Gene Expression, Vol. 10, pp. 109–114
Printed in the USA. All rights reserved.
1052-2166/02 $20.00 + .00
Copyright © 2002 Cognizant Comm. Corp.
www.cognizantcommunication.com

Genetic Analysis of the Basis of Translation in the −1 Frame of an Unusual Non-ORF Sequence Isolated From Phage Display

JENNIFER ZEMSKY, WLODEK MANDECKI,[1] AND EMANUEL GOLDMAN[2]

Department of Microbiology & Molecular Genetics, New Jersey Medical School, University of Medicine & Dentistry of New Jersey, 185 South Orange Avenue, Newark, NJ 07103

An unusual peptide-encoding sequence, called H10, and several derivatives of this sequence were previously isolated from a random peptide library screened by phage display during drug discovery protocols. The H10 family of sequences had the unusual property of being expressed despite the absence of an open reading frame. When these sequences were fused to a reporter *lacZ* gene in all three frames, β-galactosidase was expressed not only from the parental non-open reading frame, consistent with the original isolations, but also from the frame −1 to the parental. This unexpected translation in a second reading frame could result from either a recoding event or from an internal translation initiation event. In order to elucidate which type of event, a genetic approach was selected to eliminate a potential downstream initiator site within the H10 sequence. This report provides strong evidence that translation in the −1 frame in this family of sequences is indeed originating from a downstream translation initiation event. Unexpectedly, the mutation eliminating the downstream initiation event in the −1 frame simultaneously elevated expression in the original non-open reading frame.

E. coli protein synthesis Recoding Readthrough of UGA codons
Programmed translational frameshifts Internal initiation or reinitiation

DURING the process of translation, specific elements in the mRNA signal form an mRNA–ribosome complex and result in translation initiation at the initiator methionine codon. Translation continues by sequentially reading each codon until a stop codon is reached, resulting in termination of translation. Deviations from this scheme (e.g., misreading a codon, frameshifting, or ignoring a stop codon) are taken to be errors in translation [reviewed in (14)]. However, an increasing number of examples of programmed events have been discerned where these deviations are necessary for synthesis of the final protein and are "recoded" for in the mRNA structure [reviewed in (8)].

A number of sequences that appear to use translational recoding in the synthesis of their proteins were isolated during a drug discovery search (4,15). A random peptide library had been synthesized and used to create a phage display library. The peptides generated by these sequences were tested against several hormone receptors in the hope of finding pharmacologically useful peptides. Surprisingly, 50% of the identified sequences required some form of non-open reading frame (non-ORF) event to generate the phage-displayed peptide.

One sequence, designated H10, selected in the panning experiments against the human growth hormone binding protein, was chosen for further study. The H10 sequence was inserted into a β-galactosidase expression vector in all three frames to measure the frequency of the recoding event (10). In addition, a family of mutant sequences of the parental H10 was also tested to further assess the nature of the event.

Accepted November 20, 2001.
[1]Current address: PharmaSeq, Inc., 1 Deer Park Drive, Suite F, Monmouth Junction, NJ 08852.
[2]Address correspondence to Emanuel Goldman, Department of Microbiology & Molecular Genetics, New Jersey Medical School, University of Medicine & Dentistry of New Jersey, 185 South Orange Avenue, Newark, NJ 07103. Tel: (973) 972-4367; Fax: (973) 972-3644; E-mail: egoldman @umdnj.edu

Consistent with the original isolations, the non-ORF originally identified from the phage display library yielded β-galactosidase expression. However, unexpectedly, the frame −1 to the non-ORF frame also yielded significant β-galactosidase expression. As a result, expression in both the original non-ORF and the newly identified frame −1 to the original frame needed to be accounted for. It was possible that expression in both frames was due to recoding, but examination of the sequence in the frame −1 to the original non-ORF frame suggested that a new translation initiation event in this frame could account for the synthesized product. This frame contained an ATG codon without any stop codons prior to the junction to the reporter, and also possessed a consensus Shine-Dalgarno sequence (GGAGG) appropriately distanced just upstream of this ATG. There is precedent for internal initiation or reinitiation during translation [e.g., (5,12,13,18)]. In this report, we demonstrate that mutating the putative start codon in the frame −1 to the original non-ORF frame does in fact eliminate translation from that frame. Interestingly, this mutation simultaneously elevates expression of the original non-ORF frame as well.

