Structural basis for non-AUG translation regulation by 5MPs

Ximena Zottig[1,2*], Chun-Ying Huang[1*], Zahra Seraj[1], Nikolaus Grigorieff[1,2#], Andrei A. Korostelev[1#]

[1] RNA Therapeutics Institute, UMass Chan Medical School, 368 Plantation Street, Worcester, MA 01605, USA.

[2] Howard Hughes Medical Institute, UMass Chan Medical School, 368 Plantation Street, Worcester, MA 01605, USA.

* These authors contributed equally

Co-corresponding authors: niko@grigorieff.org; Andrei.Korostelev@umassmed.edu

Abstract

The cellular proteome is regulated by translation initiation on AUG or non-canonical (non-AUG) start codons[1-3]. Non-AUG initiation remodels proteome during stress and is implicated in cancer and other diseases[4-6]. The eIF5-mimic proteins (5MPs) restrict non-AUG start codon usage and thereby reprogram proteoform expression from mRNAs with alternative start sites, such as the oncogenic c-Myc[7-10]. The mechanism by which 5MPs induce such translational reprogramming remains unknown. Here, using *in extracto* cryo-electron microscopy (cryo-EM) and biochemical assays, we report that translational repression by 5MP strongly depends on the sequence context near the AUG or non-AUG codons. Cryo-EM structures of 5MP-bound 48S pre-initiation complexes (PICs) from native cell extracts reveal that 5MP binds at the A site of the small ribosomal subunit, stabilizing an expanded open-head conformation of the PIC scanning along mRNA. The N-terminal region of 5MP blocks the A site, whereas the C-terminal domain docks at eIF2β and the initiator $tRNA^{Met}$ outside the P site (*i.e.*, P_{out}). These findings indicate that 5MP protein directly biases the initiating 48S complexes toward the open conformation, promoting mRNA scanning and inhibiting initiation at suboptimal start codons.

Main

Translation initiation is a major regulatory step of gene expression, and its dysregulation contributes to a wide range of human diseases, including cancer and neurodegeneration[11-16]. During initiation, the small ribosomal 40S subunit, together with initiation factor proteins (eIFs) and initiator $tRNA^{Met}$, collectively termed the 48S pre-initiation complex (PIC), scans the 5' untranslated region (UTR) of an mRNA until it encounters the initiation (start) codon[17-19]. On the start codon, the small subunit associates with the large 60S subunit to commence protein synthesis. Although the AUG triplet is the predominant start codon, recent ribosome profiling and other studies revealed extensive initiation from non-AUG codons, expanding proteome complexity and adaptive responses to stress[20-25]. Moreover, switching between AUG and non-AUG codons within a single mRNA can control the output of distinct proteoforms[1,26,27]. For example, the increased initiation stringency, favouring AUG codons, in c-Myc and other mRNAs enhances translation of tumor-promoting transcripts[9,28-32]. Likewise, AUG codons in suboptimal contexts (*i.e.*, sequences near the start codon, also known as Kozak sequences, as detailed below) are regulated by initiation stringency, further shaping the proteome[33].

The eIF5-mimic proteins 5MP1 and 5MP2, conserved across eukaryotes except for yeasts and nematodes[34], and encoded by *BZW2/1* genes in humans, regulate the selection of AUG start codons by suppressing initiation at non-AUG codons[8,35,36]. Their C-terminal W2/HEAT3 domain is homologous to that of eIF5, an initiator factor promoting 48S progression toward subunit joining upon start-codon recognition. Accordingly, 5MP proteins (5MPs) are thought to antagonize eIF5 and bind the W2-interacting eIF2β subunit of eIF2, whose rearrangement is critical for 48S PIC switching from the scanning to start-codon recognition mode[7,8,35-37]. 5MP1 and 5MP2 sequences and cellular phenotypes are similar, indicating that they perform similar molecular roles while differing in cell/tissue expression, localization, and/or other aspects of cellular control[8,34,36,38](GTEx; Human Protein Atlas). In most cultured cells, the dominant paralog is expressed at levels comparable to those of core initiation factors under homeostatic conditions[39]. Elevated expression of 5MPs correlates with tumorigenesis, metastasis, and poor prognosis[9,40-44].

The oncogenic activity of 5MPs is attributed at least in part to the direct upregulation of c-Myc, an oncogene that drives malignant progression. Specifically, increased 5MP1 levels suppress the long c-Myc isoform translation from a CUG start codon and upregulate expression of a shorter, more tumorigenic isoform from a downstream AUG codon[9]. Furthermore, the critical roles of 5MPs in start-site selection have been documented in embryonic development, neuronal homeostasis, and neurodegeneration[33,36,45].

Despite the critical cellular functions and therapeutic relevance, the molecular basis of 5MPs action remains unknown due to the absence of structural data on 5MP interactions with initiation factors or 40S subunits. To elucidate the molecular and structural mechanisms of human 5MPs, we measured their activities in distinct mRNA contexts and performed *in extracto* cryogenic electron microscopy (cryo-EM) of translating mammalian cell lysates. In the 2.7-Å structure of a 48S pre-initiation complex, 5MP1 stabilizes the open, scanning-competent conformation of the PIC through coordinated interactions with the 40S, initiator $tRNA^{Met}$, and initiation factors eIF2β and eIF1A. Consistent with biochemical findings, 5MP proteins bias the open 48S conformation to prevent the commitment on suboptimal initiation sites, thereby favoring AUG-initiated isoforms of c-Myc and other mRNAs. Together, these findings reveal the structural basis for modulation of initiation fidelity by 5MP proteins and may pave the way for therapeutic interventions to correct dysregulated translation initiation.

Results and discussion

5MP binds 48S and inhibits initiation on mRNAs with non-AUG and poor Kozak-context starts

To visualize native interactions of the translation initiation machinery with endogenous cellular components, we performed cryo-EM of mammalian cell extracts (hence, *in extracto* cryo-EM) with high-resolution two-dimensional template matching (2DTM), enabling particle recognition in low-contrast environments[46,47]. We recently used this approach to discover novel interactions in near-atomic-resolution 80S ribosome reconstructions[48,49]. In this work, we aimed to characterize initiation complexes formed by 40S subunits in mammalian cell extracts (Methods). 3D classification of particles identified by 2DTM, using a 40S body template, revealed a class whose A site contained density that could not be assigned to known translation factors (Extended Data Fig. 1a). This class corresponds to a 48S initiation complex, as indicated by the presence of initiation factors, initiator $tRNA^{Met}$, and mRNA (Extended Data Fig. 1b; Methods). Backbone and side-chain tracing with ModelAngelo[50], followed by comparison with the AlphaFold-predicted structure library[51,52], indicated that the density can be accounted for by a 5MP protein (Extended Data Fig. 1c-d; Methods). This finding suggests that 5MPs may directly and efficiently control translation initiation from AUG or non-AUG codons by binding to the A site. To corroborate this idea and further elucidate their mechanisms, we purified recombinant human 5MPs (Extended Data Fig. 2) to perform biochemical and cryo-EM analyses.

To quantify 5MP-mediated regulation, we measured translation of a nanoluciferase (NLuc) reporter mRNA with an AUG or CUG start codon within the strong, moderate, or weak Kozak contexts[53,54]. For example, a strong Kozak context with -3 purine and +4 guanosine (where +1 is the first position of the start codon) results typically in efficient translation, whereas -3 and +4 pyrimidines render the start codon inefficient[53,55]. We added mRNA reporters to a rabbit reticulocyte lysate (RRL) translation system in the presence or absence of recombinant 5MPs and measured luminescence (Fig. 1a). Translation from an AUG codon in a strong Kozak context was unaffected in the presence of 1 μM 5MP, underscoring the robustness of canonical initiation (IC_{50} > 5 μM). By

contrast, translation from moderate and weak AUG contexts was noticeably reduced in the presence of 1 µM 5MP (Fig. 1b). 5MP1 inhibited translation from a weak AUG context more efficiently than did 5MP2 (IC_{50} values of 114 nM and 410 nM, respectively) (Fig. 1b, d). Translation from CUG codons—especially in moderate and weak contexts—was even more efficiently inhibited by 5MPs at low nanomolar IC_{50} values (5MP1: 35–11 nM; 5MP2: 131–79 nM; Fig. 1d; Extended Data Fig. 3a). The inhibitory effect of 5MP1 can therefore be >100-fold stronger on CUG codons than on the strong AUG codon.

5MP proteins autoregulate their translation from an AUG in a conserved weak Kozak context[35]. We therefore measured the ability of 5MP1 to inhibit translation from the UUUAUGA context of 5MP1 (*BZW2*) mRNA (Extended Data Fig. 3b). Indeed, 5MP1 inhibited translation from its own AUG with an IC_{50} of ~253 nM, underscoring the ability of 5MP1 to repress a moderate to weak AUG context (Extended Data Fig. 3b-d). In cells, this feedback mechanism likely controls 5MP1 concentrations[56] to enable differential translation from AUG and non-AUG codons across the transcriptome.

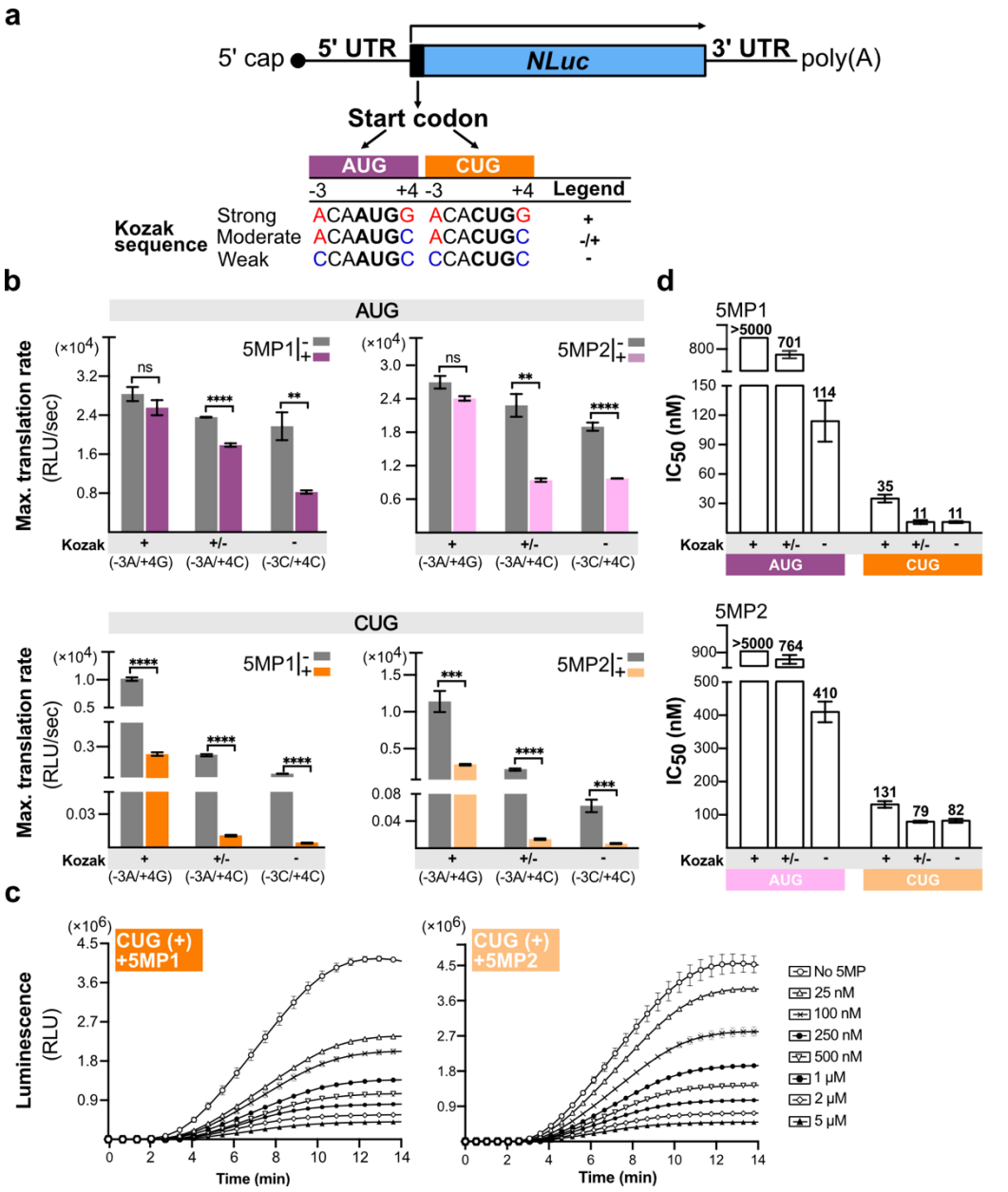


Figure 1: 5MP proteins inhibit translation initiation in a Kozak sequence-dependent manner.

(a) Schematic of mRNA constructs containing either AUG or CUG start codons within a Kozak sequence context, positioned upstream of the nanoluciferase (NLuc) coding sequence, tested in a mammalian translation system. **(b)** Effects of 5MPs at 1 μM on translation from AUG (purple) and CUG (orange) start codons in strong (+, –3A/+4G), moderate (+/-, –3A/+4C), and weak (-, –3C/+4C) Kozak contexts (mean with SEM, n=3). The statistical difference between the translation rates is shown, where ****P < 0.0001, ***P < 0.0002, **P < 0.0021, *P < 0.00332, and ns (non-significant) according to the unpaired two-tailed t-test. Left: 5MP1; Right: 5MP2. **(c)** Time progress curves are shown for strong Kozak CUG initiation at increasing 5MP concentrations (25 nM, 100 nM, 250 nM, 500 nM, 1 μM, 2 μM, and 5 μM). **(d)** Inhibitory concentrations (IC_{50}) of 5MP1 (top) and 5MP2 (bottom) for AUG and CUG initiation in strong (+, –3A/+4G), moderate (+/-, –3A/+4C), and weak (-, –3C/+4C) Kozak contexts.

5MP regulates proteoform output by start-site selection

5MP1 drives oncogenesis, at least in part, by reprogramming expression of c-Myc from a p67 proteoform initiated at a CUG start codon to an oncogenic p64 proteoform initiated at an in-frame downstream AUG codon (Fig. 2a)[9]. The “p67” CUG start codon is in a strong Kozak context, and the “p64” AUG start codon is in a moderate Kozak context. To dissect the effects of 5MP on each start codon, we generated a reporter mRNA comprising the c-Myc 5′UTR and first 48 nucleotides of the p64 coding sequence fused to 3×FLAG and NLuc, and we measured luminescence (Fig. 2b-e) and proteoform levels (Fig. 2f-h) in the absence or presence of increasing concentrations of 5MP1. Consistent with studies in cell lines[9,36], 5MP1 increased the total expression of the c-Myc construct (Fig. 2b,c) and shifted expression toward the shorter AUG-dependent proteoform by inhibiting the expression from the p67 CUG start codon (Fig. 2f). Mutations that disrupt the AUG or CUG codon individually confirm this trend (Fig. 2d-e,g-h; Extended Data Fig. 4a). In the absence of the AUG codon, 5MP1 suppresses CUG translation at micromolar concentrations (Fig. 2d). By contrast, AUG initiation from its native position in the c-Myc 5′UTR is only mildly affected by micromolar 5MP1 (Fig. 2e). Together, our data demonstrate that 5MP efficiently changes the ratio of CUG and AUG proteoforms of c-Myc at 5MP concentrations that are above its autoregulatory IC_{50} (Fig. 2c, 2f).

On the c-Myc untranslated region with both start codons, 5MP1 not only repressed the production of the CUG isoform but also increased the AUG isoform output (Fig. 2f). We wondered whether the suppression of the AUG isoform in the absence of 5MP1 might be caused by inefficient initiation on the CUG start codon, inefficient elongation between the CUG and AUG codons (*e.g.*, due to suboptimal, or “rare”, codons), or both. Mutation of the CUG start codon to AUG (*i.e.*, AUG-p67) increased the production of the longer isoform by >2-fold to a level that was nearly equivalent to that of the short AUG-p64 (Extended Data Fig. 4b-c). This comports with biochemical studies showing that AUG initiation is at least twice as efficient as initiation on near-cognate UUG and CUG start codons in the same sequence context[57-62]. This is consistent with the less efficient initiation from the CUG-p67 start codon, resulting in the lower overall c-Myc production. In summary, our data demonstrate that 5MP proteins can both *(i)* switch translation from

a less efficient to a more efficient start codon and *(ii)* increase overall production of the shorter isoform.

