

The unreasonable effectiveness of equilibrium gene regulation through the cell cycle

Jose M. G. Vilar^{1,2,*} and Leonor Saiz^{3,4,5,*}

¹ Biofisika Institute (CSIC, UPV/EHU), University of the Basque Country (UPV/EHU), P.O. Box 644, 48080 Bilbao, Spain

² IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

³ Department of Biomedical Engineering, University of California, 451 E. Health Sciences Drive, Davis, CA 95616, USA

⁴ Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

⁵ Center for Systems Biology Dresden, Dresden, Germany

* Correspondence: lsaiz@ucdavis.edu; j.vilar@ikerbasque.org

Abstract

Systems like the prototypical *lac* operon can reliably hold the repression of transcription upon DNA replication across cell cycles with just ten repressor molecules per cell and, in addition, behave as if they were at equilibrium. The origin of this type of phenomena is still an unresolved question of major implications. Here, we develop a general theory to analyze strong perturbations in quasi-equilibrium systems and use it to quantify the effects of DNA replication in gene regulation. We find a scaling law that connects actual transcription with its predicted equilibrium values in terms of a single kinetic parameter. We show that even the simplest, exceptionally reliable natural system functions beyond the physical limits of naïve regulation through compensatory mechanisms that suppress nonequilibrium effects. We validate the approach with both *in vivo* cell-population and single-cell characterization of the *lac* operon. Analyses of synthetic systems without adjuvant activators, such as the cAMP receptor protein (CRP), do not show this reliability. Our results provide a rationale for the function of CRP, beyond just being a tunable activator, as a mitigator of cell cycle perturbations.

Keywords:

Systems Biology; Gene regulation; DNA replication effects; Nonequilibrium binding; Transcription

Introduction

Gene regulation depends to a large extent on equilibrium binding interactions (Ptashne and Gann, 2002). As such, equilibrium approaches, although do not consider explicitly many essential processes, have performed exceedingly well in both qualitative and even quantitatively understanding the major features of gene expression. They have been successfully widespread in modeling gene regulation since early pioneering work (Ackers et al., 1982; von Hippel et al., 1974) and have been applied to both prokaryotes (Guharajan et al., 2021; Kinney et al., 2010; Kuhlman et al., 2007; Lagator et al., 2022; Vilar and Leibler, 2003; Vilar and Saiz, 2005) and eukaryotes (Bashor et al., 2019; Carey et al., 1990; Gertz et al., 2009; Ptashne and Gann, 2002; Vilar and Saiz, 2011). The underlying assumption is that the equilibrium binding patterns, such as those observed in biochemical *in vitro* experiments, can be extrapolated to a living growing and multiplying cell. These patterns determine the transcriptional state of the system and dictate the amplitude of the outcomes of downstream nonequilibrium processes.

There are multiple cases in which it is required to move towards nonequilibrium kinetic approaches to obtain sensible results (Coulon et al., 2013; Doudna et al., 2017; Wong and Gunawardena, 2020). These include, for instance, cyclical processes based on both activating and repressive epigenetic mechanisms (Metivier et al., 2003), alternating transcription factor (TF) phosphorylation patterns (Kim and O'Shea, 2008), or even the long-term coexistence of induction states in the *lac* operon (Novick and Weiner, 1957; Ozbudak et al., 2004; Vilar et al., 2003). But, on the other hand, there are many complex situations with underlying energy expenditure, such as the patterning of drosophila embryos (Sayal et al., 2016; Segal et al., 2008), the cooperative binding of TFs as a result of nucleosome displacement (Mirny, 2010), and transcription regulation by biomolecular condensates (Hnisz et al., 2017), that have successfully been tackled with equilibrium-based approaches.

The effectiveness of equilibrium approaches has been quantified extensively in the *E. coli lac* operon and its synthetic derivatives. They have been tested across cell-population experiments

for different strains, mutations, and synthetic constructs over a dynamic range in transcription rate changes larger than 10,000 (Garcia and Phillips, 2011; Kuhlman et al., 2007; Oehler et al., 2006; Saiz and Vilar, 2008; Vilar, 2010; Vilar and Leibler, 2003). Over this whole range, the typical discrepancies from the equilibrium predictions are, on average, just a factor 1.7 above or below the actual values (Vilar and Saiz, 2013a). Essentially this magnitude, a factor of 1.4, has also been observed in direct single-cell measurements in fast-growing microcolonies, which has been reasoned to be likely due to nonequilibrium effects in transcription initiation (Hammar et al., 2014).

In all cases, an unavoidable major far-from-equilibrium event is DNA replication, in which the promoter and DNA regulatory elements disappear momentarily upon the passage of the replication fork and reappear in duplicate (Fraser and Nurse, 1978; Voichek et al., 2016). During this time interval, there are no regulatory sequences to bind nor promoter to transcribe. After the promoter reconstitution, TFs and other regulatory molecules should find their duplicated binding sites fast enough for the equilibrium approach to hold (Fig. 1). Experimental evidence is being accumulated that indicates that the underlying processes are not sufficiently fast, even in the simplest systems (Elf et al., 2007; Hammar et al., 2014; Larson et al., 2011; Wheat et al., 2020).

This paradoxical evidence is present prominently in the high efficiency observed in systems with just a few copies of the regulatory molecules, as in the *lac* operon with just ~10 repressors per cell, and the plethora of experimental conditions which have been captured in detail with equilibrium models. The replication fork speed at about 1 kbp/s (Elshenawy et al., 2015; Pham et al., 2013) and size of 1.2kbp, as measured by the Okazaki fragment length (Balakrishnan and Bambara, 2013), indicate that in about 1 second the promoter and the regulatory elements are duplicated. Yet, the time for the repressors to find the operator site has been visualized to be about 30 seconds (Hammar et al., 2014), which would result in the production of ~10 transcripts per promoter after replication at the usual transcription rate of 1 transcript every 3 seconds (Kennell and Riezman, 1977).

