Repeat-expanded C9orf72 mRNA engages the human ribosome to initiate non-AUG translation

Jinlun Kylian Zhang[1,3], Xurui Shen[1,3]*, You Wu[1,3], Qichao Lu[1], Yuliya Gordiyenko[2], Li Li[1], Zining Guo[1], Xinmin Wang[1], Huiwen Zhong[1], Xueyan Li[2]*, Tao Wang[1]*, Hanting Yang[1]*

[1] Institute for Translational Brain Research, State Key Laboratory of Brain Function and Disorders and MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200032, China.

[2] Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, England, UK

[3] These authors contributed equally to this work.

* Corresponding author: yanght@fudan.edu.cn (H.Y.); wang_tao@fudan.edu.cn (T.W.); xueyan@mrc-lmb.cam.ac.uk (X.L.); xuruishen@fudan.edu.cn (X.S.).

Abstract

Expanded hexanucleotide repeats in *C9orf72* produce toxic proteins through repeat-associated non-AUG (RAN) translation, yet how repeat RNAs engage the eukaryotic initiation machinery has remained unresolved. Here we determine cryo-electron microscopy structures of human translation initiation complexes assembled on a $(G4C2)_{96}$ repeat mRNA. The transcript enters the canonical initiation pathway but accumulates in a normally transient 48S intermediate. A near-cognate CUG codon is decoded within a Kozak-like context in the standard decoding centre, while the repeat RNA simultaneously forms an anchoring contact with the 40S ribosomal head outside the mRNA channel. Together these features prolong the lifetime of a scanning-to-recognition transition intermediate and convert it into a productive start state without bypassing canonical initiation. RAN translation therefore arises not from an alternative initiation mechanism but from stabilisation of a latent non-AUG-competent state intrinsic to the ribosome. This framework links repeats expansion pathology to modulation of initiation dynamics and provides a structural basis for selectively suppressing pathogenic protein synthesis.

Introduction

Repetitive sequences are widespread in the human genome and are generally tolerated without deleterious consequences. However, pathological expansions of specific repeat elements disrupt cellular homeostasis and drive disease pathogenesis[1-3]. Expanded repeat RNAs interfere with multiple aspects of RNA metabolism and give rise to toxic repeat-derived proteins[4-9], collectively underlying a diverse group of repeat expansion disorders with prominent neurological involvement[3,10,11].

A major example is amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), most commonly caused by expansion of a GGGGCC (G4C2) hexanucleotide repeat within the first intron of *C9orf72*[12-17]. One hallmark consequence of this expansion is repeat-associated non-AUG (RAN) translation, which produces toxic dipeptide repeat (DPR) proteins[18-20]. RAN translation can initiate within or upstream of repetitive sequences on both sense and antisense transcripts without an AUG start codon, generating multiple DPR species including poly-GA, poly-GR, poly-PA, poly-PR and poly-GP[20-24]. Although these DPRs share broadly similar anatomical distributions in patient tissue, their abundance differs markedly, with poly-GA being most prevalent, indicating that initiation efficiency varies across reading frames and transcript contexts[24,25].

Despite extensive work on DPR toxicity, the molecular basis of RAN translation initiation remains unclear[10,22]. In canonical cap-dependent eukaryotic translation initiation, the eIF4F complex engages the 5′m^{7}G cap, together with auxiliary factors, remodels the 5′ UTR to promote recruitment of the 43S pre-initiation complex (40S subunit bound to eIF1, eIF1A, eIF3 and the eIF2-GTP-Met-tRNA$_i$ ternary complex). The resulting 48S initiation complex then scans the 5′ UTR in a 5′→3′ direction and selects the start site upon recognition of an AUG codon in a permissive sequence context[26-30]. Whether repeat-containing mRNAs engage this canonical machinery during RAN translation, how initiation sites are selected in the absence of AUG codons, and how expanded repeat RNAs interact with the ribosome at this stage remain unresolved, largely due to the absence of high-resolution structural information for

RAN initiation complexes.

More broadly, translation initiation is often described as a scanning process in which ribosomes recognise a start codon within a favourable sequence context. Yet near-cognate initiation events occur in many transcripts, indicating that start-site selection is inherently probabilistic rather than strictly deterministic[31,32]. How productive initiation sites are stabilised while transient scanning intermediates are rejected, and whether alternative initiation configurations exist within the canonical pathway, remain fundamental unresolved questions in translation.

Repeat expansion transcripts therefore provide an opportunity to examine these principles under conditions of extreme start-site ambiguity. The continued requirement for canonical initiation factors in RAN translation suggests that repeat RNAs may not bypass the initiation pathway but instead stabilise an alternative initiation configuration that is normally too transient to observe.

Here we use cryo-electron microscopy to visualise human translation initiation complexes assembled on a repeat-expanded *C9orf72* mRNA containing $(G4C2)_{96}$ repeats. We capture both scanning-competent and start-site-engaged intermediates and find that the transcript progresses through the canonical initiation pathway but stabilises a normally transient initiation configuration. Two cooperative determinants are identified in this work that enable non-AUG start selection on repeat-expanded transcripts: a Kozak-like context that supports CUG decoding in the canonical centre, and a repeat-ribosome contact outside the mRNA channel that reinforces ribosomal engagement. Together, these features capture a scanning intermediate that is otherwise short-lived and channel it into productive initiation through the standard machinery. We therefore propose that RAN translation exploits inherent flexibility of the ribosome's initiation trajectory, rather than operating through an alternative mechanism.

Result

Repeat-expanded *C9orf72* mRNA assembles a canonical human 48S initiation complex

To determine how repeat-expanded C9orf72 RNA engages the translation initiation

machinery, we visualised human initiation complexes assembled on a $(G4C2)_{96}$ transcript using cryo-electron microscopy. Particle classification revealed discrete states along the canonical initiation pathway, including the 43S pre-initiation complex and an mRNA-engaged 48S complex (Fig. 1a; Fig. S1). Notably, the 48S population accumulated in a closed conformation characterised by pronounced head closure and stable accommodation of the mRNA within the decoding channel, consistent with a scanning-to-recognition transition intermediate rather than a transient recruitment state (Fig. 1b; Fig. S2-S3).

Within this 48S complex, continuous density allowed tracing of the repeat mRNA through the decoding centre, where the backbone forms stabilising contacts with ribosomal proteins RPS2, RPS3, RPS14 and RPS26 and interacts with initiation factors eIF1A and eIF2α (Fig. 1c). The architecture includes the octameric eIF3 complex, the eIF2-GTP-Met-$tRNA_i$ ternary complex and eIF4A at the mRNA entry channel, indicating that RAN translation proceeds through a canonical initiation assembly rather than an alternative ribosomal state.

To obtain structurally homogeneous complexes, we established a hybrid reconstitution strategy combining purified native human 40S subunits and recombinant initiation factors with a defined repeat transcript. The in vitro-transcribed mRNA exhibited size homogeneity and supported RAN translation, producing detectable DPR proteins in cells and in a cell-free system (Fig. 1d-h), confirming that the captured complexes correspond to a biologically competent substrate.

Consistent with regulation at the initiation stage, arsenite treatment, which induces eIF2α phosphorylation, increased poly-GA production from the repeat reporter (Fig. 1i; Fig. S4), supporting a link between initiation-state stabilisation and DPR output. However, how a defined translational reading frame is established within this canonical complex in the absence of an AUG codon remained unclear.

RAN initiation decodes a near-cognate CUG codon within the canonical decoding centre

A key unresolved question in repeat-associated translation is how a defined reading

frame is established in the absence of an AUG codon. Inspection of the repeat mRNA-bound 48S complex revealed a single registered P-site codon-anticodon interaction in which a near-cognate CUG codon, located 24 nucleotides upstream of the repeat tract, pairs with initiator Met-tRNA$_i$ (Fig. 2a). The resulting mini-helix is stabilised by canonical decoding-centre elements, including the monitoring base C1701 of 18S rRNA and A36 of the tRNA anticodon loop, indicating recognition within the standard initiation architecture rather than through a specialised decoding mode.

Superposition with a canonical AUG initiation complex[29] showed highly similar geometry of the P-site environment, including positioning of the tRNA anticodon loop and surrounding rRNA nucleotides (Fig. S5a,b). Together with the local-resolution features at the CUG region (Fig. S5c), these data indicate that RAN initiation proceeds through the canonical decoding centre despite the near-cognate start codon.

Consistent with this structural observation, mutation of the CUG codon to CCG markedly reduced poly-GA production (Fig. 2b,c), demonstrating that the identified register functions as the dominant initiation site in the reporter context.

