Translational Efficiency of a Non-AUG Initiation Codon Is Significantly Affected by Its Sequence Context in Yeast*

Shun-Jia Chen1, Grace Lin1, Kuang-Jung Chang1, Lu-Shu Yeh9, and Chien-Chia Wang1†

From the 1Department of Life Science, National Central University, Jung-li, Taiwan 32001 and the 9Department of Life Science, Tzu-Chi University, Hua-lien, Taiwan 97041

Previous studies have shown that translation of mRNA for yeast glycyl-tRNA synthetase is alternatively initiated from UUG and a downstream AUG initiation codon. Evidence presented here shows that unlike an AUG initiation codon, efficiency of this non-AUG initiation codon is significantly affected by its sequence context, in particular the nucleotides at positions −3 to −1 relative to the initiation codon. A/A/R (R represents A or G) and C/G/C appear to be the most and least favorable sequences at these positions, respectively. Mutation of the native context sequence −3 to −1 from AAA to CGC reduced translation initiation from the UUG codon up to 32-fold and resulted in loss of mitochondrial respiration. Although an AUG initiation codon is, in general, unresponsive to context changes in yeast, an AAA (−3 to −1) to CGC mutation still reduced its initiating activity up to 8-fold under similar conditions. These results suggest that sequence context is more important for translation initiation in yeast than previously appreciated.

Aminoacyl-tRNA synthetases are a group of primordial enzymes, each of which catalyzes the attachment of a specific amino acid to its cognate tRNAs. Typically there are 20 aminoacyl-tRNA synthetases in prokaryotes, one for each amino acid (1–4). In eukaryotes, protein synthesis occurs not only in the cytoplasm but also in organelles such as mitochondria and chloroplasts (5). Thus, eukaryotes such as yeast need two distinct sets of enzymes for each aminoacylation activity, one localized to the cytoplasm and the other localized to the mitochondria. Each set aminoacylates the isoaccepting tRNAs within its respective cell compartment and is sequestered from the isoacceptors confined in other compartments. In most cases, the cytoplasmic and mitochondrial aminoacyl-tRNA synthetase activities are encoded by two distinct nuclear genes, regardless of the cellular compartments to which they are confined. However, two Saccharomyces cerevisiae genes, HTS1 (the gene encoding histidyl-tRNA synthetase) (6) and VAS1 (the gene encoding valyl-tRNA synthetase) (7), specify both the mitochondrial and cytosolic forms through alternative initiation from two in-frame AUG codons. Each of these genes encodes mRNAs with distinct 5′-ends. The mitochondrial form of the enzyme is translated from the first AUG on the “long” mRNAs, whereas the cytosolic form is translated from the second AUG on the “short” mRNAs that have their 5′-ends located between the first and second AUG codons. Hence, the mitochondrial enzymes have the same polypeptide sequences as their cytosolic counterparts, except for a short amino-terminal mitochondrial targeting sequence. The transit peptide is subsequently cleaved away upon import into the mitochondria. Because the isozymes are targeted to different compartments, the two isoforms of valyl-tRNA synthetase, for example, cannot functionally substitute for each other in vivo (7, 8). A similar scenario has been observed for the genes encoding the mitochondrial and cytoplasmic forms of Arabidopsis thaliana alanyl-tRNA synthetase (AlaRS), 2 threonyl-tRNA synthetase, and valyl-tRNA synthetase (9). Although two isoforms of AlaRS are generated in a similar manner in Candida albicans, the longer form per se appears dual-targeted and thus can be functional in both compartments (10). Recently, two rare cases of one gene encoding both activities have been reported in yeast in which the protein isoforms are produced by alternative use of two in-frame initiation codons: an upstream non-AUG initiator and a downstream AUG initiator (11, 12).

Translation initiation in eukaryotes is a stringent process requiring not only initiator tRNA but also many protein factors, including eIF1, eIF2, eIF3, eIF4F, and eIF5. Upon binding to the cap structure, the 43 S preinitiation complex, composed of the 40 S ribosome and initiation factors, moves along the mRNA in a 3′ direction until it encounters the first AUG codon. At this point, GTP hydrolysis that leads to dissociation of the eIF2-GDP complex from the initiator tRNA in the preinitiation complex and subsequent P0 release signifies a 3-bp codon/anticodon interaction between Met-tRNAMet and the start codon (13). It has recently been shown that eIF1 plays a critical role in start codon selection. Mutations of eIF1 may lead to reduced interaction of this initiation factor with 40 S subunits and increased initiation at UUG codons (14). Previous studies on CYC1 (15) in S. cerevisiae suggested that AUG is the only codon recognized as the translational initiator and that the AUG codon nearest the 5′-end of an mRNA serves as the start site for translation. If the first AUG codon is mutated, then initiation can begin at the next available AUG from the 5′-end of the mRNA. The same

