Check for updates

STRUCTURAL BIOLOGY

Structural basis for start codon selection by eIF5-mimic proteins

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

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Start codon selection extensively regulates the cellular proteome and is implicated in cancer and other diseases. The eIF5-mimic proteins (5MPs: 5MP1/BZW2 and 5MP2/BZW1) restrict non-AUG start codon usage and thereby reprogram proteoform expression from mRNAs with alternative start sites, such as the oncogenic c-Myc. Here, using in extracto cryo–electron microscopy (cryo-EM) and biochemical assays, we report how 5MP represses translation dependent on the sequence context at start codons. Cryo-EM structures of 48*S* preinitiation complexes (PICs) reveal 5MP1 bound at the A site of the small ribosomal subunit. 5MPs act as initiation factors that stabilize an open PIC conformation, promoting mRNA scanning and inhibiting initiation at suboptimal start codons. We propose to designate 5MP as eIF1S (eukaryotic initiation factor 1 scanning). These findings reveal the structural basis for proteome modulation by 5MPs and may help correct dysregulated translation in disease.

INTRODUCTION

Translation initiation is a major regulatory step of gene expression, and its dysregulation contributes to a wide range of human diseases, including cancer (*1*–*3*). During initiation, the small ribosomal 40*S* subunit, together with initiation factor proteins (eIFs) and initiator $tRNA^{Met}$, collectively termed the 48*S* preinitiation complex (PIC), scans the 5′ untranslated region (5′UTR) of an mRNA until it encounters the initiation (start) codon (*4*–*7*). At the start codon, the small subunit associates with the large 60*S* 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, notably from upstream open reading frames (ORFs), expanding proteome complexity (*8*, *9*). A recent study has shown that thousands of ORFs lie outside annotated coding sequences and encode for microproteins, strengthening the case for studying upstream ORFs and mechanisms controlling their translation (*10*, *11*). Moreover, switching between AUG and non-AUG codons within a single mRNA can control the output of distinct proteoforms (*12*–*14*). For example, the increased initiation stringency, favoring AUG codons, in *c-Myc* and other mRNAs enhances translation of tumor-promoting transcripts (*15*–*19*). 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 (*20*).

The eIF5-mimic proteins 5MP1 and 5MP2 are similar and conserved across eukaryotes, except in yeasts and nematodes (*21*), are encoded by the *BZW2* and *BZW1* genes in humans, respectively, and regulate the selection of AUG start codons by suppressing initiation at non-AUG codons (*22*–*24*). Their C-terminal W2/HEAT3 domain is homologous to that of eIF5, an initiation factor promoting 48*S* progression toward subunit joining upon start codon recognition, but the N-terminal domain (NTD) bears no homology with eIF5. 5MP is thought to antagonize eIF5 and bind the W2-interacting eIF2β subunit of eIF2, whose rearrangement is critical for 48*S* PIC switching from the scanning to start codon recognition mode (*22*–*26*). 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 (*21*–*23*). 5MP2 mRNA and protein are highly expressed in most tissues, achieving levels comparable to or exceeding those of core initiation factors (e.g., eIF1A and eIF5), whereas 5MP1 is predominantly enriched in heart and muscle tissues (GTEx; Human Protein Atlas). Elevated expression of 5MP correlates with tumorigenesis, metastasis, and poor prognosis; furthermore, 5MP1 is thought to contribute to the malignant phenotype in colorectal cancer (*19*, *27*–*29*). The oncogenic activity of 5MPs is attributed, at least in part, to the direct up-regulation of the transcription factor c-Myc, a well-characterized oncogene that can drive malignant progression (*30*). Specifically, increased 5MP levels suppress translation of the long c-Myc isoform from a CUG start codon and up-regulate expression of a shorter, more tumorigenic isoform from a downstream AUG codon (*19*). Furthermore, the critical roles of 5MP in start site selection have been documented in embryonic development, neuronal homeostasis, and neurodegeneration (*23*, *31*, *32*).

Despite their critical cellular functions and therapeutic relevance, the molecular basis of 5MP action remains unknown due to a lack of structural data. Using in extracto cryo–electron microscopy (cryo-EM)—a method we recently developed for high-resolution visualization of cell lysates—we identified 5MP among endogenous translation initiation factors bound to the 40*S* subunit (*33*). In the resulting 2.7-Å structure of a 48*S* PIC, 5MP resides in the A site and stabilizes the open, scanning-competent conformation of the PIC through interactions with the 40*S* subunit, initiator $tRNA^{Met}$, and the initiation factors eIF2β and eIF1A. Our mutational analyses reveal that the N-terminal tail (NTT) in the mRNA tunnel plays a critical role in the stringency of codon selection. Comparison of numerous start site contexts shows that 5MPs act by preventing commitment to suboptimal initiation sites, resulting in AUG-initiated isoforms of *c-Myc* and other mRNAs. Together, these findings establish 5MP as a critical initiation factor that defines start codon fidelity and reveal the structural basis of 5MP-mediated control of translation initiation.

[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.
*Corresponding author. Email: niko@grigorieff.org (N.G.); andrei.korostelev@umassmed.edu (A.A.K.)
†These authors contributed equally to this work.

Downloaded from https://www.science.org on November 03, 2026

RESULTS AND DISCUSSION

5MP binds 48*S* 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 combined with two-dimensional template matching (2DTM), an approach we refer to as in extracto cryo-EM (*33*). 2DTM is a computational approach that we developed to identify individual particles in noisy, low-contrast environments (*34*, *35*), such as cellular lysates. Traditional single-particle cryo-EM requires purifying the macromolecule(s) of interest from cellular components and/or reconstitution of a complex from a defined subset of purified factors. To allow for identification of previously unidentified cellular complexes, in extracto cryo-EM uses a known fragment of a molecular template (e.g., a part of a ribosomal subunit) to detect complexes of interest in the context of cellular components. This approach was recently used to resolve near–atomic resolution structures of endogenous 80*S* complexes in cell lysates, demonstrating that in extracto cryo-EM can recover a new set of bound proteins that were inaccessible in purified or reconstituted systems (*33*). In the present work, we used the 40*S* body as the 2DTM template because it is present in all PICs and undergoes fewer conformational changes than other regions of the 40*S* subunit, such as the head. 3D classification of particles identified by 2DTM in cellular lysates revealed a class whose A site exhibited density that could not be assigned to known translation factors (Fig. 1A, fig. S1, and Materials and Methods). This class corresponds to a 48*S* initiation complex, as indicated by the presence of initiation factors, initiator $tRNA^{Met}$, and mRNA (fig. S1B and Materials and Methods). Backbone and side-chain tracing with ModelAngelo (*36*) in our 3-Å resolution map, followed by comparison with the AlphaFold-predicted structure library (*37*), indicated that the density can be accounted for by a 5MP protein (Fig. 1A, fig. S1D, and Materials and Methods). This finding suggests that 5MPs may directly and specifically control translation initiation from AUG or non-AUG codons by binding to the A site. To corroborate this idea and further elucidate the mechanisms by which 5MP regulates start codon selection, we purified recombinant human 5MPs and structure-guided mutants (fig. S2) for biochemical and additional 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 strong, moderate, or weak Kozak contexts (*38*, *39*). 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 (*38*). We added mRNA reporters to a rabbit reticulocyte lysate (RRL) translation system in the presence or absence of recombinant 5MPs and measured luminescence (Fig. 1B). We estimated that the endogenous 5MP concentration in RRL is below 25 nM (fig. S2C). Translation from an AUG codon in a strong Kozak context was unaffected in the presence of 1 μM 5MPs, underscoring the robustness of canonical initiation [median inhibitory concentration (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 and 410 nM, respectively) (Fig. 1, C and D). As 5MP1 and 5MP2 share ~70% overall sequence identity (Expasy; SIM alignment), including fully conserved residues mediating PIC engagement, the difference in potency may be due to different interactions with other cellular components via less conserved regions or other factors to be established (see Discussion). 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 to 11 nM; 5MP2: 131 to 79 nM; Fig. 1, C and D, and fig. S3, A and B). Overall, the inhibitory effect of 5MP can therefore be >100-fold stronger on CUG codons than on the strong AUG codon. These findings align with previously reported results in a cellular context, indicating that the observed regulatory effect is not specific to RRL (*25*). For example, cellular levels in human embryonic kidney 293T cells are estimated at ~750 and ~820 nM for 5MP1 and 5MP2, respectively (*40*), well above the IC_{50} values measured for weak AUG and CUG contexts, suggesting that this inhibitory activity is physiologically relevant in vivo.

5MP proteins autoregulate their translation from an AUG in a conserved weak Kozak context (*24*). We therefore measured the ability of 5MP to inhibit translation from the UUUAUGA context of 5MP1 (BZW2) mRNA (fig. S3C). 5MP1 inhibited translation from its own AUG with an IC_{50} of ~253 nM, underscoring the ability of 5MP to repress a moderate to weak AUG context (fig. S3, D and E). In cells, this feedback mechanism likely regulates 5MP concentrations, enabling differential translation from AUG and non-AUG codons across the transcriptome (*24*).

5MP1 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 (*19*). The "p67" CUG start codon is in a strong Kozak context, and the "p64" AUG start codon is in a moderate Kozak context (Fig. 2A). To dissect the effects of 5MP1 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 measured luminescence (Fig. 2, A and B, and fig. S4A) and proteoform levels (Fig. 2, C and D) in the absence or presence of increasing concentrations of 5MP1. Consistent with studies in cell lines (*19*, *23*), 5MP1 increased the total expression of the *c-Myc* construct (Fig. 2, B to D) and shifted expression toward the shorter AUG-dependent proteoform by inhibiting the expression from the p67 CUG start codon (Fig. 2, C and D). Mutations that disrupt the AUG or CUG codon individually confirm this trend (Fig. 2, B to D, fig. S4, A and B). In the absence of the AUG codon, 5MP1 suppresses CUG translation at a concentration of 1 μM (Fig. 2, C and D). By contrast, AUG initiation from its native position in the *c-Myc* 5′UTR is robust even at 15 μM 5MP1. Together, our data demonstrate that 5MP1 efficiently changes the ratio of CUG and AUG proteoforms of *c-Myc* at 5MP1 concentrations that are above its autoregulatory IC_{50} (fig. S3E).

On the *c-Myc* UTR with both start codons, 5MP1 not only repressed the production of the CUG isoform but also increased the AUG isoform output (Fig. 2, C and D). 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. To distinguish between these possibilities, we generated a series of *c-Myc* reporter mutants and measured their translation (fig. S4, C and D). Mutating the upstream CUG while leaving the downstream AUG (mutant 1) had no effect on translation. By contrast, mutating the AUG while retaining the upstream CUG and/or another potential start codon GUG (mutants

Downloaded from https://www.science.org on November 03, 2026

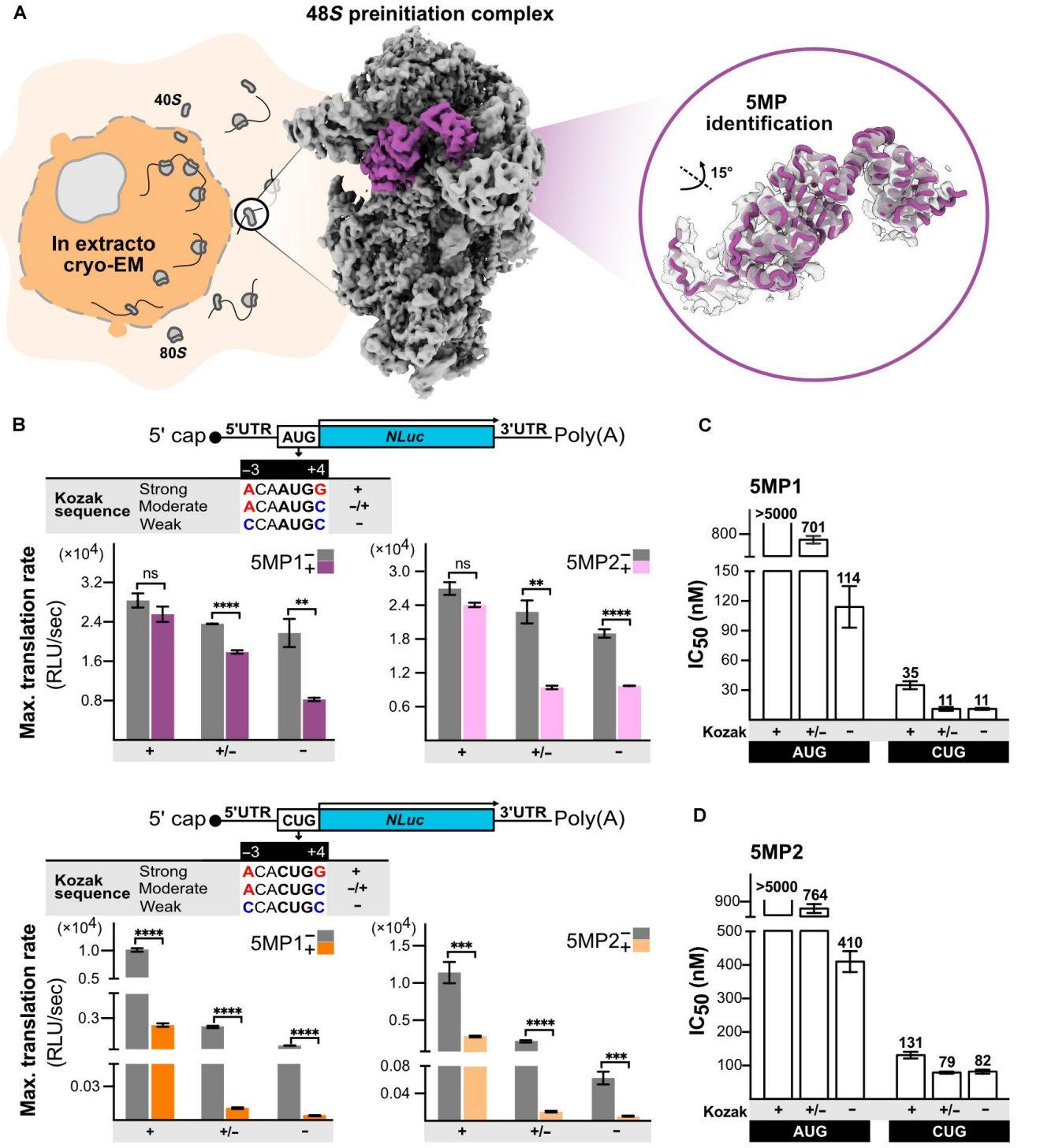


Fig. 1. 5MP proteins bind the cellular 48*S* PIC and inhibit translation initiation in a Kozak sequence–dependent manner. (**A**) Schematic of in extracto cryo-EM and reconstruction of the 48*S* PIC, highlighting unassigned density overlaid with a refined AlphaFold2-predicted model of eIF5-mimic protein (5MP) (close-up). (**B**) Schematic of mRNA constructs and 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 ****$P < 0.0001$, ***$P < 0.0002$, **$P < 0.0021$, and ns (nonsignificant) according to the unpaired two-tailed t test. Left: 5MP1. Right: 5MP2. (**C**) Inhibitory concentrations (IC_{50}) of 5MP1 (top) and (**D**) 5MP2 (bottom) for AUG and CUG initiation in strong (+, −3A/+4G), moderate (+/−, −3A/+4C), and weak (−, −3C/+4C) Kozak contexts.

Downloaded from https://www.science.org on November 03, 2026

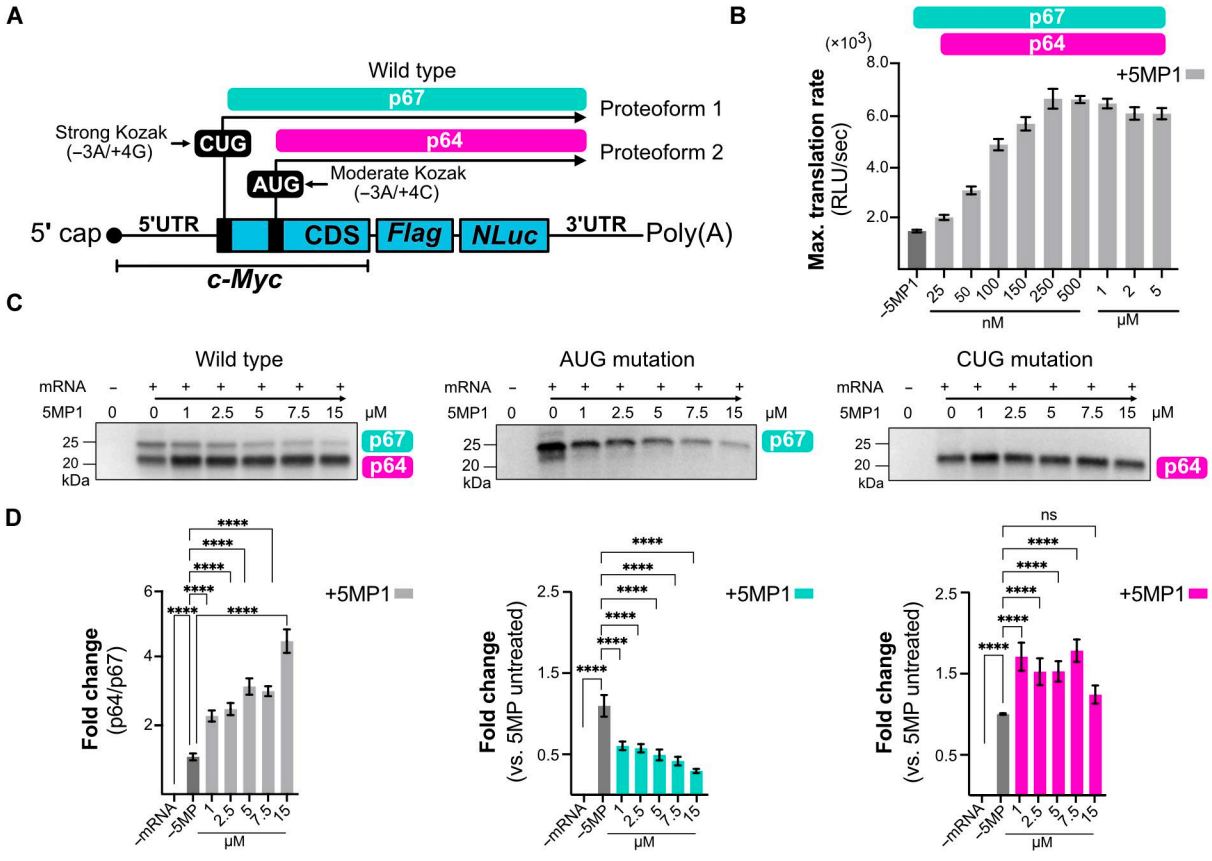


Fig. 2. 5MP1 reprograms initiation from the *c-Myc* 5′UTR. (**A**) Schematic of the mRNA reporter with *c-Myc* 5′UTR containing CUG and AUG start codons in distinct Kozak contexts. AUG-initiated p64 products are indicated in pink and CUG-initiated p67 products in green. (**B**) Maximum translation rates plotted at different 5MP1 concentrations for the *c-Myc* wild-type reporter (see also fig. S4). Data represent mean ± SD values ($n = 3$). (**C**) Translation products from wild-type, AUG-mutated, and CUG-mutated constructs. Representative Western blotting and (**D**) quantifications are shown: (left) fold change in proteoform ratio (p64/p67); (middle and right) p67 and p64 fold changes relative to 5MP1 untreated. Statistical significance was assessed by one-way analysis of variance (ANOVA); ****$P < 0.0001$.

