Front. Microbiol.* 12:654370. doi: 10.3389/fmicb.2021.654370

Dao Duc, K., Batra, S. S., Bhattacharya, N., Cate, J. H. D., and Song, Y. S. (2019). Differences in the path to exit the ribosome across the three domains of life. *Nucleic Acids Res.* 47, 4198–4210. doi: 10.1093/nar/gkz106

Davis, J. H., and Williamson, J. R. (2017). Structure and dynamics of bacterial ribosome biogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160181. doi: 10.1098/rstb.2016.0181

de la Cruz, J., Karbstein, K., and Woolford, J. L. (2015). Functions of Ribosomal Proteins in Assembly of Eukaryotic Ribosomes In Vivo. *Annu. Rev. Biochem.* 84, 93–129. doi: 10.1146/annurev-biochem-060614-033917

Dennis, P. P., Tripp, V., Lui, L., Lowe, T., and Randau, L. (2015). C/D box sRNA-guided 2′-O-methylation patterns of archaeal rRNA molecules. *BMC Genomics* 16:632. doi: 10.1186/s12864-015-1839-z

Díaz-López, I., Toribio, R., Berlanga, J. J., and Ventoso, I. (2019). An mRNA-binding channel in the ES6S region of the translation 48S-PIC promotes RNA unwinding and scanning. *Elife* 8:e48246.

Durovic, P., and Dennis, P. P. (1994). Separate pathways for excision and processing of 16S and 23S rRNA from the primary rRNA operon transcript from the hyperthermophilic archaebacterium Sulfolobus acldocaldarius: similarities to eukaryotic rRNA processing. *Mol. Microbiol.* 13, 229–242. doi: 10.1111/j.1365-2958.1994.tb00418.x

Ebersberger, I., Simm, S., Leisegang, M. S., Schmitzberger, P., Mirus, O., von Haeseler, A., et al. (2014). The evolution of the ribosome biogenesis pathway from a yeast perspective. *Nucleic Acids Res.* 42, 1509–1523. doi: 10.1093/nar/gkt1137

Ferreira-Cerca, S. (2017). "Life and Death of Ribosomes in Archaea," in *RNA Metabolism and Gene Expression in Archaea*, ed. B. Clouet-d'Orval (Cham: Springer International Publishing), 129–158. doi: 10.1007/978-3-319-65795-0_6

Ferreira-Cerca, S., Pöll, G., Kühn, H., Neueder, A., Jakob, S., Tschochner, H., et al. (2007). Analysis of the In Vivo Assembly Pathway of Eukaryotic 40S Ribosomal Proteins. *Mol. Cell* 28, 446–457. doi: 10.1016/j.molcel.2007.09.029

Formenoy, L. J., Cunningham, P. R., Nurse, K., Pleij, C. W. A., and Ofengand, J. (1994). Methylation of the conserved A1518-A1519 in *Escherichia coli* 16S ribosomal RNA by the ksgA methyltransferase is influenced by methylations around the similarly conserved U1512-G1523 base pair in the 3′ terminal hairpin. *Biochimie* 76, 1123–1128. doi: 10.1016/0300-9084(94)90040-x

Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S., and Woese, C. R. (1977). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. U. S. A.* 74, 4537–4541. doi: 10.1073/pnas.74.10.4537

Fujii, K., Susanto, T. T., Saurabh, S., and Barna, M. (2018). Decoding the Function of Expansion Segments in Ribosomes. *Mol. Cell* 72, 1013–1020.e6.

Gerbi, S. (1996). "Expansion segments: regions of variable size that interrupt the universal core secondary structure of ribosomal RNA," in *Ribosomal RNA Structure, Evolution, Processing, and Function in Protein Biosynthesis*, eds R. A. Zimmermann and A. E. Dahlberg (Boca Raton, FL: Telford-CRC Press), 71–87.

Gerbi, S. A. (1986). The evolution of eukaryotic ribosomal DNA. *Biosystems* 19, 247–258. doi: 10.1016/0303-2647(86)90001-8

Ghalei, H., Trepreau, J., Collins, J. C., Bhaskaran, H., Strunk, B. S., and Karbstein, K. (2017). The ATPase Fap7 Tests the Ability to Carry Out Translocation-like Conformational Changes and Releases Dim1 during 40S Ribosome Maturation. *Mol. Cell* 67, 990–1000.e3.

Green, R., and Noller, H. F. (1999). Reconstitution of Functional 50S Ribosomes from in Vitro Transcripts of Bacillus stearothermophilus 23S rRNA. *Biochemistry* 38, 1772–1779. doi: 10.1021/bi982246a

Grosjean, H., Breton, M., Sirand-Pugnet, P., Tardy, F., Thiaucourt, F., Citti, C., et al. (2014). Predicting the Minimal Translation Apparatus: lessons from the Reductive Evolution of Mollicutes. *PLoS Genet.* 10:e1004363. doi: 10.1371/journal.pgen.1004363

Grosjean, H., Gaspin, C., Marck, C., Decatur, W. A., and de Crécy-Lagard, V. (2008). RNomics and Modomics in the halophilic archaea Haloferax volcanii: identification of RNA modification genes. *BMC Genomics* 9:470. doi: 10.1186/1471-2164-9-470

Grünberger, F., Knüppel, R., Jüttner, M., Fenk, M., Borst, A., Reichelt, R., et al. (2020). Exploring prokaryotic transcription, operon structures, rRNA maturation and modifications using Nanopore-based native RNA sequencing. *BioRxiv* [Preprint]. doi: 10.1101/2019.12.18.880849

Hadjiolov, A. A. (1985). *The Nucleolus and Ribosome Biogenesis*. Vienna: Springer-Verlag Wien.

Hage, A. E., and Tollervey, D. (2004). A Surfeit of Factors: why is Ribosome Assembly So Much More Complicated in Eukaryotes than Bacteria? *RNA Biol.* 1, 9–14. doi: 10.4161/rna.1.1.932

Henras, A. K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* 6, 225–242. doi: 10.1002/wrna.1269

Higo, K., Held, W., Kahan, L., and Nomura, M. (1973). Functional correspondence between 30S ribosomal proteins of *Escherichia coli* and Bacillus stearothermophilus. *Proc. Natl. Acad. Sci. U. S. A.* 70, 944–948. doi: 10.1073/pnas.70.3.944

Jüttner, M., Weiß, M., Ostheimer, N., Reglin, C., Kern, M., Knüppel, R., et al. (2020). A versatile cis-acting element reporter system to study the function, maturation and stability of ribosomal RNA mutants in archaea. *Nucleic Acids Res.* 48, 2073–2090. doi: 10.1093/nar/gkz1156

Kadumuri, R. V., and Janga, S. C. (2018). Epitranscriptomic Code and Its Alterations in Human Disease. *Trends Mol. Med.* 24, 886–903. doi: 10.1016/j.molmed.2018.07.010

Klappenbach, J. A., Saxman, P. R., Cole, J. R., and Schmidt, T. M. (2001). rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Res.* 29, 181–184. doi: 10.1093/nar/29.1.181

Klinge, S., and Woolford, J. L. (2019). Ribosome assembly coming into focus. *Nat. Rev. Mol. Cell Biol.* 20, 116–131. doi: 10.1038/s41580-018-0078-y

Knüppel, R., Christensen, R. H., Gray, F. C., Esser, D., Strauß, D., Medenbach, J., et al. (2018). Insights into the evolutionary conserved regulation of Rio ATPase activity. *Nucleic Acids Res.* 46, 1441–1456. doi: 10.1093/nar/gkx1236

Knüppel, R., Trahan, C., Kern, M., Wagner, A., Grünberger, F., Hausner, W., et al. (2021). Insights into synthesis and function of KsgA/Dim1-dependent rRNA modifications in archaea. *Nucleic Acids Res.* 49, 1662–1687. doi: 10.1093/nar/gkaa1268

Krogh, N., Jansson, M. D., Häfner, S. J., Tehler, D., Birkedal, U., Christensen-Dalsgaard, M., et al. (2016). Profiling of 2′-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.* 44, 7884–7895. doi: 10.1093/nar/gkw482

Lafontaine, D., Delcour, J., Glasser, A.-L., Desgres, J., and Vandenhaute, J. (1994). The DIM1 Gene Responsible for the Conserved m62Am62A Dimethylation in the 3′-Terminal Loop of 18 S rRNA is Essential in Yeast. *J. Mol. Biol.* 241, 492–497. doi: 10.1006/jmbi.1994.1525

Lafontaine, D. L. J., and Tollervey, D. (1998). Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem. Sci.* 23, 383–388. doi: 10.1016/s0968-0004(98)01260-2

Lecompte, O., Ripp, R., Thierry, J., 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

Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z., and Ettema, T. J. G. (2021). Innovations to culturing the uncultured microbial majority. *Nat. Rev. Microbiol.* 19, 225–240. doi: 10.1038/s41579-020-00458-8

Li, S., and Mason, C. E. (2014). The Pivotal Regulatory Landscape of RNA Modifications. *Annu. Rev. Genom. Hum. Genet.* 15, 127–150. doi: 10.1146/annurev-genom-090413-025405

Littlefield, J. W., and Dunn, D. B. (1958). Natural Occurrence of Thymine and Three Methylated Adenine Bases in Several Ribonucleic Acids. *Nature* 181, 254–255. doi: 10.1038/181254a0

Liu, K., Santos, D. A., Hussmann, J. A., Wang, Y., Sutter, B. M., Weissman, J. S., et al. (2021). Regulation of translation by methylation multiplicity of 18S rRNA. *Cell Rep.* 34:108825. doi: 10.1016/j.celrep.2021.108825

Liu, Y., Makarova, K. S., Huang, W.-C., Wolf, Y. I., Nikolskaya, A. N., Zhang, X., et al. (2021). Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 593, 553–557. doi: 10.1038/s41586-021-03494-3

Londei, P., Teichner, A., Cammarano, P., De Rosa, M., and Gambacorta, A. (1983). Particle weights and protein composition of the ribosomal subunits of the extremely thermoacidophilic archaebacterium Caldariella acidophila. *Biochem. J.* 209, 461–470. doi: 10.1042/bj2090461

Londei, P., Teixidò, J., Acca, M., Cammarano, P., and Amils, R. (1986). Total reconstitution of active large ribosomal subunits of the thermoacidophilic archaebacterium Sulfolobus solfataricus. *Nucleic Acids Res.* 14, 2269–2285.

Mangiarotti, G., and Chiaberge, S. (1997). Reconstitution of Functional Eukaryotic Ribosomes fromDictyostelium discoideum Ribosomal Proteins and RNA. *J. Biol. Chem.* 272, 19682–19687. doi: 10.1074/jbc.272.32.19682

Márquez, V., Fröhlich, T., Armache, J.-P., Sohmen, D., Dönhöfer, A., Mikolajka, A., et al. (2011). Proteomic Characterization of Archaeal Ribosomes Reveals the Presence of Novel Archaeal-Specific Ribosomal Proteins. *J. Mol. Biol.* 405, 1215–1232. doi: 10.1016/j.jmb.2010.11.055

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* 19, 560–567. doi: 10.1038/nsmb.2313

Melnikov, S., Manakongtreecheep, K., and Söll, 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

Mendler, K., Chen, H., Parks, D. H., Lobb, B., Hug, L. A., and Doxey, A. C. (2019). AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* 47, 4442–4448. doi: 10.1093/nar/gkz246

Mullineux, S.-T., and Lafontaine, D. L. J. (2012). Mapping the cleavage sites on mammalian pre-rRNAs: where do we stand? *Biochimie* 94, 1521–1532. doi: 10.1016/j.biochi.2012.02.001

Nakao, A., Yoshihama, M., and Kenmochi, N. (2004). RPG: the Ribosomal Protein Gene database. *Nucleic Acids Res.* 32, D168–D170.

Nierhaus, K. H. (1991). The assembly of prokaryotic ribosomes. *Biochimie* 73, 739–755. doi: 10.1016/0300-9084(91)90054-5

Nierhaus, K. H., and Dohme, F. (1974). Total reconstitution of functionally active 50S ribosomal subunits from *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 71, 4713–4717. doi: 10.1073/pnas.71.12.4713

Nierhaus, K. H., and Lafontaine, D. L. (2004). "Ribosome Assembly," in *Protein Synthesis and Ribosome Structure*, esd K. H. Nierhaus and D. N. Wilson (KGaA: Wiley-VCH Verlag GmbH & Co), 85–143. doi: 10.1002/3527603433.ch3

Nikolay, R., Hilal, T., Qin, B., Mielke, T., Bürger, J., Loerke, J., et al. (2018). Structural Visualization of the Formation and Activation of the 50S Ribosomal Subunit during In Vitro Reconstitution. *Mol. Cell* 70, 881–893.e3.

Nomura, M. (1999). Regulation of Ribosome Biosynthesis in *Escherichia coli* and Saccharomyces cerevisiae: diversity and Common Principles. *J. Bacteriol.* 181, 6857–6864. doi: 10.1128/jb.181.22.6857-6864.1999

Nürenberg-Goloub, E., Kratzat, H., Heinemann, H., Heuer, A., Kötter, 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.

Omer, A. D., Ziesche, S., Decatur, W. A., Fournier, M. J., and Dennis, P. P. (2003). RNA-modifying machines in archaea. *Mol. Microbiol.* 48, 617–629. doi: 10.1046/j.1365-2958.2003.03483.x

Parker, M. D., Collins, J. C., Korona, B., Ghalei, H., and Karbstein, K. (2019). A kinase-dependent checkpoint prevents escape of immature ribosomes into

the translating pool. *PLoS Biol.* 17:e3000329. doi: 10.1371/journal.pbio.3000329

Parker, M. S., Sallee, F. R., Park, E. A., and Parker, S. L. (2015). Homoiterons and expansion in ribosomal RNAs. *FEBS Open Bio.* 5, 864–876. doi: 10.1016/j.fob.2015.10.005

Penev, P. I., Fakhretaha-Aval, S., Patel, V. J., Cannone, J. J., Gutell, R. R., Petrov, A. S., et al. (2020). Supersized Ribosomal RNA Expansion Segments in Asgard Archaea. *Genome Biol. Evol.* 12, 1694–1710. doi: 10.1093/gbe/evaa170

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.

Piekna-Przybylska, D., Decatur, W. A., and Fournier, M. J. (2008). The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res.* 36, D178–D183.

Qi, L., Li, J., Jia, J., Yue, L., and Dong, X. (2020). Comprehensive analysis of the pre-ribosomal RNA maturation pathway in a methanoarchaeon exposes the conserved circularization and linearization mode in archaea. *RNA Biol.* 17, 1427–1441. doi: 10.1080/15476286.2020.1771946

Ramesh, M., and Woolford, J. L. Jr. (2016). Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. *RNA* 22, 1153–1162. doi: 10.1261/rna.056705.116

Roth, H. E., and Nierhaus, K. H. (1980). Assembly Map of the 50-S Subunit from *Escherichia coli* Ribosomes, Covering the Proteins Present in the First Reconstitution Intermediate Particle. *Eur. J. Biochem.* 103, 95–98. doi: 10.1111/j.1432-1033.1980.tb04292.x

Ruggero, D., Ciammaruconi, A., and Londei, P. (1998). The chaperonin of the archaeon Sulfolobus solfataricus is an RNA-binding protein that participates in ribosomal RNA processing. *EMBO J.* 17, 3471–3477. doi: 10.1093/emboj/17.12.3471

Sanchez, E. M., Londei, P., and Amils, R. (1996). Total reconstitution of active small ribosomal subunits of the extreme halophilic archaeon Haloferax mediterranei. *Biochim. Biophys. Acta* 1292, 140–144. doi: 10.1016/0167-4838(95)00179-4

Sanchez, M. E., Urena, D., Amils, R., and Londei, P. (1990). In vitro reassembly of active large ribosomal subunits of the halophilic archaebacterium Haloferax mediterranei. *Biochemistry* 29, 9256–9261. doi: 10.1021/bi00491a021

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

Schmid, G., and Böck, A. (1982). The ribosomal protein composition of five methanogenic bacteria. *Zentralblatt Für Bakteriologie Mikrobiologie Und Hygiene: I. Abt. Originale C: Allgemeine, Angewandte Und Ökologische Mikrobiologie* 3, 347–353. doi: 10.1016/s0721-9571(82)80013-6

Schwarz, T. S., Berkemer, S. J., Bernhart, S. H., Weiß, M., Ferreira-Cerca, S., Stadler, P. F., et al. (2020). Splicing Endonuclease Is an Important Player in rRNA and tRNA Maturation in Archaea. *Front. Microbiol.* 11:594838. doi: 10.3389/fmicb.2020.594838

Seistrup, K. H., Rose, S., Birkedal, U., Nielsen, H., Huber, H., and Douthwaite, S. (2017). Bypassing rRNA methylation by RsmA/Dim1during ribosome maturation in the hyperthermophilic archaeon Nanoarchaeum equitans. *Nucleic Acids Res.* 45, 2007–2015.

Shajani, Z., Sykes, M. T., and Williamson, J. R. (2011). Assembly of Bacterial Ribosomes. *Annu. Rev. Biochem.* 80, 501–526. doi: 10.1146/annurev-biochem-062608-160432

Shankar, V., Rauscher, R., Reuther, J., Gharib, W. H., Koch, M., and Polacek, N. (2020). rRNA expansion segment 27Lb modulates the factor recruitment capacity of the yeast ribosome and shapes the proteome. *Nucleic Acids Res.* 48, 3244–3256. doi: 10.1093/nar/gkaa003

Sharma, S., and Lafontaine, D. L. J. (2015). 'View From A Bridge': a New Perspective on Eukaryotic rRNA Base Modification. *Trends Biochem. Sci.* 40, 560–575. doi: 10.1016/j.tibs.2015.07.008

Sleiman, S., and Dragon, F. (2019). Recent Advances on the Structure and Function of RNA Acetyltransferase Kre33/NAT10. *Cells* 8:1035. doi: 10.3390/cells8091035

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K.-D., Lafontaine, D. L. J., and Bohnsack, M. T. (2016). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781

Sobolev, A., Solovyev, M., Ivanova, V., Kochetkova, T., Merkel, A., Razin, S., et al. (2021). 3C-seq-captured chromosome conformation of the hyperthermophilic archaeon Thermofilum adornatum. *BioRxiv* [Preprint]. doi: 10.1101/2021.04.30.439615

Spang, A., Saw, J. H., Jorgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., et al. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179. doi: 10.1038/nature14447

Stepanov, V. G., and Fox, G. E. (2021). Expansion segments in bacterial and archaeal 5S ribosomal RNAs. *RNA* 27, 133–150. doi: 10.1261/rna.077123.120

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., and Schmidt, T. M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* 43, D593–D598.

Strunk, B. S., Loucks, C. R., Su, M., Vashisth, H., Cheng, S., Schilling, J., et al. (2011). Ribosome Assembly Factors Prevent Premature Translation Initiation by 40S Assembly Intermediates. *Science* 333, 1449–1453. doi: 10.1126/science.1208245

Strunk, B. S., Novak, M. N., Young, C. L., and Karbstein, K. (2012). A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell* 150, 111–121. doi: 10.1016/j.cell.2012.04.044

Takemata, N., and Bell, S. D. (2021). Multi-scale architecture of archaeal chromosomes. *Mol. Cell* 81, 473–487.e6.

Tang, T. H., Rozhdestvensky, T. S., d'Orval, B. C., Bortolin, M.-L., Huber, H., Charpentier, B., et al. (2002). RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res.* 30, 921–930. doi: 10.1093/nar/30.4.921

Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., et al. (2018). Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* 46, 9289–9298. doi: 10.1093/nar/gky811

Teixidò, J., Altamura, S., Londei, P., and Amils, R. (1989). Structural and functional exchangeability of 5 S RNA species from the eubacterium E.coli and the thermoacidophilic archaebacterium Sulfolobus solfataricus. *Nucleic Acids Res.* 17, 845–851. doi: 10.1093/nar/17.3.845

Thomson, E., Ferreira-Cerca, S., and Hurt, E. (2013). Eukaryotic ribosome biogenesis at a glance. *J. Cell Sci.* 126, 4815–4821. doi: 10.1242/jcs.111948

Tirumalai, M. R., Kaelber, J. T., Park, D. R., Tran, Q., and Fox, G. E. (2020). Cryo-electron microscopy visualization of a large insertion in the 5S ribosomal RNA of the extremely halophilic archaeon Halococcus morrhuae. *FEBS Open Bio.* 10, 1938–1946. doi: 10.1002/2211-5463.12962

Traub, P., and Nomura, M. (1968). Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc. Natl. Acad. Sci. U. S. A.* 59, 777–784. doi: 10.1073/pnas.59.3.777

Van Buul, C. P. J. J., Hamersma, M., Visser, W., and Van Knippenberg, P. H. (1984). Partial methylation of two adjacent adenosines in ribosomes from Euglena gracilis chloroplasts suggests evolutionary loss of an intermediate stage in the methyl-transfer reaction. *Nucleic Acids Res.* 12, 9205–9208. doi: 10.1093/nar/12.23.9205

Veith, T., Martin, R., Wurm, J. P., Weis, B. L., Duchardt-Ferner, E., Safferthal, C., et al. (2012). Structural and functional analysis of the archaeal endonuclease Nob1. *Nucleic Acids Res.* 40, 3259–3274. doi: 10.1093/nar/gkr1186

Vogel, D. W., Hartmann, R. K., Bartsch, M., Subramanian, A. R., Kleinow, W., O'Brien, T. W., et al. (1984). Reconstitution of 50 S ribosomal subunits from Bacillus stearothermophilus with 5 S RNA from spinach chloroplasts and low-M r RNA from mitochondria of Locusta migratoria and bovine liver. *FEBS Lett.* 169, 67–72. doi: 10.1016/0014-5793(84)80291-4

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* 24, 437–440. doi: 10.1016/s0968-0004(99)01460-7

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., 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

Woolford, J. L., and Baserga, S. J. (2013). Ribosome Biogenesis in the Yeast Saccharomyces cerevisiae. *Genetics* 195, 643–681. doi: 10.1534/genetics.113.153197

Yip, W. S. V., Vincent, N. G., and Baserga, S. J. (2013). Ribonucleoproteins in Archaeal Pre-rRNA Processing and Modification. *Archaea* 2013:614735.

Yutin, N., Puigbò, 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., Bäckström, 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

Zorbas, C., Nicolas, E., Wacheul, L., Huvelle, E., Heurgué-Hamard, V., and Lafontaine, D. L. J. (2015). The human 18S rRNA base methyltransferases DIMT1L and WBSCR22-TRMT112 but not rRNA modification are required for ribosome biogenesis. *Mol. Biol. Cell* 26, 2080–2095.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 September 2021
doi: 10.3389/fmicb.2021.739000

Tracing Eukaryotic Ribosome Biogenesis Factors Into the Archaeal Domain Sheds Light on the Evolution of Functional Complexity

*Mehmet Birikmen[1], Katherine E. Bohnsack[2], Vinh Tran[1], Sharvari Somayaji[1], Markus T. Bohnsack[2,3] and Ingo Ebersberger[1,4,5]**

[1]Applied Bioinformatics Group, Institute of Cell Biology and Neuroscience, Goethe University Frankfurt, Frankfurt, Germany, [2]Department of Molecular Biology, University Medical Center Göttingen, Göttingen, Germany, [3]Göttingen Center for Molecular Biosciences, Georg-August University, Göttingen, Germany, [4]Senckenberg Biodiversity and Climate Research Center (S-BIK-F), Frankfurt, Germany, [5]LOEWE Center for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt, Germany

OPEN ACCESS

Edited by:
Anna La Teana, Polytechnic University of Marche, Italy

Reviewed by:
Fabian Blombach, University College London, United Kingdom
Arlen Johnson, University of Texas at Austin, United States

***Correspondence:**
Ingo Ebersberger
ebersberger@bio.uni-frankfurt.de

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

Received: *09 July 2021*
Accepted: *17 August 2021*
Published: *16 September 2021*

Citation:
Birikmen M, Bohnsack KE, Tran V, Somayaji S, Bohnsack MT and Ebersberger I (2021) Tracing Eukaryotic Ribosome Biogenesis Factors Into the Archaeal Domain Sheds Light on the Evolution of Functional Complexity. Front. Microbiol. 12:739000. doi: 10.3389/fmicb.2021.739000

Ribosome assembly is an essential and carefully choreographed cellular process. In eukaryotes, several 100 proteins, distributed across the nucleolus, nucleus, and cytoplasm, co-ordinate the step-wise assembly of four ribosomal RNAs (rRNAs) and approximately 80 ribosomal proteins (RPs) into the mature ribosomal subunits. Due to the inherent complexity of the assembly process, functional studies identifying ribosome biogenesis factors and, more importantly, their precise functions and interplay are confined to a few and very well-established model organisms. Although best characterized in yeast (*Saccharomyces cerevisiae*), emerging links to disease and the discovery of additional layers of regulation have recently encouraged deeper analysis of the pathway in human cells. In archaea, ribosome biogenesis is less well-understood. However, their simpler sub-cellular structure should allow a less elaborated assembly procedure, potentially providing insights into the functional essentials of ribosome biogenesis that evolved long before the diversification of archaea and eukaryotes. Here, we use a comprehensive phylogenetic profiling setup, integrating targeted ortholog searches with automated scoring of protein domain architecture similarities and an assessment of when search sensitivity becomes limiting, to trace 301 curated eukaryotic ribosome biogenesis factors across 982 taxa spanning the tree of life and including 727 archaea. We show that both factor loss and lineage-specific modifications of factor function modulate ribosome biogenesis, and we highlight that limited sensitivity of the ortholog search can confound evolutionary conclusions. Projecting into the archaeal domain, we find that only few factors are consistently present across the analyzed taxa, and lineage-specific loss is common. While members of the Asgard group are not special with respect to their inventory of ribosome biogenesis factors (RBFs), they unite the highest number of orthologs to eukaryotic RBFs

in one taxon. Using large ribosomal subunit maturation as an example, we demonstrate that archaea pursue a simplified version of the corresponding steps in eukaryotes. Much of the complexity of this process evolved on the eukaryotic lineage by the duplication of ribosomal proteins and their subsequent functional diversification into ribosome biogenesis factors. This highlights that studying ribosome biogenesis in archaea provides fundamental information also for understanding the process in eukaryotes.

Keywords: domain architecture evolution, asgard group, phylogenetic profiles, orthology assignment, evolutionary traceability, pathway complexity, large subunit maturation

INTRODUCTION

Due to the essential role of ribosomes in producing all cellular proteins, their synthesis is among the few pathways that are universally necessary for organismic life. Despite substantial differences in the ways that the ribosomal proteins (RPs) and ribosomal RNAs (rRNAs) are assembled into the small and large ribosomal subunits (SSU and LSU, respectively), the fundamental basis of this process in the three domains of life (Woese et al., 1990) presumably dates back to the last universal common ancestor (LUCA). Ribosome assembly is most simple and best understood in bacteria. Bacterial ribosomes can be assembled *in vitro* without the requirement for any non-ribosomal proteins while up to approximately 30 assembly factors make the process much faster and more accurate *in vivo* (Shajani et al., 2011). In contrast, the functional network mediating the same process in eukaryotes is many-fold more extensive, and it is to date best studied in yeast (*Saccharomyces cerevisiae*; Hage and Tollervey, 2004; Henras et al., 2008; Sloan et al., 2017; Klinge and Woolford, 2019). A recent study listed 255 ribosome biogenesis factors with a confirmed or suspected direct role in this process (Ebersberger et al., 2014). Assigning these factors to different age layers revealed that precisely how eukaryotes mediate ribosome assembly is not cast into stone. Instead, an evolutionarily old set of core functions, whose emergence predates the diversification of eukaryotes, was extended and probably fine-tuned in a lineage-specific manner (Ebersberger et al., 2014). Archaea seem to assume an intermediate position between eukaryotes and bacteria with respect to the complexity of their ribosome assembly pathways (Lecompte et al., 2002; Londei and Ferreira-Cerca, 2021). Shifting focus toward the archaeal domain therefore has the potential to further disentangle the building principles of ribosomes that already evolved prior to the emergence of eukaryotes and were already established in the last eukaryotic common ancestor (LECA). Thus far, phylogenetic profiling indicated that 38 yeast RBFs have counterparts in the archaeal domain (Ebersberger et al., 2014), in parts representing functional sub-clusters that seem specifically involved in the late steps of ribosome maturation. A subset of these factors has been subsequently confirmed as RBFs by functional studies in individual archaeal models (reviewed in Londei and Ferreira-Cerca, 2021). However, we are far from fully comprehending the extent to which archaeal ribosome biogenesis resembles that of eukaryotes, where archaea have implemented alternative strategies, and the degree of diversity within the archaeal domain.

Much of the uncertainty about the common grounds of ribosome biogenesis in archaea and eukaryotes is connected to methodological issues in the large-scale profiling studies performed thus far. The objective of such studies is easily specified: "Identify the functionally equivalent archaeal protein to a eukaryotic RBF, if it is present." Its realization, however, bears numerous pitfalls. Unidirectional searches for sequences with a significant local similarity e.g., with BLAST (Altschul et al., 2009) or Diamond (Buchfink et al., 2015) rapidly identify archaeal RBF candidates at a high sensitivity, however, the specificity is low (Chen et al., 2007). Considering only such proteins as RBF candidates that diverged no longer than the species they reside in (orthologs; e.g., Koonin et al., 2001; Ebersberger et al., 2014) reduces the false positive rate, because orthologs are the best guesses when searching for functionally equivalent proteins across species (Tatusov et al., 1997; Altenhoff et al., 2012). However, the higher specificity comes at the cost of a decrease in sensitivity (Altenhoff et al., 2016). Moreover, also orthologs may have diverged in function (Ebersberger et al., 2014). The latter problem can be ameliorated by scoring the similarity of protein domain architectures (Koestler et al., 2010), which comprise features, such as Pfam (El-Gebali et al., 2019) or Smart (Letunic et al., 2009) domains, transmembrane regions, signal peptides, or regions of biased amino acid composition. Differences in domain architectures can then indicate alterations in the respective functional spectra of the compared proteins (Jiang et al., 2020). In contrast to the ample means to cope with the spurious inference of archaeal RBF candidates based on homologous proteins, false negatives, i.e., proteins overlooked in the homolog searches because their sequences have diverged to an extent that they are no more similar than it is expected by chance, have received very little attention. Such proteins can represent missing links that are essential for concluding on the presence of functional (sub-)network based on phylogenetic profiling analyses (Jain et al., 2019). Their identification requires a targeted increase of the search sensitivity accompanied by a careful downstream curation to validate the results.

