nature biotechnology

ARTICLES

Engineered riboregulators enable post-transcriptional control of gene expression

Farren J Isaacs[1,2,3,4], Daniel J Dwyer[1,2,5], Chunming Ding[2,3], Dmitri D Pervouchine[1,2], Charles R Cantor[2,3,4] & James J Collins[1,2,3,4]

Recent studies have demonstrated the important enzymatic, structural and regulatory roles of RNA in the cell. Here we present a post-transcriptional regulation system in *Escherichia coli* that uses RNA to both silence and activate gene expression. We inserted a complementary *cis* sequence directly upstream of the ribosome binding site in a target gene. Upon transcription, this *cis*-repressive sequence causes a stem-loop structure to form at the 5′–untranslated region of the mRNA. The stem-loop structure interferes with ribosome binding, silencing gene expression. A small noncoding RNA that is expressed in *trans* targets the *cis*-repressed RNA with high specificity, causing an alteration in the stem-loop structure that activates expression. Such engineered riboregulators may lend insight into mechanistic actions of endogenous RNA-based processes and could serve as scalable components of biological networks, able to function with any promoter or gene to directly control gene expression.

RNA molecules are most often thought of as messengers of information from genes to the proteins they encode[1,2]. RNAs can also assume various structural, regulatory and enzymatic roles[3]. These noncoding RNAs (ncRNA) have diverse functions including synthesizing proteins, splicing and editing RNA, modifying rRNA[1–3] and catalyzing biochemical reactions[4–6]. Small regulatory RNAs (sRNA), a subset of ncRNAs, have emerged as important regulators in both prokaryotes[7,8] and eukaryotes[9], and have enhanced our understanding of how cells undergo development and respond to changes in the environment[3,10]. More recently, RNAs have been shown to act as environmental sensors of vitamin cofactors and temperature, enabling them to transduce signals to regulate gene expression[11–17]. Regulatory RNAs operate by sensing environmental cues or other RNA molecules to either repress or, more rarely, activate[18,19] translation. Such natural mechanisms, which involve post-transcriptional regulation, provide a basis for the development of synthetic riboregulators.

A number of approaches that exploit RNA's structural dynamics and sequence-specific binding abilities are used in RNA-mediated control of gene expression[20]. Antisense strategies for gene silencing, in which an antisense RNA binds and inhibits a target RNA, are actively being pursued[21,22]. Nucleic acid–based molecules[20] and, more recently, the discovery of RNA interference (RNAi)[23] have offered additional approaches for regulating gene expression. These methods typically require sequence-specific design and have proved valuable in post-transcriptional gene silencing. However, there exists a need for modular systems that can be integrated into biological networks to function with a wide array of genes. In this study, we use RNA's versatility to design artificial riboregulators that both repress and activate translation *in vivo*, enabling precise control of gene expression through highly specific RNA-RNA interactions.

RESULTS

Artificial riboregulation design

In contrast to previous engineered schemes for prokaryotic post-transcriptional regulation, in which repression is achieved through antisense RNA or *trans*-acting ribozymes[24,25], our approach obtains effective repression by formation of an RNA secondary structure that sequesters the ribosome binding site (RBS). A short nucleotide sequence complementary to the RBS is introduced in the DNA directly upstream of the RBS such that the 5′–untranslated region (UTR) of the mRNA sequence naturally folds to form a stem-loop structure encompassing the RBS (**Fig. 1**). The resulting transcript, which we refer to as *cis*-repressed mRNA (crRNA), blocks recognition of the RBS by the 30S ribosomal subunit. Inhibiting this interaction prevents subsequent steps required for translation of a functional protein. Activation of gene expression is achieved by a *trans*-acting, noncoding RNA, referred to as *trans*-activating RNA (taRNA), produced from a second promoter. The taRNA was designed to target and hybridize to the stem-loop of the crRNA message. The resulting RNA duplex causes a conformational change in the crRNA that unfolds the stem-loop, exposing the RBS and permitting translation.

We aimed to construct a modular post-transcriptional system able to function with numerous transcriptional and regulatory components. In this regard, because the *cis* element is located in the 5′-UTR, the crRNA and taRNA sequences that we designed do not target gene-specific sequences or require specific promoters. Several aspects of

[1]Center for BioDynamics, [2]Center for Advanced Biotechnology, [3]Bioinformatics Program, [4]Department of Biomedical Engineering, [5]Department of Biology, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA. Correspondence should be addressed to J.J.C. (jcollins@bu.edu)

Published online 20 June 2004; doi:10.1038/nbt986

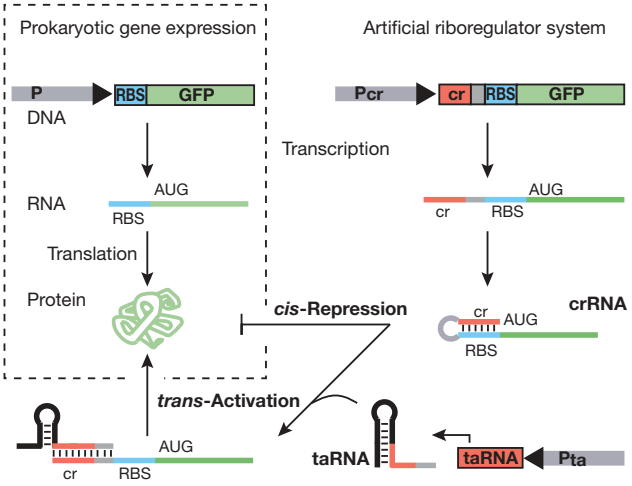

Figure 1 The artificial riboregulator system used to control post-transcriptional gene regulation. Basic steps of native prokaryotic gene expression are illustrated in the box. A promoter, P, drives the expression of a gene (GFP). After transcription, mRNA is present with a ribosome binding site (RBS) available for ribosome docking. After ribosome binding, translation of a functional protein occurs. In the artificial riboregulator, a small sequence (cr), complementary to the RBS, is inserted downstream from a promoter (P_{cr}) and upstream from the RBS. After transcription, a stem-loop is formed at the 5′ end of the mRNA, which blocks ribosome docking and translation (*cis* repression). The resulting mRNA is referred to as *cis*-repressed RNA (crRNA). A second promoter, P_{ta}, expresses a small, noncoding RNA (*trans*-activating RNA, taRNA) that targets the crRNA with high specificity. The taRNA and crRNA undergo a linear-loop interaction that exposes the obstructed RBS and activates expression. Different variants of crRNA and taRNA were constructed (see **Figs. 2** and **3**). Two riboregulator variants using different promoters were also studied to test the generality of the system. The first riboregulator system consisted of $P_{LtetO-1}$ (ref. 29) (P_{cr}) and P_{BAD} (P_{ta}). The second riboregulator was composed of $P_{LlacO-1}$ (ref. 29) (P_{cr}) and $P_{LtetO-1}$ (P_{ta}).

