NAR Genomics and Bioinformatics, 2024, **6**, lqae065
https://doi.org/10.1093/nargab/lqae065
Advance access publication date: 11 June 2024
Standard Article

Critical *cis*-parameters influence STructure assisted RNA translation (START) initiation on non-AUG codons in eukaryotes

Antonin Tidu, Fatima Alghoul, Laurence Despons, Gilbert Eriani and Franck Martin*

Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire, Architecture et Réactivité de l'ARN, CNRS UPR9002, 2 allée Konrad Roentgen, F-67084 Strasbourg, France

*To whom correspondence should be addressed. Tel: ++ 33 03 88 41 70 42; Fax: ++33 03 88 60 22 18; Email: f.martin@ibmc-cnrs.unistra.fr
Present addresses:
Antonin Tidu, Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Bern, Switzerland.
Fatima Alghoul, Department of Cell Biology, Harvard Medical School, Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

Abstract

In eukaryotes, translation initiation is a highly regulated process, which combines *cis*-regulatory sequences located on the messenger RNA along with *trans*-acting factors like eukaryotic initiation factors (eIF). One critical step of translation initiation is the start codon recognition by the scanning 43S particle, which leads to ribosome assembly and protein synthesis. In this study, we investigated the involvement of secondary structures downstream the initiation codon in the so-called START (STructure-Assisted RNA translation) mechanism on AUG and non-AUG translation initiation. The results demonstrate that downstream secondary structures can efficiently promote non-AUG translation initiation if they are sufficiently stable to stall a scanning 43S particle and if they are located at an optimal distance from non-AUG codons to stabilize the codon-anticodon base pairing in the P site. The required stability of the downstream structure for efficient translation initiation varies in distinct cell types. We extended this study to genome-wide analysis of functionally characterized alternative translation initiation sites in *Homo sapiens*. This analysis revealed that about 25% of these sites have an optimally located downstream secondary structure of adequate stability which could elicit START, regardless of the start codon. We validated the impact of these structures on translation initiation for several selected uORFs.

Introduction

In eukaryotes, translation initiation is a highly regulated process, which combines *cis*-regulatory sequences located on the messenger RNA along with *trans*-acting factors like eukaryotic initiation factors (eIF) (1). The canonical translation initiation mechanism starts with the recognition of the 5′ m^7G cap by eIF4E, which is associated with eIF4G and eIF4A to form the eIF4F complex. The interaction between eIF4G and 43S-bound eIF3 enables the recruitment of the 43S particle on the m^7G cap. Then, the 43S particle scans the 5′UTR thanks to the ATP-dependent RNA helicase activity of eIF4A, which is enhanced through its interactions with eIF4G, eIF4B (and/or eIF4H) and eIF3 (2,3). During scanning, the 43S particle is maintained in a so-called open conformation by the combined actions of eIF1A and eIF1 (4), which destabilize the codon-anticodon interaction in the P-site, inhibit spontaneous hydrolysis of eIF2-bound GTP (5) and therefore prevent ribosomal subunits joining (6–8). The codon–anticodon interaction involves the nucleotide triplet from the codon being analyzed in the P-site and the anticodon of the initiator Met-$tRNA^{Met}i$, which is bound to eIF2-GTP in the 43S particle (9). When a stable enough codon–anticodon base-pairing is established during scanning, it displaces eIF1 from the P-site and triggers the switch from the open to the close conformation of the 43S. In mammals, this results in the activation of the GTPase activity of eIF2 by eIF5 and in the subsequent hydrolysis of eIF2-bound GTP into GDP (10–13,15). As GTP hydrolysis goes on, eIF2 loses its affinity for the Met-$tRNA^{Met}i$ and is released from the 48S particle (14). A distinct timing for the hydrolysis of eIF2-bound GTP has been described in yeast where the hydrolysis of eIF2-bound GTP occurs progressively during scanning with the assistance of eIF5 which functions as a GTPase activating protein (GAP) (15–17). However, the position of eIF1 in the close conformation of the yeast scanning complex prevents Pi release. It is the structural rearrangements of the scanning complex following start-codon recognition that promote further hydrolysis of eIF2-bound GTP and Pi release by eIF5, leading ultimately to the assembly of the 48S complex (15,18). Overall, if the hydrolysis of eIF2-bound GTP and Pi release is a requirement for the assembly of the 48S complex upon start codon recognition, the kinetics of its release seems to differ among eukaryotes (15). Subunits joining is mediated by eIF5B (19) and its interaction with eIF1A (20), which contributes to the release eIF2, eIF1, eIF1A and eIF5B from the initiation complex through the hydrolysis of eIF5B-bound GTP. The assembled 80S ribosome then starts protein synthesis by incorporating the first amino-acyl-tRNA in its vacant A-site.

Start-codon selection by the scanning 43S particle is the most critical step of initiation as it determines the amino-acid sequence of the protein that will be synthesized, and therefore has drastic consequences for its function(s). Incorrect start site selection can result in N-terminal extensions or

Received: March 6, 2024. **Revised:** April 18, 2024. **Editorial Decision:** May 22, 2024. **Accepted:** May 23, 2024

deletions, leading to erroneous protein isoforms, or to out-of-frame protein synthesis, leading to an unrelated and non-functional translation product. Among several parameters, the most critical feature of start codon selection is the stability of the codon-anticodon interaction. Start-codon selection stringency is mediated by eIF1 and eIF1A which prevent unstable codon-anticodon interactions like those involving non-cognate codons (6–8). However, even AUG codons that establish a perfect Watson-Crick base-pairing with the anti-codon loop of the Met-tRNA^{Met}i are sometimes not recognized as the initiation codon. Moreover, several studies (21–25) along with ribosome profiling data (26–28) demonstrated that non-AUG codons can be used for translation initiation in eukaryotes. Therefore start-codon selection is influenced by additional *trans*- and *cis*-acting factors (1). Several reports, starting with the seminal studies by Marylin Kozak (29,30), demonstrated that the start codon nucleotide context plays a major role on start codon recognition efficiency. Distal secondary structures elements have also been shown to modulate start codon selection stringency on individual mRNAs (21,22,24,25,31). At the translatome level, it has been shown that the impairment of the RNA helicase Ded1p is linked to increased uORFs translation in yeast (32). Furthermore, bioinformatics studies have shown the broad presence of secondary structures downstream of initiation sites (33–35). These observations show the involvement of downstream RNA secondary structures in start codon selection. Similarly, proteins bound to an RNA element downstream of an initiation site can promote start codon recognition. An example is the Drosophila protein Sex-Lethal (SXL) that promotes uORF translation of male specific lethal two mRNA by stalling 43S complexes upstream the main open reading frame (36,37). Based on these observations, we previously described a mechanism of translation initiation called STructure Assisted RNA Translation or START (35,38). START initiation relies on the presence of a downstream structure that enhances translation initiation on both AUG and non-AUG codons by stalling the scanning 43S particle with those codons in the P-site, leading to 80S assembly and protein synthesis on these sub-optimal initiation codons.

In this work, we aimed to determine to what extent non-AUG translation initiation (39,40) relies on a START mechanism in mammalian systems. We have precisely investigated three critical parameters for START in the presence of a downstream RNA structure to the initiation codon: the distance between the structure and the P-site, the structure stability, and the type of the start codon. We show that downstream secondary structures can efficiently promote non-AUG translation initiation if they are sufficiently stable to stall a scanning 43S particle and if they are located at an optimal distance from non-AUG codons to stabilize the codon-anticodon base pairing in the P site. In addition, the efficiency of START relies on the start codon type along with its nucleotide environment and is also highly influenced by cell-specific *trans*-acting factors. We have previously shown the broad presence of downstream RNA structures in annotated main open reading frames of various organisms (35). Here, we further extended this study by analyzing genome-wide alternative translation initiation sites in human transcripts that were previously determined by ribosome profiling (28) and found the broad presence of moderately stable downstream RNA structures within the distance range we have determined being optimal for START. Functional analysis of some examples confirmed the role of downstream structures on translation efficiency at these alternative translation initiation sites.

Materials and methods

Oligonucleotides information

All the oligonucleotides' sequences used in this study are provided in supplementary material (Supplementary Table S1). Oligonucleotides were purchased from Integrated DNA Technologies company. In the text, oligonucleotides are referred to by their red numbers in the Supplementary Table S1.

Rabbit reticulocyte lysate

Untreated rabbit reticulocyte lysate (RRL) was prepared as previously described (41).

HEK293FT and SH-SY5Y cell lysates from cells cultured in physiological conditions

HEK293FT cells were seeded at 0.03 × 10^6 cells/cm^2 and grown for 2 days at 37°C in a 5% CO_2 humidified atmosphere. The culture medium is Gibco Dulbecco's modified Eagle's medium + GlutaMAX (Life Technologies) which is supplemented with 10% of inactivated fetal bovine serum (Life Technologies). Neuroblastoma SH-SY5Y cells were cultured in the same conditions and the same medium but supplemented with 1 mM of non-essential amino acids (Gibco). The total culture surface was 4500 cm^2, leading approx. to 200–400 × 10^6 cells before harvesting. Cells were harvested at room temperature in their culture medium by centrifugation (300g) at 4°C and washed two times with a cold buffer containing 20 mM HEPES–KOH pH 7.5, 100 mM potassium acetate, 2 mM magnesium acetate, 1 mM DTT. After washing, the cell pellets were resuspended in the same buffer supplemented or not (if TEV protease is used, see later) with 1× Halt™ Protease Inhibitor Cocktail EDTA-free (Thermo Scientific™) to reach a concentration of 50–100 × 10^6 cells/ml. Cells were lysed by nitrogen cavitation with a Cell Disruption Bomb (Parr Instrument Company) after a one-hour incubation under a pressure of 30 bar at 4°C. The lysate was cleared by centrifugations at 10 000g at 4°C, aliquoted, flash-frozen in liquid nitrogen and stored at −80°C. Total protein concentration was determined by Bradford assay (Biorad).

HEK293FT cell lysates from cells cultured in stress conditions

To prepare cell-free translation extracts from HEK293FT cells cultured in stress conditions, the same procedure was used except that the culture medium was supplemented with 5 mM DTT 3 h before harvesting to induce endoplasmic reticulum (ER) stress, which triggers the unfolded protein response (UPR) (42,43). This pathway leads in particular to the phosphorylation of the eIF2 α-subunit by the protein kinase R-like endoplasmic reticulum kinase (PERK) (44,45).

Measurements of eIF2 α-subunit phosphorylation by Western Blot

Aliquots of *in vitro* translation reactions were run on 12% SDS-PAGE. Proteins were transferred on 0.45 μm PVDF membranes (Immobilon®-P Transfer Membrane) for 1 h at 10 V. Membranes were saturated with [PBS 1×, Tween20 0.5%, BSA 50 mg/ml]. The primary antibodies (@eIF2α: #9722 and

@eIF2α-phosphorylated: #3597, Cell Signaling Technology) were diluted 10 thousand times in PBS 1×, Tween-20 0.5%, BSA 50 mg/ml and hybridized overnight at 4°C. Membranes were then washed three times in PBS 1×, Tween20 0.5% before the 2-h hybridization of the secondary antibody (@rabbit A120-101-P, Bethyl Laboratories) which was diluted ten thousand times in PBS 1×, Tween-20 0.1%. Membranes were finally washed three times in PBS 1×, Tween-20 0.1% before the addition of the ECL substrate (Clarity Western ECL Substrate #1705061). Chemiluminescence signals were measured with a Biorad ChemiDoc (MP) apparatus and resulting TIFF images were analyzed with ImageJ software.

