3980–3988 Nucleic Acids Research, 1995, Vol. 23, No. 19

Sequences 5′ of the first upstream open reading frame in *GCN4* mRNA are required for efficient translational reinitiation

Chris M. Grant+, Paul F. Miller§ and Alan G. Hinnebusch*

Section on Molecular Genetics of Lower Eukaryotes, Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, Bethesda, MD 20892, USA

Received May 11, 1995; Revised and Accepted August 6, 1995

ABSTRACT

Translation of yeast *GCN4* mRNA occurs by a reinitiation mechanism that is modulated by amino acid levels in the cell. Ribosomes which translate the first of four upstream open reading frames (uORFs) in the mRNA leader resume scanning and can reinitiate downstream. Under non-starvation conditions reinitiation occurs at one of the remaining three uORFs and *GCN4* is repressed. Under starvation conditions, in contrast, ribosomes bypass the uORFs and reinitiate at *GCN4* instead. The high frequency of reinitiation following uORF1 translation depends on an adequate distance to the next start codon and particular sequences surrounding the uORF1 stop codon. We present evidence that sequences 5′ to uORF1 also strongly enhance reinitiation. First, reinitiation was severely inhibited when uORF1 was transplanted into the position of uORF4, even though the native sequence environment of the uORF1 stop codon was maintained, and this effect could not be accounted for by the decreased uORF1–*GCN4* spacing. Second, insertions and deletions in the leader preceding uORF1 greatly reduced reinitiation at *GCN4*. Sequences 5′ to uORF1 may influence the probability of ribosome release following peptide termination at uORF1. Alternatively, they may facilitate rebinding of an initiation factor required for reinitiation prior to resumption of the scanning process.

INTRODUCTION

The *GCN4* gene of the yeast *Saccharomyces* is regulated by a unique translational control mechanism involving four short open reading frames (uORFs) located upstream of the protein coding sequence in *GCN4* mRNA (reviewed in 1,2). When yeast cells are grown under conditions of amino acid abundance *GCN4* translation is repressed to very low levels, primarily by the third and fourth of these uORFs (numbered from the 5′-end). In response to starvation of one or more amino acids the inhibitory effect of these uORFs is partially overcome, allowing increased translation of the *GCN4* coding sequences. The GCN4 protein thus produced activates transcription of multiple amino acid biosynthetic genes and thereby restores an abundant supply of amino acids.

Extensive genetic analysis of the *GCN4* mRNA leader has provided a detailed model for *GCN4* translational control. According to this model ribosomes scan from the 5′-end of the mRNA and translate uORF1, after which ~50% of the ribosomes remain attached to the mRNA and resume scanning downstream. Under non-starvation conditions essentially all of these ribosomes reinitiate at uORFs 2, 3 or 4 and, after translating these uORFs, dissociate from the mRNA. Thus *GCN4* translation is prevented. In amino acid-deprived cells it appears that ~50% of the ribosomes that resume scanning following translation of uORF1 scan past the start codons at uORFs 2–4 without initiating translation and reinitiate farther downstream at *GCN4* instead (2–4).

The differential utilization of the start sites at uORFs 2–4 and at *GCN4* during the reinitiation events that follow translation of uORF1 appears to be governed by the availability of eIF-2–GTP–Met-tRNA$_i^{Met}$ ternary complexes in the cell. It is thought that under non-starvation conditions these ternary complexes are very abundant and rapidly rebind to ribosomes following termination at uORF1. This makes the ribosomes competent to reinitiate translation by the time they reach uORFs 2, 3 or 4. Under starvation conditions ternary complexes are less abundant and more time is required to rebind to the ribosomes scanning downstream from uORF1. Consequently, a significant number cannot reinitiate at uORFs 2–4 because they lack the ternary complex when they arrive at the start codons of these uORFs. It appears that nearly all such ribosomes that bypass uORFs 2–4 rebind the ternary complex while scanning the interval between uORF4 and *GCN4*, enabling them to reinitiate at *GCN4* (3). The concentration of ternary complexes is reduced in amino acid-starved cells because the subunit of eIF-2 is phosphorylated on Ser51 by the protein kinase GCN2 (2). It is thought that phosphorylated eIF-2 inhibits the recycling of eIF-2 from its inactive GDP-bound state to the active GTP-bound form in yeast cells, just as occurs in mammalian cells (reviewed in 5). GCN2 is present in a latent form under non-starvation conditions and

* To whom correspondence should be addressed
Present addresses: +School of Biochemistry and Molecular Genetics, University of New South Wales, Sydney 2052, Australia and §Infectious Diseases Department, Pharmaceutical Research Division, Warner-Lambert Co., Ann Arbor, MI 48106-1047, USA

appears to be activated by uncharged tRNA which accumulates in amino acid-deprived cells (2,6).

The ability of ribosomes to remain attached to the mRNA and resume scanning after translating uORF1 is a critical aspect of the *GCN4* regulatory mechanism, because only those ribosomes in the process of reassembling an initiation complex after translating uORF1 can bypass uORFs 2–4 when ternary complex levels are diminished under starvation conditions. Thus virtually no *GCN4* expression occurs with a construct containing uORF4 but lacking uORF1. An equally important aspect of this control mechanism is the nearly complete inability of ribosomes to reinitiate at *GCN4* following translation of uORF4. This extremely low level of reinitiation is typical of effects seen with other uORFs introduced into yeast mRNAs (7,8), underscoring the unique properties of uORF1 that enable ~50% of the ribosomes which translate this sequence to resume scanning and reinitiate downstream (9,10). The much lower reinitiation efficiency of uORF4 versus uORF1 cannot be attributed merely to the fact that uORF4 is 200 nt closer than uORF1 to the *GCN4* start site, because increasing the distance between uORF4 and *GCN4* by this amount has little effect on *GCN4* expression (3). Moreover, uORF4 remains much more inhibitory than uORF1, even when a 40 nt segment containing uORF4 and its immediately flanking nucleotides is inserted in the leader in place of uORF1 (10). Exploiting this last finding, we showed that the final codon of uORF1 and 10 nt immediately 3′ to its stop codon are very important in determining the ability of ribosomes to reinitiate at *GCN4* following translation of uORF1 (10). A recent mutational analysis has shown that numerous A+U-rich sequences at these 13 nt positions are compatible with efficient reinitiation downstream following translation of uORF1. This last observation led to the idea that formation of G-C base pairs between nucleotides in rRNA, tRNA or elsewhere in *GCN4* mRNA and those surrounding the uORF1 stop codon will enhance ribosome dissociation from the mRNA and thereby impair reinitiation downstream at *GCN4* (11).