endogenous riboregulators[18,26,27] were used to guide the construction of the crRNA component. First, the *cis*-repressive sequence, a 19-nucleotide (nt) reverse complement sequence of the RBS, is strategically placed on the 5′-UTR directly downstream of the promoter and upstream of the RBS sequence (**Fig. 1**). Importantly, the introduced *cis* sequence does not alter the coding frame of the targeted gene and does not affect native transcription rates. Second, a short nucleotide sequence, placed between the *cis*-repressive sequence and the RBS, forms the loop and enables preferential formation of a stem-loop structure. Third, a motif that directs taRNA-crRNA binding through a linear-loop intermolecular interaction is included. This motif, a YUNR (pYrimidine-Uracil-Nucleotide-puRine) consensus sequence, has been shown to be a critical target for intermolecular RNA complexes in several endogenous systems[27]. The taRNA possesses a nucleotide sequence that is complementary to the *cis*-repressive sequence, and contains an RBS region. Therefore, it is critical that the RBS-containing sequence is sequestered in the taRNA stem structure to prevent aberrant titration of ribosomes. Because the intermolecular RNA interactions rely on specific RNA structures, we used Mfold[28] to generate predicted RNA secondary structures. Only sequences with a single predicted Mfold secondary structure for each variant were used to guide the construction of each RNA sequence.

Cis repression: intramolecular RNA pairings

To assess the *in vivo* repressive ability of the 5′-UTR *cis* element, we constructed four crRNA variants (crRL, crR7, crR10 and crRB) on *E. coli* plasmids. These crRNA variants (**Fig. 2a**) had different degrees of stem sequence complementarity to the RBS (**Table 1**). This allowed us to determine the extent of sequence complementarity required for post-transcriptional repression, and to investigate whether alternative RNA secondary structures (RNA duplexes (crRL), inner loops (crR7, crRB) and bulges (crR10)) destabilize the stem loop to help generate an open complex when targeted for activation by taRNA. A 25-nt *cis*-repressive sequence was inserted 27 nt downstream of the $P_{LtetO-1}$ promoter[29], so that this sequence is present in the 5′-UTR of the mRNA (crRNA). The *cis* sequence consisted of two sections: a 19-nt stem sequence, complementary to the RBS, and a 6-nt loop region. A ribosome binding site from the pZ plasmid system[29] and the *gfp-mut3b*[30] gene were inserted directly downstream of the *cis* sequence. Single-cell flow cytometric measurements of green fluorescent protein (GFP, product of *gfpmut3b*) were used to monitor the expression

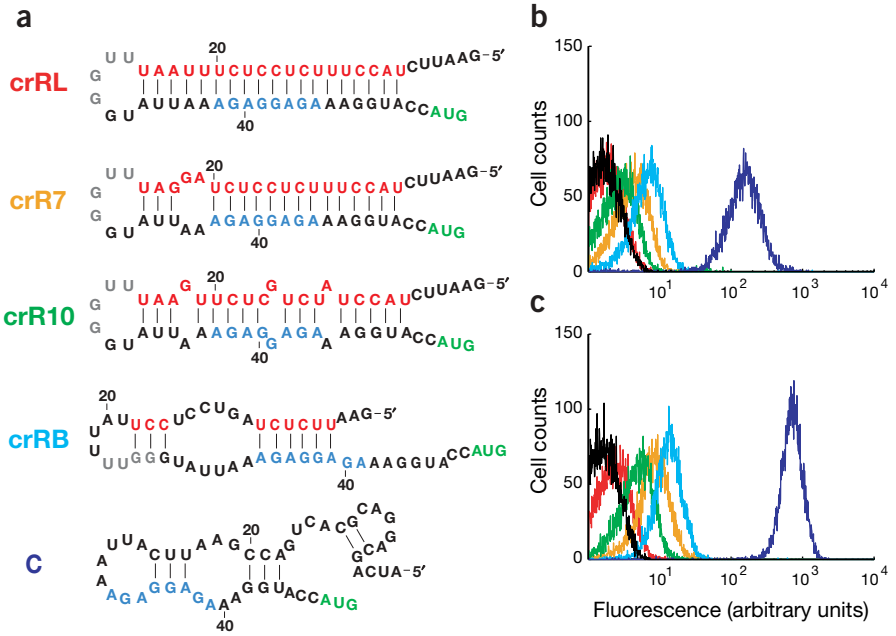

Figure 2 Results of *cis* repression of crRNA variants: crRL (red), crR7 (orange), crR10 (green), crRB (light blue) and control (dark blue). (**a**) Mfold-predicted[28] structures of the crRNA variants and the control structure, which has an arbitrary sequence upstream from the ribosome binding site in place of the *cis* element. The ribosome binding sites are shown in blue; YUNR recognition motif of loop in gray; *cis*-repressive (cr) sequence in red; start codon (AUG) in green. (**b**,**c**) Flow-cytometric results of the crRNA variants driving the expression of *gfpmut3b* at intermediate (**b**) and high (**c**) transcription rates. Histograms represent GFP expression of cultures containing each construct, color-coded in **a**. Black curve shows fluorescence measurement of cells containing plasmids that lack GFP (autofluorescence measurement).