In the repeat mRNA-engaged complex, eIF4A occupies the mRNA entry channel and contacts the transcript, consistent with a role in maintaining productive scanning and start-site positioning. Inhibition of eIF4A with rocaglamide A (RocA)[33] abolished defined P-site density in cryo-EM reconstructions (Fig. 2d) and reduced poly-GA output (Fig. 2e,f), indicating that stable near-cognate decoding requires active mRNA engagement rather than passive ribosome stalling.

We additionally observe ion-like density adjacent to the P site (Fig. S6), consistent with occupancy by a cation present in the reconstitution buffer (K^+ or Mg^{2+}). Although its identity cannot be assigned unambiguously, its position suggests local stabilisation of the initiation geometry. Because several near-cognate codons occur upstream of the repeat, this raised the question of how this specific CUG site is preferentially selected.

A Kozak-like context licenses near-cognate start-site selection

We next examined why the ribosome preferentially selects this near-cognate start site. The sequence upstream of the repeat contains a Kozak-like context surrounding the

CUG codon, with guanines at the hallmark −3 and +4 positions (Fig. 3a). Cryo-EM density resolves a continuous mRNA register spanning nucleotides −4 to +4, revealing an extensive interaction network between the flanking nucleotides, the 40S subunit and initiation factors (Fig. 3b).

Upstream of the codon, −4 A packs against RPS26 H80, −3 G forms hydrogen bonds with 18S rRNA G961 and eIF2α Arg55, and −1 U contacts 18S rRNA G1207 (Fig. 3b). Downstream, the +4 G interacts with 18S rRNA A1825 and eIF1A Lys67, supported by stacking against eIF1A Trp80 (Fig. 3c). These contacts stabilise the mRNA geometry across the P-site and compensate for the reduced pairing strength of the non-AUG codon-anticodon interaction while preserving canonical decoding-centre architecture.

Consistent with a context-driven mechanism, a nearby near-cognate candidate (AGG) located 15 nucleotides upstream lacks a favourable Kozak environment and is not captured in the initiation register (Fig. S7). Thus, start-site selection is governed primarily by context-dependent stabilisation of the scanning intermediate rather than by codon identity alone. This suggested that additional features of the repeat transcript might further stabilise the initiation state once a licensed register is reached.

Repeat RNA anchoring stabilises a scanning-to-recognition transition initiation intermediate

We therefore asked whether the repeat RNA itself interacts with the ribosome to stabilise the initiation complex. Beyond the decoding centre, the G4C2 repeat tract engages the ribosome at a site external to the canonical mRNA channel. In the repeat mRNA-bound 48S complex we resolve a prominent density extension emanating from the repeat that contacts the ES9S expansion segment of 18S rRNA on the solvent-exposed side of the 40S head, adjacent to RACK1 and surrounding head-domain ribosomal proteins (Fig. 4a; Fig. S8). The positioning places the repeat RNA outside the decoding path yet mechanically coupled to head movements that accompany initiation.

Local model building indicates that a GCCGGG segment within the repeat forms a short near-complementary helix with a GCAGCCC tract in ES9S (Fig. 4b). Several

positions are consistent with Watson-Crick base pairing, whereas the central region deviates from an ideal duplex and is instead stabilised by stacking interactions involving ES9S nucleotides G1420 and A1421 together with a cytidine from the repeat RNA. These mixed canonical and non-canonical contacts permit anchoring without requiring perfect complementarity and are compatible with heterogeneous repeat register while maintaining a defined attachment site.

The location of this tether relative to the decoding centre links the repeat interaction to start-site selection observed upstream. Within this anchoring context, the near-cognate CUG codon is licensed within a Kozak-like context and decoded in a canonical P-site geometry, indicating that start-site recognition proceeds through standard initiation architecture rather than an alternative decoding mode. The ES9S contact therefore does not create a new initiation pathway but instead constrains transcript positioning once the ribosome reaches a competent register.

The repeat-ES9S tether provides a structural basis for repeat-dependent stabilisation of a scanning-paused 48S intermediate, which may constitute a kinetically gated step within canonical initiation. In line with this kinetic interpretation, DPR output is modulated by perturbations of canonical initiation reported earlier. Under RocA treatment, which clamps eIF4A onto mRNA and disrupts productive recruitment and scanning, P-site mRNA density is not resolved in our reconstructions and poly-GA production is reduced. Conversely, conditions associated with increased eIF2α phosphorylation correlate with elevated DPR output (Fig. 4c). Together, these observations suggest that RAN efficiency is set by the stability and occupancy of a specific intermediate within the canonical initiation pathway.

Integrating these findings with the preceding sections, the data support a model in which two cooperative features stabilise a productive initiation state: a Kozak-licensed near-cognate CUG codon establishes a canonical decoding-centre configuration, while repeat-dependent ES9S anchoring restricts transcript mobility and biases the ribosome towards a scanning-to-recognition transition 48S intermediate. In this framework, RAN translation arises from stabilisation of a normally transient initiation configuration inherent to the ribosome, rather than from engagement of a distinct initiation

mechanism.

Discussion

Repeat-associated non-AUG (RAN) translation has been widely invoked to explain how expanded repeat RNAs generate toxic proteins in neurodegenerative disease, yet its relationship to canonical translation initiation has remained unclear. The prevailing view has implicitly assumed that RAN translation represents a specialised or aberrant initiation pathway. Our structural and functional analyses instead indicate that the ribosome does not bypass canonical initiation rules on repeat transcripts. Rather, repeat-expanded *C9orf72* mRNA exploits intrinsic flexibility within the normal initiation pathway to stabilise a productive non-AUG initiation state.

At the decoding centre, the near-cognate CUG codon is accommodated within an otherwise standard initiation architecture. The codon-anticodon helix, monitoring rRNA nucleotides and initiator tRNA adopt geometries closely resembling AUG recognition, demonstrating that start-site specification does not require a dedicated decoding mode. Instead, initiation fidelity is modulated quantitatively: a Kozak-like context compensates for the reduced pairing optimality of the non-AUG codon, licensing entry into the initiation pathway while preserving its structural logic. This supports a model in which non-AUG initiation is not categorically distinct from canonical initiation but represents an energetically marginal state normally disfavoured during scanning.

The repeat tract introduces a second, orthogonal influence on initiation dynamics. By forming an anchoring interaction with ES9S on the 40S head, the transcript becomes mechanically coupled to ribosomal movements while remaining outside the mRNA channel. This interaction is positioned to restrict transcript mobility after start-site recognition, thereby prolonging the lifetime of a scanning-to-recognition transition 48S intermediate. In this framework, RAN translation arises when a licensed but normally transient near-cognate initiation configuration is kinetically stabilised. The ribosome therefore does not mis-recognise the start codon; rather, it fails to disengage from a permissible but weakly favoured initiation state.

This mechanism reconciles several puzzling properties of repeat translation. Frame-biased DPR production can emerge from differential stabilisation of nearby near-cognate registers rather than from frame-specific decoding mechanisms. Sensitivity to eIF4A inhibition and enhancement under conditions associated with eIF2α phosphorylation follow naturally from modulation of initiation dwell time rather than activation of a dedicated RAN machinery. More broadly, the data suggest that repeat RNAs act as kinetic modulators of initiation rather than alternative templates interpreted by specialised ribosomes.

The principle revealed here may extend beyond *C9orf72*. Many repeat expansion transcripts form structured elements or engage ribosomal surfaces outside the decoding channel, features that could similarly stabilise marginal initiation states. RAN translation can therefore be understood as an emergent property of the normal initiation landscape when RNA-encoded interactions reshape its energy barriers. This view places repeat expansion disorders within a continuum of translational control mechanisms rather than in a mechanistically isolated category.

Finally, the ribosome-repeat interface defined here provides a conceptual and practical target for therapeutic intervention. Because the mechanism depends on stabilisation of a specific initiation intermediate rather than global translation, selectively destabilising this state may suppress DPR production while preserving bulk protein synthesis. More generally, our findings reveal that the fidelity of eukaryotic translation initiation is governed not only by decoding accuracy but also by the lifetime of intermediate states, which structured RNAs can tune to redirect proteome output in disease.

Methods

Plasmids

The CMV or T7 promoter-driven $(G4C2)_{96}$ repeat reporters were built without ATG start codon and G4C2 repeats were preceded by triple stop codons. GFP or Flag were downstream of the $(GGGGCC)_{96}$ repeats in frame with GA. The DNA sequence was confirmed by Sanger sequencing.

RNA synthesis and purification

Plasmids were linearized with XhoI (NEB) and used as templates for in vitro transcription using the HiScribe T7 High Yield RNA Synthesis Kit (NEB). The anti-reverse cap analogue 3′-O-Me-m7GpppG (ARCA) was added at a fourfold molar excess over GTP. Reactions were incubated at 37 °C for 2 h, followed by DNase I treatment at 37 °C for 15 min. mRNAs were purified using the RNA Clean & Concentrator Kit (Zymo Research). RNA concentration was measured by NanoDrop, and integrity was verified on a denaturing formaldehyde agarose gel.

