Balancing gene expression without library construction via a reusable sRNA pool

Citation: Ghodasara, Amar and Christopher A. Voigt. "Balancing Gene Expression Without Library Construction via a Reusable sRNA Pool." Nucleic Acids Research 45, 13 (June 2017): 8116–8127 © 2017 The Author(s)

As Published: http://dx.doi.org/10.1093/NAR/GKX530

Publisher: Oxford University Press (OUP)

Persistent URL: http://hdl.handle.net/1721.1/114565

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

8116–8127 Nucleic Acids Research, 2017, Vol. 45, No. 13 Published online 13 June 2017
doi: 10.1093/nar/gkx530

Balancing gene expression without library construction via a reusable sRNA pool

Amar Ghodasara[1] and Christopher A. Voigt[1,2,*]

[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and [2]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA

Received February 27, 2017; Revised June 06, 2017; Editorial Decision June 07, 2017; Accepted June 07, 2017

ABSTRACT

Balancing protein expression is critical when optimizing genetic systems. Typically, this requires library construction to vary the genetic parts controlling each gene, which can be expensive and time-consuming. Here, we develop sRNAs corresponding to 15nt 'target' sequences that can be inserted upstream of a gene. The targeted gene can be repressed from 1.6- to 87-fold by controlling sRNA expression using promoters of different strength. A pool is built where six sRNAs are placed under the control of 16 promoters that span a ∼10^3-fold range of strengths, yielding ∼10^7 combinations. This pool can simultaneously optimize up to six genes in a system. This requires building only a single system-specific construct by placing a target sequence upstream of each gene and transforming it with the pre-built sRNA pool. The resulting library is screened and the top clone is sequenced to determine the promoter controlling each sRNA, from which the fold-repression of the genes can be inferred. The system is then rebuilt by rationally selecting parts that implement the optimal expression of each gene. We demonstrate the versatility of this approach by using the same pool to optimize a metabolic pathway (β-carotene) and genetic circuit (XNOR logic gate).

INTRODUCTION

Expressing proteins at the wrong levels is a common reason why genetic designs fail. For a metabolic pathway, suboptimal enzyme concentrations could lead to problems, such as the accumulation of toxic intermediates, titration of cofactors, or overburdening the host (1–5). Similarly, when building a genetic circuit, the expression of regulatory proteins needs to be balanced in order to perform the desired computational function (6,7). Across these projects, the challenge is that the correct expression levels are often not known *a priori* and, even if they were, there is uncertainty in selecting genetic parts to achieve target expression levels (8,9). Further, the optimal expression of one gene often depends on the expression levels of other genes in the system, thus creating a 'rugged' search space where multiple genes have to be changed simultaneously to achieve an improvement (10–13). For these reasons, optimization often requires the creation of libraries by mutagenesis and screening.

Libraries of multi-gene systems can be built using guided and unguided approaches. Many variants of a pathway can be constructed by substituting genetic parts controlling each gene; for example, promoters, ribosome binding sites (RBSs), and RNA stability elements (1,5,14–19). Beyond part substitution, libraries can be built that alter the gene order, orientation, and operon occupancy (20–22). These approaches are often blind searches that require large libraries whose size is limited by assay throughput. To address this, mathematical modeling and combinatorial optimization algorithms have been used to guide the search towards a design objective, such as improved metabolic flux (23–28). In the absence of a model, the construction of pathway variants can be guided by multivariate statistical methods, including multifactor design of experiments and Bayesian approaches (12,29–32). At the other extreme, random approaches (e.g. mutagenesis and recombination) can be applied across the entire construct or targeted to particular regions (14,33–37). These approaches are low cost because they do not require sequence verification of each construct; however, they are limited by low mutagenesis rates that limit the simultaneous optimization of multiple genes (38). Further, the quality of the library needs to be monitored by sequencing to prevent nucleotide substitution biases and mutations as stop codons can cause a large fraction to be non-functional. Whether guided or unguided, the construction and sequence verification of the library is a slow and expensive step that has to be repeated for each system to be optimized.

We sought to develop a method where the expression levels of multiple genes could be simultaneously tuned without the need to rebuild a library for each system. Variation is achieved with a separate library, built once, that contains regulators expressed at different levels. Each regulator

*To whom correspondence should be addressed. Tel: +1 617 324 4851; Email: cavoigt@gmail.com

would control a different gene in the system so that changes in regulator expression lead to changes in the target gene. This approach requires regulators that are:

- *Orthogonal* so that each regulator only controls its target
- *Wide in dynamic range* to be able to sweep across expression levels,
- *Programmable* so that the regulator can be targeted to different genes and
- *Non-toxic* so that the regulators themselves do not influence the system.

We selected sRNAs (39–42) as they satisfy these criteria. sRNAs are potent negative regulators that do not require heterologous proteins to function (43). When expressed, sRNAs bind their cognate mRNA in the 5' untranslated region (UTR) via Watson–Crick base pairing with assistance from Hfq, a ubiquitous RNA chaperone protein. Once bound, translation initiation is inhibited and RNA degradation by RNase E is accelerated (44,45). Most synthetic sRNAs studied to date have involved fusing new target sequences to natural sRNA scaffolds and minor scaffold modifications (41,46–48). The best-engineered sRNAs have been shown to achieve up to 85-fold repression without cross-reactions (41,46).

Since each RNA element targets a sequence specific to its cognate gene, the RNA must be designed *ad hoc* for each target and screening of multiple designs is often required (49,50). This is usually successful because of the programmability of nucleotide interactions, however, it requires optimization for each gene (48,51). We wanted to simplify this process so that sRNAs could be optimized once and then used to control arbitrary genes. To this end, we designed sRNA 'target sequences' that are placed in the 5′ untranslated region (5'-UTR) of the gene to be tuned (Figure 1A). Fixing the sRNA target site avoids the need to redesign the sRNA for each target gene. The sRNA is optimized to increase the fold-repression against this target and verified with multiple genes. This is repeated to create a set of six sRNA:target pairs that are orthogonal to each other and do not impact the host even when highly expressed. Further, it is demonstrated that the repression can be controlled in a graded manner by using constitutive promoters of different strengths to control the expression level of the sRNA (Figure 1B). All six sRNAs are used to build a combinatorial pool of $\sim 10^7$ differentially expressed sRNAs (Figure 1C). This pool is capable of tuning genetic systems with up to six different genes over a wide expression space (Figure 1D). As a proof of concept, we applied the same sRNA pool to tune very different systems: a metabolic pathway (β-carotene) and a genetic circuit (XNOR logic). This only requires building a single construct for each system, where the target sequence is placed upstream of each gene to be tuned. Once the optimal construct is identified, the reliance on sRNA expression can be removed by performing limited part substitution in order to recover the same expression levels discovered with the sRNA screen.

MATERIALS AND METHODS

Strains and media

Escherichia coli DH10b [F-*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L *nup*G λ-] (52) was used in all cloning procedures and experiments unless stated otherwise. *Escherichia coli* K-12 MG1655* [F-k-ilvG-rfb-50 rph-1 Δ(araCBAD) Δ(LacI)] (53) was used for iron starvation assays. *Escherichia coli* K-12 JW4130-1 [F-Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) λ-rph-1 Δ(rhaD-rhaB)568 Δhfq-722::kan, hsdR514] and BW25113 [F-Δ(araD-araB)567 ΔlacZ4787(::rrnB-3), λ-rph-1 Δ(rhaD-rhaB)568 hsdR514] from the KEIO collection (54) were used for hfq knockout experiments. MegaX DH10B T1 Electrocomp Cells (ThermoFisher, C640003) were used for sRNA pool construction. 'Clonetegration' as described by St-Pierre *et al.* (55) was used to integrate sRNA-A-tagged RFP at attP21. The following *Pseudomonas* and *Salmonella* strains were used: *Pseudomonas protegens* Pf-5 (ATCC BAA-477), and *Salmonella typhimurium* LT2 (ATCC 700720). Cells were grown in LB Miller broth (Difco, 90003-350) or SOC (SOB (Teknova, S0210) supplemented with 0.4% glucose) for cloning. Supplemented M9 minimal media (1x M9 Salts (Sigma-Aldrich, M6030), 2 mM $MgSO_4$, 100 μM $CaCl_2$, 0.4% glucose, 0.2% casamino acids, 340 mg/l thiamine) was used for knockdown assays and iron starvation assays. Chloramphenicol (34 μg/ml) (Alfa Aesar, AAB20841–14), kanamycin (50 μg/ml) (GoldBio, K-120-10) or spectinomycin sulfate (50 μg/ml) (MP Biomedicals LLC, 158993) was supplemented where appropriate. IPTG (Sigma-Aldrich, I6758) and aTC (37919) were used as inducers. Three fluorescence proteins, GFPmut3 (56), mRFP1 (57) and YFP (56), were used as reporters.

