

In-depth analysis of *cis*-determinants that either promote or inhibit reinitiation on *GCN4* mRNA after translation of its four short uORFs

STANISLAVA GUNIŠOVÁ, PETRA BEZNOSKOVÁ, MAHABUB PASHA MOHAMMAD, VLADISLAVA VLČKOVÁ, and LEOŠ SHIVAYA VALÁŠEK

Laboratory of Regulation of Gene Expression, Institute of Microbiology AS CR, Prague 142 20, Czech Republic

ABSTRACT

Translational control in eukaryotes is exerted by many means, one of which involves a ribosome translating multiple cistrons per mRNA as in bacteria. It is called reinitiation (REI) and occurs on mRNAs where the main ORF is preceded by a short upstream uORF(s). Some uORFs support efficient REI on downstream cistrons, whereas some others do not. The mRNA of yeast transcriptional activator *GCN4* contains four uORFs of both types that together compose an intriguing regulatory mechanism of its expression responding to nutrients' availability and various stresses. Here we subjected all *GCN4* uORFs to a comprehensive analysis to identify all REI-promoting and inhibiting *cis*-determinants that contribute either autonomously or in synergy to the overall efficiency of REI on *GCN4*. We found that the 3' sequences of uORFs 1–3 contain a conserved AU₁₋₂A/UUAU₂ motif that promotes REI in position-specific, autonomous fashion such as the REI-promoting elements occurring in 5' sequences of uORF1 and uORF2. We also identified autonomous and transferable REI-inhibiting elements in the 3' sequences of uORF2 and uORF3, immediately following their AU-rich motif. Furthermore, we analyzed contributions of coding triplets and terminating stop codon tetranucleotides of *GCN4* uORFs showing a negative correlation between the efficiency of reinitiation and efficiency of translation termination. Together we provide a complex overview of all *cis*-determinants of REI with their effects set in the context of the overall *GCN4* translational control.

Keywords: translational control; reinitiation; uORF; *cis*-regulation; *GCN4*

INTRODUCTION

Translational control of one of yeast's most influential stress-related transcription factors, *GCN4*, represents undoubtedly the best-studied model of eukaryotic translation reinitiation (REI) (Hinnebusch 2005; Gunišová and Valášek 2014). REI is a gene-specific regulatory mechanism exploiting the presence of short upstream uORFs in mRNA leaders (i.e., 5' untranslated regions—5' UTRs) of various genes. The molecular key to this potentially abundant regulation (Davuluri et al. 2000; Iacono et al. 2005; Calvo et al. 2009; Hood et al. 2009; Zhou et al. 2010) is the ability of some of these short uORFs (in yeast up to five codons in length [Vilela et al. 1998; Rajkowitsch et al. 2004; Szamecz et al. 2008], in plants up to 16 [von Arnim et al. 2014], and in mammals up to 30 codons [Kozak 2005]) to retain the 40S ribosomal subunit on the same mRNA molecule even after they have been translated and the large 60S subunit has been recycled by the ribosome recycling factors (for review, see Jackson et al. 2012;

Valášek 2012). Such post-termination 40S subunits are then able to resume scanning downstream, and upon acquisition of the new ternary complex (TC), composed of Met-tRNA^{Met} and eukaryotic initiation factor eIF2 in its GTP form, they are able to recognize the AUG start codon of the next ORF and reinitiate translation thereon.

Generally speaking, short uORFs in principle impose a functional barrier for sufficient expression of a downstream main ORF. This repressive effect of uORFs can be, however, alleviated under specific conditions, such as various types of stress, in order to boost expression of some regulatory uORF-containing mRNAs that help the cell to cope with sudden environmental changes. It has been shown that the efficiency of REI depends on four main factors: (i) time required for uORF translation, which is determined by the relative length of uORF and the translation elongation rate; (ii) its 5' and 3' flanking sequences, which contain specific *cis*-acting features

Corresponding authors: gunisova@biomed.cas.cz, valasekl@biomed.cas.cz

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.055046.115>.

© 2016 Gunišová et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

with poorly understood molecular roles; (iii) translation initiation factors (eIFs) involved in the primary initiation event such as the eIF3 and eIF4F complexes, which are believed to remain associated with the ribosome throughout the short elongation as well as termination and recycling phases; and (iv) its distance to the next open reading frame, which determines the likelihood of acquisition of the new TC by the post-termination 40S ribosome that has resumed scanning (Kozak 1987; Dever et al. 1992; Pöyry et al. 2004; Szamecz et al. 2008; Cuchalová et al. 2010; Roy et al. 2010; Munzarová et al. 2011).

The classical REI-dependent mRNA of *GCN4* containing a total of four short uORFs has been studied in great detail for several decades; it has been found to be very sensitive to the TC levels that are changing in response to different nutrient conditions and to rely mainly on the first REI-permissive uORF1 and the last REI-nonpermissive uORF4 (reviewed in Hinnebusch 2005 and recently revised in Gunišová and Valášek 2014). Briefly, the first of the four uORFs is efficiently translated under both nutritional replete and deplete conditions, and after its translation the post-termination 40S subunit remains attached to the mRNA and resumes scanning downstream for REI at the next AUG. In nonstressed cells, where the TC levels are high, nearly all of the rescanning ribosomes can rebind the TC before reaching one of the last two distant uORFs (uORFs 3 and 4), neither of which supports efficient REI. As a result, ribosomes terminating on one of these two uORFs undergo the full ribosomal recycling step, which prevents them from reaching and translating the main *GCN4* ORF (Supplemental Fig. S1).

Under starvation conditions, the GCN2 kinase phosphorylates eIF2, which suspends formation of new TCs in the cytoplasm. Consequently, post-termination 40S ribosomes traveling from the uORF1 stop codon downstream will require more time to rebind the TC to be able to recognize the next AUG start codon. This will allow a large proportion of them to bypass uORF3 and uORF4 and reacquire the TC downstream from uORF4 but still upstream of the *GCN4* start codon (Supplemental Fig. S1). Thus, whereas the global protein synthesis is significantly down-regulated under nutrient deplete conditions, protein expression of the *GCN4* transcriptional activator is concurrently induced.

The second REI-permissive uORF, uORF2 with ~80%–90% of the uORF1 REI activity, occurs only 56-nt downstream from uORF1 and serves as a backup of uORF1 to capture all ribosomes that leaky scanned past the uORF1 AUG (Gunišová and Valášek 2014), especially during stress conditions that seem to increase the frequency of leaky scanning in general (Lee et al. 2009; Raveh-Amit et al. 2009; Palam et al. 2011; Sundaram and Grant 2014). This ensures that the maximum capacity of this intriguing regulatory system is met. Similarly, two consecutive uORFs with minimal or no REI-promoting potential occurring further downstream (uORF3 that allows approximately five times less efficient REI than the first two uORFs, but still approximately four times more efficient REI than REI-nonpermissive uORF4) also prevent

“leakiness” of this system but during nutrient replete conditions (Gunišová and Valášek 2014). Hence the tightness of *GCN4* translational control is ensured by a fail-safe mechanism that effectively prevents or triggers *GCN4* expression under nutrient replete or deplete conditions, respectively.

The exceptionally high REI potential of uORF1 and uORF2 has been ascribed to (i) their 5′ sequences (Grant et al. 1995) containing several REI-promoting elements (RPEs) that together make up the so-called 5′ enhancer with a specific structural arrangement (Munzarová et al. 2011; Gunišová and Valášek 2014), (ii) the stimulatory role of N-terminal domain (NTD) of the α/TIF32 subunit of the eukaryotic initiation factor eIF3 (Szamecz et al. 2008) that sits near the 40S mRNA exit channel (Valášek et al. 2003; Kouba et al. 2012; Aylett et al. 2015), and, in the case of uORF1, also to (iii) the first 10 nt immediately following the uORF1 stop codon (Grant and Hinnebusch 1994), and (iv) the third coding triplet of uORF1 in combination with its 3′ UTR (Grant and Hinnebusch 1994).

With respect to known molecular functions of these features, some of the RPEs were shown to functionally interact with the α/TIF32-NTD, and this interaction, the exact nature of which still awaits determination, was implicated in stabilizing the post-termination 40S subunits on stop codons of both REI-permissive uORFs in order to facilitate the resumption of scanning downstream (Szamecz et al. 2008; Munzarová et al. 2011; Gunišová and Valášek 2014). The stretch of the first 10 nt past the uORF1 stop codon was owing to its relatively low GC content proposed to allow the 40S subunit to promptly resume scanning, in contrast to the corresponding GC-rich sequence of uORF4, because it would not permit strong base-pairing interactions with the 40S subunit (Grant and Hinnebusch 1994). However, we recently noted that the 3′ sequences of the other two *GCN4* uORFs (2 and 3) have a very similar AU content to uORF1, yet uORF3 permits only poor REI (Munzarová et al. 2011). Hence the idea that it is simply the high AU content of the uORF1 3′ sequence that makes it so robustly REI permissive does not seem to be so unequivocal anymore (Jackson et al. 2012).

With the recent outburst of new findings on the *GCN4* translation control that in some cases clarified and/or extended the earlier observations but in some others led to revised hypotheses, we wished to systematically analyze all potential *cis*-determinants that either promote or inhibit REI on *GCN4* after translation of its uORFs in one complex study. Based on this study we also postulate “reinitiation rules” that can be generalized for any short uORF in yeast (please see the Conclusion section).

RESULTS AND DISCUSSION

Proper functioning of the 3′ sequences immediately following stop codons of *GCN4* uORFs is not determined by their AU content

A considerable difference in the efficiency of resumption of scanning following translation of the REI-permissive

uORF1 vs. REI-nonpermissive uORF4 in the *GCN4* mRNA leader was found to be attributable to the distinct 3' sequences following the termination codons of these two uORFs. Replacing the sequence of 25-nt downstream from the uORF1 stop codon with the corresponding nucleotides from uORF4 disabled the uORF1-promoted REI on *GCN4* by a great deal (Miller and Hinnebusch 1989). It was originally proposed that this difference is due to a varying AU content of these 3' sequences, with uORF1 being more AU-rich (the AU content of 25 nt after uORF1 stop codon is 80%) than uORF4 (60%) (Fig. 1A; Grant and Hinnebusch 1994).

In our recent work, however, we noted that the 3' sequences of the other two uORFs of *GCN4* are similarly AU-rich as the uORF1 3' sequence by itself; the second REI-permissive uORF2 shows the highest AU content (88%), while the AU content of uORF3 that is only modestly REI-permissive is comparable to that of uORF1 (76%) (Fig. 1A; Munzarová et al. 2011; Gunišová and Valášek 2014). Therefore, if the degree of AU-richness is the sole factor determining the REI-promoting activity of the 3' sequences in general, at least a part of the REI potential of uORF2 and probably the whole REI potential of uORF3 could be explained by employment of the REI-promoting 3' sequences of these two uORFs.

To test this possibility, we substituted 25 nt immediately following stop codons of uORF2 and uORF3 with the corresponding sequence of REI-nonpermissive uORF4 in the same way as it was done previously for uORF1. In all our constructs, only the uORF under study was functional while the start codons of the other three uORFs were mutated out. Note that since the uORF4 start codon is separated from the uORF3 stop codon by only 13 nt, the 3' sequence of the *uORF3-3333* construct contains the mutated out uORF4 start codon (Fig. 1A). This mutation, however, does not change the overall AU content of the uORF3 3' sequence.

Also note that we divided the sequences constituting and flanking all four uORFs into four segments, A through D, for each uORF (Fig. 1B) according to Munzarová et al. (2011). In short, segment A is 166 bp in length (from position -181 to -16 relative to the uORF1-4 AUG start codons) and corresponds to the 5' REI-promoting sequences of uORF1 (Szamecz et al. 2008; Munzarová et al. 2011); segment B designated previously as the linker is 15-bp long (-15 to -1); segment C contains coding triplets and the stop codon; and segment D encompasses 25 bp immediately following the stop codon that corresponds to the 3' REI-promoting sequence of uORF1 (Miller and Hinnebusch 1989; Munzarová et al. 2011; Gunišová and Valášek 2014). Mutual replacements of these segments among all four uORFs that feature in the whole study are always indicated by the sequence of ABCD segments expressed as numbers 1–4, indicating which uORF each particular segment comes from: For example, a hypothetical *uORF1-1234* construct would correspond to segment A from uORF1 followed by B from uORF2, C from uORF3, and D from uORF4, all situated in the uORF1 position.

