Model-based design of synthetic antisense RNA for predictable gene repression

Tae Seok Moon

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

Correspondence should be addressed to:

Tae Seok Moon

One Brookings Dr., Box 1180

St. Louis, MO 63130, USA

Tel: +1 (314) 935-5026

Email: tsmoon@wustl.edu

Abstract

Our enhanced understanding of RNA folding and function has increased the use of small RNA regulators. Among these RNA regulators, synthetic antisense RNA (asRNA) is designed to contain an RNA sequence complementary to the target mRNA sequence, and the formation of double-stranded RNA (dsRNA) facilitates gene repression due to dsRNA degradation or prevention of ribosome access to the mRNA. Despite the simple complementarity rule, however, predictably tunable repression has been challenging when synthetic asRNAs are used. Here, the protocol for model-based asRNA design is described. This model can predict synthetic asRNA-mediated repression efficiency using two parameters: the change in free energy of complex formation (ΔG_{CF}) and percent mismatch of the target binding region (TBR). The model has been experimentally validated in both Gram-positive and Gram-negative bacteria as well as for target genes in both plasmids and chromosomes. These asRNAs can be created by simply replacing the TBR sequence with one that is complementary to the target mRNA sequence of interest. In principle, this protocol can be applied to design and build asRNAs for predictable gene repression in various contexts, including multiple target genes and organisms, making asRNAs predictably tunable regulators for broad applications.

Key Words: synthetic biology; predictive model; RNA regulator; antisense RNA; gene repression

Running Head: Design of antisense RNA for predictable gene repression

1 Introduction

In many biological studies and biotechnological applications, gene regulation has been achieved using regulatory proteins [1], including transcription factors and sigma factors [2-6]. These proteins are repurposed to perform their native function, and their use generally leads to large dynamic ranges in synthetic circuits. Despite this advantage, protein regulators have several drawbacks. First, it is challenging to design new protein regulators with novel functions due to the difficulty in predicting protein structure and function. Second, protein regulators must be transcribed and translated, costing cellular resources. Third, many protein regulators interact with a target DNA sequence, meaning that the target sequence should be introduced into the chromosome to regulate chromosomal genes. To simultaneously regulate multiple chromosomal genes, genome engineering often requires cumbersome experimental procedures, especially in non-model organisms, and can sometimes cause undesirable context effects on neighboring genes [7,8].

As an alternative strategy, RNA regulators have been developed, including small transcription activating RNA [9], attenuator [10], CRISPR interference [11], CRISPR activation [12], toehold switch [13], synthetic trans-acting antisense RNA (asRNA) [14,15], riboswitch [16,17], RNA thermosensor [18], and aptazyme [19,20]. Among these RNA regulators, asRNA contains a sequence complementary to the target mRNA sequence, and its binding to the target mRNA leads to gene repression through the prevention of ribosome access to the mRNA or RNA degradation [21-24]. Additionally, asRNA has been repurposed to control other RNA regulators [25,26] and combined with protein regulators to build complex genetic circuits [27,28].

As a regulator, asRNA has many advantages over proteins. First, asRNA is easy to design. By simply transcribing the reverse complement of the target mRNA, a new regulator can be generated. Second, because of this simple base-pairing mechanism, asRNAs are generally orthogonal [27], such that a given regulator usually represses its target gene only. Third, unlike many RNA regulators that require specifically-designed targets [9,13,16-20] and thus modification of the chromosome, the asRNA sequence is tailored specifically to the existing DNA sequence. This design strategy for asRNA does not require modification of the chromosome when chromosomal genes are the targets, making asRNA ideal for regulating multiple chromosomal genes at the same time. Fourth, like other RNA regulators, asRNA is not translated into a protein, potentially reducing cellular costs. Finally, asRNAs have been found in many microbes [29-42], allowing them to be developed as regulators with broad applicability in a diverse set of hosts.

Despite these advantages, asRNA had been used in a relatively limited number of species as a synthetic tool to regulate gene expression. This scarcity was partly due to the difficulty in achieving predictably tunable repression of target genes. Recently, a mathematical model has been developed and experimentally validated for predictable asRNA-mediated repression in diverse organisms, including *Escherichia coli* DH10B, *E. coli* Nissle 1917, and *Bacillus subtilis* 168 [15]. This prediction model is based on two design parameters that have been identified by Hoynes-O'Connor and Moon [27]. One parameter is a change in free energy of complex formation (ΔG_{CF}), and the other is a percent mismatch. ΔG_{CF} is the change in ΔG when the target

binding region (TBR) of an asRNA binds to the mRNA. It is calculated using the equation $\Delta G_{asRNA:mRNA} - \Delta G_{asRNA} - \Delta G_{mRNA}$, where each ΔG is estimated using NUPACK [43] as described in the Methods section in detail. The percent mismatch is calculated by dividing the number of mismatched nucleotides by the total number of the TBR nucleotides. Notably, ΔG_{CF} and percent mismatch are not correlated, allowing these two predictors to be used for multiple linear regression analysis. Using 434 different strain-asRNA combinations, this multivariate model has been built and experimentally validated, enabling predictable tunability of asRNA-mediated repression [15].

