

Nonequilibrium thermodynamics of the RNA-RNA interaction underlying a genetic transposition program

Lucas Goiriz and Guillermo Rodrigo ^{*}

Institute for Integrative Systems Biology (I2SysBio), CSIC – University of Valencia, Paterna 46980, Spain



(Received 1 August 2020; accepted 19 March 2021; published 14 April 2021)

Thermodynamic descriptions are powerful tools to formally study complex gene expression programs evolved in living cells on the basis of macromolecular interactions. While transcriptional regulations are often modeled in the equilibrium, other interactions that occur in the cell follow a more complex pattern. Here, we adopt a nonequilibrium thermodynamic scheme to explain the RNA-RNA interaction underlying IS10 transposition. We determine the energy landscape associated with such an interaction at the base-pair resolution, and we present an original scaling law for expression prediction that depends on different free energies characterizing that landscape. Then, we show that massive experimental data of the IS10 RNA-controlled expression are better explained by this thermodynamic description in nonequilibrium. Overall, these results contribute to better comprehend the kinetics of post-transcriptional regulations and, more broadly, the functional consequences of processes out of the equilibrium in biology.

DOI: [10.1103/PhysRevE.103.042410](https://doi.org/10.1103/PhysRevE.103.042410)

I. INTRODUCTION

Gene regulation is essential for the cell to adjust its physiology against environmental changes and then persist. What lies behind this picture is a series of macromolecular interactions that allow a precise orchestration of multiple gene expressions. In particular, cells rely on protein-DNA interactions to regulate transcription, as seminally revealed for the lactose utilization network in the 1960s [1]. Since then, many other gene expression programs based on transcription regulation have been deciphered, with functional implications [2], and even synthetic circuits have been engineered [3]. However, other types of regulations exist in the cell after RNA is made, which are supposed to complement that transcriptional control in order to fine-tune expression or allow faster responses [4]. Indeed, regulations based on RNA-RNA interactions are pervasive in biology. Among other examples, they can be found within the mechanisms that bacteria exploit in response to stress, especially to regulate globally acting elements [5], that mammalian and plant cells employ to decoy functional RNAs [6], or that retroviruses follow to dimerize their genomes, such as in the case of the human immunodeficiency virus [7]. One such type of regulation consists in a small RNA (sRNA), naked or together with a protein, able to target a given messenger RNA (mRNA) to regulate its translation or stability [5]. Even so, the quantitative description of all these regulations that occur *in vivo* from simple, accurate, and formal mechanistic models is still a challenge.

Thermodynamic models (or, alternatively, models based on statistical mechanics) have been already exploited to quantitatively understand transcription regulations (protein-DNA interactions) [8–10]. There, the system is assumed in

equilibrium and its dynamic regulatory range (or fold change in expression, denoted by r) scales with the Boltzmann factor of the net free energy release associated with the intermolecular interaction (denoted by ΔG_{inter}) in relatively good agreement. Mathematically (considering the regulator as a repressor of gene expression), it can be written

$$r = \frac{\text{regulated protein expression}}{\text{non-regulated protein expression}} = \frac{1}{1 + e^{-\beta(\Delta G_{\text{inter}} - \mu)}}, \quad (1)$$

where β is the reciprocal of the temperature times the gas constant, and μ represents a kind of chemical potential of the regulator (note that $e^{\beta\mu}$ scales with the concentration of the regulator) [10]. In absence of saturation, it can be written $r \propto e^{\beta\Delta G_{\text{inter}}}$. Importantly, r is a mesoscopic parameter, in the range [0, 1], that is accessible both theoretically and experimentally. Moreover, $1-r$ can be seen as the probability of having a regulator tightly bound to each target gene ($r = 1 - P_{\text{bound}}$). In this work, the scaling law shown in Eq. (1) was considered as a null model, which entails that ΔG_{inter} is the unique predictor energy of r .

However, this assumption of thermodynamic equilibrium appears to be inappropriate (or at least incomplete) to model post-transcriptional regulations based on RNA-RNA interactions, as previous results have pointed out that only ΔG_{inter} is not sufficient to predict r [11–13]. Instead, and mostly from heuristic approaches, other Boltzmann factors relative to additional free energies characterizing the interaction, such as the free energy of seed pairing (denoted by ΔG_{seed}), need to be considered in a combinatorial way to achieve relatively good estimates [11,12]. This certainly challenges the conventional wisdom and forces us to find an alternative thermodynamic picture from which to derive a formal relationship. Further-

^{*}Corresponding author: guillermo.rodrido@csic.es

more, there are different kinds of riboswitches (i.e., structured RNAs able to control gene expression in *cis*) that have been shown to function either thermodynamically (at equilibrium) or kinetically (out of equilibrium) [14,15]. Interestingly, two molecular features mainly mark the difference between transcriptional and post-transcriptional regulations *in vivo*. First, RNAs are typically short-lived macromolecules, while DNA and proteins are not [16]. Second, RNA-RNA interactions rely on a lesser regulator-target ratio than protein-DNA interactions, which leads to lower dynamic ranges [17].

Here, we adopted a nonequilibrium thermodynamic scheme to explain RNA-RNA interactions (a nonequilibrium scheme because we considered further energies characterizing the system apart from ΔG_{inter} to derive a different scaling law). As a working example, we considered the IS10 transposition system that occurs in bacteria [18]. There, a sRNA (dubbed RNA-OUT) interacts with the 5' untranslated region (dubbed RNA-IN) of a given mRNA, encoding a transposase, to repress the protein synthesis. RNA-OUT is a structured sRNA with a stem-loop conformation (with an external loop of 6 nucleotides (nt) and a long stem with three small internal loops, as already proposed [18] and here assumed), while RNA-IN is a nonstructured element in which the Shine-Dalgarno box [19] and the start codon are exposed to the solvent. In the following, we describe the energy landscape associated with the interaction at the base-pair resolution and, in turn, we derive a simple but general statistical model that emerges from a regulatory scenario in steady state but out of equilibrium [20].

