Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation

Pierre-Damien Coureux1, Christine Lazennec-Schurdevin1, Sophie Bourcier2, Yves Mechulam1 & Emmanuelle Schmitt1*

Archaeal translation initiation occurs within a macromolecular complex containing the small ribosomal subunit (30S) bound to mRNA, initiation factors aIF1, aIF1A and the ternary complex aIF2:GDPNP:Met-tRNAiMet. Here, we determine the cryo-EM structure of a 30S:mRNA:aIF1A:aIF2:GTP:Met-tRNAiMet complex from Pyrococcus abyssi at 3.2 Å resolution. It highlights archaeal features in ribosomal proteins and rRNA modifications. We find an aS21 protein, at the location of eS21 in eukaryotic ribosomes. Moreover, we identify an N-terminal extension of archaeal eL41 contacting the P site. We characterize 34 N4-acetylcytidines distributed throughout 16S rRNA, likely contributing to hyperthermostability. Without aIF1, the 30S head is stabilized and initiator tRNA is tightly bound to the P site. A network of interactions involving tRNA, mRNA, rRNA modified nucleotides and C-terminal tails of uS9, uS13 and uS19 is observed. Universal features and domain-specific idiosyncrasies of translation initiation are discussed in light of ribosomal structures from representatives of each domain of life.
Translation initiation universally occurs with accurate selection of the start codon that defines the reading frame on the mRNA. The mechanism involves a macromolecular complex composed of the small ribosomal subunit, the mRNA, a specialized methionylated initiator tRNA and initiation factors (IFs). Even if the process is universal, the molecular mechanisms are different in the three domains of life. In bacteria, the ribosome generally binds in the vicinity of the AUG start codon through interaction of 16S RNA with the Shine-Dalgarno (SD) sequence on mRNA. The initiator tRNA is formulated and only three IFs, IF1–IF3 assist the start codon selection mechanism. In eukaryotes, translation initiation is more complex with matured mRNAs and many IFs. The canonical mechanism involves a pre-initiation complex (PIC) comprising the ternary complex eIF2: GTP:Met-tRNAiMet (TC), the two small IFs eIF1 and eIF1A, as well as eIF5 and eIF3. In the presence of factors belonging to the eIF4 family, the PIC is recruited near the 5′-capped end and scans the mRNA until an AUG codon in a correct context (Kozak motif) is found. AUG recognition stops scanning, provokes factor release and the assembly of an elongation-competent 80S complex through large subunit joining, with the help of eIF5B and eIF1A. eIF1 and eIF1A are key factors of the scanning process favoring a P OUT conformation of the TC, where tRNA is not fully accommodated in the P site. The multimeric factor eIF3 also stimulates attachment of the PIC to the mRNA and scanning. Finally, eIF5 is the guanine activating protein of eIF2 that stimulates GTP hydrolysis during scanning of the mRNA.

In archaea, genomic analyses have shown that three IFs homologous to their eukaryotic counterparts, aIF1, aIF1A, and aIF2 are found. Moreover, archael ribosomal proteins are either universal or shared with eukaryotes showing the proximity of the two ribosomes. However, there is no long-range scanning because mRNAs have SD sequences or very short 5′ untranslated regions (UTR) that allow pre-positioning of the IC in the vicinity of the start codon. Despite this, a local search of the mRNA by the IC (termed “local scanning” in ref. 10) is necessary to allow precise positioning of the start codon in the P site. Thus, even if the recruitment of the PIC on the mRNA is different in eukaryotes and archaea, start codon selection is achieved within a common structural core made up of the small ribosomal subunit, the mRNA, the methionylated initiator tRNA (Met-tRNAiMet) and the three IFs e/aIF1, e/aIF1A and e/aIF2. e/aIF2 is a specific eukaryotic and archaeal heterotrimeric protein that binds the initiator tRNA in the presence of GTP. Previous biochemical and structural studies of a full IC from Pyrococcus abyssi (30S:mRNA:TC:aIF1:aIF1A) identified two conformations with the initiator tRNA either in a remote position (IC0-PREMOTE) or bound to the P site (IC1-Pm). This led us to propose that conformational changes of the TC may participate in start codon selection.

In our model (named “spring force model” in ref. 11), interaction of aIF2 with h44 of the 30S would counteract accommodation of the tRNA in the P site. However, formation of correct codon–anticodon pairing in the P site would compensate for the restoring force exerted by aIF2 on the tRNA. This would allow a longer stay of the initiator tRNA in the P site and trigger further events, including aIF1 departure because of steric hindrance and release of aIF2 in its GDP bound form. The role of aIF1-induced dynamics of the IC in the start codon selection was supported by toeprinting experiments. In the absence of aIF1, the IC becomes more stable, as observed by a restricted toeprinting signal. However, no structural view of an archael IC illustrating a state following aIF1 departure has been described to date.

In the present study, we determine the cryo-EM structure of an archael IC (IC2, 30S:mRNA:aIF1A:aIF2-GTP:Met-tRNAiMet (TC)) from P. abyssi devoid of aIF1 at an overall resolution of 3.2 Å. Full reconstruction of an atomic model of the small ribosomal subunit highlights archael features in ribosomal proteins and rRNA modifications. We find a previously unidentified archael ribosomal protein aS21, at the location of eS21 in eukaryotic ribosomes. Moreover, a previously unobserved N-terminal extension of eL41 contacts the P site. We also identify a set of 34 N4-acyctylcytidines distributed throughout the 16S rRNA. These base modifications likely participate in the hyperthermostability of this ribosome. In the absence of aIF1, the 3OS head is no longer mobile and the initiator tRNA becomes stably bound to the P site. A network of interactions involving rRNA modified nucleotides and the C-terminal tails of three universal ribosomal proteins, uS9, uS13, and uS19 is observed. Universal features and domain-specific idiosyncrasies of translation initiation are discussed in light of ribosomal structures from representatives of each domain of life.

Results

Overview of the IC2 cryo-EM structure. In order to study the impact of aIF1 departure during translation initiation, we prepared an initiation complex (IC2) without this factor. IC2 contains archael 30S subunits from P. abyssi (Pa-30S), Pa-aIF1A, the ternary complex (Pa-aIF2-GDPNP:Met-tRNAiMetA1A1U1U12) and a synthetic 26 nucleotide-long mRNA. Cryo-EM images were collected on a Titan Krios microscope (Table 1). After image processing, 218 k particles were used for refinement without classification. A density map to 3.2 Å resolution was obtained, showing a very good structural homogeneity of the complex. Density subtractions of the head or the body parts of the 30S further improved map quality (Supplementary Fig. 1). The high resolution of the electron density map allowed complete reconstruction of the 30S, as described below. After density subtraction and classification in RELION, one class showed very weak electron density for aIF1A and was therefore not further refined (IC2C, see Methods). The other classes showed two conformations of the initiation complex, named IC2A (34k particles, 4.2 Å resolution, Fig. 1a) and IC2B (142k particles, 3.3 Å resolution, Fig. 1b). The corresponding models were refined in PHENIX (Supplementary Figs. 1–3, Table 1 and Supplementary Table 1). In the two conformations, the initiator tRNA and the mRNA are firmly bound to the 30S. Moreover, the position of the 30S head does not detectably change as compared to IC1-PIN. As already observed for ribosomal initiation complexes, local resolution of the electron density is higher for the ribosome core than for the peripheral IFs (Supplementary Fig. 3). As a consequence, aIF2-subunits and aIF1A were placed in the density by rigid body fitting of crystallographic structures (Table 1 and Methods). The two mobile wings of aIF2 (aIF2b core domain and aIF2a domain 1–2) are poorly defined in the two conformations and their positions were only tentatively modeled. As observed in IC0 and IC1, aIF1A is located in the A site. Its position corresponds to that observed for eukaryotic eIF1A in PIC, and aIF2 is bound to the 3′ aminoacylated end of the initiator tRNA. However, whereas domain III of aIF2 (aIF2yDII1) contacts h44 in IC2A, the contact is lost in IC2B. Departure of aIF2y from h44 is accompanied by a local movement of the rRNA helix (Supplementary Fig. 4). Because the non-hydrolyzable GTP analog, GDPNP, has been used in complex preparation, aIF2 is not fully released (Fig. 1).

