SHORT COMMUNICATION

Translation initiation from sequence variants of the bacteriophage T7 g10RBS in *Escherichia coli* and *Agrobacterium fabrum*

Alex B. Benedict[1] · Joshua D. Chamberlain[1] · Diana G. Calvopina[1] · Joel S. Griffitts[1]

Abstract

Background The bacteriophage T7 gene *10* ribosome binding site (g10RBS) has long been used for robust expression of recombinant proteins in *Escherichia coli*. This RBS consists of a Shine–Dalgarno (SD) sequence augmented by an upstream translational "enhancer" (Enh) element, supporting protein production at many times the level seen with simple synthetic SD-containing sequences. The objective of this study was to dissect the g10RBS to identify simpler derivatives that exhibit much of the original translation efficiency.

Methods and results Twenty derivatives of g10RBS were tested using multiple promoter/reporter gene contexts. We have identified one derivative (which we call "CON_G") that maintains 100% activity in *E. coli* and is 33% shorter. Further minimization of CON_G results in variants that lose only modest amounts of activity. Certain nucleotide substitutions in the spacer region between the SD sequence and initiation codon show strong decreases in translation. When testing these 20 derivatives in the alphaproteobacterium *Agrobacterium fabrum*, most supported strong reporter protein expression that was not dependent on the Enh.

Conclusions The g10RBS derivatives tested in this study display a range of observed activity, including a minimized version (CON_G) that retains 100% activity in *E. coli* while being 33% shorter. This high activity is evident in two different promoter/reporter sequence contexts. The array of RBS sequences presented here may be useful to researchers in need of fine-tuned expression of recombinant proteins of interest.

Keywords Recombinant protein expression · Ribosome binding site · Translation

Introduction

Recombinant protein expression in *Escherichia coli* is often optimized by engineering two control points: transcription (promoter optimization) and translation (ribosome binding site, or RBS, optimization) [1, 2]. The RBS from the bacteriophage T7 gene *10* (g10RBS) has been used extensively in commercially available plasmids to stimulate recombinant protein expression in *E. coli*. It was previously shown that when g10RBS is placed upstream of recombinant genes, protein expression is increased 40-fold or more than when a synthetic consensus RBS sequence is used [3]. Within this 45-nucleotide (nt) sequence, an A/U-rich 9-nt enhancer element (Enh) upstream of the Shine–Dalgarno (SD) region is

found to stimulate translation, even when its position relative to the SD region is changed [4]. Several studies have examined the effects of this Enh element on translation rates [3–10]. Homology between the Enh and nucleotides 458–466 in the 16S rRNA led to an initial proposal that base-pairing underlies the increased translation rates [4, 8]. Modifications to nucleotides 458–466 of the 16S rRNA, however, did not support that hypothesis [6]. Subsequent proposals suggest, instead, that Enh or similar elements interact with ribosomal protein S1 [11–13]. Beyond *E. coli*, the g10RBS has been shown to support high expression of recombinant proteins in other gammaproteobacteria such as *Pseudomonas*, *Erwinia*, and *Serratia* [14]. The objective of the present study is to more thoroughly dissect this phage-derived RBS to identify shorter derivatives that retain translational activity in *E. coli*, and to test these derivatives in the more distantly related alphaproteobacterial species, *Agrobacterium fabrum*.

✉ Joel S. Griffitts
  joelg@byu.edu

[1] Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT 84602, USA

Results and discussion

The full-length g10RBS used in commercial vectors to enhance translation is referred to in this study as "FL_A," which includes (from the 5' end) an XbaI site, an 11-nt A/U-rich sequence, the 9-nt Enh element, a 4-nt potential "standby" site, the SD sequence (GGAGAT), and the SD-initiation codon spacer. The g10RBS and its potential to hybridize with the *E. coli* 16S rRNA are depicted in Fig. 1a. The anti-Enh region within the 16S sequence is not found in the alphaproteobacterial rRNAs from *A. fabrum* and *Sinorhizobium meliloti* (Fig. 1b).