Next to the methodological issues outlined above, at least three further aspects leave the current understanding of common concepts in eukaryotic and archaeal ribosome biogenesis incomplete. In recent years, numerous yeast proteins have been discovered as RBFs (e.g., Fujiyama-Nakamura et al., 2009; van Tran et al., 2019) whose presence in the archaea has not been tested in a comprehensive screen, thus far. Moreover, structure- and mass spectrometry-based approaches have shed light on the order of selected events during the assembly process and

the identity of the participating proteins (see for example, Wu et al., 2016; Barandun et al., 2017; Kater et al., 2017; Sanghai et al., 2018; Klingauf-Nerurkar et al., 2020; Liang et al., 2020; Nieto et al., 2020). This resource, which provides an excellent basis for identifying functional sub-networks shared between eukaryotes and archaea, is largely untapped. Moreover, the focus on yeast, a highly derived model organism that lost many genes essential in other species (Peter et al., 2018), is a limiting factor itself. Knowledge on the yeast RBFs is now complemented by large-scale, RNAi-based screens in human cells, which have revealed several 100 proteins that may be either directly or indirectly required for human ribosome biogenesis (Wild et al., 2010; Tafforeau et al., 2013; Badertscher et al., 2015; Farley-Barnes et al., 2018). For a non-negligible fraction of these RBF candidates a yeast homolog is elusive (Fujiyama-Nakamura et al., 2009; Tafforeau et al., 2013; van Tran et al., 2019; Ameismeier et al., 2020). It is conceivable that at least some of these factors act as RBFs in archaea too. Lastly, both the number of archaeal taxa and their phylogenetic diversity in the public sequence databases has dramatically increased in the past few years. This data provides an excellent, yet largely unused, basis for a highly-resolved analysis of the representation of eukaryotic RBFs in the archaeal domain. Among others, it allows to test whether members of the Asgard group share more RBFs with the eukaryotes than other archaea, which would be in line with their suspected placement as the closest relative of the eukaryotes (Zaremba-Niedzwiedzka et al., 2017; Imachi et al., 2020; Liu et al., 2021).

Here, we re-address the evolutionary history of eukaryotic ribosome biogenesis and trace the deep evolutionary roots of this pathway that are shared with the archaea. We base our analysis on a set of 301 manually curated yeast and human RBFs comprising the, to date, most comprehensive collection of eukaryotic ribosome biogenesis factors. Phylogenetic profiles for each protein across more than 900 taxa, among them 727 archaea including representatives of the Asgard group, using targeted and domain architecture aware ortholog searches provide insights into the evolution of this pathway at an unprecedented resolution. The analysis is complemented by identifying RBFs that evolve too quickly to facilitate ortholog identification over longer evolutionary time scales. This helps reconciling the discrepancy between large-scale phylogenetic profiling of RBFs using ortholog/homology assignments claiming the absence of an RBF and experimental evidences showing the presence of the corresponding function. In the same line, it can direct the attention to missing functional links that may require searches at higher sensitivity to warrant their detection.

MATERIALS AND METHODS

Selection of Ribosome Biogenesis Factors

We compiled an initial non-redundant set of 307 yeast proteins putatively involved into ribosome biogenesis. About 255 RBF candidates were obtained from Ebersberger et al. (2014), 41 from the KEGG Brite database (KO3009; Kanehisa et al., 2016), and additionally 11 from recent publications focusing on molecular details of yeast ribosome biogenesis. The candidate list together with the references is given in **Supplementary Table S1**. For the human RBF collection, we seeded the set with 198 proteins that are involved into human ribosome biogenesis according to KEGG (Kanehisa et al., 2016). This data were complemented with 488 proteins with at least a suspected involvement into human ribosome biogenesis according to large scale screening studies (40S/60S: Wild et al., 2010; pre-rRNA processing: Tafforeau et al., 2013; 40S: Badertscher et al., 2015; and regulators: Farley-Barnes et al., 2018). Finally, we added eight RBFs that emerged from a literature screen (Yang et al., 2006; Freed et al., 2012; Wandrey et al., 2015; van Tran et al., 2019; Ameismeier et al., 2020). The non-redundant list of 695 proteins excluding ribosomal proteins is provided in **Supplementary Table S2**. The phylogenetic profiles of all human candidates were determined and served as basis to assess their evolutionary age (see below).

Taxon Collection

We determined the phylogenetic profiles for the RBF candidates across 982 taxa comprising 232 eukaryotes, a diverse collection of 23 bacteria representing 16 phyla, and 727 archaea. Archaeal gene sets were retrieved from the RefSeq partition of the NCBI Genome database (O'Leary et al., 2016; December 2020). The 78 Quest for Orthologs reference proteomes[1] complemented with 189 fungal taxa served as representatives for the bacterial and eukaryotic taxa. The taxon list is provided in **Supplementary Table S3**.

Pathway Analysis

Information about pathway organization and complex composition was obtained from KEGG (Kanehisa et al., 2021) and Reactome v76 (Fabregat et al., 2018).

Domain-Aware Phylogenetic Profiling and Gene Age Estimation

RBF orthologs were identified with the targeted ortholog search tool fDog (Jiang et al., 2020)[2] using the following parameter settings: –checkCoorthologsRef, −-countercheck, −-minDist = family, and −-maxDist = kingdom. Protein domain architectures were compared pair-wise between each seed protein and its orthologs, and an architecture similarity score was computed with FAS[3] implemented into fDOG. In brief, FAS compares the domain architectures of two proteins, using one architecture as the reference and the second architecture as the query. The FAS score ranges between 0 (architectures are completely dissimilar) and 1 (the architecture of the reference is at least a sub-architecture of the query; Koestler et al., 2010). Because the score is not symmetric, we computed FAS scores once using the seed protein as reference (FAS forward), and once using the ortholog as reference (FAS backward). The Domain architecture-aware phylogenetic profiles generated by

[1]https://www.ebi.ac.uk/reference_proteomes/
[2]https://github.com/BIONF/fDog/
[3]https://github.com/BIONF/FAS

fDog were visualized and analyzed with PhyloProfile (Tran et al., 2018). The evolutionary emergence of individual RBF candidates was dated using an LCA algorithm implemented into PhyloProfile. In brief, the last common ancestor (LCA) of the two most distantly related species harboring an ortholog to an RBF candidate was used as a minimal age estimate for the corresponding gene. Comparisons of FAS score distributions and domain architecture comparisons between two taxonomic groups were performed with the group comparison function implemented into PhyloProfile.

Computation of Evolutionary Traceability

Orthologs of quickly evolving proteins may lose a sufficiently high sequence similarity to warrant their detection already over short evolutionary distances. The evolutionary traceability index (Ti) is a simulation-based score on the interval [0,1] that captures, for a protein and a given evolutionary distance, whether an ortholog likely still shares a sufficiently high sequence similarity that allows its detection, or whether it likely has diverged beyond recognition (Jain et al., 2019). We extracted, for each yeast protein, the corresponding traceability indices across 273 taxa distributed across the tree of life from Jain et al. (2019). Traceability indices on a kingdom level were computed as the mean Ti across all members in the kingdom. An RBF was considered as low traceable in case the mean traceability index fell below 0.75 (Jain et al., 2019). Note, we limited this computation to the yeast proteins, since the human RBFs were pre-selected based on the presence of orthologs already in evolutionary distantly related taxa.

Phylogenetic Analyses

Multiple Sequence alignments were generated with Muscle v3.8.1551 (Edgar, 2004) using default settings. Sequence logos were generated with WebLogo (Crooks et al., 2004) provided as a web service available at https://weblogo.berkeley.edu. For maximum likelihood phylogenetic tree reconstruction with IQ-TREE using the LG+G+I model of sequence evolution (Nguyen et al., 2015), alignments were post-processed by removing alignment columns with more than 50% gaps. Phylogenetic trees were visualized with ITol (Letunic and Bork, 2021).

RESULTS

Ribosome Biogenesis From Two Perspectives

Production of ribosomes is essential for organismic life. Here, we set out to identify the common basis of ribosome biogenesis in eukaryotes and archaea using a manually curated list of eukaryotic RBFs (**Figure 1**). The most comprehensive inventory of protein factors involved in eukaryotic ribosome biogenesis is that of the yeast *S. cerevisiae* S288C (Woolford and Baserga, 2013; Bassler and Hurt, 2019). An initial screen for yeast RBFs retrieved 307 candidates (see **Figure 1** for numbers and references). To consider also eukaryotic RBFs that are either absent in yeast or have a function other than RBF, we used an initial set of 686 potential human RBFs obtained largely from comprehensive RNAi-based screens (Wild et al., 2010; Tafforeau et al., 2013; Badertscher et al., 2015; Farley-Barnes et al., 2018) as a second starting point of the analysis.

The phylogenetic profile for each RBF candidate was subsequently determined by searching for orthologs in 232 eukaryotes, 727 archaea, and 23 bacteria. For each detected ortholog, we then computed the domain architecture similarity to that of the corresponding yeast or human seed protein. On this basis, a two-stage curation procedure was devised to extract the final set of RBF candidates for further analysis. In the evolutionary-motivated first stage (**Supplementary Figure S1**; **Supplementary Table S4**), the subset of RBFs that can be traced back to the LCA of all eukaryotes (LECA) was identified. We then removed all proteins with a yeast ortholog that is considered already in the yeast RBF collection leaving 227 human proteins. We implemented this first stage for the human RBF candidates to reduce the number of proteins that enter the second stage of manual curation. In the subsequent curation step, we retained only proteins that fulfill at least one of the following criteria: (i) have a known function in ribosome biogenesis; (ii) have been identified associated with pre-ribosomal particles; or (iii) their depletion or deletion causes a defect in ribosome assembly. This retained 278 yeast proteins (RBF_{yeast}) and 17 human proteins (RBF_{human}). We complemented this data with six known human RBFs, ILF2, ILF3, TMA16, NOL11, NKRF, and ZCCHC4, most likely younger than the LECA. The final set, RBF_{euk}, comprised 301 factors. We added two subunits of the RNA polymerase II, Rpb5, and Rpo21, as positive controls for our profiling approach. The two proteins are evolutionarily highly conserved and orthologs should be identifiable in the majority of the species analyzed here. Thus, the total number of analyzed proteins sums up to 303.

PhyloRBF: Interactive Access to the Data

The phylogenetic profiles for proteins in the RBF_{euk} set across 982 taxa provide the first unifying resource for tracing eukaryotic ribosome biogenesis factors across the organismal diversity. To facilitate easy and interactive exploration and analysis of these data, we provide two options. An online instance of PhyloProfile (Tran et al., 2018), PhyloRBF, provides web-access to these data *via* the URL: https://applbio.biologie.uni-frankfurt.de/phylorbf. Users can display the full data set (**Figure 2A**), customized subsets of RBFs and taxa (**Figure 2B**), zoom in on individual ortholog pairs (**Figure 2C**), and ultimately display the domain architectures of the yeast or human RBF and of its ortholog (**Figure 2D**). Interactive links connect the information about taxon, protein sequence and Pfam (El-Gebali et al., 2019), or SMART domains (Letunic et al., 2009) with the corresponding public databases. For an offline analysis of the RBF_{euk} phylogenetic profiles with a local installation of PhyloProfile, the input data are provided as **Supplementary Data 2**.

The Evolutionary Trajectory of Eukaryotic Ribosome Biogenesis Factors

Figure 3A provides an overview of the phylogenetic profiles for the RBF_{euk} set and the two positive controls. Within

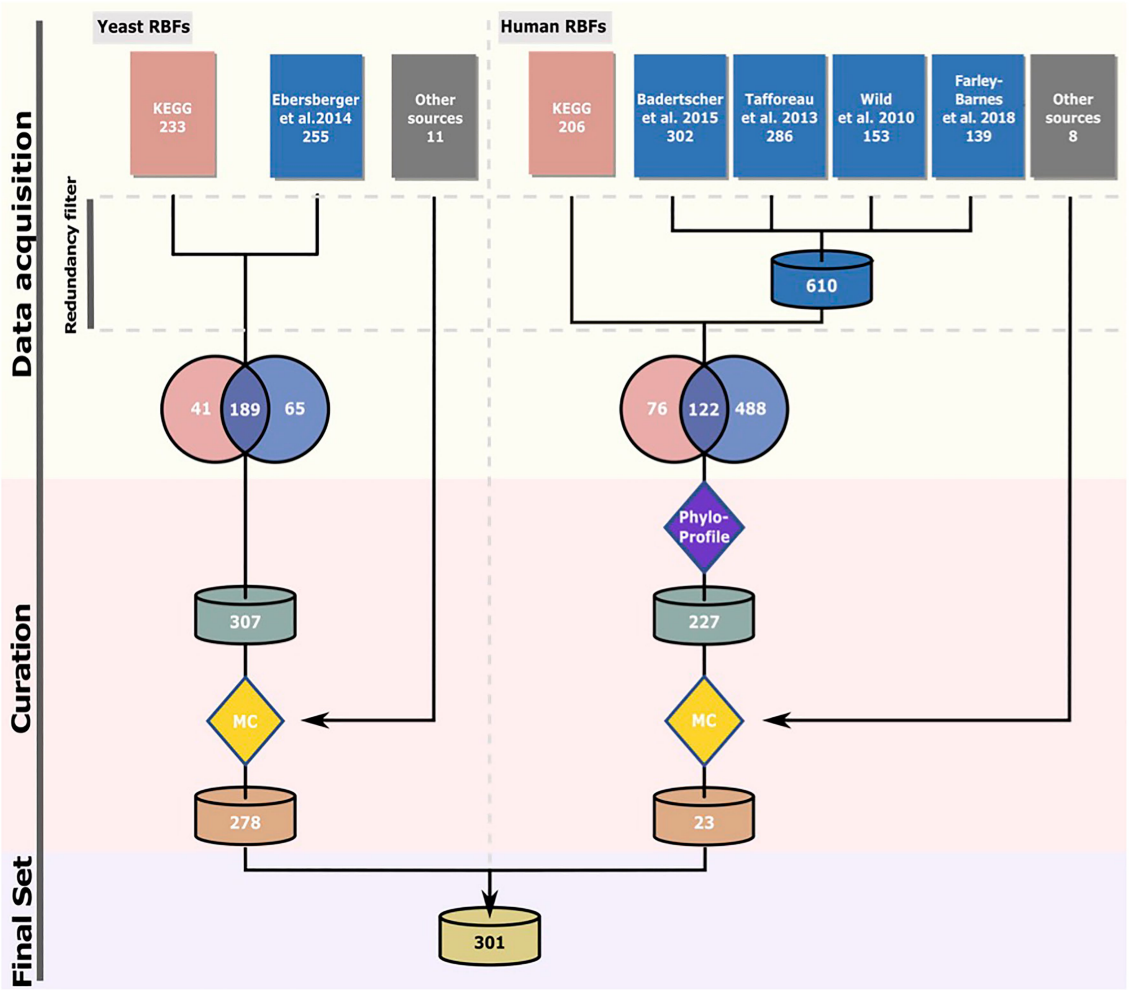


FIGURE 1 | Compilation of the RBF set. The "Phylostratigraphy" filter removed proteins younger than last eukaryotic common ancestor (LECA) and proteins with an ortholog in yeast. In the manual curation (MC) step, only proteins for which experimental evidence supports their association with pre-ribosomal particles or requirement for ribosome assembly were retained. See text for details. The sources for the candidate proteins are specified in **Supplementary Table S1** (yeast) and **Supplementary Table S2** (human).

eukaryotes, orthologs of most RBFs are represented in all investigated lineages. Losses of RBFs are rare, but nevertheless they seem to occur. Examples exist of yeast RBFs lacking orthologs in either all or at least most of the Pezizomycotina, the sister clade of the Saccharomycotina, although orthologs are found in more distantly related fungi (**Supplementary Figure S2**). Substantially more pronounced is the apparent loss of RBFs in the microsporidia, obligate intracellular fungal parasites, which lack orthologs to 56 yeast RBFs that can be traced back to LECA (**Supplementary Table S4**). This reveals that the overall trend of microsporidia to evolve toward a highly simplified variant of a eukaryotic organism (Keeling et al., 2010) extends also to ribosome biogenesis. Likewise, 17 human RBFs are represented in a diverse set of eukaryotes, but orthologs are missing either in all fungi, or specifically on the lineage toward *S. cerevisiae* (see below).

In contrast to the overall conservation of RBFs in the eukaryotic domain, the representation of orthologs in archaea and bacteria is substantially sparser. Bacteria lack orthologs for most the eukaryotic RBFs, and only individual factors are detected consistently across the sampled taxa. The picture in archaea is more differentiated; archaea possess orthologs of a considerable number of RBFs. For individual factors, orthologs are consistently detected across the individual archaeal groups, whereas for others, they are confined to individual lineages or they occur only sporadically. Overall, very few factors are ubiquitously present across all three domains of life or are consistently present in the bacterial representatives but not in the archaea.

To provide a more quantitative view of these observations, we stratified the proteins into different age layers. In the most permissive setting (**Figures 3B–D**), we assessed the minimal evolutionary age of a protein by assigning it to the LCA that the seed species shares with the most distantly related species in which an ortholog was detected. However, such unfiltered data may suffer from overestimates due to spurious orthology

FIGURE 2 | The PhyloRBF platform. **(A)** Data overview; **(B)** Custom selection of genes and proteins for display; **(C)** Detail information for seed-ortholog pairs; and **(D)** Domain architecture view.

assignments and gene sets contaminated with sequences from other species (Steinegger and Salzberg, 2020). We therefore introduced two filters to reconstruct the phylostrata with two increasing levels of confidence (see **Supplementary Figure S3** for a visualization of the filtering effect). An abundance filter was established, building on the observation that secondary losses of RBFs are rare. It is, thus, expected that an RBF predicted to be present in the LCA of a systematic group, should be detected in all, or at least the majority, of its descendants. In turn, orthologs that show low prevalence in a group are likely spurious or represent contaminations that can be disregarded. For the second filter, we additionally propose that functionally equivalent orthologs participating in ribosome biogenesis in different organisms should have similar domain architectures (Koestler et al., 2010). We, thus, applied the abundance filter on the phylogenetic profiles where we kept only orthologs with a domain architecture that is similar to that of the seed protein (delta-FAS < 0.25). The resulting phylostratigraphies after application of the filters are shown in **Figure 3B**, and the corresponding assignment of the RBFs to the individual phylostrata is provided as **Supplementary Table S5**. As expected, the proportion of RBFs assigned to evolutionarily younger phylostrata increases with filtering strength, and even the two positive controls were assigned only to LECA in the very stringent combined abundance and FAS filter. The results for the human RBFs are shown in **Figures 3C,D**. We note, however, that an *ad hoc* specification of the appropriate filtering criteria is difficult. The phylostratigraphies shown in **Figure 3** should be interpreted such that they allow dynamic analysis strategies. The most stringent filtering serves to reconstruct the evolutionary scaffold of ribosome biogenesis across the tree of life at high confidence. This scaffold can then be successively extended, where appropriate, with results obtained only with the more permissive filters. This clearly indicates where a focus on subsequent data curation and/or experimental validation should be placed.

Absence of Human RBFs in the Fungal Lineage – Functional Plasticity or Methodological Artefact?

Figure 3A reveals that evidence of lineage-specific losses of RBFs within the eukaryotes is rare, with the notable exception of the microsporidia. This makes the existence of 17 human RBFs that can be traced back at least to LECA, but either

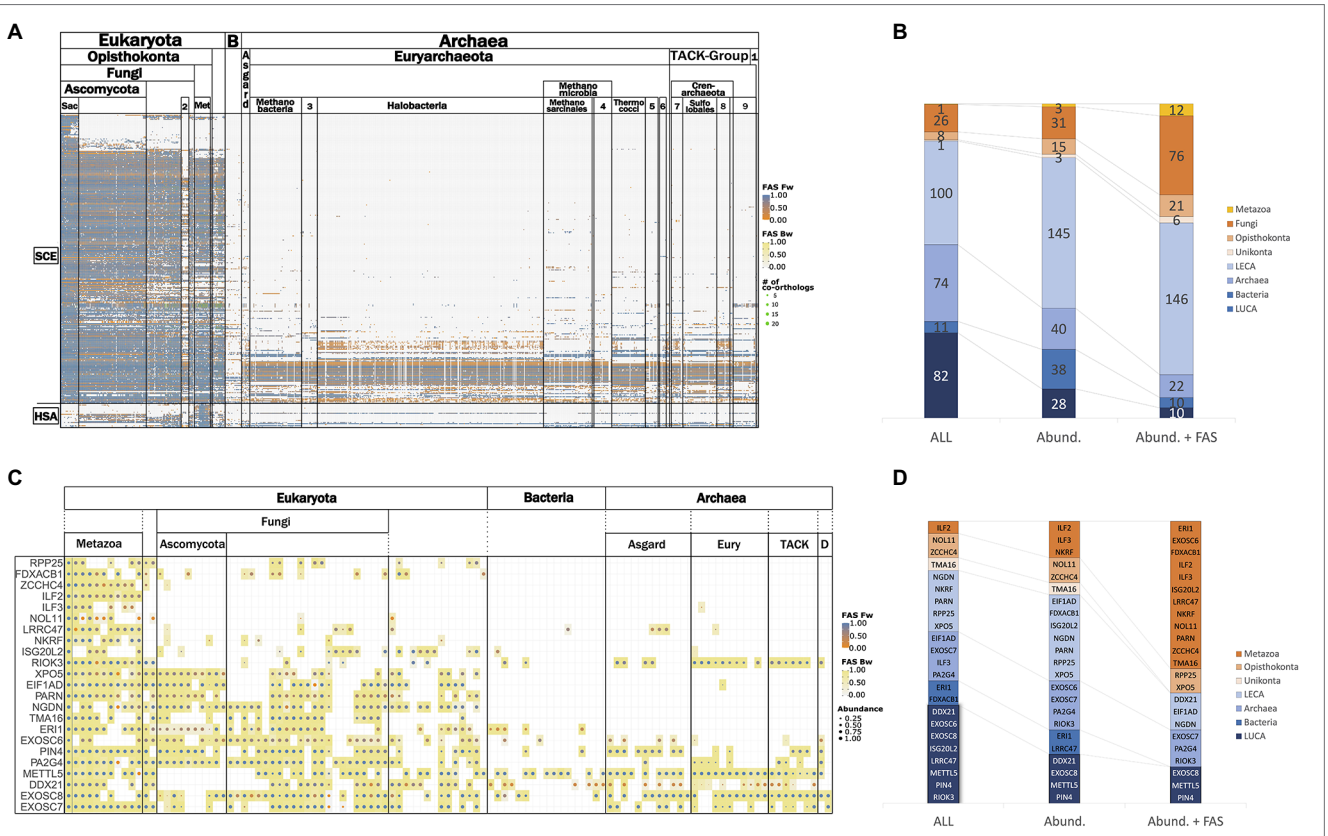


FIGURE 3 | Representation of RBF_{euk} orthologs across the three domains of life. **(A)** Phylogenetic profile of the RBF_{euk} set summarized on the class-level. Proteins are represented by the rows and columns represent the analyzed taxa. A dot in the matrix indicates that an ortholog for the respective RBF was found in this taxon. The dot color informs about the domain architecture similarity between the detected ortholog and the seed protein. **(B)** Phylostratigraphy of the RBF_{euk} set based on the data shown in **(A)** filtered at varying levels of stringency (see main text). LECA – last eukaryotic common ancestor; Archaea – orthologs to an RBF are present in the archaea but not in the bacteria; Bacteria – orthologs to an RBF are present in bacteria but not in archaea; and last universal common ancestor (LUCA) – an ortholog was present both in archaea and bacteria. The two positive controls are placed in the Archaea stratum for the All and the Abundance filter, but are placed in LECA in the Abund. + FAS filter. **(C)** Phylogenetic profiles of the RBF_{human} set summarized on the class level. The corresponding phylostratigraphies are shown in **(D)**.

lack an ortholog in yeast, or their ortholog has diverged in function, an intriguing observation. Analyses of these proteins over, in the scope of this study, the moderate evolutionary distances between human and yeast serves two important needs. They can highlight that even evolutionarily old mechanisms of eukaryotic ribosome biogenesis can change on individual lineages. Probably even more important, they can shed light on methodological issues that become relevant when extending the analysis to the substantially more distantly related archaea.

METTL5 and PIN4 (Parvulin 14) represent two particularly prominent examples of gene loss on the evolutionary lineage leading to *S. cerevisiae*. Both proteins are assigned to the evolutionarily oldest RBF stratum under all filtering conditions (*cf.* **Figures 3C,D**). Within fungi, METTL5 is confined to the early branching lineages with only three putative orthologs in the Basidiomycota. METTL5 is an RNA methyltransferase that mediates N^6-methylation of adenosine 1832 in the 18S rRNA in humans (van Tran et al., 2019), and a similar function has been described for METTL5 in *Haloferx vulcanii*, where it installs m^6A1432 in the 16S rRNA (Kowalak et al., 2000; Grosjean et al., 2008). Yeast, in turn, has no m^6A present at the corresponding position of the 18S rRNA (Taoka et al., 2016; Sergiev et al., 2018). The loss of METTL5 in the course of fungal diversification is, thus, an evolutionarily unique event that changed a long-standing event in ribosome biogenesis. PIN4, a peptidyl-prolyl cis-trans isomerase required for pre-rRNA processing (Fujiyama-Nakamura et al., 2009), displays a different timing of gene loss. Like METTL5, PIN4 is also present in the archaea (**Figures 3C,D**), which is in line with previous findings (Jaremko et al., 2011; Hoppstock et al., 2016). In contrast to METTL5, PIN4 is prevalent in fungi. A sequence comparison between fungal and animal PIN4 orthologs indicates the presence of the sequence motif relevant for the association of this protein with pre-ribosomal complexes (**Supplementary Figure S4**). The loss of PIN4 immediately predates the diversification of the Saccharomycetales. The functional consequences of this loss are, to our knowledge, not yet explored. It will be interesting to see whether yeast and its close relatives utilize a non-homologous protein to functionally replace PIN4, whether its action is not necessary in the context of yeast ribosome biogenesis, or whether an alternative mechanism exists to compensate this loss of function.

Domain Architecture Changes Indicate Lineage-Specific Loss of RBF Function

The example of XPO5 reveals that events subtler than gene loss also need to be considered when tracing factors involved in ribosome biogenesis across the organismic diversity. XPO5 is a human export factor that shuttles a diverse set of cargos, most prominently non-coding RNAs, such as tRNAs and pre-miRNA, but also ribosomal subunit precursors out of the nucleus (Bohnsack et al., 2002; Calado et al., 2002; Yi et al., 2003; Bohnsack et al., 2004; Lund et al., 2004; Wild et al., 2010; Leisegang et al., 2012). The corresponding ortholog in yeast serves also as an export factor, but with a more restricted set of cargoes that does not include ribosomal subunits (Moy and Silver, 1999; Stage-Zimmermann et al., 2000; Sloan et al., 2016). Notably, the domain architectures of human XPO5 and its ortholog in yeast, Msn5, are substantially different (**Figure 4**). Two N-terminal Pfam domains, Xpo1 (Pfam ID: PF08389) and IBN_N (Pfam ID: PF03810), that are characteristic for human XPO5 and also other proteins involved in nuclear import and export, e.g., yeast Xpo1 (P30822), are missing in yeast Msn5. Notably, we find the human domain architecture for XPO5 reflected in orthologs from early branching fungal lineages, suggesting that here the contribution of this protein to ribosome biogenesis could still be maintained. Although experimental proof that the apparent loss of the two domains causes the shift in the functional spectrum of Msn5 has to delivered, this result suggests that changes in domain architecture can highlight RBF orthologs that may have altered their function.

Evolutionary Traceability of RBFs

During integration of the RBF_{yeast} and RBF_{human} sets, no human protein was considered that has a yeast ortholog also annotated as an RBF. Interestingly, we noted that individual human factors, in particular, six of the 10 protein components of the RNase MRP complex, were initially retained because no yeast ortholog could be detected (**Figure 5A**). While this is consistent with the notion that yeast has substantially modified its RNase MRP complex because orthologs of many of its components are confined to yeast and its close relatives (Ebersberger et al., 2014), it is at odds with experimental evidence that functional equivalents exist for these factors in yeast and the function of the RNase MRP complex in pre-rRNA processing is conserved between yeast and humans (Chu et al., 1994; Lygerou et al., 1996; Rosenblad et al., 2006; Goldfarb and Cech, 2017; **Figure 5B**). To resolve this contradiction, we first confirmed that the absence of a yeast ortholog is independently supported by the InParanoid database (Ostlund et al., 2010), one of the most sensitive and specific public databases for ortholog assignments between pairs of species (Altenhoff et al., 2016). Thus, the missing of orthologs for these proteins in the respective species seems inherent to the ortholog search itself, and is not dependent on the ortholog search tool. Investigating the six proteins in greater detail revealed that all are short with lengths between 150 and 200 amino acids, and Rmp1 and Pop8 are additionally devoid of any Pfam domain (**Figure 5C**; **Supplementary Figure S5**). Both characteristics in combination are indicative for proteins that lose a significant sequence similarity to their orthologs already over small evolutionary distances (Jain et al., 2019). Indeed, the evolutionary traceability indices for these proteins in humans (see Materials and Methods) indicate that human orthologs of four proteins likely have diverged beyond recognition (*cf.* **Figure 3B**), and thus escape detection in large-scale phylogenetic profiling studies. In conclusion, the discrepancy between the *in silico* approaches to assess phylogenetic distribution and evolutionary age of the yeast RNase MRP components, and the experimental evidence, can be explained in at least four out of six cases by the limited sensitivity of the ortholog search.