Table 1 Protein (GFP) and mRNA results of crRNA constructs

		Control	crRL	crR7	crR10	crR12
% RBS Sequence Complementarity[a]		–	100	89	84	84
– aTc:	mRNA[b]	0.364 ± 0.077	0.135 ± 0.014	0.154 ± 0.022	0.152 ± 0.022	0.144 ± 0.033
	RNA normalized[c]	1	0.37	0.42	0.42	0.40
	FL1[d]	113.10 ± 15.8	2.55 ± 0.02	3.75 ± 0.19	3.41 ± 0.03	2.91 ± 0.05
	FL1 normalized[c]	1	0.022	0.033	0.030	0.026
+aTc:	mRNA[b]	1.53 ± 0.176	0.611 ± 0.113	0.629 ± 0.043	0.0628 ± 0.096	0.540 ± 0.098
	RNA normalized[c]	1	0.40	0.41	0.41	0.35
	FL1[d]	640.5 ± 25	4.06 ± 0.19	13.61 ± 1.12	10.05 ± 0.12	6.55 ± 0.14
	FL1 normalized[c]	1	0.006	0.021	0.016	0.010
$^{+aTc}/_{-aTc}$:	RNA[e]	4.2	4.5	4.1	4.1	3.8
	FL1[e]	5.7	1.6	3.6	2.9	2.2

[a]Percent RBS sequence complementarity of *cis* element to the RBS sequence. [b]Concentrations of mRNA obtained from competitive PCR coupled with MALDI-TOF mass spectrometry[49] are internally normalized to 16S rRNA levels within each sample. [c]Normalized RNA and FL1 values correspond to samples normalized to the control (crRNA/C), which lacks the *cis* sequence. [d]FL1 values (arbitrary units) represent measured GFP expression levels obtained by flow cytometry, where the autofluorescence FL1 mean value was measured to be 2.06 ± 0.03. [e]Fold induction (+aTc/–aTc) depicts the change in RNA and FL1 levels between high and low transcription rates within each column. See **Supplementary Notes** online for details on the experimental acquisition of these values.

state of this post-transcriptional control system. A control plasmid that lacks the *cis* element and contains an arbitrary sequence upstream of the RBS was also constructed.

Fluorescence data from *tetR*$^+$ cells containing control plasmids showed elevated GFP expression at intermediate (no anhydrotetracycline (aTc)) and high (30 ng/ml aTc) transcription rates (**Fig. 2b,c**). Cells possessing plasmids with upstream *cis*-repressive elements (crRNAs) were grown under the same conditions. At intermediate and high transcription rates, moderate GFP expression was detected in cultures containing the crRB variant, which has reduced *cis* sequence complementarity to the RBS. Thus, the crRB variant (**Fig. 2a**) was not investigated further. At intermediate transcription rates (**Fig. 2b**), crRL cultures showed repressed levels of GFP expression indistinguishable from autofluorescence (measured in cells containing plasmids that

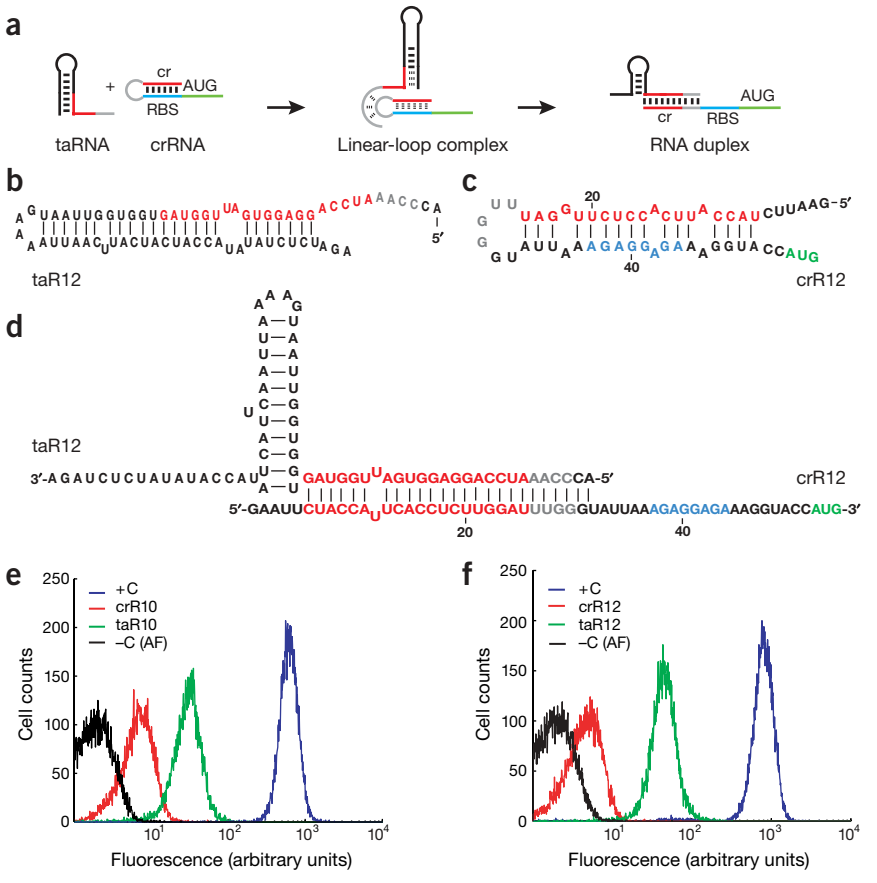

Figure 3 *Trans*-activation mechanism and results. (**a**) The artificial riboregulator system has the following proposed mechanism: (i) the 5′ linear region of the taRNA (gray) recognizes a YUNR consensus sequence (UUGG)[27] on the loop (gray) of crRNA, (ii) pairing between complementary nucleotides occurs in the presence of an unstable loop-tail complex and destabilizes the hairpin stem-loop that obstructs ribosomal recognition of the RBS (blue) and (iii) a stable intermolecular RNA duplex structure forms. The resulting RNA duplex exposes the RBS and allows translation to occur. (**b**,**c**) Mfold-predicted[28] structures of taR12 (**b**) and crR12 (**c**) variants (same color scheme as **Fig. 2**). (**d**) Proposed taR12-crR12 interaction that exposes the RBS, which is 5–6 bp downstream of the taRNA-crRNA duplex formation. (**e**,**f**) Flow-cytometric results of taR10-crR10 (**e**) and taR12-crR12 (**f**) riboregulator systems. Autofluorescence measurements (–C, negative control; cells lacking GFP) are in black and GFP expression of positive control (+ C; cells without *cis* sequence) cultures are in blue. The red curve represents *cis*-repressed cultures (no arabinose, 30 ng/ml aTc) and the green curve depicts cells containing high levels of taRNA (0.25% arabinose) and crRNA (30 ng/ml aTc). Of note, the taR12-crR12 riboregulator (**f**) showed both greater *cis* repression and higher *trans* activation than the taR10-crR10 riboregulator (**e**). Interestingly, both riboregulator variants possess the same sequence and predicted structure in the loop and share 12 of the first 13 potential duplex pairs in the *cis* stem, indicating that specificity of interaction emanates from slight changes in sequences of the *cis* elements. In the **Supplementary Notes** online, we describe various rational attempts to increase the dynamic range of the taR12-crR12 riboregulator pair.