A total of 21 RBS sequences, including FL_A, were ligated individually into a test plasmid that replicates in both *E. coli* and *A. fabrum*, and encodes the mScarlet-I fluorescent protein [15], driven by a synthetic *lac*T5 promoter. Because the plasmid does not encode the LacI repressor, the *lac*T5 promoter behaves constitutively in these strains. Fluorescent output for each RBS was measured in *E. coli* and *A. fabrum* (Fig. 2) and are presented as normalized values with the FL_A measurement in *E. coli* calibrated to "100". All other fluorescence measurements presented are normalized to that standard.

The SD sequence within FL_A (GGAGAT) does not match perfectly with the anti-SD at the 3' end of 16S in *E. coli* or *A. fabrum*, so the A nucleotide was changed to G to make a perfect match (GGAGGT). We refer to this derivative as "FL_G." For FL_G and FL_A, a series of three 5' truncations were made. Next, the 11-nt A/U-rich
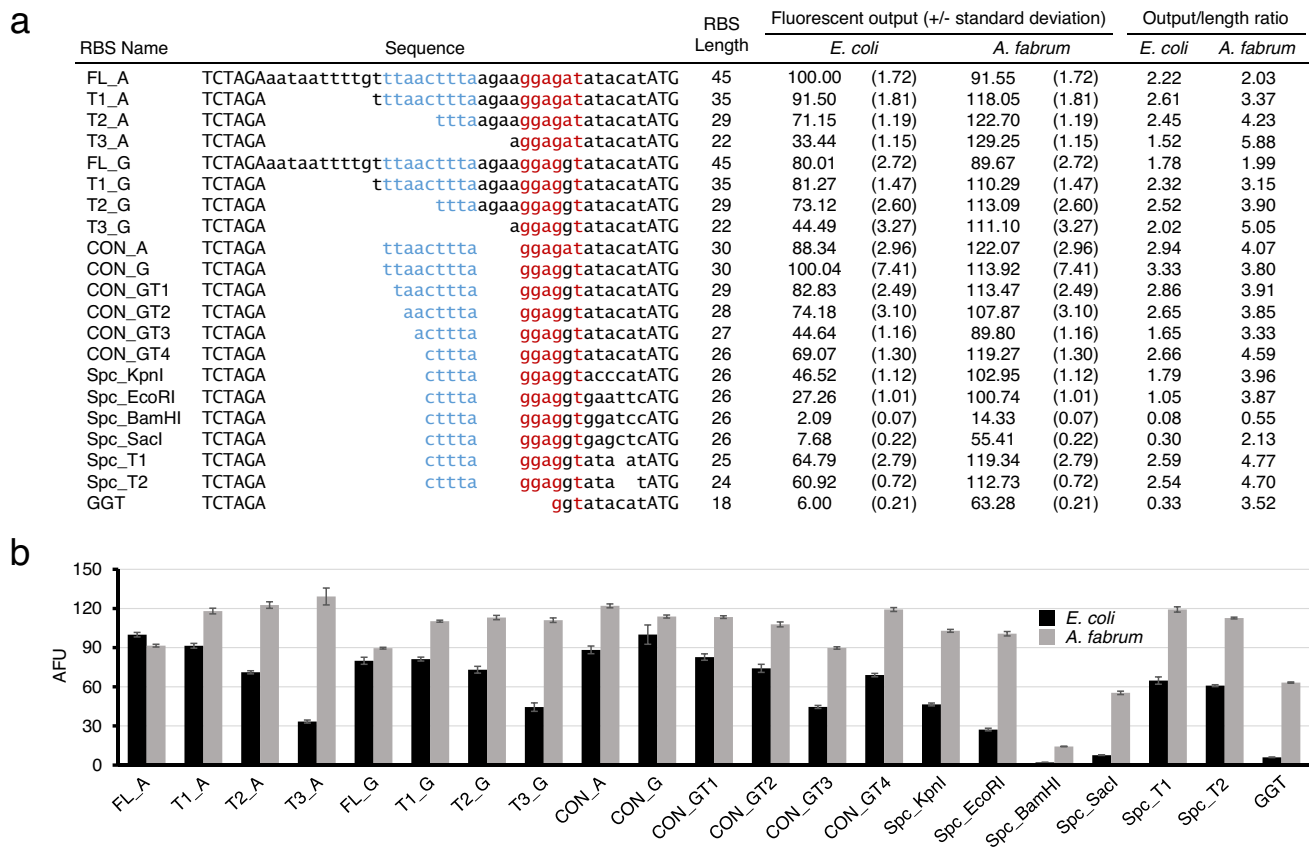
sequence upstream of Enh and the 4-nt potential standby site were simultaneously removed from FL_A and FL_G to form a condensed RBS. These are called "CON_A" and "CON_G", respectively. Then, four single-base truncations were made from the 5' end of Enh in the FL_G derivative. Finally, in the derivative with 4 bases truncated from the 5' end of Enh (termed "CON_GT4"), we made several modifications to the spacer between the SD and the initiation codon by either removing bases or replacing them to incorporate various restriction sites (see Fig. 2).