To see whether other proteins in the RBF_{euk} set may suffer from the same limitations, in particular, when extending the scope of the ortholog search to the archaeal domain, we subsequently computed the traceability indices across the full range of taxa considered here (**Figure 6A**). This identified 42 proteins that evolve at rates likely to hinder detection of orthologs in evolutionarily more distant lineages (**Figure 6B**). For these proteins, it is important to consider that the phylogenetic profiles may underestimate the taxonomic range in which orthologs are present, and hence will result in underestimated evolutionary ages. Among them, we find a further factor, Arx1, which associates with late LSU particles in the nucleus and facilitates their export (Bradatsch et al., 2007; Hung et al., 2008). PA2G4, which is a member of the RBF_{human} set, was initially suggested to be its functional equivalent in humans, but we failed to establish orthology relationships between Arx1 and PA2G4 (**Supplementary Figure S6**). Meanwhile, evidences accumulated that PA2G4 differs in regions that are implicated in nucleoporin interaction in Arx1 (Squatrito et al., 2004; Bradatsch et al., 2007; Wild et al., 2010; Bhaskar et al., 2021). While this suggests a functional diversification between the two proteins, their precise evolutionary relationships remain unclear. A phylogenetic analysis revealed gene duplication paired with a substantial acceleration of the evolutionary rate on the

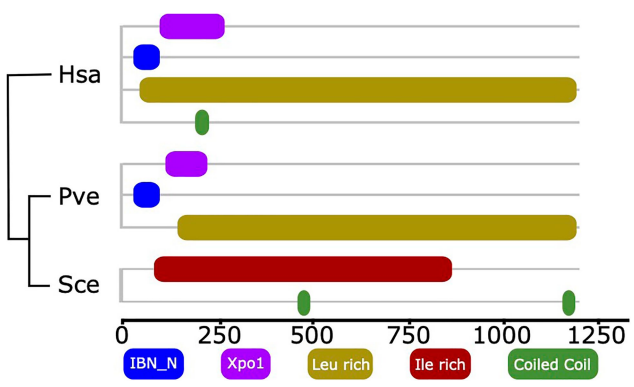


FIGURE 4 | Domain architecture comparison of Xpo5 between human and fungal orthologs. The two Pfam domains linked to import/export activity of human Xpo5, IBN_N (Pfam ID: PF03810) and Xpo1 (Pfam ID: PF08389) are still present in an early branching fungus and were secondarily lost in yeast. The phylogenetic relationships of the three taxa are indicated by the tree. Hsa, *Homo sapiens*; Pve, *Podila verticillata* (Mucoromycota); and Sce, *Saccharomyces cerevisiae*.

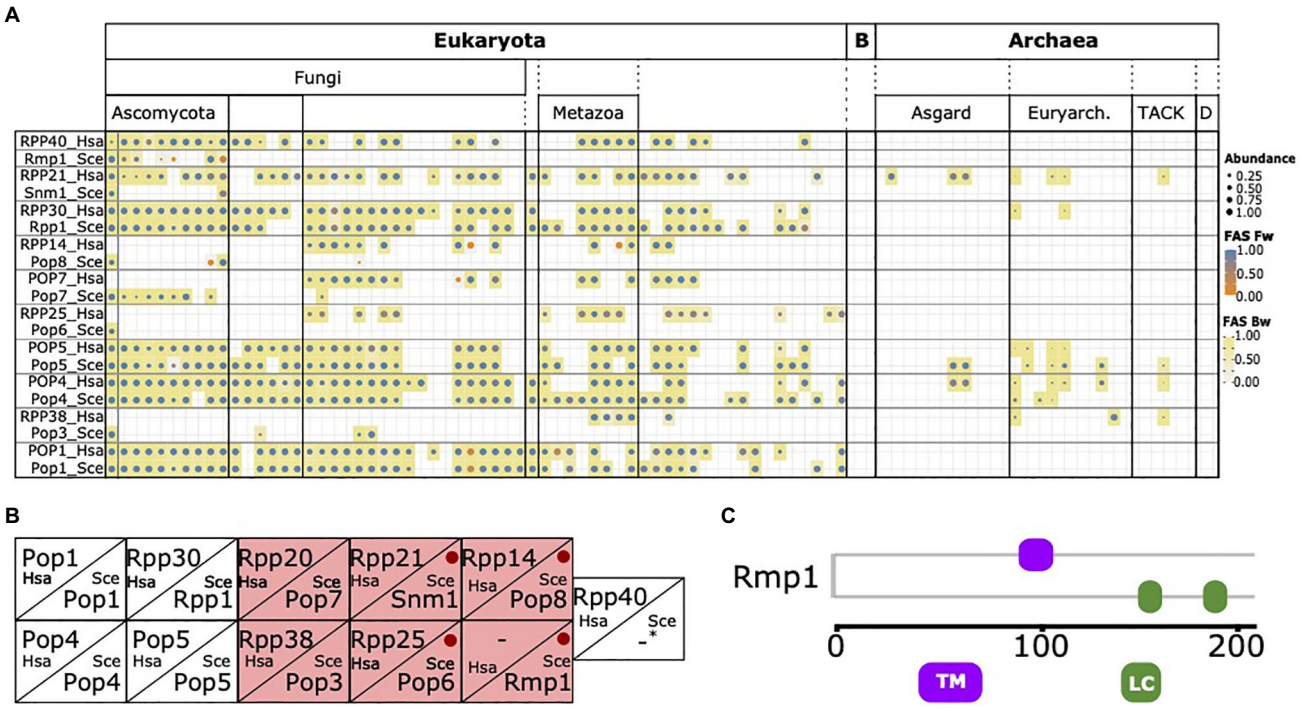


FIGURE 5 | The phylogenetic profile of the RNase MRP components. **(A)** The phylogenetic profiles of the 10 yeast and human RNase MRP components. Human and yeast components considered functionally equivalent are placed in one box. **(B)** Corresponding components of RNase MRP complex in yeast and human according to Rosenblad et al. (2006). Yeast proteins with a missed human ortholog in **(A)** are highlighted in red. Red dots indicate yeast proteins with a traceability index below 0.75 in humans, indicating that orthologs most likely have diverged beyond recognition (Snm1 = 0,58; Pop6 = 0,48; Pop8 = 0,58; and Rmp1 = 0). For human RPP40 and yeast Rmp1 no corresponding protein in the other species are known. **(C)** Domain architecture of yeast Rmp1. TM, transmembrane domain; LC, low complexity region. See **Supplementary Figure S4** for the domain architectures of the remaining five proteins.

Arx1 lineage of Saccharomycotina and Taphrinomycotina (**Supplementary Figure S7**). This lends support to the hypothesis that similar to the components of the RNase MRP complex, the orthology between PA2G4 and Arx1 was overlooked thus far.

Eukaryotic RBFs in the Archaeal Domain

The analysis thus far has set the methodological stage for projecting concepts of eukaryotic ribosome biogenesis into the archaeal domain. Subsequently, we focused on the representation of the RBF_{euk} in archaea. Sampling more than 700 taxa covering the full known archaeal diversity, including the Asgard group that are proposed as the closest relatives to the eukaryotes (Zaremba-Niedzwiedzka et al., 2017), constitutes an unprecedented basis for this analysis. Most importantly, it provides the first comprehensive overview of which factors are represented by orthologs in the archaeal domain, together with information about their lineage-specific prevalence (**Supplementary Figure S8**). For 156 factors in the RBF_{euk} set, we found in the unfiltered data at least one ortholog in the archaeal domain (*cf.* **Figure 3B**; **Supplementary Figure S8**). The number of orthologs per species varies considerably (**Supplementary Table S5**), and *Candidatus Prometheoarchaeum syntrophicum* (Asgard group) harbors the highest number of RBF orthologs (67) across the sampled taxa. Applying an abundance filter (see above) to select only proteins consistently seen in the archaea reduces the number to 17 factors that are represented by an ortholog in more than 85% of the investigated taxa (**Supplementary Table S7**), among the two positive controls. This number reduces to only four when setting the inclusion threshold at 90%. These nearly ubiquitously represented proteins comprise two RNA methyltransferases, METTL5, and Nop1/Fibrillarin, the pseudouridine synthase Pus7 that catalyzes pseudouridylation within the 5S rRNA, and one of our two positive controls, the largest subunit of the RNA polymerase II, Rpo21. The presence of many of these proteins, and particular of these latter factors, in archaea is already well acknowledged (e.g., Jaremko et al., 2011; Ebersberger et al., 2014; Londei and Ferreira-Cerca, 2021). Their prevalence in our data set suggests that insights gained from detailed experimental analyses in individual taxa can, in some cases, be projected to the entire archaeal domain. Factors that are present only in a subset of the taxa, in turn, can provide insights into lineage-specific differences of presumably related to ribosome biogenesis within the archaea, an area that is still considerably uncharted.

Lineage-Specific Presence of Eukaryotic RBFs in Archaea

From the archaeal section of **Figure 3A**, we selected the subset of RBF_{euk} that is confined to individual archaeal clades and grouped these proteins into seven functional

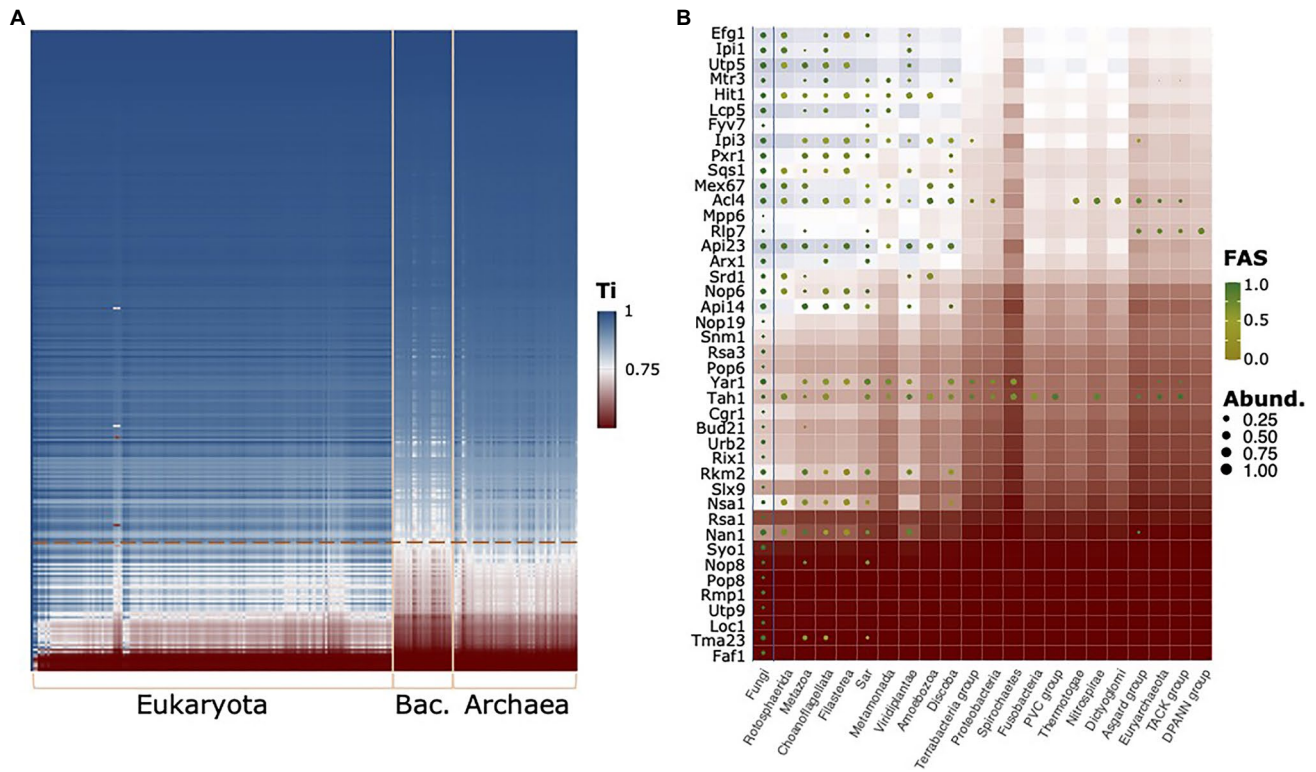


FIGURE 6 | The evolutionary traceability of the yeast RBFs. **(A)** The heat map reveals proteins and taxa for which the ortholog search sensitivity likely becomes a limiting factor (white to red areas). The phylogenetic profiles for proteins with a mean kingdom-wide traceability index (Ti) below 0.75 are shown in **(B)** summarized on a kingdom level. Abund – Fraction of taxa subsumed in a kingdom harboring an ortholog; FAS – maximum domain-architecture similarity between a yeast protein and the orthologs subsumed in this group. Background color represents the evolutionary traceability for the respective yeast protein.

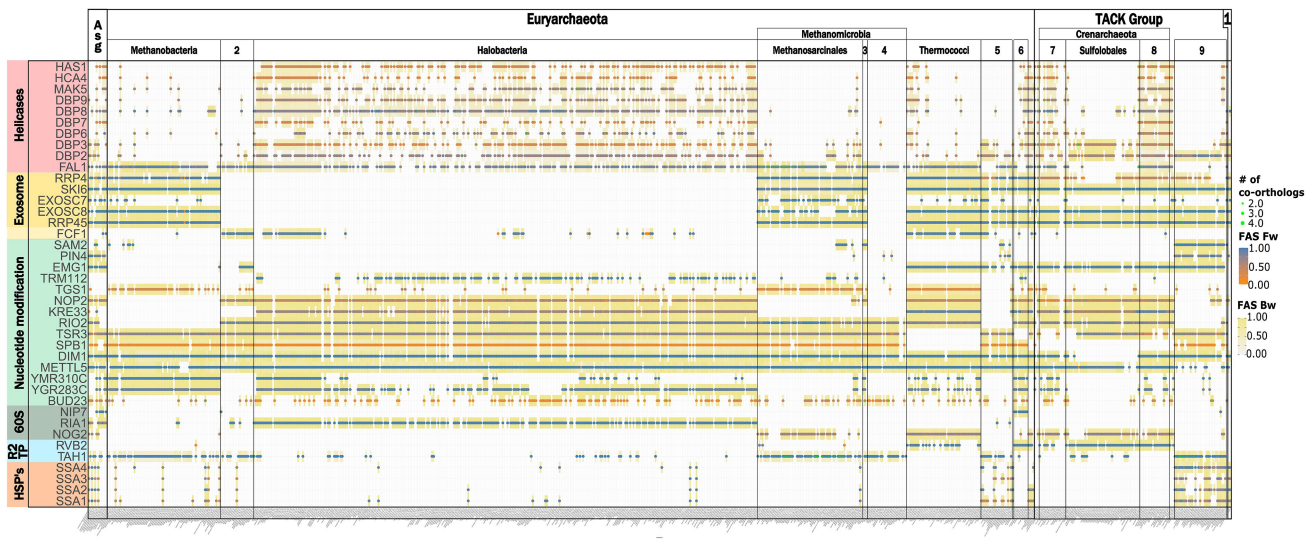


FIGURE 7 | The differential presence-absence pattern of eukaryotic RBFs in the archaeal domain summarized on the class-level. 1: DPANN, 2: Methanococci, 3: Methanocellales, 4: Methanomicrobiales, 5: Thermoproteales, 6: Archaeoglobi, 7: Thermoproteii, 8: Desulfurococcales, and 9 Thaumarchaeota.

categories (**Figure 7**). This revealed a loss of several components of the RNA exosome, a major RNA decay machinery (Sloan et al., 2012), once in Halobacteria and once in Methanomicrobiales. Two small-scale analyses hinted toward the possible absence of the exosome complex in these clades (Koonin et al., 2001; Phung et al., 2020). The concerted absence

of the exosome components across all members of the two clades seen here, provides substantial evidence that the RNA exosome has indeed been lost on these lineages. A similarly prominent signal is seen for the RNA helicases. However, we note that the domain architecture similarities between the RBFs and their archaeal orthologs are typically small for these proteins (**Figure 7**; **Supplementary Figure S9**). As RNA helicases are involved in a plethora of different processes in a cell, it is thus too early to conclude on the existence of these RBFs in the archaeal domain. The other functional categories do not share a consistent phylogenetic profile. The HSP70-family chaperones SSA1-4 are no exception here, as they all share the same archaeal proteins as orthologs, indicating that gene duplications on the eukaryote lineage gave rise to these four distinct RBFs. This directs further attention to the presence/absence pattern of individual proteins, most of them involved in nucleotide modification. Here, the missing of orthologs to Dim1, a dimethyladenosine transferase involved in rRNA modification in the Sulfolobales is, at first glance, an intriguing finding. Dim1 is a protein that is otherwise almost fully conserved across taxon collection. We detected an ortholog in 99% of the eukaryotic taxa and it is present in all but one (*Aquifex aeolicus*) of the bacterial species. The domain architectures of both eukaryotic and bacterial/archaeal orthologs are virtually identical to that of yeast Dim1 (**Supplementary Figure S10**), and the traceability of Dim1 is one across all taxa. Despite this convincing indication for a gene loss, the finding is at odds with the detection of an N^6-dimethyladenosine in the 16S rRNA (Noon et al., 1998) and a recent characterization of a Dim1-like enzyme in *Sulfolobus acidocaldarius* (Knuppel et al., 2021). In the first step, we confirmed that a BlastP search with the alleged Dim1 from *S. acidocaldarius* revealed no hit in yeast (not shown). To increase the sensitivity of the ortholog search, we used a "stepping-stone" approach. In brief, we performed a second search for Dim1 orthologs, this time seeding the search with the Dim1 of *S. acidocaldarius*. This served to reduce the evolutionary distance between seed protein and possible orthologs in the archaea (Martin-Duran et al., 2017). This search identified orthologs throughout the archaeal domain and sporadically also in the eukaryotes (**Supplementary Figure S11**). In all investigated cases, these orthologs were the same as those obtained with yeast Dim1 as a seed sequence. Based on their overlapping phylogenetic profiles, we conclude that yeast Dim1 and the protein in *S. acidocaldaricus* are indeed orthologs, although they share no significant sequence similarity. A further comparison revealed that Dim1 in Sulfolobales is only about 200 aa in length, while Dim1 in other taxa has a length of around 300 amino acids (**Supplementary Figure S12**). In summary, one of the evolutionarily most conserved proteins known to date has been modified in one archaeal lineage comprising largely extremophiles, such that the orthologs have diverged, from a eukaryotic viewpoint, beyond recognition. It will be interesting to see the precise functional consequences that are likely accompanied with this change. In contrast to the spurious absence of Dim1, there is strong evidence for a lineage specific loss of the kinase/ATPase Rio2, the RNA helicase/acetyltransferase Kre33, and the RNA methyltransferase Emg1 within the archaea. For these proteins, the stepping-stone approach provided no indication of overlooked orthologs (**Supplementary Figure S13**), and we conclude that the gaps in their phylogenetic profiles represent the genuine absence of these proteins in the respective taxa.

Large Subunit Processing in Archaea

The analyses thus far resulted in a high-resolution overview of the prevalence and distribution of eukaryotic RBFs in the archaeal domain. However, they also revealed that a face-value interpretation of phylogentic profiles can be misleading with the erroneous assumption of factor (and function) absence and presence. Careful curation of the profiles *via* a comparison of domain architectures, the consideration of the evolutionary traceability and case-by-case, more sensitive analyses *via* the stepping-stone approach, or even *via* the search for proteins harboring only a characteristic Pfam domain (see Jain et al., 2019), can substantially increase the reliability of the conclusions drawn. Such comprehensive analyses are challenging to apply on large data sets and are performed best in a context of functional sub-clusters. Here, we focused on late steps during maturation of the 60S subunit regulated by the GTPase Nog1 (Klingauf-Nerurkar et al., 2020) as a showcase example (**Figure 8A**). The phylogenetic profiles of 16 RBFs involved are shown in **Figure 8B** and reveal an intriguing pattern. Only two factors are consistently present in the archaea with domain architecture similarities that leave little doubt of functional equivalence to the yeast proteins, Tif6 (Benelli et al., 2009) and Sdo1. These proteins functionally interact, where Sdo1, in combination with Ria1 (see below), promotes removal of Tif6 from pre-60S particles. Orthologs to Nog1 are found in all major archaeal lineages, except in the two representatives of the DPANN group. However, their domain architecture similarity scores are low (**Supplementary Figure S14**), which indicates either a spurious ortholog assignment or a change in protein function during the evolution of Nog1. In strong support of the latter hypothesis, we first confirmed that in fungi, animals, and archaea, typically only a single gene in the gene set of an organism encodes a protein carrying a Nog1 Pfam domain (Pfam Id: PF06858; **Supplementary Figure S15**). A subsequent comparison of the domain architectures revealed that archaeal Nog1 is only half the size of its eukaryotic counterpart (**Figure 8C**). While the N-terminal part harboring the GTPase activity is present, archaeal Nog1 lacks the C-terminal portion that mediates the interaction with Arx1. The situation is similar for Rlp24, which was suggested to act as placeholder for the ribosomal protein eL24 during the maturation process (Warner, 2015; **Figures 8B,D**), but with two key differences. Almost all archaeal orthologs of Rlp24 are approximately 70 amino acids in length and harbor only the L24e Pfam domain (PF01246). The C-terminal tail that is essential for recruiting Drg1 is missing. But surprisingly, we find individual taxa within the Halobacteria and the Thermoplasmata whose Rlp24 ortholog is slightly longer. Their domain architectures match with that of yeast Rlp24, and they also appear in possession of the C-terminal tail of the protein found in yeast despite the absence of Drg1 orthologs in these clades. The second difference is the functional annotation of the archaeal orthologs. In all cases, we identified the ribosomal protein L24 as the ortholog

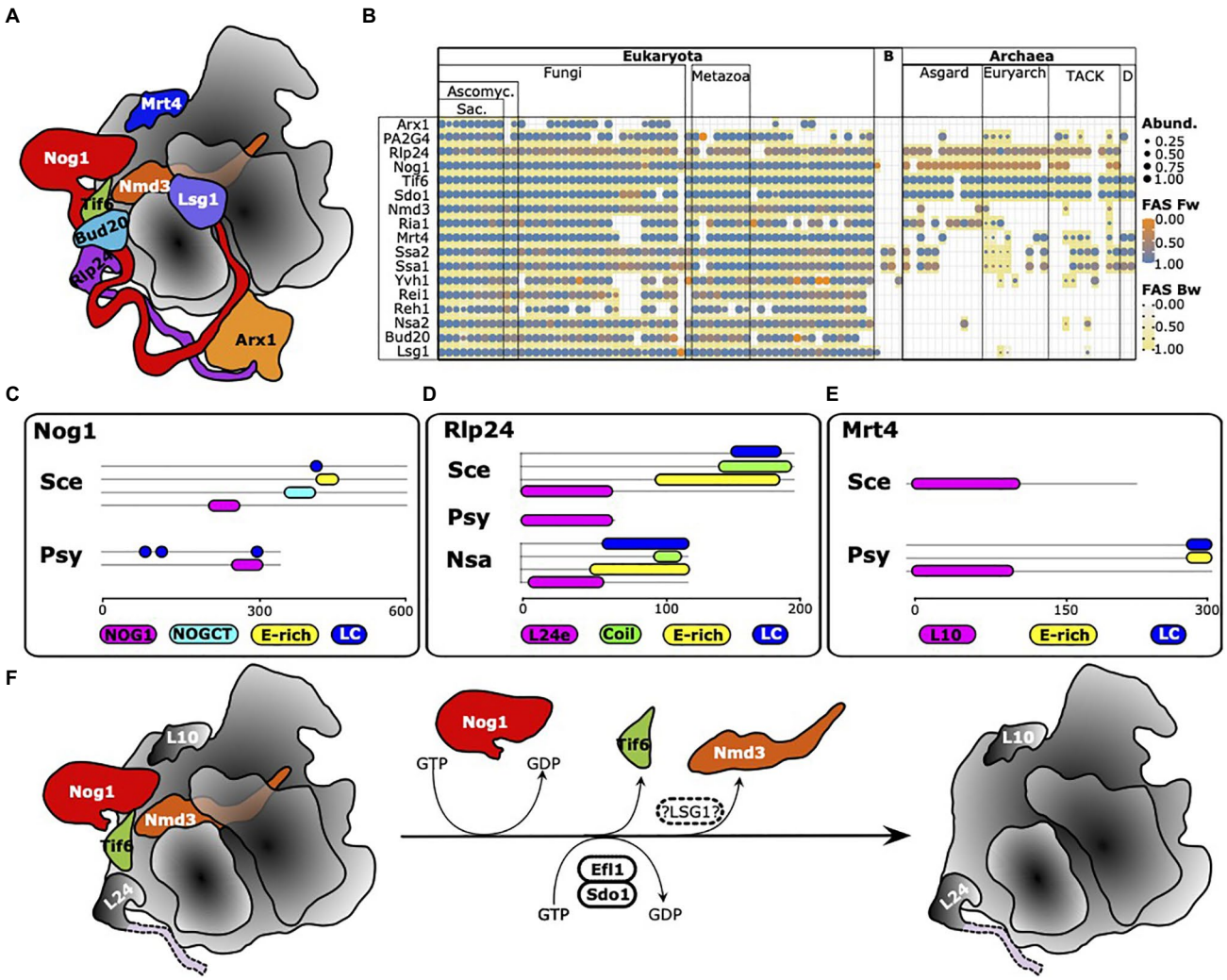


FIGURE 8 | Maturation of the large ribosomal subunit in archaea is a simplified version of the process in eukaryotes. **(A)** Ribosome biogenesis factors involved in eukaryotic pre-60S maturation. Figure re-drawn from Klingauf-Nerurkar et al. (2020). **(B)** Phylogenetic profiles of the factors involved in yeast 60S maturation. PA2G4 is a human functional equivalent to Arx1 in yeast (see text for details). **(C-E)** Domain architectures of yeast Nog1, Rlp24, and Mrt4 and of a representative archaean ortholog. Sce, *S. cerevisiae*; Psy, *Prometheoarchaeum syntrophicum*; and Nsa, *Natronomonas salsuginis*. Protein ids: $Nog1_{Sce}$ – Q02892; $Nog1_{Psyn}$ – WP_147661298; $Mrt4_{Sce}$ – P33201; $Mrt4_{Psyn}$ – WP_147662854; $Rlp24_{Sce}$ – Q07915; $Rlp24_{Psyn}$ – WP_147664435; and $Rlp24_{Nsa}$ – WP_137275029. **(F)** Essential steps of large subunit maturation likely to be conserved between eukaryotes and archaea. Thus far, we have no evidence for the presence of LSG1 in archaea (hatched oval).

to Rlp24, and this is the only protein encoded in archaeal genomes that harbors the L24e Pfam domain, again leaving no scope for overlooked orthologs (**Supplementary Figure S16**). This is perfectly in line with a scenario where eukaryotic Rlp24 and eL24 emerged by a gene duplication of an ancient ribosomal protein. One copy retained the ancestral function as a ribosomal protein, whereas the other copy evolved into an RBF. Interestingly, the same scenario applies to Mrt4 (**Figures 3A,E**), which is known to function as a placeholder for the acidic 60S ribosomal protein uL10 (Warner, 2015). The archaeal orthologs are annotated as Rpl10, and again each of the investigated archaeal taxa harbors exactly one protein with the L10e domain (**Supplementary Figure S17**). Notably, the archaeal orthologs can be longer than yeast Mrt4. They often harbor an acidic (E-rich) stretch and a low complexity region at the C-terminus (**Supplementary Figure S18**), similar to what is seen for yeast Rlp24 and Nog1 (**Figures 8C,D**). Judging by the domain architecture, they seem to resemble the eukaryotic ribosomal protein P0 rather than Mrt4, indicating that the architecture of Mrt4 is the evolutionarily-derived form (**Supplementary Figure S19**). Note, the considerably sparse representation of Mrt4 orthologs must be attributed to methodological issues (see Dim1 above), as all archaeal gene sets we checked contained exactly one protein with the L10e Pfam domain (*cf.* **Supplementary Figure S17**). Validating the presence of the other proteins involved in late 60S maturation in archaea provided unambiguous evidence only for Nmd3 (**Supplementary Figure S20**), a protein that, in eukaryotes, contributes to pre-60S nuclear export by acting as an adaptor for the nuclear transport receptor Crm1 (Ho et al., 2000).

Efl1/Ria1, which interacts with Sdo1 in the release of Tif6, identifies the elongation factor EF2 in archaea as an ortholog. Similar to Mrt4 and Rlp24, also eukaryotic EF2 and Efl1/Ria1 likely emerged by a gene duplication in LECA (see **Supplementary Figures S21, S22** for details). All other proteins appear absent in the archaea. In case an ortholog was identified in our search, it turned out to resemble a remote paralog of the RBF with an evolutionarily highly conserved function, e.g., MAP2 in the case of PA2G4 (*cf.* **Supplementary Figure S6**). A single factor, however, remains elusive, Lsg1. Its GTPase activity is required for the release of Nmd3 as one of the last steps of 60S maturation (Hedges et al., 2005). Searches at higher sensitivity identified potential homologs, but neither sequence similarity nor domain architecture conservation sufficed to conclude on the presence of this protein in the archaea. Taken together, these data suggest that many late steps of large subunit maturation in archaea follow the same principles as in eukaryotes. However, instead of using the placeholders Mrt4 and Rlp24, archaea most likely directly install the ribosomal proteins RPP0 and L24. As a consequence, the downstream machinery required for the successive release of Rlp24 and Mrt4 by Drg1 (Pertschy et al., 2007) and Yvh1 (Kemmler et al., 2009; Lo et al., 2009), and the incorporation of corresponding ribosomal proteins, is not necessary. Likewise, the lack of an ortholog to Arx1 in the archaea is in line with the findings that the C-terminus of Nog1, which is essential for interaction with Arx1, is absent from the archaeal proteins. Interestingly, Mrt4 also represents one of the factors that is implicated in pre-ribosome surveillance in eukaryotes (reviewed in Karbstein, 2013; Klinge and Woolford, 2019), as it was reported to block ribosomal stalk assembly until it is released from pre-ribosomal complexes (Kemmler et al., 2009; Lo et al., 2009; Rodriguez-Mateos et al., 2009). Its absence in archaea might therefore indicate a simplification of the process in this clade. In essence, it appears that the late steps of large subunit maturation in contemporary archaea largely resemble the primordial pathway in the LCA of archaea and eukaryotes. This basic functionality was later extended in the eukaryotic lineage to facilitate nuclear export and to implement various quality control steps.

DISCUSSION

Eukaryotic ribosome biogenesis is currently best understood in the yeast model system, *S. cerevisiae*, but knowledge on the human pathway is growing (Bohnsack and Bohnsack, 2019). However, the increasing availability of genome sequences of species from the remotest corners of cellular life now enables investigation of the conservation and plasticity of this pathway across the eukaryotic and the archaeal domains. Integrating these diverse efforts into a comprehensive view of how ribosome biogenesis is accomplished across the organismal diversity strongly benefits from a unifying data basis. This allows an interpretation of the outcomes of functional studies on individual RBFs or complexes contributing to ribosome biogenesis in the context of their evolutionary trajectory across the organismic diversity. Various public databases partially fulfil this requirement. For example, KEGG (Kanehisa et al., 2016) or REACTOME (Fabregat et al., 2018) provide access to pathways and interaction networks of proteins involved in ribosome biogenesis together with a representation of pathway components in other taxa. Dedicated ortholog databases, such as InParanoid (Ostlund et al., 2010), OMA (Altenhoff et al., 2018), and orthoDB (Zdobnov et al., 2017) provide comprehensive collections of orthologous groups across the full proteomes of hundreds to thousands of species. Despite the wealth of information contained in these resources, they also have limitations. Neither KEGG nor REACTOME consider the full set of factors currently considered as RBFs. Furthermore, the ortholog databases have no focus on a dedicated pathway, they provide no direct access to phylogenetic profiles for individual or groups of proteins, and they do not facilitate a comparison of domain architectures or a scoring of architecture similarities. Last but not least, it is almost impossible to extend the analysis to custom factors or species. Here, we have combined data from various resources to compile a manually curated, non-redundant set of 301 eukaryotic RBFs (*cf.* **Figure 1**). Their domain architecture-aware phylogenetic profiles across more than 900 taxa are a first step to close these gaps in the currently available resources. The data can be interactively explored and analyzed *via* a light-weight web-portal, or for more in-depth analyses, downloaded and analyzed offline. They can be visualized as a whole to explore the concerted evolutionary behavior of groups of proteins (*cf.* **Figure 7**), or subsets of taxa and proteins can be extracted for in depth analyses down to exploring the lineage-specific fate of individual protein domains (Roustan et al., 2016). This allows to discern factors whose functions likely have changed between orthologs from those where no such indication exists (Jiang et al., 2020). Eventually, the data basis is flexible. Our approach makes it straightforward to extend these profiles with further taxa or factors in custom fashion. We therefore hope that the broad community interested in ribosome biogenesis will benefit from this data and it could serve as template for further pathway-specific analyses.