2 to 4) reduced translation twofold relative to the wild-type *c-Myc* transcript. Converting the upstream CUG to an AUG (mutant 5) restored translation to nearly wild-type levels, indicating that the baseline suppression of downstream AUG is primarily due to inefficient initiation at the upstream CUG and GUG codons. Furthermore, replacing the suboptimal elongation codons between the CUG and AUG start codons with synonymous optimal codons (mutant 4) did not increase translation from the CUG codon, arguing against inefficient elongation. These findings are consistent with biochemical studies showing that AUG initiation is at least twice as efficient as initiation at near-cognate UUG and CUG start codons in the same sequence context (*41*–*46*). Accordingly, the less efficient initiation from the CUG-p67 start codon results in lower overall c-Myc production. Notably, the ~400-nt long *c-Myc* 5′UTR harbors nearly 20 in-frame and out-of-frame near-cognate start codons that, in the absence of 5MP1, may compete with the AUG codon for ribosome engagement (table S1). More generally, the probability of encountering an NUG triplet at any 5′UTR position is ~1/16, making the scanning process on an average mRNA inefficient and prone to low-fidelity initiation without 5MP. Suppression of these competing initiation events by 5MP would be necessary for 48*S* to efficiently scan through to the AUG codon, thereby explaining the observed increase in AUG isoform output. In summary, our data demonstrate that when near-cognate and canonical start codons compete for ribosome engagement, 5MP1 proteins shift the isoform ratio toward the AUG-initiated isoform and increase its output.

5MP1 stabilizes scanning-competent 48*S* PIC

To investigate the structural basis of 5MP1-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 nonhydrolyzable GTP analog guanosine 5′-[β,γ-imido]triphosphate (GMPPNP) to capture initiation complexes with guanosine triphosphatase initiation factors eIF2 and/or eIF5B (*47*–*51*). To enable particle detection despite the low contrast of lysate micrographs, we used 2DTM (*34*, *35*, *52*) using the 40*S* body as the search template [Protein Data Bank (PDB): 7R4X; Materials and Methods and table S2]. The detected particles comprised a range of structurally distinct PICs and 80*S* complexes, indicating that 2DTM was mostly unbiased by the incomplete-subunit template. Cryo-EM data classification yielded predominantly 48*S* PICs (~31%), containing eIF1, eIF1A, eIF2, and eIF3. Nearly half of PICs contained 5MP1,; of all the 48S PICs 37% are 48*S* complexes and 12% are 48S-like assemblies containing the NTD of eIF3c subunit near eIF1 but missing the rest of eIF3 (figs. S5, B to D, and S7). Specifically, a previously unresolved segment of eIF3c forms an α

Downloaded from https://www.science.org on November 03, 2026

helix (Thr^{250} to Asp^{290}), whose conserved FLKK motif 3 inserts into a hydrophobic pocket of eIF1 (fig. S7D), consistent with the recent report of eIF3c/eIF1 interfaces (*53*). These observations suggest a correlation between eIF3c stabilization in a PIC and 5MP1 binding (*23*). Other classes included free 40*S* subunits and 40*S* complexes containing initiator tRNA^{Met}, eIF1, eIF1A, and eIF2 (fig. S5B).

A 2.7-Å reconstruction of the predominant 48*S*•5MP1 complex features 5MP1 at the A site (Fig. 3), placed identically to the endogenous 5MP1 detected in cell lysates described above (Fig. 1 and fig. S1). Notably, cryo-EM data contain both open-head (i.e., scanning-competent) and closed-head (i.e., codon recognition) 48*S* PICs, but 5MP1 is bound to 48*S* PIC only in the open-head conformation (Fig. 4 and figs. S5 to S7). In canonical open-head initiation complexes, the head is tilted ~13° away from the body, relative to that in closed-head (codon recognition) complexes (*54*–*56*). In 5MP1-bound complexes, the 40*S* 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 (Materials and Methods) (*54*, *55*). The initiator tRNA^{Met} is positioned ~7 Å away from its codon-engaged position, nearly identical to P_{out} conformations observed in scanning-competent 48*S* PICs (Fig. 4B) (*57*, *58*). Although relatively continuous mRNA density is visible in the mRNA tunnel within the 5MP1-bound complexes (fig. S8), individual nucleotides are not resolved near the P site, consistent with the heterogeneity observed in scanning 48*S* complexes (Fig. 4C and fig. S8) (*4*, *7*). This contrasts with a well-resolved AUG codon in our 2.8-Å reconstruction of codon recognition closed-head 48*S* complexes lacking 5MP1, which is consistent with observations in previous studies of canonical AUG initiation (Fig. 4C and fig. S8) (*4*, *7*).

5MPs consist of two domains, which contact eIF2, initiator tRNA^{Met}, eIF1A, and the mRNA tunnel (Fig. 5A). The C-terminal domain (CTD; residues 248 to 407) docks onto eIF2β and initiator tRNA^{Met}. Here, Tyr^{351} and Ile^{354} of the W2/HEAT3 domain—previously implicated in interactions with eIF2β (*23*, *26*)—bind the universally conserved His^{230} of eIF2β through π-π stacking (Tyr^{351}) and hydrophobic packing (Ile^{354}) (Fig. 5B). A Tyr^{351}Ala/Ile^{354}Ala double mutant (Y351A/I354A) of 5MP1 renders the protein ~4-fold less inhibitory than wild-type 5MP1 on a transcript containing a CUG start codon in a strong Kozak context (Fig. 5F and fig. S9), highlighting the importance of this contact for 5MP function. Residues Phe^{401}/Val^{402}/Trp^{404}/Leu^{405} of 5MP1, identified as part of the 7A mutation and implicated in the interaction with the eIF2β N terminus (*23*), are resolved in our map, and nearby densities may correspond to eIF2β N terminus; however, unambiguous assignment is not possible due to insufficient resolution, leaving open the possibility of transient interactions (Fig. 3 and figs. S7C and S10). This possibility is consistent with suggestions that the intrinsically disordered NTT of eIF2β engages 5MP through dynamic and competitive contacts at its lysine-rich boxes (*26*, *59*). Residues in the previously

Downloaded from https://www.science.org on November 03, 2026

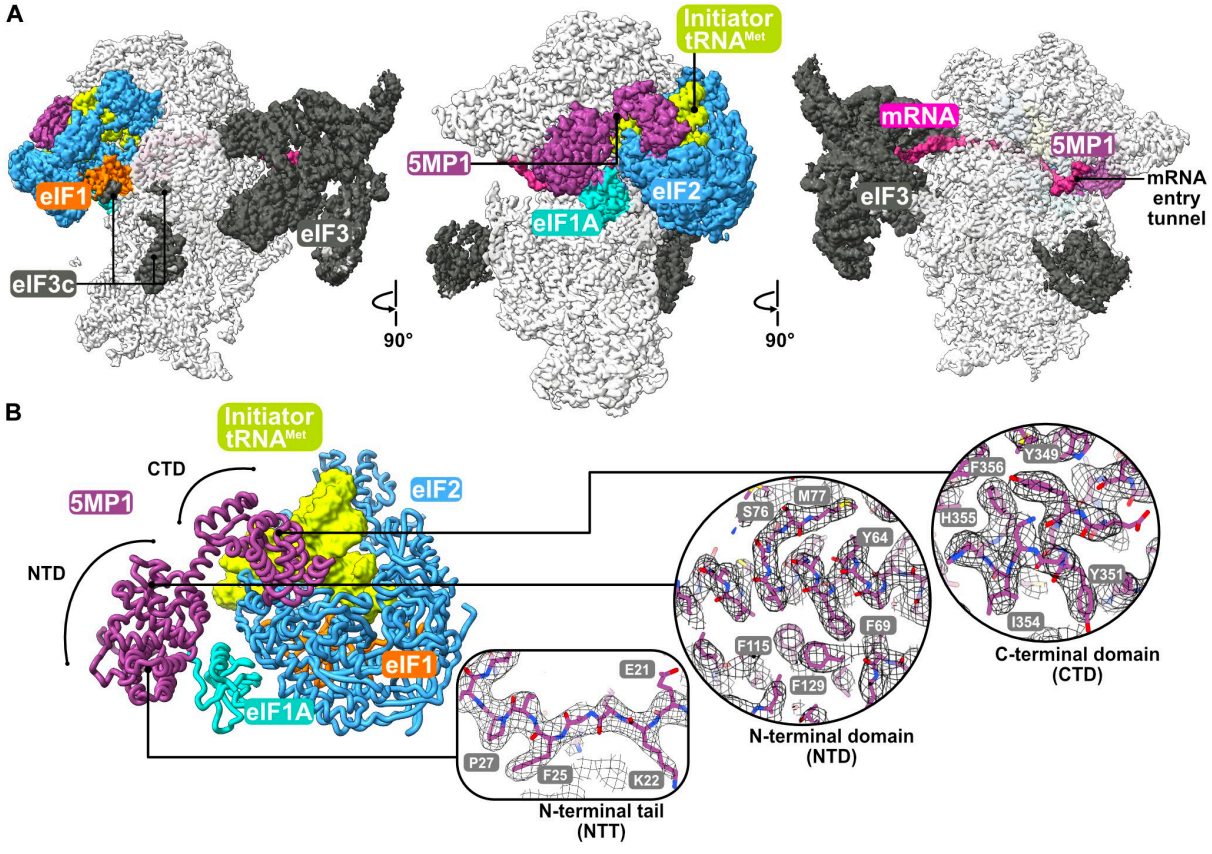


Fig. 3. Cryo-EM of 5MP1 bound to 48*S* in RRL. (**A**) A 2.7-Å cryo-EM map of the 48*S*•5MP1 initiation complex, shown in different orientations. Segmented densities show 5MP1 in purple, eIF2 in blue, eIF1 in orange, eIF1A in turquoise, and the initiator tRNA^{Met} in green. The middle view shows 5MP1 binding at the A site. The left and right show the intersubunit and solvent sides of the 48*S* initiation complex. (**B**) 5MP1 interactions with initiation factors and initiator tRNA^{Met}. Insets show the NTT, NTD, and CTD; cryo-EM density is shown as mesh.

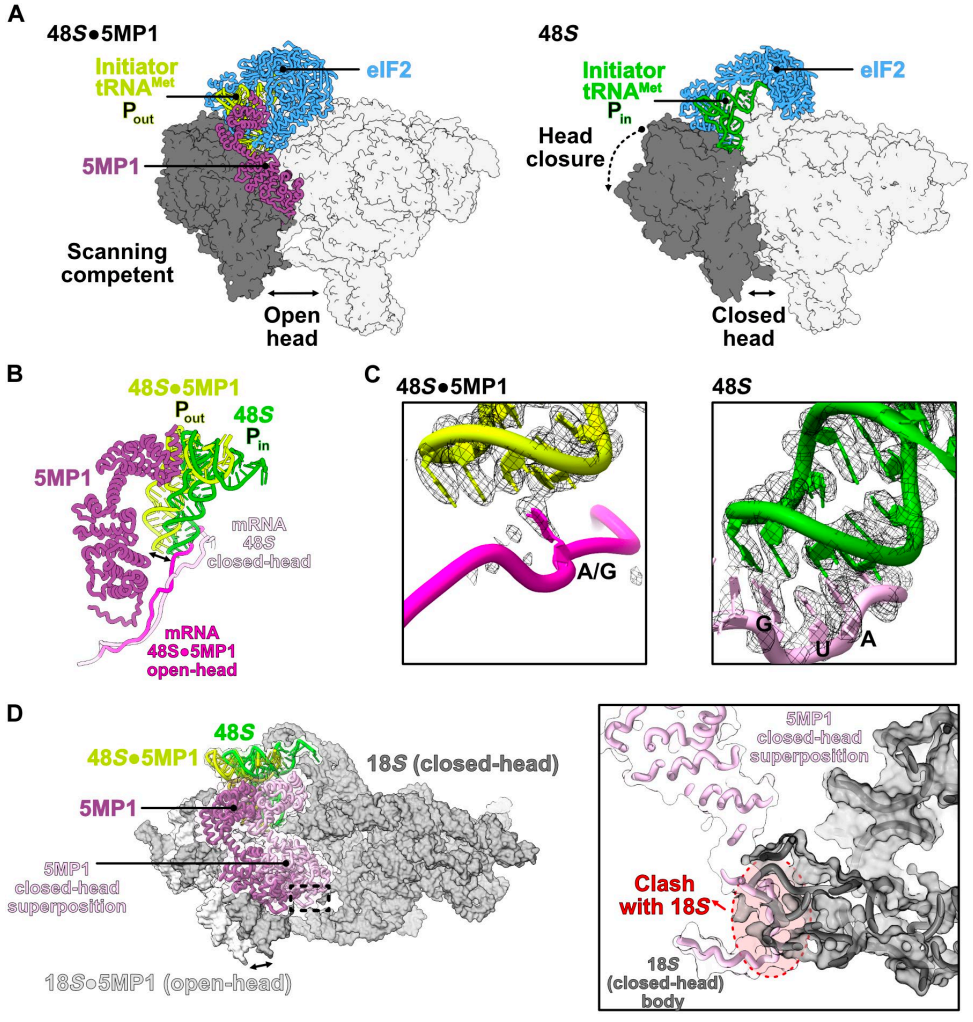


Fig. 4. 5MP1 promotes the open-head conformation of the 48*S* complex. (**A**) The 48*S* complex containing 5MP1 adopts an open conformation (left), whereas the 48*S* complex without 5MP1 is predominantly closed (right). In the 5MP1-bound structure, eIF2 (blue), initiator $tRNA^{Met}$ (green), and 5MP1 (purple) are shown as cartoons. Initiator $tRNA^{Met}$ in the 48*S* complex without 5MP1 is shown in dark green. (**B**) In the presence of 5MP1, the initiator $tRNA^{Met}$ adopts a P_{out} conformation (left), 7 Å away (arrow) from the P_{in} conformation in the closed 48*S* complex without 5MP1. The mRNA in the 5MP1-bound complex is shown in pink and in the 48*S* complex without 5MP1 in light pink. (**C**) The close-up shows no codon-anticodon pairing in the 5MP1-bound complex, in contrast to the well-resolved anticodon-paired AUG in the 48*S* complex without 5MP1. (**D**) An overlay of the 5MP1 model with the 48*S* closed-head conformation (superposition) shows a steric clash between the 5MP1 NTT and helix 18 of 18*S* rRNA. Experimental 5MP1 model is shown in purple and superimposed 5MP1 in lilac.

characterized BN1 mutant (Lys^{380}/Lys^{383}/His^{386}/Lys^{391}) (*23*) are located near eIF2β (fig. S10), and the poorly resolved density noted above for 5MP1-7A and eIF2β also lies in proximity to these residues. Alternatively, the unresolved density may correspond to the N-terminal segment of eIF3c rather than the eIF2β NTT, as the former was shown to stabilize 5MP1 in the 48*S* complex (*23*). Given the ambiguity and low resolution of this density, we also cannot exclude that the density reflects noise rather than one of these interactions. Flexible low-occupancy interactions are challenging to resolve by cryo-EM. Crosslinking coupled to mass spectrometry or other approaches would help corroborate these observations and map the contact residues.