In this protocol, the detailed procedures for the design and construction of asRNAs are described. Because various gene cloning methods can be used to construct plasmids that contain synthetic asRNA expression cassettes, this protocol focuses on the detailed steps for the model-based asRNA design instead of detailed cloning steps. An asRNA sequence consists of TBR, the Hfq binding site (MicF M7.4), and a transcription terminator. MicF M7.4 was shown to facilitate high asRNA-mediated repression with minimal off-target effect [15,27]. As long as the described design rules are correctly followed, predictable asRNA-mediated repression can be achieved. The construction of a plasmid itself is relatively simple, only requiring the replacement of an existing TBR with one that is complementary to the target mRNA sequence while retaining the rest sequence of the platform plasmid (*see* the Materials and Methods sections) [15].

2 Materials

2.1 Materials for asRNA Design

1. The target mRNA (or the corresponding DNA) sequence (e.g., 5'TAGCGAATTCACTTATTAAAGAACAGGAGTAAGTAATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAA3').

2. The TBR (or the corresponding DNA) sequence (e.g., 5'CTCCTTTTCTCATTTCTTACTCCTCTTCTTTAATAAGTGA3').

3. The asRNA (or the corresponding DNA) sequence (e.g., 5'CTCCTTTTCTCATTTCTTACTCCTCTTCTTTAATAAGTGACGTCCCGCAAGGATGCGGGTCTGTTTACCCCTATTTCAACCGGCCGCCTCGCGGCCGGTTTTTTTTTGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCA3').

4. NUPACK [43], an RNA secondary structure prediction software.

2.2 Materials for asRNA Plasmid Construction

1. pG16, the platform plasmid (ColE1 origin and ampicillin resistance; 3199 base pairs) [15] that transcribes asRNA using the aTc-inducible P_{Tet} promoter (aTc, anhydrotetracycline). This plasmid contains the asRNA sequence that includes the original TBR (5'ATAAGTGAATTCGCTA3'), the MicF M7.4 Hfq binding site, and the transcription terminator.

2. Forward asRNA primer (e.g.,

5’TCCTCTTCTTTAATAAGTGACGTCCCGCAAGGATGC3’).

3. Reverse asRNA primer (e.g.,

5’GTAAGAAATGAGAAAAGGAGAGATGTGCTCAGTATCTCTATCACTGATAG3’).

4. Phusion DNA polymerase and 5X Phusion HF Buffer.

5. 10 mM dNTPs.

6. Dimethyl sulfoxide (DMSO).

7. DpnI, T4 polynucleotide kinase (T4 PNK), T4 DNA ligase, and 10X ligation buffer.

8. Gel extraction kit, PCR purification kit, and plasmid miniprep kit.

9. *E. coli* DH10B competent cells.

10. LB agar plates with ampicillin (100 µg/mL) and LB liquid media with ampicillin (100 µg/mL).

11. Purified water (DNase and RNase free, ddH_2O).

12. Sequencing primer (5’CGACCTCATTAAGCAGCTCTAATG3’).

13. SYBR Safe DNA Gel Stain, 1 kb DNA ladder, and gel loading dye.

14. Agarose.

2.3 Materials for GFP Repression Assays

1. Test strain with the target gene (e.g., *E. coli* DH10B variant strain containing *gfpmut3* in the chromosome, called *E. coli* DH10B-GFP [15]).

2. *E. coli* DH10B (no-GFP control).

3. LB agar plates without and with ampicillin (100 μg/mL) and LB liquid media without and with ampicillin (100 μg/mL).

4. Anhydrotetracycline (aTc).

5. Phosphate-buffered saline (pH 8.0).

6. Kanamycin (2 mg/mL) in phosphate-buffered saline (pH 8.0).

7. Deep 96-well plate and breathable sealing membrane.

3 Methods

In this Methods section, the *gfpmut3* gene integrated into the *E. coli* chromosome is used as an example target gene. The same design procedure can be applied to repress other genes in other species (both chromosomal genes and plasmid-encoded genes), as demonstrated in the previous report [15].

3.1 Model-based asRNA Design

Various asRNAs with different predicted repression efficiencies can be designed. Here, one asRNA with 56.6% predicted repression efficiency is designed as an example (Fig. 1).

