Proximity of the Start Codon to a Leaderless mRNA's 5' Terminus Is a Strong Positive Determinant of Ribosome Binding and Expression in *Escherichia coli*[▽]

Karthik M. Krishnan,[#] William J. Van Etten III,[†] and Gary R. Janssen^{*}

Department of Microbiology, Miami University, Oxford, Ohio 45056

Received 29 June 2010/Accepted 4 October 2010

An AUG start codon is an important determinant of ribosome binding and expression of leaderless mRNAs in *Escherichia coli*. Using reporter constructs encoding mRNAs where the AUG start codon is preceded by untranslated leaders of various length and sequence, we find that close proximity of the start codon to the 5' terminus and the leader sequence are strong determinants of both ribosome binding and expression.

Ribosome recognition and binding to mRNA is a rate-limiting step of translation and a major determinant of gene expression. Although a Shine-Dalgarno (SD) sequence (25) is known to speed the formation of initiation complexes and result in higher expression, prokaryotes also contain noncanonical mRNAs, many of which lack a SD-type sequence but are highly expressed (6, 9, 24, 26). Leaderless mRNAs carrying a start codon at their 5' terminus are a group of noncanonical mRNAs found in all domains of life (5, 8, 15, 23). Much of our understanding about ribosome binding to a 5'-terminal AUG comes from studies of *Escherichia coli* with leaderless λ cI mRNA (1, 3, 17–19, 29), whereby *in vitro* ribosome binding and translation assays have successfully demonstrated the 70S ribosome's ability to bind and initiate translation of leaderless mRNA (16, 17, 28). A mutant initiator tRNA with compensating anticodon mutations restored expression of leadered, but not leaderless, mRNAs with UAG start codons, indicating that codon-anticodon complementarity was insufficient for the translation of leaderless mRNA (29). These data suggest that a cognate AUG initiation codon specifically serves as a stronger and different translational signal in the absence of an untranslated leader. In this report, we examine the influence of AUG start codon proximity to the 5' terminus and the influence of upstream nucleotides, if present, on 70S ribosome binding and on expression in *E. coli*.

In order to determine if efficient ribosome recognition of leaderless mRNA requires the AUG to be at the 5' terminus, we prepared cI-lacZ fusions in which the distance from the transcriptional start site to the start codon was increased by the addition of "TC" multimers. Transcription *in vivo* results in mRNA containing a "UC" repeat sequence upstream (5') to the start codon and can be used to assess the effects of the short leaders and recessed AUGs on expression. Transcription *in vitro* provides mRNAs with short leaders and recessed

AUGs for use in ribosome binding studies; transcription with T7 RNA polymerase results in the addition of a G as the initiating nucleotide to all *in vitro*-synthesized mRNAs. The choice of "UC" for the untranslated region was made to avoid resemblance to a typical purine-rich Shine-Dalgarno-type sequence. β -Galactosidase assays (14) were performed using *E. coli* RFS859 (21) as a host. Ribosome purification (*E. coli* MRE600) (30), primer extension inhibition (toeprint) assays, and *in vitro* synthesis and purification of mRNAs were performed as described previously (2, 13); toeprint reaction products were separated on denaturing polyacrylamide gels, and toeprint signals were quantified with a Molecular Dynamics (Storm 800) phosphorimager. Relative toeprint complexes (RTC) were calculated as described previously (2).