MATERIALS AND METHODS

Mutation of Residue 74G to 74T in Plasmid H10

Residue 74G was mutated to 74T in a two-step PCR procedure based upon a protocol described by Higuchi (11). In the first step, two overlapping PCR fragments were generated that contained the G to T mutation at residue 74. These PCR products were gel purified using the Qiagen Qiaex II Gel Extraction Kit. The products were combined and heated to 95°C for 10 min and allowed to slowly cool to room temperature. The annealed products were then used as the templates for generating a full-length copy of the mutated H10 sequence attached to the reporter in both the 0 and −1 frames, respectively, using outer PCR primers. PCR oligonucleotide primers were generated at the New Jersey Medical School Molecular Resource Facility.

Cloning of PCR Products to Generate Plasmids With Reporter Attached in 0 or −1 Frames

The PCR products included the *Hin*dIII and *Bam*HI restriction endonuclease sites present in the parental plasmids. PCR products and plasmid vector pJC27 were digested with *Hin*dIII restriction endonuclease at 37°C for 2 h. The pJC27 vector was treated with calf intestinal alkaline phosphatase (CIAP) at 50°C for 5 min. The products were phenol/chloroform extracted and ethanol precipitated, and then ligated using the Rapid DNA Ligation Kit from Boehringer Mannheim for 5 min at room temperature. The ligation products were again phenol/chloroform extracted and ethanol precipitated, and then digested with *Bam*HI restriction endonuclease for 2 h at 37°C. Products were again extracted, precipitated, and ligated as above. Ligation products were transformed into *Escherichia coli* strain MY411 (Δ[lac-pro], supE, thi/ F′lacI^{Q} ZΔM15, proA^{+}B^{+}) (6). Miniprep DNA was purified from resultant colonies using the Qiagen Spin Miniprep Kit. Plasmids giving appropriately sized inserts were then purified from 100-ml cultures using the Qiagen Plasmid Midi Kit and submitted to the NJ Medical School Molecular Resource Facility for DNA sequencing. Restriction endonucleases were purchased from Gibco BRL.

β-Galactosidase Assays

Cell growth and assay procedures were based on standard protocols (17). Five milliliters of 1× medium A supplemented with 0.001 M $MgSO_4 \cdot 7H_2O$, 0.2% dextrose, 0.00005% vitamin B1, and 20 μg/ml chloramphenicol (1) were inoculated 1:50 from overnight LB cultures containing 20 μg/ml chloramphenicol. Cultures were grown to an optical density of approximately 0.4 at 600 nm. Fusion protein expression was induced by adding 1 mM IPTG for 1 h. A_{600} measurements were taken and 100 μl of culture was added to 900 μl of Z buffer (0.06 M $Na_2HPO_4 \cdot 7H_2O$, 0.04 M $NaH_2PO_4 \cdot H_2O$, 0.01 M KCl, 0.001 M $MgSO_4 \cdot 7H_2O$, 0.05 M β-mercaptoethanol). Two drops of chloroform and one drop of 0.1% SDS were added to the reaction mixture, which was vortexed for 10 s. The reaction mixtures were then allowed to equilibrate to 28°C. *o*-Nitrophenyl-β-D-galactopyranoside (0.2 ml) in water (4 mg/ml) was added to the reaction mixtures at 1-min intervals. When a yellow color developed, the reaction was stopped by the addition of 0.5 ml of 1 M Na_2CO_3. A_{550} and A_{420} measurements were taken. β-Galactosidase activity was calculated based upon the following formula: β-galactosidase activity = $1000 \times (A_{420} - 1.75 \times A_{550})/t \times v \times A_{600}$, where t is equal to the time in minutes, and v is the milliliter culture volume used in the assay (0.1 ml in these experiments).