sRNA repression assay

Colonies were picked and inoculated into 300 μl LB with antibiotics and grown 16 h at 37°C and 900 RPM in a Multitron Pro incubator shaker (In Vitro Technologies). Three microliters of this overnight culture were then added to 147 μl of M9 media with antibiotics and grown for 3 h in an ELMI shaker (Elmi Ltd.) at 37°C and 1000 RPM, resulting in an OD_{600} of 0.2–0.4. Twenty microliters of the cultures were diluted into 180 μl 1× phosphate buffered saline (PBS) with 2 mg/ml kanamycin. For sRNA optimization experiments, sRNAs were expressed using promoter BBa_J23119 and terminator BBa_B0010 from a pBR322 plasmid (pRNA) while the reporter (GFP or RFP) was expressed from a p15A plasmid using promoter BBa_J23101, RBS BBa_B0032, terminator BBa_B0012, and RiboJ, an insulating ribozyme (pREP, Supplementary Figure S1) (58).

sRNA growth assay

Plasmid pMAX, in which all sRNAs are expressed from the strongest promoters used in the pool (P_{A16}, P_{B16}, P_{C16}, P_{D16}, P_{E16} and P_{F16}), was used to quantify growth defects from sRNA overexpression (Supplementary Figure S2 and Table S1). *Escherichia coli* DH10B cells harboring control plasmid pEMPTY or pMAX were grown for 18 h in LB,

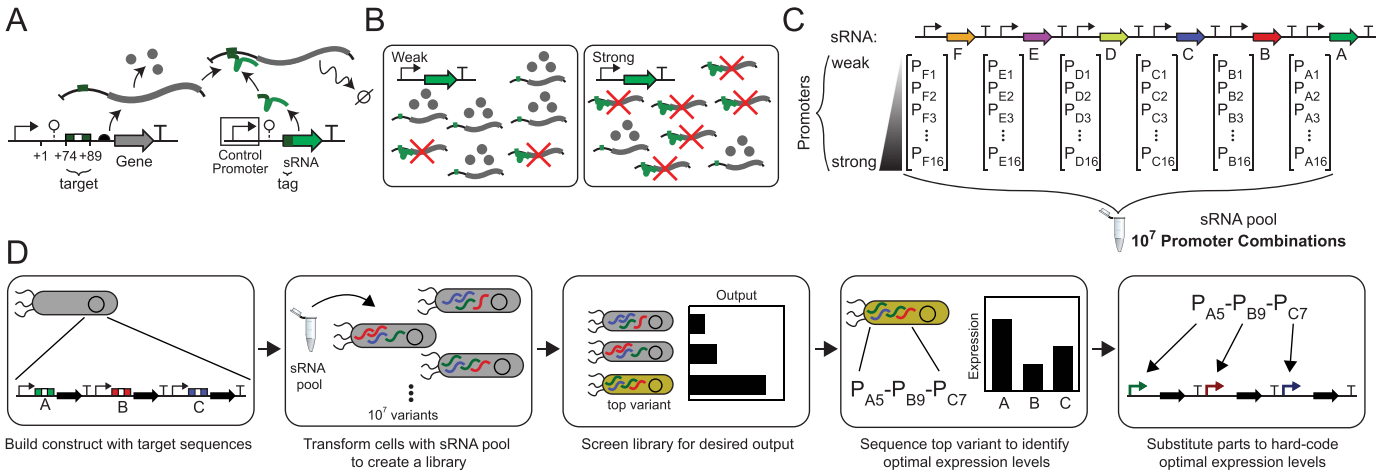

Figure 1. Combinatorial repression of target genes with sRNAs. (**A**) Targeting of genes by engineered sRNAs. Raised arrows are promoters, circles on stems are ribozymes, hemispheres are RBSs, and 'T's are terminators. Colored rectangle with a white square represents the target sequence in the 5' UTR of the target gene. (**B**) Effect of sRNA expression level on target knockdown. (**C**) Combinatorial pool of differentially expressed sRNAs. Up to six different sRNAs driven by 16 unique promoters are expressed from one plasmid. Sequences for pool parts are provided in Supplementary Table S1. (**D**) The process of using the sRNA pool to optimize a genetic system. Optimization of a three gene metabolic pathway in which promoters (colored) are replaced to recapitulate optimal gene expression (identified from sRNAs expression levels) is shown.

normalized to an OD_{600} of 1, then diluted 1:200 in LB and grown at 37°C for 3 h. Cultures were then diluted 1:200 again in LB into microtiter plates and growth was measured for 16 h. OD_{600} measurements were performed on a Cary 50 Bio UV-Vis spectrophotometer (Varian). Cultures were diluted with sterile water until the OD_{600} fell between 0.10 and 0.60. The same protocol was used for iron starvation growth assays with the exception that *E. coli* MG1655 cells and supplemented M9 media without iron were used.

Promoter strength measurements

The strengths of all constitutive promoters were characterized by using them to drive YFP expression from a reporter plasmid (pPROM, Supplementary Figure S1). Reported promoter strengths are the geometric mean of the measured fluorescence output. The sRNA repression assay was followed with the exception that cells were grown in LB instead of M9 media. T7 promoter characterizations were performed by co-transforming the T7RNAp expression plasmid N249 (59) with the reporter plasmid. For these assays, colonies were grown overnight and diluted 1:200 into LB with antibiotics and 1 mM IPTG and grown for 6 h at 37°C, then analyzed using the flow cytometer.

Flow cytometry

Fluorescence was measured using a MACSQuant VYB (Miltenyi Biotec) with a 488-nm laser for GFP and YFP excitation and 561-nm for RFP excitation. For each sample at least 5×10^4 events were recorded using a flow rate of 0.5 μl/s. FlowJo v10 (TreeStar Inc.) was used to analyze the data. All events were gated by forward scatter and side scatter. Fluorescence values are shown as the geometric mean. Fold-repression is calculated as the geometric mean of the fluorescence measured with an empty vector (pEMPTY, Supplementary Figure S1) divided by the geometric mean of the fluorescence measured with an sRNA expression vector.

Computational methods

All Random DNA sequences were generated using a random DNA sequence generator (http://www.faculty.ucr.edu/~mmaduro/random.htm), with a GC content probability parameter of 0.5. All sRNA tags were evaluated against the *E. coli* genome using BLASTn (http://www.ecogene.org/ecoblast) (60); sRNA tags exhibiting more than 11bp chromosomal homology or homology within 200 bp of annotated coding sequences on either strand were omitted. The search was performed on the *E. coli* K-12 MG1655 U00096.3 genome database with an expect threshold of 10 and the BLOSUM62 matrix. Nupack (www.nupack.org) (61) was used to predict RNA secondary structures. The following parameters were used for all RNA folding simulations: Nucleic acid type: RNA; temperature: 37°C; number of strand species: 1; maximum complex size: 1; RNA energy parameters: 'Mathews *et al.*, 1999'; Dangle treatment: some. RBS sequences for the pathway and circuit were generated using the RBS Calculator (https://www.denovodna.com/software/) (62). For all calculations, the following parameters were used unless otherwise noted: pre-sequence: (insulating ribozyme sequence followed by an sRNA tag); protein coding sequence: (full coding sequence from start to stop codon); target translation initiation rate: 50 000; free energy model: version 2.0; organism: *E. coli* str. K-12 substr. MG1655 (ACCTCCTTA). RBSs for BetI and HlyRII were designed with the Target Translation Initiation Rate set to 25 000.

sRNA library construction and sorting

All degenerate nucleotide libraries were built using circular polymerase extension cloning (CPEC) (63) with oligos

containing N's (Integrated DNA Technologies) at appropriate positions. The CPEC product was column purified (DNA Clean & Concentrator-5, ZymoResearch) and transformed into cells harboring the appropriate reporter plasmid. After 1 h outgrowth in SOC, cells were diluted in to 5 ml LB with antibiotics and grown overnight at 37°C. The sRNA repression assay was performed with the following two modifications: (i) 100 μl of culture was used instead of 3 μl, (ii) after 3 h of growth in M9 media cells were pelleted and resuspended in 2 ml 1× PBS with 1 mg/ml chloramphenicol. Cells were sorted on a FACS Aria II (BD Biosciences). Sorting gates were set to sort cells with lower GFP expression than that achieved by the previous sRNA scaffold design. Sorting was stopped after 5×10^4 cells were collected. Sorted cells were allowed to recover in SOC for 2 h then plated and allowed to form colonies. Colonies were then streaked out and individually assayed by flow cytometry.

sRNA pool construction

The CombiGEM (64) protocol was used to assemble the sRNA pool (pPOOL, Supplementary Figure S1) with the following modifications. (i) The AvrII restriction site was replaced with XbaI. (ii) Ligation products were transformed into MegaX *E. coli* DH10B T1 Electrocomp Cells and after 1 h outgrowth in SOC, cells were diluted in to 5 ml LB with antibiotics and grown overnight at 37°C; the following morning, the culture was miniprepped and the process was iterated to until all six sRNAs were cloned.

High-throughput sequencing

sRNA pool sequencing barcodes were PCR amplified using 100 ng of the pool as template. Barcode PCR products were sent to the Massachusetts General Hospital DNA Core where the samples were processed and run on a Mi-Seq producing ∼200 000 reads. Custom software was written to search the raw sequencing FASTQ file against the list of all 96 sequencing barcodes (Supplementary Table S1) to count the number of occurrences of each corresponding promoter in the pool.