In Vitro translation assay

Preparation of HEK293T cell extracts: HEK293T cells were harvested by trypsinization and washed in 5 packed cell volumes (PCV) PBS, followed by 5 PCV hypotonic buffer. Cells were resuspended in 3 PCV RNase-free hypotonic buffers. Cells were lysed by 10 passages through a 27-gauge needle and clarified by centrifugation (4 °C, 10 min). The supernatant was supplemented to final buffer containing potassium acetate, magnesium acetate, creatine phosphate, creatine kinase, spermidine, DTT and 1 U/μL RNaseOUT.

In Vitro translation reactions: Reactions were programmed with 5 nM in vitro–transcribed RNA. RNAs were heated at 65 °C for 5 min and chilled on ice before addition. Translation was performed at 33 °C for 90 min and stopped on ice. Samples were mixed with SDS loading buffer, heated at 98 °C for 10 min, and analyzed by western blotting.

Cell culture, transfection, and stress treatment

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), 100 U/mL penicillin, and 100

μg/mL streptomycin. Cells were transiently transfected with *C9orf72* plasmid DNA using Lipo8000® transfection reagent according to the manufacturer's instructions and incubated for 48 h post-transfection. For sodium arsenite treatment, cells were treated with 250 μM sodium arsenite for 2 h, followed by 6 h recovery period under standard culture conditions. Cells were then harvested, and whole-cell lysates were prepared in RIPA buffer supplemented with protease and phosphatase inhibitors for subsequent Western blot analysis. For Rocaglamide (RocA) treatment, cells were treated with 4 μM RocA or vehicle for 3 h, followed by 6h recover period under standard DMEM.

Western blot

HEK293T cells were lysed in RIPA buffer supplemented with 6 M urea and a cOmplete protease inhibitor cocktail (Roche). Cell lysates were mixed with SDS loading buffer and denatured at 98°C for 10 min. Proteins were resolved by 10% SDS-PAGE and transferred onto nitrocellulose membranes (Cytiva). Membranes were blocked with 5% non-fat milk in Tris-buffered saline containing 0.1% Tween-20 (TBST) for 1 h at room temperature and subsequently probed with primary antibodies (Proteintech, 1:1,000) overnight at 4°C. After washing with TBST, membranes were incubated with HRP-linked secondary antibodies (Proteintech, 1:10,000) or fluorescent secondary antibodies (IRDye 800CW, LI-COR Biosciences) for 2 h at room temperature. Immunoreactive bands were visualized using ECL Plus Reagent (Lablead) on a ChemiDoc Imager (Bio-Rad) or detected via a dual-mode infrared imaging system (LI-COR). Densitometric analysis was performed using FIJI software, with target protein levels normalized to the corresponding loading controls.

In vitro reconstitution of human 48S complexes for cryo-EM

Initially, eIF2, $tRNA_i^{Met}$, and GMPPCP were mixed to form the ternary complex (TC) mixture. The TC mixture was then incubated with eIF1, eIF1A, eIF3, AMPPNP, and the 40S ribosomal subunit at 32 °C for 10 min to assemble the 43S PIC in buffer containing 20 mM HEPES–KOH (pH 7.5), 97 mM potassium acetate, 2.5 mM magnesium acetate, 0.3% glycerol, 0.1 mM spermidine, and 1 mM dithiothreitol. The

48S initiation complex was reconstituted by mixing the 43S PIC with eIF4A, eIF4B, eIF4E, eIF4G (557-1599) and $(G4C2)_{96}$ mRNA, and incubate at 37 °C. For preparation of RocA-treated samples, RocA was added simultaneously with eIF4A to a final concentration of 0.5 mM.

Cryo-EM sample preparation

A total of 3.3 µl of the Assembled 48S initiation complex sample was applied to UltrAuFoil R1.2/1.3 300-mesh gold grids pre-coated with graphene oxide (Sigma) prepared in house. Grids were glow-discharged prior to sample application, followed by blotting for 5 or 5.5 s using a Vitrobot (Thermo Fisher Scientific).

Cryo-EM data collection

Cryo-EM datasets were collected on Titan Krios G4 cryo-electron microscope operated at 300 kV, equipped with a Falcon 4i Direct Electron Detector and a Selectris X energy filter (Thermo Fisher Scientific). Movie stacks were automatically collected using EPU at a magnification of 130,000× with a pixel size of 0.959 Å for a total dose per EER (electron event representation) movie of 50 $e^{-}/Å^2$. The defocus range was set between -1.0 to -1.6 µm.

Cryo-EM data processing

All datasets were processed in cryoSPARC (v4.5.3). For the untreated condition, a total of ~8,000 micrographs were collected. Movie frames were motion-corrected using MotionCor2 with 5 × 5 patch alignment, and CTF parameters were estimated for each micrograph using Gctf (v1.18). Particles were picked using Gautomatch (v0.56; https://github.com/JackZhang-Lab/Gautmatch), and the motion-corrected micrographs together with particle coordinates were imported into cryoSPARC for downstream processing. Particles were extracted and subjected to an initial round of 2D classification to remove non-particle features and enrich a 40S-dominated particle set. Masks were then generated and used for 3D classification to separate 40S, 43S, and 48S classes. Selected particle subsets were refined by homogeneous refinement, followed by local refinement, to obtain the final high-resolution reconstructions.

For the RocA-treated dataset, an analogous workflow was applied, including 2D classification, mask-guided 3D classification to separate 40S/43S/48S classes, and

subsequent homogeneous and local refinements to obtain detailed maps.

Model building, refinement and validation

Models were built using PDB 8OZ0 as the starting template. Coot was used for manual model adjustment and mutation, followed by real-space refinement with phenix.real_space_refine in PHENIX.

Acknowledgments

We thank Dejian Zhou, Anqi Dong, Hui Zhao, and Zhenguo Chen at the Center of cryoEM at Fudan University for their support with cryo-EM data collection, and members of the Yang lab for discussions. This work was supported by the grants from the National Science and Technology Innovation 2030 Major Projects of China (STI2030-Major Projects 2022ZD0212600), and EMBO Postdoctoral Fellowship (ALTF 937-2022).

Author Contributions

Conceptualization: H.Y. and T.W.; Methodology: J.K.Z., X.S., X.L., H.Y., T.W., Y.W., Q.L., Y.G., L.L., Z.G., X.W. and H.Z.; Investigation: J.K.Z., X.S., Y.W., Q.L., Y.G., L.L., Z.G., X.W. and H.Z.; Visualization: J.K.Z., X.S., Y.W. and H.Y.; Funding acquisition: X.L., H.Y. and T.W.; Supervision: H.Y. and T.W.; Writing – original draft: J.K.Z., X.S. and H.Y.; Writing – review & editing: X.L. and H.Y.

Competing interests

The authors declare no competing interests.

Reference

1 Erwin, G. S. *et al.* Recurrent repeat expansions in human cancer genomes. *Nature* **613**, 96-102, doi:10.1038/s41586-022-05515-1 (2023).

2 Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243-247, doi:10.1038/nature22386 (2017).

3 Malik, I., Kelley, C. P., Wang, E. T. & Todd, P. K. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nature reviews. Molecular cell biology* **22**, 589-607, doi:10.1038/s41580-021-00382-6 (2021).

4 Guo, P. & Han, D. Targeting Pathogenic DNA and RNA Repeats: A Conceptual Therapeutic Way for Repeat Expansion Diseases. *Chemistry – A European Journal* **28**, doi:10.1002/chem.202201749 (2022).

5 Kawabe, Y., Mori, K., Yamashita, T., Gotoh, S. & Ikeda, M. The RNA exosome complex degrades expanded hexanucleotide repeat RNA in C9orf72 FTLD/ALS. *The EMBO journal* **39**, e102700, doi:10.15252/embj.2019102700 (2020).

6 Li, Y. *et al.* Globally reduced N(6)-methyladenosine (m(6)A) in C9ORF72-ALS/FTD dysregulates RNA metabolism and contributes to neurodegeneration. *Nature neuroscience* **26**, 1328-1338, doi:10.1038/s41593-023-01374-9 (2023).

7 Lu, J., Pan, Y., Feng, X. & Lu, B. Phase transition and lysosomal degradation of expanded CAG repeat RNA suppress global protein synthesis. *Autophagy* **20**, 451-453, doi:10.1080/15548627.2023.2274208 (2024).

8 Ouyang, J. P. T., Shukla, S., Bensalah, M. & Parker, R. DM1 repeat-expanded RNAs confer RNA toxicity as individual nuclear-retained RNAs. *Cell reports* **44**, 115582, doi:10.1016/j.celrep.2025.115582 (2025).