Secondary Modification of Ribosome Biogenesis in Fungi

Large-scale screens in humans paired with targeted identification of proteinaceous factors *via* the characterization of ribosomal complexes have the potential to complement the yeast perspective on eukaryotic ribosome biogenesis (Wild et al., 2010; Tafforeau et al., 2013; Badertscher et al., 2015; Farley-Barnes et al., 2018). Interestingly, the integration of this data revealed only 23 human factors that lack an ortholog in yeast. We note that this number is most likely an underestimate. Manual curation determining whether a protein was already shown to associate with pre-ribosomal particles/factors or has a known direct function in the pathway was reserved to human proteins at least as old as LECA. Thus, many candidates in the screens have simply not been analyzed in enough detail yet to know if they fulfil these criteria. Still, 17 evolutionarily old RBFs are missing in yeast and in part also in other fungal taxa. This indicates that traditional concepts of ribosome biogenesis have been modified during fungal diversification, despite the overall evolutionary conservation of this process (Wild et al., 2010;

Tafforeau et al., 2013; Badertscher et al., 2015; Farley-Barnes et al., 2018; Bohnsack and Bohnsack, 2019). With METTL5 and XPO5, we have highlighted one prominent example each for loss-of-function and alteration of function, respectively. Specifically, we could date the loss of METTL5 most likely to the $LCA_{Dikarya}$. Thus, the absence of m^6A that has been described for yeast most likely applies to all dikarya. We could correlate a change in the domain architecture of Msn5, the yeast ortholog of XPO5, with an expansion of its functional repertoire. On closer inspection, we observed a recurrent change in domain architecture within the ascomycetes for this protein, and it will be interesting to see to what extent this affects the cargo(s) Msn5/XPO5 can transport. Few other human RBFs are absent in fungi, and the effect of their alleged loss in fungi should be investigated in greater detail and related with their functions in other eukaryotes.

Sensitivity of the Ortholog Search as a Limiting Factor

The integration of the human data has opened up a further aspect that has so far received little attention in the interpretation of presence/absence patterns in phylogenetic profiles. A number of human factors for which functionally equivalent proteins exist in yeast, appear to lack yeast orthologs. Taken at face value, this must be interpreted as evidence for non-orthologous functional replacements (Koonin et al., 1996). However, for quickly evolving proteins, the grey zone where orthologs are no more similar than it is expected by chance (Rost, 1999; Jain et al., 2019) is within reach, even for the evolutionary distances between humans and yeast, and even more so when extending analyses into the archaeal domain. Here, we have provided evidence that for 42 yeast RBFs, the sensitivity of the ortholog search can become a limiting factor. Among these we found many proteins whose functional equivalent in humans is not identified as an ortholog, eventually reconciling the findings from targeted functional and the large-scale evolutionary approaches. Increasing the search sensitivity, e.g., by switching to unidirectional profile-based searches scanning for proteins sharing the same Pfam domains as the target protein, is an obvious solution. This, however, comes at the costs of a substantially reduced specificity making careful downstream validation of the findings necessary (Jain et al., 2019). However, the example of Dim1 shows that traceability indices can also be positively misleading. The underlying simulation approach is based on the assumption that evolutionary rates and constraints for a protein do not change over time. Our results indicate that this does not apply to Dim1 in the Sulfolobales. The lineage-specific change of both rate and mode of sequence evolution for this otherwise ubiquitously conserved protein, which causes the orthologs to be missed, is intriguing. Whether this is a consequence of an altered function of Dim1 and/or a change of the selective constraints resting on the corresponding gene remains to be determined. While experimental evidence exists for the presence of Dim1 in the Sulfolobales (Knuppel et al., 2021); this is not necessarily the case for other proteins. In such cases, the stepping-stone approach that serves to reduce the evolutionary distance over which orthologs need to be identified can be used to identify even evolutionarily diverged orthologs.

Eukaryotic Ribosome Biogenesis Factors in the Archaeal Domain

Catalogues of eukaryotic ribosome biogenesis factors that are also present in the archaeal domain have been compiled before (e.g., Ebersberger et al., 2014). These earlier studies sufficed to identify the presence/absence of factors in individual archaeal species. Yet, the limited taxon sampling did not allow to differentiate between signal and noise with respect to varying abundance patterns across the analyzed taxa. Along the same line, the co-occurrence of eukaryotic RBFs in individual taxa, an obvious pre-requisite for but also an indication of their functional interaction, could not be exhaustively tested. Here, we could show that, in contrast to the situation in the eukaryotes, very few factors are consistently found throughout the archaeal domain. This can be an effect of limited search sensitivity (see above). However, even the application of the stepping-stone approach indicated the lineage-specific absence of otherwise essential factors, such as Kre33, Rio2, and Emg1. These findings can serve as starting points to elucidate whether ribosome biogenesis in the archaea is more plastic than is the case for the eukaryotes, whether individual factors are not essential (e.g., Knuppel et al., 2021), or whether they convey their activity to a different pathway.

Within our taxon sampling, we considered also all sequenced representatives of the Asgard group, which have been proposed to be the closest relatives to eukaryotes (Zaremba-Niedzwiedzka et al., 2017). This suggested close relationship is not reflected by the presence of any eukaryotic RBF exclusively in members of this clade. However, we note that among all analyzed archaeal taxa, the recently sequenced *Candidatus Prometheoarchaeum synthrophicum* (Asgard group) harbors the largest number of orthologs to eukaryotic RBFs (**Supplementary Table S9**). Whether this is co-incidence or if this indicates that ribosome biogenesis in this species is indeed more similar to that of eukaryotes remains to be determined.

Large Subunit Maturation in Archaea

Over the past years, the structural characterization of pre-ribosomal complexes together with studies on the precise functions of the proteins involved have elucidated the roles and interplay of many RBFs in the stepwise formation and maturation of the ribosomal subunits (e.g., Saveanu et al., 2003; Woolford and Baserga, 2013; Henras et al., 2015; Wu et al., 2016; Bassler and Hurt, 2019; Bohnsack and Bohnsack, 2019; Kargas et al., 2019; Zhou et al., 2019; Klingauf-Nerurkar et al., 2020; Liang et al., 2020). Clusters of RBFs provide an unprecedented opportunity to *in silico* "assemble" archaeal RBF orthologs into a comprehensive picture. This allows well-informed and testable predictions about differences and similarities of the corresponding processes in archaea. Projecting the steps in late maturation of the 60S subunit (Kargas et al., 2019; Zhou et al., 2019; Klingauf-Nerurkar et al., 2020) into the archaeal domain revealed that archaea follow, by and large, similar principles compared to eukaryotes. Yet this process is significantly more simplistic. The absence of C-terminal extensions in both archaeal Nog1, and L24, the archaeal ortholog

of both the eukaryotic RBF Rlp24 and the ribosomal protein eL24, indicates that the parts of the maturation pathway that require protein-protein interaction mediated *via* these extensions (Klingauf-Nerurkar et al., 2020) are missing in the archaea. Moreover, the ribosomal proteins uL10 and eL24 appear associated with the 50S ribosomal complex directly, without the need for previous recruitment of placeholder proteins. The apparent simplicity of this process probably contributes to the fact that the 50S subunit of various archaeal species can be spontaneously reconstituted *in vitro* with RBFs (Londei et al., 1986; Sanchez et al., 1990). The eukaryotic RBFs Mrt4 and Rlp24 that serve as placeholders for the RPPs uL10 and eL24, respectively, emerged by a duplication of the genes encoding the corresponding ribosomal proteins only in the LECA. Consequently, some factors that are required for steps leading up to the installation of these ribosomal proteins in eukaryotes were not identified in archaea and it is probably safe to consider them absent. In essence, archaea seem to bypass various assembly and surveillance steps during large subunit maturation that are characteristic for eukaryotes. Overall, this represents a showcase example of the evolution of complexity *via* the duplication and subsequent functional diversification of ancient genes (Ohno, 1970; Taylor and Raes, 2004) paired with a lineage-specific modification of evolutionary old genes.

Two aspects however remain puzzling. Firstly, the mechanism of release of the pre-60S biogenesis and export factor Nmd3, which is ubiquitously found in archaea, is elusive, as we did not find any evidence for an archaeal counterpart to the GTPase Lsg1, which releases Nmd3 from the 60S complex in yeast (Malyutin et al., 2017). While its eukaryotic function in pre-60S export will not be present in archaea due to the absence of a nucleus, archaeal Nmd3 might share its role in uL16 loading onto pre-ribosomes with its eukaryotic counterpart (Hedges et al., 2005; Hofer et al., 2007; Kargas et al., 2019; Zhou et al., 2019). Future experiments will have to elucidate the precise role of Nmd3 orthologs in the archaea and whether a different protein may serve as the functional equivalent to Lsg1. Secondly, our observation that the ribosomal protein L24 in archaea comes in two flavors is intriguing. The most widespread form is a short protein harboring exclusively the Ribosome_L24e Pfam domain. Representatives of Halobacteria and of Thermoplasmata, however, possess a protein that resembles a miniaturized version of the yeast RBF Rlp24. Their L24 includes a C-terminal extension with exactly the same features as in the yeast protein, a coiled-coil domain, an E-rich stretch and a low-complexity region. These proteins have a second interesting characteristic. Eukaryotic Rlp24 share with archaeal but not with eukaryotic L24e the presence of four conserved cysteine residues at the N-terminus (Saveanu et al., 2003). Three of the four Cysteine residues have been mutated in the L24e with the eukaryote-like domain architecture (**Supplementary Figure S23**). So far, we have no evolutionary explanation for these observations and it will be interesting to investigate the functional role of these changes.

In summary, we established a common data base for integrating research on ribosome biogenesis in eukaryotes and archaea. We have highlighted the potential, as well as possible pitfalls, in the interpretation of these data and have highlighted the cascade of *in silico* analyses that is necessary to faithfully reconstruct the evolutionary trajectory of individual or groups of RBFs from these data. Tracing eukaryotic RBFs into the archaeal domain revealed a surprisingly fragmented abundance pattern. We typically found highly conserved factors missing in individual lineages without any evidence for a methodological artefact. This passes the batton to the experimentalists to determine if, and to what extent, these factors are indeed involved in ribosome biogenesis. Using the late steps in large subunit maturation as an example, we provide evidence that archaea follow fundamentally similar but, at least in parts, highly simplified strategies to assemble their ribosomes. This underlines the role of archaea as potential model systems to elucidate general concepts of ribosome biogenesis and highlights evolutionary strategies to increase the fidelity of this process on the eukaryotic lineage.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://applbio.biologie.uni-frankfurt.de/download/RibosomeBiogenesis.

AUTHOR CONTRIBUTIONS

IE conceived and planned the study and drafted the manuscript with help from MB. MTB and KB curated the data. MB, VT, and SS performed the analyses. IE, MB, MTB, and KB interpreted the results. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Research Funding Program Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz (LOEWE) of the State of Hessen, Research Center for Translational Biodiversity Genomics (TBG) to IE, the Deutsche Forschungsgemeinschaft (SFB1190 to MTB, SFB860 to KB, and EB-285-2/2 to IE), and the University Medical Centre Göttingen (to MTB and KB).

ACKNOWLEDGMENTS

The authors wish to thank all researchers for making annotated genome sequences available to the public domain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.739000/full#supplementary-material

REFERENCES

Altenhoff, A. M., Boeckmann, B., Capella-Gutierrez, S., Dalquen, D. A., DeLuca, T., Forslund, K., et al. (2016). Standardized benchmarking in the quest for orthologs. *Nat. Methods* 13, 425–430. doi: 10.1038/nmeth.3830

Altenhoff, A. M., Glover, N. M., Train, C. M., Kaleb, K., Warwick Vesztrocy, A., Dylus, D., et al. (2018). The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* 46, D477–D485. doi: 10.1093/nar/gkx1019

Altenhoff, A. M., Studer, R. A., Robinson-Rechavi, M., and Dessimoz, C. (2012). Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Comput. Biol.* 8:e1002514. doi: 10.1371/journal.pcbi.1002514

Altschul, S. F., Gertz, E. M., Agarwala, R., Schäffer, A. A., and Yu, Y.-K. (2009). PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res.* 37, 815–824. doi: 10.1093/nar/gkn981

Ameismeier, M., Zemp, I., van den Heuvel, J., Thoms, M., Berninghausen, O., Kutay, U., et al. (2020). Structural basis for the final steps of human 40S ribosome maturation. *Nature* 587, 683–687. doi: 10.1038/s41586-020-2929-x

Badertscher, L., Wild, T., Montellese, C., Alexander, L. T., Bammert, L., Sarazova, M., et al. (2015). Genome-wide RNAi screening identifies protein modules required for 40S subunit synthesis in human cells. *Cell Rep.* 13, 2879–2891. doi: 10.1016/j.celrep.2015.11.061

Barandun, J., Chaker-Margot, M., Hunziker, M., Molloy, K. R., Chait, B. T., and Klinge, S. (2017). The complete structure of the small-subunit processome. *Nat. Struct. Mol. Biol.* 24, 944–953. doi: 10.1038/nsmb.3472

Bassler, J., and Hurt, E. (2019). Eukaryotic ribosome assembly. *Annu. Rev. Biochem.* 88, 281–306. doi: 10.1146/annurev-biochem-013118-110817

Benelli, D., Marzi, S., Mancone, C., Alonzi, T., la Teana, A., and Londei, P. (2009). Function and ribosomal localization of aIF6, a translational regulator shared by archaea and eukarya. *Nucleic Acids Res.* 37, 256–267. doi: 10.1093/nar/gkn959

Bhaskar, V., Desogus, J., Graff-Meyer, A., Schenk, A. D., Cavadini, S., and Chao, J. A. (2021). Dynamic association of human Ebp1 with the ribosome. *RNA* 27, 411–419. doi: 10.1261/rna.077602.120

Bohnsack, K. E., and Bohnsack, M. T. (2019). Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J.* 38:e100278. doi: 10.15252/embj.2018100278

Bohnsack, M. T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10, 185–191. doi: 10.1261/rna.5167604

Bohnsack, M. T., Regener, K., Schwappach, B., Saffrich, R., Paraskeva, E., Hartmann, E., et al. (2002). Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. *EMBO J.* 21, 6205–6215. doi: 10.1093/emboj/cdf613

Bradatsch, B., Katahira, J., Kowalinski, E., Bange, G., Yao, W., Sekimoto, T., et al. (2007). Arx1 functions as an unorthodox nuclear export receptor for the 60S preribosomal subunit. *Mol. Cell* 27, 767–779. doi: 10.1016/j.molcel.2007.06.034

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Calado, A., Treichel, N., Muller, E. C., Otto, A., and Kutay, U. (2002). Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. *EMBO J.* 21, 6216–6224. doi: 10.1093/emboj/cdf620

Chen, F., Mackey, A. J., Vermunt, J. K., and Roos, D. S. (2007). Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS One* 2:e383. doi: 10.1371/journal.pone.0000383

Chu, S., Archer, R. H., Zengel, J. M., and Lindahl, L. (1994). The RNA of RNase MRP is required for normal processing of ribosomal RNA. *Proc. Natl. Acad. Sci. U. S. A.* 91, 659–663. doi: 10.1073/pnas.91.2.659

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

Ebersberger, I., Simm, S., Leisegang, M. S., Schmitzberger, P., Mirus, O., von Haeseler, A., et al. (2014). The evolution of the ribosome biogenesis pathway from a yeast perspective. *Nucleic Acids Res.* 42, 1509–1523. doi: 10.1093/nar/gkt1137

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P., et al. (2018). Reactome graph database: efficient access to complex pathway data. *PLoS Comput. Biol.* 14:e1005968. doi: 10.1371/journal.pcbi.1005968

Farley-Barnes, K. I., McCann, K. L., Ogawa, L. M., Merkel, J., Surovtseva, Y. V., and Baserga, S. J. (2018). Diverse regulators of human ribosome biogenesis discovered by changes in nucleolar number. *Cell Rep.* 22, 1923–1934. doi: 10.1016/j.celrep.2018.01.056

Freed, E. F., Prieto, J. L., McCann, K. L., McStay, B., and Baserga, S. J. (2012). NOL11, implicated in the pathogenesis of north American Indian childhood cirrhosis, is required for pre-rRNA transcription and processing. *PLoS Genet.* 8:e1002892. doi: 10.1371/journal.pgen.1002892

Fujiyama-Nakamura, S., Yoshikawa, H., Homma, K., Hayano, T., Tsujimura-Takahashi, T., Izumikawa, K., et al. (2009). Parvulin (Par14), a peptidyl-prolyl cis-trans isomerase, is a novel rRNA processing factor that evolved in the metazoan lineage. *Mol. Cell. Proteomics* 8, 1552–1565. doi: 10.1074/mcp.M900147-MCP200

Goldfarb, K. C., and Cech, T. R. (2017). Targeted CRISPR disruption reveals a role for RNase MRP RNA in human preribosomal RNA processing. *Genes Dev.* 31, 59–71. doi: 10.1101/gad.286963.116

Grosjean, H., Gaspin, C., Marck, C., Decatur, W. A., and de Crecy-Lagard, V. (2008). RNomics and Modomics in the halophilic archaea Haloferax volcanii: identification of RNA modification genes. *BMC Genomics* 9:470. doi: 10.1186/1471-2164-9-470

Hage, A. E., and Tollervey, D. (2004). A surfeit of factors: why is ribosome assembly so much more complicated in eukaryotes than bacteria? *RNA Biol.* 1, 10–15. doi: 10.4161/rna.1.1.932

Hedges, J., West, M., and Johnson, A. W. (2005). Release of the export adapter, Nmd3p, from the 60S ribosomal subunit requires Rpl10p and the cytoplasmic GTPase Lsg1p. *EMBO J.* 24, 567–579. doi: 10.1038/sj.emboj.7600547

Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A., and Gleizes, P. E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* 6, 225–242. doi: 10.1002/wrna.1269

Henras, A. K., Soudet, J., Gérus, M., Lebaron, S., Caizergues-Ferrer, M., Mougin, A., et al. (2008). The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell. Mol. Life Sci.* 65, 2334–2359. doi: 10.1007/s00018-008-8027-0

Ho, J. H., Kallstrom, G., and Johnson, A. W. (2000). Nmd3p is a Crm1p-dependent adapter protein for nuclear export of the large ribosomal subunit. *J. Cell Biol.* 151, 1057–1066. doi: 10.1083/jcb.151.5.1057

Hofer, A., Bussiere, C., and Johnson, A. W. (2007). Mutational analysis of the ribosomal protein Rpl10 from yeast. *J. Biol. Chem.* 282, 32630–32639. doi: 10.1074/jbc.M705057200

Hoppstock, L., Trusch, F., Lederer, C., van West, P., Koenneke, M., and Bayer, P. (2016). NmPin from the marine thaumarchaeote Nitrosopumilus maritimus is an active membrane associated prolyl isomerase. *BMC Biol.* 14:53. doi: 10.1186/s12915-016-0274-1

Hung, N. J., Lo, K. Y., Patel, S. S., Helmke, K., and Johnson, A. W. (2008). Arx1 is a nuclear export receptor for the 60S ribosomal subunit in yeast. *Mol. Biol. Cell* 19, 735–744. doi: 10.1091/mbc.e07-09-0968

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

Jain, A., Perisa, D., Fliedner, F., von Haeseler, A., and Ebersberger, I. (2019). The evolutionary traceability of a protein. *Genome Biol. Evol.* 11, 531–545. doi: 10.1093/gbe/evz008

Jaremko, L., Jaremko, M., Elfaki, I., Mueller, J. W., Ejchart, A., Bayer, P., et al. (2011). Structure and dynamics of the first archaeal parvulin reveal a new functionally important loop in parvulin-type prolyl isomerases. *J. Biol. Chem.* 286, 6554–6565. doi: 10.1074/jbc.M110.160713

Jiang, Z., Carlantoni, C., Allanki, S., Ebersberger, I., and Stainier, D. Y. R. (2020). Tek (Tie2) is not required for cardiovascular development in zebrafish. *Development* 147:dev193029. doi: 10.1242/dev.193029

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. doi: 10.1093/nar/gkaa970

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Karbstein, K. (2013). Quality control mechanisms during ribosome maturation. *Trends Cell Biol.* 23, 242–250. doi: 10.1016/j.tcb.2013.01.004

Kargas, V., Castro-Hartmann, P., Escudero-Urquijo, N., Dent, K., Hilcenko, C., Sailer, C., et al. (2019). Mechanism of completion of peptidyltransferase Centre assembly in eukaryotes. *eLife* 8:e44904. doi: 10.7554/eLife.44904

Kater, L., Thoms, M., Barrio-Garcia, C., Cheng, J., Ismail, S., Ahmed, Y. L., et al. (2017). Visualizing the assembly pathway of nucleolar pre-60S ribosomes. *Cell* 171:e1514. doi: 10.1016/j.cell.2017.11.039

Keeling, P. J., Corradi, N., Morrison, H. G., Haag, K. L., Ebert, D., Weiss, L. M., et al. (2010). The reduced genome of the parasitic microsporidian Enterocytozoon bieneusi lacks genes for core carbon metabolism. *Genome Biol. Evol.* 2, 304–309. doi: 10.1093/gbe/evq022

Kemmler, S., Occhipinti, L., Veisu, M., and Panse, V. G. (2009). Yvh1 is required for a late maturation step in the 60S biogenesis pathway. *J. Cell Biol.* 186, 863–880. doi: 10.1083/jcb.200904111

Klingauf-Nerurkar, P., Gillet, L. C., Portugal-Calisto, D., Oborska-Oplova, M., Jager, M., Schubert, O. T., et al. (2020). The GTPase Nog1 co-ordinates the assembly, maturation and quality control of distant ribosomal functional centers. *eLife* 9:e52474. doi: 10.7554/eLife.52474

Klinge, S., and Woolford, J. L. Jr. (2019). Ribosome assembly coming into focus. *Nat. Rev. Mol. Cell Biol.* 20, 116–131. doi: 10.1038/s41580-018-0078-y

Knuppel, R., Trahan, C., Kern, M., Wagner, A., Grunberger, F., Hausner, W., et al. (2021). Insights into synthesis and function of KsgA/Dim1-dependent rRNA modifications in archaea. *Nucleic Acids Res.* 49, 1662–1687. doi: 10.1093/nar/gkaa1268

Koestler, T., Av, H., and Ebersberger, I. (2010). FACT: functional annotation transfer between proteins with similar feature architectures. *BMC Bioinformatics* 11:417. doi: 10.1186/1471-2105-11-417

Koonin, E. V., Mushegian, A. R., and Bork, P. (1996). Non-orthologous gene displacement. *Trends Genet.* 12, 334–336. doi: 10.1016/0168-9525(96)20010-1

Koonin, E. V., Wolf, Y. I., and Aravind, L. (2001). Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* 11, 240–252. doi: 10.1101/gr.162001

Kowalak, J. A., Bruenger, E., Crain, P. F., and McCloskey, J. A. (2000). Identities and phylogenetic comparisons of posttranscriptional modifications in 16 S ribosomal RNA from Haloferax volcanii. *J. Biol. Chem.* 275, 24484–24489. doi: 10.1074/jbc.M002153200

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

Leisegang, M. S., Martin, R., Ramirez, A. S., and Bohnsack, M. T. (2012). Exportin T and Exportin 5: tRNA and miRNA biogenesis—and beyond. *Biol. Chem.* 393, 599–604. doi: 10.1515/hsz-2012-0146

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res.* 37, D229–D232. doi: 10.1093/nar/gkn808

Liang, X., Zuo, M. Q., Zhang, Y., Li, N., Ma, C., Dong, M. Q., et al. (2020). Structural snapshots of human pre-60S ribosomal particles before and after nuclear export. *Nat. Commun.* 11:3542. doi: 10.1038/s41467-020-17237-x

Liu, Y., Makarova, K. S., Huang, W. C., Wolf, Y. I., Nikolskaya, A. N., Zhang, X., et al. (2021). Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 593, 553–557. doi: 10.1038/s41586-021-03494-3

Lo, K. Y., Li, Z., Wang, F., Marcotte, E. M., and Johnson, A. W. (2009). Ribosome stalk assembly requires the dual-specificity phosphatase Yvh1 for the exchange of Mrt4 with P0. *J. Cell Biol.* 186, 849–862. doi: 10.1083/jcb.200904110

Londei, P., and Ferreira-Cerca, S. (2021). Ribosome biogenesis in archaea. *Front. Microbiol.* 12:686977. doi: 10.3389/fmicb.2021.686977

Londei, P., Teixido, J., Acca, M., Cammarano, P., and Amils, R. (1986). Total reconstitution of active large ribosomal subunits of the thermoacidophilic archaebacterium Sulfolobus solfataricus. *Nucleic Acids Res.* 14, 2269–2285. doi: 10.1093/nar/14.5.2269

Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* 303, 95–98. doi: 10.1126/science.1090599

Lygerou, Z., Allmang, C., Tollervey, D., and Seraphin, B. (1996). Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science* 272, 268–270. doi: 10.1126/science.272.5259.268

Malyutin, A. G., Musalgaonkar, S., Patchett, S., Frank, J., and Johnson, A. W. (2017). Nmd3 is a structural mimic of eIF5A, and activates the cpGTPase Lsg1 during 60S ribosome biogenesis. *EMBO J.* 36, 854–868. doi: 10.15252/embj.201696012

Martin-Duran, J. M., Ryan, J. F., Vellutini, B. C., Pang, K., and Hejnol, A. (2017). Increased taxon sampling reveals thousands of hidden orthologs in flatworms. *Genome Res.* 27, 1263–1272. doi: 10.1101/gr.216226.116

Moy, T. I., and Silver, P. A. (1999). Nuclear export of the small ribosomal subunit requires the ran-GTPase cycle and certain nucleoporins. *Genes Dev.* 13, 2118–2133. doi: 10.1101/gad.13.16.2118

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nieto, B., Gaspar, S. G., Moriggi, G., Pestov, D. G., Bustelo, X. R., and Dosil, M. (2020). Identification of distinct maturation steps involved in human 40S ribosomal subunit biosynthesis. *Nat. Commun.* 11:156. doi: 10.1038/s41467-019-13990-w

Noon, K. R., Bruenger, E., and McCloskey, J. A. (1998). Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus. *J. Bacteriol.* 180, 2883–2888. doi: 10.1128/JB.180.11.2883-2888.1998

Ohno, S. (1970). *Evolution by Gene Duplication*. New York: Springer.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Ostlund, G., Schmitt, T., Forslund, K., Kostler, T., Messina, D. N., Roopra, S., et al. (2010). InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* 38, D196–D203. doi: 10.1093/nar/gkp931

Pertschy, B., Saveanu, C., Zisser, G., Lebreton, A., Tengg, M., Jacquier, A., et al. (2007). Cytoplasmic recycling of 60S preribosomal factors depends on the AAA protein Drg1. *Mol. Cell. Biol.* 27, 6581–6592. doi: 10.1128/MCB.00668-07

Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergstrom, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5

Phung, D. K., Etienne, C., Batista, M., Langendijk-Genevaux, P., Moalic, Y., Laurent, S., et al. (2020). RNA processing machineries in archaea: the 5′-3′ exoribonuclease aRNase J of the beta-CASP family is engaged specifically with the helicase ASH-Ski2 and the 3′-5′ exoribonucleolytic RNA exosome machinery. *Nucleic Acids Res.* 48, 3832–3847. doi: 10.1093/nar/gkaa052

Rodriguez-Mateos, M., Garcia-Gomez, J. J., Francisco-Velilla, R., Remacha, M., de la Cruz, J., and Ballesta, J. P. (2009). Role and dynamics of the ribosomal protein P0 and its related trans-acting factor Mrt4 during ribosome assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 37, 7519–7532. doi: 10.1093/nar/gkp806

Rosenblad, M. A., Lopez, M. D., Piccinelli, P., and Samuelsson, T. (2006). Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides further evidence of homology between the yeast and human enzymes. *Nucleic Acids Res.* 34, 5145–5156. doi: 10.1093/nar/gkl626

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng.* 12, 85–94.

