Sequences that surround the stop codons of upstream open reading frames in GCN4 mRNA determine their distinct functions in translational control

Paul F. Miller and Alan G. Hinnebusch

Unit on Molecular Genetics of Lower Eukaryotes, Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892 USA

Translational control of the GCN4 gene of Saccharomyces cerevisiae requires at least two of the four short upstream open reading frames (URFs) in the leader of GCN4 mRNA. URF4 is a strong negative element that is sufficient for repression of GCN4 expression in normal growth conditions. URF1 is ~30-fold less effective as a translational barrier when it is the single URF present in the mRNA leader and is required upstream from URF4 for efficient derepression of GCN4 expression under amino acid starvation conditions. We show that the last codon plus 10 bp immediately after the stop codon of URF4 are sufficient to convert URF1 into a strong translational barrier when it is present as a solitary URF. This result suggests that the characteristics of translation termination at URF4 are responsible largely for its strong inhibitory effect on translation initiation at the GCN4 AUG codon. Introduction of the same URF4 sequences at URF1 also reduces GCN4 expression under derepressing conditions when URF1 is upstream from URF4. This fact suggests that URF1 translation normally is compatible with efficient scanning and initiation downstream and that this property is required for its ability to overcome the translational barrier at URF4. These findings are consistent with the idea that ribosomes must first translate URF1 and then resume scanning in order to traverse URF4 sequences under starvation conditions. Our results indicate that nucleotides located 3' to the stop codon can be as important as those located 5' to the start site in determining the inhibitory effect of an URF on translation initiation downstream.

Key Words: GCN4; Saccharomyces; translational control; stop codons; reinitiation

Received March 30, 1989; revised version accepted May 23, 1989.

In most eukaryotic transcripts, the translation initiation codon is the 5' proximal AUG triplet. Insertion of an additional AUG codon into the mRNA leader generally leads to reduced translation of downstream-coding sequences, even when a stop codon located in frame with the first AUG codon precedes the downstream initiation site. The strong preference for 5' proximal AUG codons as translational start sites has been explained by the proposition that the eukaryotic initiation complex assembles at the 5' end of the mRNA and scans in the 3' direction until an AUG codon is encountered, whereupon translation begins. The fact that an in-frame stop codon generally does not abolish the inhibitory effect of an upstream AUG codon suggests that reinitiation at internal start sites is usually inefficient (Sherman and Stewart 1983; Johansen et al. 1984; Kozak 1984; Liu et al. 1984). Translational inhibition by an upstream open reading frame (URF) can be reduced by altering the sequence context of its AUG codon in a way that is expected to lower its efficiency as an initiation site. Presumably, such alterations allow the scanning initiation complex to bypass the first AUG codon and to initiate translation at a downstream start site (Kozak 1981, 1986). The inhibitory effect of a URF also has been reduced by moving it further upstream from the start site of the protein-coding sequences on the mRNA. To account for the latter effect, it was suggested that the efficiency of reinitiation was elevated in response to a greater separation between the two coding sequences, perhaps because of increased time available to reassemble an initiation complex before the second start site was encountered (Kozak 1987a).

Four short URFs two or three codons in length mediate translational control of GCN4 expression in Saccharomyces cerevisiae. GCN4 encodes a transcriptional activator of amino acid biosynthetic genes that is repressed translationally under conditions of amino acid sufficiency. The removal of all four URFs, or a mutation in a GCD negative regulatory gene, abolishes repression of GCN4 expression and leads to constitutively derepressed transcription of the many structural genes that are under GCN4 positive control. In wild-type cells,
GCN4 protein synthesis is stimulated in response to starvation for any amino acid by overcoming the translational barrier exerted by the URFs. This derepression requires the positive regulators GCN1, GCN2, and GCN3, thought to function indirectly by antagonism of GCD factors (for review, see Hinnebusch 1988).

When only one of the four URFs is present in the mRNA leader, GCN4 expression is almost completely unregulated, increasing very little in response to starvation. However, the absolute level of GCN4 expression observed in the presence of a single URF depends greatly on the URF in question. URF4 is sufficient for a wild-type level of repression, reducing GCN4 expression 50- to 100-fold below the level observed when no URFs are present in the leader. URF1 is much less inhibitory as a solitary URF, reducing GCN4 expression only two- to threefold. Surprisingly, when situated upstream from URF4, URF1 functions as a positive control element and leads to increased GCN4 expression in response to starvation. Thus, it appears that translation of the weakly inhibitory URF1 sequence, combined with activation of GCN1, GCN2, and GCN3 under starvation conditions, permits scanning ribosomes to overcome the strong translational barrier at URF4 and to initiate translation at the GCN4 start codon (Mueller and Hinnebusch 1986).

The different distances that separate URFs 1 and 4 from the GCN4 AUG codon do not appear to be responsible for the large differences between the two URFs as negative control sites [Williams et al. 1988]. Therefore, we set out to identify the particular nucleotide sequences that: [1] make URF1 less effective than URF4 as a translational barrier when each is present singly in the mRNA leader, and [2] allow URF1 to function as a positive control element in overcoming the translational barrier at URF4. Our results indicate that both functions are determined primarily by sequences that surround the stop codons of the two URFs. On the basis of this finding, we propose that translation termination occurs differently at these two sites. Termination at URF4 seems to occur in a way that prevents all initiation events downstream. By contrast, termination at URF1 is compatible with considerable initiation downstream, and this property is critical for ribosomes to overcome the translational barrier at URF4.