Residues Lys^{309} and Lys^{310} 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). This interaction appears to be as important for 5MP1 function as the contact with eIF2β described above. Specifically, substituting the Lys^{309}/Lys^{310}/Glu^{311}/Glu^{312} (KKEE) motif with Glu/Glu/Gln/Gln (EEQQ) to neutralize this positively charged surface reduced 5MP1-mediated inhibition by nearly fourfold on the same CUG reporter (Fig. 5F and fig. S9; KKEE mut). Because eIF5 shares a homologous C-terminal W2/HEAT3 domain and binds eIF2β (*22*, *43*), we asked whether the eIF5 CTD would similarly interact with the 48*S* scanning-competent complex and whether the contacts with eIF2β and initiator $tRNA^{Met}$ reported here are compatible. Superposition of eIF5 CTD onto 5MP1 CTD within the 48*S*•5MP1 complex suggests similar interactions with the 48*S* complex (fig. S11A), consistent with their competition at this site (*60*). However, despite the overall similarity, there are differences between 5MP CTD and eIF5 CTD interactions within the 48*S*: (i) eIF5 residues 366 to 368 can make additional contacts with eIF2β that are not observed with 5MP1, and (ii) the eIF5 helix reaching toward initiator $tRNA^{Met}$ is shorter compared to the one in 5MP and is likely incompatible with interactions with the initiator $tRNA^{Met}$ (fig. S11B). These differences highlight that

Downloaded from https://www.science.org on November 03, 2026

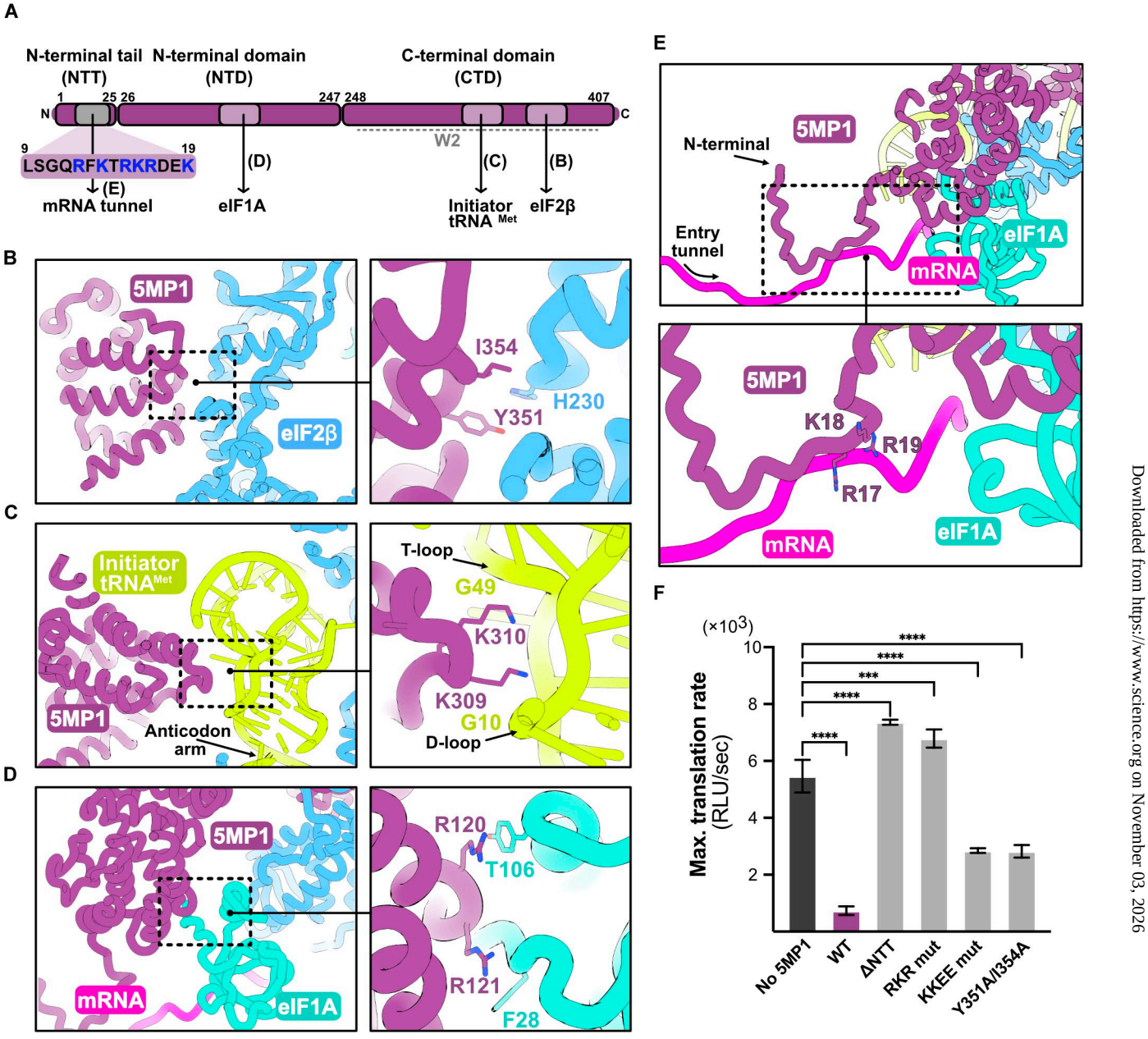


Fig. 5. Interactions of 5MP1 within the 48*S* PIC. (**A**) Schematic representation of 5MP1 domain organization, interactions within the 48*S* complex and the conserved sequence in the NTT. (**B**) Close-up view of the 5MP1-eIF2β, (**C**) 5MP1 initiator tRNA^{Met}, and (**D**) 5MP1-eIF1A interfaces, highlighting the interacting residues. (**E**) Close-up view of NTT highlights the Arg-Lys-Arg motif facing the mRNA backbone (right). (**F**) Effects of structure-guided mutations on 5MP1-mediated inhibition of translation from a transcript containing a CUG start codon in a strong Kozak context, measured at 1 μM (mean ± SEM, $n = 3$): wild-type (WT) 5MP1; the NTT deletion mutant (ΔNTT; residues 1 to 25); the RKR(17-19)–to–AAA mutant (RKR mut); the KKEE(309-312)–to–EEQQ mutant (KKEE mut); and the Y351A/I354A double mutant. Data represent mean ± SD ($n = 3$). Statistical significance was assessed by one-way ANOVA; ****$P < 0.0001$, and ***$P < 0.0002$. 5MP1 is shown in purple, eIF2 in blue, eIF1A in turquoise, and the initiator tRNA^{Met} in green.

the CTD of 5MP1 and eIF5 engage the 48*S* through distinct molecular interactions despite their shared domain architecture. In addition, 5MP1 NTD interacts with the 48*S* complexes (see discussion below), which may be essential for stabilizing this interaction, consistent with the inability to resolve eIF5 CTD in numerous 48*S* complexes that have a clearly resolved eIF5 NTD (*26*).

Last, the 40*S* A site accommodates the 5MP1 NTD next to eIF1A, stabilized by cation-π interactions between 5MP1 (Arg^{120} and Arg^{121}) and eIF1A (Tyr^{106} and Phe^{28}; Fig. 5D). The NTT extends into the mRNA tunnel (Fig. 5E), where the terminal residues form a β strand with protein uS3 and the conserved Arg^{17}-Lys^{18}-Arg^{19} motif may be involved in contacts with 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 18*S* ribosomal RNA (rRNA) (Fig. 4D) (*55*). This steric hindrance is consistent with the absence of 5MP1 in

Downloaded from https://www.science.org on November 03, 2026

the closed-head maps. To test the functional importance of the NTT, we generated two mutants. Deleting the entire NTT (ΔNTT; residues 1 to 25) or mutating the $Arg^{17}/Lys^{18}/Arg^{19}$ motif to three alanines alone (RKR mut) abolished 5MP1's inhibitory activity on a transcript containing a CUG start codon in a strong Kozak context, relative to wild-type 5MP1 (Fig. 5F and fig. S9). The comparable magnitudes of these effects suggest that the RKR motif alone accounts for the functional contribution of the NTT to 5MP1-mediated inhibition. The RKR motif is among the most highly conserved elements across 5MP homologs, from early-diverging unicellular lineages to vertebrates (Materials and Methods and fig. S12), emphasizing that NTT-mediated engagement in the mRNA tunnel represents an ancient and conserved mechanism for controlling start codon fidelity.

Together, these findings reveal that 5MP1 interacts with the 48*S* PIC through multiple contacts—spanning the NTT, NTD, and CTD—that collectively bias the complex toward the open-head, scanning-competent state. The residues directly mediating 5MP1's contacts within the 48*S* PIC are fully conserved in 5MP2 (fig. S9D), indicating that 5MP2 is poised to engage the 48*S* PIC through the same mechanism. The most pronounced sequence divergence is located in a short segment (residues ~88 to 93) immediately following the NTD (fig. S9). Superposition of an AlphaFold2-predicted structure of 5MP2 suggests that this region may form a flexible loop near the 18*S* rRNA of the 40*S* beak (fig. S9E) that may influence 5MP2 binding or interaction of the NTT with the mRNA tunnel, in keeping with the higher IC_{50} for 5MP2 (Fig. 1, C and D). Posttranslational modifications represent another layer of functional regulation: Both 5MP paralogs contain a conserved casein kinase 2 phosphorylation site within their CTD, homologous to a site in eIF5 whose phosphorylation modulates eIF2β binding (*26*). These functional consequences remain to be explored in future work. Together, the unique structural contacts emphasize that 5MP1 function extends beyond being just an eIF5 competitor and define it as a bona fide initiation factor that stabilizes the scanning-competent 48*S* state.

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 (*24*).

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 binds exclusively within the open-head 48*S* complexes, suggesting a straightforward structural mechanism of 5MP action (Figs. 4 and 6). To reduce initiation at a suboptimal start

Downloaded from https://www.science.org on November 03, 2026

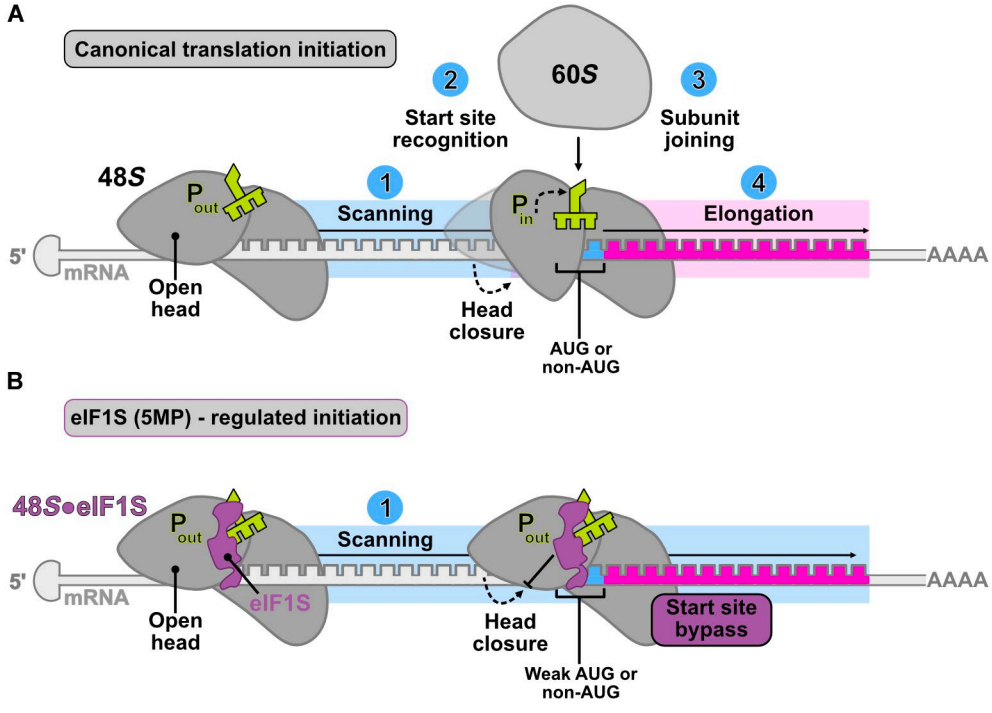


Fig. 6. Schematic of eIF1S (5MP)-mediated regulation of start codon recognition. (**A**) During canonical initiation, the 48*S* complex scans the mRNA until it recognizes a start codon. Both canonical AUGs and suboptimal codons, including weak AUGs and non-AUGs, can trigger scanning arrest. This arrest is accompanied by a conformational shift of the 48*S* 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 60*S* joining and elongation. (**B**) By contrast, eIF1S (5MP) stabilizes the 48*S* complex in a scanning-competent state, maintaining the 48*S* head in the open conformation and the initiator $tRNA^{Met}$ 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.

site, 5MP shifts the equilibrium between 48*S* conformations toward a widely open, scanning-competent state. Within 48*S*, 5MP interacts with eIF2β, initiator $tRNA^{Met}$, eIF1A, and the mRNA tunnel (Fig. 5). The NTT of 5MP is folded within the mRNA entry tunnel widened by the head opening (Fig. 5), and head closure is incompatible with this position due to a steric hindrance (Fig. 4D), highlighting the NTT as the functional core of 5MPs. The transition of 48*S* into the closed codon recognition state (e.g., on a strong AUG codon) requires eIF2β displacement (*7*), dissociation of eIF1, and the closure of the 40*S* head. This must coincide with or succeed 5MP dissociation, as is indicated by the absence of 5MP in closed 48*S* states and by the incompatibility of the NTT with the closed 48*S*. One possibility is that 5MPs associate with the scanning 48*S* reversibly and that its binding is prevented when a strong codon recognition biases fast head closure (*7*, *55*). Alternatively, for a stably bound 48*S*•5MP scanning complex, an early interaction of the eIF2•initiator $tRNA^{Met}$ and/or other initiation factors with the AUG codon and Kozak context may induce a transient conformational rearrangement that reduces the affinity of 5MP for the 48*S*. This is consistent with prior work showing that eIF1 can dissociate from the PIC upon codon-anticodon pairing at a strong AUG codon (*61*). These scenarios may explain why 5MPs have little effect on strong AUG/Kozak contexts, where codon recognition outcompetes 5MP binding, while they strongly inhibit initiation at weak or non-AUG codons subject to a slower initiation (Fig. 1B).

Together, our findings elucidate how 5MPs add a robust layer of stringency to canonical initiation factors and shift our understanding of their structural roles in translation. Rather than acting solely through competition with eIF5, 5MP functions as a bona fide initiation factor that binds the 48*S* complex and biases it toward scanning, preventing commitment at suboptimal start codons. To reflect the direct roles of 5MPs within the 48*S* PIC, we propose designating them as eIF1S (eukaryotic initiation factor 1 scanning; eIF1S1 for 5MP1 and eIF1S2 for 5MP2). This mechanistic and structural insight into 5MP-mediated proteoform regulation may lead to therapeutic developments targeting diseases driven by or dependent on dysregulated initiation.

MATERIALS AND METHODS

BSC-1 cell culture and cell lysate preparation

BSC-1 cells (American Type Culture Collection), a well-established epithelial cell line historically central to virology and membrane trafficking research, were cultured in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM GlutaMAX (Gibco), and 1% (v/v) of 100× penicillin-streptomycin (Gibco) (*62*). Cells were seeded in a 75-cm^2 flask at a density of ~20,000 cells per flask and cultured to confluence. After being washed twice with prewarmed phosphate-buffered saline, the cells were detached with trypsin-EDTA (Gibco) and collected by centrifugation at 300*g* for 4 min. The cell pellet was resuspended in 100 μl of semipermeabilization buffer [25 mM Hepes, 110 mM KOAc, 15 mM $Mg(OAc)_2$, 1 mM dithiothreitol (DTT), 0.015% digitonin, 2× protease inhibitor cocktail (Roche), 20 U of SUPERase-In RNase (ribonuclease) inhibitor (Thermo Fisher Scientific, AM2696), and 1 mM EGTA] at 4°C for 5 min. Following centrifugation at 1000*g* for 5 min, 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), 5MP2 (UniProt: Q7L1Q6), 5MP1 ΔNTT mutant (deletion of residues 1 to 25 and mutation of Pro^{26}Ala), 5MP1 RKR mutant [RKR(17-19)–to–AAA], 5MP1 KKEE mutant [KKEE(309-312)–to–EEQQ], and Y351A/I354A double mutant 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 *Escherichia coli* BL21(DE3), and cultures were grown in LB medium at 37°C to an optical density at 600 nm of 0.6, induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside, 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, and 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, and 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–polyacrylamide gel electrophoresis (SDS-PAGE), and peak fractions were pooled. The pooled Ni-NTA eluate was dialyzed overnight against a storage buffer [20 mM tris-HCl (pH 7.8), 125 mM NaCl, and 0.5 mM DTT], concentrated using 10-kDa Amicon filters and concentration estimated using ultraviolet (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 preequilibrated [20 mM tris-HCl (pH 7.8 and 8.0), 125 mM NaCl, and 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, and 0.5 mM DTT], and concentrated using 10-kDa Amicon filters, and concentration was 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, and 1 mM TCEP. For 5MP1, fractions corresponding to elution peaks 6 to 9 were collected; for 5MP2, fractions 3 to 6 were recovered. The final preparations showed a major peak migrating at ~48 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 to 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 table S3. 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 (*63*). 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).

Downloaded from https://www.science.org on November 03, 2026

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, were 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 table S3. All plasmids were sequenced to verify the sequences.

In vitro transcription of NanoLuciferase reporter mRNAs

DNA templates for in vitro transcription of mRNAs were polymerase chain reaction (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′) TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAATGAAAATAAATTTCCTTTATTAGCC.

PCR was performed using Phusion High-Fidelity DNA Polymerase (NEB, M0530L) according to the 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 adenosine 5′-triphosphate (ATP), guanosine 5′-triphosphate (GTP), cytidine 5′-triphosphate, and uridine 5′-triphosphate; and RNase Inhibitor (40 U/μl; 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*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*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 nuclease-free Milli-Q water. mRNA concentration was determined from UV absorbance at 260 nm using a NanoDrop One Spectrophotometer (Thermo Fisher 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 was 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 triplicate. The maximum translation rates were calculated from the maximum rate of luminescence change (max ΔRLU/s), 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/s) 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 rate at 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 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 prestained 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, and 16% (v/v) glycerol] and boiled at 98°C for 10 min. The total proteins in RRL samples were separated on 4 to 20% Mini-Protean TGX Precast Protein gels (Bio-Rad) alongside a protein marker (Bio-Rad) at 130 V for 2.5 hours in 1× tris-tricine-SDS running buffer, and proteins were transferred to a polyvinylidene difluoride membrane using a Trans-Blot Turbo transfer system (Bio-Rad).

