Cryo-EM study of start codon selection during archaeal translation initiation

Pierre-Damien Coureux, Christine Lazennec-Schurdevin, Auriane Monestier, Eric Larquet, Lionel Cladie`re, Bruno P. Klaholz, Emmanuelle Schmitt, & Yves Mechulam

Eukaryotic and archaeal translation initiation complexes have a common structural core comprising e/aIF1, e/aIF1A, the ternary complex (TC, e/aIF2-GTP-Met-tRNA\textsuperscript{Met}) and mRNA bound to the small ribosomal subunit. e/aIF2 plays a crucial role in this process but how this factor controls start codon selection remains unclear. Here, we present cryo-EM structures of the full archaeal 30S initiation complex showing two conformational states of the TC. In the first state, the TC is bound to the ribosome in a relaxed conformation with the tRNA oriented out of the P site. In the second state, the tRNA is accommodated within the peptidyl (P) site and the TC becomes constrained. This constraint is compensated by codon/anticodon base pairing, whereas in the absence of a start codon, aIF2 contributes to swing out the tRNA. This spring force concept highlights a mechanism of codon/anticodon probing by the initiator tRNA directly assisted by aIF2.
In all living cells, translation initiation allows accurate selection of the initiation codon on a messenger RNA, which then defines the reading frame of the protein to be synthesized. In eukaryotes, this process is highly complicated and is the target of many regulations. The mechanism begins with the assembly of the ternary complex (TC), consisting of initiator methionyl-transfer RNA (Met-tRNA) anchored to the GTP-bound form of the heterotrimeric initiation factor 2 (eIF2). The TC then binds to the small ribosomal subunit (40S) in the presence of two small initiation factors eIF1 and eIF1A. TC binding is assisted by eIF5 (ref. 4) and by the multifactor eIF3 (ref. 5). eIF5 favours reversible GTP hydrolysis on eIF2. In the presence of factors belonging to the eIF4 family, the pre-initiation ribosomal complex (PIC) is recruited near the 5′-capped end of the messenger RNA (mRNA) and scans the mRNA to search for an AUG codon in an appropriate context. AUG recognition stops scanning, leads to factor release and to the assembly of an elongation-proficient 80S complex through large subunit joining, with the help of eIF5B. A large amount of data from genetics and biochemical studies has been collected to decipher the molecular steps ensuring the correct selection of the AUG codon on a messenger RNA. In particular, isolation of yeast mutants able to initiate translation on a non-AUG codon highlighted the key roles of eIF1 and eIF2 in the selection of the start codon. In the current model, AUG recognition triggers a conformational change of the 40S subunit, which leads to eIF1 departure, and then to the release of inorganic phosphate (Pi) from eIF2 (ref. 7). In its GDP-bound form, eIF2 loses its affinity for the initiator tRNA and therefore leaves the 40S subunit (ref. 11,12). Recent structural descriptions of eukaryotic translation initiation complexes (ICs) have highlighted some of the structural changes that lead to a closed state with the initiator tRNA base-paired to the start codon (refs. 13–16). These structures have shown how accommodation of the initiator tRNA at the P site, from a POUT to a PIN position (ref. 7), is stabilized by eIF1A and is accompanied by a closure of the mRNA channel in the 40S subunit as well as by the destabilization of eIF1 binding. However, these structures do not give any clue on how the departure of eIF1 triggers the release of the Pi group from eIF2 and thereby the release of eIF2-GDP. Moreover, the absence of contact between the central GTP-binding γ subunit of eIF2 and the ribosome remains intriguing according to hydroxyl radical probing experiments that suggest interaction of domain III of eIF2γ with h44 in tRNA.

In archaea, long-range scanning does not occur because of the occurrence of Shine-Dalgarno (SD) sequences or of very short 5′-untranslated region on mRNA. However, archaeal translation initiation uses three initiation factors, aIF1, aIF1A and aIF2, homologous to their eukaryotic counterparts. This suggests that the archaeal factors fulfill similar functions as the eukaryotic factors. Such a parallel relies on the idea that the detection of the start codon in archaea proceeds through a local scanning mechanism within a short mRNA region compatible with maintenance of SD:antiSD pairing. This leads to the definition of a structural core, common to eukaryotes and archaea, that controls translation initiation accuracy. This core is made of the small ribosomal subunit, the mRNA, the Met-tRNA<sup>Met</sup> and the three factors e/aIF1, e/aIF1A and e/aIF2.

To further address the role of e/aIF2, we have analysed the full archaeal 30S initiation complex (30S IC) by cryo-EM. We isolated two conformational states of the 30S IC at 5.34 and 7.5 Å overall resolutions, respectively. All initiation factors are visible, even though at a generally worse resolution. In the first complex, the tRNA is oriented out of the peptidyl (P) site in a standby position (the complex is therefore called IC0–PREMOTE) and the heterotrimeric aIF2 strongly interacts with the 30S subunit and with aIF1. In the second complex, the initiator tRNA is accommodated within the P site (the complex is therefore called IC1–P<sub>IN</sub>) and partially releases interactions with aIF2, reflecting the structural transition from the standby position to the decoding position. Comparison of IC0–PREMOTE and IC1–P<sub>IN</sub> shows that the TC is constrained when the tRNA is engaged in the P site. However, on start codon recognition this constraint is compensated by codon/anticodon base pairing. In the absence of a start codon, aIF2 contributes to swing out the tRNA from the P site. This spring force concept provides a novel view of the role of e/aIF2 in start codon selection. Moreover, we observe contacts between aIF1 and the aIF2γ switch regions known to control its nucleotidic state. This gives a structural framework to explain how aIF1 departure provokes the release of Pi from aIF2. According to previous biochemical and genetic studies in eukaryotic systems, the features of the two novel states of the archaeal core IC are likely to be relevant for eukaryotic translation initiation.

