Added value of autoregulation and multi-step kinetics of transcription initiation

Mahendra Kumar Prajapat\textsuperscript{1} and Andre S. Ribeiro\textsuperscript{1,2,3}

\textsuperscript{1}Laboratory of Biosystem Dynamics, Faculty of Biomedical Sciences and Engineering, BioMediTech Institute, and \textsuperscript{2}Multi-scaled Biodata Analysis and Modelling Research Community, Tampere University of Technology, 33101 Tampere, Finland
\textsuperscript{3}CA3 CTS/UNINova, Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal

Bacterial gene expression regulation occurs mostly during transcription, which has two main rate-limiting steps: the close complex formation, when the RNA polymerase binds to an active promoter, and the subsequent open complex formation, after which it follows elongation. Tuning these steps' kinetics by the action of e.g. transcription factors, allows for a wide diversity of dynamics. For example, adding autoregulation generates single-gene circuits able to perform more complex tasks. Using stochastic models of transcription kinetics with empirically validated parameter values, we investigate how autoregulation and the multi-step transcription initiation kinetics of single-gene autoregulated circuits can be combined to fine-tune steady state mean and cell-to-cell variability in protein expression levels, as well as response times. Next, we investigate how they can be jointly tuned to control complex behaviours, namely, time counting, switching dynamics and memory storage. Overall, our finding suggests that, in bacteria, jointly regulating a single-gene circuit's topology and the transcription initiation multi-step dynamics allows enhancing complex task performance.

1. Introduction

Bacterial cells can tune their gene expression profile in response to environmental changes \cite{1–8}. E.g. in \textit{Escherichia coli}, this adaptability is made possible by, among other things, fine-tuning the transcription kinetics of its genes \cite{9}. This is enhanced by the multi-step nature of transcript initiation \cite{10–13}, whose steps can be individually or jointly controlled by promoter-specific external signals (e.g. transcription factors), global regulators such as $\sigma$ factors, etc. \cite{10,14–19}. Evidence suggests that both the mean rate and noise in this dynamics can be tuned \cite{20}. 

© 2018 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.
One way to halt a promoter’s activity is the intervention of a transcription factor (TF), capable of negative regulation [11,21]. Other transcription factors can act as activators [10,21,22]. Relevantly, the transcription dynamics is sequence dependent because e.g. the promoter sequence affects the kinetics of the rate-limiting steps in the initiation, altering the mean and cell-to-cell variability in RNA, and thus, protein numbers [15,18,23–25].

Almost 60% of the TFs produced by *E. coli* are autoregulators [26–28]. Autoregulation allows genes to behave as molecular clocks, switches or memory storage units that assist cells in better controlling response levels and times, etc. [29]. The outcome of introducing TFs in the cytoplasm is usually a change in the kinetics of the rate-limiting steps in transcription initiation of a specific gene(s) [11]. The result of this intervention depends on which rate-limiting step(s) is affected. Namely, acting on a longer-lasting step is likely to have a stronger effect on the RNA production rate than when acting on the shorter-lasting step [23].

In autoregulated genes, the regulatory mode (repression versus activation and the strength of the regulation) is of importance in the resulting changes in mean and cell-to-cell variability in protein numbers [15,30–34], as these affect the fitness of microbial populations [1,9,35,36]. Based on past knowledge (see e.g. [23]) we predict that the effects of autoregulation depend not only on the mode and strength of this regulation but also on the dynamics of transcription initiation of the gene under regulation.

Here, using parameter values extracted from recent single-cell measurements of transcription and translation kinetics in live cells, we designed stochastic models of gene expression controlled by different regulatory modes to explore how the combination of regulation by the action of TFs and regulation of the rate-limiting steps in transcription initiation expands the state space of possible behaviours of autoregulated and externally regulated single-gene circuits. For this, we perform stochastic simulations for varying inducer concentrations, relative durations of the rate-limiting steps in transcription initiation, and binding strengths of the activator/repressor transcription factors (to tune the feedback strength in autoregulation). From these simulations, we assess the mean expression levels at ‘steady state’ (i.e. after a long time period), the cell-to-cell variability in gene expression products and the response times of our circuit in model cells. We note that the term ‘steady state’ here refers to ‘noisy attractors’ [37] because, technically, stochastic models do not have steady states. Finally, we also implement model constitutive genes as null-models, so as to provide a point of reference for quantifying the effects of the autoregulation mechanisms on the gene expression kinetics.