With respect to the aforementioned substitutions, replacing the 3' sequence (segment D) of uORF1 with that of uORF4 (*uORF1-1114*) reduced its REI potential down to ~31% (Fig. 1C), as reported before (Miller and Hinnebusch 1989). Replacing also the segment C along with D of uORF1 with those of uORF4 (*uORF1-1144*) further reduced its REI activity, almost to the basal level corresponding to the activity of REI-nonpermissive uORF4 (*uORF1-4444*), as also observed before (Fig. 1C; Miller and Hinnebusch 1989; Munzarová et al. 2011; Gunišová and Valášek 2014). This illustrates the magnitude of the 3' sequence contribution to the REI potential of uORF1, as its loss leads to a remarkable impairment (~69% drop in this experimental setup) of the uORF1 ability to allow resumption of scanning of post-termination ribosomes. Similar replacements of uORF2 and uORF3 in *uORF2-2224* and *uORF3-3334* did not decrease but actually increased REI activities of both uORFs (Fig. 1C). These findings indicate that (i) the REI potential of uORF2 and uORF3 does not depend on their AU-rich 3' sequences, and thus (ii) the degree of AU content per se does not play the key role in rendering the 3' sequences active in promoting REI. They also contradict our earlier suggestion that the GC-rich 3' sequence of uORF4 functions as the REI inhibitor that can suppress the resumption of scanning from any REI-permissive uORF (Gunišová and Valášek 2014). The fact that the presence of the GC-rich segment D of uORF4 allows even more efficient REI on the middle two uORFs, mainly on uORF3, will be explained below.

The 3' sequences of uORF1 and uORF2 operate autonomously in the position-specific manner

Since uORF1 seems to be the only uORF utilizing the 3' sequence for its REI-promoting activity, we next asked whether the 3' sequences of uORF2 and uORF3 with a similar AU content would also work for uORF1. Interestingly, only the presence of the uORF2 3' sequence (*uORF1-1112*) was able to functionally replace the uORF1 3' sequence (Fig. 1D). The 3' sequence of uORF3 (*uORF1-1113*) decreased the uORF1-mediated REI activity by nearly the same margin as the 3' sequence of uORF4 (*uORF1-1114*), which will also be explained below. The fact that uORF1 can also utilize the 3' sequence of uORF2 suggested that its function could be position specific. To examine this, we first attempted to activate the REI enhancing potential of the uORF2 3' sequence by placing the whole uORF2 ABCD block in the position of uORF1 (*uORF1-2222*). As a control, we replaced segment 2D with that of uORF4 in *uORF1-2224*. However, this replacement had nearly the same effect as when uORF2 occurred in its authentic position—its REI activity mildly increased but did not decrease (cf. Supplemental Figs. S1C, S2A), which further illustrates that uORF2 cannot utilize its 3' sequence to increase its REI potential. Since comparison of ratios of reinitiation vs. initiation rates for uORF1 and uORF2 showed that uORF1 resumes scanning with

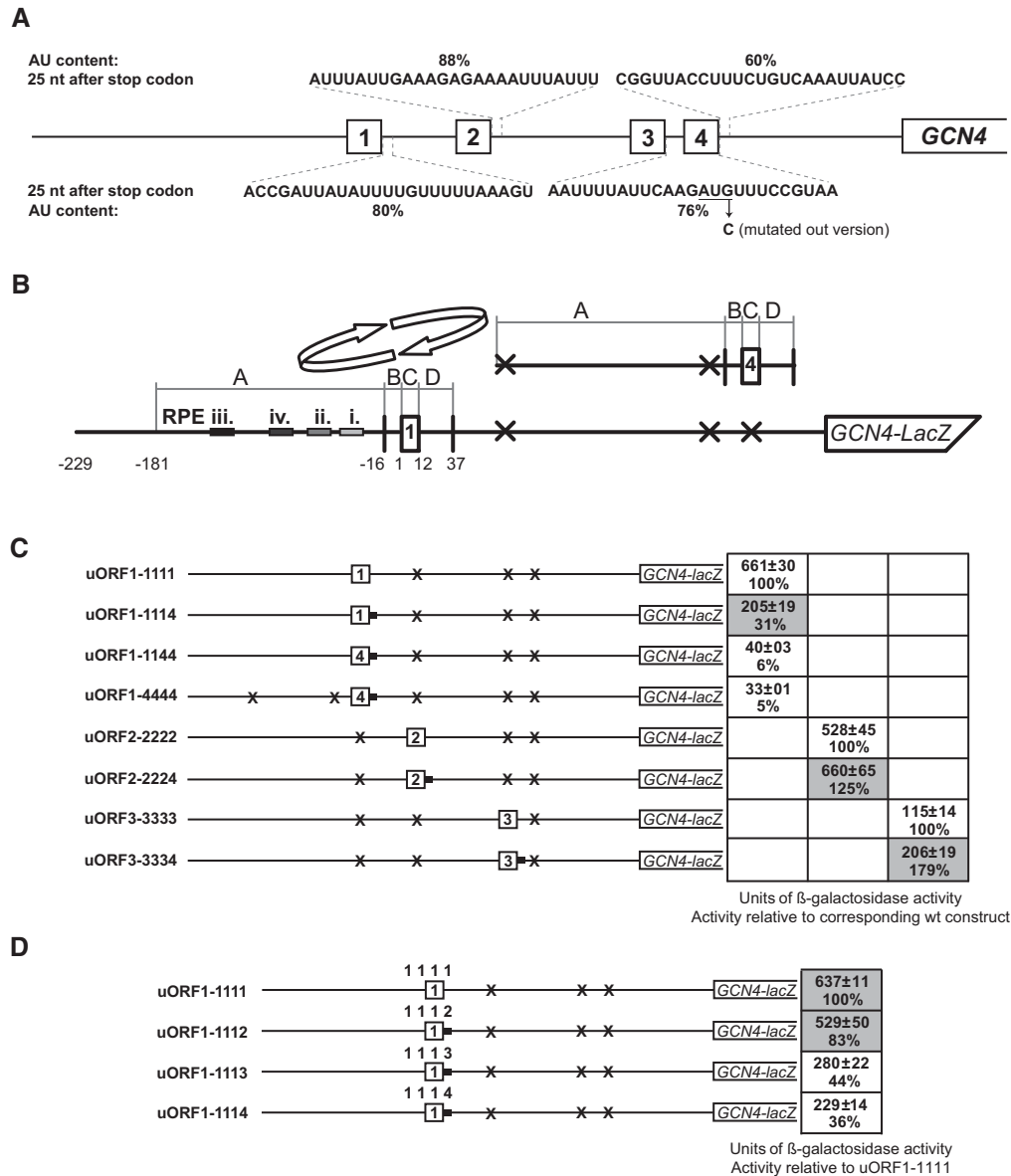


FIGURE 1. uORF1 is the only uORF utilizing its 3' flanking sequence for its REI-promoting activity. (A) Schematic of the *GCN4* leader containing four short uORFs (1–4). The first 25 nt of 3' flanking sequences (segments D) immediately following stop codon of each uORF are depicted together with the percentage of their AU content. In the uORF3 3' sequence, the start codon of uORF4 is underlined and arrow indicates G to C substitution used to mutate it out (see text for further details). (B) Schematic showing the *GCN4-lacZ* construct containing solitary uORF1, the surrounding sequences of which were divided into four separate segments (A1–D1; see text for further details). Arrows indicate replacements of these segments with the corresponding segments (A4–D4) surrounding uORF4, shown to the right of the arrows. Crosses indicate positions of other uORFs in which start codons were mutated out. Gray bars indicate sequence positions of individual uORF1-specific RPEs. Replacements with the corresponding segments surrounding uORF2 and uORF3 that are not depicted were done in the analogous way; adapted from Munzarová et al. (2011). (C) Solitary uORFs from the *GCN4* mRNA leader in their authentic positions and their derivatives with their 3' sequences replaced by the corresponding segment of uORF4 (indicated by the black bar) were introduced into the YSG2 strain. Additional replacements of indicated uORF1 segments by those of uORF4 were also prepared as a control. The resulting transformants were precultured in minimal media overnight, diluted to OD₆₀₀ ~ 0.35, and grown for an additional 6 h; the β-galactosidase activities were measured in the WCEs and expressed in units of *o*-nitrophenyl-β-d-galactopyranoside hydrolyzed/min/mg of protein. The mean values and standard deviations obtained from at least three independent measurements with five independent transformants and activities of the respective constructs relative to their corresponding wt constructs are given in the table. Differences in β-galactosidase activities between uORF2–2224 and its wt uORF2–2222 were analyzed by Student's *t*-test and the calculated *P* = 0.1335. (D) The 3' sequence of solitary uORF1 from the *GCN4* mRNA leader in its authentic position was replaced by corresponding segments of the other three uORFs and the resulting constructs were introduced into the YSG2 strain and analyzed as described in panel C. Differences in β-galactosidase activities between uORF1–1112 and its wt uORF1–1111 were analyzed by Student's *t*-test and the calculated *P* = 0.0911.

~14% higher efficiency than uORF2 (Supplemental Fig. S2B), we propose that it could be in fact the ability of uORF1 to utilize the 3' sequence expressed in its higher efficiency of resumption of scanning that to a great deal lies behind the small but reproducible difference (from 10% to 20%) in the overall REI efficiency between both uORFs that we routinely observe (Gunišová and Valášek 2014).

Next we tested whether or not the 3' sequence of uORF1, and as a proof of principle also that of uORF2, can operate from the position of uORF1 regardless of the nature of the preceding three segments. If yes, it would indicate (i) its autonomy and (ii) position specificity. To do that, we moved the whole uORF4 ABCD block into the position of uORF1 (*uORF1-4444*), and replaced segment D with the corresponding segments from the remaining three uORFs (*uORF1-4441*, *uORF1-4442*, and *uORF1-4443*) (Fig. 2A). As a control, we carried out the same D replacements (*uORF4-4441*, *uORF4-4442*, and *uORF4-4443*) also in the authentic position of uORF4 (*uORF4-4444*) (Fig. 2B). Remarkably, uORF1, and

to a slightly smaller degree also uORF2 3' sequences, did increase the REI potential of uORF4 (by approximately two-fold), but only when situated in the position of uORF1. The 3' sequence of uORF3, on the other hand, had the opposite effect on both setups, which is in accord with our findings shown in Figure 1C (see *uORF3-3333* vs. *uORF3-3334*), and we explain it below. Taken together, these results indicate that the REI-promoting 3' sequence of uORF1, and analogously also that of uORF2, does function autonomously, at least to a certain degree, but its activity is position-restricted. To further support this conclusion, we carried out the D replacements of uORF3 in its authentic position (*uORF3-3333*) and showed that neither uORF1 nor uORF2 3' sequences (*uORF3-3331* and *uORF3-3332*) had any stimulatory impact on REI allowed by this uORF. Predictably, the authentic 3' sequence of uORF3 showed the weakest REI activity of all (Supplemental Fig. S3).