Identification of archael specificities of Pa-30S. Complete building of the Pa-30S subunit was performed in the 3.2 Å resolution cryo-EM map using the medium resolution archael 30S structures as guides (Fig. 2a). Statistics of the refined structure and final secondary structure diagram of Pa-16S rRNA
**Table 1 Cryo-EM statistics.**

| Data collection and processing | IC2 body | IC2 head | IC2A | IC2B |
|-------------------------------|----------|----------|------|------|
| EMDB-10323 PDB 6SWD          |          |          |      |      |
| Magnification                | ×128,440 |          |      |      |
| Voltage (kV)                 | 300      |          |      |      |
| Electron exposure (e−/Å²)    |          |          |      |      |
| Defocus range (μm)           | −0.5 to −5 |          |      |      |
| Pixel size (Å)               | 1.09     |          |      |      |
| Symmetry imposed             | C1       |          |      |      |
| Initial particle images (no.)|          |          |      |      |
| Final particle images (no.)  | −218,000 | −218,000 | −34,000 | −142,000 |
| Resolution (unmasked, Å)     | 3.2      | 3.1      | 4.8  | 3.5 |
| Resolution (masked, Å)       | 3.2      | 3.1      | 4.2  | 3.3 |
| Initial models used (PDB codes) | 4V6U, 5JB3 | 4V6U, 5JB3, 4RD4, 2QMU, 4MNO |
| Model resolution (Å)         | 3.2      | 3.1      | 4.2  | 3.3 |
| Map sharpening B factor (Å²)  | −86      | −88      | −107 | −92 |
| Model composition            |          |          |      |      |
| Nonhydrogen atoms            | 41,559   | 22,797   | 70,662 | 70,737 |
| Protein residues/nucleotides | 2293/1039 | 1525/479 | 4508/1560 | 4518/1560 |
| Ligands                       | 55       | 17       | 74   | 74  |
| B factors (Å²)               |          |          |      |      |
| Protein                      | 26       | 52       | 135  | 57  |
| Nucleic acid                 | 38       | 44       | 106  | 52  |
| Ligand                       | 22       | 35       | 77   | 33  |
| R.m.s. deviations            |          |          |      |      |
| Bond lengths (Å)             | 0.012    | 0.010    | 0.011 | 0.010 |
| Bond angles (°)              | 1.012    | 0.937    | 1.207 | 0.895 |
| Validation                   |          |          |      |      |
| MolProbity score             | 1.91     | 2.00     | 2.68  | 2.13 |
| Clashscore                   | 5.65     | 7.31     | 29.24 | 9.42 |
| Poor rotamers (%)            | 1.12     | 0.55     | 0.80  | 0.05 |
| Ramachandran plot            |          |          |      |      |
| Favored (%)                  | 89.29    | 88.04    | 80.8  | 86.62 |
| Allowed (%)                  | 10.62    | 11.96    | 19.04 | 13.35 |
| Disallowed (%)               | 0.09     | 0.00     | 0.00  | 0.02 |

Data collection, processing, and refinement statistics for IC2A, IC2B, IC2 body, and IC2 head datasets.

**Fig. 1 Cryo-EM maps of two conformations of the initiation complex 30S:mRNA:aIF1A:TC (IC2).** The figure describes the two refined conformations, IC2A and IC2B, resulting from 3D classification. These two conformations essentially differ by the position of αIF2 with respect to h44 (see text). **a** Cryo-EM map of IC2A at a nominal resolution of 4.2 Å. **b** Cryo-EM map of IC2B at a nominal resolution of 3.3 Å. Regions of the maps are colored by component to show the 30S subunit (pale yellow), αIF1A (orange), Met-tRNA<sub>Met</sub> (yellow), αIF2γ (green), αIF2β (gray), and αIF2α (cyan). The same color code is used in all figures. The mRNA is not visible in this orientation. The view was drawn using Chimera<sup>90</sup>. For panel **b**, a composite made of the IC2B map contoured at two different levels is represented in order to allow visualization of the IFs together with a sufficient level of details for the 30S.
Fig. 2 Identified archaeal specificities of Pa-30S. a Overall view of the IC2B structure. N^4-acetylcytidines identified in this study are shown in blue spheres and other nucleotide modifications are in red spheres. eL41 is in magenta and h44 is in black. The view shows that N^4-acetylcytidines are distributed throughout the 16S rRNA. b Representative primer extension analysis of 16S rRNA to map ac^4C residues in 16S rRNA as described32. Lane 1, 16S rRNA treated with NaBH4 (100 mM, 37 °C, 1 h), lane 2 control 16S rRNA (37 °C, 1 h), lane 3, intact 16S rRNA. The RT stops are indicated to the left of the view. RT primer is number 8 (Supplementary Table 4). Other gels are shown in Fig. S10. c Structure of Pa-eL41. The newly identified N-terminal extension (1–14) and the eukaryotic-archaeal core helix are shown in magenta. A sequence alignment of S. cerevisiae and P. abyssi eL41 is shown below. d The N-terminal extension of eL41 is shown in magenta sticks. Modified bases are shown in red sticks except ac^4C1479 shown in blue sticks. mRNA is in blue and tRNA is in yellow. Start codon and anticodon nucleotides are labeled. e Structure of the archaeal S21 ribosomal protein (pink, left) compared to eS21 (blue, middle, PDB: 6FYX). The two superimposed structures are shown on the right view. The resulting sequence alignment is shown below with structurally superimposed regions underlined. Figures 2–4 were drawn with Pymol98.
are shown in Table 1 and in Supplementary Fig. 5, respectively. Overall, the structure is similar to that of *Pyrococcus furiosus* 30S\(^{26}\). However, many modifications have been rendered possible by the higher resolution of the electron density maps. Two peculiarities concerning the ribosomal proteins deserve special attention. The first one concerns archaeal eL41 (the 2014 system for naming ribosomal proteins is used throughout\(^{8}\)). As previously noted from the yeast 80S structure\(^{23}\), eL41 is more strongly associated to the small ribosomal subunit than to the large one. Consistent with this observation, eL41 was indeed found in the structures of several 40S\(^{17,22}\). In the present archaeal structure, eL41 is also found in the small ribosomal subunit, as in\(^{12}\). However, in contrast with previous annotations of the genomes of *P. furiosus* or *P. abyssi*, the electron density shows that eL41 has an N-terminal extension (Fig. 2c). The protein contains 37 amino acids with an N-terminal peptide, buried in a cavity lined by h27, h4, and h45, followed by a 20 residue long helix (Fig. 2c, d and Supplementary Fig. 6a). A protein of this length has indeed been annotated as eL41 in several archaeal genomes, arguing in favor of a conservation of the N-terminal extension in this domain of life (Supplementary Fig. 6b, c). In eukaryotes, eL41 was indeed found in the yeast 80S structure\(^{21}\), eL41 is more strongly associated to the small ribosomal subunit than to the large one. Importantly, we confirmed the presence of all rRNA modifications introduced in the model (Tables 2 and 3) using liquid chromatography/high-resolution mass spectrometry (LC-HRMS) (Supplementary Fig. 10, Supplementary Table 2 and Methods). The only exception is hm\(^{3}\)C, tentatively modeled at position 1378 (Supplementary Fig. 8 and Table 3).