β-Carotene production quantification

The β-carotene pathway construct (pCAROTENE, Supplementary Figure S1) was co-transformed with plasmid N249 (59). Cells were grown overnight, made electrocompetent (65) and transformed with 100 ng of the sRNA pool and plated. Colonies harboring all three plasmids were grown in 1 ml LB with antibiotics at 30°C for 18 h. Cultures were then diluted to an OD_{600} of 0.2 into 1 ml Terrific Broth (TekNova, T7060) with appropriate antibiotics and 1 mM IPTG and grown for 48 h in a Multitron Pro at 30°C in 96-well deep plates (USA Scientific, 1896-2000). Cells were harvested for quantification by centrifugation at 4000g for 10 min. They were then washed with 0.6 ml water, and re-centrifuged. Cell pellets were then resuspended in 0.5 ml acetone and incubated at 55°C for 30 min with frequent vortexing. Pigment extraction mixtures were then centrifuged at 4000g for 15 min. 0.2 ml of supernatant was transferred to a polypropylene 96-well plate (Greiner, 655201) for quantification. Samples were measured using a H1 Synergy plate reader (BioTek Instruments) at an absorbance wavelength of 470 nm. A standard curve (Supplementary Figure S3), using purified β-carotene (Sigma-Aldrich, 22040), was established to correlate absorbance to amount of β-carotene produced. The following equation was used: $\text{titer} = 120.6\frac{\text{mg}}{\text{L}}(\text{measured absorbance} - \text{background absorbance})$. Background absorbance was measured from a strain harboring pEMPTY (Supplementary Figure S1) in place of the pathway construct. Samples from the top producing strains were further analyzed by liquid chromatography (1260 Infinity LC System, Agilent Technologies) using a reversed phase C18 column (Phenomenex, 00A-4462) to confirm absence of lycopene and other precursor metabolites and confirm sample purity (Supplementary Figure S3).

XNOR circuit assay and sorting

Cells harboring the XNOR circuit (pCIRCUIT) and output (pOUT) plasmids (Supplementary Figure S1) were grown overnight at 37°C in 1 ml LB then diluted 1:200 into 200 μl M9 with antibiotics and grown for 3 h at 37°C. The culture was then diluted 1:700 into 200 μl M9 with antibiotics and inducers and grown for 6 h. Twenty microliters of the cultures were diluted into 180 μl 1× phosphate buffered sulfate (PBS) with 2 mg/ml kanamycin and run on the cytometer. For the first input 1 mM IPTG was used for a '1' input and no IPTG for a '0' input. For the second input, 2 ng/ml aTC was used for a '1' input and no aTC for a '0' input. The geometric mean of fluorescence in each state was used as the output. For XNOR optimization, cells harboring the XNOR circuit and output plasmid were grown overnight, made electro-competent (65), and transformed with 100 ng of the sRNA pool. After 1 h of outgrowth in SOC, cells were diluted in 5 ml LB with antibiotics and grown overnight at 37°C. The following day the culture was diluted 1:200 into 5 ml M9 with antibiotics and grown for 3 h at 37°C. Induction was commenced at this point by diluting cultures 1:700 into 3.5 ml M9 with antibiotics and appropriate inducers (1 mM IPTG and/or 2 ng/ml aTC). After 6 h of growth, cells were pelleted and resuspended in 2 ml PBS supplemented with 1 mg/ml chloramphenicol and sorted. Sort gates were selected according to the desired output (YFP-high or YFP-low) for each input condition and were designed to include 5% of all cells. Sorting was stopped after 5×10^4 cells were collected. Sorted cells were allowed to recover in SOC for 2 h then diluted into 5 ml LB with antibiotics and grown overnight. The induction and sorting process was repeated for all input conditions. After the final sort, cells were allowed to recover in SOC for 2 h then plated and allowed to form colonies. Colonies were then streaked out and individually assayed by flow cytometry.

sRNA-mediated knockdown in other Gram-negative bacteria

The sRNA expression cassette from sRNA-A and from its corresponding reporter plasmid were cloned into a pBBR1 vector (66) (pBBR-RNA, Supplementary Figure S1). The plasmid was electroporated into *Pseudomonas* and

Salmonella cells. *Pseudomonas* cells were subjected to the sRNA repression assay with the exception that cells were grown at 30°C in LB. *Salmonella* cells were subjected to the sRNA repression assay with the exception that cells were grown in LB. Fold-repression is calculated as the geometric mean of the fluorescence measured with a vector without the sRNA expression cassette (pBBR-EMPTY) divided by the geometric mean of the fluorescence measured with pBBR-RNA.

RT-qPCR

Probe-based RT-qPCR was performed on a LightCycler 480 (Roche) using PrimeTime Gene Expression Master Mix (Integrated DNA Technologies, 1055772). Custom PrimeTime qPCR assays with a 6-FAM dye and ZEN/IBFQ quencher were designed for each gene using the PrimerQuest Design Tool (https://www.idtdna.com/PrimerQuest/). The promoter measurement assay described above was followed. RNA was isolated immediately after growth using the Quick-RNA Fungal/Bacterial Microprep kit (ZymoResearch, R2010). DNA was removed by DNAse treatment with the DNA-free Kit (ThermoFisher, AM1906). The ProtoScript First Strand cDNA Synthesis Kit (New England BioLabs, E6300S) was used to generate cDNA. The *cysG*, *hcaT* and *idnT* genes were used as reference genes based on literature (67).

RESULTS

Optimization of sRNA: target pairs

Our approach is based on the ability to design a target sequence that can be inserted upstream of any gene to make it sensitive to a cognate sRNA. This required first identifying the optimal region within the 5'-UTR to place the target. Then, the sRNA was optimized to maximally reduce target gene expression. Multiple parameters, including sRNA tag length (41), Hfq scaffold sequence, terminator strength, and 5' standardization (68) and RNA stabilization (15) were optimized. Finally, a set of orthogonal sRNA:target pairs were created so that multiple genes can be simultaneously controlled.

To evaluate sRNA designs, we developed an assay involving two plasmids (Supplementary Figure S1). The first contains a constitutive promoter (J23119) driving the transcription of the sRNA. The second contains a constitutively expressed fluorescent protein in which an sRNA target sequence is inserted upstream of a strong RBS (B0032) (Materials and Methods). The plasmids are co-transformed and fluorescence quantified by flow cytometry (Materials and Methods). Fold-repression is reported by comparing the amount of sRNA-mediated knockdown to cells harboring an empty vector.

We first identified the optimal region in the 5'-UTR to target with an sRNA. For our initial tests, ten 20nt sRNA tags with 50% GC content were generated using a random DNA sequence generator (Materials and Methods). Each tag was evaluated for sequence identity with the *E. coli* genome using EcoBlast (60) and a tag with no match to the genome was selected. This tag was fused to the wild-type MicC sRNA scaffold as described by Na *et al.* (41) (sRNA-AG1) while its cognate targeting sequence (reverse complement of tag) was inserted at multiple positions in the 5'-UTR of GFP and RFP (Supplementary Figure S4). We found that placing the target sequence closer to the RBS on the 5' side improved gene repression while inserting it on the 3' of the RBS resulted in reduced repression. These results corroborate similar studies and are consistent with the mechanism of sRNA-mediated gene repression (44,69). We avoided directly targeting the RBS or coding sequences to ensure that the target can be reused to control different genes.

Once the optimal target position was identified, we tested different tags and target lengths to further improve repression. For *E. coli*, which has a genome size of ~4.6 Mb, 12nt is the minimum number nucleotides necessary to ensure perfect orthogonality to the genome. Though targeting sequences short as 6nt have been observed in natural sRNAs, shorter target sequences can lead to sRNA promiscuity requiring very high sRNA expression, and off-target effects (45). We designed and tested tags and corresponding targeting sequences ranging from 12nt to 20nt in length. Repression efficiency was similar across all tag lengths (Supplementary Figure S4). As tag length decreases, RNA hybridization is less energetically favorable (41). However, as tag length increases the probability of alternate secondary conformations for the sRNA and target mRNA increases, which may reduce repression (70). Therefore, we standardized sRNA tag/target length at 15nt as a compromise between these constraints.

Next, we evaluated different Hfq scaffolds and sought to engineer one for strong repression. Six well-studied Hfq scaffolds from natural sRNAs in *E. coli* were fused to the 15nt tag described above and assayed: DsrA (71), GcvB (72), MicC (73), MicF (74), SgrS (75) and Spot42 (76). All tested sRNAs exhibited activity, of which the Spot42-derived sRNA was the strongest achieving 68-fold repression (Figure 2A).

Most natural sRNAs include a rho-independent terminator in their Hfq scaffolds as it is essential for Hfq recruitment (77,78). Prior studies have shown that this terminator can be exchanged for other rho-independent terminators without abrogating sRNA-mediated repression (47,79). Since the secondary structure of sRNAs is essential for Hfq-binding (80), and consequently gene repression, we hypothesized that replacing the natural terminator found in the Spot42 Hfq scaffold with a stronger terminator would improve repression efficiency as this would result in fewer sRNAs with extended 3' ends that could interfere with sRNA folding. The best, T(BF1176), was selected as the core scaffold for further optimization (Figure 2C).