Results

Cassette mutagenesis of the URF1-coding sequences and termination region

We showed previously that URF1 and 4 function comparably as translation initiation sites in vivo when each is fused to lacZ protein-coding sequences and assayed for enzyme expression in the absence of all other upstream AUG codons in the GCN4 mRNA leader. This result suggests that elongation or termination steps could distinguish the regulatory functions of URFs 1 and 4. In accord with this idea, most URF1 termination codon mutations that increase the size of URF1 from 3 to 10-20 codons were found to increase the inhibitory effect of URF1 as a solitary URF in the mRNA leader and to impair its positive regulatory function when it is present upstream from URF4, thereby reducing GCN4 expression in both situations [Mueller et al. 1988].

We decided to extend our analysis of the sequence requirements for URF1 regulatory function by constructing additional mutations in the URF1-coding sequences and termination region. These mutations were generated in a GCN4–lacZ fusion contained on a low-copy number plasmid. The GCN4–lacZ construct that we employed lacked the second and third ATG codons and contained unique restriction sites that flanked URF1 to permit cassette mutagenesis of the URF1 region (Fig. 1). β-Galactosidase activity expressed from this fusion (pM23) was found to be regulated in the manner described previously for a GCN4–lacZ construct that contained the wild-type leader: derepression occurred only in response to amino acid starvation and required the

Figure 1. DNA sequence of the GCN4 mRNA leader region present in pM23, numbered from the 5′ end of the transcript. URFs 1 and 4, and the beginning of the GCN4 protein-coding sequences, are underlined. Also indicated are the unique HindIII and BglII sites introduced for cassette mutagenesis of URF1 and the EcoRI and KpnI sites introduced previously in the course of the removal of the ATG codons of URFs 2 and 3 (Mueller and Hinnebusch 1986), with the altered nucleotides shown in lower case. The A, B, and C segments that contained URF1, and its flanking nucleotides that were substituted with corresponding segments from the URF4 region, are bracketed at the two sites.
positive regulator \textit{GCN2}; \textit{GCD1} function was required for repression of enzyme activity in nonstarvation conditions (Fig. 2B). The second and third ATG codons were removed because their absence simplified the analysis of URFl mutations and produced only a slight increase in the URFl termination codon that increased or decreased URFl length by only a single codon. Both of these stop codon mutations (pM24 and pM31) had the same phenotype as the more extensive changes in URFl length mentioned above, reducing \textit{GCN4-lacZ} expression, particularly in derepressing conditions. In fact, the pM31 mutation approximated the deleterious effect of the removal of the URFl ATG codon (pM226) (Fig. 2B).

We began by introducing two new base substitutions in the ATG codons of URFl that lacked the restriction sites around URFl, and was \sim 60\% of that given by p200 under derepressing conditions. Therefore, we proceeded with the mutagenesis of the URFl region in the pM23 construct.

We began by introducing two new base substitutions

\[\begin{array}{c}
\text{A.} \\
\text{B.} \\
\text{C.} \\
\text{D.}
\end{array}\]

\[\begin{array}{c}
\text{Constructs Containing URF4} \\
\text{Units of Enzyme Activity} \\
\text{gene2'} \\
\text{gene1'} \\
\text{Complementation} \\
\text{of gene4} \\
\text{Constructs Lacking URF4} \\
\text{Units of Enzyme Activity} \\
\text{gene2'} \\
\text{gene1'} \\
\text{of gene4} \\
\end{array}\]

\[\begin{array}{c|c|c|c|c|c|c|c}
\text{Construct} & \text{URFl Sequence} & \text{wt} & \text{gen2'} & \text{gen1'} & \text{Complementation} & \text{gene2'} & \text{gene1'} \\
\hline
\text{pM23(wt)} & \text{ATG GCT TGC TAA ACC GAT} & 12 & 80 & 15 & 200 & + + + & 410 & 510 \\
\text{pM226} & \text{A_A (no URFl)} & 8 & 17 & 18 & 18 & + / - & NA & \text{NA} \\
\text{pM24} & \text{Tyr Ter TGA} & 6 & 28 & 10 & 97 & + & 120 & 250 \\
\text{pM31} & \text{Tyr Ter TAT CCG AT} & 7 & 18 & 14 & 30 & + / - & 96 & 91 \\
\text{pM35} & \text{Val GCT TGC TAA ACC GAT} & 8 & 76 & 17 & 240 & ND & ND & ND \\
\text{pM36} & \text{Gly GGT TGC TAA ACC GAT} & 8 & 62 & 10 & 170 & ND & ND & ND \\
\text{pM25} & \text{Tyr Ter TAG} & 11 & 79 & 13 & 200 & + + + & ND & ND \\
\text{pM26} & \text{Tyr Ter TAG} & 15 & 120 & 13 & 180 & + + + & ND & ND \\
\text{pM28} & \text{Tyr Ter TAG} & 13 & 70 & 11 & 135 & + + + & 210 & 290 \\
\text{pM142} & \text{Ser UCC TGC TAA ACC GAT} & 10 & 54 & 9 & 92 & + + & 260 & 310 \\
\text{pM143} & \text{Leu UCC TGC TAA ACC GAT} & 9 & 39 & 10 & 58 & + & 220 & 205 \\
\text{pM140} & \text{Tyr Ter TAG} & 10 & 37 & 10 & 54 & + + & 290 & 300 \\
\text{pM141} & \text{Tyr Ter TAG} & 10 & 30 & 8 & 35 & + & 140 & 170 \\
\end{array}\]