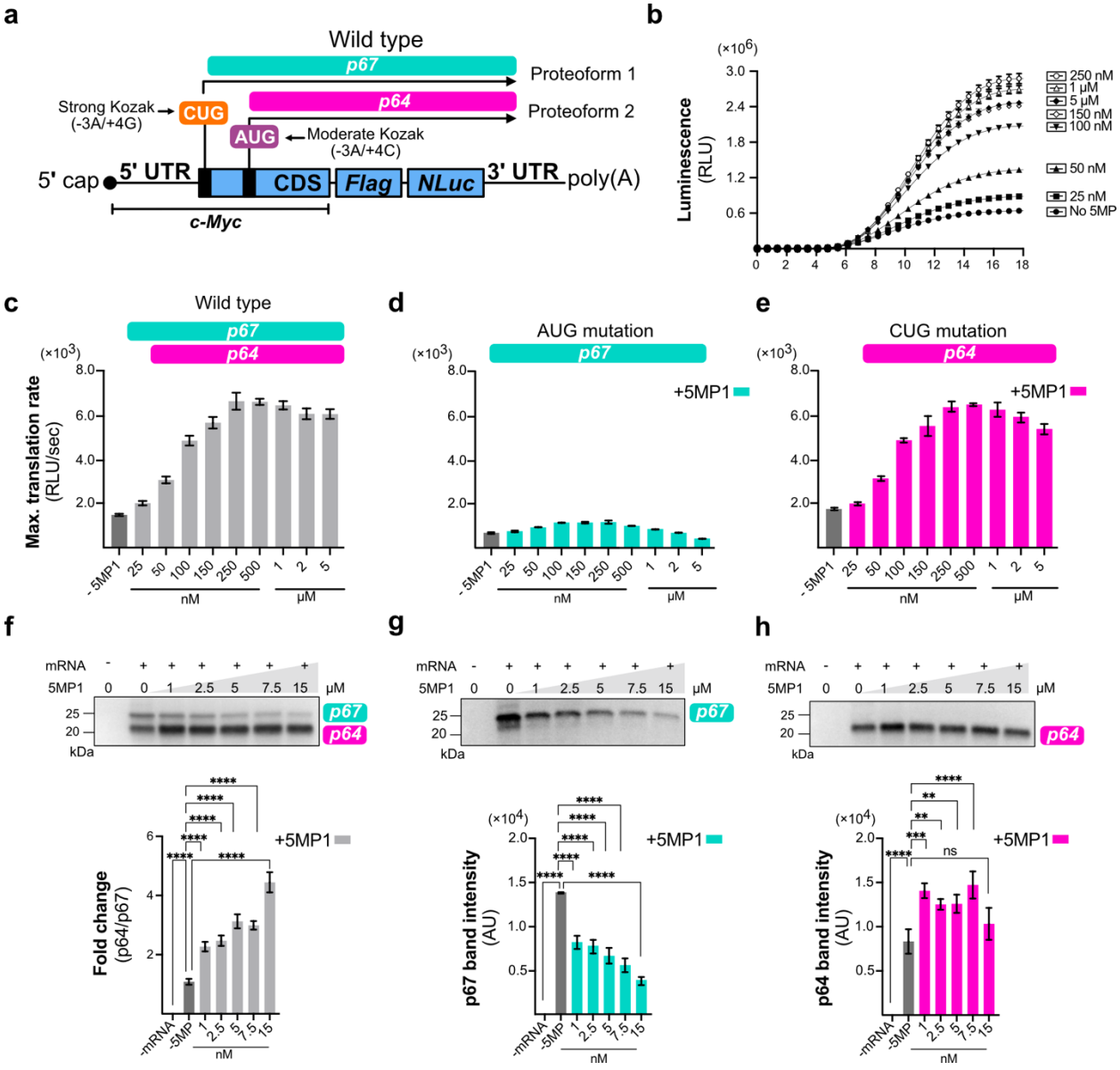


Figure 2: 5MP1 reprograms initiation from the c-Myc 5' UTR.

(a) Schematic of the mRNA reporter with c-Myc 5' UTR containing CUG and AUG start codons in distinct Kozak contexts. AUG-initiated p64 products are indicated in pink and CUG-initiated p67 products in green. **(b)** Time courses of c-Myc 5' UTR translation in the presence of increasing concentrations of 5MP1. Data represent mean ± SEM values (n = 3). **(c-e)** Maximum translation rates plotted at different 5MP1 concentrations for c-Myc wild type (c), CUG-initiated (AUG-mutation; d), and AUG-initiated (CUG mutation; e) reporters. Data represent mean ± SD values (n = 3). **(f-h)** Translation products from wild type (f), AUG-mutated (g), and CUG-mutated (h)

constructs. Representative western blotting (top) and quantifications (bottom) are shown: (left) fold change in proteoform ratio (p64/p67); (middle and right) p67 and p64 relative band intensities. Statistical significance was assessed by one-way ANOVA; ****P < 0.0001, ***P < 0.0002, **P < 0.0021, ns = not significant.

5MP1 stabilizes scanning-competent 48S PIC

To investigate the structural basis of 5MP-mediated control of initiation, we collected a cryo-EM dataset from RRL supplemented with purified recombinant human 5MP1, an mRNA bearing an AUG codon in a moderate Kozak context, and the non-hydrolyzable GTP analogue Guanosine 5′-[β,γ-imido]triphosphate (GMPPNP) to capture initiation complexes with GTPase initiation factors eIF2 and/or eIF5B. To enable particle detection despite the low contrast of lysate micrographs, we employed 2DTM[46,47] using the 40S body as the search template (PDB: 7R4X; Extended Data Table S1, Methods). The detected particles comprised a range of structurally distinct PICs and 80S complexes, indicating that 2DTM was mostly unbiased by the incomplete-subunit template. Cryo-EM data classification yielded predominantly 48S PICs (~31%), containing eIF1, eIF1A, eIF2 and eIF3. Nearly half of PICs contained 5MP1, of which 37% are 48S complexes and 12% are 48S-like assemblies containing the N-terminal domain of eIF3c subunit near eIF1 but missing the rest of eIF3 (Extended Data Fig. 5b-d and Extended Data Fig. 6). These observations are consistent with eIF3c involvement in 5MP binding to the 48S complex[36]. Other classes included free 40S subunits and 40S complexes containing tRNA, eIF1, eIF1A, and eIF2 (Extended Data Fig. 5b).

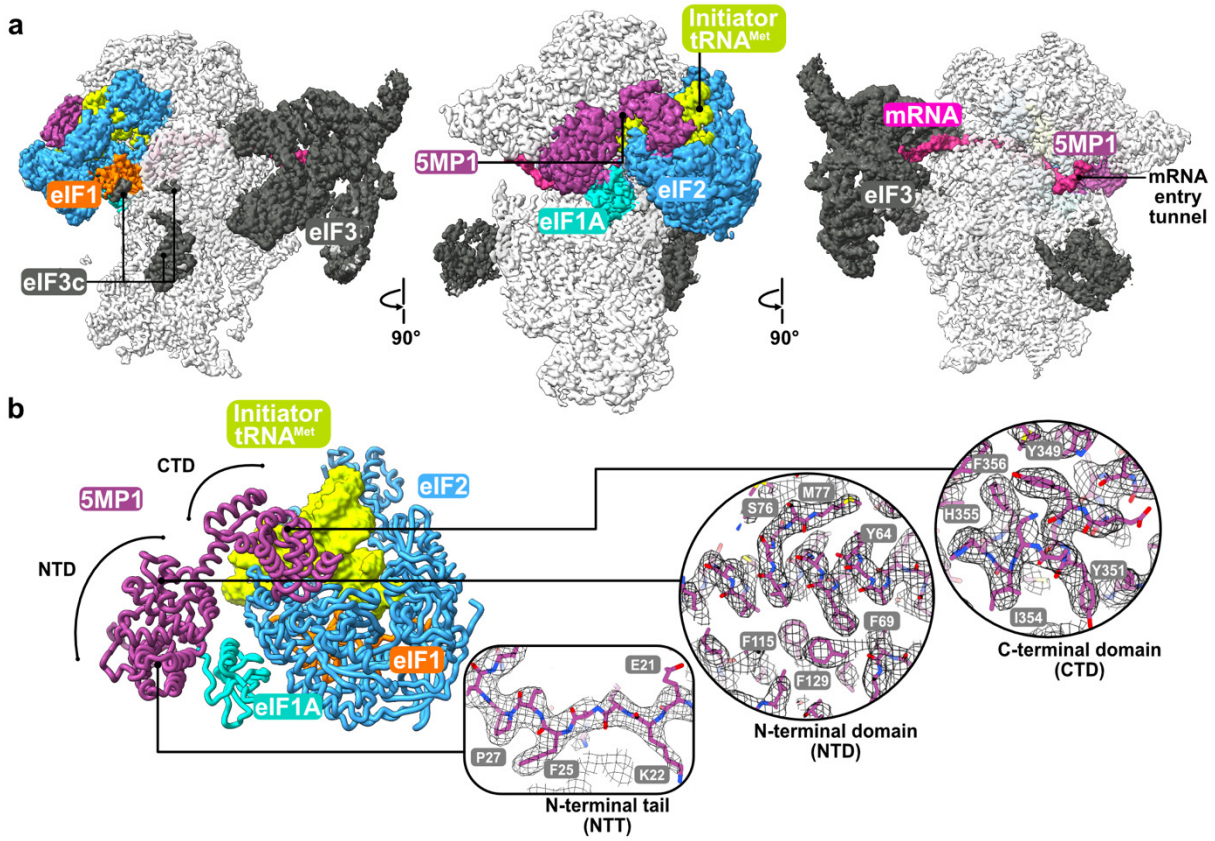


Figure 3: Cryo-EM of 5MP1 bound to 48S in RRL.

(a) Cryo-EM structure of the 48S initiation complex, shown in different orientations. Initiation factors are shown as segmented densities. 5MP1 is shown in purple, eIF2 in blue, eIF1 in orange, eIF1A in turquoise, and the initiator tRNA^{Met} in green. The middle view focuses on the 5MP binding at the A-site. The intersubunit side (right) and solvent side (left) view of the 48S initiation complex. **(b)** 5MP1 interactions with initiation factors and tRNA. Insets show the N-terminal tail (NTT), N-terminal domain (NTD), and C-terminal domain (CTD); cryo-EM density is shown as mesh.

A 2.7-Å reconstruction of the predominant 48S•5MP1 complex features 5MP1 at the A site (Fig. 3), placed identically to the endogenous 5MP detected in cell lysates described above (Extended Data Fig. 1). Notably, cryo-EM data contain both head-open (*i.e.*, scanning-competent) and head-closed (*i.e.*, codon-recognition) 48S PICs, but 5MP1 is only bound to 48S PIC in the head-open conformation (Fig. 4; Extended Data Figures 5 and 7). In canonical head-open initiation complexes, the head is tilted ~13° away from the body, relative to that in head-closed (codon-recognition) complexes [63-65]. In 5MP1-bound complexes, the 40S head is tilted away by ~15°, indicating that 5MP1 stabilizes a more

widely open state and consequently features a larger mRNA tunnel compared to the canonical scanning conformation (Methods)[63,64]. The initiator $tRNA^{Met}$ is positioned ~7 Å away from its codon-engaged position, nearly identical to P_{out} conformations observed in scanning-competent 48S PICs (Fig.4b)[66,67]. Although relatively continuous mRNA density is visible in the mRNA tunnel within the 5MP1-bound complexes, individual nucleotides are not resolved near the P site, consistent with the heterogeneity observed in scanning 48S complexes (Fig.4c; Extended Data Fig. 8)[17,68]. This contrasts with a well-resolved AUG codon in our 2.8-Å reconstruction of codon-recognition head-closed 48S complexes lacking 5MP1, which comports with that observed in previous studies of canonical AUG initiation (Fig. 4a,c; Extended Data Fig. 8)[17,68].

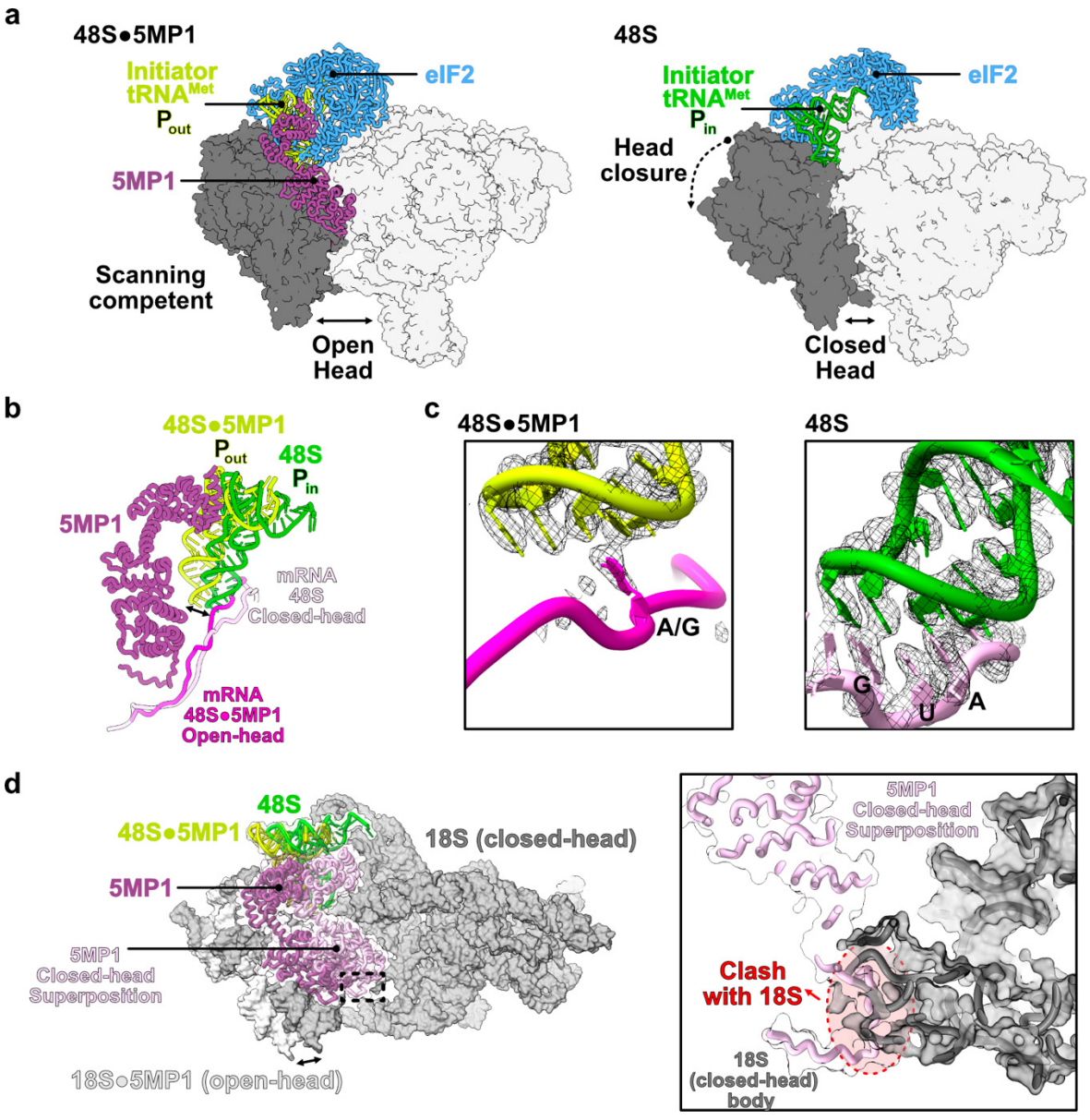


Figure 4: 5MP1 promotes the open-head conformation of the 48S complex.

(a) The 48S complex containing 5MP1 adopts an open conformation (left), whereas the 48S complex without 5MP1 is predominantly closed (right). In the 5MP1–bound structure, eIF2 (blue), tRNA (green), and 5MP1 (purple) are shown in cartoon representation. For comparison, the tRNA in the 48S complex without 5MP1 is shown in dark green. **(b)** In the presence of 5MP1, the tRNA resides in a P_{out} conformation (left), whereas in the 48S complex without 5MP1, it occupies a P_{in} conformation (right). With 5MP1 bound, the initiator $tRNA^{Met}$ sits 7 Å away from the codon-engaged position, closely resembling the canonical P_{out}. The mRNA in the 5MP1-bound complex is shown in pink, and in the 48S complex without 5MP1 in light pink. **(c)** The close-up reveals that codon–anticodon pairing is absent in the 5MP1-bound complex, with mRNA density supporting only assignment of the +1 nucleotide (purine; A/G). By contrast, in the 48S complex without 5MP1, the AUG start codon is well-resolved, consistent with codon–anticodon pairing. **(d)** Overlay of the 5MP1 model with the 48S closed-head conformation (superposition), showing a steric clash of

5MP1 N-terminal tail with helix 18 of 18S rRNA. Experimental 5MP1 model is shown in purple and superimposed 5MP1 in lilac.

5MP1 consists of two domains, which contact eIF2, initiator tRNA, eIF1A and mRNA tunnel (Fig. 5a). The C-terminal domain (residues 248-407) docks onto eIF2β and initiator $tRNA^{Met}$. Here, Tyr351 and Ile354 of the W2/HEAT3 domain—previously implicated in interactions with eIF2β[36,37]—bind the universally conserved His230 of eIF2β through π–π stacking (Tyr351) and hydrophobic packing (Ile354) (Fig. 5b). Residues 401–405 of 5MP1, implicated in the interaction with the eIF2β N-terminus[36], are resolved in our map, but there is no density for the eIF2β N-terminus, leaving open the possibility of transient interactions (Fig.3; Extended data Fig. 7c).