**Results**

**Assembly and structure of the full archaeal IC.** We assembled the full IC using archaeal 30S subunits from *Pyrococcus abyssi* (Pa-30S), initiation factors Pa-aIF1, Pa-aIF1A and Pa-aIF2, the ternary complex (Met-tRNA<sup>Met</sup> A<sub>1</sub>−U<sub>2</sub> variant from *Escherichia coli* (Supplementary Fig. 1a), GDPNP and Pa-aIF2 composed of three subunits α, β and γ and a synthetic 26 nucleotide long mRNA corresponding to the natural start region of the mRNA encoding the elongation factor aEF1A from *P. abyssi*, which contains a strong Shine-Dalgarno sequence (A<sub>17</sub> UUU GAGGAGU UWAA<sub>14</sub> UGCAAAG<sub>19</sub>). The complex was purified by affinity chromatography using N-terminally tagged versions of Pa-aIF1, Pa-aIF2α and Pa-aIF2β (Supplementary Fig. 1b). An excess of mRNA, aIF1A and TC was added before spotting onto the grids for cryo-EM analysis. Cryo-EM images were collected on an FEI Cs-corrected Titan Krios microscope operated at 300 kV (Supplementary Fig. 1c). After image processing and three-dimensional (3D) classification using RELION, the structures of two complexes were identified (Figs 1 and 2 and Supplementary Fig. 2). In one complex, the tRNA is totally out of the P site and is therefore referred to as IC0–PREMOTE. The other structure, featuring a tRNA accommodated in the P site, is named IC1–P<sub>IN</sub>. The two electron density maps were obtained at overall resolutions of 5.34 and 7.5 Å for IC0–PREMOTE and IC1–P<sub>IN</sub>, respectively, according to the 0.143 FSC criterion (Supplementary Fig. 2b, Fig. 2 and Table 1).

As already observed for other ICs, local resolution is best in the core of the 30S subunit and decreases as distance from the 30S increases (Fig. 2 and Table 1).

Model construction was mainly based on the placement of known structures of archaeal homologous components (Supplementary Fig. 3 and ‘Methods’ section). In both eukaryotes and archaea, e/aIF2 is a βγ heterotrimer. Structural data obtained with the archaeal variants of the protein have shown that the heterotrimer contains a rigid unit made of the γ-subunit, domain 3 of the α-subunit (aIF2α-D3) and the N-terminal helix of the β-subunit. Onto this rigid unit are anchored two mobile wings corresponding to domains 1 and 2 of α (aIF2α-D12) and to the main domain of aIF2β (refs 23–26). These structural features were taken into account for aIF2 fitting. In both IC0–PREMOTE and IC1–P<sub>IN</sub>, aIF1A and aIF1 are present in the A site and in front of the P site, respectively, similarly to their eIF1A and eIF1 orthologues in the previous structures of eukaryotic complexes (Fig. 3). The mRNA is clearly visible in the electron density. The interaction of the SD sequence with the anti-SD was modelled and the start codon was positioned within the P site (Supplementary Fig. 4a). The mRNA is not visible from...
In IC0–PREMOTE, all actors of translation initiation are bound to the ribosome but the initiator tRNA is located totally outside of the P site (Fig. 1a,c). The atomic model corresponding to the archaean TC23 could be readily fitted in the electron density. Therefore, on the 30S subunit, the initiator tRNA is bound to aIF2 in a manner similar to that in the free TC, without constraints induced by the ribosome. As compared with the crystal structure, only slight rotations of the rigid unit made of domains 1 and 2 of aIF2 (aIF2x–D12) are observed, accompanied by a minor movement of GDPNP is shown in sticks. This view also highlights contacts between aIF1 and aIF2γ. One contact involves the β20–β21 loop (residues 402–404) of aIF2γ with the aIF1 α2 helix. A second area of density connects the two switch regions of aIF2γ and the N-terminal domain of aIF1. (d) P site region in IC1–PIN. The density map has been low-passed filtered to 12 Å resolution. Note that positioning of the core domain of aIF2γ is only tentative.

Similarly, aIF2γ local resolution is preserved in the IC0–PREMOTE and IC1–PIN complexes. (Fig. 2) Left: surface of IC0–PREMOTE map coloured according to local resolution (see Supplementary Methods) using the colour scale indicated. Right: same for IC1–PIN.

base +5 to +9 at the 3’ extremity, as in the yeast 48S complex (py48S)14; bases –17 and –16 at the 5’ end are not visible.

IC0–PREMOTE shows a new network of interactions around aIF2. In IC0–PREMOTE, all actors of translation initiation are bound to the ribosome but the initiator tRNA is located totally outside of the P site (Fig. 1a,c). The atomic model corresponding to the archaean TC23 could be readily fitted in the electron density. Therefore, on the 30S subunit, the initiator tRNA is bound to aIF2 in a manner similar to that in the free TC, without constraints induced by the ribosome. As compared with the crystal structure, only slight rotations of the rigid unit made of domains 1 and 2 of aIF2 (aIF2x–D12) are observed, accompanied by a minor movement of the tRNA (Supplementary Fig. 3b).

Within the heterotrimeric aIF2, the γ-subunit is tightly bound to the 30S subunit, mainly held by two contact points (Figs 1 and 4). One interaction involves domain III of aIF2γ (aIF2γ–DIII). More precisely, the carboxy (C)-terminal part of the long loop between β16 and β17 as well as the loop between β18 and β19 overhang the top of the h44 bulge (Fig. 4b). Accordingly, a distortion of the helical structure of h44 is observed and the local structure of h44 had to be remodelled (Fig. 4c). Also, the C terminus of ribosomal protein uL41 points towards aIF2γ–DIII (Fig. 4b). A second contact between aIF2 and the 30S subunit involves the long β-hairpin (β10–β11) of aIF2γ–DII and the groove formed by the rRNA h27 loop on one side and by the extremity of h24 on the other side. Notably, this region of the rRNA links the decoding centre and the platform. This contact gives sense to the presence of the β10–β11 hairpin in all e/aIF2γ, whereas it is absent in elongation factors e/αEF1A, despite strong structural homology between the two proteins28. The position of aIF2x–D3 with respect to aIF2γ corresponds to that observed in the TC and in other aIF2 structures26,29,30. aIF2x–D3 is sitting on the platform close to the tip of h23 and h24 bulges. Moreover, the C terminus of aIF2x points towards h24 (Fig. 4b). Interestingly, the length of the e/aIF2x C-terminal tail varies from species to species. In eukaryotes, this tail was shown to negatively influence tRNA
Table 1 | Local resolution of initiation factors.