\textit{Figure 2.} Effects of point mutations in URFl on \textit{GCN4} expression. (A) Schematic of the \textit{GCN4-lacZ} construct in pM23 with URFls 1 and 4 shown as solid boxes and the beginning of \textit{GCN4}-coding sequences as an open rectangle. (X) Base substitutions in the ATG codons of URFls 2 and 3. (H and B) HindIII and BglII restriction sites employed in cassette mutagenesis of URFl, respectively. (B) Effects of URFl mutations on \textit{GCN4} expression. The nucleotide positions that differ from the wild-type sequence in pM23 are underlined and the novel amino acids encoded by the mutated URFl sequences are shown where applicable. The effects of the mutations were measured in two ways for plasmid-borne constructs that contained URFls 1 and 4: \textit{GCN4-lacZ} fusion enzyme activity (expressed as nanomoles of o-nitrophenyl-\beta-D-galactopyranoside hydrolyzed per minute per milligram of protein) was measured in extracts of wild-type [wt] cells grown in nonstarvation conditions (repressing, \textit{R}) or in histidine-starvation conditions (derepressing, \textit{DR}). Expression in the same two growth conditions was measured and averaged for a \textit{gen2-1} mutant that was repressed constitutively [\textit{R} in parentheses] and in a \textit{gcd1-101} mutant that was derepressed constitutively [\textit{DR} in parentheses]. The values shown are averages calculated from 2 to 3 determinations collected from independently derived transformants. The individual measurements differed from the mean values shown by 30\% or less. In the second assay, the leader mutations were inserted upstream from the authentic \textit{GCN4} protein-coding sequences and the resulting plasmids were tested for complementation of the 3-aminotriazole (3-AT) sensitivity associated with the \textit{gcd4-103} chromosomal deletion. The average growth rate on 3-AT medium of several independently derived transformants for each construct is summarized qualitatively. Little or no variation in growth rate was observed for different transformants that harbored the same plasmid. In certain cases, the effects of URFl mutations on \textit{GCN4-lacZ} fusion enzyme activity were determined in constructs in which the ATG codon of URFl was altered to an ATC codon. [NA] Not applicable; [ND] not determined.
its ability to complement a chromosomal gcn4 deletion for the failure to derepress HIS3 gene expression under amino acid starvation conditions [Fig. 2B].

Because the pM24 and pM31 mutations changed URF1 length, coding sequence, termination site, and sequences 3' to the stop codon, we varied these features independently by making additional URF1 point mutations. Four mutations were constructed in the coding region, three of which caused amino acid substitutions in the predicted URF1 tripeptide sequence. These mutations had little or no effect on GCN4 expression. The same result was seen for a fifth mutation in which TAG was substituted for the TAA termination codon of URF1 [Fig. 2B]. These results indicate that the negative effects of the pM24 and pM31 stop codon mutations on GCN4 expression do not result from a unique sequence requirement for URF1-coding sequences or for a particular termination codon. Second, these data show that it is improbable that the URF1 tripeptide has an important role in translational control. Third, because these five mutations were introduced at three different positions in the URF1 sequence, and because all have little effect on GCN4 expression, it seems unlikely that sequences downstream from the URF1 AUG codon affect the initiation rate at this start site. Consequently, the deleterious effects of the pM24 and pM31 stop codon mutations on GCN4 expression appear to result from alterations in elongation or termination steps at URF1.

Insertions of any one of eight different codons, including both rare and preferred codons [see Materials and methods], between the second and third codons of wild-type URF1 all reduced GCN4 expression by roughly the same amount. The deleterious effects of two such codon insertions shown in Figure 2C (pM142 and pM143) are comparable in degree to the effects seen with the pM24 stop codon mutation. However, none of the single-codon insertions impaired derepression to the extent seen for the stop codon mutation in pM31, which also caused URF1 to be four codons long. For this reason, and because wild-type URFs 1 and 4 are both three codons in length, we suspected that sequences downstream from the novel URF1 termination site formed in pM31 were additionally responsible for its derepression defect.

Accordingly, multiple substitutions were introduced at random into the 6 bp immediately after the URF1 termination codon in a construct that was identical to pM23 except that it contained the authentic GCN4 protein-coding sequences. [Typically, four of the six positions were substituted.] The resulting GCN4 alleles were screened for their ability to complement the chromosomal gcn4 deletion. Only five out of the ~100 alleles we examined showed reduced GCN4 expression. Of these, none was as defective as pM31, two displayed a derepression defect comparable in degree to that exhibited by pM24 (pM141 and pM72, Table 1), and the other three were less defective than pM24 (pM73, pM71, and pM140). GCN4-lacZ expression was analyzed for one allele of each category [pM141 and pM140], and the results shown in Figure 2D are in good agreement with the complementation data for these mutations shown in Table 1. These results indicate that considerable flexibility exists for the 6 bp immediately 3' to the URF1 stop codon for efficient derepression; however, it is also evident that certain nucleotide sequences at this position are incompatible with efficient URF1 positive regulatory function. Additional evidence that supports the latter conclusion will be presented in the next section.

At this point, we examined the effects of the aforementioned mutations [Fig. 2B–D] on GCN4 expression when URF1 was present singly in the mRNA leader. In this situation, URF1 normally functioned as a weak negative element. The results of this analysis showed that those mutations that reduced the ability of URF1 to stimulate GCN4 expression when located upstream from URF4, made URF1 a more inhibitory element when present alone in the leader [Fig. 2]; thus, mutations that impair URF1 positive function cause it to resemble more closely UR4 as a translational barrier. On the basis of this correlation, we reasoned that by making hybrid URFs that contained sequences from URF1 and UR4, we could identify the nucleotides responsible for the opposing functions of these two sequence elements in GCN4 translational control.