Residues Lys309 and Lys310 of 5MP1 reach toward the phosphate groups of D-loop and variable-region nucleotides 10 and 49 of initiator $tRNA^{Met}$, in the vicinity of eIF2β (Fig. 5c). Notably, 5MP1 does not overlap with the binding site of eIF5 N-terminal domain near the tRNA anticodon stem[59], indicating that these proteins do not compete for their primary sites on the 40S subunit. Nevertheless, the C-terminal domains of 5MP1 and eIF5 may compete for binding to eIF2β and thus delay the recruitment of eIF5 during codon recognition[37].

Lastly, the 40S A site accommodates the 5MP1 N-terminal domain (NTD) next to eIF1A, stabilized by cation-π interactions between 5MP1 (Arg120 and Arg121) and eIF1A (Tyr106 and Phe28; Fig. 5d). The N-terminal tail (NTT; residues 1-19) extends into the mRNA tunnel (Fig. 5e), where the terminal residues form a β-stand with protein uS3 and the conserved Arg17-Lys18-Arg19 motif contacts the mRNA backbone. Transitioning to the closed-head codon-recognition state would require displacement of NTT to prevent the positively charged Arg-Lys-Arg motif of NTT from clashing with helix 18 of the 18S rRNA (Fig. 4d) [64,69]. This steric hindrance is consistent with the absence of 5MP1 in the closed-head maps. Taken together, 5MP1 interactions within 48S bias the head-open PIC toward scanning.

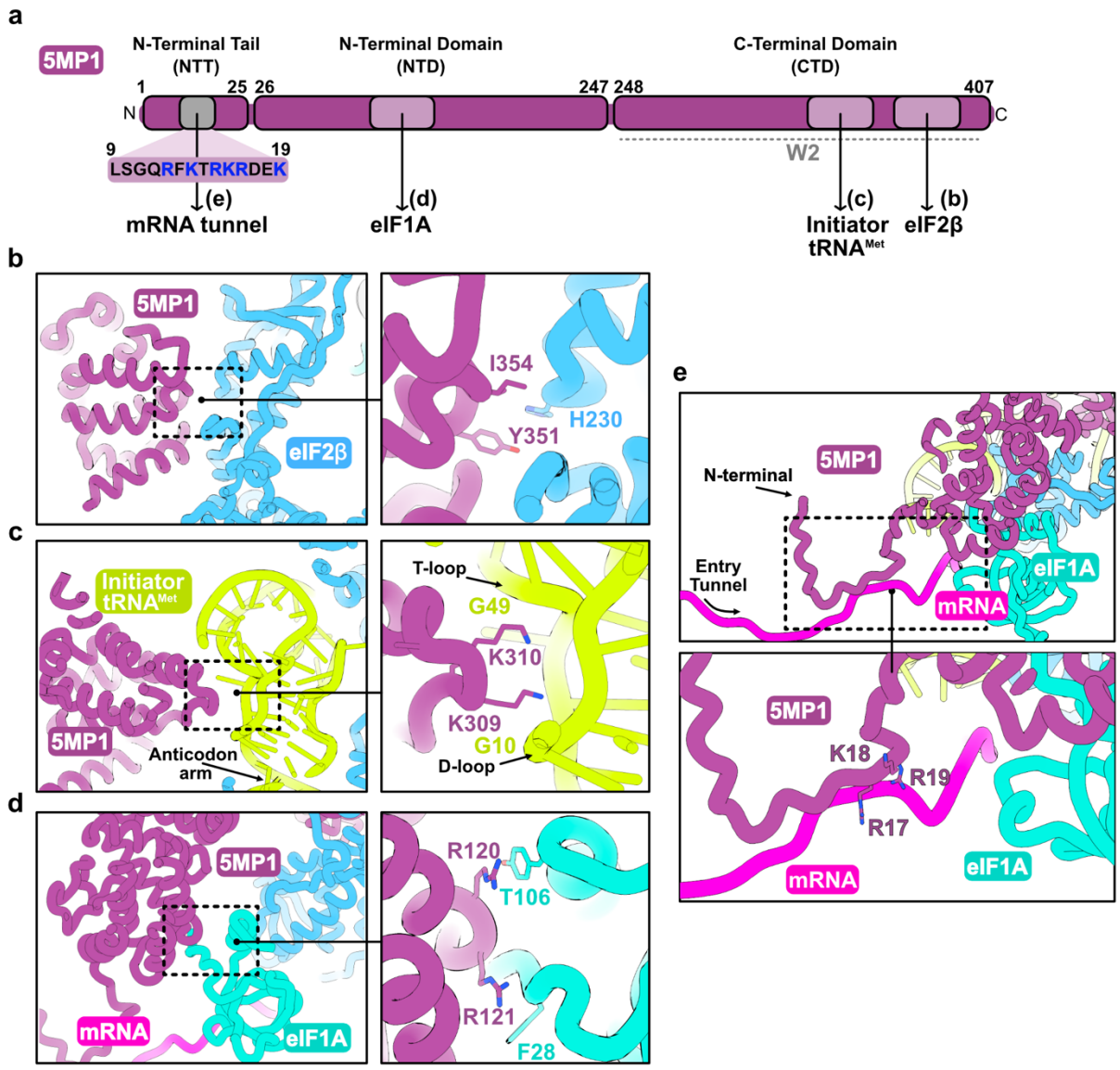


Figure 5: 5MP1 docks the scanning 48S conformation through a network of interactions.

(a) Schematic representation of 5MP1 showing its domain organization, interactions within the 48S complex and conserved sequence in the N-terminal tail (NTT). **(b)** Close-up view of the 5MP1–eIF2β, **(c)** 5MP1–Initiator tRNA^{Met} and **(d)** 5MP1–eIF1A interface, highlighting the interacting residues. **(e)** Close-up view of the N-terminal tail (NTT) highlighting the Arg-Lys-Arg motif facing the mRNA backbone (right). 5MP1 is shown in purple, eIF2 in blue, eIF1A in turquoise, and the initiator tRNA^{Met} in green.

The mechanism of initiation regulation by 5MPs

The molecular mechanism of eIF5-mimic proteins has long remained unknown, despite their key roles in regulating translation initiation and in diseases such as cancer. Our work shows that 5MP potently enforces start-codon fidelity by inhibiting initiation at alternative start sites, such as non-AUG codons or AUG in suboptimal Kozak contexts (Fig. 1). In the c-Myc context, this regulation shifts translation from a mixture of AUG- and non-AUG-initiated proteoforms toward predominantly the AUG-initiated form, while also enhancing overall translation, a dual effect with potential relevance in disease (Fig. 2). The cellular concentrations of 5MPs are regulated by the weak start codons in 5MP-encoding genes, offering a surveillance mechanism for accurate selection of AUG codons and switching between proteoforms.

To further delineate 5MP function in the context of cellular components, we performed *in extracto* cryo-EM, achieving near-atomic resolution for native 5MP-bound initiation complexes (Fig.3). Cryo-EM analyses reveal that 5MP acts as an additional initiator factor bound exclusively within the head-open 48S complexes, suggesting a straightforward structural mechanism of 5MP action (Fig. 4; Fig. 6). To reduce initiation at a suboptimal start site, 5MP shifts the equilibrium between 48S conformations toward a widely open, scanning-competent state. Within 48S, 5MP interacts with eIF2β, initiator $tRNA^{Met}$, eIF1A and mRNA tunnel (Fig. 5). The transition of 48S into the closed codon-recognition state (*e.g.*, on a strong AUG codon) requires eIF2β displacement[68] and the closure of the 40S head. This must coincide with or succeed 5MP dissociation, as is indicated by the absence of 5MP1 in closed 48S states. Indeed, the NTT of 5MP is folded within the mRNA entry tunnel widened by the head opening (Fig. 5). Head closure is incompatible with this position due to a steric hindrance (Fig. 4d). Together, our findings elucidate how 5MPs provide a robust layer of stringency on top of canonical initiation factors[21,70,71].

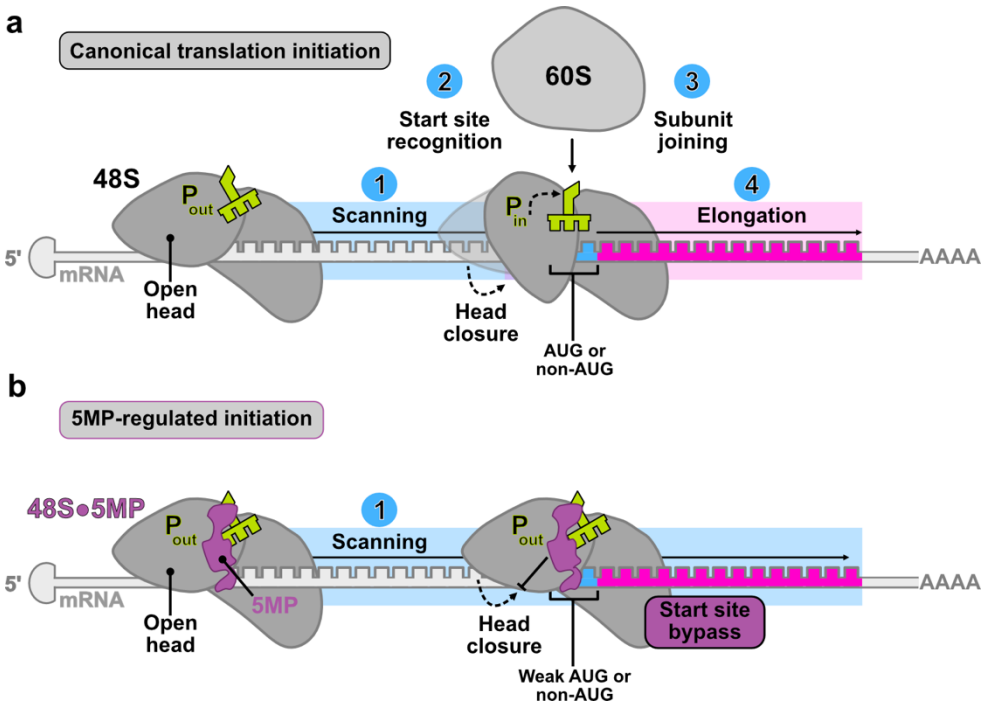


Figure 6: Schematic of 5MP-mediated regulation of start codon recognition.

(a) During canonical initiation, the 48S complex scans the mRNA until it recognizes a start codon. Both canonical AUGs and suboptimal codons, including weak AUGs and non-AUGs, can trigger scanning arrest. This arrest is accompanied by a conformational shift of the 48S head from an open to a closed state, while the initiator $tRNA^{Met}$ transitions from the P_{out} to the P_{in} position to allow codon–anticodon pairing. Successful start-site recognition then drives 60S joining and elongation. **(b)** By contrast, 5MP stabilizes the 48S complex in a scanning-competent state, maintaining the 48S head in the open conformation and the initiator tRNA in the P_{out} position. As a result, weak AUG and non-AUG codons fail to induce arrest, are bypassed during scanning, and do not initiate elongation.

Methods

BSC-1 cell culture and cell lysate preparation

BSC-1 cells (ATCC), a well-established epithelial cell line historically central to virology and membrane trafficking research, were cultured in DMEM (Invitrogen) supplemented with 10% heat-inactivated FBS, 2 mM GlutaMAX (Gibco) and 1 % (v/v) of 100× Penicillin-Streptomycin (Gibco). Cells were seeded in a 75 cm^2 flask at a density of ~20,000 cells/flask and cultured to confluence. After being washed twice with pre-warmed PBS, the cells were detached with Trypsin-EDTA (Gibco) and collected by centrifugation at 300×g for 4 minutes. The cell pellet was resuspended in 100 µL of semi-permeabilization buffer (25 mM HEPES, 110 mM KOAc, 15 mM $Mg(OAc)_2$, 1 mM DTT, 0.015% digitonin, 2× protease inhibitor cocktail (Roche), 20U SUPERase-In RNase inhibitor (Thermo Fisher Scientific: AM2696), 1 mM EGTA) at 4°C for 5 min. Following centrifugation at 1,000×g for 5 minutes, the supernatant was collected for grid preparation (see below). RNA concentration was quantified before grid preparation as a quality control measure using a NanoDrop One Spectrophotometer (A260/A280 and A260/A230; Thermo Fisher Scientific) without diluting the extract.

5MP protein expression and purification

Coding sequences for human 5MP1 (UniProt:Q9Y6E2) and 5MP2 (UniProt: Q7L1Q6) were synthesized (GenScript) and cloned into pET29a(+) using NdeI–XhoI restriction sites, generating constructs with a C-terminal His_6 tag (GenScript). All plasmids were sequence-verified before use. Plasmids were transformed into *E. coli* BL21(DE3) and cultures were grown in LB medium at 37 °C to an OD600 of 0.6, induced with 0.5 mM IPTG, and expressed at room temperature for 5 hours. Cells were harvested by centrifugation and stored at −80 °C until purification. The cell pellets were resuspended in a lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.5 mM DTT, 2× protease inhibitor cocktail (Roche),) and lysed by sonication on ice. Lysates were clarified by centrifugation at ~40,000×g for 40 min at 4 °C, and supernatants were filtered through a 0.22 µm filter before chromatography. The clarified lysate was loaded onto a pre-

equilibrated Ni-NTA affinity column, washed with 10 column volumes of binding buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.5 mM DTT), and eluted on an ÄKTA FPLC (Cytiva) system using a linear gradient up to 1 M imidazole. Fractions were analyzed by SDS-PAGE, and peak fractions pooled. The pooled Ni-NTA eluate was dialyzed overnight against a storage buffer (20 mM Tris-HCl, pH 7.8, 125 mM NaCl, 0.5 mM DTT), concentrated using centrifugal filters (10 kDa; Amicon) and concentration estimated using UV absorbance at 280 nm on a NanoDrop One Spectrophotometer (Thermo Fisher Scientific). A total of 100 mg of protein was loaded into a HiTrap Q HP anion-exchange column pre-equilibrated (20 mM Tris-HCl, pH 7.8, pH 8.0, 125 mM NaCl, 0.5 mM DTT). Proteins were eluted with a linear NaCl gradient. Fractions containing single bands corresponding to 5MP1 or 5MP2 by SDS–PAGE were pooled, dialyzed overnight against a storage buffer (10 mM HEPES pH 7.5, 125 mM NaCl, 0.5 mM DTT), concentrated using 10kDa Amicon filters and concentration estimated using UV absorbance at 280 nm on a NanoDrop One Spectrophotometer (Thermo Fisher Scientific). Proteins were subjected to a final purification using Superdex 200 Increase 10/300 GL column equilibrated in 10 mM HEPES pH 7.5, 125 mM NaCl, 1 mM TCEP. For 5MP1, fractions corresponding to elution peaks 6–9 were collected; for 5MP2, fractions 3–6 were recovered. The final preparations showed a major peak migrating at ~42 kDa. Protein concentration was determined by UV absorbance at 280 nm using theoretical extinction coefficients and a NanoDrop One Spectrophotometer (Thermo Fisher Scientific), with purity confirmed by SDS–PAGE (NuPage Bis-Tris Mini Protein Gels 4-12%, MOPS SDS buffer kit, SeeBlue Pre-stained protein standard (Invitrogen)). Aliquots were flash-frozen in liquid nitrogen and stored at −80 °C until use. All the buffers used were filtered with 0.5 µm filters.

Plasmids for mRNA preparations

Plasmids used for this mRNA preparation are listed in Extended Data Table S2. A plasmid carrying the coding sequence for NanoLuciferase (NLuc) flanked by the 5′ and 3′ UTRs of rabbit β-globin was synthesized by Azenta (vector: pUC-GW-Kan), as designed and described[72]. Variants of the plasmid containing the mutations in the AUG or CUG start

codon, and their Kozak contexts, were also synthesized by Azenta (vector: pUC-GW-Kan).

The plasmids containing the 5′UTR human *c-MYC* (NM_002467.6) with in-frame start codons (CUG and AUG) and the coding sequence for the first 48 nucleotides of c-MYC, fused to a 3×FLAG and NLuc CDS was synthesized by Azenta (vector: pUC-GW-Kan). Variants of the plasmid containing the mutations in the start codons and their Kozak contexts were also synthesized by Azenta (vector: pUC-GW-Kan). The constructs are listed in Extended Data Table S2. All plasmids were sequenced to verify the sequences.

In vitro transcription of NanoLuciferase reporter mRNAs

DNA templates for *in vitro* transcription of mRNAs were PCR-amplified from the plasmids using primers flanking the template region for transcription. The T7 promoter sequence and a 30-nt poly(A) tail were added through overhangs of the forward and reverse primers. The following forward primer sequences were used: *TTTTTTAATACGACTCACTATAGGGAGA*ACACTTGCTTTTGACACAACTGTG (NLuc constructs, 5′-3′);

*TTTTTTAATACGACTCACTATAGGGAGA*AACTCGCTGTAGTAATTCCAGCGAG (c-Myc constructs, 5′-3′); sequence with the T7 promoter site is in italic.