If the previous considerations were all true, repression would be 10 times less efficient than observed for the natural *lac* operon, which on average leads to fewer than 1 transcript per hour (Kennell and Riezman, 1977). Under these circumstances, equilibrium approaches would not work. Namely, without a much faster dynamics than observed, the kinetic limitations of the observed rates would prevent a very strong binding to work as effectively as needed before it is reset by the replication fork.

Therefore, the observed effectiveness of equilibrium approaches seems to underlie the experimentally observed efficiency. Strikingly, this paradoxical behavior is not just implied but has also been observed directly at the single-cell level, with transcription behaving as if the cell cycle were not present (Chang et al., 2022). These data showed not only that the completely repressed state in the natural, wild type (WT) system can be propagated across multiple generations without any transcriptional leakiness but also that the observed statistics of transcriptionally silent intervals and burst size statistics is consistent with the equilibrium dynamics quantified across independent experiments (Chang et al., 2022; Hammar et al., 2014).

Here, we develop a general theory for analyzing the impact of strong nonequilibrium perturbations on quasi-equilibrium systems and apply it to quantify the recurrent effects of cell cycle events in gene regulation systems. The results show that the cell cycle imposes strong limitations to the TF binding patterns that can be achieved over time and, at the same time, also provide an avenue, with short transient deactivation of transcription at the promoter after replication, that would allow the system to bypass these limitations. Interestingly, the *lac* operon is under both negative and positive control. Therefore, it is appealing to consider the potential role of CRP as the mechanism responsible for the delayed activation of the promoter. With values for the *in vivo* rates that are consistent across experiments, we show that the observed efficiency in synthetic systems that do not have active CRP falls at predictable nonequilibrium levels limited by the rate of dislodging of the TFs from DNA by the replication fork.

Results

The development of the theory considers a state-based description of the system (Saiz and Vilar, 2006; Vilar and Saiz, 2013b) in which the fast-relaxing modes reach rapid equilibration between them (van Kampen, 2007) and collectively follow the slower dynamics of nonequilibrium states (Haken, 1983).

We consider the system characterized by an ensemble of N states, $S = (0, 1, \dots, N - 1)$. The reference state is indexed as $i = 0$, which we use also as the nonequilibrium state. The other states $i > 0$, with $i \in S$, are assumed to reach fast equilibration between them. The probability for the system to be in a state is denoted by P_i and the corresponding equilibrium value, by P_i^{eq} .

Quasi-equilibrium transcription

For the states $i > 0$ in quasi-equilibrium between them, the detailed balance principle (Beard and Qian, 2008) implies $P_i/P_j = P_i^{eq}/P_j^{eq}$, $\forall i, j > 0$, which together with the probability normalization expressed as $\sum_{i>0} P_i = 1 - P_0$ leads to general expression

$$P_i = \frac{1 - P_0}{1 - P_0^{eq}} P_i^{eq}, \quad \forall i > 0. \quad (1)$$

This expression is approach agnostic; namely, P_i^{eq} can generally be expressed in terms of equivalent TF and DNA properties, such as concentrations and equilibrium dissociation constants, concentrations and standard free energies, and number of molecules and free energy changes (Saiz, 2012).

The model of transcription considers that transcription at each state i takes place at a rate γ_i so that the overall transcription rate,

$$\Gamma = \sum_i \gamma_i P_i, \quad (2)$$

is expressed as the average over all the states. The quasi-equilibrium assumption for the states $i > 0$ implies $\Gamma = \gamma_0 P_0 + \frac{1 - P_0}{1 - P_0^{eq}} \sum_{i>0} \gamma_i P_i^{eq}$, which can be rewritten as

$$\Gamma = \gamma_0 P_0 + \frac{1 - P_0}{1 - P_0^{eq}} (\Gamma_{eq} - \gamma_0 P_0^{eq}), \quad (3)$$

where $\Gamma_{eq} = \sum_i \gamma_i P_i^{eq}$ is the usual equilibrium transcription rate. This general expression we have obtained is extremely powerful as it provides an explicit quantification of the nonequilibrium transcription rate from its overall equilibrium value and the departure from equilibrium. Namely, the details of the intra quasi-equilibrium dynamics are not relevant provided that the overall effect is known.

Nonequilibrium cell-cycle effects

We generally consider that the state $i = 0$ is out of equilibrium following the dynamics $dP_0/dt = -k_{in} P_0 + k_{out} (1 - P_0)$, indicating that the system can switch in and out of the state $i = 0$ with rates k_{in} and k_{out} , respectively. In terms of a repressive system like the *lac* operon, this nonequilibrium state would be the promoter with no repressor bound to the operators, k_{in} would be the association rate of the repressor to any of the operators, and k_{out} would be the rate at which the operators end with no repressor. This equation can be rewritten as

$$\frac{dP_0}{dt} = \tilde{k}_{in}(P_0^{eq} - P_0), \quad \text{with } \tilde{k}_{in} = \frac{k_{in}}{1 - P_0^{eq}}, \quad (4)$$

Noting that $\frac{dP_0^{eq}}{dt} = 0$ at equilibrium generally implies $k_{out} = k_{in} \frac{P_0^{eq}}{1 - P_0^{eq}}$.

To simplify the modeling of the cell cycle effects, we track just one of the daughter strands. We consider that the replication fork crosses the promoter at time 0 and that the effects are to clear the system from transcriptional regulators. In mathematical terms, this implies that at time zero $P_0 = 1$ and $P_i = 0, \forall i > 0$, which results in

$$P_0 = e^{-\tilde{k}_{in}t} (1 - P_0^{eq}) + P_0^{eq}. \quad (5)$$

This dynamic characterization implies that the system will reach equilibrium on a time scale of $1/\tilde{k}_{in}$. The dynamic evolution of the probabilities, $P_i, \forall i > 0$, for the other states of the system follows straightforwardly from Eqs. (1) and (5).