1. Select the target mRNA sequence in the translation initiation region (TIR) of *gfpmut3* (*see* **Note 1**).

2. Decide the TBR sequence that is complementary to the target mRNA sequence and paste it upstream of the MicF M7.4 Hfq binding site, as shown in Fig. 1 (*see* **Note 2**).

3. Calculate ΔG using the NUPACK's analysis tool (setting the nucleic acid type to RNA and the temperature to 37°C) [43]. ΔG_{asRNA} (-65.5 kcal/mol) is estimated by entering the entire asRNA sequence (224 nucleotides in Fig. 1; *see* **Note 3**). ΔG_{mRNA} (-5.5 kcal/mol) is estimated by entering the target mRNA sequence and one extra nucleotide at the 3'-end (41 nucleotides in Fig. 1; *see* **Note 4**). The extra nucleotide is added to consider the stacking contribution of neighboring base pairs [44]. Similarly, $\Delta G_{asRNA:mRNA}$ (-111.38 kcal/mol, change in free energy of asRNA-mRNA complex) is estimated by entering the two sequences used above (1:1 concentration ratio). Finally, ΔG_{CF} (-40.38 kcal/mol) is calculated by using the equation: $\Delta G_{CF} = \Delta G_{asRNA:mRNA} - \Delta G_{asRNA} - \Delta G_{mRNA}$.

4. Determine the percent mismatch (7.5%), as shown in Fig. 1. The percent mismatch is calculated by dividing the number of mismatched nucleotides (3 nucleotides) by the total number of the TBR nucleotides (40 nucleotides).

5. Calculate the predicted repression efficiency (F=0.566) using the experimentally validated model (*see* **Note 5**): $F(X_1,X_2) = [0.3848 - 0.0068X_1 - 0.0125X_2 + \varepsilon]$ (R^2=0.685, p-value < 0.001)

where X_1 is ΔG_{CF} (in kcal/mol), X_2 is percent mismatch (in %), and ε is the standard error (ε=0.123). For reliable prediction, ΔG_{CF} should be from -59.88 to -6.58 kcal/mol, and percent mismatch should be from 0 to 32.1% (*see* **Note 6**).

3.2 Construction of the asRNA Plasmid Using Inverse PCR

1. Design primers for inverse PCR (*see* Section 2.2 for their sequences). The forward primer contains the 3' half of the TBR (20 nucleotides) and anneals to the part of pG16 (i.e., the part of the MicF M7.4 Hfq binding site, 16 nucleotides). The reverse primer contains the 5' half of the TBR (reverse complement, 20 nucleotides) and anneals to the part of pG16 (i.e., the part of the P_{Tet} promoter (reverse complement), 30 nucleotides).

2. Perform inverse PCR using the following conditions (*see* **Note 7**).

Water (ddH_2O)	37 µL
5X HF Phusion buffer	10 µL
10 mM dNTPs	1 µL
25 µM forward primer	0.5 µL
25 µM reverse primer	0.5 µL
Template DNA (pG16)	0.5 µL
Phusion DNA Polymerase	0.5 µL

Thermal cycle temperature	Duration	Repeat
98°C	30 sec	1 cycle
98°C	10 sec	30 cycles
60°C	20 sec	
72°C	80 sec	
72°C	5 min	1 cycle

3. Purify the size-confirmed, 3223 base-pair PCR products (or DpnI-treated PCR products; *see* **Note 8**) using a gel extraction kit or a PCR purification kit (following the manufacturer's instruction).

4. Perform simultaneously phosphorylation and blunt-end ligation at room temperature for 1 h (*see* **Note 9**).

T4 DNA ligase	1 μL
10X ligation buffer	1 μL
Purified DNA	7.8 μL
T4 PNK	0.2 μL

5. Transform competent *E. coli* cells, select ampicillin-resistant colonies using LB agar plates with ampicillin (100 μg/mL), and prepare the constructed asRNA plasmid (called pG43B2; 3223 base pairs) using a plasmid miniprep kit according to the manufacturer's instruction (*see* **Note 10**). The DNA sequence should be confirmed using the sequencing primer (*see* Section 2.2 for the primer sequence), which can cover the entire P_{Tet} promoter, the entire asRNA region (the replaced TBR sequence, the MicF M7.4 Hfq binding site, and the lambda T0 transcription terminator), and the extra terminator (*rrnB* T1 terminator).

3.3 GFP Repression Assays

1. Transform *E. coli* DH10B-GFP (*see* **Note 11**) using the sequence-confirmed asRNA plasmid (e.g., pG43B2).

2. Incubate the transformed cells on LB agar plates containing ampicillin (100 µg/mL) overnight at 37°C. Incubate *E. coli* DH10B (no-GFP control) on LB agar plates without any antibiotic overnight at 37°C.