Table 1. Effect of sequence alterations downstream from the URF1 termination codon on complementation of a gcn4 deletion

| Plasmid-borne GCN4 allele | Sequence 3' to URF1 | Complementation of gcn4A |
|--------------------------|---------------------|--------------------------|
| pM23                     | TAA ACGGAT          | + + +                    |
| pM58                     | TAA AAATAT          | + + +                    |
| pM59                     | TAA ACGGTT          | + + +                    |
| pM60                     | TAA AGTGGT          | + + +                    |
| pM61                     | TAA ATTCGC          | + + +                    |
| pM62                     | TAA AAAAT           | + + +                    |
| pM63                     | TAA TCTTAC          | + + +                    |
| pM48                     | TATGTT              | + + +                    |
| pM53                     | TAA ATCTGC          | + + +                    |
| pM141                    | TAA TAGAGG          | +                       |
| pM72                     | TAA TAAAGG          | +                       |
| pM73                     | TAA AACTAA          | +                       |
| pM71                     | TAA GGTAGC          | +                       |
| pM140                    | TAA GCGAGA          | +                       |
| p227                     | [no URF1; solitary URF4] | + / -                  |

The sequence of the 6 nucleotides immediately 3' to the URF1 stop codon is shown for each GCN4 allele that was examined for its ability to complement a gcn4 deletion for sensitivity to 3-aminotriazole [3-AT]. The degree of 3-AT resistance is a measure of the extent of GCN4 derepression. The average growth rate on 3-AT medium for several independently derived transformants of each allele is summarized qualitatively. The p227 construct lacked URFs 1, 2, and 3 as a result of point mutations in their respective ATG codons [Mueller and Hinnebusch 1986]. The complementation response of p227 is shown here for comparison of the deleterious effects on GCN4 expression of altering sequences 3' to the URF1 stop codon with removal of URF1 altogether.

Construction of URF1/URF4 hybrids

The two URFs were divided into three segments: 16 bp upstream from the URF start codon (segment A), the
coding sequences and termination codon (segment B), and 25 bp downstream from the URF stop codon (segment C) [Fig. 1]. Hybrid URFs were constructed by the substitution of URF1 segments with the corresponding segment derived from URF4, and the hybrid elements were introduced in place of URF1 upstream from wild-type URF4. When the URF1 A–B–C segment was replaced with URF4 A–B–C (pM98), GCN4 expression in derepressing conditions was very low compared to the wild-type construct [Fig. 3B]. Because this substitution also reduced expression under repressing conditions, a significant derepression response remained intact; however, the URF4 A–B–C segment clearly functioned less efficiently as a positive control site than the corresponding wild-type URF1 sequences. Substitution of only the upstream [A] sequences at URF1 with those found at URF4 (4–1–1 construct) had no detectable effect on GCN4 expression. Given that sequences upstream from the start codon are the primary determinants of the initiation rate (Kozak 1986; Baim and Sherman 1988; Cigan and Donahue 1988), this result is consistent with our previous conclusion that URF1 and URF4 have similar initiation efficiencies (Mueller et al. 1988). Substitution of the coding region (B) of URF1 with that of URF4 had a small, but reproducible, deleterious effect on URF1 positive regulatory function (1–4–1); however, the URF4 A–B–C segment clearly functioned less efficiently as a positive control site than the corresponding sequence from URF4. The results in Table 2 show that the 10 bp immediately downstream from URF4 are sufficient to confer the full derepression defect associated with URF4 region C, whereas the 6 bp that follow the URF4 stop codon confer only a partial inhibitory effect on GCN4 expression when introduced at URF1.

To define more precisely the sequences downstream from the URF stop codons that are important for their distinct regulatory functions, we replaced 10- and 6-bp segments 3’ to URF1 with the corresponding nucleotides from URF4. The results in Table 2 show that the 10 bp downstream from URF4 are sufficient to confer the full derepression defect observed when the complete URF1-coding region (Segment B) was replaced by that from URF4. The results in Table 2 show that this is indeed the case [cf. pM122 with pM133 and pM134]. Moreover, the replacement of the URF1 third codon with the rare CCC proline codon (pM121) led to a similar reduction in URF1 positive function, whereas substitutions that introduced the preferred CCA or CCT proline codons (pM119 and pM120) had no effect on GCN4 expression. (Although...

![Figure 3](image-url)
expression of replacing the third codon and sequences immediately following the stop codon at URF1 with the corresponding nucleotides at URF4.