9 Pan, Y. *et al.* Gelation of cytoplasmic expanded CAG RNA repeats suppresses global protein synthesis. *Nature chemical biology* **19**, 1372-1383, doi:10.1038/s41589-023-01384-5 (2023).

10 Castelli, L. M., Huang, W.-P., Lin, Y.-H., Chang, K.-Y. & Hautbergue, G. M. Mechanisms of repeat-associated non-AUG translation in neurological microsatellite expansion disorders. *Biochemical Society Transactions* **49**, 775-

792, doi:10.1042/bst20200690 (2021).

11 Nguyen, L., Cleary, J. D. & Ranum, L. P. W. Repeat-Associated Non-ATG Translation: Molecular Mechanisms and Contribution to Neurological Disease. *Annual Review of Neuroscience* **42**, 227-247, doi:10.1146/annurev-neuro-070918-050405 (2019).

12 Cattaneo, M., Giagnorio, E., Lauria, G. & Marcuzzo, S. Therapeutic Approaches for C9ORF72-Related ALS: Current Strategies and Future Horizons. *Int J Mol Sci* **26**, doi:10.3390/ijms26136268 (2025).

13 Green, K. M. *et al.* RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nature Communications* **8**, doi:10.1038/s41467-017-02200-0 (2017).

14 Jiang, X. *et al.* Blocking RAN translation without altering repeat RNAs rescues C9ORF72-related ALS and FTD phenotypes. *Science* **391**, eadv2600, doi:10.1126/science.adv2600 (2026).

15 McEachin, Z. T. *et al.* Molecular impact of antisense oligonucleotide therapy in C9orf72-associated ALS. *Cell* **188**, 6424-6435.e6417, doi:10.1016/j.cell.2025.07.045 (2025).

16 Mizielinska, S. *et al.* Amyotrophic lateral sclerosis caused by hexanucleotide repeat expansions in C9orf72: from genetics to therapeutics. *The Lancet. Neurology* **24**, 261-274, doi:10.1016/s1474-4422(25)00026-2 (2025).

17 Simón-Sánchez, J. *et al.* The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. *Brain : a journal of neurology* **135**, 723-735, doi:10.1093/brain/awr353 (2012).

18 Ash, P. E. *et al.* Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639-646, doi:10.1016/j.neuron.2013.02.004 (2013).

19 Mori, K. *et al.* Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta neuropathologica* **126**, 881-893, doi:10.1007/s00401-013-1189-3 (2013).

20 Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in C9ORF72

ALS and frontotemporal dementia. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E4968-4977, doi:10.1073/pnas.1315438110 (2013).

21 Wojciechowska, M., Olejniczak, M., Galka-Marciniak, P., Jazurek, M. & Krzyzosiak, W. J. RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. *Nucleic Acids Research* **42**, 11849-11864, doi:10.1093/nar/gku794 (2014).

22 Kearse, M. G. & Wilusz, J. E. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev* **31**, 1717-1731, doi:10.1101/gad.305250.117 (2017).

23 Gendron, T. F. *et al.* Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta neuropathologica* **126**, 829-844, doi:10.1007/s00401-013-1192-8 (2013).

24 Mackenzie, I. R. A. *et al.* Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta neuropathologica* **130**, 845-861, doi:10.1007/s00401-015-1476-2 (2015).

25 Tabet, R. *et al.* CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat Commun* **9**, 152, doi:10.1038/s41467-017-02643-5 (2018).

26 Eliseev, B. *et al.* Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucleic Acids Res* **46**, 2678-2689, doi:10.1093/nar/gky054 (2018).

27 Brito Querido, J. *et al.* Structure of a human 48S translational initiation complex. *Science* **369**, 1220-1227, doi:10.1126/science.aba4904 (2020).

28 Brito Querido, J., Díaz-López, I. & Ramakrishnan, V. The molecular basis of translation initiation and its regulation in eukaryotes. *Nature Reviews Molecular Cell Biology* **25**, 168-186, doi:10.1038/s41580-023-00624-9 (2023).

29 Brito Querido, J. *et al.* The structure of a human translation initiation complex

reveals two independent roles for the helicase eIF4A. *Nature Structural & Molecular Biology* **31**, 455-464, doi:10.1038/s41594-023-01196-0 (2024).

30 Lapointe, C. P. *et al.* eIF5B and eIF1A reorient initiator tRNA to allow ribosomal subunit joining. *Nature* **607**, 185-190, doi:10.1038/s41586-022-04858-z (2022).

31 Andreev, D. E. *et al.* Non-AUG translation initiation in mammals. *Genome biology* **23**, 111, doi:10.1186/s13059-022-02674-2 (2022).

32 Kearse, M. G. & Wilusz, J. E. Non-AUG translation- a new start for protein synthesis in eukaryotes. *GENES & DEVELOPMENT*, doi:10.1101/gad.305250 (2021).

33 Iwasaki, S., Floor, S. N. & Ingolia, N. T. Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. *Nature* **534**, 558-561, doi:10.1038/nature17978 (2016).

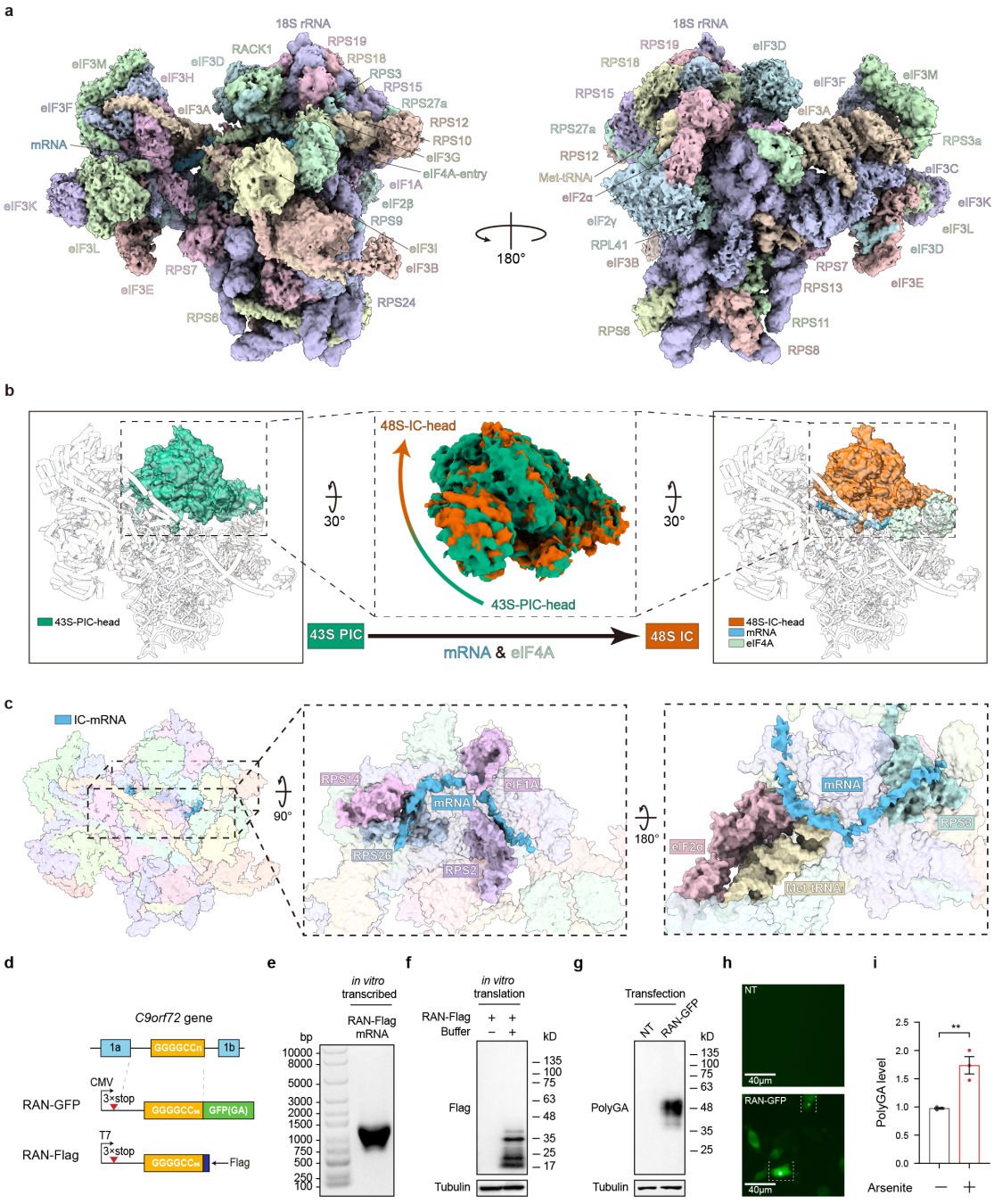

Figure 1. Reconstitution and structural characterization of a G4C2 repeat mRNA-bound human 48S RAN initiation complex