As mentioned above, when a small segment of the leader containing uORF4 was inserted in place of the corresponding segment containing uORF1 the transplanted copy of uORF4 retained its strong inhibitory effect on *GCN4* translation (10). From this finding it seemed possible that the sequences immediately surrounding the stop codons of uORF1 and uORF4 would be the sole determinants of the widely different abilities of these elements to allow reinitiation at *GCN4*. This simple possibility has been eliminated by the results presented in this report. We found that uORF1 is unable to support high level reinitiation at *GCN4* when inserted in the mRNA leader in place of uORF4, even when the transplanted copy of uORF1 contains the critical 13 nt that normally surround its stop codon. In pursuing an explanation for this observation we made the unexpected finding that sequences located 5′ to uORF1 make an important contribution to the high frequency of reinitiation following uORF1 translation, whereas the native sequences located upstream of uORF4 are inhibitory to reinitiation at *GCN4*. Together with previous findings, these results indicate that the *GCN4* mRNA leader is highly specialized to ensure the unusually high efficiency of reinitiation following uORF1 translation that is critical for regulating *GCN4* expression. In addition, the identification of a sequence element located 5′ to uORF1 that enhances reinitiation downstream of the uORF may help uncover new modes of regulating the behavior of ribosomes during translation termination events.

MATERIALS AND METHODS

Construction of mutant *GCN4* alleles

Plasmid constructions were generated by following standard procedures (12) or by PCR with oligonucleotides specific for *GCN4* sequences. All constructs are derivatives of plasmid-borne *GCN4* alleles contained on the *Escherichia coli*–yeast shuttle vector YCp50, which contains the yeast *URA3*, *ARS1* and *CEN4* sequences for selection and single copy maintenance in *S.cerevisae*, as described previously (3,4,9–11). Plasmids p235, p237 and p238 were constructed previously (9), as were pM199 and pG30 (4). p235, p237 and p238 contain point mutations in the ATG codons of uORFs 2–4, 1–3 and 1–4 respectively. pM199 contains the *GCN4* leader sequence extending from the 5′-end to a *Bgl*II site, introduced by a point mutation 30 nt downstream of the uORF1 stop codon (10), joined to a *Hin*dIII site introduced by a point mutation 44 nt downstream of the uORF4 stop codon (3) by a linker which restores the wild-type length separating uORF4 from the *GCN4* ATG codon (4). pG30 is identical to pM199 except for a point mutation that removes the uORF1 ATG codon (4).

Plasmid pM189 was constructed by first replacing the 78 nt *Sna*BI–*Hin*dIII segment of pA44 (3) containing uORF4 with the analogous sequences from uORF1 using synthetic oligonucleotides. The 5′ copy of uORF1 was removed from the new plasmid, pM183, by exchanging its *GCN4 Eco*RI fragment with the corresponding fragment from p238 (9), creating pM189. Plasmid pG148 was derived from pM189 by replacing the *Sna*BI–*Hin*dIII leader fragment with identical sequences except for G→C and C→A substitutions that changed the uORF1 ATG codon to ATC and created a *Cla*I restriction site.

The 5′ nested deletions in pM189 were constructed by replacing the *Sal*I–*Sna*BI leader fragment with PCR-generated leader fragments having progressively shortened 3′-ends. These were made by using a common 5′ PCR primer containing the *Sal*I site and 3′ *Sna*BI primers that deleted the following lengths of leader upstream from the *Sna*BI site in pM189: pM207, 42 nt; pG152, 89 nt; pG160, 144 nt; pG153, 166 nt; pG154, 210 nt. The uORF1–*GCN4* overlap constructs pG150 and pG151 were made using construct pM226 (4), in which the uORF1 TAA codon was mutated by insertion of a T residue between the two A residues and two in-frame stop codons were mutated, resulting in an elongated uORF1 of 93 codons. The 344 nt *Sna*BI–*Bam*HI leader fragments of pM189 and pG154 were substituted with a similar PCR-generated fragment from pM226, which was made using a 5′ primer that introduces a *Sna*BI site 26 nt upstream of uORF1 and a 3′ primer that contains the *Bam*HI site within the *GCN4* coding region. Thus uORF1 overlaps and is out of frame with *GCN4* in plasmids pG150 and pG151.

The insertions upstream from uORF1 in pM199 were made by inserting one (pM211), two (pM212) or three (pM213) copies of the S1 oligonucleotide (3) into the *Hin*dIII site of pM199. Insertions in the correct orientation were verified by restriction enzyme analysis, since the oligonucleotide was designed so that the upstream *Hin*dIII cloning junction was destroyed. pG149, generated by PCR, is identical to pM213 except for a T→A substitution which changes the ATG codon of uORF1 to AAG. pG155 was constructed essentially as described previously for pM23 (10) except that pG155 contains only the *Hin*dIII site introduced upstream of uORF1 and lacks the *lacZ* fragment.

