

Document downloaded from:

<http://hdl.handle.net/10251/69697>

This paper must be cited as:

Rodrigo Tarrega, G.; Majer, E.; Prakash, S.; Daros Arnau, JA.; Jaramillo, A.; Poyatos, JF. (2015). Exploring the Dynamics and Mutational Landscape of Riboregulation with a Minimal Synthetic Circuit in Living Cells. *Biophysical Journal*. 109(5):1070-1076. doi:10.1016/j.bpj.2015.07.021.



The final publication is available at

<http://dx.doi.org/10.1016/j.bpj.2015.07.021>

Copyright Biophysical Society

Additional Information

# Exploring the dynamics and mutational landscape of riboregulation with a minimal synthetic circuit in living cells

Guillermo Rodrigo,<sup>1,2,\*</sup> Eszter Majer,<sup>3</sup> Satya Prakash,<sup>4</sup>  
José-Antonio Daròs,<sup>1</sup> Alfonso Jaramillo,<sup>2,4</sup> and Juan F. Poyatos<sup>5</sup>

<sup>1</sup>*Instituto de Biología Molecular y Celular de Plantas, CSIC – UPV, 46022 Valencia, Spain*

<sup>2</sup>*Institute of Systems and Synthetic Biology, CNRS – UEVE, 91030 Évry, France*

<sup>3</sup>*Instituto de Biología Molecular y Celular de Plantas, CSIC – UPV, 46022 Valencia, Spain*

<sup>4</sup>*School of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom*

<sup>5</sup>*Logic of Genomic Systems Laboratory, CNB – CSIC, 28049 Madrid, Spain*

Regulation of gene expression triggered by conformational changes in RNA molecules is widely observed in cellular systems. Here, we examine this mode of control by means of a model-based design and construction of a fully synthetic riboregulatory device. We present a theoretical framework that rests on a simple energy model to predict the dynamic response of such a system. Following an equilibrium description, our framework integrates thermodynamic properties –anticipated with an RNA physicochemical model– with a detailed description of the intermolecular interaction. The theoretical calculations are confirmed with an experimental characterization of the action of the riboregulatory device within living cells. This illustrates, more broadly, the predictability of genetic robustness on synthetic systems, and the faculty to engineer gene expression programs from a minimal set of first principles.

## Introduction

It is now widely recognized that phenotypic evolution often occurs through genetic changes in regulatory (non-coding) regions of the genome that affect fundamental parameters of gene expression [1]. Most attention on how these changes alter expression focused on sequences associated to transcriptional regulation [2], but sequences related to other types of control should also matter. In particular, non-coding regions linked to regulatory RNA molecules are being progressively identified as instrumental modulating agents at work in many taxonomically diverse genomes [3, 4]. In the specific case of riboregulation, the capacity to change expression typically relies on the assembly of flexible structures constituted by combinations of interacting RNAs [4], i.e., between a small non-coding RNA (sRNA) and a messenger RNA (mRNA). In this situation, it is important to first understand how the species involved determine different expression features to then inspect how sequence mutations could reshape these parameters.

Notably, some of these matters are beginning to be examined in recent studies on natural riboregulation in bacteria, studies that identified a number of principles. For example, it is now confirmed the presence of an activation threshold from which the system responds [5]; that the action of the sRNA on target genes is fast and linear, which imposes a moderate dynamic range [5–7]; and that the levels of gene expression appear to be correlated with the predicted free energies of the system [8, 9]. However, more work is needed to fully recognize the design principles of RNA-based control, e.g., the impact of species stoichiometry or interaction strength on its function, and how the corresponding sequences encode this information. To what extent these principles are not influenced by more complex processes linked to the intricate regulatory circuitry of the cell is also not entirely known.

To investigate how several basic aspects of riboregulation predictably determine function, we followed in this manuscript a bottom-up approach complementary to the analysis of natural systems. We engineered a simple synthetic riboregulatory device, in which the sequences of its RNA species were designed computationally (using energetic and conformational criteria) [10]. Synthetic approaches are successfully contributing to appreciate many fundamental aspects of gene regulation [11] –by constructing tunable systems that limit any unexpected interplay with the hosting cell– and they are expected to be equally productive in the case of RNA, with many practical implications [12]. We aimed to characterize quantitatively how the conformations, energetics and concentrations determine expression in the synthetic system, and how this information is encoded in the nucleotide sequences. This consists in solving the equilibrium and simulating the intra- and intermolecular structures of the species involved with the use of energy models [13, 14]. The validity of this class of models is expected from its effective prediction of macromolecular structures even at atomic accuracy [15].