Beside rRNA modifications already characterized in other domains of life, the electron density showed the presence of many N\(^{4}\)-acetylcytidines (ac\(^{4}\)C; Table 2, Supplementary Figs. 5 and 9) distributed throughout the 16S rRNA. A total of 34 N\(^{4}\)-acetylcytidines were identified. The high level of posttranscriptional

### Table 2 N\(^{4}\)-acetylcytidines in *P. abyssi* 16S rRNA of IC2.

| rRNA structural element | Position |
|-------------------------|----------|
| h1-uS5                  | ac\(^{4}\)C\(^{1}\) |
| h4                     | ac\(^{4}\)C\(^{5}\) |
| h11                    | ac\(^{4}\)Cm\(^{25}\) |
| h11                    | ac\(^{4}\)C\(^{28}\) |
| h12                    | ac\(^{4}\)C\(^{30}\) |
| h12                    | ac\(^{4}\)C\(^{39}\) |
| h15                    | ac\(^{4}\)C\(^{37}\) |
| h15                    | ac\(^{4}\)C\(^{39}\) |
| h18                    | ac\(^{4}\)C\(^{47}\) |
| h18                    | ac\(^{4}\)C\(^{51}\) |
| h20                    | ac\(^{4}\)C\(^{54}\) |
| h21                    | ac\(^{4}\)C\(^{59}\) |
| h22                    | ac\(^{4}\)C\(^{62}\) |
| h22                    | ac\(^{4}\)C\(^{63}\) |
| h23                    | ac\(^{4}\)C\(^{64}\) |
| h23                    | ac\(^{4}\)C\(^{70}\) |
| h24                    | ac\(^{4}\)C\(^{71}\) |
| h24                    | ac\(^{4}\)C\(^{73}\) |
| h26                    | ac\(^{4}\)C\(^{82}\) |
| h26                    | ac\(^{4}\)C\(^{83}\) |
| h25-eS12               | ac\(^{4}\)C\(^{84}\) |
| h19-eS12               | ac\(^{4}\)C\(^{85}\) |
| h27                    | ac\(^{4}\)C\(^{86}\) |
| h27                    | ac\(^{4}\)C\(^{95}\) |
| h32                    | ac\(^{4}\)C\(^{99}\) |
| h32                    | ac\(^{4}\)C\(^{101}\) |
| h34                    | ac\(^{4}\)C\(^{102}\) |
| h36                    | ac\(^{4}\)C\(^{104}\) |
| h40                    | ac\(^{4}\)C\(^{114}\) |
| h34                    | ac\(^{4}\)C\(^{184}\) |
| h32                    | ac\(^{4}\)C\(^{193}\) |
| h41                    | ac\(^{4}\)C\(^{233}\) |
| h34                    | ac\(^{4}\)C\(^{239}\) |
| h45-eL41               | ac\(^{4}\)C\(^{479}\) |

### Table 3 16S rRNA modifications other than N\(^{4}\)-acetylcytidines in *P. abyssi* 30S.

| rRNA structural element | Position |
|-------------------------|----------|
| Hairpin loop:h13        | Am\(^{3}\)73 |
| P site                  | m\(^{1}\)Ψ\(^{38}\) |
| P site                  | m\(^{1}\)C\(^{39}\) |
| P site-uS13             | m\(^{1}\)C\(^{120}\) |
| P site-mRNA             | Cm\(^{13}\)76 |
| P site                  | hm\(^{3}\)C\(^{137}\) |
| P site-rsmE             | m\(^{3}\)C\(^{14}\)67 |
| P site bulge loop:h44   | m\(^{3}\)A\(^{14}\)69 |
| P site-KsgA universal   | m\(^{3}\)A\(^{14}\)87 |
| P site-KsgA universal   | m\(^{3}\)A\(^{14}\)88 |

All modified nucleotides were observed by LC-HRMS, except for hm\(^{3}\)C, which was tentatively modeled at position 1378. Note that an unknown modification was identified at the corresponding position in some archaeal\(^{20}\).
modifications and the presence of N4-acetylcytidines observed here is reminiscent of previous studies concerning two hyperthermophilic crenarchaea (Sulfolobus solfataricus and Pyrodictium occultum)\textsuperscript{30,31}. In order to confirm the positions of ac\textsuperscript{4}C in the 16S rRNA sequence, we performed reverse transcription mapping on borohydride-reduced 16S rRNA as described\textsuperscript{32} (Fig. 2b and Supplementary Fig. 11). N4-acetylcytidine modifications systematically target the second cytosine of a 5' CCG 3' sequence inside or at extremities of 16S rRNA helices (Supplementary Fig. 5). Remarkably, this is also true for previously identified N4-acetylcytidines in eukaryotic 18S rRNA\textsuperscript{24,33,34}. However the CCG motif was not highlighted at this time because too few modified sites were available. Ac\textsuperscript{4}C have been modeled in their preferred proximal conformation as observed in the ac\textsuperscript{4}C nucleoside crystal