A prior structural study showed that all but six nucleotides (not including the terminator stem-loop) of the Spot42 scaffold are involved in Hfq-binding, and consequentially essential for sRNA-mediated repression (81). Therefore, we created a targeted mutagenesis library in which these six nucleotides were randomized resulting in a 4096-variant library (Figure 2C, arrows). Potent scaffold variants were identified by fluorescence-activated cell sorting (FACS) (Materials and Methods). The best one (sRNA-AG4) yields 115-fold repression with all positions mutated (Figure 2A). Folding predictions of these high-

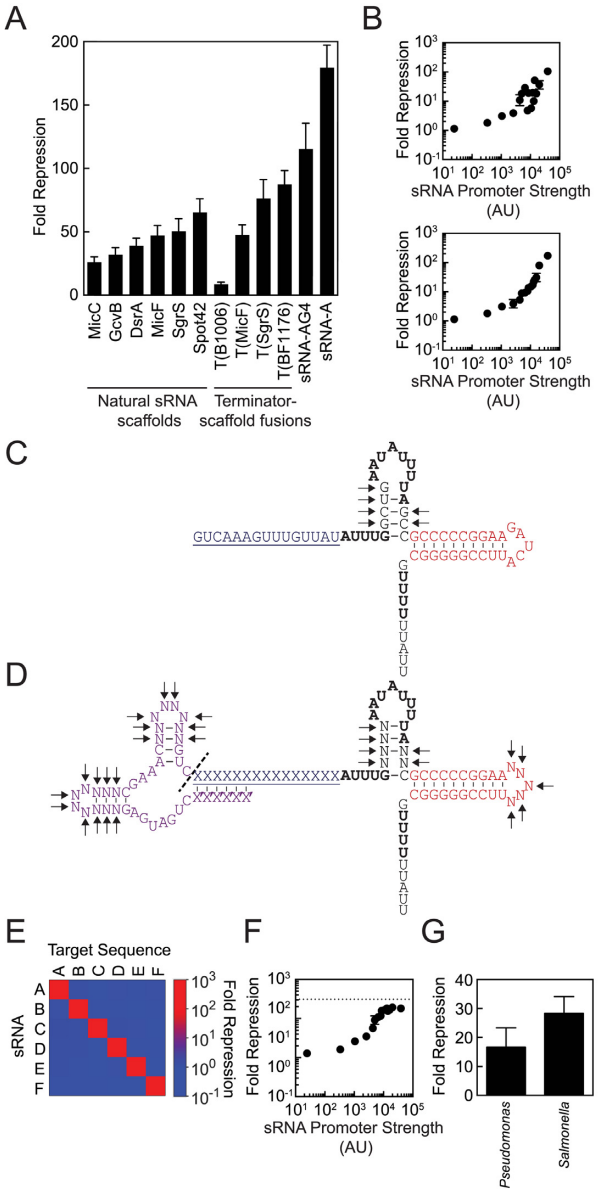

Figure 2. Optimization of sRNA-mediated gene knockdown against target sequences. (**A**) Results of sRNA scaffold engineering. The reporter and sRNA are expressed from different plasmids as shown in Supplementary Figure S1. Sequences for all sRNA variants are provided in Table 1 and Supplementary Table S3. (**B**) Repression response of sRNAs expressed without (top panel) and with (lower panel) a ribozyme from 16 constitutive promoters of increasing strength. Sequences of all promoters are provided in Supplementary Table S1. (**C**) Sequence and structure of sRNA T(BF1176). The sRNA tag is shown in blue and underlined. Bold nucleotides are essential for Hfq-binding (81). Red nucleotides indicate the terminator stem loop of terminator BF1176. Arrows indicate nucleotides that were randomized to screen for improved scaffolds. (**D**) Design of library used to identify and diversify ribozymes and scaffolds for each sRNA tag. Purple nucleotides indicate the ribozyme; dashed line represents ribozyme cleavage. 'X' denotes sRNA tag sequence while 'X" denotes ribozyme nucleotides complementary to the tag. Arrows indicate nucleotides that were randomized to screen for improved, diverse ribozymes and sRNA scaffolds. (**E**) Orthogonality matrix for all six sRNAs and their target sequences. Error bars are shown in Supplementary Figure S6. (**F**) sRNA targeting a chromosomally integrated reporter. The dashed line denotes complete gene repression (background level). (**G**) sRNA-A activity in other Gram-negative bacteria. For all data in this figure error bars correspond to the standard deviation of three experiments performed on different days.

performing scaffold variants revealed that the predicted secondary structure of the Spot42 Hfq-binding region was preserved (Materials and Methods). As expected, this repression is Hfq-dependent (Supplementary Figure S5).

A challenge with using promoters to control sRNA expression is that they can generate a distribution of transcription start sites (14,82). As a result, the initial attempt to control gene expression using different strength promoters produced a poor correlation between promoter strength and repression (Figure 2B, top panel). Using the published transcription start sites of the promoters (14), we performed folding predictions after adding the 5' overhang to the sRNA and found that 10/16 disrupted folding. Ribozymes were chosen to standardize the 5' end of the sRNA. Using a hammerhead ribozyme (HHR) scaffold (83), we designed a random library in which 18 non-essential nucleotides of the ribozyme (Figure 2D, purple nucleotides with arrows) were replaced with degenerate bases resulting in a theoretical library of 10^{10} variants fused to sRNA-AG4. This library was then screened using FACS. Approximately 3% of the screened variants exhibited improved repression from which the top-performing HHR/sRNA design (sRNA-A), capable of 175-fold repression, was discovered. The repression achieved with this variant was found to strongly correlate with promoter strength (Figure 2B, lower panel).

We then created a set of sRNAs that target different tags to enable the control of multiple genes within a cell. Redesigning the sRNA scaffold for a new target simply requires changing the 15nt tag on the sRNA and inserting the reverse complement of the cognate target sequence preceding the Shine Delgarno sequence on the gene to be controlled. A bioinformatic approach was pursued to generate a list of tags that avoid off-target effects in *E. coli* (Materials and Methods). In addition, secondary structure predictions were performed to ensure that the addition of the target sequence to the sRNA would not disrupt folding of the Hfq-binding region. From this list, 25 tags were selected, fused to sRNA-A, HHR appropriately modified for each, and tested for repression against corresponding reporter constructs. Those were eliminated that exhibited toxicity (9), crosstalk (3), or produced less than 100-fold repression (7). This led to a set of 6 sRNAs that efficiently target different sequences without impacting the host (Supplementary Table S2).

The optimized sRNA scaffold (without the tag) is 55nt and the HHR is 42nt. Simultaneously using all six sRNAs presents a problem as sequence repeats of >25 bp, especially in RecA+ strains, can lead to genetic instability due to recombination (84). To overcome this, we diversified the scaffold of each sRNA by mutagenizing 11 non-essential positions (6 in the Hfq-binding stem and 5 in the terminator loop) along with the 18 positions in the ribozyme and screening by FACS, as before (Figure 2D, arrows). The best performing sRNA for each tag was selected (Table 1), and all sRNAs have a maximum of 19nt of continuous shared sequence identity. The six final sRNAs exhibited similar repression against their cognate target sequence ranging from 120- to 175-fold and showed near perfect orthogonality (Figure 2E and Supplementary Figure S6).

We also evaluated the engineered sRNA against a chromosomally integrated target (Figure 2F). The RFP reporter gene harboring the targeting sequence for sRNA-A from

Table 1. sRNA sequences

sRNA	DNA sequence[a]
sRNA-A	TTTGACCTGATGAGTCCGTGAGGACGAAACGAGCTAGCTCGTC**GTCAAAGTTTGTTAT***ATTTGT* *AGAAATATTTTATTCGCCCCCGGAAGATCATTCCGGGGGCTTTTTTATT*
sRNA-B	GATGAGCTGATGAGCCCCGCACGGCGAAACTCAGCAGTTCGTC**CTCATCATTAAATTT***ATTTGT* *GTAAATATTTTACGCGCCCCCGGAAGGCGCTTCCGGGGGCTTTTTTATT*
sRNA-C	CTACACTGATGAGGTATGACCCACGAAACGTCCCTCTTCGTCT**GTAGTATAGTGTGAA***TTTGTTAT* *AATATTTTAACGGCCCCCGGATGAGTCTTCCGGGGGCTTTTTTATT*
sRNA-D	GCGTTACTGATGAGTCGCCTTTACCGAAACTTGGAGCCTCGTC**TAACGCCTTAGAATA***ATTTGCGC* *TAATATTTTAGCGGCCCCCGGAATTACGTTCCGGGGGCTTTTTTATT*
sRNA-E	AGTTTACTGATGAGCCCGCACCAACGAAACGCACCAGCTCGTC**TAAACTACCAGGAAT***ATTTGG* *TCTAATATTTTAGCCGCCCCCGGAATCGGTTTCCGGGGGCTTTTTTATT*
sRNA-F	GCCAAACCTGATGAGCGACCGTGCTCGAAACGGTAGTCCTCGTC**GTTTGGCTGTAGAGA***ATTTGA* *CTAAATATTTTAGTCGCCCCCGGAACAACTTTCCGGGGGCTTTTTTATT*

[a] Underline indicates ribozyme, bold indicates sRNA tag, and italics indicates sRNA scaffold.

pREP was integrated at site attP21 in *E. coli* DH10B cells (Supplementary Figure S1 and Materials and Methods). Once constructed, the reporter strain was tested against a wide range of RNA levels spanning 18–21,000 au from promoters P_{A1}-P_{A16} with the same expression plasmids used in Figure 2B (Supplementary Table S1). Even though the sRNA was characterized with a target expressed from a multi-copy plasmid (pREP and Figure 2B), target repression on the genome, constituting a drastic change in target copy number, was predictable. However, since the starting level of the reporter was low, repression saturates as reporter levels approach background (Figure 2F).