In the following, we initially discuss the theoretical framework required to characterize the response of the (synthetic) riboregulatory system. We focused on an essential mechanism that achieves control of protein concentration through a conformational change affecting the interaction of a mRNA with the ribosome [16]. An sRNA interacts with the 5' untranslated region (UTR), which codes for the gene acting as the output of the system (Fig. 1). This allowed us to anticipate how RNA abundances primarily determine the dynamic response, and how response becomes modified by mutations that reshape the core sRNA-mRNA interacting capacities. We then present experimental results testing the framework in *Escherichia coli*, with an RNA switch that follows the theoretical proposal (we termed this switch RAJ11). The system incorporates a green fluorescent protein (GFP) as riboregulated element, and also incorporates the capacity to modulate the expression of the RNA molecules involved with the presence of external inducers in the appropriate bacterial genetic background (Fig. S1) [17].

## Materials and Methods

### *Calculation of RNA free energies and secondary structures*

The synthetic riboregulatory system RAJ11 was studied in this work (Fig. S1), which was obtained by computational design [10], together with manually designed sequence mutants, in order to derive an energy-based model for predicting riboregulatory activity. The natural riboregulatory systems IS10 [9] and RyhB [8] and the synthetic system RR12 [16] (together with the corresponding mutants) were also considered. To compute the free energies and secondary structures of the different RNA species of the system (intra- and intermolecular) the ViennaRNA package was used [18].

*Plasmids, strains and media*

All plasmids characterized in this work were constructed from plasmids pRAJ11 and pRAJ11m, coding for the riboregulatory device RAJ11 [10]. Mutations were introduced in both the sRNA and 5' UTR, and all new mutants were sequenced. As negative control, a plasmid pBlueScript (pBS) was used. As positive control, a plasmid with GFP under the control of promoter  $P_{LtetO1}$  was used. To perform the stationary characterization, *E. coli* DH5 $\alpha$  was transformed. This allowed expressing constitutively the sRNA and mRNA. However, to perform the dynamic characterization, *E. coli* MG1655Z1 [17] was transformed. This allowed controlling the expression of the sRNA with anhydrotetracycline (aTc), and the expression of the mRNA with isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG). Medium LB was used for overnight cultures, and minimal medium M9 (1x M9 salts, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 0.4% glucose) for characterization cultures. IPTG was used at the concentration of 1 mM and aTc varied from 0 to 100 ng/mL. Ampicillin was used at the concentration of 50  $\mu$ g/mL. See Supplementary Material for details on transcript quantification and genome integration.

*Characterization by fluorometry*

Cultures inoculated from single colonies (three biological replicates) were grown overnight in medium LB at 37 °C and 220 rpm. Cultures were then diluted 1:100 in minimal medium M9, and were grown in the same conditions to reach higher OD<sub>600</sub>. Subsequently, a multiwell plate was loaded with 200  $\mu$ L for each sample, which was assayed in a Victor X5 (Perkin Elmer) to measure absorbance (600 nm absorbance filter) and fluorescence (485/14 nm excitation filter, 535/25 nm emission filter, for GFP). Background values of absorbance and fluorescence, which corresponded to medium M9, were subtracted to correct the signals, and the normalized fluorescence was calculated as the ratio of fluorescence and absorbance (Supplementary Material for more details). Growth rates were obtained as the slope of a linear regression between the log values of the corrected absorbance and time.

**Results and Discussion**

*Theoretical prediction of dynamic behavior*

We examined the fold change in expression of the riboregulated protein ( $f$ ), i.e., the ratio in protein concentration at steady state when the sRNA is present ( $p_1$ ) or absent ( $p_0$ ), as an effective score to quantify the performance of the system ( $f = p_1/p_0$ ) [19]. This required to put forward a simple differential equation to describe protein dynamics,  $\frac{dp}{dt} = R - \delta p$ . Here,  $p$  denotes the concentration of the protein,  $R$  the translation rate (determined by the interaction between the the sRNA and mRNA; see below), and  $\delta$  the effective protein degradation rate (a function of the rates of protein degradation and maturation and cell growth; defined in the Supplementary Material) [20].

