

On noise and silence in small RNA regulation

Erel Levine^{*}, Min Huang^{† ‡}, Yingwu Huang^{§ ¶}, Thomas Kuhlman^{*}, Hualin Shi[‡], Zhongge Zhang^{||}, and Terence Hwa^{**}
^{*} [†]

^{*}Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, CA 92093-0374, and [†]Center for Advanced Study, Tsinghua University, Beijing, China

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Fluctuations in the number of proteins and RNA molecules can have strong effect on the well-being of a living cell. These molecules are sometimes held in small numbers, and are involved in complex regulatory networks that may amplify small fluctuations. The recently discovered members of the regulatory network, small regulatory RNAs, are sometimes used to keep target mRNAs at extremely low copy number, making it prone to genetic noise. Here we study in detail noise properties of a gene targeted by a bacterial small RNA. This small RNA belongs to a class of regulators which bind specifically to the 5'-UTR of target mRNAs, promoting cleavage of both RNA molecules. Using a simplified mathematical model, we predict that sRNA-mediated gene silencing is generally effective in attenuating noise despite the low number of its target mRNA. In contrast, the onset of gene expression is dominated by anomalous fluctuations, with a surprising noise-induced phenotypic diversity. Even in the absence of any feedback circuit, the mutual interaction of sRNA and its target is predicted to give rise to a bi-modal distribution of population, with cells exhibiting either negligible or substantial gene expression. Using a plasmid-borne synthetic sRNA-target pair, we present experimental evidences supporting the occurrence of this predicted bi-modality in growing *E. coli* cells.

small regulatory RNA | stochastic fluctuations | stochastic gene expression
| post-transcriptional regulation

Living cells often retain key molecules, such as proteins and mRNA, at very low copy number [1]. Fluctuations can therefore be significant, and may have implications on cellular regulation [2, 3] and lead to phenotypic diversity [4, 5, 6]. Sources of genetic noise are typically divided into two components, intrinsic and extrinsic [7, 8, 3]. The first is attributed to the stochastic nature of mRNA and protein synthesis and degradation [9, 10, 11], while the latter stems from global factors within the cell, such as ribosome concentration or degree of DNA supercoiling [7, 8]. Advances in single-cell technologies of recent years allowed researchers to probe the degree of intrinsic vs extrinsic noise at the single-cell level [7, 12], as well as to characterize sources of intrinsic noise [13, 14, 15, 16, 17, 18]. These studies were accompanied by theoretical modeling of genetic noise at the level of a single gene [9, 10] or small modules of gene networks [8, 11, 19, 20, 21].

Most of the studies mentioned above were done in the context of transcriptional initiation control (except for [19], which studied theoretically protein mediated post-transcription regulation). In recent years the central role of small regulatory RNA (sRNA) molecules in gene regulation has been unveiled. Organisms that employ small RNAs for gene silencing range from bacteria to vertebrates [22, 23, 24, 25]. While the pathways and details are different, in all cases small RNAs target specifically messenger RNAs of target genes, based on base-pair complementarity and secondary structure. Upon binding – and with the aid of auxiliary proteins – a target mRNA is either directed for cleavage or its translation is inhibited. Interestingly, some cases are known where bacteria use small RNAs as activators of gene expression [26, 27]. However, a eukaryotic counterpart to the

latter is yet to be discovered.

Is gene silencing by small RNA efficient in terms of suppressing copy-number fluctuations of its target genes? Here we address this question in the context of RyhB, one of the best-characterized small RNAs in *E. coli* [28, 29]. This small RNA, involved in the control of iron homeostasis and oxidative stress [28, 30], represses the expression of several target genes by targeting the 5'-UTR of their mRNAs. Binding of the sRNA to its target promotes degradation of both molecules [29, 31]. Previously, we described the threshold-linear response that characterizes target genes under this mode of regulation [32]. Here we extend our study to investigate the effect of stochastic fluctuations both theoretically and experimentally.

When the synthesis rate of the small RNA significantly exceeds that of its target mRNA, the average copy number of the latter is kept extremely low. Under these circumstances, one might naively expect that mRNA level should fluctuate considerably. However, our model predicts that sRNA regulation keeps the noise level extremely low by reducing the burstiness of translation, consistent with the notion that bursty translation is a major source of intrinsic noise [10, 13, 16, 17].

Average mRNA level is expected to remain low even when synthesis of sRNA and mRNA occur at comparable rates. However, in this case our model predicts overwhelming fluctuations, which may increase mRNA level, even to its unrepressed level. Our model suggest that these broad fluctuations may take the form of a bi-modal distribution, with some cells possessing high mRNA copy number, while other possessing almost none. We tested this prediction experimentally, by constructing synthetic sRNA-target pairs on plasmids, and transforming them into *E. coli* cells. Synthetic sRNAs were based on RyhB, while targets consisted of fusion of a 5'-control-region of a RyhB target to a GFP reporter. Measuring single-cell fluorescence using flow cytometry allowed us to identify population diversity at the threshold of gene expression, as predicted by the model.

Results

Model for sRNA regulation. Consider a simplified model for a small regulatory RNA, which represses expression of a target gene by targeting its 5'-UTR. The two RNA species are coded by independent genes on the bacterial chromosome, and their synthesis can be independently regulated, e.g. by transcription factors, at the two promoters. Small RNAs and target mRNAs are therefore synthesized at two independent rates, denoted α_s and α_m respectively. Typical mRNAs in bacteria are subject to active degradation, which we assume to be linear and characterize by a rate β_m . Small RNAs may

Conflict of interest footnote placeholder

This paper was submitted directly to the PNAS office.

Abbreviations: sRNA, small regulatory RNA;

** To whom correspondence should be addressed. E-mail: hwa@ucsd.edu

©2006 by The National Academy of Sciences of the USA

also be subject to active degradation, or may be diluted due to cell growth. Both processes are coded as a linear depletion of with rate β_s . Finally, binding of the small RNA to its target occludes ribosome binding and prevents translation initiation, and promotes degradation of both molecules. Within our model, this is summarized as an irreversible second-order binding process with rate k , after which the mRNA and sRNA molecules no longer participate in any gene expression process. The values of these parameters can be estimated for various sRNAs [33, 32].