To compute the average transcription $\hat{\Gamma} = \frac{1}{T} \int_{\Delta t}^T \Gamma dt$ over a cell cycle of length T , we generally consider that transcription can resume after a time Δt from the replication fork crossing the promoter, which after substitution of Eq. (5) in Eq. (3) leads to

$$\hat{\Gamma} = (\gamma_0 - \Gamma_{eq})\alpha + \Gamma_{eq} \left(1 - \frac{\Delta t}{T}\right), \quad \text{with } \alpha = \frac{e^{-\tilde{k}_{in}\Delta t} (1 - e^{-\tilde{k}_{in}(T-\Delta t)})}{T\tilde{k}_{in}}. \quad (6)$$

Repression of transcription

The repression level, denoted by R , is the inverse normalized transcription and is formally defined as the maximal transcription, which corresponds to the fully unrepressed system when the repressor is not present or cannot bind to the promoter region, $\Gamma_{max} = \gamma_0(1 - \Delta t/T)$, over the actual transcription $\hat{\Gamma}$:

$$R = \frac{\Gamma_{max}}{\hat{\Gamma}}. \quad (7)$$

Combining the expression of the average transcription over a cell cycle with the expression of the equilibrium repression level $R^{eq} = \gamma_0/\Gamma_{eq}$ leads to the actual nonequilibrium repression level

$$R = \frac{R_{eq}}{(R_{eq} - 1)\alpha + 1} \quad (8)$$

expressed in terms of the equilibrium value and the nonequilibrium parameter α .

In contrast to the equilibrium value, the kinetic analysis shows that the maximum achievable repression level is bounded from above by $1/\alpha \simeq T\tilde{k}_{in}e^{\tilde{k}_{in}\Delta t}$ as a result of the cell-cycle nonequilibrium effects. In general, \tilde{k}_{in} depends on the number of TF molecules, n_{TF} , and under typical conditions, it is expected to increase proportionally to this number as $\tilde{k}_{in} = k_{on}n_{TF}$, which makes the limit tunable at the cellular (e.g., regulation of the number of TFs), environmental (e.g. regulation of the binding of the TFs), and experimental (overexpressing or knocking down TFs) levels.

Experimental validation with the lac operon

The *lac* operon has been the best and most comprehensively characterized genetic system since the discovery of gene regulation (Jacob and Monod, 1961). Its response to lactose relies on three regulatory DNA sites, a main operator O1 and two auxiliary operators O2 and O3, where the LacI repressor can bind. LacI's binding to O1 prevents transcription by the RNA polymerase irrespective of its binding to O2 or O3 (Oehler et al., 1990). Allolactose, a derivative of lactose, and other analogous molecules prevent the specific binding of LacI, leading to the expression of the three genes involved in lactose metabolism. The auxiliary operators are required for the formation of a

higher-order structure of DNA, a DNA loop between the main and an auxiliary operator bound simultaneously to a single LacI tetramer, to efficiently repress transcription using only a handful of LacI molecules. Looping is possible because LacI functions as a tetramer with two dimeric DNA binding domains (Lewis et al., 1996). As in the expression of other metabolic systems, the promoter needs activation by the complex CRP-cAMP to achieve significant transcription. The overall DNA looping-repressor-activator regulation motif is widely present among prokaryotes and eukaryotes (Cournac and Plumbridge, 2013).

To quantify the limitations imposed by the cell cycle on the naïve functioning of gene regulation systems, we apply the approach to the lac operon in *E. coli* with different combinations of operator deletions and sequence changes, WT and increased intracellular LacI numbers, changes in CRP activation, inhibition of repressors by inducers, and WT and synthetic promoters.

The scaling law, Eq. (8), connects equilibrium with actual values in terms of a single nonequilibrium parameter α . Therefore, we use previously computed equilibrium repression levels (Saiz and Vilar, 2008). For the value of α , we consider a characteristic value for the cell division time of $T = 90$ min and the association rate constant for the LacI tetramer binding to an operator of $k_{\text{on}} = 0.3 \text{ molec}^{-1}\text{min}^{-1}$, as inferred from the bursting statistics of single-cell experiments at 30°C (Chang et al., 2022). The value of k_{on} is consistent with the upper and lower bounds established for the LacI dimer at 37°C and 25°C, respectively (Hammar et al., 2014) and is in line with those observed for other TFs in more complex organisms (Larson et al., 2011). Therefore, high repression, $P_0^{eq} \simeq 0$, leads to $\tilde{k}_{in} = 0.3 n_R \text{ min}^{-1}$, where n_R is the number of repressors.

Cell-population activity assays

The early experiments that led to the discovery of gene regulation already measured that the activity of the *lac* operon increased more than 1,000-fold upon mutations in the DNA repressor region (Jacob and Monod, 1961). The key point is, in quantitative terms, that if transcription can resume as soon as the promoter is duplicated, the promoter would remain active, on average, for the value of \tilde{k}_{in}^{-1} ; namely, 20 seconds for the WT case with 10 LacI tetramers per cell. Therefore, the TF kinetics limits the maximum repression level to $\tilde{k}_{in}T = 270$ for WT, which is a factor 5 smaller than the observed value of 1,300 (Jacob and Monod, 1961; Oehler et al., 1990). In general, this limit would not significantly affect the single-operator synthetic systems with the auxiliary operators deleted, which, with a repression level of 20 (Oehler et al., 1990), is expected to be active for 4.5 min during

the whole cell cycle. Nor would it have a substantial impact in the synthetic systems with unnaturally high numbers of LacI tetramers, such as those that express 900 LacI tetramers per cell with an estimated overall search time \tilde{k}_{in}^{-1} for the operators of 0.22 seconds.

The detailed analysis through Eq. (8) shows that, indeed, as the delay in resuming transcription increases, the behavior of the system becomes closer to the expected equilibrium value (Fig. 2). At WT number of repressors, the equilibrium value is generally higher than the measured one and there is a characteristic delay of about 40 seconds that exhibits the closest agreement with the measured values (Fig. 2). For systems with a single operator, repression is generally low and all the kinetic effects collapse within the equilibrium values for all range of repressor molecules per cell (Fig. 2).