The reverse primer including a poly(A)-encoding sequence in all constructs was (5′-3′): TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAATGAAAATAAATTTCCTTTATTAGCC.

PCR was performed using Phusion High-Fidelity DNA Polymerase (NEB; M0530L) according to NEB protocol. After PCR, DNA templates were purified by phenol/chloroform extraction and dissolved in nuclease-free Milli-Q water. DNA concentrations were measured using a NanoDrop One Spectrophotometer (Thermo Fisher Scientific). The presence of a single band as a DNA template was confirmed by agarose gel electrophoresis (1% (w/v) agarose in TAE buffer). *In vitro* transcription reactions were carried out using 4 µg of purified DNA templates and purified recombinant T7 polymerase in transcription buffer (166 mM HEPES–KOH, pH 7.5; 20 mM $MgCl_2$; 40 mM DTT, 2 mM spermidine, 25 mM each of ATP, GTP, CTP, and UTP; and 40 U/µl RNase Inhibitor (NEB;

M0314S)) in an 80 µl reaction. After incubation at 37°C for 3.5 hours, magnesium pyrophosphate precipitate was removed by centrifugation (14,000 x g, 5 min), and mRNA was precipitated from the supernatant by adding 7.5 M LiCl (2.5 M final concentration) and incubating at −80°C overnight. The next day, mRNA was pelleted by centrifugation (21,300 x g, 15 min at 4°C), washed with cold 80% ethanol and pelleted again. This washing step was repeated three times. After discarding the supernatant, the mRNA pellet was air-dried and dissolved in nuclease-free Milli-Q water. To attach 5′cap, capping reactions were performed using the Vaccinia Capping System (NEB; M2080S). The 5′capped mRNA was then purified by LiCl precipitation as described above and dissolved in nucleases-free Milli-Q water. mRNA concentration was determined from UV absorbance at 260 nm using a NanoDrop One Spectrophotometer (Thermo Scientific). The size and integrity of the *in vitro* transcribed mRNA were examined by denaturing agarose gel electrophoresis (1% (w/v) agarose in MOPS buffer with 1.11% (v/v) formaldehyde) alongside an ssRNA ladder (NEB). The stock solution of *in vitro* transcribed mRNAs was stored at −80 °C.

***In vitro* mRNA translation in RRL**

In vitro translation was performed using a commercial micrococcal-nuclease-treated RRL (Promega; L4960) with modifications as described below. Translation reactions were carried out in the presence of 50% RRL, 30 mM HEPES–KOH (pH 7.5), 50 mM KOAc, 1.0 mM $Mg(OAc)_2$, 0.2 mM ATP, 0.2 mM GTP, 0.04 mM of 20 amino acids (Promega), 5 mM DTT, and 1% furimazine NanoLuciferase substrate (Promega; N113A). To initiate translation, 11 µL of reactions were preincubated at 30°C for 3 min, followed by addition of *in vitro* transcribed mRNAs encoding Nanoluciferase (10 nM final) to a total volume of 12 µL. Translation kinetics were monitored by continuously recording NanoLuciferase luminescence at 30°C for 20 min using an Infinite m1000 pro microplate reader (Tecan). The luminescence values were plotted over time using Prism 10 (GraphPad Software). Progress curves were recorded for experiments in triplicates. The maximum translation rates were calculated from the maximum rate of luminescence change

(Max ΔRLU/Δsec), which were obtained as the peak values of the first derivative of the RLU curves calculated in Prism 10 (GraphPad Software)

Determination of IC_{50} of 5MP1/2

To derive the IC_{50} values for 5MP1/2-mediated translation inhibition from luminescence curves, maximum translation rates (Max ΔRLU/Δsec) were plotted as a function of 5MP1/2 concentrations. The IC_{50} values were determined by fitting the data using a four-parameter inhibition model (Y=Bottom + (Top-Bottom)/[1+(IC_{50}/X)^HillSlope]) with nonlinear regression (curve fitting) in Prism 10. The average translation rate in the absence of 5MP1/2 was defined as 0% inhibition (maximum translation activity), whereas the average rateat the highest 5MP1/2 concentration was set as 100% inhibition (minimum translation activity).

Detection of RNase contamination within recombinant 5MP1/2 purified protein solutions

To test for RNase contamination in the purified recombinant 5MP1/2 proteins, NLuc mRNAs (final 140 nM) were incubated with and without the purified recombinant 5MP1/2 (final 1 μM) at 37 °C for 1 hour, *i.e.,* exceeding the extents of translation experiments in this work. As a control, incubation reactions were also set up with 40 units (U) of RNase Inhibitor (NEB; M0314S) for a final volume of 20 μl. The reactions were stopped by the addition of 2× RNA loading dye (NEB; B0363S), and the samples were heated at 70 °C for 10 min. mRNA integrity was assessed by 1% denaturing formaldehyde agarose gel pre-stained with SYBR Safe DNA gel stain (Invitrogen; S33102).

Western blotting

Equal volumes of RRL translation reactions were mixed with 2× urea sample dye (80 mM Tris-HCl, pH 6.8; 138.75 mM SDS; 8 M urea; 3 mM bromophenol blue; 16% (v/v) glycerol) and boiled at 98 °C for 10 min. The total proteins in RRL samples were separated on 4–20% Mini-Protean TGX Precast Protein gels (Bio-Rad) alongside a protein marker (Bio-

Rad) at 130 V for 2.5 h in 1× Tris-Tricine-SDS running buffer, and proteins were transferred to a PVDF membrane using a Trans-Blot Turbo transfer system (Bio-Rad) .

To reduce non-specific signal, the membranes were blocked in blocking buffer (1× PBST containing 0.1% Tween-20 and 5% (v/v) non-fat dry milk (Research Products International; M17200)) at room temperature for 30 min. The membranes were then incubated either with 1:3000 mouse anti-FLAG M2 IgG1 monoclonal antibody (Sigma-Aldrich; F1804) or 1:5000 rabbit anti-BZW2 IgG polyclonal antibody (GeneTex; GTX106985) in blocking buffer at 4 °C overnight. After washing the membranes by shaking in 1× PBST for 10 min three times, the membranes were incubated either with 1:6000 goat anti-mouse IgG secondary antibody conjugated to horseradish peroxidase (Invitrogen) or with 1:10000 goat anti-rabbit IgG secondary antibody conjugated to horseradish peroxidase (Invitrogen) in blocking buffer at room temperature for 1 h. After washing the membranes three times, 10 min each, the blots were detected using SuperSignal West Atto Ultimate Sensitivity Substrate (Thermo Fisher) and imaged on the ChemiDoc MP system (Bio-Rad). The intensities of Western blot bands were quantified using ImageJ.

RRL sample preparation for cryo-EM

Purified 5MP1 (5 μM final) was added to 60 % commercial micrococcal-nuclease-treated RRL (Promega; L4960) supplemented with 30 mM HEPES–KOH (pH 7.5), 50 mM KOAc, 1.0 mM $Mg(OAc)_2$, 0.2 mM ATP, 0.6 mM Guanosine 5′-[β,γ-imido]triphosphate (GMPPNP), 0.04 mM of 20 amino acids (Promega) and 20U SUPERase-In RNase inhibitor (Thermo Fisher Scientific: AM2696), maintained on ice until mRNA addition. NLuc (AUG within a moderate Kozak sequence) was added to the 5MP1-RRL-supplemented mixture and incubated for 10 min at 30 °C under continuous agitation (400 rpm). The mixture was directly added to the grids and vitrified.

Grid preparation and plunge freezing

For BSC-1 cellular lysates, Quantifoil Au 300 mesh R2/2 holey SiO_2 grids (Electron Microscopy Sciences) were glow-discharged in an EMITECH K100X (25 mA, negative

polarity, 45 s). Grids were back-blotted for 8 s and plunge-frozen in liquid ethane at −184 °C using a Leica EM GP (15 °C, 85% relative humidity). Vitrified grids were stored in liquid nitrogen until imaging.

For 5MP1-RRL samples, Quantifoil R2/1 holey-carbon grids coated with a thin continuous carbon layer (Electron Microscopy Sciences) were glow-discharged in a PELCO easiGlow (20 mA, negative polarity, 30 s). A Vitrobot Mark IV (Thermo Fisher Scientific) was equilibrated to 4 °C and 100% relative humidity, with blot force set to zero. For each sample, 2.5 µL of lysate was applied, immediately blotted (2–3 blots, 3–7 s each, force 0), and plunge-frozen in liquid nitrogen-cooled ethane. Grids were stored in liquid nitrogen.

Cryo-EM data collection and analysis

Datasets for BSC-1 cell lysate and 5MP1-RRL samples were collected on a Titan Krios electron microscope (Thermo Fisher Scientific) operating at 300 kV, equipped with a Gatan Imaging Filter (20-eV slit width) and a K3 direct electron detector (Gatan). Images were recorded at a nominal magnification corresponding to a super-resolution pixel size of 0.415 Å (physical pixel size 0.83 Å), using a target defocus range of 1.0–1.5 µm. Automated data acquisition was performed in SerialEM[73,74] using beam-image shift to collect 5 movies per hole across 9 holes at each stage position. Zero-loss peak (ZLP) refinement was performed every 90 mins at a predetermined location to avoid darkened areas. The BSC-1 dataset consisted of 162,583 movies and the 5MP1-RRL dataset of 104,673 movies; each movie containing 30 frames with a dose of ~1 e^{-} $Å^{-2}$ per frame (total exposure: ~30 e^{-} $Å^{-2}$). Beam-induced motion correction, gain reference application, dose weighting, and 2×/4× binning (to pixel sizes of 0.83 Å and 1.66 Å, respectively) were performed using MotionCor2[75]. CTF parameters were estimated using CTFFIND[76] via the *cis*TEM GUI[77]. For 2D template matching (2DTM), atomic coordinates for the 40S body from PDB 7R4X were converted into a 3D map using the simulate program in *cis*TEM[78], with a pixel size of 1.66 Å, 336×336×336 box, and a linear B-factor of 10. 2DTM was executed on 4× binned images (pixel size 1.66 Å), processing data in blocks of 2,000 images. This yielded 523,233 targets for BSC-1

lysate and 1,573,843 targets for 5MP1-RRL, using in-plane angular sampling of 1.5°, out-of-plane sampling of 2.5°, and no defocus search. Template-matched coordinates, Euler angles, and CTFFIND5-derived defocus values were exported using the MT package in *cis*TEM GUI to generate .star files, which were merged using an in-house Python script; paths to the 4× binned images were replaced with those for 2× binned images (pixel size 0.83 Å). The updated .star file was re-imported in the MT package, and particle stacks (box size 560³ pixels) were generated using the *cis*TEM refine pipeline. Initial 3D reconstructions were computed using Generate 3D, and final alignment parameters and defocus values were exported to Frealign format (Extended Data Fig. 5). Collection and processing pixel sizes were as follows: super-resolution, 0.415 Å; physical, 0.83 Å (final reconstructions); 2× bin, 1.66 Å (CTF estimation and 2DTM); 4× bin, 3.32 Å (3D classification); and 8× bin, 6.64 Å (3D classification).

Cryo-EM data classification

All classifications steps were performed in Frealign v.9.11[79] on binned data (8×, 4×, 2×), with particle stacks prepared using the program resample in *cis*TEM package[77]. Binning of the particle stacks was performed using the program resample from cisTEM. Initial unmasked 3D maximum-likelihood classification separated elongating ribosomes (80S) and junk from pre-initiation complexes (8× binned data). For BSC-1 cell lysates, pre-initiation complex (PIC) classes were merged using an in-house Python script, preserving micrograph paths and generating a .star file compatible with cisTEM (available in YAFW: https://github.com/GrigorieffLab/yafw). A particle stack of PIC was generated in *cis*TEM, and further manual refinements were performed, including cycles of x,y alignment, Euler angles, and defocus/beam-tilt refinement. Focus-masked 3D maximum-likelihood subclassification was then applied to the head region (mask diameter 90 Å) to separate subpopulations. Classes containing the unknown density at the A-site were merged using YAFW, and 3D reconstructions were performed from unbinned data (pixel size 0.83 Å) in *cis*TEM. Local resolution estimation and map denoising were performed using RELION-3.1[80].

For 5MP1-RRL samples, an initial classification into 100 classes was performed. Representative PIC classes were divided into six complex types and merged using YAFW scripts. Particle alignment files (.star) were used to generate stacks in *cis*TEM for further refinements, as described above. Focus-masked subclassification focused on 5MP1, the head, eIF2, or mRNA/initiator tRNA codon–anticodon sites did not reveal additional heterogeneity, even when tested on 2×, 4×-, and 8× binned data. The 2DTM and classification workflow are illustrated in Extended Data Fig. 5a. 3D reconstructions with unbinned data (pixel size 0.83 Å) were performed in *cis*TEM. Local resolution estimation and map denoising were performed using RELION-3.1.

Protein identification in a cryo-EM map

The map containing the unknown density was low-pass filtered (4 Å) and segmented, retaining only the density corresponding to the unknown feature. ModelAngelo[50]implemented in RELION-5.0[81] was run without the sequence module (no_seq option) to generate an initial model based on HMM profiles. These profiles were then searched against the reference proteome using the hmm_search option. Hits for 5MP1 and 5MP2 showed the highest significance, with E-values of 1.43×10^{-9} and 1.24×10^{-37}, respectively. The predicted AlphaFold2 structures of these proteins (AF-Q9Y6E2-F1) were fitted into the density to validate the identification[51,52].

Model building

Maps were sharpened using DeepEMhancer[82]. The initial models were derived from the *Oryctolagus cuniculus* 80S ribosome (PDB 9BDL)[48], the *H. sapiens* 48S complex with eIF4F/eIF4A (PDB 8OZ0)[17], and the *H. sapiens* 48S complex in the open scanning state (PDB 8PJ1)[68]. The 18S rRNA from 9BDL was first rigid-body-fitted to the density. Then, ribosomal proteins and initiation factors from 8OZ0 or 8PJ1 were independently rigid-body-fitted using ChimeraX[83]. The 5MP1 model was generated using AlphaFold2 prediction (AF-Q9Y6E2-F1) was rigid-body fitted into the density. Following initial placement, all components were refined using ISOLDE[84] within ChimeraX, applying

segment-wise interactive molecular dynamics. Final refinements were performed with phenix.real_space_refine, yielding models with good stereochemical quality and high correlation with the cryo-EM maps (Extended Data Table S1). Model-map FSC was calculated using Phenix.

Head rotation measurements

To compare head rotation between initiation complexes, we generated a closed-head model by rigid-body fitting the human 48S complex in the AUG-recognition state (PDB 8PJ2) into our 2.8-Å map using ChimeraX[83], followed by a single round of real-space refinement in Phenix[85,86]. For both the 5MP1•48S (open-head) and the refined closed-head model, we aligned the 18S body (nucleotides 1–1213 and 1691–1869) using Matchmaker in ChimeraX and saved the aligned coordinates. The 18S head (nucleotides 1214–1690) was then independently aligned, and the resulting angular displacement was calculated in ChimeraX. Head rotation between our 48S•5MP1 complex and the previously reported 48S open-head structure (PDB 8PJ1) was measured using the same procedure[68].