Roustan, V., Jain, A., Teige, M., Ebersberger, I., and Weckwerth, W. (2016). An evolutionary perspective of AMPK-TOR signaling in the three domains of life. *J. Exp. Bot.* 67, 3897–3907. doi: 10.1093/jxb/erw211

Sanchez, M. E., Urena, D., Amils, R., and Londei, P. (1990). In vitro reassembly of active large ribosomal subunits of the halophilic archaebacterium Haloferax mediterranei. *Biochemistry* 29, 9256–9261. doi: 10.1021/bi00491a021

Sanghai, Z. A., Miller, L., Molloy, K. R., Barandun, J., Hunziker, M., Chaker-Margot, M., et al. (2018). Modular assembly of the nucleolar pre-60S ribosomal subunit. *Nature* 556, 126–129. doi: 10.1038/nature26156

Saveanu, C., Namane, A., Gleizes, P.-E., Lebreton, A., Rousselle, J.-C., Noaillac-Depeyre, J., et al. (2003). Sequential protein association with nascent 60S ribosomal particles. *Mol. Cell. Biol.* 23, 4449–4460. doi: 10.1128/MCB.23.13.4449-4460.2003

Sergiev, P. V., Aleksashin, N. A., Chugunova, A. A., Polikanov, Y. S., and Dontsova, O. A. (2018). Structural and evolutionary insights into ribosomal RNA methylation. *Nat. Chem. Biol.* 14, 226–235. doi: 10.1038/nchembio.2569

Shajani, Z., Sykes, M. T., and Williamson, J. R. (2011). Assembly of bacterial ribosomes. *Annu. Rev. Biochem.* 80, 501–526. doi: 10.1146/annurev-biochem-062608-160432

Sloan, K. E., Gleizes, P. E., and Bohnsack, M. T. (2016). Nucleocytoplasmic transport of RNAs and RNA-protein complexes. *J. Mol. Biol.* 428, 2040–2059. doi: 10.1016/j.jmb.2015.09.023

Sloan, K. E., Schneider, C., and Watkins, N. J. (2012). Comparison of the yeast and human nuclear exosome complexes. *Biochem. Soc. Trans.* 40, 850–855. doi: 10.1042/BST20120061

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J., and Bohnsack, M. T. (2017). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781

Squatrito, M., Mancino, M., Donzelli, M., Areces, L. B., and Draetta, G. F. (2004). EBP1 is a nucleolar growth-regulating protein that is part of pre-ribosomal ribonucleoprotein complexes. *Oncogene* 23, 4454–4465. doi: 10.1038/sj.onc.1207579

Stage-Zimmermann, T., Schmidt, U., and Silver, P. A. (2000). Factors affecting nuclear export of the 60S ribosomal subunit in vivo. *Mol. Biol. Cell* 11, 3777–3789. doi: 10.1091/mbc.11.11.3777

Steinegger, M., and Salzberg, S. L. (2020). Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol.* 21:115. doi: 10.1186/s13059-020-02023-1

Tafforeau, L., Zorbas, C., Langhendries, J. L., Mullineux, S. T., Stamatopoulou, V., Mullier, R., et al. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. *Mol. Cell* 51, 539–551. doi: 10.1016/j.molcel.2013.08.011

Taoka, M., Nobe, Y., Yamaki, Y., Yamauchi, Y., Ishikawa, H., Takahashi, N., et al. (2016). The complete chemical structure of *Saccharomyces cerevisiae* rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res.* 44, 8951–8961. doi: 10.1093/nar/gkw564

Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A genomic perspective on protein families. *Science* 278, 631–637. doi: 10.1126/science.278.5338.631

Taylor, J. S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38, 615–643. doi: 10.1146/annurev.genet.38.072902.092831

Tran, N. V., Greshake Tzovaras, B., and Ebersberger, I. (2018). PhyloProfile: dynamic visualization and exploration of multi-layered phylogenetic profiles. *Bioinformatics* 34, 3041–3043. doi: 10.1093/bioinformatics/bty225

van Tran, N., Ernst, F. G. M., Hawley, B. R., Zorbas, C., Ulryck, N., Hackert, P., et al. (2019). The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Res.* 47, 7719–7733. doi: 10.1093/nar/gkz619

Wandrey, F., Montellese, C., Koos, K., Badertscher, L., Bammert, L., Cook, A. G., et al. (2015). The NF45/NF90 heterodimer contributes to the biogenesis of 60S ribosomal subunits and influences nucleolar morphology. *Mol. Cell. Biol.* 35, 3491–3503. doi: 10.1128/MCB.00306-15

Warner, J. R. (2015). Twenty years of ribosome assembly and ribosomopathies. *RNA* 21, 758–759. doi: 10.1261/rna.050435.115

Wild, T., Horvath, P., Wyler, E., Widmann, B., Badertscher, L., Zemp, I., et al. (2010). A protein inventory of human ribosome biogenesis reveals an essential function of exportin 5 in 60S subunit export. *PLoS Biol.* 8:e1000522. doi: 10.1371/journal.pbio.1000522

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.

Woolford, J. L. Jr., and Baserga, S. J. (2013). Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* 195, 643–681. doi: 10.1534/genetics.113.153197

Wu, S., Tutuncuoglu, B., Yan, K., Brown, H., Zhang, Y., Tan, D., et al. (2016). Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature* 534, 133–137. doi: 10.1038/nature17942

Yang, X. C., Purdy, M., Marzluff, W. F., and Dominski, Z. (2006). Characterization of 3'hExo, a 3′ exonuclease specifically interacting with the 3′ end of histone mRNA. *J. Biol. Chem.* 281, 30447–30454. doi: 10.1074/jbc.M602947200

Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* 17, 3011–3016. doi: 10.1101/gad.1158803

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

Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simao, F. A., Ioannidis, P., et al. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* 45, D744–D749. doi: 10.1093/nar/gkw1119

Zhou, Y., Musalgaonkar, S., Johnson, A. W., and Taylor, D. W. (2019). Tightly-orchestrated rearrangements govern catalytic center assembly of the ribosome. *Nat. Commun.* 10:958. doi: 10.1038/s41467-019-08880-0

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 October 2021
doi: 10.3389/fmicb.2021.742806

Elucidation of the Translation Initiation Factor Interaction Network of *Haloferax volcanii* Reveals Coupling of Transcription and Translation in Haloarchaea

*Franziska Schramm[1], Andreas Borst[1], Uwe Linne[2] and Jörg Soppa[1]**

[1] Institute for Molecular Biosciences, Biocentre, Goethe-University, Frankfurt, Germany, [2] Mass Spectrometry Facility, Department of Chemistry, Phillipps University Marburg, Marburg, Germany

OPEN ACCESS

Edited by:
Paola Londei, Sapienza University of Rome, Italy

Reviewed by:
Gregor Blaha, University of California, Riverside, United States
Emmanuelle Schmitt, UMR7654 Bases moléculaires et régulation de la biosynthèse protéique, France

***Correspondence:**
Jörg Soppa
soppa@bio.uni-frankfurt.de

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

Received: *16 July 2021*
Accepted: *29 September 2021*
Published: *26 October 2021*

Citation:
Schramm F, Borst A, Linne U and Soppa J (2021) Elucidation of the Translation Initiation Factor Interaction Network of Haloferax volcanii Reveals Coupling of Transcription and Translation in Haloarchaea. Front. Microbiol. 12:742806. doi: 10.3389/fmicb.2021.742806

Translation is an important step in gene expression. Initiation of translation is rate-limiting, and it is phylogenetically more diverse than elongation or termination. Bacteria contain only three initiation factors. In stark contrast, eukaryotes contain more than 10 (subunits of) initiation factors (eIFs). The genomes of archaea contain many genes that are annotated to encode archaeal homologs of eukaryotic initiation factors (aIFs). However, experimental characterization of aIFs is scarce and mostly restricted to very few species. To broaden the view, the protein–protein interaction network of aIFs in the halophilic archaeon *Haloferax volcanii* has been characterized. To this end, tagged versions of 14 aIFs were overproduced, affinity isolated, and the co-isolated binding partners were identified by peptide mass fingerprinting and MS/MS analyses. The aIF–aIF interaction network was resolved, and it was found to contain two interaction hubs, (1) the universally conserved factor aIF5B, and (2) a protein that has been annotated as the enzyme ribose-1,5-bisphosphate isomerase, which we propose to rename to aIF2Bα. Affinity isolation of aIFs also led to the co-isolation of many ribosomal proteins, but also transcription factors and subunits of the RNA polymerase (Rpo). To analyze a possible coupling of transcription and translation, seven tagged Rpo subunits were overproduced, affinity isolated, and co-isolated proteins were identified. The Rpo interaction network contained many transcription factors, but also many ribosomal proteins as well as the initiation factors aIF5B and aIF2Bα. These results showed that transcription and translation are coupled in haloarchaea, like in *Escherichia coli*. It seems that aIF5B and aIF2Bα are not only interaction hubs in the translation initiation network, but also key players in the transcription-translation coupling.

Keywords: *Haloferax volcanii*, translation initiation, aIF, ribosome, RNA polymerase, interaction networks, transcription, coupling

INTRODUCTION

Translation is a very important step in the process of expression of the genome information into the phenotype of cells and organisms. Translation is evolutionary very old, and ribosomes were already present in the Last Universal Common Ancestor of all living beings from the three domains of life (Fox, 2010; Opron and Burton, 2018; Bowman et al., 2020). In fact, comparison of the 16S/18S rRNA has led to the proposal that a third domain of life exist, the archaea, which are not closely related to the second group of prokaryotes, the bacteria (Woese and Fox, 1977). Initially this was based on very few species of methanogenic archaea, however, the molecular distinction between archaea and bacteria based on rRNA sequences has held true after the isolation of hundreds of new species and thousands of rRNA sequences generated by metagenomics. Recently, the three domain concept of life has been challenged, but this does not concern the dichotomy of archaea and bacteria. Instead, the recent discovery of many new groups of archaea currently makes it more likely that the eukaryotes evolved from within the archaea, and thus, that only two major primary domains exist (Eme et al., 2017; Liu et al., 2021). While in evolution different phylogenetic groups added additional subdomains into the rRNA sequences and added lineage-specific ribosomal proteins, a structural core of the ribosomal RNA exists that is shared by archaea, bacteria, and eukaryotes, and the majority of ribosomal proteins are universal (Bernier et al., 2018).

Translation is comprised of the steps' initiation, elongation, termination, and ribosome recycling. Initiation of translation is phylogenetically most diverse among these four steps, and at least five different mechanisms exist. In eukaryotes, canonical translation initiation involves recognition of the 5'-cap structure of mRNAs and scanning of the small 40S ribosomal subunit along the mRNA, until the start codon is reached. Then, the large 60S ribosomal subunit joins, and translation elongation can start. An alternative translation initiation mechanism in eukaryotes involves Internal Ribosome Entry Sites (IRES). These are specific structures within the 5'-UTRs of transcripts (or in intergenic regions of bicistronic viral transcripts) that are recognized by specific proteins, so-called IRES Trans-Acting Factors, which attract the 40S subunit to the internal sites. Various eukaryotic translation initiation factors (eIFs) are involved in and essential for translation initiation (see below). Several reviews summarize different aspects of translation initiation in eukaryotes (Dever et al., 2016; Andreev et al., 2017; Aylett and Ban, 2017; Hinnebusch, 2017; Guca and Hashem, 2018; Shirokikh and Preiss, 2018; Weisser and Ban, 2019).

In bacteria, canonical translation initiation involves base-pairing between the so-called Shine Dalgarno (SD) motif in the mRNA, which is localized a few nucleotides upstream of the start codon, and the anti-SD motif, which is localized at the 3'-end of the 16S rRNA. Thereby, the start codon is localized in the P-site of the small 30S rRNA, and the large 50S subunit can join, before elongation can start. The internal recognition of start sites enables the formation of polycistronic mRNAs, which contain several to many genes. Also in bacteria, initiation factors (IFs) are involved in the process (see below). In addition to the canonical transcripts, also non-canonical transcripts exist in bacteria, which either contain a 5'-UTR lacking a SD motif or lack a 5'-UTR and are leaderless. The fractions of the three groups of transcripts differ widely in different phylogenetic groups of bacteria. For example, the SD mechanism for translation initiation is not functional at all in Bacteriodetes (Accetto and Avguštin, 2011), and the fractions of SD-led genes are rather low in Chlamydia and cyanobacteria (Huber et al., 2019). Translation initiation in bacteria has been reviewed intensively (Marintchev and Wagner, 2004; Kaberdin and Bläsi, 2006; Simonetti et al., 2009; Malys and McCarthy, 2011; Milón and Rodnina, 2012; Duval et al., 2015; Gualerzi and Pon, 2015).

Archaea contain the same three types of transcripts as bacteria, i.e., (1) canonical transcripts with 5'-UTRs and SD motif, (2) non-canonical transcripts with 5'-UTRs lacking an SD motif, and (3) leaderless transcripts. The distribution is very different in various groups of archaea. For example, transcripts in methanogenic archaea typically have very long 5'-UTRs with SD motifs, while, in stark contrast, transcripts in haloarchaea and *Sulfolobales* are typically leaderless. A dRNA-Seq study has shown that 72% of all transcripts of *Haloferax volcanii* are leaderless and the fraction of transcripts with SD motifs is extremely low (Babski et al., 2016). In addition, SD motifs are non-functional for translation initiation at 5'-UTRs in *H. volcanii* (Kramer et al., 2014). Several reviews summarize various aspects about translation initiation in archaea, and compare it with initiation in bacteria and eukaryotes (Londei, 2005; Benelli and Londei, 2009, 2011; Schmitt et al., 2019, 2020).

In stark contrast to the similarities in the classes of transcripts, the numbers of translation initiation factors are totally different in bacteria and archaea. Bacteria contain only three initiation factors, IF1, IF2, and IF3. IF1 is homologous to the archaeal factor aIF1A and the eukaryotic factor eIF1A, and it is thus universally conserved. The second universally conserved factor is IF2, which is homologous to the archaeal factor aIF5B and the eukaryotic factor eIF5B. The third bacterial factor, IF3, has some structural similarities with the factors aIF1 and eIF1, but the sequences and topologies are different, and thus, bacterial IF3 and aIF1/eIF1 are not homologues.

Archaea and eukaryotes share several additional factors, which are not present in bacteria. A central factor is the heterotrimeric factor aIF2/eIF2, which binds the initiator tRNA and brings it into the P-site of the ribosome. aIF1 and eIF1 are homologous and are found in the preinitiation complex of archaea and eukaryotes together with aIF2/eIF2 and aIF1A/eIF1A, thus enhancing the accuracy of start codon selection (Schmitt et al., 2020). The eukaryotic factor eIF4F consists of three subunits, a homolog of one of which, a/eIF4A, is encoded in many archaeal genomes. However, it cannot have the same function as the eukaryotic factor. The eukaryotic factor eIF4F binds to the cap of eukaryotic transcripts and brings the preinitiation complex to the mRNA 5'-end. However, archaeal transcripts do not have a 5'-cap, and thus there is no use for

a cap-binding factor. Therefore, it is not clear whether aIF4A has a function in translation initiation at all.

Another initiation factor in eukaryotes is eIF2B. It consists of a catalytic subcomplex of two subunits (eIF2Bγ, eIF2Bε) and a regulatory subunit of three subunits (eIF2Bα, eIF2Bβ, eIF2Bδ). The whole complex is a decamer, because there are two copies of each of the five subunits in the complex (Bogorad et al., 2014; Schoof et al., 2021). eIF2B binds to the central factor eIF2 and catalyzes the exchange of GDP against GTP. The two catalytic subunits are not encoded in archaea, while there are genes for the three regulatory subunits in many archaeal genomes. The biological role of these "regulatory proteins" in the absence of their catalytic binding partners is unknown. It could be shown that a presumed aIF2B subunit from three species, *Pyrococcus horikoshii*, *Pyrococcus furiosus*, and *Thermoplasma acidophilum*, binds to the alpha subunit of aIF2 of the cognate species *in vitro* (Dev et al., 2009), indicating that the aIF2B subunits might have some – as yet unknown – function in archaeal translation initiation.

In recent years considerable progress in the experimental characterization of translation initiation has been obtained, albeit the number of studies is much lower than the number of studies in eukaryotes or in bacteria (in Pubmed the numbers of studies with "translation initiation" AND archae*, bacteri*, or eukaryote* in Title/Abstract are 187, 892, and 4,890, respectively). By far the highest number of studies have been performed with the Crenarchaeon *Sulfolobus solfataricus* (La Teana et al., 2013; Schmitt et al., 2020). Structures of preinitiation complexes have been solved with constituents from *Pyrococcus abyssi* (Coureux et al., 2016, 2020). The number of studies with halophilic archaea is very low. As mentioned above, it could be shown that the SD motif is non-functional for translation initiation *in vivo* (Kramer et al., 2014), and that a novel mechanism for translation initiation operates (Hering et al., 2009). It was also revealed that the 5'-ends and 3'-ends of *H. volcanii* transcripts functionally interact *in vivo* (Brenneis and Soppa, 2009). In a very comprehensive study with 14 genes that were annotated to encode (subunits of) translation initiation factors, all nine non-essential genes were deleted and all five essential genes were conditionally depleted, and the consequences for the phenotype of the mutants were characterized (Gäbel et al., 2013). In the present study we have extended this approach, and 14 proteins with the annotation to be (subunits of) translation initiation factors of *H. volcanii* were overproduced as tagged variants. After affinity purification, co-isolated binding partners were identified by peptide mass fingerprinting and MS–MS analyses. Thereby, the protein–protein interaction network of haloarchaeal translation initiation could be resolved. The unexpected co-isolation of several subunits of the RNA polymerase prompted us to extend the project further. To this end, seven subunits of the RNA polymerase were overproduced, and the RNA polymerase interaction network was also elucidated using co-affinity isolation and MS as well as MS/MS analyses. Together, we report here a very comprehensive analysis of the interaction network of 21 proteins, which was controlled by two very strict negative controls, i.e., cultures containing an empty vector and cultures overproducing a metabolic enzyme.

MATERIALS AND METHODS

Strains, Media and Culture Conditions

The strain *H. volcanii* H26 was obtained from Thorsten Allers (Nottingham, United Kingdom), it is a *pyrE* deletion strain lacking the plasmid pHV2. The deletion of the *dhfr* (dihydrofolate reductase) gene HVO_1279 has been described previously (Maurer et al., 2018). The deletion strains of genes encoding translation initiation factors have been generated in a previous study (Gäbel et al., 2013). Multi cycle PCRs were used to confirm that all mutants were still homozygous. The sequences of oligonucleotides are listed in **Supplementary Table S1**. In a few cases the mutants could not be regrown from permanent cultures, therefore, they were regenerated as described (Gäbel et al., 2013) using the oligonucleotides listed in **Supplementary Table S1**. All overproduction strains have been generated in this study (see below).

Haloferax volcanii strains were grown in complex medium with 50 μg/ml uracil as previously described (Dambeck and Soppa, 2008). The cultures were grown in Erlenmeyer flasks at 42°C with shaking at 250 rpm. Growth was either measured spectroscopically at 600 nm or cells were counted using a Neubauer counting chamber.

The *E. coli* strain XL1-blue MRF' (Agilent Technologies, Waldbronn, Germany) was used for cloning. It was grown in complex SOB medium under standard conditions (Hanahan, 1983).

Generation of Overproduction Strains

For overproduction of proteins, the respective genes were cloned into the shuttle vector pSD1/R1-6 under the control of a strong synthetic constitutive promoter (Danner and Soppa, 1996). The genes were amplified using the primers listed in **Supplementary Table S2** using genomic DNA from *H. volcanii* as template. The primers added the sequences for an N-terminal hexahistidine tag to the genes. The plasmids were isolated from *E. coli* and the sequences were verified before they were used to transform *H. volcanii*. All non-essential aIFs were overproduced in the respective deletion stains, while all essential aIFs were over-produced in the strain H26 Δ1279 that was used as wild-type strain concerning all proteins of this study. **Table 1** gives an overview of the overproduced aIFs and the production strains. The dihydrofolate reductase (DHFR) was used as a negative control protein that is not involved in translation initiation. All subunits of the RNA polymerase were assumed to be essential without any testing, and, therefore, they were over-produced in the wildtype strain H26 Δ1279 (**Table 2**). All strains with expression plasmids derived from pSD1 were grown in the presence of Novobiocin (0.5 μg/ml).

TABLE 1 | Overview of aIF-genes and strains used in this study.

Name in this study	Gene ID	Production-Strain	Accession	MW [kDa]	Gene Name (Halolex)	Protein Name (Halolex)
dhfr	HVO_1279	WT	L9UT07	18.0	hdrA, folA1	dihydrofolate reductase
aIF1	HVO_1946	WT	D4GTH5	11.0	tif1a	translation initiation factor aIF-1 (SUI1 protein, bacterial-type IF3)
aIF1A1	HVO_0136	WT Δ0136	D4GZ79	11.5	tif1A1	translation initiation factor aIF-1A
aIF1A2	HVO_A0637	WT ΔA0637	D4GRU5	11.2	tif1A2	translation initiation factor aIF-1A
aIF2α	HVO_0699	WT Δ0699	D4GT46	29.5	tif2a	translation initiation factor aIF2 alpha subunit
aIF2β-1	HVO_1678	WT Δ1678	D4GZP2	15.0	tif2b1	translation initiation factor aIF2 beta subunit
aIF2β-2	HVO_2242	WT Δ2242	D4GVV8/L9VAS4	22.2	tif2b2	translation initiation factor aIF2 beta subunit / probable RNA-binding protein
aIF2γ	HVO_1901	WT	D4GTD4	44.0	tif2c	translation initiation factor aIF2 gamma subunit
aIF2Bsu	HVO_1934	WT Δ1934	D4GTG3	43.2	–	NUDIX family hydrolase/eIF-2B domain protein
aIF2Bα	HVO_0966	WT	L9USK7	35.0	–	ribose-1,5-bisphosphate isomerase
aIF2Bδ	HVO_2706	WT Δ2706	D4GW08/L9V7F9	30.8	–	eIF-2B domain protein
eIF4A-homolog	HVO_1333	WT Δ1333	D4GXK1/L9UST9	104.5	lhr2	ATP-dependent DNA helicase
aIF5A	HVO_2300	WT	D4GWG6/L9V7A1	14.2	tef5A	translation elongation factor aEF-5A
aIF5B	HVO_1963	WT	D4GTJ2	65.4	tif5B	translation initiation factor aIF-5B (bacterial-type IF2)
aIF6	HVO_0117	WT	D4GYW3/L9UI67	23.0	tif6	translation initiation factor aIF-6

WT = H26 Δ HVO_1279.

Characterization of Growth Curves

Growth curves were generated for the wild-type, for all deletion mutants, for all production strains with expression plasmids, and, as controls, for all respective strains containing the empty vector. In each case, exponentially growing pre-cultures were used to inoculate the test cultures. For each condition, 150 µl medium was inoculated in triplicates in 96-well plates to an OD_{600} of 0.05. The outermost wells were filled with 1 M NaCl to inhibit evaporation from the inner wells containing the test cultures. The OD_{600} was determined frequently using a microtiter plate photometer (Spectramax 340, Molecular Devices). Average values and their standard deviations were used to generate growth curves.

Co-affinity Isolation of Proteins

Haloferax volcanii production cultures were grown overnight in complex medium, and the exponentially growing cells were harvested by centrifugation (4,700 rpm, 30 min., 4°C). The pellet was suspended in 4 ml of ice-cold binding buffer (2.1 M NaCl, 20 mM HEPES, 20 mM imidazole) and the cells were lysed by sonication on ice (3 × 30 s, 50% duty cycle, output strength three). The lysate was subsequently centrifuged to remove cell debris and membranes (13,000 rpm, 30 min., 4°C) and to generate a cytoplasmic extract. 30 µl aliquots were removed for analysis by SDS PAGE, the remaining supernatants were used for co-affinity isolation.

To this end, 500 µl 50% Nickel Chelating Sepharose® Fast Flow beads (NCS, GE Healthcare) were pelleted and resuspended in 1 ml 0.2 M $NiCl_2$ solution. After incubation for 5 min, the NCS was pelleted (13,000 rpm, 30 s.), washed three times in aqua bidest., and suspended in binding buffer (2.1 M NaCl, 20 mM HEPES, 20 mM imidazole).

Two hundred and fifty micro liter of 50% NCS was pelleted and resuspended in 1.6 ml cytoplasmic extract and incubated at room temperature with mixing to enable binding of protein complexes *via* his-tagged bait proteins. The NCS was pelleted, and an aliquot was removed from the supernatant for SDS-PAGE analysis of unbound proteins. The beads were washed four times with 1.6 ml wash buffer (2.1 M NaCl, 20 mM HEPES, 30 mM imidazole). Bound proteins were eluted by the incubation of the NCS in 0.1 ml elution buffer (2.1 M NaCl, 20 mM HEPES, 700 mM imidazole). After centrifugation and removal of the supernatant, a second elution step was performed with 0.1 ml elution buffer. The eluates were dialyzed against 25 mM Tris/HCl, pH7.2 on 13 mm plates (Merck) or using a MEMBRA-CEL® 3.5 kDa tube. Aliquots representing all steps of the co-affinity isolation procedure were analyzed by SDS-PAGE. For normal-sized proteins standard SDS-PAGE was used, for small proteins below 10 kDa Tricine-SDS-PAGE was used instead (Jiang et al., 2016). Suitable elution fractions were used to identify the protein composition by peptide mass fingerprinting and MS–MS analyses.

Identification of Proteins by Mass Spectrometry

Samples were digested by the addition of Sequencing Grade Modified Trypsin (Serva) and incubated at 37°C overnight.

Peptides were desalted and concentrated using Chromabond C18WP spin columns (Macherey-Nagel, Part No. 730522). Finally, peptides were dissolved in 25 µl of water with 5% acetonitrile and 0.1% formic acid.

The mass spectrometric analysis of the samples was performed using an Orbitrap Velos Pro mass spectrometer (ThermoScientific). An Ultimate nanoRSLC-HPLC system (Dionex), equipped with a custom end-fritted 50 cm × 75 µm ID C18 RP column filled with 2.4 µm beads (Dr. Maisch) was connected online to the mass spectrometer through a Proxeon nanospray source. 1–15 µl (depending on peptide concentration

TABLE 2 | Overview of rpo-genes and strains used in this study.

Name in this study	Gene ID	Production-Strain	Accession	MW [kDa]	Gene Name (Halolex)	Protein Name (Halolex)
rpoA1	HVO_0349	WT	D4GZX6	108.8	rpo1n, rpoA1	DNA-directed RNA polymerase subunit Rpo1N
rpoA2	HVO_0350	WT	D4GZX7	46.1	rpo1c, rpoA2	DNA-directed RNA polymerase subunit Rpo1C
rpoB1	HVO_0348	WT	L9UJM2	67.7	rpo2c, rpoB1	DNA-directed RNA polymerase subunit Rpo2C
rpoB2	HVO_0347	WT	L9UK99	58.9	rpo2n, rpoB2	DNA-directed RNA polymerase subunit Rpo2N
rpoD	HVO_2781	WT	L9V5W2	28.1	rpo3, rpoD	DNA-directed RNA polymerase subunit Rpo3
rpoH	HVO_0346	WT	D4GZX3	8.5	rpo5, rpoH	DNA-directed RNA polymerase subunit Rpo5
rpoL	HVO_1042	WT	D4GVL8	10.4	rpo11, rpoL	DNA-directed RNA polymerase subunit Rpo11

WT = H26 Δ HVO_1279.

and sample complexity) of the tryptic digest were injected onto a 1 cm × 300 µm ID C18 PepMap pre-concentration column (Thermo Scientific). Automated trapping and desalting of the sample was performed at a flowrate of 6 µl/min using water/0.05% formic acid as solvent.

Separation of the tryptic peptides was achieved with the following gradient of water/0.05% formic acid (solvent A) and 80% acetonitrile/0.045% formic acid (solvent B) at a flow rate of 300 nl/min: holding 4% B for 5 min, followed by a linear gradient to 45% B within 30 min and linear increase to 95% solvent B in additional 5 min. The column was connected to a stainless steel nanoemitter (Proxeon, Denmark) and the eluent was sprayed directly towards the heated capillary of the mass spectrometer using a potential of 2,300 V. A survey scan with a resolution of 60,000 within the Orbitrap mass analyzer was combined with at least three data-dependent MS/MS scans with dynamic exclusion for 30 s either using CID with the linear ion-trap or using HCD combined with orbitrap detection at a resolution of 7,500. Data analysis was performed using Proteome Discoverer 2.2 (ThermoScientific) with SEQUEST search engine.

The identification of the protein compositions in elution fractions was performed for 14 aIFs and 7 subunits of the RNA polymerase. Two different negative controls were included, i.e., (1) *H. volcanii* cultures containing the empty vector, and (2) cultures overproducing the metabolic enzyme DHFR. The bioinformatic workflow for the removal of contaminants and false positives and identification of proteins that specifically bind to the overproduced aIFs or Rpos is discussed in the **Results** section.

Generation of Phylogenetic Trees

After the genome of *H. volcanii* had been sequenced (Hartman et al., 2010), HVO_0966 was annotated to be a subunit of the translation initiation factor aIF2B. Later, the annotation was changed to the metabolic enzyme ribose-1,5-bisphosphate isomerase (R15BI). The three regulatory subunits of the eukaryotic translation initiation factor eIF2B (eIF2Bα, eIF2Bβ, eIF2Bδ) are homologous to one another and to a family of sugar phosphate isomerases methylthiophosphoribose isomerases (MTPI). It was decided to analyze the phylogeny of HVO_0966 with the aim to find indications whether it is more likely to be an aIF2B subunit than a metabolic enzyme.

BLAST searches at the website of the European Bioinformatics Institute were used to retrieve, in total, 34 sequences of proteins that are homologous to HVO_0966.[1] At first, the taxonomic subset "human" was searched with HVO_0966, and the human sequences of eIF2Bα, eIF2Bβ, eIF2Bδ and of MTPI were retrieved. BLAST searches with these four protein sequences in the phylogenetic subsets "mammals," "rodents," "arthropoda," "plants," and "fungi" were used to retrieve the most similar non-human homologue of each group. Thereby, 24 eukaryotic sequences were retrieved (including the human proteins). In addition, the four human proteins were used for BLAST searches in the taxonomic subset "archaea," to retrieve archaeal proteins that are similar to the four human protein families. At last, HVO_0966 was used for a BLAST search in the taxonomic subset of "archaea," and proteins from different phylogenetic groups of archaea were retrieved, yielding a total set of 35 proteins (including HVO_0966).

The program "MEGA X" (Kumar et al., 2018) was used to generate a multiple sequence alignment (MSA) and calculate phylogenetic trees. All protein sequences were loaded individually into the program, and an MSA was generated. MEGA X allows to visualize and edit the MSA. This was used to remove positions that are phylogenetically un-informative, e.g., the N-terminal region that is exclusively present in the eIF2Bδ subfamily as well as insertions that are present in only one or very few sequences and that are obviously non-conserved. The resulting MSA was used to calculate three phylogenetic trees, which are based on the Maximum Likelihood, the Neighbor Joining, and the Maximum Parsimony algorithm. In each case 1,000 bootstrap replications were performed, and the fractions that were retrieved at each node were written to the respective

[1]https://www.ebi.ac.uk/Tools/sss/ncbiblast/

nodes (% values). Capital letters were added to selected nodes to facilitate the discussion in the **Results** and **Discussion** sections.

Databases and Programs

Bioinformatic analyses of the *H. volcanii* genome were performed at the website Halolex (Pfeiffer et al., 2008). The Halolex database is freely available, but currently usage is restricted to registered users. To request access, send a mail to halolex@rzg.mpg.de. The Integrated Genome Browser (Freese et al., 2016) was used to visualize the genome annotation, as well as the results of the dRNA-Seq study (Babski et al., 2016) and a recent RNA-Seq study (Laass et al., 2019). The program "Clone Manager"[2] was used for the design of primer sequences and cloning experiments. The EMBL-EBI website[3] was used for BLAST searches and to retrieve protein sequences.