***In vivo* single-transcript imaging**

In vivo single-molecule imaging of transcript production from the *lac* system with its genes replaced by an RNA reporter has provided direct evidence of increased transcription after the replication fork (Wang et al., 2019). The probability of transcription in a 5-minute window after replication, with a value of 0.116, is about twice as much as that of previous and subsequent time windows, with an average value of 0.067. These values imply that 8.4% of the mRNA content has been produced upon replication. The observed values of ~ 0.1 transcripts per cell with a degradation time for the transcripts of 2 min, results in a total mRNA production of 10 transcripts per cycle. Therefore, on average, one would expect the replication fork to have induced the production of $\Delta m = 0.84$ transcripts. This quantity, with $\tilde{k}_{in} = 3 \text{ min}^{-1}$ and a basal level of $\gamma_0 = 20$ transcripts/min, leads to a delay in transcription reactivation of

$$\Delta t = -\frac{1}{\tilde{k}_{in}} \ln \left(\Delta m \frac{\tilde{k}_{in}}{\gamma_0} \right) = 41 \text{ s}, \quad (9)$$

which closely follows those inferred in cell population experiments.

Synthetic promoters without activator

The *lac* operon has both positive and negative control of transcription. Therefore, a potential source of delay in restarting transcription after replication is the activator CRP. We tested whether

replacement of the natural CRP-dependent promoter for the *lacUV5* promoter, which is constitutively active, revealed the underlying nonequilibrium effects.

The *lacUV5* promoter has been widely used in multiple applications as it provides constitutively high transcription independently of CRP (Buc and McClure, 1985). Our analysis shows that if the system is far from the kinetic limit imposed by cell cycle perturbations, $R_{eq}\alpha \ll 1$, both actual and equilibrium values are consistent with each other. Indeed, this is the case for the WT promoter with a delay in reactivating transcription of $\Delta t \simeq 40$ s. If the delay is associated with CRP, discrepancies with the equilibrium values should be higher for the *lacUV5* promoter, especially for regulation of the *lacUV5* promoter with strong operators coupled through DNA looping at low to medium Lacl numbers (Muller et al., 1996).

The single loop case for a main operator with a strong auxiliary operator leads to an equilibrium repression level,

$$R_{eq} \simeq 1 + (e^{-g_L/k_B T} + n_R)e^{-g_{O1}/k_B T} \simeq e^{-(g_{O1}+g_L)/k_B T}, \quad (10)$$

that is independent of Lacl expression and the precise binding to the auxiliary operator (Vilar and Saiz, 2013b). It depends mainly on the free energy of binding to $O1$, g_{O1} , and the free energy of looping, g_L , which is substantially lower than $-k_B T \ln n_R$, with $k_B T$ being the thermal energy. However, this independence in Lacl values is not observed for the experimental values of repression (Fig. 3).

The nonequilibrium repression level accurately reproduces the observed behavior and implies just a minor delay of 5.9 seconds in reactivating transcription for the WT values of g_{O1} ($= -0.85 k_B T$) and g_L . The WT value of $g_L = -8.57 k_B T$ for the short $O3$ - $O1$ loop should be expected to be a lower bound. It includes the stabilization effects of CRP-cAMP, which binds within the loop in the WT system but not in the synthetic *lacUV5* promoter. The best agreement for the strongest auxiliary operator (O_i) replacing $O3$ is achieved with an increased free energy of looping of $g_L = -8.07 k_B T$ and a similar delay of 7.1 seconds. This increase in free energy is consistent with an upper bound for the missing stabilization of the loop by binding of CRP to its DNA site between $O1$ and $O3$ (Hudson and Fried, 1990). In both cases, the discrepancy between actual and equilibrium values of the repression level is about a factor 10 for low numbers of repressors (20 Lacl tetramers),

whereas it is negligible for the high number of repressors (50 LacI tetramers) after incorporating the lack of stabilization of the O3-O1 loop by CRP-cAMP.

Induction without CRP-cAMP activation

The role of CRP in the *lac* operon has been a long-standing question. In the earlier literature, it was assumed that levels of the complex CRP-cAMP, the active form that binds DNA, were high in the absence of glucose and low otherwise (Pastan and Perlman, 1970), but subsequent studies found that the major mechanism from preventing the activation of the *lac* operon in the presence of glucose is inducer exclusion; namely, glucose inhibits the transport of lactose inside the cell (Inada et al., 1996). Consistently with these findings, it has been shown in precise detail that CRP-cAMP is naturally present at saturating levels under multiple growing conditions, including either glucose and glycerol, and that it enhances transcription by a factor ~ 200 for fully induced systems and by a factor ~ 50 for fully repressed systems (Kuhlman et al., 2007).

To analyze the induction process in synthetic systems without cAMP, we use the repression values of the system with saturating exogenous cAMP levels (Kuhlman et al., 2007), which closely mimics the WT case, as a first approximation to the equilibrium values R_{eq} . Binding of one IPTG inducer molecule to one LacI monomer deactivates the specific DNA binding of the corresponding dimeric domain (Oehler et al., 2006). Considering $p_F = 1/(1 + [IPTG]/K_I)$ as the probability that one monomer is free from inducer, the number of fully active LacI tetramers is $n_T = n_R p_F^4$ and the number of tetramers with one binding domain impaired is $n_D = n_R(4p_F^3(1 - p_F) + 2p_F^2(1 - p_F)^2)$, which leads to

$$\tilde{k}_{in} = k_{on}n_T + \frac{1}{2}k_{on}n_D. \quad (11)$$

As value of the inducer dissociation constant, we use explicitly $K_I = 7.2 \mu\text{M}$ from the experimentally determined tight range of $6.4 - 8.2 \mu\text{M}$ (Oehler et al., 2006), which closely reproduces multiple equilibrium and kinetic induction situations (Vilar and Saiz, 2013a). The resulting active transcription after replication ($\Delta t = 0$) show an almost perfect agreement with the observed repression level (Fig. 4).

Because the observed repression values for the system with saturating values of cAMP are expected to be lower than the corresponding equilibrium values, we computed the kinetic limit of

the repression level as well (Fig. 4). Strikingly, the kinetic limit provides a perfect match to the experimental values, indicating that the binding at the promoter is sufficiently long-lived for the induction process to be entirely determined by the recurrent nonequilibrium kinetics induced by dislodging of the repressor by the DNA replication fork.