2. Material and methods

2.1. Models of gene expression and regulation mechanisms

We model gene expression when constitutive (acting as a ‘null model’) and when externally or autoregulated by activator/repressor TFs, which act on transcription initiation. The models are depicted in figure 1. In all, we assume a multi-step model of active transcription [38,39], validated in [12]. We note that this model should be applicable to plasmid-borne and chromosome-integrated promoters, provided that the latter are not located in highly expressed operons and, thus, are not strongly influenced by promoter halting due to the accumulation of positive supercoiling build-up [20].

From figure 1a, constitutive promoters are always active [40,41] (i.e. in the ON state). Thus, their expression rate is regulated by the binding/unbinding rates of RNA polymerases (RNAPs) [41]. Constitutive gene expression levels usually depend mostly on the cell growth rate [17,41–43], as this rate can affect the RNAP concentration.

Meanwhile, promoters subject to regulation by TFs can have their activity reduced/enhanced during the time period when repressors/activators are present in the system [44,45]. For that, we assume that when bound by an activator/repressor their activity is affected accordingly (enhanced/reduced). We note that, given the use of a two-step model of transcription (reactions 2.5 and 2.6 below), this model behaviour is not identical to that of a simple on–off switch. These interactions are modelled, respectively, by reaction 2.1 and reaction 2.2:

\[P_{\text{OFF}} + \text{activator} \xrightarrow{k_b} k_u P_{\text{ON}}\]  \hspace{1cm} (2.1)

and

\[P_{\text{ON}} + \text{repressor} \xrightarrow{k_b} k_u P_{\text{OFF}}.\]  \hspace{1cm} (2.2)
In reactions 2.1 and 2.2, the unbinding of activators/repressors occurs at the constant rate $k_U$, while their binding occurs at the rate $k_B$, which depends on inducer/repressor concentration and their effective binding affinity ($k_B$) [46] (estimated by equations (2.3) and (2.4)).

The inducer concentration is represented by $I$. The maximum and minimum affinities of activator/repressor to the promoter are represented by, respectively, $C_{\text{max}}$ and $C_{\text{min}}$ (see electronic supplementary material in [46]). Meanwhile, $K_I$ is a half-maximum concentration of inducers for activators (in % w/v) and repressors (in mM). The binding strength of activators/repressors can be tuned by altering factor...
which is the relative ratio of TF binding rates, which we use as a measure of the feedback strength of autoregulated genes.

\[
k_B = f \cdot k_U \cdot \left( \frac{(C_{\text{max}} - C_{\text{min}})}{C_{\text{max}}/C_{\text{min}}} \right) \left( \frac{1}{1 + (I/K_i)^2} \right)^2 + C_{\text{min}}
\]  

(2.3)

and

\[
k_B = f \cdot k_U \cdot \left( \frac{C_{\text{max}} - C_{\text{min}}}{1 + (I/K_i)^2} \right) + C_{\text{min}}.
\]  

(2.4)

In all models, transcription starts with the binding of a free RNAP to an active promoter (\(P_{\text{ON}}\)), forming a closed complex, \(P_{\text{CC}}\) \([10,19,22,47,48]\) (reaction 2.5). As this step is reversible, multiple closed complex formations can occur between two consecutive RNA production events \([12]\) (reaction 2.5). When ‘successful’, it follows the irreversible open complex formation \([10,49,50]\). Once complete, an mRNA will be produced and the promoter becomes again available to RNAPs \([10,51]\) (reaction 2.6). In reaction 2.6, \(k_2\) represents the inverse of the mean time-length for the open complex to be complete, once initiated:

\[
P_{\text{ON}} \xrightarrow{k_1, \text{RNAP}} P_{\text{CC}}
\]  

\[
P_{\text{CC}} \xrightarrow{k_2, \text{mRNA} + P_{\text{ON}}.}
\]  

(2.5)

(2.6)

The model does not include transcription elongation nor termination because, in normal conditions, these steps are much faster than the rate-limiting steps in initiation \([11,52–59]\), not affecting significantly the time intervals between RNA production events.