1 To whom correspondence should be addressed. Tel.: 886-3-426-0840; Fax: 886-3-422-8482; E-mail: dukewang@cc.ncu.edu.tw.

2 The abbreviations used are: AlaRS, alanyl-tRNA synthetase; ADH, alcohol dehydrogenase; GlyRS, glycyyl-tRNA synthetase; YPG, yeast extract-peptide-glycerol; eIF, eukaryotic initiation factor; RT, reverse transcription; β-gal, β-galactosidase.
rules apply to all eukaryotes. However, there are many examples in higher eukaryotes, where cellular and viral mRNAs can initiate from non-AUG codons that differ from AUG by just one nucleotide (16). The relatively weak base pairing between a non-AUG codon and the anticodon of an initiator tRNA appears to be compensated for by interactions with nearby nucleotides, in particular a purine (A or G) at position −3 and a G at position +4 (17, 18). Similarly, mutations in the sequence surrounding the first AUG can lead to its inefficient utilization as an initiator and subsequent use of an AUG at a downstream location. In addition to sequence context, a stable hairpin structure located 12–15 nucleotides downstream from the initiator can also facilitate recognition of a weak initiator by the 40 S ribosomal subunit (19). Although some reports suggest that sequences immediately preceding the initiation codon may also play a role in modulating the efficiency of translation initiation in yeast, the magnitude of this context effect appears relatively insignificant (20–22). Perhaps for that reason, yeast cannot efficiently use non-AUG codons as the translation start site (23, 24). Nonetheless, three yeast genes, GRS1 (one of the two glycyl-tRNA synthetase (GlyRS) genes in S. cerevisiae) (11), ALA1 (the only AlaRS gene in S. cerevisiae) (12), and CARP2A (the gene coding for acidic ribosomal protein P2A in C. albicans) (25) have recently been reported to use naturally occurring non-AUG triplets as translation initiators.

In the case of CARP2A, a non-AUG codon, i.e. GUG, serves as the exclusive translation initiator, whereas in the cases of ALA1 and GRS1, non-AUG codons act as alternative translation initiators that are accompanied by a downstream in-frame AUG initiation codon. Although two homologous GlyRS genes, namely GRS1 and GRS2, have been identified in the yeast genome, only one, i.e. GRS1, is functional and provides both cytoplasmic and mitochondrial aminoacylation activities, whereas the other, i.e. GRS2, appears to be dispensable for growth (26). Further studies by site-specific mutagenesis indicated that two functionally exclusive protein isoforms are alternatively generated from GRS1. A short form that is responsible for the cytoplasmic activity is translationally initiated from a classical AUG initiator, whereas a longer isoform that provides the mitochondrial activity is initiated from an upstream in-frame UUG codon (11). Expression of ALA1 follows a similar scenario. The mitochondrial form of AlaRS is initiated from two successive in-frame ACG triplets 69 nucleotides upstream of the AUG initiator of the cytoplasmic form (12). This feature of redundancy of ACG initiators may in itself represent a novel mechanism to improve the efficiency of a poor initiation event (27). However, in general the detailed molecular mechanism that enables the ACG (in ALA1) and UUG (in GRS1) codons to serve as translation initiation sites remains poorly understood. In particular, if other factors such as sequence context do not play a role in recognition of non-AUG initiators, the initiating ribosome should not be able to differentiate an authentic non-AUG initiation codon from other random non-AUG codons. In the work described here, we looked for the most and least favorable sequence contexts for the UUG initiator of GRS1 and compared the effects of these contexts on the efficiency of translation initiation. Contrary to many previous reports on context effect, our results argued that sequence context, in particular the nucleotides at relative position −3 to −1, is very important to the efficiency of the non-AUG initiator in GRS1. Mutation of the nucleotides at its relative position −3 to −1 from AAA to CGC reduced its initiating activity up to 32-fold. As a result, a GRS1 construct carrying this mutation was unable to maintain mitochondrial function because of drastically lowered protein levels. Although AUG was relatively insensitive to context changes under similar conditions, an AAA (−3 to −1) to CGC mutation still reduced its initiating activity up to 8-fold, suggesting that an AUG initiation codon could be subject to the same rule of modulation, albeit to a lesser extent. In this sense, a non-AUG initiator could perhaps serve as a better model in identifying the sequence and structural elements important for translation initiation in yeast.

**EXPERIMENTAL PROCEDURES**

**Construction of Plasmids—Cloning of GRS1 from S. cerevisiae** into pRS315, a low copy number yeast shuttle vector (28), and pADH, a high copy number yeast shuttle vector with a constitutive ADH promoter, has previously been described (11). Construction of various GRS1-lexA fusions for Western blot analysis was as previously described (11, 27). To clone a wild-type lexA sequence in pADH, the open reading frame of lexA was amplified by PCR as an NdeI-XhoI fragment and cloned in the appropriate sites in pADH. To clone an initiator mutant of lexA (designated here as lexA*), the Ndel site was mutated by site-directed mutagenesis to a Spel site. Therefore, in the wild-type lexA clone, the ATG initiator of lexA is part of the Ndel site, whereas in lexA*, the native ATG initiator is mutated to ACT and becomes part of the Spel site. A wild-type or mutant GRS1 sequence containing base pairs −88 to −12 relative to ATG1 was amplified by PCR as a PstI-Ndel or PstI-Spel fragment and cloned in-frame into the 5′-end of the wild-type lexA or its initiator mutant. These constructs were expressed under the control of a constitutive ADH promoter (29). Construction of GRS1-lacZ fusions followed a similar strategy. Basically, the lexA* portion in GRS1-lexA* fusions was replaced by an initiator mutant of lacZ. Thus, the only translation initiation site for the GRS1-lacZ fusions is TTG (−23) (or its derivatives) located in the GRS1 portion.