RESULTS

The Co-affinity Isolation Approach: Experimental Design and Data Analysis

The aim of this study was to unravel the protein–protein interaction network of translation initiation factors from the halophilic archaeon *H. volcanii*. In total, 14 genes are annotated in the genome of *H. volcanii* to encode (subunits of) translation initiation factors, which are summarized in **Table 1**. In a previous study, it was attempted to delete all these genes (Gäbel et al., 2013). Nine single-gene *in frame* deletion mutants could be successfully generated, while five genes turned out to be essential. In the present study, eight of the nine deletion mutants were used as background strains for the overproduction of the respective initiation factor. The remaining five proteins and one additional protein were overproduced in the wild-type. **Supplementary Figure S1A** gives an overview of the experimental workflow of co-affinity isolation. After overproduction of his-tagged versions of the proteins, cells were harvested and re-suspended in a high salt solution. Haloarchaea use the so-called salt-in strategy for osmotic adaptation, and the cytoplasmatic salt concentration equals that of the high-salt environment. Therefore, the co-affinity purification has to be performed under the native high salt conditions to prevent the dissociation of protein complexes and the unfolding of proteins, which is a typical problem when dealing with haloarchaeal proteins under low salt conditions.

The cells were lysed by sonication, cell debris was removed by centrifugation, and nickel chelating sepharose was used for the affinity purification of his-tagged proteins. **Figure 1A** gives an overview of the affinity isolation of the DHFR. The major band in the elution fractions was the DHFR, showing that overproduction and affinity purification were successful (see red arrow in **Figure 1A**). A second protein of around 70 kDa was also highly enriched (black arrow in **Figure 1A**). This is PitA, a native *H. volcanii* protein with a histidine-rich N-terminus, which binds with high affinity to the nickel chelating sepharose (Allers et al., 2010). Co-purification of PitA could be prevented by replacing the gene with a variant lacking the histidine-rich stretch (Allers et al., 2010). However, in this study we used PitA as an internal control for the success of the affinity isolation.

The same workflow was used for the affinity purification of the 14 aIFs. In all cases, three biological replicates were performed. A strain containing an empty vector was used as a second negative control in addition to the DHFR overproduction strain. Therefore, in total 48 cultures were used for the overproduction and affinity purification of the 14 aIFs and the two negative controls. In all cases, purification gels like the example shown in **Figure 1A** were used to guarantee that the last wash fraction was protein-free, and to estimate the pattern of co-purified proteins in the elution fractions. Typical elution fractions, representing one of the three biological replicates, are shown in **Figure 1B** for the four subunits of aIF2 and in **Supplementary Figure S2** for all other aIFs.

The elution fractions containing the bait proteins and the mixtures of co-isolated proteins were dialyzed against a low salt buffer to enable downstream analyses. The protein mixtures after co-affinity isolation were identified by peptide mass fingerprinting. The proteins were digested with trypsin, and LC–MS/MS was used to analyze the resulting peptide mixtures. The peptide masses were used to search an *in silico* peptide library that was generated based on a genome annotation of *H. volcanii* that was supplied by Friedhelm Pfeiffer (MPI of Biochemistry, Martinsried, Germany). Result lists of identified proteins were obtained, which were sorted by the parameter "peptide spectrum matches" (PSM). The PSMs are semi-quantitative approximations of the amounts of co-isolated proteins, which are influenced by the affinities between bait and co-isolated proteins and the intracellular concentrations of the proteins.

Supplementary Figure S1B gives an overview of the bioinformatic workflow that was used for the identification of proteins that were specifically co-purified with the 14 aIFs of *H. volcanii*. In short, all proteins not encoded in the genome of *H. volcanii* were removed, e.g., trypsin and contaminations with human proteins. Then all proteins were removed that had not at least two "unique peptide hits." Proteins that were identified in one or both of the two negative controls were also removed, including PitA and other proteins that can directly bind to the nickel chelating sepharose. Next, all proteins were removed that were not found in all three biological replicates. This workflow generated lists of trusted proteins that were specifically co-purified with aIFs. It should be noted that co-purification can be based on direct physical interactions, but can also be indirect based on bridging proteins or RNAs. An RNase step, which is sometimes included in the analyses of protein interaction networks, was deliberately not included in our co-affinity purification workflow. The project aimed at characterization of translation initiation networks, which require the presence of mRNAs as well as ribosomes. To facilitate further analyses, metabolic enzymes and hypothetical proteins

[2]https://scied.com/pr_cmbas.htm
[3]https://www.ebi.ac.uk/

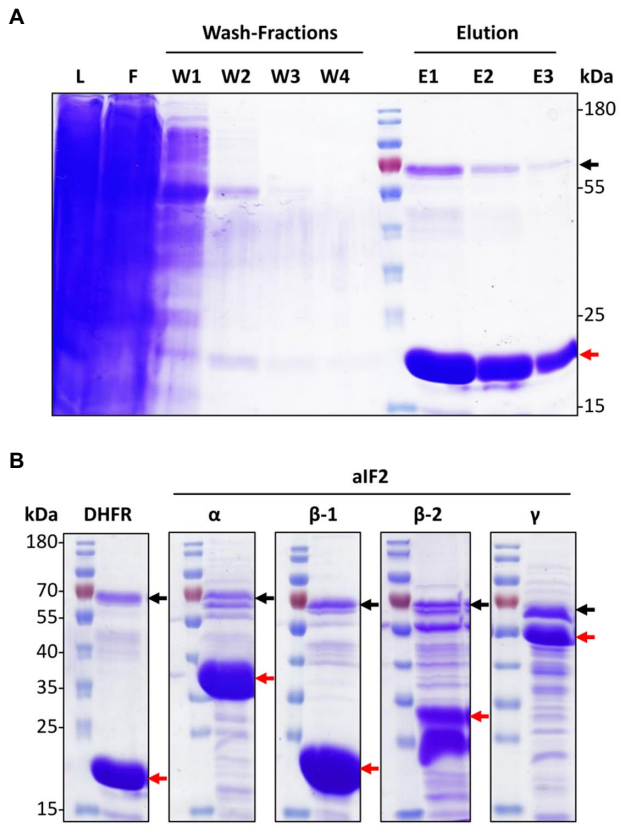


FIGURE 1 | Overproduction and co-affinity purification of N-His_6-tagged proteins. **(A)** SDS-Page representing different steps of the co-affinity purification of the negative control protein N-His_6 DHFR. The red arrow points to N-His_6 DHFR, the black arrow to an endogenous histidine-rich protein of *H. volcanii*, PitA. L – lysate, F – flow through, W – washing fractions, E – elution fractions. **(B)** SDS-Page of the elution fraction 1 from the co-affinity purifications of N-His_6-tagged DHFR and the four indicated aIF2 subunits. Black and red arrows point to PitA and the produced His_6-tagged respective aIF2 subunits. The figure shows one typical example of three biological replicates for each protein.

were also removed, and, thereby, the analyses were concentrated on proteins involved in the biological processes translation, transcription, replication and repair, RNA and protein turnover, and protein folding. In the following paragraphs, various aspects of the results from co-affinity purification with the 14 haloarchaeal aIFs are discussed.

The Ternary Initiation Factor aIF2

The heterotrimeric translation initiation factor aIF2 is of exceptional importance for translation initiation, because it brings the initiator tRNA to the P-site of the ribosome. All three subunits of the eukaryotic homolog eIF2 in yeast are essential. Unexpectedly, in *H. volcanii* only the γ-subunit is essential, while the deletion mutant missing the α-subunit is viable (Gäbel et al., 2013). *H. volcanii* contains two paralogous genes for the β-subunit, which can be individually deleted. However, a double mutant could not be obtained, and, thus, also the β-subunit is essential.

The three deletion mutants as well as the *dhfr*-deletion mutant were transformed with expression plasmids containing the respective genes, and the wild-type was transformed with an expression plasmid for the overproduction of the γ-subunit. As a control, all strains were also transformed with the empty vector. **Figure 2** shows growth curves of all plasmid-lacking and plasmid-containing strains. The plasmid-free cultures (dotted lines) were used to verify that the deletion mutants had the same phenotypes as previously reported (Gäbel et al., 2013). And indeed, deletion of *aIF2α* resulted in a very severe growth defect (**Figure 2B**), in contrast to the deletion of the genes for either of the two beta subunits (**Figures 2C,D**), and also deletion of *dhfr* did not compromise growth (**Figure 2A**). In all cases, the presence of the empty vector in the wildtype (black solid lines) led to a growth defect in comparison to the vector-free cultures (black dotted lines). Obviously, the Novobiocin resistance gene on the vector could not fully restore growth in Novobiocin-containing medium to the level in antibiotic-free medium. Therefore, growth characteristics of plasmid-free and plasmid-containing strains cannot be compared, and production strains have to be compared with the respective controls containing the empty vector. Overproduction of aIF2α and of aIF2β-1 did not increase (or decrease) the growth rate, but resulted in a slight increase in growth yield (**Figures 2B,C**). In contrast, overproduction of aIF2β-2 and of aIF2γ increased the growth yield considerably (**Figures 2C,E**). Because aIF2γ was overproduced in the wildtype, this indicates that the native concentration of aIF2γ might be rate-limiting.

Next, all four tagged-subunits were used for co-affinity purification, as described above. The patterns of co-isolated proteins are shown in **Figure 1B** in comparison to each other and the negative control producing DHFR. The following results were obtained: (1) a large number of proteins could be co-isolated with the aIF2 subunits, in contrast to the negative control, (2) the patterns of co-isolated proteins is different for the four aIF subunits, and (3) the number of co-isolated proteins is much higher for aIF2β-2 than for aIF2β-1. A higher importance of aIF2β-2 compared to aIF2β-1 is consistent with the growth analyses (**Figures 2C,D**) and with previous observations (Gäbel et al., 2013).

The co-isolated proteins were identified by peptide mass fingerprinting, as described above. Importantly, reciprocal co-isolation between the three subunits aIF2α, aIF2β-2, aIF2γ was observed, indicating that the three proteins form a heterotrimeric complex, as in eukaryotes and other archaea. The subunit aIF2β-1 was not co-isolated with any of the other subunits. However, with aIF2β-1 as bait, aIF2α and aIF2γ were co-isolated, indicating that two alternative heterotrimeric aIF2 complexes exist in *H. volcanii* (**Figure 3**).

Further proteins that could be co-isolated with one or more of the aIF2 subunits are listed in **Supplementary Table S3**. Translation initiation occurs at the ribosome, therefore, it was not surprising that 14 ribosomal proteins were co-isolated. However, aIF1 and aIF1A, which are part of the archaeal preinitiation complex together with aIF2 (Coureux et al., 2020), were not co-isolated. A translation initiation factor that could be co-isolated was the universally conserved aIF5B (eIF5B in

eukaryotes, and IF2 in bacteria). Another protein that could be co-isolated with aIF2α as well as with aIF2γ is HVO_0966, which is annotated as the enzyme R15BI. However, for reasons that are discussed below, we propose that in fact HVO_0966 is a translation initiation factor and should be renamed to aIF2Bα. In eukaryotes, eIF2B is a regulator and GDP/GTP exchange factor for eIF2.

Unexpectedly, also several proteins could be co-isolated with aIF2 that are involved in transcription initiation, transcription regulation, and DNA repair. The relevance of these results will be discussed in the **Discussion** section.

The aIF–aIF Interaction Network

Co-affinity isolations to identify interaction partners were performed with ten further translation initiation factors, in addition to the four aIF2 subunits. This allowed to unravel the aIF–aIF protein interaction network. The results are summarized in **Figure 4** and in **Supplementary Table S4**. The growth curves of all deletion mutants and production strains are shown in **Supplementary Figure S3**. Most interactions were observed only in one direction, however, in six cases co-isolation occurred in both directions, further underscoring their importance (red double arrows in **Figure 4**). Two hubs with many interactions are clearly visible in the aIF–aIF interaction network, i.e., aIF5B and aIF2Bα (HVO_0966). The universally conserved factor aIF5B could be co-isolated by 11 other aIFs, while aIF5B as bait led only to the co-isolation of aIF2Bα. At first sight this asymmetry seems to be unexpected, and reciprocal co-isolation should be expected whenever two proteins interact. However, notably, the co-isolation experiments with two proteins are not symmetrical, i.e., only the bait protein is overproduced, while the co-isolated interaction partner has its native intracellular concentration. In addition, only the bait protein carries an N-terminal tag, which might interfere with complex formation. For example, if the N-terminus of aIF5B would be important for the interaction with many other proteins, the N-terminal tag would inhibit co-isolation with aIF5B as bait, while the untagged aIF5B can easily be co-isolated with tagged other proteins as baits.

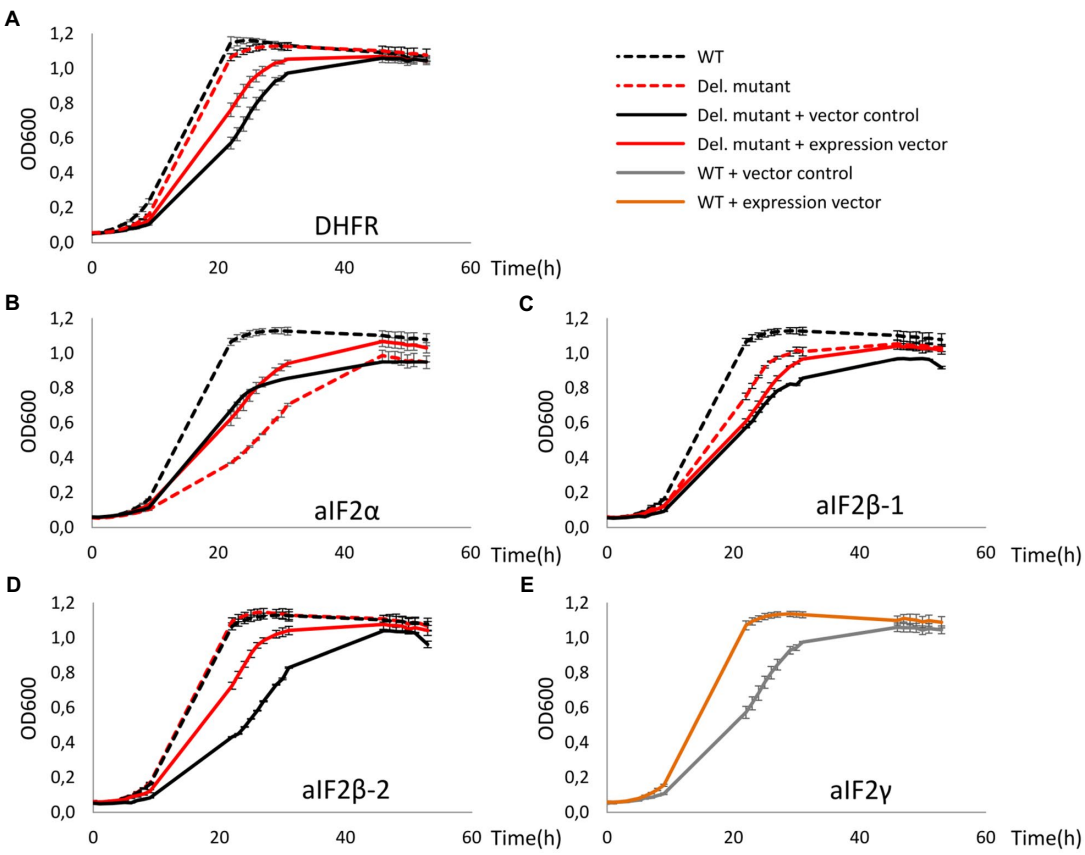


FIGURE 2 | Growth curves of deletion mutants of *dhfr* and genes for aIF2 subunits as well as of the respective overproduction strains. All cultures were grown in triplicates under optimal growth conditions in complex medium in 96-well plates. The OD600 was measured frequently, and average values and standard deviations are shown. **(A-D)** Growth curves of the wildtype are shown in black, growth curves of the deletion mutants are shown in red. Dotted lines – vector-free cultures (medium without antibiotic), solid lines – vector-containing cultures (medium with 0.5 μg/ml Novobiocin). The missing proteins in the respective deletion mutants are indicated at the bottom. **(E)** The essential aIF2 γ-subunit was overproduced in the wild type. Gray – wildtype with the empty vector, orange – wildtype with the expression plasmid.

The second major hub in the aIF–aIF interaction network is the protein that we would like to re-annotated to aIF2Bα (HVO_0966), albeit it is currently annotated as an enzyme. It was co-isolated with 10 different aIFs, while with aIF2Bα as bait co-isolated five aIFs, including two subunits of aIF2. This very central position of HVO_0966 in the aIF–aIF interaction network prompted us to study its phylogeny, with the aim to unravel its connections to other archaeal and eukaryotic initiation factors and enzymes (see below).

Two genes in the *H. volcanii* genome are annotated to encode putative subunits of aIF2B, i.e., aIF2Bδ (HVO_2706) and aIF2Bsu (HVO_1934). In eukaryotes, eIF2B is composed of two subcomplexes, the subunits eIF2Bα, eIF2Bβ, and eIF2Bδ form a regulatory subcomplex, while the subunits eIF2Bγ and eIF2Bε form a catalytic subcomplex (Bogorad et al., 2014). Genes for the catalytic subunits are not present in archaeal genomes, while often genes with similarities to the three regulatory subunits are present, so that the presence of a ternary complex in archaea has been postulated (Dev et al., 2009). However, our analysis of the aIF–aIF interaction network did not give strong support for the presence of such a ternary complex in *H. volcanii*. Reciprocal co-isolation was not observed between any of the three putative subunits, in contrast to the aIF2 ternary complex, and co-isolation between the putative aIF2Bδ and aIF2Bsu was not observed at all. In addition, in stark contrast to the many interactions of aIF2Bα, only one or two, respectively, interactions were found for the other two proteins. Taken together, these results argue against the existence of a heteromeric complex of the three proteins, at least in haloarchaea.

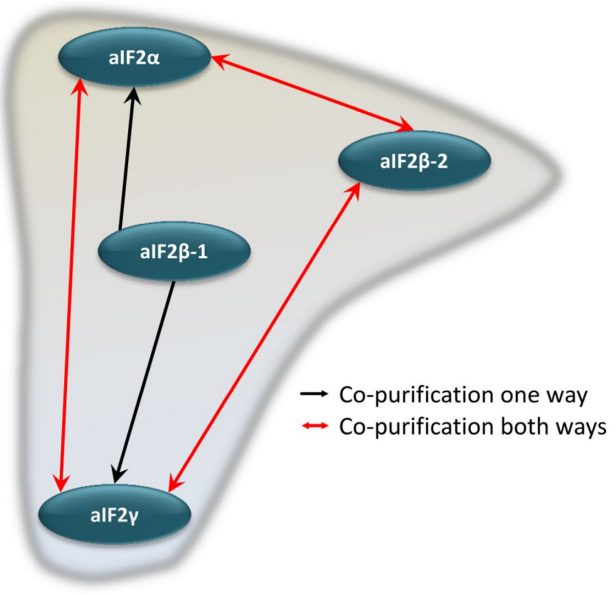


FIGURE 3 | Schematic overview of the internal aIF2 subunit interaction network. Red arrows denote reciprocal co-purification, while black arrows depict one-directional co-purification.

As discussed above, direct interactions between aIF1 and aIF1A and aIF2 were not found, however, aIF1 as well as both paralogs of aIF1A interact with the central hub protein aIF2Bα, which, in turn, interacts with aIF2. Therefore, an indirect interaction does exist within the aIF–aIF interaction network. In addition, a Cryo-EM structure of the preinitiation complex of *P. abyssi* revealed that the prominent interactions of aIF1, aIF1A, and aIF2 are formed with ribosomal proteins, and not with each other (Coureux et al., 2020).

The Interaction Network Between aIFs and the Ribosome

Translation initiation occurs at the ribosome, and, therefore, many interactions between translation initiation factors and ribosomal proteins can be expected. And indeed, a large number of ribosomal proteins could be co-isolated with the 14 (subunits of) translation initiation factors as bait proteins. The results are summarized in **Figure 5** and in **Supplementary Table S5**. The interaction with an aIF and the ribosome was defined as "extensive," when at least five ribosomal proteins were co-isolated (red arrows in **Figure 5**). Based on this definition, two proteins had extensive interactions with the large as well as with the small ribosomal subunit. One of the proteins is aIF2Bα, which underscores that aIF2Bα is a central hub in the translation initiation protein interaction network. In contrast, the results are very different for the two proteins that are annotated as putative aIF2B subunits, again not strengthening the idea that a ternary aIF2B complex might exist in haloarchaea.

The other protein with extensive interactions with both ribosomal subunits is aIF1. In contrast, one of the two aIF1A paralogs has only limited to medium interactions, while not a single ribosomal protein was co-isolated with the other paralog, aIF1A-2, as bait. In the structure of the preinitiation complex of *P. abyssi* the interactions of both aIF1 and aIF1A with the ribosome seemed to be similar (Coureux et al., 2016).

As mentioned above, many ribosomal proteins were co-isolated with two subunits of aIF2 as baits. The interactions of both proteins were much more extensive to the large subunit, which is unexpected, because aIF2 is thought to leave the preinitiation complex before the large subunits joins and the elongation phase begins. The universally conserved factor aIF5B, which was found to be a central hub in the aIF–aIF network, led only to a limited to medium co-isolation of ribosomal proteins. Possible implications of the observed aIF–ribosome interactions are considered in the **Discussion** section.

HVO_0966, an Important Translation Initiation Factor or a Metabolic Enzyme?

The results of the co-affinity isolation approach described above indicated that HVO_0966 is a central hub in the translation initiation protein interaction network. On the other hand, HVO_0966 is annotated as the enzyme R15BI. A BLAST search with HVO_0699 in the domain Archaea retrieved exclusively proteins that are also annotated as R15BI. In contrast, a BLAST search in the domain of Eukaryotes retrieved methylthioribose-1-phosphate isomerase as well as the alpha subunit of translation

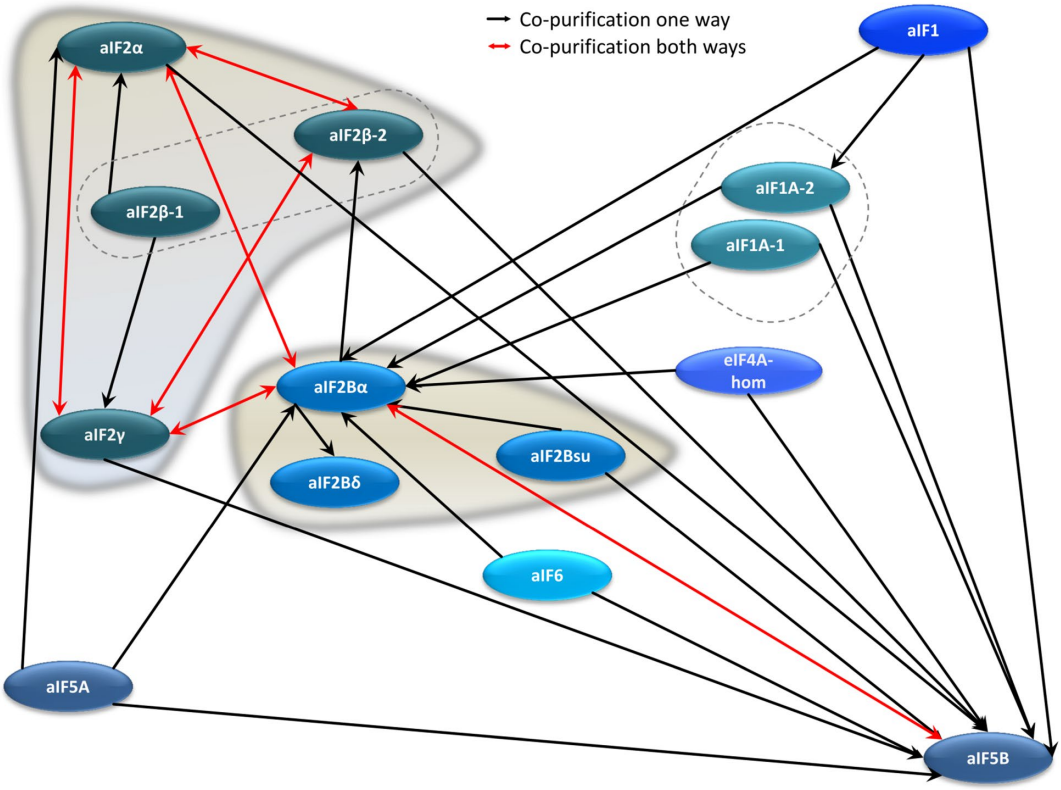


FIGURE 4 | Schematic overview of the aIF–aIF interaction network. Red arrows denote reciprocal co-purification, while black arrows depict one-directional co-purification.

initiation factor eIF2B. A BLAST search in the domain of Bacteria retrieved methylthioribose-1-phosphate isomerase, with very few exceptions.

The annotation of the archaeal proteins is based on a publication by Sato et al., who proposed that in *T. kodakarensis* three enzymes convert AMP to two molecules of 3-phosphoglycerate, which is part of the central metabolism (Sato et al., 2007). This gave a biological function to the third enzyme, ribulose-1.5-bisphosphate carboxylase/oxygenase (RuBisCO; TK2290), the role of which in archaea had been enigmatic before. The second enzyme is R15BI (TK0185), which isomerizes ribose to ribulose, and thus yields the substrate for RuBisCO. Before, the protein had been annotated to be an aIF2B subunit, and later this and all homologous archaeal proteins were re-annotated to be R15BIs. In a later publication, the enzymatic function of TK0185 was verified and the enzyme kinetic characteristics were characterized (Aono et al., 2012).

Haloferax volcanii contains all three genes, and, thus, could also use this pathway to funnel AMP into the central metabolism. In addition, the genes for the first two enzymes are adjacent (HVO_0965 and HVO_0966), and the third gene is close by (HVO_0970). Analysis of results from a RNA-Seq and a dRNA-Seq study (Babski et al., 2016; Laass et al., 2019) revealed that all three genes have independent promoters, and that the transcript level of HVO_0970 is much higher than that of HVO_0966 and HVO_0965. Nevertheless, the existence and close neighborhood of the three genes indicates that HVO_0966 might have the enzymatic function of a R15BI, like the homolog from *T. kodakarensis*. Therefore, the MS results were checked whether the two other enzymes of the AMP salvage pathway were co-isolated with HVO_0966, but this was not the case. However, it should be noted that a successful co-isolation of the two other enzymes with HVO_0966 would have been a strong indication for the existence of the AMP salvage pathway also in *H. volcanii*, but that the lack of co-isolation does not indicate its absence, because enzymes of metabolic pathways often do not form heteromeric complexes, but work as independent modules.

Taken together, strong arguments for both putative functions of HVO_0966 exist. Therefore, we decided to generate phylogenetic trees to analyze the evolution of HVO_0966 and its archaeal and eukaryotic homologs, with the aim to clarify whether HVO_0966 is more likely to be a haloarchaeal translation initiation factor or an enzyme. The three regulatory subunits of the eukaryotic eIF2B are paralogs, and they are homologs to the enzyme MTPI. Therefore, the sequences of eIF2Bα, eIF2Bβ, eIF2Bδ and MTPI from humans and one representative of rodents, Arthropoda, plants and fungi were retrieved from protein sequence databases. In addition, BLAST searches were performed with HVO_0966 as well

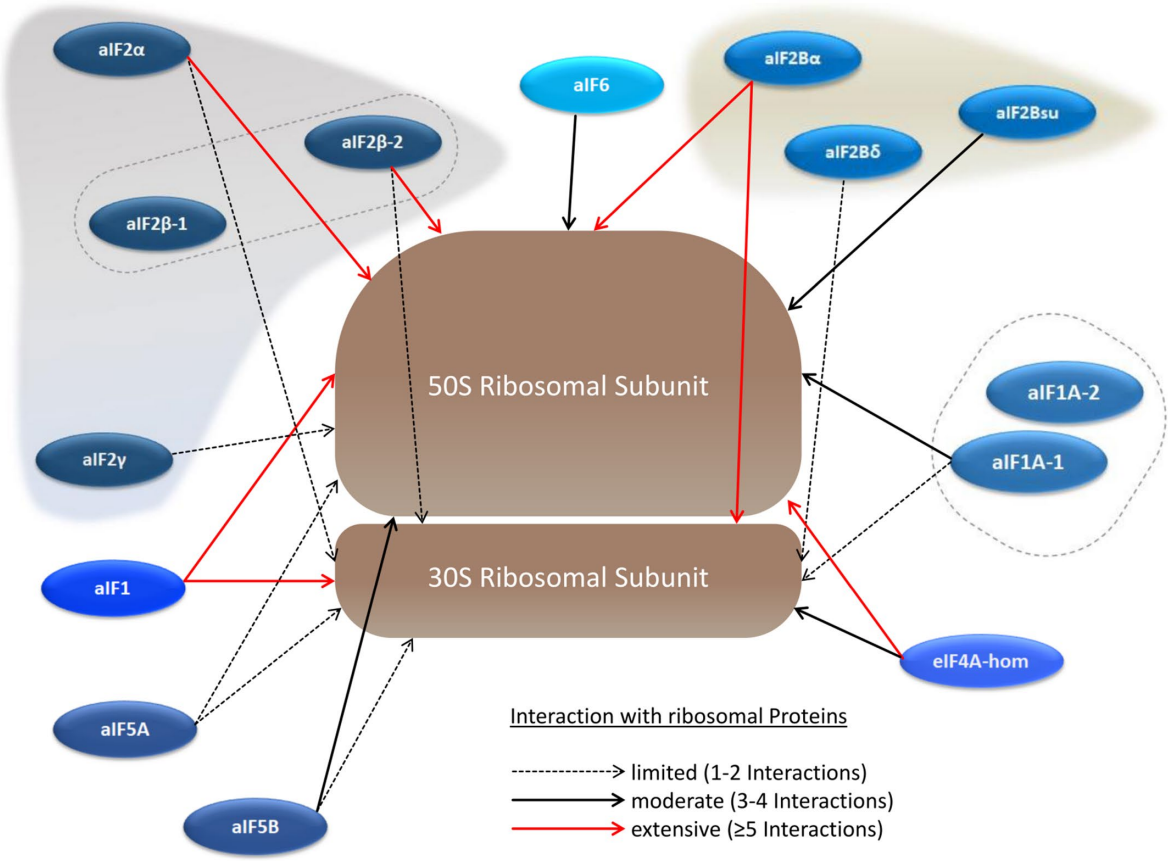


FIGURE 5 | Schematic overview of the aIF–ribosome interaction network. The large and the small ribosomal subunits are shown schematically instead of including all ribosomal proteins individually. The numbers of co-purified ribosomal proteins are indicated by red arrows (at least 5 co-purified proteins), solid black arrows (3–4 co-purified proteins), and dotted black arrows (1–2 co-purified proteins).

as with the four human proteins in the domain of Archaea, and the most similar archaeal sequences were retrieved irrespective of their annotation. In total, 35 sequences were used to generate a MSA. The MSA was manually edited to remove regions without phylogenetic information, e.g., non-conserved long N-termini of the β and δ subunits of eIF2B and insertions in single sequences. The resulting MSA was used to generate phylogenetic trees with MEGA X (Kumar et al., 2018). Three different approaches were used, i.e., Maximum Likelihood (ML), Neighbor Joining (NJ), and Maximal Parsimony (MP). In all three cases 1,000 bootstrap replications were performed to estimate the confidence of different parts of the tree.