Table 2 Measured *in vitro* and *in vivo* results obtained from intermolecular (taRNA-crRNA) and intramolecular (crRNA-crRNA) pairings

	crR7		crR10		crR12	
	K_D[a]	ΔFL1[b]	K_D[a]	ΔFL1[b]	K_D[a]	ΔFL1[b]
taR7	0.03	1×	12.5	1×	1.03	1×
taR10	1.00	1×	0.09	8×	0.77	1×
taR12	1.50	1×	7.70	1×	0.08	19×
$\Delta G_{crR\text{-}crR}$[c]	−76.0		−90.0		−100.1	

[a]K_D: equilibrium dissociation constants (μmol) measured by *in vitro* experiments. [b]ΔFL1 depicts the normalized fold change of fluorescence in the presence of taRNA (+arabinose/–arabinose). *In vitro* results are consistent with *in vivo* measurements. Observed differences may be caused by disparate conditions between *in vitro* and *in vivo* studies. See **Supplementary Notes** online for an explanation of the protocols used to calculate the equilibrium dissociation constants. [c]crRNAs, consisting of the stem-loop region, were isolated, and *in vitro* experiments were done to measure the equilibrium constants ($\Delta G_{crR\text{-}crR}$) of the intramolecular pairings (*cis*-repressive sequence-RBS interactions) for the 7, 10 and 12 crRNAs. The resulting $\Delta G_{crR\text{-}crR}$ values (kcal/mol) of crRNA stem-loop variants were obtained from RNA thermal melting curves (see **Supplementary Fig. 4** online).

lack GFP). The crR7 and crR10 cultures also showed dramatic silencing of gene expression at intermediate transcription rates (**Fig. 2b**). At high transcription rates (**Fig. 2c**), we observed low GFP expression values for all variants, indicating that the *cis*-repressive element provides striking suppression of post-transcriptional expression. Our results also indicate that the degree of repression is correlated with measured $\Delta G_{crR\text{-}crR}$ values (**Table 2**) and base-pairing in the stem region (crR7, crR10 and crR12 (see below)). The observed repression of >96% (intermediate transcription) and >97% (high transcription) provides greater silencing (98% repression is observed when autofluorescence is subtracted) compared to alternative antisense and *trans*-ribozyme systems[24,25], which exhibit 75% repression[25] .

To confirm that the observed silencing is due to the presence of translational repression by the *cis* sequence, we measured cellular mRNA concentrations. Total cell RNA was isolated from cultures containing each crRNA variant and the control plasmid. **Table 1** lists quantitative measurements of mRNA concentrations. We consistently observed a fourfold increase in mRNA concentration upon shifting from intermediate to high transcription rates (+aTc/–aTc). Additionally, the crRNA variants are present at 40% of the mRNA levels measured in the control cultures. This could be a result of premature transcription termination downstream of the stem-loop structure or targeted degradation by RNases that cleave double-stranded RNAs[31,32], and warrants further exploration. Despite the moderate drop in cellular mRNA concentrations, silencing observed at the protein level (GFP expression) was far greater (>96% repression) than the measured loss of RNA. Together with the GFP data, these results are consistent with the interpretation that the stem-loop, which is formed by the upstream *cis* sequence, prevents ribosome binding at the RBS and interferes with post-transcriptional gene expression.

Trans activation: intermolecular RNA pairings

To determine whether the silenced crRNAs can be structurally manipulated to expose the RBS and initiate translation, we produced *trans*-activating RNAs (taRNAs) to direct linear-loop (taRNA-crRNA) RNA pairing. This mode of RNA-RNA interaction (**Fig. 3a**) was modeled on several well-characterized natural RNA systems[18,27], in particular, the *hok/sok* post-segregational killing system of plasmid R1 (ref. 27). The resulting intermolecular RNA complex (**Fig. 3a**) exposes the RBS, which permits ribosomal recognition and translation. Because the final RNA complex could contain 26 consecutive base pairs, two mismatches were intentionally introduced to provide immunity from RNase III cleavage of RNA duplexes[31,32]. To assess the activation ability of each crRNA variant, we designed unique taRNA structures for each crRNA target, ensuring that the final duplex structures all contain 24 base pairs and two mismatches.

Three taRNA-crRNA cognate pairs (taRL-crRL, taR7-crR7 and taR10-crR10) were investigated, and the resulting activation of GFP expression in the presence of the small taRNAs was measured. These taRNA molecules were produced *in vivo* from the P_{BAD} promoter of the arabinose operon. Cultures containing the crRL and crR7 variants showed no detectable increase in GFP expression at high arabinose (0.25%) induction of taRL and taR7, respectively. However, upon induction of taR10, cultures containing crR10 exhibited a fivefold increase in GFP expression (see below). In light of these results, we constructed another taRNA-crRNA pair: taR12 and crR12 (**Fig. 3b,c**). The taR12-crR12 pair was designed both to verify that dispersed bulges render crRNA molecules susceptible to open complex formation in the presence of its target taRNA and to determine whether in-