In *E. coli*, the A-to-G base substitution in the SD sequence (FL_A vs FL_G) led to a 20% drop in expression. Each of the three truncations within the FL_A sequence ("T1_A", "T2_A", and "T3_A") led to 9%, 29%, and 67% reductions in expression, respectively. In contrast, the same three truncations within the FL_G sequence ("T1_G", "T2_G", and "T3_G") led to a 2% increase, and 9% and 44% decreases, respectively. When comparing CON_A and CON_G to FL_A, we observed, surprisingly, that the CON_G derivative yielded expression levels nearly identical to the parental FL_A sequence despite being 33% (15 nt) shorter, while expression from the CON_A derivative was 12% lower than FL_A. This observation is consistent with data from two high-throughput studies on the 5' untranslated regions of mRNAs in *E. coli*, where it was determined that there is a preference for a G nucleotide eight bases upstream of the initiation codon [16, 17]. For this reason, subsequent RBS modifications were made to the highly efficient CON_G sequence. Four single-base deletions from the 5' end of the enhancer element in CON_G ("CON_GT1", "CON_GT2",
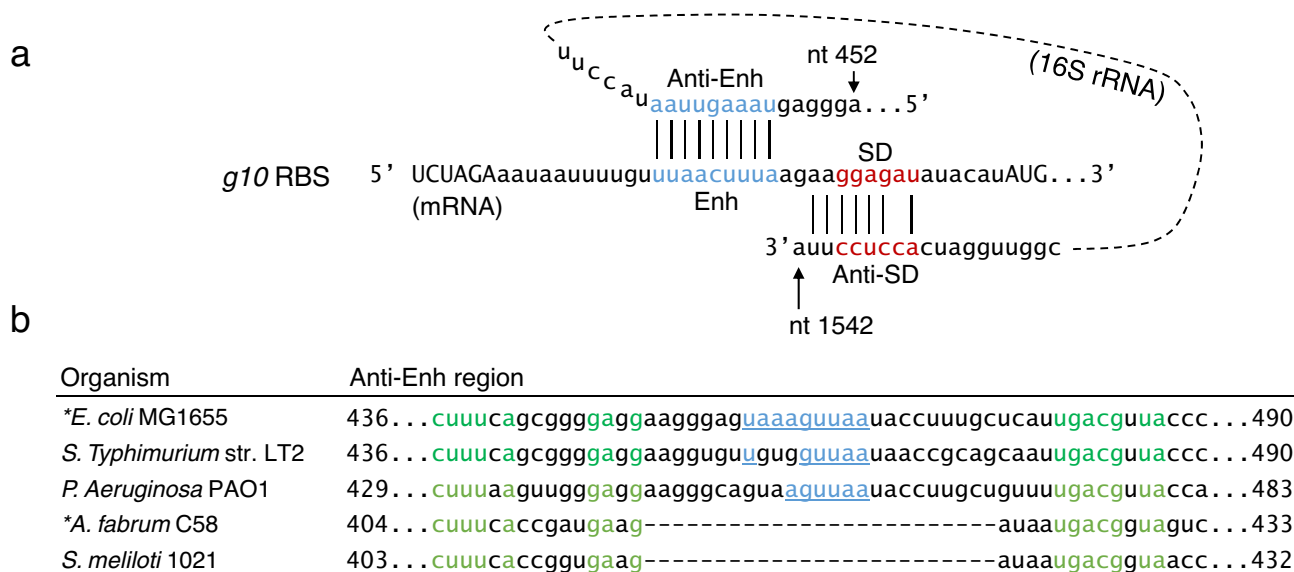
Fig. 1 Homology between g10RBS and two positions within the 16S rRNA. **a** Blue and red coloring indicates the Enh/Anti-Enh and SD/Anti-SD sequences, respectively. **b** Alignment of the anti-Enh region in several species. Green coloring indicates residues conserved in all species and blue, underlined, text indicates homology with the Enh. (Color figure online)