Quantification of the transcription reactivation delay

Overall, our analysis has shown that promoters positively regulated through CRP-cAMP behave as if they were at equilibrium whereas promoters that lack this activation behave strongly limited by the nonequilibrium dynamics resulting from the replication fork dislodging the associated TFs. Precise quantification of different *lac* promoter and operator configurations in *E. coli* (Fig. 5) show that the equilibrium-like behavior is associated with transcription being inactive after replication for 39.7 seconds on average and that nonequilibrium effects are prominent for fast restoring of transcription after replication, with an average delay of 4.4 seconds. These values are higher and lower, respectively, than the 20 seconds ($\sim 1/\tilde{k}_{in}$) needed by the set of 10 TFs to find a functional target on average.

Discussion

Naturally occurring genetic control systems reliably function through the recurrent perturbations of the cell cycle, including nonequilibrium phenomena as diverse as dislodging of TFs from DNA by the replication fork, duplication of the gene, and partition of the TFs into daughter cells. The *lac* operon in *E. coli*, a paradigm of efficiency in gene regulation and the best characterized genetic system, can do so with just a few repressor molecules per cell.

The general theory we have developed has provided a general avenue to analyze the effects of nonequilibrium perturbations in gene regulation. Applied to the cell cycle effects, our results have led to an explicit scaling law that relates observed transcription with its expected equilibrium value in terms of a single parameter that combines the slowest time scale of the system and the generation time. This expression has revealed the existence of major physical constraints in the limits of repression, showing that efficient repression is severely limited by the kinetics of the TFs. Paradoxically, mechanistic predictive models accurately account for a substantial body of the observed average-cell behavior, but these models remarkably fail when they are improved to consider the molecular perturbations associated with the cell cycle (Chang et al., 2022). In this

regard, the *lac* operon is operating 10 times more efficiently than naïvely allowed by the mechanisms considered up to now.

We have identified a class of situations in which it is possible to reconcile equilibrium models with the observed behavior by considering that transcription gets momentarily deactivated upon DNA replication. In the *lac* operon, these situations include the natural system and other situations with the WT *lac* promoter under the positive control of CRP-cAMP. In contrast, regulation without an adjuvant activator, as we have shown for both the *lacUV5* promoter and systems without CRP-cAMP activation, is generally associated with results that are not compatible with equilibrium approaches, up to a factor 10 as observed experimentally, and that need to consider nonequilibrium cell cycle effects. In general, inefficient systems are not subject to these limitations, as for instance, artificial promoters regulated with weak sites. In these cases, the effects of the cell cycle are expected to be present but get masked by the relatively high overall transcription at low numbers of TFs and are not present at high numbers of TFs, which lead to a fast dynamics resulting from high concentrations.

We have analyzed two lines of evidence, cell-population activity assays and *in vivo* transcript imaging, in the light of our approach to quantify the deactivation of transcription conclusion. These two completely different approaches provide the same results.

If the TFs had a much faster kinetics, the nonequilibrium effects would not be apparent. The experimental evidence, however, shows that they do not. The main proof is the direct measurement of the *in vivo* association and dissociation rate constants for the LacI repressor dimer (Elf et al., 2007; Hammar et al., 2014). Concomitantly, the time scale of repressor search kinetics has been confirmed by the statistics of transcriptionally silent intervals and protein burst sizes in single cells for multiple regulatory setups (Chang et al., 2022).

Further supporting the main conclusion, we have shown that highly repressible regulatory systems without CRP-cAMP activation are not consistent with the expected equilibrium behavior, which would rule out much faster kinetics (Figs. 3 and 4). In addition, in these systems, the results get closer to the equilibrium values when the number of LacI molecules increases, as expected as the overall TF target search times decrease (Fig. 4).

Deactivation of transcription after replication has fundamental implications at the physiological level. It has been observed in living cells at the single-molecule level that infrequent events of complete dissociation of LacI from DNA result in large bursts of permease expression that

trigger induction of the *lac* operon and that partial dissociation of LacI from looped DNA leads only to basal expression (Choi et al., 2008). Therefore, the *lac* operon without deactivation of transcription would be induced frequently after replication rather than stochastically and infrequently. In more complex organisms, like yeast and other eukaryotes, it is well established that there are long delays in restarting transcription after DNA replication and that the delay depends on the type of gene (Stewart-Morgan et al., 2019). Therefore, the suppression of the effects of major nonequilibrium perturbations through transcriptional deactivation could also be present in these more complex systems.

Overall, our results show that the observed consistency with equilibrium approaches is not because of a fast dynamic behavior that moves the system close to equilibrium sooner but the consequence of the underlying cellular strategy of shutting down transcription after replication for as long as nonequilibrium effects are present. In this regard, the observed unreasonable effectiveness of equilibrium approaches in natural systems is not coincidental but the result of an efficient, beneficial behavior selected through evolution, as other key properties of natural systems (Dekel and Alon, 2005).

Acknowledgments

J.M.G.V. acknowledges support from Ministerio de Ciencia e Innovación under grants PGC2018-101282-B-I00 and PID2021-128850NB-I00 (MCI/AEI/FEDER, UE). L.S. is indebted to the Max Planck Institute for the Physics of Complex Systems for hosting her for a sabbatical stay and to the ELBE visiting faculty program.

Author contributions

J.M.G.V and L.S conceived, designed, and performed the research.

References

- Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci U S A* 79, 1129-1133.
- Balakrishnan, L., and Bambara, R.A. (2013). Okazaki fragment metabolism. *Cold Spring Harb Perspect Biol* 5, a010173.