For generating various mutations in the 5′-untranslated region of the GRS1 gene, a short GRS1 sequence containing bp −300 +300 (relative to ATG1) was PCR-amplified from the full-length GRS1 clone as an Eagl-XbaI fragment and cloned in the appropriate sites of pBluescript II SK (Stratagene, La Jolla, CA). This plasmid containing the 600-bp GRS1 sequence was subsequently used as template for mutagenesis. Mutagenesis was carried out following standard protocols provided by the manufacturer (Stratagene, La Jolla, CA). After mutagenesis, the GRS1 fragment (−300+260 bp) carrying the desired mutation was removed from the plasmid by Eagl/Bsu361 digestion and then fused in-frame to the 5′-end of a truncated version of GRS1 lacking the Eagl-Bsu361 segment. Note that Bsu361 is a native cutting site between bp +260 and +266 of GRS1.

**RT-PCR—**To estimate the relative levels of specific GRS1-lexA mRNAs generated from the fusion constructs, an RT-PCR experiment was carried out following the protocols provided by the manufacturer (Invitrogen). The relative amounts of GRS1-
Effect of Sequence Context

Production of GlyRS Isoforms through Leaky Scanning—Our previous study showed that two distinct protein isoforms are generated from the yeast GRS1 gene via alternative initiation of translation. The longer form that accounts for the mitochondrial activity is initiated from UUG(−23) (the number “−23” in UUG(−23) refers to the 23rd codon triplet that precedes AUG1), whereas the short form that is responsible for the cytoplasmic activity is initiated from AUG1 (Fig. 1) (11). The leader peptide that is encoded by the nucleotide sequence between the two initiators acts as a targeting signal for the mitochondrial precursor form of GlyRS. Because only one mRNA species having its 5′-end located at position −88 relative to AUG1 was transcribed from GRS1 (11), and UUG(−23) is a poor initiation codon, translation of the GRS1 mRNA was expected to follow the mechanism of “leaky scanning” proposed by Kozak (30).

To provide direct evidence that scanning ribosomes can skip the relatively poor UUG initiation site and initiate at a downstream AUG, we employed Western blot analysis. Because the processed form of mitochondrial GlyRS is almost indistinguishable in size from its cytoplasmic counterpart, it is unfeasible to directly assay the relative amounts of the isofoms using Western blots. To circumvent this problem, we chose to use LexA as a reporter to measure the relative initiation activities of the alternative initiators of GRS1 (11, 31). In addition, to avoid mitochondrial import and cleavage of the signal peptide, only part of the leader sequence of GRS1 (bp −88 to −12) was used in construction of the GRS1-LexA fusions. Signal peptide deduced from the truncated leader sequence (containing amino acid residue Met(−23) to Val(−5)) lacks the predicted matrix processing peptidease cleavage site (Fig. 1) and was shown to be unable to target GlyRS into mitochondria (data not shown). As

lexA mRNAs were determined using a set of primers complementary to −40 to −20 nucleotides of GRS1 (5′-TTACTCTCAGATTGTTAAA-3′) and +370 to +390 nucleotides of lexA (5′-CAAGTCACCATCCATAATGCC-3′), respectively. For a more accurate comparison, cDNA products for each construct were obtained from three different cycle numbers of PCR amplification as indicated in the figure. As a control, the relative amounts of actin-specific mRNAs were determined using a set of primers complementary to +537 to +560 nucleotides (5′-ACCAACTTGGGACGATATGGAAAAG-3′) and +696 to +719 nucleotides (5′-TTGGATGGAAACGTAGAAGGC-TGG-3′) of actin, respectively.

Complementation Assays for the Mitochondrial Function of GRS1—The yeast GRS1 knockout strain, RJT3/II-1, has been described previously (26). To test the mitochondrial function of the GRS1 constructs, RJT3/II-1 was first transformed with a test plasmid (carrying a LEU2 marker) and plated on a selection medium lacking uracil and leucine. Following 5-fluoroorotic acid selections, a single colony of the transformants was picked and grown to stationary phase in a synthetic medium lacking a marker and 2 mM phenylmethylsulfonyl fluoride. Aliquots of the protein extracts were prepared from each of the transformants do not grow on YPG plates unless a functional mitochondrion survives on glycerol without functional mitochondria, the transformants are obtained from three different cycle numbers of PCR amplification as indicated in the figure. As a control, the relative amounts of actin-specific mRNAs were determined using a set of primers complementary to +537 to +560 nucleotides (5′-ACCAACTTGGGACGATATGGAAAAG-3′) and +696 to +719 nucleotides (5′-TTGGATGGAAACGTAGAAGGC-TGG-3′) of actin, respectively.

Western Blot—The protein expression patterns of the GRS1-lexA fusions were determined by a chemiluminescence-based Western blot analysis. The lexA fusion constructs were first transformed into INVScl (Invitrogen), and the transformants were subsequently grown in a selection medium lacking uracil. Total protein extracts were prepared from each of the transformants with a buffer containing 50 mM Tris, pH 7.4, 150 mM NaCl, 0.5% SDS, 0.5% Triton X-100, 10 mM EDTA, and 1 mM phenylmethylsulfonyl fluoride. Aliquots of the protein extracts (10–40 μg) were loaded onto a mini gel containing 10% polyacrylamide (size: 8 cm × 10 cm) or a medium gel containing 18% polyacrylamide (size: 8 cm × 20 cm) and electrophoresed at 100 volts for 2–4 h. Following electrophoresis, the resolved proteins were transferred using a semi-dry transfer device to a polyvinylidene difluoride membrane in a buffer containing 30 mM glycine, 48 mM Tris base, pH 8.3, 0.037% SDS, and 20% methanol. The membrane was probed with a horse-radish peroxidase-conjugated anti-LexA antibody (Santa Cruz, Santa Cruz, CA) and then exposed to an x-ray film following addition of the appropriate substrates. The Western blot data were obtained from at least three independent experiments (only one blot was shown in each figure), and the relative intensities of the protein bands shown were presented as a percentage using the mean value of the intensity of the protein band translated from the native initiator or an initiator with the native context as a reference.

