

# De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells

Guillermo Rodrigo<sup>a,b,1</sup>, Thomas E. Landrain<sup>a,1</sup>, and Alfonso Jaramillo<sup>a,2</sup>

<sup>a</sup>Institute of Systems and Synthetic Biology, Genopole – Université d'Évry Val d'Essonne – Centre National de la Recherche Scientifique (CNRS UPS3509), 91030 Évry, France; and <sup>b</sup>Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas – Universidad Politécnica de Valencia, 46022 Valencia, Spain

Edited by David Baker, University of Washington, Seattle, WA, and approved August 6, 2012 (received for review March 20, 2012)

**A grand challenge in synthetic biology is to use our current knowledge of RNA science to perform the automatic engineering of completely synthetic sequences encoding functional RNAs in living cells. We report here a fully automated design methodology and experimental validation of synthetic RNA interaction circuits working in a cellular environment. The computational algorithm, based on a physicochemical model, produces novel RNA sequences by exploring the space of possible sequences compatible with predefined structures. We tested our methodology in *Escherichia coli* by designing several positive riboregulators with diverse structures and interaction models, suggesting that only the energy of formation and the activation energy (free energy barrier to overcome for initiating the hybridization reaction) are sufficient criteria to engineer RNA interaction and regulation in bacteria. The designed sequences exhibit nonsignificant similarity to any known noncoding RNA sequence. Our riboregulatory devices work independently and in combination with transcription regulation to create complex logic circuits. Our results demonstrate that a computational methodology based on first-principles can be used to engineer interacting RNAs with allosteric behavior in living cells.**

post-transcriptional regulation | evolutionary computation | computational RNA design | RNA synthetic biology

The understanding of RNA interactions in living cells and their subsequent exploitation as regulators is providing new synthetic biology applications (1). RNA regulation is being studied from natural systems by the analysis of the interactions of small RNAs (sRNAs) with messenger RNAs (mRNAs) (2), proteins (3) or molecules (4). However, it is also possible to follow a forward engineering approach and attempt the de novo design of RNA regulators. Rational design techniques have been applied, in both prokaryotes and eukaryotes, for repression or activation of translation (5–10), mRNA degradation (11), riboswitches and ribozymes (12–14), transcription attenuation (15–18), and scaffolding (19). On the other hand, computational methods allowed designing nucleic-acid-based logic circuits in vitro (20–23), including the redesign of allosteric ribozymes (23), hence, challenging the current knowledge of nucleic acid structure and function. Previous RNA design approaches, however, have been mostly developed to work with in vitro systems or required incorporating fragments of natural sequences.

We now propose a fully automated sequence selection methodology to design general circuits based on RNA interactions to operate in living cells. Previous computational methodologies relied on the dominance of the Watson–Crick interactions (24), but they were not adapted to in vivo operations where RNA could be very unstable as it occurs in bacteria. Our approach consists of stabilizing RNA molecules by enforcing a given structure, as done in the inverse folding problem (25–27), together with targeted interactions and conformational changes. The analysis of natural systems unveils another challenge: The kinetics of RNA interactions is rate-limited by the initial interaction of solvent-accessible nucleotides from each binding partner, as illustrated in the kissing-loop mechanism (2). To make manageable the computa-

tional problem, here we also rely on Watson–Crick interactions (i.e., canonical purine-pyrimidine pairs plus the G–U pair), although it is possible to extend our approach to other types of interactions (28). Therefore, we are faced with the problem of designing a set of RNA species with predefined structures and with unspecified intermolecular interactions able to produce the intended allosteric regulation. Here, we show it is possible to solve such a combinatorial problem by developing a fully automated procedure that exploits physicochemical principles and structural constraints and that outputs the RNA sequences implementing the predefined interactions.