- Bashor, C.J., Patel, N., Choubey, S., Beyzavi, A., Kondev, J., Collins, J.J., and Khalil, A.S. (2019). Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science* *364*, 593-597.
- Beard, D.A., and Qian, H. (2008). *Chemical biophysics : quantitative analysis of cellular systems* (Cambridge, UK ; New York: Cambridge University Press).
- Buc, H., and McClure, W.R. (1985). Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. *Biochemistry* *24*, 2712-2723.
- Carey, M., Lin, Y.S., Green, M.R., and Ptashne, M. (1990). A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* *345*, 361-364.
- Chang, C., Garcia-Alcala, M., Saiz, L., Vilar, J.M.G., and Cluzel, P. (2022). Robustness of DNA looping across multiple cell divisions in individual bacteria. *Proc Natl Acad Sci U S A* *119*, e2200061119.
- Choi, P.J., Cai, L., Frieda, K., and Xie, X.S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* *322*, 442-446.
- Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat Rev Genet* *14*, 572-584.
- Cournac, A., and Plumbridge, J. (2013). DNA looping in prokaryotes: experimental and theoretical approaches. *J Bacteriol* *195*, 1109-1119.
- Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature* *436*, 588-592.
- Doudna, J., Bar-Ziv, R., Elf, J., Noireaux, V., Berro, J., Saiz, L., Vavylonis, D., Faulon, J.-L., and Fordyce, P. (2017). How will kinetics and thermodynamics inform our future efforts to understand and build biological systems? *Cell Systems* *4*, 144-146.
- Elf, J., Li, G.W., and Xie, X.S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* *316*, 1191-1194.
- Elshenawy, M.M., Jergic, S., Xu, Z.Q., Sobhy, M.A., Takahashi, M., Oakley, A.J., Dixon, N.E., and Hamdan, S.M. (2015). Replisome speed determines the efficiency of the Tus-Ter replication termination barrier. *Nature* *525*, 394-398.
- Fraser, R.S., and Nurse, P. (1978). Novel cell cycle control of RNA synthesis in yeast. *Nature* *271*, 726-730.
- Garcia, H.G., and Phillips, R. (2011). Quantitative dissection of the simple repression input-output function. *Proc Natl Acad Sci U S A* *108*, 12173-12178.
- Gertz, J., Siggia, E.D., and Cohen, B.A. (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* *457*, 215-218.
- Guharajan, S., Chhabra, S., Parisutham, V., and Brewster, R.C. (2021). Quantifying the regulatory role of individual transcription factors in *Escherichia coli*. *Cell Rep* *37*, 109952.
- Haken, H. (1983). *Synergetics : an introduction : nonequilibrium phase transitions and self-organization in physics, chemistry and biology*, 3rd rev. and enl. edn (Berlin ; New York: Springer).

- Hammar, P., Wallden, M., Fange, D., Persson, F., Baltekin, O., Ullman, G., Leroy, P., and Elf, J. (2014). Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nat Genet* *46*, 405-408.
- Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* *169*, 13-23.
- Hudson, J.M., and Fried, M.G. (1990). Co-operative interactions between the catabolite gene activator protein and the lac repressor at the lactose promoter. *J Mol Biol* *214*, 381-396.
- Inada, T., Kimata, K., and Aiba, H. (1996). Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. *Genes Cells* *1*, 293-301.
- Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* *3*, 318-356.
- Kennell, D., and Riezman, H. (1977). Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *J Mol Biol* *114*, 1-21.
- Kim, H.D., and O'Shea, E.K. (2008). A quantitative model of transcription factor-activated gene expression. *Nat Struct Mol Biol* *15*, 1192-1198.
- Kinney, J.B., Murugan, A., Callan, C.G., Jr., and Cox, E.C. (2010). Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci U S A* *107*, 9158-9163.
- Kuhlman, T., Zhang, Z., Saier, M.H., Jr., and Hwa, T. (2007). Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc Natl Acad Sci U S A* *104*, 6043-6048.
- Lagator, M., Sarikas, S., Steinrueck, M., Toledo-Aparicio, D., Bollback, J.P., Guet, C.C., and Tkacik, G. (2022). Predicting bacterial promoter function and evolution from random sequences. *Elife* *11*, e64543.
- Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* *332*, 475-478.
- Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* *271*, 1247-1254.
- Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* *115*, 751-763.
- Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. *Proc Natl Acad Sci U S A* *107*, 22534-22539.
- Muller, J., Oehler, S., and Muller-Hill, B. (1996). Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol* *257*, 21-29.
- Novick, A., and Weiner, M. (1957). Enzyme Induction as an All-or-None Phenomenon. *Proc Natl Acad Sci U S A* *43*, 553-566.
- Oehler, S., Alberti, S., and Muller-Hill, B. (2006). Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Res* *34*, 606-612.

- Oehler, S., Eismann, E.R., Kramer, H., and Muller-Hill, B. (1990). The three operators of the lac operon cooperate in repression. *EMBO J* 9, 973-979.
- Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427, 737-740.
- Pastan, I., and Perlman, R. (1970). Cyclic adenosine monophosphate in bacteria. *Science* 169, 339-344.
- Pham, T.M., Tan, K.W., Sakumura, Y., Okumura, K., Maki, H., and Akiyama, M.T. (2013). A single-molecule approach to DNA replication in *Escherichia coli* cells demonstrated that DNA polymerase III is a major determinant of fork speed. *Molecular microbiology* 90, 584-596.
- Ptashne, M., and Gann, A. (2002). *Genes & signals* (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).
- Saiz, L. (2012). The physics of protein-DNA interaction networks in the control of gene expression. *J Phys Condens Matter* 24, 193102.
- Saiz, L., and Vilar, J.M. (2006). Stochastic dynamics of macromolecular-assembly networks. *Mol Syst Biol* 2, 2006 0024.
- Saiz, L., and Vilar, J.M.G. (2008). Ab initio thermodynamic modeling of distal multisite transcription regulation. *Nucleic Acids Res* 36, 726-731.
- Sayal, R., Dresch, J.M., Pushel, I., Taylor, B.R., and Arnosti, D.N. (2016). Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early *Drosophila* embryo. *Elife* 5, e08445.
- Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 451, 535-540.
- Stewart-Morgan, K.R., Reveron-Gomez, N., and Groth, A. (2019). Transcription Restart Establishes Chromatin Accessibility after DNA Replication. *Mol Cell* 75, 408-414.
- van Kampen, N.G. (2007). *Stochastic processes in physics and chemistry*, 3rd edn (Amsterdam ; New York: Elsevier).
- Vilar, J.M.G. (2010). Accurate prediction of gene expression by integration of DNA sequence statistics with detailed modeling of transcription regulation. *Biophys J* 99, 2408-2413.
- Vilar, J.M.G., Guet, C.C., and Leibler, S. (2003). Modeling network dynamics: the lac operon, a case study. *J Cell Biol* 161, 471-476.
- Vilar, J.M.G., and Leibler, S. (2003). DNA looping and physical constraints on transcription regulation. *J Mol Biol* 331, 981-989.
- Vilar, J.M.G., and Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev* 15, 136-144.
- Vilar, J.M.G., and Saiz, L. (2011). Control of gene expression by modulated self-assembly. *Nucleic Acids Res* 39, 6854-6863.
- Vilar, J.M.G., and Saiz, L. (2013a). Reliable prediction of complex phenotypes from a modular design in free energy space: an extensive exploration of the lac operon. *ACS Synth Biol* 2, 576-586.