β-Gal Assay—Yeast cells were pelleted by centrifugation at 12,000 × g for 30 s and resuspended in 100 μl of breaking buffer (100 mM Tris-HCl, pH 8.0, 1 mM dithiothreitol, 10% glycerol, and 2 mM phenylmethylsulfonyl fluoride) and 100-μl beads.
Effect of Sequence Context

schematized in Fig. 2, a wild-type or mutant GRS1 fragment containing bp −88 to −12 was PCR-amplified and fused in-frame to the 5′-end of lexA, yielding various GRS1-lexA fusions. These fusions were cloned in the yeast shuttle vector pADH and expressed under the control of a constitutive ADH promoter. As expected, two distinct protein bands were generated from the parental GRS1-lexA fusion (Fig. 2, number 1). Inactivation of the TTG or ATG initiator on the fusions specifically impaired the production of the upper and lower bands, respectively (compare numbers 2 and 4), suggesting that the upper and lower bands were initiated from the TTG and ATG codons, respectively. Consistently, inactivation of both initiators impaired the production of both bands (see number 5). Interestingly but not surprisingly, mutation of TTG(−23) to ATG strongly increased the levels of the upper band and completely abolished the production of the lower band (see number 3). Further mutation of the downstream ATG initiator codon to ACT in this construct did not alter the pattern of protein expression (see number 6). These results suggest that TTG(−23) can be readily bypassed by scanning ribosomes, and two protein isoforms can be generated from the GRS1 mRNA through leaky scanning.

Scanning for Sequence Elements Important for Non-AUG Initiation—The question arose as to why UUG(−23), and not other non-AUG codons that also differ from AUG by a nucleotide in the leader sequence, can be recognized by the preinitiating complex as an alternative translation initiator in GRS1. One possibility is that this non-AUG initiation site contains an unusual sequence context that can effectively compensate for

the poor interaction between the initiator codon and initiator tRNA. To test this hypothesis, a trinucleotide scanning strategy was carried out to map the approximate sequence elements important for non-AUG initiation. The mutations were first aimed at the nucleotide sequences flanking the initiating site. In practice, the 5′-flanking sequences that are rich in T and C were mutated to 5′-flanking sequences that are rich in A were mutated to T in a fashion of one codon at a time, whereas the 3′-flanking sequences that are rich in T and C were mutated to A in a similar fashion (Fig. 3A). Fig. 3B shows that mutations of the nucleotides at positions −3 to −1 relative to the TTG initiator to TTG had the strongest effect among all the mutations tested and reduced the initiating activity up to 10-fold (Fig. 3B, number 4), whereas mutations at other positions, including positions −9 to −7 (to TTT), −6 to −4 (to TTT), +4 to +6 (to AAA), +7 to +9 (to AAA), and +10 to +12 (to AAA), altered the initiating activity less than 2-fold (Fig. 3B, numbers 2, 3, and 5−7). These results suggest that the nucleotides at position −3 to −1 relative to the TTG initiator are the most important sequence elements for translation among the flanking sequences tested.
Previous studies in higher eukaryotes showed that recognition of an AUG initiator is significantly modulated by the nucleotides at its relative positions $-3$ and $+4$ (16). In contrast, our current study showed that mutations of the nucleotides at positions $+4$ to $+6$ had barely detectable effect in yeast (Fig. 3). To rule out the possibility that triple A mutations cannot faithfully reflect the effect of individual nucleotides on initiation, nucleotides at these positions were individually mutated and tested. As shown in Fig. 4A, except for mutation of T($+4$) to A, which increased the initiating activity ~2-fold, mutations at $+4$, $+5$, and $+6$ had little effect on initiation. It should be noted that changing C($+5$) to A would create an in-frame stop codon, TAG, and in turn block protein translation. To avoid this unfortunate situation, the nucleotide at position $+6$ was concomitantly mutated to C when testing the effect of the nucleotide $+5$. Taken together, the result suggests that the nucleotides at positions $+4$ to $+6$ are not as significant to the initiating activity as those at position $-3$ to $-1$, and the $+4$ nucleotide appears to be the only one with a discernible effect on initiation among the three.

Comparing the Effect of Various Sequence Contexts on AUG and Non-AUG Initiators—To examine whether sequence context has a similar effect on translation from various initiation codons in yeast, we next mutated the TTG initiation codon to ACG (another non-ATG initiation codon found in yeast) and ATG (a canonical initiation codon) and then tested their responses to various contexts. As shown in Fig. 5A, ACG and ATG could each substitute for the initiator function of TTG($-23$) in GRS1 (Fig. 5A, numbers 1, 5, and 9). However, when nucleotides $-3$ to $-1$ relative to the TTG or ACG initiation site were mutated from AAA (presumably the best context) to CGC (presumably the worst context), the resultant constructs lost much of their initiating activity and failed to provide the necessary mitochondrial GlyRS function to the knockout strain (Fig. 5A, numbers 2 and 10). This finding, although interesting, was not totally unexpected in view of the fact that the TTG initiator is highly sensitive to its context changes (Fig. 4). However, it was surprising to find that a similar mutation also affected the complementing activity of the ATG initiator, albeit to a much lesser extent (Fig. 5A, numbers 5 and 6). In contrast to the dramatic effect caused by the mutation of AAA to CGC, mutation of AAA to CAA or TTT had only mild or little effect on the complementing activity of the resultant constructs (Fig. 5A, numbers 3, 4, 7, 8, 11, and 12).