To reduce nonspecific signal, the membranes were blocked in blocking buffer [1× PBST containing 0.1% Tween 20 and 5% (v/v) nonfat 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 immunoglobulin G1 (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

Downloaded from https://www.science.org on November 03, 2026

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 1:10,000 goat anti-rabbit IgG secondary antibody conjugated to horseradish peroxidase (Invitrogen) in blocking buffer at room temperature for 1 hour. After washing the membranes three times, 10 min each, the blots were detected using SuperSignal West Atto Ultimate Sensitivity Substrate (Thermo Fisher Scientific) 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 GMPPNP (*47*–*51*), 0.04 mM of 20 amino acids (Promega), and 20 U of SUPERase-In RNase inhibitor (Thermo Fisher Scientific, AM2696), maintained on ice until mRNA addition. Using GMPPNP in the RRL allowed us to capture ~3 times more initiation complexes than in the endogenous BSC-1 lysate. 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 (two to three blots, 3 to 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 to 1.5 μm. Automated data acquisition was performed in SerialEM (*64*, *65*) using beam-image shift to collect five movies per hole across nine holes at each stage position. Zero-loss peak refinement was performed every 90 min at a predetermined location to avoid darkened areas. The BSC-1 dataset consisted of 162,583 movies, and the 5MP1-RRL dataset consisted of 104,673 movies; each movie containing 30 frames with an exposure of ~1 e^- $Å^{-2}$ per frame (total exposure: ~30 e^- $Å^{-2}$). Beam-induced motion correction, gain reference application, exposure weighting, and 2×/4× binning (to pixel sizes of 0.83 and 1.66 Å, respectively) were performed using MotionCor2 (*66*). CTF parameters were estimated using CTFFIND5 (*67*) via the *cis*TEM GUI (*68*). For 2DTM, atomic coordinates for the 40*S* body from PDB 7R4X were converted into a 3D map using the simulate program in *cis*TEM (*69*), with a pixel size of 1.66 Å, 336 by 336 by 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 2000 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 reimported in the MT package, and particle stacks (box size, 560^3 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 (fig. S5). 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). Empiar ID-13853 and 13577 (both BSC-1 cell lysate) and Empiar ID 13852 (RRL).

Cryo-EM data classification

All classifications steps were performed in Frealign v.9.11 (*70*) on binned data (8×, 4×, and 2×), with particle stacks prepared using the program resample in the *cis*TEM package (*68*). Binning of the particle stacks was performed using the program resample from cisTEM. Initial unmasked 3D maximum-likelihood classification separated elongating ribosomes (80*S*) and junk from PICs (8× binned data). For BSC-1 cell lysates, PIC classes were merged using an in-house Python script, preserving micrograph paths and generating a .star file compatible with *cis*TEM (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 (*71*).

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 refinement, as described above. Focus-masked subclassification focused on 5MP, the head, eIF2, or mRNA/initiator $tRNA^{Met}$ 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 fig. S5A. 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. Maps are deposited in EMDB, EMD-78242 (48*S*•5MP1), EMD-78324 (48*S* closed head), and EMD-78325 (48*S*-like•5MP1).

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

Downloaded from https://www.science.org on November 03, 2026

unknown feature. ModelAngelo (*36*) implemented in RELION-5.0 (*72*) 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 and AF-Q7L1Q6) were fitted into the density to validate the identification (*37*, *73*).

Model building

Maps were sharpened using DeepEMhancer (*74*). The initial models were derived from the *Oryctolagus cuniculus* 80*S* ribosome (PDB 9BDL) (*75*), the *Homo sapiens* 48*S* complex with eIF4F/eIF4A (PDB 8OZ0) (*4*), and the *H. sapiens* 48*S* complex in the open scanning state (PDB 8PJ1) (*7*). The 18*S* 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 (*76*). The 5MP1 model was generated using AlphaFold2 prediction (AF-Q9Y6E2) and was rigid-body fitted into the density. Following initial placement, all components were refined using ISOLDE (*77*, *78*) 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 (table S2). The model map FSC was calculated using Phenix (*79*). The PDB identifier is 37KM.

Head rotation measurements

To compare head rotation between initiation complexes, we generated a closed-head model by rigid-body fitting the human 48*S* complex in the AUG recognition state (PDB 8PJ2) into our 2.8-Å map using ChimeraX (*76*), followed by a single round of real-space refinement in Phenix (*78*, *79*). For both the 5MP1•48S (open-head) and the refined closed-head model, we aligned the 18*S* body (nucleotides 1 to 1213 and 1691 to 1869) using Matchmaker in ChimeraX and saved the aligned coordinates. The 18*S* head (nucleotides 1214 to 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 48*S* open-head structure (PDB 8PJ1) was measured using the same procedure (*7*).

Multiple sequence alignment

Protein sequences used for comparative analysis were obtained from a previously published phylogenetic dataset (*22*) and supplemented with mammalian organisms with their accession numbers as follows: *H. 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); *Monosiga brevicollis* (UniParc: UPI000163B8DD); *Hydra magnipapillata* (UniParc: UPI0002B4BB35); *Capitella teleta* (UniParc: UPI00032C0488); *Aplysia californica* (UniParc: UPI0003593758); *Branchiostoma floridae* (UniParc: UPI0001861DC4); *Danio rerio* (UniParc: UPI000000BC41); *Xenopus tropicalis* (UniParc: UPI00004D064B); *Chrysemys picta bellii* (UniParc: UPI000388ED85); *Taeniopygia guttata* (UniParc: UPI000194BDA7); *Gallus gallus* (UniParc: UPI000044358E); *Melopsittacus undulates* (UniParc: UPI000383468D); *Amblyomma variegatum* (UniParc: UPI0002009EA1); *Caligus clemensi* (UniParc: UPI000198AD9A); *Lepeophtheirus salmonis* (UniParc: UPI000198A411); *Caligus rogercresseyi* (UniParc: UPI000198A05F); *Daphnia pulex* (UniParc: UPI0001FEB3EB); *Acyrthosiphon pisum* (UniParc: UPI0002061392); *Riptortus pedestris* (UniParc: UPI00032B4F83); *Pediculus humanus corporis* (UniParc: UPI000186E225); *Tribolium castaneum* (UniParc: UPI0000D56980); *Dendroctonus ponderosae* (UniParc: UPI00020FF929); *Bombyx mori* (UniParc: UPI0000E5D066); *Helicoverpa armigera* (UniParc: UPI00019203AC); *Danaus plexippus* (UniParc: UPI000239C638); *Nasonia vitripennis* (UniProt: UPI00015B411E); *Camponotus floridanus* (UniProt: UPI0001E7B62E); *Megachile rotundata* (UniParc: UPI000258DA7A); *Bombus terrestris* (UniParc: UPI00021A85D9); *Bombus impatiens* (UniParc: UPI00022CAAA0); *Apis mellifera* (UniParc: UPI0000DB7C8B); *Aedes aegypti* (UniParc: UPI0000D76E88); *Culex quinquefasciatus* (UniParc: UPI00016D8A73); *Ceratitis capitata* (UniParc: UPI00032984C4); *Drosophila willistoni* (UniParc: UPI00017D8E71); *Drosophila grimshawi* (UniParc: UPI00017C6DDE); *Drosophila mojavensis* (UniParc: UPI00017C7884); *Drosophila virilis* (UniParc: UPI00017D4931); *Drosophila melanogaster* (UniParc: UPI000000741C); *Ovis aries* (RefSeq: XDA75500.1); *Mus musculus* (RefSeq: XP_006515229.1); *O. cuniculus* (RefSeq: XP_051688519.2); *Pan troglodytes* (RefSeq: PNI66831.1); and *Manis pentadactyla* (RefSeq: KAI5178856.1).

All sequences were aligned using MUSCLE (*80*) with default parameters. The resulting alignment was manually inspected, and conserved residues and motifs were visualized using ESPript 3.2 (*81*, *82*).

Supplementary Materials

This PDF file includes:

Figs. S1 to S12
Tables S1 to S3

REFERENCES

1. A. Sendoel, J. G. Dunn, E. H. Rodriguez, S. Naik, N. C. Gomez, B. Hurwitz, J. Levorse, B. D. Dill, D. Schramek, H. Molina, J. S. Weissman, E. Fuchs, Translation from unconventional 5′ start sites drives tumour initiation. *Nature* **541**, 494–499 (2017).
2. L.-R. He, H.-Y. Zhao, B.-K. Li, Y.-H. Liu, M.-Z. Liu, X.-Y. Guan, X.-W. Bian, Y.-X. Zeng, D. Xie, Overexpression of eIF5A-2 is an adverse prognostic marker of survival in stage I non–small cell lung cancer patients. *Int. J. Cancer* **129**, 143–150 (2011).
3. J. Nakamura, S. Aoyagi, I. Nanchi, S.-I. Nakatsuka, E. Hirata, S. Shibata, M. Fukuda, Y. Yamamoto, I. Fukuda, N. Tatsumi, T. Ueda, F. Fujiki, M. Nomura, S. Nishida, T. Shirakata, N. Hosen, A. Tsuboi, Y. Oka, R. Nezu, M. Mori, Y. Doki, K. Aozasa, H. Sugiyama, Y. Oji, Overexpression of eukaryotic elongation factor eEF2 in gastrointestinal cancers and its involvement in G2/M progression in the cell cycle. *Int. J. Oncol.* **34**, 1181–1189 (2009).
4. J. Brito Querido, M. Sokabe, S. Kraatz, Y. Gordiyenko, J. M. Skehel, C. S. Fraser, V. Ramakrishnan, Structure of a human 48S translational initiation complex. *Science* **369**, 1220–1227 (2020).
5. Y. Hashem, A. des Georges, V. Dhote, R. Langlois, H. Y. Liao, R. A. Grassucci, C. U. T. Hellen, T. V. Pestova, J. Frank, Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell* **153**, 1108–1119 (2013).
6. Y. Hashem, A. des Georges, V. Dhote, R. Langlois, H. Y. Liao, R. A. Grassucci, T. V. Pestova, C. U. T. Hellen, J. Frank, Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature* **503**, 539–543 (2013).
7. V. Petrychenko, S.-H. Yi, D. Liedtke, B.-Z. Peng, M. V. Rodnina, N. Fischer, Structural basis for translational control by the human 48S initiation complex. *Nat. Struct. Mol. Biol.* **32**, 62–72 (2025).
8. R. W. Beerman, T. A. Jongens, A non-canonical start codon in the *Drosophila* fragile X gene yields two functional isoforms. *Neuroscience* **181**, 48–66 (2011).

Downloaded from https://www.science.org on November 03, 2026

9. N. T. Ingolia, L. F. Lareau, J. S. Weissman, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
10. J. A. Curran, K. A. J. Curran, M. A. Inchingolo, P. Jaquier-Gubler, The microprotein SEP53BP1: Its bizarre mode of translational expression and intracellular behaviour. bioRxiv 722586 [Preprint] (2026). https://doi.org/10.64898/2026.05.04.722586.
11. A. Wacholder, E. W. Deutsch, L. W. Kok, J. T. van Dinter, J. Lee, J. C. Wright, S. Leblanc, A. H. Jayatissa, K. Jiang, I. Arefiev, K. Cao, F. Bourassa, F.-A. Trifiro, M. Bassani-Sternberg, P. V. Baranov, A. Bogaert, S. Chothani, I. Fierro-Monti, D. Fijalkowska, K. Gevaert, N. Hubner, J. M. Mudge, J. Ruiz-Orera, J. Schulz, J. A. Vizcaíno, J. R. Prensner, M. A. Brunet, T. F. Martinez, S. A. Slavoff, X. Roucou, J. S. Choudhary, S. van Heesch, R. L. Moritz, A.-R. Carvunis, Detection of human unannotated microproteins by mass spectrometry-based proteomics: A community assessment. bioRxiv 639069 [Preprint] (2025). https://doi.org/10.1101/2025.02.19.639069.
12. S. Vagner, C. Touriol, B. Galy, S. Audigier, M. C. Gensac, F. Amalric, F. Bayard, H. Prats, A. C. Prats, 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).
13. P. Trulley, G. Snieckute, D. Bekker-Jensen, M. B. Menon, R. Freund, A. Kotlyarov, J. V. Olsen, M. D. Diaz-Muñoz, M. Turner, S. Bekker-Jensen, M. Gaestel, C. Tiedje, Alternative translation initiation generates a functionally distinct isoform of the stress-activated protein kinase MK2. *Cell Rep.* **27**, 2859–2870.e6 (2019).
14. P. J. Lee, Y. Sun, A. R. Soares, C. Fai, M. R. Picciotto, J. U. Guo, Alternative translation initiation produces synaptic organizer proteoforms with distinct localization and functions. *Mol. Cell* **84**, 3967–3978.e8 (2024).
15. S. R. Hann, M. W. King, D. L. Bentley, C. W. Anderson, R. N. Eisenman, 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).
16. I. P. Ivanov, A. E. Firth, A. M. Michel, J. F. Atkins, P. V. Baranov, Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* **39**, 4220–4234 (2011).
17. S.-M. Shen, C. Zhang, M.-K. Ge, S.-S. Dong, L. Xia, P. He, N. Zhang, Y. Ji, S. Yang, Y. Yu, J.-K. Zheng, J.-X. Yu, Q. Xia, G.-Q. Chen, PTENα and PTENβ promote carcinogenesis through WDR5 and H3K4 trimethylation. *Nat. Cell Biol.* **21**, 1436–1448 (2019).
18. R. Weber, U. Ghoshdastider, D. Spies, C. Duré, F. Valdivia-Francia, M. Forny, M. Ormiston, P. F. Renz, D. Taborsky, M. Yigit, M. Bernasconi, H. Yamahachi, A. Sendoel, Monitoring the 5′UTR landscape reveals isoform switches to drive translational efficiencies in cancer. *Oncogene* **42**, 638–650 (2023).
19. K. Sato, T. Masuda, Q. Hu, T. Tobo, S. Gillaspie, A. Niida, M. Thornton, Y. Kuroda, H. Eguchi, T. Nakagawa, K. Asano, K. Mimori, Novel oncogene 5MP1 reprograms c-Myc translation initiation to drive malignant phenotypes in colorectal cancer. *EBioMedicine* **44**, 387–402 (2019).
20. I. P. Ivanov, J. A. Saba, C.-M. Fan, J. Wang, A. E. Firth, C. Cao, R. Green, T. E. Dever, Evolutionarily conserved inhibitory uORFs sensitize Hox mRNA translation to start codon selection stringency. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117226119 (2022).
21. H. Hiraishi, J. Oatman, S. L. Haller, L. Blunk, B. McGivern, J. Morris, E. Papadopoulos, W. Gutierrez, M. Gordon, W. Bokhari, Y. Ikeda, D. Miles, J. Fellers, M. Asano, G. Wagner, L. Tazi, S. Rothenburg, S. J. Brown, K. Asano, Essential role of eIF5-mimic protein in animal development is linked to control of ATF4 expression. *Nucleic Acids Res.* **42**, 10321–10330 (2014).
22. C. R. Singh, R. Watanabe, D. Zhou, M. D. Jennings, A. Fukao, B. Lee, Y. Ikeda, J. A. Chiorini, S. G. Campbell, M. P. Ashe, T. Fujiwara, R. C. Wek, G. D. Pavitt, K. Asano, Mechanisms of translational regulation by a human eIF5-mimic protein. *Nucleic Acids Res.* **39**, 8314–8328 (2011).
23. C. R. Singh, M. R. Glineburg, C. Moore, N. Tani, R. Jaiswal, Y. Zou, E. Aube, S. Gillaspie, M. Thornton, A. Cecil, M. Hilgers, A. Takasu, I. Asano, M. Asano, C. R. Escalante, A. Nakamura, P. K. Todd, K. Asano, Human oncoprotein 5MP suppresses general and repeat-associated non-AUG translation via eIF3 by a common mechanism. *Cell Rep.* **36**, 109376 (2021).
24. G. Loughran, A. E. Firth, J. F. Atkins, I. P. Ivanov, Translational autoregulation of BZW1 and BZW2 expression by modulating the stringency of start codon selection. *PLOS ONE* **13**, e0192648 (2018).
25. L. Tang, J. Morris, J. Wan, C. Moore, Y. Fujita, S. Gillaspie, E. Aube, J. Nanda, M. Marques, M. Jangal, A. Anderson, C. Cox, H. Hiraishi, L. Dong, H. Saito, C. R. Singh, M. Witcher, I. Topisirovic, S. B. Qian, K. Asano, Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. *Nucleic Acids Res.* **45**, 11941–11953 (2017).
26. P. A. Wagner, M. Song, P. Doran, A. Şeker, R. Ficner, B. Kuhle, A. Marintchev, Molecular basis for the interactions of eIF2β with eIF5, eIF2B, and 5MP1 and their regulation by CK2. *RNA* **31**, 1419–1435 (2025).
27. S. Li, Z. Chai, Y. Li, D. Liu, Z. Bai, Y. Li, Y. Li, Z. Situ, BZW1, a novel proliferation regulator that promotes growth of salivary muocepodermoid carcinoma. *Cancer Lett.* **284**, 86–94 (2009).
28. S. Agarwal, F. Afaq, P. Bajpai, M. Behring, H. G. Kim, A. Varambally, D. S. Chandrashekar, S. Peter, S. A. Diffalha, M. Khushman, A. Seeber, S. Varambally, U. Manne, BZW2 inhibition reduces colorectal cancer growth and metastasis. *Mol. Cancer Res.* **21**, 698–712 (2023).
29. K. Jin, Y. Li, R. Wei, Y. Liu, S. Wang, H. Tian, BZW2 promotes malignant progression in lung adenocarcinoma through enhancing the ubiquitination and degradation of GSK3β. *Cell Death Discov.* **10**, 105 (2024).
30. C. V. Dang, MYC on the path to cancer. *Cell* **149**, 22–35 (2012).
31. S. Lee, M. Nahm, M. Lee, M. Kwon, E. Kim, A. D. Zadeh, H. Cao, H.-J. Kim, Z. H. Lee, S. B. Oh, J. Yim, P. A. Kolodziej, S. Lee, The F-actin-microtubule crosslinker Shot is a platform for Krasavietz-mediated translational regulation of midline axon repulsion. *Development* **134**, 1767–1777 (2007).
32. J. Zhang, S.-B. Pi, N. Zhang, J. Guo, W. Zheng, L. Leng, G. Lin, H.-Y. Fan, Translation regulatory factor BZW1 regulates preimplantation embryo development and compaction by restricting global non-AUG Initiation. *Nat. Commun.* **13**, 6621 (2022).
33. Z. Seraj, X. Zottig, C.-Y. Huang, A. B. Loveland, S. Diggs, E. Sholi, N. Grigorieff, A. A. Korostelev, In extracto cryo-EM reveals eEF2 as a major hibernation factor on 60S and 80S particles. *eLife* **15**, RP110114 (2026).
34. B. A. Lucas, B. A. Himes, L. Xue, T. Grant, J. Mahamid, N. Grigorieff, Locating macromolecular assemblies in cells by 2D template matching with cisTEM. *eLife* **10**, e68946 (2021).
35. J. P. Rickgauer, N. Grigorieff, W. Denk, Single-protein detection in crowded molecular environments in cryo-EM images. *eLife* **6**, e25648 (2017).
36. K. Jamali, L. Käll, R. Zhang, A. Brown, D. Kimanius, S. H. W. Scheres, Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
37. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
38. M. Kozak, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
39. M. Kozak, Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol. Cell. Biol.* **9**, 5073–5080 (1989).
40. N. H. Cho, K. C. Cheveralls, A.-D. Brunner, K. Kim, A. C. Michaelis, P. Raghavan, H. Kobayashi, L. Savy, J. Y. Li, H. Canaj, J. Y. S. Kim, E. M. Stewart, C. Gnann, F. M. Carthy, J. P. Cabrera, R. M. Brunetti, B. B. Chhun, G. Dingle, M. Y. Hein, B. Huang, S. B. Mehta, J. S. Weissman, R. Gómez-Sjöberg, D. N. Itzhak, L. A. Royer, M. Mann, M. D. Leonetti, OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**, eabi6983 (2022).
41. Y.-N. Cheung, D. Maag, S. F. Mitchell, C. A. Fekete, M. A. Algire, J. E. Takacs, N. Shirokikh, T. Pestova, J. R. Lorsch, A. G. Hinnebusch, Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. *Genes Dev.* **21**, 1217–1230 (2007).
42. A. K. Saini, J. S. Nanda, P. Martin-Marcos, J. Dong, F. Zhang, M. Bhardwaj, J. R. Lorsch, A. G. Hinnebusch, 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 Res.* **42**, 9623–9640 (2014).
43. J. L. Llácer, T. Hussain, A. K. Saini, J. S. Nanda, S. Kaur, Y. Gordiyenko, R. Kumar, A. G. Hinnebusch, J. R. Lorsch, V. Ramakrishnan, Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *eLife* **7**, e39273 (2018).
44. S. E. Kolitz, J. E. Takacs, J. R. Lorsch, Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA* **15**, 138–152 (2009).
45. P. Martin-Marcos, J. Nanda, R. E. Luna, G. Wagner, J. R. Lorsch, A. G. Hinnebusch, 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).
46. A. Thakur, A. G. Hinnebusch, 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).
47. A. Simonetti, J. Brito Querido, A. G. Myasnikov, E. Mancera-Martinez, A. Renaud, L. Kuhn, Y. Hashem, eIF3 peripheral subunits rearrangement after mRNA binding and start-codon recognition. *Mol. Cell* **63**, 206–217 (2016).
48. A. Simonetti, E. Guca, A. Bochler, L. Kuhn, Y. Hashem, Structural insights into the mammalian late-stage initiation complexes. *Cell Rep.* **31**, 107497 (2020).
49. A. Bochler, J. B. Querido, T. Prilepskaja, H. Soufari, A. Simonetti, M. L. del Cistia, L. Kuhn, A. R. Ribeiro, L. S. Valášek, Y. Hashem, Structural differences in translation initiation between pathogenic trypanosomatids and their mammalian hosts. *Cell Rep.* **33**, 108534 (2020).
50. E. Guca, R. Alarcon, M. Z. Palo, L. Santos, S. Alonso-Gil, M. Davyt, L. H. F. de Lima, F. Boissier, S. Das, B. Zagrovic, J. D. Puglisi, Y. Hashem, Z. Ignatova, N^6-methyladenosine in 5′ UTR does not promote translation initiation. *Mol. Cell* **84**, 584–595.e6 (2024).