In models where gene expression is regulated by TFs, but feedback reactions are absent (figure 1b,c), the production of activators/repressors is assumed to occur at a basal rate (reaction 2.7):

\[
\emptyset \xrightarrow{k_{\text{basal}}} \text{Activator} \\text{Repressor}.
\]  

(2.7)

Model autoregulated genes (figure 1d,e) produce their activators/repressors, which establish a feedback regulatory system (reactions 2.1 and 2.2). Activators/repressors are produced from the RNAs (reaction 2.8):

\[
mRNA \xrightarrow{k_p} \text{mRNA} + \text{protein/activator/repressor}.
\]  

(2.8)

Finally, we assume constant degradation rates of mRNAs (reaction 2.9) and TFs (reaction 2.10) \([60–62]\):

\[
mRNA \xrightarrow{d_{\text{mRNA}}} \emptyset
\]  

(2.9)

and

\[
\text{activator/repressor/protein} \xrightarrow{d_{\text{protein}}} \emptyset.
\]  

(2.10)

The parameter values used here are based on empirical data or have been fitted to physiologically realistic ranges (shown in table 1, unless stated otherwise). Finally, as an approximation, the models assume only one copy of the gene of interest in the cell \([12]\).

2.2. Transcription initiation kinetics and interval between transcription events

In \([12]\), a method was proposed for dissecting the in vivo kinetics of the rate-limiting steps in active transcription (figure 1f). Shortly, from measurements of intervals between consecutive transcription events in individual cells (\(\Delta t\)) at different RNAP concentrations, one can infer the mean duration of the events prior to (\(\tau_{\text{prior}}\)) and after (\(\tau_{\text{after}}\)) the commitment to open complex formation. This is possible as the value of \(\tau_{\text{prior}}\) differs with a concentration of RNAP, while \(\tau_{\text{after}}\) does not \([10]\). Here, for simplicity, we assume that TFs also only affect the kinetics of the first step.
According to the models, the mean interval between consecutive RNA production events ($D_t$) equals, when regulated by activators or repressors, respectively (equations (2.11) and (2.12)):

$$D_t = \left(\frac{k_B + k_{UL}}{k_1 \cdot k_2 \cdot k_B} + \frac{1}{k_2}\right)$$  \hspace{2cm} (2.11)$$

and

$$D_t = \left(\frac{k_B + k_{UL}}{k_1 \cdot k_2 \cdot k_B} + \frac{1}{k_2}\right).$$  \hspace{2cm} (2.12)$$

In constitutively expressed genes, $D_t$ equals (equation (2.13)):

$$D_t = \left(\frac{(k_{-1} + k_2)}{k_1 \cdot k_2} + \frac{1}{k_2}\right).$$  \hspace{2cm} (2.13)$$

### 2.3. Simulations and dynamics evaluation

To simulate the models, we use SGNSim (Stochastic Gene Networks Simulator) [74], a simulator of chemical reaction systems according to the Gillespie’s stochastic simulation algorithm [75,76]. In addition, this simulator allows for the reaction rates to be calculated from complex functions or from physical parameters, when necessary. SGNSim was designed to e.g. model specific genetic circuits and systems of chemical reactions. It further allows for perturbations during simulations, including the introduction of new components in the system.

We focus on how tuning the kinetics of transcription initiation affects the behaviour of the model circuits. For this, instead of changing the mean transcription rate, we alter the fraction of time spent in the events prior to and after initiation of the open complex formation. Namely, we vary $\tau_{\text{after}}/\Delta t$ between 0.05 and 0.95, to cover the wide diversity of empirical values reported in [13,18]. For this, $\Delta t$ is kept constant, and $k_1$ and $k_2$ are changed according to equations (2.12)–(2.15), depending on the