a

RBS Name	Sequence	RBS Length	Fluorescent output (+/- standard deviation)				Output/length ratio	
			E. coli		A. fabrum		E. coli	A. fabrum
FL_A	TCTAGAaataattttgtttaactttaagaaggagatatacatATG	45	100.00	(1.72)	91.55	(1.72)	2.22	2.03
T1_A	TCTAGA tttaactttaagaaggagatatacatATG	35	91.50	(1.81)	118.05	(1.81)	2.61	3.37
T2_A	TCTAGA tttaagaaggagatatacatATG	29	71.15	(1.19)	122.70	(1.19)	2.45	4.23
T3_A	TCTAGA aggagatatacatATG	22	33.44	(1.15)	129.25	(1.15)	1.52	5.88
FL_G	TCTAGAaataattttgtttaactttaagaaggaggtatacatATG	45	80.01	(2.72)	89.67	(2.72)	1.78	1.99
T1_G	TCTAGA tttaactttaagaaggaggtatacatATG	35	81.27	(1.47)	110.29	(1.47)	2.32	3.15
T2_G	TCTAGA tttaagaaggaggtatacatATG	29	73.12	(2.60)	113.09	(2.60)	2.52	3.90
T3_G	TCTAGA aggaggtatacatATG	22	44.49	(3.27)	111.10	(3.27)	2.02	5.05
CON_A	TCTAGA ttaacttta ggagatatacatATG	30	88.34	(2.96)	122.07	(2.96)	2.94	4.07
CON_G	TCTAGA ttaacttta ggaggtatacatATG	30	100.04	(7.41)	113.92	(7.41)	3.33	3.80
CON_GT1	TCTAGA taacttta ggaggtatacatATG	29	82.83	(2.49)	113.47	(2.49)	2.86	3.91
CON_GT2	TCTAGA aacttta ggaggtatacatATG	28	74.18	(3.10)	107.87	(3.10)	2.65	3.85
CON_GT3	TCTAGA acttta ggaggtatacatATG	27	44.64	(1.16)	89.80	(1.16)	1.65	3.33
CON_GT4	TCTAGA cttta ggaggtatacatATG	26	69.07	(1.30)	119.27	(1.30)	2.66	4.59
Spc_KpnI	TCTAGA cttta ggaggtacccatATG	26	46.52	(1.12)	102.95	(1.12)	1.79	3.96
Spc_EcoRI	TCTAGA cttta ggaggtgaattcATG	26	27.26	(1.01)	100.74	(1.01)	1.05	3.87
Spc_BamHI	TCTAGA cttta ggaggtggatccATG	26	2.09	(0.07)	14.33	(0.07)	0.08	0.55
Spc_SacI	TCTAGA cttta ggaggtgagctcATG	26	7.68	(0.22)	55.41	(0.22)	0.30	2.13
Spc_T1	TCTAGA cttta ggaggtata atATG	25	64.79	(2.79)	119.34	(2.79)	2.59	4.77
Spc_T2	TCTAGA cttta ggaggtata tATG	24	60.92	(0.72)	112.73	(0.72)	2.54	4.70
GGT	TCTAGA ggtatacatATG	18	6.00	(0.21)	63.28	(0.21)	0.33	3.52

b

Fig. 2 Effects of RBS modifications on fluorescent output. **a** Sequence and length of each RBS is displayed along with its relative strength. Output/length ratio was calculated by dividing the relative strength value by the length of the RBS sequence. **b** Relative fluorescence in each strain was determined by using the average of nine cultures (three technical replicates from each of three biological replicates). The wildtype g10RBS (FL_A) measurement from *E. coli* was calibrated to "100" with all other fluorescence values normalized to that standard. Error bars indicate the standard deviation