The master equation for the distribution of mRNA, sRNA and protein copy number (m , s and g , resp.) in this model is given by

$$\begin{aligned} \frac{d}{dt}P(m, s, g) = & [\{\alpha_m(a_m - 1) + \beta_m(a_m^+ - 1)m\} \\ & + \{\alpha_s(a_s - 1) + \beta_s(a_s^+ - 1)s\} + k(a_m^+ a_s^+ - 1)ms \quad [1] \\ & + \{\gamma m(a_g - 1) + \delta(a_g^+ - 1)g\}] P(m, s, g), \end{aligned}$$

where the lowering and raising operators are defined through $a_x h(x) = h(x - 1)$, $a_x h(0) = 0$, and $a_x^+ h(x) = h(x + 1)$ for any function $h(x)$. In this study we focus on the steady-state properties of gene expression, and therefore all time derivatives are set hereafter to zero.

Multiplying the master equation by either m or by s , and summing over m , s and g yields

$$\alpha_m - \beta_m \langle m \rangle - k \langle ms \rangle = 0, \quad \alpha_s - \beta_s \langle s \rangle - k \langle ms \rangle = 0. \quad [2]$$

Here $\langle \cdot \rangle$ denotes a *steady-state* ensemble average over different realizations of the stochastic process. As usual, one assumes ergodicity and identifies this average with the population average, namely the quantity measured in culture.

Mass-action model. As a first approximation, we ignore correlations between m and s , and replaces the term $\langle ms \rangle$ on the right hand side of [2] by $\langle m \rangle \langle s \rangle$. The equations then have the same structure as the mass-action model presented in [32]. It is now straightforward to obtain the steady-state concentrations of all molecule species, and investigate their dependence on the different parameters. Let us briefly review these results.

The expression of a target gene depends strongly on the ratio between the transcription rate of the target, α_m , and that of the small RNA, α_s . When the sRNA is produced at much higher rate, all mRNAs are rapidly targeted for degradation, and protein synthesis cannot take place. Gene expression is then repressed ($w\alpha_s$)-fold, where $w \equiv k/(\beta_m\beta_s)$ measures the efficiency of the coupled degradation channel. Based on available data, we estimate that for *E. coli* in

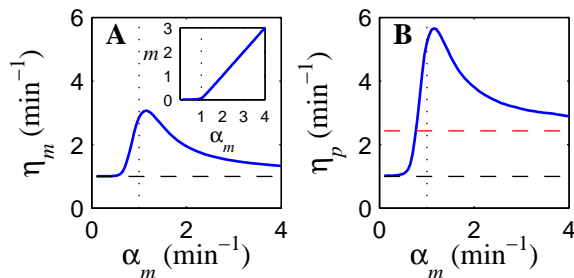


Fig. 1. Noise indices for target mRNA and protein. **A.** For mRNA, noise approaches the 'bare' Poissonian limit $\eta_m = 1/\text{min}$ (dashed line) for high and low transcription rates, but shows enhanced fluctuations near the threshold (dotted line, $\alpha_m \simeq \alpha_s = 1/\text{min}$). **B.** For proteins, the 'bare' noise level depends on the burst size (red dashed line, see text). One observes in addition that noise is attenuated at low expression rates. Here $\beta_m = \beta_s = 0.02/\text{min}$ and $k = 0.1/(\text{nM}\cdot\text{min})$. Inset: Threshold-linear behavior of the mass-action model.

the exponential growth phase, $w\alpha_s$ can be as large as $10^2 - 10^3$ for RyhB [32]. In contrast, if the transcription rate of the target gene significantly exceeds that of the small RNA, repression is effectively relieved. The concentration of mRNA is then given by $(\alpha_m - \alpha_s)/\beta_m + \mathcal{O}(w^{-1})$. In summary, for large w gene expression is effectively silenced below the threshold $\alpha_m \simeq \alpha_s$, and increases linearly with the promoter activity above it. The sharpness ('sensitivity') of the transition at the threshold increases with w . This *threshold linear* behavior is depicted in the inset of figure 1. This behavior was verified experimentally for RyhB and another small RNA in *E. coli* [32].

Noise in target gene expression. We now turn to describe intrinsic noise in the expression of a gene regulated by a small RNA. We first take a bird's eye view of the parameter space, in order to see where noise may play an important part. To this end, we focus on a noise index $\eta_x \equiv C_{x,x}/\langle x \rangle$, where $C_{x,y} = \langle xy \rangle - \langle x \rangle \langle y \rangle$ is the steady-state correlation between the copy numbers of molecules x and y (so $C_{x,x}$ is the variance of x). Note that η_x has the same dimensions as a molecule number.

Multiplying the master equation by m^2 or s^2 , summing over all variables, and using [2], one finds that at steady state

$$\alpha_m - \beta_m C_{m,m} - k C_{m,s} = 0, \quad \alpha_s - \beta_s C_{s,s} - k C_{s,m} = 0. \quad [3]$$

Again, one can make the approximation where m and s are uncorrelated, in which case the variance of each species follows the same steady-state equation as its mean (cf. equation [2]). For the mRNA, these equations correspond to a birth-death process, with birth rate α_m and death rate $\beta_m^* = \beta_m + k\langle s \rangle$. Under this approximation $\eta_m = \eta_s = 1$, just as for a Poisson process. However, it is natural to suspect that at least for some choices of parameters m and s should be highly correlated, and this approximation should not hold. In figure 1A we plot the noise index η_m as obtained from exact numerical integration of the master equation [1]. When the target transcription rate is very high or very low compared with that of the small RNA, η_m indeed approaches 1. However, when the two rates are comparable, correlations between the two species cannot be neglected. In fact, strong anti-correlation between the two results in a large noise index. This phenomenon will be addressed theoretically and experimentally below.