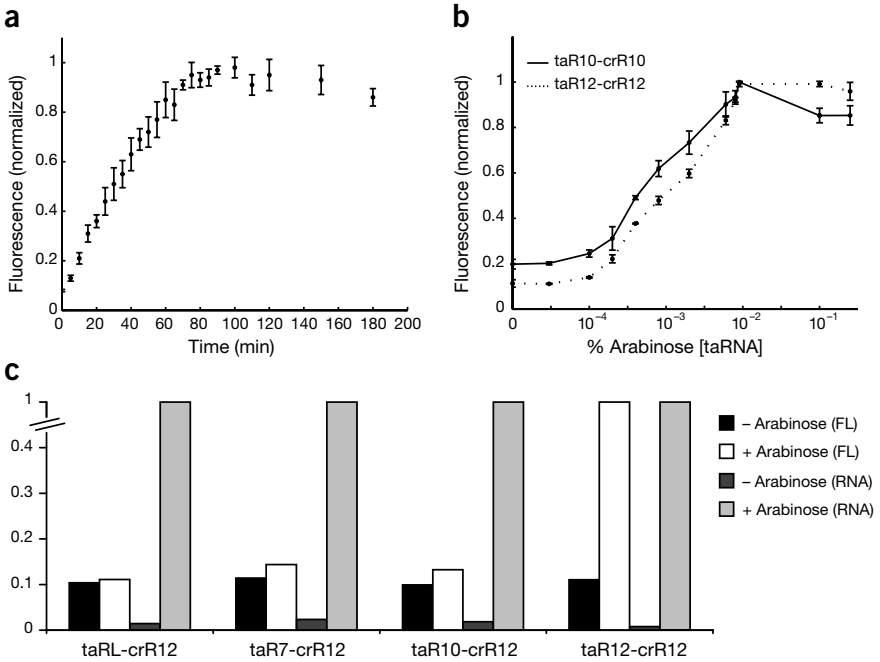

Figure 4 Transient and steady-state responses and specificity results of riboregulator systems. (**a**) Normalized transient response of taR12-crR12 riboregulator system at corresponding time points, where 0.25% arabinose was added at 0 min. (**b**) Normalized dose-response curves of taR10-crR10 (solid line) and taR12-crR12 (dotted line) riboregulators at corresponding concentrations of arabinose. The population averages in **a** and **b** were obtained from a uniform population of cells. (**c**) GFP fluorescence (black and white bars) and taRNA concentrations (light and dark gray bars) of four riboregulator variants (taRL-crR12, taR7-crR12, taR10-crR12, taR12-crR12) at low (black and dark gray) and high (white and light gray) arabinose concentrations. All presented data are normalized to high GFP and RNA levels.

creased dynamic range can be achieved. Whereas the crR12 variant, similar in structure to crR10, also contains three predicted dispersed bulges in its secondary structure, three important distinctions exist. First, the middle bulge on crR10 contains an exposed G that is a highly conserved nucleotide of the RBS and thus an important nucleotide in ribosome binding (**Fig. 2a**). Instead of having the conserved G exposed, the middle bulge in crR12 has the adjacent nucleotide, A, exposed (**Fig. 3c**). Here, the rationale was to leave a less recognizable nucleotide exposed to potential ribosome binding. Second, the crR12 contains an additional G-C base pairing in the stem (six G-C pairs versus five G-C pairs in crR10) to make crR12 a more stable structure than crR10. Lastly, the number of base pairings in the final intermolecular taR12-crR12 interaction was increased from 24 (taR10-crR10) to 25 (**Fig. 3d**). Together, these modifications were designed to increase the dynamic range in the taR12-crR12 pair as compared to the fivefold increase in GFP expression in the taR10-crR10 pair (**Fig. 3e**).

In the absence of arabinose, cells containing crR12 showed low fluorescence (near autofluorescence). Upon arabinose induction, GFP expression increased tenfold (**Fig. 3f**). These results suggest that partial helix destabilization (that is, presence of bulges) in crRNA is necessary for the taRNA to form a stable intermolecular RNA duplex, enabling protein translation. When background autofluorescence is subtracted, we observed eightfold and 19-fold increases in GFP expression for the taR10-crR10 and taR12-crR12 riboregulators, respectively. Interestingly, our engineered riboregulators exhibited greater levels of activation than the endogenous DsrA-RpoS system, which has only threefold activation[18].

The transient and steady-state responses of *trans* activation of GFP expression were also investigated. A high concentration (0.25% arabinose) of taR12 was induced in cultures containing high levels (30 ng/ml aTc) of crR12, enabling GFP expression to be measured as a function of time (**Fig. 4a**). An increase in GFP expression was detected at the first time point (5 min after arabinose induction), indicating an immediate response to the presence of taRNA. Maximal GFP expression was observed within 70 min, and this state was maintained for the duration of the experiment. The steady-state taRNA dependence of *trans* activation of GFP expression as modulated by arabinose concentrations is shown in **Figure 4b**. Qualitatively similar dose-response curves were obtained for both taR10-crR10 and taR12-crR12 riboregulator pairs. There was no activation at low taRNA concentrations ($<10^{-4}$% arabinose), followed by a rise in activation at intermediate taRNA concentrations (10^{-4}–10^{-2}%), and finally a high state that plateaus at elevated levels of taRNA ($>10^{-2}$%). These data show tunable activation of protein expression through the controlled introduction of taRNA.

Highly specific riboregulators

We investigated all 16 combinations (L,7,10,12) of the taRNA-crRNA constructs to determine whether the artificial riboregulator pairs are specific. **Figure 4c** shows results from separate cultures containing crR12 and four different taRNAs (taRL,7,10,12). With no arabinose induction, the *cis*-repressed GFP expression was near autofluorescence levels, and the concentration of taRNA was nearly undetectable. Upon arabinose induction, there was a strong increase in RNA concentration of all taRNA variants, but tenfold activation of protein expression was seen only in the taR12-crR12 cognate pair (**Fig. 4c**). Equivalent specificity was seen with the taR10-crR10 cognate pair. No activation was seen with any crRL- or crR7-taRNA pairs. The equilibrium constants of the intramolecular (crRNA-crRNA) and intermolecular (taRNA-crRNA) pairings were determined *in vitro* (**Table 2**). The dissociation constants of cognate pairs (*e.g.*, taR10-crR10) were at least ten times lower than those of noncognate pairs (*e.g.*, taR10-crR7), indicating a good correlation with *in vivo* GFP measurements (**Figs. 2** and **3**, **Table 1**). Taken together, these results indicate that taRNA-crRNA interactions that expose the RBS require highly specific cognate RNA pairings.