| Initiation factor | IC0-PRoM | IC1-PIN |
|-------------------|---------|---------|
| alf1A             | 6.0     | 9.3     |
| alf1              | 6.9     | 8.5     |
| tRNA              | 8.7     | 10.0    |
| alf2\-D3-Nterβ    | 8.3     | 15.6    |
| alf2\-D12         | 15.0    | 15.0    |
| alf2β             | ND      | ND      |

ND, not determined. The table shows the local resolutions in Å of the components of the initiation complexes as determined with RELION using a soft mask (with a 15-pixel soft edge) surrounding the region of interest as described.51.

binding, and it was suspected that this negative effect might be cancelled on interaction of the TC with the ribosome. In this context, it is notable that a contact between the C-terminal tail of eIF2\(\gamma\) and h24 can be inferred from the position of alf2\-D3 observed here. The release of such a contact may help the release of eIF2 from the tRNA on start codon recognition.

On the side of alf2\-D12, electron density suggests contacts with ribosomal protein eS1. Notably, in the case of py48S, eIF2\-D12 was located in the E site, close to the \(-2\) and \(-3\) positions of mRNA (Fig. 3c,d). This position is suitable for a participation of eIF2\(\gamma\) in sensing the favourable environment of the correct start codon during mRNA scanning. The different orientation of alf2\-D12 observed here is possibly related to the pre-positioning of the 3OS subunit on the mRNA due to the recognition of the SD sequence.

Finally, the electron density also suggests interaction of the z\(\beta\) domain of alf2β with the D-stem and D-loop of the tRNA (Figs 1c and 4d). Such a position of alf2β with respect to the tRNA and to eIF2\(\gamma\) is similar to that observed in eukaryotic translation ICs.48 (Fig. 3).

Hence, on the whole, the initiator tRNA appears to be wrapped up by the three alf2 subunits leaving the anticodon loop free of interaction (Figs 1 and 4d). Notably, the observed position of the ternary complex on the 3OS subunit prevents uncontrolled tRNA access to the ribosomal E site.

In front of the alf2\-D11 domain, alf1 occupies a position corresponding to that previously described for eIF1 in eukaryotic ICs, on h44 close to the P site (Figs 1 and 3). As compared with the structure of the isolated factor, the conformation of the basic L1 loop had to be remodelled to account for the electron density. The loop comes close to the P site codon on the mRNA, suggesting that alf1 contributes to its stabilization in the mRNA channel. As compared with eukaryotic complexes,13,14,22, a novel feature of the present structure is that alf1 and alf2 interact with each other (Figs 1b and 3). One contact between both initiation factors involves the \(\beta20\)–\(\beta21\) loop of alf2\-D11 (402–404) with the top of the alf1 \(\gamma2\) helix and the following loop 91–95. A supplementary density at the N terminus of alf1 corresponds to a part of the 23 N-terminal residues missing in the alf1 crystal structure, and shows the occurrence of a contact between the N-terminal domain of alf1 and the two switch regions of alf2\(\gamma\).

Finally, the electron density of alf1 in IC0-PRoM is as strong as that observed for the ribosome, reflecting its tight binding (Figs 1 and 2).

alf1A is positioned within the A site as its eIF1A orthologue in eukaryotic PICs.15,16,22 (Fig. 3 and Supplementary Fig. 4b). Interactions between h44 and the tight turn 34–36 are visible. To avoid clashes with alf1A, bases 1447 and 1448 (1492 and 1493 in E. coli numbering) must flip out from h44, as previously observed for eukaryotic complexes.15,16,22. Notably, the channel leading to the P site, corresponding to that occupied by the functionally important N-terminal extremity of eIF1A in the py48S complex, is empty in the present structure (Supplementary Fig. 4b). The N-terminal tail of Pa-alf1A is 22-residues long, 10 residues shorter than that of the eukaryotic factor. Residues 13 to 22 form an elongated peptide with its extremity facing the h32–h34 junction in the 3OS head but the first 12 residues are not visible within the electron density.

IC1-\(\text{P}_{\text{PIN}}\) shows a constrained state of the TC. In IC1-\(\text{P}_{\text{PIN}}\), the initiator tRNA is positioned in the P site (Fig. 1b,d). The density is consistent with the interaction of the anticodon with the start codon on mRNA whereas the 3′ methionylated extremity remains bound to alf2\(\gamma\). The structure of the P site initiator tRNA as observed in the Thermus thermophilus 70S complex could not be readily rigid-body fitted in the density. Indeed a bending of the tRNA molecule at the level of the hinge region was necessary (Fig. 5b and Supplementary Fig. 5c). The acceptor end was also remodelled to fit the 3′ methionyl-adenosine in its pocket on alf2\(\gamma\)-D1 (Supplementary Fig. 6a). Notably, alf2\(\gamma\) moves only very slightly in IC1-\(\text{P}_{\text{PIN}}\) as compared with IC0-PRoM and remains bound to h44 via its domain III (Figs 1 and 3a,b). Hence, in IC1-\(\text{P}_{\text{PIN}}\), the tRNA molecule interacts with alf2 in a manner different from that in free TC or in IC0-PRoM (Fig. 1). Overall, both the tRNA structure and the tRNA:alf2\(\gamma\) interface are constrained in IC1-\(\text{P}_{\text{PIN}}\) (Fig. 5a). Remarkably, induced structural constraints have already been observed during accommodation of the EF1α:tRNA complex in the A site on codon recognition.55,56 (Supplementary Fig. 6). The remodelling of the TC is accompanied by the loosening of some contacts between alf2 and the 3OS subunit such as those involving the long L2 loop of alf2\(\gamma\)-D11 and the alf2\(\gamma\)-D3 domain (Supplementary Fig. 5). Moreover, concomitant with its adjustment within the P site, the contacts of the tRNA with alf2\(\gamma\)-D3 are lost. Accordingly, the electron density of alf2\(\gamma\)-D3