In most cases, it is the number of proteins rather the number of transcribed RNA that affects the behavior of the cell. Indeed, the known natural targets of RyhB all encode functional proteins [28, 30], and accordingly the synthetic reporter we will use below to test our model is the green fluorescent protein (GFP). We therefore direct our attention to η_p , the noise index for protein, plotted in figure 1B (solid

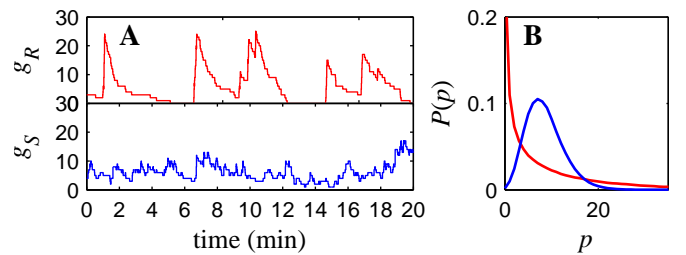


Fig. 2. Noise properties in the silenced state. **A.** Typical temporal evolution for a transcriptionally repressed gene (g_R , red) and an sRNA target (g_S , blue), as obtained from Monte-Carlo simulations. Here $\alpha_m = 0.084(g_R)$ and $0.3(g_S)$, $\alpha_s = 1$, $\beta_m = 0.1$, $\beta_s = 0.02$, $k = 0.1$, $\gamma = 2$ and $\delta = 0.02$, all in min^{-1} . **B.** Steady-state distributions for the two genes, obtained from Monte-Carlo simulations.

line). In this figure, we also plot η_p for proteins that is not regulated by a small RNA, but otherwise have the same characteristics (red dashed line). For such a protein it is straightforward to show that $\eta_p = 1 + b/(1 + \delta/\beta_m)$, where $b \equiv \gamma/\beta_m$ is the size the burst of proteins translated from a single mRNA [10].

Two features of the mRNA noise index propagate to the protein level. First, when the target transcription rate is much larger than that of the sRNA, the effect of the sRNA regulation on noise diminishes, as manifested by the approach of the solid line to the red dashed line in Fig. 1B. Also, as the transcription levels become comparable, the strong noise in the mRNA level is also observed for the protein. One new feature is found as the mRNA transcription level becomes even lower. Here, the noise in the protein level is significantly reduced as compared with an sRNA-free gene.

Noise attenuation in the silenced regime. When gene expression is silenced by the small RNA, the noise index for mRNA is predicted to approach the limit of ‘bare’ Poisson noise. For proteins however, noise is considerably reduced, a remarkable feature given the fact that in this region the average number of molecules is kept very low.

To gain insight into the mechanism of noise attenuation, let us compare the nature of this basal expression – namely gene expression in the repressed state – for two genes. Expression of the first, which we call gene g_R , is repressed at the transcription level by a protein transcription repressor. The second, named gene g_S , is repressed post-transcriptionally by a small RNA. An efficient transcription factor can keep the mRNA synthesis rate (which, at steady state, is the same as the rate of transcription initiation) very low; the small RNA makes the life-time of its target mRNA very low. We can therefore conduct the numerical experiment where the two have the same effect, yielding equal mean mRNA copy numbers for the two genes.

In figure 2A we present typical results of Monte-Carlo simulations [34] for the two genes. Expression of gene g_R is characterized by rare ‘bursts’ of protein, embedded in long periods with not a single protein in the cell. Each burst of proteins corresponds to a single transcription event. The stray mRNA is then amplified by a series of translation events, until the mRNA molecule is degraded. The burst size is therefore given by $b = \gamma/\beta_m$. This type of kinetics has been recently observed experimentally, by tracking production of Venus-Tsr fusion protein, expressed by a lacR-controlled promoter [17].

Let us now follow the kinetics of gene g_S . Here the expression level is much smoother, in the sense that no two distinct behaviors are observed. Rather, gene expression seems to fluctuate around some mean. This is understood by the fact that mRNA transcripts are being constantly synthesized. In the silenced regime, the noise index $\eta_m \simeq 1$, which means that m and s are weakly correlated¹. In this case, as discussed following equation [3], the moments of m are described by the same equations as for a birth-death process, with an effective death rate $\beta_m^* \equiv \beta_m + k\langle s \rangle$. The mRNA transcripts are short-lived, resulting in a decreased burst size $b^* = \gamma/\beta_m^*$. This manifests itself in a reduced noise index, $\eta_p = 1 + b^*/(1 + \delta/\beta_m^*) \simeq 1 + b^*$. Note that the burst size b – and with it the noise index η_p – is now increasing with α_m : an increased target transcription rate results in a smaller sRNA pool (s), and therefore in larger bursts.

Another possible manifestation of the different kinetics in the two cases is the shape of the probability distribution, demonstrated in figure 2B. For a translationally repressed gene, the maximum probability occurs at zero protein expression. For an sRNA repressed gene with the same average, the probability may be maximized at low but finite protein number, as in the example given here.

Can one probe these noise properties experimentally? Naturally,

with multi-copy plasmid, noise is expected to become smoother even in the transcription-regulated case, as the burst frequency is approximately linear in the plasmid copy number. A single chromosomal reporter target is required, accompanied by a single-molecule resolution, as in [17]. However, the temporal resolution of a system based on fluorescent reporter is estimated at about 3 minutes [17]. While this resolution is good enough to track rare bursts, it is not expected to suffice for the smooth kinetics of an sRNA target.

Noise amplification at the onset of gene expression – Transcriptional noise. According to the threshold-linear picture presented above (inset of figure 2, [32]), whenever transcription of the two genes – the small RNA and its target – occurs at similar rates, the steady-state average of both species is considerably suppressed. In this case, where all participants in the process are at low numbers, fluctuations can dominate. This situation has been studied in another context in [36] within a linear noise approximation. This approximation proved useful in quantifying the strong anti-correlation between species, but cannot be applied to study more refined issues such as the population diversity.