Downloaded from https://www.science.org on November 03, 2026

51. L. Jia, T.-T. Nguyen, S. Uematsu, Y. Gu, S. Shi, Y. Hashem, S.-B. Qian, Programmable initiation of mRNA translation by trans-RNA. *Nat. Biotechnol.*, 10.1038/s41587-025-02897-1 (2025).
52. B. A. Lucas, K. Zhang, S. Loerch, N. Grigorieff, In situ single particle classification reveals distinct 60S maturation intermediates in cells. *eLife* **11**, e79272 (2022).
53. S. Anand, A. Marintchev, NMR assignments and secondary structure analysis of the eIF1-interacting fragment of human eIF3c. bioRxiv 675972 [Preprint] (2025). https://doi.org/10.1101/2025.09.13.675972.
54. L. A. Passmore, T. M. Schmeing, D. Maag, D. J. Applefield, M. G. Acker, M. A. Algire, J. R. Lorsch, V. Ramakrishnan, The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell* **26**, 41–50 (2007).
55. S.-H. Yi, V. Petrychenko, J. E. Schliep, A. Goyal, A. Linden, A. Chari, H. Urlaub, H. Stark, M. V. Rodnina, S. Adio, N. Fischer, Conformational rearrangements upon start codon recognition in human 48S translation initiation complex. *Nucleic Acids Res.* **50**, 5282–5298 (2022).
56. A. Jobe, Z. Liu, C. Gutierrez-Vargas, J. Frank, New insights into ribosome structure and function. *Cold Spring Harb. Perspect. Biol.* **11**, a032615 (2019).
57. J. L. Llácer, T. Hussain, L. Marler, C. E. Aitken, A. Thakur, J. R. Lorsch, A. G. Hinnebusch, V. Ramakrishnan, Conformational differences between open and closed states of the eukaryotic translation initiation complex. *Mol. Cell* **59**, 399–412 (2015).
58. T. Hussain, J. L. Llácer, I. S. Fernández, A. Munoz, P. Martin-Marcos, C. G. Savva, J. R. Lorsch, A. G. Hinnebusch, V. Ramakrishnan, Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* **159**, 597–607 (2014).
59. K. Asano, T. Krishnamoorthy, L. Phan, G. D. Pavitt, A. G. Hinnebusch, Conserved bipartite motifs in yeast eIF5 and eIF2Bε, GTPase-activating and GDP–GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2. *EMBO J.* **18**, 1673–1688 (1999).
60. B. S. de la Fuente, L. Villamayor-Belinchón, J. R. Ciges-Tomas, M. López-Redondo, C. Espinosa, E. Boeri Erba, J. L. Llácer, Novel insights into the assembly of the yeast translation initiation multifactor complex: The critical role of eIF5. *Protein Sci.* **35**, e70550 (2026).
61. D. Maag, C. A. Fekete, Z. Gryczynski, J. R. Lorsch, A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol. Cell* **17**, 265–275 (2005).
62. E. N. Salgado, B. G. Rodriguez, N. Narayanaswamy, Y. Krishnan, S. C. Harrison, Visualization of calcium ion loss from rotavirus during cell entry. *J. Virol.* **92**, e01327 (2018).
63. D. Susorov, S. Egri, A. A. Korostelev, Termi-Luc: A versatile assay to monitor full-protein release from ribosomes. *RNA* **26**, 2044–2050 (2020).
64. D. N. Mastronarde, SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183 (2003).
65. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
66. S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
67. J. Elferich, L. Kong, X. Zottig, N. Grigorieff, CTFFIND5 provides improved insight into quality, tilt, and thickness of TEM samples. *eLife* **13**, RP97227 (2024).
68. T. Grant, A. Rohou, N. Grigorieff, cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
69. B. Himes, N. Grigorieff, Cryo-TEM simulations of amorphous radiation-sensitive samples using multislice wave propagation. *IUCrJ* **8** (Pt. 6), 943–953 (2021).
70. D. Lyumkis, A. F. Brilot, D. L. Theobald, N. Grigorieff, Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388 (2013).
71. S. H. W. Scheres, Amyloid structure determination in RELION-3.1. *Acta Crystallogr. D Struct. Biol.* **76** (Pt. 2), 94–101 (2020).
72. A. Burt, B. Toader, R. Warshamanage, A. von Kügelgen, E. Pyle, J. Zivanov, D. Kimanius, T. A. M. Bharat, S. H. W. Scheres, An image processing pipeline for electron cryo-tomography in RELION-5. *FEBS Open Bio* **14**, 1788–1804 (2024).
73. H. Liu, A. Laiho, P. Törönen, L. Holm, 3-D substructure search by transitive closure in AlphaFold database. *Protein Sci.* **34**, e70169 (2025).
74. R. Sanchez-Garcia, J. Gomez-Blanco, A. Cuervo, J. M. Carazo, C. O. S. Sorzano, J. Vargas, DeepEMhancer: A deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
75. A. B. Loveland, C. S. Koh, R. Ganesan, A. Jacobson, A. A. Korostelev, Structural mechanism of angiogenin activation by the ribosome. *Nature* **630**, 769–776 (2024).
76. E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
77. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* **74** (Pt. 6), 519–530 (2018).
78. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74** (Pt. 6), 531–544 (2018).
79. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, N. Echols, J. J. Headd, L.-W. Hung, S. Jain, G. J. Kapral, R. W. Grosse Kunstleve, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
80. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
81. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
82. P. Gouet, X. Robert, E. Courcelle, ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).

Acknowledgments: We thank C. Xu, K. Song, and C. Ouch for support with cryo-EM data collection at UMass Chan Medical School; J. Elferich, D. Zhu, and S. Diggs for valuable scientific input; and members of the Grigorieff and Korostelev laboratories for constructive feedback on this study. We also thank A. Marintev and T. Hainthaler for helpful discussions on datasets. **Funding:** N.G. is an investigator of the Howard Hughes Medical Institute. This study was supported by the US National Institutes of Health R35GM127094 and U54AI170856 (CHEETAH) awarded 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. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. Plasmids generated in this study are available upon request from the corresponding authors N.G. and A.A.K. at niko@grigorieff.org and andrei.korostelev@umassmed.edu, respectively.

Submitted 18 February 2026
Accepted 16 September 2026
Published 23 October 2026
10.1126/sciadv.aeg5840

Downloaded from https://www.science.org on November 03, 2026

Science Advances
AAAS

Supplementary Materials for

Structural basis for start codon selection by eIF5-mimic proteins

Ximena Zottig *et al.*

Corresponding author: Nikolaus Grigorieff, niko@grigorieff.org;
Andrei A. Korostelev, andrei.korostelev@umassmed.edu

Sci. Adv. **12**, eaeg5840 (2026)
DOI: 10.1126/sciadv.aeg5840

This PDF file includes:

Figs. S1 to S12
Tables S1 to S3

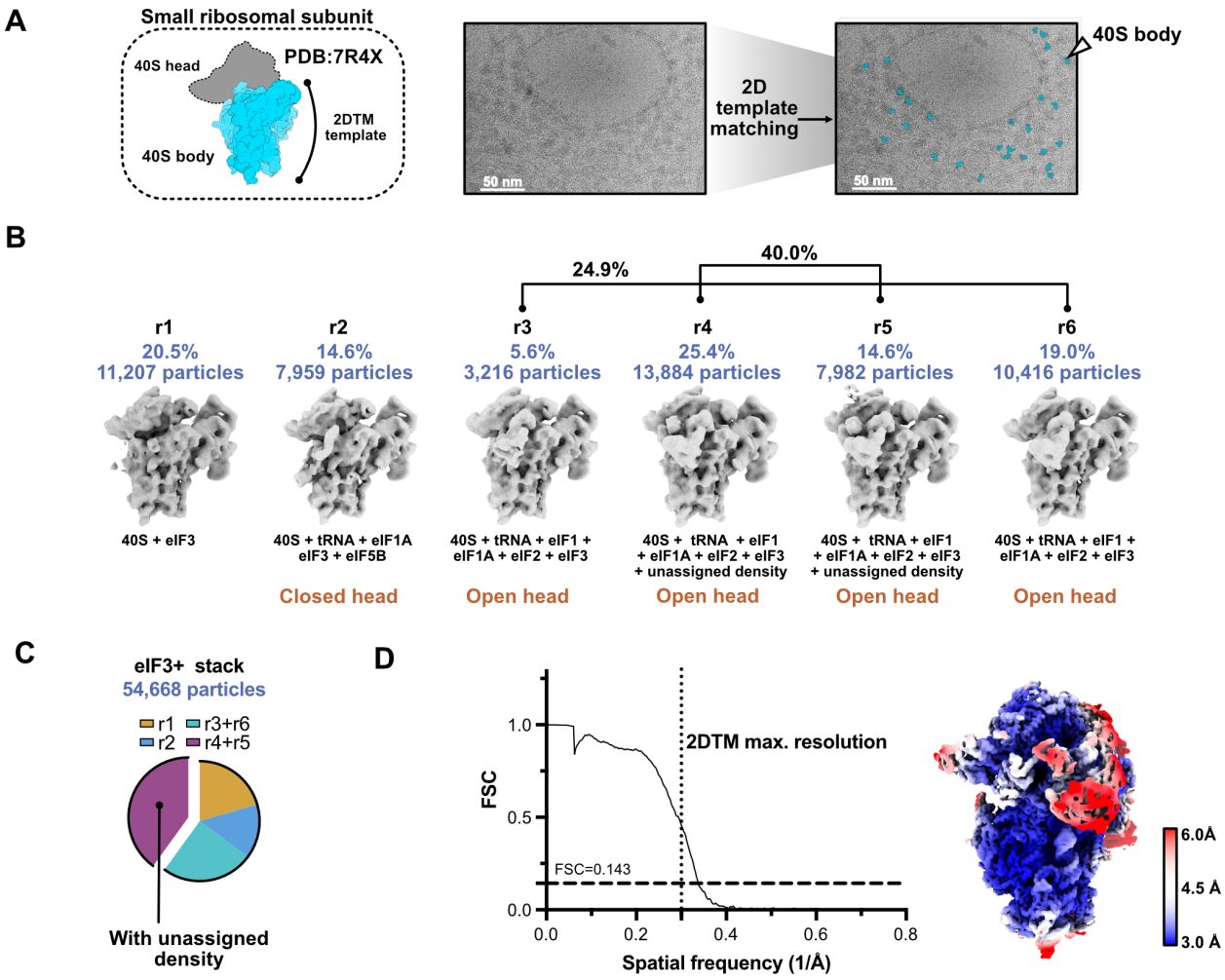


Fig. S1. Identification of translation pre-initiation complexes in BSC-1 lysates.
(A) Schematic representation of the template used for 2D template matching (2DTM), representative micrograph of BSC-1 cellular lysate and an illustration of 2DTM detections (blue). Template matching was carried out using 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. **(C)** Pre-initiation complexes 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. **(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.

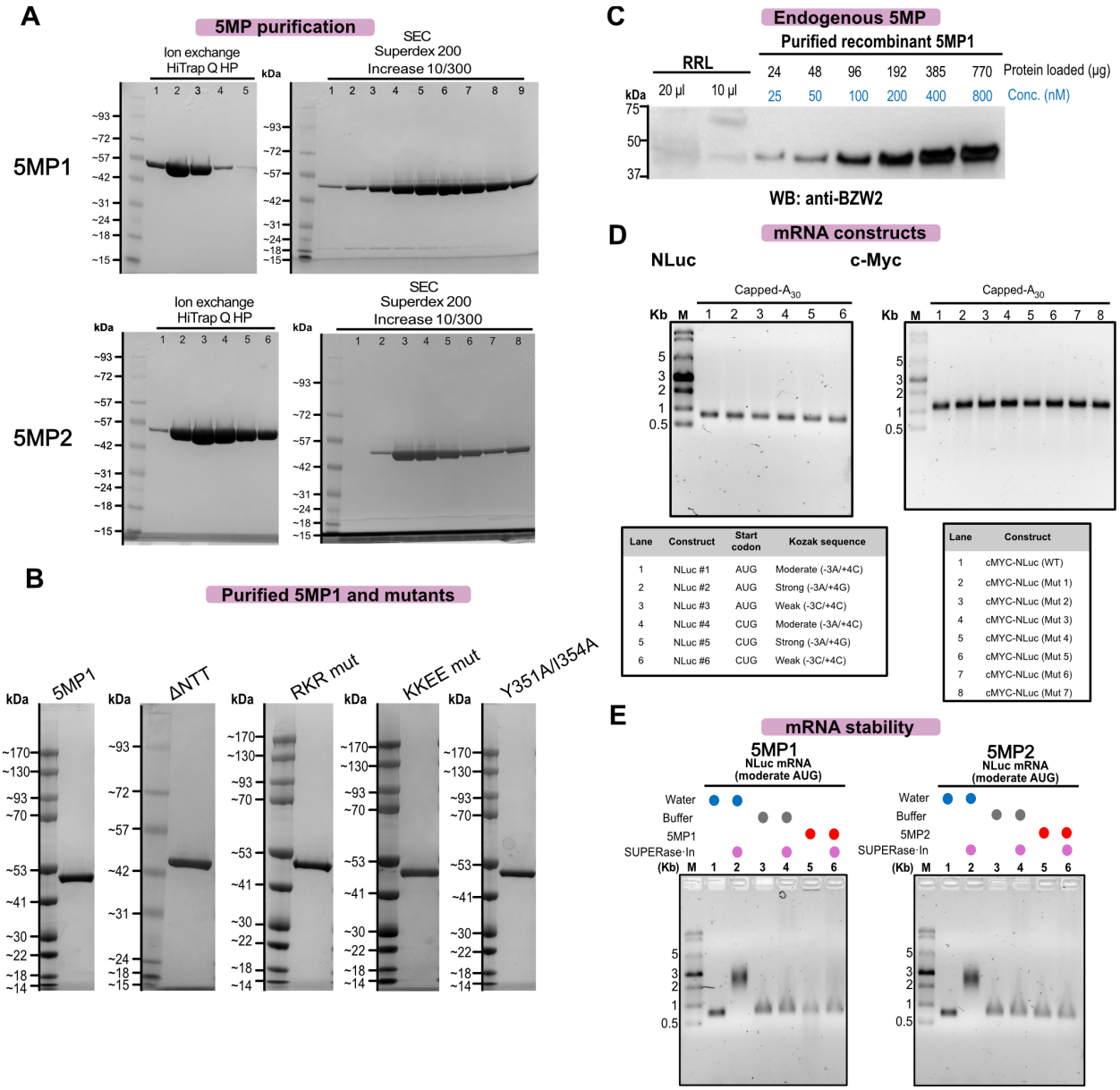
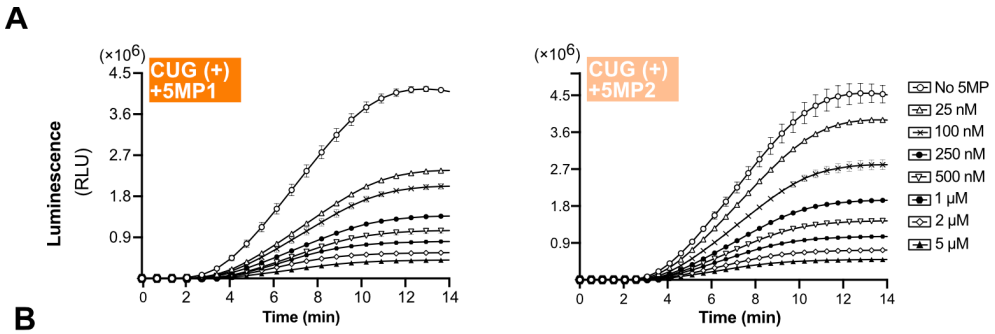


Fig. S2. Preparation and validation of recombinant 5MP proteins and reporter mRNAs. **(A)** Recombinant eIF5-mimic proteins 5MP1 (top) and 5MP2 (bottom) were purified by His-tag affinity chromatography, followed by ion-exchange separation on a HiTrap Q HP column, and a final purification step by size-exclusion 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)** SDS-PAGE of purified 5MP1 and mutants used for functional assays (2 μg per well): the NTT-deletion residues 1-25 mutant (ΔNTT), the RKR(17-19)-to-AAA mutant (RKR mut), the KKEE(309-312)-to-EEQQ mutant (KKEE mut), and the Y351A/I354A double mutant. **(C)** 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. **(D)** Capped mRNA constructs with A_{30} tails were produced by *in vitro* transcription for translation assays and cryo-EM. **(E)** 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).