## Results and Discussion

**Computational Design of sRNA Circuits.** Our methodology (see details in *SI Appendix*) starts by choosing well-defined structures for all single species of the circuit. Because we focus on Watson–Crick interactions, we implicitly assume that the secondary structure already determines a stable three-dimensional architecture (29). We also hypothesize that the interaction between two species is nucleated by their unpaired nucleotides (Fig. 1A). In this interaction model, an initial intermolecular pairing driven by a small sequence (toehold or seed site) nucleates a downhill reaction where the size of the intermolecular pairs (represented as a reaction coordinate in Fig. 1A) rapidly increases until the hybridization ends. Initially, the algorithm assigns random nucleotides to the sequences of each RNA species, while respecting their designated secondary structure, which ensures a low  $\Delta G_{\text{form}}^i$  by solving an inverse folding problem (Fig. 1A). It then explores the space of allowed nucleotide sequences by using an objective function and a Monte Carlo simulated annealing (MCSA) search algorithm (30) (Fig. 1B). The convergence of our algorithm is coupled to the existence of large networks of neutral paths (of common structure) able to connect highly different sequences (31), and the algorithm scatters several initial random sequences along these networks to perform an efficient exploration of the sequence space. In addition, to enlarge such neutral paths and then improve the optimization, we allow non-neutral mutations perturbing, up to three base pairs, the structures specified for the single species.

The objective function is defined to minimize by MCSA two competing design goals: (i) free energy of complex formation and (ii) activation energy of complex formation. The first term accounts for the free energy difference between the interacting and free species for all possible interactions in the circuit ( $\Delta G_{\text{form}}$ ). For the second term we consider a magnitude related

Author contributions: G.R., T.E.L., and A.J. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

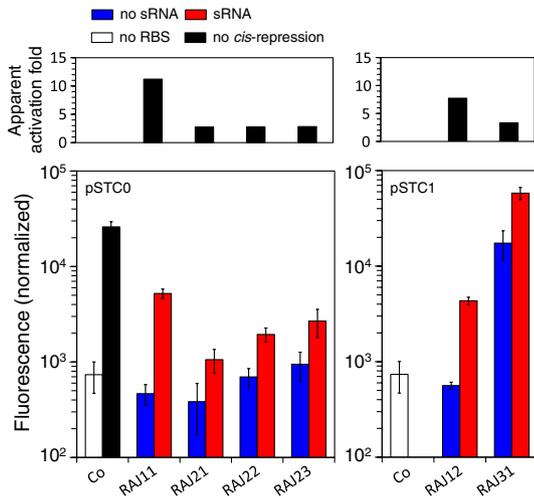
This article is a PNAS Direct Submission.

<sup>1</sup>G.R. and T.E.L. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. E-mail: alfonso.jaramillo@issb.genopole.fr.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1203831109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1203831109/-DCSupplemental).





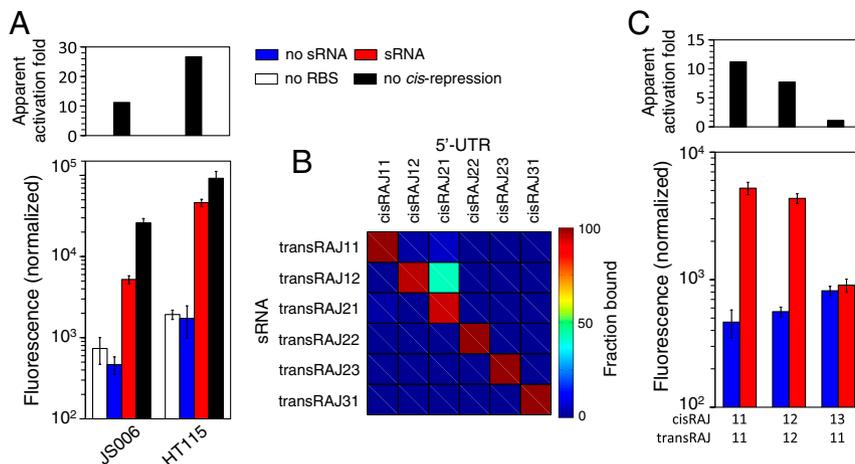
**Fig. 3.** Experimental characterization of the RNA devices. Normalized fluorescence of the devices together with the apparent activation fold (*Top*), measured as the ratio of fluorescence in presence and absence of the riboregulator (see also *SI Appendix, Fig. S15*, for flow cytometry characterizations).