Multiple sequence alignment

Protein sequences used for comparative analysis were obtained from a previously published phylogenetic dataset[8] and supplemented with mammalian organisms with their accession numbers as follows:

Homo Sapiens (UniProt:Q9Y6E2); Giardia intestinalis (*UniProt:*C6LQG9); *Chlamydomonas reinhardtii (RefSeq:*XP_001691456, *UniProt:*A8IP37); *Volvox carteri (UniProt:*D8TS79*)*; *Chlorella variabilis (UniProt:*E1Z485*)*; *Coccomyxa subellipsoidea (UniProt:*I0Z2Z0), *Selaginella moellendorffii (UniParc:*UPI0001E02D4B*)*; *Physcomitrella patens* (*UniParc*: UPI000162096B); *Ricinus communis (UniParc:*UPI0001940808*)*; *Arabidopsis thaliana* (*UniParc*:UPI00000A2E64); *Triticum aestivum (UniParc:*UPI00029B5BFD*)*; *Zea mays* (*UniParc*:UPI00017B6F18); *Chondrus crispus (UniParc:*UPI000331C39F*)*; *Schizophyllum commune* (*UniParc*: UPI0001DF4DBE);

606 *Monosiga brevicollis (UniParc:*UPI000163B8DD); *Hydra magnipapillata*
607 *(UniParc:*UPI0002B4BB35*)*; *Capitella teleta* (*UniParc:*UPI00032C0488); *Aplysia*
608 *californica* (*UniParc*:UPI0003593758); *Branchiostoma floridae*
609 (*UniParc*:UPI0001861DC4*)*; *Danio rerio* (*UniParc*:UPI000000BC41), *Xenopus tropicalis*
610 (*UniParc*:UPI00004D064B); *Chrysemys picta bellii* (*UniParc*:UPI000388ED85);
611 *Taeniopygia guttata* (*UniParc*: UPI000194BDA7); *Gallus gallus*
612 (*UniParc*:UPI000044358E); *Melopsittacus undulates* (*UniParc*:UPI000383468D);
613 *Amblyomma variegatum* (*UniParc*:UPI0002009EA1); *Caligus clemensi*
614 (*UniParc*:UPI000198AD9A); *Lepeophtheirus salmonis* (*UniParc*:UPI000198A411);
615 *Caligus rogercresseyi* (*UniParc*:UPI000198A05F) *Daphnia pulex*
616 (*UniParc*:UPI0001FEB3EB); *Acyrthosiphon pisum* (*UniParc*:UPI0002061392); *Riptortus*
617 *pedestris* (*UniParc:*UPI00032B4F83), *Pediculus humanus corporis*
618 (*UniParc*:UPI000186E225); *Tribolium castaneum* (*UniParc*:UPI0000D56980);
619 *Dendroctonus ponderosae* (*UniParc*:UPI00020FF929); *Bombyx mori*
620 (*UniParc*:UPI0000E5D066); *Helicoverpa armigera* (*UniParc*:UPI00019203AC); *Danaus*
621 *plexippus* (*UniParc*:UPI000239C638); *Nasonia vitripennis* (*UniProt*:UPI00015B411E);
622 *Camponotus floridanus* (*UniProt*:UPI0001E7B62E); *Megachile rotundata*
623 (*UniParc*:UPI000258DA7A); *Bombus terrestris* (*UniParc*:UPI00021A85D9); *Bombus*
624 *impatiens* (*UniParc*:UPI00022CAAA0); *Apis mellifera (UniParc:UPI0000DB7C8B)*; *Aedes*
625 *aegypti* (*UniParc*:UPI0000D76E88); *Culex quinquefasciatus* (*UniParc*:UPI00016D8A73);
626 *Ceratitis capitata* (*UniParc*:UPI00032984C4); *Drosophila willistoni*
627 (*UniParc*:UPI00017D8E71); *Drosophila grimshawi* (*UniParc*:UPI00017C6DDE);
628 *Drosophila mojavensis* (*UniParc*:UPI00017C7884); *Drosophila virilis*
629 (*UniParc*:UPI00017D4931); *Drosophila melanogaster* (*UniParc*:UPI000000741C); *Ovis*
630 *aries* (*RefSeq*: XDA75500.1); *Mus musculus* (*RefSeq*: XP_006515229.1); *Oryctolagus*
631 *cuniculus* (*RefSeq*: XP_051688519.2); *Pan troglodytes* (*RefSeq*: PNI66831.1); *Manis*
632 *pentadactyla* (*RefSeq*: KAI5178856.1).

633 All sequences were aligned using MUSCLE[87] with default parameters. The resulting
634 alignment was manually inspected, and conserved residues and motifs were visualized
635 using ESPript 3.2[88,89].

636

637 **Figures**

638 Figures were prepared using ChimeraX 1.9 and Affinity Designer 2.

639

640

Acknowledgments

We thank Chen Xu, Kangkang Song, and Christina Ouch for support with cryo-EM data collection at UMass Chan Medical School; Johannes Elferich, Dongjie Zhu, and Stephen Diggs for valuable scientific input; and members of the Grigorieff and Korostelev laboratories for constructive feedback on this study. This study was supported by Howard Hughes Medical Institute to N.G. and the US National Institutes of Health R35GM127094 to A.A.K.

Author contributions

X.Z. performed and analyzed biochemical experiments**,** prepared cell cultures and cryo-EM samples; collected and analyzed cryo-EM data; prepared illustrations and drafted the manuscript. **C-Y.H.** performed and analyzed biochemical experiments; prepared cryo-EM samples; prepared illustrations and drafted manuscript. **Z.S.** developed cell lysate protocols. **N.G.** designed research, supervised data analysis, drafted manuscript and secured funding. **A.A.K.** designed research, supervised data analysis, drafted the manuscript and secured funding. All authors revised the manuscript.

References

1 Vagner, S. *et al*. Translation of CUG- but not AUG-initiated forms of human fibroblast growth factor 2 is activated in transformed and stressed cells. *J Cell Biol* **135**, 1391-1402 (1996). https://doi.org:10.1083/jcb.135.5.1391

2 Diaz de Arce, A. J., Noderer, W. L. & Wang, C. L. Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic Acids Res* **46**, 985-994 (2018). https://doi.org:10.1093/nar/gkx1114

3 Eisenberg, A. R. *et al*. Translation Initiation Site Profiling Reveals Widespread Synthesis of Non-AUG-Initiated Protein Isoforms in Yeast. *Cell Syst* **11**, 145-160.e145 (2020). https://doi.org:10.1016/j.cels.2020.06.011

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

5 Mahé, M., Rios-Fuller, T., Katsara, O. & Schneider, R. J. Non-canonical mRNA translation initiation in cell stress and cancer. *NAR Cancer* **6** (2024). https://doi.org:10.1093/narcan/zcae026

6 Williams, T. D. & Rousseau, A. Translation regulation in response to stress. *The FEBS Journal* **291**, 5102-5122 (2024). https://doi.org:https://doi.org/10.1111/febs.17076

7 Tang, L. *et al*. Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. *Nucleic Acids Research* **45**, 11941-11953 (2017). https://doi.org:10.1093/nar/gkx808

8 Singh, C. R. *et al*. Mechanisms of translational regulation by a human eIF5-mimic protein. *Nucleic Acids Research* **39**, 8314-8328 (2011). https://doi.org:10.1093/nar/gkr339

9 Sato, K. *et al*. Novel oncogene 5MP1 reprograms c-Myc translation initiation to drive malignant phenotypes in colorectal cancer. *EBioMedicine* **44**, 387-402 (2019). https://doi.org:10.1016/j.ebiom.2019.05.058

10 Cheng, D.-D., Li, S.-J., Zhu, B., Yuan, T., Yang, Q.-C. & Fan, C.-Y. Downregulation of BZW2 inhibits osteosarcoma cell growth by inactivating the Akt/mTOR signaling pathway. *Oncol Rep* **38**, 2116-2122 (2017). https://doi.org:10.3892/or.2017.5890

11 Wek, R. C., Jiang, H.-Y. & Anthony, T. G. Coping with stress: eIF2 kinases and translational control. *Biochemical Society Transactions* **34**, 7-11 (2006). https://doi.org:10.1042/bst0340007

12 Tejada, S. *et al*. Eukaryotic Initiation Factors (eIF) 2α and 4E Expression, Localization, and Phosphorylation in Brain Tumors. *Journal of Histochemistry & Cytochemistry* **57**, 503-512 (2009). https://doi.org:10.1369/jhc.2009.952929

13 Krishnamoorthy, G. P. *et al*. EIF1AX and RAS Mutations Cooperate to Drive Thyroid Tumorigenesis through ATF4 and c-MYC. *Cancer Discovery* **9**, 264-281 (2019). https://doi.org:10.1158/2159-8290.Cd-18-0606

14 Sendoel, A. *et al*. Translation from unconventional 5' start sites drives tumour initiation. *Nature* **541**, 494-499 (2017). https://doi.org:10.1038/nature21036

15 He, L.-R. *et al*. Overexpression of eIF5A-2 is an adverse prognostic marker of survival in stage I non–small cell lung cancer patients. *International Journal of Cancer* **129**, 143-150 (2011). https://doi.org:https://doi.org/10.1002/ijc.25669

16 Nakamura, J. *et al*. Overexpression of eukaryotic elongation factor eEF2 in gastrointestinal cancers and its involvement in G2/M progression in the cell cycle. *International journal of oncology* **34**, 1181-1189 (2009).

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

18 Pestova, T. V. & Kolupaeva, V. G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev* **16**, 2906-2922 (2002). https://doi.org:10.1101/gad.1020902

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

20 Andreev, D. E., Loughran, G., Fedorova, A. D., Mikhaylova, M. S., Shatsky, I. N. & Baranov, P. V. Non-AUG translation initiation in mammals. *Genome Biology* **23**, 111 (2022). https://doi.org:10.1186/s13059-022-02674-2

21 Hinnebusch, A. G. & Lorsch, J. R. The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harb Perspect Biol* **4** (2012). https://doi.org:10.1101/cshperspect.a011544

22 Beerman, R. W. & Jongens, T. A. A non-canonical start codon in the Drosophila fragile X gene yields two functional isoforms. *Neuroscience* **181**, 48-66 (2011).

23 Ingolia, Nicholas T., Lareau, Liana F. & Weissman, Jonathan S. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**, 789-802 (2011). https://doi.org:10.1016/j.cell.2011.10.002

24 Fedorova, A. D., Kiniry, S. J., Andreev, D. E., Mudge, J. M. & Baranov, P. V. Thousands of human non-AUG extended proteoforms lack evidence of evolutionary selection among mammals. *Nature Communications* **13**, 7910 (2022). https://doi.org:10.1038/s41467-022-35595-6

25 Meiron, M., Anunu, R., Scheinman, E. J., Hashmueli, S. & Levi, B.-Z. New isoforms of VEGF are translated from alternative initiation CUG codons located in its 5' UTR. *Biochemical and Biophysical Research Communications* **282**, 1053-1060 (2001).

26 Trulley, P. *et al*. Alternative Translation Initiation Generates a Functionally Distinct Isoform of the Stress-Activated Protein Kinase MK2. *Cell Rep* **27**, 2859-2870.e2856 (2019). https://doi.org:10.1016/j.celrep.2019.05.024

27 Lee, P. J., Sun, Y., Soares, A. R., Fai, C., Picciotto, M. R. & Guo, J. U. Alternative translation initiation produces synaptic organizer proteoforms with distinct localization and functions. *Mol Cell* **84**, 3967-3978.e3968 (2024). https://doi.org:10.1016/j.molcel.2024.08.032

28 Hann, S. R., King, M. W., Bentley, D. L., Anderson, C. W. & Eisenman, R. N. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52**, 185-195 (1988).

29 Ivanov, I. P., Firth, A. E., Michel, A. M., Atkins, J. F. & Baranov, P. V. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic acids research* **39**, 4220-4234 (2011).

30 Shen, S.-M. *et al*. PTENα and PTENβ promote carcinogenesis through WDR5 and H3K4 trimethylation. *Nature cell biology* **21**, 1436-1448 (2019).

31 Pancsa, R., Andreev, D. E. & Dean, K. The implication of non-AUG-initiated N-terminally extended proteoforms in cancer. *RNA Biol* **22**, 1-18 (2025). https://doi.org:10.1080/15476286.2025.2498203

32 Weber, R. *et al*. Monitoring the 5'UTR landscape reveals isoform switches to drive translational efficiencies in cancer. *Oncogene* **42**, 638-650 (2023). https://doi.org:10.1038/s41388-022-02578-2

33 Ivanov, I. P. *et al*. Evolutionarily conserved inhibitory uORFs sensitize Hox mRNA translation to start codon selection stringency. *Proceedings of the National Academy of Sciences* **119**, e2117226119 (2022). https://doi.org:doi:10.1073/pnas.2117226119

34 Hiraishi, H. *et al*. Essential role of eIF5-mimic protein in animal development is linked to control of ATF4 expression. *Nucleic Acids Research* **42**, 10321-10330 (2014). https://doi.org:10.1093/nar/gku670

35 Loughran, G., Firth, A. E., Atkins, J. F. & Ivanov, I. P. Translational autoregulation of BZW1 and BZW2 expression by modulating the stringency of start codon selection. *PLOS ONE* **13**, e0192648 (2018). https://doi.org:10.1371/journal.pone.0192648

36 Singh, C. R. *et al*. Human oncoprotein 5MP suppresses general and repeat-associated non-AUG translation via eIF3 by a common mechanism. *Cell Rep* **36**, 109376 (2021). https://doi.org:10.1016/j.celrep.2021.109376

37 Wagner, P. A. *et al*. Molecular basis for the interactions of eIF2β with eIF5, eIF2B, and 5MP1 and their regulation by CK2. *RNA* **31**, 1419-1435 (2025). https://doi.org:10.1261/rna.080652.125.

38 Uhlén, M. *et al*. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). https://doi.org:doi:10.1126/science.1260419

39 Schwanhäusser, B. *et al*. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011). https://doi.org:10.1038/nature10098

40 Li, Z. M. *et al*. Optimization of extended Kozak elements enhances recombinant proteins expression in CHO cells. *J Biotechnol* **392**, 96-102 (2024). https://doi.org:10.1016/j.jbiotec.2024.06.020

41 Li, S. *et al*. BZW1, a novel proliferation regulator that promotes growth of salivary muocepodermoid carcinoma. *Cancer Letters* **284**, 86-94 (2009). https://doi.org:https://doi.org/10.1016/j.canlet.2009.04.019

42 Agarwal, S. *et al*. BZW2 Inhibition Reduces Colorectal Cancer Growth and Metastasis. *Mol Cancer Res* **21**, 698-712 (2023). https://doi.org:10.1158/1541-7786.Mcr-23-0003

43 Kozel, C. *et al*. Overexpression of eIF5 or its protein mimic 5MP perturbs eIF2 function and induces ATF4 translation through delayed re-initiation. *Nucleic Acids Res* **44**, 8704-8713 (2016). https://doi.org:10.1093/nar/gkw559

44 Jin, K., Li, Y., Wei, R., Liu, Y., Wang, S. & Tian, H. BZW2 promotes malignant progression in lung adenocarcinoma through enhancing the ubiquitination and degradation of GSK3β. *Cell Death Discovery* **10**, 105 (2024). https://doi.org:10.1038/s41420-024-01879-7

45 Lee, S. *et al*. The F-actin-microtubule crosslinker Shot is a platform for Krasavietz-mediated translational regulation of midline axon repulsion. *Development* **134**, 1767-1777 (2007). https://doi.org:10.1242/dev.02842

46 Lucas, B. A., Himes, B. A., Xue, L., Grant, T., Mahamid, J. & Grigorieff, N. Locating macromolecular assemblies in cells by 2D template matching with cisTEM. *Elife* **10** (2021). https://doi.org:10.7554/eLife.68946

47 Lucas, B. A., Zhang, K., Loerch, S. & Grigorieff, N. In situ single particle classification reveals distinct 60S maturation intermediates in cells. *eLife* **11**, e79272 (2022). https://doi.org:10.7554/eLife.79272

48 Loveland, A. B., Koh, C. S., Ganesan, R., Jacobson, A. & Korostelev, A. A. Structural mechanism of angiogenin activation by the ribosome. *Nature* **630**, 769-776 (2024). https://doi.org:10.1038/s41586-024-07508-8

49 Seraj, Z. *et al*. In extracto cryo-EM reveals eEF2 as a major hibernation factor on 60S and 80S particles. *bioRxiv*, 2025.2011.2025.690450 (2025). https://doi.org:10.1101/2025.11.25.690450

50 Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D. & Scheres, S. H. W. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450-457 (2024). https://doi.org:10.1038/s41586-024-07215-4

51 Liu, H., Laiho, A., Törönen, P. & Holm, L. 3-D substructure search by transitive closure in AlphaFold database. *Protein Science* **34**, e70169 (2025). https://doi.org:https://doi.org/10.1002/pro.70169

52 Jumper, J. *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org:10.1038/s41586-021-03819-2

53 Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-292 (1986). https://doi.org:10.1016/0092-8674(86)90762-2

54 Kozak, M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* **9**, 5073-5080 (1989). https://doi.org:10.1128/mcb.9.11.5073-5080.1989

55 Kozak, M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol* **196**, 947-950 (1987). https://doi.org:10.1016/0022-2836(87)90418-9

56 Cho, N. H. *et al*. OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**, eabi6983 (2022). https://doi.org:doi:10.1126/science.abi6983

57 Cheung, Y. N. *et al*. Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. *Genes Dev* **21**, 1217-1230 (2007). https://doi.org:10.1101/gad.1528307

58 Saini, A. K. *et al*. Eukaryotic translation initiation factor eIF5 promotes the accuracy of start codon recognition by regulating Pi release and conformational transitions of

the preinitiation complex. *Nucleic Acids Research* **42**, 9623-9640 (2014). https://doi.org:10.1093/nar/gku653

59 Llácer, J. L. *et al*. Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *eLife* **7**, e39273 (2018). https://doi.org:10.7554/eLife.39273

60 Kolitz, S. E., Takacs, J. E. & Lorsch, J. R. Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *Rna* **15**, 138-152 (2009). https://doi.org:10.1261/rna.1318509

61 Martin-Marcos, P., Nanda, J., Luna, R. E., Wagner, G., Lorsch, J. R. & Hinnebusch, A. G. β-Hairpin loop of eukaryotic initiation factor 1 (eIF1) mediates 40 S ribosome binding to regulate initiator tRNA(Met) recruitment and accuracy of AUG selection in vivo. *J Biol Chem* **288**, 27546-27562 (2013). https://doi.org:10.1074/jbc.M113.498642

62 Thakur, A. & Hinnebusch, A. G. eIF1 Loop 2 interactions with Met-tRNA(i) control the accuracy of start codon selection by the scanning preinitiation complex. *Proc Natl Acad Sci U S A* **115**, E4159-e4168 (2018). https://doi.org:10.1073/pnas.1800938115