"CON_GT3", and "CON_GT4") led to 17%, 26%, 55%, and 31% reductions in expression, respectively. Modifying the spacer between the SD and the initiation codon to incorporate restriction sites ("Spc_KpnI", "Spc_EcoRI", "Spc_BamHI", and "Spc_SacI") led to 33%, 61%, 97%, and 89% reductions in expression, respectively. Removing 1 and 2 bases ("Spc_T1" and "Spc_T2") from the spacer led to 6% and 12% reductions in expression, respectively (see Fig. 2).

In *A. fabrum*, FL_A and FL_G gave nearly identical expression levels and truncations of these yielded even higher levels of translation. In fact, the 22-nt T3_A sequence supported the highest level of translation of all 21 sequences tested. The CON_A and CON_G RBS sequences yielded 33% and 24% (respectively) more fluorescence than FL_A, with CON_A exhibiting 94% of the activity as T3_A. The CON_GT series of Enh truncations had similar effects in *A. fabrum* as in *E. coli*, with the CON_GT4 variant unexpectedly giving higher expression than CON_GT3. Levels of expression in *A. fabrum* remained high in most cases when modifying the spacer between the SD and initiation codon,

two exceptions being the Spc_BamHI and Spc_SacI variants, which led to 88% and 54% decreases in expression, respectively. These are also the variants that gave the lowest expression values in *E. coli*. Interestingly, cultures of *A. fabrum* harboring a few of the most active RBS sequences (T1_A, T2_A, T3_A, and CON_A) initially had high levels of fluorescence and somewhat decreased cell density. In subsequent passages of these cultures, normal cell density was restored, but fluorescence decreased—suggesting a fitness cost associated with such high-level expression of the reporter gene. Consistent with this, colonies from these high-expressing strains appeared slightly smaller compared to the other strains (see Supplementary Fig. S1a–c). Suppression of fitness defects was possibly due to mutations in the promoter, RBS, reporter gene, or plasmid origin of replication, though the precise cause was not investigated.

To determine if the effect of RBS variants on reporter gene expression is in part due to flanking sequences, we selected four RBS variants to test in a new promoter/reporter context. The RBS sequences were selected to represent a

wide range of translation enhancement. Whereas the original system employs the *lac*T5 promoter and mScarlet-I reporter, the alternative system uses the constitutive Pkan promoter (from the Tn5 transposon) and a monomeric superfolder GFP (msfGFP) reporter. The dissimilarity of flanking sequences in these two systems is shown in Fig. 3a. In *E. coli*, no differences in relative RBS activity could be seen between these sequence contexts (Fig. 3b). In *A. fabrum*, the relative values differ somewhat. For example, CON_GT4 gives slightly higher translation than CON_G in the mScarlet-I context (a ∼ 5% increase), but substantially lower activity in the msfGFP context (a 33% decrease; see Fig. 3c).

To test the influence of the four RBS variants on protein expression and purification from *E. coli*, the msfGFP gene was modified to encode a *C*-terminal His$_6$ tag in each expression plasmid. Expression and purification according to conventional methods (see Materials and Methods) was carried out, and gel analysis of purified products is shown in Fig. 3d. Consistent with the fluorescence-based analysis, CON_G and CON_GT4 gave the highest yield of recombinant protein, with T3_G yielding a lower amount, and GGT yielding undetectable levels of recombinant protein.