Toeprint assays revealed signals at +16 relative to A (+1) of the AUG start codon, and the toeprint signal decreased in intensity as the start codon was recessed inward by the addition of "UC" leaders (Fig. 1A). Quantification of the toeprint signals revealed that the relative toeprint signal decreased from 100% with leaderless LL-cI-lacZ mRNA to 74%, 26%, 4%, and 4% binding after the addition of (UC)1, (UC)2, (UC)3, and (UC)6 leaders, respectively, upstream to the start codon (Fig. 1B). In order to determine if the recessed AUG effects on *in vitro* ribosome binding correlated with *in vivo* expression, we constructed p(UC)*n*-cI-lacZ fusions by PCR-directed additions of (TC)*n* (where *n* = 1, 2, 3, or 6) upstream of the cI start codon in pLL-cI-lacZ (18). β -Galactosidase assays revealed that expression from the p(UC)*n*-cI-lacZ fusions decreased in parallel with the decrease in ribosome binding, with negligible expression or ribosome binding observed when the 5'-terminal AUG start codon was recessed by the addition of (UC)3 or (UC)6 (Fig. 1B). These results suggest that AUG proximity to the mRNA's 5' terminus is a strong determinant for ribosome recognition and expression of leaderless mRNA.

Sequences immediately downstream of the start codon are also determinants of ribosome binding and expression of leadered mRNAs (2, 11, 13). In order to test whether the inhibitory effect of UC leaders on expression was independent of downstream coding sequence, we constructed p(UC)*n*-lacZ and p(UC)*n*-uidA reporter constructs (where *n* = 1, 2, 3, 6, or 9) that were deleted of the DNA encoding the lacZ and uidA untranslated leaders. Primer extension analyses of

^{*} Corresponding author. Mailing address: Department of Microbiology, Miami University, Oxford, OH 45056. Phone: (513) 529-5422. Fax: (513) 529-2431. E-mail: janssegr@muohio.edu.

[#] Present address: Pathobiology and Molecular Medicine, University of Cincinnati, Cincinnati, OH 45267.

[†] Present address: BioTeam, Inc., Middleton, MA 01949.

[▽] Published ahead of print on 22 October 2010.

FIG. 1. (A) Ribosome binding (toeprint) assays for leaderless *cI-lacZ* (LL-*cI-lacZ*) or (UC)*n-cI-lacZ* (where *n* = 1, 2, 3, or 6) mRNAs. Toeprint assays were performed with the indicated *in vitro*-synthesized mRNA (22 nM), 70S ribosomes (66 nM), and initiator tRNA (132 nM). The full-length (FL) cDNA product is shown, and the arrowhead indicates the position of the tRNA^{Met}-dependent toeprint signal (+16) relative to the first position of the AUG start codon (+1). (B) Relative β -galactosidase activity expressed from cells containing leaderless (pLL-*cI-lacZ*) or the p(UC)*n-cI-lacZ* (where *n* = 1, 2, 3 or 6) constructs; also plotted, for comparison, are the relative ribosome binding activities, averaged from results of three independent assays, as shown in 1A. LacZ activity expressed from LL *cI-lacZ* was 10,704 Miller units (= 100%).

(UC)*n-lacZ* (where *n* = 1, 2, 3, 6, or 9) transcriptional start sites revealed that transcription in each case initiated at the predicted uridine residue of the UC upstream sequence (data not shown). We assume that transcription initiates at the same position of the UC leader when found upstream of the *cI-lacZ* and *uidA-lacZ* fusions. β -Galactosidase (14) and β -glucuronidase assays (10) revealed that short leaders of (UC)1, (UC)2, (UC)3, (UC)6, or (UC)9 upstream to the AUG start codon decreased expression from 100% (with no added "UC"), to 37%, 15%, 3%, 1%, or 1% β -galactosidase activities (LacZ; Fig. 2A) and 95%, 30%, 7%, 1%, or 1% β -glucuronidase activities (UidA; Fig. 2B), respectively, indicating that recessing the AUG inward by the addition of the UC multimers dramatically decreased expression regardless of the downstream coding sequence.