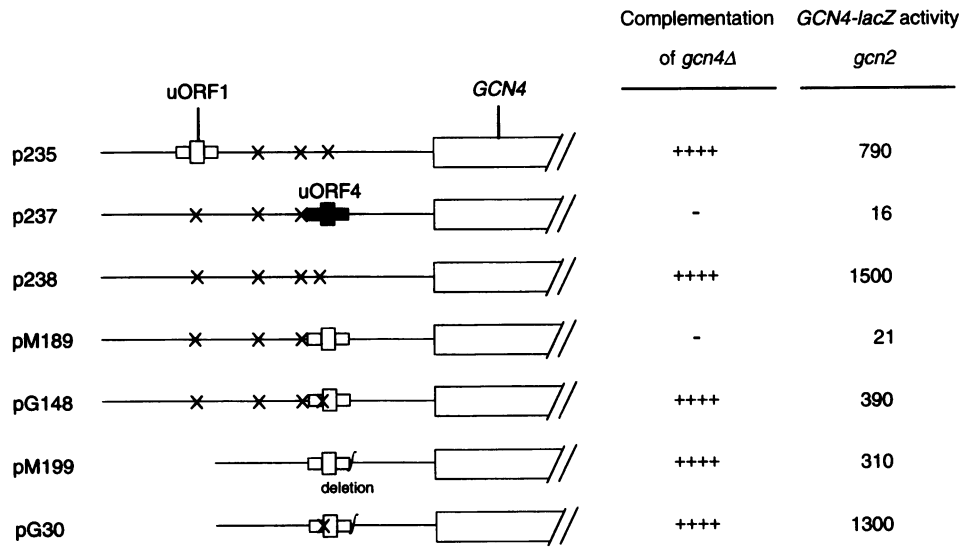

Figure 1. Efficient reinitiation following uORF1 translation is abolished by placing uORF1 at the position in the leader normally occupied by uORF4. Schematic showing the *GCN4* mRNA leader. uORF1 and *GCN4* are shown as open boxes and uORF4 as a black box. The rectangles attached to uORF1 and uORF4 designate the 16 and 25 nt flanking the uORFs on the 5′ and 3′ sides respectively. Xs indicate point mutations in the ATG codons of uORF1–uORF4. The wavy lines in constructs pM199 and pG30 signify deletion junctions. Constructs are drawn approximately to scale. In construct pM189 uORF1 and 41 flanking nt are inserted in place of the corresponding nucleotides of uORF4. pG148 is identical to pM189 except for a point mutation removing the uORF1 ATG codon. PM199 and pG30 contain a deletion between uORF1 and *GCN4* which moves uORF1 to a position 140 nt upstream of the *GCN4* start site. *GCN4* expression from these constructs was quantified in two different ways. First, *GCN4* constructs were tested for complementation of a chromosomal *gcn4* deletion in strain H384 by measuring the growth rate of transformants after replica plating to medium supplemented with 3-AT. Growth was scored after 2 or 3 days at 30°C. Second, β-galactosidase activities expressed from the corresponding *GCN4–lacZ* fusions were measured in two to four independent transformants of the non-derepressible *gcn2-1* strain H15. The individual measurements differed from the mean value by <28%. For comparison, *GCN4–lacZ* expression in the *gcn2* strain from constructs containing the wild-type leader sequence or one containing uORFs 1 and 4 alone due to point mutations in the start codons of uORFs 2–3, is 5 and 15 U respectively (9).

Insertions upstream from uORF1 in pG155 were made by replacing the *Sal*I–*Hin*dIII leader fragment with the analogous fragments from the pM199 series just described. pG159 was constructed by inserting a PCR-generated *Hin*dIII fragment, extending from 11 nt downstream of the 5′-end of the *GCN4* mRNA (13) to the *Hin*dIII site upstream from uORF1, into the *Hin*dIII site of pG155. This insertion resulted in a duplication of 200 nt of sequences preceding uORF1.

The 5′ nested deletions in pG36 (11) were made by replacing the 400 nt *Sal*I–*Hin*dIII leader fragment with PCR-generated leader fragments having progressively shortened 3′-ends. These were made by using a common 5′ primer containing the *Sal*I site and 3′ *Hin*dIII primers that deleted the following lengths of leader: pG164, 80 nt; pG163, 120 nt; pG162, 160 nt. pG166, generated by PCR, is identical to pG162 except for G→C and C→A changes.

GCN4–lacZ fusion derivatives of all of the above plasmids were made by inserting a 3.2 kb *Bam*HI fragment containing codons 9–1023 of *lacZ* at the *GCN4 Bam*HI site (13).

Assays of *GCN4* expression

GCN4 expression was measured in two different ways. First, plasmid-borne mutant *GCN4* alleles were introduced by transformation (14) into the *GCN4* deletion strain H384 (*Matα his1-29 gcn4-103 ura3-52*) and tested for the ability to restore growth on minimal medium lacking histidine (–His) and containing 3-aminotriazole (3-AT), as previously described (9). The *his1-29* allele has a leaky mutation that does not prevent growth on –His medium provided that *GCN4*-mediated derepression of *his1-29* transcription can occur. 3-AT inhibits the activity of the *HIS3*-encoded enzyme in the histidine biosynthetic pathway and derepression of *HIS3* transcription by *GCN4* is required for growth on media containing 3-AT. In the second method we measured β-galactosidase activity expressed from plasmid-borne *GCN4–lacZ* fusions containing the relevant leader mutations in the non-derepressible *gcn2* mutant H15 (*Matα gcn2-1 leu2-3 leu2-112 ura3-52*) and in the constitutively derepressed *gcd1* mutant F98 (*Matα gcd1-101 ura3-52*), all as previously described (9). [β-Galactosidase activity is expressed as nmol *o*-nitrophenyl-β-D-galactopyranoside (ONPG) hydrolyzed/min/mg total protein (U).] The level of 3-AT resistance conferred by the *GCN4* alleles in strain H384 generally correlated well with expression of the corresponding *GCN4–lacZ* fusions in strain F98.

RNA blot hybridization analysis

Total RNA was isolated from transformants of strain H384 grown in minimal medium in the presence of sulfometuron methyl to impose starvation for leucine, isoleucine and valine, exactly as described previously (15). RNA was fractionated by formaldehyde–agarose gel electrophoresis and subjected to blot hybridization analysis with radiolabeled probes for *GCN4* and pyruvate kinase (*PYK1*) mRNAs, all as described previously (16). The following constructs were analyzed: p235, pG165, pG162,

pM189, pG154, pG151, pM199, pM213, pG149, pG159, pM212, pM211, pG152, pG160, pG153, pG36, pG164 and pG163.

RESULTS

Sequences surrounding the uORF1 stop codon are not the sole determinants of reinitiation following uORF1 translation

A nearly wild-type pattern of *GCN4* translational control can be observed in the absence of uORFs 2 and 3, provided that uORFs 1 and 4 are present at their normal locations in the mRNA leader (9). To study the ability of uORF1 or uORF4 to allow ribosomes to resume scanning and reinitiate downstream following their own translation we analyzed *GCN4* constructs of even simpler design in which these uORFs were present alone in the leader, the other three having been removed by point mutations in their ATG start codons. We showed previously that the presence of uORF4 alone reduces *GCN4* translation to only ~1% of the level that occurs in the absence of all four uORFs (compare p237 and p238 in Fig. 1), whereas uORF1 alone decreases *GCN4* translation by only ~50% (compare 235 and p238 in Fig. 1) (9). Given that essentially no ribosomes scan past uORF1 without initiating translation (4), the high level of *GCN4* expression that occurs in the presence of uORF1 indicates that reinitiation is much more efficient following translation of uORF1 than uORF4. In previous studies we demonstrated that the efficiency of reinitiation is strongly affected by nucleotides immediately surrounding the uORF1 stop codon (10,11,17). To determine whether any other sequences flanking uORF1 are required for efficient reinitiation at *GCN4* a small segment containing the uORF1 coding region plus 16 nt 5′-flanking and 25 nt 3′-flanking sequences was inserted in place of the corresponding sequence at uORF4, in a leader construct lacking the start codons at uORF1, uORF2 and uORF3 (Fig. 1, pM189). If the only nucleotides in the vicinity of uORF1 that affect reinitiation are those surrounding its stop codon we would expect to observe high level *GCN4* expression when the uORF1 cassette was inserted in place of uORF4 in pM189, comparable with that given by construct p235 containing uORF1 at its normal location. Instead we found that *GCN4* expression from pM189 was very low and more similar to that seen for the p237 construct containing wild-type uORF4 (Fig. 1).