Quantitative analysis of the initiating activity by Western blots showed that these three initiators, TTG, ATG, and ACG, had a relative initiating activity of 1:10:0.5 (Fig. 5B, numbers 1, 5, and 9). When the context sequences $-3$ to $-1$ were mutated from AAA to TTT or CAA, the initiating activity of the TTG codon was reduced by 10–12-fold (Fig. 5, B and C, compare numbers 1, 3, and 4). In contrast, when the context sequences were mutated to CGC, the initiating activity of the TTG codon was reduced up to ~32-fold (Fig. 5, B and C, compare numbers 1 and 2). This result may explain why only pSJ198 lost its complementing function, whereas pSJ215 and pKJ30 did not (Fig. 5A, compare numbers 2–4). A similar hierarchy of context preference was found for the ACG initiator (Fig. 5, B and C; numbers 9–12). In contrast to the non-ATG initiators, an ATG

Identifying the Best and Worst Sequence Contexts—The trinucleotide scanning strategy provided a rough picture about which triplet might be important for non-AUG initiation. However, it was still not clear which nucleotide of the triplet affects the initiating activity the most and to what extent. Pursuant to this objective, the nucleotides at position $-3$ to $-1$ (relative to TTG($-23$)) were each mutated from A to T, G, and C, and the resultant mutants were analyzed for their initiating activities by Western blots. To obtain a more accurate estimate of the relative levels of protein expression, protein extracts prepared from the wild-type construct were 2-fold serially diluted (indicated at the bottom of the blot) and compared with those from the mutants. The quantitative data for the relative initiating activities of the mutants are shown as a separate diagram at the bottom.

Effect of Sequence Context

FIGURE 4. Effect of sequence context on the native TTG initiator. A, assay of initiating activity by Western blots. For clarity, the wild-type GRS1 sequence (Wt) and the nucleotide positions relative to the TTG initiator are shown on the top. The numbers and nucleotides shown on top of the blots denote the positions and mutations of the nucleotides, respectively. Upper panel, GlyRS-LexA fusion; lower panel, PGK (as loading controls). The junction between GRS1 and lexA sequences is a Spel site. B, assay of relative initiating activity by Western blots. To obtain a more accurate estimate of the relative levels of protein expression, protein extracts prepared from the wild-type construct were 2-fold serially diluted (indicated at the bottom of the blot) and compared with those from the mutants. The quantitative data for the relative initiating activities of the mutants are shown as a separate diagram at the bottom.
initiator appeared to be less sensitive toward its context changes; mutation of the context sequences −3 to −1 from AAA to CAA or TTT had no discernible effect on its initiating activity (Fig. 5B, compare numbers 5, 7, and 8). However, when the context sequences were mutated to CGC, the initiating activity of the ATG initiator was reduced by 8-fold (Fig. 5, B and C, compare numbers 5 and 6), suggesting that an ATG initiator is subjected to the same rule of modulation, albeit to a much lesser extent. Interestingly, although the ATG initiator with the worst context had an initiating activity similar to that of the TTG or ACG initiator, the transformants carrying this compromised ATG initiator showed a relatively poor growth phenotype on YPG (Fig. 5A, compare numbers 6, 1, and 9). To check whether the fusions shown in Fig. 5B express similar levels of mRNAs, a semi-quantitative RT-PCR experiment was conducted. Three different cycle numbers of PCR amplification were carried out for each construct. Fig. 5D shows that similar levels of the cDNA products were generated from these fusions, suggesting that these mutations did not compromise the stability of the specific mRNAs.

Context Effect Determined Using lacZ as a Reporter Gene—To verify whether the quantitative assays shown in Fig. 5 truthfully reflect the initiating activities of the initiation codons with various contexts, we next used lacZ as a reporter. The lacZ portion in each of the GRS1-lexA fusions (Fig. 5B, numbers 1–8) was replaced by a lacZ reporter gene where the native initiator codon ATG had been mutated to ACT. Consequently, the initiator codon TTTG (−23) (or its derivatives) in the GRS1 portion becomes the only initiation site for the production of the GlyRS-LacZ fusion proteins (schematized in Fig. 6). Consistent with the results obtained using lexA as a reporter, the GRS1-lacZ fusions showed a similar pattern of expression. Mutation of the context sequence −3 to −1 from AAA to CGC, TTT, and CAA reduced the initiating activities of the TTG codon 30-, 6-, and 9-fold, respectively (Fig. 6, compare numbers 1–4). Similar mutations reduced the initiating activities of the ATG codon 4-, 1-, and 1-fold, respectively (compare numbers 5–8). In addition, the initiating activity of the TTG codon (with AAA context) was only 3-fold lower than that of the ATG codon with the same context (compare numbers 1 and 5). Although the relative levels of the LacZ fusions were slightly different from those of the LexA fusions, the overall patterns of