Modular riboregulators

Finally, we investigated the modular nature of this system by replacing the P_{BAD} and $P_{LtetO-1}$ promoters with the $P_{LtetO-1}$ and $P_{LlacO-1}$ promoters, respectively. In the new scheme, $P_{LlacO-1}$ drives the expression of crR12 whereas $P_{LtetO-1}$ produces taR12. As with the original riboregulator variant, we observed GFP fluorescence levels from crR12 that were near autofluorescence. In this riboregulator system, taR12 was transcribed from six different positions relative to the transcription start site[29] of $P_{LtetO-1}$: +1, +3, +5, +19, +21 and +23 (spacer sequences are provided in **Supplementary Table 1** online). No detectable activation was observed in the +1, +19, +21 and +23 variants, whereas the +3 and +5 variants exhibited 9-fold and 13-fold increases in GFP expression, respectively. These data reveal an important mechanistic feature of this system: the taRNA, which targets the consensus loop of the crRNA, sensitively depends on an accessible 5′ linear complementary sequence. More specifically, the variants that do not exhibit any activation suggest that an elongated (+19, +21, +23) or truncated (+1) 5′ taRNA end interferes with taRNA-crRNA interaction, preventing stable intermolecular duplex formation. In this regard, we investigated the predicted secondary structures of all the spacer regions using Mfold software[28]. We found that these regions are not highly structured and could interact with another region of the taR12 molecule or the crRNA component to form alternative conformations that might interfere with taRNA-crRNA interactions. These mechanistic features could be important in natural RNA-based systems[10] and warrant further exploration.

DISCUSSION

This study, which details positive and negative post-transcriptional control, elucidates the action of *cis*- and *trans*-acting regulatory RNAs. We found that conformational changes in RNA structures and stable duplex formation not only depend on the initial recognition complex but also on the ability of *trans* activators to bind to nucleotides in the partially destabilized stem structure. In our system, the specificity of intermolecular RNA interaction arises from unique sequences in the crRNA stem and not from the consensus sequence of the recognition loop. Studies of artificial riboregulators of this sort could be useful for characterizing potential modes of action of sRNAs, which have been implicated as regulators of transcription and translation and as modulators of developmental switches[3,10]. In addition, this work may complement ongoing sequence- and structure-based efforts[33] to identify and characterize novel sRNAs, particularly *trans* activators, in both prokaryotes and eukaryotes. Ultimately, the versatility of artificial riboregulators may yield insights into RNA-based cellular processes and RNA's evolutionary role in biology[1,2].

From a biotechnology standpoint, artificial riboregulators can be used with both synthetic (*e.g.*, $P_{LtetO-1}$ (ref. 29)) and endogenous promoters. With regard to endogenous promoters, our system is capable of producing physiologically relevant levels of a target gene from its native promoter, a feature lacking with current inducible prokaryotic gene expression systems that require specific promoters. Furthermore, with our system, tighter and tunable control of

gene expression is achieved at the post-transcriptional level. The *cis*-repressed element (crRNA), coupled to inducible promoters, provides near-absolute repression that results in reduced leakage of gene expression, which may be of value in studies such as those involving toxic genes. In addition, the small taRNAs are capable of acting on single and multiple crRNA targets, conserve more energy than protein regulators as they do not require translational resources[10,18,34], and compared to DNA-protein networks, exhibit quicker response times to environmental and biological stimuli[10,18,34], which could be advantageous for a variety of biosensing and biocomputing applications.

Our artificial riboregulators also could be used as RNA-based modules in genetic circuits to control and investigate gene regulation at the post-transcriptional level. Because artificial riboregulators can be used with any prokaryotic promoter or gene, they are compatible with a wide array of transcriptional and regulatory components, expanding the potential of synthetic gene networks, which have been limited to well-characterized transcription factors[35]. For example, RNA-based modules could be used to selectively perturb networks, at physiologically relevant levels, to reveal functional properties of large-scale genetic networks[36]. Although the current scheme characterizes a small number of taRNA-crRNA pairs, the number of elements and their range of function can be expanded considerably by *in vitro* selection techniques[37–43], given their scalability and specificity of interaction. Because 26 nucleotides participate in the taRNA-crRNA duplex, greater than 10^{15} unique sequence pairs can potentially be constructed, creating a near-infinite library of interactive riboregulators. Such an assembly, combined with improved oligonucleotide synthesis and methods of producing whole genome assemblies[44], could generate networks of highly specific riboregulators *in vivo*, increasing the complexity of biological networks based solely on DNA-protein components[45,46]. The design of endogenous or synthetic networks based on artificial riboregulators could thus lead to engineered cellular control and a better understanding of complex cellular processes.

METHODS

Plasmid construction, cell strains, reagents. Basic molecular biology techniques were implemented as previously described[47]. Two riboregulator systems were constructed, in which each system used separate promoters to drive the expression of the *cis*-repressed RNAs (crRNA) and *trans*-activating RNAs (taRNA), respectively. In the first system, $P_{LtetO-1}$ produces crRNA and P_{BAD} produces taRNA. In the second system, $P_{LlacO-1}$ drives the expression of crRNA and $P_{LtetO-1}$ produces taRNA. We chose the constitutive $P_{LtetO-1}$ promoter[29], a modified version of the native Phage λ P_L promoter containing two TetR operator sites, so that the rate of transcription could be modulated by the TetR protein and the chemical inducer of the *tetR* operon, anhydrotetracycline (aTc). Similarly, the rate of transcription from the $P_{LlacO-1}$ promoter is regulated by the LacI protein and the chemical inducer of the *lac* operon, isopropyl-β-D-thiogalactopyranoside (IPTG). The taRNA molecules produced from the P_{BAD} promoter of the arabinose operon could be modulated by the presence of arabinose in the *araC*$^+$, *tetR*$^+$ *E. coli* strain. All plasmids (see **Supplementary Table 2** and **Supplementary Fig. 1** online) contained the ColE1 origin of replication and genes coding for either ampicillin or kanamycin resistance (see **Supplementary Notes** online). Oligonucleotide primers were purchased from Amitof Biotech and Integrated DNA Technologies. All genes and promoters were PCR amplified using the PTC-200 PCR machine (MJ Research) with *PfuTurbo* DNA Polymerase (Stratagene). DNA sequences were obtained as follows: *gfpmut3b* gene from pJBA113 (ref. 48), $P_{LlacO-1}$ promoter from pZE12-luc[29], $P_{LtetO-1}$ promoter and ribosome binding site (RBS) sequence from pZE21[29] and the arabinose operon (P_{BAD}) from pBAD-HisA (Invitrogen). *Cis* and *trans* sequences were constructed by custom oligonucleotide design (see **Supplementary Notes** online).