One can then generally write the translation rate as  $R = \rho_0(x_m - y) + \rho_1 y$ , a combination of translation rates when sRNA is absent ( $\rho_0$ ) or present ( $\rho_1$ ); with  $y$  being the concentration of the sRNA-mRNA complex, and  $x_m$  denoting the total concentration of mRNA (so that  $x_m - y$  denotes concentration of free mRNA species). While  $\rho_1$  is of course an average over the ensemble of conformational states for the complex (i.e.,  $\rho_1 = \sum_i \rho_{1i} \Pi_i$ , where  $\rho_{1i}$  is the translation rate of the  $i$ -th conformational state and  $\Pi_i$  the corresponding Boltzmann probability), we considered here the ensemble dominated by the optimal state [10] as the sequences in RAJ11 were computationally optimized.

To model the concentrations of the corresponding molecular species, we considered a hybridization reaction between sRNA and mRNA with an intermediary state, and followed a law of mass action characterizing the resulting equilibrium state. The concentration of the complex is given by  $y = \frac{x_s + x_m + K}{2} \left( 1 - \sqrt{1 - \frac{4x_s x_m}{(x_s + x_m + K)^2}} \right)$  (see Supplementary Material), where  $K$  is the effective dissociation constant of the hybridization reaction, and  $x_s$  denoted the total concentrations of the sRNA species (recall that  $x_m$  indicates the equivalent for mRNA). To derive this expression, we discarded the formation of homodimers, as the sequences of the RNA species were designed to minimize this type of interactions [10].

Note also that  $K$  represents a threshold value with respect to the amount of sRNA. This threshold-linear behavior (at low sRNA concentrations there is no interaction; beyond the threshold the expression of the riboregulated gene

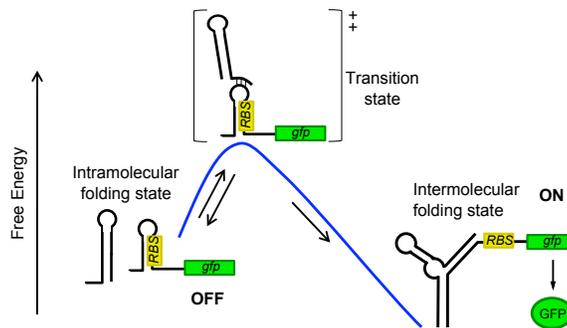


FIG. 1. Scheme of a simple (synthetic) riboregulation system. A *cis*-repressed gene, where the ribosome-binding site (RBS) is trapped, is *trans*-activated by a small RNA (sRNA) that induces a conformational change to release the RBS. The output of the regulation resides in the expression of a *gfp* gene.

increases linearly) was initially observed in studies of natural systems [5, 7]. To overcome this threshold, the concentrations of the RNA molecules at play must be high (in contrast to the dozens or even hundreds of molecules required to regulate transcription with proteins). Thus, certain sRNA molecules have evolved sequence domains able to recruit RNA chaperones to help in the intermolecular interaction (e.g., Hfq) [21]. This is achieved in our model by lowering the value of  $K$ .

For the synthetic riboregulatory system RAJ11, in which the sequences were designed by only accounting for structural and energetic considerations, Hfq and eventually other host helpers are expected to play a minor role, i.e.,  $K$  presents even a higher value. We thus approximated at first order the expression for  $y$ , with  $K \gg x_m + x_s$ , to obtain  $y \simeq x_s x_m / K$ . This imposes a condition of high gene expression to secure hybridization between the RNA species, which could be achieved by using a high-copy number plasmid and strong promoters [10, 22].

Finally, and given that in our experimental system  $\rho_0 \rightarrow 0$  (i.e.,  $p_0 \rightarrow 0$ ; very efficient *cis*-repression), we can write  $p_1 \simeq \frac{\rho_1 y}{\delta} = \frac{\rho_1 x_s x_m}{\delta K}$ . The very low value of  $p_0$  limits on the other hand the confident experimental determination of  $f$ . We thus introduced a modified expression of  $f$  that we termed *apparent* dynamic range ( $f_{\text{app}}$ ). This new measure controls for the strong *cis*-repression of the riboregulated gene (GFP) by including explicitly a parameter that denotes the autofluorescence ( $\theta$ , such that  $p_0 + \theta \simeq \theta$ , i.e., the level of fluorescence of cells expressing the mRNA alone is very similar to the one detected in control cells, see Supplementary Material). This is defined as

$$f_{\text{app}} = \frac{p_1 + \theta}{p_0 + \theta} \simeq 1 + \frac{p_1}{\theta}. \quad (1)$$

We can substitute the previous expression of  $p_1$  to obtain a full theoretical prediction for  $f_{\text{app}}$  that we can also quantify experimentally.