FIG. 2. Relative β -galactosidase (A and C) and β -glucuronidase activities (B). (A) Expression from cells containing the leaderless *lacZ* (pLL-*lacZ*) or p(UC)*n-lacZ* (where *n* = 1, 2, 3, 6, or 9). (B) Expression from cells containing the leaderless *uidA* (pLL-*uidA*) or p(UC)*n-uidA* (where *n* = 1, 2, 3, 6, or 9). (C) Expression from cells containing the leaderless *cI-lacZ* (pLL-*cI-lacZ*) or *cI-lacZ* fusions containing the three-nucleotide leaders GAU (pGAT-*cI-lacZ*), GAA (pGAA-*cI-lacZ*), GGC (pGGC-*cI-lacZ*), AAA (pAAA-*cI-lacZ*), or AGG (pAGG-*cI-lacZ*) upstream (5') to the *cI* start codon. LacZ activity expressed from pLL-*lacZ* and pLL-*cI-lacZ* were 225 and 12,956 Miller units (= 100%), respectively. Glucuronidase activity expressed from pLL-*uidA* was 790 Gus units (= 100%).

Because the nucleotide sequence immediately upstream of a leadered mRNA's start codon can contribute to expression levels (7), it is necessary to study the influence of a short leader's nucleotide sequence on expression and distinguish it from the effect of the AUG's proximity to the 5' terminus. From a large pool of -1 triplet-containing constructs (data not shown), we identified five *cI-lacZ* fusion plasmids that showed

FIG. 3. Ribosome binding (toeprint) assays for leaderless *yceD* (LL-*yceD*) and GCCU-*yceD* mRNAs. Toeprint assays were performed with the indicated *in vitro*-synthesized mRNA (22 nM), 70S ribosomes (66 nM), and initiator tRNA (132 nM). The full-length (FL) cDNA product is indicated, and the arrowhead indicates the position of the tRNA^{fMet}-dependent toeprint signal (+16) relative to the first position of the AUG start codon (+1).

a range of effects on expression. LacZ assays of cells expressing *cI-lacZ* fusions containing the three-nucleotide leaders GAU (pGAT-*cI-lacZ*), GAA (pGAA-*cI-lacZ*), GGC (pGGC-*cI-lacZ*), AAA (pAAA-*cI-lacZ*), or AGG (pAGG-*cI-lacZ*) upstream (5') to the *cI* start codon show that expression from mRNAs containing the GAU, GAA, GGC, AAA, or AGG "–1 triplets" decreased from 100% (with no added "–1 triplet") to 56%, 22%, 15%, 10%, or 3% β -galactosidase activities, respectively (Fig. 2C). These results indicate that recessing the 5'-terminal AUG inward generally decreases expression but the amount of decrease is influenced greatly by the sequence of the short leader.

Since none of the mRNAs we constructed are native to *E. coli*, we considered it necessary to test whether the observed influence of AUG's proximity to the 5' terminus extends to an endogenous *E. coli* gene as well. *E. coli*'s *yceD* gene is reported to contain four transcriptional start sites, resulting in mRNAs with untranslated leader lengths of 144, 3, 2, or 0 nucleotides (20, 27). Examination of the ribosome binding strengths for the leaderless (LL-*yceD*) and the three-nucleotide leader (CCU-*yceD*) mRNAs represents an opportunity to compare ribosome binding for an *E. coli* endogenous mRNA that occurs both as a leaderless and a short leadered mRNA. Toeprint analysis of *in vitro*-synthesized *yceD* mRNAs revealed that ribosomes bind AUGs of both LL-*yceD* and CCU-*yceD* mRNAs, but ribosome binding to CCU-*yceD* mRNA was decreased by 5-fold compared with that to LL-*yceD* mRNA (Fig. 3), indicating that a

recessed start codon is detrimental also for ribosome binding to a naturally occurring endogenous *E. coli* mRNA.