The low level *GCN4* expression observed with pM189 did not result simply from a novel junction which inhibits ribosomal scanning or lowers the mRNA level introduced in making this construct, because a point mutation that removed the uORF1 ATG codon in pM189 restored high level *GCN4* expression (Fig. 1, pG148). Moreover, the mRNA levels for constructs p235 and pM189 were not significantly different (Fig. 2, lanes 1 and 4). Thus it appears that most ribosomes translate the uORF1 sequence in pM189 and fail to reinitiate at *GCN4*. It is noteworthy, however, that *GCN4* expression from pG148 was lower than that seen for p238, a construct that contains the wild-type leader sequence with point mutations in the start codons of all four uORFs (Fig. 1). This last comparison suggests that the novel junction sequence or a structure created by the mutation in pM189 does inhibit scanning to some degree independently of reinitiation events. We can eliminate this unintended effect from consideration by comparing *GCN4–lacZ* expression from pM189 with that given by pG148, because pG148 contains the same novel junctions as pM189 but lacks the uORF1 ATG codon. This

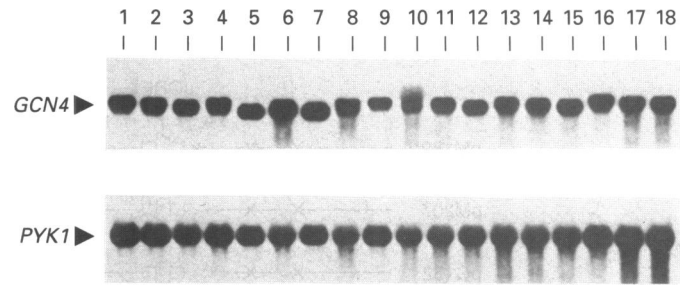

Figure 2. RNA blot hybridization analysis of selected *GCN4* constructs. Total RNA was isolated from cells grown under starvation conditions and subjected to blot hybridization analysis with radiolabeled probes for *GCN4* and pyruvate kinase (*PYK1*) mRNAs. The following constructs were analyzed: 1, p235; 2, pG165; 3, pG162; 4, pM189; 5, pG154; 6, pG151; 7, pM199; 8, pM213; 9, pG149; 10, pG159; 11, pM212; 12, pM211; 13, pG152; 14, pG160; 15, pG153; 16, pG36; 17, pG164; 18, pG163. pG149 (lane 9) was the only construct that showed a significant reduction in *GCN4* mRNA, relative to *PYK1* mRNA, compared with the p235 transcript, which is expressed at wild-type levels (9). The pG151 construct (lane 6) appeared to give somewhat higher than wild-type levels of *GCN4* mRNA. In this blot most of the pG159-encoded transcript was not larger than the p235 transcript (lanes 1 and 10). However, in replicate blots carried out prior to this one essentially all of the pG159 transcript migrated with the expected lower mobility characteristic of the minority fraction visible in lane 10. We presume that the pG159 transcript underwent degradation *in vitro* during the course of this analysis.

comparison indicates that only 5% of the ribosomes (21 U/390 U) reinitiate at *GCN4* following translation of uORF1 when the uORF1 segment is inserted in place of uORF4. Thus the uORF1 sequence in pM189 does not block reinitiation as completely as uORF4 does in p237, allowing 5 versus 1% reinitiation, however, it clearly permits much less reinitiation than occurs when uORF1 is present at its wild-type location in p235 (50%).

The reduced level of reinitation seen with construct pM189 cannot be explained by the smaller distance separating uORF1 from *GCN4* compared with the wild-type uORF1–*GCN4* spacing. This conclusion was established by the fact that construct pM199, in which uORF1 was moved downstream into the position of uORF4 by deleting all sequences between uORF1 and uORF4 (4), gives a much higher level of *GCN4* expression than does pM189 (Fig. 1). By comparing expression from pM199 with that of pG30, a derivative of pM199 with a point mutation in the uORF1 start codon, we calculated that 24% of the ribosomes that translate uORF1 in the pM199 construct can reinitiate at *GCN4* (310 U/1300 U). This level of reinitiation is 50% of that calculated for p235, in which uORF1 is present at its normal location. The lower reinitiation efficiency observed for pM199 versus p235 has been attributed to the decreased distance scanned by ribosomes following translation of uORF1 before reaching *GCN4* in the pM199 construct (4). It is important to note that the efficiency of reinitiation for pM199 is still 5-fold higher than that of pM189 (24 versus 5%), even though the distance between uORF1 and *GCN4* is identical in these two constructs (Fig. 1). This last comparison suggests that sequences 5′ of uORF1 that differ between pM199 and pM189 have a substantial effect on the efficiency of reinitiation.

It seemed likely that nucleotides normally found between uORF1 and uORF4 that are now juxtaposed 5′ to uORF1 in pM189 are responsible for the low reinitiation efficiency of this construct relative to pM199. According to this hypothesis the novel sequences present 5′ to uORF1 would largely offset the ability of

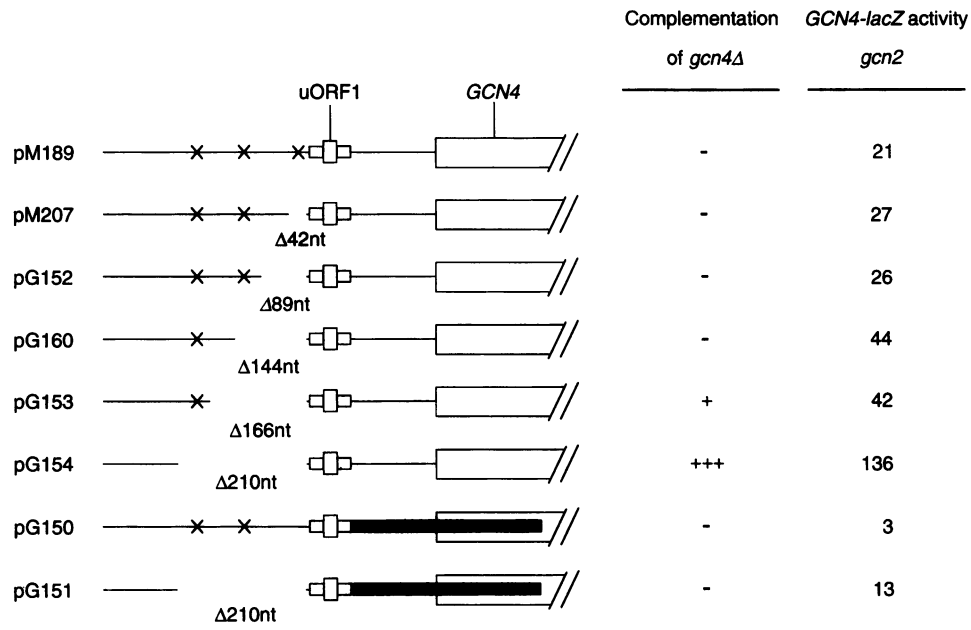

Figure 3. The sequences present between uORF1 and uORF4 inhibit reinitiation following uORF1 translation when they are located 5′ of uORF1. Schematic of pM189, already depicted in Figure 1, and constructs pM207, pG152, pG160, pG153 and pG154 in which 42, 89, 144, 166 and 210 nt respectively normally present upstream of uORF4 have been deleted. pG151 is identical to pG154 except for a 1 bp insertion in the stop codon of uORF1 as well as point mutations in two downstream in-frame termination codons, which together lengthen uORF1 (by the sequences shown as shaded) and cause it to overlap the *GCN4* start codon. Analysis of *GCN4* expression was conducted exactly as described in the legend to Figure 1. The individual β-galactosidase measurements differed from the mean value by <16%.