#### Theoretical prediction of effects of mutations

In addition, we analyzed the impact of the thermodynamic properties of the system, which affect the value of  $K$ . In this way, we can anticipate the changes in riboregulatory activity with mutations. We assumed a thermodynamic equilibrium of the system, which has two states according to the reaction coordinate: sRNA and mRNA separate, or sRNA and mRNA hybridized. Then, the free energy of each state determines the probability of finding the system in that state. For this calculation, we predicted the associated free energies [14], and then followed the many particle partition function [23] to get  $\frac{1}{K} = \frac{1}{K_0} W_{sm} \left( \frac{Z_{sm}}{Z_s Z_m} - 1 \right)$ , where  $K_0$  is a constant that ultimately represents an asymptotic effective dissociation constant (see Supplementary Material).  $W_{sm}$  is an entropic factor that accounts for the initiation of the reaction.  $Z_s$  is the partition function of the sRNA,  $Z_m$  of the mRNA, and  $Z_{sm}$  of the complex formed between the sRNA and mRNA.

Even though a living cell integrates multiple dynamic processes of different natures, the high number of molecules in the system of interest (the amount of sRNAs and mRNAs is estimated in hundreds of molecules per gene copy), the difference in time-scales between processes (intramolecular folding faster than sRNA–mRNA interaction, which in turn is faster than translation), and the dominance of certain processes over others (which serve to simplify the great

complexity of biological systems) ultimately make useful thermodynamic descriptions to model and predict experimental results. Indeed, thermodynamic models have been proven quite helpful to study transcription regulation [24], and here we applied the same framework in a riboregulatory context.

In a living cell, the intermolecular interaction starts with highly accessible, unstructured regions in both species [25], forming an intermediary state. We considered that the free energy of activation depends on the length and composition of the seed sequence. This consideration makes that  $K$ , in case of riboregulation, is approached not only by the free energy of hybridization, but also by the free energy of activation –note that previous models considered a sequence-independent activation barrier, but recent data has shown that the dynamic range is well explained by a composition of these two free energies [9].

Therefore, if we calculate the free energy of the seed (or toehold) interaction ( $\Delta G^\#$ ), the term  $W_{sm}$  can be approached by  $e^{-\beta\Delta G^\#}$ . In addition, because we considered the ensembles dominated by the optimal states (and given  $\Delta G$  the free energy of the whole hybridization), we can approach the ratio of partition functions by  $e^{-\beta\Delta G}$ . This gave the following energy model  $\frac{K_0}{K} = e^{-\beta\Delta G^\#} (e^{-\beta\Delta G} - 1)$ .

### Experimental validation of dynamic behavior

We next assessed experimentally the preceding theoretical framework. We initially examined the influence of RNA concentration (of both species and of sRNA). To this end, we expressed  $f_{\text{app}}$  as

$$f_{\text{app}} = 1 + \phi SD^2, \quad (2)$$

with  $S$  denoting the ratio between the sRNA and mRNA concentrations ( $S = x_s/x_m$ ),  $D$  indicating the normalized gene dosage ( $D = x_m/M$ , with  $M$  representing the highest mRNA concentration attainable from plasmid expression), and  $\phi$  an empirical constant given by  $\phi = \frac{\rho_1 M^2}{\theta \delta K}$  (see details in Supplementary Material). At maximal expression of the RNA species for system RAJ11 (i.e., mounted in a plasmid;  $D = 1$ ), we obtained the experimental value of  $f_{\text{app}} = 7.9$ . Under these conditions, we also quantified  $S$  by measuring transcript abundance, obtaining  $S \simeq 1$  (Fig. S2; experimental details in Supplementary Material; note that other riboregulatory systems may present however different relative concentrations, even when the species share gene copy number). With the previous estimates, the value of  $\phi$  was simply calculated (not fitted) as  $f_{\text{app}} - 1 = 6.9$  (see also Fig. S5). Note that  $\phi$  includes all complexity of the dynamic model (i.e., it integrates all dynamic parameters used to construct it), and then only one empirically-calculated parameter was used to make predictions.

In this way, we can predict the behavior of  $f_{\text{app}}$  using Eq. (2) as a function of gene dosage and sRNA expression,  $D$  and  $S$  (solid lines in Figs. 2a,b). We tested the model by mounting alternatively the system RAJ11 into the bacterium chromosome (with expected 100 fold decrease in gene dosage with respect to the plasmid; experimental details in Supplementary Material). As predicted, we observed no significant activation by the sRNA in this case (Figs. 2a and S3). Moreover, by varying the concentration of one of the inducers (anhydrotetracycline), we could modulate sRNA expression (Fig. S1). We employed a new synthetic construction to obtain, with the use of a *gfp* gene, the values of  $S$  for different concentrations of anhydrotetracycline (Fig. S4; experimental details in Supplementary Material). Experimental results were in tune with the prediction (Fig. 2b), highlighting that the system was not in saturation.