To gain insight into the nature of these fluctuation, we show in figure 3A typical temporal evolution of the sRNA and target copy number in our model, as obtained from Monte-Carlo simulations [34]. While the average mRNA number $\langle m \rangle$ (as obtained by numerical solution of the master equation and compared with the simulation) is not very different from the one predicted by the mass-action model (compare the two horizontal lines), these averages do not capture the dynamical behavior of the mRNA copy number. In fact, simulations suggest

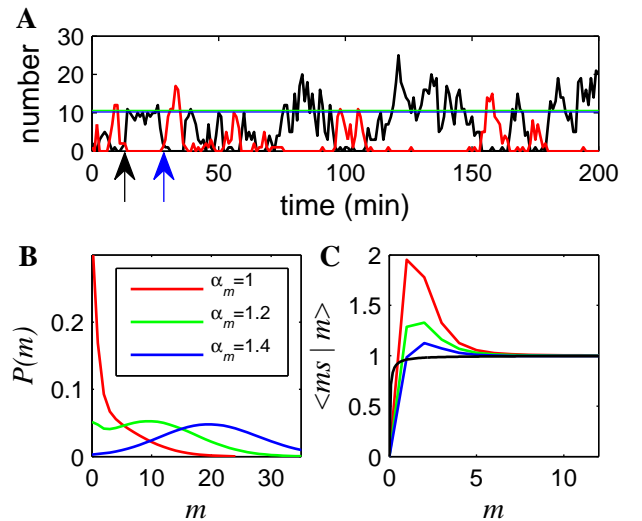


Fig. 3. Diversity in gene expression near the transition. **A.** Temporal evolution of mRNA (black) and sRNA (red) expression levels in a Monte-Carlo simulation. Here $\alpha_m = 1.2, \alpha_s = 1, \beta_m = \beta_s = 0.02, k = 0.5$. Horizontal lines: the steady-state mean number of mRNA molecules, as predicted by the mass action model (blue) or by integrating the master equation (green). The two are barely distinguishable. **B.** Probability distribution for the mRNA number at steady-state, calculated by direct numerical integration of the master equation. The coexistence of two populations at the onset of gene expression appears here as two modes in the distribution function (green). **C.** Correlations between the numbers of mRNA and sRNA molecules amplify the coupled degradation at low but finite m . Color scheme is the same as in B. For comparison, black line is the coupled degradation strength as predicted by the mass-action model.

¹Using a linear noise approximation [35] one can show that for $\alpha_m \ll \alpha_s$ the correlation coefficient $C_{m,s}/(\sigma_m \sigma_s) \simeq \sqrt{w^{-1} \alpha_m / \alpha_s^2}$.

that the cell goes through alternating periods of complete mRNA suppression (with high level of sRNA) and mRNA expression (with the sRNA suppressed). Starting from a state where both RNA species are suppressed (marked by a black arrow in figure 3A), a small fluctuation makes the mRNA the majority species. In this state, most mRNAs decay through the sRNA-independent degradation pathway. At the same time, the developing mRNA pool keeps the sRNA level suppressed. After a while, fluctuations drive the mRNA pool back to comparable level to that of the sRNA (blue arrow in figure 3A). Now, similar fluctuation can drive the cell to another round of abundant mRNA, or – as it happens in this particular realization – to a round of abundant sRNA, accompanied by mRNA suppression.

These simulations suggest an appealing scenario, where the transition from the silenced regime ($\alpha_m \ll \alpha_s$) to the active regime ($\alpha_m \gg \alpha_s$) occurs via separation of the probability $P(m)$ into two modes: One mode is a reminiscent of the silenced state, while the other is the bud of the active state. This scenario is depicted in figure 3B, where $P(m)$ – as obtained from numerical solution of the master equation [1] – is followed through the transition. Just above the transition one indeed finds a regime where the population of cells is divided into two sub-populations (see the doubly peaked distribution in Fig.3b), which do or do not express the target gene.

Further understanding of the mechanism behind this behavior is provided by the following picture. The steady-state dynamics of the mRNA molecule number, $m(t)$, can be described as a random walker on the positive axis. This number is increased by 1 (“hop” to the right) due to transcription event, with rate α_m . It is decreased by 1 (“hop” to the left) whenever an mRNA is degraded by a ribonuclease, independent of the small RNA interaction, a process that occurs with rate $\beta_m m(t)$. Finally, m also decreases by 1 whenever an sRNA binds to an mRNA, a process that occurs with rate $km(t)s(t)$. At steady state, it is tempting to use the mass-action equations and let $km(t)s(t) = \alpha_s [m/(m + \beta_s/k)] < \alpha_s$. Under this description the random walker is biased to the right, and is only kept bound by the independent degradation. However, at small and intermediate values of m , correlations between the small RNA and its target cannot be neglected. A better estimate for the left-hopping rate due to coupled degradation would therefore be $k\langle ms|m \rangle$. This function is plotted in figure 3C (colored lines), and compared with the mass-action result (black line). While the two estimates approach α_s at high values of m , taking correlations into account significantly increases the coupled degradation rate at low – but finite – values of m . Thus, (anti-) correlations between the sRNA and its target form an effective potential

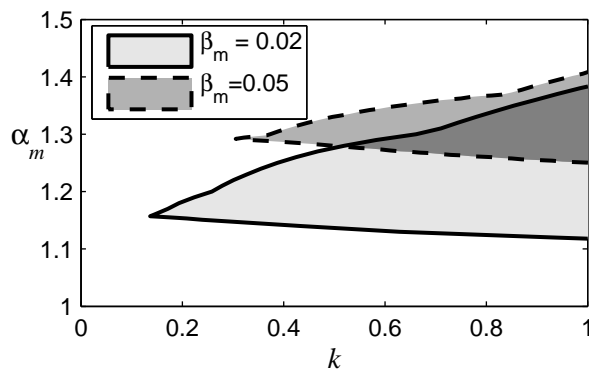


Fig. 4. Phase Diagram. The domains of bimodality (filled areas) where determined through exact numerical integration of the master equation, with $\alpha_s = 1/\text{min}$ and $\beta_s = 0.02/\text{min}$. The two different domains shown in the figure correspond to two different levels of mRNA stability.

barrier, which makes it hard for $m(t)$ to approach or escape the origin. This potential barrier shapes the bimodal distribution (Supporting figure 1).

To study the dependence of this bimodal behavior on the model parameter, we draw in figure 4 the ‘phase diagram’ of the model. Bimodality takes place just above the threshold $\alpha_m \gtrsim \alpha_s$. Strong interaction between the sRNA and its target (large k) increases the anti-correlations between that two species, and extends the range of α_m at which bimodality occurs. In addition, slower degradation rate of the target makes this range of α_m values larger, and also allows for bimodality at smaller k .