What could determine the position specificity for the REI-promoting action of the 3' sequence of uORF1? In theory it

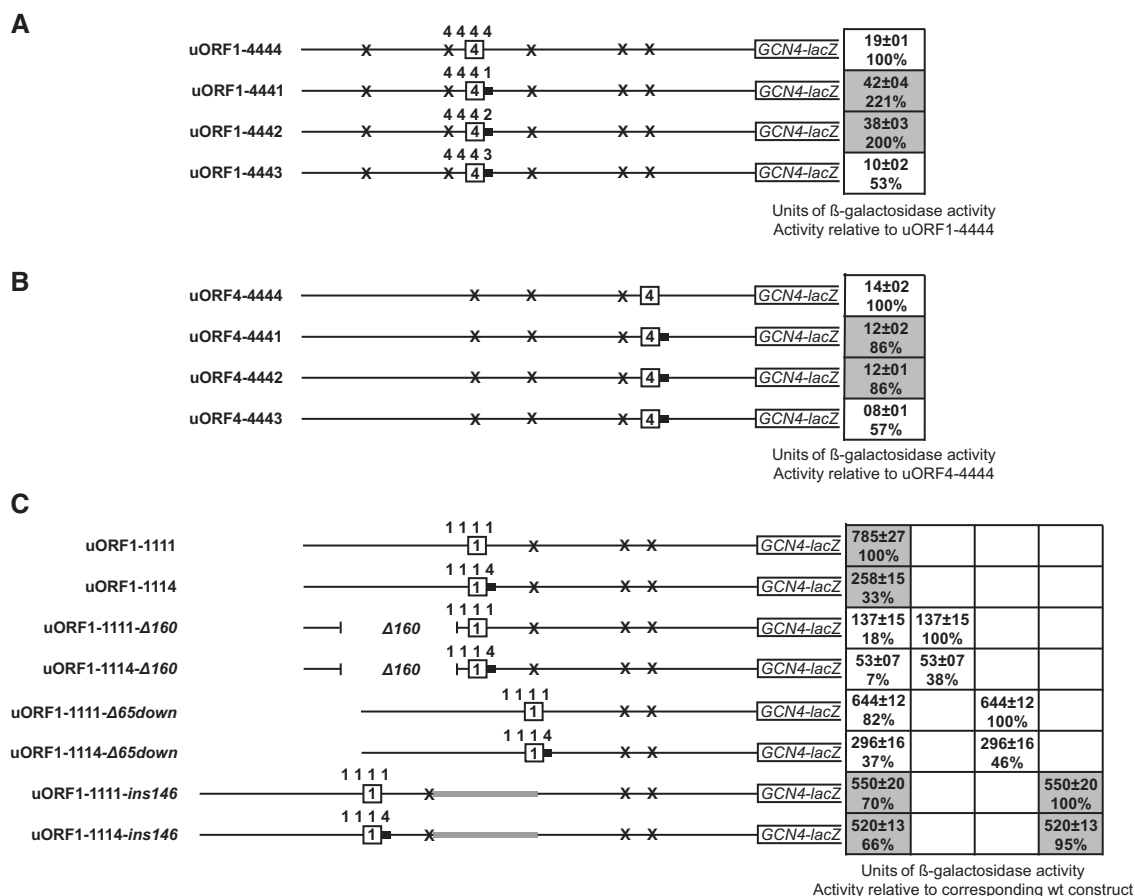


FIGURE 2. Function of the REI-promoting 3' sequence is position specific. (A) The 3' sequence of solitary uORF4 from the *GCN4* mRNA leader with its authentic flanking sequences (A4–D4) situated in the position of uORF1 was replaced by corresponding segments of the other three uORFs and analyzed as described in Figure 1C. (B) Same as in A except that the 3' sequence substitutions were done in the authentic position of uORF4. (C) Solitary uORF1 or its derivative containing the 3' sequence substitution for the corresponding segment of uORF4 situated in the authentic uORF1 position were modified as indicated and analyzed as described in Figure 1C. Differences in β -galactosidase activities between *uORF1-1111- Δ 65down*, *uORF1-1111-*ins*146*, or *uORF1-1114-*ins*146* and their wt *uORF1-1111* were analyzed by Student's *t*-test and the calculated *P* = 0.0014, 0.0001, or 0.0001, respectively.

could simply be its distance either from the 5' end of *GCN4* mRNA or from the *GCN4* main gene. To investigate this, we measured (i) the impact of shortening the distance between uORF1 start codon and the 5' end, or (ii) shortening or (iii) prolonging the distance between the uORF1 3' sequence and the *GCN4* start codon on the REI potential of wild-type uORF1 using the 1114 versions of all constructs designed for this purpose as a tool (Fig. 2C). We assumed that if any of our manipulations eliminate the REI-promoting activity of the 3' sequence of uORF1, their replacement with the 3' sequence of uORF4 would then show genetic epistasis; i.e., would not lead to any further decrease in the REI activity.

Shortening the distance from the 5' end by the deletion of 160-nt upstream of uORF1 containing all four REI-promoting elements (RPEs) (*uORF1-1111-Δ160*) dramatically decreased the REI potential of uORF1, as expected (Szamecz et al. 2008), but had only marginal, if any, impact on the autonomous activity of the uORF1 3' sequence (*uORF1-1114-Δ160*) (Fig. 2C, compare 100% vs. 33% and 100% vs. 38%). Hence the distance of the uORF1 3' sequence from the 5' end of *GCN4* mRNA does not seem to determine the position-specific phenomenon of the 3' sequences.

Next we examined the potential importance of the distance of the uORF1 3' sequence from the *GCN4* ORF. First we deleted 65 nt downstream from the uORF1 segment D, placing uORF1 directly into the uORF2 position (*uORF1-1111-Δ65down*). This deletion decreased REI on *GCN4* by ~18% compared to the wild-type *uORF1-1111* construct (Fig. 2C). It can be explained by two effects: (i) impaired function of the 3' sequence of uORF1 due to its altered position; (ii) shortening the distance that the 40S ribosomes scanning downstream from uORF1 need to reacquire the new ternary complex (TC) in order to become REI-competent, as demonstrated before (Dever et al. 1992). Since the same deletion in the *uORF1-1114-Δ65down* construct reduced REI on *GCN4* to a smaller degree than in the wild-type *uORF1-1114* construct (Fig. 2C, compare 46% to 33%), we believe that the observed drop in the REI activity of *uORF1-1111-Δ65down* is the combination of both aforementioned effects.

In a complementary approach, we prolonged the distance of uORF1 from the *GCN4* start by insertion of 146 nt representing two copies of the previously identified segment S1 downstream from the 3' sequence of uORF1 (Abastado et al. 1991) (*uORF1-1111-ins146*). Hypothetically, increasing the uORF1–*GCN4* distance should increase the REI on *GCN4*, as the scanning 40S ribosomes would gain even more time (thanks to a longer distance) to reacquire the TC on their way downstream from uORF1. Disruption of a prospective position-specific role of the uORF1 3' sequence would counteract this effect, producing either no increase of the REI activity on *GCN4* or even a decrease. Replacement of the uORF1 3' sequence with that of uORF4 should then be epistatic; and this is exactly what we observed (Fig. 2C). The double S1 insertion in the *uORF1-1111-ins146* construct reduced REI on *GCN4* by ~30% and had virtually the same

effect; i.e., the reduction by ~34% was also seen with the *uORF1-1114-ins146* construct. These observations thus demonstrate that moving the uORF1 3' sequence further away from the *GCN4* start codon completely abrogates its REI-promoting activity. Taken together, both shortening and lengthening approaches further support our theory that the action of the uORF1 3' sequence is, within some limited range of tolerance, position-specific.

The question that still remains unanswered is how this feature operates at the molecular level. Taking into account our earlier suggestion that the action of the 3' sequence of uORF1 is a prerequisite for the action of 5'-based RPEs of uORF1 (Munzarová et al. 2011), we propose the following two options. The RPEs are, in cooperation with eIF3, believed to stabilize the post-termination 40S subunit on the *GCN4* mRNA after the 60S subunit has been recycled in order to facilitate its resumption of scanning. If true, the 3' sequence of uORF1 that during termination is buried in the mRNA binding channel could, for example, expedite the ribosomal recycling of the large subunit by stabilizing the post-termination 80S ribosomes at the stop codon. It was recently shown that post-termination 80S ribosomes are not stably anchored at the stop codon and can migrate in both directions to codons that are cognate to the P-site deacylated tRNA (Skabkin et al. 2013). Such instability may lead to an undesirable accumulation of nonrecycled aberrant ribosomal complexes in the vicinity of the stop codon, which would indeed dampen the efficiency of REI. With this role, however, it would be hard to reconcile the position specificity of the uORF1 3' sequence; i.e., its functional requirement to be situated within a defined distance from the *GCN4* start codon.

This requirement brings us to the second option, as it could suggest that the uORF1 3' sequence interacts with some downstream sequence and/or *trans*-acting factor bound to a downstream sequence. And since it is buried in the terminating 80S ribosome, as hinted above, this prospective interaction could not be formed before completion of the first recycling step involving the 60S subunit dissociation (Pisarev et al. 2007). The purpose of this contact could be to prevent the last recycling step; i.e., dissociation of mRNA from the 40S subunit. Only if this is ensured, the RPEs in co-operation with eIF3 could further strengthen this stabilization effect and enable resumption of scanning of post-termination 40S subunits downstream. Changing the position of the uORF1 3' sequence with respect to the *GCN4* start site may prevent formation of this interaction, resulting in diminished REI efficiency of uORF1, which we observed.

The 3' sequences of the first three uORFs contain the conserved, REI-promoting AU-rich motif and at the same time, in the case of uORFs 2 and 3, also inhibitory elements further downstream

The puzzling similarity of the AU-richness of the 3' sequences of uORF1–3 combined with their differential effects on

GCN4 REI observed in Figure 1D prompted us to inspect the composition of the 3' sequences of these three uORFs in greater detail. We revealed a potential AU-rich motif ($AU_{1-2}A/UUAU_2$) specifically occurring within the first 12 nt of segment D of only the first three uORFs (Fig. 3A). To examine whether this motif solely carries the REI-promoting activity of the 3' sequences of uORFs 1 and 2, and perhaps even of uORF3, we replaced either the first 1–12 nt or the immediately following 13–25 nt of the uORF1 segment D with the corresponding sequences of the other three uORFs and measured the β -galactosidase activities of the resulting constructs. In the case of uORF2, we not only observed that its $AU_{1-2}A/UUAU_2$ motif-containing 1 through 12 nt sequence behaved as efficiently as the native uORF1 3' sequence ($\sim 102\%$ for $uORF1-1112^{1-12}$), but also that its 13 through 25 nt sequence ($uORF1-1112^{13-25}$) had a modest inhibitory effect on the REI activity of uORF1 (Fig. 3). In the case of uORF3, the first 12 nt containing the AU-rich motif showed only modestly decreased activity compared to the native uORF1 3' sequence ($\sim 79\%$ for $uORF1-1113^{1-12}$), whereas its 13 through 25 nt sequence ($uORF1-1113^{13-25}$) displayed

a robust inhibitory effect. Strikingly, this inhibitory effect is dominant over the stimulatory effect of the AU-motif, which perfectly explains why replacing the entire segment D of uORF1 (Fig. 1D) or of uORF4 in the uORF1 position (Fig. 2A) with the same segment of uORF2 or uORF3 reduced the efficiency of REI by varying degree despite their similar AU content. These findings also suggest that the modest dominant inhibitory impact of 13–25 nt of uORF2 3' sequence (further documented in Fig. 1C; Supplemental Fig. S2A) further contributes to the difference in the overall REI efficiency between uORF1 and uORF2.

Replacing the first 12 nt of the uORF1 segment D with the same uORF4 sequence ($uORF1-1114^{1-12}$) completely eliminated the stimulatory effect of the uORF1 3' sequence, whereas the replacement of the following 13–25 nt ($uORF1-1114^{13-25}$) had no effect whatsoever (Fig. 3B), clearly suggesting that the 13 through 25 nt segment of uORF4 has no inhibitory effect. This observation explains why the D segment replacement of uORF2 and uORF3 with that of uORF4 increased their overall REI activity (Fig. 1C). Importantly, it was previously suggested that the first 10 (and not 12) nt of