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
& \textbf{E. coli} & \textbf{T. thermophilus} & \textbf{P. abyssi} & \textbf{S. cerevisiae} & \textbf{H. sapiens} \\
\hline
m\textsuperscript{2}G966 (B) & m\textsuperscript{2}G944 (B) & m\textsuperscript{2}G938 (AE) & m\textsuperscript{2}G939 & m\textsuperscript{2}G939 (B) & m\textsuperscript{2}G939 (B) \\
\hline
m\textsuperscript{5}C967 & C945 & m\textsuperscript{5}C939 & m\textsuperscript{5}C920 (AE) & C1192 & C1249 \\
\hline
C1228 & C1210 & m\textsuperscript{5}C1202 & m\textsuperscript{5}C1202 & C1461 & A1522 \\
\hline
C1400 & m\textsuperscript{5}C1383 & C1374 & C1374 & C1637 & C1701 \\
\hline
\textsuperscript{m}C1402 (B) & m\textsuperscript{5}C1385 (B) & m\textsuperscript{5}C1376 (AE) & Cm1376 (AE) & Cm1639 (AE) & Cm1703 (AE) \\
\hline
C1404 & m\textsuperscript{5}C1387 & C1381 & C1381 & C1641 & C1701 \\
\hline
\textsuperscript{m}C1407 (B) & m\textsuperscript{5}C1390 (B) & hm\textsuperscript{5}C1378 & hm\textsuperscript{5}C1378 & C1644 & C1708 \\
\hline
\textsuperscript{m}U1498 & \textsuperscript{m}U1476 & \textsuperscript{m}U1467 & \textsuperscript{m}U1467 & U1761 & \textsuperscript{m}U1830 \\
\hline
A1500 & A1478 & A1469 (AE) & A1469 (AE) & A1763 (AE) & m\textsuperscript{6}A1832 (AE) \\
\hline
C1510 & C1488 & ac\textsuperscript{4}C1479 (AE) & ac\textsuperscript{4}C1479 (AE) & ac\textsuperscript{4}C1773 (AE) & ac\textsuperscript{4}C1842 (AE) \\
\hline
m\textsuperscript{2}G1516 (B) & G1494 (B) & G1485 & G1485 & U1779 & U1848 \\
\hline
m\textsuperscript{6}A1518 & m\textsuperscript{6}A1492 & m\textsuperscript{6}A1487 & m\textsuperscript{6}A1487 & m\textsuperscript{6}A1871 & m\textsuperscript{6}A1850 \\
\hline
m\textsuperscript{5}A1519 & m\textsuperscript{5}A1493 & m\textsuperscript{5}A1488 & m\textsuperscript{5}A1488 & m\textsuperscript{5}A1782 & m\textsuperscript{5}A1851 \\
\hline
\end{tabular}
\caption{16S rRNA modifications in P site vicinity in the three domains of life.}
\end{table}
structure and also as calculated using quantum chemistry methods. This conformation allows canonical Watson-Crick base pairing interactions. Moreover, the acetyl group reinforces π-π stacking with adjacent bases that increases stability of the base pair. Therefore, this modification distributed throughout the 16S rRNA might largely contribute to the hyperthermostability. Some ac4C residues also interact with ribosomal proteins. One notable interaction involves the eukaryotic-conserved ac4C1479 and R15 of eL41 (Fig. 2d). This position corresponds to acC1773 and acC1842 previously characterized in S. cerevisiae and in human ribosomes, respectively. An archaeal orthologue of the eukaryotic Nat10/Kre33 enzyme could be responsible for these modifications. Taking into account the small RNA guided mechanism of Nat10/Kre33 modifications of 18S rRNA, we hypothesized that small RNA guides would assist Nat10/Kre33 in archaea. Small RNAs would also explain the non-systematic modification of CCG sequences. Finally, because conservation of the corresponding CCG sequences in archaean 16S rRNA sequences is not obvious, the presence of N4-acetylations in other archaea has to be studied in each case.

**Structure of the mRNA.** An mRNA corresponding to the start region of the highly expressed elongation factor eEF1A from *P. abyssi* was used in IC2. It contains a strong SD sequence with a spacing of 10 nucleotides to the AUG start codon (Fig. 4a and ref. 37). The SD duplex is extended to 9 nucleotides and is not obvious (http://www.rna.icmb.utexas.edu/), the presence of N4-acetylations in other archaea has to be studied in each case.

The initiator tRNA is stably bound to the P site. In IC2A and B, the initiator tRNA is stably bound to the P site and base paired with the mRNA AUG start codon. The interactions of the initiator tRNA within the P site will hereafter be described from the IC2B structure because of its higher resolution.

A set of rRNA modifications contribute to the stabilization of the codon-anticodon duplex (Fig. 3a, b) as previously observed in bacteria and eukaryotes. The base pair C34:G3 of the 3′-end of the 16S rRNA (Fig. 4 and Supplementary Fig. 7b). The SD helix is positioned in the exit chamber delineated by uS11, eS3, and h26 on one side and by uS7, eS28, h28, and h37, on the other side. The SD duplex is extended to 9 nucleotides and is not obvious (http://www.rna.icmb.utexas.edu/), the presence of N4-acetylations in other archaea has to be studied in each case.

**Structure of the mRNA.** An mRNA corresponding to the start region of the highly expressed elongation factor eEF1A from *P. abyssi* was used in IC2. It contains a strong SD sequence with a spacing of 10 nucleotides to the AUG start codon (Fig. 4a and ref. 37). The SD duplex is extended to 9 nucleotides and is not obvious (http://www.rna.icmb.utexas.edu/), the presence of N4-acetylations in other archaea has to be studied in each case.

The initiator tRNA is stably bound to the P site. In IC2A and B, the initiator tRNA is stably bound to the P site and base paired with the mRNA AUG start codon. The interactions of the initiator tRNA within the P site will hereafter be described from the IC2B structure because of its higher resolution.

A set of rRNA modifications contribute to the stabilization of the codon-anticodon duplex (Fig. 3a, b) as previously observed in bacteria and eukaryotes. The base pair C34:G3 of the 3′-end of the 16S rRNA (Fig. 4 and Supplementary Fig. 7b). The SD helix is positioned in the exit chamber delineated by uS11, eS3, and h26 on one side and by uS7, eS28, h28, and h37, on the other side. The SD duplex is extended to 9 nucleotides and is not obvious (http://www.rna.icmb.utexas.edu/), the presence of N4-acetylations in other archaea has to be studied in each case.

The initiator tRNA is stably bound to the P site. In IC2A and B, the initiator tRNA is stably bound to the P site and base paired with the mRNA AUG start codon. The interactions of the initiator tRNA within the P site will hereafter be described from the IC2B structure because of its higher resolution.

A set of rRNA modifications contribute to the stabilization of the codon-anticodon duplex (Fig. 3a, b) as previously observed in bacteria and eukaryotes. The base pair C34:G3 of the 3′-end of the 16S rRNA (Fig. 4 and Supplementary Fig. 7b). The SD helix is positioned in the exit chamber delineated by uS11, eS3, and h26 on one side and by uS7, eS28, h28, and h37, on the other side. The SD duplex is extended to 9 nucleotides and is not obvious (http://www.rna.icmb.utexas.edu/), the presence of N4-acetylations in other archaea has to be studied in each case.

The initiator tRNA is stably bound to the P site. In IC2A and B, the initiator tRNA is stably bound to the P site and base paired with the mRNA AUG start codon. The interactions of the initiator tRNA within the P site will hereafter be described from the IC2B structure because of its higher resolution.

A set of rRNA modifications contribute to the stabilization of the codon-anticodon duplex (Fig. 3a, b) as previously observed in bacteria and eukaryotes. The base pair C34:G3 of the 3′-end of the 16S rRNA (Fig. 4 and Supplementary Fig. 7b). The SD helix is positioned in the exit chamber delineated by uS11, eS3, and h26 on one side and by uS7, eS28, h28, and h37, on the other side. The SD duplex is extended to 9 nucleotides and is not obvious (http://www.rna.icmb.utexas.edu/), the presence of N4-acetylations in other archaea has to be studied in each case.
Fig. 4 Structure of the mRNA and comparison of the mRNA exit channels in the three domains of life. 

**a** Surface representation of IC2B mRNA in blue. The anti-SD sequence is shown in orange sticks. The view emphasizes the sharp turn of the mRNA downstream from $U_{-4}$ and the kink between adjacent codons at the A, P, and E sites. The $\beta$-hairpin of uS7 is in yellow. Magnesium ions are shown in green spheres.

**b** Surface representation of the mRNA exit channel in IC2B. Ribosomal proteins and rRNA helices are labeled. mRNA is in blue and the anti-SD sequence is in orange. The color code is the same for the four **b–e** views. The four structures are shown in the same orientation.