Table 2. Effects on GCN4-lacZ expression of replacing the third codon and sequences immediately following the stop codon at URF1 with the corresponding nucleotides at URF4

| GCN4-lacZ allele | Origin of URF sequences | Units of enzyme activity |
|------------------|-------------------------|--------------------------|
|                  | A | B | C | R | DR | gcna | gcdn |
| pm23             | 1 | 1 | 1 (25) | 12 | 80 | 15 | 200 |
| pm100            | 1 | 1 | 4 (25) | 5 | 25 | 10 | 56 |
| pm99             | 1 | 4 | 4 (25) | 2 | 16 | 5 | 22 |
| pm121            | 1 | 1 | 4 (10) | 8 | 28 | 14 | 64 |
| pm162            | 1 | 4 | 4 (10) | 4 | 21 | 6 | 27 |
| pm121            | 1 | 1 | 4 (10) | 7 | 42 | 8 | 84 |
| pm121            | 1 | 1 (CCG) | 4 (10) | 5 | 16 | 7 | 26 |
| pm119            | 1 | 1 (CCU) | 4 (10) | 6 | 17 | 7 | 27 |
| pm120            | 1 | 1 (CCU) | 4 (10) | 7 | 27 | 8 | 44 |

All constructs contained wild-type URF4 sequences present at the normal location in the leader. The origin of sequences at URF1 is indicated as in Fig. 3, except that the number in parentheses after the C-segment designation gives the number of nucleotides derived from URF4 that are inserted immediately 3' to the URF1 stop codon, the remainder of the 25 nucleotides present in the C segment were derived from URF1. For pm122, pm121, pm119, and pm120, the nucleotides in parentheses after the B-segment designation indicate the proline codon used to replace the third codon in URF1; the first and second codons present in the B segment of these alleles are those found normally at URF1. GCN4-lacZ enzyme activity was measured under the same conditions described in Fig. 2. The data for the first three constructs was taken from Fig. 3 and is shown here for comparison.

GCN4-lacZ expression under derepressing conditions from the CCG- or CCC-containing versus the CCA- or CCT-containing constructs differs by only ~2-fold, these differences are significant because the standard errors on the measurements are only ~15%.) These results could indicate that the rare proline codon contributes to the inhibitory effect of URF4 on GCN4 expression by reducing the rate of translation elongation in the URF. Alternatively, the sequence context immediately 5' to the stop codon may be important for URF4 negative regulatory function.

RNA blot hybridization analysis was conducted to measure the sizes and steady-state amounts of transcripts produced by most of the GCN4 alleles shown in Figures 2 and 3B [see Materials and methods for complete list]. In accord with our previous results on GCN4 alleles that contained point mutations in the leader region, no differences in size were observed for any of the mutant transcripts compared with wild-type GCN4 mRNA. In addition, the abundance of the mutant transcripts was found to vary from that of the wild-type mRNA by only 50% or less [data not shown]. These results indicate that differences in GCN4 expression observed among these constructs result primarily from changes in their translational efficiency.

Discussion

Substitutions in the coding region and in sequences 3' to the URF1 stop codon that increase its strength as a translational barrier

The aim of the experiments described here was to identify the nucleotides at URFs 1 and 4 that are responsible for the different functions of these two elements in translational control of GCN4 expression. When present singly at its normal location in the mRNA leader, URF4 reduces GCN4 expression by ~99% from the high level observed when no URFs are present. In this situation it appears that most ribosomes initiate at URF4 and fail to reinitiate downstream. When present alone at its normal location, URF1 is about 30-fold less inhibitory to GCN4 expression than URF4, which indicates that many ribosomes either bypass this site or reinitiate downstream after URF1 translation [Mueller and Hinnebusch 1986]. When a 60-bp segment that contained URF4 was moved upstream into the sequence environment of URF1, GCN4-lacZ expression was about threefold higher than that observed when URF4 was present alone at its normal location [Fig. 2, p226, and Fig. 3, pm98 lacking URF4]. This result suggests that closer proximity to the GCN4 start site is responsible partly for the greater inhibitory effect of URF4 compared to URF1. Likewise, URF1 is several fold more inhibitory to GCN4 expression when moved downstream into the position occupied normally by URF4 [Williams et al. 1988]. A similar relationship between the proximity of an URF and its inhibitory effect on translation downstream was described for the preproinsulin transcript. To explain this relationship, Kozak [1987a] proposed that the efficiency of reinitiation increases with greater separation between the URF and the protein-coding sequences on the mRNA [Kozak 1987a]. We favor this explanation for our results also and consider it likely that reinitiation at the GCN4 start site can occur after translation of URFs 1 and 4, at least when these elements are located far upstream in the mRNA leader. No evidence for reinitiation has emerged from studies on various yeast CYC1 transcripts that contained URFs [Sherman and Stewart 1983]; however, in these cases, the URFs were located very close to the start site for CYC1 protein synthesis.

Despite the effects of position on URF function, URF4 remained about 10-fold more inhibitory than URF1 when the two sequence elements were compared at the
same location in the mRNA leader [Fig. 3]; thus, particular nucleotides present at URFs 1 and 4 are more important than their distances from the GCN4 start site in determining the different efficiencies of these elements as translational barriers to GCN4 expression. The most striking result of our experiments was that a combination of the CCG proline codon and 10 bp immediately downstream from the stop codon at URF4 was sufficient to confer upon URF1 the strong inhibition of GCN4 expression that is characteristic of URF4. By contrast, substitution of sequences upstream from URF1 with the corresponding sequences from URF4 had little or no effect on URF1 function. The latter result suggests that the differences between URF1 and URF4 as negative control elements are unrelated to their initiation efficiencies. In fact, the AUG codons of URFs 1 and 4 functioned equally well as translation initiation sites when the two URFs were fused in frame to lacZ-coding sequences [Mueller et al. 1988]. The fact that sequences that surround the stop codons of the two URFs are the principal determinants of their inhibitory effects on translation suggests that termination events occur differently at the two sites, allowing continued scanning and initiation downstream in the case of URF1, and interfering with these processes in the case of URF4. The sequences that surround the URF4 stop codon may reduce the flow of scanning ribosomes to downstream start sites by promoting ribosome dissociation, or by inhibiting reassociation of the necessary initiation factors, after the termination event. Alternatively, these sequences may lead to inefficient termination and cause ribosomes to stall at the URF4 stop codon. Perhaps the rare proline codon at URF4 enhances the effects of sequences 3′ to the stop codon by inducing a translational pause just prior to the termination step.