![FIGURE 5. Effect of sequence context on ATG and non-ATG initiators. A, complementation assays for the mitochondrial function on YPG plates. For clarity, the initiation codon (underlined) and its 5′-flanking sequence (positions −3 to −1) are shown for each construct. B, assay of initiating activity by Western blots. Upper panel, GlyRS-LexA fusion; lower panel, PGK (as loading controls). The junction between GRS1 and lexA sequences is a SpeI site. C, assay of relative initiating activity by Western blots. Protein extracts prepared from constructs with the native context were 2-fold serially diluted and then compared with those prepared from constructs with a mutant context. The quantitative data for the relative initiating activities of the mutants are shown as a separate diagram at the bottom. D, RT-PCR. The relative amounts of specific GRS1-lexA mRNAs generated from each construct were determined by RT-PCR (see “Experimental Procedures”). As a control, the relative amounts of actin mRNAs were also determined and only the cDNA products of 24 cycles of PCR amplification were shown. In C and D, the numbers 1–12 (circled) denote constructs shown in B. The GRS1 sequences used in GRS1-lexA fusion constructs 1–12 of B were transferred from constructs 1–12 shown in A, respectively.]

![FIGURE 6. Context effect determined using lacZ as a reporter. Efficiencies of translation initiation from TTG or ATG codon with various contexts were determined by introducing the GRS1-lacZ constructs into yeast and testing the relative β-gal activities in extracts prepared from the transformants. The open and solid boxes represent the GRS1 and lacZ sequences, respectively. For clarity, the initiation codons (underlined) and their 5′ context sequences are labeled on top of the schematics. The data were obtained from three independent experiments, and the relative β-gal activity for each construct was presented as the mean ± 2× S.D.) using the β-gal activity of the construct carrying AAATTG as a reference.]

**Effect of Sequence Context**
their protein expression were consistent. Thus, the results obtained from the LexA fusions appear consistent.

**DISCUSSION**

We presented evidence here that recognition of the alternative translation initiators of GRS1, UUG(−23) and AUG1, could be mediated by a mechanism known as leaky scanning (30), wherein only a small portion of the scanning ribosomes (∼10% in this case) initiate from the first initiator, i.e., UUG(−23), whereas the rest of the ribosomes skip this start site and continue scanning toward the 3’-end of the mRNA until they encounter the next available initiator, i.e., AUG1. As a consequence, mutations that replaced the upstream UUG initiator codon with AUG effectively impaired the production of the short protein form (Fig. 2). On the other hand, because only a minor portion of the scanning ribosomes start from the native upstream initiator, inactivation of this initiator did not result in a drastic increase in initiation from the downstream AUG initiator. Although leaky scanning might also be involved in the synthesis of AlaRS isoforms in S. cerevisiae (12, 31), no solid experimental evidence has been presented so far to confirm this hypothesis. In addition, it should be noted that three distinct mRNAs (with their 5’-ends located at nucleotide positions −117, −105, and −54 relative to the AUG initiator, respectively) are alternatively transcribed from the AlaRS gene. Thus, the cytoplasmic form of AlaRS could be translated from the longer two mRNAs via leaky scanning or directly from the shortest mRNA (12). In this sense, it is interesting to mention that expression of two yeast genes, MOD5 (coding for isopen-tenyl pyrophosphatetRNA isopenetyl transferase) (32) and CCA1 (coding for ATP (CTP):tRNA nucleotidylyltransferase) (33), also involves leaky scanning. However, leaky scanning occurs in these two genes not because the upstream initiator is a non-AUG codon or an initiator with a suboptimal sequence context, but because the first AUG codon in the two genes is positioned too close to the 5’-end of the mRNA. In contrast to the few cases of leaky scanning found in yeast, this mechanism has been observed frequently in mammals (34–37).

Many cellular and viral mRNAs have been shown to use non-AUG codons, such as AUU, CUG, and GUG, as translation start sites (30). Occasionally, non-AUG codons act as exclusive translation initiators, such as in the mRNAs derived from the Arabidopsis AGAMOUS gene (38) and the latently expressed kaposi locus of Kaposi’s sarcoma-associated herpesvirus (39). However, in most cases, non-AUG codons serve as alternative translation start sites that are accompanied by a downstream in-frame AUG initiator (17). Recognition of the non-AUG initiators in these mRNAs by initiating ribosomes appears to be compensated for by interactions with nucleotide sequences flanking the initiators. A recent study suggested that components of the 48 S translation initiation complex, in particular eIF2α and 18 S rRNA, might be involved in specific recognition of the context nucleotides at positions −3 and +4 (18). So, an optimal sequence context is important to the efficient recognition of a non-AUG initiator. Likewise, recognition of a canonical AUG initiator can be severely compromised by a suboptimal sequence context. In contrast to the active participation of sequence context in translation initiation in higher eukaryotes, many reports argued that sequence context does not appear to substantially affect the initiating activity of an AUG initiator in yeast (21, 22, 40). Mostly for this reason, yeast was previously thought to be unable to use non-AUG codons as translation start sites (24). Surprisingly, two genes in S. cerevisiae (ALA1 and GRS1) and one gene in C. albicans (CARP2A) have recently been reported to use naturally occurring non-AUG triplets as translation initiators. Even more surprising was the finding reported here that the UUG initiator of GRS1 is strongly affected by its 5’-flanking sequences (Figs. 4–6). As it turned out, this noncanonical initiator happens to have a very good, if not the best, sequence context, which may explain why only this particular UUG triplet, and not other non-AUG codons that also differ from AUG by a nucleotide elsewhere in the leader sequence, can function as a translation start site (11). Perhaps for that same reason, an ACG triplet can functionally substitute for the TTG initiator (Fig. 5A, number 9). In addition, it is interesting to point out that the yeast AlaRS gene, which uses ACG as an alternative initiation codon, also contains the optimal 5’ sequence context AAR (AAG in this case), lending further support to the findings of this report.