**c** Surface representation of *T. thermophilus* 30S bound to an mRNA allowing a “free choice” of the start codon (PDB: 4V4Y [83] Supplementary Table 3).

**d** Surface representation of *K. lactis* 40S bound to a 49-nt unstructured mRNA (PDB: 6FYX [17]).

**e** Surface representation of *T. thermophilus* 30S bound to an mRNA containing the GGAGGUAAAAAUG sequence (PDB: 5LMQ [5]).
Fig. 5 Steps of translation initiation in *P. abyssi*. Surface representation of successive *P. abyssi* translation initiation complexes. ICO-PREMOTE and IC1-PIN are from ref. 12. tRNA accommodation, h44 adjustments and aIF2 release from h44 are associated with the release of aIF1. After start codon recognition, aIF1 is released and the initiator tRNA is fully accommodated. Release of aIF1 would cause both Pi release from aIF2 and h44 adjustments leading to irreversible aIF2 release (see discussion).

In IC2A and B, the anticodon stem of the tRNA is tightly bound to the P site partly thanks to interactions with the C-terminal tails of the universal proteins uS9, uS13 and uS19. The role of uS9 C-tail in fidelity was previously shown by studies with bacterial18–61 and yeast systems62,63. Moreover, in eukaryotes, uS9 was shown to favor the recruitment of the TC on the ribosome64. The present archaean structure suggests that uS9 tail is universally related to fidelity. The C-tail of uS13 was only modeled in a bacterial IC51 and the C-tail of uS19 was, to our knowledge, never observed in initiation complexes before this study. In the present structure, uS13 and uS19 interact with G30 of the second base pair of the almost universally conserved three GC base pairs in the initiator tRNA anticodon stem (Fig. 3a, b and Supplementary Fig. 7c). The present observations in the archaean system agree with previous studies in *E. coli* showing that the central GC pair and particularly base G30 was the most crucial nucleotide for translation fidelity64–66.

Interestingly, several studies in yeast identified allosteric information pathways connecting functional centers in the large ribosomal subunit (LSU) to the decoding center in the SSU through the B1a and B1b/c intersubunit bridges67,68. In eukaryotes and archaea, uS13 participates in B1b/c bridge and uS19 is part of the B1a bridge. One can therefore imagine that some molecular information is relayed through these two proteins to facilitate LSU joining after the accommodated state of the initiator tRNA in the P site has been sensed by the C-tails of uS13 and uS19.

Around the P site, we observed a series of rRNA modifications. In bacteria, m2G966 and m5C967 have been shown to participate in fine-tuning of initiation58,69. The variations in the three domains of life of the residue corresponding to *E. coli* 16S rRNA-G966 and of its modifications (Table 4) are likely related to evolution of translation initiation. eL41 is connected to the network of rRNA modifications close to the P site by the bulge loop between h44 and h45. Importantly, the C-terminus of eL41 also contacts h44 and aIF2yDIII. eL41 may therefore relay some structural information from the P site to the yDIII binding site and participate in the control of aIF2 release after start codon recognition. The question of the phylogenetic conservation of eL41 is rather puzzling. As evidenced here, archaean representatives contain 37 residues (Supplementary Fig. 6c). However, eL41 has not been identified in all archaean genomes5. One possibility is that its identification has been hampered by the reduced size of the protein. In higher eukaryotes, eL41 has been annotated as a 25 aminoacid residues protein named eL41delity64.

In eukaryotes, it was shown that full release of eIF2 is linked to the release of Pi coming from GTP hydrolysis on eIF256. In archaea, we previously described possible contacts between the N-terminal domain of aIF1 and the switch regions controlling the nucleotide state of aIF2y12. These contacts may link aIF1 release to Pi release. Because we used GDPNP in IC2 complex preparation, aIF2 is not fully released and contacts between aIF2 and the acceptor stem of the initiator tRNA still exist. Motion of h44 close to the P site may be caused by aIF1 release (Supplementary Fig. 12), as previously proposed in eukaryotes18,19. Readjustments of the position of h44 in the bulge region could explain the relaxation of the contacts between this helix and yDIII. Both these movements and the release of contacts between aIF1 and aIF2y observed in IC0 and IC1 would explain how aIF2 is detached from the ribosome after start codon recognition and aIF1 release. Interestingly, contacts between h44 and other translation factors were observed, such as those involving ABCE1 during recycling57. Therefore, conformational adjustments of h44 may be a general mechanism controlling binding and release of factors during the translation cycle.
bS22 conserved in actinobacteria located at the eL41 binding site. Finally, in human and yeast mitochondria, the N-terminal part of the mitochondria-specific m538 protein occupies this position. m538 is proposed to be related to translation of mitochondrial mRNAs lacking typical SD sequences. Overall, the whole data indicate that variability of eL41 might be related to evolution of translation initiation.

The present structure was also compared to the structure of a Kluyveromyces lactis initiation complex in which the N-terminal domain of eIF5 is found at the location vacated by eIF1, in front of the anticodon stem of the initiator tRNA. This Kl-PI-C structure represents a step of initiation occurring after start codon recognition. The initiator tRNA is stabilized by interactions with eIF5-ND and the N-terminal extension of eIF1A. As compared to the present structure, the initiator tRNA adopts a position slightly more tilted towards the head of the SSU. Interestingly, eIF1A is more deeply bound to the A site and its N-terminal extension occupies the position observed here for the C-terminal tail of uS19. Moreover in Kl-PI-C, the C-terminal tail of uS13 is not visible. According to what is observed in IC2, a step following the one illustrated by the Kl-PI-C structure may be a repositioning of eIF1A and a relocation of the C-terminal tails of uS13 and uS19 to further stabilize the initiator tRNA in its accommodated state preceding eIF5B loading.

In bacteria and in some archaea, it has been demonstrated that the SD sequence played an important role in the formation of the IC by base-pairing with the anti-SD sequence at the 3′ end of 16S rRNA. In eukaryotes however, canonical translation involves scanning of the PIC searching for AUG in an optimal context. The current structure highlights similarities and differences in the exit mRNA channel in the three domains of life. The present position of the SD duplex in the chamber corresponds to the archaeal domain. In this view, it is notable that eS26 is absent in the archaeal exit channel with the bacterial one shows that bS6 and bS18 are found in place of eS3. eS28 is absent in bacteria and uS2 possesses a supplementary C-terminal domain, at the location of eS17. Therefore, archaeal and bacterial exit channels appear as two structural solutions for binding the SD duplex. The spacing between the AUG codon and the SD sequence changes the position of the duplex in the chamber, probably explaining how it influences translation initiation efficiency.

The present study of an archaeal initiation complex fills a gap in high-resolution structures of ribosomes representative of the three domains of life. It provides new structural information useful in support of evolution models based on sequence and experimental data. Universal elements located in the SSU common core and in the decoding center ensure stabilization of universal features of the initiation tRNA. The mRNA exit channel that does not map a part of the SSU core indeed shows evolution related to domain-specificities for mRNA binding. On the other hand, the present study also demonstrates the occurrence of a large number of N3-acetylatedcytidines in the rRNA of a hyperthermophilic organism. N3-acetylation of cytidines is of increasing interest since its discovery in mRNA coding sequences where it promotes stability and translation efficiency. Our study shows that the targeted sequence in the 16S rRNA from P. abyssi is systematically CaC. This rule also applies to the acG identified to date in 18S rRNA. In this view, our findings also bring information important to understand a possible general mechanism of acG modification.