In addition to the predictions at steady state, we studied the transient response. The goal here was to examine to what extent the dynamics of the sRNA–mRNA interaction influences the protein dynamics (i.e., if it changes its shape and scale, incorporates delays, etc.) [26]. This is important because it quantifies the rate at which the phenotype changes, either as a result of an environmental perturbation (change in the concentration of one species) or a mutational effect. By assuming a fast scale of RNA interactions, we have a dynamic equilibrium where the system maintains the linear dependence of  $y$  with  $x_s$  upon changes in transcription with time. Then, the differential equation is analytically solvable, resulting  $p(t) = \frac{\rho_1 M}{K} e^{-\delta t} \int_0^t e^{\delta \tau} x_s(\tau) d\tau$ , with the initial condition of  $p(0) = 0$ . The time scale of the sRNA dynamics is much faster than the one of the protein (i.e., the sRNA is quickly degraded). To model an experiment of induction at  $t = 0$  with maximal concentration of anhydrotetracycline, we assumed constant sRNA expression (i.e., quasi-steady state) to solve the integral, obtaining

$$f_{\text{app}}(t) = 1 + \phi SD^2 (1 - e^{-\delta t}). \quad (3)$$

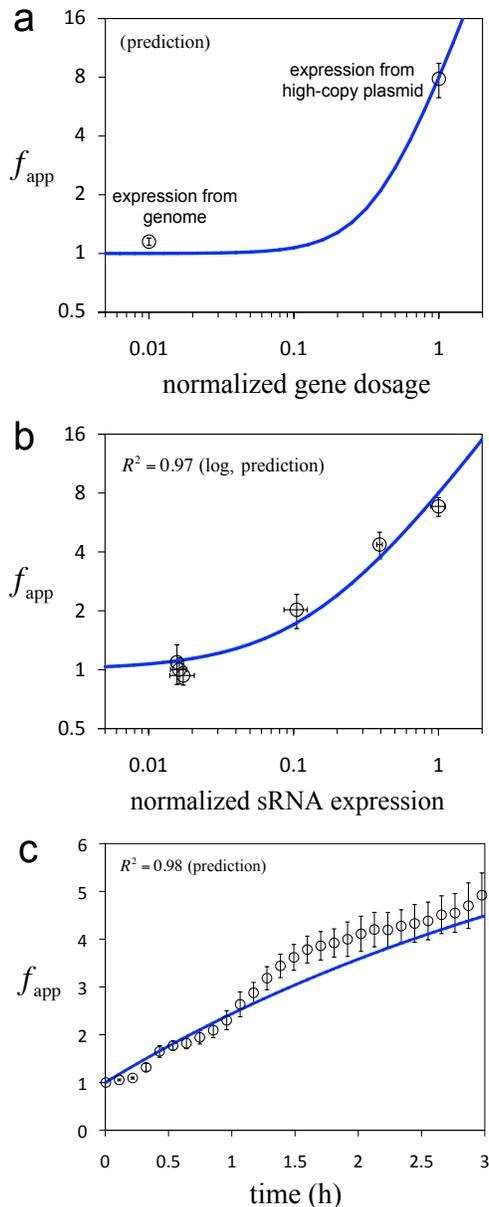


FIG. 2. Dynamics of the riboregulated gene. Dynamic range ( $f_{app}$ ) as a function of normalized gene dosage (a), normalized sRNA expression (b), and time (c). Circles (average of biological replicates) and error bars (standard deviation) correspond to experimental data, whereas the solid lines are predictions from the models in Eqs. (2, 3).  $R^2$  quantifies the agreement between theoretical and experimental calculations (no fitting).

To predict the dynamic response, we needed an estimation of  $\delta$ . For that, we first measured the cell growth rate in our experimental conditions, obtaining  $0.18 \text{ h}^{-1}$ , and then we took advantage of the measurement of the GFP maturation time from earlier work [27], known to be  $\sim 10 \text{ min}$ . In this case, GFP lacks any additional degradation tag, so it is as stable as  $\sim 1 \text{ day}$ . Taken together these values, we calculated  $\delta = 0.23 \text{ h}^{-1}$  (see Supplementary Material). We performed a time-dependent experiment to monitor GFP upon induction with anhydrotetracycline, showing similar characteristic times ( $\sim 1/\delta$ ) in the theoretical and experimental dynamics (Fig. 2c; note that in this case  $S = 1$  and  $D = 1$ ). We also analyzed the effect of different maturation rates and protein stabilities in Fig. S6). This showed that the sRNA–mRNA interaction is fast and then can be treated in quasi-steady state.