the native sequences surrounding the uORF1 stop codon to promote reinitiation at *GCN4*. In an effort to locate the putative inhibitory sequences we made progressive deletions of the uORF1–uORF4 interval in construct pM189. As shown in Figure 3, deleting the sequences immediately upstream from the uORF1 cassette in pM189 led to higher levels of *GCN4* expression, without any increase in the mRNA levels (Fig. 2, lanes 4, 5 and 13–15). The results in Figure 3 could be interpreted as indicating that two different sequence elements contribute to the inefficient reinitation that occurs following translation of uORF1 in construct pM189. In this view one inhibitory element would occur between the deletion junctions of pG152 and pG160 and the second would be located between the deletion junctions of pG153 and pG154. Alternatively, it is possible that each of the different sequences juxtaposed 5′ to uORF1 in the series of constructs pM189 to pG153 is incompatible with reinitiation and in construct pG154 the sequences 5′ to uORF1 have been restored nearly to the wild-type configuration, which is permissive for reinitiation. We noted that *GCN4* expression from pG154 (Fig. 3) was lower than that given by construct pM199 (Fig. 1), which contains very similar but not identical sequences in the vicinity of uORF1. This difference cannot be explained by a reduction in the mRNA level, because pG154 produces essentially wild-type levels of *GCN4* mRNA (Fig. 2, lanes 1,4,5). Instead, we attribute it to the aforementioned inhibitory junction produced when the uORF1-containing segment was inserted at uORF4 in pM189. Correcting for this effect by comparing pG154 with pG148, we estimate that the reinitiation efficiency for construct pG154 is 35%, similar to the value of 24% calculated for pM199.

An alternative explanation for the increased *GCN4* expression seen with pG154 versus pM189 (Fig. 3) is that ribosomes fail to recognize uORF1 in construct pG154, due to a sequence alteration 5′ to the uORF1 start site, and instead initiate at *GCN4* as the first initiation event on the mRNA. This possibility was eliminated by removing the stop codon of uORF1 in pG151, thereby elongating uORF1 and making it overlap the beginning of *GCN4* by 140 nt (Fig. 3). If the increased expression from pG154 versus pM189 resulted from 'leaky' scanning past the uORF1 start codon, the level of *GCN4* expression should not be greatly reduced by changing the termination site of uORF1 in construct pG154. Instead we found that elongating uORF1 in construct pG154 (yielding pG151) led to a dramatic decrease in *GCN4* expression. This indicates that the majority of ribosomes scanning the pG154 transcript initiate translation at uORF1, rather than bypassing the uORF1 start codon; these ribosomes reinitiate at *GCN4* when uORF1 is only three codons long (in construct pG154), but cannot do so when uORF1 is elongated and extensively overlaps the beginning of *GCN4* in pG151. We conclude that by deleting the uORF1–uORF4 interval from construct pM189 we increased *GCN4* translation by restoring efficient reinitiation following uORF1 translation.

Sequences 5′ to uORF1 promote reinitiation following uORF1 translation

As an additional test of the idea that sequences upstream of uORF1 affect the efficiency of reinitiation we made insertion mutations at a site 21 nt upstream from uORF1 in construct pM199 (Fig. 4A). As indicated above, uORF1 was moved into the position of uORF4 in construct pM199 by deleting the entire uORF1–uORF4 interval. We presume that this construct exhibits high level reinitiation because wild-type sequences are still present both upstream of uORF1 and immediately surrounding its stop codon. One to three copies of a 72 nt sequence called S1 (3),