All plasmids were constructed using restriction endonucleases and T4 DNA Ligase from New England Biolabs. Plasmids were introduced into the *E. coli* XL-10 strain (Stratagene; Tet^r Δ (mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, recA1, gyrA96, relA1, lac Hte [F proAB lacI^q ZDM15 Tn10 (Tet^r) Amy Cam^r]) using standard heat-shock, transformation and storage solution, transformation protocols[47]. The *E. coli* XL-10 strain, DH5αpro strain (Clontech; deoR, endA1, gyrA96, hsdR17(rk$^-$m-k$^+$), recA1, relA1, supE44, thi-1, Δ(lacZYA-argF)U169, ϕ80δlacZ Δ M15, F$^-$, λ–, PN 25/tetR, P_{lacI}q/laci, Sp^r), 2.300 strain (Genetic Stock Center no. 5002, λ–, lacI22, rpsL135 and thi-1) and wild-type K-12 strain (Genetic Stock Center no. 4401) were used for all experiments. All cells were grown in selective medium: Luria-Bertani (DIFCO) and either 30 μg/ml kanamycin or 100 μg/ml ampicillin (Sigma). Plasmid isolation was done using Perfectprep Plasmid Isolation Kits (Eppendorf). Subcloning was confirmed by restriction analysis. Plasmid modifications were verified by sequencing, using the PE Biosystem ABI Prism 377 sequencer.

Gene expression experiments. For all experiments, cells were grown overnight in the appropriate conditions, diluted 1:1,000, and regrown before collecting RNA samples and measuring GFP expression by flow cytometry. All RNA and GFP measurements were obtained during logarithmic growth at OD_{600} 0.4–0.6, measured by a SPECTRAFluor Plus (Tecan). A positive control, pZE21G, was constructed such that the promoter drives the expression of *gfpmut3b* without the repressive *cis* element. *Cis* experiments were conducted under two conditions: no anhydrotetracycline (aTc) and 30 ng/ml aTc. The expression of TetR in the XL-10 strain was insufficient to obtain full repression of GFP contained on high-copy number ColE1 plasmids. Therefore, in control experiments (**Fig. 2b**), we observed intermediate levels of GFP expression, which correspond to intermediate transcription rates. DH5αpro cells contain higher cellular levels of TetR, and thus demonstrated a lower expression state at no aTc induction. *Cis/trans* experiments were conducted under four conditions: (i) no aTc, no arabinose, (ii) no aTc, 0.25% arabinose, (iii) 30 ng/ml aTc, no arabinose and (iv) 30 ng/ml aTc, 0.25% arabinose. In these experiments, aTc modulates the transcription of crRNA and arabinose modulates the expression of taRNA. We measured the expression of the riboregulator systems in two additional *tetR*$^-$ strains: 2.300 strain and wild-type K-12 strain. In these strains, we grew cultures containing riboregulator systems in the absence and presence of arabinose, and obtained results consistent with those of the other strains (XL-10 and DH5αpro).

GFP quantification using the flow cytometer. All expression data were collected using a Becton Dickinson FACSCalibur flow cytometer with a 488-nm argon laser and a 515- to 545-nm emission filter (FL1) at low flow rate. Before analysis, cells were pelleted and resuspended in filtered PBS, pH 7.2 (Life Technologies) immediately after each time point. Calibrite Beads (Becton Dickinson) were used to calibrate the flow cytometer. Each fluorescent measurement of gene expression was obtained from >100,000 cells. Flow data were converted to ASCII format using MFI software (E. Martz, University of Massachusetts, Amherst). Matlab (Mathworks) software was used to filter (in a narrow forward scatter (FSC) range) and analyze a homogenous (based on FSC, an indicator of cell size) population of cells in each sample.

Quantification of cellular RNA concentrations: rcPCR gene expression analysis. Quantitative measurements of mRNA levels used PCR coupled with matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometry[49] (see **Supplementary Notes** online). The reported concentrations of crRNA in **Table 1** and taRNA in **Figure 4c** are expressed as a percentage of 16S rRNA concentration within each sample. The assay designs for 16S rRNA, taRNA and crRNA and the steps of rcPCR are described in the **Supplementary Notes** and **Supplementary Table 3** online.

***In vitro* experiments.** All taRNA-crRNA pairs corresponding to the 7, 10 and 12 variants were investigated to assess the *in vitro* specificity of interactions. The equilibrium constants for complexes between the *cis*-repressed and *trans*-activating RNAs were measured through *in vitro* synthesis of RNA constructs from the T7 promoter. We use an approach based on the property of

some reverse transcriptases to stall and terminate on stable RNA duplexes (see **Supplementary Notes** online). When hybridized to crRNA, taRNA creates an obstacle for the reverse transcriptase, yielding a truncated product (taRNA-crRNA complex). The amount of truncated transcripts and full-length (crRNA only) transcripts are assayed by PAGE, permitting equilibrium dissociation constants to be calculated (see **Supplementary Figs. 2** and **3** online). The steps used to obtain the equilibrium constants in **Table 2** are described in the **Supplementary Notes** online.

Parallel *in vitro* experiments were done to measure the equilibrium constants of the stem-loop (*cis*-repressive element and the RBS) in the crRNA constructs. RNA thermal melting curves (see **Supplementary Fig. 4** online) were obtained on a UV-visible spectrophotometer (Cary-Varian) equipped with a water-circulated temperature-controlled cell holder. Samples were first heated at 90 °C for 2 min, chilled on ice and dissolved in buffer (50 mM Tris/HCl, pH 7.5; 100 mM NaCl) to a final concentration of 0.15 μM of crRNA (44-mer stem-loop RNA oligonucleotides purchased from Integrated DNA Technologies) before absorbance measurements. RNA thermal denaturation was monitored by measuring the absorbance of UV light at 260 nm in a quartz cuvette (Starna Cells, Catalog no. 28B9-Q-10) with a standard 1-cm path length. Absorbance was measured over a continuous temperature range from 10–95 °C at a rate of approximately 1 °C /min, sampling at each 1 °C . The equilibrium constants ($\Delta G_{crR\text{-}crR}$) in **Table 2** were determined from RNA melting curves and calculated as previously described[50].