II. RESULTS

A. Energy landscape

To resolve the energy landscape associated with the sRNA-mRNA interaction with precision, the different gains and losses of free energy that occur as the reaction coordinate progresses need to be computed [Fig. 1(a)]. We defined the reaction coordinate as the number of intermolecular base pairs. In the particular case of the IS10 system, this ranges from 0 (no interaction; state 1) to 33 base pairs (bp) (final complex; state 3). Notably, the different secondary structures that are progressively formed in that interaction range (base pair by base pair) can be evaluated by hand according to a simplified physicochemical model with the stacking and looping free energies previously determined at 37 °C [21,22]. To this end, we followed a decomposition principle by which the free energy of a given structure is the sum of the individual contributions of each stack and loop; that is,

$$G(\text{structure}) = \sum_i G_i(\text{stack or loop}). \quad (2)$$

Initially, the species are just folded intramolecularly [Fig. 1(b)], as this process occurs much faster than the eventual interaction (in the order of seconds or even minutes) [23]. The free energy of RNA-IN is directly $G_{\text{IN}} = 0$ and the free energy of RNA-OUT $G_{\text{OUT}} = -20.5$ Kcal/mol (from 20 stacks and 4 loops). Then, the free energy of the state 1 reads $G_1 = G_{\text{IN}} + G_{\text{OUT}} = G_{\text{OUT}}$.

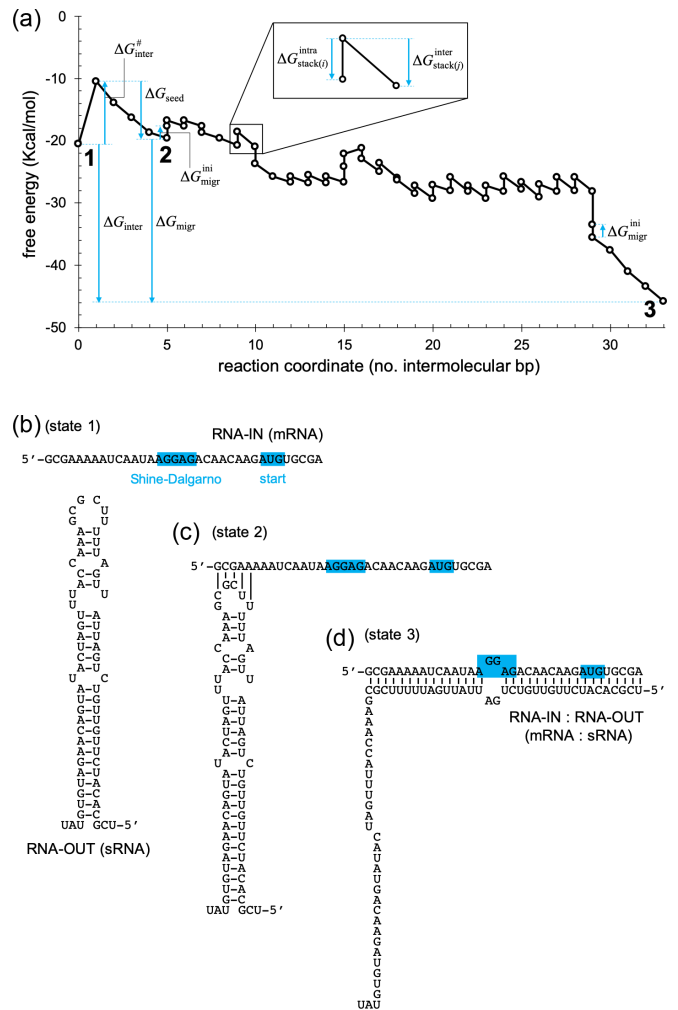


FIG. 1. (a) Energy landscape of the sRNA-mRNA interaction in the IS10 system at the base-pair resolution. The intramolecular folding state is labeled as 1, the just-the-seed-paired intermolecular folding state as 2, and the final intermolecular folding state as 3. Vertical arrows (in blue) mark the different free energies that characterize the interaction. (b) Intramolecular RNA secondary structures in state 1. (c) RNA secondary structure in the state 2 with intra- and intermolecular contacts. (d) Intermolecular RNA secondary structure in state 3. The Shine-Dalgarno box and the start codon are shaded (in blue).

The sRNA-mRNA interaction can be divided into three different stages. In first place, the two species have to meet in order to start the interaction. This association clearly incurs in a thermodynamic penalty (denoted by $\Delta G_{\text{inter}}^{\#}$) due to entropic considerations [21]. Here, we considered $\Delta G_{\text{inter}}^{\#} \approx 10$ Kcal/mol without loss of generality (Fig. 1(a); note that in conventional physicochemical models $\Delta G_{\text{inter}}^{\#} \approx 4$ Kcal/mol [21,22], but here we followed recent estimations of this energy barrier [24]). This value can be assumed, to some extent, independent of the sequence, and it is expected to vary from *in vitro* to *in vivo* due to molecular crowding effects or spurious interactions that can take place in the cell [25]. In second place, the seed regions of both RNAs interact to form a partly-stabilized complex [state 2; Fig. 1(c)]. Here, the seed region consists of 5 nt, from which we calculated