![Figure 3](image)

**Figure 3** | Comparison of archaean and eukaryotic initiation complexes.

(a) IC0-PRoM. (b) IC1-\(\text{P}_{\text{PIN}}\). (c) Eukaryotic open complex (PDB entry 3JAQ; ref. 14). (d) Eukaryotic closed complex (PDB entry 3JAP; ref. 14). Colour code for the initiation factors is the same as that used in Fig. 1 except that h44 is in brown and the small ribosomal subunit in pink. Colour code for eukaryotic factors corresponds to that used for archaean orthologues. The eIF3 subunits are in black. The figure shows that archaean alf1 and alf1A occupy similar positions as their eukaryotic orthologues elf1 and elf1A. In contrast, the positions of the archaean TC in either IC0-PRoM or IC1-\(\text{P}_{\text{PIN}}\) differ from the positions of the eukaryotic TC observed in the eukaryotic structures.
Figure 4 | Contacts between aIF2 and the 30S ribosomal subunit as observed in IC0–PREMOTE. (a) aIF2γ–DIII is in contact with the h44 bulge in 16S RNA. These contacts involve two regions, 351–354 and 372–375, as labelled in green. (b) The view is rotated by ~90° as compared with a to highlight the contact of the L2 loop of aIF2γ–DIII with h24 (723–725; 769–771 in E. coli numbering) and h27 (856–860; 898–902 in E. coli numbering), at the hinge between the decoding centre and the platform. Contacts between aIF2x and the ribosome are also visible (see the text for details). (c) Fit of h44 (pink) into IC0–PREMOTE map is shown. The view shows that the structure of h44 from Pf-30S subunit (blue, PDB entry 4V6U; ref. 48) had to be remodelled to account for the density. A1378 and A1438 (Pf numbering) correspond to A1418 and A1483 in Ec numbering, respectively. (d) View showing how aIF2 subunits wrap the initiator tRNA in IC0–PREMOTE. The cryo-EM map is low-pass filtered to 8.5 Å.

Discussion

Comparison of IC0–PREMOTE and IC1–PIN highlights the structural transition that occurs on start codon recognition. The tRNA accommodation can be described as an isomerization of the IC0–PREMOTE complex whereby the tRNA molecule, surrounded by the α- and β-subunits of aIF2, swings into the P site using the 3′ methionyl-adenosine as a pivot point (Fig. 6 and Supplementary Movie 1). This movement is accompanied by the head rotation allowing locking of the tRNA anticodon stem in the P site. Because the 3′ methionyl-adenosine remains bound to aIF2γ, the establishment of the codon:anticodon interaction induces a structural constraint in the TC. This suggests that aIF2 exerts a spring force on the tRNA adjusted in the P site, which brings back the tRNA in the ‘remote’ conformation if there is no correct start codon in the mRNA available. On the contrary, recognition of the AUG start codon and formation of correct codon–anticodon pairing, would compensate for the spring force on the tRNA, allowing the tRNA to remain longer in the P site thus triggering further events related to start codon recognition (Fig. 6).

In the POUT state described for the eukaryotic IC14,22, the tRNA is already engaged in the P site. The PREMOTE complex observed here was rather unexpected. Therefore, to further probe the existence of the IC0–PREMOTE conformation, we used the S1 nuclease, known to cleave initiator tRNAs in their anticodon loops38,39. Indeed, the PREMOTE state features a tRNA having an anticodon loop expected to be more accessible to nuclease cleavage than that of a tRNA engaged in the P site. We therefore prepared two complexes. The first one corresponds to the complex studied by cryo-EM except that the mRNA contains a CAU codon in place of the AUG start codon, to favour the PREMOTE conformation. The second complex contains the AUG bearing mRNA, but to favour accommodation of the tRNA in the P site, aIF1 was omitted. As shown in Supplementary Fig. 8, the rate of cleavage in the anticodon loop by the S1 nuclease was higher in the former complex than in the latter (2.1 ± 0.07% min−1 versus 1.2 ± 0.08% min−1). This result is
The occurrence of the spring force mechanism in archaea raises the question of whether it could be a general mechanism in eukaryotes during the start codon selection process. Interestingly, the rather slow rate of mRNA scanning during CAP-dependent initiation (5–10 nt s⁻¹; ref. 41) makes it likely that a constrained state of the TC is also pivotal in eukaryotes. Because constrained states are intrinsically less stable, it is possible that the previously described py48S closed complex 14 represents a later step of the process in which elf2 has been released from its constrained position on h44 (Fig. 5c). Indeed, in py48S-closed, the TC is unconstrained and no contact between elf2γ and the ribosome is visible. Here, the use of a hyperthermophilic model organism may have helped trap the constrained complex as an important intermediate. Notably, contacts between elf2 and helix h44, as well as contacts between elf2 and elf1, although not observed in the previously determined structures 13,14,22, would make sense in eukaryotic complexes. As a matter of fact, directed hydroxyl radical probing experiments have suggested close vicinity of elf2γ-DIII and h44 in yeast 18. Moreover, biochemical studies have proposed that elf1 release triggers the release of the PI group from elf2 (ref. 7). In this view, it is remarkable that the regions of contact between elf2γ and alf1, as identified here, are the targets of Sui mutations in Saccharomyces cerevisiae orthologues 8,42,43 suggesting that such interactions may be at the origin of communication between elf2 and elf1 during eukaryotic translation initiation. Taken together, we propose that the constrained Pnin state observed here represents a new key intermediate during start codon search involving alf2-assisted codon probing by the initiator tRNA. This provides a novel view of the central role that e/alf2 plays in start codon selection.