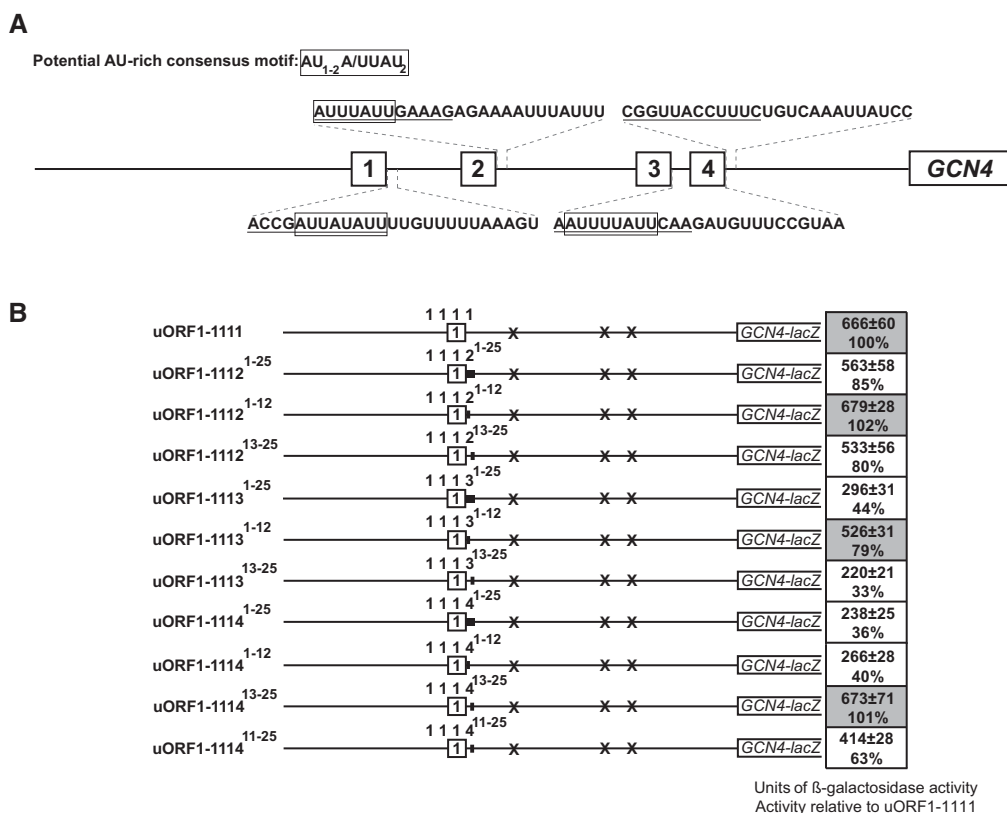


FIGURE 3. The presence of the specific AU-rich motif determines the REI-promoting activity of the 3' sequence. (A) Schematic of the GCN4 mRNA leader adapted from Figure 1A. The specific AU-rich motif (shown in frame) is present within the first 12 nt of the 3' sequences of uORF1, 2, and 3; the first 12 nt of the 3' sequence of each uORF are underlined. (B) The "full-length" 3' sequence (1–25) of solitary uORF1 from the GCN4 mRNA leader in its authentic position or its parts represented by the first 12 nt (1–12) or second 13 nt (13–25), or 15 nt in the case of uORF4 (11–25), were replaced by corresponding segments of the other three uORFs and analyzed as described in Figure 1C. Differences in β -galactosidase activities between $uORF1-1112$, $uORF1-1112^{13-25}$, $uORF1-1113^{1-12}$, or $uORF1-1114^{11-25}$ and their wt $uORF1-1111$ were analyzed by Student's *t*-test and the calculated *P* = 0.25, 0.1469, 0.0722, or 0.0052, respectively.

the 3' sequence of uORF1 suffice for its full function (Miller and Hinnebusch 1989), which leaves out the last 2 nt from the AU-rich motif. To reconcile these and our observations, we replaced the 11–25 nt of the uORF1 3' sequence with the corresponding region of uORF4 (*uORF1-1114*^{11–25}) and observed that the REI activity on *GCN4* dropped down to 63%, in contrast to the 13–25 nt replacement showing 100% activity (Fig. 3B). Hence at least in our experimental setup, it is not the first 10 but the first 12 nt that encompass the entire AU₁₋₂A/UUAU₂ motif needed for the full REI potential of the uORF1 3' sequence.

Taken together, our findings strongly suggest that the specific AU-rich motif occurring within the first 12 nt of segment D of uORF1, uORF2, and uORF3 is solely responsible for the REI-promoting function of the 3' sequence, provided that it is situated at or near the authentic position of uORF1. As can be seen, however, sequences immediately following this motif also play an important role in this regulatory mechanism as they, at least in the case of uORFs 2 and 3, negatively influence their overall REI potential. Since uORF2 and uORF3 stimulate REI in the AU-rich motif-independent manner (Fig. 1C), and since these inhibitory sequences are transferable also to the ABCD blocks of uORF1 and uORF4, it is conceivable that they also act autonomously and impact a different REI step than that involving the action of the AU₁₋₂A/UUAU₂ motif.

The exact length of uORF1 and uORF2 coding sequences is critical for their maximal REI activity that shows a partially autonomous character

In our effort to systematically assess individual contributions of sequences flanking the REI-permissive uORFs to their REI capacity, we also wished to examine the potential contribution of the uORF1 and uORF2 coding sequences (CDS) per se. Early experiments done in the Hinnebusch laboratory suggested that the last coding triplet, which is represented by TGC coding cysteine for uORF1 and CCG coding proline for uORF4, does somehow contribute to the remarkable difference in the efficiency of REI between these two uORFs (Miller and Hinnebusch 1989; Grant and Hinnebusch 1994). Interestingly, the second REI-permissive uORF, uORF2, also contains a cysteine codon (TGT) as its last sense codon as uORF1, only it is one codon shorter (Fig. 4A). The same kind of similarity for the last sense codon also applies to uORF3 and REI-nonpermissive uORF4, as uORF3, like uORF4, contains the CCG triplet coding proline (Fig. 4A).

In analogy with our previous approach (Fig. 1D), we first replaced the CDS of uORF1 with CDSs of the other three uORFs (*uORF1-1121*, *uORF1-1131*, and *uORF1-1141*). None of them could work in place of uORF1 as efficiently as uORF1 by itself, not even the Cys codon-containing REI-permissive uORF2; the REI activity of these three replacements was reduced by ~51% to 70% (Fig. 4B). This is in agreement with an earlier study where the activity of

uORF1-1141 was measured in the strain deleted for the *GCN4* pathway-specific GCN2 kinase (Miller and Hinnebusch 1989). Hence we propose that not only the character of the last coding triplet but also the exact length of the uORF1 CDS is critical for its function. In support, extending the length of the two-codon uORF2 by insertion of the uORF1 GCT Ala codon immediately after the uORF2 start codon (*uORF2-2212*) reduced its REI activity by 40% (Fig. 4C). Our findings thus clearly imply that the authentic length of both REI-permissive uORFs has to be maintained for their optimal activity. This could matter, for example, due to the fact that even slight alterations of their length may interfere with proper functioning of their 5' and/or 3' sequences.

To find out whether or not the CDSs of uORF1 and uORF2 can at least to some degree stimulate REI independently of their flanking, REI-promoting sequences, we replaced the uORF4 segment C from the uORF4 ABCD block situated either in the position of uORF1 (*uORF1-4444*) (Fig. 4D) or in its authentic position (*uORF4-4444*) (Fig. 4E) with those of uORF1 (*uORF1-4441* and *uORF4-4441*) or uORF2 (*uORF1-4442* and *uORF4-4442*). Regardless of the placement of the uORF4 ABCD block, the presence of CDSs of both REI-permissive uORFs did increase the efficiency of REI by ~26% to 44%. However, since segment C replacements in the uORF1 ABCD block showed >50% decrease in its REI activity (Fig. 4B), it is intuitive that a fully autonomous contribution of CDSs of both REI-permissive uORFs should lead to a larger, approximately twofold increase, which we did not observe. Hence we conclude that both uORFs per se have some inherent ability to promote REI independently of other sequences that is not position-specific, in contrast to the 3' AU-rich motif. Not surprisingly, however, they also functionally interact with their flanking sequences to maximize their REI potential. Generally speaking, a REI-permissive uORF can be viewed as the complex unit with several independent and/or mutually interdependent contributors with peculiar attributes such as the CDS length.

The CCG triplet coding proline negatively impacts REI efficiency and at least in yeast has a predictable value in this respect

To investigate the causality of the aforementioned occurrence of the Cys triplet as the ultimate coding triplet in REI-permissive uORFs 1 and 2, and of the Pro triplet as the ultimate coding triplet in uORF3 and uORF4, we first inspected the *GCN4* mRNA leaders of closely or distantly related yeast species to *Saccharomyces cerevisiae* (Fig. 5A). With the exception of *Candida glabrata*, there is a clear conserved tendency for all other species to preserve Cys and Pro as the last coding triplets in at least the first and last short uORF, respectively. In the case of *C. glabrata*, the Cys triplet occurs in uORF3 and is followed by long uORF4 that is, owing to its length, by definition strictly REI-nonpermissive (Vilela et al. 1998; Szamecz et al. 2008). Therefore, this regulatory setup simply

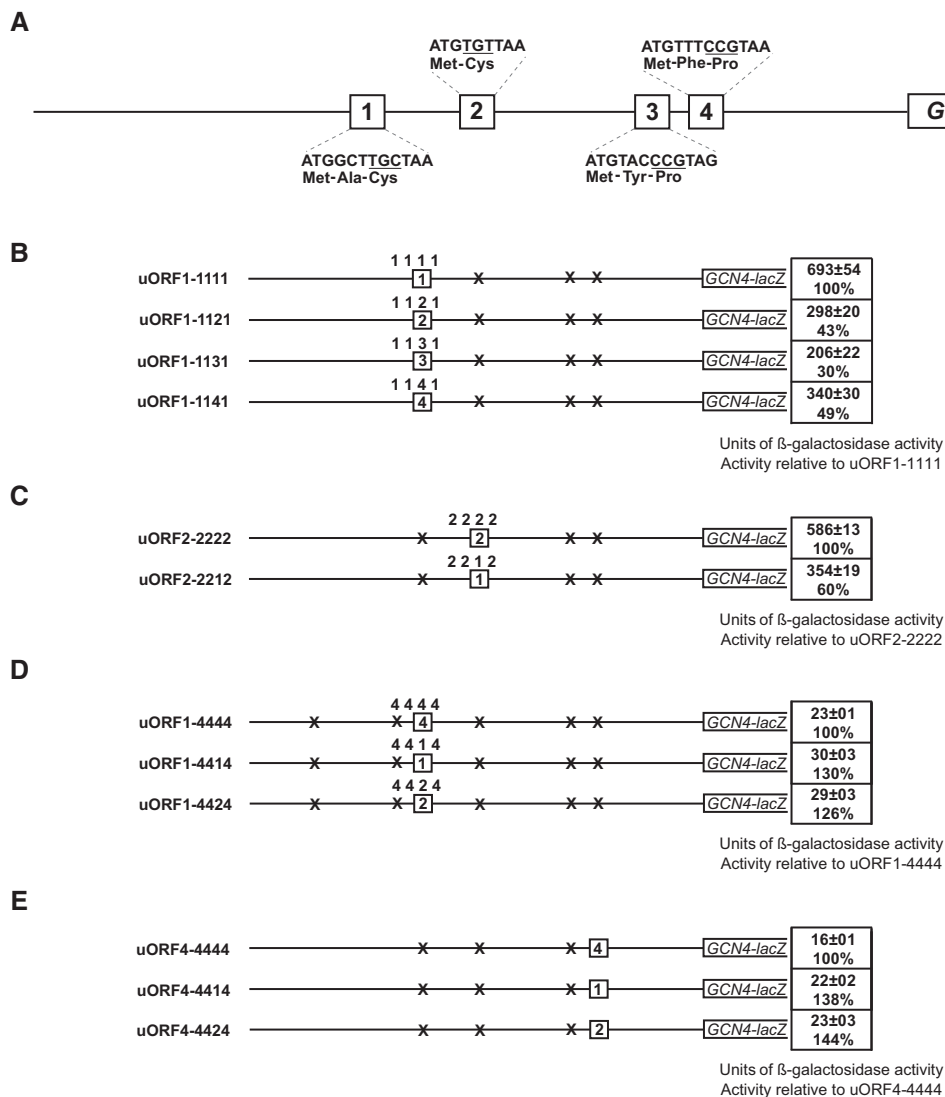


FIGURE 4. Sequence composition and length but not the position of uORF1 and uORF2 CDSs define their REI permissiveness. (A) Schematic of the *GCN4* leader containing four short uORFs (1–4). For each uORF, nucleotide as well as amino acid sequences are shown. The triplets of the last sense codon are underlined. (B) The CDS (segment C) of solitary uORF1 from the *GCN4* mRNA leader in its authentic position was replaced by corresponding segments of the other three uORFs and analyzed as described in Figure 1C. (C) The CDS of solitary uORF2 from the *GCN4* mRNA leader in its authentic position was replaced by corresponding segment of uORF1 and analyzed as described in Figure 1C. (D) The CDS of solitary uORF4 from the *GCN4* mRNA leader with its authentic flanking sequences (A4–D4) situated in the position of uORF1 was replaced by corresponding segments of uORF1 and 2 and analyzed as described in Figure 1C. Differences in β -galactosidase activities between uORF1-4414 or uORF1-4424 and their wt uORF1-4444 were analyzed by Student's *t*-test and the calculated $P = 0.0572$ or 0.094 , respectively. (E) Same as in D except that the CDS substitutions were tested in the authentic position of uORF4. Differences in β -galactosidase activities between uORF4-4414 or uORF4-4424 and their wt uORF4-4444 were analyzed by Student's *t*-test and the calculated $P = 0.0275$ or 0.0581 , respectively.

does not require featuring Pro as the ultimate coding triplet of the last uORF.