$\Delta G_{\text{seed}} = -9.1$ Kcal/mol, and in the case of RNA-OUT they are located in the external loop of its structure (scenario of kissing loop). Then, the free energy of the state 2 reads $G_2 = G_1 + \Delta G_{\text{inter}}^{\#} + \Delta G_{\text{seed}}$. Note that G_2 might even be lesser than G_1 if the seed region were long enough [24]. In third place, the intermediate complex starts a process of branch migration to break the intramolecular base pairs and form the intermolecular ones, ending in a stable hybridization state between the two RNAs [Fig. 1(d)]. By aggregating the free energies of all 27 stacks and one loop of this final complex, we obtained $G_{\text{IN:OUT}} = -55.8$ Kcal/mol; the free energy of the state 3 is then $G_3 = \Delta G_{\text{inter}}^{\#} + G_{\text{IN:OUT}}$. For branch migration to start, nonetheless, the system incurs in a thermodynamic penalty (denoted by $\Delta G_{\text{migr}}^{\text{ini}}$), as the resulting secondary structures that need to be formed are suboptimal. We considered $\Delta G_{\text{migr}}^{\text{ini}} \approx 2$ Kcal/mol, which roughly corresponds to the gain in free energy if the external loop of RNA-OUT is enlarged. This small gain is released once the intramolecular structure is completely destroyed [at a reaction coordinate of 29 bp; Fig. 1(a)]. For simplicity, we neglected potential intramolecular interactions occurring in the 3' end of RNA-OUT upon hybridization [Fig. 1(d)]. Denoting by ΔG_{migr} the free energy gap between the states 2 and 3 ($G_3 = G_2 + \Delta G_{\text{migr}}$), we were able to write

$$\Delta G_{\text{inter}} = \Delta G_{\text{inter}}^{\#} + \Delta G_{\text{seed}} + \Delta G_{\text{migr}}, \quad (3)$$

with $\Delta G_{\text{inter}} = -25.3$ Kcal/mol. Clearly, our calculations agree with a decomposition of the intermolecular interaction into incremental steps, i.e., $\Delta G_{\text{migr}} + \Delta G_{\text{seed}} = G_{\text{IN:OUT}} - G_{\text{OUT}}$ is satisfied [26].

This detailed energy landscape can be abstracted into a simple figure if we consider the whole branch migration as a single process. For that, its different elementary steps were assumed fast and reversible so that detailed balance can be applied [27]. Accordingly, for $6 \leq r \leq 29$, it can be written $k_{r,r+1}/k_{r+1,r} \propto e^{-\beta(\Delta G_r^{\text{inter}} - \Delta G_r^{\text{intra}})}$, where $\Delta G_r^{\text{intra}}$ corresponds to the broken intramolecular base pair at the reaction coordinate r and $\Delta G_r^{\text{inter}}$ to the formed intermolecular base pair, thereby increasing the reaction coordinate from r to $r+1$. Consequently, by considering a Markov chain the whole branch migration, we defined an effective free energy barrier between the states 2 and 3 as $\Delta G_{\text{migr}}^{\#} = \Delta G_{\text{migr}}^{\text{ini}} - G_1$ (i.e., the total free energy gain as the intramolecular structure of RNA-OUT is destroyed, noting that $|G_{\text{OUT}}|$ is indeed the gross gain). In this way, the resulting energy landscape has three stable states and two energy barriers (Fig. 2), and a practical mathematical formulation can then be developed. According to the transition state theory, a given kinetic rate is proportional to the Boltzmann factor of the free energy barrier faced [28]. Thus,

$$\begin{aligned} k_{12} &\propto e^{-\beta \Delta G_{\text{inter}}^{\#}}, & k_{21} &\propto e^{\beta \Delta G_{\text{seed}}}, \\ k_{23} &\propto e^{-\beta \Delta G_{\text{migr}}^{\#}}, & k_{32} &\propto e^{\beta(\Delta G_{\text{migr}} - \Delta G_{\text{migr}}^{\#})}, \end{aligned} \quad (4)$$

where k_{ij} is the kinetic constant to go from the state i to the state j (the different ΔG s correspond to free energy gaps between transition and stable states, *sensu* Arrhenius).

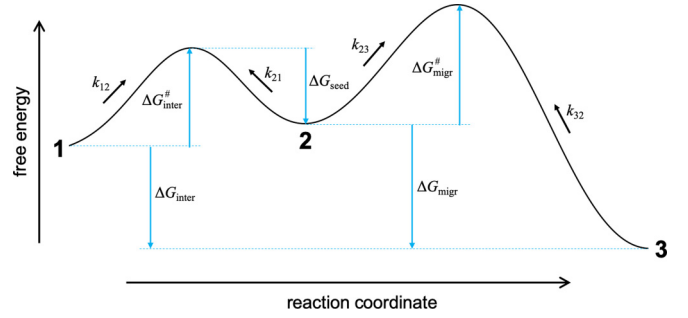


FIG. 2. Abstracted energy landscape of the sRNA-mRNA interaction in the IS10 system. There are two transition states and three stable states (labeled as 1, 2, and 3). Vertical arrows (in blue) mark the different free energies that characterize the interaction. Oblique arrows over the landscape (in black) designate the kinetic constants.

B. Kinetic modeling

Our aim was to derive a scaling law relating r with the free energies characterizing the system, as this is instrumental to predict gene expression changes directly from the sequence (i.e., sequence-to-function mapping), as the RNA folding problem is solved. To this end, we developed a kinetic model, where we considered transcription and degradation of RNA, as well as reversible transitions among the different conformational states. If we denote by x_i the concentration of the RNA species in the folding state i (x_1 for the mRNA and x'_1 for the sRNA in the state 1), the system of nonlinear ordinary differential equations that govern the sRNA-mRNA interaction reads

$$\begin{aligned} \dot{x}_1 &= \alpha - k_{12}x_1x'_1 - \delta x_1 + k_{21}x_2, \\ \dot{x}'_1 &= n\alpha - k_{12}x_1x'_1 - \delta x'_1 + k_{21}x_2, \\ \dot{x}_2 &= k_{12}x_1x'_1 - (k_{21} + k_{23} + \delta)x_2 + k_{32}x_3, \\ \dot{x}_3 &= k_{23}x_2 - (k_{32} + \delta)x_3, \end{aligned} \quad (5)$$

where α is transcription rate of the mRNA, n is the relative transcription rate of the sRNA with respect to the mRNA (e.g., due to different gene copy numbers or promoter strengths), and δ the RNA degradation rate (for simplicity, assumed equal for all species, single or complex). Of note, if we consider the total amount of mRNA, given by $y = x_1 + x_2 + x_3$, it turns out that $\dot{y} = \alpha - \delta y$, which entails mass conservation. The dynamic regulatory range (ratio of protein concentrations with and without sRNA) can be defined as $r = 1 - \frac{x_3}{y}$ (i.e., $P_{\text{bound}} = \frac{x_3}{y}$), assuming that the translation rate per mRNA is equal and substantial in states 1 and 2, but 0 in state 3.