a, Cryo-EM structure of the human 48S initiation complex (IC) assembled on G4C2 repeat mRNA. The reconstruction resolves the 40S subunit, the octameric eIF3 complex, the ternary complex (eIF2-GTP-Met-tRNA$_i$) and eIF4A positioned at the mRNA entry channel, together defining the architecture of the repeat mRNA-bound IC. **b,** Conformational transition from the 43S pre-initiation complex (PIC) to the 48S IC upon mRNA recruitment. Comparison of the two states reveals a swivel of the 40S head domain associated with mRNA and eIF4A engagement, consistent with a shift from an open recruitment state to a more closed initiation state. **c,** Interactions stabilizing the repeat mRNA within the 48S IC. **Left,** Overview of the resolved mRNA density (cyan) traversing the mRNA channel. **Right,** Close-up views

highlighting contacts between the mRNA backbone and ribosomal proteins (RPS2, RPS3, RPS14 and RPS26), as well as initiation factors eIF1A and eIF2α. **d,** Schematic of the C9orf72 locus containing the G4C2 repeat expansion and design of RAN translation reporters. The cellular reporter (RAN-GFP) and the in vitro reporter (RAN-Flag) contain G4C2 repeats preceded by triple stop codons to minimize canonical upstream translation. **e,** Quality assessment of in vitro-transcribed RAN-Flag mRNA by gel electrophoresis, showing the expected transcript size and integrity. **f,** Cell-free translation of in vitro-transcribed RAN-FLAG mRNA yields FLAG-immunoreactive products in cell lysates. Vehicle buffer was added as indicated Tubulin, loading control. **g,** Representative immunoblot showing poly(GA) production in HEK293T cells transfected with the RAN-GFP reporter compared with non-transfected (NT) controls. Tubulin was used as a loading control. **h,** Formation of poly(GA) aggregates derived from G4C2 repeat expansions. Representative fluorescence images of HEK293T cells. Top: Non-transfected control cells. Bottom: Cells transfected with a RAN-GFP reporter. White dashed lines outline poly(GA) inclusions. Scale bar = 40 μm. **i,** Quantification confirms increased CTG-dependent poly(GA) upon arsenite treatment. Densitometric analysis normalized to tubulin and expressed relative to the untreated condition. Dots represent independent biological replicates; bars show mean ± s.e.m. Statistical significance was assessed using a two-sided unpaired t-test. **P < 0.01; ***P < 0.001.

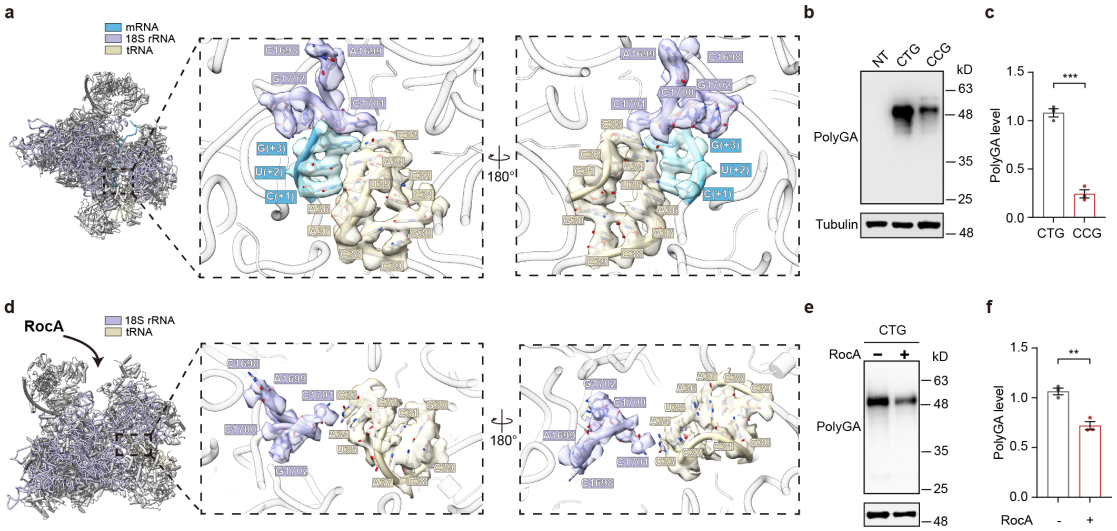

Figure 2. RAN translation initiates at a near-cognate CUG codon

a, Cryo-EM density at the 40S decoding centre showing a defined register in which a near-cognate CUG codon (cyan) is engaged by the initiator Met-tRNA$_i$ anticodon (gold) in the P site (peptidyl site). The codon–anticodon interaction is further stabilized by surrounding nucleotides, including C1701 of 18S rRNA (purple) and A36 of the tRNA. **b,c,** Mutation of the near-cognate start codon (CTG to CCG) markedly reduces poly (GA) production. **b,** representative immunoblot; **c,** quantification normalized to tubulin and to the CTG condition. **d,** Rocaglamide A (RocA) perturbs stable decoding-centre engagement. In the RocA-treated reconstruction, mRNA density at the P site is no longer resolved. **e,f,** RocA suppresses CTG-dependent poly (GA) production. **e,** Representative immunoblot of poly (GA) expression following RocA treatment. **f,** Quantification of poly (GA) levels normalized to tubulin and relative to vehicle control. For **c,f,** dots represent independent biological replicates; bars show mean ± s.e.m. Statistical significance was assessed using a two-sided unpaired t-test. **P < 0.01; ***P < 0.001.

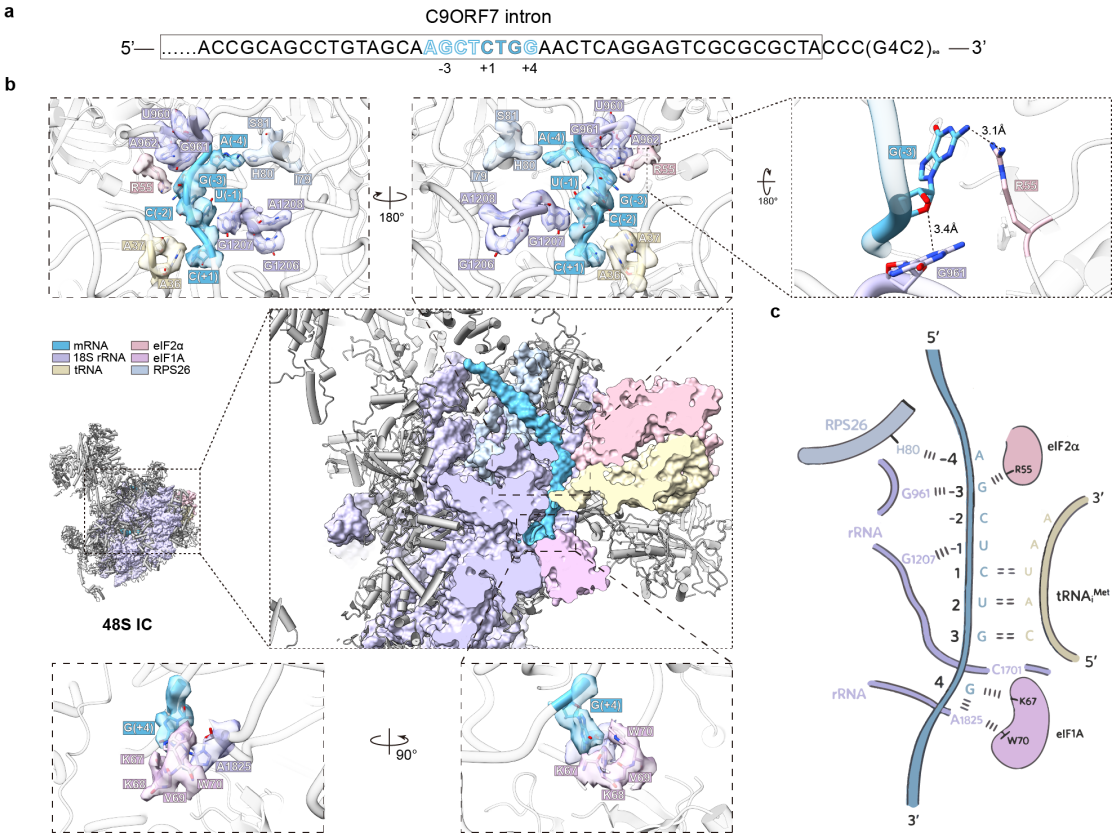

Figure 3. An endogenous Kozak-like context licenses non-AUG initiation

a, Sequence context of the C9ORF72 intronic repeat reporter showing a Kozak-like element upstream of the (G4C2)n tract. The near-cognate start codon CUG (cyan highlighted) is annotated by position relative to the first codon nucleotide (+1), with the hallmark Kozak positions -3 and +4 indicated. **b,** Cryo-EM density at the 40S decoding centre from the repeat mRNA-bound 48S initiation complex, revealing a defined mRNA register spanning nucleotides -4 to +4 around the CUG codon. Flanking nucleotides engage the 18S rRNA and ribosomal protein S26 (RPS26) and contact initiation factors eIF2α and eIF1A. **Right Top,** close-up view showing stabilizing hydrogen-bonding interactions of the -3 nucleotide with 18S rRNA G961 and eIF2α R55. **c,** Schematic summary of the interaction network that stabilizes the Kozak-like context and the CUG codon in the Peptidyl site, including contacts with 18S rRNA (G961, G1207, C1701 and A1825), RPS26 (H80), eIF2α (Arg55), eIF1A (K67 and W80) and pairing with the initiator Met-tRNA$_i$ anticodon.