Among the most notable RBS variants tested in this study, the SD-initiator spacer variants were rather surprising, though consistent with recent work indicating that adenosine

nucleotides in this region can enhance translation, while cytidine nucleotides are unfavorable [18]. The most detrimental spacer substitutions in our analysis were Spc_SacI (ATACAT to GAGCTC) and Spc_BamHI (ATACAT to GGATCC). Both of these changes reduce the number of A nucleotides and increase the number of C nucleotides. We acknowledge, however, that the effect of spacer sequence on translation efficiency is idiosyncratic, as recent high-throughput studies identified efficient RBS sequences that deviate from this rule [19, 20].

In both *E. coli* and *A. fabrum*, we expected the CON_GT1/2/3/4 truncation series (which removes 1 to 4 nucleotides from Enh) to exhibit a downward trend in expression level compared to CON_G. While this was partly true, CON_GT4 showed a rebounded level of expression compared to CON_GT3. This observation bolsters the model in which the effect of Enh does not necessarily depend on its propensity to hybridize to the complementary sequence on 16S [6]. High levels of expression in *A. fabrum* for T3_A and T3_G indicate the dispensability of Enh in this organism.

The g10RBS derivatives tested in this study display a range of observed activity, including a minimized version (CON_G) that retains 100% activity in *E. coli* while being 33% shorter. This high activity is evident in two different promoter/reporter sequence contexts. *A. fabrum* appears
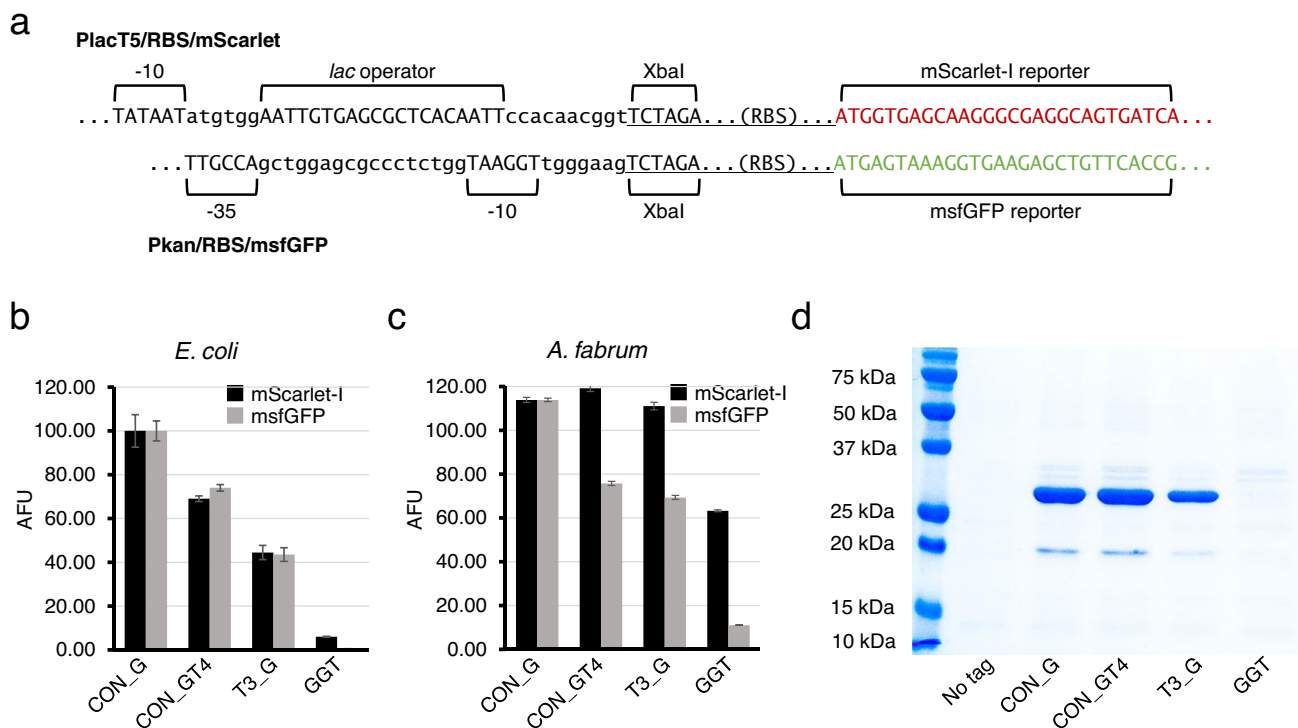
Fig. 3 Effects of selected RBS sequences on fluorescent output in a new promoter/reporter gene context. **a** The two RBS sequence contexts are displayed. **b** and **c** Relative mScarlet and GFP fluorescence in each strain was determined by using the average of nine cultures (three technical replicates from each of three biological replicates). Error bars indicate the standard deviation. **d** His$_6$ tagged GFP was purified from *E. coli* cells harboring the selected RBS sequences