Numerical simulations revealed how, upon the expression of the sRNA, the concentrations of the different species evolve to reach a stationary picture in which all folding states are represented [Fig. 3(a)]. According to kinetic parameter values inferred from experimental data (in studies of kissing loop interactions) [23], this process takes about half an hour, the system is far from being completely in the final intermolecular folding state (about two-thirds of mRNAs are paired to sRNAs), and the just-the-seed-paired intermolecular folding state is the least represented. Our simulations confirm that r improves by promoting the formation of the final intermolecular complex [i.e., increasing k_{23} and reducing k_{21} ; Fig. 3(b)],

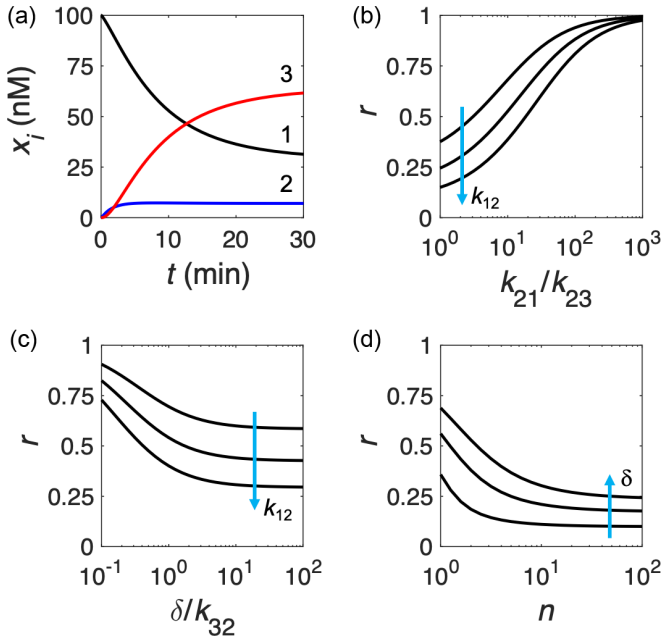


FIG. 3. Numerical simulations of the RNA-based regulated response with a kinetic model. (a) Time-dependent concentrations (x_i) of the different RNA species that correspond to the three stable states (labeled as 1, 2, and 3). The initial condition corresponds to a scenario of absence of sRNA. The kinetic parameters are $n = 1$, $\alpha = 10$ nM/min, $\delta = 0.1$ min $^{-1}$, $k_{12} = 0.06$ nM $^{-1}$ min $^{-1}$, $k_{21} = 6$ min $^{-1}$, and $k_{32} = 0.01$ min $^{-1}$, and the initial condition is $x_1 = 100$ nM and $x_2 = x_3 = 0$. (b)–(d) Dynamic regulatory range (r) as a function of distinct kinetic parameters. k_{12} varies from 0.01 to 0.1 nM $^{-1}$ min $^{-1}$ in (b) and (c), and δ varies from 0.1 to 0.3 min $^{-1}$ in (d).

by enforcing the irreversibility of this process [i.e., increasing δ and reducing k_{32} ; Fig. 3(c)], and by working with an excess of sRNA in the cell [i.e., increasing n ; Fig. 3(d)]. However, the improvement is limited due to the nonlinear nature of the system, a consequence of accounting for intermediate state 2. Indeed, this acts as a buffer to balance sRNA release (to go back to state 1) and branch migration (to go forward to state 3).

The system can be analytically solved in steady state if we consider that the sRNA is much more expressed than the mRNA ($n \gg 1$). In this regard, the mRNA amount limits the reaction and $x'_1 \approx \frac{n\alpha}{\delta}$ is a good approximation. Thus, we obtained

$$x_1 \approx \frac{\alpha/\delta}{1 + \frac{n\alpha}{\delta} \frac{k_{12}}{k_{21}} \left(\frac{k_{23}}{k_{32} + \delta} \right)}, \quad (6)$$

and by exploiting the relationships shown in Eqs. (4) we finally found

$$r \approx \frac{1}{1 + [e^{\beta(\Delta G_{\text{inter}} - \mu)} + e^{\beta(\Delta G_{\text{trans}} - \mu')}]^{-1}}, \quad (7)$$

where $\Delta G_{\text{trans}} = \max(\Delta G_{\text{seed}}, \Delta G_{\text{seed}}^{\text{wild-type}}) + \Delta G_{\text{migr}}^{\#}$ (a variable > 0). As in the case of Eq. (1), μ (also μ') represents a kind of chemical potential of the regulator ($e^{\beta\mu} \propto e^{\beta\mu'} \propto x'_1$). To obtain a compact expression in energetic terms, we considered k_{21} as the fastest kinetics in the system.