B

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 µ M	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 µ M	764±87	410±31	131±10	79±3	82±6

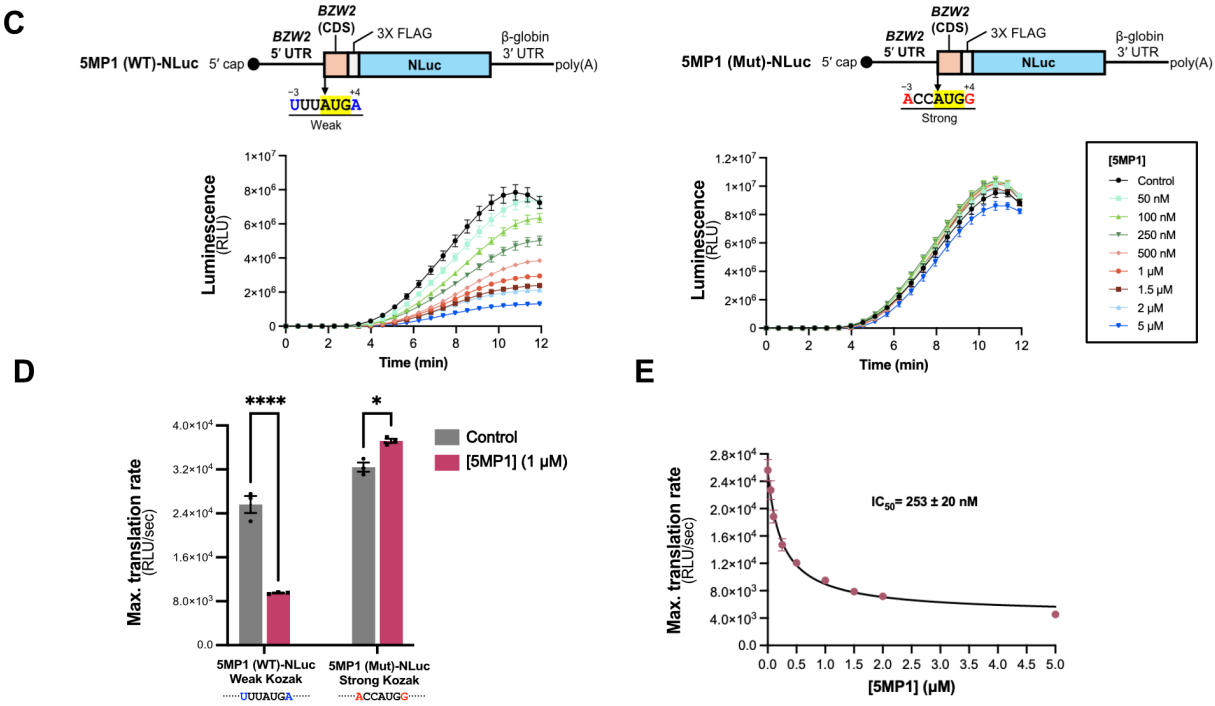


Fig. S3. 5MP-Mediated Translation Inhibition and Self-Regulation.
(A) Time progress curves for strong Kozak CUG initiation at increasing 5MP concentrations (25 nM, 100 nM, 250 nM, 500 nM, 1µM, 2 µM and 5 µM) **(B)** IC_{50} values for translation inhibition of NLuc constructs with different start codons by 5MP1 and 5MP2 in rabbit reticulocyte lysate (RRL); see additional data in main-text Fig. 1. **(C)** 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. **(D)** 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. **(E)** Translation inhibition curve showing the 5MP1 inhibitory concentration (IC_{50}) for the mRNA carrying the WT-5MP1 start codon in RRL. Data represent mean ± SEM (n=3) (C) and mean ± SD (n=3) (D, E). Statistical significance on maximum translation rates was assessed by two-way ANOVA; ****$P < 0.0001$, ***$P < 0.0002$, **$P < 0.0021$, ns = not significant.

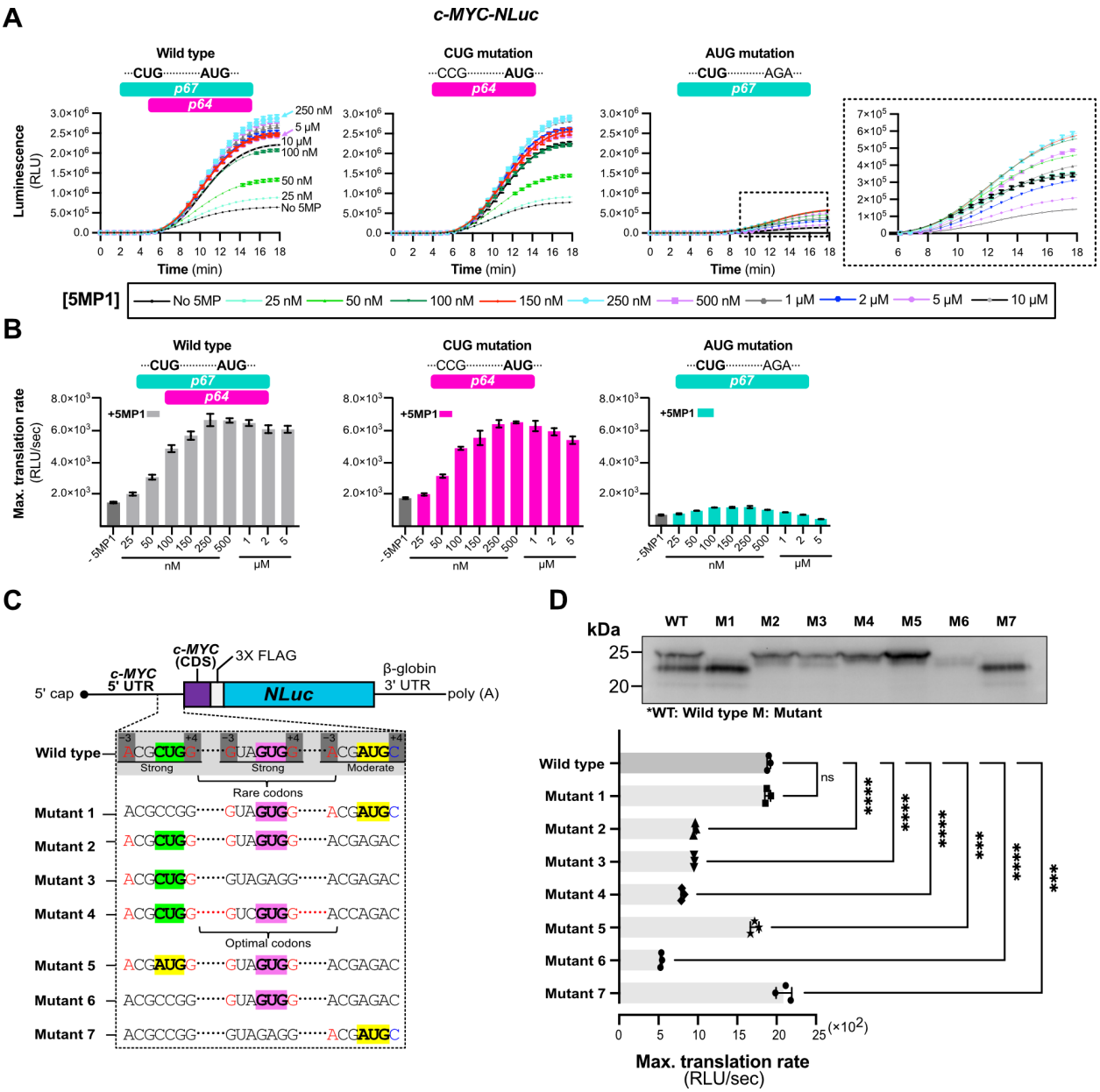


Fig. S4. 5MP1 regulation of translation from *the c-Myc* 5′ UTR and inefficient upstream CUG initiation

(A) Time course analysis of in vitro RRL translation of the c-Myc-NLuc wild type construct, the CUG-mutant (Mut 1), and the AUG-mutant (Mut 2). (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)** Schematic of the *c-Myc* 5′ UTR reporter showing the in-frame initiation sites, Kozak contexts, and mutants. **(D)** Translation products detected by anti-FLAG western blotting (top) and maximum

translation rates for *c-Myc* wild-type 5′ UTR and the mutants (bottom). Data represent mean ± SEM (n=3) (A, B) and mean ± SD (n=3) (D). Statistical significance of maximum translation rates was assessed by one-way ANOVA; ****P < 0.0001, ***P < 0.0002, **P < 0.0021; ns, not significant.

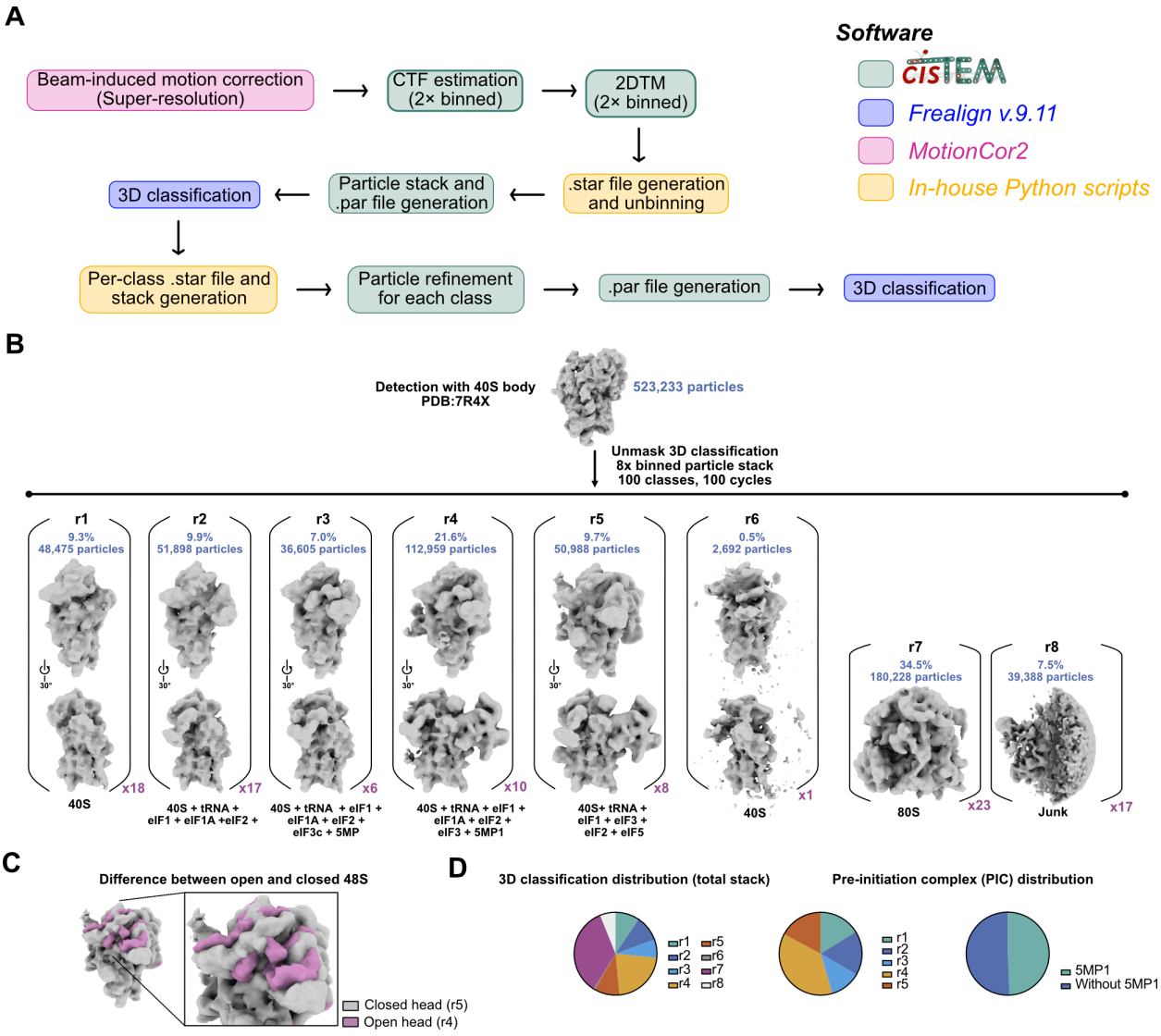


Fig. S5. Cryo-EM image processing and classification workflow for RRL supplemented with recombinant 5MP1.

(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.

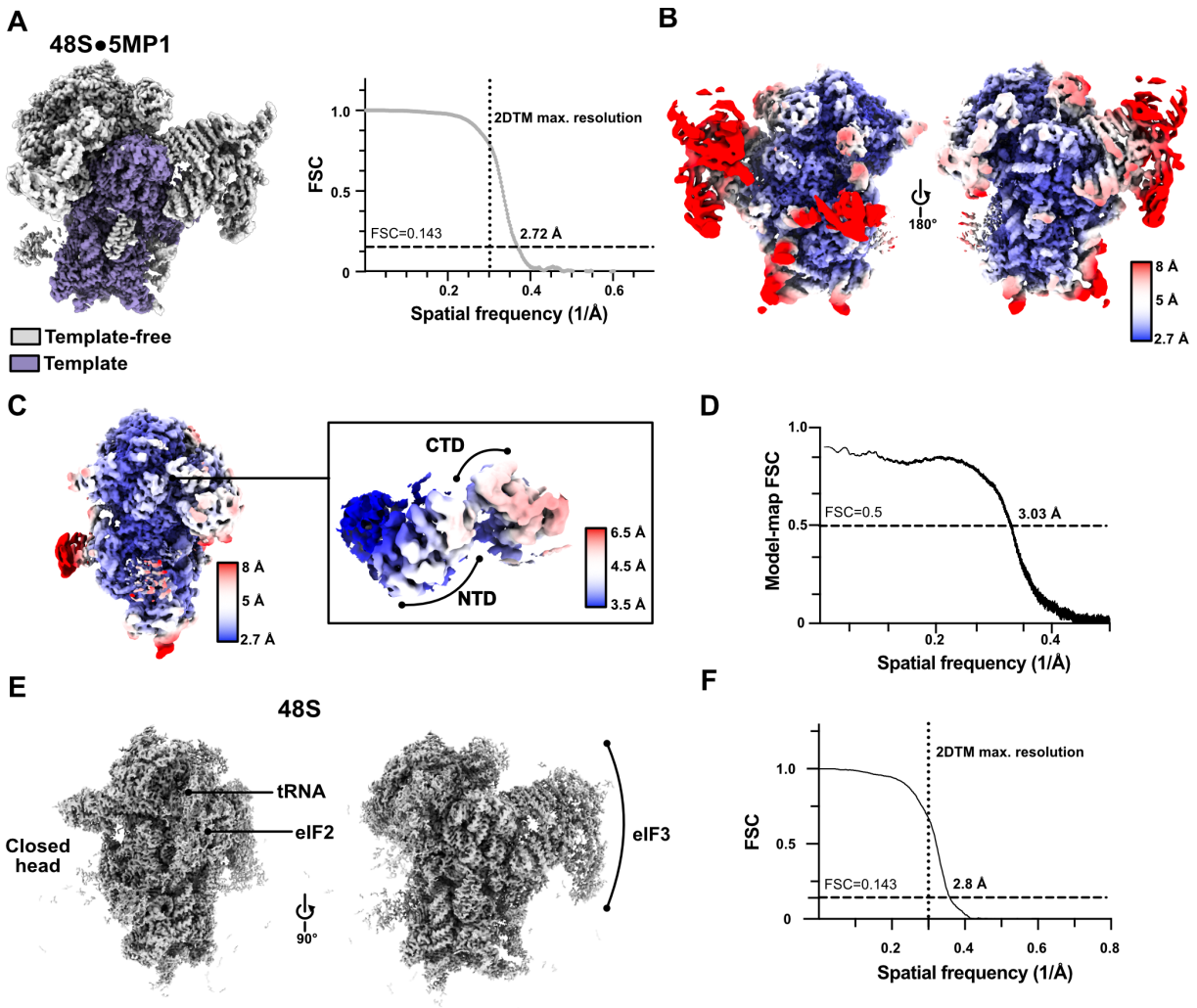


Fig. S6. Assessment of average resolution, template bias and local resolution of the 48S•5MP1 and 48S pre-initiation complex.