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>values</th>
<th>refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{\text{basal}}$</td>
<td>basal synthesis rate of activator/repressor</td>
<td>0.07 protein min$^{-1}$</td>
<td>*</td>
</tr>
<tr>
<td>$k_0$</td>
<td>unbinding of activator/repressor from promoter</td>
<td>1.8 min$^{-1}$</td>
<td>[46]</td>
</tr>
<tr>
<td>$C_{\text{max}}$</td>
<td>max. affinity of activators to the binding site</td>
<td>1 molecule$^{-1}$</td>
<td>[46]</td>
</tr>
<tr>
<td>$C_{\text{min}}$</td>
<td>max. affinity of repressors to the binding site</td>
<td>0.2 molecule$^{-1}$</td>
<td>[46]</td>
</tr>
<tr>
<td>$k_I$</td>
<td>half-maximal concentration of inducer for activator</td>
<td>2.5% (w/v)</td>
<td>[46]</td>
</tr>
<tr>
<td>$k_{R}$</td>
<td>half-maximal concentration of inducer for repressor</td>
<td>0.035 mM</td>
<td>[46]</td>
</tr>
<tr>
<td>$k_1$</td>
<td>binding affinity of the RNAP to the promoter</td>
<td>+</td>
<td>[12]</td>
</tr>
<tr>
<td>$k_{-1}$</td>
<td>unbinding affinity of the RNAP from the promoter</td>
<td>60 min$^{-1}$</td>
<td>[12]</td>
</tr>
<tr>
<td>$k_2$</td>
<td>effective rate of events after transcription initiation until mRNA synthesis</td>
<td>+</td>
<td>[12]</td>
</tr>
<tr>
<td>$k_F$</td>
<td>rate of translation</td>
<td>27 protein min$^{-1}$</td>
<td>[63–65]</td>
</tr>
<tr>
<td>$d_{\text{mRNA}}$</td>
<td>rate of mRNA degradation</td>
<td>0.12 min$^{-1}$</td>
<td>[60,66]</td>
</tr>
<tr>
<td>$d_{\text{protein}}$</td>
<td>rate of protein/activator/repressor degradation</td>
<td>0.0231 min$^{-1}$</td>
<td>[67]</td>
</tr>
<tr>
<td>RNAP</td>
<td>number of free RNA polymerases</td>
<td>1000</td>
<td>[68–70]</td>
</tr>
<tr>
<td>$I$</td>
<td>inducer concentration for activators</td>
<td>[0–2.5] % (w/v)</td>
<td>[23,42,47]</td>
</tr>
<tr>
<td>$I^*$</td>
<td>inducer concentration for repressor</td>
<td>[0–1] mM</td>
<td>[23,42,47]</td>
</tr>
<tr>
<td>$\Delta t$</td>
<td>avg. duration of transcription intervals</td>
<td>[1–20] min</td>
<td>[12,71–73]</td>
</tr>
</tbody>
</table>

According to the models, the mean interval between consecutive RNA production events ($\Delta t$) equals, when regulated by activators or repressors, respectively (equations (2.11) and (2.12)):
circuit’s topology. By keeping $\Delta t$ constant, the RNA production kinetics (e.g. its noise) is changed due to changing the quantitative relationship between $\tau_{\text{prior}}$ and $\Delta t$, rather than due to changing the mean rate of transcription (which would require changing $\Delta t$). This is because changes in $\Delta t$ are limited by biophysical constraints such as intracellular concentration of RNA polymerases, promoter affinity biophysical limitations, etc. while empirical evidence suggests that $\tau_{\text{prior}}/\Delta t$ can be changed from almost 0 to almost 1 [13,18].

The rates $k_1$ and $k_2$ for constitutive genes were estimated using equations (2.14) and (2.15). Since, in this model, the kinetics of the steps following initiation of the open complex formation does not depend on the regulatory molecules, $k_2$ of externally regulated genes (thus, without feedback) is also estimated using equation (2.15).

$$k_1 = \frac{\Delta t \cdot k_{-1} - \Delta t \cdot (\tau_{\text{prior}}/\Delta t) \cdot k_{-1} + 1}{\Delta t \cdot \text{RNAP} \cdot (\tau_{\text{prior}}/\Delta t)}$$

(2.14)

and

$$k_2 = \frac{1}{\Delta t \cdot [1 - (\tau_{\text{prior}}/\Delta t)]}.$$  

(2.15)

Meanwhile, $k_1$ depends on the inducer and TF intracellular concentrations and, thus, is calculated using equation (2.16) for positive regulation and equation (2.17) for negative regulation. To change the induction strength, while maintaining $\Delta t$ and $\tau_{\text{prior}}/\Delta t$ constant, we change $k_1$, in accordance with equation (2.16) (for activation) and equation (2.17) (for repression), for both externally regulated and self-regulated genes.