Reporter RNA synthesis for *in vitro* translation

All plasmids containing the 5′UTR and protein reporter sequences of interest were prepared using the NEBuilder® HiFi DNA Assembly Cloning Kit (#E5520S). Site-directed mutagenesis was performed with NEB Q5® Site Directed Mutagenesis Kit (#E0554S). Then, the corresponding T7-5′UTR-reporter DNA construct was PCR-amplified using forward primer n°1 (capped reporters) or n°2 (IGR reporters) and reverse primer n°3 (Supplementary Table S1). Corresponding RNA reporters (Supplementary File 1) were synthesized in a 100 µl *in vitro* transcription reaction using 0.1 µM of T7-DNA template, 5 mM Tris–HCl pH 8, 30 mM $MgCl_2$, 1 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 5 mM of each ribonucleotide (ATP, CTP, GTP, UTP pH 7.5), 0.5 U/µl RNase inhibitor (Promega) and 0.125 mg/ml of home-made recombinant T7 RNA polymerase. The reaction was incubated for 1h at 37°C. For co-transcriptional capping, the reaction was started in the absence of GTP and in the presence of 0.5 mM of anti-reverse m^7G-cap analog (NEB #S1411) for 10 min. GTP was then gradually added up to 5 mM at 1h incubation. After 1h incubation, 0.02 mg/ml pyrophosphatase (Merck) was added and after 30 min, the DNA template was degraded by 1 h incubation with 0.2U/µl DNaseI (Roche). Transcription products were loaded on a 1 ml G-25 Superfine Sephadex column (Cytiva) and the resulting eluate was phenol-extracted and ethanol-precipitated. The resulting RNA pellets were resuspended in 30 µl milli-Q water, frozen, and quantified by 260 nm absorbance measurement. All reporter RNA sequences used in this work are provided in Supplementary File 1.

In vitro translation assays

RRL: The final composition of the reaction was 50% RRL, 100 mM potassium acetate, 1 mM magnesium acetate, 0.1 mM of all amino acids but methionine, 0.5 U/µl RNase inhibitor (Promega), 0.125 µCi/µl ^{35}S-methionine (Perkin-Elmer) and 0.2 µM reporter RNA in a total volume of 20 µl. The reaction was incubated for 1 h at 30°C or 2 h at 30°C when TEV protease was used for co-translational cleavage.

HEK293FT and SH-SY5Y extracts: The final composition of the reaction was 15% cell extract (1.8 µg/µl total protein), 100 mM potassium acetate, 1 mM magnesium acetate, 0.1 mM of all amino acids but methionine, 20 mM HEPES KOH pH 7.5, 0.5 mM spermidine, 1 mM DTT, 0.8 mM ATP, 0.1 mM GTP, 8 mM phospho-creatine, 0.1 µg/µl creatine phospho-kinase, 0.5 U/µl RNase inhibitor (Promega), 0.125 µCi/µl ^{35}S-methionine (Perkin-Elmer) and 0.2 µM reporter RNA in a total volume of 20 µl. The reaction was incubated for 1 h at 30°C or 2 h at 30°C when TEV protease was used for co-translational cleavage.

Co-translational cleavage of in vitro synthesized reporter proteins with TEV protease: Cleavage was triggered by the addition of recombinant home-made TEV protease to *in vitro* translation reactions at a final concentration of 0.035 µg/µl.

Assessment of *in vitro* translation products levels

^{35}S-signal acquisition: 5 µl of *in vitro* translation reactions were analyzed by 12% SDS-PAGE. ^{35}S-labelled translation products bands were quantified with a phosphorimager (Typhoon FLA 7000). Image analysis was conducted with ImageJ software using the resulting TIFF files.

Luciferase assay: The amount of *in vitro* synthetized Renilla luciferase was assessed upon injection of 100 µl coelenterazine 0.25 µmol/ml (Synchem) into 10 µl of *in vitro* translation reaction using a luminometer (Varioskan LUX, Thermo Scientific™). Subsequent photon emission was measured for 10 seconds with the default measurement parameters.

Raw luminescence and ^{35}S-signal were processed with a self-made Python script using the relevant functions of *numpy* (v 1.24.3), *pandas* (v 2.1.1) and *seaborn* (v 0.12.2) python packages. For luminescence signals, as each sample was assessed in triplicates, descriptive (mean, standard deviation) and inferential statistics (Student *P*-values) of the ratio of sample A (triplicates a1, a2, a3) over sample B (triplicates b1, b2, b3) were determined using the nine possible ratios between (a1, a2, a3) and (b1, b2, b3). For ^{35}S-signal, as signals were normalized within each 3 gels independently, this resulted in three possible ratios for assessing the statistics of A/B. Statistical tests and annotations were performed with the *statannotations* (v 0.4.4), *scipy* (v 1.11.3) *and statsmodels* (v 0.14.0) python packages. Statistical analysis of global null hypothesis and corresponding corrected p-values for all *in vitro* translation assays are provided in Supplementary File 4.

Predictions and free energy calculations of RNA secondary structures

These calculations were conducted using the Mfold V2.3 for RNA (46) on the various a11 structure mutants used in this work. Constraints were implemented according to the probing data that enabled us to draw the 2D model of the wild-type structure (47).

Automated secondary structures predictions in transcripts

RNA transcript sequence recovery from human smORF data (28) was performed using two homemade Python scripts available as supplementary data. They must be executed in the following order: (I) format_smORFs_seq.py, (II) select_smORF_sequences.py. Nucleotide sequences between the +16 and +65 positions relative to the initiator codon of smORFs were extracted. When the smORF coding sequences were shorter than 65 bases, BLASTN searches (48) were run against local human RNA databases to find the corresponding transcripts. Two additional scripts were used to search and then analyze smORF structures: (III) search_2Dstruct_in_smORFs.py, (IV) analyse_2Dstruct_in_smORFs.py. Scripts III and IV are based on scripts used in a previous study (35) and modified to suit the prediction or analysis of secondary structures in the +16 to +65 region of the smORF transcripts. The structure predic-

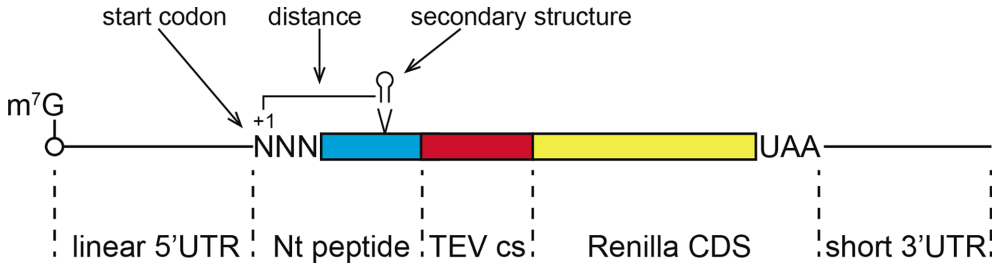

Figure 1. Schematic representation of the RNA reporters used in the study. The three *cis*-acting parameters studied are (i) the distance between a downstream structure and the start codon, (ii) the stability of the downstream structure, (iii) the nature of the start codon. TEV cs: Tobacco Etch Virus protease cleavage site, Nt peptide: N-terminal fusion peptide, Renilla CDS: Renilla luciferase coding sequence.

tion algorithm implemented in these scripts is RNAfold from the ViennaRNA package 2 (49). RNAfold program calculates the minimum free energy (MFE) of RNA secondary structures such as stem-loops and G-Quadruplexes. The Python module named toolbox.py is imported by some of these scripts (Supplementary File 5: Python_scripts.zip). Further analyses were made using the relevant functions of *numpy* (v 1.24.3), *pandas* (v 2.1.1) and *seaborn* (v 0.12.2) python packages. In particular, comparisons of distributions were performed using two statistical tests. The Mann–Whitney *U* test was used to determine if the AUG and the non-AUG distributions are equal. The Mood's median test was used to determine whether the AUG and the non-AUG distributions have the same median and, under the assumption that all distributions are left-tailed and centered on their median, whether they are similarly centered. The probability of superiority (PS) and the Hodges–Lehmann estimator (HL) were used for effect size assessment. In Figure 7, the probability of superiority (PS) corresponds to the probability that a randomly selected observation from the AUG distribution has a mfe superior to the mfe of a randomly selected observation from the non-AUG distribution it is compared with. The Hodges–Lehmann estimator is defined by the median of the pairwise differences between all observations from the compared distributions.

Gene ontology (GO) terms analysis

GO-terms analysis was performed using the relevant functions from the GOATools package, version 1.3.11 (50). For GO-terms enrichment analysis, *P*-values from Fisher exact tests were corrected using the Holm method. All data is provided in Supplementary File 3. The gene ontology file (go-basic.obo) was downloaded from Download ontology (geneontology.org) in its latest version to date (release from 15/11/2023). Human annotation file (goa_human.gaf) was downloaded from Download Annotations | Gene Ontology Consortium in its latest version to date (Panther version v.17.0, GO version 2023-10-09).

Results

A model mRNA reporter system to investigate the impact of downstream secondary structures on start codon recognition

To test the impact of several critical parameters for START, we synthesized a set of reporter mRNAs that encode the Renilla luciferase fused to variable N-terminal peptides resulting from the translation of various RNA structures downstream of the start codon (Figure 1). To avoid any putative artifacts due to variable N-terminal sequences that would influence the luciferase activity of the Renilla protein, we introduced a TEV protease cleavage upstream of the Renilla coding sequence. The quantification of luciferase activity was performed after TEV cleavages which yield identical Renilla proteins produced by all the reporter mRNAs. The sequences of all the reporter RNA used in this study are shown in Supplementary File 1. More precisely, these reporter mRNAs feature a linear CAA-rich 5′UTR to minimize the effect of structures on 5′-3′ scanning, which is followed by a start codon of interest. These reporter mRNAs were co-transcriptionally m^7G-capped, and their translation efficiency was determined by Renilla luciferase activity measurements and/or ^{35}S-methionine incorporation. With these reporters, we investigated the influence of both the location and the structure stability on translation initiation efficiency, as well as the nature of the start codon and its nucleotide context. We confirmed that the cell-free translation extracts used in this study are cap-dependent (Supplementary Figure S1).

As a model secondary structure, we used a 50-nucleotide GC-rich secondary structure from the 5′UTR of Hox a11 mRNA. We have previously determined its secondary structure by probing and shown that this structure is very stable ($\Delta G = -21.6$ kcal/mol) as it contains 16 G–C base pairs. Importantly, the a11 structure can stall a scanning complex (47).