(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.

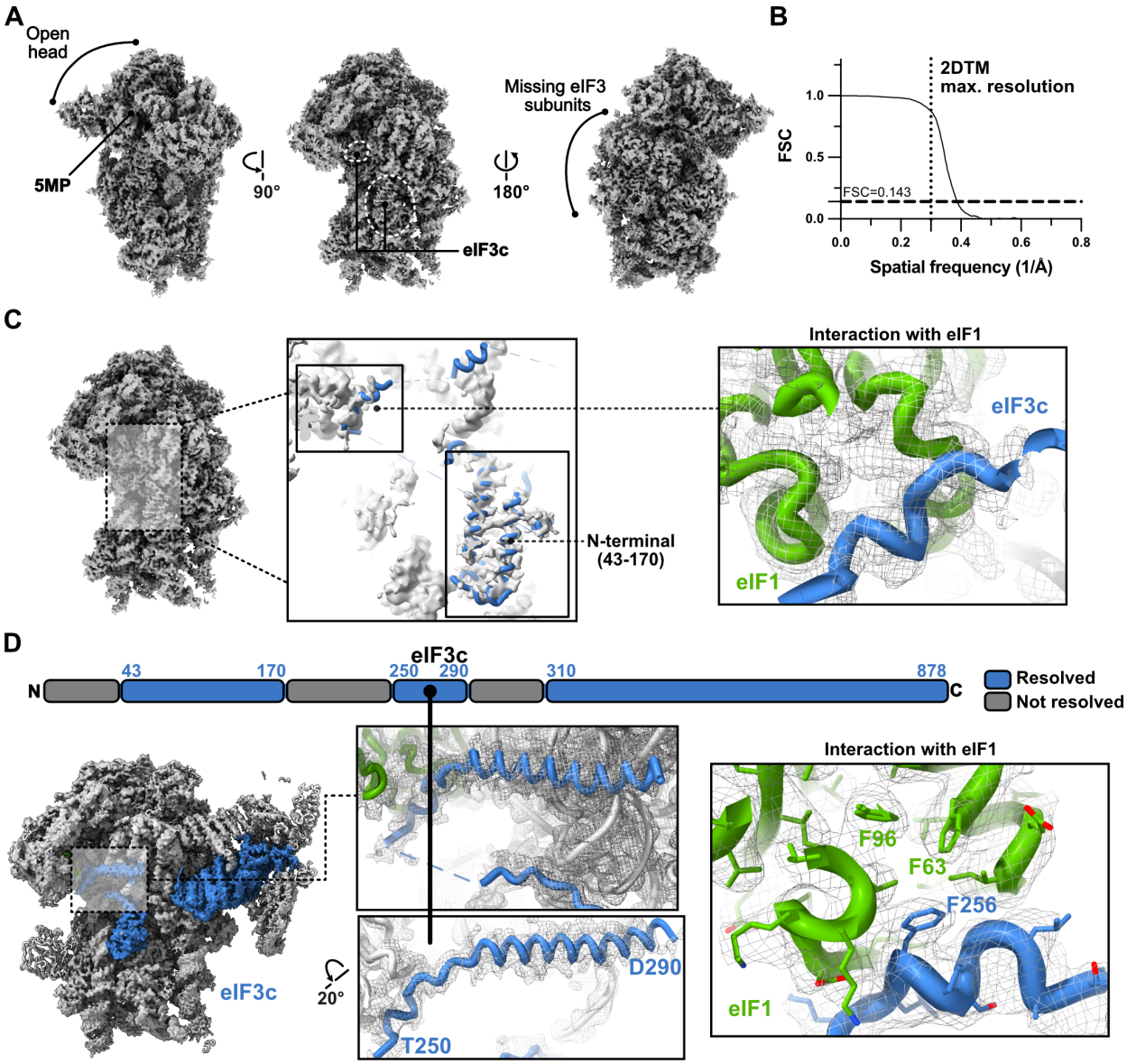


Fig. S7. The N-terminal domain of eIF3c contacts eIF1.
(A) 2.8 Å map of 5MP1 bound to a 48S-like pre-initiation complex showing densities consistent with the eIF3c N-terminal region. Density is observed for residues 43-260 of eIF3c, whereas no density is detected for the remaining eIF3 subunits. **(B)** Fourier shell correlation (FSC) curve for the reconstruction shown in panel A. The 2DTM resolution limit is indicated by a dotted line. **(C)** Densities corresponding to the eIF3c helix interacting with eIF1 (also see box panel D for reference). **(D)** A scheme and densities in the 48S complex with eIF3, showing the resolved segments of eIF3c in blue and eIF1 model in green. The interaction supports a role of the eIF3c N-terminal domain in 5MP1-mediated regulation of initiation stringency *(50)*. Residues T250–D290 (middle) were initially modeled using AlphaFold2 DB and refined using Isolde and Phenix. The structure revealed that the Phe256 of the FLKK motif 3 of eIF3c *(50)* is inserted into a hydrophobic pocket on the surface of eIF1 (right).

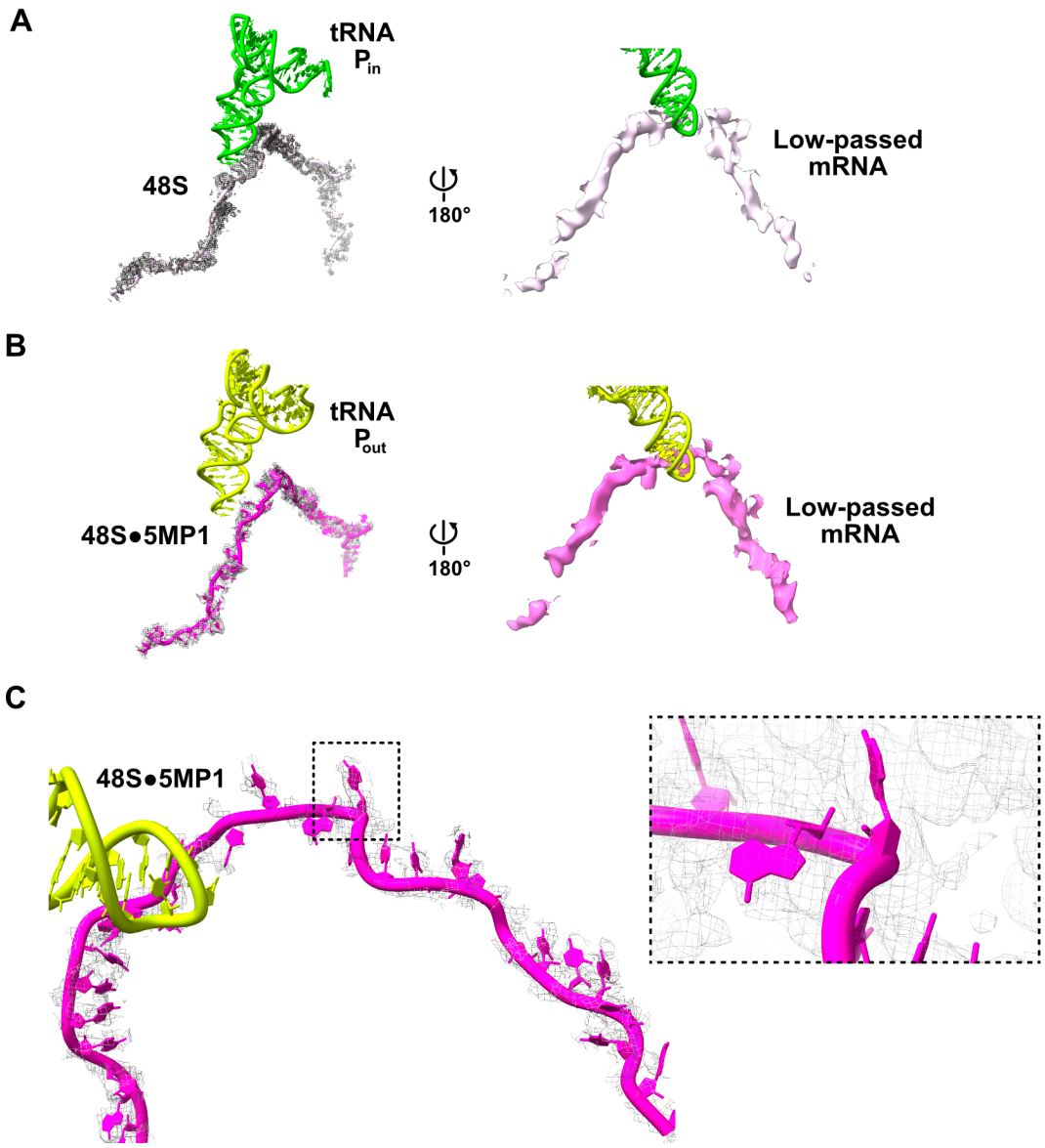


Fig. S8. Scanning 48S•5MP1 initiation complexes exhibits continuous density for the mRNA.

(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 a 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)

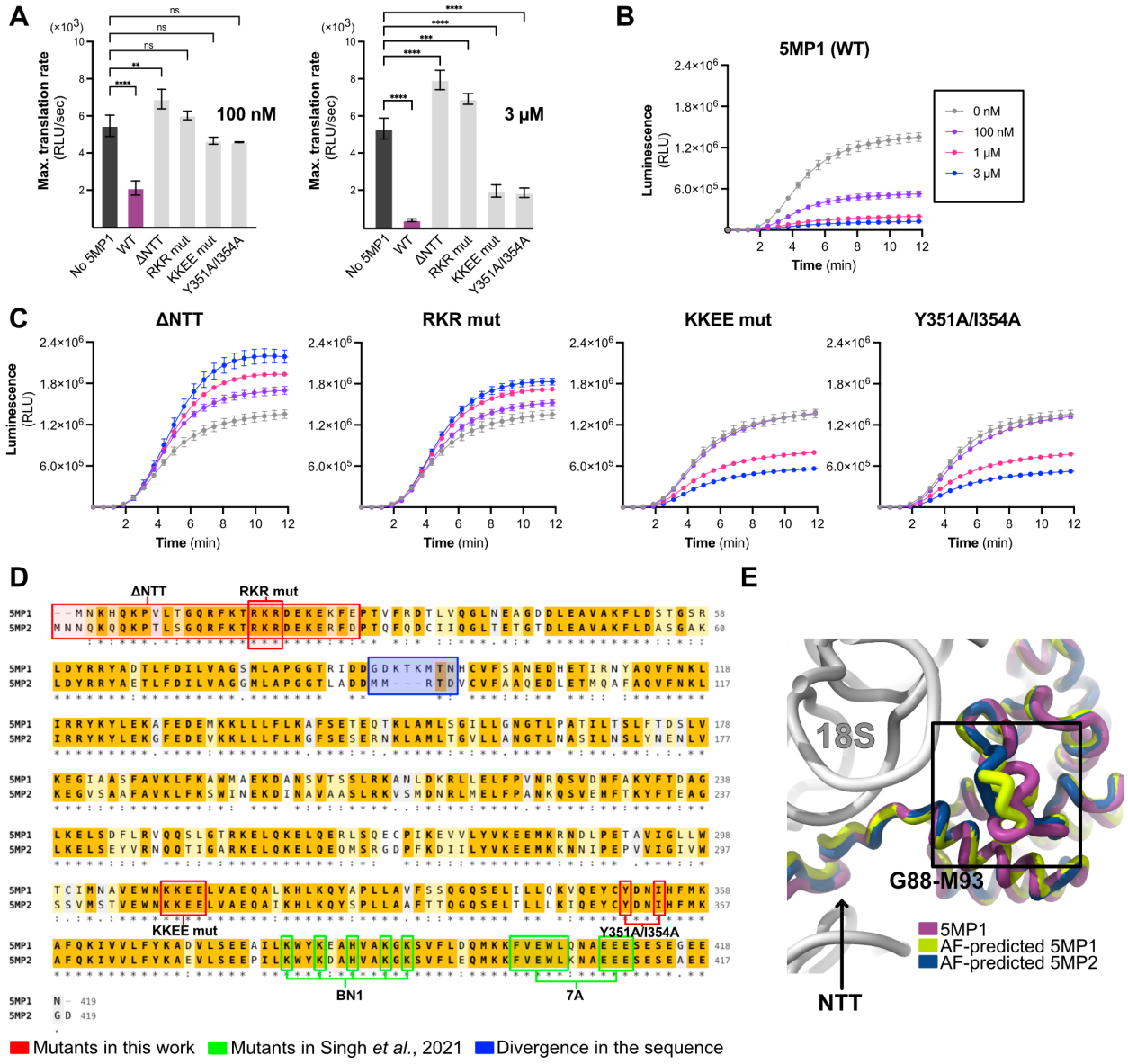


Fig. S9. Effects of structure-guided mutations of 5MP1 on translation.

(A) Maximal rates of translation of the transcript containing a CUG start codon in a strong Kozak context (see construct in Fig 1B), at 100 nM and 3 µM of 5MP1 WT or mutant proteins (mean ± SD, n=3). The rates are calculated from data in panels B and C. **(B)** Time progress curves for wild-type 5MP1 and **(C)** 5MP1 mutants (from left to right): the NTT-deletion residues 1-25 mutant (ΔNTT), the RKR(17-19)-to-AAA mutant (RKR mut), the KKEE(309-312)-to-EEQQ mutant (KKEE mut), and the Y351A/I354A double mutant, at 100 nM, 1 µM, and 3 µM (mean ± SEM, n=3). Statistical significance was assessed by one-way ANOVA; ****P < 0.0001, ***P < 0.0002, **P < 0.0021, ns = not significant. **(D)** Sequence alignment of 5MP1/BZW2 (UniProt: Q9Y6E2) and 5MP2/BZW1 (UniProt: Q7L1Q6) performed using Clustal Omega. Mutations generated in this study are boxed in red, mutations from Singh et al., 2021 are boxed in green, and the most divergent sequence region is boxed in blue (Gly88–Asn95). **(E)** Close-up view of the divergent sequence mapped onto the structure (with the divergent Gly88-Met93 region clearly visible). This

region is located in the N-terminal domain (NTD). Our 5MP1 model is shown in purple, the AlphaFold prediction for 5MP1 (Q9Y6E2) in yellow-green, and for 5MP2 (Q7L1Q6) in navy blue.

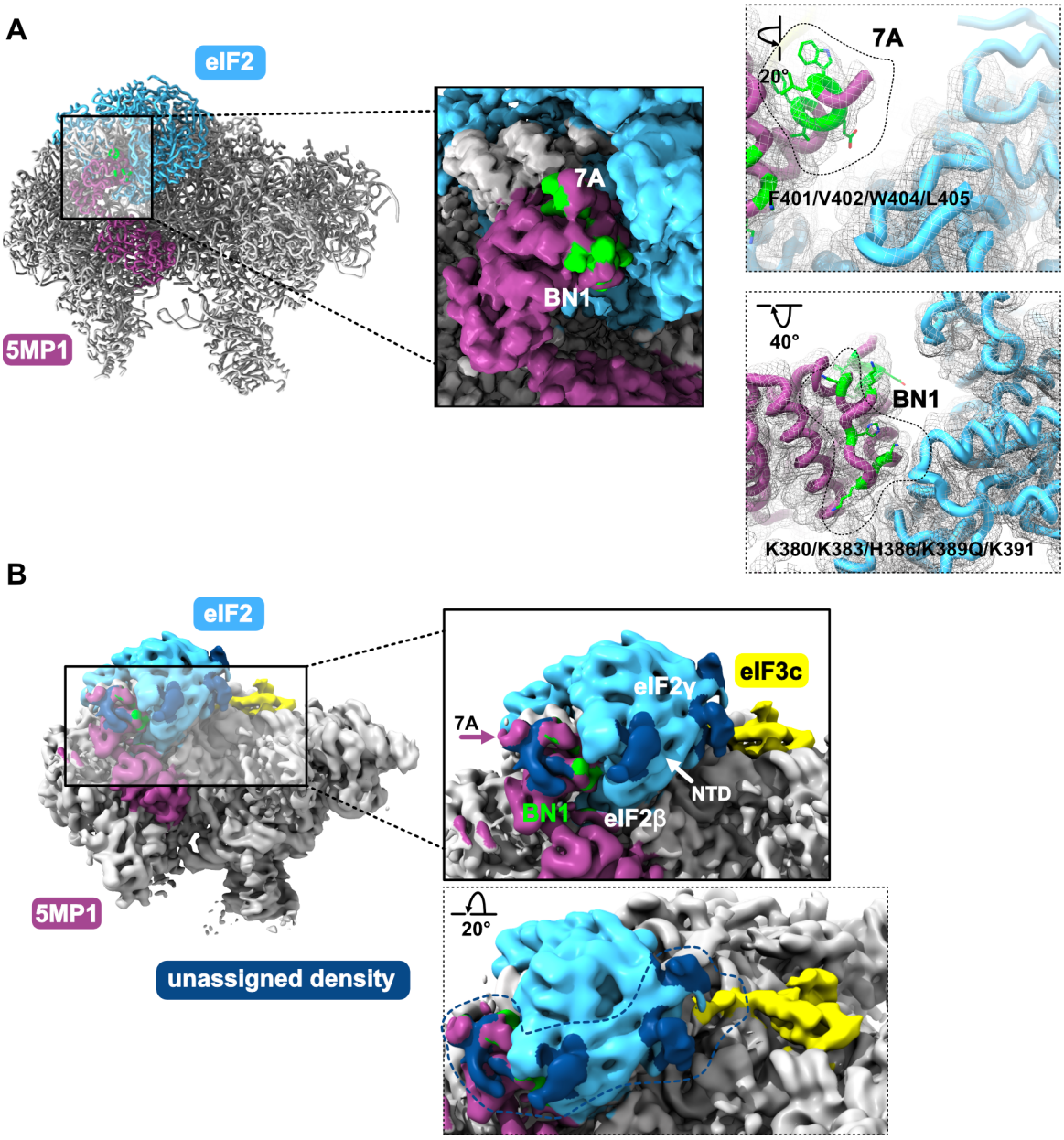


Fig. S10. Identification of 7A and BN1 mutations in 5MPs and of unassigned densities that potentially interact with the amino acids altered in these mutants.
(A) Identification of the mutations by Asano's group (Singh et al., 2021), BN1 and 7A, mapped onto our model and cryo-EM density map. Both mutation clusters (green) are located at the interface with eIF2β **(B).** Unassigned densities (dark blue) in an 8 Å low-passed filtered map displayed at a low contour level. These densities could be attributed to eIF2β N-terminal tail (densities near 5MP1 and eIF2β) or to eIF3c N-terminal (all densities shown in dark blue). Noise can not be excluded as an explanation for these densities. 5MP1 is shown in purple, eIF2 in light blue and eIF3c in yellow. 7A and BN1 mutations are shown in green. NTD stand for N-terminal domain, here identified for eIF2β.