Earlier reports suggest that *E. coli* 70S ribosomes preferably bind an mRNA's 5' terminus in a sequence-specific manner that does not relate entirely to codon anticodon complementarity (3, 12, 18, 29). Although steady-state mRNA levels between our constructs may vary slightly, the close correlation observed between *in vitro* ribosome binding and *in vivo* expression, as well as primer extension confirmation of many of the predicted transcriptional start sites, allows us to conclude that ribosome recognition and binding to an AUG can occur when an AUG is present within six nucleotides of an mRNA's 5' terminus, with most efficient binding when the AUG is at the 5' terminus. Furthermore, the efficiency with which the AUG of a leaderless mRNA is recognized and bound by ribosomes relates to its proximity to the 5' terminus and the sequence of any upstream nucleotides; additional effects could relate to mRNA secondary structures that influence AUG availability for ribosome binding. The variable effect of upstream nucleotides on ribosome binding and expression could relate to ribosomal E-site preferences (22), the increased stability of a 4-bp codon-anticodon interaction (4), or the ability of some nucleotides to interfere with ribosome recognition of the start codon. Our demonstration that ribosomes continue to recognize a start codon after the addition of nucleotides to either a naturally leaderless mRNA (*cI*) or to mRNAs engineered to be leaderless (*lacZ*, *uidA*) suggests that a 5' AUG is a distinct translational signal recognized by the ribosome and argues that the *E. coli* transcriptome should be examined for mRNAs containing 5' AUGs that could function in ribosome binding and/or expression.

This research was supported by National Institutes of Health grant GM065120 to G.R.J.

REFERENCES

- Balakin, A. G., E. A. Skripkin, I. N. Shatsky, and A. A. Bogdanov. 1992. Unusual ribosome binding properties of mRNA encoding bacteriophage lambda repressor. *Nucleic Acids Res.* **20**:563–571.
- Brock, J. E., R. L. Paz, P. Cottle, and G. R. Janssen. 2007. Naturally occurring adenines within mRNA coding sequences affect ribosome binding and expression in *Escherichia coli*. *J. Bacteriol.* **189**:501–510.
- Brock, J. E., S. Pourshahian, J. Giliberti, P. A. Limbach, and G. R. Janssen. 2008. Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5'-terminal AUG. *RNA* **14**:2159–2169.
- Esposito, D., J. P. Fey, S. Eberhard, A. J. Hicks, and D. B. Stern. 2003. In vivo evidence for the prokaryotic model of extended codon-anticodon interaction in translation initiation. *EMBO J.* **22**:651–656.
- Hering, O., M. Brenneis, J. Beer, B. Suess, and J. Soppa. 2009. A novel mechanism for translation initiation operates in haloarchaea. *Mol. Microbiol.* **71**:1451–1463.
- Horinouchi, S., K. Furuya, M. Nishiyama, H. Suzuki, and T. Beppu. 1987. Nucleotide sequence of the streptothricin acetyltransferase gene from *Streptomyces lavendulae* and its expression in heterologous hosts. *J. Bacteriol.* **169**:1929–1937.
- Hui, A., J. Hayflick, K. Dinkelspiel, and H. A. de Boer. 1984. Mutagenesis of the three bases preceding the start codon of the beta-galactosidase mRNA and its effect on translation in *Escherichia coli*. *EMBO J.* **3**:623–629.
- Janssen, G. R. 1993. Eubacterial, archaeobacterial and eukaryotic genes that encode leaderless mRNA, p. 59–67. In R. H. Baltz, G. D. Hegeman, and P. L. Scatrud (ed.), *Industrial microorganisms: basic and applied molecular genetics*. ASM Press, Washington, DC.
- Janssen, G. R., J. M. Ward, and M. J. Bibb. 1989. Unusual transcriptional and translational features of the aminoglycoside phosphotransferase gene (*aph*) from *Streptomyces fradiae*. *Genes Dev.* **3**:415–429.
- Jefferson, R. A. 1989. The GUS reporter gene system. *Nature* **342**:837–838.
- Looman, A. C., J. Bodlaender, L. J. Comstock, D. Eaton, P. Jhurani, H. A. de Boer, and P. H. van Knippenberg. 1987. Influence of the codon following the AUG initiation codon on the expression of a modified *lacZ* gene in *Escherichia coli*. *EMBO J.* **6**:2489–2492.