**Methods**

**Growth of P. abyssi and preparation of ribosomal particles.** *P. abyssi* cells were grown under anaerobic conditions 44. Mass cultivations were carried out at 90 °C in a 160 l fermenter (Laboratoire de microbiologie industrielle UMR FARE 614, Reims, France). The cells were collected by centrifugation, aliquots were flash-frozen in liquid nitrogen and stored at −80 °C until further use.

To isolate archaeal ribosomes, 3 g of cell pellets were dissolved in 16 ml of buffer A (10 mM Hepes pH 7.5, 100 mM NH4Cl, 10.5 mM magnesium acetate, 0.1 mM EDTA, 6 mM 2-mercaptoethanol, 0.1 mM PMSEF, 0.1 mM benzamidine) at 4 °C, mixed with 16 ml of glass beads (5 μm, Sigma) and disrupted by using a Vibrogen-Zellmißhle apparatus (Edmund Bühler, Fisher Scientific). After extensive washing of the glass beads and centrifugations, the crude extract (34 ml) was loaded onto a 35 ml sucrose cushion (10 mM Hepes pH 7.5, 1.1 M sucrose, 1 M NH4Cl, 10.5 mM magnesium acetate, 0.1 mM EDTA, 6 mM 2-mercaptoethanol, 0.1 mM PMSEF, 0.1 mM benzamidine). After centrifugation at 235,000g for 19 h at 4 °C, the pellet was dissolved in 3 ml of buffer A and loaded onto a Sephacryl S400 column (11 mm × 60 cm, GE Healthcare). The fractions containing ribosomal particles were pooled, dialysed for 3 h against buffer B (same as buffer A but 2.5 mM magnesium acetate) and then loaded on the top of sucrose gradients (10–30% sucrose in buffer B made using BioComp Gradient Master, Frederinton, Canada). After a 19 h centrifugation at 70,000g (SW32.1 rotor, Beckman), the gradients were fractionated. Fractions containing 30S subunits were pooled and the magnesium concentration was increased to 10.5 mM. After centrifugation (230,000g, 20 h), the pellet was briefly rinsed with buffer A to remove residual sucrose and then dissolved in buffer A. About 400 μl of 30S subunits at 2.7 μM were usually obtained. The 30S subunits obtained using this procedure were quite pure, as shown by gel electrophoresis analyses (Supplementary Fig. 1b). However, a contaminating band was systematically visible on SDS gels. This band was identified as phosphoenolpyruvate synthase by peptide mass fingerprinting. This high molecular weight protein results from the octameric assembly of a 91 kDa monomer.

**Purification of initiation factors and of aminoacyl-tRNAs.** Pa-alf2—Cultures (230 ml) of cells overproducing each of the three subunits (native alf2γ), N-terminally tagged versions of alf2γ and alf2β) were harvested, mixed in 40 ml of buffer C (500 mM NaCl, 10 mM HEPES pH 7.5, 10 mM 2-mercaptoethanol, 0.1 mM EDTA, 0.1 mM PMSEF, 0.1 mM benzamidine) and disrupted by sonication. After centrifugation, the supernatant was heated for 10 min at 80 °C. After pelleting the nonhosterstables proteins by centrifugation, the supernatant was directly loaded on a Q-HiLoad column (10 mm × 4 cm; GE Healthcare) equilibrated in buffer C. The flow-through was recovered and diluted twofold with buffer D (10 mM HEPES pH 7.5, 10 mM 2-mercaptoethanol, 0.1 mM EDTA, 0.1 mM PMSEF, 0.1 mM benzamidine) and loaded on a Superdex 16/60 column (16/60 × 160 cm; GE Healthcare) equilibrated in buffer D. After a further centrifugation at 100,000g for 20 h, the proteins were collected by fractionation in buffer D.
Figure 6 | Model of the role of aIF2 in start codon selection. (a) The positions of the TC in IC0–PREMOTE and IC1–PIN are compared after superimposition of the 30S bodies. Colour code for IC0–PREMOTE is as follows: ribosome in pink, tRNA in light yellow, aIF2γ in light green, aIF2x in light cyan, aIF1 in light pink, aIF1A in light orange and mRNA in blue. Colour code for IC1–PIN is as follows: 30S head in grey, tRNA in yellow, aIF2γ in green, aIF2x in cyan, aIF1 in magenta, aIF1A in orange and mRNA in blue. For the sake of clarity, aIF2β is not represented. The view shows that aIF2 does not significantly move during the IC0–PREMOTE to IC1–PIN transition. Therefore, the TC is distorted in IC1–PIN. The view also shows that accommodation of the tRNA in the P site is accompanied by a rotation of the 30S head. (b) Surface representation of IC0–PREMOTE: colour code is same as in (a) with aIF2β in light blue. The TC has the same conformation as unbound to 30S and is therefore unconstrained. (c) Surface representation of IC0–PREMOTE superimposed on IC1–PIN. The tRNA can explore the P site although this results in a constrained TC conformation. (d) Surface representation of IC1–PIN. Colour code is same as in (a) with aIF2β in dark blue. If an AUG codon is found in the P site, the constrained conformation of the TC is stabilized by base pairing. The lower part of the figure shows schemes illustrating the spring force model.

0.1 mM benzamidine). The sample was then loaded onto an S-Hiload column (10 mm × 4 cm; GE Healthcare) equilibrated in buffer D containing 250 mM NaCl. Assembled heterotrimer was eluted by applying a step of buffer D containing 700 mM NaCl. The recovered sample was then concentrated to 1 ml. The final heterotrimer preparation was obtained after purification by molecular sieving on a Superdex 200 HR column (10 mm × 30 cm, GE Healthcare). At the end of purification, the protein was concentrated and 1 mM Gpp(NH)p-Mg2⁺ was added. Flash-frozen samples were stored at −80 °C.