### Methods

#### IC2 complex preparation and cryo-EM analysis

The strategy used to prepare the IC2 complex was adapted from a previously described one. In brief, archaeal 30S subunits from P. abyssi (Pa-30S) were purified and mixed with IFs Pa-aIF1A, the ternary complex Pa-aIF2:GDPNP:Met-tRNA, and a synthetic 26 nucleotide-long mRNA corresponding to the natural start region of the mRNA encoding the elongation factor aEF1A from P. abyssi (AC−17')UUGAGGAAGU−GUAUUAAAC−1(U5GCAAGAAG−1)13. The complex was purified by affinity chromatography (TALON, Clontech) using N-terminally tagged versions of Pa-30S and Pa-aIF2A. An excess of mRNA, aIF1A and TC was added before dilution and spotting onto Quantifoil R2/2 grids with an extra 3 nm carbon layer (Quantifoil R2/2, FEI Titan Krios microscope operated at 300 kV at the eBIC center, Diamond Light Source, England (Table 1)).

#### Data processing

Image processing was performed with RELION 2.1. A total of 2174 images was selected for the autopicking tool. Several rounds of 2D classification and an initial 3D refinement with 330 K particles gave a 3.3 Å resolution density map. While the density for the 30S subunit showed high-resolution details, the density allocated to aIF2 and aIF1A was less resolved. Several rounds of 3D classification and particle polishing gave an intermediate density at 3.2 Å resolution with 218 K particles but the bound factors remained poorly resolved. Density subtraction of different parts of the complex gave final maps of the 30S head alone, the 30S head with associated factors and 30S body alone at 3.1, 3.7, and 3.2 Å resolution, respectively (Supplementary Fig. 1). Further 3D classification on the aIF2 binding region gave 7.2, 3.4, and 5.6 Å resolution. The three classes obtained correspond to different positions of aIF2 relative to the 30S and were called IC2A, IC2B, and IC2C (Supplementary Fig. 1). Because very weak electron density for aEF1A was observed in IC2C, this class was considered as representative of an IC2 conformation and was therefore not further refined. Finally, the particles belonging to the IC2A and IC2B classes were re-extracted from raw images and reprocessed to obtain final density maps for IC2A and IC2B complexes at 4.2 Å (34 K particles) and 3.3 Å resolution (142 K particles), respectively, according to the 0.143 FSC gold-standard criterion (Supplementary Fig. 2). Local resolution estimation was carried out using the program RESMIP.

#### 30S model building

A full atomic model of the P. abyssi 30S has been built using an iterative approach including sequence alignment with RNA and proteins of known structure (PDBe: 4V61, 5BJ3, Table 1) rigid body fitting in...
CHIMERA86, real space refinement and geometry regularization in COOT91 combined with real space refinement in PHENIX. All model components (rRNA, ribosomal proteins) were first fitted in the map densities of the 30S head or body alone with simultaneous optimization of stereochromic properties and correction of intra- and intermolecular steric clashes. Secondary structure restraints for further refinement were determined directly from the model using the web interface of 3DNA-DSSR82 and recalculated at each round of refinement. For the three RNAs present in these models, hydrogen-bonding, base-pair and stacking restraints were applied. Adequate restraint files were constructed for modified nucleotides. The high-resolution limit was set during refinement to match the nominal resolution obtained by postprocessing in RELION. The initial structure obtained was then subjected to rigid body refinement as implemented in philex-rr in the IC2B density map and further refined. The crystallographic structure of aIF1A from P. abyssi (PDB: 4MNO12) was manually rigid-body fitted in the cryo-EM map with COOT. Moreover, because the quality of the electron density of aIF2 did not allow side-chain fitting, we used the crystallographic structure of aIF2 from S. solfataricus (PDBs: 4RD493, 3V1155, 2QMU94) to perform rigid body fitting even if aIF2 from P. abyssi was used during complex preparation. The model of the initiator tRNA was fully reconstructed. The anticonid stem-loop region was very well defined. In contrast, according to the higher mobility of the solvent side of the aIF2:tRNA complex, the electron density of the acceptor helix was weaker. The same was observed for the anticodon loop region according to the tRNA density map. The full IC2A and IC2B models were then subjected to cycles of systematic inspection and manual corrections using COOT followed by refinement in PHENIX. Finally, the entire structure was validated using MOLPROBITY95 and correlation coefficients were calculated (Supplementary Table 1). Various programs from CCP-EM were also used throughout the study. The final refinement statistics are provided in Table 1 and in Supplementary Table 1.

rRNA modifications and ribosomal proteins. The full atomic model of the P. abyssi 30S allowed us to identify the N-terminal extension of eL14 and an archaeal version of eS21, as described in the Results section. rRNA modifications visible in the cryo-EM maps have been modeled taking into account information coming from high-resolution cryo-EM and X-ray structures82–85, RNA modifications databases (http://mods.rna.albany.edu/, http://modomics.genesilico.pl/), and literature. The presence of the rRNA modifications was then verified using mass spectrometry as well as primer extension analysis for N4-acetylcytidines. A total of 44 rRNA modifications have been identified including 34 N4-acetylcytidines (Tables 2 and 3).

Primer extension analysis of N4-acetylcytidines. 16S rRNA was prepared from purified P. abyssi 30S subunits84 using a standard phenol–ethanol extraction protocol. 16S rRNA reduction was performed as described83. Total, 3 µg of rRNA was incubated with either sodium borohydride (100 mM in H2O) or water (control sample) in a final reaction volume of 50 µL. Samples were incubated 60 min at 37 °C, quenched with 7.5 µL of 1 M HCl and neutralized with 7.5 µL of 1 M Tris-HCl pH 8.0. Reactions were then performed at 95 °C for 5 min and cooled directly on ice. After annealing, 1.3 µL of 100 mM dNTPs (dATP, dCTP, dGTP, dTTP) were added and incubated in a final volume of 10 µL. Reactions were then heated at 5 min at 65 °C and cooled directly on ice. After annealing, 1.3 µL of 100 mM Mg Acetate and 0.75 µL of dNTPs (3.75 mM each) were added and reverse transcription was started by adding 1 µL of AMV reverse transcriptase (3 U/µL, Promega). After 15 min incubation at 51 °C, the reaction was stopped with 3 µL of a solution containing 95% formamide and blue dextran. Reverse transcripns were analyzed on a Licor 4200 DNA sequencer. Sequencing reactions made from a PCR amplified sample of the 16S rRNA and the RT primers were loaded in order to identify the stop positions. Typical experiments are shown in Fig. 2b and in Supplementary Fig. 11.