Certainly, in the case of structured sRNAs, $\Delta G_{\text{migr}}^{\#} > -\Delta G_{\text{seed}}$ typically holds, meaning that once the just-the-seed-paired intermolecular folding state is reached, the reaction tends to preferentially go back (i.e., $k_{21} > k_{23}$) [24]. In addition, if n is moderate, $k_{21} > \frac{n\alpha}{\delta} k_{12}$ also holds. The contribution of ΔG_{seed} is saturated by $\Delta G_{\text{seed}}^{\text{wild-type}}$ to account for cases in which $k_{21} < \frac{n\alpha}{\delta} k_{12}$ (this saturation effect is in tune with what follows from DNA reactions *in vitro*, i.e., saturation of toehold-mediated kinetics [27]). Interestingly, ΔG_{trans} is in essence the free energy gap between the two transition states in Fig. 2, which makes Eq. (7) a satisfying combination of a standard Gibbs formulation with the Curtin-Hammett principle [29].

It is worthwhile to note that for systems in which there is a sufficient free energy gap between states 2 and 3 (e.g., $\Delta G_{\text{migr}} < -20$ Kcal/mol), once the final intermolecular complex is formed, the probability of going back is negligible, because the RNA will be degraded before this process can take place (i.e., $\delta \gg k_{32}$). Hence, $x_3 \approx \frac{k_{23}}{\delta} x_2$, which already breaks the conventional scheme of thermodynamic equilibrium. In this case, the dynamic range is independent of ΔG_{inter} and the model predicts a maximal performance of $r \approx e^{\beta(\Delta G_{\text{trans}} - \mu')}$. Consequently, a stable seed pairing (e.g., enriched in GC content) and weak intramolecular structures (e.g., with internal loops and low GC content) are requirements to achieve a high dynamic range.

Equation (7) corresponds to a scenario of nonequilibrium and differs from Eq. (1), the conventional statistics at equilibrium [10] that was here considered as the null model. However, the system of Eqs. (5) is general enough so that Eq. (1) can also be derived from it. If we consider that the sRNA-mRNA interaction is a process much faster than the transcription and degradation of these RNA species, we can write $k_{12}x'_1x_1 = k_{21}x_2$ and $k_{23}x_2 = k_{32}x_3$, which means detailed balance, and Eq. (1) naturally arises after some calculations. Note also that in the limit $\mu' \rightarrow +\infty$ (or $\Delta G_{\text{trans}} \ll \mu'$), we get Eq. (1) directly from Eq. (7), which makes equilibrium an asymptote of a more complex scenario.

Equation (1) predicts that even if there is no accessibility (i.e., no seed region, $\Delta G_{\text{seed}} = 0$), the protein can be down-regulated provided the free energy of interaction is low enough (i.e., $\Delta G_{\text{inter}} < \mu$). However, this does not occur with the scaling law here presented, because $\Delta G_{\text{seed}} = 0$ entails $\Delta G_{\text{trans}} > \mu'$ and then the term $e^{\beta(\Delta G_{\text{trans}} - \mu')}$ always contributes to have high values of r (i.e., poor regulation), even if $e^{\beta(\Delta G_{\text{inter}} - \mu)} \approx 0$. In other words,

$$\lim_{\Delta G_{\text{inter}} \rightarrow -\infty} r = 0 \text{ (equilibrium)},$$

$$\lim_{\Delta G_{\text{inter}} \rightarrow -\infty} r = \frac{1}{1 + e^{-\beta(\Delta G_{\text{trans}} - \mu')}} \text{ (nonequilibrium)}. \quad (8)$$

C. Experimental validation

To validate our theory, we considered the dynamic ranges reported for a large series of mutants of the IS10 system (mutations introduced in both the sRNA and the mRNA), which were experimentally measured by changes in fluorescence [11]. For each system, we calculated the values of ΔG_{inter} with the RNAfold routine of the ViennaRNA package [30].

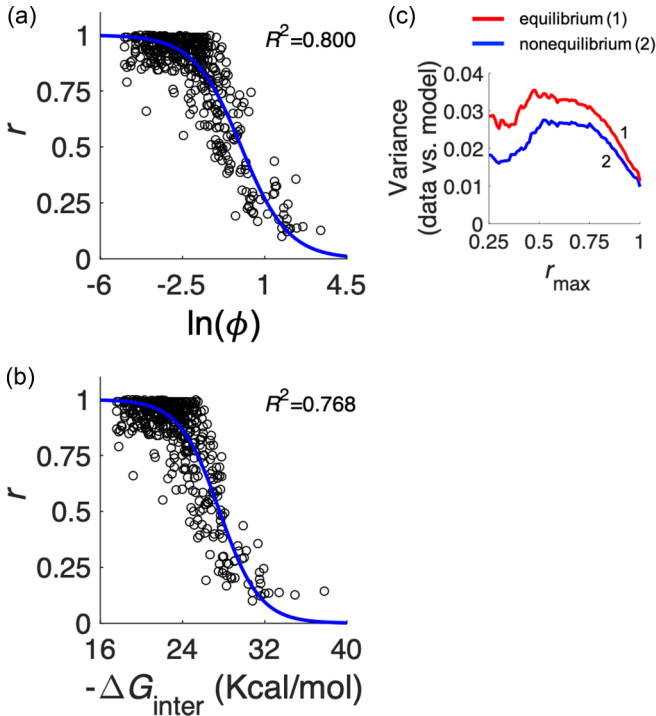


FIG. 4. Correlation of experimentally measured dynamic range (r) against a computationally predicted free energy for different mutants of the IS10 system (529 data points). (a) Nonequilibrium thermodynamic model (r vs ϕ), fitted with $\beta = 0.545$ mol/Kcal, $\mu = -26.6$ Kcal/mol, and $\mu' = 20.7$ Kcal/mol. (b) Equilibrium thermodynamic model (r vs ΔG_{inter} , this corresponds to the null model), fitted with $\beta = 0.514$ mol/Kcal and $\mu = -27.6$ Kcal/mol. (c) Variance of experimental data vs model predictions for data with a maximal dynamic range (r_{max}). $r_{\text{max}} = 1$ means that all data are considered, while $r_{\text{max}} = 0.25$ means that only those systems with substantial regulatory activity are considered.