3. Pick colonies and grow them in 1 mL LB liquid media with or without ampicillin (100 µg/mL) overnight at 37°C and 250 rpm. For all liquid cultures, use deep 96-well plates with breathable sealing membranes.

4. Transfer the overnight cultures (1% v/v) into 1 mL fresh LB media with or without ampicillin (100 µg/mL) and grow them for 2 h at 37°C and 250 rpm.

5. Transfer the subcultures (1.67% v/v) into 0.6 mL fresh LB media with or without ampicillin (100 µg/mL) and aTc (250 ng/mL; *see* **Note 12**). Grow them for 8 h at 37°C and 250 rpm.

6. Pellet the cells by centrifugation (for 10 min at 2200*g*) and resuspend them in 0.2 mL phosphate-buffered saline containing 2 mg/mL kanamycin (*see* **Note 13**).

7. Measure the population fluorescence ($F_{experimental}$) and the absorbance at 600 nm ($Abs_{experimental}$) using a microplate reader. The fluorescence is measured with excitation at 483 nm and emission at 530 nm.

8. Normalize the measured fluorescence value by the equation: $F_{norm} = (F_{experimental} / Abs_{experimental}) - (F_{no\text{-}GFP\ control} / Abs_{no\text{-}GFP\ control})$ where the no-GFP control is *E. coli* DH10B without a plasmid (i.e., the parent DH10B strain without *gfpmut3* integrated into the genome). Calculate the repression efficiency by the equation: $1 - (F_{aTc+} / F_{aTc-})$ where F_{aTc-} and F_{aTc+} are the normalized fluorescence values without and with aTc, respectively (*see* **Note 14**).

4 Notes

1. In this protocol, the target region is restricted to 75 nucleotides that cover the translation initiation region (i.e., the sequence from -35 to +40 with the ATG start codon's A as +1). The TIR includes the upstream of Shine-Dalgarno sequence (USD), the ribosome binding site (RBS or the SD sequence), and the start codon of the target gene. Targeting the TIR is essential to generate asRNAs with high and reliable repression efficiencies [15].

2. Mismatches can be introduced in a TBR sequence through base substitutions (e.g., T to A).

3. While both bacteriophage lambda T0 (the sequence shown in Fig. 1) and *rrnB* T1 (not shown in Fig. 1) terminators are used to ensure transcription termination of the asRNA, the actual asRNA is likely to contain only the upstream terminator (lambda T0), not both. Thus, ΔG_{asRNA} is estimated by entering the sequence of the TBR, the MicF M7.4 Hfq binding site, and the lambda T0 terminator, which excludes the *rrnB* T1 terminator sequence.

4. It is assumed that the coupling of transcription and translation in bacteria leads to local folding [45]. Thus, only the target mRNA region plus one extra nucleotide at the 3'-end (41 nucleotides) is considered in this model, as opposed to the entire mRNA sequence.

5. These two parameters had been selected as the main contributors for asRNA-mediated gene repression from the 12 initially-considered parameters, including target location, double-stranded RNA length, Hfq binding site, ribosome interaction, and YUNR motif [15,27]. To prevent the target location effect on repression, the target region should be restricted to the TIR, as mentioned above. The double-stranded RNA length and ΔG_{CF} show strong multi-collinearity,

requiring the elimination of one of the two parameters in the model. As discussed below, this predictive equation should be used only for asRNAs containing MicF M7.4 in *E. coli* and asRNAs lacking an Hfq binding site sequence in *B. subtilis*. In *E. coli*, this model should not be used for asRNAs lacking an Hfq binding site sequence. The rest initially-considered parameters, as well as the newly visited parameters (e.g., GC-contents of the paired asRNA:mRNA sequence and target accessibility), did not affect the model predictability.

6. Considering the desired target repression efficiency and off-target repression levels, the users can design TBR sequences with different lengths/target locations (ΔG_{CF} = -59.88 to -6.58 kcal/mol) and percent mismatch (0 to 32.1%). However, this data-driven model must be used only within the tested parameter ranges (ΔG_{CF} = -59.88 to -6.58 kcal/mol; 0 to 32.1% mismatch) to obtain a reliable prediction. As mentioned above, this model must also be used only for TIR-targeting asRNAs. Thus, it is unlikely to accurately predict potential off-target repression levels in which the parameter values are most likely to be outside of the acceptable ranges. Notably, the users can run BLAST searches against the entire genome of interest using the target sequence or the TBR sequence to identify potential off-target sites. Predicting these sites or their off-target repression levels is interesting future work and beyond the scope of this protocol.