12. Marquez, V., D. N. Wilson, W. P. Tate, F. Triana-Alonso, and K. H. Nierhaus. 2004. Maintaining the ribosomal reading frame: the influence of the E site during translational regulation of release factor 2. *Cell* **118**:45–55.
13. Martin-Farmer, J., and G. R. Janssen. 1999. A downstream CA repeat sequence increases translation from leadered and unleadered mRNA in *Escherichia coli*. *Mol. Microbiol.* **31**:1025–1038.
14. Miller, J. 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
15. Moll, I., S. Grill, C. O. Gualerzi, and U. Blasi. 2002. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Microbiol.* **43**:239–246.
16. Moll, I., G. Hirokawa, M. C. Kiel, A. Kaji, and U. Blasi. 2004. Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res.* **32**:3354–3363.
17. O'Donnell, S. M., and G. R. Janssen. 2002. Leaderless mRNAs bind 70S ribosomes more strongly than 30S ribosomal subunits in *Escherichia coli*. *J. Bacteriol.* **184**:6730–6733.
18. O'Donnell, S. M., and G. R. Janssen. 2001. The initiation codon affects ribosome binding and translational efficiency in *Escherichia coli* of cI mRNA with or without the 5' untranslated leader. *J. Bacteriol.* **183**:1277–1283.
19. Ptashne, M., K. Backman, M. Z. Humayun, A. Jeffrey, R. Maurer, B. Meyer, and R. T. Sauer. 1976. Autoregulation and function of a repressor in bacteriophage lambda. *Science* **194**:156–161.
20. Salmon, K., S. P. Hung, K. Mekjian, P. Baldi, G. W. Hatfield, and R. P. Gunsalus. 2003. Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J. Biol. Chem.* **278**:29837–29855.
21. Schleif, R. 1972. Fine-structure deletion map of the *Escherichia coli* L-arabinose operon. *Proc. Natl. Acad. Sci. U. S. A.* **69**:3479–3484.
22. Schuwirth, B. S., J. M. Day, C. W. Hau, G. R. Janssen, A. E. Dahlberg, J. H. Cate, and A. Vila-Sanjurjo. 2006. Structural analysis of kasugamycin inhibition of translation. *Nat. Struct. Mol. Biol.* **13**:879–886.
23. Sharma, C. M., S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiss, A. Sittka, S. Chabas, K. Reiche, J. Hackermuller, R. Reinhardt, P. F. Stadler, and J. Vogel. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**:250–255.
24. Shean, C. S., and M. E. Gottesman. 1992. Translation of the prophage lambda cl transcript. *Cell* **70**:513–522.
25. Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U. S. A.* **71**:1342–1346.
26. Skorski, P., P. Leroy, O. Fayet, M. Dreyfus, and S. Hermann-Le Denmat. 2006. The highly efficient translation initiation region from the *Escherichia coli* rpsA gene lacks a Shine-Dalgarno element. *J. Bacteriol.* **188**:6277–6285.
27. Tanaka, Y., A. Tsujimura, N. Fujita, S. Isono, and K. Isono. 1989. Cloning and analysis of an *Escherichia coli* operon containing the rpmF gene for ribosomal protein L32 and the gene for a 30-kilodalton protein. *J. Bacteriol.* **171**:5707–5712.
28. Udagawa, T., Y. Shimizu, and T. Ueda. 2004. Evidence for the translation initiation of leaderless mRNAs by the intact 70 S ribosome without its dissociation into subunits in eubacteria. *J. Biol. Chem.* **279**:8539–8546.
29. Van Etten, W. J., and G. R. Janssen. 1998. An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in *Escherichia coli*. *Mol. Microbiol.* **27**:987–1001.
30. Wade, H. E., and H. K. Robinson. 1966. Magnesium ion-independent ribonucleic acid depolymerases in bacteria. *Biochem. J.* **101**:467–479.