Vilar, J.M.G., and Saiz, L. (2013b). Systems biophysics of gene expression. *Biophys J* *104*, 2574-2585.

Voichek, Y., Bar-Ziv, R., and Barkai, N. (2016). Expression homeostasis during DNA replication. *Science* *351*, 1087-1090.

von Hippel, P.H., Revzin, A., Gross, C.A., and Wang, A.C. (1974). Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. *Proc Natl Acad Sci U S A* *71*, 4808-4812.

Wang, M., Zhang, J., Xu, H., and Golding, I. (2019). Measuring transcription at a single gene copy reveals hidden drivers of bacterial individuality. *Nat Microbiol* *4*, 2118-2127.

Wheat, J.C., Sella, Y., Willcockson, M., Skoultchi, A.I., Bergman, A., Singer, R.H., and Steidl, U. (2020). Single-molecule imaging of transcription dynamics in somatic stem cells. *Nature* *583*, 431-436.

Wong, F., and Gunawardena, J. (2020). Gene Regulation in and out of Equilibrium. *Annu Rev Biophys* *49*, 199-226.

Figures

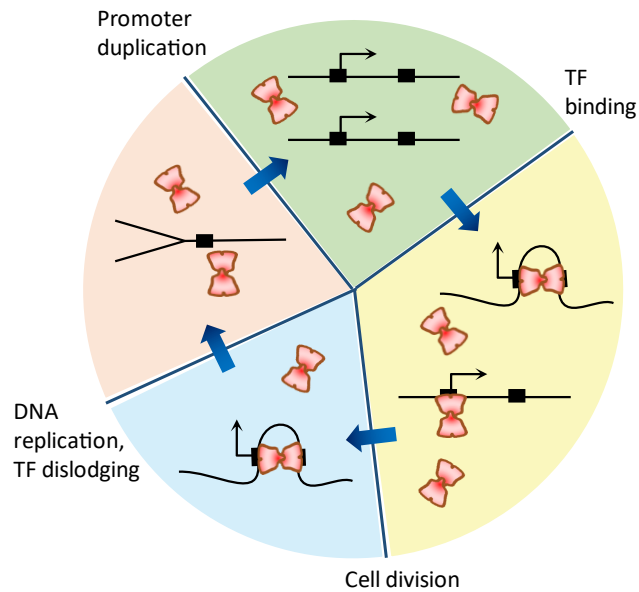


Figure 1. The cell cycle leads to strong recurrent nonequilibrium perturbations to gene regulation.

The cartoon illustrates four sequential recurrent stages in gene regulation (clockwise direction). After cell division, there is a single copy of the promoter (black-angled arrow) and a set of TF binding sites (black rectangles) where the divalent TFs (red figures) can bind, potentially looping DNA (black line). DNA replication leads to dislodging of the TFs from the promoter region. After the passage of the replication fork, the promoter and DNA regulatory elements reappear in duplicate. Afterward, TFs should find their duplicated binding sites to reinstate regulation.

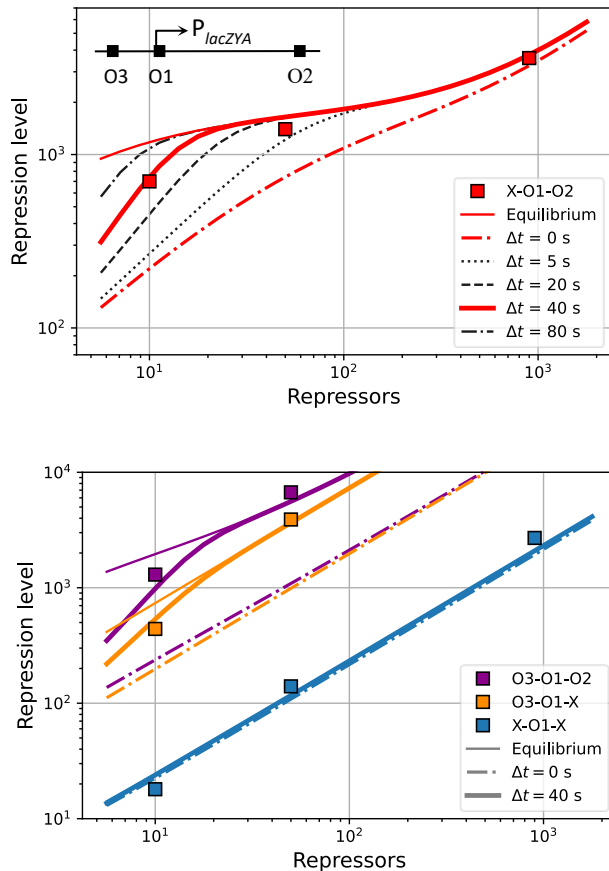


Figure 2. Transcription deactivation after replication for natural promoters makes regulation systems behave as if they were at equilibrium. Measured repression levels for the *lac* operon (Oehler et al., 1990) (colored square symbols) follow with minor differences the corresponding equilibrium predictions (Saiz and Vilar, 2008) (colored continuous thin lines) for different numbers of repressors per cell. The systems considered include the WT operator configuration (O3-O1-O2, bottom panel), O3 deletion (X-O1-O2, top panel), O2 deletion (O3-O1-X, bottom panel), and O2 and O3 deletions (X-O1-X, bottom panel). (The relative positions of the operators and promoter and shown on the top panel inset.) The nonequilibrium results from applying Eq. (8) to the equilibrium predictions with 40 seconds deactivation of transcription after replication (colored thick lines, $\Delta t = 40$ s) are even closer to the measured repression levels. With no deactivation (colored dot-dashed lines, $\Delta t = 0$ s), the nonequilibrium effects would prevent efficient repression. As the deactivation of transcription Δt increases (multiple black lines, top panel), the system behaves as it were closer to equilibrium.