7. Different labs use different reagents, and the users can follow the manufacturer's instructions to determine their PCR conditions. DMSO (up to 8%) can be added when using DNA templates with GC-rich sequences or secondary structures. When DMSO is used, the annealing temperature should be lowered (up to ~5^{o}C).

8. If no DMSO (or enzyme-denaturing chemical) is used for the PCR step, DpnI can be added directly into the PCR reaction mixture after the PCR step (1 μL per 50 μL PCR mixture). DpnI

treatment at 37ºC for 1 h would be sufficient to degrade the template plasmid. If enzyme-denaturing chemicals are used, the PCR products should be purified first, followed by DpnI treatment and additional purification according to the manufacturer's instruction.

9. Because primers are usually synthesized without 5'-phosphorylation, T4 PNK treatment is necessary for the ligation reaction. While T4 PNK's optimum temperature is 37ºC, this simultaneous phosphorylation/ligation reaction is usually efficient at room temperature for blunt-end ligation of inverse PCR products.

10. These steps can be modified according to each lab's practice and protocol as long as the DNA sequence of the constructed plasmid is verified. The confirmed sequence should cover the entire P_{Tet} promoter, the entire asRNA region, and the extra terminator.

11. The prediction model has been experimentally validated for asRNAs containing MicF M7.4 in *E. coli* (DH10B and Nissle 1917) and asRNAs lacking an Hfq binding site sequence in *B. subtilis* 168 [15]. The repression targets can be either chromosomal genes or plasmid-encoded genes if the expressed asRNA level is in excess of the target mRNA level. The repression targets can also be native genes other than fluorescent reporter genes, as the model was experimentally validated in the previous report [15]. Similar models can be applied to other strains than the above-mentioned strains, but the constants in $F(X_1, X_2)$ would need to be determined for accurate prediction, as described in the previous report (i.e., a, b, and c should be determined in $F = a - bX_1 - cX_2$) [15]. For the above-mentioned strains, the experimentally-measured repression efficiencies of 70% of the tested asRNAs (117 out of 168) were within a range of one standard error (1ε) of the predicted repression efficiencies (95% within a range of 2ε and 100% within 3ε) [15]. For non-TIR targeting asRNAs, this model cannot be used.

12. As shown in the previous report [15], the target gene repression level increases with the asRNA transcript level. The prediction model was built on the premise that the asRNA would be in excess of the target mRNA such that asRNA abundance would not be a determining factor for the repression efficiency. To ensure that asRNA levels are in excess of target mRNA levels, a high-copy number plasmid (e.g., pG43B2 with ColE1 origin) should be used, and the P_{Tet} promoter should be induced maximally.

13. Kanamycin (2 mg/ml) is used to stop bacterial translation (e.g., GFP synthesis) after sampling. Cells resuspended in phosphate-buffered saline containing kanamycin can be stored at 4°C before fluorescence measurement. Some fluorescent proteins require a long maturation time, especially when aeration is inadequate. Pipetting cells multiple times during the resuspension step and storing them at 4°C before analysis would help fluorescent proteins' maturation without further bacterial translation.

14. The repression efficiency (RE) should be defined as $1 - (F_{aTc+} / F_{aTc-})$ for accurate comparison with the model's predicted value. This definition gives conservative RE values in that a basal asRNA level (without aTc) due to promoter leakiness reduces F_{aTc-} and thus RE. If the RE is alternatively defined as $1 - (F_{asRNA+} / F_{asRNA-})$, where F_{asRNA-} (e.g., DH10B-GFP without any "leakiness" effect) and F_{asRNA+} (e.g., DH10B-GFP + an asRNA plasmid) are the normalized fluorescence values without and with an asRNA plasmid containing a constitutively and maximally expressed asRNA cassette, the RE value would be higher than that of the original definition ($1 - (F_{aTc+} / F_{aTc-})$). This alternative definition allows for reporting higher RE values, but the fluorescence ratio (F_{asRNA+} / F_{asRNA-}) is obtained by comparing genetically different strains (DH10B-GFP vs. DH10B-GFP + an asRNA plasmid), as opposed to the genetically identical strain (DH10B-GFP + pG43B2; without vs. with aTc). Additionally, inducible asRNA

expression, instead of constitutive asRNA expression, would make a better tool for gene expression control. Thus, considering the fair comparison (conservative reporting) and the RNA tool's utility, this model has been developed based on the experimental data obtained by using the original definition. The same point was discussed in a previous report covering a topic of engineering toehold switches in which RNA regulators are also used [46].

Acknowledgments

This work was supported by the National Science Foundation (MCB-1714352 and MCB-2001743). The author declares no conflict of interest.