Pa-aIF1—The gene encoding aIF1 from P. abyssi was amplified from genomic DNA and cloned into pET15b/p to produce an N-terminally his-tagged version of the factor. A culture (250 ml) of cells overproducing the tagged version of aIF1 was collected, re-suspended in 10 ml of buffer D plus 100 mM NaCl and heated for 10 min at 70 °C. After pelleting non-thermostable proteins by centrifugation, the supernatant was directly loaded onto a Q-Hiload column (10 mm × 4 cm; GE Healthcare) equilibrated in buffer D containing 250 mM NaCl. aIF1A was eluted by applying an NaCl gradient (250 to 700 mM). The sample was then loaded onto an S-Hiload column (10 mm × 4 cm; GE Healthcare) equilibrated in buffer A containing 250 mM NaCl. aIF1A was eluted by applying an NaCl gradient (250 to 800 mM) in buffer A. The recovered sample was then concentrated, flash-frozen and stored at −80 °C. Routinely, 4 mg purified protein were obtained from a 250 ml culture.

Pa-aIF1A—The gene encoding aIF1A from P. abyssi was amplified from genomic DNA and cloned into pET3a/p to produce a native version of the factor. A culture (250 ml) of cells overproducing the native version of aIF1A was collected, resuspended in 10 ml buffer A containing 500 mM NaCl and disrupted by sonication. After centrifugation, the supernatant was heated for 10 min at 80 °C. After pelleting the non-thermostable proteins by centrifugation, the supernatant was directly loaded onto a Q-Hiload column (10 mm × 4 cm; GE Healthcare) equilibrated in buffer A plus 500 mM NaCl. The flow-through was recovered, diluted twice to obtain a final concentration of 250 mM NaCl and loaded onto an S-Hiload column (10 mm × 4 cm; GE Healthcare) equilibrated in buffer A containing 250 mM NaCl. aIF1A was eluted by applying an NaCl gradient (250 to 800 mM) in buffer A. The recovered sample was then concentrated, flash-frozen and stored at −80 °C. Routinely, 4 mg purified protein were obtained from a 250 ml culture.

Preparations of methionylated tRNAs. tRNAMetA1–U72 was produced in E. coli from a cloned gene, purified and aminoacylated as described25. As previously shown, tRNAMetA1–U72 from E. coli is 79% identical to Pa-tRNAiMet and the two methionyl-tRNAs bind Pa-aIF2 with almost the same affinity28. Moreover, SAXS data have shown that an aIF2–tRNA complex formed with either Met-tRNAiMet from E. coli or archaeal Met-tRNAiMet display identical diffusion curves, which strongly suggest similar structures31.

Purification of IC. The IC was assembled as follows: 864 picomoles of Pa-30S were mixed to a twofold excess of Pa-aIF2 (1,728 picomoles), fivefold excess of Pa-aIF1 (4,322 picomoles), Pa-aIF1A, met-tRNAiMetA1–U72, synthetic mRNA (A(A–17)
5UUGGAGGGAGUAAAUAGC (ThermoScientific) in the presence of GDPNP to a final concentration of 1 mM. The sequence of this synthetic mRNA was identical except for the natural termination codon from the mRNA containing elongation factor aIF1 from P. abyssi. Note that the ternary complex was preformed first by mixing aIF2, GDPNP, Met-tRNAfMet–U72, and incubating 30–50 °C. The IC mixture was then incubated for 5 min at 51 °C in IC buffer (10 mM Hepes pH 7.5, 100 mM NH4Cl, 10 mM magnesium acetate, 2 mM ATP, 10 mM EDTA, 0.03% xylene cyanol, 0.03% bromophenol blue). The samples were heated for 3 min at 95 °C before loading onto a 12% polyacrylamide gel containing 5% urea. The gel was stained with ethidium bromide and then analysed with Image J (ref. 46). As previously shown38,39, S1 nuclease cleaves in the anticono loop. The percentage of RNA cleavage was plotted as a function of the time and rate of cleavage deduced from linear regression (Supplementary Fig. 8).

Electron microscopy. Before spotting onto the grids, an excess of mRNA, aIF1A, TC (each 1:1 regarding IC) and GDPNP (0.5 mM) was added and the IC was heated 1 min at 50 °C. After heating, the samples were diluted to a final concentration of ~0.2 μM in IC buffer containing 0.5 mM GDPNP and immediately used for grid preparations. Carbon-coated Quantifoil R2/2 grids (Quantifoil Micro Tools GmbH) were blotted with 4 μl of the purified samples and vitrified in liquid ethane at 100 K on a Vitrobot Mark IV (FEI). Low-dose images were collected at a calibrated magnification of ×129,629 on an FEI Titan Krios with a Cs corrector operating at 300 kV, using a 4 × 4 keV FEI Falcon II camera (Netherlands Centre for Electron Nanoscopy, Leiden). The pixel size was 1.12 Å on the camera. De-focus values ranged from ~0.5 to ~2.8 μm. All the images were recorded using the EPU software (FEI) for automatic data collection.

Image processing. The low-dose images were taken with a total dose of 44 e− Å−2 summed on 16 frames. The first frame and the last eight frames were discarded to minimize beam-induced movement and irradiation. The remaining frames corresponding to the total dose of 19 Å−2 were first aligned with DRIFTCORR47, averaged and then sorted according to best power spectrum and ice quality. The particles from the selected images (2,154) were automatically picked with the RELION package48. An initial data set of ~192,000 particles was sorted using several rounds of classical two-dimensional classification protocols implemented in RELION. Then, a 3D-classification in RELION using a 70 Å filtered map of the archaeal 30S (extracted from the Pyrococcus furiosus 70S structure49) helped to isolate several well-resolved classes that were used for further 3D classification. The following sorting procedure allowed us to discard classes without or with poorly resolved aIF1, aIF1A or RNA. Some of the remaining classes showed a mixture of conformations of the IC–Pf607 subunit and IC–Pf607 and needed further 3D-classification. Finally, well-resolved classes representing unambiguous positions of the RNA were regrouped and used for individual 3D refinement (Supplementary Fig. 2a). Two final data sets of ~12,600 and ~5,500 particles were used to get the final 3D reconstructions called IC–Pf607remote and IC–Pf607, respectively. The resolution of the final 3D reconstructions were estimated, using gold-standard Fourier Shell Correlation, to 9.5 and 12.6 Å for IC–Pf607remote and IC–Pf607, respectively. A movie processing approach49 was also attempted in RELION but did neither increase the maps quality nor their resolution. Finally, a post-processing approach also implemented in RELION gave higher resolution estimates of 5.34 and 7.5 Å (Supplementary Fig. 2b) for IC–Pf607remote and IC–Pf607, respectively. Reported overall resolutions were calculated with the gold-standard 0.143 FSC cut-off criterion50 and were corrected for the effects of a soft mask on the FSC curve with high-resolution noise substitution51. The final density maps were corrected for the modulation transfer function of the detector and sharpened using a negative factor that was estimated using automatic sharpening procedures51. Sharpening B-values were −132 Å2 for IC–Pf607remote and −208 Å2 for IC–Pf607.