The question is, what stands behind this phenomenon? Is it the Cys/Pro triplet or the identity of these two residues per se? To examine that, we systematically substituted the ultimate coding triplet individually in all four *GCN4* uORFs and measured the resulting REI activities. Recoding the cysteine triplet of uORF1 [uORF1-1111-Cys (TGT)] or uORF2 [uORF2-2222-Cys (TGC)] showed almost no impact on REI of these uORFs (Fig. 5B,C), which contrasts with the

remarkable increase in REI seen when the analogous uORF1 recoding was tested in early days, however, in the artificial setup with the genuine uORF1 3' sequence replaced with the corresponding 3' sequence of uORF4 (Grant and Hinnebusch 1994). In fact, our measurements are perfectly consistent with another work by the same group, where the REI activity was measured for the synonymous uORF1 cysteine substitutions under starvation conditions using a minimalistic construct lacking uORF2 and uORF3 and no significant changes were found (Miller and Hinnebusch

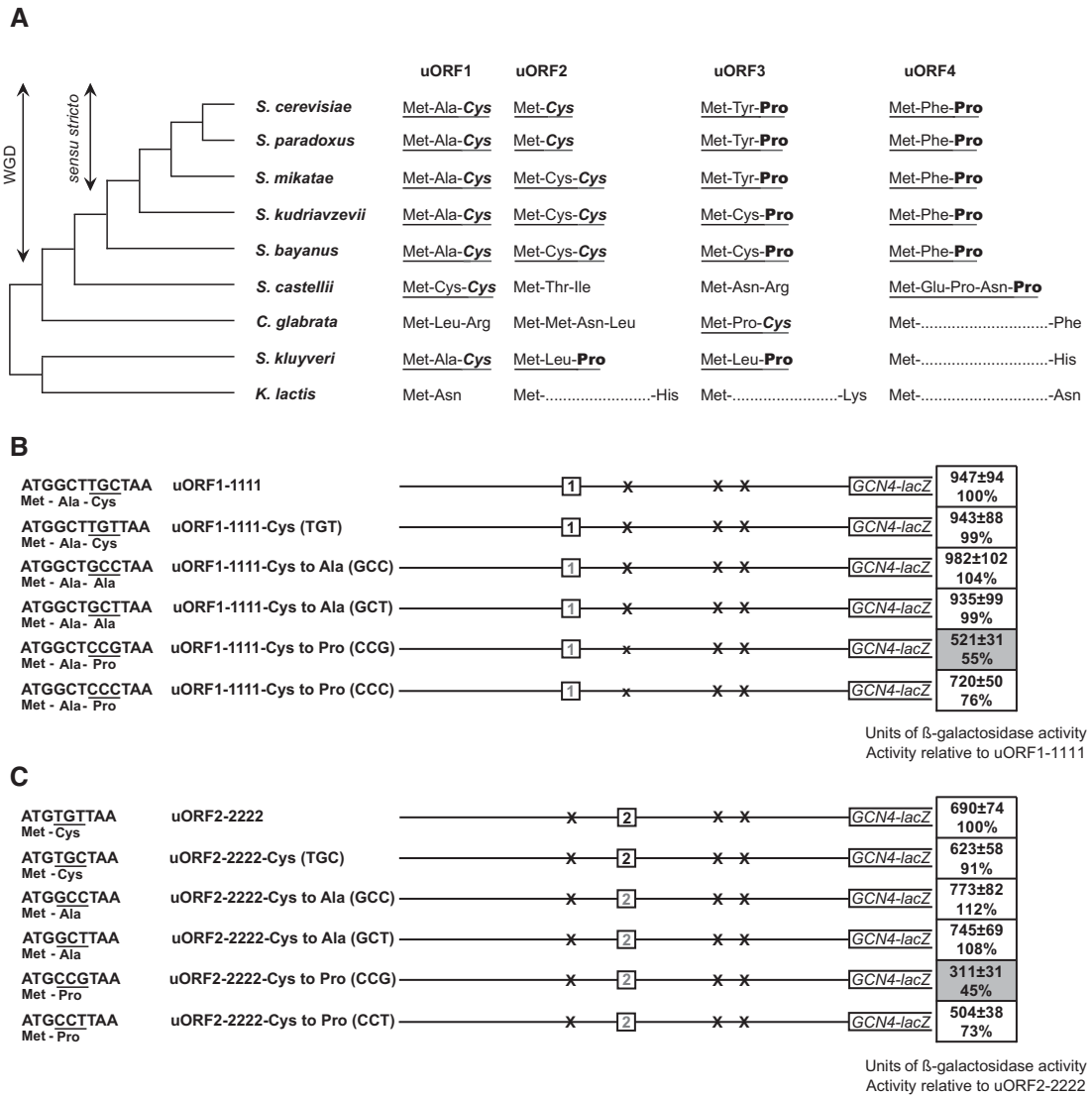


FIGURE 5. Insertion of the CCG triplet as the last sense codon of uORF1 and uORF2 inhibits their REI activity. (A) The conservation of Cys and Pro codons in short uORFs from the *GCN4* mRNA leaders of yeast species related to *S. cerevisiae*. The uORFs are presented as identified in Cvijović et al. (2007). uORFs containing Cys (in italics) and Pro (in bold) triplets as their last sense codons are underlined. In longer uORFs, the sequence of their CDSs is represented by a dotted line. WGD, whole-genome duplication. (B) The TGC cysteine codon of solitary uORF1 from the *GCN4* mRNA leader in its authentic position was replaced by indicated cysteine, alanine, and proline substitutions and analyzed as described in Figure 1C. Differences in β -galactosidase activities between uORF1-1111-Cys to Pro (CCC) and its wt uORF1-1111 were analyzed by Student's *t*-test and the calculated $P = 0.0648$. (C) The TGT cysteine codon of solitary uORF2 from the *GCN4* mRNA leader in its authentic position was replaced by indicated cysteine, alanine, and proline substitutions and analyzed as described in Figure 1C. Differences in β -galactosidase activities between uORF2-2222-Cys to Pro (CCT) and its wt uORF2-2222 were analyzed by Student's *t*-test and the calculated $P = 0.0554$.

1989). Hence we conclude that both cysteine codons of uORF1 and uORF2 utilizing the same Cys-tRNA are similarly efficient in promoting REI. In addition, our findings further support the idea that it is most probably the length of the CDSs of uORF1 and uORF2 and not the differing Cys triplet that stands behind their inability to replace each other in their authentic environment (Fig. 4B,C).

To examine whether it is the cysteine residue per se that lies behind this phenomenon, we substituted the cysteine codons of uORF1 (TGC) and uORF2 (TGT) with two different ala-

nine codons: (i) GCC [uORF1-1111-Cys to Ala (GCC) and uORF2-2222-Cys to Ala (GCC)], and (ii) GCT [uORF1-1111-Cys to Ala (GCT) and uORF2-2222-Cys to Ala (GCT)]. Like the cysteine substitutions, neither of the Ala substitutions had any effect on either uORF (Fig. 5B,C). Hence the REI permissiveness of uORF1 and uORF2 CDSs does not exclusively depend on the presence of a cysteine codon as the last coding triplet, as suggested before (Grant and Hinnebusch 1994). It also does not seem to strictly depend on the AU vs. GC content, as proposed by Grant

and Hinnebusch (1994), because the changes that increased or decreased the AU content showed no effect (Fig. 5B,C). This implies that the third codon can most probably tolerate a wide range of codons as long as these do not interfere with the resumption of scanning.

Finally, we substituted Cys codons in both REI-permissive uORFs with the supposedly inhibitory proline CCG triplet from uORFs 3 and 4 [*uORF1-1111-Cys to Pro (CCG)* and *uORF2-2222-Cys to Pro (CCG)*] as well as with the other two proline triplets; CCC for uORF1 [*uORF1-1111-Cys to Pro (CCC)*] and CCT for uORF2 [*uORF2-2222-Cys to Pro (CCT)*] (Fig. 5B,C). Whereas the latter CCT and CCC substitutions led to ~24%–27% drop in the REI activity on both uORFs, the CCG triplet produced the most robust ~45%–55% reduction. A similar decrease was also seen when the CCG substitution in uORF2 was made in the construct where the entire uORF2 ABCD block was placed in the position of uORF1 (data not shown). These findings correspond well with our uORF1 segment C replacement data (Fig. 4B), where introduction of the corresponding uORF4 segment in *uORF1-1141* also reduced the REI efficiency by ~51%. This suggests that at least for uORF1 the entire negative effect could be solely attributable to the presence of the last coding CCG triplet.

Importantly, we cannot explain the observed difference among the three tested Pro triplets by their varying codon optimality, because it simply does not correlate with the amounts of distinct types of Pro-tRNAs specific to these triplets. Hence, we propose that both the presence of proline as the ultimate residue and also some peculiar attributes of the CCG codon beyond codon optimality act together to maximize the inhibitory impact on REI after translation of short uORFs that are not meant to promote efficient resumption of scanning. Consistently, replacing the CCG Pro triplet in uORF3 and uORF4 with the uORF1 TGC cysteine codon relieved the CCG-mediated inhibition of REI after uORF3 and uORF4 translation to some degree (Supplemental Fig. S4).

The proline residue is known for its unusual conformation that, due to its limited spatial rotation during peptide bond formation, slows down the incorporation of this residue into the nascent polypeptide chain and therefore negatively influences the speed of translation (Wohlgemuth et al. 2008; Pavlov et al. 2009). Since uORF1 and uORF2 are very short, the presence of the slowing down Pro as the last residue could theoretically interfere with the time-sensitive formation of the interaction between uORF1 and uORF2 RPEs and their interacting factor eIF3. In support, prolonging the time of uORF1 proteosynthesis by extending its coding region by up to five codons gradually decreased its REI potential (Szamecz et al. 2008). The difference observed among the three tested Pro triplets could then be caused by a specific character of the CCG sequence that might, for example, interfere with proper stop codon recognition for effective translation termination.

The sequence composition of the uORF3 and uORF4 second codon and markedly differing termination efficiency makes the difference in the REI activity between these two uORFs

As mentioned above, we previously established that the inherent REI activity of uORF3 represents ~13%–18% of the maximum REI capacity allowed by uORF1, whereas uORF4 goes down to ~3%–5% (Munzarová et al. 2011; Gunišová and Valášek 2014). Therefore, we wished to uncover what determines the approximately fourfold higher REI activity of uORF3 compared to uORF4. In our earlier work, we demonstrated that the modest REI potential of uORF3 does not depend on any of the uORF1- and/or uORF2-specific RPEs and, accordingly, it does not require intact eIF3 (Gunišová and Valášek 2014). Here we ruled out the possibility that this difference could be caused by the uORF3's ability to utilize the AU-rich motif contained within its 3' sequence (Fig. 1C; Supplemental Fig. S3). Hence the only option left was the CDS of both uORFs, where only two differences can be found: (i) the second codon that is represented by TAC encoding Tyr in uORF3 and TTT encoding Phe in uORF4, and (ii) the stop codon that is represented by TAG in uORF3 and TAA in uORF4 (Fig. 6A). Looking at their 12-nt sequences per se, these two uORFs differ only by 3 nt.

To investigate what lies behind this difference, we prepared the uORF3- and uORF4-based constructs, where each aspect at which they differ was individually or in combinations swapped between them and examined for REI activity (Fig. 6B). The Tyr to Phe substitution of uORF3 [*uORF3-3333-Tyr to Phe (TTT)*] markedly decreased its activity (down to ~55%) (Fig. 6B). A milder decrease down to ~71% was obtained when the stop codon of uORF3 was replaced with that of uORF4 (*uORF3-3333-TAA*), however, a combination of both of these substitutions in *uORF3-3333-Tyr to Phe (TTT)-TAA* showed an additive effect by allowing only ~42% of the intrinsic REI activity of wild-type uORF3 (Fig. 6B). Interestingly, whereas the stop codon swap in uORF4 (*uORF4-4444-TAG*) showed no effect, the Phe to Tyr substitution of uORF4 [*uORF4-4444-Phe to Tyr (TAC)*] increased the REI activity by ~1.7-fold and their subsequent combination in *uORF4-4444-Phe to Tyr (TAC)-TAG* also led to an additive effect by increasing the intrinsic REI potential of uORF4 by approximately twofold (Fig. 6B). Our results thus indicate that the differences in the composition of the uORF3 and uORF4 CDSs are indeed at least partially responsible for their different REI permissiveness.