The deleterious effects of the single-codon insertions that introduce rare codons into URF1 [pM143, Fig. 2] may be attributed to a reduced rate of elongation in URF1, except that insertions of preferred codons at the same site also reduce GCN4 expression [pM142, Fig. 2]. Because there is no obvious reason why a three-codon URF would be required for efficient elongation or termination, we suggest that the three-codon length may facilitate resumed scanning and reinitiation downstream after URF1 translation. For example, initiation factors may remain associated with the ribosome for the first few elongation steps, and their continued presence after termination at URF1 may enhance the frequency of reinitiation downstream [Kozak 1984]. Of interest in this connection is the observation that certain antibiotics, such as verrucarin A, inhibit elongation only during formation of the first few peptide bonds, which suggests a mechanistic difference between early and advanced elongation steps and a possible overlap between initiation and elongation processes [Vazquez 1979; Warner 1983].

Five alleles that contained multiple substitutions in the 6 bp that immediately follow the URF1 stop codon showed reduced GCN4 expression; however, the majority of such constructs had no mutant phenotype. These results suggest that a large family of sequences can mimic the effect of the 6 bp normally found 3′ to URF1 in allowing efficient initiation to occur downstream. Because the random substitutions made 3′ to URF1 were examined in the absence of the CCG proline codon, however, only the most inhibitory sequences may have been detected in our analysis. In addition, we found that the introduction of the corresponding 6 nucleotides from URF4 was insufficient to convert URF1 to a strong negative element [Table 2], which indicates that residues at positions 7–10 contribute to the inhibitory effect of URF4 also. Elucidation of the exact sequence requirements for the ability of URF1 to allow efficient initiation downstream will require additional mutagenesis experiments.

The novel sequences introduced 3′ to URF1 in 5 alleles that did show reduced GCN4 expression [Table 1], plus the 10 nucleotides normally found 3′ to URF4, represent a second family of sequences that interfere with efficient initiation at downstream start sites. These sequences have no obvious similarity to one another, nor do they appear to have the potential to form secondary structures with surrounding nucleotides. Of course, they need not increase translational inhibition by URF1 for the same reasons. For example, pM141 and pM72 each contain a second stop codon that immediately follows the normal URF1 termination site: the presence of two adjacent stop codons may reduce the probability that ribosomes remain attached to the mRNA and resume scanning after termination at URF1. Interestingly, the 10 nucleotides normally found downstream from wild-type URF4 are complementary in 8 out of 10 positions to a sequence present at the 3′ end of S. cerevisiae 18S rRNA [positions 1767–1776] that occurs in a stem–loop structure that is highly conserved between eukaryotes and prokaryotes [Stiegler et al. 1981; Hogan et al. 1984]. In Escherichia coli, this structure has been implicated in binding initiation factor IF3 [Ehresmann et al. 1986]. Perhaps interaction between the 10 nucleotides 3′ to the URF4 stop codon and complementary nucleotides in 18S rRNA blocks the association of a factor analogous to IF3 with the ribosome, reducing the probability of reinitiation events downstream from URF4 under normal conditions. We intend to pursue this possibility by additional mutagenesis experiments.

**URF1 as a positive translational control element**

The mutations that increase the inhibitory effect of URF1 when this sequence element is present alone in the mRNA leader also reduce its ability to act as a positive control site when situated upstream from URF4. This correlation suggests strongly that URF1 positive regulatory function depends both on its ability to be well-translated and on its ability to permit a steady stream of ribosomes (or 40S subunits) to advance downstream from its termination site. One way to account for these characteristics is to propose that only those ribosomes that translate URF1 and resume scanning are able to subsequently traverse URF4 sequences [Fig. 4]. This requirement may exist if the mechanism of reinitiation at internal AUG codons differs in some important re-

---

**GENES & DEVELOPMENT 1223**
Figure 4. Hypothetical molecular mechanism for the interactions between URFs 1 and 4 in controlling translation initiation at the GCN4 start site. The GCN4 mRNA is shown schematically with only URFs 1 and 4 present upstream from the GCN4-coding sequences. 40S subunits and 80S ribosomes are indicated. In the top two constructs (I), URF1 is missing as a result of a point mutation in its AUG codon. As a result, all 40S subunits scanned to URF4 and initiated translation. After translation termination at URF4, no reinitiation occurred at the GCN4 start site in either repressing (R, Gcn−) or derepressing (DR, Gcd−) conditions. When URF1 was present upstream (II), most ribosomes translated URF1 and scanning could resume after termination at this site. In repressing conditions, reinitiation occurred at URF4 but no reinitiation followed termination at URF4. In derepressing conditions, the behavior of 40S subunits that engaged in prior translation of URF1 was altered as the result of reduced GCD activity in the cell (hatched subunits). These altered initiation complexes were able to traverse URF4 and reinitiate at the GCN4 start site, either by skipping over the URF4 AUG codon or by more efficient reinitiation after URF4 translation. [The latter explanation is supported more favorably by recent experiments in which the effect of URF1 on translation of an URF4–lacZ fusion was measured in repressing and derepressing conditions [Hinnebusch et al. 1988]]. In a related model (Tzamarias and Thireos 1988), scanning resumed after URF1 translation only in derepressing conditions and the reinitiation complexes generated in these conditions traversed URF4 by either of the two mechanisms just described.