The sRNAs were then evaluated in other Gram-negative bacteria where Hfq is conserved (81) (Figure 2G). A single broad host range plasmid was constructed, from which sRNA-A and the corresponding GFP reporter gene are expressed (pBBR-RNA). The same vector without the sRNA expression cassette was also constructed and used for fold repression calculations (pBBR-EMPTY). Though the parts in pBBR-RNA are native to *E. coli*, they were previously evaluated and observed to exhibit activity in *Pseudomonas protegens* PF-5 and *Salmonella typhimurium* LT2. The sRNA repression assay was performed for both *Pseudomonas* and *Salmonella* cells harboring pBBR-RNA (Materials and Methods). sRNA activity was observed in both organisms (Figure 2G).

Combinatorial sRNA pool construction and validation

A library was built that varies the expression levels of each of the sRNAs. Ninety-six promoters (16 different promoters for each of 6 sRNAs) spanning strengths of 18–40 000 au were gleaned from a published set (Supplemental Table S1) (8). Each of the 96 promoter–sRNA combinations were evaluated for repression of both GFP and RFP to capture any variability due to the target gene sequence (Figure 3A and Supplementary Figure S7). For each sRNA, there is a strong correlation between promoter strength and repression (Figure 3A). Intermediate promoter strengths lead to a graded response. There is only a slight reduction in growth rate when all six sRNAs are maximally expressed (Supplementary Figure S2 and Materials and Methods).

The next step was to build the sRNA pool by randomly combining the promoter–sRNA pairs so that each member of the pool expresses the six sRNAs at different levels. The sRNA pool was constructed using CombiGEM assembly (64). Each of the 96 sRNA assembly plasmids contain strategically placed restriction sites that encompass a 10nt sequencing barcode associated with a specific promoter/sRNA pair, unique constitutive promoter, sRNA, and a strong terminator (85) (Figure 3B, Supplementary Table S1). The assembly reactions are then performed, the result of which is a barcode string that can be used to determine the promoter identity at each sRNA position with a 90 bp sequencing read (Figure 3C, Materials and Methods). Deep sequencing of the final pool resulted in the recovery of 80/96 promoters from ~200 000 reads and an even distribution of promoter strengths (Figure 3D).

All of the sRNAs are oriented in the same direction on the construct and thus could be influenced by transcriptional readthrough. To evaluate this effect, we built constructs that permuted the order of the sRNAs. No effect of ordering the sRNA genes on repression was observed, likely due to the use of strong terminators (Supplementary Figure S8).

Application to a metabolic pathway: optimization of β-carotene

The β-carotene pathway was selected as a model system to test our sRNA pool because it is known that enzyme balancing is important to increase titer (5,37,86–88). A single construct was built that includes the four-gene pathway to β-carotene (*crtE*, *crtB*, *crtI*, crtY) as well as the *E. coli* genes *dxs* and *idi*, which have been shown to improve titers when overexpressed (Figure 4A and B) (37,89). Each enzyme gene was placed under the control of a strong T7 promoter and is preceded by an insulating ribozyme, strong RBS designed for that gene (Materials and Methods), and sRNA target sequence (A–F) (Figure 4B and Supplementary Table S3). A unique strong terminator was placed after each gene. Transcription is induced from the promoters by expressing T7 RNA polymerase (RNAP) from a separate plasmid under IPTG-inducible control (59).

The sRNA pool was transformed into a strain harboring the pathway, generating a library of variants. More than 1200 colonies were picked, induced with 1 mM IPTG, and assayed for β-carotene production (Materials and Methods). Approximately half of the variants improved yield

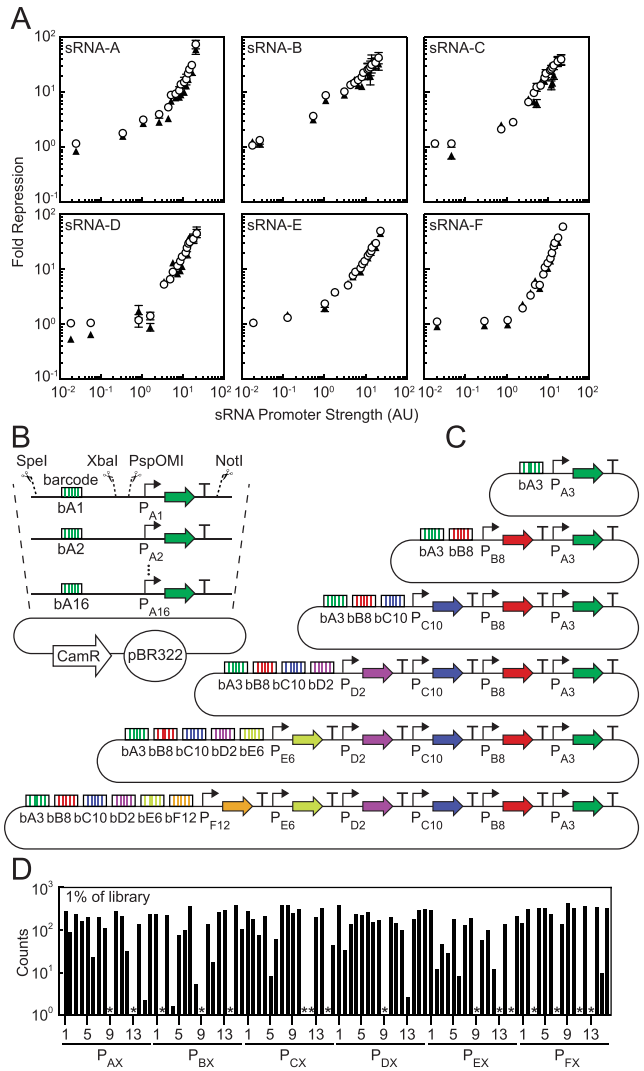

Figure 3. Pool assembly from differentially expressed sRNAs. (**A**) Repression response for all six sRNAs against GFP (circles) and RFP (triangles) genes containing targets in their 5'-UTR. Each sRNA is expressed from a set of 16 promoters unique to each sRNA. Sequences of all promoters are provided in Supplementary Table S1. Error bars correspond to the standard deviation of three experiments performed on different days. (**B**) Example of pool assembly plasmids with restriction sites for CombiGEM assembly. The barcode symbol depicts a unique 10nt sequencing barcode corresponding to each promoter in the pool. Sequences for promoters and barcodes are provided in Supplementary Table S1. (**C**) Iterative assembly example of a six sRNA pool variant. (**D**) Deep sequencing results for promoter coverage in the assembled pool containing six RNAs (1% pool coverage with ~200 000 reads); '*' signifies that the promoter was not detected. In the promoter names, X refers to the promoter number above the horizontal line.

from the starting titer of 43 ± 7 mg/l, with the best one producing a titer of 200 ± 30 mg/l (Figure 4C).

sRNA tuning plasmids from the top ten strains (161–204 mg/l) were isolated and barcodes sequenced in order to determine the promoters controlling the expression of each sRNA. From this, the repression of each gene can be elucidated for these strains (Figure 4D). These data suggest that the pathway is most sensitive to expression levels of Dxs and CrtY. In all ten high titer strains, Dxs is expressed at a low level (>10-fold repression). This enzyme is known to cause growth defects when overexpressed (90), Conversely, the final enzyme in the pathway CrtY is highly expressed (<2.5-fold repression).

It is not ideal to have to continuously transcribe a set of sRNAs in order to achieve the optimal balance. Instead, we sought to recapitulate this balance without sRNAs by constructing a pathway with T7 promoters that correspond to the optimal expression level for each enzyme. Once the top strain was identified using the sRNA library, the fold-repression of each gene due to the sRNA was inferred from the data in Figure 3A (Figure 4D). Then, using a previously-characterized T7 RNAP promoter library, we selected a promoter that generates the optimal expression level for each gene (Figure 4E and Supplementary Table S4). After these substitutions and with no additional tuning, the pathway produces 160 ± 20 mg/l β-carotene (Figure 4F).

The T7 promoter substitutions were based on the optimal repression of each gene, as inferred indirectly using a fluorescent reporter. RT-qPCR was performed to validate whether these indirect measurements correspond to the reduction in the expression of β-carotene genes after promoter substitution (Supplementary Figure S9). Indeed, there observed change in transcript levels closely correlate with the responses measured with a fluorescent reporter.

Application to a genetic circuit

A genetic circuit was chosen to demonstrate that the same pool could be applied to the optimization of a very different type of system. A two-input XNOR circuit was constructed using repressor-based NOT and NOR gates (Figure 5). Previously, we have demonstrated that we can computationally match the response functions of the gates to build this circuit (7). However, as a proof-of-principle, we took a different approach here, where strong RBSs were used to control the expression of each repressor (Materials and Methods). The sRNA target sequences (targets A–D) were placed between the ribozyme insulators and the RBS of each gene. As would be expected, this 'blindly' constructed XNOR circuit does not produce the correct logic operation (Figure 5D).