Next we analyzed the effect of the second sense codon in greater detail because it seemed to have a bigger influence on efficiency of resumption of scanning over the identity of the stop codon. To find out whether it is the encoded amino acid or a particular triplet that makes the difference, we tested both existing Tyr codons (original TAC and synonymous TAT) and both existing Phe codons (original TTT and synonymous TTC) in both uORFs of interest. Luckily, the

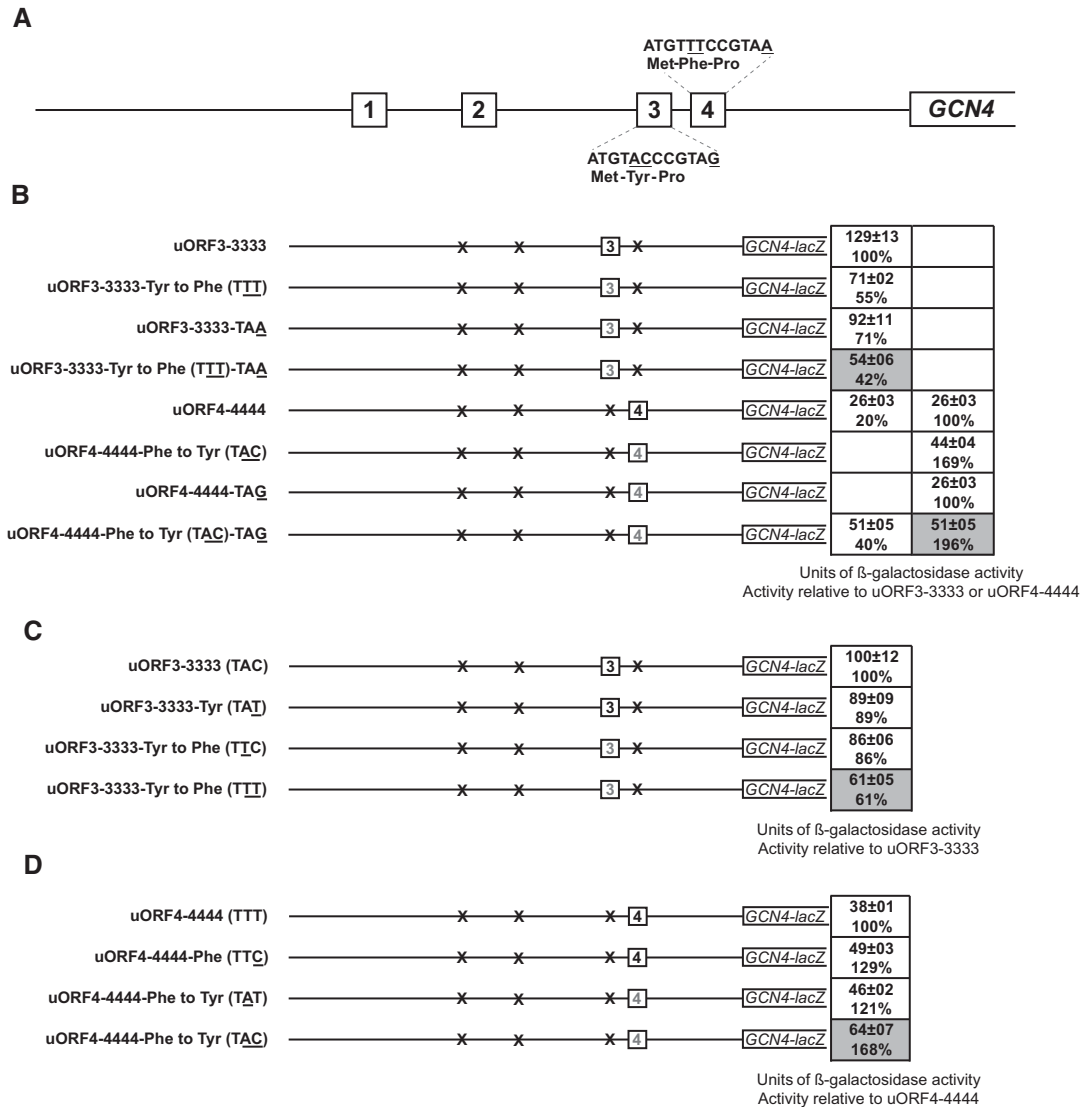


FIGURE 6. The differences in the sequence composition of the uORF3 and uORF4 CDSs are responsible for their varying REI permissiveness. (A) Schematic of the GCN4 leader adapted from Figure 4A. Nucleotide as well as amino acid sequences of only uORF3 and uORF4 are depicted. The different nucleotides between these two uORFs are underlined. (B) The CDSs of solitary uORF3 and uORF4 from the GCN4 mRNA leader in their authentic positions were modified by indicated substitutions and analyzed as described in Figure 1C. Differences in β -galactosidase activities between uORF3-3333-TAA and its wt uORF3-3333, or uORF3-3333-Tyr to Phe (TTT) and uORF3-3333-Tyr to Phe (TTT)-TAA, or uORF4-4444-Phe to Tyr (TAC) and uORF4-4444-Phe to Tyr (TAC)-TAG were analyzed by Student's *t*-test and the calculated *P* = 0.0609, 0.0275, or 0.3033, respectively. (C) Same as in B except that a different set of solitary uORF3 substitutions was used. Differences in β -galactosidase activities between uORF3-3333-Tyr to Phe (TTT) and its wt uORF3-3333 were analyzed by Student's *t*-test and the calculated *P* = 0.0171. (D) Same as in B except that a different set of solitary uORF4 substitutions was used.

interpretation of these analyses is simplified by the fact that both existing Tyr and Phe codons are recognized by a single Tyr- or Phe-tRNA, respectively. Therefore, there should be no unspecific influence on the effect of our substitutions. Out of these three possible substitutions only Phe (TTT), changing two out of three original bases in uORF3-3333-Tyr to Phe (TTT), produced by far the most dramatic reduction in the uORF3 REI activity (down to ~61%) (Fig. 6C). Similarly, the Tyr (TAC) substitution, changing two out of three original bases in uORF4-4444-Phe to Tyr (TAC), pro-

duced by far the largest increase (by ~1.7-fold) in the uORF4 REI activity (Fig. 6D). All other substitutions that changed only one out of three bases of both uORFs [uORF3-3333-Tyr (TAT) and uORF3-3333-Tyr to Phe (TTC) in Fig. 6C, and uORF4-4444-Phe (TTC) and uORF4-4444-Phe to Tyr (TAT) in Fig. 6D] altered the efficiency of REI only modestly. Since the same amino acid encoded by two different triplets that are, however, recognized by the same tRNA produced a marked difference in REI activity, we can conclude that it is not the identity of an amino acid

residue or of a tRNA by which it is delivered, but the sequence composition of the uORF3 and uORF4 second codon that contributes to the difference in the REI potential between these two uORFs.

Because the full remake of uORF3 into uORF4 in the authentic position of uORF3 [*uORF3-3333-Tyr to Phe (TTT)-TAA*] reduces its REI activity only approximately twofold, and at the same time the full remake of uORF4 into uORF3 in the authentic position of uORF4 [*uORF4-4444-Phe to Tyr (TAC)-TAG*] increases its REI activity also only approximately twofold, there must clearly be some other factor outside of their CDSs that further contributes to the approximately fourfold difference in REI activity between uORF3 and uORF4. We hypothesized that this factor could lie in the identity of the base immediately following the stop codon (the so-called +4 nucleotide), which is known to strongly influence the efficiency of stop codon recognition (see, for example, Beznosková et al. 2015b). In particular, cytosine at the +4 position is the leakiest base with respect to efficiency of translation termination among all four bases at all three stop codons (Beznosková et al. 2015a; Dabrowski et al. 2015). Hence, we rationalized that an uORF with a poorer stop codon context allowing increased frequency of stop codon readthrough would extend its translation beyond its genuine stop codon, which would in turn decrease its efficiency of REI. Interestingly, the TAA stop codon of uORF4 is followed by C and TAG of uORF3 by A (Figs. 1A, 7A), and indeed our dual-luciferase reporter assay revealed that uORF4 has ~3.7-fold increased frequency of readthrough compared to uORF3 (Fig. 7B). As a control, we used the TGA stop codon with the C at the +4 position (TGA-C), which is known to allow relatively high levels of readthrough (Beznosková et al. 2015b). Notably, readthrough on uORF4 is even ~1.5-fold higher than readthrough on this very rare terminating tetranucleotide with the so-called programmed readthrough. Hence we propose that the fact that ~3.7-times more ribosomes fail to terminate at uORF4 compared to uORF3 does further contribute to its overall REI nonpermissiveness and thus to the difference in REI activity between these two uORFs.

If true, the combination of the full remake of uORF4 into uORF3 [*uORF4-RT-Phe to Tyr (TAC)-TAG*] with a substitution of the +4 nucleotide C of uORF4 for A of uORF3 (*uORF4-RT-TAAA*) into one construct [*uORF4-RT-Phe to Tyr (TAC)-TAGA*] should reduce the frequency of readthrough down to the levels observed for wt uORF3; and this is exactly what we observed. Whereas individual mutations reduced readthrough by similar ~2.6-fold, their combination displayed an approximately fourfold decrease [Fig. 7B; compare 24% of *uORF4-RT-Phe to Tyr (TAC)-TAGA* with 27% of *uORF3-RT*]. Consistently, the same trend but in a reverse order was seen when we substituted the +4 nucleotide C of uORF4 for A of uORF3 and measured the REI efficiency using the *GCN4-LacZ* construct (*uORF4-4444-TAAA*). It increased the overall REI activity of uORF4 by ~1.6-fold

(164%) and the same substitution in the full remake of uORF4 into uORF3 [*uORF4-4444-Phe to Tyr (TAC)-TAGA*] increased the overall REI activity from approximately twofold (209%) to approximately threefold (291%), which is now close to the aforementioned approximately fourfold difference between both uORFs (Fig. 7C). We thus conclude that the combination of the sequence composition of the uORF3 and uORF4 second codon and the identity of the stop codon tetranucleotide is what makes the major contribution to the difference in the REI potential between these two uORFs.

Individual and combined effects of all *cis*-elements that determine REI permissiveness of uORF1 and uORF2

To capture our in-depth analysis, we gradually combined mutations in all identified REI-promoting *cis*-determinants of uORF1 as well as uORF2 in an effort to compare their individual contributions to the overall REI potential of both REI-permissive uORFs. First, we side-by-side evaluated individual contributions of (i) the RPEs i.–iv. that are situated in segment A of uORF1 (Munzarová et al. 2011; Gunišová and Valášek 2014), (ii) the CDS of uORF1 and in particular of its third sense codon, and (iii) the 3' sequence of uORF1. The biggest contribution comes from the RPEs (*uORF1-1111-Δ160*; ~79% drop), some of which like RPE i. and iv. require intact eIF3 for their function. The RPEs are followed by the autonomous 3' sequence (*uORF1-1114*; ~67% drop), and the uORF1 third sense codon [*uORF1-1111-Cys to Pro (CCG)*; ~49% drop] (Fig. 8A). Concurrent removal of the two biggest contributors in *uORF1-1114-Δ160* displayed a robust decrease in the REI activity (down to ~8%), strongly suggesting that their contributions are independent ($0.21 \times 0.33 = 0.07$, which pretty much equals the obtained 8%). A relatively smaller additive effect was observed when the removal of all RPEs was combined with the Cys to Pro substitution in *uORF1-1111-Δ160-Cys to Pro (CCG)* (16%), indicating their partial interdependence ($0.21 \times 0.51 = 0.11$, which is smaller than the obtained 16%). Testing the combination of the Cys to Pro substitution with the removal of the uORF1 3' sequence in the hypothetical *uORF1-only-1114-Cys to Pro (CCG)* construct is meaningless, because we previously showed that the otherwise autonomous activity of RPEs follows and in fact requires a prior action of the uORF1 3' sequences (Munzarová et al. 2011). Therefore, with respect to its minimal activity (~6%), the *uORF1-1114* construct shown in Figure 1C basically mimics the elimination of all three contributors in *uORF1-1114-Δ160-Cys to Pro (CCG)* reaching the uORF4-like basal REI activity (~5%) (Fig. 8A).