Noise amplification at the onset of gene expression – Extrinsic noise. So far, we considered noise that is attributed to the discrete nature of gene synthesis (“intrinsic noise”), in particular the transcription events. Our analysis suggests that this source of noise suffices to generate a broad distribution of gene products, manifested as bi-modal distribution. However, the cell offers many other sources of noise in gene expression (“extrinsic noise”), such as fluctuations in the number of ribosomes or polymerases. For sRNA regulation, mediator proteins, such as Hfq, can also vary in number from cell to cell. Another source of noise is the number of sRNA or target genes. This number fluctuates due to chromosome duplication (for chromosome-encoded genes) or due to plasmid copy number fluctuations (for plasmid-borne genes). Its possible importance stems from the fact that in both cases the mean number may be very small (on the order of 1 in the first case and 10 in the latter).

The threshold-linear response has the capacity to amplify this noise. Consider, for example, fluctuations in the transcription rate of the target gene. These fluctuations can be due to copy number fluctuation in a transcriptional regulator or the coding gene. It is then possible that during a certain time, the effective transcription rate of the target would be lower than that of the sRNA, resulting in effective gene silencing, while during other periods the target transcription rate would be high enough to cross the threshold for gene expression.

We performed a variety of Monte-Carlo simulations to consider this effect. To obtain the results shown in Fig. 5, we allowed for the target gene copy number g to increase (decrease) with rates $1 \pm \delta_g |g - g_0|$. Motivated by the estimated copy number of the plasmid we use in our experiments (described in the next section), we take $g_0 = 50$. Our results show no qualitative difference with those that lack extrinsic noise

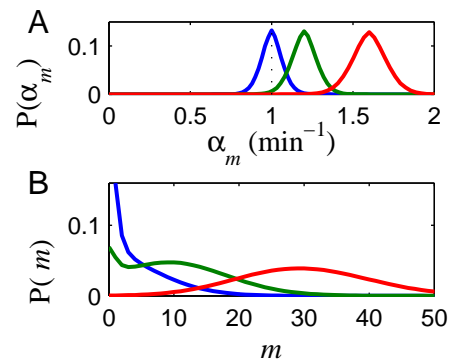


Fig. 5. Effect of noise in gene copy number. Monte-Carlo simulations were carried out as in Fig. 3A, with additional fluctuation in gene copy number g as described in the text. The effective transcription rate is $\alpha_m = \alpha_0 g / g_0$, with $g_0 = 50$ and $\delta_g = 10 \text{ min}^{-1}$. **A.** Probability distribution for gene copy number at steady-state for $\alpha_0 = 1, 1.2, 1.5 \text{ min}^{-1}$, represented by the blue, green, and red lines respectively. **B.** Corresponding distribution of mRNA number. Bi-stability is only observed for $\langle \alpha_m \rangle \gtrsim \alpha_s$.

(Fig. 3B). In particular, bi-modality is only observed when the mean transcription rate of the target is slightly above that of the sRNA. In other sets simulations (data not shown) we took broader (uni-modal) distributions, or considered fluctuations in other parameters (such as sRNA transcription level, or mRNA degradation). None of these simulations showed bi-modality when average parameters were outside the region of bi-modality expected from transcriptional noise alone (Fig. 4). These numerical results support the above prediction, that a broad (bi-modal) distribution should be expected at the regime where the target transcription rate is somewhat above that of the sRNA.

Population diversity at the threshold – Experiment. The model presented above is consistent with available data about RyhB, an endogenous small RNA in *E. coli*, involved in regulation of iron homeostasis and oxidative stress [29, 31, 24]. Expression of the *ryhB* gene is regulated by Fur, the ferric uptake regulator, which inhibits gene expression in response to high cellular levels of free Fe^{2+} ions. At least 18 genes are directly repressed by RyhB [28, 29, 30]. Transcription of these genes is independent from that of RyhB, and is controlled by other signals. For some targets of RyhB it has been shown that binding of the small RNA to its mRNA target leads to degradation of both molecules through an RNase E dependent process [29, 37, 31].

We constructed a synthetic pair of small RNA and its target to test the results of our model. The *ryhB* gene was cloned from the *E. coli* chromosome onto a medium-copy plasmid (p15A ori). Expression of *ryhB* was driven by the $P_{Ltet-O1}$ promoter [38], repressed by TetR which is constitutively expressed (see Supporting Table 1). This allowed us to control the synthesis rate of the small RNA (α_s) via the addition of an inducer, anhydrotetracycline (aTc). As a target, we fused the control region of *sodB*, one of the strongest known targets of RyhB, to a *gfp* gene. The target gene was encoded on another medium copy plasmid (colE1 ori; see Supporting Table 1), and was driven by the $P_{Llac-O1}$ promoter. Thus the transcription rate of the target (α_m) can be controlled by the inducer Isopropyl beta-D-1-thiogalactopyranoside (IPTG) [38]. The mean expression of the target protein was estimated by measuring GFP fluorescence per OD₆₀₀ from a culture grown in 48-well microplate. As expected, fluorescence level increases with IPTG concentration, and decreases when aTc is added to the medium (figure 6A, left panel). With no aTc added to the media, expression of RyhB was fully repressed, and expression of the target GFP was the same as in a strain with no RyhB source. Adding 1 ng/ml allowed for an intermediate level of RyhB expression, and resulted in strong repression of GFP expression at low IPTG concentrations (below 0.2 mM). Finally, aTc at concentration 5 ng/ml fully repressed GFP expression at all IPTG concentrations (data not shown).

Addressing fluctuations refined our predictions for the *threshold-linear* response of a target gene to activity at its promoter. For $\alpha_m \ll \alpha_s$, gene expression is suppressed in all cells. At the onset of gene expression, $\alpha_m \gtrsim \alpha_s$, a snapshot of a cell population at a given point in time is predicted to capture a sub-population of cells expressing the target gene, and another sub-population which does not. Finally, above the threshold, $\alpha_m \gg \alpha_s$, almost full expression is expected in all cells.

To test this prediction, we consider the case where expression of RyhB is set by adding 1 ng/ml aTc to the media (filled circles in figure 6A). With no IPTG added to the media, we interpret the low GFP fluorescence as a result of very small α_m , much smaller than α_s . Also, IPTG concentration above 0.5mM corresponds to $\alpha_m \gg \alpha_s$. The threshold is expected around $[\text{IPTG}] \simeq 0.2$ mM. It is around this concentration of IPTG that we expect to find two sub-populations of cells.