We obtained values that differ a bit from our previous manual calculations as a consequence of a different value of $\Delta G_{\text{inter}}^{\#}$ (this is not important because it is a systematic discrepancy that only affects the fitted value of μ) and the consideration of additional physicochemical features [20]. Moreover, we calculated the corresponding values of ΔG_{seed} by following the aforementioned manual procedure, considering the sequence motifs previously described in the species RNA-IN and RNA-OUT [11]. In the case of an intermolecular structure between the seed regions with a loop or bulge, only the most stable consecutive stackings were considered. To estimate $\Delta G_{\text{migr}}^{\#}$, we calculated $|G_{\text{OUT}}|$ with the RNAfold routine [30] by imposing the seed region to be unpaired, and then we added the value of $\Delta G_{\text{migr}}^{\text{ini}}$. For RNA-OUT mutants in which the external loop is very large, we set $\Delta G_{\text{migr}}^{\text{ini}} \approx 0$, as the base pairing that follows the seeding is not a process of branch migration but between unpaired regions [26].

We then represented the experimental values of r vs the computed free energies. For representation purposes, we defined $\phi = [e^{\beta(\Delta G_{\text{inter}} - \mu)} + e^{\beta(\Delta G_{\text{trans}} - \mu')}]^{-1}$, revealing a good agreement between theory and experiments with our scaling law from nonequilibrium thermodynamics [Fig. 4(a)]. ΔG_{inter} still exhibits substantial prediction ability [Fig. 4(b)], through

the application of the null model], as $e^{\beta(\Delta G_{\text{inter}} - \mu)}$ is the dominant term in ϕ for most of the mutant systems. Nonetheless, an F test showed a statistically significant better prediction ability in the former case ($F_{1,526} = 81.7$, $P < 10^{-6}$). By using partial least squares regression, Arkin and co-workers found empirically that ΔG_{inter} and ΔG_{seed} explain most of the variance in repression fold, ΔG_{inter} being the main predictor [11]. According to our results, obtained by following a bottom-up approach, ΔG_{inter} can account for up to 76.8% of the variance, while when both ΔG_{inter} and ΔG_{trans} (which depends on ΔG_{seed}) are considered the percentage increases 3.2%. The explained variance increases more (with respect to the null model) when only data for systems with significant regulatory activity are considered, i.e., when the term $e^{\beta(\Delta G_{\text{inter}} - \mu)}$ is close to 0 and then the contribution of the term $e^{\beta(\Delta G_{\text{trans}} - \mu')}$ is manifested [Fig. 4(c)].

In theory, the value of β should be fixed. At 37 °C, we would have $\beta = 1.62$ mol/Kcal. However, previous work that correlated gene expression data with computationally predicted free energies revealed values of β much lower ($\beta \approx 0.5$ mol/Kcal) [31,32]. This might indicate that the physicochemical models used for RNA folding overestimate the free energies of the processes that occur *in vivo* and/or that the reduction of the whole ensemble to a single folding (for tractable purposes, the minimal free energy structure) leads to lose the extant functional variability. In this regard, here we let the value of the β variable to be fitted against the data, in order to work with computational predictions of energies and structures, obtaining $\beta \approx 0.5$ – 0.6 mol/Kcal irrespective of Eqs. (1) or (7). In any case, the observed exponential trend between r and predicted ΔG s indeed reflects that thermodynamic principles are applicable to formally describe these systems.

III. CONCLUSION

This work represents a significant step forward in our basic understanding of RNA-based regulation. An original scaling law for expression prediction from computational energetic and structural calculations is derived by following a nonequilibrium thermodynamic scheme in steady state. Although it is particularized for the IS10 transposition system, we expect its validity to enable the study of further riboregulatory systems in which an sRNA acts as a repressor of translation (it might also be adapted straightforwardly to deal with an activator of a *cis*-repressed mRNA) [5]. This law is dependent on different free energies, not only the free energy of formation, as a result of RNA degradation. The more efficient this enzymatic process (as it is in bacteria), the further from a scenario of equilibrium. Accordingly, ΔG_{inter} will be a suitable predictor of regulatory activity when the final complex is not stable enough, as degradation will not be the dominant process at that moment, while ΔG_{trans} will govern the regulation when the global reaction is very favorable. Importantly, our model can be exploited to study natural RNA-based regulation or even to guide the design of synthetic systems [17,33]. Of note, our results contribute to appreciate the importance of the seed region, disclosing from *ab initio* calculations its impact on regulatory activity [11]. In addition, our results reveal a trade-off in gene regulatory systems based on RNA-RNA

interactions. On the one hand, the sRNA needs to be structured to enhance its stability in a cellular context, avoiding a premature degradation by the action of ribonucleases [34]. On the other hand, that structure contributes to difficult the process of branch migration by imposing an effective free energy barrier [27]. In sum, as we realize that enzymatic activity manifests into breaking detailed balance over some biochemical processes, we expect a wider application of nonequilibrium thermodynamics to predictably map genotype and phenotype, especially if this leads to amenable mathematical expressions connecting microscopic descriptors of the molecular world with mesoscopic parameters [20].

ACKNOWLEDGMENTS

Work was supported by Grant No. PGC2018-101410-B-I00 (SYSY-RNA) from the Spanish Ministry of Science, Innovation, and Universities (cofinanced by the European Regional Development Fund). L.G. was supported by CSIC JAE Intro Fellowship No. JAEINT-20-00714.