63 Passmore, L. A. *et al*. The Eukaryotic Translation Initiation Factors eIF1 and eIF1A Induce an Open Conformation of the 40S Ribosome. *Molecular Cell* **26**, 41-50 (2007). https://doi.org:https://doi.org/10.1016/j.molcel.2007.03.018

64 Yi, S. H. *et al*. Conformational rearrangements upon start codon recognition in human 48S translation initiation complex. *Nucleic Acids Res* **50**, 5282-5298 (2022). https://doi.org:10.1093/nar/gkac283

65 Jobe, A., Liu, Z., Gutierrez-Vargas, C. & Frank, J. New Insights into Ribosome Structure and Function. *Cold Spring Harb Perspect Biol* **11** (2019). https://doi.org:10.1101/cshperspect.a032615

66 Llácer, Jose L. *et al*. Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Molecular Cell* **59**, 399-412 (2015). https://doi.org:https://doi.org/10.1016/j.molcel.2015.06.033

67 Hussain, T. *et al*. Structural Changes Enable Start Codon Recognition by the Eukaryotic Translation Initiation Complex. *Cell* **159**, 597-607 (2014). https://doi.org:https://doi.org/10.1016/j.cell.2014.10.001

68 Petrychenko, V., Yi, S.-H., Liedtke, D., Peng, B.-Z., Rodnina, M. V. & Fischer, N. Structural basis for translational control by the human 48S initiation complex. *Nature Structural & Molecular Biology* **32**, 62-72 (2025). https://doi.org:10.1038/s41594-024-01378-4

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

70 Fletcher, C. M., Pestova, T. V., Hellen, C. U. T. & Wagner, G. Structure and interactions of the translation initiation factor eIF1. *The EMBO Journal* **18**, 2631-2637-2637 (1999). https://doi.org:https://doi.org/10.1093/emboj/18.9.2631

71 Korostelev, A. A. The Structural Dynamics of Translation. *Annu Rev Biochem* **91**, 245-267 (2022). https://doi.org:10.1146/annurev-biochem-071921-122857

72 Susorov, D., Egri, S. & Korostelev, A. A. Termi-Luc: a versatile assay to monitor full-protein release from ribosomes. *Rna* **26**, 2044-2050 (2020). https://doi.org:10.1261/rna.076588.120

73 Mastronarde, D. N. SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. *Microscopy and Microanalysis* **9**, 1182-1183 (2003). https://doi.org:10.1017/s1431927603445911

74 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005). https://doi.org:10.1016/j.jsb.2005.07.007

75 Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y. & Agard, D. A. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017). https://doi.org:10.1038/nmeth.4193

76 Elferich, J., Kong, L., Zottig, X. & Grigorieff, N. CTFFIND5 provides improved insight into quality, tilt, and thickness of TEM samples. *Elife* **13**, RP97227 (2024).

77 Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018). https://doi.org:10.7554/eLife.35383

78 Himes, B. & Grigorieff, N. Cryo-TEM simulations of amorphous radiation-sensitive samples using multislice wave propagation. *IUCrJ* **8**, 943-953 (2021). https://doi.org:10.1107/s2052252521008538

79 Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J Struct Biol* **183**, 377-388 (2013). https://doi.org:10.1016/j.jsb.2013.07.005

80 Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr D Struct Biol* **76**, 94-101 (2020). https://doi.org:10.1107/s2059798319016577

81 Burt, A. *et al*. An image processing pipeline for electron cryo-tomography in RELION-5. *FEBS Open Bio* **14**, 1788-1804 (2024). https://doi.org:10.1002/2211-5463.13873

82 Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J. M., Sorzano, C. O. S. & Vargas, J. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Communications Biology* **4**, 874 (2021). https://doi.org:10.1038/s42003-021-02399-1

83 Meng, E. C. *et al*. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* **32**, e4792 (2023). https://doi.org:10.1002/pro.4792

84 Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530 (2018). https://doi.org:10.1107/s2059798318002425

85 Adams, P. D. *et al*. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94-106 (2011). https://doi.org:10.1016/j.ymeth.2011.07.005

86 Afonine, P. V. *et al*. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018). https://doi.org:10.1107/s2059798318006551

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

88 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320-W324 (2014). https://doi.org:10.1093/nar/gku316

89 Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research* **31**, 3320-3323 (2003). https://doi.org:10.1093/nar/gkg556

SUPPLEMENTARY INFORMATION FOR:

Structural basis for non-AUG translation regulation by 5MPs

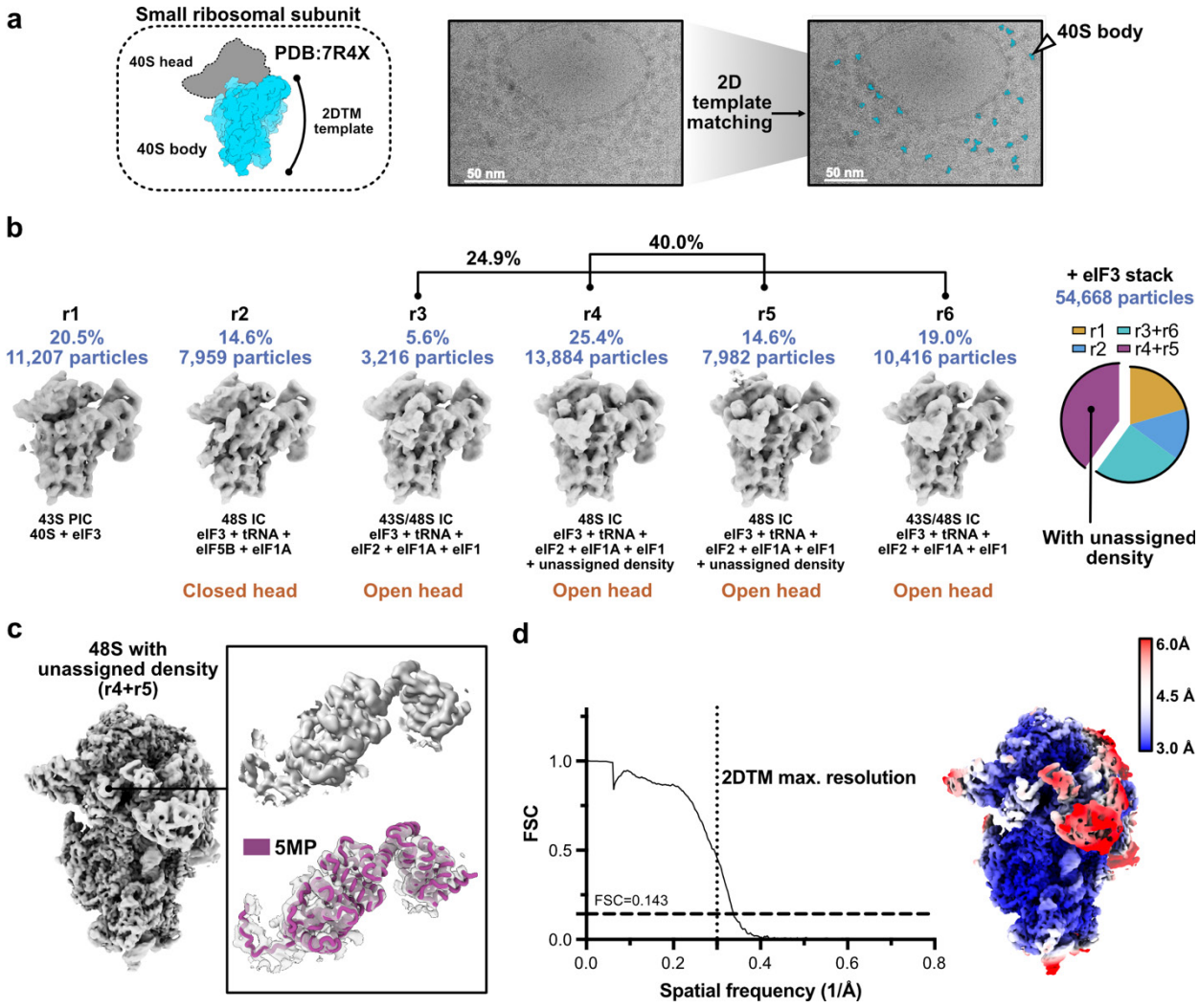
Ximena Zottig[1,2*], Chun-Ying Huang[1*], Zahra Seraj[1], Nikolaus Grigorieff[1,2#], Andrei A. Korostelev[1#]

[1] RNA Therapeutics Institute, UMass Chan Medical School, 368 Plantation Street, Worcester, MA 01605, USA.

[2] Howard Hughes Medical Institute, UMass Chan Medical School, 368 Plantation Street, Worcester, MA 01605, USA.

* These authors contributed equally

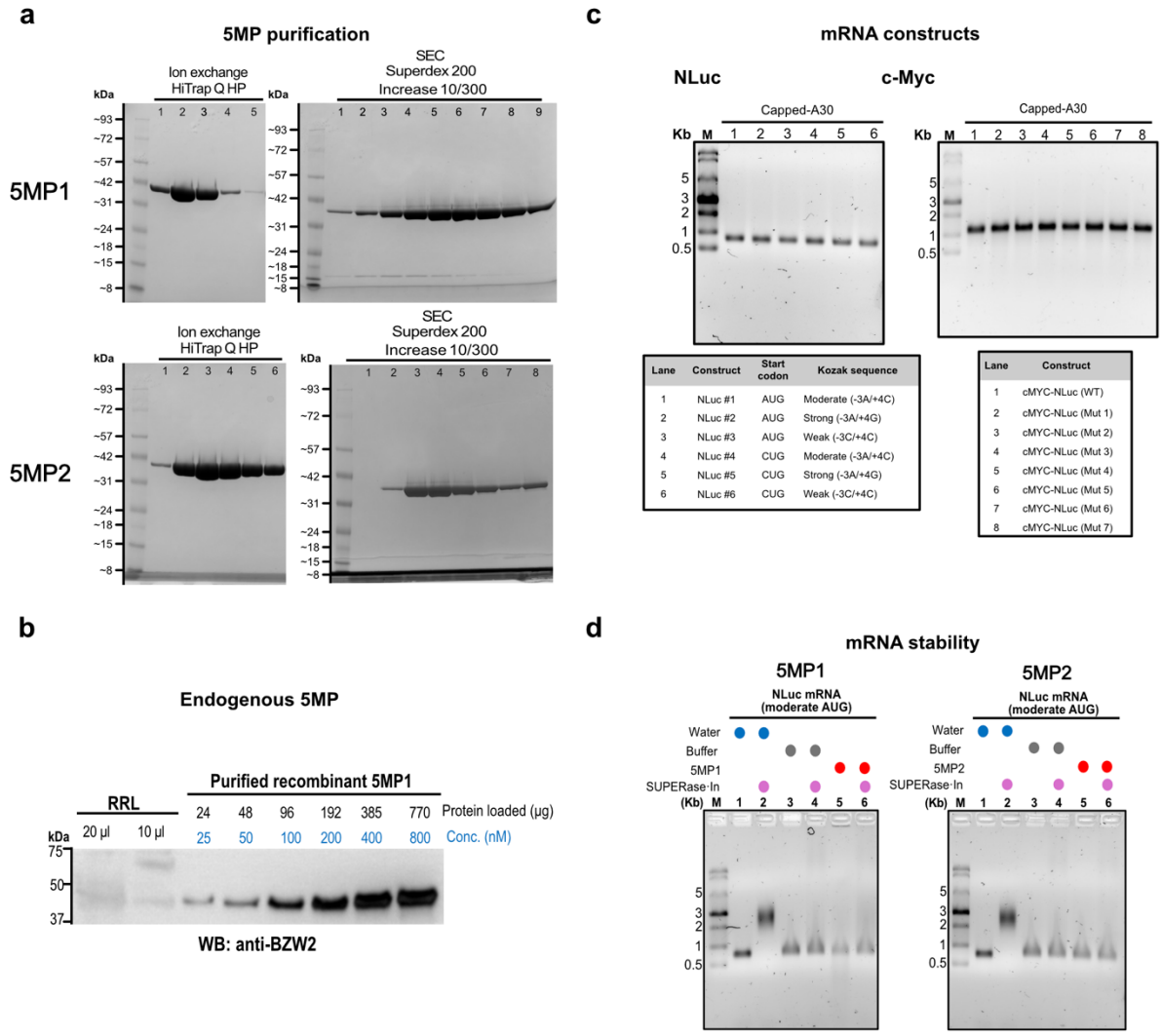
Co-corresponding authors: niko@grigorieff.org; Andrei.Korostelev@umassmed.edu



Extended Data Figure 1: Identification of translation pre-initiation complexes in BSC-1 lysates.

(a) Schematic representation of the template used for 2D template matching (2DTM), representative micrograph of BSC-1 cellular lysate and an illustration of 2DTM detections (blue). Template matching was carried out using the 40S body from PDB 7R4X. **(b)** Pre-initiation complex (PIC) classes obtained after 3D classification of 2DTM-identified particles (8× binned data). Prior particle sorting removed 80S and non-ribosomal particles. PICs account for ~11% of all particles detected in BSC-1 lysate. Among these, ~40% of the particles adopt an open-head conformation containing an additional unassigned density, which we identified as a member of the 5MP protein family. Approximately 25% of particles exhibit an open-head conformation without the additional density, representing a mixture of 48S and 43S complexes. By contrast, ~15% of 48S initiation complexes display a closed-head conformation, and no additional density is observed in these particles. **(c)** Reconstruction of the 48S pre-initiation complex containing the unassigned density. Shown are a close-up view (top) and its overlay with the AlphaFold2-predicted eIF5-mimic protein 1 (5MP1) model refined into our density map (bottom). **(d)** Fourier shell correlation (FSC) curve showing the global resolution. The

resolution limit imposed during 2DTM is indicated by the dotted line (~3.2 Å). A local resolution map of the 48S•5MP reconstruction is shown, with higher-resolution regions indicated in blue (~3.0 Å) and lower-resolution regions in red (~6.0 Å). The map was denoised using RELION. The FSC, as well as local resolution estimates in regions overlapping with the 40S template, are biased toward higher values due to the use of a 2DTM template during particle detection.



Extended Data Figure 2: Preparation and validation of recombinant 5MP proteins and reporter mRNAs.

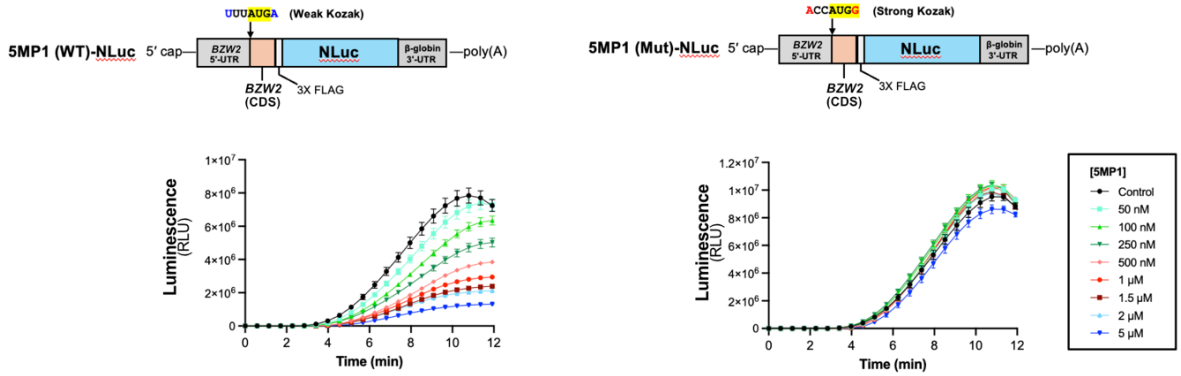
(a) Recombinant eIF5-mimic proteins 5MP1 (top) and 5MP2 (bottom) were purified by His-tag affinity chromatography, followed by ion-exchange separation on a HiTrap Q HP column, and a final purification step by size-exclusion chromatography using a Superdex 200 Increase 10/300 column. SDS-PAGE results are shown. For 5MP1, fractions 6–9 from the size-exclusion step were pooled, quantified, and used for translation assays and cryo-EM. For 5MP2, fractions 3–6 were collected for subsequent experiments. **(b)** The concentration of endogenous 5MP in rabbit reticulocyte lysate (RRL) is estimated to be < 25 nM. Titrations were performed using recombinant 5MP1, and detection was performed with an anti-BZW2 antibody. **(c)** Capped mRNA constructs with A_{30} tails were produced by *in vitro* transcription for translation assays and cryo-EM. **(d)** Purified 5MP1 and 5MP2 showed no detectable RNase contamination: mRNA stability in the presence of purified

5MPs was assessed by incubating the moderate-AUG reporter mRNA (the construct used for cryo-EM) for 1 hour at 37 °C, *i.e.*, substantially longer than the typical reaction time in this work (~20 minutes).