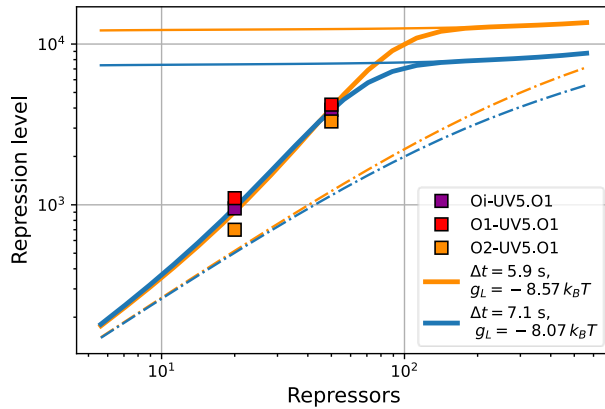


Figure 3. Regulation of the synthetic *lacUV5* promoter is strongly determined by nonequilibrium effects. Measured repression levels (Muller et al., 1996) (colored square symbols) strongly diverge from the corresponding equilibrium predictions obtained through Eq. (10) (continuous thin lines: orange for $g_L = -8.57 k_B T$, blue for $g_L = -8.07 k_B T$) as the number of repressors per cell decreases. The systems considered include the *lacUV5* promoter with O1 as the main operator for three different auxiliary operators [Oi (Oi-UV5.O1), O1 (O1-UV5.O1), and O2 (O2-UV5.O1)]. The nonequilibrium results from applying Eq. (8) to the equilibrium predictions with 5.9 seconds ($g_L = -8.57 k_B T$) and 7.1 seconds ($g_L = -8.07 k_B T$) deactivation of transcription after replication (thick lines) closely follow the measured repression levels. For low numbers of repressors, the repression levels converge to the results with no deactivation (dot-dashed lines, $\Delta t = 0 s$).

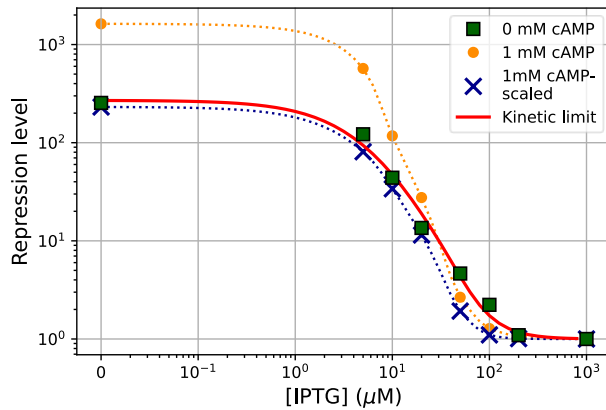


Figure 4. Regulation of the natural *lac* promoter without CRP activation is kinetically determined by nonequilibrium effects. The measured repression levels for strain TK310 with 0 mM cAMP (Kuhlman et al., 2007) (square symbols) for different inducer concentrations are closely approximated by scaling the measured repression levels with 1 mM cAMP (Kuhlman et al., 2007) (yellow dots), as proxies of R^{eq} , through Eq. (8) with $\Delta t = 0$ (crosses). The kinetic limit (red continuous line), obtained with $R^{eq} = \infty$ and $\Delta t = 0$ in Eq. (8), closely follows the measured repression for TK310 with 0 mM cAMP.

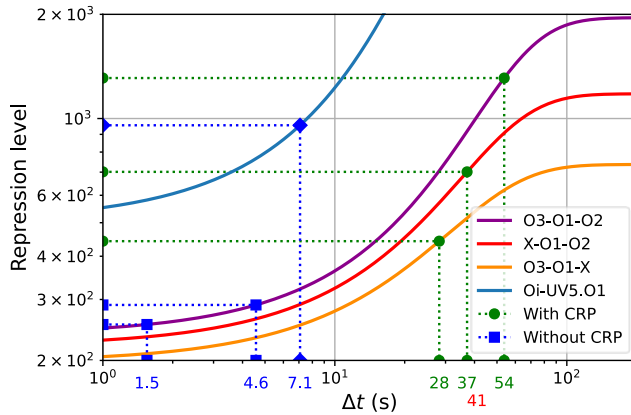


Figure 5. Transcription deactivation after replication is 10-fold longer in WT regulation than in the absence of active CRP. The duration of transcription deactivation is obtained through Eq. (8) as the time Δt at which the computed repression level (continuous colored lines) coincides with the measured value (symbols). O3-O1-O2, X-O1-O2, and O3-O1-X correspond to systems in Fig. 2 with 10 repressors/cell (green dots); Oi-UV5.O1 corresponds to the system in Fig. 3 with 20 repressors/cell (blue diamonds). O3-O1-O2 systems without active CRP (blue squares) include strains TK230 and TK310 (0 mM cAMP) (Kuhlman et al., 2007). Deactivation times Δt for WT regulation (green values; $\langle \Delta t \rangle = 39.7$ s) are on average 10-fold longer than for regulation without active CRP (blue values; $\langle \Delta t \rangle = 4.4$ s). The deactivation time inferred from single-transcript imaging (red value) agrees with the values for WT regulation.