We experimentally validated in *E. coli* six solutions corresponding to different interaction models (Fig. 2) by exploring a space of  $10^{40}$  sequences (*SI Appendix, Table S7*). The *cis*-repressing elements tightly reduced the protein expression to about 1–4% of the maximal expression (*SI Appendix, Fig. S13*). These strong repressions certainly support the formation of the predicted structures, which prevent the exposure of the RBS to the solvent. The designed riboregulators activated translation (monitored by fluorometry) in a range going from 2.8 (RAJ21) to 11.2 (RAJ11) of apparent activation fold (Fig. 3A). These values are similar to those reported for natural systems (35); therefore, our devices could be readily used to reprogram bacterial behavior. A flow cytometry analysis also revealed a statistically significant sRNA activation within the cell population in all designs (*SI Appendix, Fig. S15*), with average values presenting quantitative agreement with fluorometry results (*SI Appendix, Fig. S12*). The significant activation of protein expression in all cases supports the sRNA-mRNA interaction and the intended conformational change in the 5'-UTR of the mRNA<sub>GFP</sub>. Furthermore, manual design can be applied on top of our computational designs to enlarge the number of different devices (16). For

instance, in system RAJ11 (*SI Appendix, Fig. S6*) a single point mutation would reduce the activity in 52% of the cases following our objective function (*SI Appendix, Figs. S8 and S9*). In particular, we identified that two compensatory mutations in the toehold would be sufficient to create an orthogonal device with similar specificity (*SI Appendix, Table S8*).

**Context Analysis of Engineered Riboregulation.** Our design methodology does not consider cellular factors that could drive RNA interactions, such as the specific and nonspecific binding to endogenous RNAs or proteins. On the one hand, we could incorporate into the designed sequences specific recognition sites to known proteins, such as Hfq, provided its role in sRNA stabilization and catalysis of sRNA-mRNA pairing (38, 39). Instead, by expressing both RNAs within the same plasmid we may already promote local coexpression, avoiding intracellular RNA diffusion. Other cellular factors involve ribonucleases (RNases), which could be incorporated if their cleavage sequences are known. On the other hand, nonspecific interactions could jeopardize our predictions. For instance, bacterial RNase III is a potent and fast RNase that targets double stranded regions and we evaluated the effect of such an RNase on this system by replicating our characterizations in the corresponding knockout strain. Although we found an increase about the double in the apparent activation fold (from 11.2 to 26.7) when using a  $\Delta$ RNase III strain (Fig. 4A), the real-time quantitative PCR (RT-qPCR) quantification of the ratio sRNA/mRNA<sub>GFP</sub> did not change in such a strain and it warrants further exploration (see *SI Appendix*).

Because our methodology finds novel nucleotide sequences stabilizing a set of RNA structures and interactions, we would expect each sequence was sufficiently dissimilar to the others making cross-talk unlikely. Moreover, a Basic Local Alignment Search Tool showed that all sequences of our devices display no significant similarity to any known noncoding RNA sequence (*SI Appendix, Table S6*). To investigate the orthogonality of our riboregulatory devices, we checked the hybridization ability between possible combinations of *cis*-repressing and *trans*-activating RNAs. We estimated computationally the relative levels of the interaction complex at the equilibrium, showing that our devices, despite the homologies already present in the sequences and structures due to imposing a common RBS sequence, displayed low interactions between noncognate pairs (Fig. 4B; *SI Appendix, Table S5*; degree of orthogonality of the 98%). We also investigated the effect of higher RNA concentrations on the orthogonality, showing a notable dose-dependent cross-talk between devices RAJ12 and RAJ21 (*SI Appendix, Fig. S11*). We then



**Fig. 4.** Context analysis of the RNA devices. (A) Comparison of the activity of the device RAJ11 in the regular strain (JS006) and in a strain  $\Delta$ RNase III (HT115). (B) Orthogonality analysis of the devices, showing the fraction of complex formation at the equilibrium (values provided in *SI Appendix, Table S5*; see also *SI Appendix, Fig. S11*). (C) Experimental validation of the orthogonality between the devices RAJ11 and RAJ12.



enzymatic expression is *cis*-repressed (10) and the riboregulator is even combined with aptamers that sense a given metabolite (17, 41). As a result, the synthetic circuit can trigger the expression of enzymes for regulating the activity of the pathway. Also in bacteria, we could exploit other mechanisms such as ribozymes (14) or antiterminators (15), and high-throughput analyses could be addressed using a library of automatically designed sRNAs. Our methodology does not require any ribosome involvement, where the RBS sequence could be replaced by the corresponding binding sequence motif of an RNA-binding protein (or molecule) if known (19, 42). In addition, we could improve our methodology (see *SI Appendix*) by incorporating cellular factors (e.g., Hfq, RNase III, RNase E, or 16S rRNA) into the physicochemical model, by considering non-Watson–Crick interactions (28), or by adding predictors of translation initiation by 5'-UTR sequences (43). Furthermore, we could design riboregulators for eukaryotic hosts by replacing the RBS by the Kozak sequence, or alternatively, if known, by the internal ribosome entry site (44). In yeast, the sRNA would interact with the 5'-UTR of certain mRNA to alter the cap-dependent scanning of the 40S subunit (45). In higher eukaryotes, the sRNA could target a given RNA, leading to a conformational change that triggers the intended function (9). Yet, the proposed automatic algorithm to design sRNA circuits in living cells will open new venues for RNA synthetic biology (1) and for quantitative testing of our assumptions about RNA function.