Because the circuit output is a fluorescent protein, fluorescence-activated cell sorting (FACS) can be used to rapidly screen for improvements. The sRNA pool was co-transformed with the circuit construct and then the library was serially sorted for correct output under the different combinations of inputs (1 mM IPTG and 2 ng/ml aTc) (Figure 5E) (Materials and Methods). After sorting, 96 colonies were individually assayed for the XNOR function, and 21 exhibited proper functionality. The best performing circuit exhibited the correct XNOR logic (Figure 5F). The corresponding sRNA construct was sequenced to determine the barcodes that identify the promoters controlling each sRNA, from which the fold-repression could be inferred (Figure 5G).

While the circuit now works, it is not ideal to have to continue to transcribe the sRNAs for proper function. In contrast to the β-carotene pathway, the circuit's promoters cannot easily be changed to different strengths because they contain operators that bind to the repressors. Therefore, RBS substitutions were used to recapitulate the op-

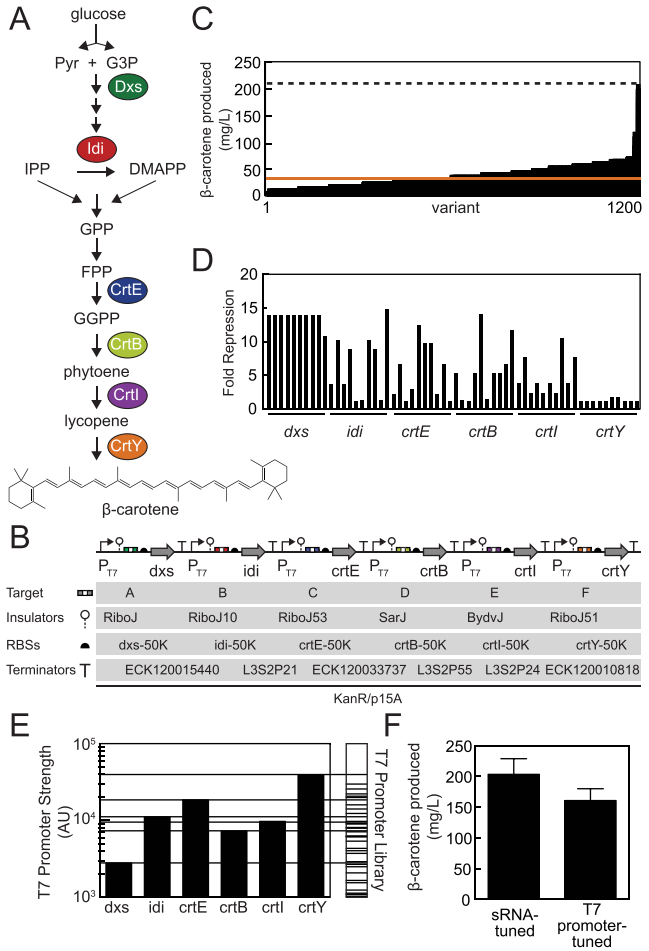

Figure 4. Optimization of β-carotene production using the sRNA pool. (**A**) The metabolic pathway for β-carotene production is shown. Enzymes that are essential or known to improve β-carotene production are circled and colored. (**B**) The genetic design for the β-carotene production pathway including the insertion of sRNA targets. The plasmid map and part sequences are provided in Supplementary Figure S1 and Table S3. (**C**) β-carotene production for 1200 screened variants. The orange line represents the titer from the un-tuned β-carotene pathway where an empty vector is used in place of a sRNA pool plasmid. Quantification of titers is described in Materials and Methods. (**D**) The repression of each gene for the ten highest β-carotene producing strains (in order from left to right) inferred from the promoter strength controlling each sRNA. (**E**) Calculated optimal T7 promoter strength for each gene (left axis). The box to the right shows T7 promoters available in the library. Solid lines extended across the graph represent T7 promoters from the library that were selected to hard-code the optimal expression levels. T7 promoter sequences are provided in Supplementary Table S4. (**F**) β-carotene production for the top sRNA-tuned pathway and its corresponding reconstructed pathway with substituted T7 promoters. Error bars correspond to the standard deviation of three experiments performed on different days.

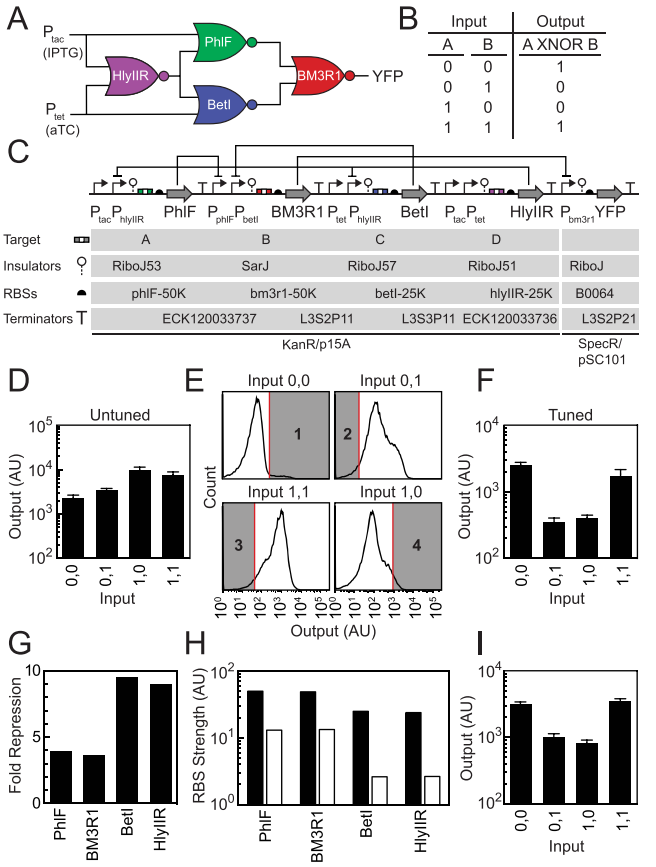

Figure 5. Optimization of a XNOR circuit with the sRNA pool. (**A**) The circuit diagram for the XNOR function is shown. Inside each NOR gate, the repressor name is shown. (**B**) Truth table for a two-input XNOR gate. (**C**) The genetic design for the XNOR circuit is shown. Blunt ended lines represent the repression of promoters. The plasmid map and part sequences are provided in Supplementary Figure S1 and Table S3. (**D**) Response of an untuned XNOR circuit where empty vector is used in place of a sRNA pool plasmid. Inputs correspond to the absence or presence of 1 mM IPTG (left '0' / '1') and aTC (2 ng/ml; right '0' / '1'). (**E**) Sort gates (shown in gray) used for each input combination are shown and numbered in the order in which they were applied to the population. The sorting procedure is described in Materials and Methods. (**F**) Response of the best sRNA-tuned XNOR circuit. (**G**) The sRNA-mediated knockdown of each repressor gene in the tuned circuit. (**H**) Calculated RBS strengths for the original untuned circuit (black bars) and for the reconstructed (new RBSs) tuned circuit (white bars) (Materials and Methods). (**I**) Response of the circuit with the substituted RBSs. For all data, error bars correspond to the standard deviation of three experiments performed on different days.

timal levels. The fold-repression of each gene achieved by the sRNAs was inferred from Figure 3A. The RBS Calculator was used to predict the strengths of the unoptimized RBSs and these were divided by the fold-repressions in order to obtain the desired strengths of the new RBSs (Figure 5H). The RBS Calculator was then used to generate RBS sequences corresponding to these target strengths. The circuit was reconstructed with the new RBSs and was found to be functional (Figure 5I).

DISCUSSION

We developed a method to tune genetic systems by exploring large gene expression spaces without having to build *ad hoc* libraries. This method can be used to tune a variety of multi-gene systems, and presents a novel approach to genetic system optimization by using trans-acting RNA regulators that target standardized target sequences. Further, once a functional variant is discovered from the sRNA pool, this can be used to determine the optimal gene expression levels, which can then be hard-coded by selecting corresponding promoters or RBSs to build the final construct. This speaks to the value of having libraries of standardized genetic parts that have been sufficiently characterized such

that they can be rationally selected to achieve a target design objective.

One limitation of our approach is that genes can only be repressed, as opposed to being up-regulated. This constrains the expression space to the maximum level of the initial promoters. This can be problematic if any of the genes are toxic. This was averted in the β-carotene pathway by using T7 promoters and the expression of T7 RNAP on a separate plasmid. This avoids the problem of toxicity because the genes will not be expressed until T7 RNAP is induced. Up-regulating sRNAs have been discovered (91,92) and may be amenable to a tuning pool; however, systematic characterization and engineering of these sRNAs is still required.