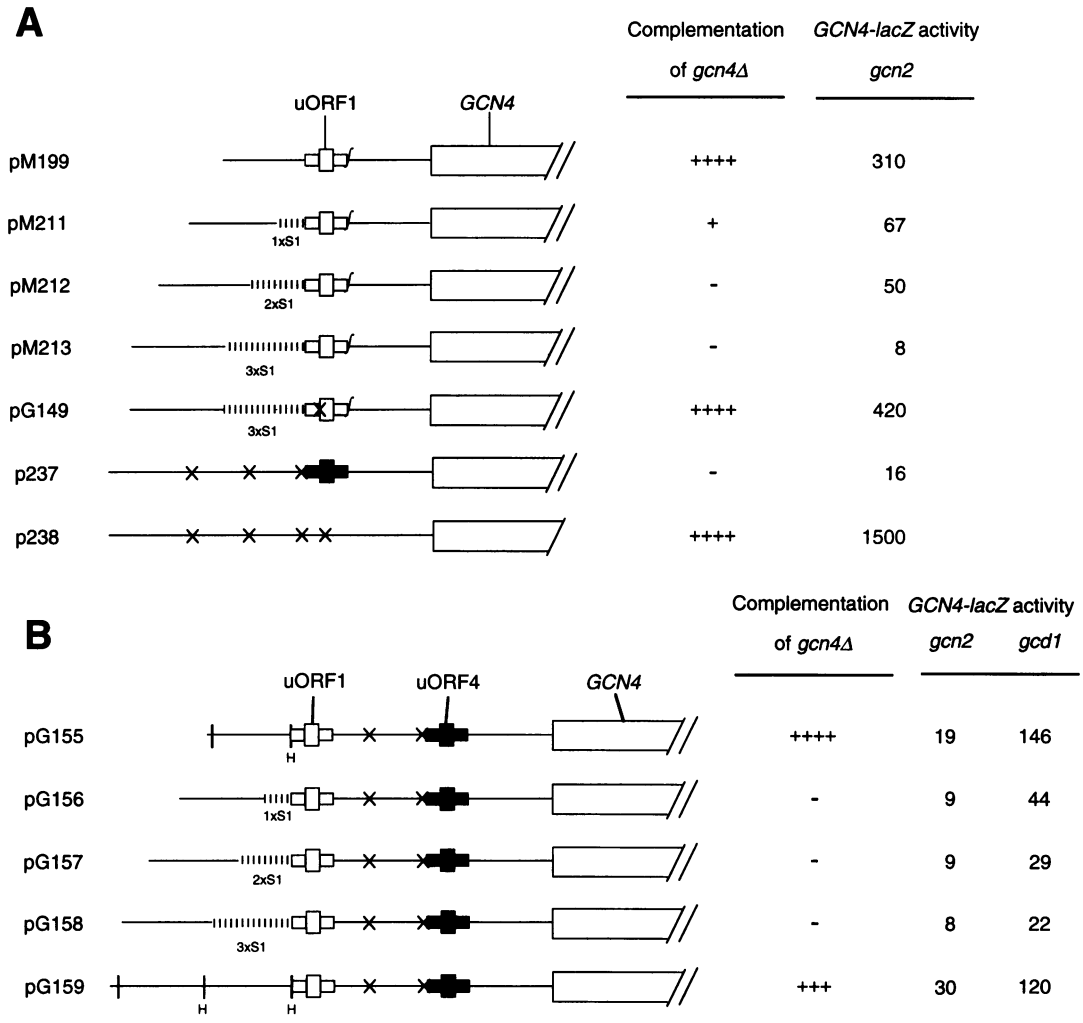

Figure 4. Insertions between uORF1 and the leader sequences normally present upstream of it abolish reinitiation at *GCN4* following uORF1 translation. (**A**) Schematic of construct pM199, already depicted in Figure 1, and constructs pM211, pM212 and pM213 in which one, two or three copies of a 73 nt spacer oligonucleotide (designated as a string of vertical lines) are inserted upstream of uORF1. The deletion junction 3′ of uORF1 is shown as a wavy line. pG149 is identical to pM213 except for a T→A substitution that changes the ATG codon of uORF1 to AAG. p237 and p238, already shown in Figure 1, are included for comparison. (**B**) Constructs equivalent to those in (A) except that they contain uORF4. H, a *Hin*dIII site used in constructing the duplication of 200 nt of sequences preceding uORF1 in pG159. Analysis of *GCN4* expression was conducted exactly as described in the legend to Figure 1. In addition, *GCN4–lacZ* fusions were assayed in the constitutively derepressed *gcd1-101* strain F98 for the constructs shown in (B). The individual β-galactosidase measurements differed from the mean value by <17%.

an A+T-rich sequence found between uORF4 and *GCN4*, were inserted upstream of uORF1 in the native orientation. The insertions of one or two copies of S1 reduced *GCN4* expression by a factor of 5–6, whereas insertion of three copies of S1 lowered expression by a factor of 40 (Fig. 4A). The strong inhibitory effect of the three copies of S1 in pM213 could be attributed primarily to a reduction in reinitiation following uORF1 translation, as opposed to non-specific inhibition of scanning, because elimination of the uORF1 ATG codon in pM213 restored high level *GCN4* expression (pG149, Fig. 4A). Comparing expression from pM213 and pG149 in Figure 4A we calculated that the frequency of reinitiation following uORF1 translation was reduced to 2% (8 U/420 U) by the S1 insertions in pM213, from the value of 24% calculated above for pM199. Note that the 2% reinitiation following uORF1 translation calculated for pM213 is only slightly higher than that deduced for construct p237 (1%) containing wild-type uORF4 alone at its normal location (see p237 and p238 in Fig. 4A).

Insertions of the S1 sequence that reduced the efficiency of reinitiation at *GCN4* in pM199 containing uORF1 alone had the same effect on construct pG155 containing both uORF1 and uORF4 in a *gcd1* mutant, where reinitiation at *GCN4* occurs at high levels constitutively (Fig. 4B). For the parental construct pG155 *GCN4* expression was ~8-fold higher in the *gcd1* strain compared with the non-derepressible *gcn2* mutant. The insertion of one to three copies of the S1 sequence in pG155 upstream of uORF1 led to a progressive decrease in *GCN4* expression. For pG158, containing three copies of S1 inserted upstream of uORF1, *GCN4* expression under derepressing conditions was reduced by a factor of 7 relative to the parental construct (Fig. 4B). According to our model ribosomes must translate uORF1 and resume scanning in order to bypass uORF4 and reinitiate at

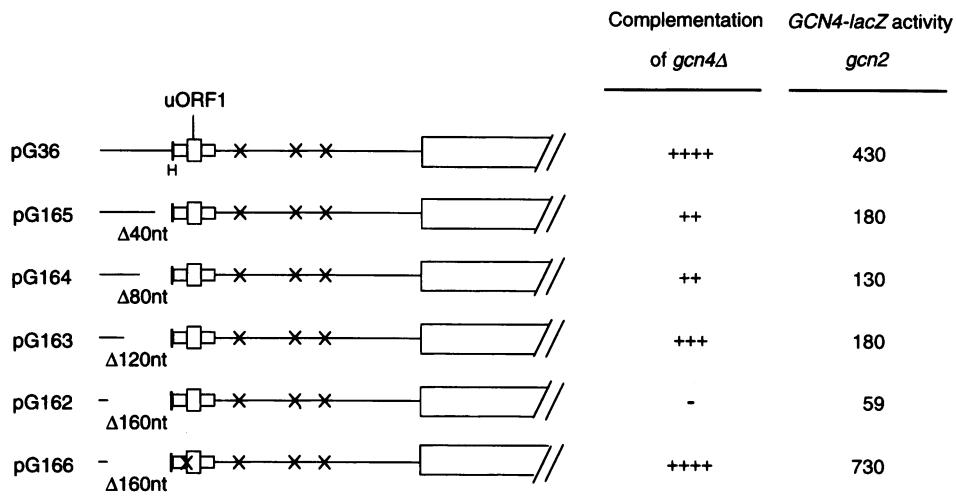

Figure 5. Sequence elements located 5′ of uORF1 contribute to the high efficiency of reinitiation following uORF1 translation. Schematic of pG36, containing uORF1 alone, and pG165, pG164, pG163 and pG162, in which 40, 80, 120 or 160 nt respectively normally present upstream of uORF1 have been deleted. pG166 is identical to pG162 except for a G→C substitution which changes the ATG codon of uORF1 to ATC. The constructs were drawn as described in the legend to Figure 1. H, *Hin*dIII site. Analysis of *GCN4* expression was conducted exactly as described in the legend to Figure 1. The individual β-galactosidase measurements differed from the mean value by <10%.

GCN4, thus the inhibitory effect of the S1 insertions on *GCN4* expression seen here with constructs containing uORFs 1 and 4 is consistent with the idea that the insertions decrease reinitiation following uORF1 translation.

It was conceivable that insertion of the S1 sequence upstream of uORF1 decreased reinitiation at *GCN4* because it increased the length of the leader preceeding uORF1, rather than disrupting an important sequence element. The former possibility was made unlikely by the results obtained with construct pG159, in which the length of the leader preceding uORF1 was increased by making a tandem duplication of the 200 nt sequence normally found upstream of uORF1. This produced a leader sequence 5′ to uORF1 similar in length to that found in pG158, but containing the wild-type sequence for 216 nt immediately upstream of uORF1 (Fig. 4B). In contrast to the greatly reduced expression seen for pG158 under derepressing conditions the pG159 construct showed only a modest reduction relative to the parental construct pG155. Thus merely elongating the leader segment upstream from uORF1 did not interfere with reinitiation following uORF1 translation. Instead the sequence or structure of the leader preceding uORF1 appears to be required for the high frequency of reinitiation following translation of uORF1.