We collected single-cell fluorescence data using flow-cytometry. Typical histograms are plotted in figure 6A. In the absence of IPTG, all cells show only background level of fluorescence, while at high level (1 mM) of IPTG virtually all cells show high fluorescence level. However, at intermediate levels of IPTG (3rd and 4th columns of figure. 6), the histogram is bi-modal. One mode corresponds to cells whose fluorescence is at or near the background level; the other mode corresponds to cells expressing some amount of GFP.

In order to verify that these effects depend on the interaction between the small RNA and the reporter target, we performed a single mutation in the 5'-control region of our target. This mutation, C \rightarrow G, is located one nucleotide away from the start codon, in the core recognition sequence of RyhB and the SodB mRNA [31] (see Supporting Table 2 for sequences). Plasmids harboring this mutated target, *crsodB*-gfp*, were transformed into Δ *ryhB* cells, with either a *ryhB*-harboring plasmid (strain ZZS2*3) or a control plasmid (ZZS2*1). Bulk fluorescence from the two strains was indistinguishable even with saturating concentrations of aTc (10 ng/ml), which allow for full expression of the small RNA (figure 6B, left column). These data suggest that interaction between RyhB and the target has been completely removed.

Figure 6B also shows histograms of *crsodB*-gfp* expression. These histograms suggest a homogeneous population of cells at all IPTG levels. As promoter activity is turned on, the population mean and variance of protein-concentrations is smoothly increased. Similar results were obtained by removing the control-region altogether (supporting figure 2). Thus, interaction between the target and RyhB is required for the bi-modality at the onset of gene expression. For a comparison of our full data sets of the two targets, see supporting figure 3.

RyhB is an endogenous small RNA in *E. coli* that interacts with many targets. Those interactions affect the level of RyhB in a complex manner. It is therefore possible that the observed RyhB-dependent bi-modality is an indirect effect of these interactions. To address this possibility, we mutated a single nucleotide in *ryhB*. This mutation was made at the position complementary to the mutation in *crsodB* (see Supporting Table 2 for sequences). Plasmids harboring this small RNA gene, *ryhB**, were transformed into Δ *ryhB* cells along with plasmids coding for either the target with the wild-type control region, *crsodB-gfp* (strain ZZS23*), or the one with the mutated control-region, *crsodB*-gfp* (ZZS2*3*). Measurements of bulk fluorescence indicate that RyhB* interacts with *crsodB*-gfp* (figure 6C, left column), in the same threshold-linear manner that characterizes the interaction between RyhB and *crsodB-gfp*. The effect of aTc was also similar, with 2 ng/ml resulting in strong repression only at low IPTG concentrations (below 0.2 mM), and 5 ng/ml resulting in full repression. On the other hand, RyhB* does not affect the fluorescence of *crsodB-gfp*. Using RT-PCR we also verified that RyhB* expression has also no effect on the mRNA levels of chromosomal *sodB* and *fumA*, two endogenous targets of RyhB [28] (data not shown). Taken together, these results suggest that RyhB* does not interact with a wild-type control regions targeted by RyhB.

Single-cell data for the strain expressing the mutated target gene *crsodB*-gfp* and its cognate sRNA *ryhB** are presented in figure 6C. As for the wild-type small RNA and control region, we find again homogeneous populations of cells in the low and high ends of gene expression, and inhomogeneous populations at intermediate levels. To test our predictions further, we make use of the fact that RyhB* is expressed from the $P_{Ltet-O1}$ promoter, inducible by aTc. Thus, keeping the IPTG level constant at some intermediate level (i.e., constant α_m), we can tune α_s by changing the concentration of aTc in