The reusable sRNA pool presented in this work establishes a new paradigm for genetic system optimization. Our approach is to build a library of regulators once and then use this same library to rapidly optimize different systems. Combinatorial optimization in genetic engineering has proven very successful, but the corresponding libraries are expensive, technically difficult to build, and require extensive sequence verification to ensure quality. It is particularly challenging if the constructs are large, reducing the efficiency of transformation, or have to be inserted chromosomally. Our approach only requires that the construct be built – including the sRNA target sequences – and introduced into the cell once. This one construct can be co-transformed with the plasmid-borne pool and screened for the optimal construct. It is sufficiently simple to be routinely or systematically applied when building constructs for new pathways, for example when mining pathways from metagenomic datasets. Often, the necessary expression levels in a heterologous host are unknown and the pooled approach allows many 'shots on goal' to be taken rather than risking the selection of potentially suboptimal parts for the control of each gene. Further, this method can be applied simply, without the need for DNA construction robotics and automation platforms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

US Defense Advanced Research Projects Agency [DARPA HR0011-15-c-0084, DARPA HR0011-12-C-4016 to C.A.V. and A.G.]; Office of Naval Research Multidisciplinary University Research Initiative [N00014-16-1-2388].
Conflict of interest statement. None declared.

REFERENCES

1. Smolke,C.D., Martin,V.J.J. and Keasling,J.D. (2001) Controlling the metabolic flux through the carotenoid pathway using directed mRNA processing and stabilization. *Metab. Eng.*, **3**, 313–321.
2. Stephanopoulos,G. (1999) Metabolic fluxes and metabolic engineering. *Metab. Eng.*, **1**, 1–11.
3. San,K.-Y., Bennett,G.N., Berríos-Rivera,S.J., Vadali,R.V., Yang,Y.-T., Horton,E., Rudolph,F.B., Sariyar,B. and Blackwood,K. (2002) Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in Escherichia coli. *Metab. Eng.*, **4**, 182–192.
4. Lee,M.E., Aswani,A., Han,A.S., Tomlin,C.J. and Dueber,J.E. (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.*, **41**, 10668–10678.
5. Zelcbuch,L., Antonovsky,N., Bar-Even,A., Levin-Karp,A., Barenholz,U., Dayagi,M., Liebermeister,W., Flamholz,A., Noor,E., Amram,S. *et al.* (2013) Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.*, **41**, e98.
6. Hooshangi,S., Thiberge,S. and Weiss,R. (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 3581–3586.
7. Nielsen,A.A.K., Der,B.S., Shin,J., Vaidyanathan,P., Paralanov,V., Strychalski,E.A., Ross,D., Densmore,D. and Voigt,C.A. (2016) Genetic circuit design automation. *Science*, **352**, aac7341.
8. Mutalik,V.K., Guimaraes,J.C., Cambray,G., Lam,C., Christoffersen,M.J., Mai,Q.-A., Tran,A.B., Paull,M., Keasling,J.D., Arkin,A.P. *et al.* (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods*, **10**, 354–360.
9. Brophy,J.A.N. and Voigt,C.A. (2014) Principles of genetic circuit design. *Nat. Methods*, **11**, 508–520.
10. Lee,M.E., Aswani,A., Han,A.S., Tomlin,C.J. and Dueber,J.E. (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.*, **41**, 10668–10678.
11. Slusarczyk,A.L., Lin,A. and Weiss,R. (2012) Foundations for the design and implementation of synthetic genetic circuits. *Nat. Rev. Genet.*, **13**, 406–420.
12. Zhou,H., Vonk,B., Roubos,J.A., Bovenberg,R.A.L. and Voigt,C.A. (2015) Algorithmic co-optimization of genetic constructs and growth conditions: application to 6-ACA, a potential nylon-6 precursor. *Nucleic Acids Res.*, **43**, 10560–10570.
13. Kauffman,S.A. (1993) *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press, NY.
14. Kosuri,S., Goodman,D.B., Cambray,G., Mutalik,V.K., Gao,Y., Arkin,A.P., Endy,D. and Church,G.M. (2013) Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 14024–14029.
15. Carrier,T.A. and Keasling,J.D. (1999) Library of synthetic 5′ secondary structures to manipulate mRNA stability in Escherichia coli. *Biotechnol. Progr.*, **15**, 58–64.
16. Pfleger,B.F., Pitera,D.J., Smolke,C.D. and Keasling,J.D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotech.*, **24**, 1027–1032.
17. Farasat,I., Kushwaha,M., Collens,J., Easterbrook,M., Guido,M. and Salis,H.M. (2014) Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.*, **10**, 731.
18. Nowroozi,F.F., Baidoo,E.E.K., Ermakov,S., Redding-Johanson,A.M., Batth,T.S., Petzold,C.J. and Keasling,J.D. (2014) Metabolic pathway optimization using ribosome binding site variants and combinatorial gene assembly. *Appl. Microbiol. Biotechnol.*, **98**, 1567–1581.
19. Coussement,P., Maertens,J., Beauprez,J., Van Bellegem,W. and De Mey,M. (2014) One step DNA assembly for combinatorial metabolic engineering. *Metab. Eng.*, **23**, 70–77.
20. Smanski,M.J., Bhatia,S., Zhao,D., Park,Y., Woodruff,L.B.A., Giannoukos,G., Ciulla,D., Busby,M., Calderon,J., Nicol,R. *et al.* (2014) Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.*, **32**, 1241–1249.
21. Woodruff,L.B.A., Gorochowski,T.E., Roehner,N., Mikkelsen,T.S., Densmore,D., Gordon,D.B., Nicol,R. and Voigt,C.A. (2017) Registry in a tube: multiplexed pools of retrievable parts for genetic design space exploration. *Nucleic Acids Res.*, **45**, 1567–1568.
22. Casini,A., Storch,M., Baldwin,G.S. and Ellis,T. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.*, **16**, 568–576.
23. Almquist,J., Cvijovic,M., Hatzimanikatis,V., Nielsen,J. and Jirstrand,M. (2014) Kinetic models in industrial biotechnology – improving cell factory performance. *Metab. Eng.*, **24**, 38–60.
24. Link,H., Christodoulou,D. and Sauer,U. (2014) Advancing metabolic models with kinetic information. *Curr. Opin. Biotechnol.*, **29**, 8–14.

