

Translation enhancement by a *Dictyostelium* gene sequence in *Escherichia coli*

Tomo Kondo* and Shigehiko Yumura

Graduate School of Sciences and Technology for Innovation, Yamaguchi University,
Yamaguchi 753-8512, Japan

*Correspondence to Tomo Kondo: Phone: +81-83-933-5717, E-mail:
tomokond@yamaguchi-u.ac.jp, ORCID: 0000-0003-2507-3399

Abstract

Methods for heterologous protein production in *Escherichia coli* have revolutionized biotechnology and the bioindustry. It is ultimately important to increase the amount of protein product from bacteria. To this end, a variety of tools, such as effective promoters, have been developed. Here, we present a versatile molecular tool based on a phenomenon termed “translation enhancement by a *Dictyostelium* gene sequence” (“TED”) in *E. coli*. We found that protein expression was increased when a gene sequence of *Dictyostelium discoideum* was placed upstream of the Shine-Dalgarno sequence located between the promoter and the initiation codon of a target gene. The most effective sequence among the genes examined was *mlcR*, which encodes the myosin regulatory light chain, a subunit of myosin II. Serial deletion analysis revealed that at least 10 bases of the 3' end of the *mlcR* gene enhanced the production of green fluorescent protein in cells. We applied this tool to a T7 expression system and found that the expression level of the proteins tested was increased when compared with the conventional method. Thus, current protein production systems can be improved by combination with TED.

Key words

Protein production, *E. coli*, *Dictyostelium*, Translation

Introduction

Escherichia coli is one of the most powerful workhorses in biological science, including biotechnology, because of its low cost, tractability, and the availability of conventional methods (Rosenberg et al. 1987; Qing et al. 2004; Structural Genomics Consortium et al. 2008). Over the last four decades, a great variety of proteins, including somatostatin (Itakura et al. 1977), insulin (Goeddel et al. 1979), interferon (Taniguchi et al. 1980; Derynck et al. 1980), interleukin (Devos et al. 1983), and Cas9 (Jinek et al. 2012; Gasiunas et al. 2012), have been bacterially synthesized in the laboratory and at the industrial scale. More recently, bacterial synthesis of antibody fragments (e.g., Fab and nanobody) has been reported to minimize the cost of and the sacrifice of animals for production (Spadiut et al. 2014; Buser et al. 2018). Thus, protein production using genetically engineered *E. coli* benefits the understanding of fundamental cellular processes as well as the development of biopharmaceuticals.

To increase protein yield, it is effective to increase the absolute amount of protein production in the cell. The first approach is to increase the transcription of the target gene by using strong promoters, such as *lac* and its derivatives (Dickson et al. 1975; de Boer et al. 1983). In addition, expression levels can be regulated by three factors of the sequence upstream of the initiation codon of the target gene: (i) the strength of the Shine-Dalgarno (SD) sequence (Shine and Dalgarno 1974), (ii) the distance between the SD sequence and the initiation codon (Singer et al. 1981; Shepard et al. 1982; Ringquist et al. 1992; Chen et al. 1994a), and (iii) the sequence upstream of the SD sequence (Roberts et al. 1979; Stanssens et al. 1985; Coleman et al. 1985; McCarthy et al. 1985; Olins et al. 1988). In factors (i) and (ii), optimum patterns can be determined by a complementary sequence of the 3' end region of 16S rRNA and the relative position of the 30S ribosome, respectively. The third factor is known to affect mRNA stability and the preferred interaction of mRNA with the ribosomal S1 protein (Boni et al. 1991; Komarova et al. 2002; Komarova et al. 2005; Takahashi et al. 2013). According to these studies, an A/U-rich sequence, with typically more than 70% AT content at the 5'-untranslated region (UTR), is empirically effective, but which sequences can efficiently induce protein expression remains elusive.

Here, we report that gene sequences from the social amoeba *Dictyostelium discoideum*, which has an A/T-rich genome (Eichinger et al. 2005), enhance expression

in *E. coli*. We tested genes having 60–73% AT content and found that even a sequence an AT content of less than 70% has the ability to enhance expression. The most effective sequence was the *mlcR* gene, which encodes the myosin regulatory light chain, an evolutionarily conserved subunit protein that regulates the function of myosin II (Uyeda and Spudich 1993; Chen et al. 1994b; Liu et al. 1998; Kondo et al. 2011, 2012, 2015). As proof-of-principle, we demonstrate the usefulness of this sequence for expression enhancement by assessing the expression of proteins, including green fluorescent protein (GFP), in *E. coli*.

Materials and methods

E. coli culture

E. coli strains HST08 Premium (Takara Bio, Shiga, Japan), DH5 α (Cosmo Bio, Tokyo, Japan), and BL21(DE3) (Cosmo Bio) were used. The cells were cultured at 37°C in LB medium (Nippon Genetics, Tokyo, Japan) supplemented with 100 μ g/ml ampicillin (Wako Pure Chemical Industries, Osaka, Japan) or 20 μ g/ml kanamycin (Wako Pure Chemical Industries). Luria-Bertani (LB) agar was prepared by solidifying LB medium with 1.5% agar (Wako Pure Chemical Industries). For expression induction by isopropyl β -D-thiogalactopyranoside (IPTG) (Nacalai Tesque, Kyoto, Japan), 1/100 volume of precultured cells were inoculated in fresh LB medium containing appropriate antibiotics and were cultured at 37°C until an optical density at 600 nm of 0.5–0.6 was reached. Then, IPTG was added at the final concentration of 0.5 mM, followed by culture at 22°C for 16 h or at 37°C for 3 h.

Plasmids and cDNA cloning

Isolation of plasmid DNA, preparation of DNA fragments, ligation, and transformation of *E. coli* cells were carried out using standard techniques. The genes of *D. discoideum* were amplified from cDNA (Robinson and Spudich 2000) or genomic DNA of the Ax2 strain using PrimeStar Max DNA polymerase (Takara Bio) and the designed primers, and were cloned into pUC19 (Yanisch-Perron et al. 1985) (Takara Bio). The cDNA was kindly provided by Dr. Douglas Robinson (Johns Hopkins University School of Medicine, Baltimore, MD, USA). Genomic DNA was purified from Ax2 cells with the NucleoSpin Tissue kit (Takara Bio). The following plasmids were gifted by Dr. Scott Gradia: pET His6 GST TEV LIC cloning vector (1G) (Addgene plasmid # 29655), pET His6 MBP TEV LIC cloning vector (2M-T) (Addgene plasmid # 29708), pET His6 Sumo TEV LIC cloning vector (2S-T) (Addgene plasmid # 29711), and pET GFP LIC cloning vector (2GFP-T) (Addgene plasmid # 29716). The reverse complementary sequence of *mlcR* was synthesized by Integrated DNA Technologies, Inc. (Coralville, IA, USA).