In conclusion, our findings indicate that sequences surrounding the URF4 stop codon are an order of magnitude more effective than those present at URF1 in preventing ribosomes from reaching the GCN4 start site following translation of the URF. Importantly, the different effects of these two sequence elements are observed under normal growth conditions and in mutants that lack specific regulatory factors required for translational control of GCN4 expression. These facts suggest that our results may be relevant to many other transcripts in yeast and higher eukaryotes that contain one or more URF sequences (Cigan and Donahue 1987; Kozak 1987b). Therefore, nucleotides that surround the stop codon of an URF frequently may be as important as the sequence located 5′ to its initiation codon, or as the distance that separates the URF from the next start site, in determining the efficiency of scanning and translation initiation downstream.

Materials and methods

Point mutations that generated the HindIII and BglII restriction sites flanking URF1 in plasmid pM23 (Fig. 1) were introduced by oligonucleotide-directed mutagenesis, as already described [Zoller and Smith 1982; Mueller and Hinnebusch 1986]. With the exception of these mutations and the point mutations in the ATG codons of URFs 2 and 3, pM23 is identical to plasmid p180 [Hinnebusch 1985], containing a GCN4–lacZ translational fusion and URA3, ARS1, and CEN4 sequences for selection of transformants and low copy number, autonomous replication in yeast. Cassette mutagenesis of URF1 was carried out by the insertion of double-stranded synthetic oligonucleotides of the desired sequence between the HindIII and BglII sites in pM23. The nucleotide sequence of each insert was determined subsequently by use of the dideoxyribonucleotide chain-termination technique [Sanger et al. 1977]. In the construction of alleles with single codon insertions in URF1, oligonucleotides were prepared to contain three nucleotides, chosen at random, that were inserted between the alanine and cysteine codons of URF1. Eight of the resulting constructs were selected at random and found to contain the following single codon insertions, listed along with the encoded amino acid and the relative frequency with which that codon occurs in known S. cerevisiae genes [Sharp et al. 1986]: ACU, Thr, high; UAC, Tyr, high; UCC, Ser, high; UGU, Cys, high; UUA, Leu, moderate; UUU, Phe, moderate; CUC, Leu, low; UCA, Ser, low. Derivatives of pM23 that contained a point mutation in the URF4 initiation codon were constructed by the replacement of an EcoRI fragment extending from mutated URF2 to the 3′ end of the GCN4–lacZ-coding sequences with the analogous fragment from p227, a plasmid-derivative that contained point mutations in the ATG codons of all four URFs [Mueller and Hinnebusch 1986]. Plasmids that contained the various leader mutations located upstream from authentic GCN4-coding sequences were constructed by the elimination of the lacZ–BamHI fragment from pM23-derivative constructs. Multiple substitutions were introduced immediately 3′ to the URF1 stop codon by the preparation of double-stranded oligonucleotides in which all four nucleotides were introduced at random at the six positions that immediately follow the URF1 stop codon. This oligonucleotide mixture was inserted between the HindIII and BglII sites of a plasmid identical to pM23 except in that it contained authentic GCN4 protein-coding sequences in place of GCN4–lacZ (pM37).

GCN4–lacZ enzyme expression was assayed in transformants that contained pM23 or mutant derivatives after growth for 6 hr in minimal medium (nonstarvation conditions) or in

1224 GENES & DEVELOPMENT
minimal medium supplemented with 10 mM 3-aminotriazole ([Lucchini et al. 1984], an inhibitor of histidine biosynthesis that functions as a competitive inhibitor of the HIS3 product. Expression of each construct was measured in several different transformants of wild-type (F113), gcn2-1 (H15), and gcd1-101 (F98) strains described previously (Hinnebusch 1985). GCN4 alleles that contained leader mutations were tested for their ability to complement the gcn4-103 deletion in strain H384 [Mueller and Hinnebusch 1986] by the scoring of growth of transformants that contained the plasmid-borne alleles on solid medium containing 3-aminotriazole (Hinnebusch and Fink 1983). RNA blot hybridization analysis of GCN4 transcripts in H384 transformants was conducted as described previously (Hinnebusch and Fink 1983; Mueller and Hinnebusch 1986). Transcripts were analyzed for GCN4 alleles that contained the following leader mutations: pM22-26, pM28, pM31, pM98–103, pM140–143, and p226.

Acknowledgments
We thank David Garfinkel, Ernest Hannig, Charles Moehle, Jeffrey Stratham, and Ronald Wek for their helpful comments on the manuscript, and Kathy Shoobridge for help in its preparation. This work was done while P.F.M. held a National Research Council—National Institutes of Health Research Associateship.

References
Bairn, S.B. and F. Sherman. 1988. mRNA structures influencing translation in the yeast Saccharomyces cerevisiae. Mol. Cell. Biol. 8: 1591–1601.

Cigan, A.M. and T.F. Donahue. 1987. Sequence and structural features associated with translational initiator regions in yeast—A review. Gene 59: 1–18.

Cigan, A.M., E.K. Fabich, and T.F. Donahue. 1988. Mutational analysis of the HIS4 translational initiator region in Saccharomyces cerevisiae. Mol. Cell. Biol. 8: 2964–2975.