The optimal position of the secondary structure for translation on a CUG codon ranges from +23 to +26

To determine the optimal distance between the a11 secondary structure and the start codon, *in vitro* translation assays using untreated rabbit reticulocyte lysate were conducted with reporter mRNAs containing an AUG or a CUG start codon with the a11 secondary structure located at positions ranging from +11 to +35 (Figure 2A). First, the position of the secondary structure affected translation on both AUG and CUG codons (Figure 2B, C and Supplementary File 4). However, while AUG translation was only mildly sensitive to the position of the secondary structure (Figure 2B, C), CUG translation was much more sensitive (Figure 2C). CUG translation is of note impaired when the structure is located at positions +11 or +35. In addition, a stronger increase in relative translation efficiency is observed for CUG compared to AUG at position +26, indicating the optimal position of the structure would be close to +26 for CUG translation initiation. We then determined the optimal position of the structure for CUG translation (Figure 3A). Translation efficiency gradually decreases when the structure is located before +17 or beyond +32 from the CUG codon (Figure 3B, C). This

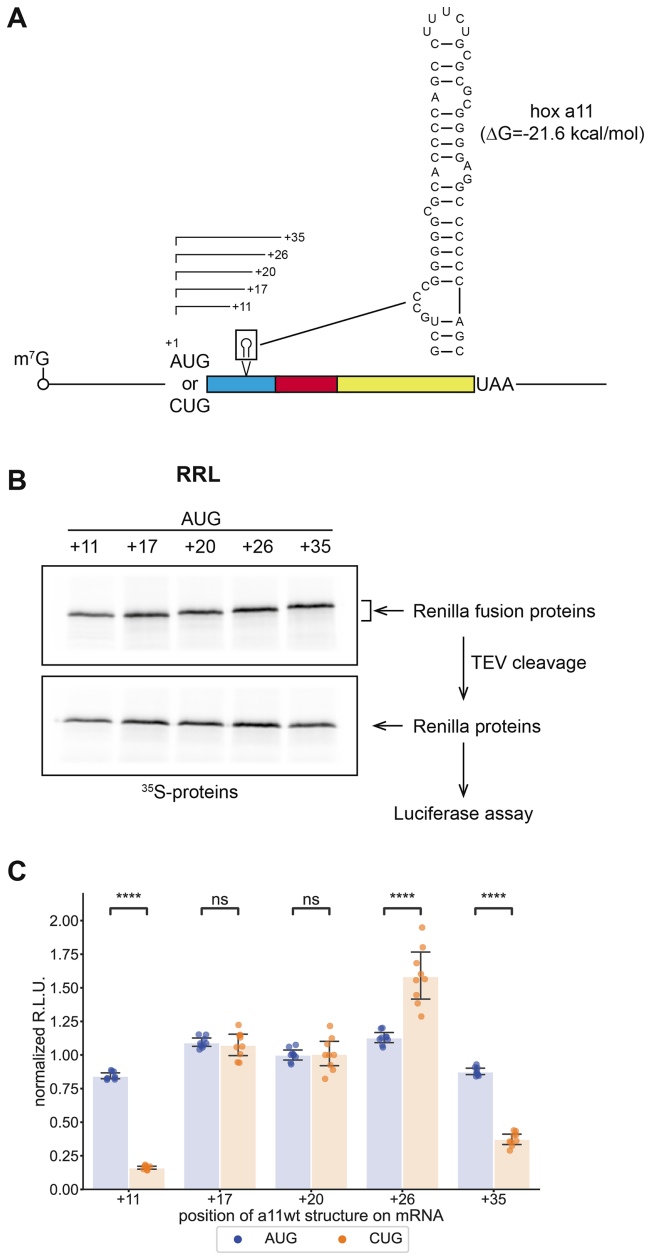

Figure 2. Translation initiation on CUG is sensitive to the position of the downstream secondary structure. (**A**) Schematic representation of the RNA reporters highlighting the two AUG or CUG initiation codons used and the wild-type a11 hairpin, which is inserted at +11, +17, +20, +26 or +35 positions. (**B**) Representative SDS-PAGE of *in vitro* translation products obtained with AUG reporters in RRL. The upper panel shows the ^{35}S-labelled products before TEV cleavage, the lower panel shows the ^{35}S-labelled products after TEV cleavage. Proteins produced by TEV cleavage were used for Luciferase assays. (**C**) Quantification of Renilla luciferase luminescence produced in RRL from a11-RNA reporters initiating with an AUG codon (blue) or with a CUG codon (orange). Relative luminescence units (R.L.U.) were normalized to those obtained with the +20 AUG reporter for all AUG reporters or with the +20 CUG reporter for all CUG reporters. Bar heights show mean normalized ratios and error bars represent the corresponding 99% confidence interval using the *t*-distribution. *P*-values were calculated using a Student *t*-test for independent samples. *: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$, ****: $P < 0.0001$

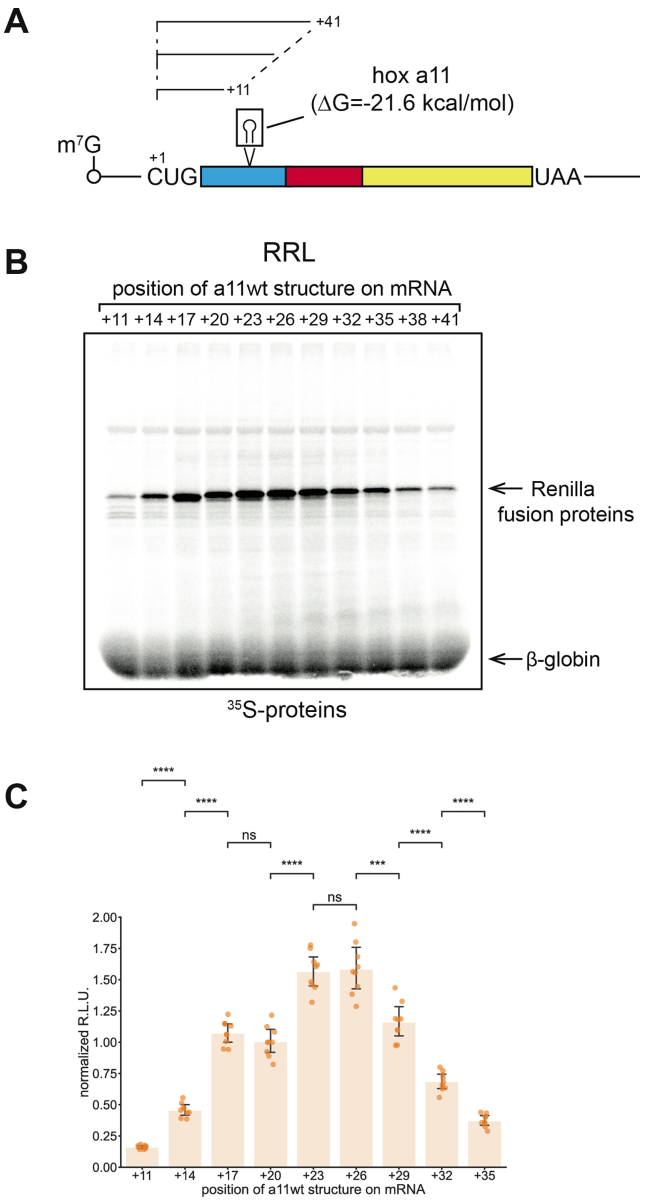

Figure 3. The optimal position of the secondary structure for initiation on a CUG codon is between +23 to +26. (**A**) Schematic representation of the RNA reporters initiating on a CUG codon upstream of the wild-type a11 hairpin inserted at positions +11, +17, +20, +23, +26, +29, +32, +35, +38 or +41. (**B**) Representative SDS-PAGE of *in vitro*^{35}S radiolabelled translation products obtained with AUG reporters in RRL. (**C**) Quantification of Renilla luciferase luminescence produced in RRL from a11 RNA reporters initiating with a CUG codon. Relative luminescence units (R.L.U.) were normalized to those obtained with the +20 CUG reporter. Bar heights show the mean normalized ratios and error bars show the 99% confidence intervals calculated with the t-distribution. *P*-values were calculated using a Student *t*-test for independent samples. *: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$, ****: $P < 0.0001$.

behavior suggests that CUG initiation is enhanced by the optimal positioning of the structure (i.e. by a START mechanism) rather than by the alleviation of its potential inhibitory effect when it is too close to the initiation site. START enhancement of CUG initiation is further confirmed by the fact that in the absence of structure at +20, CUG initiation is dramatically impaired indicating a strong requirement for a

sufficiently stable structure (Figure 4A, –4 kcal/mol 'structure'). Overall, these results show that while efficient AUG initiation does not rely on stable downstream secondary structures located between +11 and +35, CUG initiation requires a structure located between +23 and +26. Although not tested here, it could be reasonably assumed that these effects are further modulated by the start codon nucleotide context. Translation initiation enhancement by START would indeed be more pronounced for start codons (AUG or not) flanked by a rather inefficient nucleotide context for initiation.

The optimal stability of the secondary structure is influenced by the cell type

Next, we examined the required stability of the secondary structure for efficient translation initiation on a CUG codon. We used reporters with an AUG as a control and a CUG start codon upstream of the a11 secondary structure that is localized at the rather optimal position +20, as determined in the previous section. We chose a highly efficient but suboptimal position so that our conclusions would not be restrained to the most optimal situation regarding structure positioning on the mRNA. The inserted a11 structures contain mutations that progressively decrease the number of G–C base pairs, resulting in predicted stabilities ranging from –21.6 to –4.0 kcal/mol (Figure 4A). Silent mutations were used to avoid any putative co-translational side effects due to amino acid substitutions, except for the –4 kcal/mol mutant structure that contains mutations leading to three amino acids changes in the N-terminal fusion of the translated reporter protein. In addition, since TEV cleavage was performed co-translationally, all the translated mRNAs produced the same Renilla luciferase protein. We therefore eliminated potential N-terminal induced co-translational artifacts on the activity of the Renilla luciferase protein.

We first confirmed with rabbit reticulocyte lysates that translation of the AUG reporters does not rely on downstream structures while CUG translation does (Figure 4B and Supplementary File 4). For the AUG reporters, the secondary structure may even have an inhibitory effect on initiation as translation of the AUG reporters is more efficient when a structure less stable than –15.3 kcal/mol is present (Figure 4B, C). An explanation would be that the most stable versions of the secondary structure would interfere with start codon recognition by the scanning 43S particle, and/or with the subsequent molecular events leading to 80S assembly. On the contrary, initiation on the CUG codon decreases as soon as the secondary structure is less stable than –21 kcal/mol, meaning that CUG initiation requires a stable downstream structure. When the stability is gradually diminished, translation efficiency decreases stepwise. This dependency on stable structures for efficient initiation on the CUG codon is observed despite the potential inhibitory effect of the stable structures previously observed with the AUG codon, which would reasonably also apply with the CUG. Quantitative analysis of the effect of structure stabilities on translation efficiency (calculated by the ratio $\frac{\Delta\ luminescence}{\Delta\ stability}$ in Figure 4C, D) reveals two stability thresholds in RRL above which translation efficiency dramatically increases for the CUG codon: first between –12.8 and –15.3 kcal/mol and second between –18.4 and –21.6 kcal/mol. Altogether, this data shows the importance of a downstream secondary structure with a stability of at least –15 kcal/mol for efficient translation initiation on a CUG codon using these experimental conditions.

We have developed and characterized cell-free *in vitro* translation extracts from human embryonic kidney (HEK293FT) (Supplementary Figures S1 and S2) and neuroblastoma (SH-SY5Y) cell lines, which makes it possible to analyze START in the presence of cell-type-specific *trans*-acting factors. Translation of the reporter RNAs using these human cell extracts resulted in the same trend as with RRL (Figure 4E–H and Supplementary File 4). However, the minimal stability required to reach efficient translation initiation on a CUG codon is different with these extracts: initiation on a CUG codon strictly requires an a11 secondary structure with a stability of –21.6 kcal/mol (Figure 4F, G, I, J). When the a11 structure stability is gradually diminished, translation from the CUG is exponentially reduced, in contrast to the gradual reduction observed with RRL. This indicates that HEK293FT and SH-SY5Y extracts are much more sensitive to the structure stability for START than RRL. This difference with RRL might be attributed to a distinct content of RNA helicases or RNA binding proteins that could unwind or stabilize the downstream structure required for CUG initiation.