25. Saha,R., Chowdhury,A. and Maranas,C.D. (2014) Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr. Opin. Biotechnol.*, **29**, 39–45.
26. Cooling,M.T., Rouilly,V., Misirli,G., Lawson,J., Yu,T., Hallinan,J. and Wipat,A. (2010) Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics*, **26**, 925–931.
27. Haseltine,E.L. and Arnold,F.H. (2007) Synthetic gene circuits: design with directed evolution. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 1–19.
28. Pirie,C.M., De Mey,M., Prather,K.L.J. and Ajikumar,P.K. (2013) Integrating the protein and metabolic engineering toolkits for next-generation chemical biosynthesis. *ACS Chem. Biol.*, **8**, 662–672.
29. Bordbar,A., Monk,J.M., King,Z.A. and Palsson,B.O. (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.*, **15**, 107–120.
30. Long,C.P. and Antoniewicz,M.R. (2014) Metabolic flux analysis of Escherichia coli knockouts: lessons from the Keio collection and future outlook. *Curr. Opin. Biotechnol.*, **28**, 127–133.
31. Rocha,I., Maia,P., Evangelista,P., Vilaça,P., Soares,S., Pinto,J.P., Nielsen,J., Patil,K.R., Ferreira,E.C. and Rocha,M. (2010) OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst. Biol.*, **4**, 45.
32. Xu,P., Rizzoni,E.A., Sul,S.-Y. and Stephanopoulos,G. (2017) Improving metabolic pathway efficiency by statistical model-based multivariate regulatory metabolic engineering. *ACS Synth. Biol.*, **6**, 148–158.
33. Brophy,J.A. and Voigt,C.A. (2016) Antisense transcription as a tool to tune gene expression. *Mol. Syst. Biol.*, **12**, 854.
34. Klein,J.C., Lajoie,M.J., Schwartz,J.J., Strauch,E.-M., Nelson,J., Baker,D. and Shendure,J. (2016) Multiplex pairwise assembly of array-derived DNA oligonucleotides. *Nucleic Acids Res.*, **44**, e43.
35. Bonde,M.T., Kosuri,S., Genee,H.J., Sarup-Lytzen,K., Church,G.M., Sommer,M.O.A. and Wang,H.H. (2015) Direct mutagenesis of thousands of genomic targets using microarray-derived oligonucleotides. *ACS Synth. Biol.*, **4**, 17–22.
36. Stanton,B.C., Nielsen,A.A.K., Tamsir,A., Clancy,K., Peterson,T. and Voigt,C.A. (2014) Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.*, **10**, 99–105.
37. Wang,H.H., Isaacs,F.J., Carr,P.A., Sun,Z.Z., Xu,G., Forest,C.R. and Church,G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.
38. Voigt,C.A., Kauffman,S. and Wang,Z.-G. (2001) Rational evolutionary design: the theory of in vitro protein evolution. *Adv. Protein Chem.*, **55**, 79–160.
39. Kang,Z., Wang,X., Li,Y., Wang,Q. and Qi,Q. (2012) Small RNA RyhB as a potential tool used for metabolic engineering in Escherichia coli. *Biotechnol. Lett.*, **34**, 527–531.
40. Liu,Y., Zhu,Y., Li,J., Shin,H.-d., Chen,R.R., Du,G.,, Liu,L. and Chen,J. (2014) Modular pathway engineering of Bacillus subtilis for improved N-acetylglucosamine production. *Metab. Eng.*, **23**, 42–52.
41. 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. Biotech.*, **31**, 170–174.
42. Yakandawala,N., Romeo,T., Friesen,A.D. and Madhyastha,S. (2008) Metabolic engineering of Escherichia coli to enhance phenylalanine production. *Appl. Microbiol. Biotechnol.*, **78**, 283–291.
43. Gottesman,S. (2004) The small RNA regulators of Escherichia coli: roles and mechanisms*. *Annu. Rev. Microbiol.*, **58**, 303–328.
44. Lalaouna,D., Simoneau-Roy,M., Lafontaine,D. and Massé,E. (2013) Regulatory RNAs and target mRNA decay in prokaryotes. *Biochim. Biophys.*, **1829**, 742–747.
45. Storz,G., Vogel,J. and Wassarman,K.M. (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell*, **43**, 880–891.
46. Sharma,V., Yamamura,A. and Yokobayashi,Y. (2012) Engineering artificial small RNAs for conditional gene silencing in Escherichia coli. *ACS Synth. Biol.*, **1**, 6–13.
47. Sakai,Y., Abe,K., Nakashima,S., Yoshida,W., Ferri,S., Sode,K. and Ikebukuro,K. (2013) Improving the gene-regulation ability of small RNAs by scaffold engineering in Escherichia coli. *ACS Synth. Biol.*, **3**, 152–162.
48. Man,S., Cheng,R., Miao,C., Gong,Q., Gu,Y., Lu,X., Han,F. and Yu,W. (2011) Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Res.*, **39**, e50.
49. Amman,F., Flamm,C. and Hofacker,I. (2012) Modelling translation initiation under the influence of sRNA. *Int. J. Mol. Sci.*, **13**, 16223–16240.
50. 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.
51. 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.
52. Durfee,T., Nelson,R., Baldwin,S., Plunkett,G., Burland,V., Mau,B., Petrosino,J.F., Qin,X., Muzny,D.M. and Ayele,M. (2008) The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *J. Bacteriol.*, **190**, 2597–2606.
53. Blattner,F.R., Plunkett,G., Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K. and Mayhew,G.F. (1997) The complete genome sequence of Escherichia coli K-12. *Science*, **277**, 1453–1462.
54. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, doi:10.1038/msb4100050.
55. St-Pierre,F.o., Cui,L., Priest,D.G., Endy,D., Dodd,I.B. and Shearwin,K.E. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.*, **2**, 537–541.
56. Cormack,B.P., Valdivia,R.H. and Falkow,S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, **173**, 33–38.
57. Campbell,R.E., Tour,O., Palmer,A.E., Steinbach,P.A., Baird,G.S., Zacharias,D.A. and Tsien,R.Y. (2002) A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7877–7882.
58. Lou,C., Stanton,B., Chen,Y.-J., Munsky,B. and Voigt,C.A. (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotech.*, **30**, 1137–1142.
59. Temme,K., Hill,R., Segall-Shapiro,T.H., Moser,F. and Voigt,C.A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.*, **40**, 8773–8781.
60. Zhou,J. and Rudd,K.E. (2013) EcoGene 3.0. *Nucleic Acids Res.*, **41**, D613–D624.
61. 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.
62. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotech.*, **27**, 946–950.
63. Quan,J. and Tian,J. (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.*, **6**, 242–251.
64. Cheng,A.A., Ding,H. and Lu,T.K. (2014) Enhanced killing of antibiotic-resistant bacteria enabled by massively parallel combinatorial genetics. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 12462–12467.
65. Dower,W.J., Miller,J.F. and Ragsdale,C.W. (1988) High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res.*, **16**, 6127–6145.
66. Antoine,R. and Locht,C. (1992) Isolation and molecular characterization of a novel broad-host-range plasmid from Bordetella bronchiseptica with sequence similarities to plasmids from Gram-positive organisms. *Mol. Microbiol.*, **6**, 1785–1799.
67. Zhou,K., Zhou,L., Lim,Q.E., Zou,R., Stephanopoulos,G. and Too,H.-P. (2011) Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR. *BMC Mol. Biol.*, **12**, 18.
68. Gao,Y. and Zhao,Y. (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.*, **56**, 343–349.
69. Park,H., Bak,G., Kim,S.C. and Lee,Y. (2013) Exploring sRNA-mediated gene silencing mechanisms using artificial small RNAs derived from a natural RNA scaffold in Escherichia coli. *Nucleic Acids Res.*, **41**, 3787–3804.
70. Mückstein,U., Tafer,H., Hackermüller,J., Bernhart,S.H., Stadler,P.F. and Hofacker,I.L. (2006) Thermodynamics of RNA–RNA binding. *Bioinformatics*, **22**, 1177–1182.

71. Sledjeski,D.D., Gupta,A. and Gottesman,S. (1996) The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli. *EMBO J.*, **15**, 3993–4000.
72. Urbanowski,M.L., Stauffer,L.T. and Stauffer,G.V. (2000) The gcvB gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in Escherichia coli. *Mol. Microbiol.*, **37**, 856–868.
73. Chen,S., Zhang,A., Blyn,L.B. and Storz,G. (2004) MicC, a second small-RNA regulator of Omp protein expression in Escherichia coli. *J. Bacteriol.*, **186**, 6689–6697.
74. Andersen,J., Delihas,N., Ikenaka,K., Green,P.J., Pines,O., Ilercil,O. and Inouye,M. (1987) The isolation and characterization of RNA coded by the micF gene in Escherichia Coli. *Nucleic Acids Res.*, **15**, 2089–2101.
75. Vanderpool,C.K. and Gottesman,S. (2004) Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol. Microbiol.*, **54**, 1076–1089.
76. Sahagan,B.G. and Dahlberg,J.E. (1979) A small, unstable RNA molecule of Escherichia coli: spot 42 RNA. *J. Mol. Biol.*, **131**, 573–592.
77. Aiba,H. (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.*, **10**, 134–139.
78. Otaka,H., Ishikawa,H., Morita,T. and Aiba,H. (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 13059–13064.
79. Ishikawa,H., Otaka,H., Maki,K., Morita,T. and Aiba,H. (2012) The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3′ poly(U) tail. *RNA*, **18**, 1062–1074.
80. Wang,W., Wang,L., Wu,J., Gong,Q. and Shi,Y. (2013) Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA. *Nucleic Acids Res.*, **41**, 5938–5948.
81. Møller,T., Franch,T., Højrup,P., Keene,D.R., Bächinger,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.
82. Mendoza-Vargas,A., Olvera,L., Olvera,M., Grande,R., Vega-Alvarado,L., Taboada,B., Jimenez-Jacinto,V., Salgado,H., Juárez,K., Contreras-Moreira,B. *et al.* (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE*, **4**, e7526.
83. Ruffner,D.E., Stormo,G.D. and Uhlenbeck,O.C. (1990) Sequence requirements of the hammerhead RNA self-cleavage reaction. *Biochemistry*, **29**, 10695–10702.
84. Lovett,S.T., Hurley,R.L., Sutera,V.A., Aubuchon,R.H. and Lebedeva,M.A. (2002) Crossing over between regions of limited homology in Escherichia coli: RecA-dependent and RecA-independent pathways. *Genetics*, **160**, 851–859.
85. Chen,Y.-J., Liu,P., Nielsen,A.A.K., Brophy,J.A.N., Clancy,K., Peterson,T. and Voigt,C.A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods*, **10**, 659–664.
86. Farmer,W.R. and Liao,J.C. (2000) Improving lycopene production in Escherichia coli by engineering metabolic control. *Nat. Biotech.*, **18**, 533–537.
87. Kim,S.-W. and Keasling,J.D. (2001) Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in Escherichia coli enhances lycopene production. *Biotechnol. Bioeng.*, **72**, 408–415.
88. Sun,T., Miao,L., Li,Q., Dai,G., Lu,F., Liu,T., Zhang,X. and Ma,Y. (2014) Production of lycopene by metabolically-engineered Escherichia coli. *Biotechnol. Lett.*, **36**, 1515–1522.
89. Yoon,S.-H., Park,H.-M., Kim,J.-E., Lee,S.-H., Choi,M.-S., Kim,J.-Y., Oh,D.-K., Keasling,J.D. and Kim,S.-W. (2007) Increased β-carotene production in recombinant Escherichia coli harboring an engineered isoprenoid precursor pathway with mevalonate addition. *Biotechnol. Progr.*, **23**, 599–605.
90. Matthews,P.D. and Wurtzel,E.T. (2000) Metabolic engineering of carotenoid accumulation in Escherichia coli by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl. Microbiol. Biotechnol.*, **53**, 396–400.
91. Soper,T., Mandin,P., Majdalani,N., Gottesman,S. and Woodson,S.A. (2010) Positive regulation by small RNAs and the role of Hfq. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 9602–9607.
92. Fröhlich,K.S. and Vogel,J. (2009) Activation of gene expression by small RNA. *Curr. Opin. Microbiol.*, **12**, 674–682.