Mass spectrometry analysis of hydrolyzed 16S rRNA. Fully hydrolyzed 16S rRNA suitable for mass spectrometry analysis was prepared as described27. Samples were then diluted in 0.1% formic acid (FA) prior to analysis. Chromatographic grade solvents (99.99% purity), acetonitrile (MeOH) and FA, were purchased from Sigma Aldrich. LC-HRMS analyses were performed with the timsTOF mass spectrometer coupled with an Elute HPLC system (Bruker Daltonics, Bremen, Germany). The sample (10 µL, 2.5 µg digested 16S rRNA) was injected and separated on an Atlantis T3 column (3 µm, 150 × 2.1 mm; Waters, Saint Quentin, France). The effluent was introduced at a flow rate of 0.2 ml·min⁻¹ into the interface with a gradient increasing from 10% of solvent B to 50% in 6 min to achieve 70% at 8 min (A: water with 0.1% FA; B: methanol with 0.1% FA). From 8 min to 12 min, the percentage of solvent increased up to 90% of B. The flow was then set to 10% of B for the last 6 min. Electrospray ionization was operated in the positive ion mode. Capillary and end plate voltages were set at −4.5 and −0.5 kV, respectively. Nitrogen was used as the nebulizer and drying gas at 2 bar and 8 L/min, respectively, with a drying temperature of 220 °C. In MS/MS experiments, the precursor ion was selected with an isolation window of 1 Da and the collision-induced dissociation was performed using collision energies (eV) ranging from 7 to 25 eV. Tuning mix (Agilent, France) was used for calibration. The elemental compositions of all ions were determined with the instrument software Data analysis; the precision of mass measurement was better than 3 ppm. All nucleosides have been characterized by both molecular (MH+) and fragment ions (BH+) masses (Supplementary Table 2). The only exception was ac4Gm for which only the H+ ion was identified.

Statistics and reproducibility. The cryo-EM data reported here come from a single experiment and a single grid was used for data collection. Individual images with either bad ice, as shown by visual inspection, or too much motion or astigmatism as shown by power spectra, were excluded from the dataset. Data collection, processing and refinement statistics (see Table 1) were calculated in RELION88, RESMA89, and PHENIX16.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The EMDataBank accession numbers for the EM maps reported in this paper are EMD-10320 (IC2A), EMD-10322 (IC2B), EMD-10323 (IC2 body) and EMD-10324 (IC2 head). The coordinates of the models fitted in the maps have been deposited in the Protein Data Bank (PDB: 6SW9, IC2A, PDB: 6SWC, IC2B, PDB: 6SWD, IC2 body, PDB: 6SWE, IC2 head).

Received: 27 September 2019; Accepted: 17 January 2020; Published online: 06 February 2020

References
1. Rodnina, M. V. Translation in prokaryotes. Cold Spring Harb. Perspect. Biol. 10, a032664 (2018).
2. Hershey, J. W., Sonenberg, N. & Mathews, M. B. Principles of translational control: an overview. Cold Spring Harb. Perspect. Biol. 4, a011528 (2012).
3. Hinnebusch, A. G. & Lorsch, J. R. The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb. Perspect. Biol. 4, a011544 (2012).
4. Pestova, T. V. et al. Molecular mechanisms of translation initiation in eukaryotes. Proc. Natl Acad. Sci. USA 98, 7029–7036 (2001).
5. Kyriides, N. C. & Woese, C. R. Universally conserved translation initiation factors. Proc. Natl Acad. Sci. USA 95, 224–228 (1998).
6. Kyriides, N. C. & Woese, C. R. Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families. Proc. Natl Acad. Sci. USA 95, 3726–3730 (1998).
7. Lecompte, O., Ripp, R., Thierry, J. C., Moras, D. & Poch, O. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res. 30, 5382–5390 (2002).
8. Yutin, N., Puigbo, P., Koonin, E. V. & Wolf, Y. I. Phylogenomics of prokaryotic ribosomal proteins. PLoS One 7, e39792 (2012).
9. Ban, N. et al. A new system for naming ribosomal proteins. Curr. Opin. Struct. Biol. 24, 165–169 (2014).
10. Schmitt, E., Couroux, P. D., Monestier, A., Dubiez, E. & Mechulam, Y. Start codon recognition in eukaryotic and archaeal translation initiation: a common structural core. Int. J. Mol. Sci. 20, 939 (2019).
11. Schmitt, E., Naveau, M. & Mechulam, Y. Eukaryotic and archaeal translation initiation factor 2: a heterotrimeric tRNA carrier. FEBS Lett. 584, 405–412 (2010).
12. Coureux, P. D. et al. Cryo-EM study of start codon selection during archaeal translation initiation. Nat. Commun. 7, 13366 (2016).
13. Monestier, A. et al. The structure of an E. coli tRNAmetal A1-U72 variant shows an unusual conformation of the A1-U72 base pair. RNA 23, 673–682 (2017).
14. Monestier, A., Lazennec-Schurdevin, C., Couroux, P. D., Mechulam, Y. & Schmitt, E. Role of eIF1 in Pyrococcus abyssi translation initiation. Nucleic Acids Res. 46, 11061–11074 (2018).
15. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. elife 7, e42186 (2018).
16. Afonine, P. V., et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol. Crystallogr. 68, 352–367 (2012).
17. Llacer, J. I. et al. Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. Elife 7, e39273 (2018).
18. Lomakin, I. B. & Steitz, T. A. The initiation of mammalian protein synthesis and mrNA scanning mechanism. Nature 500, 307–311 (2013).
19. Weisgerber, M. Vogel-Hoffmann, F. Rabi, J. L. and Ban, N. The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF3 and eIFIA. Nat. Struct. Mol. Biol. 20, 1015–1017 (2013).
20. Armache, J. P. et al. Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. Nucleic Acids Res., 41, 12329–1239 (2013).
21. Ben-Shem, A. et al. The structure of the eukaryotic ribosome at 3.0 A resolution. Science 334, 1524–1529 (2011).
22. Acosta-Reyes, F., Neupane, R., Frank, J. A. Oohio and Fernandez, I. S. A. Oohio The Israeli acute parallel virus IRES captures host ribosomes by mimicking a ribosomal state with hybrid RNAs. EMBO J. 38, e102226 (2019).
23. Shrivastava, S. K. et al. Shaping the ribosome: The influence of RNA modification on eukaryotic ribosome biogenesis and function. RNA Biol. 14, 1138–1152 (2017).
24. Natchiar, S. K., Myiasnikov, A. G., Krazat, H., Hazemann, I. & Klholaz, B. P. Visualization of chemical modifications in the human 80S ribosome structure. Nature 551, 472–477 (2017).
25. Fischer, N. et al. Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM. Nature 520, 567–570 (2015).
26. Polikanov, Y. S., Melnikov, S. V., Soll, D. & Steitz, T. A. Structural insights into the role of RNA modifications in protein synthesis and ribosome assembly. Nat. Struct. Mol. Biol. 22, 342–344 (2015).
27. Ling, W., Weng, Y., Qi, Y., Zhang, X. & Frank, J. A. The crystal structure of the ribosome for Haloferax volcanii. J. Biol. Chem. 275, 24484–24489 (2000).
28. O’Farrell, H. C., Pulicherla, N., Desai, P. M. & Rife, J. P. Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved through evolution. Nat. Struct. Mol. Biol. 17, 725–733 (2010).
29. Seistrup, K. H. et al. Bypassing RNA methylation by RsmA/Dim1 during ribosome maturation in the hyperthermophile Haloarcula marismortui. EMBO J. 34, 3335–3346 (2015).
30. Seistrup, K. H. et al. Bypassing RNA methylations in 16S ribosomal RNA from Methanococcus jannaschii. Nucleic Acids Res. 43, 7352–7363 (2015).
31. Seistrup, K. H. et al. Nucleic acid modification by RsmA/Dim1 during ribosome maturation in the hyperthermophile Haloarcula marismortui. EMBO J. 34, 3335–3346 (2015).
32. Schmid, A. M. et al. Ribosomal protein uS19 mutants reveal its role in near-cognate translation initiation complex. Mol. Cell 59, 399–412 (2015).
33. Schmitt, E. et al. Structure of the ternary initiation complex eIF2-GDPNP-methionylated initiator tRNA. Nat. Struct. Mol. Biol. 19, 450–454 (2012).
34. Algire, M. A., Maag, D. & Lorsch, J. R. PI release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. Mol. Cell 20, 231–242 (2005).
35. Polikanov, Y. S., Tampé, R. and Cate, J. H. Ribosome maturation in the hyperthermophilic archaeon Pyrodictium occultum. Proc. Natl Acad. Sci. USA 101, 12439–12443 (2004).
36. Noller, H. F., Hoang, L. and Fredrick, K. The 30S ribosomal P-site of human ribosome. FEBS Lett. 579, 855–858 (2005).
37. Jindal, S., Gosh, A., Ismail, A., Singh, N. and Komar, A. A. The role of the uS5/uS16 C-terminal tail of the 50S protein in the fidelity of initiation of translation in Escherichia coli. Nucleic Acids Res. 41, 4963–4975 (2013).
38. Arora, S. et al. 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. 43, 4963–4975 (2013).
39. Shetty, S., Shah, R. A., Chembazhi, U. V., Sah, S. and Varshney, U. Two highly conserved features of bacterial initiator tRNA license them to pass through distinct checkpoints in translation initiation. Nucleic Acids Res. 45, 2040–2050 (2015).
40. Ayyub, S. A. et al. Evolution of the translational machinery optimizes initiation with unusual initiator tRNAs and initiation codons in mycoplasmas. RNA Biol. 15, 70–80 (2018).
41. Samhita, L. Shetty, S. and Varshney, U. Unconventional initiator tRNAs sustain Escherichia coli. Proc. Natl Acad. Sci. USA 109, 13058–13063 (2012).
42. Bowen, A. M. et al. Ribosomal protein uS19 mutants reveal its role in coordinating ribosome structure and function. Translation 3, e1177703 (2015).
43. Rhodin, M. H. and Dinman, J. D. An extensive network of information flow through the B1/bc intersubunit bridge of the yeast ribosome. PLoS One 6, e180048 (2011).
44. Burakovskiy, D. E. et al. Implications for stability of m2G966/m5C967 in 16S RNA on bacterial fitness and translation initiation. Nucleic Acids Res. 40, 7885–7895 (2012).
45. Wong, W. et al. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. Elife 3, e03080 (2014).
46. Hentschel, J. et al. The complete structure of the Mycobacterium smegmatis 70S ribosome. Cell Rep. 20, 149–160 (2017).
47. Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. and Ramakrishnan, V. Ribosome. The structure of the human mitochondrial ribosome. Science 348, 477–481 (2015).
48. Greber, B. J. et al. Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. Science 348, 303–308 (2015).
74. Mays, J. N. et al. The mitoribosome-specific protein mS38 is preferentially required for synthesis of cytochrome c oxidase subunits. *Nucleic Acids Res.* **47**, 5746–5760 (2019).