Finally, removal of the eIF3-independent RPE ii. together with the uORF2-specific, eIF3-dependent RPE v. (Gunišová and Valášek 2014), situated in the uORF2 segment A (*uORF2-2222-Δ233*), reduced the efficiency of REI by a robust ~77%, whereas the Cys to Pro substitution of the

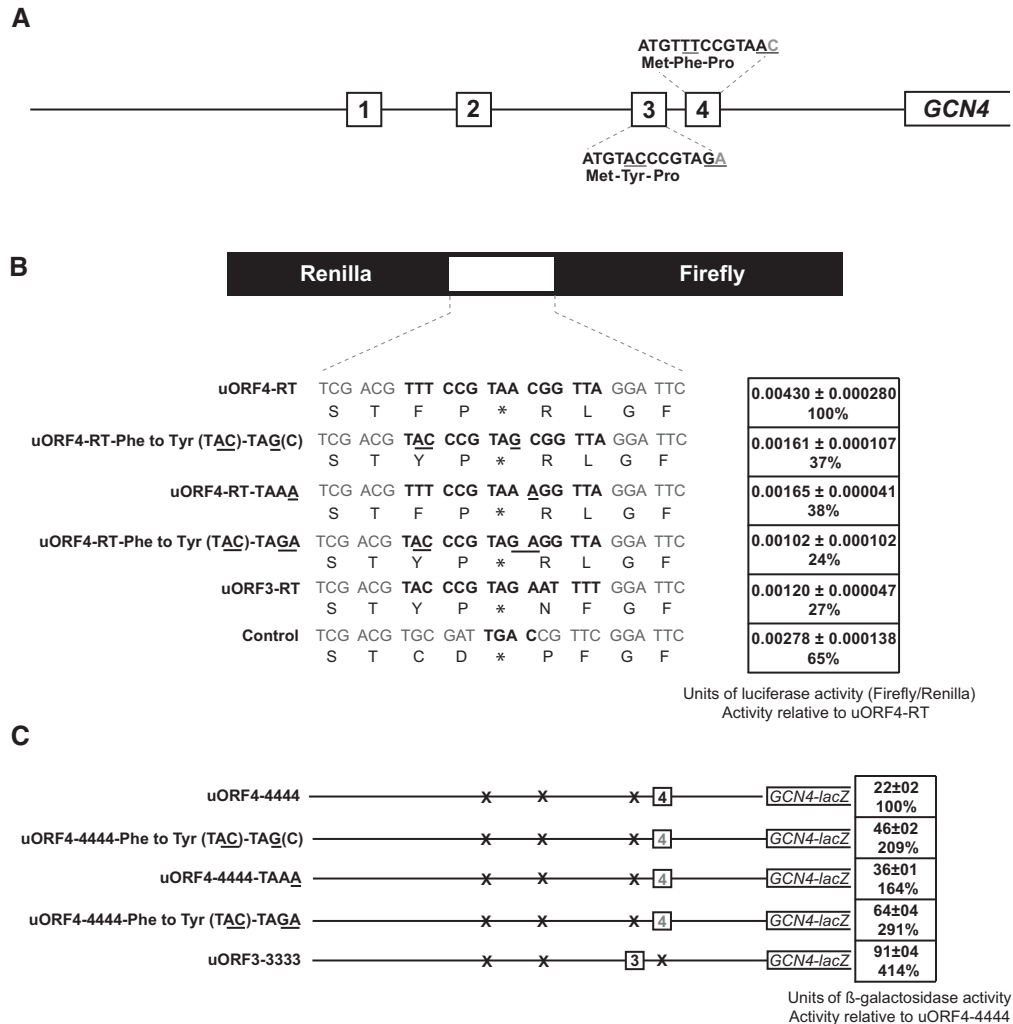


FIGURE 7. Poor termination efficiency at the uORF4 stop codon strongly contributes to the difference in the REI efficiency between uORF3 and uORF4. (A) Modified schematic from Figure 6A, in addition depicting the nucleotide following the stop codon of uORF3 and uORF4 (the +4 base; in gray). (B) Schematic of the standard dual luciferase readthrough reporter constructs adapted from Keeling et al. (2004). In the linker region, the 15-nt long termination sequences of uORF3 or uORF4 with their derivatives, as well as the control TGA-C terminating tetranucleotide, are shown in bold. The uORF4 derivatives have all their substitutions gradually turning them into wt uORF3 underlined. All indicated constructs were introduced into the PBH156 strain and the resulting transformants were grown in SD and processed for stop codon readthrough measurements as described in Materials and Methods. Values of luciferase activity are represented as mean values and standard deviations obtained from five independent transformants, and each experiment was repeated at least three times. Differences in luciferase activities between uORF4-RT-Phe to Tyr (TAC)-TAG or uORF4-RT-TAAA and their wt uORF4-RT, and between uORF4-RT and control TGAC plasmid were analyzed by Student's *t*-test and the calculated *P* = 0.0039 or 0.0004, and 0.0012, respectively. (C) The difference in the nucleotide following the stop codon of uORF3 and uORF4 (the +4 base) contributes to their varying REI permissiveness. Solitary uORF4 from the GCN4 mRNA leader in its authentic position was modified by indicated substitutions from uORF3 and analyzed as described in Figure 1C. Differences in β-galactosidase activities between uORF4-4444-TAAA and its wt uORF4-4444 were analyzed by Student's *t*-test and the calculated *P* = 0.0002.

uORF2 last coding triplet [uORF2-2222-Cys to Pro (CCG)] showed a similar reduction to that observed for uORF1 (~56% drop in REI) (Fig. 8B). Hence, as in the case of uORF1, the major contributor to the uORF2 REI potential is represented by its RPEs. Elimination of both of these *cis*-determinants at the same time in uORF2-2222-Δ233-Cys to Pro (CCG) expectedly reached the uORF4-like basal level of REI (~5%) (Fig. 8B), clearly illustrating the absolute independence of the uORF2 REI activity on its 3' sequences, as shown above.

Conclusion

Our systematic analysis of all potential *cis*-determinants that either promote or inhibit reinitiation on GCN4 mRNA revealed the following attributes of individual uORFs (summarized in Fig. 9). The 3' sequences of uORFs 1–3, in particular the first 12 nt immediately following their stop codons, contain a conserved AU₁₋₂A/UUAU₂ motif that promotes REI independently of other REI-promoting elements but only when situated at the defined distance from the GCN4 AUG

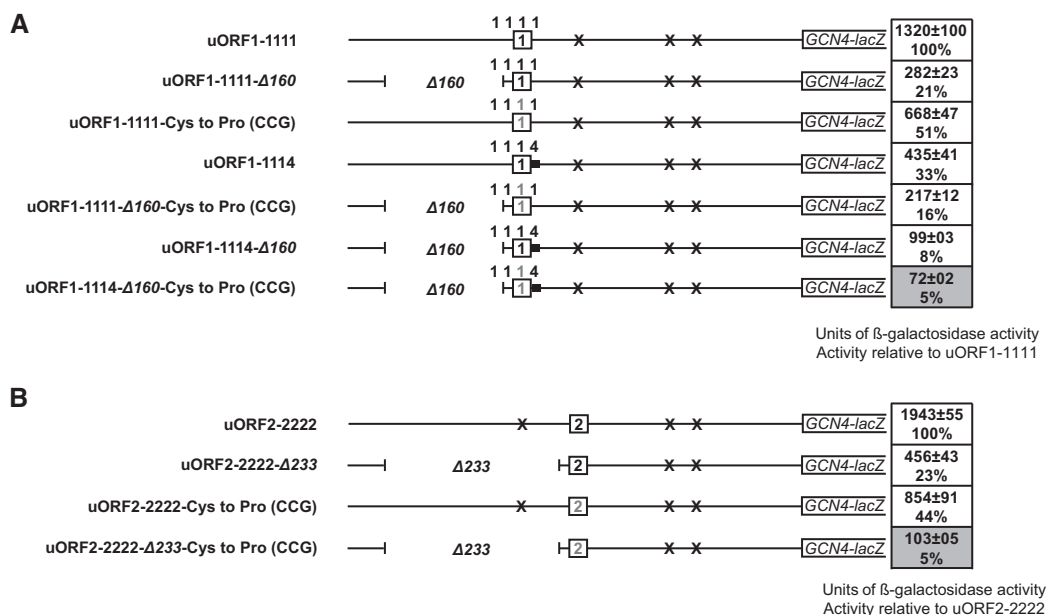


FIGURE 8. Combined effects of mutations in *cis*-determinants of uORF1 and uORF2 on their REI permissiveness. (A) The REI permissiveness of uORF1 is determined by three *cis*-factors: RPEs situated in the 5' sequence, CDS, and the 3' sequence. Solitary uORF1 from the *GCN4* mRNA leader in its authentic position and its derivatives containing the indicated mutations were analyzed as described in Figure 1C. Differences in β -galactosidase activities between uORF1-1111- Δ 160 and uORF1-1111- Δ 160-Cys to Pro (CCG), or uORF1-1114- Δ 160 and uORF1-1114- Δ 160-Cys to Pro (CCG) were analyzed by Student's *t*-test and the calculated *P* = 0.0369 or 0.0001, respectively. (B) REI permissiveness of uORF2 is determined by two *cis*-factors: RPEs situated in the 5' sequence and CDS. Solitary uORF2 from the *GCN4* mRNA leader in its authentic position and its derivatives containing the indicated mutations were analyzed as described in Figure 1C.

start codon, in principle corresponding to the position of uORF1. Hence, despite carrying this autonomous motif in their 3' sequences, uORF2 and uORF3 do not utilize it. Intriguingly, the 3' sequences of specifically these two uORFs in addition contain inhibitory elements that immediately follow the AU-rich motif and decrease the REI potential of these two uORFs. Interestingly, these inhibitory elements are transferable and function irrespectively of their distance from the *GCN4* start codon. Furthermore, we also revealed that the authentic length of both REI-permissive uORFs has to be maintained for their optimal activity and that the last coding triplet can most probably tolerate a wide range

of codons with the exception of the REI-inhibiting proline CCG triplet. Indeed, specifically this Pro triplet occurs as the last triplet in uORF3 and uORF4 and, in fact, features in all ultimate uORFs in the *GCN4* mRNA leaders across yeast species. Finally, we show that the approximately four-fold difference between the REI potential of modestly REI-permissive uORF3 and REI-nonpermissive uORF4 does not lie in the supposedly inhibitory 3' sequence of uORF4, as suggested before (Gunišová and Valášek 2014), but is manifested through the specific effects of the sequence composition of their second codon and of the identity of their stop codon tetranucleotide, which together impact the efficiency of stop

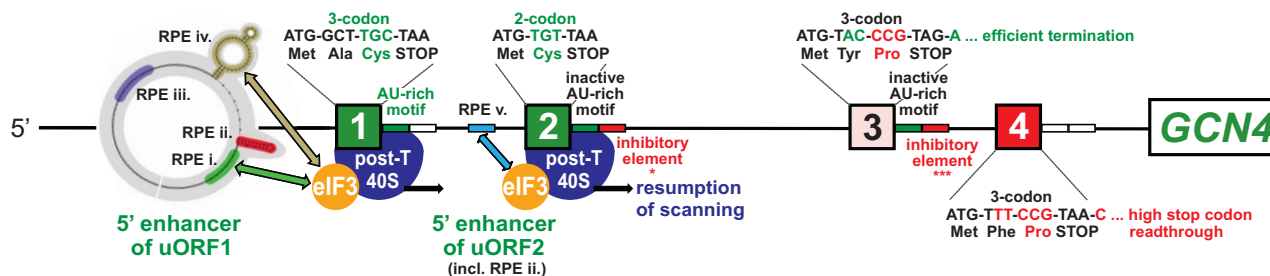


FIGURE 9. Summary of all *cis*-determinants that either promote or inhibit reinitiation on *GCN4* after translation of its four short uORFs. Schematics of the 5' enhancers of uORF1 and 2 containing their respective RPEs, some of which functionally interact with eIF3 to promote resumption of scanning, were taken from Gunišová and Valášek (2014). Green color-coding generally indicates stimulatory effects of the corresponding *cis*-factors on efficiency of REI, whereas red color-coding indicates inhibitory effects (with the exception of RPE ii. of uORF1, which is also stimulatory); the number of asterisks below the inhibitory elements of the uORF2 and uORF3 3' sequences illustrates the degree of their inhibition as determined experimentally. Please see text for further details.

codon recognition in a positive (uORF3) or negative (uORF4) way. In other words, we demonstrate for the first time that there is a direct negative correlation between the efficiency of reinitiation and efficiency of translation termination. Collectively this comprehensive approach pictures an intriguing complexity of this delicate regulatory system that depends on several REI-promoting as well as inhibiting features that mutually fine-tune their often autonomous effects on the overall efficiency of REI on *GCN4* mRNA in order to keep it as low as possible during nonstarvation conditions or as high as possible during starvation/stress conditions.