*Experimental validation of effects of mutations*

We next examined the effect of mutations in the RNAs involved in the riboregulation. To address this experimentally, we considered maximal expression of both species (hence in the following  $S = 1$  and  $D = 1$ ). Mutations modify the effective dissociation constant of the system ( $K$ ) and consequently  $\phi$  in Eq. (2). Thus, by introducing  $\phi_0 = \frac{\rho_1 M^2}{\theta \delta K_0}$ , we have

$$f_{\text{app}} = 1 + \phi_0 S D^2 e^{-\beta \Delta G^\#} (e^{-\beta \Delta G} - 1). \quad (4)$$

Note then that  $\phi_0$  is a parameter whose value is not affected by the mutations introduced. The value of the parameter  $\beta$  is then critical to make predictions. We here assumed that  $\beta$  is a global constant that can be employed to characterize bacterial riboregulation. To estimate it in an independent way, we considered previous experimental data on natural and synthetic riboregulation [8, 9, 16]. Eq. (4) can be reorganized to explicit the linear dependence of  $\log(f_{\text{app}} - 1)$  with  $\Delta G^\# + \Delta G$  (assuming  $\beta \Delta G \ll 0$ ), given  $-\beta$  the slope. We obtained  $\beta = 0.38$  mol/Kcal with a linear regression between the logarithm of relative  $f$  and  $\Delta G^\# + \Delta G$  (Fig. S7; note that without loss of generality  $f$  was used instead of  $f_{\text{app}}$ ). We also performed a K-fold cross-validation to analyze the estimation of  $\beta$  (Fig. S8). With this value and having calculated  $\Delta G^\# = -10.3$  Kcal/mol and  $\Delta G = -14.5$  Kcal/mol for system RAJ11 (recall  $f_{\text{app}} = 7.9$ ), we calculated (not fitted) the value of the empirical constant  $\phi_0 = 5.7 \cdot 10^{-4}$  (see details in Supplementary Material).

To assess the predictability of our energy model, Eq. (4), we constructed different genotypes from the riboregulatory system RAJ11 and characterized experimentally the resulting apparent dynamic ranges ( $f_{\text{app}}$ ). We mutated the seed region in both sRNA and 5' UTR to get six different genotypes (Fig. S9; experimental details in Supplementary Material). Experimental results nicely followed the predictions (Fig. 3, see Supplementary Material for details on the free energy calculations). We also observed that our model could roughly be used to predict activity with free energy calculations using different empirical parameterizations of the RNA physicochemical model (Fig. S10). Additional predictions showed that point mutations in the seed region contribute in higher extent to reduce activity (Fig. S11). Note in this case that only two empirically-calculated parameters ( $\beta$  and  $\phi_0$ ) are used in the energy model.

We also analyzed the degree of *cis*-repression of this synthetic system. We engineered different mutants of the 5' UTR in order to alter the predicted free energy of it ( $G_m$ ) (see Supplementary Material). Experimentally, we calculated the ratio between fluorescence and autofluorescence for each mutant system, which is a measure of  $1 + p_0/\theta$ . Indeed, autofluorescence is given by  $\theta$ , and mutations in the 5' UTR lead to different levels of *cis*-repression, which changes the value of apparent fluorescence in absence of sRNA ( $p_0 + \theta$ ). According to previous work that examined multiple 5' UTR sequences [28], and having the value of  $\beta$ , we proposed an empirical exponential trend ( $p_0 \propto e^{\beta G_m}$ ) to model changes in  $p_0$  with  $G_m$  (Fig. S12).

## Conclusions

To sum up, our results showed that interacting RNAs need to face up thermodynamic equilibria with large effective dissociation constants in a cellular context (e.g., in comparison with transcription factors). This imposes several constraints on their gene expression attributes. For instance, in a fully synthetic riboregulatory system (e.g., RAJ11), it was essential to have high gene dosage (i.e., reach high concentrations) to obtain a significant dynamic range. Those equilibria also entail a linear effect with the sRNA expression on the response genes. The use of a synthetic system was otherwise instrumental to take out the contribution of the protein machinery [21] to the equilibrium, serving to confirm that mere physicochemical criteria are sufficient to explain the observed behavior. Note nevertheless that the binding sites for those chaperones can be exploited in synthetic systems to achieve higher dynamic range [29]. To circumvent the lack of helping-protein-binding sites, interacting RNAs could evolve towards higher free energy gaps, both in hybridization and activation, to properly work in a living cell. This strategy has been used to design riboregulatory devices with very high dynamic ranges [22].