5 References

1. Chappell J, Watters KE, Takahashi MK, Lucks JB (2015) A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. Current opinion in chemical biology 28:47-56. doi:10.1016/j.cbpa.2015.05.018

2. Rhodius VA, Segall-Shapiro TH, Sharon BD, Ghodasara A, Orlova E, Tabakh H, Burkhardt DH, Clancy K, Peterson TC, Gross CA, Voigt CA (2013) Design of orthogonal genetic switches based on a crosstalk map of sigmas, anti-sigmas, and promoters. Molecular systems biology 9:702. doi:10.1038/msb.2013.58

3. Shopera T, Henson WR, Moon TS (2017) Dynamics of sequestration-based gene regulatory cascades. Nucleic acids research 45:7515-7526. doi:10.1093/nar/gkx465

4. Nielsen AA, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA (2016) Genetic circuit design automation. Science New York, NY 352 (6281):aac7341. doi:10.1126/science.aac7341

5. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403 (6767):339-342

6. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403 (6767):335-338

7. Venturelli OS, Egbert RG, Arkin AP (2016) Towards Engineering Biological Systems in a Broader Context. Journal of molecular biology 428 (5, Part B):928-944. doi:http://dx.doi.org/10.1016/j.jmb.2015.10.025

8. Lou C, Stanton B, Chen YJ, Munsky B, Voigt CA (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. Nature biotechnology 30 (11):1137-1142. doi:10.1038/nbt.2401

9. Chappell J, Takahashi MK, Lucks JB (2015) Creating small transcription activating RNAs. Nat Chem Biol 11:214-220. doi:10.1038/nchembio.1737

10. Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP (2011) Versatile RNA-sensing transcriptional regulators for engineering genetic networks. Proc Natl Acad Sci USA 108:8617-8622. doi:10.1073/pnas.1015741108

11. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell 152:1173-1183. doi:10.1016/j.cell.2013.02.022

12. Bikard D, Jiang WY, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41:7429-7437. doi:10.1093/nar/gkt520

13. Green AA, Silver PA, Collins JJ, Yin P (2014) Toehold Switches: De-Novo-Designed Regulators of Gene Expression. Cell 159:925-939. doi:10.1016/j.cell.2014.10.002

14. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY (2013) Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. Nat Biotechnol 31:170-174. doi:10.1038/nbt.2461

15. Lee YJ, Kim SJ, Amrofell MB, Moon TS (2019) Establishing a Multivariate Model for Predictable Antisense RNA-Mediated Repression. ACS Synth Biol 8 (1):45-56. doi:10.1021/acssynbio.8b00227

16. Winkler W, Nahvi A, Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419:952-956. doi:10.1038/nature01145

17. Jang S, Jang S, Yang J, Seo SW, Jung GY (2018) RNA-based dynamic genetic controllers: development strategies and applications. Current opinion in biotechnology 53 (Supplement C):1-11. doi:https://doi.org/10.1016/j.copbio.2017.10.005

18. Hoynes-O'Connor A, Hinman K, Kirchner L, Moon TS (2015) De novo design of heat-repressible RNA thermosensors in E. coli. Nucleic acids research 43 (12):6166-6179. doi:10.1093/nar/gkv499

19. Win MN, Smolke CD (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci USA 104:14283-14288. doi:10.1073/pnas.0703961104

20. Carothers JM, Goler JA, Juminaga D, Keasling JD (2011) Model-driven engineering of RNA devices to quantitatively program gene expression. Science New York, NY 334 (6063):1716-1719. doi:10.1126/science.1212209

21. Prevost K, Desnoyers G, Jacques JF, Lavoie F, Masse E (2011) Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. Genes & development 25 (4):385-396. doi:10.1101/gad.2001711

22. Mizuno T, Chou MY, Inouye M (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). Proceedings of the National Academy of Sciences of the United States of America 81 (7):1966-1970

23. Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation. Molecular cell 32 (6):827-837. doi:10.1016/j.molcel.2008.10.027

24. Lee YJ, Moon TS (2018) Design rules of synthetic non-coding RNAs in bacteria. Methods 143:58-69. doi:https://doi.org/10.1016/j.ymeth.2018.01.001

25. Lee YJ, Kim S-J, Moon TS (2018) Multilevel Regulation of Bacterial Gene Expression with the Combined STAR and Antisense RNA System. ACS Synthetic Biology 7 (3):853-865. doi:10.1021/acssynbio.7b00322

26. Lee Y, Hoynes-O'Connor A, Leong MC, Moon TS (2016) Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. Nucleic Acids Res 44:2462–2473

27. Hoynes-O'Connor A, Moon TS (2016) Development of Design Rules for Reliable Antisense RNA Behavior in E. coli. ACS Synthetic Biology 5 (12):1441-1454. doi:10.1021/acssynbio.6b00036