As a final test of the idea that sequences upstream of uORF1 promote reinitiation at *GCN4* sequences were progressively deleted from a construct containing uORF1 alone at its normal position in the leader, beginning at a position 21 nt upstream of uORF1 (Fig. 5). The parental construct pG36 exhibited the high level *GCN4* expression expected for a construct containing uORF1 alone at its normal location in the full-length leader. The deletion 5′ of uORF1 in construct pG165 led to a substantial reduction in *GCN4* expression, suggesting that sequences within 60 nt of the uORF1 ATG make an important contribution to reinitiation efficiency. A further reduction in *GCN4* expression occurred in response to the largest deletion, which removed 160 nt (pG162). In an effort to rule out the possibility that novel junctions formed by these deletions reduce *GCN4* expression by inhibiting scanning, rather than impairing reinitiation, we showed that removing the ATG codon of uORF1 in construct pG162 restored *GCN4* expression to very high levels (Fig. 5, pG166). Using the results obtained from pG162 and pG166 we calculated that reinitiation following uORF1 translation was reduced to 8% by the deletion in pG162, compared with the 50% frequency of reinitiation characteristic of the wild-type leader in pG36.

DISCUSSION

Evidence that sequences upstream of uORF1 are required for reinitiation at *GCN4*

The uORF1 and uORF4 sequences in the *GCN4* mRNA leader play different roles in regulating the flow of ribosomes to the *GCN4* start codon. uORF4 is a strong translational barrier that can prevent essentially all ribosomes from reaching *GCN4* when present alone in the mRNA leader. In contrast, solitary uORF1 is a weak barrier, but it must be present upstream of uORF4 to obtain increased *GCN4* translation under derepressing conditions. The different functions of uORFs 1 and 4 in regulating *GCN4* translation can be explained by the fact that uORF4 allows essentially no ribosomes to resume scanning and reinitiate at *GCN4* following its own translation, whereas ~50% of the ribosomes that translate uORF1 can reinitiate downstream. Suppressing the ability of these ribosomes to reinitiate at uORFs 2, 3 and 4 following translation of uORF1, thus allowing them to continue scanning downstream to *GCN4*, is responsible for increasing *GCN4* expression under amino acid starvation conditions (3). The differences in the frequencies of reinitiation following translation of uORF1 versus uORF4 is partly attributable to the different sequences surrounding the termination codons of these two uORFs. Replacing the last codon and 10 nt downstream of the uORF1 stop codon with the corresponding nucleotides from uORF4 was sufficient to make uORF1 as inhibitory as uORF4 for reinitiation at *GCN4* (10). In the present study we

found that sequences located in the leader region >20 nt upstream from uORF1 are also required for efficient reinitiation at *GCN4*. When these sequences were deleted or displaced from their normal location uORF1 functioned like uORF4 and blocked efficient reinitiation downstream at *GCN4*.

When uORF1 is present at the wild-type position 350 nt upstream from *GCN4* we estimate that 53% of the ribosomes reinitiate at *GCN4* following uORF1 translation (Fig. 1, compare p235 with p238; 790 U/1500 U). When uORF1 was moved 200 nt downstream into the position of uORF4 by deleting all the sequences in the uORF1–uORF4 interval reinitiation following uORF1 translation fell to 24% (Fig. 1, compare pM199 with pG30; 310 U/1300 U). This reduction is in accord with the requirement for a certain extended period of scanning following uORF1 translation for efficient reinitiation at *GCN4* (4). In contrast, transplanting uORF1 and its surrounding sequences into the normal position of uORF4 reduced the frequency of reinitiation to only 5% (Fig. 1, compare pM189 with pG148; 21 U/390 U). These results indicate that sequences normally present between uORFs 1 and 4 are incompatible with efficient reinitiation when they are situated upstream from uORF1, even when uORF1 contains the native sequences surrounding its termination codon. It is noteworthy that uORF1 in construct pM189 allows 5% reinitiation at *GCN4*, whereas wild-type uORF4 allows only 1% of the ribosomes to reinitiate following uORF4 translation (Fig. 1, compare p237 with p238, 16 U/1500 U). This comparison indicates that the uORF1 termination region can promote a significantly higher level of reinitiation than that conferred by the corresponding uORF4 sequences when each is juxtaposed 3′ to the inhibitory sequence element(s) in the uORF1–uORF4 interval. Thus the sequences surrounding the stop codon and those located 5′ to uORF1 appear to make additive contributions to its ability to support a high frequency of reinitiation at *GCN4*.

There are two ways to view the effects of the deletions that progressively removed sequences from the uORF1–uORF4 interval upstream of the transplanted copy of uORF1 in pM189 (Fig. 3). One possibility is that they removed two distinct inhibitory elements from the uORF1–uORF4 interval that interfere with the ability of uORF1 to promote reinitiation. The more potent of these elements would be located just downstream from the native position of uORF1, between the deletion junctions of pG153 and pG154. A second possibility is that the uORF1–uORF4 interval inhibits reinitiation at *GCN4* by separating uORF1 from sequences normally located upstream from it that stimulate reinitiation. In this view each of the deletion constructs replaces these stimulatory sequences with different segments of the uORF1–uORF4 interval, none of which can promote reinitiation as effectively as do the native sequences upstream of uORF1. This second explanation implies that the postulated stimulatory sequences must be located immediately 5′ of uORF1 in order to stimulate reinitiation.

Inserting one or more copies of the S1 sequence at a site 20 nt upstream of uORF1 greatly reduced reinitiation at *GCN4* (Fig. 4). These insertions could impair reinitiation either by disrupting the postulated stimulatory element or by displacing it from its normal location. The fact that inserting three copies of the S1 sequence reduced *GCN4* expression to a greater extent than occurred with only one or two copies of the inserted sequence could indicate that the postulated stimulatory element is gradually inactivated as it is moved progressively farther upstream. Our results are not compelling on this point, however, because part of the inhibitory effect of the triple insertion in pG149 is observed in the absence of uORF1 (Fig. 4A, compare pG149 with p238 and pG30; Fig. 1). Thus the insertions probably have a non-specific inhibitory effect on scanning unrelated to reinitiation, the magnitude of which may increase with the number of inserted copies of S1. It is possible that inserting one, two or three copies of the S1 sequence equally impairs the ability of the postulated sequence element 5′ of uORF1 to stimulate reinitiation at *GCN4*. The results obtained with construct pG159 in Figure 4B indicate that the insertions impair reinitiation by disrupting an important sequence element and not by lengthening the distance between the 5′-end of the mRNA and uORF1.