Of relevance, our energy-based dynamic model accurately predicted the effect of those features on riboregulatory activity in *E. coli*. In particular, the model correctly explained changes in the observed dynamic range due to changes in RNA expression and mutations introduced in the sequences of the two species. Mutations in the seed region were shown to impact much more in activity due to its double contribution in  $\Delta G$  and  $\Delta G^\#$ . This clearly highlights distinct regions on the RNA molecules with different selection pressures. These results pointed out, together with previous observations in natural systems [5, 9], general design principles that can be exploited to engineer more sophisticated riboregulatory systems [30].

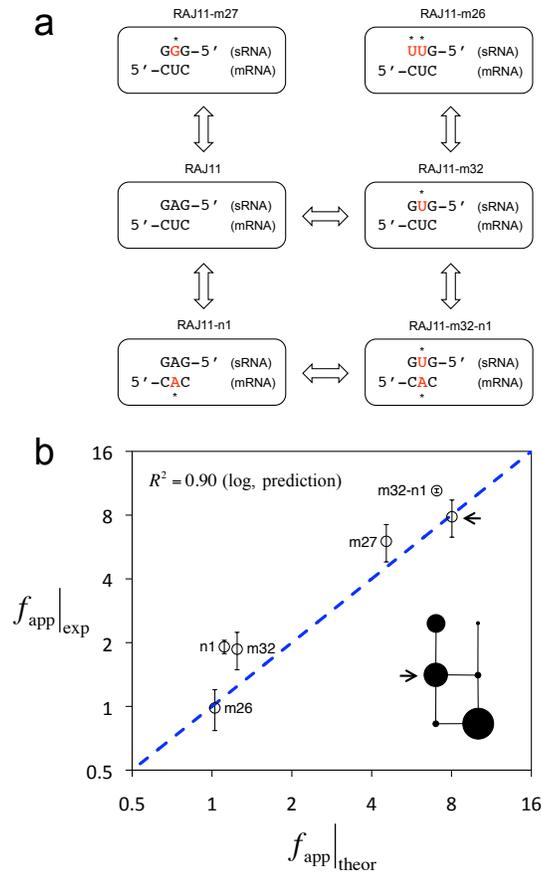


FIG. 3. Sequence-to-function map. (a) Mutated systems from the native RAJ11. We show three nucleotides of the seed region (for both sRNA and 5' UTR of the mRNA) chosen to contain mutations (marked by an asterisk). (b) Experimental versus theoretical dynamic range ( $f_{app}$ ) for different mutant systems with different values of  $\Delta G$  and  $\Delta G^\#$ . Note that Eq. (4) is used here to calculate  $f_{app}$ . Dashed line represents equal experimental and theoretical values. The inset shows the network of mutants (edges connect genotypes differing in a single mutation, and node size corresponds to the value of dynamic range). Arrows point to the native system RAJ11.  $R^2$  quantifies the agreement between theoretical and experimental calculations (no fitting).

Although we focused on positive riboregulation, this model could also be applied to account for repressors of translation, and even combined with other bottom-up approaches to account, for instance, for RNA-small-molecule interactions [31, 32] or transcription control [19, 33] to be relevant to study layered and integrative regulatory mechanisms. Yet, the combination of systems biology principles together with synthetic biology platforms is expected to illustrate the mechanistic bases and limitations of function, robustness and evolvability of different riboregulatory circuits.

### Author Contributions

GR designed the research. GR, EM, SP, JAD, and AJ performed the research and contributed with materials/tools. GR and JFP analyzed the data and wrote the paper.

### Acknowledgements

This work was supported by the AXA Research Fund to GR, the Spanish Ministry of Education, Culture and Sports FPU fellowship AP2012-3751 to EM, the Spanish Ministry of Economy and Competitiveness grants AGL2013-49919-EXP and BIO2011-26741 to JAD and BFU2011-24691 to JFP, and the EU grant FP7-KBBE-613745 (PROMYS) to

A.J.