The ML tree is shown in **Figure 6**, the NJ tree is shown in **Supplementary Figure S4**, and the MP tree is shown in **Supplementary Figure S5**. While the three trees are not identical, the following major results are the same for all three approaches: (1) the three eukaryotic regulatory eIF2B subunits form monophyletic groups with very high bootstrap support (nodes A-C in **Figure 6**). The only exception is one protein from fungi, which was retrieved with a BLAST search with the human eIF2Bβ, but is found in the δ subtree. (2) Also the five MTPIs form one monophyletic group with a very high bootstrap value (node D). (3) Three archaeal proteins group together with the eukaryotic MTPIs (node E). All three are annotated as homologs to eIF2B subunits, but might well be enzymes based on their phylogenetic relationship to MTPIs. (4) Eight archaeal sequences form one phylogenetic group with high bootstrap support (node F). It is tempting to speculate that all members of this group have the same function, in spite of the very mixed annotations. Clearly, the annotations of the archaeal proteins are as yet unresolved and do not help to understand their function. This phylogenetic group includes HVO_0966, which seems to be a central hub of the translation initiation network based on the results presented above, as well as protein TK0185 from *T. kodakarensis*, which was shown to be a R12BI. (5) The archaeal group is between the eukaryotic eIF2B groups and the eukaryotic MTPI group. All deep nodes (G, H, I, J) have very low bootstrap supports, therefore, the phylogenetic analysis does not help to decide whether HVO_0966 and the other members of the archaeal group are more likely to be enzymes or translation initiation factors. In the **Discussion**, we will argue that it is possible that

members of this protein family could have both functions and are moonlighting proteins. (6) The other two *H. volcanii* proteins that are annotated to be putative eIF2B subunit homologs (HVO_1934, HVO_2706) cluster together and far from HVO_0966. Their position in the tree depends on the algorithm used and cannot be clarified.

The Interaction Network of aIFs With Translation and Transcription Proteins

A full list of co-isolated proteins (after filtering, see **Materials and Methods**) is shown in **Supplementary Table S6**. In addition to the aIFs and the ribosomal proteins discussed above, many more proteins could be co-isolated. These included eight subunits of the RNA-polymerase (Rpo), two TATA box-binding proteins, and 27 transcription factors, indicating that translation initiation and transcription are not independent in *H. volcanii*. In bacteria, coupling of transcription and translation has been discussed for a long time, and, very recently, a direct interaction between RNA-polymerase and the ribosome has been reported (see **Discussion**). One experimental study exists, which indicated that coupling of transcription and translation might also occur in archaea (French et al., 2007). Electron microscopy was used to show that polysomes are close to DNA in lysed cells of *T. kodakarensis*. In this Special Issue of Frontiers in Microbiology, Weixlbaumer et al. propose that transcription and translation in archaea might be coupled, based on bioinformatic comparisons of bacterial and archaeal proteins (Weixlbaumer et al., 2021). To gain further insight into a putative coupling of transcription and translation in haloarchaea, we decided to study the interaction between the RNA polymerase and translation initiation factors further.

The Interaction Network of RNA Polymerase Subunits

The genes for seven subunits of the RNA polymerase were cloned into an expression vector, and the proteins were produced with an N-terminal hexahistidine tag (**Table 2**). All subunits of the catalytic core were included (RpoA1, RpoA2, RpoB1, RpoB2), two subunits of the assembly platform (RpoD, RpoL), and one auxiliary subunit (RpoH). Subunits of the RNA polymerase can be assumed to be essential, therefore, it was not attempted to generate deletion mutants, but the tagged proteins were produced in the wildtype, in addition to the native proteins encoded on the chromosome. In all seven cases co-isolation of other proteins turned out to be possible, showing that the tagged proteins were incorporated into the multi-subunit RNA polymerase complex. The interaction network between RNA polymerase subunits is shown in **Figure 7** and in **Supplementary Table S7**. In nearly all cases, reciprocal co-isolation of the seven bait subunits was successful. Only the affinity isolation of RpoD did not result in the co-isolation of RpoB1 and RpoL, which can readily be explained by the very low production level of RpoD, in contrast to the other six subunits (see **Supplementary Figure S6A**). The effect of overexpression on the growth of *H.* volcanii is shown in **Supplementary Figure S6B**. Three further RNA polymerase subunits could be co-isolated with one or several of the bait subunits, i.e., the two assembly platform subunits RpoN and RpoP were co-isolated with six and five subunits, respectively, and the auxiliary subunit RpoF was co-isolated with RpoH. The only two subunits that were not co-isolated with any of the seven bait subunits were the two auxiliary subunits RpoE and RpoK. Possible reasons include loss of these two auxiliary subunits during the affinity purification procedure, which includes intensive washing before elution, and/or failure to detect the co-isolated subunits *via* peptide mass fingerprinting (RpoK is very small with only 58 amino acids). In any case, the results showed that the seven tagged bait proteins were integrated into the complex and enabled the affinity isolation of a (nearly) complete RNA polymerase, which should allow to unravel the RNA polymerase interaction network.

The Interaction Network Between RNA Polymerase Subunits and aIFs

Co-affinity isolations with 14 aIFs as bait proteins and seven Rpos as bait proteins enabled to gain a comprehensive view of the interconnections between RNA polymerase and translation initiation. The results are shown in **Figure 8** and **Supplementary Tables S8 and S9**. Obviously, two interaction hubs exist, which are characterized by a high number of interactions and reciprocal interactions, i.e., aIF2Bα has eight interactions with Rpos, six of which are reciprocal, and aIF5B has six interactions, two of which are reciprocal. No other aIF shows a reciprocal interaction with a Rpo subunit. Only two additional aIFs are co-isolated with a Rpo as bait, namely aIF2α with RpoH and aIF2Bsu with RopB2. As discussed above, eight different aIFs as baits led to the co-isolation of one or more RNA polymerase subunits (**Supplementary Table S9**). Two possible reasons for an only uni-directional co-isolation have been discussed above, for an essential multi-subunit complex like the RNA polymerase two additional reasons apply: (1) The overproduction of one subunit does not lead to the overproduction of the whole complex, because all other subunits are encoded on the chromosome and produced under the control of the native promoters, and (2) only a subpopulation of complexes carries the tag and can be co-purified, which results in a lower concentration during the isolation and washing steps. In any case, the high number of (reciprocal) co-purifications between RNA polymerase subunits and the initiation factors aIF5B and aIF2Bα underscores that transcription and translation initiation are not independent processes in haloarchaea.

Interactions of RNA Polymerase Subunits With Further Translation and Transcription Proteins

Many additional proteins could be co-isolated with RNA polymerase subunits as bait proteins, in addition to the four

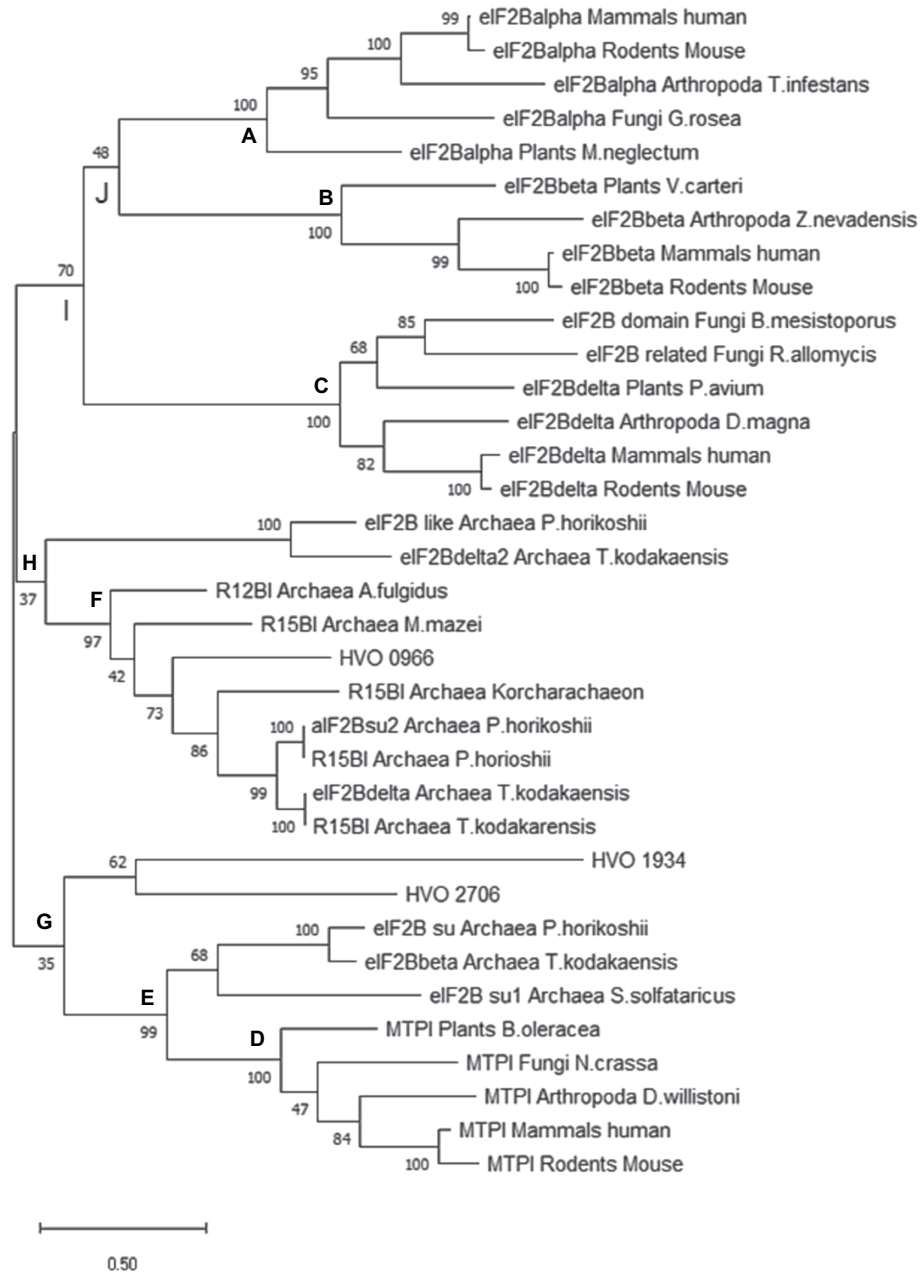


FIGURE 6 | Phylogenetic tree of selected archaeal and eukaryotic proteins of the a/eIF2B superfamily. Criteria for selection of the 35 proteins are explained in the text. A maximum likelihood (ML) tree was calculated using the program Mega X (Kumar et al., 2018). 1,000 bootstrap replications were performed, and the bootstrap values are shown at each node (%). Specific nodes of interest were labeled with A to G to facilitate the discussion (see text).

aIFs (**Supplementary Table S10**). Notably, this includes 21 ribosomal proteins, which – again – underscores the interdependence of transcription and translation in *H. volcanii*.

RNA polymerase subunits as baits also led to the co-isolation of one TATA box binding protein and of 21 transcription factors (TFs). Of these 21 TFs, 19 were also co-isolated with aIFs as baits, underlining that these are real positives and that they are present in the complexes that also contain aIFs and Rpos (**Supplementary Table S10**).

DISCUSSION

Elucidation of Protein–Protein Interaction Networks *via* Co-affinity Isolation of Complexes

Most if not all biological processes rely on the interaction of biomolecules, e.g., proteins and nucleic acids. Therefore, experimental approaches to unravel single interactions or complete interaction networks are of utmost importance. Different

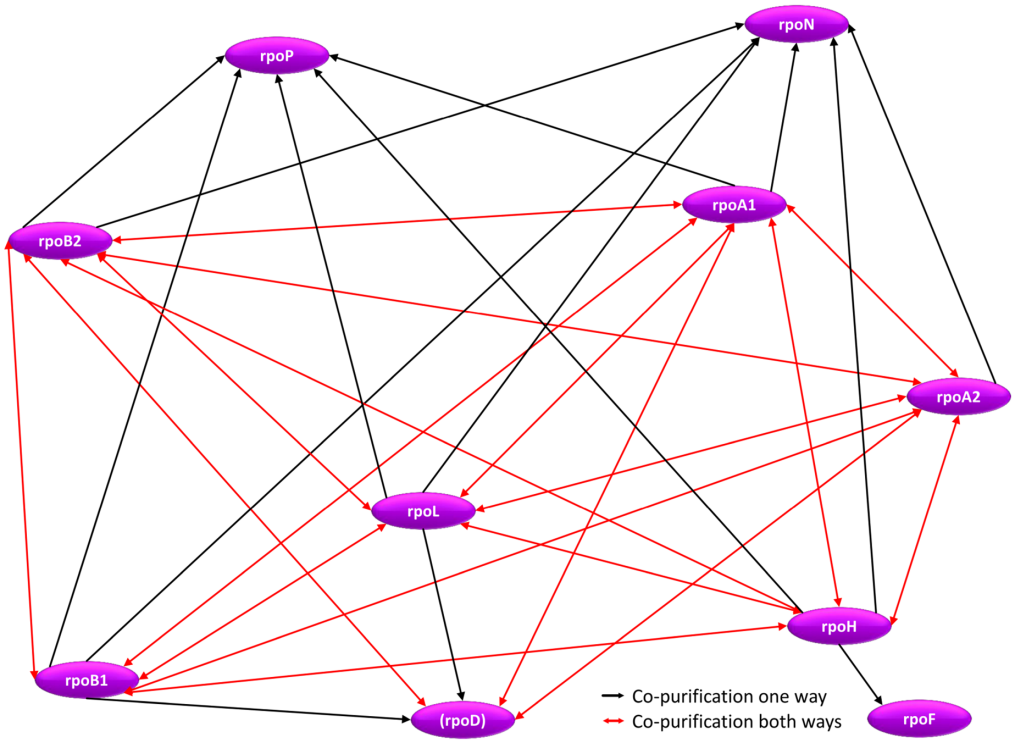


FIGURE 7 | Schematic overview of the internal RNA polymerase subunit interaction network. Red arrows denote reciprocal co-purification, while black arrows depict one-directional co-purification.

aspects of experimental approaches to characterize protein–protein interaction networks have been reviewed recently (Koh et al., 2012; Yang et al., 2015).

Co-affinity isolation of protein complexes using bait proteins followed by the identification of co-isolated binding partners has been applied very extensively using model species of bacteria and eukaryotes. However, also a few projects have been reported that unraveled protein–protein interaction networks in archaea. For example, the DNA replication protein network of *T. kodakarensis* was analyzed using 19 hexahistidine-tagged replication proteins as baits (Li et al., 2010). In another study, the protein interaction network of the taxis signal transduction pathway of *Halobacterium salinarum* was studied using 18 tagged bait proteins (Schlesner et al., 2012). A third study analyzed the protein–protein network of genomic maintenance in *P. abyssi* with 22 tagged bait proteins (Pluchon et al., 2013). Here, we report the fourth study that used co-affinity isolation of protein complexes with tagged bait proteins coupled to MS analysis to unravel protein interaction networks in archaea.

Initially, three different tags were compared, which were fused to the DHFR, i.e., the hexahistidine tag (His_6), the CBD tag, and the streptavidin tag. The His_6 tag turned out to be highly superior to the other two tags, and consequently, it was used for all proteins analyzed in this study (Schramm and Soppa, unpublished results). In total, 14 aIFs, seven Rpo subunits, and the metabolic enzyme DHFR as negative control were tagged, affinity isolated, and co-isolated binding partners were identified. Three biological replicates were performed to guarantee reproducibility and minimize false-positive hits. Taken together, we present here one of only extremely few studies that unraveled a protein–protein interaction network in an archaeon, and with 22 analyzed proteins its size is second to none of the three previous studies, which concentrated on other biological processes.

The bait proteins were overproduced, as in the three other studies with archaea mentioned above and many studies with bacteria and eukaryotes. Notably, this will lead to the formation of non-native complexes with additional proteins, which have lower affinities to the bait proteins than the native interaction partners (false positives). However, cell disruption leads to a very large dilution of the cytoplasm, so that low affinity complexes will dissociate again, and only native high-affinity complexes will remain. In addition, we applied intensive washing to guarantee that all proteins that bind non-specifically to the bait proteins or to the nickel-chelating column were removed. If we had expressed the bait proteins at their native levels, the high dilution and stringent washing could have led to the dissociation of native protein complexes of intermediate affinity (false negatives). Taken together, we think that the experimental design of bait overproduction and omission of a crosslinking step has a high probability to keep the false-positive as well as the false-negative rate low.

It should be noted that the observed interactions are not restricted to direct physical binding partners, but also includes

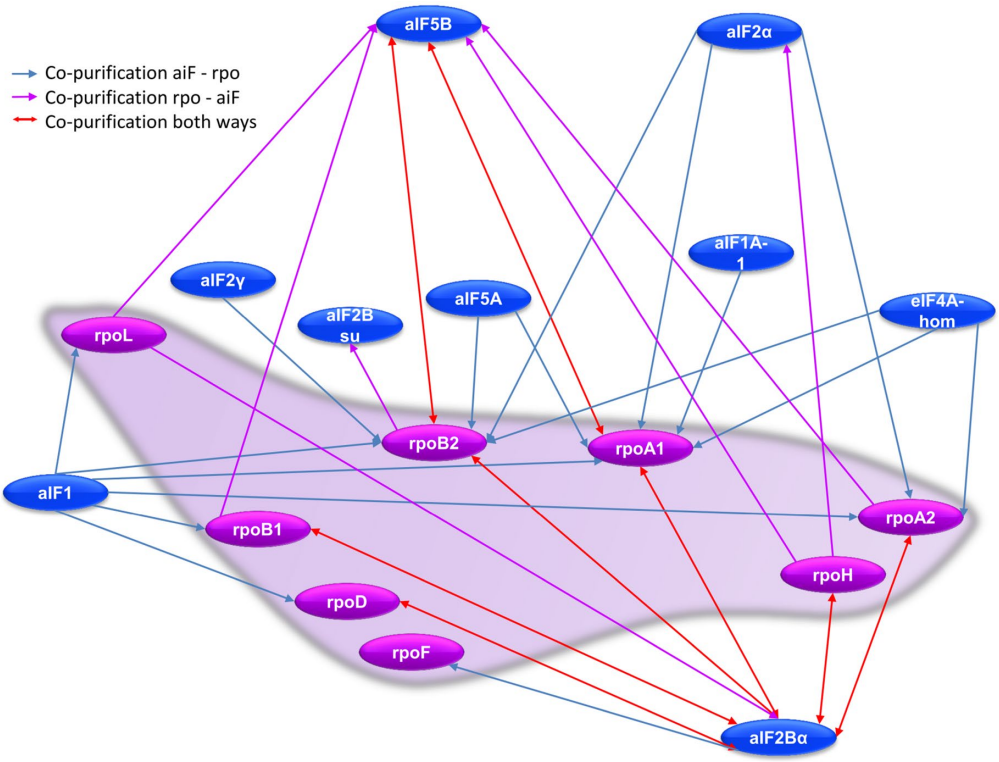


FIGURE 8 | Schematic overview of the RNA polymerase subunit–aIF interaction network. Red arrows denote reciprocal co-purification, blue arrows indicate co-purification of a Rpo subunit with a tagged aIF bait, while purple arrows indicated co-purification of an aIF with a tagged Rpo subunit as bait.

indirect interactions. For example, if a heteromeric complex exists, it can be expected that many or all subunits can be co-isolated, even those that do not directly physically interact with the bait protein. In the present case, translation initiation involves the mRNA as an essential constituent, and, therefore, also indirect protein–mRNA–protein interactions occur. We have deliberately not included an RNase step in our experimental design, which would have destroyed this form of indirect interaction and would have focused the analysis on protein–protein complexes alone, because the aim was to get a comprehensive overview of the protein–protein interaction network, even if some of the interactions might be indirect and RNA mediated. A parallel analysis of the putative RNA components of the affinity isolated complexes was beyond the scope of this project.

Characterization of Translation Initiation in Archaea

As mentioned in the **Introduction**, in recent years considerable progress has been obtained in the analysis of translation initiation in archaea. A very recent review gives an excellent overview of the progress (Schmitt et al., 2020). Breakthrough results were the structure determinations of preinitiation complexes from *P. abyssi* (Coureux et al., 2016, 2020). The complexes contained the ribosome, a short RNA, the initiation factors aIF1 and aIFA, and the ternary complex of the central factor aIF2 with the initiator tRNA and a GTP analog. Importantly for this study, the structures show the interaction of aIF1, aIF1A, and aIF2 of *P. abyssi* with one another and with the ribosome.

By far the highest number of studies on archaeal translation initiation have been performed with *Sulfolobus solfataricus*, which belongs to the kingdom of Crenarchaeota. Notably, it was found that the central heterotrimeric factor aIF2 in *Sulfolobus* does not only fulfill the homologous function of eIF2 to bring the initiator tRNA to the ribosome, but that the subunit aIF2γ has an additional role as stand-alone protein. It can bind to the 5'-end of transcripts and thereby it stabilizes transcripts and shields them from exonucleolytic degradation (Hasenöhrl et al., 2008). These results underscore the necessity to characterize archaeal translation initiation factors. Even if they are homologous to the eukaryotic factors, the two domains are separated in evolution by more than one billion of years, and the archaeal factors might have fewer or additional functions, different functions, or might even not be involved in translation initiation at all, despite their primary sequence similarities to eIFs.

In addition, the biological function of annotated aIFs might be different in different groups of archaea. The various phylogenetic groups of archaea are separated by billions of years, and it might well be that the functions of homologous proteins have evolved differently in different groups. This would be similar to bacteria, for which biodiversity in the mechanisms

of translation initiation have become obvious (Malys and McCarthy, 2011).

The Central Initiation Factor aIF2

The factor aIF2/eIF2 is present in archaea and eukaryotes, but it is not present in bacteria. It is a central factor because it guides the initiator tRNA to the P-site of the ribosome. In accordance with the crucial role of eIF2, all three subunits are essential in eukaryotes. Very unexpectedly, it was found that in *H. volcanii* only two subunits are essential (aIF2β and aIF2γ), while the gene for aIF2α could be deleted (Gäbel et al., 2013). This result indicated that either the complex is not heterotrimeric in haloarchaea, or that the aIF2βγ dimer lacking the native alpha subunit retains a residual function in translation initiation. In the present study, reciprocal co-isolation of all three subunits aIFα, aIFβ-2 and aIFγ was observed (**Figure 3**), showing that aIF2 is a heterotrimer also in *H. volcanii*, like in other archaea and eIF2 in eukaryotes. aIF2γ is the central and largest subunit (44.0 kDa), while aIF2α and aIF2β-2 are somewhat smaller (29.5 kDa, 22.2 kDa). The structure of the preinitiation complex of *P. abyssi* revealed that aIF2γ is tightly bound to the 30S ribosomal subunit, while ribosomal binding of aIF2α and aIF2β is more loose (Schmitt et al., 2019). Therefore, it seems feasible that an aIF2βγ dimer retains the capability of binding the initiator tRNA, GTP, and the 30S subunit. In accordance with this view, in yeast it was shown that the alpha subunit contributed only slightly to the binding affinity of the initiator tRNA (Nika et al., 2001). However, initiator tRNA binding was found to be different in aIF2 from *S. solfataricus* and *P. abyssi* (Yatime et al., 2004; Schmitt et al., 2012; Naveau et al., 2013). In these archaeal systems the alpha subunit contributed to the tRNA binding, while the beta subunit had only a minor role. The contribution of the alpha and beta subunits to the tRNA binding of the haloarchaeal aIF2 is unclear. In any case, the alpha subunit is very important for the full function of aIF2 also in *H. volcanii*, because the deletion mutant has a severe growth defect under all tested conditions (Gäbel et al., 2013).

Haloferax volcanii contains two paralogs of the beta subunit, aIF2β-1 and aIF2β-2, in contrast to other archaea. They can functionally replace one another to some extent, because single deletion mutants of both genes could be constructed, while generation of a double deletion mutant was not possible (Gäbel et al., 2013). However, they are not equivalent. For example, a much higher number of proteins could be co-purified with aIF2β-2 than with aIF2β-1 (**Figure 1B**). In addition, only aIF2β-2 could be co-purified with both other subunits, not aIF2β-1 (**Figure 3**). However, aIF2β-1 could co-purify both other subunits, indicating that it is also part of a heterotrimeric aIF2 complex. The specific differential roles of the two aIF2β subunits remain to be determined.

aIF2β-1 and aIF2β-2 share an N-terminal region of about 130 amino acids (aa), and aIF2β-2 has an additional C-terminal domain of about 70 aa that is not present in aIF2β-1. A BLAST search with these extra 70 aa revealed that aIF2β-2 is widely distributed in halophilic and methanogenic archaea. However, it is also confined to these archaeal groups, so that the most plausible explanation is a gene duplication of an ancient version of aIF2 in the common ancestor of halophilic and methanogenic archaea, which was followed by the addition of 70 extra aa to only one of the two copies. The 70 aa have limited similarities to small proteins that are annotated as TRAM domain proteins. About 20 years ago a bioinformatic analysis identified a conserved domain that occurred in two families of tRNA modifying enzymes, other proteins associated with translation, and a family of small, uncharacterized archaeal proteins and proposed the acronym TRAM (Anantharaman et al., 2001). Recently, it was shown that single domain small TRAM proteins in two psychrophilic methanogenic archaea have RNA-binding activity and are cold-shock proteins (Taha et al., 2016; Zhang et al., 2017). Deletion of a gene for a small TRAM domain protein in *Methanococcus maripaludis* reduced the growth rate and altered the levels of 55% of all transcripts (Li et al., 2019). Many 5'-UTRs were identified as potential targets of this protein, and three representative 5'-UTRs were unfolded by this TRAM protein *in vitro*. Taken together, it seems that after gene duplication one copy of the primordial aIF2β was fused with a small protein that added additional RNA-binding and RNA chaperone function and had a preference for 5'-UTRs.

To get a further insight into the evolution of a/eIF2β, a MSA was generated containing both paralogs from *H. volcanii* and selected proteins from haloarchaea, other archaea, and eukaryotes (**Supplementary Figure S7**). Obviously, the primordial protein had a size of about 130 aa and contained a CXXC and a CXXCG motif in its C-terminus. The four cysteines (red in **Supplementary Figure S7**) are highly conserved. They were shown to bind zinc (Gutiérrez et al., 2002), and the structures of several of these zinc fingers from several archaeal proteins have been solved (Gutiérrez et al., 2004; Nikonov et al., 2021). Later in evolution, eukaryotes have added an additional N-terminal domain of about 150 aa, which contains three poly-lysine stretches (blue in **Supplementary Figure S7**). These poly-lysine stretches are involved in mRNA-binding and the presence of at least on stretch is essential for function (Laurino et al., 1999; Salton et al., 2017). Taken together, a/eIF2β is an excellent example for protein evolution, in which the primordial protein of about 130 aa has been optimized by the addition of further RNA-binding domains either at the N-terminus (eukaryotes) or the C-terminus (halophilic and methanogenic archaea). This added functionality nicely explains the higher importance of aIF2β-2 compared to aIF2β-1 that was revealed in the present study.

In eukaryotes, eIF2α can be phosphorylated at a highly conserved serine residue. This is a key event in the integrated stress response of many or all eukaryotes, and it results in a downregulation of translation initiation (Kashiwagi et al., 2019; Marintchev and Ito, 2020). Phosphorylation of Ser51 (yeast numbering) of eIF2α inhibits the interaction of eIF2 with eIF2B, and, therefore, the GDP–GTP exchange is blocked and translation initiation stops (Gordiyenko et al., 2019). There are strong arguments that this regulatory step is not conserved in archaea, mainly (1) they do not contain homologs to the

catalytic subunits of eIF2B, (2) there is no serine at the position homologous to yeast serine 51, and (3) the *Sulfolobus* aIF2 does not need an auxiliary factor for GDP/GTP exchange (Pedullà et al., 2005). In contrast to these arguments, it has been shown that aIF2α from *P. horikoshii* can be phosphorylated at a serine *in vitro* (Tahara et al., 2004). In archaeal aIF2 a serine is highly conserved at a position that is adjacent to the highly conserved serine in eukaryotes. We have exchanged this serine (serine 46 in *H. volcanii*) against alanine and aspartate to mimic the non-phosphorylated and the phosphorylated state (Schramm and Soppa, unpublished results). Both mutants grew nearly identical to the wildtype. In addition, the MS results after affinity isolation of aIF2α were searched, but a phosphorylated peptide could not be detected. These results indicate that regulation of translation initiation *via* phosphorylation of the conserved serine 46 of aIF2α does not occur in haloarchaea, in contrast to differential phosphorylation of the conserved serine 51 in eIF2α.

The aIF Interaction Network With aIFs and With Other Proteins

In the present study, the protein–protein interaction network of translation initiation in *H. volcanii* could be unraveled (**Figures 3–5** and **Supplementary Tables S3–S6**). A high number of interactions between aIFs as well as between aIFs and the ribosome could be detected. These results indicate that the bioinformatic annotation, which is based on protein sequence similarity to eukaryotic translation initiation factors, is correct, and that the aIFs are indeed involved in translation initiation in *H. volcanii*. Unexpectedly, also subunits of the RNA polymerase and other transcription proteins could be co-isolated with aIFs, indicating that translation and transcription are not independent in *H. volcanii* (discussed below). Two interaction hubs were identified in the aIF–aIF and the aIF–Rpo interaction networks, i.e., aIF5B and HVO_0966, which we propose to rename to aIF2Bα for reasons that are discussed in the next paragraph. The high importance of aIF5B was not unexpected, because it is one of only two universally conserved translation initiation factors (IF2 in bacteria and eIF5B in eukaryotes). It is involved in later stages of translation initiation and promotes binding of the large ribosomal subunit (Schmitt et al., 2020). Experimental evidence for this function also in archaea has been obtained with aIF5B from *Aeropyrum pernix* (Murakami et al., 2018). It should be noted that aIF5B binds to the initiator tRNA, like aIF2. Therefore, the co-purification strategy without an RNase step does not only include bait protein-mRNA-protein complexes (as discussed above), but might also include bait protein-tRNA-protein complexes.