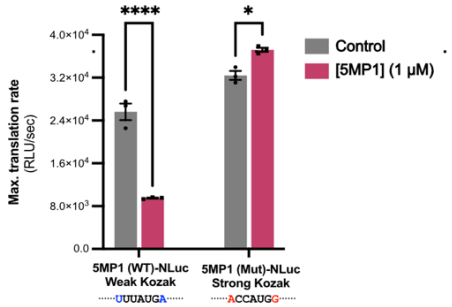
a

5MP1						
Start codon	AUG	AUG	AUG	CUG	CUG	CUG
Kozak context	Strong (−3A/+4G)	Moderate (−3A/+4C)	Weak (−3C/+4C)	Strong (−3A/+4G)	Moderate (−3A/+4C)	Weak (−3C/+4C)
IC_{50}(nM, mean ± SEM)	≫ 5 μM	701±66	114±21	35±4	11±2	11±1
5MP2						
Start codon	AUG	AUG	AUG	CUG	CUG	CUG
Kozak context	Strong (−3A/+4G)	Moderate (−3A/+4C)	Weak (−3C/+4C)	Strong (−3A/+4G)	Moderate (−3A/+4C)	Weak (−3C/+4C)
IC_{50}(nM, mean ± SEM)	≫ 5 μM	764±87	410±31	131±10	79±3	82±6

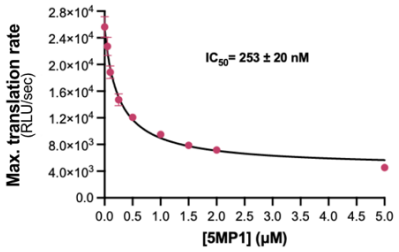
b



c



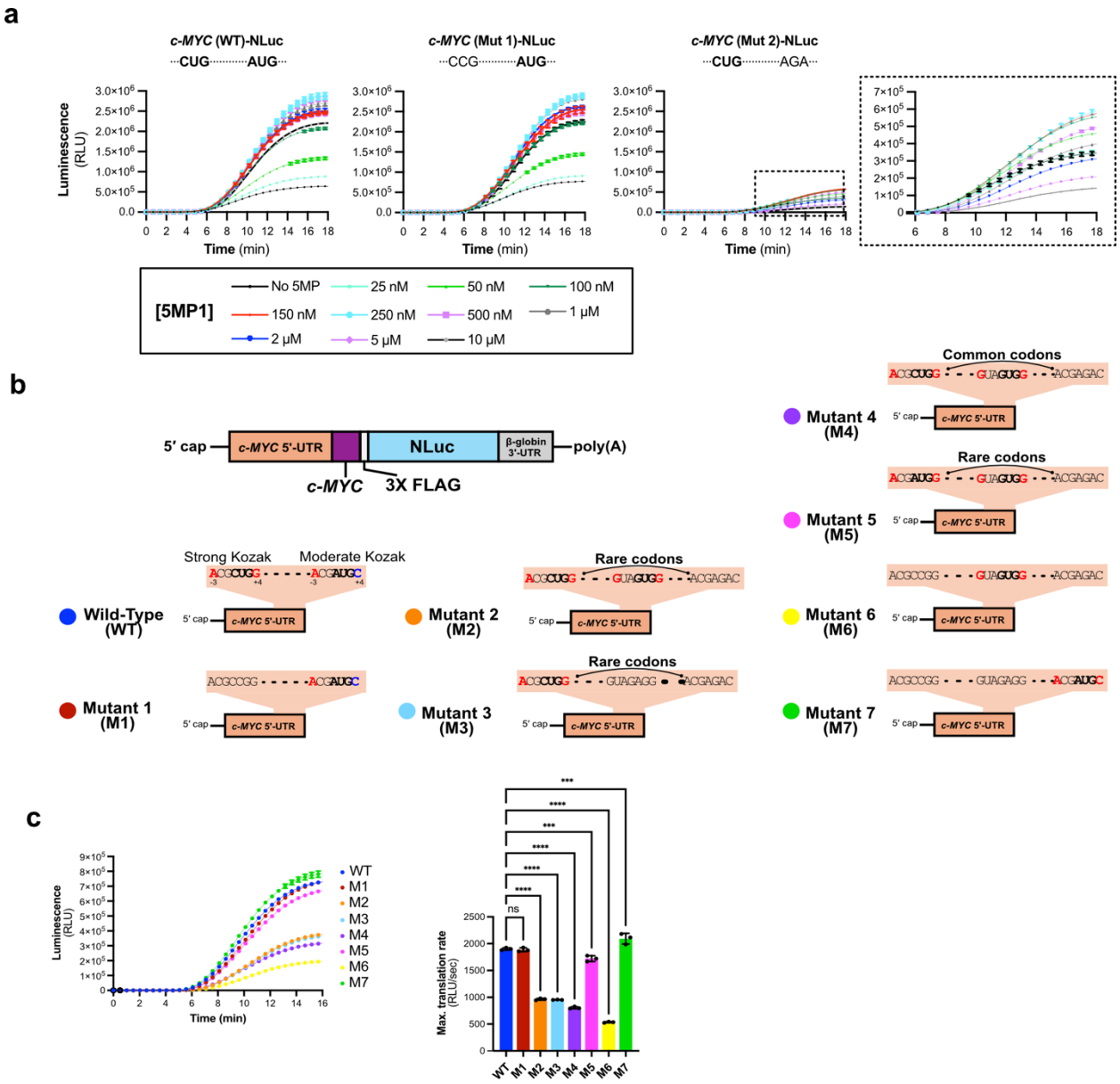
d



Extended Data Figure 3: IC_{50} values for translation inhibition by 5MP1 and 5MP2 in RRL and self-regulation of 5MP1 through a weak initiation context

(a) IC_{50} values for translation inhibition of NLuc constructs with different start codons by 5MP1 and 5MP2 in rabbit reticulocyte lysate (RRL); see additional data in main-text Figure 1. **(b)** Schematic representation of the mRNA construct containing the 5'UTR and coding sequence (CDS) of the 5MP1 (*BZW2*) transcript, followed by a 3×FLAG tag and a nanoluciferase (NLuc) reporter. To assess self-regulation, constructs were generated with either the native translation initiation site (wild type; WT), which contains an AUG codon in a weak Kozak context (−3U/+4A, left), or a mutant version (Mut) in which AUG is placed in a strong Kozak context (−3A/+4G, right). Time-course analysis of *in vitro* RRL translation with increasing concentrations of 5MP1 for the native (WT; left) and mutant (Mut; right) constructs, demonstrating the self-regulatory effect of 5MP1. **(c)** Bar chart

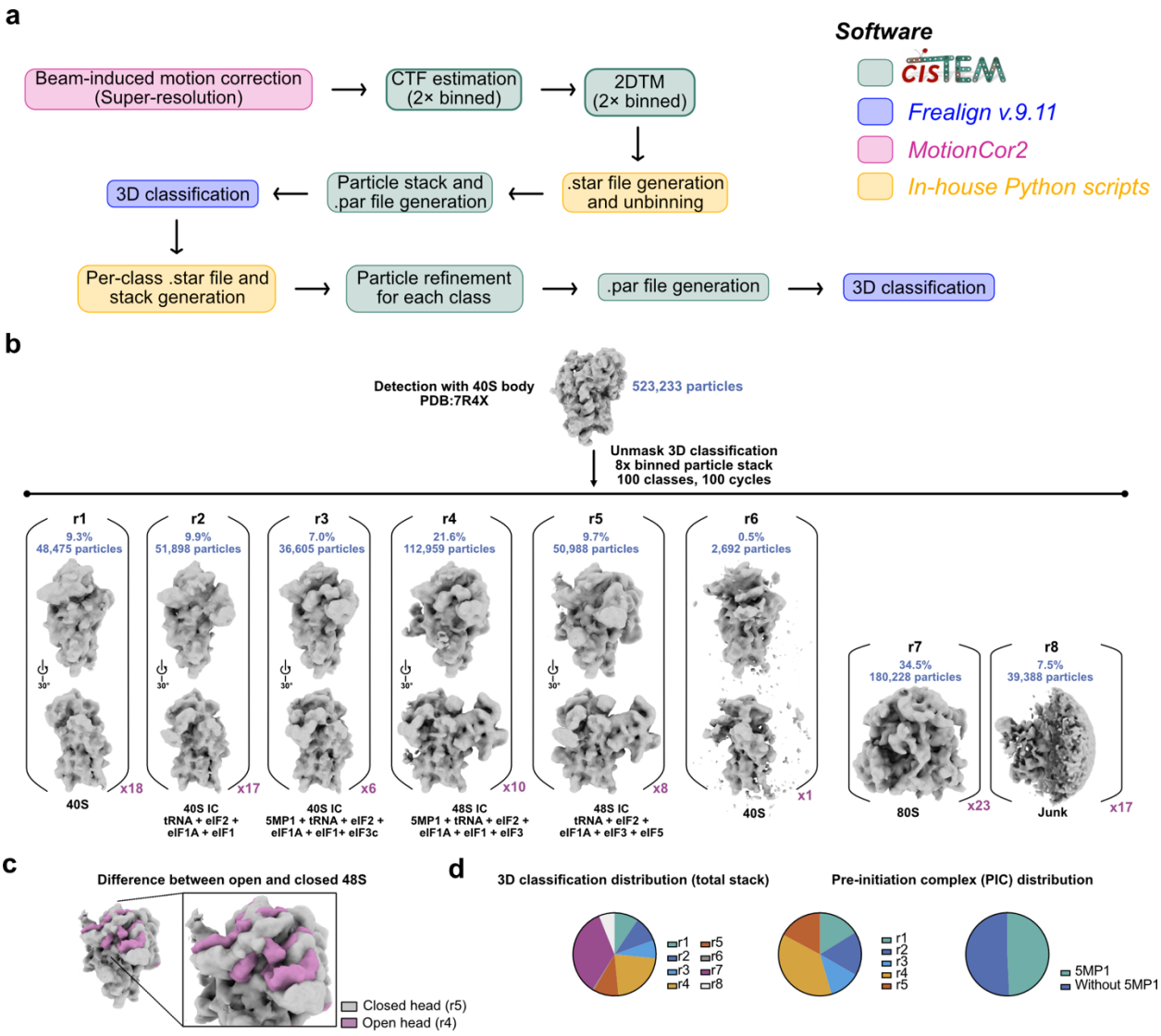
showing the effect of 5MP1 (1 μM) on translation of the native (WT; weak Kozak, left) and mutant (Mut; optimal Kozak, right) 5MP1-NLuc mRNAs. **(d)** Translation inhibition curves showing the 5MP1 inhibitory concentration (IC_{50}) for the native initiation context. Statistical significance on maximum translation rates was assessed by two-way ANOVA; ****P < 0.0001, ***P < 0.0002, **P < 0.0021, ns = not significant.



Extended Data Figure 4: 5MP1 regulation of translation from the c-Myc 5′ UTR and inefficient upstream CUG initiation

(a) Time course analysis of *in vitro* RRL translation of c-Myc-NLuc wild-type construct, CUG-Mutant (Mut 1) and AUG-Mutant (Mut2). (a-right) Close-up view of the translation curves. **(b)** Schematic representation of the c-Myc reporter constructs. Previously reported translation initiation sites included an upstream CUG within a strong Kozak context and a downstream AUG within a moderate Kozak context. Additional constructs were designed to modify initiation efficiency by mutating the CUG codon, replacing rare codons between the CUG and AUG start codons with frequent codons, or mutating the GUG codon between CUG and AUG . **(c)** Time course analysis of *in vitro* RRL translation for the indicated mutants (right) and corresponding maximum translation

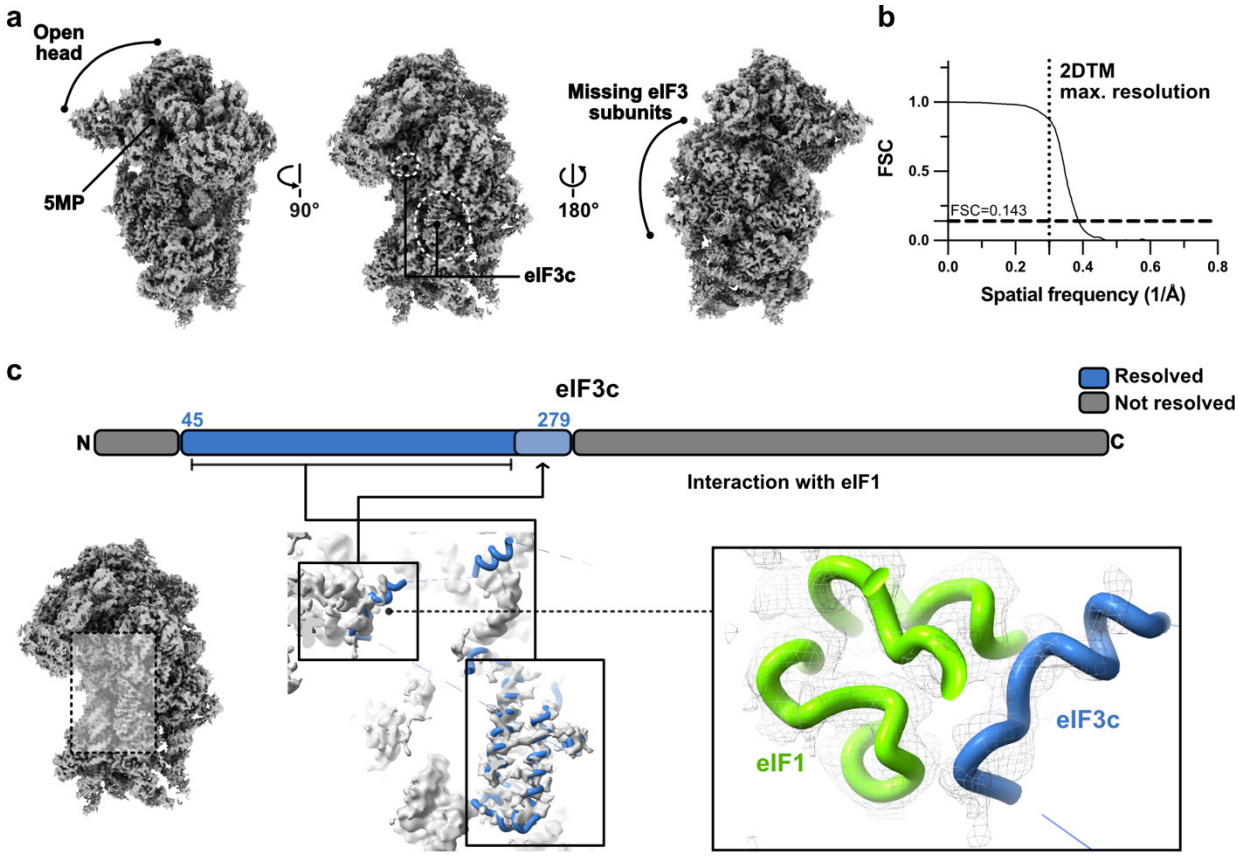
rates (left). Data represent mean ± SEM (n=3).Statistical significance of maximum translation rates was assessed by one-way ANOVA; ****P < 0.0001, ***P < 0.0002, **P < 0.0021; ns, not significant.