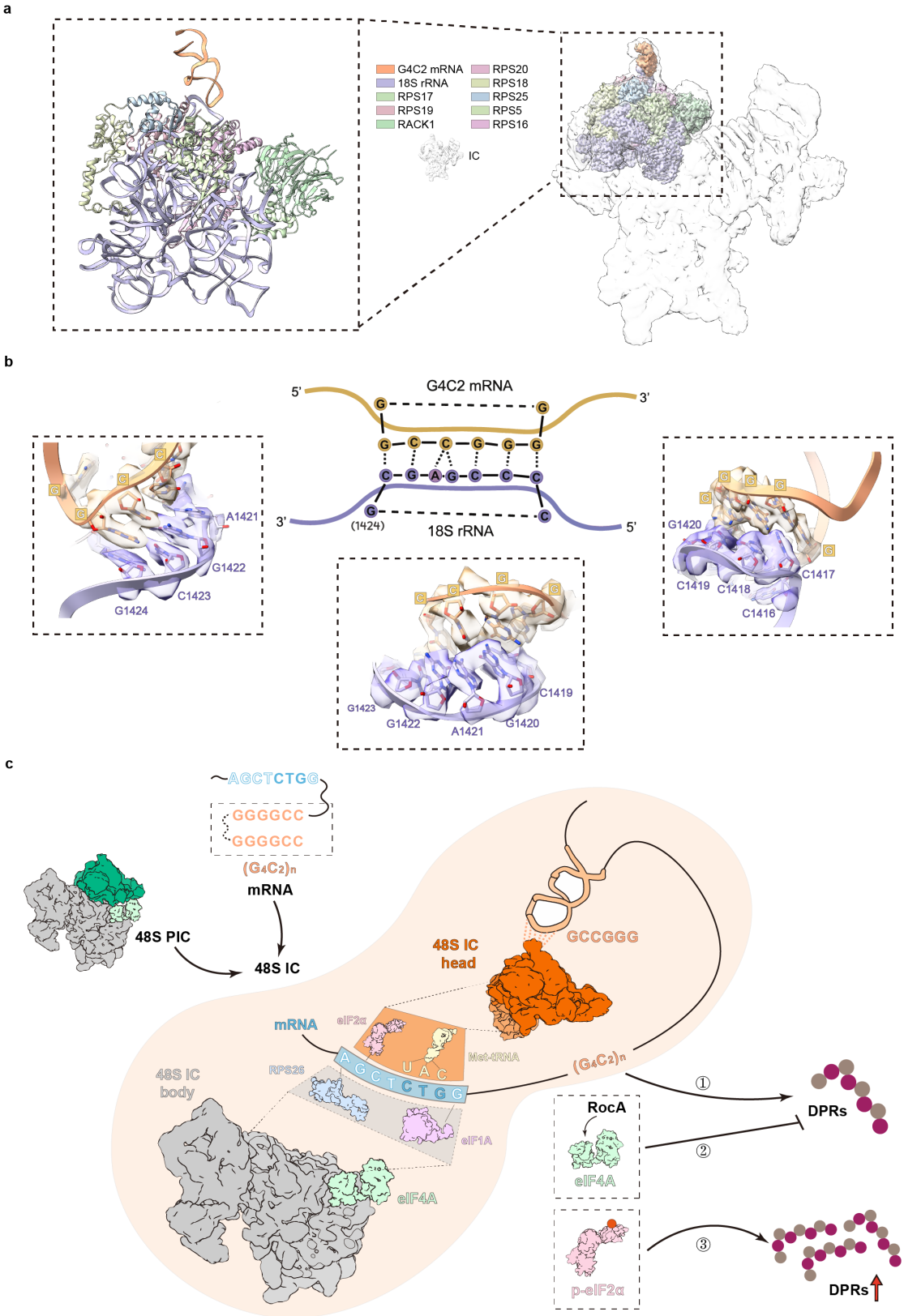

Figure 4. Repeat mRNA directly engages the ribosome and arrests the pre-initiation complex

a, Cryo-EM density and structure of the repeat mRNA-bound human 48S IC highlighting a contact between the G4C2 repeat mRNA (orange) and ES9S region of 18S rRNA on the 40S

head (purple). Ribosomal proteins surrounding this site (RPS2, RPS5, RPS16, RPS17, RPS18, RPS19, RPS20 and RPS25) and RACK1 are indicated as in the colour key. **b,** Structural basis for mRNA-rRNA anchoring. Centre, schematic of a short RNA-RNA helix formed by a near-complementary base pairing between a GCCGGG segment of the G4C2 repeat mRNA and GCAGCCC tract within the ES9S region of 18S rRNA. Insets, atomic model fitted into the corresponding cryo-EM density, highlighting Watson-Crick base pairs (C1417-C1419 and G1422-C1423) and stabilizing stacking interactions involving among G1420 and A1421 of 18S rRNA and C from G4C2 repeat mRNA. **c,** Purposed model linking rRNA anchoring to initiation-site selection and RAN translation output. A Kozak-like sequence upstream and a near-cognate CUG codon stabilizes mRNA engagement in the 48S IC. In addition, the G4C2 repeat can anchor the transcript through interact with the ES9S region of 18S rRNA, thereby stabilizing a trapped RAN initiation state conducive to DPR production. RocA reduces DPR output, consistent with disruption of eIF4A-dependent mRNA engagement, whereas eIF2α phosphorylation correlates with increased DPR production.