The START mechanism alone does not promote efficient translation initiation on all AUG-like codons

Next, we extended these experiments to other AUG-like codons (CUG, GUG, UUG, ACG and AUC) with the –15.3 kcal/mol a11 structure located at position +20 in RRL or the wt a11 structure (–21.6 kcal/mol) for experiments with HEK293FT extracts (Figure 5A). Overall, the AUG-like codons are generally less efficiently recognized by the ribosome. In addition, all AUG-like codons are not equivalently recognized and follow the hierarchy ACG > UUG > AUC > GUG ≈ 0. The result obtained with GUG must however be nuanced, as the inefficiency of GUG initiation is most likely linked to the reporter RNA sequence around the initiation site. When a GUG is introduced in the initiation context of these reporters, it introduces an upstream out-of-frame *AUG* that overlaps the GUG (CGUA*AUG* UGGAC). Such a nucleotide context does not allow efficient initiation on GUG but rather favors initiation on this upstream *AUG* or leaky scanning to other initiation codons further downstream, whose resulting translation products can be detected (Figure 5B). In fact, all the AUG-like codons tested promote significant leaky scanning. In the presence of the a11wt structure using HEK293FT extracts, the same hierarchy is observed. Comparison between unstressed and stressed HEK293FT cell extracts revealed only minor differences in initiation patterns, although the UUG codon was less efficiently utilized than the AUG (Figure 5C). This suggests that stress has a minor effect on the *cis*-regulation of translation initiation on non-AUG codons in our synthetic reporter system.

Next, we checked whether the previously established hierarchy of AUG-like codons would be the same using another nucleotide context with the a11wt reporters. For that purpose, the context CGUAAUNNNGAC that was used in previous experiments was changed to GCCACCNNNGCG (Figure 6A). This context is the average nucleotide context of the AUG initiation codon in the human genome, also known as

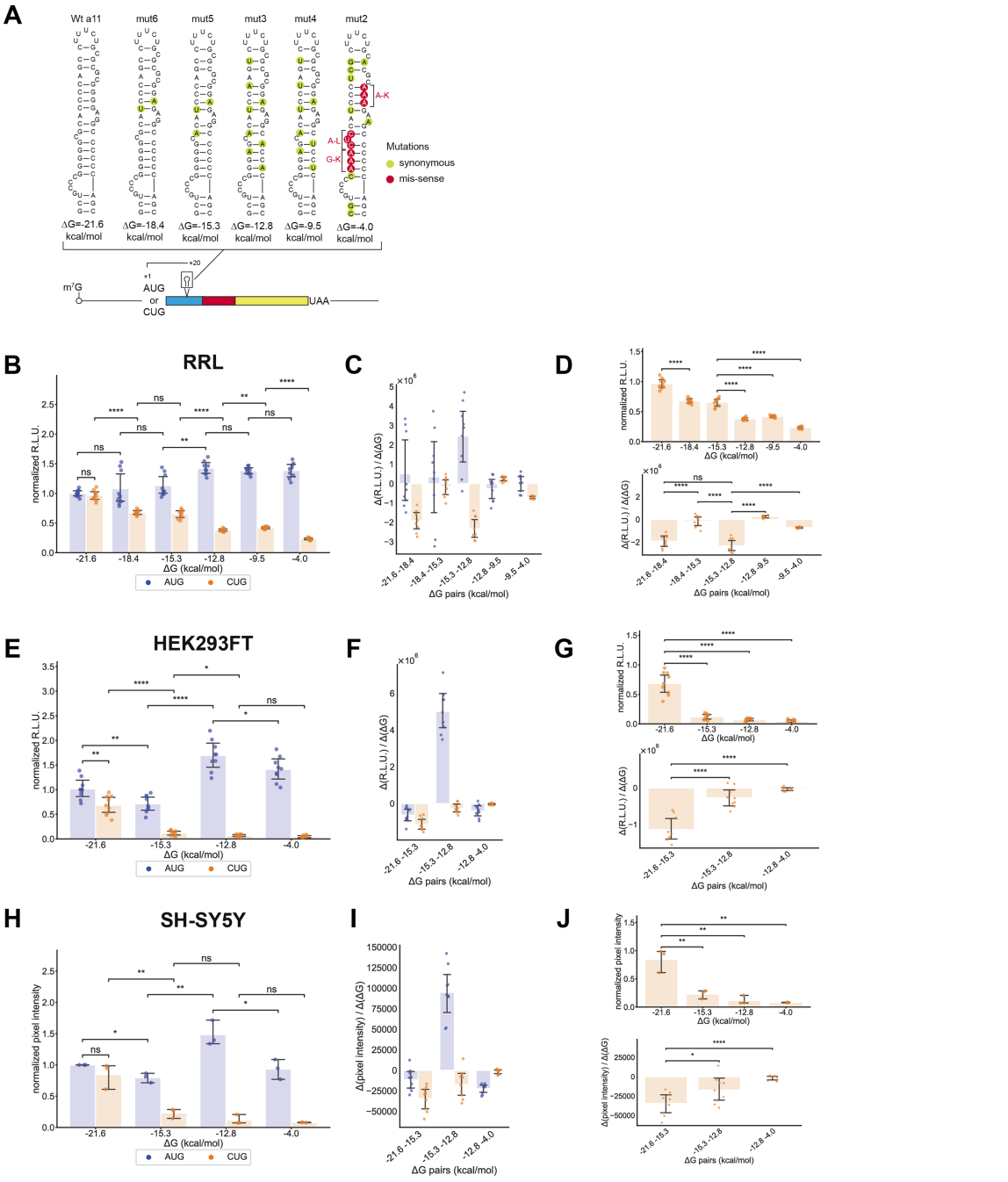

Figure 4. Secondary structure stability determines translation initiation efficiency in different cell extracts. (**A**) Schematic representation of the RNA reporter and sequences of the stem–loops used during the study. The initiation codons were AUG or CUG. Wild-type and mutant a11 hairpins were inserted at +20 position. Silent mutations introduced to lower the a11 structure stability are shown in green circles whereas missense mutations are shown in red circles. (**B**) Quantification of Renilla luciferase luminescence produced in RRL from variable a11 RNA reporters initiating with AUG or CUG codon. Relative luminescence units (R.L.U.) were normalized to those obtained with the reporter containing the wild-type a11 structure and initiating from an AUG codon. Bar heights show the mean normalized ratios and error bars show the 99% confidence intervals calculated with the *t*-distribution. *P*-values were calculated using a student t-test for independent samples. *: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$, ****: $P < 0.0001$. (**C**) The ratio $\Delta(\text{R.L.U.})/\Delta(\Delta G)$ quantifies the effect of a ΔG variation ($\Delta(\Delta G)$) on translation efficiency for AUG and CUG codons in RRL. (**D**) Stability threshold determination for initiation on a CUG codon in RRL. The top graph corresponds to the luminescence data for the CUG codon from (B), and the lower graph corresponds to the data for the CUG codon from (C). The set of pairwise comparisons presented enable to define two stability thresholds for CUG initiation at –21.6 kcal/mol and –15.3 kcal/mol. *P*-values were calculated using a Student *t*-test for independent samples. *: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$, ****: $P < 0.0001$. (**E**) Quantification of Renilla luciferase luminescence produced in HEK293FT cell extracts from variable a11 RNA reporters initiating with AUG or CUG codon. Results are presented as in (B). (**F**) Effect calculation of a ΔG variation on translation efficiency for the AUG and CUG reporters as in (C) in HEK293FT cell extracts. (**G**) Stability threshold determination for initiation on a CUG codon in HEK293FT cell extracts as in (D). The set of pairwise comparisons presented enable to define the stability threshold for CUG initiation at –21.6 kcal/mol. (**H**) Quantification of ^{35}S incorporation in Renilla luciferase proteins produced in SH-SY5Y cell extracts. Results are presented as in (B) and (E). (**I**) Effect calculation of a ΔG variation on translation efficiency for the AUG and CUG reporters as in (C) in SH-SY5Y cell extracts. (**J**) Stability threshold determination for initiation on a CUG codon in SH-SY5Y cell as in (D). The set of pairwise comparisons presented enable to define the stability threshold for CUG initiation at –21.6 kcal/mol.

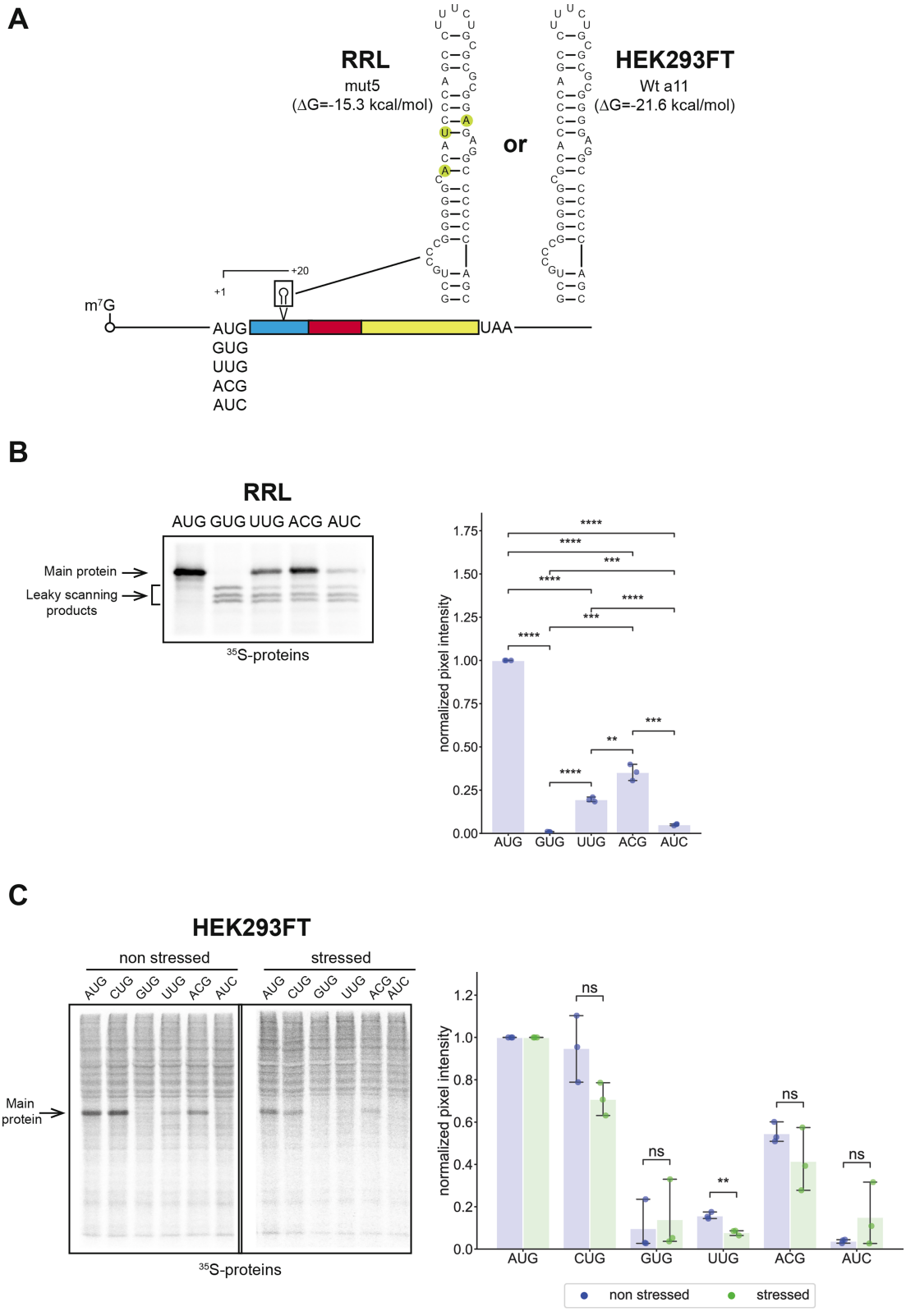

Figure 5. Stable secondary structures do not allow the initiation of translation for all AUG-like codons. (**A**) Schematic representation of the RNA reporters and sequences of the stem-loops used during the study. The initiation codons are AUG, GUG, UUG, ACG or AUC codons upstream of the mut5 a11 hairpin for RRL and wt a11 hairpin for HEK293FT which are inserted at +20 position. The nucleotide context of the initiation codons is CGUAAUNNNGAC. (**B**) Representative SDS-PAGE of *in vitro* ^{35}S radiolabeled translation products obtained with RRL and quantification of pixel intensities of the corresponding gels. Pixel intensities were normalized to those obtained with the AUG reporter. Bar heights show the mean normalized ratios and error bars show the 99% confidence intervals calculated with the *t*-distribution. *P*-values were calculated using a Student *t*-test for independent samples. *: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$, ****: $P < 0.0001$. (**C**) Representative SDS-PAGE of *in vitro* ^{35}S radiolabeled translation products obtained with unstressed and stressed HEK293FT extracts and quantification of pixel intensities of the corresponding gels. Pixel intensities were normalized to those obtained with the AUG reporter for each extract separately and results are presented as in (B).