RESULTS AND DISCUSSION

The H10 sequence had previously been cloned into the pJC27 vector, which contains a reporter *lacZ* gene (10). The sequence was fused in parallel constructions such that each reading frame in H10 made an in-frame fusion with the *lacZ* reporter. The 0 frame fusion was designated with the suffix .1, the +1 frame

fusion was designated with the suffix .2, and the −1 frame fusion was designated with the suffix .3. Expression of β-galactosidase was observed in both the 0 and −1 frame constructs. The difference between the 0 frame clones and the −1 frame clones is that the −1 frame clones have a single base pair deletion in the junction sequence that connects the *lacZ* reporter to the −1 reading frame (Fig. 1). Expression in the 0 frame appears to require either a readthrough or frameshift event because of the presence of two in-frame UGA termination codons. Expression from the −1 frame either requires a frameshift from the 0 to the −1 frame or some other novel event. Examination of the sequence in the −1 frame indicates an ATG codon in that frame with a canonical Shine-Dalgarno sequence spaced appropriately upstream, suggesting a potential translation initiator site (10) (Fig. 1). We undertook a genetic approach to investigate if expression in the −1 frame was due to an initiation event at this downstream ATG.

Mutation of Residue 74G to 74T

In order to determine if protein synthesis initiation was occurring at the internal ATG in the −1 frame of the H10 sequence, we decided to mutate this ATG. The A residue is the last code letter of the 0 frame TGA, and therefore could not be mutated for this purpose. Of the remaining residues, the G residue was chosen for mutagenesis. The G to T mutation was deemed to make the most conservative change in the 0 frame while eliminating the potential initiator codon.

A two-step PCR mutagenesis protocol was employed (see Materials and Methods). The resulting PCR product was cloned back into the parental pJC27 plasmid. Products were sequenced, verifying the 74G to T mutations, and that no extraneous mutations had been introduced.

Analysis of the β-Galactosidase Expression of Mutant Versus Parental Plasmids

As in the parental clones, the 0 and −1 frames of H10-74T (H10-74T.1 and H10-74T.3, respectively) generated sequences that expressed β-galactosidase fusion proteins. The 0 and −1 frames of clones H10 (H10.1 and H10.3, respectively), H10-74T, and control vector pJC27 (pJC27.1 and pJC27.3, respectively) were assayed for the β-galactosidase activity of their fusion proteins.