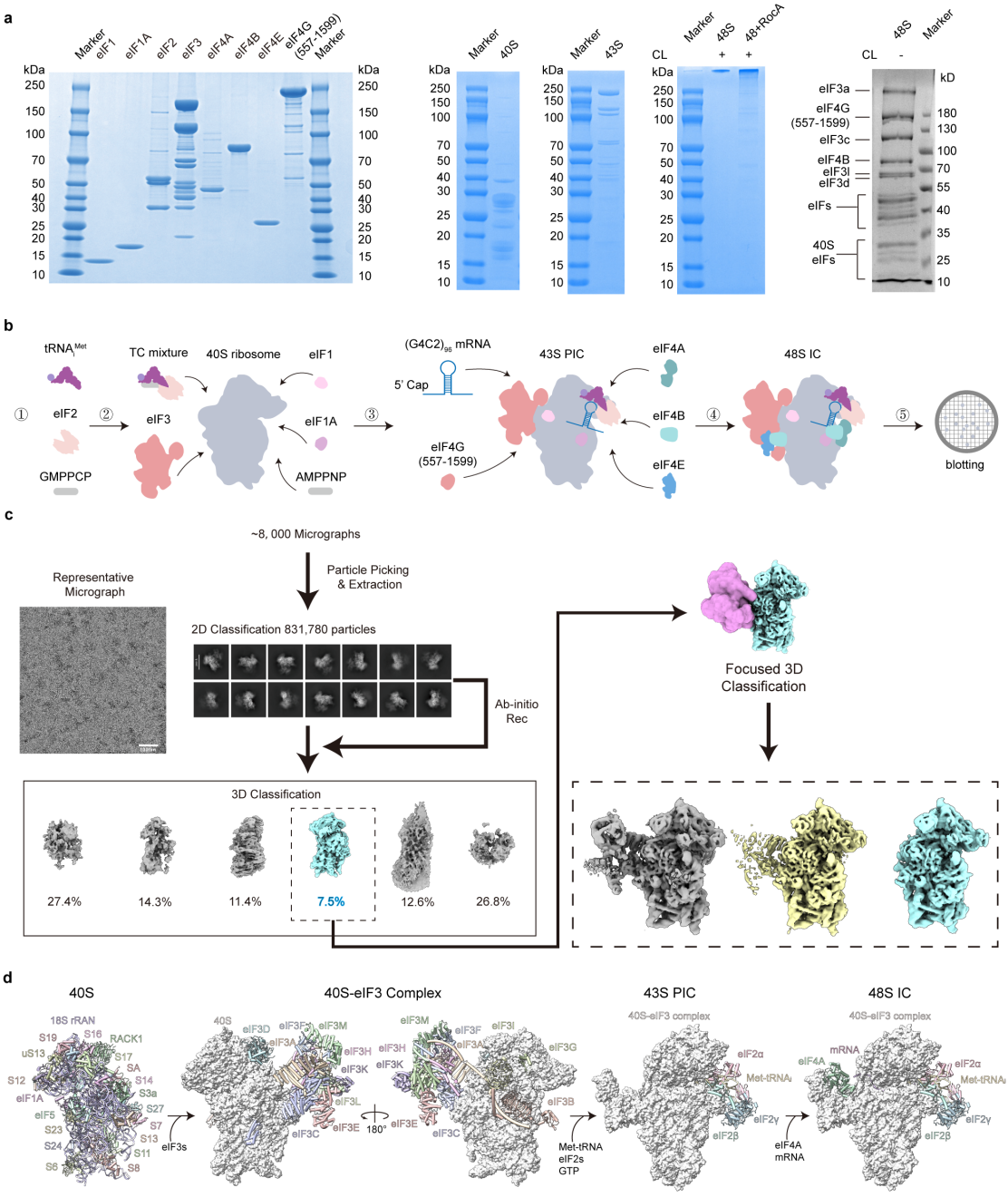

Figure S1. Hybrid reconstitution and cryo-EM classification of human RAN initiation complexes

a, Biochemical validation of the hybrid reconstitution system. Coomassie-stained SDS–PAGE of purified human initiation factors, native 40S subunits and reconstituted initiation complexes. **b,** Schematic of stepwise assembly of human initiation complexes on capped $(G4C2)_{96}$ repeat mRNA, from 43S pre-initiation complex (PIC) formation to repeat mRNA–bound 48S initiation complex (IC). **c,** Cryo-EM data-processing workflow used to resolve distinct initiation-complex states. **d,** Representative cryo-EM reconstructions and fitted models for the major classes, including free 40S, the 40S–eIF3 complex, the 43S PIC and the repeat mRNA–bound 48S IC. Key components are labelled.

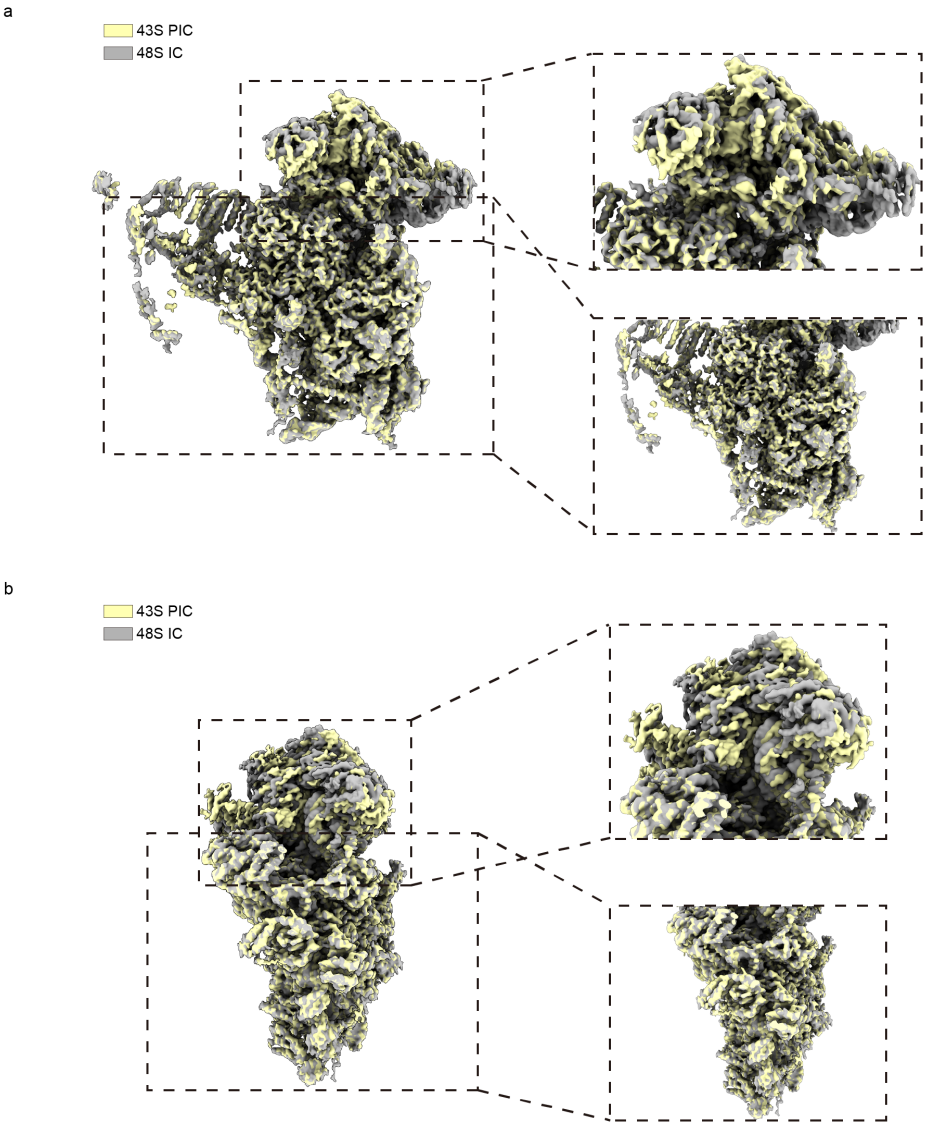

Figure S2. Cryo-EM density comparison of 43S PIC and 48S IC

a,b, Two orthogonal views of the superposed 43S PIC (yellow) and 48S IC (grey) highlight similar positioning of the 40S body but a pronounced conformational difference in the 40S head (zoomed views, right).

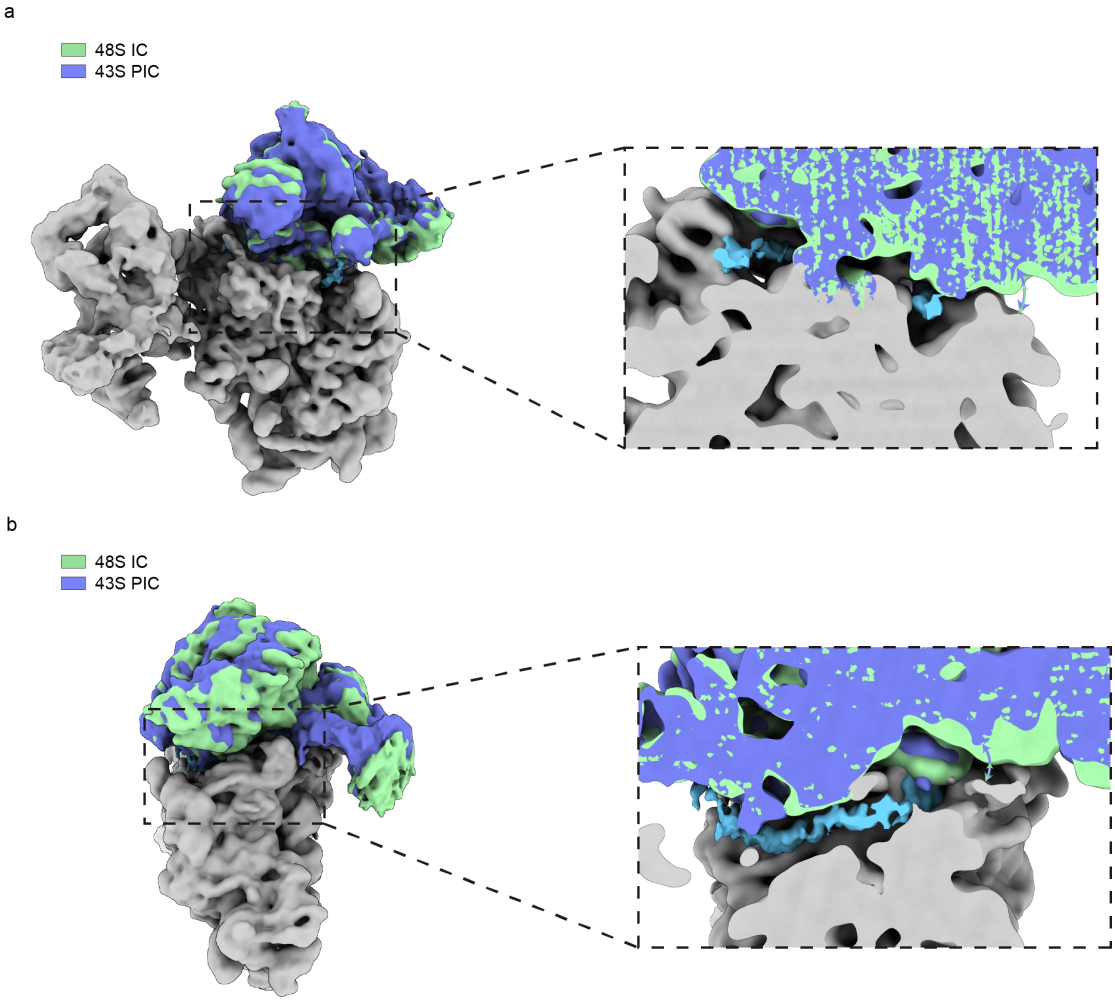

Figure S3. Comparison of the 43S PIC and 48S IC at the mRNA channel.

a,b, Two views of the superposed 40S heads from the 43S PIC (purple) and 48S IC (green) reveal distinct head conformations that alter the geometry of the mRNA channel.