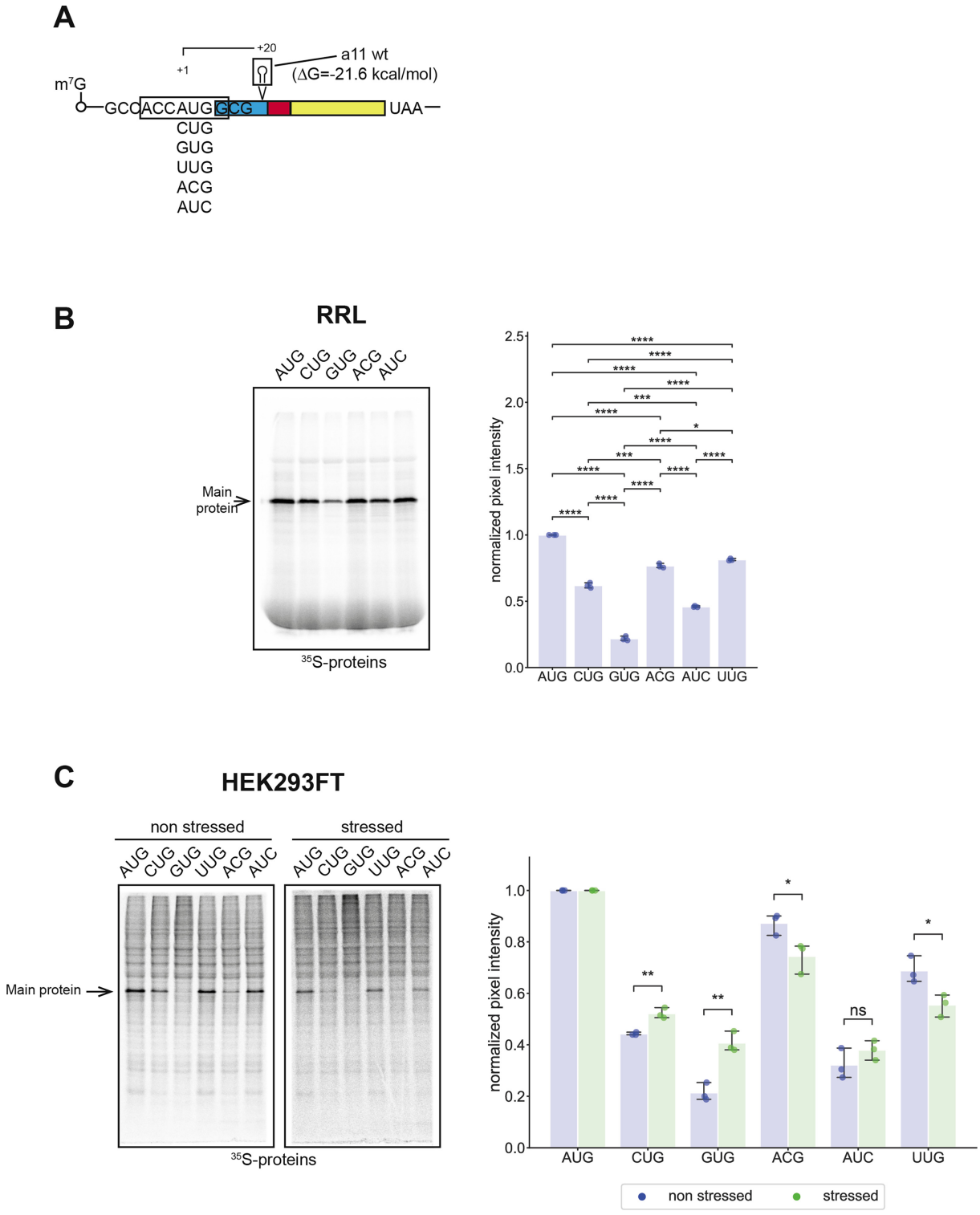

Figure 6. Changing the start-codon nucleotide context influences the efficiency of initiation rescue by START. (**A**) Schematic representation of the RNA reporters used in the study. The initiation codons are AUG, CUG, GUG, UUG, ACG or AUC codons upstream of wt a11 hairpin for RRL and HEK293FT which is inserted at +20. The nucleotide context of the tested initiation codons is GCCACCNNNGCG (boxed). (**B**) Representative SDS-PAGE of *in vitro* ^{35}S radiolabeled translation products obtained with RRL and quantification of pixel intensities of the corresponding gels. Pixel intensities were normalized to those obtained with the AUG reporter. Bar heights show the mean normalized ratios and error bars show the 99% confidence intervals calculated with the *t*-distribution. *P*-values were calculated using a Student *t*-test for independent samples. *: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$, ****: $P < 0.0001$. (**C**) Representative SDS-PAGE of *in vitro* ^{35}S radiolabeled translation products obtained with unstressed and stressed HEK293FT extracts and quantification of pixel intensities of the corresponding gels. Pixel intensities were normalized to those obtained with the AUG reporter for each extract separately and are presented as in (B).

the Kozak sequence (30,51). As expected, changing the context rescued initiation on a GUG codon because the Kozak context with a GUG does not contain an out-of-frame AUG codon. However, the GUG codon remains the least efficiently used codon (Figure 6B, C). For the other non-AUG codons, the hierarchy is different from the one established with the previous nucleotide context and is consistent regardless of the cell-free translation extract used for *in vitro* translation: AUG > UUG ≈ AUC>>CUG > ACG > GUG.

Secondary structures are found downstream of alternative translation initiation sites in human transcripts

To evaluate the involvement of the START mechanism in non-AUG translation in human cells, we investigated genome-wide the presence of RNA structures located at the appropriate distance from start codons. To do that, we used the previously described secondary structure prediction algorithm (35) to search for such structures at alternative translation initiation sites in a set of translated human small ORFs (smORFs: upstream ORFs, downstream ORFs and 'non-coding' ORFs) identified by ribosome profiling (28). Briefly, we searched for secondary structures within the +15/+65 window downstream of these ORFs start site, with AUG and non-AUG initiation codon.

Among the 5743 recovered smORFs (i.e. where we could recover the sequence of +16/+65 region) from the dataset of Chothani *et al*. (Supplementary File 2), we found moderately stable downstream structures ($\triangle G$ between −15.6 and −7.1 kcal/mol) in 50% of the analyzed ORFs (Figure 7A). In addition, 50% of these structures have moderate to high predicted stability (more stable than −10.9 kcal/mol). Of note 25% are more stable than −15.6 kcal/mol, which approximatively corresponds to the lower stability threshold that we found in RRL for efficient initiation on a CUG codon (Figure 7A). These overall statistics suggest that the majority of the translated smORFs (at least 75%) do not contain RNA structures sufficiently stable on their own to promote START according to our experimental data within the +16/+65 window. If START is involved in the initiation of these ORFs, it is likely with the assistance of RNA binding proteins involved in stabilizing those moderately stable RNA structures.

Furthermore, the presence of stable secondary structures is not specific to whether initiation codons are AUG or non-AUG, as their distributions are similar (Figure 7B). Although the calculated p-values would rather suggest the opposite, effect sizes (probabilities of superiority and Hodges–Lehmann estimator) are very mild as they are in a similar range of those observed when comparing for instance the AUG and CUG distributions (Figure 7D) where p-values clearly do not provide evidence that these distributions and their medians are different (Figure 7C). In other words, although p-values suggest that the AUG and non-AUG distributions are likely different, effect sizes calculations show that the magnitude of that difference is not significant. More generally, we found that 29.3% and 26% of the smORFs starting with an AUG or a non-AUG respectively contain secondary structures more stable than −15 kcal/mol (Figure 7B). Even when analyzing each start codon individually, the distributions are very similar. (Figure 7C). Overall, the main differences between codons lie in the most negative mfe values. Of note, ACG is the start codon with the highest proportion (40.7%) of downstream structures more stable than −15 kcal/mol (Figure 7C), then followed by CUG, GUG (37.2% for both) and AUG (29.3%). Although the AUG codon has more outlier mfe values than the other codons, this is most likely due to a larger population size, which ultimately increases the counts of rare events such as finding very stable downstream structures. The same would also apply, to a lesser extent, for the CUG and GUG codons. More precisely, comparing the density distributions of structure stabilities of all non-AUG codons to the AUG codon shows that they can be classified into three groups (Figure 7D). The first group (boxed in blue) includes the distributions of the CUG, GUG and AGG codons, which cannot be considered different from that of AUG after *P*-values calculations. Of note, the mfe distributions of CUG and GUG codons almost perfectly overlap with AUG codons. For the AGG codon, it is not clear whether the lack of overlap of its distribution with the AUG's for the most negative mfe values (i.e. more stable structures) is due to a reduced sample size compared to CUG or GUG codons. The discovery of more ORFs starting with AGG codons will help resolve the left tail of the AGG mfe distribution and reveal whether it is similar to the AUG's. This actually applies to all studied non-AUG codons. The second group (boxed in green) includes the distributions of the UUG, AAG and AUN (N≠G) codons which are likely different from the AUG mfe distribution as suggested by the *P*-values. According to the comparison of the medians through Mood's *P*-values and the Hodges–Lehmann effect size, which is at least 3 times higher than those calculated for the first group, the central tendency of the mfe distributions of these non-AUG codons is shifted to less stable structures, which would indicate that those non-AUG codons have less stable downstream structures than AUG. This is further supported by the lower probabilities of superiority calculated for this group compared to the first group, as they indicate that a randomly selected AUG smORF is less likely to have a downstream structure with a mfe superior (i.e. less stable) to the mfe of downstream structure from a randomly selected UUG, AAG or AUN (N≠G) smORF. The last group only contains only ACG codon (boxed in red). The same analysis suggests that the ACG distribution is shifted to lower mfe values (more stable structures) than AUG, although the probability of superiority is not that high. Nonetheless, contrary to the other codons, the strongest overlap of its distribution with the AUG's is observed for the most negative mfe values (most stable structures) despite a total number of ACG-starting ORFs (167) similar to the previously analyzed AUN (N≠G) codons (113, 163). In conclusion, apart from the striking exception of ACG codons, the presence of stable structures is not more biased for specific non-AUG than for AUG initiation codons in the analyzed smORFs. AUN (N≠G) codons along with UUG and AAG codons may even feature less stable downstream structures than AUGs.