Extended Data Figure 5: Cryo-EM image processing and classification workflow for RRL supplemented with recombinant 5MP1

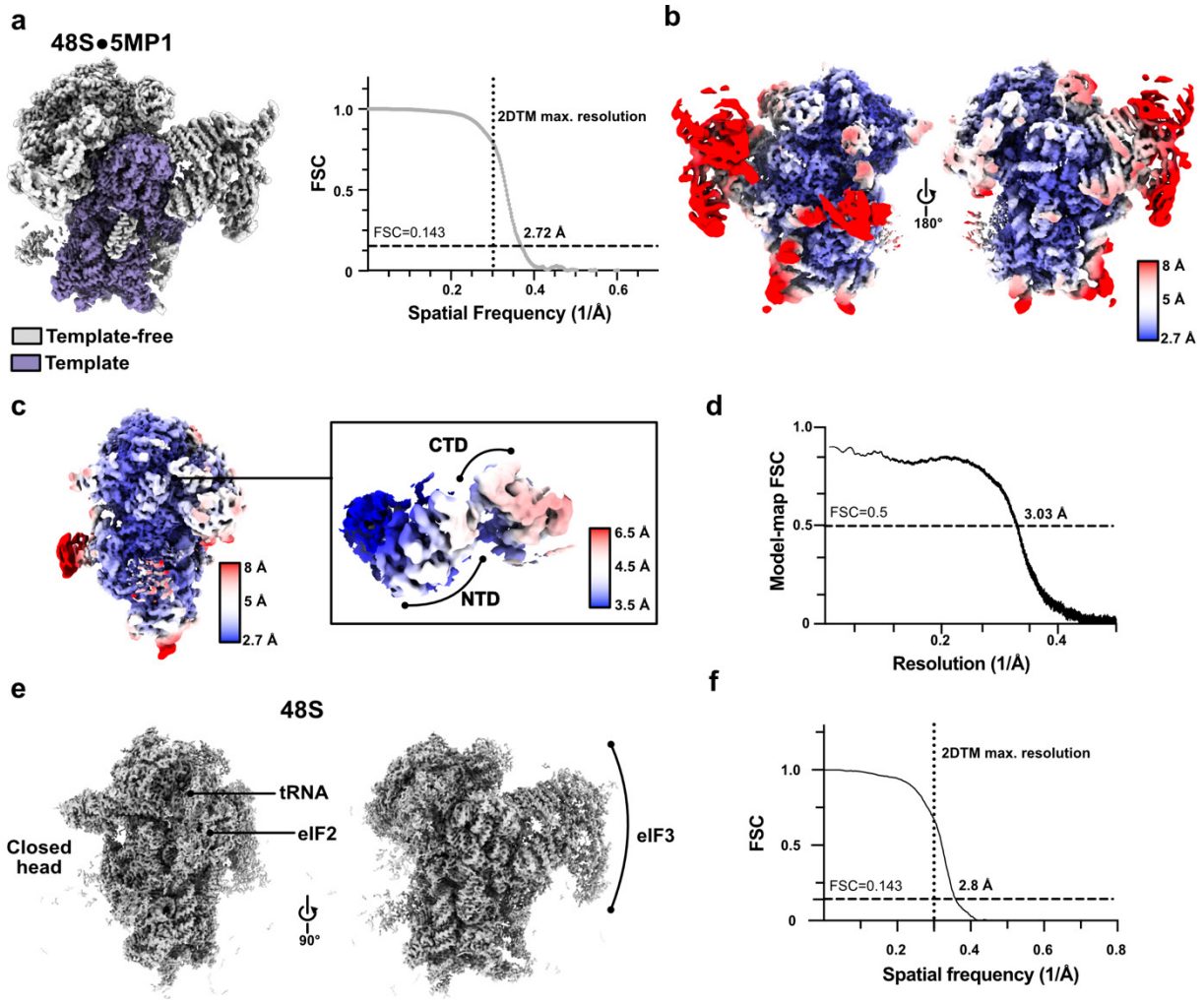
(a) Beam-induced motion correction was performed on dose-fractionated movies using MotionCor2 at super-resolution. Contrast transfer function (CTF) parameters were estimated from 2× binned micrographs using *cis*TEM, followed by 2D template matching (2DTM) on binned data. Particle coordinates were unbinned using in-house Python scripts, and corresponding particle stacks and .par files were generated. Unbinning refers to generating particle stacks from the original micrographs at the original pixel size. Initial 3D classification was carried out in Frealign v9.11. For each resulting class, per-class particle stacks and .star files were generated and subjected to independent particle refinement. Refined parameters were then used to create updated .par files for subsequent rounds of focused or global 3D classification (see Methods). Software packages used at each step are indicated by color coding. In-house scripts are available

in YAFW: https://github.com/GrigorieffLab/yafw. **(b)** 3D maximum-likelihood classification was performed in Frealign v9.11. Subclassification was carried out on binned data. **(c)** Structural differences between the open- and closed-head conformations are shown for classes r4 and r5, respectively. **(d)** Distribution of particle classes across the whole dataset and pre-initiation complexes (PICs), which correspond to ~58% of particles. Within the PIC population, ~50% of particles contain 5MP1. Among these, 25% correspond to 43S PICs with only the eIF3c subunit visible (class r3), and the rest correspond to 48S complexes containing 5MP1 (class r4). A closed-head 48S conformation (class r5) accounts for 9.7% of particles.



Extended Data Figure 6: 5MP1 binds 48S-like pre-initiation complex with weak eIF3c density and missing the rest of eIF3.

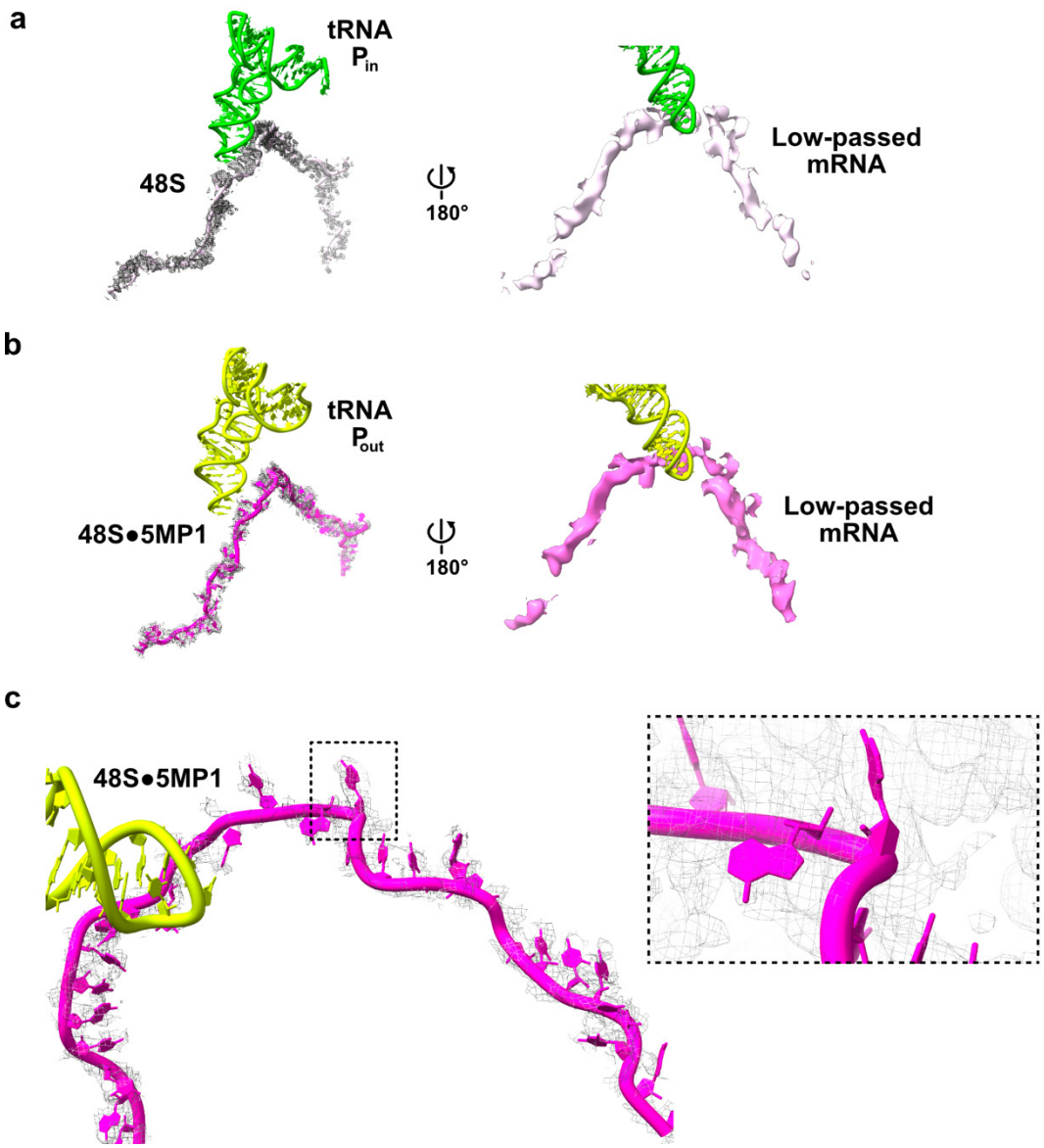
(a) Reconstruction of a 2.8 Å map of 5MP1 bound to a 48S-like pre-initiation complex showing densities consistent with the eIF3c N-terminal region. The head adopts an open conformation, similar to that observed in the 48S•5MP1 maps. Density is observed for residues 45–279 of eIF3c, whereas no density is detected for the remaining subunits of eIF3. **(b)** Fourier shell correlation (FSC) curve to estimate global resolution. The 2DTM resolution limit is indicated by a dotted line. **(c)** Densities corresponding to the eIF3c helix interacting with eIF1 are resolved. A scheme showing the resolved segments of eIF3c in blue and unresolved in gray. In the close-up view, the eIF3c model is shown in blue, and eIF1 in green. This complex supports a role of the eIF3c N-terminal domain in 5MP1-mediated regulation of initiation stringency.



Extended Data Figure 7: Assessment of average resolution, template bias and local resolution of the 48S•5MP1 and 48S pre-initiation complex.

(a) Reconstruction of the 48S•5MP1 initiation complex (class r4) following manual refinement and post-processing with DeepEMenhancer. Regions included in the template are shown in purple, and omitted areas are in gray. The Fourier shell correlation (FSC) curve used for global resolution estimation is shown. Template matching was performed on 2× binned data (1.66 Å/pixel), imposing a resolution limit of 3.32 Å (right). Therefore, density features at resolutions higher than ~3.3 Å cannot be attributed to template bias. **(b-c)** Local resolution estimation of the 48S•5MP1 map and close-up view of the local resolution for 5MP1; the N-terminal domain (NTD) and C-terminal domain (CTD) are labelled. **(d)** FSC between the model and the map was calculated for the model refined against the map. **(e)** Cryo-EM density map of the closed-head 48S complex lacking 5MP1, refined to 2.8 Å resolution. The reconstruction shows density for the initiator $tRNA^{Met}$, the eIF2 ternary complex, and eIF3. Density corresponding to mRNA in the mRNA tunnel is also present (see Fig. 4 and Extended Data Figure 11). Two orthogonal views are shown, highlighting the closed conformation of the 40S head relative to the body and the presence of eIF3. **(f)** Fourier shell correlation (FSC) curve for the global

reconstruction of the closed-head 48S complex. The 2DTM resolution limit is indicated with a dotted line.



Extended Data Figure 8: Scanning 48S•5MP1 initiation complexes exhibits continuous density for the mRNA.

(a) In the 48S complex, mRNA density is continuous throughout the mRNA tunnel, and the initiator $tRNA^{Met}$ adopts a P_{in} conformation, consistent with an arrested scanning state. **(b)** In the 48S•5MP1 complex, mRNA density is present within the mRNA tunnel, while the initiator $tRNA^{Met}$ adopts a P_{out} conformation, consistent with an active scanning state. **(c)** Close-up view of the mRNA tunnel in the 48S•5MP1 complex. Nucleotide densities are not well-resolved, consistent with increased heterogeneity in the scanning state, although density is present for some nucleotide positions, as shown in the close-up view, similar to those reported for the scanning 48S complex (EMD-17696; PDB 8PJ1).

Extended Data Table S1. Cryo-EM data collection, 2DTM results, and model refinement and validation statistics.

Data Collection	**BSC-1**	**RRL + mRNA + 5MP1**
Microscope	Titan Krios	Titan Krios
Voltage (kV)	300	300
Magnification	105000	105000
Pixel size (Å)	0.83	0.83
Detector	K3	K3
Defocus range (μm)	1.0-1.5	1.0-1.5
Total electron exposure ($e^{-}Å^{-2}s^{-}$)	~30	~30
Data collection software	SerialEM	SerialEM
Micrograph collected (no.)	162 583	104 673
Sample Geometry		
Average sample thickness (nm)*	143 ± 13	148 ±12
2DTM		
Total number of particles (no.)	1 573 843	523 233
Average SNR*	8.1	8.0
Model refinement		**48S•5MP1**
Number of particles (no.)		112959
Average map resolution (Å) (FSC threshold=0.143)		2.7
Map resolution range (Å)		2.7-16
Symmetry imposed		*C1*
Atomic model refinement		
Initial model used (PDB code)		Methods
Model resolution (Å)		2.8
Map sharpening B-factor ($Å^2$)		-50
CCmask		0.84
Model composition, no. atoms		112238
RNA residues		1813
Protein residues		9536
B-factors ($Å^2$)		
RNA residues		38.44/398.50/113.12
Proteins		37.30/487.39/159.19
Root-mean square deviations		
Bond lengths (Å)		0.015
Bond angles (°)		1.126
Validation		
MolProbity score		2.20
Clashscore, all atoms		3.30
Good sugar pucker (%)		98.62

Ramachandran plot	
Favored (%)	94.0

*Over 10 representative micrographs

Extended Data Table S2: Sequences of NLuc reporters and constructs with c-Myc 5' UTR.

Construct	Sequence (5'-3')
Strong AUG (−3A/+4G)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagacaatggatcatcatcatcatcatggctcgagcggcgtcttcacact cgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcg gggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgag cggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactgg taatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacag ggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgac cggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttc tggctaataaaggaaatttattttcattgc
Moderate AUG (−3A/+4C)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagacaatgcatcatcatcatcatcatggctcgagcggcgtcttcacact cgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcg gggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgag cggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactgg taatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacag ggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgac cggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttc tggctaataaaggaaatttattttcattgc
Weak AUG (−3C/+4C)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagccaatgcatcatcatcatcatcatggctcgagcggcgtcttcacact cgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcg gggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgag cggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactgg taatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacag ggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgac cggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttc tggctaataaaggaaatttattttcattgc
Strong CUG (−3A/+4G)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagacactggatcatcatcatcatcatggctcgagcggcgtcttcacact cgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcg gggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgag cggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactgg taatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacag

	ggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgac cggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttc tggctaataaaggaaatttattttcattgc
Moderate CUG (−3A/+4C)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagacactgcatcatcatcatcatcatggctcgagcggcgtcttcacact cgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcg gggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgag cggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactgg taatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacag ggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgac cggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttc tggctaataaaggaaatttattttcattgc
Weak CUG (−3C/+4C)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagccactgcatcatcatcatcatcatggctcgagcggcgtcttcacact cgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcg gggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgag cggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactgg taatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacag ggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgac cggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttc tggctaataaaggaaatttattttcattgc
c-Myc-NLuc (WT)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgg gcgtcctgggaagggagatccggagcgaataggggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaaccctt gccgcatccacgaaactttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgc ggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgctggat ttttttcgggtagtggaaaaccagcagcctcccgcgacgatgcccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcg agcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtctt cacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcag aatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaag gtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggca cactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgt aacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga

	gtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctg acttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (Mut 1)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgg gcgtcctgggaagggagatccggagcgaataggggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccctt gccgcatccacgaaactttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgc ggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgccggat ttttttcgggtagtggaaaaccagcagcctcccgcgacgatgcccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcg agcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtctt cacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcag aatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaag gtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggca cactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgt aacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga gtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctg acttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (Mut 2)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgg gcgtcctgggaagggagatccggagcgaataggggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccctt gccgcatccacgaaactttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgc ggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgctggat ttttttcgggtagtggaaaaccagcagcctcccgcgacgagaccccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcg agcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtctt cacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcag aatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaag gtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggca cactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgt aacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga gtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctg acttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (Mut 3)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgg gcgtcctgggaagggagatccggagcgaataggggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccctt gccgcatccacgaaactttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgc

	ggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgctggat ttttttcgggtagaggaaaaccagcagcctcccgcgacgagaccccctcaacgttagcttcaccaacaggaactatgacctcgactacggctc gagcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtc ttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttca gaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaa ggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggc acactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcact gtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacg gagtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagcccttgagcat ctgacttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (Mut 4)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgg gcgtcctgggaagggagatccggagcgaataggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttt gccgcatccacgaaactttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgc ggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgctggac ttcttccgcgtcgtggagaaccagcagcccccccgccaccagaccccctcaacgttagcttcaccaacaggaactatgacctcgactacggctc gagcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtc ttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttca gaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaa ggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggc acactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcact gtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacg gagtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagcccttgagcat ctgacttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (Mut 5)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgg gcgtcctgggaagggagatccggagcgaataggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttt gccgcatccacgaaactttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgc ggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgatggat ttttttcgggtagtggaaaaccagcagcctcccgcgacgagaccccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcg agcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtctt cacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcag aatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaag

	gtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggca cactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgt aacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga gtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctg acttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (Mut 6)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgg gcgtcctgggaagggagatccggagcgaataggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttt gccgcatccacgaaactttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgc ggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgccggat ttttttcgggtagtggaaaaccagcagcctcccgcgacgagaccccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcg agcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtctt cacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcag aatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaag gtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggca cactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgt aacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga gtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctg acttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (Mut 7)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgg gcgtcctgggaagggagatccggagcgaataggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttt gccgcatccacgaaactttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgc ggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgccggat ttttttcgggtagaggaaaaccagcagcctcccgcgacgatgccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcg agcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtctt cacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcag aatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaag gtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggca cactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgt aacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga gtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctg acttctggctaataaaggaaatttattttcattgc

Color code for NLuc constructs: green, *β-globin* 5' and 3' untranslated regions (UTRs) (*Oryctolagus cuniculus*; NM_001082260.3); red, start codon (Kozak context); brown, GSSG linker; blue, NanoLuc (NLuc) coding sequence (pNL1.1 [NLuc]; Promega).

Color code for *c-Myc* constructs: black, 5' untranslated region (UTR) of *c-MYC* (*Homo sapiens*; NM_002467.6); red, start codons (common codons); purple, *c-MYC* coding sequence (*Homo sapiens*; NM_002467.6); brown, GSSG linker; green, 3×FLAG tag; blue, NanoLuc (NLuc) coding sequence (pNL1.1 [NLuc]; Promega); gray, *β-globin* 3' untranslated region (UTR) (*Oryctolagus cuniculus*; NM_001082260.3).