Imaging and quantification

For excitation of GFP in bacteria grown on an agar plate, a custom-made blue light-emitting diode (LED) illuminator was used. For imaging of *E. coli*, culture was dropped on a coverslip (24 × 60 mm; Matsunami Glass, Osaka, Japan) and overlaid with a block of 1.5% agar (8 × 8 × 1 mm; Dojindo Laboratories, Kumamoto, Japan) in distilled water. The specimens were observed with an inverted microscope (DIAPHOT300; Nikon, Tokyo, Japan) equipped with a camera (Orca-ER C4742-80-12AG; Hamamatsu Photonics, Hamamatsu, Japan) and a mercury lamp (Nikon). Images were acquired five times using a 100× objective lens (Nikon), and maximum intensity projection was performed using Fiji (Schindelin et al. 2012). To obtain a Z score, the cytoplasmic fluorescence intensity was normalized by subtracting the mean intensity from 10 background regions, and the value was divided by the standard deviation of the intensity of the background regions. The statistical significance of differences between groups for a dataset was tested by the Wilcoxon rank sum test (“ranksum” function in MATLAB; MathWorks, Inc., Natick, MA, USA).

Prediction of RNA secondary structure

The local secondary structure of RNA was predicted using RNAfold (Gruber et al. 2008) with default settings. Centroid structures, which are the secondary structures with minimal base pair distance to all other secondary structures in the Boltzmann ensemble, are shown in relevant figures.

Results

Protein expression in *E. coli* is enhanced by *Dictyostelium* gene sequences

We cloned the *mlcR* gene from a cDNA library of *D. discoideum* Ax2 and linked it with *gfp* in the pUC19 vector (Fig. 1a). These two genes were connected with nucleotides composed of an SD sequence-spacer (Fig. S1). After transformation of the plasmid into *E. coli* (strain HST08) and overnight culture on LB agar in the dark, we found that the colonies showed fluorescence upon excitation with blue LED light, even in the absence of the expression inducer IPTG (Fig. 1b, bottom-right). In contrast, no visible fluorescence appeared in cells harboring the plasmid containing only the SD sequence-spacer and *gfp* (Fig. 1b, top-left). GFP expression did not adversely affect the growth rate (Fig. S2). A similar level of fluorescence was detected in another strain, DH5 α (Fig. S3). Moreover, fluorescence could be observed in cells expressing not only GFP, but also the red fluorescence protein mRuby3 (Bajar et al. 2016) (Fig. 1c). Microscopic observation revealed that fluorescence was observed in the cytoplasm of individual cells (Figs. 1d and e). Thus, the *mlcR* sequence enhances the expression of the downstream gene.

Next, we investigated whether gene sequences other than *mlcR* (68% AT content) also induce expression enhancement. Previous studies have reported translation enhancement, which is induced by A/U-rich sequences upstream of the SD sequence (McCarthy et al. 1985; Komarova et al. 2002). Thus, we additionally tested arbitrarily selected 8 genes, including *H1*, *eb1*, *tubC*, *tom7*, *tom20*, DDB_G0288629, *tom40*, and *tom70*, having AT contents of 60–73% (Fig. 2a, Table S1). All of the genes tested showed expression enhancement activity of various strengths (Fig. 2b). Hereafter, we refer to this phenomenon as “TED” (“translation enhancement by a *Dictyostelium* gene sequence”).

The *mlcR* gene (68% AT content, 483 bp) had the highest TED activity among the genes examined. TED activity was measured as the ratio increase in GFP fluorescence. As shown in Figs. 2b and c, there was no obvious association between the strength of TED ability and the AT content or the length of the gene. The *H1* gene (60% AT content, 540 bp) and the *tom7* gene (73% AT content, 165 bp) exhibited relatively high TED activity. In contrast, some of the genes that had an AT content of more than 70% (i.e., DDB_G0288629 [71%, 789 bp], *tom40* [71%, 942 bp], and *tom70* [70%,

1605 bp]) or were similar to *mlcR* in AT content (i.e., *ebf1* [67%, 1518 bp] and *tubC* [67%, 1386 bp]) showed relatively weak activity. These data suggest that TED activity cannot be simply explained by the simple parameters AT content and gene length. Consistent herewith, the complementary sequence of *mlcR* no longer had a translation enhancement effect (Fig. 2a). Thus, the simple concept of using an AU-rich sequence for enhancing expression was proven insufficient.

TED activity can be achieved with a 3' end of at least 10 bp of the *mlcR* sequence

Hereafter, we focus on *mlcR*, which had the strongest TED activity. To analyze the mechanism of expression enhancement, we identified the sequence that is indispensable for enhanced protein expression by *mlcR* deletion analysis. As the inducer agent (i.e., IPTG) was not added to the medium, we hypothesized that the basal activity of the *lac* promoter with the catabolite activator protein (CAP) site contributes to the expression of *gfp*, as previously described (Kennell and Riezman 1977; Yu and Reznikoff 1984; Bellis and Schwartz 1990; Wilson et al. 2007; Gatti-Lafranconi et al. 2013). As expected, GFP fluorescence disappeared for all examined patterns that lacked both the *lac* promoter (*Plac*) and the *lac* operator (*lacO*), only *Plac*, or the CAP site (Fig. 3a), suggesting that *Plac*-dependent transcription, not a potential cryptic promoter in the *mlcR* sequence, is responsible for *gfp* expression.

Next, we investigated which region of the *mlcR* sequence is required for GFP expression. The sequence was deleted serially from the 5' end of full-length *mlcR* cDNA (483 bp) (Figs. 3b and S1). In the *mlcR* sequence, there are eight potential initiation codons (ATG) in the frame that can express *gfp*. However, there are no sites with an SD sequence (e.g., AGGA, GAGG, or GGAG) upstream (at least 15 bp) of the initiation codon. Thus, the SD sequence downstream of the *mlcR* sequence is the only sequence to express *gfp*. We found that even only 10 bp of the 3' end of *mlcR* (*mlcR10*) had TED ability. The 25-bp fragment of the 3' end of *mlcR* (*mlcR25*) showed the highest activity. Interestingly, the 5-bp longer construct (*mlcR30*) showed the lowest activity. These data suggest that TED activity can be controlled by using at least 10 bp or longer of the *mlcR* sequence.

Predicted local secondary structure around the ribosome-binding site

The secondary structure of RNA affects the translation rate (Iserentant and Fiers 1980; Gold et al. 1981; de Smit and Duin 1990; de Smit and van Duin 1990; Osterman et al. 2013). To gain insight into the variation in strength of TED activity using parts of the *mlcR* sequence of various lengths, we examined the predicted local structure of RNA. To focus on the property of the inserted sequence, the inserted *mlcR* sequence and the initiation codon were used for calculation. The sequences with relatively high expression upregulation (*mlcR10* [Fig. 4a], *mlcR20* [Fig. 4b], *mlcR25* [Fig. 4c], *mlcR40* [Fig. 4e], *mlcR45* [Fig. 4f], and *mlcR74* [Fig. 4h]) were expected to have a stem-loop structure near the initiation codon, whereas such structure was lost in *mlcR30* (Fig. 4d) and *mlcR60* (Fig. 4g), which had relatively low TED activity. These correlations suggest that the property of forming a stem loop consisting of an SD sequence and a proper length of the *mlcR* sequence is suitable for effective translation.

***mlcR25* TED improves the expression level of the classical pET expression system**

We combined TED with the pET expression system (Studier et al. 1990). The pET vectors contain the T7 phi10 promoter and related sequences upstream of the start codon of a target protein. To demonstrate the effect of TED, we replaced 22 bp of the sequence between the SD sequence and an *XbaI* site with *mlcR25* (Fig. 5a). The similarity of these sequences was 41% (9/22 nucleotides) at most. According to the prediction of the local RNA structure, both the T7 phi10 sequence (pET-native) and *mlcR25* formed a similar stem-loop structure (Fig. 5b and c). Using these vectors, we expressed several types of proteins in *E. coli* strain BL21(DE3), which is frequently used for high-yield production. The formation of protein aggregates (i.e., inclusion bodies) is generally prevented at low temperature (Schein and Noteborn 1988); therefore, low-temperature culture is usually used to increase yield (Structural Genomics Consortium et al. 2008). Thus, we tested two culture temperatures, 22°C and 37°C. After IPTG induction, GFP (Fig. 5d), GFP-tagged maltose binding protein (MBP) (Fig. 5e), small ubiquitin-like modifier (SUMO) (Fig. 5f), and glutathione-S-transferase (GST) (Fig. 5g) were successfully expressed in cells harboring *mlcR25*-inserted plasmid, and the fluorescence level was increased by approximately 10% compared to that in cells harboring the original T7 phi10-inserted plasmid, at 22°C. A similar effect was

observed in cells cultured at 37°C (Fig. 5h-k). These data suggest that the classical pET expression system can be improved by using *mlcR25* TED.

Discussion

This study demonstrated the principle of using TED activity for translation enhancement in *E. coli*. We identified nine genes that have TED activity, among which *mlcR* has the strongest activity. In addition, the activity can be tuned by deletion of part of the sequence. Using TED, we could improve the pET expression system.

We explored sequences with translation enhancement activity by examining various patterns from a preset of *Dictyostelium* gene sequences as a starting material. Alternatively, we may find sequences with a similar effect from randomly designed sequences. The latter approach requires the synthesis of various sequences because there is no template, and thus, is costly and inefficient. From the data currently available in NCBI (<https://www.ncbi.nlm.nih.gov/>), more than 13,000 coding sequences of *D. discoideum* are annotated. These sequences are promising materials for improving the efficiency of protein production in *E. coli*.

It has been reported that the secondary structure of the sequence upstream of the SD sequence influences the loading and liberation efficiency of the ribosome and subsequent protein expression (de Smit and van Duin 2003; Osterman et al. 2013; Takahashi et al. 2013; Espah Borujeni et al. 2014). By comparing the results of deletion analysis of the *mlcR* sequence with the local secondary RNA structure, we predicted that the sequence that forms a stem loop structure with the SD sequence is effective for TED activity. Based on this criterion, our data empirically indicate that the optimal minimal free energy is calculated to be approximately -3 kcal/mol. This prediction will serve as a basis for finding a better sequence.

We predicted that a sequence forming a moderately stable stem loop upstream of the initiation codon would confer effective TED activity (Fig. 4). This may seem to contradict the view on the relationship between the accessibility of ribosomes to linear mRNA and translation. Interestingly, a recent study reported that the mechanism of internal ribosome-entry site (IRES) works in bacteria as well as in eukaryotes (Colussi et al. 2015). Given that IRES RNAs are known to form intricate structures (Kieft 2008), bacterial ribosomes might recognize and bind to mRNAs having certain structures to initiate translation. Thus, the IRES-like structure may contribute to TED-dependent expression.

An alternative, but not mutually exclusive, mechanism determining the strength of TED activity is mRNA stability (Radhakrishnan and Green 2016). It is known that the features of the 5'-UTR affect the longevity of mRNA (Belasco et al. 1986; Bouvet and Belasco 1992). Furthermore, Komarova et al. (2005) have reported that an A/U-rich 5'-UTR increases mRNA longevity. Unfortunately, due to a technical problem, we could not examine this mechanism in the present study.

Intensive studies have led to a number of improved methods that are associated with potent promoters and optimization of the SD sequence, spacers, and codons (Chen et al. 1994a; Vimberg et al. 2007; Salis et al. 2009; Hanson and Collier 2018). Interestingly, the insertion of a translation-promoting sequence downstream of the initiation codon is also effective (Etchegaray and Inouye 1999; Qing et al. 2003). It is worth emphasizing that TED can be used in combination with all of these methods. One of the advantages of TED is the ease of activity regulation. As shown in Figs. 2A and 3B, even if one promoter is used, various expression levels can be tailored by simple manipulation of the gene sequence, which may be useful for analyzing genes whose phenotype changes depending on the expression level (Ward and Lutkenhaus 1985). Moreover, as the translation mechanism is largely conserved in various bacteria (Laursen et al. 2005), TED is expected to be applicable to expression systems for other bacterial species, such as *Magnetospirillum sp.* (Matsunaga et al. 2007; Uebe and Schüler 2016), *Lactococcus lactis* (Hugenholtz and Smid 2002; Song et al. 2017), and *Bacillus subtilis* (Wong 1995; Song et al. 2015).

In conclusion, tuning of the expression level of a heterologous protein was achieved by using *D. discoideum* gene sequences upstream of the SD sequence. The level of enhancement can be adjusted by the gene used and its length. This method is simple, inexpensive, and easy to perform, and it can be readily adapted for protein expression in both the laboratory and at the industrial scale.

Acknowledgments

TK was supported by the Japan Society for the Promotion of Science Research Fellowships for Young Scientists.

Funding

This research was supported by the Japan Society for the Promotion of Science KAKENHI Grant Number 16J08310 to TK.

Compliance with ethical standards

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Bajar BT, Wang ES, Lam AJ, Kim BB, Jacobs CL, Howe ES, Davidson MW, Lin MZ, Chu J (2016) Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci Rep.* 6:20889. doi: 10.1038/srep20889
- Belasco JG, Nilsson G, Gabain von A, Cohen SN (1986) The stability of *E. coli* gene transcripts is dependent on determinants localized to specific mRNA segments. *Cell* 46:245–251.
- Bellis DD, Schwartz I (1990) Regulated expression of foreign genes fused to lac: Control by glucose levels in growth medium. *Nucleic Acids Res.* 18:1311–1311. doi: 10.1093/nar/18.5.1311
- 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.
- Bouvet P, Belasco JG (1992) Control of RNase E-mediated RNA degradation by 5'-terminal base pairing in *E. coli*. *Nature* 360:488–491. doi: 10.1038/360488a0
- Buser DP, Schleicher KD, Prescianotto-Baschong C, Spiess M (2018) A versatile nanobody-based toolkit to analyze retrograde transport from the cell surface. *Proc Natl Acad Sci U S A.* 12:201801865–E6236. doi: 10.1073/pnas.1801865115
- Chen H, Bjerknes M, Kumar R, Jay E (1994a) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.* 22:4953–4957.
- Chen P, Ostrow BD, Tafuri SR, Chisholm RL (1994b) Targeted disruption of the *Dictyostelium* RMLC gene produces cells defective in cytokinesis and development. *J Cell Biol.* 127:1933–1944.
- Coleman J, Inouye M, Nakamura K (1985) Mutations upstream of the ribosome-binding site affect translational efficiency. *J Mol Biol.* 181:139–143. doi: 10.1016/0022-2836(85)90332-8

- Colussi TM, Costantino DA, Zhu J, Donohue JP, Korostelev AA, Jaafar ZA, Plank T-DM, Noller HF, Kieft JS (2015) Initiation of translation in bacteria by a structured eukaryotic IRES RNA. *Nature* 519:110–113. doi: 10.1038/nature14219
- de Boer HA, Comstock LJ, Vasser M (1983) The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc Natl Acad Sci U S A.* 80:21–25. doi: 10.1073/pnas.80.1.21
- de Smit MH, Duin JV (1990) Control of prokaryotic translational initiation by mRNA secondary structure. In: *Progress in Nucleic Acid Research and Molecular Biology.* Elsevier, pp 1–35
- de Smit MH, van Duin J (1990) Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. *Proc Natl Acad Sci U S A.* 87:7668–7672. doi: 10.1073/pnas.87.19.7668
- de Smit MH, van Duin J (2003) Translational standby sites: how ribosomes may deal with the rapid folding kinetics of mRNA. *J Mol Biol.* 331:737–743.
- Derynck R, Remaut E, Saman E, Stanssens P, De Clercq E, Content J, Fiers W (1980) Expression of human fibroblast interferon gene in *Escherichia coli*. *Nature* 287:193–197.
- Devos R, Plaetinck G, Cheroutre H, Simons G, Degrave W, Tavernier J, Remaut E, Fiers W (1983) Molecular cloning of human interleukin 2 cDNA and its expression in *E. coli*. *Nucleic Acids Res.* 11:4307–4323. doi: 10.1093/nar/11.13.4307
- Dickson RC, Abelson J, Barnes WM, Reznikoff WS (1975) Genetic regulation: The lac control region. *Science* 187:27–35.
- Eichinger L, Pachebat JA, Glöckner G, Rajandream MA, Sugang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I,

- Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbino-witsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435:43–57. doi: 10.1038/nature03481
- Espah Borujeni A, Channarasappa AS, Salis HM (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* 42:2646–2659. doi: 10.1093/nar/gkt1139
- Etchegaray JP, Inouye M (1999) Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*. *J Biol Chem.* 274:10079–10085. doi: 10.1074/jbc.274.15.10079
- Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A.* 109:E2579–E2586. doi: 10.1073/pnas.1208507109
- Gatti-Lafranconi P, Dijkman WP, Devenish SR, Hollfelder F (2013) A single mutation in the core domain of the lac repressor reduces leakiness. *Microb Cell Fact* 12:67. doi: 10.1186/1475-2859-12-67
- Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K, Riggs AD (1979) Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc Natl Acad Sci U S A.* 76:106–110.
- Gold L, Pribnow D, Schneider T, Shinedling S, Singer BS, Stormo G (1981) Translational Initiation in Prokaryotes. *Annu Rev Microbiol* 35:365–403. doi:

- 10.1146/annurev.mi.35.100181.002053
- Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The Vienna RNA Websuite. *Nucleic Acids Res.* 36:W70–W74. doi: 10.1093/nar/gkn188
- Hanson G, Collier J (2018) Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol.* 19:20–30. doi: 10.1038/nrm.2017.91
- Hughenoltz J, Smid EJ (2002) Nutraceutical production with food-grade microorganisms. *Curr Opin Biotechnol* 13:497–507.
- Iserentant D, Fiers W (1980) Secondary structure of mRNA and efficiency of translation initiation. *Gene* 9:1–12. doi: 10.1016/0378-1119(80)90163-8
- Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, Boyer HW (1977) Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 198:1056–1063.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821. doi: 10.1126/science.1225829
- Kennell D, Riezman H (1977) Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *J Mol Biol.* 114:1–21. doi: 10.1016/0022-2836(77)90279-0
- Kieft JS (2008) Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci* 33:274–283. doi: 10.1016/j.tibs.2008.04.007
- 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. doi: 10.1128/JB.187.4.1344-1349.2005
- 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. doi: 10.3791/54064

- Kondo T, Hamao K, Kamijo K, Kimura H, Morita M, Takahashi M, Hosoya H (2011) Enhancement of myosin II/actin turnover at the contractile ring induces slower furrowing in dividing HeLa cells. *Biochem J.* 435:569–576. doi: 10.1042/BJ20100837
- Kondo T, Itakura S, Hamao K, Hosoya H (2012) Phosphorylation of myosin II regulatory light chain controls its accumulation, not that of actin, at the contractile ring in HeLa cells. *Exp Cell Res.* 318:915–924. doi: 10.1016/j.yexcr.2012.02.009
- Kondo T, Okada M, Kunihiro K, Takahashi M, Yaoita Y, Hosoya H, Hamao K (2015) Characterization of myosin II regulatory light chain isoforms in HeLa cells. *Cytoskeleton* 72:609–620. doi: 10.1002/cm.21268
- Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU (2005) Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* 69:101–123. doi: 10.1128/MMBR.69.1.101-123.2005
- Liu X, Ito K, Morimoto S, Hikkoshi-Iwane A, Yanagida T, Uyeda TQ (1998) Filament structure as an essential factor for regulation of *Dictyostelium* myosin by regulatory light chain phosphorylation. *Proc Natl Acad Sci U S A.* 95:14124–14129.
- Matsunaga T, Suzuki T, Tanaka M, Arakaki A (2007) Molecular analysis of magnetotactic bacteria and development of functional bacterial magnetic particles for nano-biotechnology. *Trends Biotechnol* 25:182–188. doi: 10.1016/j.tibtech.2007.02.002
- McCarthy JE, Schairer HU, Sebald W (1985) Translational initiation frequency of atp genes from *Escherichia coli*: identification of an intercistronic sequence that enhances translation. *The EMBO journal* 4:519–526.
- 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. doi: 10.1016/0378-1119(88)90329-0
- Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA (2013) Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res.*

- 41:474–486. doi: 10.1093/nar/gks989
- Qing G, Ma LC, Khorchid A, Swapna GVT, Mal TK, Takayama MM, Xia B, Phadtare S, Ke H, Acton T, Montelione GT, Ikura M, Inouye M (2004) Cold-shock induced high-yield protein production in *Escherichia coli*. Nat Biotechnol. 22:877–882. doi: 10.1038/nbt984
- Qing G, Xia B, Inouye M (2003) Enhancement of translation initiation by A/T-rich sequences downstream of the initiation codon in *Escherichia coli*. J Mol Microbiol Biotechnol 6:133–144. doi: 10.1159/000077244
- Radhakrishnan A, Green R (2016) Connections underlying translation and mRNA stability. J Mol Biol. 428:3558–3564. doi: 10.1016/j.jmb.2016.05.025
- Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, Gold L (1992) Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. Mol Microbiol 6:1219–1229.
- Roberts TM, Kacich R, Ptashne M (1979) A general method for maximizing the expression of a cloned gene. Proc Natl Acad Sci U S A. 76:760–764. doi: 10.1073/pnas.76.2.760
- Robinson DN, Spudich JA (2000) Dynacortin, a genetic link between equatorial contractility and global shape control discovered by library complementation of a *Dictyostelium discoideum* cytokinesis mutant. J Cell Biol. 150:823–838.
- Rosenberg AH, Lade BN, Dao-shan C, Lin S-W, Dunn JJ, Studier FW (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56:125–135. doi: 10.1016/0378-1119(87)90165-X
- Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol. 27:946–950. doi: 10.1038/nbt.1568
- Schein CH, Noteborn MHM (1988) Formation of soluble recombinant proteins in *Escherichia Coli* is favored by lower growth temperature. Bio/Technology 1988 6:3

- 6:291–294. doi: 10.1038/nbt0388-291
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods*. 9:676–682. doi: 10.1038/nmeth.2019
- Shepard HM, Yelverton E, Goeddel DV (1982) Increased synthesis in *E. coli* of fibroblast and leukocyte interferons through alterations in ribosome binding sites. *DNA* 1:125–131. doi: 10.1089/dna.1.1982.1.125
- Shine J, Dalgarno L (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.
- Singer BS, Gold L, Shinedling ST, Colkitt M, Hunter LR, Pribnow D, Nelson MA (1981) Analysis in vivo of translational mutants of the rIIB cistron of bacteriophage T4. *J Mol Biol*. 149:405–432. doi: 10.1016/0022-2836(81)90479-4
- Song AA-L, In LLA, Lim SHE, Rahim RA (2017) A review on *Lactococcus lactis*: From food to factory. *Microb Cell Fact* 16:55. doi: 10.1186/s12934-017-0669-x
- Song Y, Nikoloff JM, Zhang D (2015) Improving Protein Production on the Level of Regulation of both Expression and Secretion Pathways in *Bacillus subtilis*. *J Microbiol Biotechnol* 25:963–977. doi: 10.4014/jmb.1501.01028
- Spadiut O, Spadiut O, Capone S, Capone S, Krainer F, Krainer F, Glieder A, Glieder A, Herwig C (2014) Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol* 32:54–60. doi: 10.1016/j.tibtech.2013.10.002
- Stanssens P, Remaut E, Fiers W (1985) Alterations upstream from the Shine-Dalgarno region and their effect on bacterial gene expression. *Gene* 36:211–223.
- Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Gräslund S, Nordlund P, Weigelt J, Hallberg BM,

- Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park H-W, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim S-H, Rao Z, Shi Y, Terwilliger TC, Kim C-Y, Hung L-W, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schütz A, Heinemann U, Yokoyama S, Büsow K, Gunsalus KC (2008) Protein production and purification. *Nat Methods*. 5:135–146. doi: 10.1038/nmeth.f.202
- 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. doi: 10.1021/ja405967h
- Taniguchi T, Guarente L, Roberts TM, Kimelman D, Douhan J, Ptashne M (1980) Expression of the human fibroblast interferon gene in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 77:5230–5233. doi: 10.1073/pnas.77.9.5230
- Uebe R, Schüler D (2016) Magnetosome biogenesis in magnetotactic bacteria. *Nat Rev Micro* 14:621–637. doi: 10.1038/nrmicro.2016.99
- Uyeda TQ, Spudich JA (1993) A functional recombinant myosin II lacking a regulatory light chain-binding site. *Science* 262:1867–1870.
- Vimberg V, Tats A, Remm M, Tenson T (2007) Translation initiation region sequence preferences in *Escherichia coli*. *BMC Mol Biol* 8:100. doi: 10.1186/1471-2199-8-100
- Ward JE, Lutkenhaus J (1985) Overproduction of FtsZ induces minicell formation in *E. coli*. *Cell* 42:941–949. doi: 10.1016/0092-8674(85)90290-9

- Wilson CJ, Zhan H, Swint-Kruse L, Matthews KS (2007) The lactose repressor system: Paradigms for regulation, allosteric behavior and protein folding. Cellular and molecular life sciences: CMLS 64:3–16. doi: 10.1007/s00018-006-6296-z
- Wong SL (1995) Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins. Curr Opin Biotechnol 6:517–522.
- Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. Gene 33:103–119. doi: 10.1016/0378-1119(85)90120-9
- Yu XM, Reznikoff WS (1984) Deletion analysis of the CAP-cAMP binding site of the *Escherichia coli* lactose promoter. Nucleic Acids Res. 12:5449–5464.

Figure captions

Fig. 1 GFP expression in *E. coli* harboring the plasmid containing an *mlcR* gene sequence upstream of the SD sequence, a spacer, and *gfp*. (a) *mlcR-gfp* was cloned between the *Xba*I and *Bam*HI sites in the pUC19 vector. The nucleotide sequence of the insert is shown in Fig. S1. The SD sequence is designated as “SD.” (b) GFP fluorescence of *E. coli* (strain HST08) on an LB-agar plate not containing IPTG. The bacteria harboring the plasmid described in (a) in the bottom right of the figure exhibit GFP fluorescence. The upper left bacteria without fluorescence contain a similar plasmid lacking the *mlcR* gene. The plate was photographed over a UV shield during irradiation with a hand-held blue-light LED illuminator. (c) Bacterial pellets with or without fluorescence. Left; no fluorescence, center; GFP fluorescence, right; mRuby3 fluorescence. (d and e) Merged fluorescence and phase-contrast microscopic images of cells expressing GFP (green) or mRuby3 (magenta).

Fig. 2 Enhancement of GFP expression in *E. coli* harboring plasmid containing the full length of the indicated gene sequence. (a) The strength of induction of GFP expression differed depending on the gene. All genes were inserted into the pUC19 vector, as shown in Fig. 1A. As a negative control, cells harboring empty pUC19 vector and only SD sequence, designated as “SD,” were tested. GFP expression can be visually confirmed with a Z score of approximately 80. Error bars represent the standard errors of the means. All statistical tests are the result of comparison with SD. $n \geq 239$. $**p < 0.01$ (Wilcoxon rank sum test). (b) Scatter plot of GFP intensity versus AT content of the indicated gene. (c) Scatter plot of GFP intensity versus nucleotide length of the indicated gene.

Fig. 3 Analysis of the region required for GFP expression in the plasmid containing *mlcR-gfp*. (a) Deletion of the *lac* promoter (*Plac*), the *lac* operator (*lacO*), and/or the CAP site reduced GFP expression. Error bars represent the standard errors of the means. $n \geq 239$. (b) Serial deletion of the *mlcR* sequence in the pUC19 vector. The number that follows “*mlcR*” is the number of bases from the 3' side that does not contain the stop codon of the *mlcR*. The value of *mlcR* is the same as that shown in (a). Error bars represent the standard errors of the means. $n \geq 344$.

Fig. 4 *mlcR* sequences with relatively high TED activity form a stem loop with the SD sequence. Predicted secondary RNA structures of *mlcR10* (a), *mlcR20* (b), *mlcR25* (c), *mlcR30* (d), *mlcR40* (e), *mlcR45* (f), *mlcR60* (g), and *mlcR74* (h). The *mlcR* sequence and initiation codon are shown in magenta bold and green italics, respectively. The calculated minimum free energies at 37°C are indicated.

Fig. 5 Protein expression in *E. coli* strain BL21(DE3) using T7 expression vectors. (a) Comparison of the T7 phi10 sequence and *mlcR25*. T7 phi10, shown in blue bold, was replaced *mlcR25* shown in magenta bold. The SD sequence is boxed. The initiation codon is shown in green italics. (b, c) Predicted structure of the sequences described in (a). The calculated minimum free energies are indicated. (d-e) Comparison of expression levels of the indicated proteins in cells cultured at 22°C. Fluorescence was measured after IPTG induction for 16 h. (h-k) Comparison of expression levels of the indicated proteins in cells cultured at 37°C. Fluorescence was measured after IPTG induction for 3 h.

Figure 1

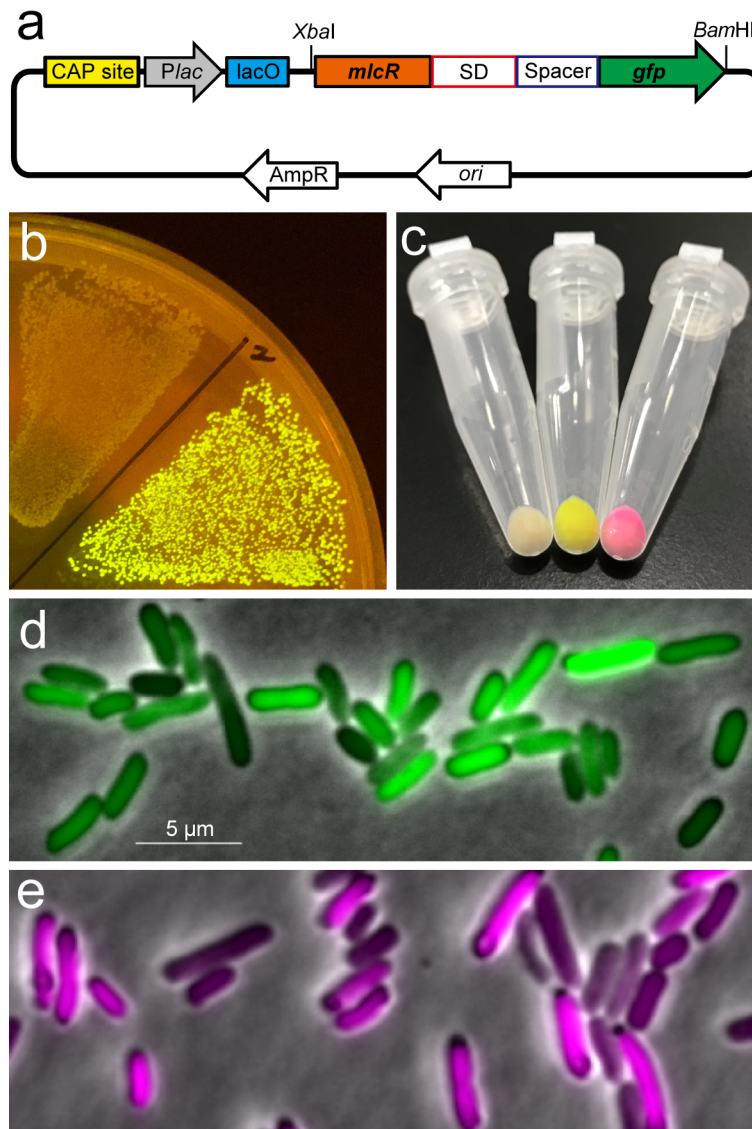


Figure 2

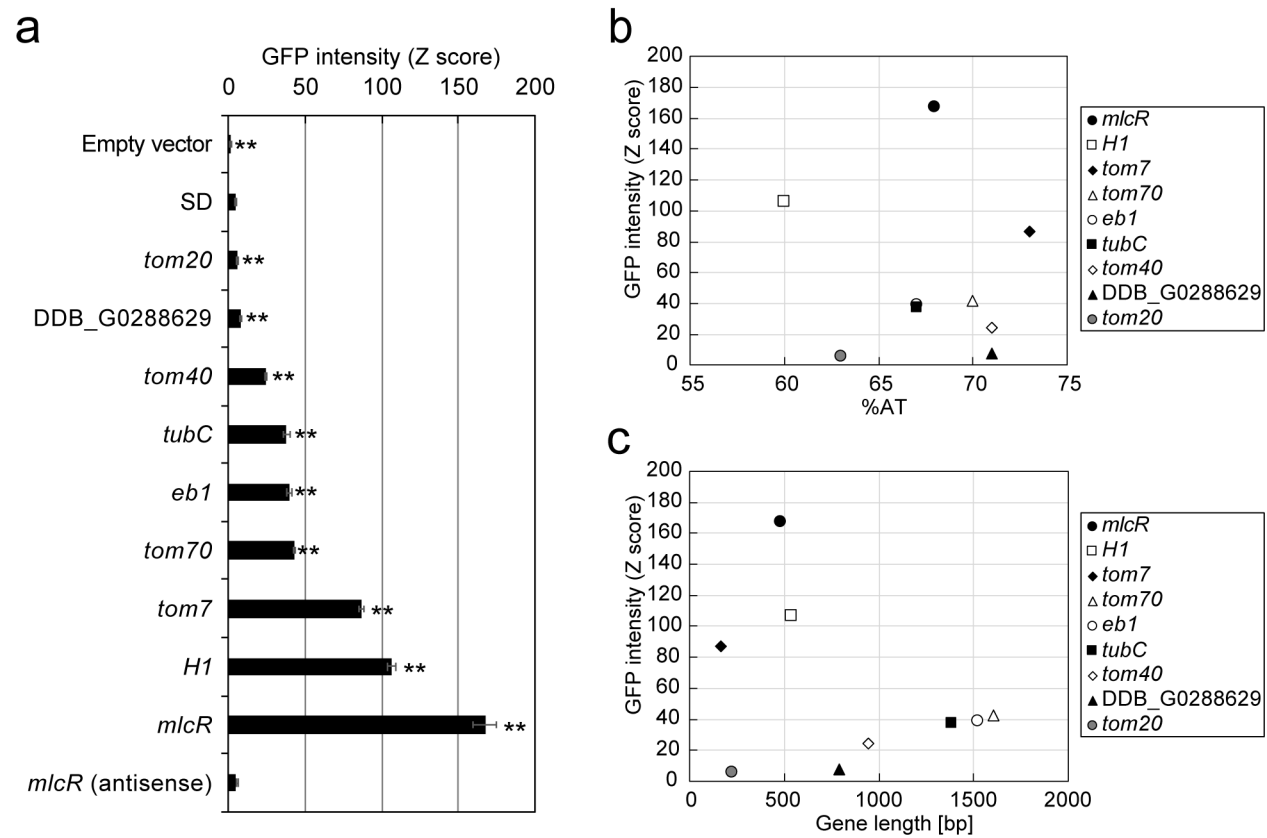
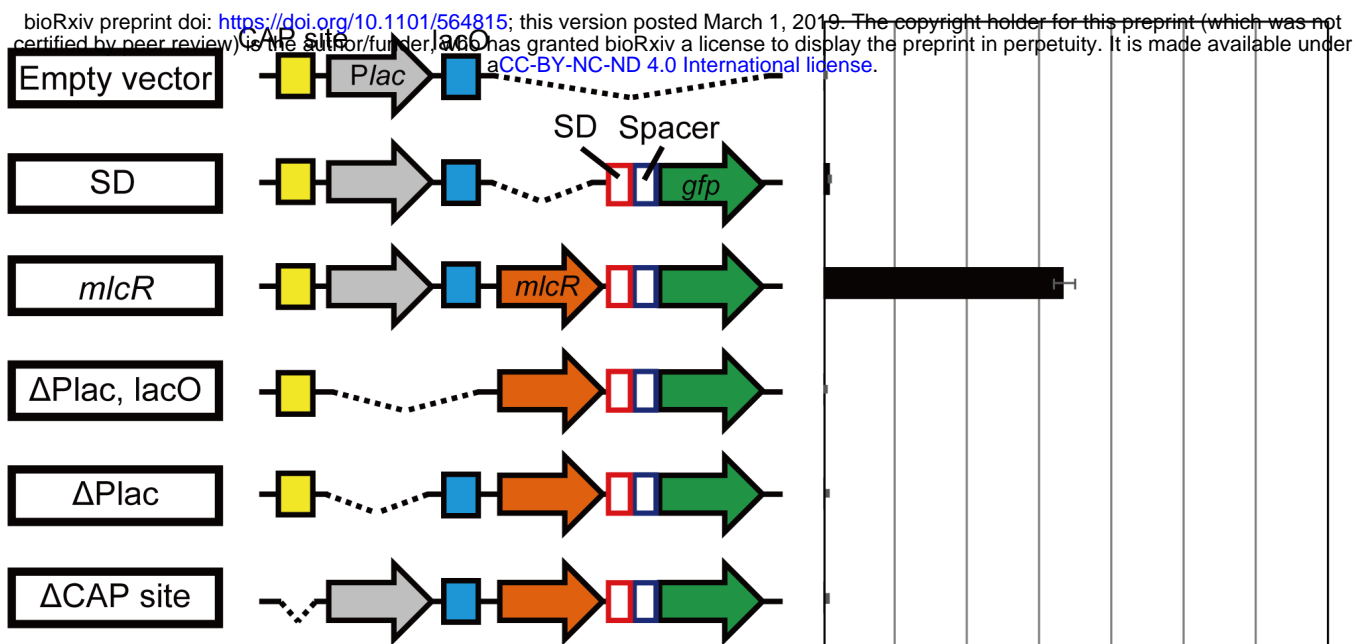
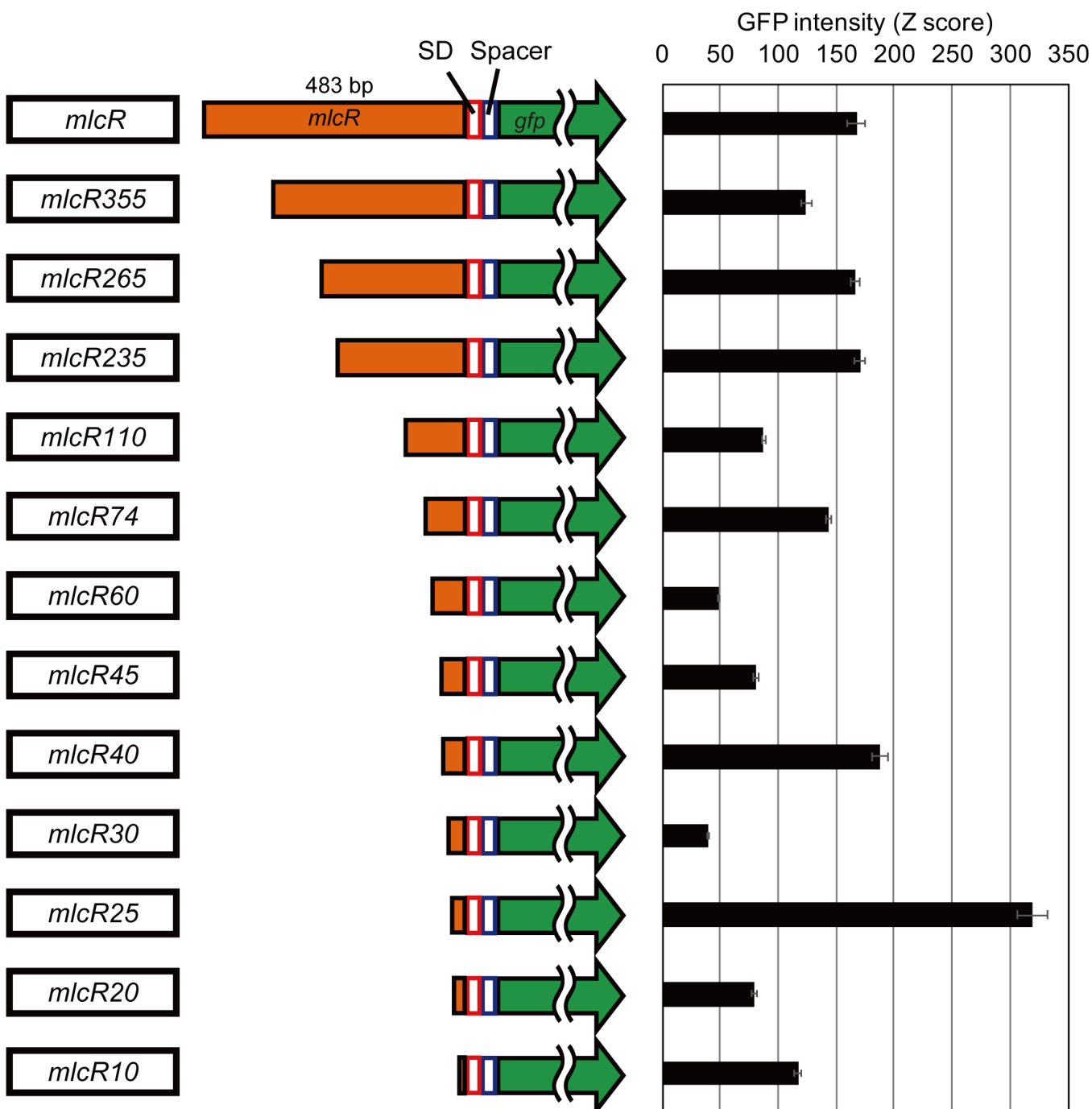


Figure 3

a



b



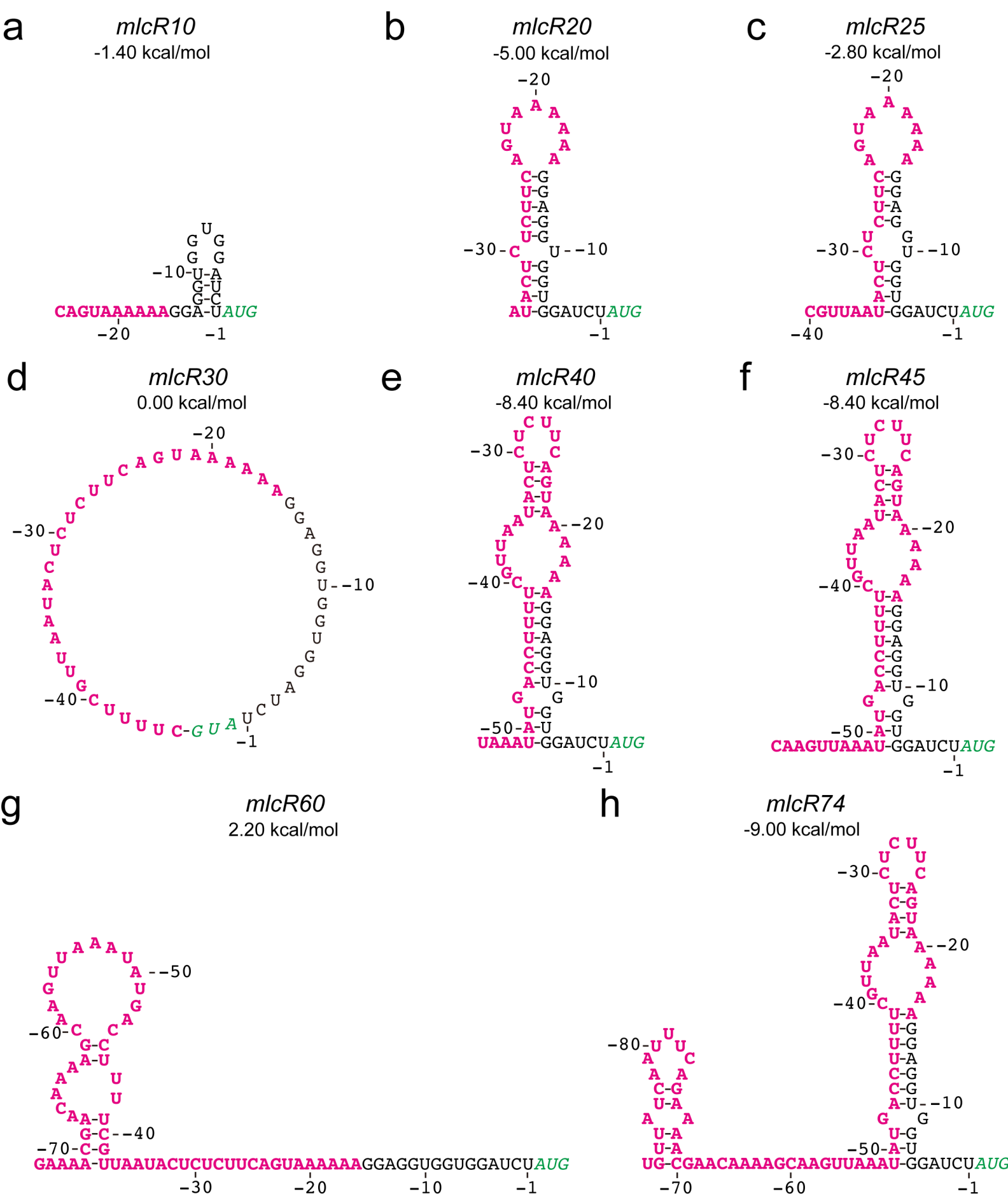


Figure 5