Plasmids were transformed into *Escherichia coli*

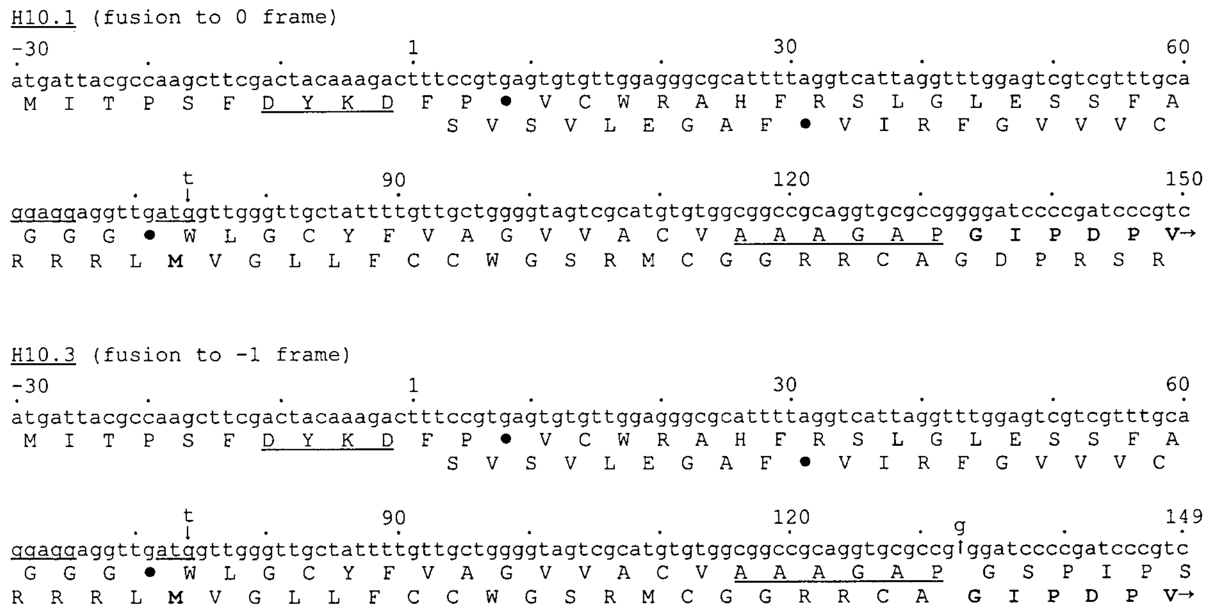

Figure 1. DNA and translation sequences of H10 fused to the *lacZ* gene in the 0 and −1 frames. The nucleotide sequence, with numbering, of parental H10 clone (lower case) is shown in the upper lines of each construct, followed below by the translation sequences (upper case) in the 0 and −1 frames. Stop codons are indicated by filled circles. The H10.3 sequence has a single base pair deletion relative to H10.1 at the *lacZ* junction site. This deletion is designated by an upward arrow in the sequence with the deleted G residue above the arrow. The putative Shine-Dalgarno sequence (GGAGG) and the putative downstream initiator ATG are underlined in the nucleotide sequence. Residue 74G, mutated to T in this study, is indicated by a downward arrow. Following the mutation, the 0 frame codon becomes TTG (Leu) and the −1 frame codon becomes ATT (Ile). The frame in which *lacZ* continues is designated by a rightward arrow. The control vectors, pJC27.1 and pJC27.3, are similar to the test sequences but lack the H10 sequence; they are fused from the DYKD- to the AAAGAP-encoding sequences (underlined).

strain MY411, which has a deletion in the *lacZ* gene (6); therefore, enzyme measurements reflect expression from the plasmid-borne *lacZ* reporter gene. β-Galactosidase assays were performed as described in Sambrook et al. (17). The two mutant clones of H10-74T were compared with the parental clones, control plasmids, and the host strain MY411. In these assays, expression from the in-frame control was approximately twofold higher than previously reported (10). Nevertheless, expression of the non-ORF test vectors remained highly significant relative to the in-frame control (Table 1).

The H10-74T mutant in the −1 frame expresses 40-fold less β-galactosidase activity than the H10 parent in this same frame (compare lines 5 and 7 of Table 1). The expression from the H10-74T mutant in the −1 frame is indistinguishable from the −1 frame control of pJC27 (lines 2 and 7 of Table 1), strongly indicating that mutation of the ATG codon eliminates expression in this frame. This implies that expression in the −1 frame is indeed the result of a second downstream initiation event and not the result of a frameshift event. Other variants of the H10 sequence were also examined in the previous study (10). Similarly to the parental H10 series, expression had been observed in both the frame isolated from phage display and the frame −1 to that. While the β-galactosidase activity was not identical in all these mutants, β-galactosidase was significantly produced. The sequences of these mutant derivatives are identical to the parental H10 in the region surrounding the internal ATG and Shine-Dalgarno sequences, which argues strongly that expression of these clones in the frame −1 to the phage display frame is also the result of a downstream initiation event.

Interestingly, the loss of expression in the −1 frame resulted in a 2.4-fold increase in expression in the non-ORF 0 frame (compare lines 4 and 6 of Table 1). The increased expression may result from an increase in transit of continuing ribosomes, which could have been inhibited by some sort of competition from the downstream initiation of other ribosomes in the parental clones. Similarly, the reason that expression in the −1 frame of H10 isn't equivalent to an in-frame control may be because of competition from elongating ribosomes that had initiated at the beginning of the sequence, in the 0 frame (3). A graphical representation of our results is presented in Figure 2, illustrating how the 74G to T mutation alters expression of the reporter in the two reading frames. For the sake of comparison, Figure 2 also shows that the third reading frame (+1 frame) is inactive for expression of reporter in parental H10, taken from measurements in Goldman et al. (10).

An alternative explanation for the increase in expression in the 0 frame could be that the in-frame UUG codon generated by the 74T mutation, immediately downstream of the UGA codon (Fig. 1), is functioning as a translation initiation signal in the 0 frame. The very next downstream codon is also UUG; thus, the increased target of two adjacent in-frame UUG codons might be more effective in attracting ribosomes for translation starts. Appropriately spaced, overlapping Shine-Dalgarno sequences are upstream of these UUG codons. However, the simple presence of a Shine-Dalgarno signal upstream of a potential start codon is not sufficient to ensure a translation start (7). Also, UUG is used for translation initiation in *E. coli* only about 1% of the time (9), and is about an order of magnitude less efficient than AUG (2,16). Nevertheless, this explanation could account for the result of increased expression in the 0 frame following the 74T mutation. In the case of parental H10 and other variants of this sequence, our