$$k_1 = \frac{k_B + k_{U_1} + \Delta t \cdot k_{U_1} \cdot k_{-1} - \Delta t \cdot (\tau_{\text{prior}}/\Delta t) \cdot k_{U_1} \cdot k_{-1} + \Delta t \cdot k_B \cdot k_{-1} - \Delta t \cdot (\tau_{\text{prior}}/\Delta t) \cdot k_B \cdot k_{-1}}{\Delta t \cdot \text{RNAP} \cdot k_B \cdot (\tau_{\text{prior}}/\Delta t)}$$

(2.16)

and

$$k_1 = \frac{k_B + k_{U_1} + \Delta t \cdot k_{U_1} \cdot k_{-1} - \Delta t \cdot (\tau_{\text{prior}}/\Delta t) \cdot k_{U_1} \cdot k_{-1} + \Delta t \cdot k_B \cdot k_{-1} - \Delta t \cdot (\tau_{\text{prior}}/\Delta t) \cdot k_B \cdot k_{-1}}{\Delta t \cdot \text{RNAP} \cdot k_B \cdot (\tau_{\text{prior}}/\Delta t)}.$$  

(2.17)

For any given set of values of variables (e.g. rate constants), we simulate 1000 individual model cells. From these simulations, we extract the mean and cell-to-cell variability in protein numbers in individual cells at steady state. We also estimate the mean activation time, defined as the time taken to reach half of the protein expression levels at steady state. Finally, as TFs and RNAP numbers only affect $\tau_{\text{prior}}$, we use $\tau_{\text{prior}}/\Delta t$ as a means to evaluate the effective influence of transcription initiation on the overall protein expression dynamics (as the dynamics of translation is identical in all models). We also explore how the features added by autoregulation, such as memory storage, bimodal activation and oscillations are affected by $\tau_{\text{prior}}/\Delta t$, feedback strength and inducer concentration. We use this to determine the optimal parameter values for performing these tasks.

We quantify noise in gene expression (variability in e.g. protein numbers over time) by the squared coefficient of variation (CV$^2$) (squared mean over standard deviation). This quantity is shown to differ with $\tau_{\text{prior}}/\Delta t$ (e.g. figure 2b).

We further quantify the uncertainty ($U$) of estimation of a given quantity ($Q$) (e.g. CV$^2$) from the simulations (due to this estimation being performed from a finite set of simulations). $U$ is here quantified by the variance (equation (2.18)), and as expected is shown to differ with $\Delta t$ (i.e. usually, the higher is $\Delta t$, the higher is $U$).

$$U = \frac{\text{Standard deviation}(Q)}{\text{Mean}(Q)}.$$  

(2.18)

To analyse oscillatory behaviours, we calculate the frequency ($F$) of the oscillations as follows, where $n$ is the number of frequency bins in the spectrum, $F_i$ is the frequency of the spectrum at bin $i$ of $n$ and $P_{xx_i} =$ power spectral density (dB Hz$^{-1}$) of spectrum at bin $i$ of $n$:

$$F_{\text{mean}} = \frac{\sum_{i=0}^{n} P_{xx_i} \cdot F_i}{\sum_{i=0}^{n} P_{xx_i}}.$$  

(2.19a)

We also calculate the spread of the amplitudes of each oscillation:

$$\text{Spread} = \frac{\text{Standard deviation}(\text{amplitude})}{\text{Mean}(\text{amplitude})}.$$  

(2.19b)
Finally, as noted, changing $k_1$ in reaction 2.5 alters $\tau_{\text{prior}}/\Delta t$. The resulting values of this parameter can be estimated from the rate constants of the models for constitutive, activated (auto- or externally) and repressed (auto- or externally) genes, as follows, respectively:

$$\frac{\tau_{\text{prior}}}{\Delta t} = \frac{1}{1 + \text{RNAP} \cdot k_1/(k_{-1} + k_2)} \quad \text{(2.20)}$$

$$\frac{\tau_{\text{prior}}}{\Delta t} = \frac{1}{1 + \text{RNAP} \cdot k_1 \cdot k_B/(k_B + k_U) \cdot (k_{-1} + k_2)} \quad \text{(2.21)}$$

$$\text{and} \quad \frac{\tau_{\text{prior}}}{\Delta t} = \frac{1}{1 + \text{RNAP} \cdot k_1 \cdot k_U/(k_B + k_U) \cdot (k_{-1} + k_2)} \quad \text{(2.22)}$$