Additional evidence for the existence of a stimulatory element upstream of uORF1 came from the fact that deletions in this part of the leader substantially reduced the efficiency of reinitiation (Fig. 5). *GCN4* expression was decreased significantly by deletion of the first 40 nt upstream of position –21 relative to uORF1 (pG165), consistent with the idea that a stimulatory element is located between 20 and 60 nt upstream of uORF1. However, this deletion and the next two in the series, which removed 80 and 120 nt upstream of uORF1 (pG164 and pG163 respectively), did not reduce *GCN4* expression to the same extent as did the S1 sequence insertions 5′ of uORF1 or when uORF1 was transplanted into the sequence environment of uORF4. One way to account for these different results would be to propose that the native sequences located between positions –60 and –160 upstream of uORF1 can partially substitute for the stimulatory element located between –20 and –60, whereas the S1 sequence or sequences in the uORF1–uORF4 interval cannot. Perhaps a particular sequence or structural feature which enhances reinitiation is reiterated in the region between positions –20 and –160 5′ of uORF1. Alternatively, this segment of the leader may be devoid of sequences or structures that inhibit reinitiation. It might be imagined that the deletion of 160 nt from the leader sequences upstream of uORF1 in construct pG162, leaving only 50 nt in this interval, reduces *GCN4* expression because the leader has been made too short for efficient recognition of uORF1. At odds with this possibility, a leader length of 50 nt has been shown to be sufficient for high level translation in yeast (8). Moreover, if ribosomes were now skipping the uORF1 start codon this would lead to higher, not lower, *GCN4* expression.

Possible mechanisms for the stimulation of reinitiation by sequences 5′ of uORF1

There are several ways in which leader sequences upstream of uORF1 could promote reinitiation following uORF1 translation. One possibility is that these sequences reduce the probability of ribosome dissociation from the mRNA following translation termination. This is the mechanism we proposed previously for sequences surrounding the uORF1 stop codon. The important feature of sequences in that location which are compatible with efficient reinitiation appears to be a high A+U content, prompting the notion that they decrease the frequency of ribosome dissociation because they interact weakly with the rRNA. We suggested that the absence of strong base pairing interactions between the rRNA and *GCN4* mRNA sequences surrounding the stop codon would allow ribosomes to resume scanning rapidly following peptide chain termination and thereby escape the action of a ribosome release factor (11). Ribosomes protect 30–35 nt of mRNA from nuclease digestion (18,19), thus ribosomes positioned over the uORF1 stop codon would be expected to interact

with 9–11 nt immediately preceding uORF1. Our results indicate that sequences >20 nt upstream of uORF1 have a strong effect on the frequency of reinitiation. It is improbable, therefore, that these sequences promote reinitiation simply by virtue of a weak interaction with rRNA in the mRNA binding track of the ribosome in the manner we suggested for the uORF1 termination region. Instead, based on their distance from uORF1, it seems more likely that the sequences 5′ of uORF1 enhance reinitiation by a more active process. This conclusion is in accord with the fact that the >200 nt leader upstream of uORF1 is unusually long compared with the majority of yeast mRNA leaders, which range in size from 20 to 60 nt (20). The atypically large size of the leader segment upstream of uORF1 suggests that important regulatory sequences reside in that region of *GCN4* mRNA.

One possibility for the positive regulatory role of the sequences upstream of uORF1 would be to interfere with the function of a ribosome release factor following peptide release at uORF1, either directly or through a protein factor which binds to the sequences. Alternatively, the stimulatory sequences could facilitate rebinding of an initiation factor, which is required for reinitiation downstream, to the ribosome following termination at uORF1. Reinitiation events involve ribosomal subunits that have just completed the elongation and termination steps of protein synthesis at an uORF. Consequently, these ribosomes should be devoid of initiation factors which are associated with the ribosome or the mRNA at the start of conventional initiation events, including eIF-3, eIF-4C, eIF-4F, eIF-4A and eIF-4B, in addition to eIF-2 and initiator $tRNA^{Met}$ (21). It may be necessary to facilitate rebinding of some or all of these factors to the ribosome following termination at uORF1 to achieve the high frequency of reinitiation characteristic of uORF1. If the stimulatory element 5′ of uORF1 comprises a high affinity binding site for the required factor(s) its relative proximity to the uORF1 stop codon could allow ribosomes to rebind the factor immediately after peptide chain termination. Distinguishing between these different models will require a more precise identification of the nucleotides upstream of uORF1 that stimulate reinitiation. Important clues about the function of this sequence and factors which interact with it might also be obtained by isolating genetic suppressors that restore high level reinitiation at *GCN4* when the upstream element is mutationally impaired.

REFERENCES

1 Hinnebusch,A.G. (1988) *Microbiol. Rev.*, **52,** 248–273.
2 Hinnebusch,A.G. (1992) In Jones,E.W., Pringle,J.R. and Broach,J.R. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 319–414.
3 Abastado,J.P., Miller,P.F., Jackson,B.M. and Hinnebusch,A.G. (1991) *Mol. Cell. Biol.*, **11,** 486–496.
4 Grant,C., Miller,P. and Hinnebusch,A. (1994) *Mol. Cell. Biol.*, **14,** 2616–2628.
5 Hinnebusch,A.G. (1994) *Trends Biochem. Sci.*, **19,** 409–414.
6 Lanker,S., Bushman,J.L., Hinnebusch,A.G., Trachsel,H. and Mueller,P.P. (1992) *Cell*, **70,** 647–657.
7 Williams,N.P., Mueller,P.P. and Hinnebusch,A.G. (1988) *Mol. Cell. Biol.*, **8,** 3827–3836.
8 Hinnebusch,A.G. and Liebman,S.W. (1991) In Broach,J.R., Pringle,J.R. and Jones,E.W. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis and Energetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 627–735.
9 Mueller,P.P. and Hinnebusch,A.G. (1986) *Cell*, **45,** 201–207.
10 Miller,P.F. and Hinnebusch,A.G. (1989) *Genes Dev.*, **3,** 1217–1225.
11 Grant,C.M. and Hinnebusch,A.G. (1994) *Mol. Cell. Biol.*, **14,** 606–618.
12 Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
13 Hinnebusch,A.G. (1984) *Proc. Natl. Acad. Sci. USA*, **81,** 6442–6446.
14 Ito,H., Fukada,Y., Murata,K. and Kimura,A. (1983) *J. Bacteriol.*, **153,** 163–168.
15 Vazquez De Aldana,C.R., Wek,R.C., San Segundo,P., Truesdell,A.G. and Hinnebusch,A.G. (1994) *Mol. Cell. Biol.*, **14,** 7920–7932.
16 Hinnebusch,A.G. (1985) *Mol. Cell. Biol.*, **5,** 2349–2360.
17 Mueller,P.P., Jackson,B.M., Miller,P.F. and Hinnebusch,A.G. (1988) *Mol. Cell. Biol.*, **8,** 5439–5447.
18 Kozak,M. (1983) *Microbiol. Rev.*, **47,** 1–45.
19 Wolin,S.L. and Walter,P. (1988) *EMBO J.*, **7,** 3559–3569.
20 Cigan,A.M. and Donahue,T.F. (1987) *Gene*, **59,** 1–13.
21 Merrick,W.C. (1992) *Microbiol. Rev.*, **56,** 291–315.