75. Condo, I., Ciammaruconi, A., Benelli, D., Ruggero, D. & Londei, P. Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **34**, 377–384 (1999).

76. Osada, Y., Saito, R. & Tomita, M. Analysis of base-pairing potentials between 16S rRNA and 5′ UTR for translation initiation in various prokaryotes. *Bioinformatics* **15**, 578–581 (1999).

77. Saito, R. & Tomita, M. Computer analyses of complete genomes suggest that some archaea bacteria employ both eukaryotic and eubacterial mechanisms in translation initiation. *Gene* **238**, 79–83 (1999).

78. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).

79. Archer, S. K., Shirokikh, N. E., Beilharz, T. H. & Preiss, T. Dynamics of 16S rRNA and 5′ UTR for translation initiation in various prokaryotes. *Bioinformatics* **15**, 578–581 (1999).

80. Ferretti, M. B., Ghalei, H., Ward, E. A., Potts, E. L. & Karbstein, K. Rps26 directs mRNA-specific translation by recognition of Kozak sequence elements. *Nat. Struct. Mol. Biol.* **24**, 790–797 (2017).

81. Schutz, S. et al. Molecular basis for disassembly of an importin:ribosomal protein complex by the escortin Tar2. *Nat. Commun.* **9**, 3699 (2018).

82. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).

83. Yusupova, G., Jenner, L., Bees, B., Moras, D. & Yusupov, M. Structural basis for messenger RNA movement on the ribosome. *Nature* **444**, 391–394 (2006).

84. Kamishin, T. et al. A snapshot of the 30S ribosomal subunit capturing mRNA via the Shine-Dalgarno interaction. *Structure* **15**, 289–297 (2007).

85. Bernier, C. R., Petrov, A. S., Kovacs, N. A., Penev, P. I. & Williams, L. D. Translation: the universal structural core of life. *Mol. Biol. Evol.* **35**, 2063–2076 (2018).

86. Arango, D. et al. Acetylation of cytidine in mRNA promotes translation efficiency. *Cell* **175**, 1872–1886 (2018).

87. Dominissini, D. & Rechavi, G. N(4)-acetylation of cytidine in mRNA by NAT10 regulates stability and translation. *Cell* **175**, 1725–1727 (2018).

88. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

89. Kucukelbir, A., Sigworth, F. J., Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63 (2013).

90. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

91. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

92. Zheng, G., Lu, X.-J. & Olson, W. Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **37**, W240–W246 (2009).

93. Dubie, E., Aleksandrov, A., Lazenec-Schurdevin, C., Mechulam, Y. & Schmitt, E. Identification of a second GTP-bound magnesium ion in archaeal initiation factor 2. *Nucleic Acids Res.* **43**, 2946–2957 (2015).

94. Yatime, L., Mechulam, Y., Blanquet, S. & Schmitt, E. Structure of an archaeal heterotrimetric initiation factor 2 reveals a nucleotide state between the GTP and the GDP states. *Proc. Natl Acad. Sci. USA* **104**, 18445–18450 (2007).

95. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

96. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Biol. Crystallogr.* **73**, 469–477 (2017).

97. Gracin, P. F. Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol.* **193**, 782–790 (1990).

98. Schrodinger, L. The PyMOL Molecular Graphics System, Version 2.0. (2017).

99. Tuoka, M. et al. The complete chemical structure of *Saccharomyces cerevisiae* rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res.* **44**, 8951–8961 (2016).

Acknowledgements

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). We thank Alistair Siebert for data collection at the electron biological cryo-imaging facility (eBIC, Diamond Light Source, UK). We thank Thomas Gaillard for helpful discussions.

Author contributions

P.D.C., Y.M. and E.S. designed the research; P.D.C., C.L.S., S.B., Y.M. and E.S. contributed to the experimental work; P.D.C., Y.M. and E.S. analyzed the data and wrote the paper. P.D.C., C.L.S., S.B., Y.M. and E.S. reviewed and edited the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s42003-020-0780-0.

Correspondence and requests for materials should be addressed to P.-D.C. or E.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.