## Materials and Methods

**Plasmid Construction.** The *SI Appendix* (Figs. S1 and S2) outlines the two plasmid templates, pSTC0 and pSTC1, used in this work. The *SI Appendix* (Table S11) shows all plasmids constructed. The target mRNA was placed in 5' sense under the control of the  $P_{LlacO-1}$  promoter (regulated by LacI and IPTG) and the sRNA in 3' sense under the  $P_{LtetO-1}$  promoter (regulated by TetR and aTc) (40). The RNA devices (from the terminator of the sRNA to the 5'-UTR of the mRNA) were made by DNA synthesis (DNA 2.0) and then inserted by ligation into the plasmid templates. From the resulting plasmids, the sRNA was removed to generate systems bearing only the mRNA operon. pSTC0 had the pMB1 origin, kanamycin resistance, and GFPmut3b as a reporter gene. pSTC1 had the pSC101 origin, kanamycin resistance, and superfolder GFP as a reporter gene (see details in *SI Appendix*).

**Strains, Reagents, and Cell Culture.** *E. coli* TOP10 (Invitrogen) was used for routine transformation, as described in the protocol (46). Characterization experiments were performed in *E. coli* K-12 JS006 cells (MG1655  $\Delta araC \Delta lacI Kan^S$ ) (47), in *E. coli* K-12 HT115 cells (W3110 *rnc-14::Tn10*) as RNase III depleted environment (48), and in *E. coli* K-12 MG1655-Z1 cells (MG1655 *lacI<sup>+</sup> tetR<sup>+</sup> araC<sup>+</sup> Sp<sup>R</sup>*) for constitutive control over the  $P_{LlacO-1}$  and  $P_{LtetO-1}$  promoters (49). Cells were grown aerobically in Luria–Bertani broth or in a modified M9 minimum media comprising M9 minimum salts (Sigma M6030) supplemented with glycerol at 0.8% (vol/vol) as the only carbon source,  $CaCl_2$  at 100  $\mu$ M,  $MgSO_4$  at 2 mM, and  $FeSO_4$  at 100  $\mu$ M for higher growth yield (50). Casamino acids were avoided because of their natural green fluorescence that produces too high a fluorescence background for fluorometer

measurements. Cultures were grown over-night at 37 °C and at 225 rpm from single-colony isolates before being diluted. The following concentrations of antibiotics were used when appropriate: kanamycin (50  $\mu$ g/mL), tetracycline (15  $\mu$ g/mL), spectinomycin (100  $\mu$ g/mL). When using *E. coli* K-12 MG1655-Z1 cells, 1 mM of IPTG was used for full activation of the  $P_{LlacO-1}$  promoter when needed, and 100 ng/mL of aTc was used for full activation of the  $P_{LtetO-1}$  promoter when needed.

**Fluorescence Quantification Using Fluorometry.** Before characterization experiments, cells were grown in M9 over two nights in order to reach stationary phase. Cultures were then diluted 200 times in 200  $\mu$ L of M9 within each well of the plate (Custom Corning Costar 96-well microplate, black transparent bottom with lid). The plate was incubated in an Infinite F500 multiwell fluorometer (TECAN) at 37 °C with shaking (orbital mode, frequency of 33 rpm, 2 mm of amplitude) and assayed with an automatically repeating protocol of absorbance measurements (600 nm absorbance filter) and fluorescence measurements (480/20 nm excitation filter–530/25 nm emission filter for GFP and 580/20 nm excitation filter–610/10 nm emission filter for RFP). Time between repeated measurements was 15 min. All samples were present in 3–6 replicated on the plate. Each measurement was repeated two to three times on independent days to verify reproducibility in the results. All data analyses were done using values harvested when cells were in exponential growth phase ( $OD_{600}$  between 0.1 and 0.4). Growth rates were obtained as the slope of a linear regression between the values of  $\log(OD_{600})$  and time. Similar growth rates were observed in all experiments, except for the HT115 strain growing naturally slower. The steady-state protein expression value was obtained as the slope of a linear regression between the values of fluorescence and  $OD_{600}$ . For the dynamical analysis of protein expression, the absolute fluorescence was divided by  $OD_{600}$  to have a magnitude per cell.