To assess the quality of the final reconstructions, a local resolution map for each reconstruction was also computed: information up to 4 Å, for the core of the 30S subunit and between 6 and 15 Å for the translation initiation factors binding regions is available in both complexes (Fig. 2 and Table 1). Such decrease in resolution as distance with the core 30S increases was already observed in other IC52. Because of this, for better clarity in the figures showing factors, maps were generally low-pass filtered to 8.5 Å (IC–Pf607remote) or 12 Å (IC–Pf607).

Model building and refinement. IC–Pf607remote. The structures of the head and body parts of the 30S subunits were extracted from the P. furiosus 70S coordinates46 (PDB entry 4V6U). The head and body were manually placed independently into the electron density map using Chimera, before automatically refining the position in the same programme. Messenger RNA was clearly visible in the density, including interactions with the 3D structure of the IC–Pf607remote complex (Supplementary Fig. 4a). The mRNA was modelled in Coot52 using the structure of the T. thermophilus 70S ribosome complexed with mRNA and tRNAs (PDB entry 1JG0)53 as a guide. The start codon was positioned in the P site. The mRNA was not visible from base +5 to base +9 (as in yeast 48S complex, PDB entry 3JAQ) at the 3′ extremity. Bases −17 and −16 at the 5′ end of mRNA were not visible. The isolated PDR models of initiation factors and tRNAs were manually placed in the electron density. The crystallographic model of the ternary IC aIF2-GDPNP-tRNA53 (PDB entry 3V11) could be readily fitted in the electron density with few modifications (see text, Fig. 4d and Supplementary Fig. 3b). Finally, a few manual model corrections were performed in Coot52 to better fit the electron density maps. These corrections concerned the following regions: 16S rRNA connection between the head and body, h44 bulge, aIF1 basic loop, h44 near aIF1A. The obtained model was then refined in Refmac54 using data up to 5.8 Å resolution. Tight geometric and external restraints (generated using ProSmart and LIB) were maintained throughout refinement to minimize overfitting, as described55–57. According to the relatively modest resolution, a B value was used for the whole model. Average Fourier shell correlation (FSC) was monitored in Refmac. For cross-validation against overfitting, the atoms of the final model were randomly replaced (with a maximum deviation of 0.5 Å), and a refinement procedure was performed against the maps that were reconstructed from only one of the two independent halves of the data used in our gold-standard FSC procedure. The FSC curves were calculated between the resulting model and the half map against which it had been refined (FSCwork) as well as the FSC curve between the model and the other half map (FSCtest)58 (Supplementary Fig. 2c). The average FSC for the final model against the 5.8 Å resolution map in refinement was 0.56 (Rfactor 0.42). The correlation coefficient between the map simulated from the model (5.3 Å resolution) and the experimental map was 82.1%, as determined using Chimera59. Further attempts to improve this value by loosening restraints during refinement systematically led to overfitting. Therefore, at the present resolution, Refmac was essentially used to fix the small clashes and to improve geometry of the initial 30S unreﬁned cryo-EM model. The number of bad bonds dropped from 5.93 to 0.06%, and the number of bad angles from 6.93 to 0.58%. The statistics for image processing and refinement are summarized in Supplementary Table 1.

IC–Pf607. The refined head, body and mRNA from IC–Pf607remote were fitted within IC1 map using Chimera. The positions of aIF1 and aIF1A were manually adjusted in the electron density. According to its known structural organization25,26, aIF2 was placed as three rigid bodies: (1) the γ-subunit bound to domain 3 of the β-subunit and to the N-terminal helix of the β-subunit, (2) domains 1 and 2 of the γ-subunit and (3) the core domain of the β-subunit. The model of the tRNA was built using the structure of the P site initiator (tRNA as observed in 4V5I (ref. 34). The tRNA was cut into two parts in the hing region (G23, A36). The two parts were fitted independently. Start codon and anticodon tRNA bases were adjusted to allow base pairing. Moreover, the acceptor CCA end was remodelled to allow fitting of the methylated 3′adenosine in its pocket on the aIF2–IIH domain. Finally, the connection between the two RNA halves was manually adjusted. As a final step, the basic loop of aIF1 was rebuilt to better fit the density. Because of the modest resolution of IC1–Pf607, Refmac refinement systematically led to overfitting and was therefore not applied to the final model. Correlation coefficient between the map simulated from the model (7.5 Å resolution) and the experimental map was 88.0%, as determined using Chimera58.

Data availability. The EMDataBank accession numbers for the EM maps reported in this paper are EMDB-8148 (IC–Pf607remote) and EMDB-8149 (IC1–Pf607). The coordinates of the models fitted in the maps have been deposited in the Protein Data Bank (accession numbers 5JB3 for IC–Pf607remote, 5JBH for IC1–Pf607). The data that support the findings of this study are available from the corresponding author on request.

References
1. Hershey, J. W., Sonenberg, N. & Mathews, M. B. Principles of translational control: an overview. Cold Spring Harb. Perspect. Biol. 4, a011528 (2012).
2. Hinnebusch, A. G. & Lorsch, J. R. The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb. Perspect. Biol. 4, a011544 (2012).
3. Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol*. 11, 113–124 (2010).

4. Akano, K. et al. Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. *EMBO J.* 20, 2326–2337 (2001).

5. Majumdar, R., Bandyopadhyay, A. & Maitra, U. Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40S preinitiation complex. *J. Biol. Chem.* 278, 6580–6587 (2003).