Ehresmann, C., H. Moine, M. Mougel, J. Dondon, M. Grunberg-Manago, J.-P. Ebel, and B. Ehresmann. 1986. Cross-linking of initiation factor IF3 to Escherichia coli 30S ribosomal subunit by trans-diaminedichloroplatinum(II): characterization of two cross-linking sites in 16S rRNA; a possible way of functioning for IF3. Nucleic Acids Res. 14: 4803–4821.

Hinnebusch, A.G. 1985. A hierarchy of trans-acting factors modulate translation of the activator of general amino acid control in Saccharomyces cerevisiae. Mol. Cell. Biol. 5: 2349–2360.

Hinnebusch, A.G. 1988. Mechanisms of gene regulation in the general control of amino acid biosynthesis in Saccharomyces cerevisiae. Microbiol. Rev. 52: 248–273.

Hinnebusch, A.G. and G.R. Fink. 1983. Repeated DNA sequences upstream from HIS1 also occur at several other co-regulated genes in Saccharomyces cerevisiae. J. Biol. Chem. 258: 5238–5247.

Hinnebusch, A.G., Jackson, B.J., and P.P. Mueller. 1988. Evidence for regulation of reinitiation in translational control of GCN4 mRNA. Proc. Natl. Acad. Sci. 85: 7279–7283.

Hogan, J.J., R.R. Gutell, and H.F. Noller. 1984. Probing the conformation of 18S rRNA in yeast 40S ribosomal subunits with kethoxal. Biochemistry 23: 3322–3330.

Johansen, H., D. Schumperli, and M. Rosenberg. 1984. Affecting gene expression by altering the length and sequence of the 5′ leader. Proc. Natl. Acad. Sci. 81: 7698–7702.

Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. 9: 5233–5252.

1984. Selection of initiation sites by eukaryotic ribosomes: Effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. Nucleic Acids Res. 12: 3878–3893.

1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283–292.

1987a. Effects of intercistronic length on the efficiency of reinitiation by eukaryotic ribosomes. Mol. Cell. Biol. 7: 3438–3445.

1987b. An analysis of 5′-noncoding sequences from 69 vertebrate messenger RNAs. Nucleic Acids Res. 15: 8125–8148.

Lucchini, G., C.C. Simonsen, and A.D. Levinson. 1984. Initiation of translation at internal AUG codons in mammalian cells. Nature 309: 82–85.

Lucchini, G.A. Hinnebusch, C. Chen, and G. R. Fink. 1984. Positive regulatory interactions of the HIS4 gene of Saccharomyces cerevisiae. Mol. Cell. Biol. 4: 1326–1333.

Moldave, K. 1985. Eukaryotic protein synthesis. Annu. Rev. Biochem. 54: 1109–1149.

Mueller, P.P. and A.G. Hinnebusch. 1986. Multiple upstream AUG codons mediate translational control of GCN4. Cell 45: 201–207.

Mueller, P.P., B.M. Jackson, P.F. Miller, and A.G. Hinnebusch. 1988. The first and fourth upstream open reading frames in GCN4 mRNA have similar initiation efficiencies but respond differently in translational control to changes in length and sequence. Mol. Cell. Biol. 8: 5439–5447.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74: 5463–5467.

Sharp, P.M., T.M.F. Tuohy, and K.R. Mosurksi. 1986. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14: 5125–5143.

Sherman, F. and J.W. Stewart. 1983. Mutations altering initiation of translation of iso-1-cytochrome c; contrasts between the eukaryotic and prokaryotic initiation process. In The molecular biology of the yeast Saccharomyces cerevisiae: Metabolism and gene expression [ed. J.N. Stratham, E.W. Jones, and J.R. Broach], pp. 301–334. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Stiegler, P., P. Carbon, J.-P. Ebel, and C. Ehresmann. 1981. A general secondary-structure model for procaryotic and eucaryotic RNAs of the small ribosomal subunits. Eur. J. Biochem. 120: 487–491.

Tzamaras, D. and G. Thioreos. 1988. Evidence that the GCN2 protein kinase regulates reinitiation by yeast ribosomes. EMBO J. 7: 3547–3551.

Vazquez, D. 1979. Inhibitors of protein synthesis. Springer-Verlag, New York.

Warner, J.R. 1983. The yeast ribosome: structure, function, and synthesis. In The molecular biology of the yeast Saccharomyces cerevisiae: Metabolism and gene expression [ed. J.N. Stratham, E.W. Jones, and J.R. Broach], pp. 529–560. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Werner, M., A. Feller, F. Messenguy, and A. Pierard. 1987. The positive regulatory interactions of the CAI gene of Saccharomyces cerevisiae, a member of the yeast 5′ proximal open reading frames in GCN4 mRNA can be mimicked by heterologous, short coding sequences. Mol. Cell. Biol. 8: 3827–3836.

Zoller, M.J. and M. Smith. 1982. Oligonucleotide-directed mutagenesis using M13-derived vectors: An efficient and general procedure for the production of point mutations in any fragment of DNA. Nucleic Acids Res. 10: 6487–6500.
Sequences that surround the stop codons of upstream open reading frames in GCN4 mRNA determine their distinct functions in translational control.

P F Miller and A G Hinnebusch

*Genes Dev.* 1989, 3:
Access the most recent version at doi:10.1101/gad.3.8.1217

---

References
This article cites 27 articles, 12 of which can be accessed free at: [http://genesdev.cshlp.org/content/3/8/1217.full.html#ref-list-1](http://genesdev.cshlp.org/content/3/8/1217.full.html#ref-list-1)

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here.](#)