APPENDIX

1. Stability analysis

The Jacobian matrix of the system of Eqs. (5) is

$$J = \begin{pmatrix} -k_{12}x'_1 - \delta & -k_{12}x_1 & k_{21} & 0 \\ -k_{12}x'_1 & -k_{12}x_1 - \delta & k_{21} & 0 \\ k_{12}x'_1 & k_{12}x_1 & -k_{21} - k_{23} - \delta & k_{32} \\ 0 & 0 & k_{23} & -k_{32} - \delta \end{pmatrix}. \quad (\text{A1})$$

In the limit $x'_1 \approx \frac{n\alpha}{\delta}$ (constant), the system can be reduced in one dimension, then arriving to the following matrix:

$$J_{\text{reduced}} = \begin{pmatrix} -k_{12}\frac{n\alpha}{\delta} - \delta & k_{21} & 0 \\ k_{12}\frac{n\alpha}{\delta} & -k_{21} - k_{23} - \delta & k_{32} \\ 0 & k_{23} & -k_{32} - \delta \end{pmatrix}. \quad (\text{A2})$$

Importantly, this matrix is strictly column diagonally dominant and all elements in the diagonal are negative. This entails that the real parts of all eigenvalues are negative (condition of stability).

2. Derivation of a thermodynamic model

In the steady state, it can be written that

$$\begin{aligned} \alpha - k_{12}x_1x'_1 - \delta x_1 + k_{21}x_2 &= 0, \\ n\alpha - k_{12}x_1x'_1 - \delta x'_1 + k_{21}x_2 &= 0, \end{aligned}$$

$$\begin{aligned} k_{12}x_1x'_1 - (k_{21} + k_{23} + \delta)x_2 + k_{32}x_3 &= 0, \\ k_{23}x_2 - (k_{32} + \delta)x_3 &= 0. \end{aligned} \quad (\text{A3})$$

First, by combining the first and second equations, we obtained $(n-1)\alpha = \delta(x'_1 - x_1)$. Then, by considering $n \gg 1$, which entails $x'_1 \gg x_1$, it turned out that $x'_1 \approx \frac{n\alpha}{\delta}$ (constant). Second, by combining the third and fourth equations, we obtained $\frac{n\alpha}{\delta}k_{12}x_1 = (k_{21} + \delta + \frac{k_{23}\delta}{k_{32} + \delta})x_2$, which was further simplified to $\frac{n\alpha}{\delta}k_{12}x_1 \approx (k_{21} + \frac{k_{23}\delta}{k_{32} + \delta})x_2$, by considering $k_{21} \gg \delta$. Therefore, we obtained $y \approx x_1 + \frac{k_{23}}{k_{32} + \delta}x_2$. Then, we were able to write $r \approx \frac{x_1 + x_2}{x_1 + \frac{k_{23}}{k_{32} + \delta}x_2}$. By doing some rearrangements, we obtained

$$r \approx \frac{1}{1 + \frac{\frac{n\alpha}{\delta}k_{12}(\frac{k_{23}}{k_{32} + \delta} - 1)}{\frac{n\alpha}{\delta}k_{12} + k_{21} + \frac{k_{23}\delta}{k_{32} + \delta}}}. \quad (\text{A4})$$

From here, we assumed $k_{21} \gg \frac{n\alpha}{\delta}k_{12} + \frac{k_{23}\delta}{k_{32} + \delta}$ and $\frac{k_{23}}{k_{32} + \delta} \gg 1$, which allowed us to write

$$r \approx \frac{1}{1 + \frac{n\alpha}{\delta} \frac{k_{12}}{k_{21}} \left(\frac{k_{23}}{k_{32} + \delta} \right)}. \quad (\text{A5})$$

Finally, by knowing that $\frac{k_{12}k_{23}}{k_{21}k_{32}} \propto e^{-\beta\Delta G_{\text{inter}}}$ and that $\frac{k_{12}k_{23}}{k_{21}} \propto e^{-\beta\Delta G_{\text{trans}}}$, we have Eq. (7). If $k_{32} \gg \delta$, we just recovered Eq. (1).

3. Formulation in terms of on and off kinetics

If we denote by k_{on} the effective kinetic constant to form the final sRNA-mRNA complex and by k_{off} the constant to dissociate it, it can be written that

$$P_{\text{bound}} = \frac{k_{\text{on}}}{k_{\text{on}} + k_{\text{off}}}, \quad (\text{A6})$$

and then

$$r = \frac{1}{1 + \frac{k_{\text{on}}}{k_{\text{off}}}}. \quad (\text{A7})$$

In a scenario of equilibrium, we have $k_{\text{on}} = \frac{n\alpha}{\delta}k_{12}k_{23}$ (i.e., the global kinetic rate to go from state 1 to state 3 according to the energy landscape shown in Fig. 2) and $k_{\text{off}} = k_{21}k_{32}$ (i.e., the global kinetic rate to go from state 3 to state 1). However, in a scenario of nonequilibrium, the degradation of the RNA molecules is considered, and then the kinetic constants cannot be directly derived from the energy landscape. In this case, there are two off processes from state 3, one to go back to state 1 and another to degrade the RNA molecules. By using Eq. (A5), it turned out $k_{\text{off}} = k_{21}k_{32} + k_{21}\delta$, maintaining the definition of k_{on} as in the scenario of equilibrium.

- [1] F. Jacob and J. Monod, Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol.* **3**, 318 (1961).
 [2] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett *et al.*, Transcriptional regulatory networks in *Saccharomyces cerevisiae*, *Science* **298**, 799 (2002).

- [3] D. Sprinzak and M. B. Elowitz, Reconstruction of genetic circuits, *Nature (London)* **438**, 443 (2005).
 [4] O. Dahan, H. Gingold, and Y. Pilpel, Regulatory mechanisms and networks couple the different phases of gene expression, *Trends Genet.* **27**, 316 (2011).