Finally, when looking at all the found secondary structures, no explicit GO term is specific for transcripts with secondary structure stabilities within the first three quartiles (Supplementary Figure S3 and Supplementary File 3), suggesting that all biological processes may be regulated by a suboptimal START mechanism. However, the fourth quartile, which contains secondary structures likely able to elicit START on their own (more stable than −15.6 kcal/mol), shows specific GO terms linked to the regulation of transcription by RNA polymerase II, signaling cascades, and protein maturation (Supplementary Figure S3E).

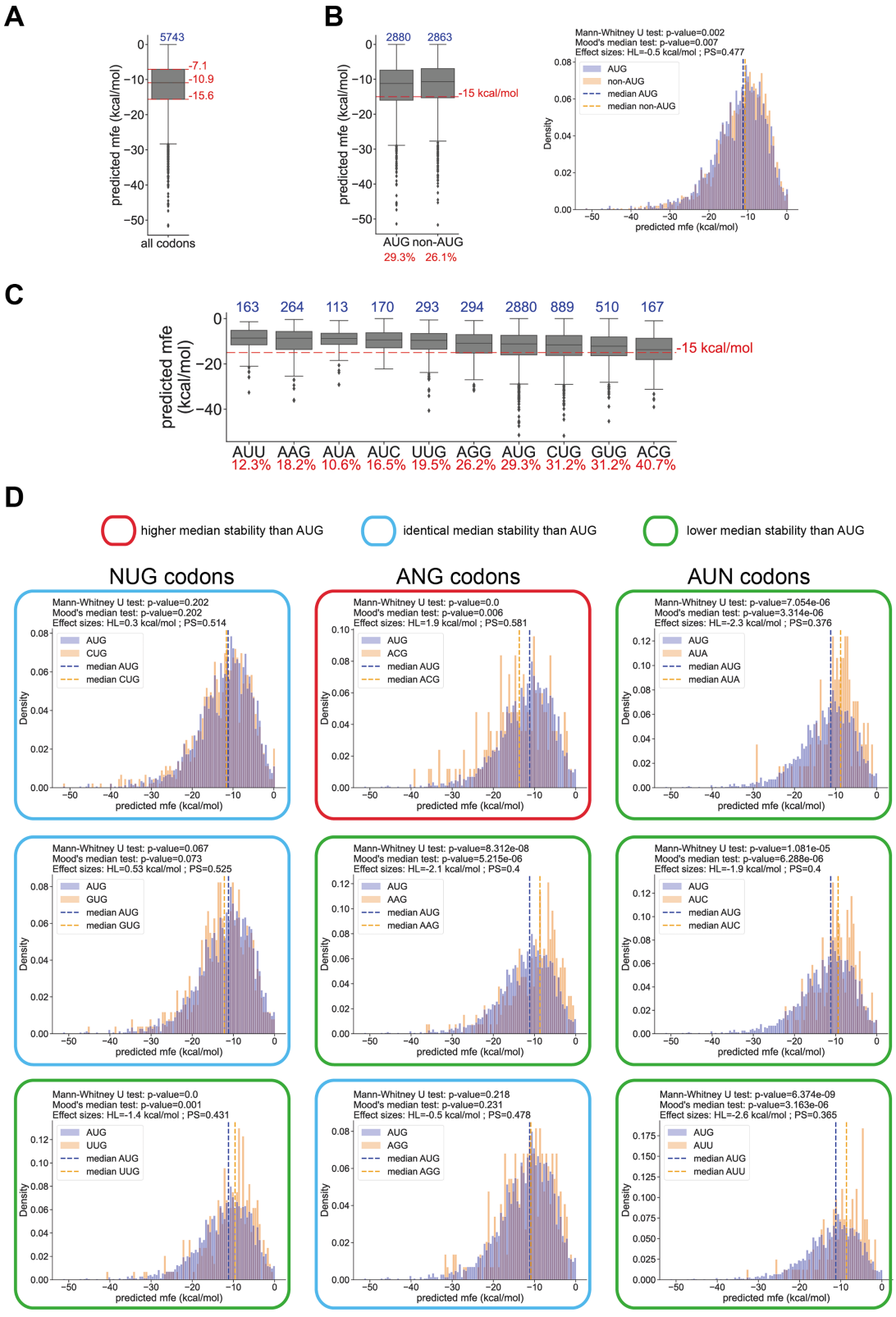

Figure 7. Analysis of ΔG distributions of downstream secondary structures within the ΔG (or mfe = minimum free energy) distributions of downstream secondary structures within +16/+65 for all codons (**A**), for AUG and non-AUG codons (**B**) and for each codon individually (**C**). Top blue numbers show the population size of each boxplot. Numbers in red show the percentage of smORFs starting with the corresponding start codon that contain a structure more stable than –15 kcal/mol (i.e. mfe inferior to –15 kcal/mol). Histograms in (B) and (D) show the density distributions corresponding to the boxplots. Histograms have a bin width of 0.5 kcal/mol, and the total area of each histogram is normalized to 1. The Mann–Whitney *U* test was used to determine if the AUG and the non-AUG distributions are equal. The Mood's median test was used to determine if the AUG and the non-AUG distributions have the same median. The probability of superiority (PS) and the Hodges–Lehmann estimator (HL) were used for effect size assessment (see Methods section).

Translation of four identified human smORF that initiate at a non-AUG codon

We then arbitrarily chose a few candidates among the transcripts having a smORF starting with a non-AUG codon and a secondary structure more stable than –15 kcal/mol to check whether their downstream structure is indeed required for translation initiation. We chose the –15 kcal/mol threshold stability according to the ones determined by *in vitro* translation assay (Figure 4), although it cannot be excluded that some RNA structures may be differentially modified *in vivo* by specific RNA-binding proteins. The selected transcripts were FAM179B, DDHD2, USP9X and RPTOR. The construction scheme for the corresponding reporter RNAs was similar to the previous ones but with the flanking sequences of the specific smORFs and their own initiation codons (Figure 8A, Supplementary File 1). The non-AUG start codons were AGG, GUG, AUA and CUG. We also designed corresponding mutated structures by introducing destabilizing silent mutations (Figure 8B). The *in vitro* translation patterns were reminiscent of those observed with the mutated versions of the a11 structure. First, we found that the four non-AUG codons tested were efficiently recognized as start codons in the cell-free extracts. Of note, translation efficiency in HEK cell extracts is more sensitive to secondary structure stability than in RRL, with drastic decreases in the translation of reporter RNAs with mutated secondary structures (Figure 8C, D). This suggests the involvement of *trans*-acting factors that could be different from one cell extract to another. We also observe that in the case of FAM179B in RRL, the presence of the mutated structure triggers significant leaky scanning, indicating that the wild-type structure is indeed required for the accuracy of translation initiation on AGG codon from FAM179B. In the case of DDHD2, translation in RRL is more efficient when the structure is mutated suggesting that the structure has a negative effect (Figure 8E), possibly in a similar way to what was observed in Figure 4B with the AUG codon for RRL. Overall, in HEK extracts, initiation on non-AUG codons strictly requires the presence of the wild-type secondary structure (Figure 8F). In addition, there is a strong decrease in the translation efficiency of FAM179B and USP9X RNAs relative to RPTOR RNA translation, compared to the relative efficiencies measured in RRL. Altogether, these experiments suggest that downstream structures do impact the initiation on the AUG-like codons and this effect is modulated by the cell's content in *trans*-acting factors.

Discussion

Our data demonstrates that the START mechanism requires an RNA structure with a cell-type dependent minimal stability to promote translation initiation on AUG-like codons and further supports the initial work on that topic (22). Considering the length of ribosome footprints provided by toe-printing and ribosome profiling data (approx. 30 nucleotides around the AUG start codon), we expected the optimal distance between the structure and the start codon to be at least 15–16 nucleotides or slightly higher. Indeed, a structure that is too close to the start codon would interfere with the progression of the scanning complex before reaching the start codon and would consequently inhibit translation initiation. In contrast, a structure located too far away from the start codon would not interact with the scanning complex and consequently lead to leaky-scanning. We found that the optimal distance to favor initiation ranges from 20 to 23 nucleotides downstream of a CUG codon, which is longer than the distance of 14 nucleotides previously described (22). As a possible explanation, it can be assumed that the optimal distance also depends on the tri-dimensional shape of the structure. A small stable structure with few base pairs might require a closer distance to the start codon than one with a more sophisticated fold that could for instance create a kink in the structure pointing toward the scanning complex. This possibility requires more investigation. Finally, the sequences of the structure loop itself may also have their own importance as they may interact specifically with either ribosomal components (ribosomal proteins or rRNA) and/or *trans*-acting factors. As an example, such interactions between helix h16 of the rRNA and nucleotides at positions +17 to +19 located immediately upstream of a stable structure have been observed in the case of histone H4 mRNA (52).

The START model implies that the stability of the structure (whether it is associated or not with an RNA-binding protein) must be high enough to stall or at least slow down a scanning complex in order to promote 80S assembly on the codon in the P-site. Therefore we chose as a model RNA structure the a11 GC-rich structure (-21 kcal/mol), as it was previously shown to promote stalling of the scanning pre-initiation complex (47). Using this structure as a starting point, we determined that the threshold stability was close to –15 kcal/mol for initiation on CUG codons with RRL, and approximately –20 kcal/mol with extracts from HEK293FT and SHSY5Y cells. These thresholds also allowed initiation on GUG, UUG, ACG and AUC codons, but with distinct efficiencies. Therefore, other parameters are critical such as the nucleotide context, specific *trans*-acting factors, and base-pairing ability with the initiator $tRNA^{Met}$. For instance, changing the start codon nucleotide context resulted in two distinct hierarchies of translation initiation efficiencies among AUG-like codons. Accordingly, previous reports also established that initiation on AUG-like codons is influenced by their nucleotide context similarly to AUG codon, with the average preferred nucleotide context being the Kozak context (53–55,51).