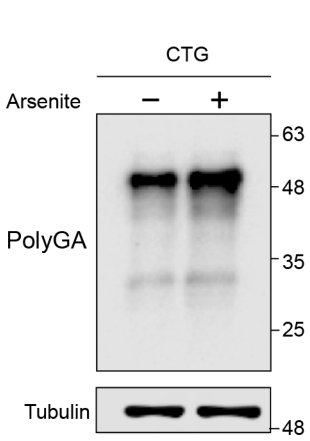

Figure S4. Arsenite enhances CTG-dependent poly(GA) production

a, Arsenite treatment increases CTG-dependent poly(GA) levels in cells expressing the reporter. Representative immunoblot showing poly(GA) and tubulin (loading control) ± sodium arsenite.

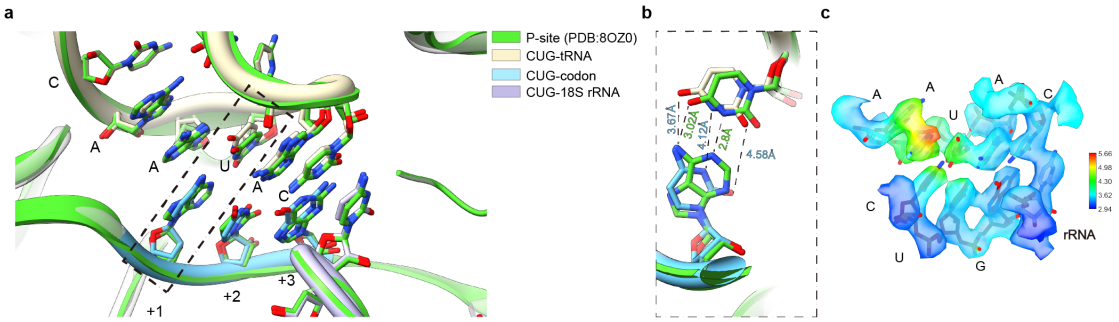

Figure S5. Comparison of P-site decoding with a canonical AUG codon and a RAN CUG codon.

a, Superposition of structures containing an AUG P-site codon and a CUG P-site codon. **b,** Zoomed views of the P-site region from a, showing the AUG (left) and CUG (right) codon–anticodon interactions. **c,** Local-resolution map of the CUG P-site region.

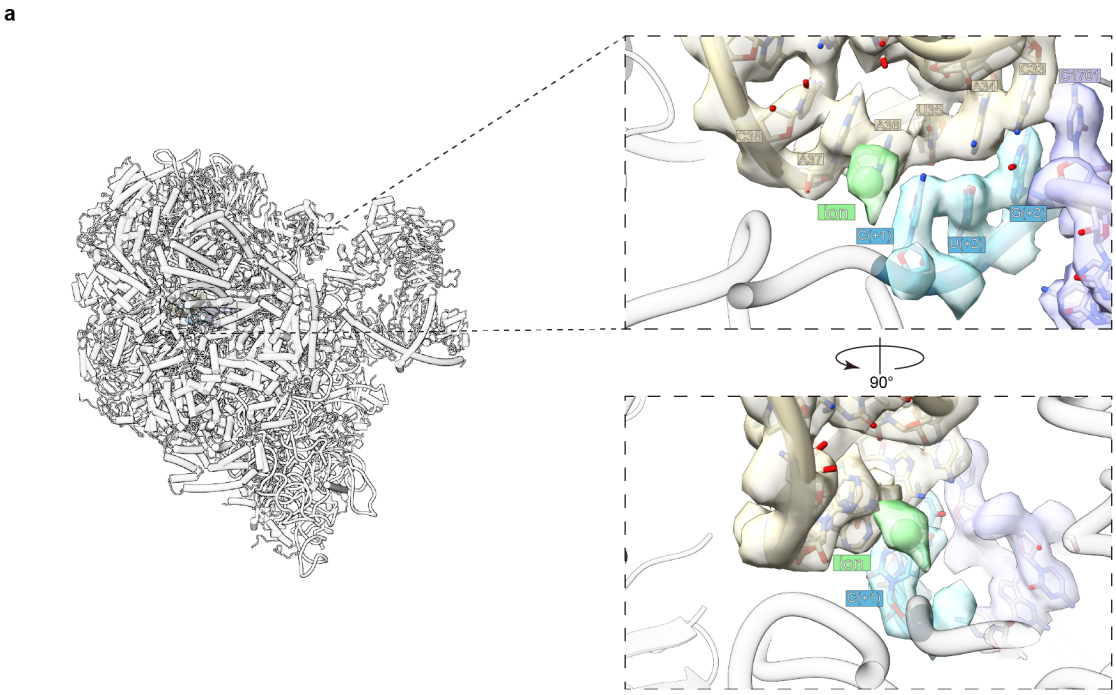

Figure S6. Model fit at the peptidyl site reveals an ion-like density.

a, Left, overall view of the 48S IC. Right, zoomed view of the peptidyl site showing the fitted atomic model and an ion-like density adjacent to the peptidyl site.

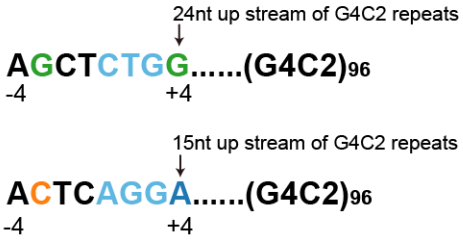

Figure S7. Co Comparison of two AUG-cognate initiation contexts upstream of the G4C2 repeat. Sequence alignment of the regions upstream of the G4C2 repeat showing the CTG context (24 nt upstream) and the AGG context (15 nt upstream).

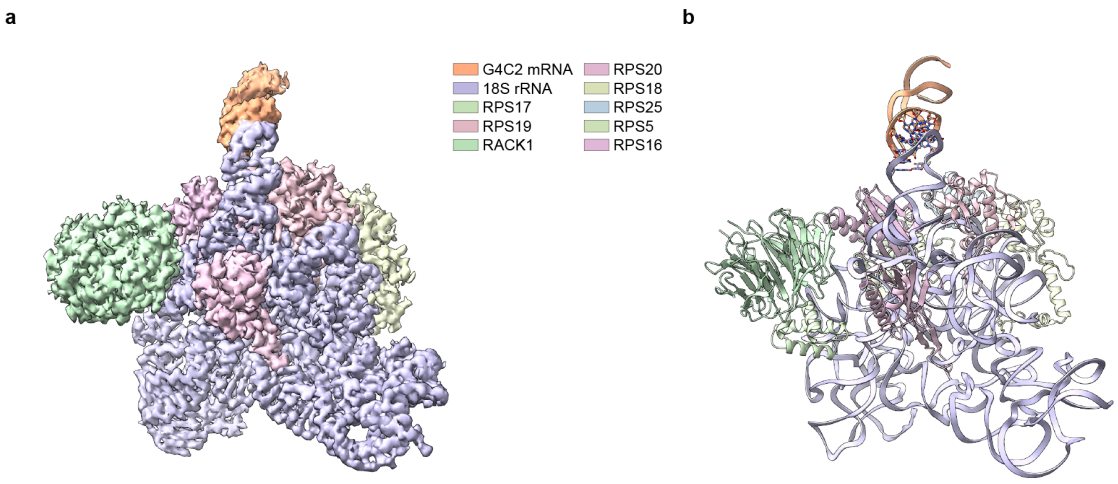

Figure S8. Alternative overall view of the G4C2 repeat mRNA-ES9S interaction

a, Cryo-EM density of the repeat mRNA-bound human 48S initiation complex (IC) highlighting the G4C2 repeat mRNA (orange) engaging the ES9S region of 18S rRNA on the 40S head (purple). **b,** Corresponding atomic model of the repeat mRNA-bound human 48S IC showing the same interaction between the G4C2 repeat mRNA (orange) and 18S rRNA ES9S (purple). Surrounding ribosomal proteins and RACK1 are coloured as indicated.