- [5] L. S. Waters and G. Storz, Regulatory RNAs in bacteria, *Cell* **136**, 615 (2009).
- [6] S. Guil and M. Esteller, RNA-RNA interactions in gene regulation: The coding and noncoding players, *Trends Biochem. Sci.* **40**, 248 (2015).
- [7] J. C. Paillart, M. Shehu-Xhilaga, R. Marquet, and J. Mak, Dimerization of retroviral RNA genomes: An inseparable pair, *Nat. Rev. Microbiol.* **2**, 461 (2004).
- [8] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Transcriptional regulation by the numbers: Models, *Curr. Opin. Genet. Dev.* **15**, 116 (2005).
- [9] R. C. Brewster, F. M. Weinert, H. G. Garcia, D. Song, M. Rydenfelt, and R. Phillips, The transcription factor titration effect dictates level of gene expression, *Cell* **156**, 1312 (2014).
- [10] R. Phillips, Napoleon is in equilibrium, *Annu. Rev. Condens. Matter Phys.* **6**, 85 (2015).
- [11] V. K. Mutalik, L. Qi, J. C. Guimaraes, J. B. Lucks, and A. P. Arkin, Rationally designed families of orthogonal RNA regulators of translation, *Nat. Chem. Biol.* **8**, 447 (2012).
- [12] G. Rodrigo, T. E. Landrain, E. Majer, J. A. Daros, and A. Jaramillo, Full design automation of multi-state RNA devices to program gene expression using energy-based optimization, *PLoS Comput. Biol.* **9**, e1003172 (2013).
- [13] N. Peterman, A. Lavi-Itzkovitz, and E. Levine, Large-scale mapping of sequence-function relations in small regulatory RNAs reveals plasticity and modularity, *Nucleic Acids Res.* **42**, 12177 (2014).
- [14] J. K. Wickiser, W. C. Winkler, R. R. Breaker, and D. M. Crothers, The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch, *Mol. Cell* **18**, 49 (2005).
- [15] S. Guedich, B. Puffer-Enders, M. Baltzinger, G. Hoffmann, C. Da Veiga, F. Jossinet, S. Thore, G. Bec, E. Ennifar, D. Burnouf, and P. Dumas, Quantitative and predictive model of kinetic regulation by E. coli TPP riboswitches, *RNA Biol.* **13**, 373 (2016).
- [16] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, Global quantification of mammalian gene expression control, *Nature (London)* **473**, 337 (2011).
- [17] G. Rodrigo, S. Prakash, S. Shen, E. Majer, J. A. Daros, and A. Jaramillo, Model-based design of RNA hybridization networks implemented in living cells, *Nucleic Acids Res.* **45**, 9797 (2017).
- [18] J. D. Kittle, R. W. Simons, J. Lee, and N. Kleckner, Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5' end of the target RNA and a loop in the anti-sense RNA, *J. Mol. Biol.* **210**, 561 (1989).
- [19] J. Shine and L. Dalgarno, Determinant of cistron specificity in bacterial ribosomes, *Nature (London)* **254**, 34 (1975).
- [20] F. S. Gnesotto, F. Mura, J. Gladrow, and C. P. Broedersz, Broken detailed balance and non-equilibrium dynamics in living systems: A review, *Rep. Prog. Phys.* **81**, 066601 (2018).
- [21] T. Xia, J. SantaLucia Jr., M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner, Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick pairs, *Biochemistry* **37**, 14719 (1998).
- [22] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA secondary structure, *J. Mol. Biol.* **288**, 911 (1999).
- [23] T. Franch, M. Petersen, E. G. H. Wagner, J. P. Jacobsen, and K. Gerdes, Antisense RNA regulation in prokaryotes: Rapid RNA/RNA interaction facilitated by a general U-turn loop structure, *J. Mol. Biol.* **294**, 1115 (1999).
- [24] N. Salim, R. Lamichhane, R. Zhao, T. Banerjee, J. Philip, D. Rueda, and A. L. Feig, Thermodynamic and kinetic analysis of an RNA kissing interaction and its resolution into an extended duplex, *Biophys. J.* **102**, 1097 (2012).
- [25] I. Schoen, H. Krammer, and D. Braun, Hybridization kinetics is different inside cells, *Proc. Natl. Acad. Sci. USA* **106**, 21649 (2009).
- [26] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.103.042410> for dataset S1 where we describe the quantitative resolution of the energy landscape at the base-pair level and provide the free energy values of all IS10 sequences analyzed.
- [27] N. Srinivas, T. E. Ouldridge, P. Sulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye, and E. Winfree, On the biophysics and kinetics of toehold-mediated DNA strand displacement, *Nucleic Acids Res.* **41**, 10641 (2013).
- [28] K. J. Laidler and M. C. King, The development of transition-state theory, *J. Phys. Chem.* **87**, 2657 (1983).
- [29] J. I. Seeman, The Curtin-Hammett principle and the Winstein-Holness equation: New definition and recent extensions to classical concepts, *J. Chem. Educ.* **63**, 42 (1986).
- [30] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, Fast folding and comparison of RNA secondary structures, *Monatsh. Chem.* **125**, 167 (1994).
- [31] Y. Hao, Z. J. Zhang, D. W. Erickson, M. Huang, Y. Huang, J. Li, T. Hwa, and H. Shi, Quantifying the sequence-function relation in gene silencing by bacterial small RNAs, *Proc. Natl. Acad. Sci. USA* **108**, 12473 (2011).
- [32] H. M. Salis, E. A. Mirsky, and C. A. Voigt, Automated design of synthetic ribosome binding sites to control protein expression, *Nat. Biotechnol.* **27**, 946 (2009).
- [33] A. A. Green, P. A. Silver, J. J. Collins, and P. Yin, Toehold switches: De-novo-designed regulators of gene expression, *Cell* **159**, 925 (2014).
- [34] M. Saramago, C. Barria, dos R. F. Santos, I. J. Silva, V. Pobre, S. Domingues, J. M. Andrade, S. C. Viegas, and C. M. Arraiano, The role of RNases in the regulation of small RNAs, *Curr. Opin. Microbiol.* **18**, 105 (2014).