In higher eukaryotes, the most critical context positions are −3 and +4 relative to the initiator, and the best (or most favorable) context contains a purine (A or G) at position −3 and a G at position +4 (30). Although sequences immediately flanking AUG initiation codons are somewhat preferred in yeast, the bias in nucleotide distribution (5’···(A/Y)A(A/Y)AaugUCY···3’) is highly diverged from the higher eukaryotic consensus (5’···CACCaugG···3’) with the exception of an A preference at the −3 position (20). In addition, many studies suggested that sequence context plays only a minor role in translation of yeast mRNAs initiated from AUG. In this sense, the yeast consensus deduced solely from sequence analysis might not actually represent the best context for translation initiation in yeast. We report here that the best context for the UUG initiator of the yeast GRS1 gene is AAR (R represents A or G) at nucleotide positions −3 to −1, and the most critical context position is position −3 (Figs. 4 and 5). Mutation of A(−3) to C alone reduced the initiating activity of the UUG initiator (or an ACG codon positioned in the same context) up to 12-fold (Figs. 4 and 5). To our knowledge, this is the most dramatic context effect that has been reported for an initiator in yeast. Consistent with these findings, many highly expressed yeast genes are found to have an A at the −3 position and an A-rich leader region (20). In contrast to position −3, the nucleotide at position +4 appears to have only minor effect on the efficiency of the UUG initiator, which is very different from the scenario seen in higher eukaryotes. Mutation of the nucleotide at this position to any other nucleotide affected the initiating activity at most by 2-fold, which is much less than the effect manifested by any of the nucleotides at positions −3, −2, and −1 (Fig. 4). Interestingly, recent studies suggested that the nature of the penultimate amino-terminal residue may influence the substrate preferences of the methionine aminopeptidase and N-α-acetyltransferase and in turn affect the stability of the produced protein (41, 42). Therefore, changes in the +4, +5, or +6 position may affect not only the context preference but also the stability of the protein. A more thorough study is needed to delineate the
Effect of Sequence Context

effect of nucleotides at these positions. So far, it is not known whether secondary structure also plays a role in modulating the efficiency of the UUG initiator. Although most of the mutations shown in Figs. 3 and 5 did not lead to formation or loss of a stable secondary structure near the initiation site, mutation of AAA (position −3 to −1 relative to the UUG initiator) to CGC of or of UCG (position +4 to +6 relative to the UUG initiator) to AAA was predicted to facilitate the formation of a weak secondary structure (with ∆G of −10 and −7 kcal/mol, respectively). In the CGC (position −3 to −1) mutant, AAGAACG (position −8 to −2) pairs with CGUUCUU (position +5 to +11) (Fig. 5, number 2), whereas in the AAA (position +4 to +6) mutant, UUUCAAA (−12 to −7) pairs with UUGAAA (+1 to +6) (Fig. 3, number 5). Both stem-loop structures include the UUG initiation site; UUG is either in the loop or in the stem. Strangely enough, the CGC mutation significantly reduced the efficiency of translation initiation (Fig. 5B, compare numbers 1 and 2), whereas the AAA mutation had little effect on UUG initiation (Fig. 3B, compare numbers 1 and 5). Further investigation is under way to check whether the weak structure formed in the CGC mutant actually plays a role in UUG initiation.

It should be noted that the context preference reported here was deduced mainly from a non-AUG initiator model (Figs. 3 and 4), and an AUG initiator appeared to be much less sensitive under similar conditions (Figs. 5 and 6). For example, mutation of A(−3) to C or AAA (−3 to −1) to UUU had little effect on an AUG initiation codon placed in the same position as UUG(−23) (Figs. 5 and 6). This feature appears to be consistent with many of the previous reports, where various mutations in the 5′ sequence context affected the translation initiation from an AUG codon at most 2–3-fold (21, 22). Nevertheless, mutations that changed the context of an AUG initiator from AAA (−3 to −1) to CGC significantly impaired its initiation activity (4–8-fold) (Figs. 5 and 6). This result is particularly noteworthy because a single nucleotide change even in the most critical position has little effect on translation from an AUG codon. Therefore, context effect is likely to be less obvious if a relatively insensitive system with AUG as the initiation codon is used as the test model (24). A non-AUG initiation codon such as the one present in GRS1 can perhaps serve as a better model in identifying the sequence, and structural elements important for translation initiation that may have previously been overlooked in yeast.

Acknowledgment—We thank Dr. Huey-Nan Wu (Academia Sinica) for technical support and helpful suggestions.