to be less influenced by the varying RBS sequences tested here which may be related to the parental sequence not being native to this species. The set of RBS sequences presented here may be useful to researchers in need of fine-tuned expression of recombinant proteins of interest in diverse species.

Materials and methods

Bacterial genetic manipulations and growth conditions

Escherichia coli DH5α (Sharon Long collection) and *A. fabrum* B527 (a streptomycin-resistant derivative of C58, Griffitts lab collection) strains were grown in lysogeny broth (LB) at 37 °C or 30 °C, respectively. Fluorescent protein expression plasmids were constructed in vitro, transformed into DH5α, sequence-verified, and transferred to *A. fabrum* by conjugation using helper strain B001 [21]. Plasmid selection was carried out in lysogeny broth (LB) containing 30 µg/ml kanamycin (GoldBio K-120–5) (*E. coli*) or 100 µg/ml neomycin (Sigma N1876-25G) (*A. fabrum*). Full sequences for the FL_A-containing starting plasmids (pJG1082 for mScarlet-I expression and pAB215 for msfGFP expression) are given as supplementary information, with the g10RBS and initiation codon capitalized. Variants of these plasmids, and primers used for their construction, are documented in Supplementary Tables S1 and S2.

Measurement of fluorescent output

Escherichia coli and *A. fabrum* strains harboring plasmids with modified RBS sequences were used to start liquid LB cultures, which were initially grown to saturation. Saturated cultures were passaged in triplicate to fresh LB and grown for 8 h (*E. coli*) or 19 h (*A. fabrum*) before taking OD_{600} measurements and adjusting all cultures to an OD of 0.5. After adjustment, each culture was then distributed in triplicate to a flat-bottom, 96-well microplate (CellStar, F-bottom), and fluorescent measurements were taken on a VICTOR Nivo Multimode Plate Reader. For mScarlet-I measurements, the excitation and emission filters were set to 540/10 nm and 595/10 nm, respectively; for msfGFP measurements, the excitation and emission filters were set to 480/30 nm and 530/30 nm, respectively. Negative control strains lacking fluorescent reporters were used to subtract background fluorescence. The wildtype FL_A measurement in *E. coli* was calibrated to "100" and all other fluorescence measurements were normalized to that standard.

Protein purification

To prepare samples for msfGFP-His$_6$ purification, 100-ml cultures were inoculated with 2 ml of overnight culture and allowed to grow at 30 °C for 7.5 h. Centrifuged cell pellets were frozen at − 80 °C overnight. Pellets were then thawed on ice and re-suspended in 1.3 ml of lysis buffer (50 mM HEPES pH 7.8 (FisherScientific BP310-500), 300 mM NaCl (Mallinckrodt Chemicals 7581-06), 0.2% Triton X-100 (Sigma T8787-100 ml), 0.5 mg/ml lysozyme (Sigma L6876-5G), 60 mM imidazole (Sigma 12399-100G), 1 mM EDTA (Sigma E4884-500G)). Lysis took place for 1 h at 4 °C. Cell lysates were then sonicated to ensure complete lysis and fragmentation of DNA. Sonicated samples were cleared by centrifugation, and supernatant transferred to a new microcentrifuge tube. This was incubated end-over-end with 50 µl of NTA-nickel agarose beads (Qiagen 1018244) at 4 °C for 30 min. Beads were washed four times with 1 ml wash buffer (50 mM HEPES pH 7.8, 300 mM NaCl, and 60 mM imidazole), and protein was eluted by incubating the beads in 50 µl of SDS-containing sample buffer at 100 °C. Eluted protein samples were resolved by SDS-PAGE (12% acrylamide, Apex Bioresearch Products 18–197) and stained using Coomassie blue prior to imaging.