28. Hoynes-O'Connor A, Shopera T, Hinman K, Creamer JP, Moon TS (2017) Enabling complex genetic circuits to respond to extrinsic environmental signals. Biotechnology and bioengineering 114:1626-1631. doi:10.1002/bit.26279

29. Dornenburg JE, Devita AM, Palumbo MJ, Wade JT (2010) Widespread antisense transcription in Escherichia coli. mBio 1 (1). doi:10.1128/mBio.00024-10

30. Lee JM, Zhang S, Saha S, Santa Anna S, Jiang C, Perkins J (2001) RNA expression analysis using an antisense Bacillus subtilis genome array. J Bacteriol 183 (24):7371-7380. doi:10.1128/jb.183.24.7371-7380.2001

31. Irnov I, Sharma CM, Vogel J, Winkler WC (2010) Identification of regulatory RNAs in Bacillus subtilis. Nucleic Acids Res 38 (19):6637-6651. doi:10.1093/nar/gkq454

32. Rasmussen S, Nielsen HB, Jarmer H (2009) The transcriptionally active regions in the genome of Bacillus subtilis. Mol Microbiol 73 (6):1043-1057. doi:10.1111/j.1365-2958.2009.06830.x

33. Csiszar K, Houmard J, Damerval T, Demarsac NT (1987) Transcriptional Analysis of the Cyanobacterial Gvpabc Operon in Differentiated Cells - Occurrence of an Antisense RNA Complementary to 3 Overlapping Transcripts. Gene 60 (1):29-37. doi:10.1016/0378-1119(87)90210-1

34. Duhring U, Axmann IM, Hess WR, Wilde A (2006) An internal antisense RNA regulates expression of the photosynthesis gene isiA. Proceedings of the National Academy of Sciences of the United States of America 103 (18):7054-7058. doi:10.1073/pnas.0600927103

35. Eisenhut M, Georg J, Klaehn S, Sakurai I, Mustila H, Zhang P, Hess WR, Aro E-M (2012) The Antisense RNA As1_flv4 in the Cyanobacterium Synechocystis sp PCC 6803 Prevents Premature Expression of the flv4-2 Operon upon Shift in Inorganic Carbon Supply. Journal of Biological Chemistry 287 (40):33153-33162. doi:10.1074/jbc.M112.391755

36. Flaherty BL, Van Nieuwerburgh F, Head SR, Golden JW (2011) Directional RNA deep sequencing sheds new light on the transcriptional response of Anabaena sp strain PCC 7120 to combined-nitrogen deprivation. Bmc Genomics 12. doi:10.1186/1471-2164-12-332

37. Gong Y, Xu X (2012) A small internal antisense RNA (aftsH) of all3642 (ftsH) in Anabaena sp PCC 7120. Chinese Science Bulletin 57 (7):756-761. doi:10.1007/s11434-011-4929-6

38. Hernandez JA, Alonso I, Pellicer S, Luisa Peleato M, Cases R, Strasser RJ, Barja F, Fillat MF (2010) Mutants of Anabaena sp PCC 7120 lacking alr1690 and alpha-furA antisense RNA show a pleiotropic phenotype and altered photosynthetic machinery. Journal of Plant Physiology 167 (6):430-437. doi:10.1016/j.jplph.2009.10.009

39. Hernandez JA, Muro-Pastor AM, Flores E, Bes MT, Peleato ML, Fillat MF (2006) Identification of a furA cis antisense RNA in the cyanobacterium Anabaena sp PCC 7120. Journal of Molecular Biology 355 (3):325-334. doi:10.1016/j.jmb.2005.10.079

40. Merino-Puerto V, Herrero A, Flores E (2013) Cluster of Genes That Encode Positive and Negative Elements Influencing Filament Length in a Heterocyst-Forming Cyanobacterium. Journal of Bacteriology 195 (17):3957-3966. doi:10.1128/jb.00181-13

41. Sakurai I, Stazic D, Eisenhut M, Vuorio E, Steglich C, Hess WR, Aro E-M (2012) Positive Regulation of psbA Gene Expression by cis-Encoded Antisense RNAs in Synechocystis sp PCC 6803. Plant Physiology 160 (2):1000-1010. doi:10.1104/pp.112.202127

42. Georg J, Hess WR (2011) cis-antisense RNA, another level of gene regulation in bacteria. Microbiology and molecular biology reviews : MMBR 75 (2):286-300. doi:10.1128/MMBR.00032-10

43. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, Dirks RM, Pierce NA (2011) NUPACK: Analysis and design of nucleic acid systems. Journal of computational chemistry 32 (1):170-173. doi:10.1002/jcc.21596