TABLE 1
EXPRESSION OF SITE-DIRECTED MUTANTS AT THE PUTATIVE ATG START IN THE −1 FRAME RELATIVE TO THE ORIGINAL CLONES

	Plasmid	β-Galactosidase Reading Frame	UGAs in 0 Frame	β-Galactosidase Units (±SD)	% In-Frame Control	Normalized % Total*
1	pJC27.1	0	0	8018 (±553)	100	99
2	pJC27.3	−1		70 (±6)	1	1
3	none (MY411)	(deletion strain)		26 (±15)†		
4	H10.1	0	2	685 (±13)	9	21
5	H10.3	−1		2472 (±306)	31	79
6	H10-74T.1	0	2	1678 (±152)	21	98
7	H10-74T.3	−1		61 (±25)	1	2

Cells were grown to A_{600} approximately equal to 0.4 in medium A, induced with 1 mM IPTG for 1 h, and β-galactosidase activity was measured as described in Sambrook et al. (17).

*Background (no plasmid) subtracted. The normalized % total represents the fraction in each frame relative to the sum of expression from both frames.

†$n = 6$; for all others, $n = 4$.

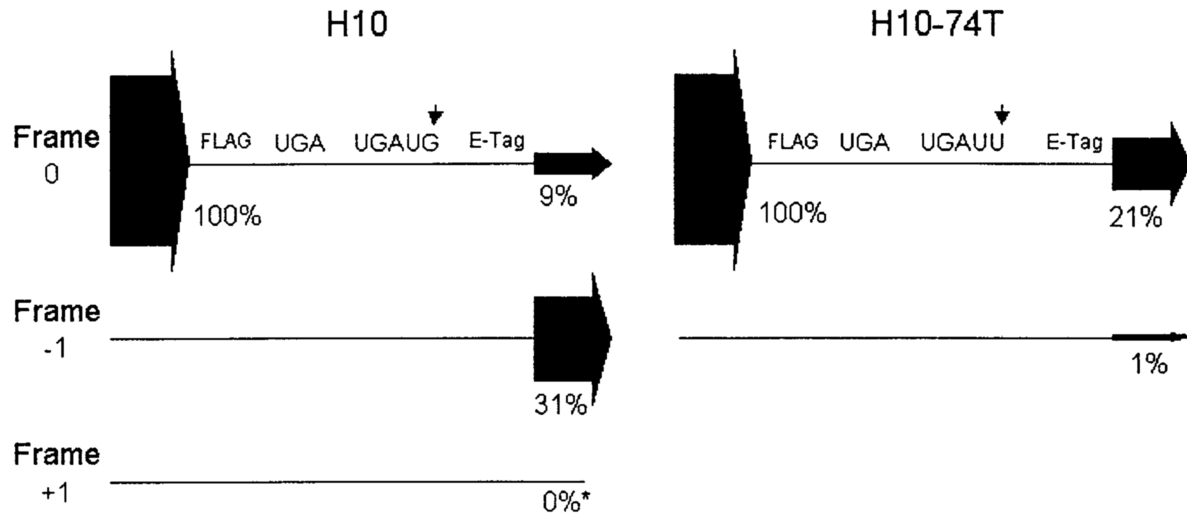

Figure 2. Graphical representation of effects of 74T mutation on reporter gene expression. The percentages are taken from the "% In-Frame Control" column in Table 1, with the exception of the value for the +1 frame under H10 (0%*), which was taken from Goldman et al. (10). "FLAG" and "E-Tag" denote portions of the amino acid epitopes (DYKD for FLAG and AAAGAP for E-Tag) that were fused at the amino- and carboxy-terminal boundaries of the H10 series sequences in the original isolations (4). The two 0 frame UGA codons in H10 are shown, as well as the overlapping −1 frame AUG in H10, mutated to AUU in H10-74T, indicated by down arrows.

previous results with site-directed mutants of the first UGA codon (10) are inconsistent with downstream initiation from the UUG codon at positions 76–78.

It is conceivable that our results are due to some other mechanism than the hypothesized inactivation of translation initiation from the ATG sequence. Such a mechanism could involve a change of the pattern of frameshifting as a result of the mutation. For example, it is possible that the 74G to T mutation enhances a recoding event required for expression in the 0 frame. Nevertheless, the experiments reported here provide strong evidence for downstream translation initiation in the H10 sequence and its derivatives, in the −1 frame relative to the expressed non-ORF frame. Because our earlier work had demonstrated translation in two of the three reading frames, this result simplifies the problem by explaining one of these frames. The basis for the non-ORF expression in the 0 reading frame, however, remains to be elucidated.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation (MCB-9981879). Thanks are given to Dr. Hieronim Jakubowski for constructive conversations. Dr. Robert Donnelly and the New Jersey Medical School Molecular Resource Facility are acknowledged for oligonucleotide synthesis and DNA sequencing.