The AUG-like codon hierarchies we have established using two different start codon nucleotide contexts are rather consistent with the usage of those codons determined by ribosome profiling (26–28). One striking exception is the GUG codon that is efficiently used in those experiments but not in ours. Thermodynamic studies of the codon-anticodon interaction with a GUG codon demonstrated an increased so-called 'free-energy penalty' due to eIF1 and eIF1A interactions that further impairs initiation on GUG codons but not on CUG codons (56). Therefore, initiation on GUG codons is more stringent than on CUG codons and requires more free-energy compensation. Such compensation might come either from a more stable downstream secondary structure and its interaction with the initiation complex, an optimal nucleotide environment, and/or from various *trans*-acting factors that either modulate the structure stability (helicases, RNA-binding proteins) or influence start codon selection stringency (such as cellular concentrations of eIF1 (8) or the balance between eIF5-mimic protein 5MP1/5MP2 and eIF5 (23,57)). Alternatively, Leu-$tRNA_{CAG}$ and Val-$tRNA_{CAC}$ have been described as alternative initiator tRNAs and could be used to initiate translation from CUG and GUG codons respectively under certain stress circumstances through mechanisms that involve

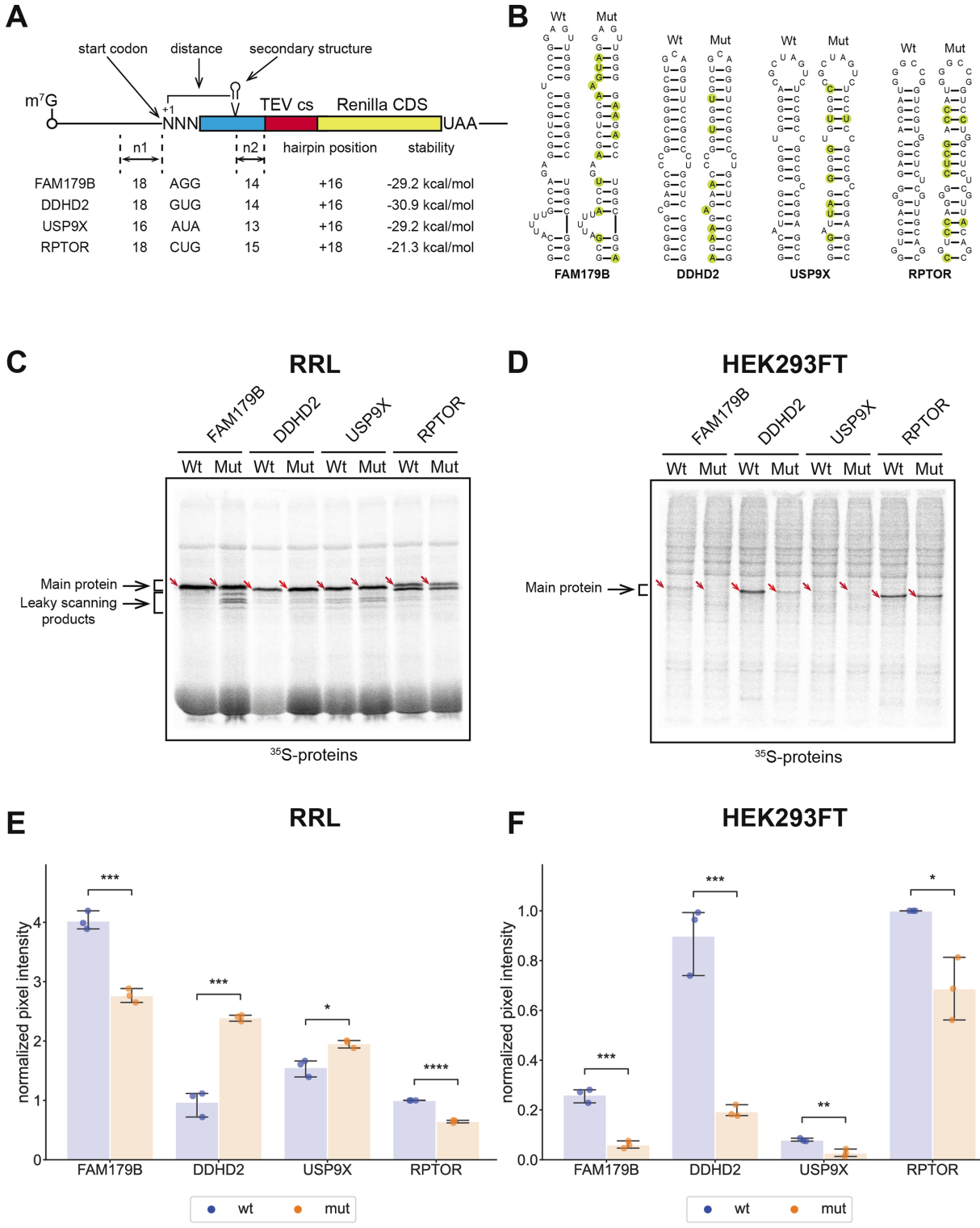

Figure 8. *In vitro* translation of RNA reporters carrying secondary structures from four identified human smORFs using a non-AUG codon. (**A**) Schematic representation of the four RNA reporters FAM179B, DDHD2, USP9X and RPTOR. The stability of the respective structures is indicated as well as the number of nucleotides upstream of the start codon that originates from the human transcript (n1), the start codon, the position of the secondary structure, the predicted stability of the secondary structure and the number of nucleotides downstream the structure that originates from the human transcript (n2). (**B**) Representations of the secondary structures of each reporter studied. The silent mutations that were introduced to lower the structure stability are shown in green circles in the Mut stem–loops. (**C**) Representative SDS-PAGE of *in vitro* ^{35}S radiolabeled translation products obtained with FAM179B, DDHD2, USP9X, RPTOR wild-type and Mut reporters in RRL. The positions of the full-length fusion proteins are shown by red arrows. Lower bands correspond to alternative translation initiation sites resulting from a leaky scanning of the main start codon. In the case of the RPTOR reporter, there is an in-frame AUG at +34. (**D**) Same as (C) but in HEK293FT extracts. (**E**, **F**) Quantification of ^{35}S incorporation in Renilla luciferase proteins produced in RRL and HEK293FT extracts. Bar heights show the mean values and error bars show the 99% confidence intervals calculated with the *t*-distribution. *P*-values were calculated using a Student *t*-test for independent samples. *: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$, ****: $P < 0.0001$.

non-canonical initiation factors eIF2A or eIF2D (39). All these variables are likely to explain the differences in the efficiencies of non-AUG initiation that are reported here compared to previous reports (21,53,58), especially considering that non-AUG initiation is suboptimal and would be therefore be much more sensitive than an AUG codon to variations in factors influencing start codon recognition.

Translation initiation on AUG-like codons was assessed in three types of cell extracts, which resulted in distinct required minimal structure stabilities. First, this suggests that a stable downstream secondary structure is a critical parameter influencing translation initiation on AUG-like codons. Since the optimal structure stability for efficient initiation on AUG-like codons is dependent on the type of extract used, we further suggest that the involvement of cell-specific *trans*-acting factors cannot be overlooked. Among these are initiation factors eIF1 and eIF1A which control the codon-anticodon interactions and start codon selection stringency (6,8), RNA helicases whose activities were shown to decrease uORFs translation in yeast (32) or RNA binding proteins (36,37). In particular, the RNA helicase content of the cell would determine the stability range of secondary structures that can be unwound and therefore modulate the use of the START mechanism. A precise assessment of the RNA helicase content in distinct cell types would then be needed to better understand the propensity of these cells to initiate translation from AUG-like codons by a START mechanism.

Along with the determination of the RNA helicase content, secondary structures predictions in 5′UTRs of mRNAs may provide an important tool to predict the use of uORFs with AUG-like codons and unveil potential new ORFs. In a previous study, we demonstrated the broad presence of downstream RNA structures in annotated main open reading frames of various organisms (35). Here, we used the same algorithm on the small ORF dataset of Chothani *et al.* obtained by ribosome profiling of human primary cells and tissues to analyze downstream secondary structures at alternative translation initiation sites. Among the selected structures, 25% are more stable than –15.6 kcal/mol and half of them are between –7.1 and –15.6 kcal/mol. Because –15.3 kcal/mol is the lower threshold stability characterized in RRL to promote START initiation on a CUG codon (this threshold stability being even 'higher' in human cell extracts, –21.6 kcal/mol), this would indicate that <25% of the identified smORFs are initiated by a START mechanism based on the sole presence of a downstream structure. Moreover, the presence of stable downstream structures is not more biased for specific non-AUG than for AUG start codons, apart from the notable exception of ACG start codons. Given that 75% of the identified downstream structures have a moderate to high stability (at least more stable than –7.1 kcal/mol), we hypothesize that the absence of '*cis*-correlation' between the nature of the start codon and stable downstream structures enables fine-tuning of START by *trans*-acting factors depending on the physiological conditions. Although most of these structures are not sufficiently stable to promote START on their own, the energetic contribution brought by these suboptimal START may nonetheless facilitate non-AUG recognition with the assistance of *trans*-acting factors that either stabilize those structures (RNA binding proteins), decrease start-codon selection stringency (increase in eIF5 or decrease in eIF1 cellular concentrations), or reduce the activity of scanning-associated RNA helicases. The regulation dynamic of these factors depending on the cell physiological status remains to be elucidated and would help quantify the involvement of START in translation initiation from specific start sites in mRNAs. Furthermore, GO terms enrichment analysis did not show any clear link between transcripts featuring stable downstream structures and subsets of biological processes, although we observed an enrichment of GO terms linked to transcription regulation and signaling cascades among transcripts featuring the most stable structures (more stable than –15.6 kcal/mol). Overall, this suggests that downstream structures may regulate a broad range of cellular functions.

However, assigning a precise molecular role to these downstream RNA structures requires further investigation. It could be assumed on one hand that some potential start codons have a rather suboptimal nucleotide context that would promote leaky scanning and therefore require a downstream structure to promote initiation. These downstream structures would be modulated by other *trans*-factors that control the structure stability and therefore translation initiation, as it has been observed in yeast (32). Integrating these other factors into uORFs translation predictions remains a real challenge, as RNA remodeling is a dynamic process that relies on many factors other than the RNA sequence itself. On the other hand, some of these structures may not be linked to a START mechanism at all and may be involved in other molecular mechanisms, such as modulating mRNA half-life. Therefore, which percentage of the identified downstream structures is linked to START remains an open question.

Altogether, we propose that the ubiquity of moderately stable downstream structures at alternative translation initiation sites in human RNAs has evolved as a mechanism for controlling translation initiation through the conditional expression of RNA helicases or RNA-binding proteins, depending on cell-type.

Conclusion

Here, we provide further insights into *cis*-regulatory parameters that influence start codon recognition. Downstream secondary structures can efficiently promote AUG-like translation initiation if they have the required stability to stall a scanning 43S particle. These downstream structures must be located at an optimal distance from this AUG-like codon to trigger codon-anticodon base-pairing in the P site. This is achieved by providing an energetic contribution to the 43S particle with suboptimal AUG-like codons in the P-site. The START mechanism may also rescue translation initiation on start codons with a suboptimal nucleotide context that would otherwise lead to leaky scanning. A better understanding of how frequently this atypical initiation mechanism is used in living cells might help to predict more accurately translation initiation sites in cellular mRNAs and give a more precise picture of the complete proteome. Overall, adjusting the stability of RNA structures downstream of potential start codons by the expression level of specific RNA helicases or RNA-binding proteins in various cell types is a way to turn on and off translation initiation on specific codons depending on the physiological conditions.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and / or its supplementary data.

Supplementary data

Supplementary Data are available at NARGAB Online.

Acknowledgements

This work is funded by *Agence Nationale pour la Recherche* (ANR-17-CE12-0025-01, ANR-17-CE11-0024, ANR-20-COVI-0078, ANR-21-CE12-0028-01, ANR-21PRRD-0001-01), by *Fondation pour la Recherche Médicale* (project CoronaIRES, ALZ201912009641 and MND202310017885), by *Fondation Bettencourt*, by University of Strasbourg and by the *Centre National de la Recherche Scientifique*. We would also like to thank Dr Sebastian Schafer for kindly providing us with DNA sequences of the 7767 high-confidence smORFs detected in the human genome (28).

Author contributions: A.T. designed the experimental strategy of this study, performed most of the experiments, created scripts to analyze the data, interpreted the data and wrote the manuscript. F.A. performed the previous experiments on hox a11 that enabled this study and proofread the manuscript. L.D. created and used the scripts required for the screening of the smORF database, G.E. interpreted the data and participated to the writing of the manuscript, F.M. designed the experimental strategy, supervised the conducted experiments, interpreted the data and wrote the manuscript.