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank T. Yoshida for providing access to the UV spectrophotometer; E. Protozanova for discussions and advice with RNA melting experiments; I. Smolina for help and advice with reverse transcription experiments; W. Blake, J. Hasty, D.H. Lee, J. Graber and members of our lab for helpful discussions and advice in preparing the manuscript. This work was supported by the National Science Foundation and Defense Advanced Research Projects Agency.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 9 January; accepted 6 May 2004
Published online at http://www.nature.com/naturebiotechnology/

1. Gesteland, R.F., Cech, T.R. & Atkins, J.F. *The RNA World,* edn. 2 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
2. Joyce, G.F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
3. Eddy, S.R. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**, 919–929 (2001).
4. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**, 147–157 (1982).
5. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The TNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
6. Doudna, J.A. & Cech, T.R. The chemical repertoire of natural ribozymes. *Nature* **418**, 222–228 (2002).
7. Steward, P., Molin, S. & Nordstrom, K. RNAs involved in copy-number control and incompatibility of plasmid R1. *Proc. Natl. Acad. Sci. USA* **78**, 6008–6012 (1981).
8. Wagner, E.G.H. & Simons, R.W. Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.* **48**, 713–742 (1994).
9. Lee, R.C., Feinbaum, R.L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
10. Gottesman, S. Stealth regulation: biological circuits with small RNA switches. *Genes Dev.* **16**, 2829–2842 (2002).
11. Gelfand, M.S., Mironov, A.A., Jomantas, J., Kozlov, Y.I. & Perumov, D.A. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet.* **15**, 439–442 (1999).
12. Winkler, W., Nahvi, A. & Breaker, R.R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–656 (2002).
13. Johansson, J. *et al.* An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* **110**, 551–561 (2002).
14. Mironov, A. *et al.* Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**, 747–756 (2002).
15. Morita, M.T. *et al.* Translational induction of heat shock transcription factor *sigma32*: evidence for a built-in RNA thermosensor. *Genes Dev.* **13**, 655–665 (2002).
16. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C. & Breaker, R.R. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**, 577–586 (2003).
17. Winkler, W., Nahvi, A., Roth, A., Collins, J.A. & Breaker, R.R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281–286 (2004).
18. Lease, R.A. & Belfort, M. A *trans*-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. *Proc. Natl. Acad. Sci. USA* **97**, 9919– 9924 (2000).
19. Majdalani, N., Hernandez, D. & Gottesman, S. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol. Microbiol.* **46**, 813–826 (2002).
20. Opalinska, J.B. & Gewirtz, A.M. Nucleic–acid therapeutics: basic principles and recent applications. *Nat. Rev. Drug Discov.* **1**, 503–514 (2002).
21. Good, L. Translation repression by antisense sequences. *Cell. Mol. LifeSci.* **60**, 854–861 (2003).
22. Ji, Y. *et al.* Identification of critical Staphylococcal gene using conditional phenotypes generated by antisense RNA. *Science* **293**, 2266–2269 (2001).
23. Dykxhoorn, D., Novina, C.D. & Sharp, P.A. Killing the messenger: Short RNAs that silence gene expression. *Nat. Rev. Mol. Cell Biol.* **4**, 457–467(2003).
24. Wagner, E.G.H. & Flardh, K. Antisense RNAs everywhere? *Trends Genet.* **18**, 223–226 (2002).
25. Engdahl, H.M., Lindell, M. & Wagner, E.G.H. Introduction of an RNA stability element at the 5′–end of an antisense RNA cassette increases the inhibition of target RNA translation. *Antisense Nucleic A.* **11**, 29–40 (2001).
26. Morfeldt, E., Taylor, D., von Gabain, A. & Arvidson, S. Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J.* **14**, 4569–4577 (1995).
27. Franch, T., Petersen, M., Wagner, E.G.H., Jacobsen, J.P. & Gerdes, K. Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J. Mol. Biol.* **294**, 1115–1125 (1999).
28. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
29. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraCI1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
30. Cormack, B.P., Valdivia, R.C. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
31. Hjalt, T.A.H. & Wagner, E.G.H. Bulged-out nucleotides protect an antisense RNA from RNase III cleavage. *Nucleic Acids Res.* **23**, 571–579 (1995).
32. Court, D. RNA processing and degradation by RNase III. *Control of Messenger RNA Stability*, (eds. Belasco, J. & Brawerman, G.) 71–116. (Academic Press, New York, 1993).
33. Rivas, E., Klein, R.J., Jones, T.A. & Eddy, S.R. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr. Biol.* **11**, 1369–1373 (2001).
34. Altuvia, S. & Wagner, E.G.H. Switching on and off with RNA. *Proc. Natl. Acad. Sci. USA* **97**, 9824–9826 (2000).
35. Hasty, J., McMillen, D. & Collins, J.J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
36. Gardner, T.S., di Bernardo, D., Lorenz, D. & Collins, J.J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
37. Bartel, D.P. & Szostak, J.W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).
38. Wilson, D.S. & Szostak, J.W. *In vitro* selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611–647 (1999).
39. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
40. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
41. Joyce, G.F. Amplification, mutation and selection of catalytic RNA. *Gene* **82**, 83–87 (1989).
42. Robertson, M.P. & Ellington, A.D. *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat. Biotechnol.* **17**, 62–66 (1999).
43. Faulhammer, D., Cukras, A.R., Lipton, R.J. & Landweber, L.F. Molecular computation: RNA solutions to chess problems. *Proc. Natl. Acad. Sci. USA* **97**, 1385–1389 (2000).
44. Smith, H.O., Hutchison, C.A. III, Pfannkoch, C. & Venter, J.C. Generating a synthetic genome by whole genome assembly: ϕx174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445 (2003).
45. Tabor, J.J. & Ellington, A.D. Playing to win at DNA computation. *Nat. Biotechnol.* **21**, 1013–1015 (2003).
46. Stojanovic, M.N. & Stefanovic, D. A deoxyribozyme–based molecular automaton. *Nat. Biotechnol.* **21**, 1069–1074 (2003).
47. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual,* edn. 2 (Cold Spring Harbor Laboratory Press, Plainview, New York, 1989).
48. Andersen, J.B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246 (1998).
49. Ding, C. & Cantor, C.R. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *Proc. Natl. Acad. Sci. USA* **100**, 3059–3064 (2003).
50. Marky, L.A. & Breslauer, K.J. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **26**, 1601–1620 (1987).