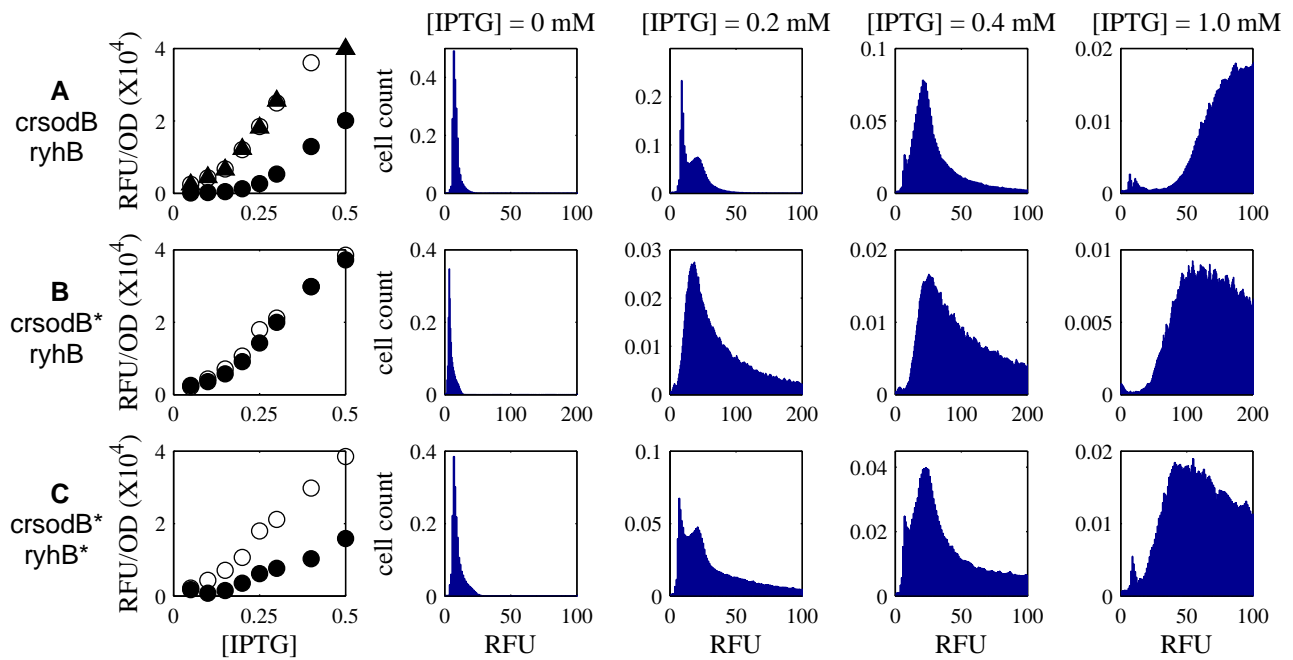


Fig. 6. Response of a reporter target to regulation by a small RNA, for the strains **A.** ZS23, **B.** ZS2*3, **C.** ZS2*3* (see labels for phenotype, and Supporting Table 1 for details). The leftmost column shows average fluorescence (in RFU/OD) as measured from a liquid culture in a microplate reader (closed symbols). The level transcription of target, expressed from the $P_{Llac-O1}$ promoter [38] is induced by IPTG, while the level of RyhB expression from the $P_{Ltet-O1}$ is controlled by aTc (A. triangles - no aTc, circles - 1 ng/ml; B. 10 ng/ml; C. 2 ng/ml; see text for details). For comparison, open circles show an identical measurement for a corresponding strain with no RyhB source (A. ZS21, B. and C. ZS2*1). Other columns show histograms of single-cell fluorescence obtained by flow-cytometry. Bimodality is observed near the threshold ([IPTG]=0.2mM) for ryhB/crsodB and ryhB*/crsodB*, but not for the control ryhB/crsodB*.

the media. Typical histograms are shown in supporting figure 4. As the small RNA transcription level is increased, the sub-population of cells which show background fluorescence increases on the expense of the other sub-population, in accordance with our prediction.

Discussion

Small regulatory RNA control many stress response pathways in bacteria, such as SOS response to DNA damage [39], oxidative stress [40] and quorum sensing [41]. Regulation of a stress response pathways by a small RNA allows for a strong suppression in the absence of stress. Here we suggest that small RNA regulation have also noise properties which are favorable for regulating stress pathways. By destabilizing their target mRNAs, small RNAs regulation results in decreased burstiness in protein production. This reduces the variance in target gene expression, and can prevent accidental switch-on of stress response programs. Coupled with other circuit motifs, a reduced variance infers higher homogeneity within the cell population [4, 6]. For example, small RNA repressing a gene which is part of a bi-stable genetic switch would reduce the rate of spontaneous switching from the ‘off’ to the ‘on’ state. Such a mechanism has also been suggested for microRNAs acting in differentiation pathways in eukaryotes [42]. These predictions call for an experimental verification, which would require single-cell imaging technology, in the spirit of [17, 16].

Recently two small RNAs have been identified as inhibitors for SOS-induced toxins in *E. coli* [43, 44]. The two toxin genes, *tisAB* and *symE*, are strongly repressed under normal conditions by LexA. Under SOS conditions, however, both genes are strongly induced. In both cases, small RNAs serve as anti-toxins. The small RNA IstR, which is constitutively transcribed, binds to *tisAB* mRNA, entailing translational inhibition and RNase III-dependent cleavage

of both molecules [43, 45]. Similarly, translation of *symE* mRNA is repressed by the abundant SymR sRNA. Our study suggests that, in both cases, the role of the small RNA is likely to serve as a safeguard mechanism to prevent mRNA translation. In controlling toxin synthesis, we can easily rationalize a functional role for the sRNA in attenuating fluctuations.

A class of small RNAs in bacteria act by binding to their target mRNA, thereby promoting cleavage of both RNA molecules. This non-catalytic mode of action provides a threshold response, which has also been suggested as a favorable property in regulating stress response [41, 32]. This mechanism however allows for circumstances where both RNA molecules are synthesized at similar rates. In such case, mRNA and sRNA would be paired and removed from the cell, leaving behind a small number of molecules of both species. This situation is dominated by fluctuations. Our model suggests that if both RNA species are relatively stable in the absence of the other species, a cell can undergo through alternating periods of substantial and negligible target expression, even in the absence of any additional feedback. This effect can be amplified by extrinsic sources of noise. Thus, at the threshold of gene expression, small RNA mediated regulation can in fact be used to generate, rather than suppress, diversity in a cell population.

Consistent with this prediction, we presented experimental data for a synthetic pair of reporter target and its cognate small RNA. It is yet to be shown whether bi-modality also occurs for endogenous small RNA and their target, and whether this feature is used by bacteria for physiological function. However, our results do suggest that this surprising feature of sRNA-mediated regulation may be used in synthetic genetic circuits to enhance nonlinear effects crucial to applications such as bistability and oscillation [46, 47, 48].

Materials and Methods

Strains & plasmids. All experiments were performed with BW-RI cells derived from *E. coli* K-12 BW25113 [49], with the transfer of the *spr-lacI-tetR* cassette from DH5 α -ZI cells [38] by phage P1 transduction. This cassette provides the constitutive expression of *lacI* and *tetR* [38]. Chromosomal *ryhB* was deleted from BW-RI using the method of Dotsenko & Wanner [49]. These strains were then transformed by the following ‘target’ and ‘source’ plasmids. All strains and plasmids used are summarized in Supporting Table 1.

Medium, growth, measurements. BW-RI strains each containing the appropriate target and/or source plasmids (as described in text) were grown in M63 minimal media with standard concentrations of the appropriate antibiotics. The overnight cultures were diluted into fresh M63 media ($OD_{600} \simeq 0.002$) containing the appropriate antibiotics as well as varying amounts of the inducers (aTc, IPTG) in the wells of 12-well plates. The plates were incubated with shaking at 37°C and taken for OD_{600} and fluorescence measurements every hour for up to 8 hours (until a final OD_{600} of 0.2 - 0.3) using a TECAN

Genios-Pro plate reader. Each measurement was repeated 3 times and the data was analyzed as described in [32].

To obtain single-cell data, 1.7 ml of each culture was spun down and resuspended in 1 ml PBS. Expression data were collected using a Becton-Dickinson FACSCalibur flow cytometer with a 488-nm argon excitation laser and a 515- to 545-nm emission filter (FL1) at a low flow rate. Photomultiplier tube (PMT) voltage was set to 950V, and a linear amplifier was set at 9.5X. Forward scatter and fluorescence values were collected for 50,000 cells.