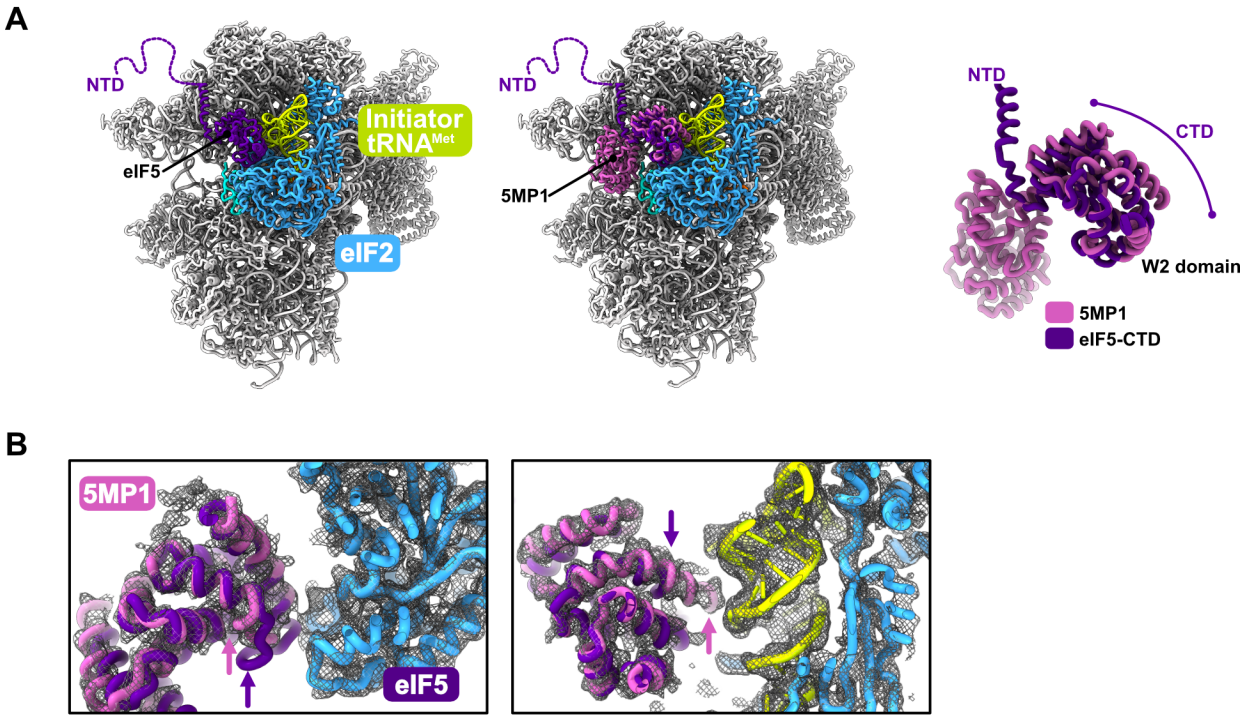


Fig. S11. Comparison of the CTDs from 5MP1 and from a predicted model of eIF5. **(A)** Superposition of the AlphaFold2-predicted structure of the eIF5 C-terminal domain (CTD; deep purple) onto the 48S•5MP1 cryo-EM structure (this work), aligned on 5MP1 CTD. Modeled densities are shown for 5MP1 in purple, eIF2 in blue, eIF1 in orange, eIF1A in turquoise, and initiator $tRNA^{Met}$ in green. The overlay suggests that the eIF5 CTD (left panel) could occupy a similar position within the 48S as the 5MP1 CTD (middle and right panels), consistent with potential competition between eIF5 and 5MP1 for binding at this site. **(B)** Close-up view of the eIF5 CTD and 5MP1 models, along with the 48S•5MP1 cryo-EM densities for eIF2β and initiator $tRNA^{Met}$, suggesting distinct molecular contacts.

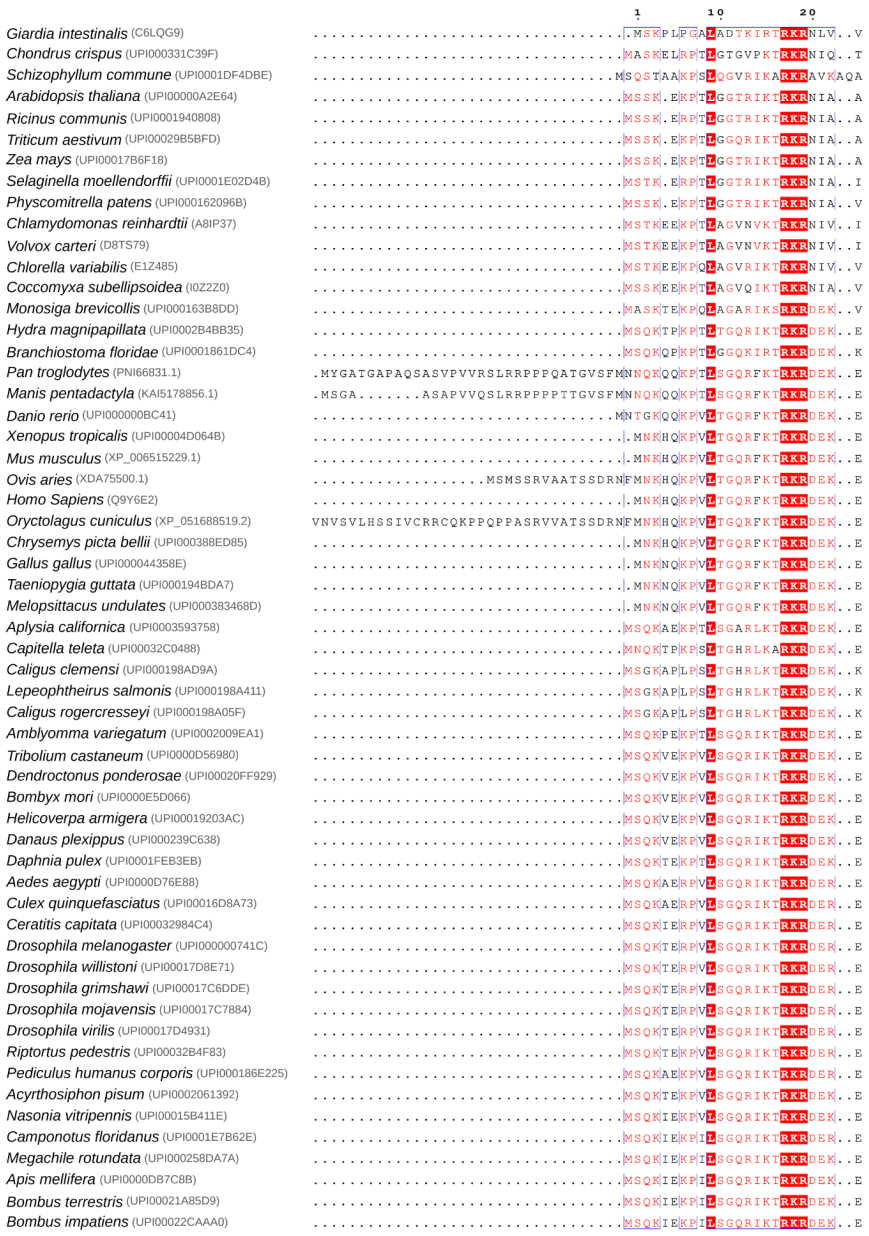


Fig. S12. Conservation of the 5MP1 N-terminal tail and RKR motif across eukaryotes.
Multiple sequence alignment of the 5MP1 N-terminal region (residues 1–20; human 5MP1 numbering, UniProt Q9Y6E2) across 57 species, ranging from unicellular eukaryotes to vertebrates, generated with MUSCLE and rendered with ESPript 3.2. Species names are shown with the corresponding UniProt/UniParc/RefSeq accession used for the alignment. Identical residues are shaded red. The Arg17-Lys18-Arg19 (RKR) motif, which contacts the mRNA within the 40S mRNA tunnel (Fig. 5E), is among the most highly conserved elements.

Table S1. Near-Cognate Start Codons in the *c-MYC* 5′UTR (nt 1–408)

#	Position	Codon	Frame	Codons upstream	Kozak			Sequence
					-3	+4	Strength	
1	7	CUG	In-frame	134	U	U	weak	ACUCG[CUG]UAGUA
2	81	CUG	Out-of-frame	109	G	C	moderate	CAGAG[CUG]CGCUG
3	86	CUG	Out-of-frame	107	G	C	moderate	CUGCG[CUG]CGGGC
4	97	CUG	In-frame	104	G	G	strong	GCGUC[CUG]GGAAG
5	136	CUG	In-frame	91	C	G	moderate	CGCCU[CUG]GCCCA
6	153	CUG	Out-of-frame	85	C	A	weak	UCCCG[CUG]AUCCC
7	231	CUG	Out-of-frame	59	G	G	strong	UUGCA[CUG]GAACU
8	287	CUG	Out-of-frame	40	A	C	moderate	CUAUU[CUG]CCCAU
9	315	CUG	Out-of-frame	31	C	C	weak	CGCCG[CUG]CCAGG
10	333	CUG	Out-of-frame	25	U	A	weak	CUUCU[CUG]AAAGG
11	353	CUG	Out-of-frame	18	C	C	weak	UGCAG[CUG]CUUAG
12	364	CUG	In-frame	15	A	G	strong	AGACG[CUG]GAUUU
13	182	UUG	Out-of-frame	75	C	C	weak	AACCC[UUG]CCGCA
14	201	UUG	Out-of-frame	69	A	C	moderate	AAACU[UUG]CCCAU
15	226	UUG	In-frame	61	A	C	moderate	GCACU[UUG]CACUG
16	295	UUG	In-frame	38	C	G	moderate	CCCAU[UUG]GGGAC
17	347	UUG	Out-of-frame	20	U	C	weak	UCUCC[UUG]CAGCU
18	58	GUG	In-frame	117	A	G	strong	CUAGG[GUG]GAAGA
19	382	GUG	In-frame	9	G	G	strong	GGGUA[GUG]GAAAA

Table S2. 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
EMPIAR ID	13853/13577	13852
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**
EM Data Bank and PDB identifiers		EMD-78242, 37KM
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		6.81
Good sugar pucker (%)		98.62
Ramachandran plot		
Favored (%)		94.0

*Over 10 representative micrographs

Table S3: DNA sequences of NLuc reporters and constructs with *c-Myc* 5′ UTR.

Construct	Sequence (5′-3′)
Strong AUG (−3A/+4G)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagacaatggatcatcatcatcatcatggctcgagcggcgtcttcacactcgaagatttcgttg gggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaa aggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttt aaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggc cgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacg gctccctgctgttccgagtaaccatcaacggagtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacat catgaagccccttgagcatctgacttctggctaataaaggaaatttatttttcattgc
Moderate AUG (−3A/+4C)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagacaatgcatcatcatcatcatcatggctcgagcggcgtcttcacactcgaagatttcgttg gggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaa aggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttt aaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggc cgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacg gctccctgctgttccgagtaaccatcaacggagtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacat catgaagccccttgagcatctgacttctggctaataaaggaaatttatttttcattgc
Weak AUG (−3C/+4C)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagccaatgcatcatcatcatcatcatggctcgagcggcgtcttcacactcgaagatttcgttg gggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaa aggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttt aaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggc cgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacg gctccctgctgttccgagtaaccatcaacggagtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacat catgaagccccttgagcatctgacttctggctaataaaggaaatttatttttcattgc
Strong CUG (−3A/+4G)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagacactggatcatcatcatcatcatggctcgagcggcgtcttcacactcgaagatttcgttg gggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaa aggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttt aaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggc cgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacg gctccctgctgttccgagtaaccatcaacggagtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacat catgaagccccttgagcatctgacttctggctaataaaggaaatttatttttcattgc
Moderate CUG	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagacactgcatcatcatcatcatcatggctcgagcggcgtcttcacactcgaagatttcgttg gggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaa

(−3A/+4C)	aggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatttattttcattgc
Weak CUG (−3C/+4C)	acacttgcttttgacacaactgtgtttacttgcaatcccccaaaacagccactgcatcatcatcatcatcatggctcgagcggcgtcttcacactcgaagatttcgttgggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (WT)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgggcgtcctgggaagggagatccggagcgaatagggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttgccgcatccacgaaacttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgcggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgctggatttttttcgggtagtggaaaaccagcagcctcccgcgacgatgcccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcgagcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggctgcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatttattttcattgc
c-Myc-NLuc (Mut 1)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgggcgtcctgggaagggagatccggagcgaatagggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttgccgcatccacgaaacttgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgcggggaggctattctgcccatttggggacacttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgccggatttttttcgggtagtggaaaaccagcagcctcccgcgacgatgcccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcgagcggcgactataaggaccacgacggagactacaaggatcatgatattgattacaaggacgacgatgacaagggaggaggaggaagcgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactg

	taacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggctg gcggctgtgcgaacgcattctggcgtaagatctttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatt tattttcattgc
c-Myc-NLuc (Mut 2)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgggcgtcctg ggaagggagatccggagcgaatagggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttgccgcatccacgaaactt tgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgcggggaggctattctgcccatttggggaca cttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgctggatttttttcgggtagtggaaaaccagcagcctcccgcgacga gacccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcgagcggcgactataaggaccacgacggagactacaaggatcatgatattgat tacaaggacgacgatgacaagggaggaggaggaagcgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagt ccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatc catgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcct gcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactg taacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggctg gcggctgtgcgaacgcattctggcgtaagatctttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatt tattttcattgc
c-Myc-NLuc (Mut 3)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgggcgtcctg ggaagggagatccggagcgaatagggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttgccgcatccacgaaactt tgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgcggggaggctattctgcccatttggggaca cttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgctggatttttttcgggtagaggaaaaccagcagcctcccgcgacg agacccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcgagcggcgactataaggaccacgacggagactacaaggatcatgatattg attacaaggacgacgatgacaagggaggaggaggaagcgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaa gtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgaca tccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatc ctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcac tgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggc tggcggctgtgcgaacgcattctggcgtaagatctttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaa atttattttcattgc
c-Myc-NLuc (Mut 4)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgggcgtcctg ggaagggagatccggagcgaatagggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttgccgcatccacgaaactt tgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgcggggaggctattctgcccatttggggaca cttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgctggacttcttccgcgtcgtggagaaccagcagccccccgccac cagacccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcgagcggcgactataaggaccacgacggagactacaaggatcatgatatt gattacaaggacgacgatgacaagggaggaggaggaagcgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggacca

	agtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgac atccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgat cctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatca ctgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccgg ctggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaagga aatttattttcattgc
c-Myc-NLuc (Mut 5)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgggcgtcctg ggaagggagatccggagcgaatagggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttgccgcatccacgaaactt tgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgcggggaggctattctgcccatttggggaca cttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgatggatttttttcgggtagtggaaaaccagcagcctcccgcgacga gaccccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcgagcggcgactataaggaccacgacggagactacaaggatcatgatattgat tacaaggacgacgatgacaagggaggaggaggaagcgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagt ccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatc catgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcct gcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactg taacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggctg gcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatt tattttcattgc
c-Myc-NLuc (Mut 6)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgggcgtcctg ggaagggagatccggagcgaatagggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttgccgcatccacgaaactt tgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgcggggaggctattctgcccatttggggaca cttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgccggatttttttcgggtagtggaaaaccagcagcctcccgcgacg agaccccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcgagcggcgactataaggaccacgacggagactacaaggatcatgatattg attacaaggacgacgatgacaaggaggaggaggaagcgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaa gtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgaca tccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatc ctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcac tgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggc tggcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaa atttattttcattgc
c-Myc-NLuc (Mut 7)	aactcgctgtagtaattccagcgagaggcagagggagcgagcgggcggccggctagggtggaagagccgggcgagcagagctgcgctgcgggcgtcctg ggaagggagatccggagcgaatagggggcttcgcctctggcccagccctcccgctgatcccccagccagcggtccgcaacccttgccgcatccacgaaactt tgcccatagcagcgggcgggcactttgcactggaacttacaacacccgagcaaggacgcgactctcccgacgcggggaggctattctgcccatttggggaca

	cttccccgccgctgccaggacccgcttctctgaaaggctctccttgcagctgcttagacgccggatttttttcgggtagaggaaaaccagcagcctcccgcgacg atgcccctcaacgttagcttcaccaacaggaactatgacctcgactacggctcgagcggcgactataaggaccacgacggagactacaaggatcatgatattga ttacaaggacgacgatgacaagggaggaggaggaagcgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagt ccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatcgacatc catgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggccagatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcct gcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactg taacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtgaccggctg gcggctgtgcgaacgcattctggcgtaagatcttttccctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatt tattttcattgc

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.