REFERENCES

1. Carter, C. W., Jr. (1993) Annu. Rev. Biochem. 62, 715–748
2. Martinis, S. A., and Schimmel, P. (1996) in Escherichia coli and Salmonella Cellular and Molecular Biology (Neidhardt, F. C., ed) 2nd Ed, pp. 887–901. American Society of Microbiology, Washington, DC
3. Gieger, R., Sissler, M., and Florentz, C. (1998) Nucleic Acids Res. 26, 5017–5035
4. Pelchat, M., and Lapointe, J. (1999) Biochem. Cell Biol. 77, 343–347
5. Dietrich, A., Weil, J. H., and Marechal-Drouard, L. (1992) Annu. Rev. Cell Biol. 8, 115–131
6. Natsoulis, G., Hilger, F., and Fink, G. R. (1986) Cell 46, 232–243
7. Chatton, B., Walter, P., Ebel, J.-P., Lacroute, F., and Fasiolo, F. (1988) J. Biol. Chem. 263, 52–57
8. Wang, C. C., Chang, K. J., Tang, H. L., Hsieh, C. J., and Schimmel, P. (2003) Biochemistry 42, 1646–1651
9. Souciet, G., Menand, B., Ovessa, J., Cosset, A., Dietrich, A., and Wintz, H. (1999) Eur. J. Biochem. 268, 848–854
10. Huang, H. Y., Kuei, Y., Chao, H. Y., Chen, S. J., Yeh, L. S., and Wang, C. C. (2006) J. Biol. Chem. 281, 31430–31439
11. Chang, K. J., and Wang, C. C. (2004) J. Biol. Chem. 279, 13778–13785
12. Tang, H. L., Yeh, L. S., Chen, N. K., Rimpiart, T., Schimmel, P., and Wang, C. C. (2004) J. Biol. Chem. 279, 49656–49663
13. Unbehaun, A., Borukhov, S. I., Hellen, C. U., and Pestova, T. V. (2004) Genes Dev. 18, 3078–3093
14. Cheung, Y. N., Maag, D., Mitchell, S. F., Fekete, C. A., Algire, M. A., Takacs, J. E., Shirokikh, N., Pestova, T., Larsch, J. R., and Hinnebusch, A. G. (2007) Genes Dev. 21, 1217–1230
15. Sherman, F., Stewart, J. W., and Schweingruber, A. M. (1980) Cell 20, 215–222
16. Kozak, M. (1989) Mol. Cell Biol. 9, 5073–5080
17. Kozak, M. (1991) J. Biol. Chem. 266, 19867–19870
18. Pisarev, A. V., Kolupaeva, V. G., Pisareva, V. P., Merrick, W. C., Hellen, C. U. T., and Pestova, T. V. (2006) Genes Dev. 20, 624–636
19. Kozak, M. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 8301–8305
20. Cigan, A. M., and Donahue, T. F. (1987) Gene (Amst.) 59, 1–18
21. Baim, S. B., and Sherman, F. (1988) Mol. Cell Biol. 8, 1591–1601
22. Cigan, A. M., Pabich, E. K., and Donahue, T. F. (1988) Mol. Cell Biol. 8, 2964–2975
23. Zitomer, R. S., Walthall, D. A., Rymond, B. C., and Hollenberg, C. P. (1984) Mol. Cell Biol. 4, 1191–1197
24. Clements, J. M., Laz, T. M., and Sherman, F. (1988) Mol. Cell Biol. 8, 4533–4536
25. Abramczyk, D., Tchorzewski, M., and Grankowski, N. (2003) Yeast 20, 1045–1052
26. Turner, R. J., Lovato, M., and Schimmel, P. (2000) J. Biol. Chem. 275, 27681–27688
27. Chang, K. J., Lin, G., Men, L. C., and Wang, C. C. (2006) J. Biol. Chem. 281, 7775–7783
28. Sikorski, R. S., and Hieter, P. (1989) Genetics 122, 19–27
29. Bennetzen, J. L., and Hall, B. D. (1982) J. Biol. Chem. 257, 3018–3025
30. Kozak, M. (1999) Gene (Amst.) 234, 187–208
31. Huang, H. Y., Tang, H. L., Chao, H. Y., Yeh, L. S., and Wang, C. C. (2006) Mol. Microbiol. 60, 189–198
32. Slusher, L. B., Gillman, E. C., Martin, N. C., and Hopper, A. K. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 9789–9793
33. Wolfe, C. L., Lou, Y. C., Hopper, A. K., and Martin, N. C. (1994) J. Biol. Chem. 269, 13361–13366
34. Acland, P., Dixon, M., Peters, G., and Dickson, C. (1990) Nature 343, 662–665
35. Saris, C. J., Domen, J., and Berns, A. (1991) EMBO J. 10, 655–664
36. Hahn, S. R., Sloan-Brown, K., and Spotts, G. D. (1992) Genes Dev. 6, 1229–1240
37. Packham, G., Brimmel, M., and Cleveland, J. L. (1997) Biochem. J. 328, 807–813
38. Riechmann, I. L., Ito, T., and Meyerowitz, E. M. (1999) Mol. Cell Biol. 19, 8505–8512
39. Sadler, R., Wu, L., Forghani, B., Renne, R., Zhong, W., Herndier, B., and Gannem, D. (1999) J. Virol. 73, 5722–5730
40. Yoon, H., and Donahue, T. F. (1992) Mol. Microbiol. 6, 1413–1419
41. Huang, S., Elliott, R. C., Liu, P. S., Koduri, R. K., Weickmann, J. L., Lee, J. H., Blair, L. C., Ghosh-Dastidar, P., Bradshaw, R. A., Bryan, K. M., Einason, B., Kendall, R. L., Kolacz, K. H., and Saito, K. (1987) Biochemistry 26, 8242–8246
42. Polevoda, B., Norbeck, J., Takakura, H., Blomberg, A., and Sherman, F. (1999) EMBO J. 18, 6155–6168