Declarations

References

1. Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. https://doi.org/10.3389/fmicb.2014.00172

2. Wilson BS, Kautzer CR, Antelman DE (1994) Increased protein expression through improved ribosome-binding sites obtained by library mutagenesis. Biotechniques 17:944–953

3. Olins PO, Devine CS, Rangwala SH, Kavka KS (1988) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. Gene 73:227–235

4. Olins PO, Rangwala SH (1989) A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacz gene in *Escherichia coli*. J Biol Chem 264:16973–16976

5. Golshani A, Golomehova V, Mironova R, Ivanov IG, AbouHaidar MG (1997) Does the epsilon sequence of phage T7 function as an initiator for the translation of cat mRNA in *Escherichia coli*? Biochem Biophys Res Commun 236:253–256

6. O'Connor M, Dahlberg AE (2001) Enhancement of translation by the epsilon element is independent of the sequence of the 460 region of 16s rRNA. Nucleic Acids Res 29:1420–1425

7. Golshani A, Kolev V, AbouHaidar MG, Ivanov IG (2000) Epsilon as an initiator of translation of cat mRNA in *Escherichia coli*. Biochem Biophys Res Commun 273:528–531

8. Golshani A, Kolev V, Mironova R, AbouHaidar MG, Ivanov IG (2000) Enhancing activity of epsilon in *Escherichia coli* and agrobacterium tumefaciens cells. Biochem Biophys Res Commun 269:508–512

9. Takahashi S, Furusawa H, Ueda T, Okahata Y (2013) Translation enhancer improves the ribosome liberation from translation initiation. J Am Chem Soc 135:13096–13106

10. Golshani A, Xu J, Kolev V, Abouhaidar MG, Ivanov IG (2002) Inability of agrobacterium tumefaciens ribosomes to translate in vivo mrnas containing non-shine-dalgarno translational initiators. Z Naturforsch C J Biosci 57:307–312

11. Boni IV, Isaeva DM, Musychenko ML, Tzareva NV (1991) Ribosome-messenger recognition: mRNA target sites for ribosomal protein s1. Nucleic Acids Res 19:155–162

12. Komarova AV, Tchufistova LS, Supina EV, Boni IV (2002) Protein s1 counteracts the inhibitory effect of the extended shine-dalgarno sequence on translation. RNA 8:1137–1147

13. Komarova AV, Tchufistova LS, Dreyfus M, Boni IV (2005) Au-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. J Bacteriol 187:1344–1349

14. Rangwala SH, Fuchs RL, Drahos DJ, Olins PO (1991) Broad host-range vector for efficient expression of foreign genes in gram-negative bacteria. Bio/Technology 9:477–479

15. Bindels DS et al (2017) Mscarlet: a bright monomeric red fluorescent protein for cellular imaging. Nat Methods 14:53–56

16. Evfratov SA et al (2017) Application of sorting and next generation sequencing to study 5'-UTR influence on translation efficiency in *Escherichia coli*. Nucleic Acids Res 45:3487–3502

17. Kuo S-T et al (2020) Global fitness landscapes of the shine-dalgarno sequence. Genome Res 30:711–723

18. Komarova ES et al (2020) Influence of the spacer region between the shine–dalgarno box and the start codon for fine-tuning of the translation efficiency in *Escherichia coli*. Microb Biotechnol 13:1254–1261

19. Shilling PJ et al (2020) Improved designs for pet expression plasmids increase protein production yield in *Escherichia coli*. Commun Biol 3:214

20. Ding N et al (2020) Programmable cross-ribosome-binding sites to fine-tune the dynamic range of transcription factor-based biosensor. Nucleic Acids Res 48:10602–10613

21. Griffitts JS et al (2008) A sinorhizobium meliloti osmosensory two-component system required for cyclic glucan export and symbiosis. Mol Microbiol 69:479–490