6. Pestova, T. V., Borukhov, S. I. & Hellen, C. U. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* 394, 854–859 (1998).

7. 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, 251–262 (2005).

8. Cheung, Y. N. et al. Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start site selection in vivo. *Genes Dev.* 21, 1217–1230 (2007).

9. Huang, H. K., Yoon, H., Hannig, E. M. & Donahue, T. F. GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev.* 11, 2396–2413 (1997).

10. Unbehaun, A., Borukhov, S. I., Hellen, C. U. & Pestova, T. V. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev.* 18, 3078–3093 (2004).

11. Aitken, C. E. & Lorsch, J. R. A mechanistic overview of translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.* 19, 568–576 (2012).

12. Voigts-Hoffmann, F., Klinge, S. & Ban, N. Structural insights into eukaryotic ribosomes and the initiation of translation. *Curr. Opin. Struct. Biol.* 22, 768–777 (2012).

13. Hashem, Y. et al. Structure of the mammalian ribosomal 43S preinitiation complex boHx29. *Cell* 153, 1108–1119 (2013).

14. Lacer, J. L. et al. Conformational differences between open and closed states of the eukaryotic translation initiation complex. *Mol. Cell* 59, 399–412 (2015).

15. Lomakin, I. B. & Steitz, T. A. The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature* 500, 307–311 (2013).

16. Weiss, M., Voigts-Hoffmann, F., Rabl, J., Leibundgut, M. & Ban, N. The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF2. *Nat. Struct. Mol. Biol.* 20, 1015–1017 (2013).

17. Saini, A. K., Nanda, J. S., Lorsch, J. R. & Hinnebusch, A. G. Regulatory elements controlling stringent selection of the AUG start codon during translation initiation. *Mol. Cell* 73, 730–736 (2011).

18. Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *J. Mol. Biol.* 240–255 (1997).

19. Scheres, S. H. A Bayesian view on cryo-EM structure determination. *Nat. Struct. Mol. Biol.* 20, 624–636 (2006).

20. Scheres, S. H. A. Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* 159, 597–607 (2014).

21. Schmitt, E. et al. Structure of the ternary initiation complex aIF2-GDPNP-tRNA(Met) ternary complex binding to the 40S ribosome. *Nat. Struct. Mol. Biol.* 18, 1227–1234 (2011).

22. Hasenfeld, D., Fabbretti, A., Londei, P., Gualerzi, C. O. & Blasi, U. Translation initiation complex formation in the crenarchaeon Sulfolobus solfataricus. *RNA* 15, 2288–2289 (2009).

23. Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* 415, 406–418 (2012).

24. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 599–605 (2012).

25. Huxsoll, T. A. et al. Translation initiation in mammalian cells: a structural and functional study of the eukaryotic 40S ribosome. *J. Biol. Chem.* 288, 27546–27562 (2013).

26. Erasoa, G. et al. Pyrococcus abyssi sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch. Microbiol.* 160, 338–349 (1993).

27. Mechulam, Y., Guillou, L., Yatime, L., Blanket, S. & Schmitt, E. Protection-based assays to measure aminoacyl-tRNA binding to translation initiation factors. *Methods Enzymol.* 430, 263–281 (2007).

28. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675 (2012).

29. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* 10, 584–590 (2013).

30. Armache, J. P. et al. Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. *Nucleic Acids Res.* 41, 1284–1293 (2013).

31. Bai, X. C., Fernandez, I. S., McMullan, G. & Scheres, S. H. Ribosome structures near-atomic resolution from thirty thousand cryo-EM particles. *Elife* 2, e00461 (2013).

32. Chen, S. et al. Near-atomic resolution electron microscopy of the 40S ribosome. *EMBO J.* 22, 914–919 (2015).

33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr.* D66, 486–501 (2010).

34. Yusupova, G. Z., Yusupov, M. M., Cate, J. H. & Noller, H. F. The path of messenger RNA through the ribosome. *Cell* 106, 233–241 (2001).

35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* D 55, 240–255 (1999).

36. Amunts, A. et al. The yeast mitochondrial large ribosomal subunit. *Science* 343, 1485–1489 (2014).

37. Brown, A. et al. Structure of the large ribosomal subunit from human mitochondria. *Science* 346, 718–722 (2014).

38. Chen, S. et al. Mechanism of eIF6 release from the nascent 60S ribosomal subunit. *Nat. Struct. Mol. Biol.* 22, 914–919 (2015).

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

40. Nguyen, T. H. et al. The architecture of the spliceosomal U4/U6/U5 tri-snRNP. *Nature* 523, 47–52 (2015).

**Acknowledgements**

This work is dedicated to the memory of our friend and colleague Nicolas Boisset. We thank F. Wei, R. Gillet, S. Jonc, A. Myasnikov for valuable discussions on electron microscopy and J.F. Menetret for help with data collection at the preliminary stages of this study. We also gratefully acknowledge R. Matareend and S. de Carlo (FEI Company)
for image acquisition at the National Center for Electron Nanoscopy in Leiden (NeCEN) and helpful advice. We thank Thomas Simonson for critical reading of the manuscript. This work was supported by the Centre National de la Recherche Scientifique, Ecole polytechnique, the Physique et Chimie du Vivant program of Agence Nationale de la Recherche (MASTIC project) and by the French Infrastructure For Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01. A.M. was supported by a ‘Bourse Monge’ doctoral fellowship from Ecole polytechnique.

**Author contributions**

P.-D.C., E.S. and Y.M. designed the research; P.-D.C., Y.M., E.S., C.L.-S., A.M., E.L. and L.C. carried out the experiments; all the authors analysed the data; P.-D.C., E.S., Y.M. and B.P.K. wrote the paper.

**Additional information**

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Couroux, P.-D. et al. Cryo-EM study of start codon selection during archaeal translation initiation. Nat. Commun. 7, 13366 doi: 10.1038/ncomms13366 (2016).

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

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/