EL and TH are grateful to Jeff Hasty and Natalie Ostroff for advice and hospitality at the UCSD Bioengineering core facility. MH and YWH are grateful to Zhongcan Ouyang, Zhirong Sun, and Weimou Zheng for advices and supervision. The experimental work carried out in Beijing was supported by the National Natural Science Foundation, China (Grant No. 10428409 and 90403140) to HS and TH. TH further acknowledges the Cheung Kong Scholars Programme for supporting visits that made this collaboration with China possible. The work at UCSD was supported by the US NSF through Grant No. MCB-0417721, and through the PFC-sponsored Center for Theoretical Biological Physics (Grant No. PHY-0216576 and PHY-0225630).

1. Guptasarma, P. (1995) *Bioessays* **17**, 987–97.
2. Raser, J. M. & O’Shea, E. K. (2005) *Science* **309**, 2010–3.
3. Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005) *Nat Rev Genet* **6**, 451–64.
4. Kussell, E. & Leibler, S. (2005) *Science* **309**, 2075–8.
5. Kearns, D. B. & Losick, R. (2005) *Genes Dev* **19**, 3083–94.
6. Suel, G. M., Garcia-Ojalvo, J., Liberman, L. M., & Elowitz, M. B. (2006) *Nature* **440**, 545–50.
7. Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002) *Science* **297**, 1183–6.
8. Swain, P., Elowitz, M., & Siggia, E. (2002) *Proc Natl Acad Sci USA* **99**, 12795–800.
9. McAdams, H. H. & Arkin, A. (1997) *Proc Natl Acad Sci U S A* **94**, 814–9.
10. Thattai, M. & van Oudenaarden, A. (2001) *Proc Natl Acad Sci U S A* **98**, 8614–9.
11. Paulsson, J. (2004) *Nature* **427**, 415–8.
12. Volfson, D., Marciniak, J., Blake, W. J., Ostroff, N., Tsimring, L. S., & Hasty, J. (2006) *Nature* **439**, 861–4.
13. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & van Oudenaarden, A. (2002) *Nat Genet* **31**, 69–73.
14. Becskei, A., Kaufmann, B. B., & van Oudenaarden, A. (2005) *Nat Genet* **37**, 937–44.
15. Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005) *Cell* **123**, 1025–36.
16. Cai, L., Friedman, N., & Xie, X. S. (2006) *Nature* **440**, 358–62.
17. Yu, J., Xiao, J., Ren, X., Lao, K., & Xie, X. S. (2006) *Science* **311**, 1600–3.
18. Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O’Shea, E., Pilpel, Y., & Barkai, N. (2006) *Nat Genet* **38**, 636–43.
19. Swain, P. S. (2004) *J Mol Biol* **344**, 965–76.
20. Walczak, A. M., Onuchic, J. N., & Wolynes, P. G. (2005) *Proc Natl Acad Sci U S A* **102**, 18926–31.
21. Pedraza, J. M. & van Oudenaarden, A. (2005) *Science* **307**, 1965–9.
22. Ambros, V. (2004) *Nature* **431**, 350–355.
23. Bartel, D. P. (2004) *Cell* **116**, 281–297.
24. Gottesman, S. (2004) *Annu. Rev. Microbiol.* **58**, 303–328.
25. Gottesman, S. (2005) *Trends Genet.* **21**, 399–404.
26. Lease, R. A., Smith, D., McDonough, K., & Belfort, M. (2004) *J Bacteriol* **186**, 6179–85.
27. Opdyke, J. A., Kang, J. G., & Storz, G. (2004) *J Bacteriol* **186**, 6698–705.
28. Masse, E. & Gottesman, S. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 4620–4625.
29. Masse, E., Escorcía, F., & Gottesman, S. (2003) *Genes. Dev.* **17**, 2374–2383.
30. Masse, E., Vanderpool, C., & Gottesman, S. (2005) *J Bacteriol.* **187**, 6962–6971.
31. Geissmann, T. & Touati, D. (2004) *EMBO J.* **23**, 396–405.
32. Levine, E., Zhang, Z., Kuhlman, T., & Hwa, T. (2007) *PLoS Biol.* **5**, e229.
33. Wagner, E. G., Altuvia, S., & Romby, P. (2002) *Adv Genet* **46**, 361–98.
34. Gillespie, D. (1977) *J. Phys. Chem* **81**, 2340–61.
35. Kampen, N. V. (2001) *Stochastic Processes in Physics and Chemistry.* (North-Holland, Amsterdam), 2 edition.
36. Elf, J., Paulsson, J., Berg, O. G., & Ehrenberg, M. (2003) *Biophys J* **84**, 154–70.
37. Moll, I., Afonyushkin, T., Vytvytska, O., Kabardin, V., & Blasi, U. (2003) *RNA* **9**, 1308–1314.
38. Lutz, R. & Bujard, H. (1997) *Nucleic Acids Res* **25**, 1203–10.
39. Vogel, J., Argaman, L., Wagner, E. G., & Altuvia, S. (2004) *Curr Biol* **14**, 2271–6.
40. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C., & Storz, G. (2002) *Mol Cell* **9**, 11–22.
41. Lenz, D., Mok, K., Lilley, B., Kulkarni, R., Wingreen, N., & Bassler, B. (2004) *Cell.* **118**, 69–82.
42. Hornstein, E. & Shomron, N. (2006) *Nat. Gen. Supp.* **38**, S20.
43. Vogel, J., Argaman, L., Wagner, E. G., & Altuvia, S. (2004) *Curr Biol* **14**, 2271–6.
44. Kawano, M., Aravind, L., & Storz, G. (2007) *Mol Microbiol* **64**, 738–54.
45. Darfeuille, F., Unoson, C., Vogel, J., & Wagner, E. G. (2007) *Mol Cell* **26**, 381–92.
46. Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000) *Nature* **403**, 339–42.
47. Elowitz, M. B. & Leibler, S. (2000) *Nature* **403**, 335–8.
48. Atkinson, M. R., Savageau, M. A., Myers, J. T., & Ninfa, A. J. (2003) *Cell* **113**, 597–607.
49. Datsenko, K. A. & Wanner, B. L. (2000) *Proc Natl Acad Sci U S A* **97**, 6640–5.