**Fluorescence Quantification Using Flow Cytometry.** All expression data were collected using a Becton Dickinson FACSCanto II flow cytometer with a 488 nm argon laser and a 530/30 nm emission filter (GFP) and a 695/40 nm emission filter (RFP). Overnight cultures in M9 were diluted 200 times in 200  $\mu$ L of fresh medium and incubated to reach an  $OD_{600}$  of about 0.1. Cells from M9 cultures were fixed using 4% paraformaldehyde (PFA)—culture cells were pelleted and washed in filtered PBS (Biorad), washed 15 min with PFA, washed in filtered PBS, and finally kept at 4 °C until measurement. Fluorescence measurement of gene expression from each sample was obtained from >20,000 cells. We analyzed the data using FCS express 4 (Denovo software), and we gated the events using narrow forward and side scatter range. We then represented the fluorescence distributions in log scale.

**ACKNOWLEDGMENTS.** We thank J. Hasty for providing the *E. coli* strain JS006, N. Guillén for the strain HT115, and M.B. Elowitz for the strain MG1655-Z1. We thank F. de la Cruz, R. Ruiz, and I. del Campo for practical support with flow cytometry measurements and S.F. Elena and M.P. Zwart for assistance in RT-qPCR assays. We also thank H. Isambert, J.J. Collins, F.J. Isaacs, B.F. Luisi, and E. Westhof for useful comments. Work was supported by the FP7-ICT-043338 (Bacterial Computing with Engineered Populations), the Actions Thématiques Incitatives de Genopole, and the Fondation pour la Recherche Médicale grants (to A.J.). G.R. is supported by an European Molecular Biology Organization long-term fellowship cofunded by Marie Curie actions (ALTF-1177-2011), and T.E.L. by a PhD fellowship from the AXA Research Fund.

- Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. *Nat Biotechnol* 24:545–554.
- Brantl S (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* 10:102–109.
- Rajkowitz L, et al. (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* 4:118–130.
- Gallivan JP (2007) Toward reprogramming bacteria with small molecules and RNA. *Curr Opin Chem Biol* 11:612–619.
- Pfleger BF, et al. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* 24:1027–1032.
- Isaacs FJ, et al. (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* 22:841–847.
- Callura JM, Dwyer DJ, Isaacs FJ, Cantor CR, Collins JJ (2010) Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc Natl Acad Sci USA* 107:15898–15903.
- Nakashima N, Tamura T (2009) Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. *Nucleic Acids Res* 37:e103.
- Venkataraman S, Dirks RM, Ueda CT, Pierce NA (2010) Selective cell death mediated by small conditional RNAs. *Proc Natl Acad Sci USA* 107:16777–16782.
- Callura JM, Cantor CR, Collins JJ (2012) Genetic switchboard for synthetic biology applications. *Proc Natl Acad Sci USA* 109:5850–5855.
- Rinaudo K, et al. (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotechnol* 25:795–801.
- Beisel CL, Bayer TS, Hoff KG, Smolke CD (2008) Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol* 4:224.
- Culler SJ, Hoff KG, Smolke CD (2010) Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* 330:1251–1255.
- Carothers JM, Goler JA, Juminaga D, Keasling JD (2011) Model-driven engineering of RNA devices to quantitatively program gene expression. *Science* 334:1716–1719.
- Dawid A, Cayrol B, Isambert H (2009) RNA synthetic biology inspired from bacteria: Construction of transcription attenuators under antisense regulation. *Phys Biol* 6:025007.
- Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP (2011) Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc Natl Acad Sci USA* 108:8617–8622.
- Qi L, Lucks JB, Liu CC, Mutalik VK, Arkin AP (2012) Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. *Nucleic Acids Res* 40:5775–5786.
- Mutalik VK, Qi L, Guimaraes JC, Lucks JB, Arkin AP (2012) Rationally designed families of orthogonal RNA regulators of translation. *Nat Chem Biol* 8:447–454.
- Delebecque CJ, Lindner AB, Silver PA, Aldaye FA (2011) Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333:470–474.