Funding

Fondation pour la Recherche Médicale [ALZ201912009641, MND202310017885]; Agence Nationale pour la Recherche [ANR-17-CE11-0024, ANR-17-CE12-0025-01, ANR-20-COVI-0078, ANR-21-CE12-0028-01, ANR-21PRRD-0001-01]; Centre National de la Recherche Scientifique; University of Strasbourg; Fondation Bettencourt.

Conflict of interest statement

None declared.

References

1. Tidu,A. and Martin,F. (2023) The interplay between cis- and trans-acting factors drives selective mRNA translation initiation in eukaryotes. *Biochimie*, **217**, 20–30.
2. Rogers,G.W., Richter,N.J., Lima,W.F. and Merrick,W.C. (2001) Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.*, **276**, 30914–30922.
3. Schütz,P., Bumann,M., Oberholzer,A.E., Bieniossek,C., Trachsel,H., Altmann,M. and Baumann,U. (2008) Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein–protein interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9564–9569.
4. Passmore,L.A., Schmeing,T.M., Maag,D., Applefield,D.J., Acker,M.G., Algire,M.A., Lorsch,J.R. and Ramakrishnan,V. (2007) The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell*, **26**, 41–50.
5. Valášek,L., Nielsen,K.H., Zhang,F., Fekete,C.A. and Hinnebusch,A.G. (2004) Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection. *Mol. Cell. Biol.*, **24**, 9437–9455.
6. Pestova,T.V., Borukhov,S.I. and Hellen,C.U.T. (1998) Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature*, **394**, 854–859.
7. Thakur,A. and Hinnebusch,A.G. (2018) 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.
8. Zhou,F., Zhang,H., Kulkarni,S.D., Lorsch,J.R. and Hinnebusch,A.G. (2020) eIF1 discriminates against suboptimal initiation sites to prevent excessive uORF translation genome-wide. *RNA*, **26**, 419–438.
9. Lomakin,I.B. and Steitz,T.A. (2013) The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature*, **500**, 307–311.
10. Paulin,F.E.M., Campbell,L.E., O'Brien,K., Loughlin,J. and Proud,C.G. (2001) Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Curr. Biol.*, **11**, 55–59.
11. Pestova,T.V. and Kolupaeva,V.G. (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.*, **16**, 2906–2922.
12. Unbehaun,A., Borukhov,S.I., Hellen,C.U.T. and Pestova,T.V. (2004) Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev.*, **18**, 3078–3093.
13. Maag,D., Fekete,C.A., Gryczynski,Z. and Lorsch,J.R. (2005) A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol. Cell*, **17**, 265–275.
14. Kapp,L.D. and Lorsch,J.R. (2004) GTP-dependent recognition of the methionine moiety on initiator tRNA by translation factor eIF2. *J. Mol. Biol.*, **335**, 923–936.
15. Yi,S.-H., Petrychenko,V., Schliep,J.E., Goyal,A., Linden,A., Chari,A., Urlaub,H., Stark,H., Rodnina,M.V., Adio,S., *et al.* (2022) Conformational rearrangements upon start codon recognition in human 48S translation initiation complex. *Nucleic Acids Res.*, **50**, 5282–5298.
16. Algire,M.A., Maag,D. and Lorsch,J.R. (2005) Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell*, **20**, 251–262.
17. Asano,K., Clayton,J., Shalev,A. and Hinnebusch,A.G. (2000) A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA ^{Met} is an important translation initiation intermediate in vivo. *Genes Dev.*, **14**, 2534–2546.
18. Hinnebusch,A.G. (2017) Structural insights into the mechanism of scanning and start codon recognition in eukaryotic translation initiation. *Trends Biochem. Sci.*, **42**, 589–611.
19. Pestova,T.V., Lomakin,I.B., Lee,J.H., Choi,S.K., Dever,T.E. and Hellen,C.U.T. (2000) The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature*, **403**, 332–335.
20. Nag,N., Lin,K.Y., Edmonds,K.A., Yu,J., Nadkarni,D., Marintcheva,B. and Marintchev,A. (2016) eIF1A/eIF5B interaction network and its functions in translation initiation complex assembly and remodeling. *Nucleic Acids Res.*, **44**, 7441–7456.
21. Peabody,D.S. (1989) Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.*, **264**, 5031–5035.
22. Kozak,M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci.*, **87**, 8301–8305.
23. Tang,L., Morris,J., Wan,J., Moore,C., Fujita,Y., Gillaspie,S., Aube,E., Nanda,J., Marques,M., Jangal,M., *et al.* (2017) Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. *Nucleic Acids Res.*, **45**, 11941–11953.
24. Liang,H., Chen,X., Yin,Q., Ruan,D., Zhao,X., Zhang,C., McNutt,M.A. and Yin,Y. (2017) PTENβ is an alternatively translated isoform of PTEN that regulates rDNA transcription. *Nat. Commun.*, **8**, 14771.
25. Tabet,R., Schaeffer,L., Freyermuth,F., Jambeau,M., Workman,M., Lee,C.-Z., Lin,C.-C., Jiang,J., Jansen-West,K., Abou-Hamdan,H., *et al.* (2018) CUG initiation and frameshifting enable production

of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat. Commun.*, **9**, 152.
26. Ingolia,N.T., Lareau,L.F. and Weissman,J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.
27. Chen,J., Brunner,A.-D., Cogan,J.Z., Nuñez,J.K., Fields,A.P., Adamson,B., Itzhak,D.N., Li,J.Y., Mann,M., Leonetti,M.D., *et al.* (2020) Pervasive functional translation of noncanonical human open reading frames. *Science*, **367**, 1140–1146.
28. Chothani,S.P., Adami,E., Widjaja,A.A., Langley,S.R., Viswanathan,S., Pua,C.J., Zhihao,N.T., Harmston,N., D'Agostino,G., Whiffin,N., *et al.* (2022) A high-resolution map of human RNA translation. *Mol. Cell*, **82**, 2885–2899.
29. Kozak,M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283–292.
30. Kozak,M. (1987) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.*, **196**, 947–950.
31. Wang,J., Shin,B.-S., Alvarado,C., Kim,J.-R., Bohlen,J., Dever,T.E. and Puglisi,J.D. (2022) Rapid 40S scanning and its regulation by mRNA structure during eukaryotic translation initiation. *Cell*, **185**, 4474–4487.
32. Guenther,U.-P., Weinberg,D.E., Zubradt,M.M., Tedeschi,F.A., Stawicki,B.N., Zagore,L.L., Brar,G.A., Licatalosi,D.D., Bartel,D.P., Weissman,J.S., *et al.* (2018) The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs. *Nature*, **559**, 130–134.
33. Kochetov,A.V., Palyanov,A., Titov,I.I., Grigorovich,D., Sarai,A. and Kolchanov,N.A. (2007) AUG_hairpin: prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinf.*, **8**, 318.
34. Robbins-Pianka,A., Rice,M.D. and Weir,M.P. (2010) The mRNA landscape at yeast translation initiation sites. *Bioinforma. Oxf. Engl.*, **26**, 2651–2655.
35. Despons,L. and Martin,F. (2020) How many messenger RNAs can Be translated by the START mechanism? *Int. J. Mol. Sci.*, **21**, 8373.
36. Beckmann,K., Grskovic,M., Gebauer,F. and Hentze,M.W. (2005) A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in drosophila. *Cell*, **122**, 529–540.
37. Medenbach,J., Seiler,M. and Hentze,M.W. (2011) Translational control via protein-regulated upstream open reading frames. *Cell*, **145**, 902–913.
38. Eriani,G. and Martin,F. (2018) START: sTructure-assisted RNA translation. *RNA Biol.*, **15**, 1250–1253.
39. Kearse,M.G. and Wilusz,J.E. (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.*, **31**, 1717–1731.
40. Andreev,D.E., Loughran,G., Fedorova,A.D., Mikhaylova,M.S., Shatsky,I.N. and Baranov,P.V. (2022) Non-AUG translation initiation in mammals. *Genome Biol.*, **23**, 111.
41. Pelham,H.R.B. and Jackson,R.J. (1976) An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.*, **67**, 247–256.
42. Kozutsumi,Y., Segal,M., Normington,K., Gething,M.-J. and Sambrook,J. (1988) The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins. *Nature*, **332**, 462–464.
43. Hetz,C., Zhang,K. and Kaufman,R.J. (2020) Mechanisms, regulation and functions of the unfolded protein response. *Nat. Rev. Mol. Cell Biol.*, **21**, 421–438.
44. Harding,H.P., Zhang,Y., Bertolotti,A., Zeng,H. and Ron,D. (2000) Perk is essential for translational regulation and cell survival during the unfolded protein response. *Mol. Cell*, **5**, 897–904.
45. Scheuner,D., Song,B., McEwen,E., Liu,C., Laybutt,R., Gillespie,P., Saunders,T., Bonner-Weir,S. and Kaufman,R.J. (2001) Translational control is required for the unfolded protein response and in vivo glucose homeostasis. *Mol. Cell*, **7**, 1165–1176.
46. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
47. Alghoul,F., Laure,S., Eriani,G. and Martin,F. (2021) Translation inhibitory elements from Hoxa3 and Hoxa11 mRNAs use uORFs for translation inhibition. *eLife*, **10**, e66369.
48. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
49. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
50. Klopfenstein,D.V., Zhang,L., Pedersen,B.S., Ramírez,F., Warwick Vesztrocy,A., Naldi,A., Mungall,C.J., Yunes,J.M., Botvinnik,O., Weigel,M., *et al.* (2018) GOATOOLS: a Python library for gene ontology analyses. *Sci. Rep.*, **8**, 10872.
51. Hernández,G., Osnaya,V.G. and Pérez-Martínez,X. (2019) Conservation and variability of the AUG initiation codon context in eukaryotes. *Trends Biochem. Sci.*, **44**, 1009–1021.
52. Martin,F., Ménétret,J.-F., Simonetti,A., Myasnikov,A.G., Vicens,Q., Prongidi-Fix,L., Natchiar,S.K., Klaholz,B.P. and Eriani,G. (2016) Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat. Commun.*, **7**, 12622.
53. Ivanov,I.P., Loughran,G., Sachs,M.S. and Atkins,J.F. (2010) Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 18056–18060.
54. Ivanov,I.P., Firth,A.E., Michel,A.M., Atkins,J.F. and Baranov,P.V. (2011) Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.*, **39**, 4220–4234.
55. Diaz de Arce,A.J., Noderer,W.L. and Wang,C.L. (2018) Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic Acids Res.*, **46**, 985–994.
56. Lind,C. and Åqvist,J. (2016) Principles of start codon recognition in eukaryotic translation initiation. *Nucleic Acids Res.*, **44**, 8425–8432.
57. Singh,C.R., Watanabe,R., Zhou,D., Jennings,M.D., Fukao,A., Lee,B., Ikeda,Y., Chiorini,J.A., Campbell,S.G., Ashe,M.P., *et al.* (2011) Mechanisms of translational regulation by a human eIF5-mimic protein. *Nucleic Acids Res.*, **39**, 8314–8328.
58. Wei,J., Zhang,Y., Ivanov,I.P. and Sachs,M.S. (2013) The stringency of start codon selection in the filamentous fungus *Neurospora crassa*. *J. Biol. Chem.*, **288**, 9549–9562.

Received: March 6, 2024. **Revised:** April 18, 2024. **Editorial Decision:** May 22, 2024. **Accepted:** May 23, 2024