\* Corresponding author: guillermo.rodrido@csic.es

- [1] G. A. Wray. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.*, 8:206–216, 2007.
- [2] G. A. Wray, M. W. Hahn, and *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.*, 20:1377–1419, 2003.
- [3] F. F. Costa. Non-coding RNAs: New players in eukaryotic biology. *Gene*, 357:83–94, 2005.
- [4] L. S. Waters and G. Storz. Regulatory RNAs in bacteria. *Cell*, 136:615–628, 2009.
- [5] E. Levine, Z. Zhang, and *et al.* Quantitative characteristics of gene regulation by small RNA. *PLoS Biol.*, 5:e229, 2007.
- [6] D. Jost, A. Nowojewski, and *et al.* Regulating the many to benefit the few: role of weak small RNA targets. *Biophys. J.*, 104:1773–1782, 2013.
- [7] R. Hussein and H. N. Lim. Direct comparison of small RNA and transcription factor signaling. *Nucleic Acids Res.*, 40:7269–7279, 2012.
- [8] Y. Hao, Z. J. Zhang, and *et al.* Quantifying the sequence-function relation in gene silencing by bacterial small RNAs. *Proc. Natl. Acad. Sci. USA*, 108:12473–12478, 2011.
- [9] V. K. Mutalik, L. Qi, and *et al.* Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.*, 8:447–454, 2012.
- [10] G. Rodrigo, T. E. Landrain, and *et al.* De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proc. Natl. Acad. Sci. USA*, 109:15271–15276, 2012.
- [11] S. Mukherji and A. van Oudenaarden. Synthetic biology: understanding biological design from synthetic circuits. *Nat. Rev. Genet.*, 10:859–871, 2009.
- [12] L. S. Qi and A. P. Arkin. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat. Rev. Microbiol.*, 12:341–354, 2014.
- [13] W. Fontana, P. F. Stadler, and *et al.* RNA folding and combinatorial landscapes. *Phys. Rev. E*, 47:2083–2099, 1993.
- [14] D. H. Mathews, M. D. Disney, and *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of rna secondary structure. *Proc. Natl. Acad. Sci. USA*, 101:7287–7292, 2004.
- [15] R. Das and D. Baker. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.*, 77:363–382, 2008.
- [16] F. J. Isaacs, D. J. Dwyer, and *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.*, 22:841–847, 2004.
- [17] R. Lutz and H. Bujard. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/II-12 regulatory elements. *Nucleic Acids Res.*, 25:1203–1210, 1997.
- [18] I.L. Hofacker, W. Fontana, and *et al.* Fast folding and comparison of rna secondary structures. *Monatsch. Chem.*, 125:167–188, 1994.
- [19] H. G. Garcia and R. Phillips. Quantitative dissection of the simple repression input-output function. *Proc. Natl. Acad. Sci. USA*, 108:12173–12178, 2011.
- [20] J. H. J. Leveau and S. E. Lindow. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J. Bacteriol.*, 183:6752–6762, 2001.
- [21] J. Vogel and B. F. Luisi. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.*, 9:578–589, 2011.
- [22] A. A. Green, P. A. Silver, and *et al.* Toehold switches: de-novo-designed regulators of gene expression. *Cell*, 159:925–939, 2014.
- [23] S. H. Bernhart, H. Tafer, and *et al.* Partition function and base pairing probabilities of rna heterodimers. *Algorithms Mol. Biol.*, 1:3, 2006.
- [24] R. Phillips. Napoleon is in equilibrium. *Annu. Rev. Condens. Matter Phys.*, 6:85–111, 2015.
- [25] A. S. Richter and R. Backofen. Accessibility and conservation: General features of bacterial small RNA-mRNA interactions? *RNA Biol.*, 9:954–965, 2012.
- [26] Y. Shimoni, G. Friedlander, and *et al.* Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol. Syst. Biol.*, 3:138, 2007.
- [27] R. Iizuka, M. Yamagishi-Shirasaki, and *et al.* Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal. Biochem.*, 414:173–178, 2011.
- [28] H. M. Salis, E. A. Mirsky, and *et al.* Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, 27:946–950, 2009.
- [29] D. Na, S. M. Yoo, and *et al.* Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat. Biotechnol.*, 31:170–174, 2013.
- [30] G. Rodrigo, T. E. Landrain, and *et al.* A new frontier in synthetic biology: automated design of small RNA devices in bacteria. *Trends Genet.*, 29:529–536, 2013.
- [31] X. Chen and A. D. Ellington. Design principles for ligand-sensing, conformation-switching ribozymes. *PLoS Comput. Biol.*, 5:e1000620, 2009.
- [32] J. M. Carothers, J. A. Goler, and *et al.* Model-driven engineering of RNA devices to quantitatively-program gene expression. *Science*, 334:1716–1719, 2011.
- [33] L. Bintu, N. E. Buchler, and *et al.* Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.*, 15:116–124, 2005.