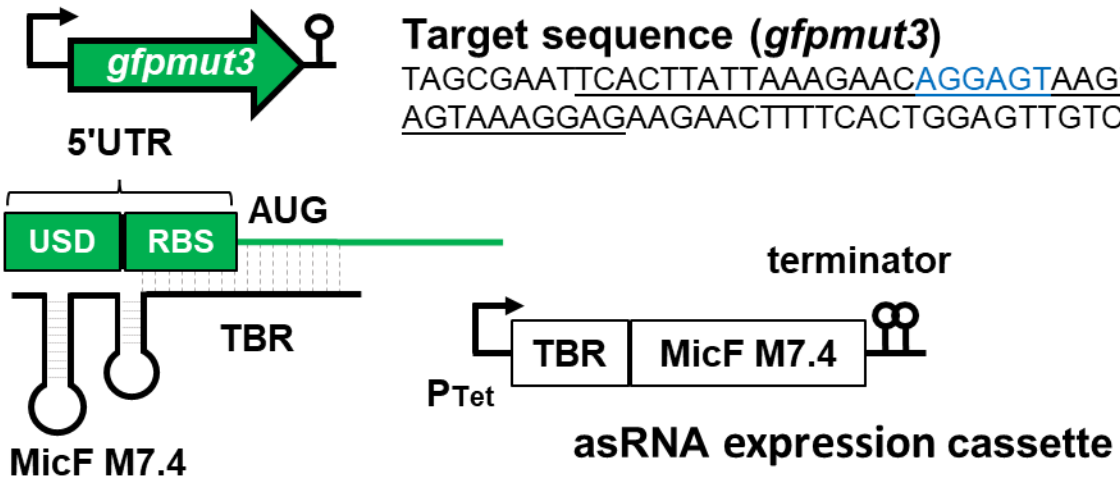
44. Vazquez-Anderson J, Mihailovic MK, Baldridge KC, Reyes KG, Haning K, Cho SH, Amador P, Powell WB, Contreras LM (2017) Optimization of a novel biophysical model using large scale in vivo antisense hybridization data displays improved prediction capabilities of structurally accessible RNA regions. Nucleic acids research 45 (9):5523-5538. doi:10.1093/nar/gkx115

45. Shao Y, Wu Y, Chan CY, McDonough K, Ding Y (2006) Rational design and rapid screening of antisense oligonucleotides for prokaryotic gene modulation. Nucleic acids research 34 (19):5660-5669. doi:10.1093/nar/gkl715

46. Kim S-J, Leong M, Amrofell MB, Lee YJ, Moon TS (2019) Modulating Responses of Toehold Switches by an Inhibitory Hairpin. ACS Synthetic Biology 8 (3):601-605. doi:10.1021/acssynbio.8b00488

Figure Captions

Figure 1 Model-based asRNA design. Different DNA sequence elements are shown in different colors: Shine-Dalgarno sequence (blue); start codon (red); the target sequence (underlined) plus one extra nucleotide on the 3'-end (highlighted in yellow); TBR (brown); MicF M7.4 Hfq binding site (green); transcription terminator (grey); mismatched nucleotide (highlighted in magenta). Abbreviation: upstream of Shine-Dalgarno sequence (USD); ribosome binding site (RBS); 5' untranslated region (5'UTR).



Target sequence (*gfpmut3*)

TAGCGAATTCACTTATTAAAGAACAGGAGTAAGTAATG
AGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAA

Target sequence (T for U)

+1 extra nucleotide

5'TCACTTATTAAAGAACAGGAGTAAGTAATGAGTAAAGGAGA3'

$$\Delta G_{mRNA} = -5.5\ \text{kcal/mol}$$

asRNA sequence (T for U)

5'CTCCTTTTCTCATTTCTTACTCCTCTTCTTTAATAAGTGACGTCCCGCAAGGAT
GCGGGTCTGTTTACCCCTATTTCAACCGGCCGCCTCGCGGCCGGTTTTTTTTT
GCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAA
CTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGG
TGAGAATCCA3'

$$\Delta G_{asRNA} = -65.5\ \text{kcal/mol}$$

$$\Delta G_{asRNA:mRNA} = -111.38\ \text{kcal/mol}$$

$$\Delta G_{CF} \text{ or } X_1 = \Delta G_{asRNA:mRNA} - \Delta G_{asRNA} - \Delta G_{mRNA} = -40.38\ \text{kcal/mol}$$

Percent mismatch or $X2$

$$= \frac{\#\ of\ mismatched\ nucleotides}{\#\ of\ nucleotides\ in\ TBR} \times 100\% = \frac{3}{40} \times 100\% = 7.5\%$$

Predicted repression efficiency

$$= 0.3848 - 0.0068X_1 - 0.0125X_2 = 0.566$$