REFERENCES

1. Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K. Current protocols in molecular biology. New York: John Wiley & Sons; 1987.
2. Barrick, D.; Villanueba, K.; Childs, J.; Kalil, R.; Schneider, T. D.; Lawrence, C. E.; Gold, L.; Stormo, G. D. Quantitative analysis of ribosome binding sites in *E. coli*. Nucleic Acids Res. 22:1287–1295; 1994.
3. Berkhout, B.; Kastelein, R. A.; van Duin, J. Translational interference at overlapping reading frames in prokaryotic messenger RNA. Gene 37:171–179; 1985.
4. Carcamo, J.; Ravera, M. W.; Brissette, R.; Dedova, O.; Beasley, J. R.; Alam-Moghé, A.; Wan, C.; Blume, A.; Mandecki, W. Unexpected frameshifts from gene to expressed protein in a phage-displayed peptide library. Proc. Natl. Acad. Sci. USA 95:11146–11151; 1998.
5. Cone, K. C.; Steege, D. A. Functional analysis of lac repressor restart sites in translational initiation and reinitiation. J. Mol. Biol. 186:733–742; 1985.
6. Curran, J. F.; Yarus, M. Use of tRNA suppressors to probe regulation of *Escherichia coli* release factor 2. J. Mol. Biol. 203:75–83; 1988.
7. Dreyfus, M. What constitutes the signal for the initiation of protein synthesis on *Escherichia coli* mRNAs? J. Mol. Biol. 204:79–94; 1988.
8. Gesteland, R. F.; Atkins, J. F. Recoding: Dynamic reprogramming of translation. Annu. Rev. Biochem. 65: 741–768; 1996.
9. Gold, L.; Stormo, G. Translation initiation. In: Neid-

hardt, F. C.; Ingraham, J. L.; Low, K. B.; Magasanik, B.; Schaechter, M.; Umbarger, E., eds. *Escherichia coli* and *Salmonella typhimurium*. Cellular and molecular biology. Washington, DC: American Society for Microbiology; 1987:1302–1307.
10. Goldman, E.; Korus, M.; Mandecki, W. Efficiencies of translation in three reading frames of unusual non-ORF sequences isolated from phage display. FASEB J. 14: 603–611; 2000.
11. Higuchi, R. Recombinant PCR. In: Innis, M. A.; Gelfand, D. H.; Sninsky, J. J.; White, T. J., eds. PCR protocols: A guide to methods and applications. San Diego: Academic Press; 1990:177–183.
12. Manley, J. L. Synthesis of internal re-initiation fragments of beta-galactosidase *in vitro* and *in vivo*. J. Mol. Biol. 125:449–466; 1978.
13. Napoli, C.; Gold, L.; Singer, B. S. Translational reinitiation in the rIIB cistron of bacteriophage T4. J. Mol. Biol. 149:433–449; 1981.
14. Parker, J. Errors and alternatives in reading the universal genetic code. Microbiol. Rev. 53:273–298; 1989.
15. Ravera, M. W.; Carcamo, J.; Brissette, R.; Alam-Moghé, A.; Dedova, O.; Cheng, W.; Hsiao, K. C.; Klebanov, D.; Shen, H.; Tang, P.; Blume, A.; Mandecki, W. Identification of an allosteric binding site on the transcription factor p53 using a phage-displayed peptide library. Oncogene 16:1993–1999; 1998.
16. Ringquist, S.; Shinedling, S.; Barrick, D.; Green, L.; Binkley, J.; Stormo, G. D.; Gold, L. Translation initiation in *E. coli*: Sequences within the ribosome-binding site. Mol. Microbiol. 6:1219–1229; 1992.
17. Sambrook, J.; Fritsch, E. F.; Maniatis, T. Molecular cloning: A laboratory manual (2nd ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989:17–35.
18. Spanjaard, R. A.; van Duin, J. Translational reinitiation in the presence and absence of a Shine and Dalgarno sequence. Nucleic Acids Res. 17:5501–5507; 1989.