Even though there is a prevailing opinion that practically each short uORF is different and there are no generalizable rules that would apply to a majority of them (Wethmar 2014), we can conclude the following. The presence of the CCG triplet encoding proline as the last coding triplet most probably signals very poor REI potential of a given uORF, at least in yeast. Conversely, the presence of the RPE i.-like and/or RPE v.-like sequence motif not far upstream of a given short uORF, as well as the presence of the AU-rich motif immediately following the stop codon, might signal increased permissiveness for REI. An additional indicator of the REI permissiveness could be the presence of a structural element resembling the secondary structure of the uORF1 RPE iv., which can also be found in the 5' enhancer of the solitary uORF preceding the *YAP1* gene (Munzarová et al. 2011).

MATERIALS AND METHODS

Yeast strains, plasmids, and other biochemical methods

Lists of strains (Supplemental Table S1), plasmids (Supplemental Table S2), and PCR primers (Supplemental Table S3) used in this study and details of their construction can be found in the Supplemental Material. β -Galactosidase assays were conducted as described previously (Mueller et al. 1987; Grant and Hinnebusch 1994). For all β -galactosidase values recorded with mutant constructs that differed from their respective wt constructs by <40%, the *P*-values were calculated and are given in the corresponding figure legends.

Stop codon readthrough assay

Dual luciferase assay was performed using the Dual Luciferase Reporter Assay System (Promega). Briefly, yeast strain PBH156 was transformed with the indicated dual luciferase reporter non-sense control plasmid pTH477 (Fig. 7A) as well as with the sense codon (CAA) plasmid pTH460 and pSG358, pSG362, pSG349, pSG365, or pSG350. The experiments and data analysis were carried out according to the Microtiter plate-based dual luciferase protocol developed by Merritt et al. (2010) and commercially distributed by Promega. Assays were done in quintuplicates ($n = 5$), the data are expressed as the mean \pm SD, and each experiment was repeated at least three times. Resulting luciferase activity in each strain was expressed as the firefly/Renilla luciferase activity (nonsense or all indicated pSG plasmids) divided by the firefly/Renilla luciferase activity (sense). For further details, please see Keeling et al. (2004).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We are thankful to Josef Pánek for computational analysis, to Olga Krýdová for technical and administrative assistance, and to the members of the Valášek laboratory for helpful suggestions. This research was supported by the Czech Science Foundation Grant P305-11-0172.

Received October 26, 2015; accepted December 18, 2015.

REFERENCES

- Abastado JP, Miller PF, Jackson BM, Hinnebusch AG. 1991. Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for *GCN4* translational control. *Mol Cell Biol* **11**: 486–496.
- Aylett CH, Boehringer D, Erzberger JP, Schaefer T, Ban N. 2015. Structure of a yeast 40S-eIF1-eIF1A-eIF3-eIF3j initiation complex. *Nat Struct Mol Biol* **22**: 269–271.
- Beznosková P, Gunisová S, Valášek LS. 2015a. Rules of UGA-N decoding by near-cognate tRNAs and analysis of readthrough on short uORFs in yeast. *RNA* doi: 10.1261/rna.054452.115.
- Beznosková P, Wagner S, Jansen ME, von der Haar T, Valášek LS. 2015b. Translation initiation factor eIF3 promotes programmed stop codon readthrough. *Nucleic Acids Res* **43**: 5099–5111.
- Calvo SE, Pagliarini DJ, Mootha VK. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci* **106**: 7507–7512.
- Cuchalová L, Kouba T, Herrmannová A, Dányi I, Chiu W-I, Valášek L. 2010. The RNA recognition motif of eukaryotic translation initiation factor 3g (eIF3g) is required for resumption of scanning of posttermination ribosomes for reinitiation on *GCN4* and together with eIF3i stimulates linear scanning. *Mol Cell Biol* **30**: 4671–4686.
- Cvijović M, Dalevi D, Bilsland E, Kemp GJ, Sunnerhagen P. 2007. Identification of putative regulatory upstream ORFs in the yeast genome using heuristics and evolutionary conservation. *BMC Bioinformatics* **8**: 295.
- Dabrowski M, Bukowy-Bieryllo Z, Zietkiewicz E. 2015. Translational readthrough potential of natural termination codons in eukaryotes—the impact of RNA sequence. *RNA Biol* **12**: 950–958.
- Davuluri RV, Suzuki Y, Sugano S, Zhang MQ. 2000. CART classification of human 5' UTR sequences. *Genome Res* **10**: 1807–1816.
- Dever TE, Feng L, Wek RC, Cigan AM, Donahue TD, Hinnebusch AG. 1992. Phosphorylation of initiation factor 2 α by protein kinase GCN2 mediates gene-specific translational control of *GCN4* in yeast. *Cell* **68**: 585–596.
- Grant CM, Hinnebusch AG. 1994. Effect of sequence context at stop codons on efficiency of reinitiation in *GCN4* translational control. *Mol Cell Biol* **14**: 606–618.
- Grant CM, Miller PF, Hinnebusch AG. 1995. Sequences 5' of the first upstream open reading frame in *GCN4* mRNA are required for efficient translational reinitiation. *Nucleic Acids Res* **23**: 3980–3988.
- Gunisová S, Valášek LS. 2014. Fail-safe mechanism of *GCN4* translational control—uORF2 promotes reinitiation by analogous mechanism to uORF1 and thus secures its key role in *GCN4* expression. *Nucleic Acids Res* **42**: 5880–5893.
- Hinnebusch AG. 2005. Translational regulation of *GCN4* and the general amino acid control of yeast. *Annu Rev Microbiol* **59**: 407–450.
- Hood HM, Neafsey DE, Galagan J, Sachs MS. 2009. Evolutionary roles of upstream open reading frames in mediating gene regulation in fungi. *Annu Rev Microbiol* **63**: 385–409.

- Iacono M, Mignone F, Pesole G. 2005. uAUG and uORFs in human and rodent 5' untranslated mRNAs. *Gene* **349**: 97–105.
- Jackson RJ, Hellen CU, Pestova TV. 2012. Termination and post-termination events in eukaryotic translation. *Adv Protein Chem Struct Biol* **86**: 45–93.
- Keeling KM, Lanier J, Du M, Salas-Marco J, Gao L, Kaenjak-Angeletti A, Bedwell DM. 2004. Leaky termination at premature stop codons antagonizes nonsense-mediated mRNA decay in *S. cerevisiae*. *RNA* **10**: 691–703.
- Kouba T, Dányi I, Gunišová S, Munzarová V, Vlčková V, Cuchalová L, Neueder A, Milkereit P, Valášek LS. 2012. Small ribosomal protein RPS0 stimulates translation initiation by mediating 40S-binding of eIF3 via its direct contact with the eIF3a/TIF32 subunit. *PLoS One* **7**: e40464.
- Kozak M. 1987. Effects of intercistronic length on the efficiency of reinitiation by eukaryotic ribosomes. *Mol Cell Biol* **7**: 3438–3445.
- Kozak M. 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**: 13–37.
- Lee YY, Cevallos RC, Jan E. 2009. An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2 α phosphorylation. *J Biol Chem* **284**: 6661–6673.
- Merritt GH, Naemi WR, Mugnier P, Webb HM, Tuite MF, von der Haar T. 2010. Decoding accuracy in eRF1 mutants and its correlation with pleiotropic quantitative traits in yeast. *Nucleic Acids Res* **38**: 5479–5492.
- Miller PF, Hinnebusch AG. 1989. Sequences that surround the stop codons of upstream open reading frames in *GCN4* mRNA determine their distinct functions in translational control. *Genes Dev* **3**: 1217–1225.
- Mueller PP, Harashima S, Hinnebusch AG. 1987. A segment of *GCN4* mRNA containing the upstream AUG codons confers translational control upon a heterologous yeast transcript. *Proc Natl Acad Sci* **84**: 2863–2867.
- Munzarová V, Pánek J, Gunišová S, Dányi I, Szamecz B, Valášek LS. 2011. Translation reinitiation relies on the interaction between eIF3a/TIF32 and progressively folded *cis*-acting mRNA elements preceding short uORFs. *PLoS Genet* **7**: e1002137.
- Palam LR, Baird TD, Wek RC. 2011. Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. *J Biol Chem* **286**: 10939–10949.
- Pavlov MY, Watts RE, Tan Z, Cornish VW, Ehrenberg M, Forster AC. 2009. Slow peptide bond formation by proline and other *N*-alkylamino acids in translation. *Proc Natl Acad Sci* **106**: 50–54.
- Pisarev AV, Hellen CU, Pestova TV. 2007. Recycling of eukaryotic post-termination ribosomal complexes. *Cell* **131**: 286–299.
- Pöry TA, Kaminski A, Jackson RJ. 2004. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev* **18**: 62–75.
- Rajkowitsch L, Vilela C, Berthelot K, Ramirez CV, McCarthy JE. 2004. Reinitiation and recycling are distinct processes occurring downstream of translation termination in yeast. *J Mol Biol* **335**: 71–85.
- Raveh-Amit H, Maissel A, Poller J, Marom L, Elroy-Stein O, Shapira M, Livneh E. 2009. Translational control of protein kinase C η by two upstream open reading frames. *Mol Cell Biol* **29**: 6140–6148.
- Roy B, Vaughn JN, Kim BH, Zhou F, Gilchrist MA, Von Arnim AG. 2010. The h subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring upstream open reading frames. *RNA* **16**: 748–761.
- Skabkin MA, Skabkina OV, Hellen CU, Pestova TV. 2013. Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol Cell* **51**: 249–264.
- Sundaram A, Grant CM. 2014. A single inhibitory upstream open reading frame (uORF) is sufficient to regulate *Candida albicans* GCN4 translation in response to amino acid starvation conditions. *RNA* **20**: 559–567.
- Szamecz B, Rutkai E, Cuchalová L, Munzarová V, Herrmannová A, Nielsen KH, Burela L, Hinnebusch AG, Valášek L. 2008. eIF3a cooperates with sequences 5' of uORF1 to promote resumption of scanning by post-termination ribosomes for reinitiation on GCN4 mRNA. *Genes Dev* **22**: 2414–2425.
- Valášek LS. 2012. 'Ribozoomin'—translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). *Curr Protein Pept Sci* **13**: 305–330.
- Valášek L, Mathew AA, Shin BS, Nielsen KH, Szamecz B, Hinnebusch AG. 2003. The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40S ribosome in vivo. *Genes Dev* **17**: 786–799.
- Vilela C, Linz B, Rodrigues-Pousada C, McCarthy JE. 1998. The yeast transcription factor genes YAP1 and YAP2 are subject to differential control at the levels of both translation and mRNA stability. *Nucleic Acids Res* **26**: 1150–1159.
- von Arnim AG, Jia Q, Vaughn JN. 2014. Regulation of plant translation by upstream open reading frames. *Plant Sci* **214**: 1–12.
- Wethmar K. 2014. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip Rev RNA* **5**: 765–778.
- Wohlgemuth WA, Safonova O, Engelhardt M, Freitag M, Wölfl K, Kirchhof K. 2008. Improvement of the quality of life concerning the health of patients with peripheral arterial disease (PAD) after successful bypass surgery. *Vasa* **37**: 338–344.
- Zhou F, Roy B, von Arnim AG. 2010. Translation reinitiation and development are compromised in similar ways by mutations in translation initiation factor eIF3h and the ribosomal protein RPL24. *BMC Plant Biol* **10**: 193.