Remarkably, the results with the second universally conserved initiation factor were totally different (aIF1A in archaea, eIF1A in eukaryotes, IF1 in bacteria). *H. volcanii* contains two paralogs, in contrast to other archaea, but none of them exhibited a high number of interactions to other aIFs or the ribosome. This is totally unexpected, because aIF1A is the third initiation factor that is part of the preinitiation complex of *P. abyssi*, in addition to aIF2 and aIF1 (Schmitt et al., 2019). It seems that either the affinities of the haloarchaeal aIF1As to the ribosome are not very high, or that the N-terminal hexahistidine tag interfered with complex formation. In any case, the two paralogs of aIF1A fulfill an essential role in *H. volcanii*, because it turned out to be impossible to delete both genes simultaneously (Gäbel et al., 2013).

The Case of HVO_0966: aIF2Bα or Metabolic Enzyme

HVO_0966 turned out to be a very special case. It was included into the study, because it was annotated as a subunit of aIF2B. Later, the annotation was changed to the metabolic enzyme R15BI. There is experimental evidence that HVO_0966 homologs from *T. kodakarensis* and *P. horikoshii* function as R15BIs (Aono et al., 2012; Gogoi and Kanaujia, 2018). An enzyme kinetic characterization of the *T. kodakarensis* protein was performed, and a K_m value of 0.6 mM for R15B was determined. The crystal structure of the *P. horikoshii* protein was determined, and R15B binding was verified. The enzyme R15BI is part of a three enzyme pathway that converts AMP into two molecules of 3-phosphoglycerate, and, thereby, funnels AMP into the central catabolic metabolism (Sato et al., 2007). The other two enzymes are AMP phosphorylase and ribulose-1,5-bisphosphate carboxylase (RuBisCO). Before this AMP salvage pathway was detected, the presence of RuBisCO in archaea had been an enigma, because RuBisCO is a central enzyme for CO_2 fixation in the Calvin cycle, which is not present in archaea.

This experimental proof that the two homologs of HVO_0966 are metabolic enzymes strongly suggests that the haloarchaeal protein has the same function. On the other hand, our finding that HVO_0966 is one of only two interaction hubs in the translation initiation network and that it has a very high number of interactions with other aIFs, with the ribosome, and with Rpo subunits suggest evenly strongly that it functions as a translation initiation factor, and we propose to rename it to aIF2Bα. Additional experimental evidence in this direction is that HVO_0966 homologs from *P. horikoshii*, *P. furiosus*, and *T. acidophilum* bind to the cognate aIF2α as well as to the eIF2α from yeast (Dev et al., 2009). The structure of a protein annotated as aIF2Bα was solved and was used to model the structure of the eukaryotic regulatory subcomplex comprised of eIF2Bα, eIF2Bβ, and eIF2Bαδ (Kakuta et al., 2004). Taken together, strong experimental evidence in both directions has been provided.

The eukaryotic homologs of HVO_0966 form four subfamilies, three are comprised of the three regulatory subunits of eIF2B (eIF2Bα, eIF2Bβ, and eIF2Bδ), and the fourth subfamily consists of eukaryotic 5-methylthioribose-1-phosphate isomerase (MTPI). The phylogenetic analyses of selected archaeal and eukaryotic proteins confirmed the existence of these four eukaryotic subfamilies with very high bootstrap values (**Figure 6**, nodes A, B, C, D, and **Supplementary Figures S4, 45**). In contrast, the archaeal proteins did not form one monophyletic group. A few proteins group with the eukaryotic MTPI subfamily (node E), and indeed, it has been reported that the protein from *P. horikoshii* has this enzymatic activity (Gogoi and Kanaujia, 2018). The

majority of the remaining archaeal proteins form one well-supported subfamily (node F). The annotation is very mixed; however, the group includes the two proteins that were shown to have the R15BI activity as well as HVO_0966 from *H. volcanii*, which is an interaction hub in the translation initiation network. Two explanations for this contradiction seem possible, (1) the archaeal proteins have evolved into the two different functions in the long time the species were separated, or (2) the archaeal proteins have both functions and are in fact moonlighting proteins.

So-called moonlighting proteins can fulfill at least two very different functions in different biological processes. A classic example is enolase, which is part of the glycolytic pathway, but which is also involved in RNA degradation (Henderson and Martin, 2013). However, the number of known moonlighting proteins has increased drastically in recent years (Jeffery, 2020; Liu and Jeffery, 2020; Singh and Bhalla, 2020; Beaufay et al., 2021; Rodríguez-Saavedra et al., 2021; Turek and Irving, 2021). An indication that the archaeal proteins might have two different functions is the recent observation that the eukaryotic eIF2Bα still has the ability to bind sugar phosphates with high affinity, e.g., fructose-6-phosphate has a K_D of 9.4 μM (Hao et al., 2021). The binding of sugar phosphate enhances the formation of the complete eIF2B decamer. The authors speculated that the high-affinity sugar binding couples the nutrient status of the cell to the translation rate. This would lead to an automatic downregulation of translation, a very costly process, when nutrients become scarce.

It could be envisaged that the archaeal proteins of the subfamily node F represent a primordial state, in which sugar metabolism is coupled to translation regulation not only *via* the sole binding of sugar phosphates, but in which they have a second role as enzymes. A first example of a moonlighting translation initiation factor has been reported, because aIF5A from *S. solfataricus* exhibits also the enzymatic function of a ribonuclease (Bassani et al., 2019). Future experiments are needed to unravel whether the archaeal enzymes indeed has two different functions. Until then we propose to rename HVO_0966 to aIF2Bα because of its central position in the translation initiation protein interaction network.

The Rpo–Rpo Interaction Network

Structure and function of archaeal RNA polymerases (Rpo) and their relationship to eukaryotic RNA polymerases have been studied very intensely (Grohmann and Werner, 2011; Werner and Grohmann, 2011; Nagy et al., 2015; Fouqueau et al., 2018). They are composed of catalytic subunits, an assembly platform, a stalk, and auxiliary subunits. The co-isolation of several Rpo subunits with aIFs prompted us to extend our study to co-isolation experiments with Rpo subunits. In total, seven subunits from different parts of the complex were chosen. With only one exception reciprocal co-isolation of Rpo subunits was observed (**Figure 7**). Additional subunits were also co-isolated, that were not used as bait proteins. In total, 10 Rpo subunits could be co-purified, which strongly indicates that the whole enzyme complex could be co-purified when a single subunit was tagged and used a bait protein. Therefore, the approach was well suited to unravel the RNA polymerase interaction network.

Transcription-Translation Coupling in Archaea and in *E. coli*

In contrast to eukaryotes, in which transcription and translation occur in two different cellular compartments, the nucleus and the cytoplasm, in prokaryotes both processes occur in the cytoplasm. Therefore, translation can begin before a gene has been fully transcribed. This allows the coupling of transcription and translation. A regulatory process called attenuation has already been described decades ago (Yanofsky, 1981). In this case, the speed of the first translating ribosome determines whether the gene is fully transcribed, or whether a transcription termination stemloop is formed.

Very recently, a much more direct coupling between transcription and translation in *E. coli* has been reported. In 2017 two structures of a complex of RNA polymerase with the 30S ribosomal subunit were solved (Demo et al., 2017; Kohler et al., 2017). Since then, several additional structures were obtained, and the current opinion is that not a single interaction complex exists, but that complex formation between the RNA polymerase and the ribosome can be very dynamic, and that the interaction can be either direct or mediated by the protein NusG, which can bind to the ribosome as well as to Rpo. Several reviews summarize the current knowledge (Artsimovitch, 2018; Conn et al., 2019; Irastortza-Olaziregi and Amster-Choder, 2020; Wang et al., 2020). However, transcription–translation coupling is not universal in bacteria, for example, it has been shown that it is not coupled in *Bacillus subtilis* and probably other gram-positive bacteria (Wang and Artsimovitch, 2021).

In contrast to the rapidly- growing experimental evidence for *E. coli* and other bacteria, nearly nothing is known about the coupling of transcription and translation in archaea. Electron microscopic observation of lysed cells of *T. kodakarensis* revealed that polysomes were very close to dispersed strands of genomic DNA, and, based on this observation, it was proposed that transcription and translation are coupled in this species (French et al., 2007). In this Research topic of Frontiers of Microbiology a theoretical paper was published that proposed transcription–translation coupling to occur in archaea based on the universal conservation of NusG/Spt5 and NusA, which connect Rpo and the ribosome in *E. coli* (Weixlbaumer et al., 2021). NusG and NusA are also encoded in the genome of *H. volcanii*. We searched for the two proteins in the MS results, but neither of them was co-isolated with any of the seven Rpo subunits used as bait proteins.

However, the reciprocal co-isolation of aIFs and Rpo subunits as well as the co-isolation of 21 ribosomal proteins with Rpo subunits as bait proteins provides very strong evidence that translation and transcription are not independent in *H. volcanii*, but the two processes are coupled (**Figure 8** and **Supplementary Tables S9 and S10**). To our knowledge this is the second experimental study that indicates that transcription and translation are coupled in archaea. Clearly, more analyses with more archaeal species are highly needed for a better

understanding of the molecular mechanism and the distribution in the domain of archaea.

CONCLUSION

In a very comprehensive approach the protein–protein interaction networks of translation initiation and the RNA polymerase of *H. volcanii* have been elucidated. Manifold aIF-aIF, aIF-ribosome, aIF-Rpo, and aIF-transcription factor interactions were observed. Two proteins turned out to be interaction hubs, i.e., the universally conserved factor aIF5B as well as HVO_0966, which we propose to rename aIF2Bα. The reciprocal co-isolation of aIFs and Rpo subunits as well as Rpo subunits and ribosomal proteins gives an additional evidence that transcription and translation are coupled in archaea.

DATA AVAILABILITY STATEMENT

The protein interactions from this publication have been submitted to the IMEx (http://www.imexconsortium.org) consortium through IntAct (Orchard et al., 2014) and assigned the identifier IM-29121. The original Excel files with the MS results have been deposited at a German repository called "Hessenbox" and can be downloaded from the following site: https://hessenbox-a10.rz.uni-frankfurt.de/getlink/fiMwRZN2jZQYjQfyWzLk7QsB/Raw%20MS%20Data%20Schramm.zip. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium vie the PRIDE partner repository (Perez-Riverol et al., 2019) with the dataset identifier PXD028666.

AUTHOR CONTRIBUTIONS

FS and JS designed the experiments. FS performed all molecular biological experiments and co-affinity purifications. UL performed all peptide mass fingerprinting and MS/MS experiments. FS, AB, UL, and JS analyzed the data. FS and AB generated all tables, figures and supplementary data. JS wrote a first draft of the manuscript. All authors contributed to the final version and approved it for publication.

FUNDING

This project was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through the grants So264/24 and So264/29 to JS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.742806/full#supplementary-material

REFERENCES

Accetto, T., and Avguštin, G. (2011). Inability of *Prevotella bryantii* to form a functional Shine-Dalgarno interaction reflects unique evolution of ribosome binding sites in Bacteroidetes. *PLoS One* 6:e22914. doi: 10.1371/journal.pone.0022914

Allers, T., Barak, S., Liddell, S., Wardell, K., and Mevarech, M. (2010). Improved strains and plasmid vectors for conditional overexpression of his-tagged proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* 76, 1759–1769. doi: 10.1128/AEM.02670-09

Anantharaman, V., Koonin, E. V., and Aravind, L. (2001). TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes. *FEMS Microbiol. Lett.* 197, 215–221. doi: 10.1111/j.1574-6968.2001.tb10606.x

Andreev, D. E., O'Connor, P. B. F., Loughran, G., Dmitriev, S. E., Baranov, P. V., and Shatsky, I. N. (2017). Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res.* 45, 513–526. doi: 10.1093/nar/gkw1190

Aono, R., Sato, T., Yano, A., Yoshida, S., Nishitani, Y., Miki, K., et al. (2012). Enzymatic characterization of AMP phosphorylase and ribose-1,5-bisphosphate isomerase functioning in an archaeal AMP metabolic pathway. *J. Bacteriol.* 194, 6847–6855. doi: 10.1128/JB.01335-12

Artsimovitch, I. (2018). Rebuilding the bridge between transcription and translation. *Mol. Microbiol.* 108, 467–472. doi: 10.1111/mmi.13964

Aylett, C. H. S., and Ban, N. (2017). Eukaryotic aspects of translation initiation brought into focus. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 372:20160186. doi: 10.1098/rstb.2016.0186

Babski, J., Haas, K. A., Näther-Schindler, D., Pfeiffer, F., Förstner, 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

Bassani, F., Zink, I. A., Pribasnig, T., Wolfinger, M. T., Romagnoli, A., Resch, A., et al. (2019). Indications for a moonlighting function of translation factor aIF5A in the crenarchaeum Sulfolobus solfataricus. *RNA Biol.* 16, 675–685. doi: 10.1080/15476286.2019.1582953

Beaufay, F., Coppine, J., and Hallez, R. (2021). When the metabolism meets the cell cycle in bacteria. *Curr. Opin. Microbiol.* 60, 104–113. doi: 10.1016/j.mib.2021.02.006

Benelli, D., and Londei, P. (2009). Begin at the beginning: evolution of translational initiation. *Res. Microbiol.* 160, 493–501. doi: 10.1016/j.resmic.2009.06.003

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

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

Bogorad, A. M., Xia, B., Sandor, D. G., Mamonov, A. B., Cafarella, T. R., Jehle, S., et al. (2014). Insights into the architecture of the eIF2Bα/β/δ regulatory subcomplex. *Biochemistry* 53, 3432–3445. doi: 10.1021/bi500346u

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., and Soppa, J. (2009). Regulation of translation in haloarchaea: 5'- and 3'-UTRs are essential and have to functionally interact in vivo. *PLoS One* 4:e4484. doi: 10.1371/journal.pone.0004484

Conn, A. B., Diggs, S., Tam, T. K., and Blaha, G. M. (2019). Two old dogs, one new trick: A review of RNA polymerase and ribosome interactions during transcription-translation coupling. *Int. J. Mol. Sci.* 20:2595. doi: 10.3390/ijms20102595

Coureux, 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

Coureux, P.-D., Lazennec-Schurdevin, C., Monestier, A., Larquet, E., Cladière, 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

Dambeck, M., and Soppa, J. (2008). Characterization of a *Haloferax volcanii* member of the enolase superfamily: deletion mutant construction, expression analysis, and transcriptome comparison. *Arch. Microbiol.* 190, 341–353. doi: 10.1007/s00203-008-0379-1

Danner, S., and Soppa, J. (1996). Characterization of the distal promoter element of halobacteria *in vivo* using saturation mutagenesis and selection. *Mol. Microbiol.* 19, 1265–1276. doi: 10.1111/j.1365-2958.1996.tb02471.x

Demo, G., Rasouly, A., Vasilyev, N., Svetlov, V., Loveland, A. B., Diaz-Avalos, R., et al. (2017). Structure of RNA polymerase bound to ribosomal 30S subunit. *elife* 6:e28560. doi: 10.7554/eLife.28560

Dev, K., Santangelo, T. J., Rothenburg, S., Neculai, D., Dey, M., Sicheri, F., et al. (2009). Archaeal aIF2B interacts with eukaryotic translation initiation factors eIF2alpha and eIF2Balpha: implications for aIF2B function and eIF2B regulation. *J. Mol. Biol.* 392, 701–722. doi: 10.1016/j.jmb.2009.07.030

Dever, T. E., Kinzy, T. G., and Pavitt, G. D. (2016). Mechanism and regulation of protein synthesis in *Saccharomyces cerevisiae*. *Genetics* 203, 65–107. doi: 10.1534/genetics.115.186221

Duval, M., Simonetti, A., Caldelari, I., and Marzi, S. (2015). Multiple ways to regulate translation initiation in bacteria: mechanisms, regulatory circuits, dynamics. *Biochimie* 114, 18–29. doi: 10.1016/j.biochi.2015.03.007

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

Fouqueau, T., Blombach, F., Cackett, G., Carty, A. E., Matelska, D. M., Ofer, S., et al. (2018). The cutting edge of archaeal transcription. *Emerg Top Life Sci* 2, 517–533. doi: 10.1042/ETLS20180014

Fox, G. E. (2010). Origin and evolution of the ribosome. *Cold Spring Harb. Perspect. Biol.* 2:a003483. doi: 10.1101/cshperspect.a003483

Freese, N. H., Norris, D. C., and Loraine, A. E. (2016). Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32, 2089–2095. doi: 10.1093/bioinformatics/btw069, PMID: 27153568

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., Näther, 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

Gogoi, P., and Kanaujia, S. P. (2018). A presumed homologue of the regulatory subunits of eIF2B functions as ribose-1,5-bisphosphate isomerase in *Pyrococcus horikoshii* OT3. *Sci. Rep.* 8:1891. doi: 10.1038/s41598-018-20418-w

Gordiyenko, Y., Llácer, J. L., and Ramakrishnan, V. (2019). Structural basis for the inhibition of translation through eIF2α phosphorylation. *Nat. Commun.* 10:2640. doi: 10.1038/s41467-019-10606-1

Grohmann, D., and Werner, F. (2011). Recent advances in the understanding of archaeal transcription. *Curr. Opin. Microbiol.* 14, 328–334. doi: 10.1016/j.mib.2011.04.012

Gualerzi, C. O., and Pon, C. L. (2015). Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell. Mol. Life Sci.* 72, 4341–4367. doi: 10.1007/s00018-015-2010-3

Guca, E., and Hashem, Y. (2018). Major structural rearrangements of the canonical eukaryotic translation initiation complex. *Curr. Opin. Struct. Biol.* 53, 151–158. doi: 10.1016/j.sbi.2018.08.006

Gutiérrez, P., Coillet-Matillon, S., Arrowsmith, C., and Gehring, K. (2002). Zinc is required for structural stability of the C-terminus of archaeal translation initiation factor aIF2beta. *FEBS Lett.* 517, 155–158. doi: 10.1016/S0014-5793(02)02610-8

Gutiérrez, P., Osborne, M. J., Siddiqui, N., Trempe, J.-F., Arrowsmith, C., and Gehring, K. (2004). Structure of the archaeal translation initiation factor aIF2 beta from *Methanobacterium thermoautotrophicum*: implications for translation initiation. *Protein Sci.* 13, 659–667. doi: 10.1110/ps.03506604

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580. doi: 10.1016/S0022-2836(83)80284-8

Hao, Q., Heo, J.-M., Nocek, B. P., Hicks, K. G., Stoll, V. S., Remarcik, C., et al. (2021). Sugar phosphate activation of the stress sensor eIF2B. *Nat. Commun.* 12:3440. doi: 10.1038/s41467-021-23836-z

Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., et al. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* 5:e9605. doi: 10.1371/journal.pone.0009605

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

Henderson, B., and Martin, A. (2013). Bacterial moonlighting proteins and bacterial virulence. *Curr. Top. Microbiol. Immunol.* 358, 155–213. doi: 10.1007/82_2011_188

Hering, O., Brenneis, M., Beer, J., Suess, B., and Soppa, J. (2009). A novel mechanism for translation initiation operates in haloarchaea. *Mol. Microbiol.* 71, 1451–1463. doi: 10.1111/j.1365-2958.2009.06615.x

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

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

Irastortza-Olaziregi, M., and Amster-Choder, O. (2020). Coupled transcription-translation in prokaryotes: An old couple with new surprises. *Front. Microbiol.* 11:624830. doi: 10.3389/fmicb.2020.624830

Jeffery, C. J. (2020). Enzymes, pseudoenzymes, and moonlighting proteins: diversity of function in protein superfamilies. *FEBS J.* 287, 4141–4149. doi: 10.1111/febs.15446

Jiang, S., Liu, S., and Wu, C. (2016). Developing protocols of tricine-SDS-PAGE for separation of polypeptides in the mass range 1-30 kDa with minigel electrophoresis system. *Int. J. Electrochem. Sci.* 11, 640–649.

Kaberdin, V. R., and Bläsi, U. (2006). Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol. Rev.* 30, 967–979. doi: 10.1111/j.1574-6976.2006.00043.x

Kakuta, Y., Tahara, M., Maetani, S., Yao, M., Tanaka, I., and Kimura, M. (2004). Crystal structure of the regulatory subunit of archaeal initiation factor 2B (aIF2B) from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3: a proposed structure of the regulatory subcomplex of eukaryotic IF2B. *Biochem. Biophys. Res. Commun.* 319, 725–732. doi: 10.1016/j.bbrc.2004.05.045

Kashiwagi, K., Yokoyama, T., Nishimoto, M., Takahashi, M., Sakamoto, A., Yonemochi, M., et al. (2019). Structural basis for eIF2B inhibition in integrated stress response. *Science* 364, 495–499. doi: 10.1126/science.aaw4104

Koh, G. C. K. W., Porras, P., Aranda, B., Hermjakob, H., and Orchard, S. E. (2012). Analyzing protein-protein interaction networks. *J. Proteome Res.* 11, 2014–2031. doi: 10.1021/pr201211w

Kohler, R., Mooney, R. A., Mills, D. J., Landick, R., and Cramer, P. (2017). Architecture of a transcribing-translating expressome. *Science* 356, 194–197. doi: 10.1126/science.aal3059

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

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

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

Laass, S., Monzon, V. A., Kliemt, J., Hammelmann, M., Pfeiffer, F., Förstner, K. U., et al. (2019). Characterization of the transcriptome of *Haloferax volcanii*, grown under four different conditions, with mixed RNA-Seq. *PLoS One* 14:e0215986. doi: 10.1371/journal.pone.0215986

Laurino, J. P., Thompson, G. M., Pacheco, E., and Castilho, B. A. (1999). The beta subunit of eukaryotic translation initiation factor 2 binds mRNA through

the lysine repeats and a region comprising the C2-C2 motif. *Mol. Cell. Biol.* 19, 173–181. doi: 10.1128/MCB.19.1.173

Li, Z., Santangelo, T. J., Cuboňová, L., Reeve, J. N., and Kelman, Z. (2010). Affinity purification of an archaeal DNA replication protein network. *MBio* 1:e00221-10. doi: 10.1128/mBio.00221-10

Li, J., Zhang, B., Zhou, L., Qi, L., Yue, L., Zhang, W., et al. (2019). The archaeal RNA chaperone TRAM0076 shapes the transcriptome and optimizes the growth of *Methanococcus maripaludis*. *PLoS Genet.* 15:e1008328. doi: 10.1371/journal.pgen.1008328

Liu, H., and Jeffery, C. J. (2020). Moonlighting proteins in the fuzzy logic of cellular metabolism. *Molecules* 25:3440. doi: 10.3390/molecules25153440

Liu, Y., Makarova, K. S., Huang, W.-C., Wolf, Y. I., Nikolskaya, A. N., Zhang, X., et al. (2021). Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 593, 553–557. doi: 10.1038/s41586-021-03494-3

Londei, P. (2005). Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol. Rev.* 29, 185–200. doi: 10.1016/j.fmrre.2004.10.002

Malys, N., and McCarthy, J. E. G. (2011). Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* 68, 991–1003. doi: 10.1007/s00018-010-0588-z

Marintchev, A., and Ito, T. (2020). eIF2B and the integrated stress response: A structural and mechanistic view. *Biochemistry* 59, 1299–1308. doi: 10.1021/acs.biochem.0c00132

Marintchev, A., and Wagner, G. (2004). Translation initiation: structures, mechanisms and evolution. *Q. Rev. Biophys.* 37, 197–284. doi: 10.1017/S0033583505004026

Maurer, S., Ludt, K., and Soppa, J. (2018). Characterization of copy number control of two *Haloferax volcanii* replication origins using deletion mutants and Haloarchaeal artificial chromosomes. *J. Bacteriol.* 200. doi: 10.1128/JB.00517-17

Milón, P., and Rodnina, M. V. (2012). Kinetic control of translation initiation in bacteria. *Crit. Rev. Biochem. Mol. Biol.* 47, 334–348. doi: 10.3109/10409238.2012.678284

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:e00067-18. doi: 10.1128/MCB.00067-18

Nagy, J., Grohmann, D., Cheung, A. C. M., Schulz, S., Smollett, K., Werner, F., et al. (2015). Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat. Commun.* 6:6161. doi: 10.1038/ncomms7161

Naveau, M., Lazennec-Schurdevin, C., Panvert, M., Dubiez, E., Mechulam, Y., and Schmitt, E. (2013). Roles of yeast eIF2α and eIF2β subunits in the binding of the initiator methionyl-tRNA. *Nucleic Acids Res.* 41, 1047–1057. doi: 10.1093/nar/gks1180

Nika, J., Rippel, S., and Hannig, E. M. (2001). Biochemical analysis of the eIF2beta gamma complex reveals a structural function for eIF2alpha in catalyzed nucleotide exchange. *J. Biol. Chem.* 276, 1051–1056. doi: 10.1074/jbc.M007398200

Nikonov, O. S., Nevskaya, N. A., Garber, M. B., and Nikonov, S. V. (2021). Structure and function of archaeal translation initiation factor 2 fragments containing Cys2-Cys2 motifs. *Biochemistry (Mosc)* 86, 1003–1011. doi: 10.1134/S0006297921080101

Opron, K., and Burton, Z. F. (2018). Ribosome structure, function, and early evolution. *Int. J. Mol. Sci.* 20:40. doi: 10.3390/ijms20010040

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42, D358–D363. doi: 10.1093/nar/gkt1115

Pedullà, N., Palermo, R., Hasenöhrl, D., Bläsi, 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

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450. doi: 10.1093/nar/gky1106

Pfeiffer, F., Broicher, A., Gillich, T., Klee, K., Mejía, J., Rampp, M., et al. (2008). Genome information management and integrated data analysis with HaloLex. *Arch. Microbiol.* 190, 281–299. doi: 10.1007/s00203-008-0389-z

Pluchon, P.-F., Fouqueau, T., Crezé, C., Laurent, S., Briffotaux, J., Hogrel, G., et al. (2013). An extended network of genomic maintenance in the archaeon *Pyrococcus abyssi* highlights unexpected associations between eucaryotic homologs. *PLoS One* 8:e79707. doi: 10.1371/journal.pone.0079707

Rodríguez-Saavedra, C., Morgado-Martínez, L. E., Burgos-Palacios, A., King-Díaz, B., López-Coria, M., and Sánchez-Nieto, S. (2021). Moonlighting proteins: The case of the hexokinases. *Front. Mol. Biosci.* 8:701975. doi: 10.3389/fmolb.2021.701975

Salton, G. D., Laurino, C. C. F. C., Mega, N. O., Delgado-Cañedo, A., Setterblad, N., Carmagnat, M., et al. (2017). Deletion of eIF2β lysine stretches creates a dominant negative that affects the translation and proliferation in human cell line: A tool for arresting the cell growth. *Cancer Biol. Ther.* 18, 560–570. doi: 10.1080/15384047.2017.1345383

Sato, T., Atomi, H., and Imanaka, T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* 315, 1003–1006. doi: 10.1126/science.1135999

Schlesner, M., Miller, A., Besir, H., Aivaliotis, M., Streif, J., Scheffer, B., et al. (2012). The protein interaction network of a taxis signal transduction system in a halophilic archaeon. *BMC Microbiol.* 12:272. doi: 10.1186/1471-2180-12-272

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

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., 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

Schoof, M., Boone, M., Wang, L., Lawrence, R., Frost, A., and Walter, P. (2021). eIF2B conformation and assembly state regulate the integrated stress response. *elife* 10:e65703. doi: 10.7554/eLife.65703

Shirokikh, N. E., and Preiss, T. (2018). Translation initiation by cap-dependent ribosome recruitment: recent insights and open questions. *Wiley Interdiscip. Rev. RNA* 9:e1473. doi: 10.1002/wrna.1473

Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., et al. (2009). A structural view of translation initiation in bacteria. *Cell. Mol. Life Sci.* 66, 423–436. doi: 10.1007/s00018-008-8416-4

Singh, N., and Bhalla, N. (2020). Moonlighting proteins. *Annu. Rev. Genet.* 54, 265–285. doi: 10.1146/annurev-genet-030620-102906

Taha, , Siddiqui, K. S., Campanaro, S., Najnin, T., Deshpande, N., Williams, T. J., et al. (2016). Single TRAM domain RNA-binding proteins in Archaea: functional insight from Ctr3 from the Antarctic methanogen *Methanococcoides burtonii*. *Environ. Microbiol.* 18, 2810–2824. doi: 10.1111/1462-2920.13229

Tahara, M., Ohsawa, A., Saito, S., and Kimura, M. (2004). In vitro phosphorylation of initiation factor 2 alpha (aIF2 alpha) from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. *J. Biochem.* 135, 479–485. doi: 10.1093/jb/mvh055

Turek, I., and Irving, H. (2021). Moonlighting proteins shine new light on molecular signaling niches. *Int. J. Mol. Sci.* 22:1367. doi: 10.3390/ijms22031367

Wang, B., and Artsimovitch, I. (2021). A growing gap between the RNAP and the lead ribosome. *Trends Microbiol.* 29, 4–5. doi: 10.1016/j.tim.2020.09.011

Wang, C., Molodtsov, V., Firlar, E., Kaelber, J. T., Blaha, G., Su, M., et al. (2020). Structural basis of transcription-translation coupling. *Science* 369, 1359–1365. doi: 10.1126/science.abb5317

Weisser, M., and Ban, N. (2019). Extensions, extra factors, and extreme complexity: ribosomal structures provide insights into eukaryotic translation. *Cold Spring Harb. Perspect. Biol.* 11:a032367. doi: 10.1101/cshperspect.a032367

Weixlbaumer, A., Grünberger, F., Werner, F., and Grohmann, D. (2021). Coupling of transcription and translation in archaea: cues from the bacterial world. *Front. Microbiol.* 12:661827. doi: 10.3389/fmicb.2021.661827

Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* 9, 85–98. doi: 10.1038/nrmicro2507

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

Yang, J., Wagner, S. A., and Beli, P. (2015). Illuminating spatial and temporal organization of protein interaction networks by mass spectrometry-based proteomics. *Front. Genet.* 6:344. doi: 10.3389/fgene.2015.00344

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. *Nature* 289, 751–758. doi: 10.1038/289751a0

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

Zhang, B., Yue, L., Zhou, L., Qi, L., Li, J., and Dong, X. (2017). Conserved TRAM domain functions as an archaeal cold shock protein via RNA chaperone activity. *Front. Microbiol.* 8:1597. doi: 10.3389/fmicb.2017.01597

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership