Establishing a multivariate model for predictable antisense RNA-mediated repression

Young Je Lee[†], Soo-Jung Kim[†], Matthew B. Amrofell[†] and Tae Seok Moon*

Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, MO, 63130, USA

[†] These authors contributed equally to the work.

* To whom correspondence should be addressed.

Tae Seok Moon

One Brookings Dr., Box 1180

St. Louis, MO 63130, USA

Tel: +1 (314) 935-5026

Fax: +1 (314) 935-7211

Email: tsmoon@wustl.edu

ABSTRACT

Recent advances in our understanding of RNA folding and functions have facilitated the use of regulatory RNAs such as synthetic antisense RNAs (asRNAs) to modulate gene expression. However, despite the simple and universal complementarity rule, predictable asRNA-mediated repression is still challenging due to the intrinsic complexity of native asRNA-mediated gene regulation. To address this issue, we present a multivariate model, based on the change in free energy of complex formation (ΔG_{CF}) and percent mismatch of the target binding region, which can predict synthetic asRNA-mediated repression efficiency in diverse contexts. First, 69 asRNAs that bind to multiple target mRNAs were designed and tested to create the predictive model. Second, we showed that the same model is effective predicting repression of target genes in both plasmids and chromosomes. Third, using our model, we designed asRNAs that simultaneously modulated expression of a toxin and its antitoxin to demonstrate tunable control of cell growth. Fourth, we tested and validated the same model in two different biotechnologically-important organisms: *Escherichia coli* Nissle 1917 and *Bacillus subtilis* 168. Last, multiple parameters, including target locations, the presence of an Hfq binding site, GC contents, and gene expression levels, were revisited to define the conditions under which the multivariate model should be used for accurate prediction. Together, 434 different strain-asRNA combinations were tested, validating the predictive model in a variety of contexts, including multiple target genes and organisms. The result presented in this study is an important step towards achieving predictable tunability of asRNA-mediated repression.

Keywords: multivariate model; RNA regulator; antisense RNA; gene repression

Control of gene expression can be achieved by using versatile and programmable RNA regulators[1]. For example, small guide RNAs (sgRNAs) of the Type II clustered regularly interspaced short palindromic repeat (CRISPR) system of *Streptococcus pyogenes* have been repurposed for synthetic gene regulation[2, 3], small transcription activating RNAs (STARs) have been developed to activate transcription of target genes[4, 5], and toehold switches have been constructed to regulate translation through toehold-mediated RNA strand displacement[6, 7]. As discussed in a recent review paper[8], the synthetic biology community has also witnessed significant progress towards predictable gene expression control via RNA regulators. However, to potentially address many fundamental biological questions, the research community would need a generalizable model that enables quantitatively predictable gene regulation in diverse gene targets and organisms, as opposed to a single gene target or one strain only.

The primary allure of RNA regulators are their relatively simple structures and interaction modes. Their folding and interactions with target DNA or RNA molecules often follow straightforward base-pairing rules and thermodynamics. This structural and behavioral simplicity has enabled the development of computational tools to design diverse RNA regulators. For example, the NUPACK software was developed to analyze RNA secondary structure for systems involving multiple interacting strands[9]; the CRISPR-ERA software was developed for automated design of sgRNAs for gene activation or repression in nine different model organisms[2, 10]; nucleic acid design algorithms were combined with a simple STAR design motif to construct large libraries of STARs[5]; and a web tool was developed to custom-design toehold switches[11]. Furthermore, a computational algorithm was used to design riboregulators by considering only the free energy of complex formation and the activation energy of complex formation[12], a biophysical model was developed to predict the function of translation-regulating riboswitches[13], and both a

mechanistic modeling and an RNA folding simulations were used to predict the behavior of aptazymes[14]. These tools and approaches could facilitate the rapid creation of a large library of high-performing and orthogonal RNA regulators.

Efforts to rationally design synthetic antisense RNAs (asRNAs) have also been aided by algorithms and models that predict asRNA:mRNA interactions[9, 15, 16]. asRNAs are a well-studied class of RNA regulators that can modulate expression of target genes through asRNA:mRNA interaction[8, 15-20]. Notable models include the inTherAcc biophysical model that describes asRNA:RNA hybridization by incorporating a series of thermodynamic terms[15, 20], and a multivariate model that shows the hybridization energy of both the entire RNA duplex and the seed region as key determinants of asRNA specificity in the *E. coli* strain TOP10[16]. These models can assist in the design of asRNAs by considering the thermodynamics of RNA-RNA interactions. Despite these advances in computational modeling of asRNA:mRNA interactions, a comprehensive model that accurately predicts the asRNA-mediated repression levels in a variety of contexts is yet to be developed.

The hybridization of an asRNA to its mRNA target, typically binding to the 5' untranslated region (UTR) or possibly to the coding region of the target mRNA, results in efficient gene silencing via mRNA degradation or prevention of ribosomal access to the mRNA[21]. Furthermore, the regulatory impact of an asRNA can be improved by introducing an Hfq binding site on the 3'-end of the target binding region (TBR) of the asRNA, in which the Hfq binding site provides a scaffold for recruiting the Hfq protein[18, 22]. The Hfq protein is an RNA chaperone protein that is proposed to enhance the stability of asRNA by preventing its degradation and to facilitate the asRNA:mRNA interaction[23-25].

Despite the simple universal complementarity rules for asRNA:mRNA hybridization, predictably tunable asRNA-mediated repression is still challenging. In our previous work, we identified three quantitative design parameters for reliable asRNA-mediated repression: thermodynamics, mismatch, and length[22]. However, a predictive model for asRNA design that can be expanded to multiple target genes and organisms has yet to be developed. The aim of this study is to develop a data-driven model for the predictable tunability of asRNA-mediated repression by considering these parameters. First, the repression efficiencies of asRNAs were evaluated in *E. coli* DH10B[26]. TBR sequences were altered to modulate the binding affinity between the asRNA and target mRNA. Second, we developed a multivariate model and validated it in different genetic contexts (e.g., target genes in the genome or plasmids). Third, we expanded the predictive model to two different organisms: *E. coli* Nissle 1917 (EcN) and *Bacillus subtilis* 168 (*B. subtilis* 168). EcN is a probiotic strain that has been clinically characterized over the past decades[27], and *B. subtilis* 168 is a well characterized Gram-positive bacterium notable for its protein production and secretion abilities[28]. Last, multiple parameters, including target locations, the presence of an Hfq binding site, GC contents, and gene expression levels, were revisited to define the conditions under which the multivariate model should be used for accurate prediction. Together, 271 unique asRNA plasmids were built, and 434 different strain-asRNA combinations were tested, validating the predictability of the simple multivariate model. This study is an important step towards achieving predictable tunability of asRNA-mediated repression by providing the research community with the simple and predictive model.

RESULTS

Development of a multivariate model in *E. coli* DH10B

A two-plasmid system was built to understand and characterize asRNA-mediated repression in *E. coli* DH10B. The first plasmid (p15A origin, medium copy number) constitutively expresses red fluorescent protein (RFP) or green fluorescent protein (GFP) as a reporter. The second plasmid (ColE1 origin, high copy number) transcribes asRNA using the aTc-inducible P_{Tet} promoter. A high copy number was used to ensure that asRNA levels are in excess of target mRNA levels (i.e., *rfp* or *gfpmut3* mRNAs). These asRNAs consist of two parts: a target binding region (TBR) and an Hfq binding site. The TBR contains a sequence that is complementary to the target gene, and an Hfq binding site provides a scaffold for native Hfq protein recruitment (Figure 1A). In this study, the engineered MicF binding site (MicF M7.4)[29] was used because it performed well with low off-target effect in our previous studies[22, 30]. The transcription of asRNAs was terminated by both bacteriophage lambda T0 and rrnB T1 terminators to ensure complete transcription termination (see Supplementary Table 2 for the sequences).

We first determined the optimum inducer concentration for asRNA expression to achieve the highest repression efficiency. It was found that 250 ng/mL aTc resulted in the highest repression efficiency for the two tested asRNAs (69.5% for G4 and 77.2% for G18; Supplementary Figure S1A). Unless stated otherwise, cells were induced with 250 ng/mL aTc for subsequent experiments. Next, the asRNA-mediated gene repression was characterized by designing 24 asRNAs that target the 5' untranslated region (UTR) and seven coding regions of *rfp* mRNA as well as 30 asRNAs that target the 5' UTR and nine coding regions of *gfpmut3* mRNA. For each target region, three asRNAs with varied TBR length (16, 28, and 40 nucleotides with 0% mismatch) were tested, and the repression efficiencies were measured. Overall, increase in the TBR length generally increased repression efficiency (Supplementary Figures S1B and S1C). Different regions showed different repression efficiencies, demonstrating an apparent target location effect. All

asRNAs targeting the 5' UTR and the start codon region (referred to as the translation initiation region; TIR; -35 to +40 with the ATG start codon's A as +1) had high repression efficiency (an average of 64.4% for *rfp* mRNA and 62.4% for *gfp* mRNA; Supplementary Figures S1B and S1C) compared to asRNAs targeting the other coding regions (an average of 29.7% for *rfp* mRNA and 30.2% for *gfp* mRNA; Supplementary Figures S1B and S1C). This result is consistent with previous findings that asRNAs are most effective when they target the TIR[18, 31, 32]. Yet, many asRNAs targeting the other coding regions had high repression efficiency (in the case of TBR length of 28 and 40 nucleotides, but not 16 nucleotides; Supplementary Figures S1B and S1C). The difference in the rates of translation initiation and elongation, due to asRNAs binding to coding regions, could result in ribosome queuing and inhibition of translation[33, 34]. Both *rfp* and *gfpmut3* mRNAs contain an RBS sequence that would support a high translation initiation rate (18142 au for *rfp* and 27867 au for *gfpmut3* on the RBS Calculator v2.1 scale)[35-37]. On the other hand, ribosomes might be unable to efficiently move past the asRNA:mRNA duplex in the coding region because the mRNA must be in a single-stranded form to be translated[38]. While the N-terminal coding region has been systemically investigated regarding translation inhibition by using systematically designed RNA hairpins[37] or small noncoding RNA[39], we found that the other coding regions showed an inconsistent repression pattern and a target location effect when targeted by asRNAs (Supplementary Figures S1 and S2; discussed later in detail). To eliminate the target location effect on repression and select a region for reliable repression, the target region was restricted to the TIR of target mRNAs for the development of the model (Figure 1A).

Previously, we identified that thermodynamics, mismatch, and double-stranded RNA length are important design parameters for reliable asRNA-mediated repression[22]. Further analysis of previous and current data revealed that thermodynamics (change in free energy of complex

formation (ΔG_{CF})) and double-stranded RNA length had a high variance inflation factor (VIF=22.63), indicating the presence of strong multicollinearity. To remove the redundant effect, the double-stranded RNA length parameter was omitted from further analysis. Both ΔG_{CF} and percent mismatch had a low variance inflation factor (VIF=1.76), indicating the absence of multicollinearity. In other words, ΔG_{CF} and percent mismatch are not correlated in a regression model. Therefore, these two predictors were used for multiple linear regression analysis to create a model for predicting asRNA repression efficiency. ΔG_{CF} is the change in ΔG when the TBR binds to the mRNA and is calculated using the equation $\Delta G_{asRNA:mRNA} - \Delta G_{asRNA} - \Delta G_{mRNA}$, with free energy changes predicted by NUPACK[9] (see Methods). Percent mismatch is the total number of mismatch as a percentage of the total TBR length. Mismatches were introduced in a TBR sequence through base substitutions (A to T; G to C) as opposed to insertions or deletions.

A total of 69 asRNAs that target either *rfp* or *gfpmut3* mRNA were designed and tested (Figure 1B). TBR sequences were altered by varying ΔG_{CF} and percent mismatch to modulate the repression efficiencies. ΔG_{CF} was varied by increasing or decreasing the TBR length (8 to 40 nucleotides), and percent mismatch was varied by substituting 0 to 9 nucleotides in the TBR sequence. A number of studies have demonstrated the seed region or the first 7 nucleotides of small RNA is critical for gene silencing[2, 40-43]. Therefore, nucleotides were randomly substituted after the seed region (i.e., starting on the eighth nucleotide from the 5' end of TBR). From multiple linear regression analysis using 69 experimental data points (each in triplicate) and the SPSS Statistics package, a multivariate model was found:

$$F(X_1, X_2) = [0.3848 - 0.0068X_1 - 0.0125X_2 + \varepsilon] \ (R^2=0.685) \quad (1)$$

where F is the predicted repression efficiency, X_1 is ΔG_{CF} (in kcal/mol), X_2 is percent mismatch (in %), and ε is the standard error (ε=0.123; Figure 1B). All coefficients had p-values less than

0.001. Next, 20 additional asRNAs that target either *rfp* or *gfpmut3* mRNA were designed and tested to validate the predictive model. The model was to be considered valid if R^2 is greater than 0.5 with respect to the y=x line when measured repression efficiencies are compared to their predicted values[44]. The measured and predicted repression efficiencies were in agreement, validating the model (R^2=0.749; Figure 1C).

Validation of the model using chromosomal gene targets

An advantage of synthetic asRNA is its ability to regulate chromosomal gene targets without requiring chromosomal modification. Several studies have already utilized asRNAs to regulate genes expressed from the chromosome of the host organisms[18, 45-47]. Once the predictive model had been developed, we tested and validated the model in different genetic contexts (e.g., target genes in the genome). A cassette consisting of the BBa_J23105 constitutive promoter and *rfp* or a cassette consisting of the BBa_J23110 constitutive promoter and *gfpmut3* was independently integrated into the *E. coli* DH10B genome (*bglA*::*rfp* and *bglA*::*gfpmut3*, respectively; see Supplementary Table 2 for sequences) by using λ Red recombinase (see Supplementary Table 3 for the primers used)[48]. The resulting *E. coli* strain contains a functional copy of either *rfp* or *gfpmut3* in the chromosome, and RFP or GFP is constitutively expressed. These strains were transformed with the 69 asRNAs that had been tested to develop the multivariate model. Each of the asRNAs was expressed by the P_{Tet} promoter, and the fluorescence levels of induced and uninduced cells were compared to calculate measured repression. The measured and predicted repression efficiencies were generally in agreement (R^2=0.621 with respect to the y=x line; Figure 2A). Importantly, this data demonstrates that our model accurately

predicts asRNA-mediated repression of chromosomally-integrated genes as well as plasmid-encoded genes.

In addition to targeting chromosomally-integrated reporter genes, asRNAs were designed to target TIRs of native chromosomal genes. The first set of asRNAs were designed to repress *lysA*, which encodes for diaminopimelate (DAP) decarboxylase. DAP decarboxylase is the last enzyme of the L-lysine biosynthesis pathway, catalyzing decarboxylation of *meso*-DAP into L-lysine. L-lysine is one of the essential amino acids required for normal bacterial growth[49]. Each asRNA was expressed, and the absorbance of induced and uninduced cells were compared to calculate repression efficiency (see Methods). To determine repression efficiency from growth data, we first note that the L-lysine generation rate is linearly proportional to the enzyme concentration. This relationship was supported by *in vitro* and *in vivo* experiments in previous reports. For example, a proportionality relationship between the rate of decarboxylation and the amount of DAP decarboxylase was reported from an *in vitro* experiment[50], and a close correlation between increases in DAP decarboxylase specific activity and increases in intracellular lysine concentration was reported from an *in vivo* experiment[51]. Additionally, a linear relationship between lysine concentration and growth was reported using an *E. coli* lysine auxotroph[52]. Overall, a linear relationship among DAP decarboxylase concentration, L-lysine concentration, and growth allowed us to estimate repression efficiency of *lysA* using growth data obtained when cells were grown in minimal media lacking L-lysine. The measured and predicted repression efficiencies were generally in agreement, further validating the applicability of the model (R^2=0.502 with respect to the y=x line; Figure 2B).

The second set of asRNAs was designed to repress *uidA*, which encodes for β-glucuronidase. This enzyme hydrolyzes X-gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic

acid; colorless) to release glucuronic acid (colorless) and 4-chloro-bromo-indigo (4-CBI; a blue precipitate; Supplementary Figure S3). The relative expression level of β-glucuronidase can be estimated by measuring the absorbance of 4-CBI at 615 nm (i.e., the 4-CBI generation rate is proportional to the enzyme concentration; Supplementary Figure S3C). Thus, the measured absorbance can be used to determine repression efficiency. Each of the asRNAs was expressed, and the changes in absorbance at 615 nm of induced and uninduced cells were compared to calculate repression. Again, the measured and predicted repression efficiencies were generally in agreement (R^2=0.684 with respect to the y=x line; Figure 2C). The DAP-decarboxylase and β-glucuronidase repression results validate that the model that was generated using fluorescent protein repression data can also predict native gene repression.

Application of the model to control a toxin-antitoxin system

mazEF is a stress-induced toxin-antitoxin module responsible for programmed cell death in *E. coli*. MazF is a toxic endoribonuclease that cleaves single-stranded mRNAs at ACA sequences; MazE is an antitoxin that counteracts the lethal effect of MazF[53]. Thus, inhibition of MazE expression increases the intracellular MazF concentration, which leads to cell death[54]. The *mazEF* operon consists of two overlapping genes, *mazE* and *mazF* (i.e., the stop codon of *mazE* mRNA overlaps with the start codon of *mazF*; Figure 3A). A set of asRNAs was designed to bind to the TIR of *mazE* mRNA. All asRNAs were designed to bind to the start codon region (+1 to +40 with the ATG start codon's A as +1) or the predicted RBS region (-28 to -1). Any ACA sequence was excluded in the TBR of asRNA to avoid cleavage by MazF. For ten *mazE* mRNA-targeting asRNAs tested, a correlation analysis was performed, and a strong negative correlation between the predicted repression efficiency and absorbance at 600 nm was observed (R^2=0.791,

p<0.01; Figure 3B). In other words, cells expressing asRNAs with high predicted repression efficiencies of *mazE* had a lower cell density than cells expressing asRNAs with low predicted repression efficiencies of *mazE*.

After testing *mazE*-targeting asRNAs, the ME3 and M1 asRNAs that resulted in the largest growth defect were selected and expressed constitutively using the BBa_J23119 promoter (see Supplementary Table 2 for the sequence). Each *mazE*-targeting asRNA was co-expressed with a *mazF*-targeting asRNA partner to tunably recover growth. The ME3 asRNA binds to the start codon region, and the M1 asRNA binds to the predicted RBS region of the *mazE* mRNA (Figure 3C). The RBS calculator was used to predict the RBS region for both *mazE* and *mazF*[35, 36]. The second set of asRNAs that binds to the start codon region or the predicted RBS region of the *mazF* mRNA were designed and simultaneously expressed using the aTc-inducible P_{Tet} promoter, with the constitutively expressed *mazE*-targeting asRNA (ME3 or M1) (Figure 3C). The co-expression of asRNAs resulted in an increase in cell density compared to cells that express the *mazE*-targeting asRNA only (i.e., ME3 or M1 only; Figure 3C). A significant increase in absorbance was determined by two-sample *t*-test ($t = 50.56$, $p < 0.01$ for MF1B2 and the "ME3 only" control; $t = 6.21$, $p < 0.05$ for M2B2 and the "M1 only" control). Furthermore, cells expressing *mazF*-targeting asRNAs with higher predicted repression efficiencies had higher cell densities compared to cells expressing *mazF*-targeting asRNAs with lower predicted repression efficiencies (Figure 3C). These results show that asRNAs can be designed using the predictive model to modulate MazE and MazF expression, and thus to achieve tunable control of cell growth.

Expanding the model to different organisms

Experiments were designed to examine the behavior of asRNA in two different organisms. The first organism is *E. coli* Nissle 1917 (EcN), a probiotic strain which has been thoroughly analyzed by means of microbiological and molecular genetic methods[55, 56]. The genome sequence is available[57], and there have been studies to engineer EcN for therapeutic purposes[58, 59]. Importantly, EcN has an annotated *hfq* gene with an amino acid sequence identical to that of *E. coli* DH10B[60]. However, use of synthetic asRNA in EcN has yet to be explored. To this end, we first determined the aTc concentration (250 ng/mL) that led to the highest repression efficiency (65.5% for G46 and 71.0% for G47; Supplementary Figure S4A). Next, we tested 20 asRNAs that target either the *rfp* or *gfpmut3* mRNAs (plasmid-encoded genes) to validate the predictive model in EcN. Interestingly, the measured and predicted repression efficiencies were in agreement (R^2=0.734 with respect to the y=x line; Figure 4A). Lastly, we validated the model in a different genetic context by testing six additional asRNAs that target the *uidA* mRNA in EcN (native chromosomal gene). Again, the measured and predicted repression efficiencies were in agreement (R^2=0.638 with respect to the y=x line; Figure 4B). These results show that the predictive model developed in *E. coli* DH10B can expand to EcN.

Intrigued by the predictability of the model, we decided to expand it to another species. *B. subtilis* 168 is a well-characterized, Gram-positive bacterium, which has been widely utilized in biotechnology for protein production[61]. Although there have been a number of native asRNAs identified in *B. subtilis*, synthetic asRNA has not been used extensively as a regulator in this organism[62]. We placed asRNA under the control of the IPTG-inducible P_{spac} promoter and tested a range of IPTG concentrations to achieve the highest repression efficiency[63]. It was found that 1 mM IPTG resulted in the highest repression efficiency (53.7% for sfG1 and 71.8% for sfG2; Supplementary Figure S4B). We also investigated the effects of an Hfq binding site on repression.

The *B. subtilis* Hfq protein is a homolog of *E.coli* Hfq[64]. We extracted potential Hfq binding sites from the following three small RNAs: the SR4 antitoxin that pairs with the BsrG toxin mRNA; the FsrA small RNA, involved in the iron-sparing response; and the CsfG small RNA, highly conserved in *Bacillaceae*[64]. The Hfq binding sequences predicted from these RNAs were fused to the sfG2 TBR to determine whether they can be used as modular *B. subtilis*-specific Hfq binding sites (see Supplementary Table 2 for the sequences). MicC, Spot42, MicF, and MicF M7.4 (*E. coli* Hfq binding sites) were also tested for comparison[22]. The fusion did not affect the original secondary RNA structure of the sfG2 TBR, as predicted by NUPACK[9]. The pDG148 plasmid (~75 copies/cell) was used to ensure that asRNA levels are in excess of target mRNA levels[65]. All asRNAs were tested in *B. subtilis* SG13, which is a derivative of *B. subtilis* 168 carrying a functional copy of the superfolder green fluorescent protein (*sfgfp*) gene in the chromosome[66]. The inclusion of an Hfq binding site decreased the repression efficiency by 24.8% on average, compared to the absence of an Hfq binding site (Supplementary Figure S5). The decrease in the repression efficiency might be due to a functional limitation of the Hfq protein in *B. subtilis*. Though *B. subtilis* Hfq protein is a homolog of *E. coli* Hfq, it contains 29 fewer amino acids at its C-terminus, and this region has been proposed to be critical for Hfq:RNA interaction in bacteria[67]. As the sfG2 asRNA without an Hfq binding site had the highest repression efficiency, subsequently designed asRNAs contained no Hfq binding site. A total of 14 asRNAs that target the *sfgfp* mRNA were tested to validate the predictive model in *B. subtilis* SG13. To our surprise, the measured and predicted repression efficiencies were in agreement (R^2=0.788 with respect to the y=x line; Figure 4C). Next, we validated the model in a different genetic context by testing seven additional asRNAs that target the *lysA* mRNA (native chromosomal gene). Again, the measured and predicted repression efficiencies were in agreement (R^2=0.704 with respect to the y=x line; Figure 4D).

These results demonstrate that the model developed in *E. coli* DH10B can be applied to predict synthetic asRNA-mediated gene repression in *B. subtilis*.

DISCUSSION

In our previous study, we had identified ΔG_{CF}, percent mismatch, and double-stranded RNA length as three design parameters for reliable asRNA-mediated repression by investigating and selecting from six categories of parameters: target location, mismatch, length, thermodynamics, ribosome interaction, and YUNR motif (total 13 individual parameters including asRNA abundance and Hfq binding site)[22]. In the current work, we created and validated a simple two-parameter model (Figure 1B) with an assumption that ΔG_{CF} and percent mismatch are two independent variables that are linearly related to the repression efficiency (Supplementary Figure S2). The two variables had a low variance inflation factor (VIF=1.76), which is lower than 10, indicating that they can be simultaneously used as independent variables. Multiple linear regression analysis had been previously used to create a predictive model for asRNA-mediated repression in *E. coli* TOP10, although the two predictors (duplex and seed hybridization free energy) showed nonlinear relationship when each was separately plotted against percent repression[16]. Similarly, our two-parameter model was built using multiple linear regression analysis within the tested parameter ranges. In contrast, our results demonstrate that ΔG_{CF} and percent mismatch are reliable parameters for asRNA design that can predict asRNA-mediated repression of multiple target genes in different organisms (Figures 1-5). Because our model is a data-driven model, it should be used only within the tested parameter ranges for reliable prediction: ΔG_{CF} = -59.88 to -6.58 kcal/mol and percent mismatch = 0 to 32.1%. Additionally, this model should be used only for TIR-targeting asRNAs as discussed below.

It is worth noting that additional parameters (other than 13 parameters previously investigated) may affect asRNA-mediated repression. First, the target accessibility can be an important parameter for small regulatory RNAs[15, 68-71]. Notably, Vazquez-Anderson et al. created a biophysical model by deriving the accessibility-based thermodynamic model to describe asRNA-RNA hybridization in bacteria[15]. This model considers the overall change of Gibbs free energy ($\Delta G_{overall}$) as a predictor of asRNA binding. In this model, free energies of the local unfolding of the target region (ΔG_{Tf}; Tf denotes the target RNA folding) were considered when calculating the $\Delta G_{overall}$. Similarly, we considered ΔG_{CF} to be a combination of the accessibility to the target region (ΔG_{mRNA}), the penalty for breaking the structure of the asRNA (ΔG_{asRNA}), and the thermodynamic driving force for intermolecular base-pairing ($\Delta G_{asRNA:mRNA}$). Thus, our model accounts for the target accessibility. Using a cell-free assay, Shao et al. found that structural accessibility ($\Delta G_{disruption}$, which is conceptually same as ΔG_{mRNA}) was an important predictor for antisense oligos' activity[71]. When we used a fluorescence reporter in bacterial cells, however, there was no positive correlation between ΔG_{mRNA} and measured repression efficiency (R^2=0.134; Supplementary Figure S6), confirming that the combined thermodynamic parameter (ΔG_{CF}) is a better predictor for asRNA-mediated repression in bacterial cells.

Second, mRNA levels or translation initiation rates might affect repression efficiency because asRNAs compete with the ribosome for binding to mRNA. To investigate this possibility, we tested asRNAs while varying the *gfpmut3* mRNA level using the 3OC12-inducible P_{Las} promoter. We found that our model accurately predicted the asRNA-mediated repression levels regardless of the tested mRNA levels (Supplementary Figure S7). We also found that our model was generally accurate regardless of the tested translation initiation rates of the target mRNA (i.e. when tested using mRNAs with different RBS sequences; Supplementary Figure S8). As

demonstrated in our previous work[22], asRNA abundance would not be a determining factor in our model because the current experiments were also designed such that asRNA transcripts (expressed maximally from a higher copy plasmid) would be in excess of the target mRNAs (expressed from a lower copy plasmid or the genome). However, insufficient asRNA transcripts led to lower repression (Supplementary Figures S1A and S4), and for accurate prediction, our model should be used in a regime where asRNA transcripts would be in excess of the target mRNAs.

Third, we investigated whether the GC-content of the paired asRNA:mRNA sequence affected the predictability. Using 40 asRNAs (with fixed TBR length of 28 nucleotides with 0 – 32.1 % mismatch) that target either *rfp* or *gfpmut3* mRNA, we found that the two-parameter model accurately predicted the measured repression efficiency over the tested GC-content range (25.9 – 52.4% when only paired nucleotides are considered; Supplementary Figure S9). This is consistent with the previous report that GC content is a poor predictor of asRNA affinity[72].

Fourth, as demonstrated in our previous study[22], asRNAs lacking an Hfq binding site sequence showed significantly lower measured repression efficiency in *E. coli* than asRNAs containing MicF M7.4. To quantify the extent of this reduction, we performed multiple linear regression analysis using 23 experimental data points obtained from asRNAs that do not contain an Hfq binding site sequence (Supplementary Figure S10A). For this case, a different multivariate model was found:

$$F(X_1, X_2) = [0.1374 - 0.0071X_1 - 0.0047X_2 + \varepsilon] \ (R^2=0.665) \quad (2)$$

where F is the predicted repression efficiency, X_1 is ΔG_{CF} (in kcal/mol), X_2 is percent mismatch (in %), and ε is the standard error (ε=0.092). All coefficients had p-values less than 0.05. From Equations 1 and 2, we can obtain ΔF to quantify the extent of repression efficiency reduction. As

expected, when Equation 1 was applied to the same 23 asRNAs lacking an Hfq binding site sequence, we obtained a low R^2 value ($R^2 = 0.025$; Supplementary Figure S10B). Thus, Equation 1 should be used only for asRNAs containing MicF M7.4 in *E. coli*, while it should be used only for asRNAs lacking an Hfq binding site sequence in *B. subtilis* SG13. Investigating the effect of diverse Hfq binding site sequences and Hfq proteins on the model prediction in different organisms would be a great future study.

For reliable forward engineering, a model should accurately predict experimental outcomes. We analyzed the repression efficiencies of 168 asRNAs (including asRNAs tested in Figures 1C, 2, and 4) and found that those of 117 asRNAs (70%) are within a range of one standard error (1ε), those of 159 asRNAs (95%) are within a range of two standard error (2ε), and those of all asRNAs (100%) are within a range of three standard error (3ε) of the predicted repression efficiencies (Figure 5). To show whether our model can be applied to non-TIR targeting asRNAs, we analyzed the repression efficiency of 82 non-TIR targeting asRNAs. As expected, the measured repression efficiencies of only 20 asRNAs (24%) were within a range of one standard error (1ε), those of 44 asRNAs (54%) were within a range of two standard error (2ε), and those of 60 asRNAs (73%) were within a range of three standard error (3ε) of the predicted repression efficiencies (Supplementary Figure S11). In other words, our model cannot be applied to non-TIR targeting asRNAs.

To show whether we can create a model for non-TIR targeting asRNAs, we performed a multiple linear regression analysis using 56 experimental data points (only asRNAs targeting the non-TIR of *gfpmut3* mRNA). For this case, another multivariate model was found:

$$F(X_1, X_2) = [0.0435 - 0.0128X_1 - 0.0066X_2 + \varepsilon] \ (R^2=0.702) \quad (3)$$

where F is the predicted repression efficiency, X_1 is ΔG_{CF} (in kcal/mol), X_2 is percent mismatch (in %), and ε is the standard error (ε=0.157). All coefficients had p-values less than 0.001 (Supplementary Figure S12A). Next, we calculated predicted repression efficiencies of 18 asRNAs targeting the non-TIR of *rfp* mRNA using the model that was created with asRNAs targeting the non-TIR of *gfpmut3* mRNA (Equation 3). As expected, we observed a weak correlation between the measured and predicted repression efficiencies (R^2 = 0.187; Supplementary Figure S12B). Thus, Equation 3 should be used only for asRNAs targeting the non-TIR of *gfpmut3*. In other words, the model for non-TIR targeting asRNAs is gene-specific.

In bacteria, small noncoding RNAs generally bind to mRNAs in the 5' UTR to prevent translation initiation[73]. Bouvier et al. proposed the "five codon window" hypothesis that small noncoding RNAs can also inhibit 30S ribosomal subunit:mRNA complex formation by base-pairing with nucleotides within the first five codons[39]. Similarly, Espah Borujeni et al. determined that the ribosomal footprint, with which mRNA hairpins overlap and cause inhibition of translation initiation, extends 13 nucleotides into the coding region[37]. At the downstream positions past the fifth codon, small noncoding RNAs would not inhibit translational initiation, but endonucleolytic mRNA destabilization, facilitated by their binding to the coding region, could be an alternative gene-repression mechanism[73]. In our current study, we designed and tested synthetic asRNAs that target the mRNA coding sequence at diverse positions (Supplementary Figures S1 and S2). Although elongating 70S ribosomes have a strong RNA helicase activity, which can disrupt a perfect 27 base-pair helix[38], we found that asRNA-mediated repression could be achieved by asRNA:mRNA base pairing at most of the coding regions tested in this work. Interestingly, asRNAs with TBR length of 28 and 40 nucleotides (with 0% mismatch) repressed gene expression quite well in coding regions of mRNAs, while asRNAs with TBR length of 16 nucleotides (with

0% mismatch) had lower repression efficiencies compared to those targeting TIR (Supplementary Figure S1). Strong asRNA:mRNA binding at the coding region, which could facilitate[74] or could be assisted by[75] Hfq binding to mRNA, may induce ribosome stalling or RNA cleavage[73, 76], leading to gene repression. However, we found that it is difficult to predict repression efficiencies of non-TIR targeting asRNAs in multiple contexts (Supplementary Figure S12). When the non-TIR is targeted, asRNA-mediated repression might be affected by many factors other than ΔG_{CF} and percent mismatch. For example, a lower upstream elongation rate due to rare codons would permit asRNA binding more likely at the nearby downstream locations, leading to higher repression efficiencies at these regions, compared to the upstream regions occupied by elongating ribosomes. Because mRNA stability may play a bigger role in repression by non-TIR targeting asRNAs, recognition sites for diverse RNases should be collectively considered. Identifying such factors would be a great future study that would require systematically designed mRNAs, along with corresponding asRNAs that target different coding regions.

CONCLUSION

In the present study, we first created a data-driven model using two independent variables, ΔG_{CF} and percent mismatch, to predict asRNA-mediated repression levels in *E. coli* DH10B (Figure 1B). Next, we validated that the model can predict repression efficiencies of asRNAs targeting genes in the genome (Figure 2) as well as plasmid-encoded genes (Figure 1C). In addition, we designed asRNAs using the predictive model to effectively control the cell density by simultaneously modulating both MazE and MazF expression (Figure 3). Furthermore, we tested the predictive model in EcN and *B. subtilis*, thus showing its applicability in multiple organisms (Figure 4). Lastly, the validity of the multivariate model was challenged to define the conditions

under which it should be used for accurate prediction. We found that the model (Equation 1) should be used only for TIR-targeting asRNAs (Supplementary Figure S11) that contain MicF M7.4 in *E. coli* (Supplementary Figure S10) and no Hfq binding site in *B. subtilis* SG13 (Figures 4C-4D and Supplementary Figure S5), while other parameters, including target gene expression levels (Figures S7 and S8) and GC-contents of the paired asRNA:mRNA sequence (Supplementary Figure S9), did not affect the model predictability within the tested parameter range. Together, this work provides the research community with the simple, quantitative model for prediction of asRNA-mediated repression levels that was validated in diverse contexts, including multiple target genes and organisms (Figure 5). This model, along with our approach to create and validate a predictive model for synthetic asRNAs, will contribute to future forward-engineering efforts via predictably tunable RNA regulators.

METHODS

Strains and culture media

E. coli DH10B was used for developing and validating the predictive model. *E. coli* Nissle 1917 (EcN; DSM 6601 from Leibniz Institute DSMZ) and *Bacillus subtilis* SG13 (*B. subtilis* SG13; Bacillus Genetic Stock Center)[66] were used to confirm the applicability of the predictive model in different bacteria. *B. subtilis* SG13 contains a functional copy of superfolder green fluorescent protein (*sfgfp*) gene in the chromosome, and sfGFP is constitutively expressed by the P_{veg} promoter (see Supplementary Table 2 for the sequences). All plasmid constructions were done in *E. coli* DH10B. *E. coli* DH10B, EcN, and *B. subtilis* SG13 were grown in LB media (10 g/L tryptone, 10 g/L NaCl, and 5 g/L yeast extract) supplemented with the appropriate antibiotics: ampicillin (100

μg/mL), kanamycin (20 μg/mL), chloramphenicol (34 μg/mL), and spectinomycin (100 μg/mL). Minimal M9 medium (6.8 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, and 0.5 g/L NaCl) with 0.4% glucose and 0.8 mM L-leucine was used for the *E. coli* DH10B DAP decarboxylase repression experiment and toxin-antitoxin repression experiment. Minimal M9 medium with 0.5% glucose and 0.3% malate was used for the *B. subtilis* SG13 DAP decarboxylase repression experiment. Inducers were used at the following concentrations: aTc (anhydrotetracycline, 0-250 ng/mL), IPTG (isopropyl β-D-1-thiogalactopyranoside, 0-1 mM), and 3OC12 (*N*-3-oxododecanoyl-L-homoserine lactone, 0-1000 nM). All chemical reagents and inducers used in this study are from Sigma-Aldrich (St. Louis, MO, USA).

asRNA design and plasmid construction

TBR sequences were designed to bind to the TIR of the target mRNA sequence unless otherwise indicated. ΔG_{CF} was varied by increasing or decreasing the TBR length (8 to 40 nucleotides; -59.88 to -6.58 kcal/mol kcal/mol). ΔG_{CF} was calculated by using the equation $\Delta G_{asRNA:mRNA} - \Delta G_{asRNA} - \Delta G_{mRNA}$, and the change in free energy is predicted by NUPACK[9]. The change in free energy of asRNA-mRNA complex was estimated by entering the entire asRNA sequence (e.g. TBR, Hfq binding site, and terminator) and the target mRNA sequence (only the region that was specifically targeted plus one extra nucleotide at the 3'-end to consider stacking contributions of neighboring base pairs[15]), setting the nucleic acid type to RNA, and predicting the complex structure and its ΔG. This ΔG value was recorded as $\Delta G_{asRNA:mRNA}$. Instead of the entire mRNA sequence, only the target mRNA region was considered with the assumption of local folding due to the coupling of transcription and translation in bacteria[71]. Likewise, ΔG_{asRNA} and ΔG_{mRNA} were estimated by entering the entire asRNA sequence and the target mRNA sequence,

respectively. Mismatches were introduced in a TBR sequence through base substitutions (A to T; G to C) and percent mismatch was varied by randomly substituting 0 to 9 nucleotides in the TBR sequence (0 to 32.1% mismatch) starting on the eighth nucleotide from the 5' end of TBR. Percent mismatch is the total number of mismatch as a percentage of the total TBR length. The total number of mismatched nucleotides was recorded by counting the total number of unpaired nucleotides from the predicted structure, and the length of the TBR was recorded by counting the number of nucleotides in the TBR.

Each TBR sequence was placed under the control of the P_{Tet} (for *E.* coli) or P_{spac} promoter (for *B. subtilis* SG13) via either blunt end ligation as described previously[30] or Gibson assembly[77]. A terminator was included downstream of the Hfq binding site (lambda T0 and rrnB T1 terminators for *E. coli* DH10B and EcN; $t_1t_2t_0$ terminator for *B. subtilis* SG13). Hfq binding sites were inserted adjacent to the TBR sequence. For Figure 3C, plasmids co-expressing two asRNAs were constructed via the Golden Gate assembly method using AarI Type IIS restriction enzymes (Thermo Fisher Scientific, St. Peters, MO)[78]. For Figures 4C and 4D, asRNAs were cloned into *B. subtilis*-*E. coli* shuttle vector pDG148[63]. BBa_J23105 and BBa_J23110 constitutive promoters expressing *rfp* and *gfpmut3*, respectively, were from the Anderson promoter collection (http://parts.igem.org/Promoters/Catalog/Anderson; Supplementary Table 2). All plasmid sequences were verified by DNA sequencing (GENEWIZ, South Plainfield, NJ). All of the oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, IA). All DNA amplicons were purified with a DNA Clean and Concentrator Kit or a Zymoclean DNA Gel Purification Kit (Zymo Research, Irvine, CA). All plasmid minipreps were performed with a Zyppy Plasmid Miniprep Kit (Zymo Research, Irvine, CA). Information of all constructed

plasmids and sequences of genetic parts (i.e., genes, promoters, Hfq binding sites, terminators, and TBRs) are summarized in Supplementary Tables 1 and 2, respectively.

Chromosomal RFP or GFPmut3 reporter strain construction

Integration of genetic parts into the *E. coli* DH10B genome was done according to the reported recombineering strategy with some modification as described previously[30, 48]. A cassette consisting of a constitutive promoter (BBa_J23105 for *rfp*; BBa_J23110 for *gfpmut3*), RBS/nearby region, reporter gene (*rfp* or *gfpmut3*), and the BBa_B0015 terminator was integrated into the *E. coli* DH10B chromosome in the middle of the *bglA* gene (Supplementary Table 2). The kanamycin resistance gene and FLP recognition target (FRT) sites were amplified from pKD13 using the bglA-1F/bglA-1R primer set (Supplementary Table 3). The cassettes were amplified from the pRFP and pGfpmut3 plasmids using the bglA-2F/bglA-2R primer set (Supplementary Table 3). Both PCR-amplified DNA fragments were fused by the overlap extension PCR. The *E. coli* DH10B cells harboring the pKD46 plasmid were grown for 2 h in 5 mL LB media (100 µg/mL ampicillin and 10 mM arabinose; arabinose induces the λ Red recombinase) at 30°C, and the cells were prepared for electroporation. Cells were transformed with 100 ng of the overlap extension PCR product and grown overnight on LB agar plate with 20 µg/mL of kanamycin at 37°C. The pKD46 plasmid was cured during the cultivation on the plate. The kanamycin resistance gene in the genome was eliminated by expressing FLP recombinase (pCP20). The pCP20 plasmid was cured by culturing the cell for 16 h in LB media at 37°C without ampicillin.

Fluorescence and absorbance measurements

E. coli DH10B, EcN, and *B. subtilis* SG13 were transformed with asRNA and reporter plasmids, and incubated overnight at 37°C on LB agar plates containing appropriate antibiotics. Colonies were picked and grown overnight in 1 mL LB media with appropriate antibiotics at 37°C and 250 rpm (New Brunswick Excella E25 shaking incubator). Overnight cultures were subcultured (1% v/v) in fresh 1 mL LB media with appropriate antibiotics and grown for 2 h at 37°C and 250 rpm. The subcultures were transferred to fresh 0.6 mL LB media (1.67% v/v), supplemented with appropriate antibiotics and inducers (at concentrations as indicated in figure captions), and grown for 8 h at 37°C and 250 rpm. All cultures were grown in deep 96-well plates (Eppendorf, Hamburg, Germany). At the end of induction, cells were centrifuged to form a cell pellet, and the cell pellets were resuspended in 0.2 mL phosphate-buffered saline (pH 8.0). The population fluorescence ($F_{experimental}$) and the absorbance at 600 nm ($Abs_{experimental}$) were measured using a Tecan microplate reader (Infinite M200 Pro). The measured fluorescence was normalized by [$F_{norm} = (F_{experimental}/Abs_{experimental}) - (F_{negative\ control}/Abs_{negative\ control})$], where the negative control is *E. coli* DH10B without a plasmid, EcN without a plasmid, or *B. subtilis* 168 without a plasmid. Repression efficiencies were calculated by [$1-(F_{inducer+}/F_{inducer-})$], where $F_{inducer-}$ is the normalized fluorescence of the target gene without inducer and $F_{inducer+}$ is the normalized fluorescence with inducer. The GFPmut3 and sfGFP fluorescence was measured with excitation at 483 nm and emission at 530 nm. The RFP fluorescence was measured with excitation at 535 nm and emission at 620 nm.

For the *E. coli* DH10B DAP decarboxylase repression experiment (Figure 2B), *E. coli* DH10B harboring the asRNA plasmid was grown overnight in 1 mL LB media with ampicillin at 37°C and 250 rpm. Overnight cultures were pelleted (5 min at 4,000g), resuspended in an equivalent volume of minimal M9 medium (with 0.4% glucose, 0.8 mM L-leucine, and ampicillin),

transferred (1% v/v) into 1 mL of fresh minimal M9 medium (with 0.4% glucose, 0.8 mM L-leucine, and ampicillin), and grown for 2 h at 37°C and 250 rpm. The subcultures were transferred (0.5% v/v) to 0.2 mL of fresh minimal M9 medium (with 0.4% glucose, 0.8 mM L-leucine, and ampicillin), supplemented with 250 ng/mL aTc in flat bottom 96-well plate to measure their growth for 21 h in a Tecan microplate reader. The measured repression efficiency was reported by calculating $[1-(Abs_{aTc+}/Abs_{aTc-})]$, where the Abs_{aTc+} was the measured absorbance (at 600 nm) of the induced cells when the uninduced cells reached the Abs of 0.1 ($Abs_{aTc-} = 0.1$). For asRNA-mediated toxin and antitoxin repression experiment (Figure 3), the absorbance was measured after cells were induced for 15 h.

For the *B. subtilis* SG13 DAP decarboxylase repression experiment (Figure 4D), *B. subtilis* SG13 harboring the asRNA plasmid was grown overnight in 1 mL LB media with spectinomycin and kanamycin at 37°C and 250 rpm. Overnight cultures were pelleted (5 min at 4,000*g*), re-suspended in an equivalent volume of minimal M9 medium (with 0.5% glucose, 0.3% malate, spectinomycin, and kanamycin), transferred (1% v/v) into 1 mL of fresh minimal M9 medium (with 0.5% glucose, 0.3% malate, spectinomycin, and kanamycin), and grown for 8 hours at 37°C and 250 rpm. The subcultures were transferred (1% v/v) to 0.2 mL of fresh minimal M9 medium (with 0.5% glucose, 0.3% malate, spectinomycin, and kanamycin), supplemented with 1 mM IPTG in flat bottom 96-well plate to measure their growth for 21 h in a Tecan microplate reader. The measured repression efficiency was reported by calculating $[1-(Abs_{IPTG+}/Abs_{IPTG-})]$, where the Abs_{IPTG+} was the measured absorbance (at 600 nm) of the induced cells when the uninduced cells reached the Abs of 0.1 ($Abs_{IPTG-} = 0.1$).

Quantification of β-glucuronidase

E. coli DH10B or EcN harboring the asRNA plasmid was induced with 250 ng/mL aTc as described above. The induced cultures were pelleted (10 min at 2,200g) and resuspended in 0.2 mL phosphate-buffered saline (pH 8.0 for *E. coli* DH10B; pH 7.2 for EcN). After measuring the absorbance at 600 nm, the resuspended cells were diluted to an absorbance of 0.5 using phosphate-buffered saline (pH 8.0), and 0.2 mL of the diluted samples (containing the same number of cells) were centrifuged again (10 min at 2,200g). The pelleted cells were resuspended in 0.4 mL of 50 mM sodium phosphate monobasic, and 25 μL permeabilization solution (9:1 acetone to toluene (v/v)) was added into the resuspended cells. After incubation at 37°C for 1 h, enzymatic reactions were carried out at room temperature for 200 min by adding 5 μL of 50 mg/mL X-gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid; Thermo Fisher Scientific, MA) (Supplementary Figure S3). Absorbance was measured at 615 nm to quantify 4-chloro-bromo-indigo (a blue precipitate). The repression efficiency was reported by calculating $[1-(\Delta Abs_{aTc+}/\Delta Abs_{aTc-})]$, where the ΔAbs was the absorbance (Abs_{615}) change over 100 min for which Abs_{615} increased linearly (t = 0 - 100 min for *E. coli* DH10B; t = 100 - 200 min for EcN; Supplementary Figure S3). To determine the absorbance wavelength of 4-chloro-bormo-indigo (615 nm), the absorbance scanning was performed in the wavelength range between 560 and 710 nm for 25 min with 5 min interval (Supplementary Figure S3).

SUPPORTING INFORMATION

Supplementary Figures S1-S13 and Supplementary Tables 1-3 are available. The sequence, parameter, and repression data of asRNAs are available in an excel file.

ACKNOWLEDGEMENT

We thank Prof. Gautam Dantas for providing *B. subtilis-E.coli* shuttle vector pDG148. This work was supported by the National Science Foundation (MCB-1714352).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

References

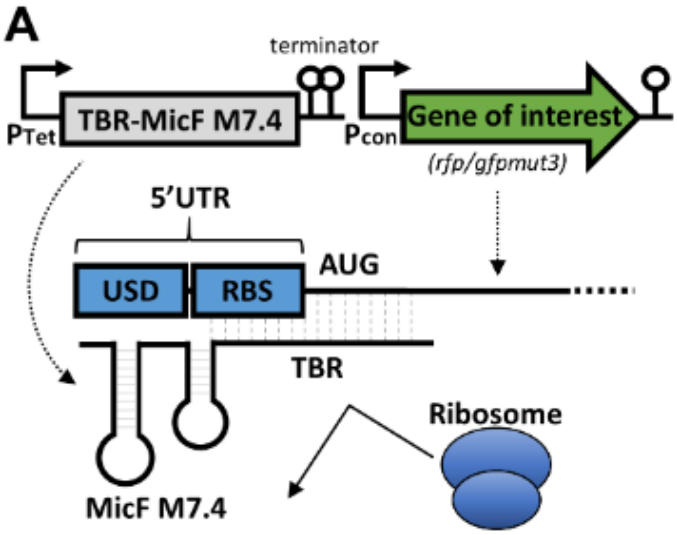
[1] Seo, S. W., and Jung, G. Y. (2013) Synthetic regulatory RNAs as tools for engineering biological systems: Design and applications, *Chem. Eng. Sci. 103*, 36-41.
[2] Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression, *Cell 152*, 1173-1183.
[3] Fontana, J., Dong, C., Ham, J. Y., Zalatan, J. G., and Carothers, J. M. (2018) Regulated Expression of sgRNAs Tunes CRISPRi in E. coli, *Biotechnol J. 13*, e1800069.
[4] Chappell, J., Takahashi, M. K., and Lucks, J. B. (2015) Creating small transcription activating RNAs, *Nat. Chem. Biol. 11*, 214-220.
[5] Chappell, J., Westbrook, A., Verosloff, M., and Lucks, J. B. (2017) Computational design of small transcription activating RNAs for versatile and dynamic gene regulation, *Nat. Commun. 8*, 1051.
[6] Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014) Toehold switches: de-novo-designed regulators of gene expression, *Cell 159*, 925-939.
[7] Green, A. A., Kim, J. M., Ma, D., Ilver, P. A. S., Collins, J. J., and Yin, P. (2017) Complex cellular logic computation using ribocomputing devices, *Nature 548*, 117-121.
[8] Lee, Y. J., and Moon, T. S. (2018) Design rules of synthetic non-coding RNAs in bacteria, *Methods 143*, 58-69.
[9] Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. (2011) NUPACK: Analysis and design of nucleic acid systems, *J. Comput. Chem. 32*, 170-173.
[10] Liu, H., Wei, Z., Dominguez, A., Li, Y., Wang, X., and Qi, L. S. (2015) CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation, *Bioinformatics 31*, 3676-3678.
[11] To, A. C., Chu, D. H. T., Wang, A. R., Li, F. C., Chiu, A. W., Gao, D. Y., Choi, C. H. J., Kong, S. K., Chan, T. F., Chan, K. M., and Yip, K. Y. (2018) A comprehensive web tool for toehold switch design, *Bioinformatics 34*, 2862-2864.
[12] Rodrigo, G., Landrain, T. E., and Jaramillo, A. (2012) De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells, *Proc. Natl. Acad. Sci. U.S.A. 109*, 15271-15276.
[13] Borujeni, A. E., Mishler, D. M., Wang, J. Z., Huso, W., and Salis, H. M. (2016) Automated physics-based design of synthetic riboswitches from diverse RNA aptamers, *Nucleic Acids Res. 44*, 1-13.
[14] Carothers, J. M., Goler, J. A., Juminaga, D., and Keasling, J. D. (2011) Model-Driven Engineering of RNA Devices to Quantitatively Program Gene Expression, *Science 334*, 1716-1719.

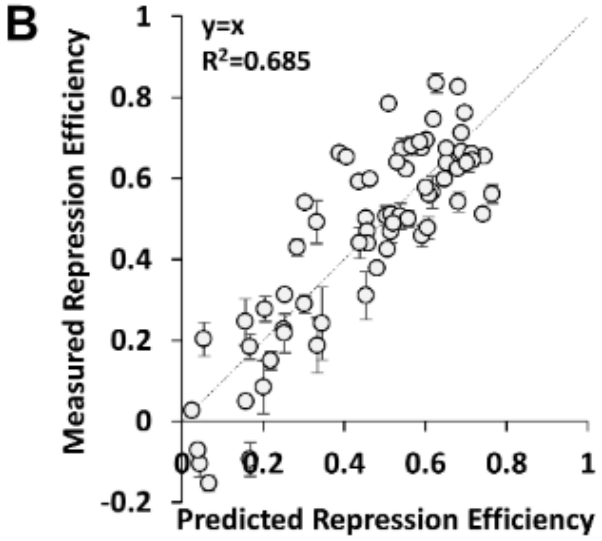
[15] Vazquez-Anderson, J., Mihailovic, M. K., Baldridge, K. C., Reyes, K. G., Haning, K., Cho, S. H., Amador, P., Powell, W. B., and Contreras, L. M. (2017) Optimization of a novel biophysical model using large scale in vivo antisense hybridization data displays improved prediction capabilities of structurally accessible RNA regions, *Nucleic Acids Res. 45*, 5523-5538.
[16] Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., and Arkin, A. P. (2012) Rationally designed families of orthogonal RNA regulators of translation, *Nat. Chem. Biol. 8*, 447-454.
[17] Lahiry, A., Stimple, S. D., Wood, D. W., and Lease, R. A. (2017) Retargeting a Dual-Acting sRNA for Multiple mRNA Transcript Regulation, *ACS Synth. Biol. 6*, 648-658.
[18] Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., and Lee, S. Y. (2013) Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs, *Nat. Biotechnol. 31*, 170-174.
[19] Rodrigo, G., Prakash, S., Shen, S. S., Majer, E., Daros, J. A., and Jaramillo, A. (2017) Model-based design of RNA hybridization networks implemented in living cells, *Nucleic Acids Res. 45*, 9797-9808.
[20] Leistra, A. N., Arnador, P., Buvanendiran, A., Moon-Walker, A., and Contreras, L. M. (2017) Rational Modular RNA Engineering Based on In Vivo Profiling of Structural Accessibility, *ACS Synth. Biol. 6*, 2228-2240.
[21] Thomason, M. K., and Storz, G. (2010) Bacterial Antisense RNAs: How Many Are There, and What Are They Doing?, *Annu. Rev. Genet. 44*, 167-188.
[22] Hoynes-O'Connor, A., and Moon, T. S. (2016) Development of Design Rules for Reliable Antisense RNA Behavior in E. coli, *ACS Synth. Biol. 5*, 1441-1454.
[23] Pulvermacher, S. C., Stauffer, L. T., and Stauffer, G. V. (2009) Role of the Escherichia coli Hfq protein in GcvB regulation of oppA and dppA mRNAs, *Microbiology 155*, 115-123.
[24] Moller, T., Franch, T., Hojrup, P., Keene, D. R., Bachinger, H. P., Brennan, R. G., and Valentin-Hansen, P. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction, *Mol. Cell 9*, 23-30.
[25] Masse, E., and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A. 99*, 4620-4625.
[26] Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., Burland, V., Mau, B., Petrosino, J. F., Qin, X., Muzny, D. M., Ayele, M., Gibbs, R. A., Csorgo, B., Posfai, G., Weinstock, G. M., and Blattner, F. R. (2008) The complete genome sequence of Escherichia coli DH10B: Insights into the biology of a laboratory workhorse, *J. Bacteriol. 190*, 2597-2606.
[27] Schultz, M. (2008) Clinical use of E-coli Nissle 1917 in inflammatory bowel disease, *Inflamm. Bowel Dis. 14*, 1012-1018.
[28] van Dijl, J. M., and Hecker, M. (2013) Bacillus subtilis: from soil bacterium to super-secreting cell factory, *Microb. Cell Fact. 12*, 3.
[29] Sakai, Y., Abe, K., Nakashima, S., Yoshida, W., Ferri, S., Sode, K., and Ikebukuro, K. (2014) Improving the Gene-Regulation Ability of Small RNAs by Scaffold Engineering in Escherichia coli, *ACS Synth. Biol. 3*, 152-162.
[30] Lee, Y. J., Kim, S. J., and Moon, T. S. (2018) Multilevel Regulation of Bacterial Gene Expression with the Combined STAR and Antisense RNA System, *ACS Synth. Biol. 7*, 853-865.
[31] Park, H., Yoon, Y., Suk, S., Lee, J. Y., and Lee, Y. (2014) Effects of different target sites on antisense RNA-mediated regulation of gene expression, *BMB Rep. 47*, 619-624.
[32] Yoo, S. M., Na, D., and Lee, S. Y. (2013) Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli, *Nat. Protoc. 8*, 1694-1707.
[33] Gorgoni, B., Marshall, E., McFarland, M. R., Romano, M. C., and Stansfield, I. (2014) Controlling translation elongation efficiency: tRNA regulation of ribosome flux on the mRNA, *Biochem. Soc. Trans. 42*, 160-165.
[34] Rodnina, M. V. (2016) The ribosome in action: Tuning of translational efficiency and protein folding, *Protein Sci. 25*, 1390-1406.
[35] Borujeni, A. E., Channarasappa, A. S., and Salis, H. M. (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.

[36] Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression, *Nat. Biotechnol. 27*, 946-950.
[37] Borujeni, A. E., Cetnar, D., Farasat, I., Smith, A., Lundgren, N., and Salis, H. M. (2017) Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences, *Nucleic Acids Res. 45*, 5437-5448.
[38] Takyar, S., Hickerson, R. P., and Noller, H. F. (2005) mRNA Helicase Activity of the Ribosome, *Cell 120*, 49-58.
[39] Bouvier, M., Sharma, C. M., Mika, F., Nierhaus, K. H., and Vogel, J. (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation, *Molecular cell 32*, 827-837.
[40] Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, *Cell 120*, 15-20.
[41] Storz, G., Vogel, J., and Wassarman, K. M. (2011) Regulations by Small RNAs in Bacteria: Expanding Frontiers, *Mol. Cell 43*, 880-891.
[42] Busch, A., Richter, A. S., and Backofen, R. (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions, *Bioinformatics 24*, 2849-2856.
[43] Tjaden, B., Goodwin, S. S., Opdyke, J. A., Guillier, M., Fu, D. X., Gottesman, S., and Storz, G. (2006) Target prediction for small, noncoding RNAs in bacteria, *Nucleic Acids Res. 34*, 2791-2802.
[44] Cohen, J. (1992) A power prime, *Psychological Bulletin 112*, 155-159.
[45] Noh, M., Yoo, S. M., Kim, W. J., and Lee, S. Y. (2017) Gene Expression Knockdown by Modulating Synthetic Small RNA Expression in Escherichia coli, *Cell Syst. 5*, 418-426.
[46] Solomon, K. V., Sanders, T. M., and Prather, K. L. J. (2012) A dynamic metabolite valve for the control of central carbon metabolism, *Metab. Eng. 14*, 661-671.
[47] Yang, Y. P., Lin, Y. H., Li, L. Y., Linhardt, R. J., and Yan, Y. J. (2015) Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products, *Metab. Eng. 29*, 217-226.
[48] Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, *Proc. Natl. Acad. Sci. U.S.A. 97*, 6640-6645.
[49] Li, X., and Ricke, S. C. (2003) Characterization of an Escherichia coli lysA insertion targeted mutant using phenotype arrays, *Bioresour. Technol. 89*, 249-253.
[50] White, P. J., and Kelly, B. (1965) Purification and Properties of Diaminopimelate Decarboxylase from Escherichia Coli, *Biochem. J. 96*, 75-84.
[51] Grandgenett, D. P., and Stahly, D. P. (1971) Control of Diaminopimelate Decarboxylase by L-Lysine during Growth and Sporulation of Bacillus-Cereus, *J. Bacteriol. 106*, 551-560.
[52] Li, X., and Ricke, S. C. (2003) Generation of an Escherichia coli lysA targeted deletion mutant by double cross-over recombination for potential use in a bacterial growth-based lysine assay, *Lett. Appl. Microbiol. 37*, 458-462.
[53] Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A. C., Engelberg-Kulka, H., and Moll, I. (2011) Selective Translation of Leaderless mRNAs by Specialized Ribosomes Generated by MazF in Escherichia coli, *Cell 147*, 147-157.
[54] Venturelli, O. S., Tei, M., Bauer, S., Chan, L. J. G., Petzold, C. J., and Arkin, A. P. (2017) Programming mRNA decay to modulate synthetic circuit resource allocation, *Nat. Commun. 8*, 15128.
[55] Blum-Oehler, G., Oswald, S., Eiteljorge, K., Sonnenborn, U., Schulze, J., Kruis, W., and Hacker, J. (2003) Development of strain-specific PCR reactions for the detection of the probiotic Escherichia coli strain Nissle 1917 in fecal samples, *Res. Microbiol. 154*, 59-66.
[56] Grozdanov, L., Zahringer, U., Blum-Oehler, G., Brade, L., Henne, A., Knirel, Y. A., Schombel, U., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., Rietschel, E. T., and Dobrindt, U. (2002) A single nucleotide exchange in the wzy gene is responsible for the semirough O6 lipopolysaccharide phenotype and serum sensitivity of Escherichia coli strain Nissle 1917, *J. Bacteriol. 184*, 5912-5925.
[57] Cress, B. F., Linhardt, R. J., and Koffas, M. A. (2013) Draft Genome Sequence of Escherichia coli Strain Nissle 1917 (Serovar O6:K5:H1), *Genome Announc. 1*, e00047-00013.

[58] Hwang, I. Y., Koh, E., Wong, A., March, J. C., Bentley, W. E., Lee, Y. S., and Chang, M. W. (2017) Engineered probiotic Escherichia coli can eliminate and prevent Pseudomonas aeruginosa gut infection in animal models, *Nat. Commun. 8*, 15028.
[59] He, L., Yang, H. J., Liu, F., Chen, Y. Y., Tang, S. J., Ji, W., Tang, J. L., Liu, Z. D., Sun, Y. J., Hu, S. B. A., Zhang, Y. M., Liu, X., Huang, W. T., Ding, X. Z., and Xia, L. Q. (2017) Escherichia coli Nissle 1917 engineered to express Tum-5 can restrain murine melanoma growth, *Oncotarget 8*, 85772-85782.
[60] Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Collado-Vides, J., Paley, S. M., Pellegrini-Toole, A., Bonavides, C., and Gama-Castro, S. (2002) The EcoCyc database, *Nucleic Acids Res. 30*, 56-58.
[61] Schumann, W. (2007) Production of recombinant proteins in Bacillus subtilis, *Adv. Appl. Microbiol. 62*, 137-189.
[62] Liu, Y. F., Zhu, Y. Q., Li, J. H., Shin, H. D., Chen, R. R., Du, G. C., Liu, L., and Chen, J. (2014) Modular pathway engineering of Bacillus subtilis for improved N-acetylglucosamine production, *Metab. Eng. 23*, 42-52.
[63] Stragier, P., Bonamy, C., and Karmazyncampelli, C. (1988) Processing of a Sporulation Sigma Factor in Bacillus-Subtilis - How Morphological Structure Could Control Gene-Expression, *Cell 52*, 697-704.
[64] Dambach, M., Irnov, I., and Winkler, W. C. (2013) Association of RNAs with Bacillus subtilis Hfq, *Plos One 8*, e55156.
[65] Espinar, L., Dies, M., Cagatay, T., Suel, G. M., and Garcia-Ojalvo, J. (2013) Circuit-level input integration in bacterial gene regulation, *Proc. Natl. Acad. Sci. U.S.A. 110*, 7091-7096.
[66] Guiziou, S., Sauveplane, V., Chang, H. J., Clerte, C., Declerck, N., Jules, M., and Bonnet, J. (2016) A part toolbox to tune genetic expression in Bacillus subtilis, *Nucleic Acids Res. 44*, 7495-7508.
[67] Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R., and Blasi, U. (2008) The C-terminal domain of Escherichia coli Hfq is required for regulation, *Nucleic Acids Res. 36*, 133-143.
[68] Abudayyeh, O. O., Gootenberg, J. S., Essletzbicher, P., Han, S., Joung, J., Belanto, J. J., Verdine, V., Cox, D. B. T., Kellner, M. J., Regev, A., Lander, E. S., Voytas, D. F., Ting, A. Y., and Zhang, F. (2017) RNA targeting with CRISPR-Cas13, *Nature 550*, 280-284.
[69] Tafer, H., Ameres, S. L., Obernosterer, G., Gebeshuber, C. A., Schroeder, R., Martinez, J., and Hofacker, I. L. (2008) The impact of target site accessibility on the design of effective siRNAs, *Nat. Biotechnol. 26*, 578-583.
[70] Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007) The role of site accessibility in microRNA target recognition, *Nat. Genetics 39*, 1278-1284.
[71] Shao, Y., Wu, Y., Chan, C. Y., McDonough, K., and Ding, Y. (2006) Rational design and rapid screening of antisense oligonucleotides for prokaryotic gene modulation, *Nucleic Acids Res. 34*, 5660-5669.
[72] Walton, S. P., Stephanopoulos, G. N., Yarmush, M. L., and Roth, C. M. (1999) Prediction of antisense oligonucleotide binding affinity to a structured RNA target, *Biotechnol. Bioeng. 65*, 1-9.
[73] Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C. D., and Vogel, J. (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation, *Nat. Struct. Mol. Biol. 16*, 840.
[74] Azam, M. S., and Vanderpool, C. K. (2018) Translational regulation by bacterial small RNAs via an unusual Hfq-dependent mechanism, *Nucleic Acids Res. 46*, 2585-2599.
[75] Wroblewska, Z., and Olejniczak, M. (2016) Hfq assists small RNAs in binding to the coding sequence of ompD mRNA and in rearranging its structure, *RNA 22*, 979-994.
[76] Ikeda, Y., Yagi, M., Morita, T., and Aiba, H. (2011) Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli, *Mol. Microbiol. 79*, 419-432.
[77] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nat. Methods 6*, 343-345.
[78] Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability, *PloS One 3*, e3647.

Figures


A
terminator
PTet
TBR-MicF M7.4
Pcon
Gene of interest
(rfp/gfpmut3)
5'UTR
USD
RBS
AUG
TBR
Ribosome
MicF M7.4


B
y=x
R²=0.685
Measured Repression Efficiency
Predicted Repression Efficiency
1
0.8
0.6
0.4
0.2
0
-0.2
0
0.2
0.4
0.6
0.8
1

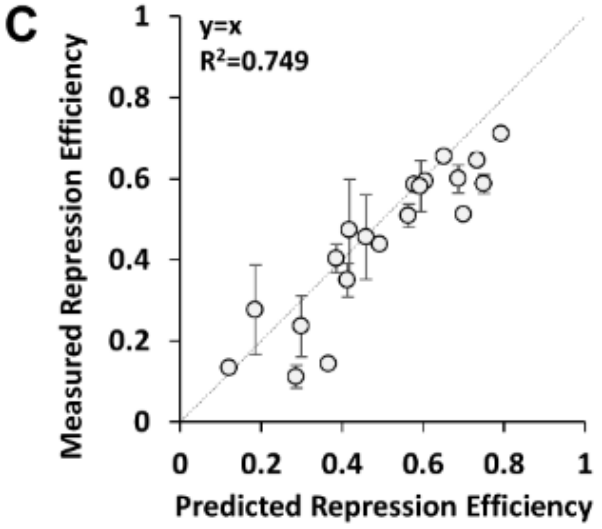

C
y=x
R²=0.749
Measured Repression Efficiency
Predicted Repression Efficiency
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1

Figure 1. Establishing a predictive model for asRNA-mediated gene repression. (A) Schematic of translational repression by asRNA composed of a target binding region (TBR) and an Hfq binding sequence (MicF M7.4)[29]. The target region was restricted to 75 nucleotides containing the upstream of Shine-Dalgarno sequence (USD), the ribosome binding site (RBS or the SD sequence), and the start codon of *rfp* or *gfpmut3* mRNA (-35 to +40 with the ATG start codon's A as +1). asRNAs are transcribed by the aTc-inducible P_{Tet} promoter (ColE1 origin, high copy number). *rfp* and *gfpmut3* are under the control of the constitutive BBa_J23105 and BBa_J23110 promoters, respectively (p15A origin, medium copy number; see Supplementary Table 2 for the sequences). **(B)** Multivariate model results shown as a scatter plot (predicted vs. measured repression efficiency). A total of 69 asRNAs targeting either *rfp* or *gfpmut3* mRNA were built, their repression efficiency was experimentally measured in *E. coli* DH10B, and the observed data was analyzed by fitting a linear equation using two independent variables ($F(X_1,X_2) = [0.3848 - 0.0068X_1 - 0.0125X_2 + \varepsilon]$, $R^2=0.685$, $p<0.001$). In this equation, F is the predicted repression efficiency, X_1 is ΔG_{CF}, X_2 is percent mismatch, and ε is the standard error. ΔG_{CF} is calculated using the equation $\Delta G_{asRNA:mRNA} - \Delta G_{asRNA} - \Delta G_{mRNA}$ (see Methods for details), and the change in free energy is estimated by NUPACK[9]. Percent mismatch is calculated by total number of mismatch as a percentage of the total TBR length as described previously[22]. **(C)** The model created in (B) was further validated by testing 20 newly designed asRNAs that target either *rfp* or *gfpmut3* mRNA in *E. coli* DH10B. For (B) and (C), cells were grown either in the absence or presence of 250 ng/mL aTc. The measured repression efficiency is reported by calculating $[1 - (F_{aTc+}/F_{aTc-})]$, where the F is the normalized fluorescence (Methods). The dashed line represents y=x. The error bars represent the standard deviation of the measured repression from three biological replicates performed on different days.

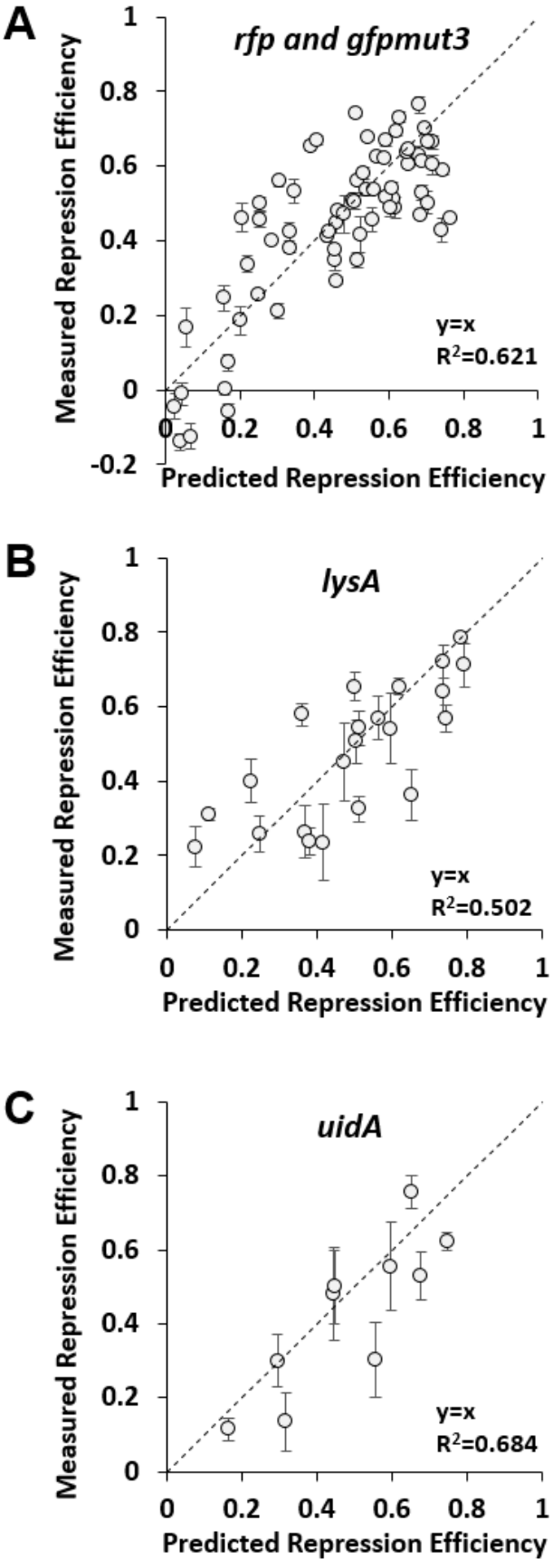

A
rfp and gfpmut3
Measured Repression Efficiency
Predicted Repression Efficiency
y=x
R2=0.621
B
lysA
Measured Repression Efficiency
Predicted Repression Efficiency
y=x
R2=0.502
C
uidA
Measured Repression Efficiency
Predicted Repression Efficiency
y=x
R2=0.684

Figure 2. Validating the model using asRNAs targeting genes in the genome. (A) Validation of the model with asRNAs targeting fluorescent reporter genes integrated into the genome. The *rfp* or *gfpmut3* cassette was integrated into the *E. coli* DH10B genome using the λ Red recombination method (*bglA*::*rfp*, constitutively expressing RFP; *bglA*::*gfpmut3*, constitutively expressing GFP)[48]. A total of 69 asRNAs from Figure 1B were tested, and their repression efficiency was calculated by $[1-(F_{aTc+}/F_{aTc-})]$, where the F is the normalized fluorescence (Methods). **(B, C)** Observed linear correlations between the measured and predicted repression efficiencies for 32 asRNAs targeting the *lysA* (B) or *uidA* (C) mRNAs. For (B), the measured repression efficiency is reported by calculating $[1-(Abs_{aTc+}/Abs_{aTc-})]$, where the Abs is the measured absorbance at 600 nm (Methods). For (C), the measured repression efficiency is reported by calculating $[1-(\Delta Abs_{aTc+}/\Delta Abs_{aTc-})]$, where the ΔAbs is the difference in the measured absorbance (615 nm) at t=100 min and t=0 min. See Supplementary Figure S3A for absorbance spectra of 4-chloro-bromo-indigo. All cells were grown either in the absence (aTc-) or presence (aTc+) of 250 ng/mL aTc. The dashed line represents y=x. The error bars represent the standard deviation of the measured repression from three biological replicates performed on different days.

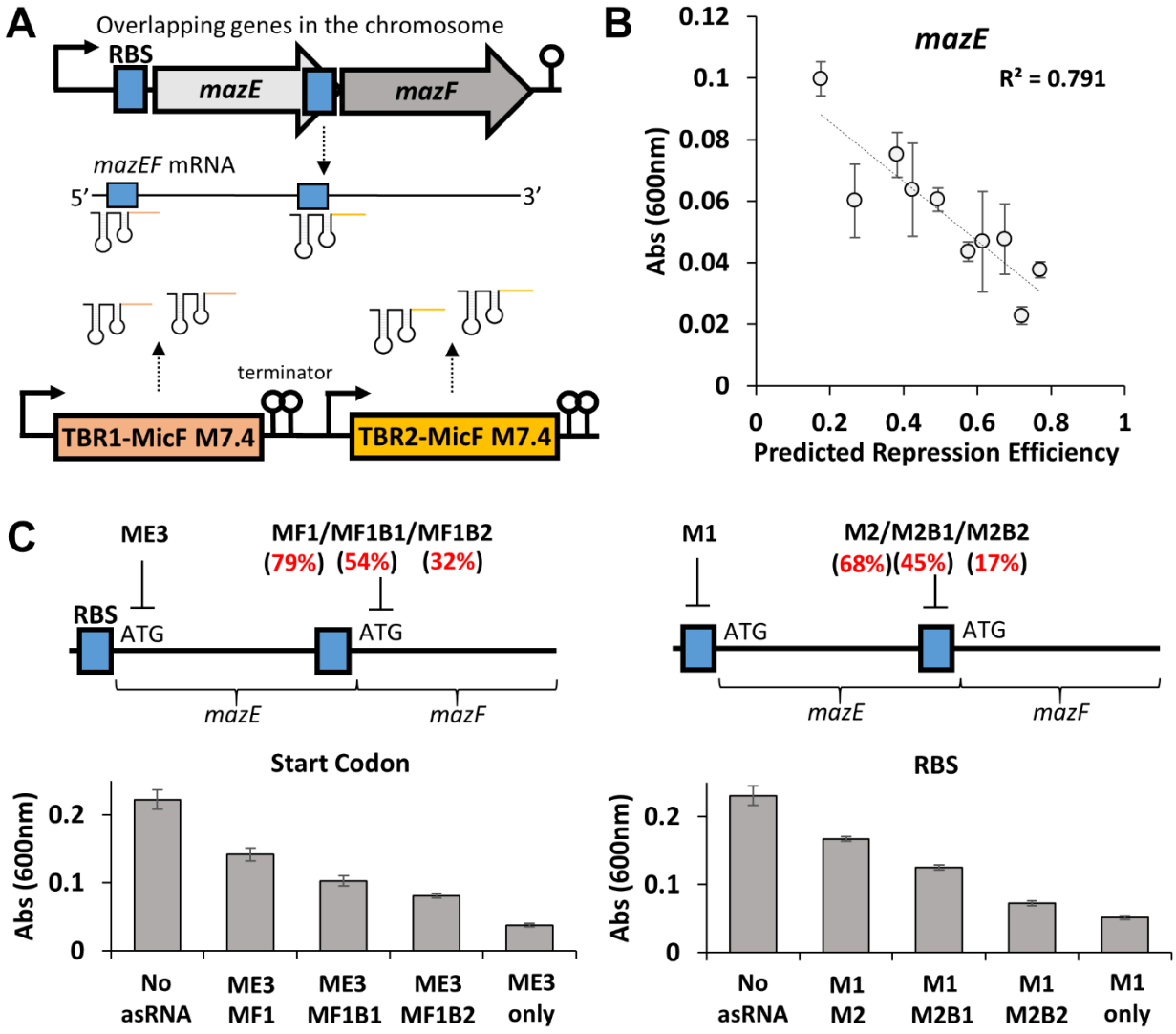


Figure 3. Controlling microbial growth with multiple asRNAs. (A) asRNAs can modulate cell growth by targeting *mazE* and *mazF* mRNAs. *mazEF* is a stress-induced toxin-antitoxin module (MazF is the toxin protein and MazE is the antitoxin protein), located on the chromosome of *E. coli*. Predicted RBSs are shown by blue boxes. asRNAs are transcribed by the aTc-inducible P_{Tet} promoter or the constitutive BBa_J23119 promoter on the same plasmid (ColE1 origin, high copy number). See Supplementary Table 2 for the sequences. **(B)** Observed negative correlation between the measured absorbance at 600 nm (Abs) and predicted repression efficiency for 10 *mazE*-targeting asRNAs. The absorbance was measured after cells were grown for 15 hours. All *mazE*-targeting asRNAs were transcribed by the aTc-inducible P_{Tet} promoter (250 ng/mL aTc). **(C)** Co-expression of the *mazE*- and *mazF*-targeting asRNAs. asRNAs were designed to bind to the start codon/downstream coding region (+1 to +40 with the start codon's A as +1; left figure) or the predicted RBS/nearby region (-28 to -1 with the start codon's A as +1; right figure). Predicted RBSs are shown by blue boxes. "No asRNA" for both graphs are cells with an empty vector. *mazF*-targeting asRNAs were expressed (induced by 250 ng/mL aTc), along with ME3 (left graph) or M1 (right graph) which was constitutively transcribed. The predicted repression efficiencies of

mazF-targeting asRNAs are indicated as % in parenthesis and highlighted in red. The absorbance was measured after cells were grown for 15 hours. The error bars represent the standard deviation of the measured repression from three biological replicates performed on different days.

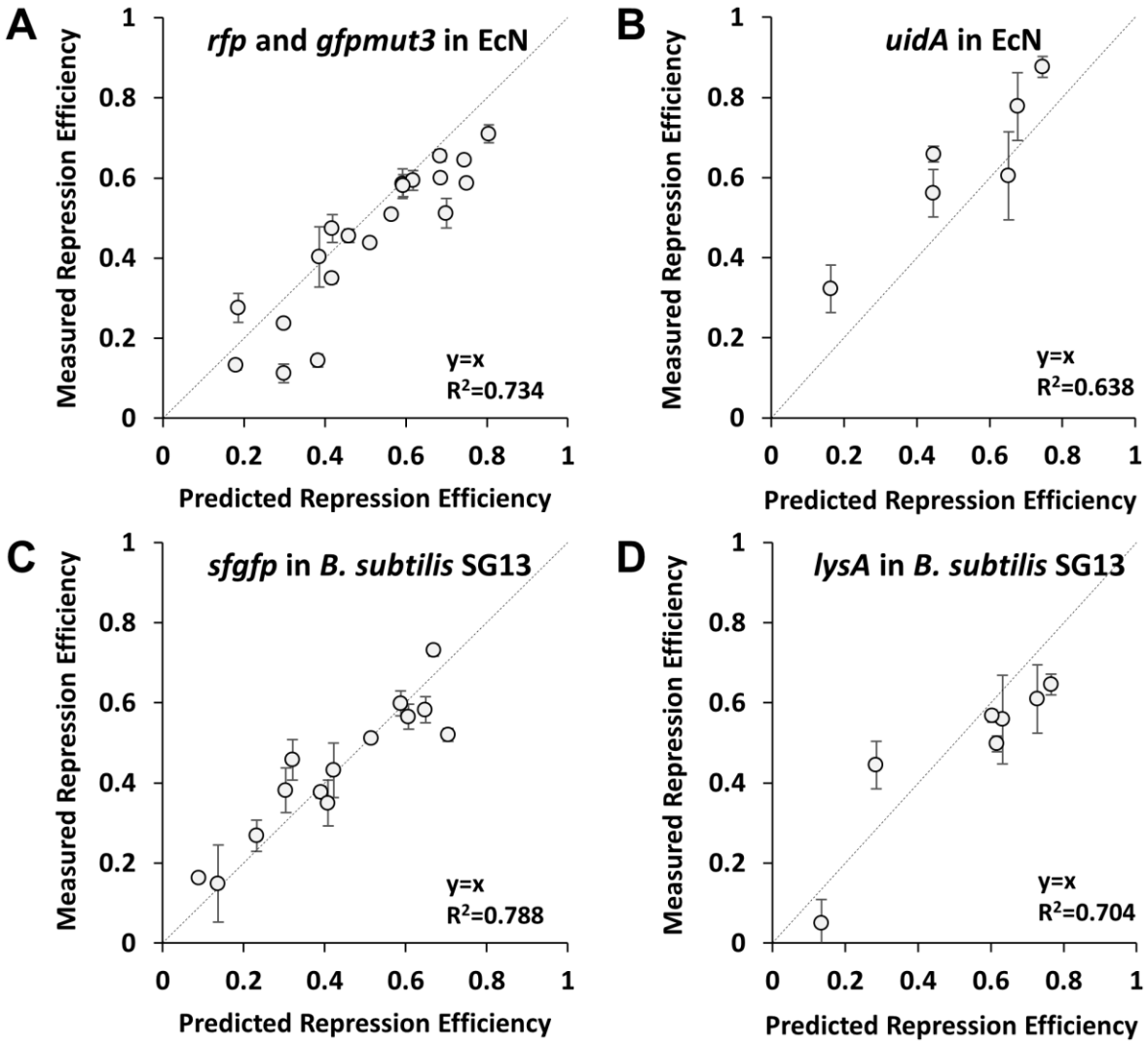
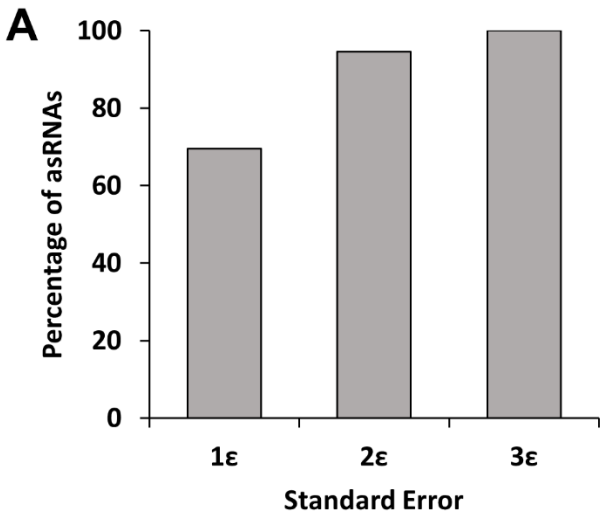


Figure 4. Expanding the predictive model to different organisms. (A) A total of 20 asRNAs that target TIR of either *rfp* or *gfpmut3* mRNA were tested in EcN. TBR was fused to MicF M7.4, and *rfp* and *gfpmut3* were plasmid-encoded genes. The dashed line represents y=x. All cells were grown either in the absence or presence of 250 ng/mL aTc. The measured repression efficiency is reported by calculating $[1-(F_{aTc+}/F_{aTc-})]$, where the F is the normalized fluorescence (Methods). **(B)** A linear correlation was observed between the measured and predicted repression efficiencies for asRNAs targeting the *uidA* mRNA in EcN. The dashed line represents y=x. All cells were

grown either in the absence or presence of 250 ng/mL aTc. The measured repression efficiency is reported by calculating [1−($\Delta Abs_{aTc+}/\Delta Abs_{aTc-}$)], where the ΔAbs is the difference in the measured absorbance (615 nm) at t=200 min and t=100 min. See Supplementary Figure S3A for absorbance spectra of 4-chloro-bromo-indigo. **(C)** A total of 14 asRNAs that target the *sfgfp* mRNA were tested in *B. subtilis* SG13. All TBRs were tested without an Hfq binding site. The dashed line represents y=x. All cells were grown either in the absence or presence of 1 mM IPTG. The measured repression efficiency is reported by calculating [1−(F_{IPTG+}/F_{IPTG-})], where the F is the normalized fluorescence (Methods). **(D)** A linear correlation was observed between the measured and predicted repression efficiencies for asRNAs targeting the *lysA* mRNA in *B. subtilis* SG13. All TBRs were tested without an Hfq binding site. The dashed line represents y=x. All cells were grown either in the absence or presence of 1 mM IPTG. The measured repression efficiency is reported by calculating [1−(Abs_{IPTG+}/Abs_{IPTG-})], where the Abs is the measured absorbance at 600 nm (Methods). The error bars represent the standard deviation of the measured repression from three biological replicates.



B

	1ε	2ε	3ε
Figure 1C	80.0	100.0	100.0
Figure 2A	68.1	89.9	100.0
Figure 2B	54.5	95.5	100.0
Figure 2C	60.0	90.0	100.0
Figure 4A	80.0	100.0	100.0
Figure 4B	33.3	100.0	100.0
Figure 4C	85.7	100.0	100.0
Figure 4D	85.7	100.0	100.0

Figure 5. Accuracy of the predictive model. (A) Accuracy of the model was reported by calculating the percentage of asRNAs whose measured repression efficiencies were within the standard error range of the predicted repression efficiencies. The standard error is from the model in Figure 1B (1ε=0.123, 2ε=0.246, and 3ε=0.369). **(B)** The table summarizes the percentage of asRNAs whose measured repression efficiencies were within the standard error range of the predicted repression efficiencies for each figure.

Graphical Table of Contents