### 3. Results and conclusion

#### 3.1. Transcription initiation kinetics affects the mean and noise in protein numbers at steady state, but not activation times of constitutive genes

Here, constitutive genes are used as a ‘null model’ to assess, by comparison, the effects of external and autoregulation by TFs. Thus, we first characterize the dynamics of this null model. We simulated the model in figure 1a for varying $\tau_{\text{prior}}/\Delta t$ (while keeping $\Delta t$ constant). Also, we changed $\Delta t$ for fixed $\tau_{\text{prior}}/\Delta t$ values. From the simulations, we extracted the mean and variability (as measured by $\text{CV}_2$) of the protein numbers in individual cells at steady state, and the mean activation times. We also estimated the uncertainty in these variables (equation (2.18)). In these simulations, the model is initialized without proteins.

Results in figure 2 show that $\tau_{\text{prior}}/\Delta t$ does not affect the protein numbers at steady state (figure 2a), as expected, because $\Delta t$ was not altered. Only the cell-to-cell variability in protein numbers is affected, which is expected because higher $\tau_{\text{prior}}/\Delta t$ allows more frequent binding and unbinding of the RNAPs to the active promoters in between transcription events (figure 2b). As such, the uncertainty in these quantities is not affected (figure 2, insets). Meanwhile, changing $\Delta t$ while keeping $\tau_{\text{prior}}/\Delta t$ constant affects the mean and cell-to-cell variability of the protein numbers at steady state. Finally, the uncertainty in these quantities increases with $\Delta t$ (figure 2, insets), due to the decrease in mean RNA and protein numbers.

#### 3.2. Rate-limiting steps in transcription initiation have different effects on autoregulated genes and externally regulated genes

Previous studies showed that the sensitivity of a promoter’s activity to TFs is affected by $\tau_{\text{prior}}/\Delta t$, when TFs do not affect identically the kinetics of the rate-limiting steps in transcription initiation [18,23] (figure 1f). For example, consider two TFs with similar repressing capabilities, with one being able to double the mean duration of the first rate-limiting step, while the other can double the duration of the second rate-limiting step. In this scenario, if e.g. the first rate-limiting step is more longer-lasting than
the second, the TF acting on the first step will have a stronger effect on the rate of RNA production. Thus, we hypothesized that the modes of regulation involving TFs (external and autoregulation) change in sensitivity with $\tau_{\text{prior}}/\Delta t$. To test this, we changed inducer concentration and $\tau_{\text{prior}}/\Delta t$ and assessed the steady state expression levels in each model. Results are shown in figure 3.

Overall, in all cases, the quantitative behaviour of the circuits depends on all three variables ($\Delta t$, $\tau_{\text{prior}}/\Delta t$ and inducer concentration). Meanwhile, the qualitative behaviour depends mostly on $\tau_{\text{prior}}/\Delta t$ and inducer concentration. Interestingly, the effects of each variable depend on the value of the other. E.g. in figure 3b,d, changing inducer concentration has stronger effects if $\tau_{\text{prior}}/\Delta t$ is large. Also, changing $\tau_{\text{prior}}/\Delta t$ has stronger effects for weak inducer levels (figure 3d).

In addition, comparing figure 3a and figure 3b, we find significant differences between external activation and autoactivation. Meanwhile, comparing figure 3c and figure 3d, we find little difference between external repression and autorepression. Further, comparing figure 3a and figure 3c, we see little difference between external activation and repression. Finally, comparing figure 3b and figure 3d, we see significant differences between autoactivation and autorepression.

Also from figure 3a,b, for weak inducer levels, decreasing $\tau_{\text{prior}}/\Delta t$ reduces protein numbers at steady state. This is due to the time window for the RNAP to bind to the unrepressed promoter being shorter. Externally activated genes change expression levels nearly monotonically with inducer concentration until reaching the plateau of full induction (figure 3a,c). Meanwhile, autoactivation causes this increase to be less monotonic (figure 3b). This would be relevant in the context of gene circuits, allowing sharper state shifting. Meanwhile, in figure 3c,d, for weak inducer levels, increasing $\tau_{\text{prior}}/\Delta t$ reduces protein numbers at steady state. I.e. it decreases leaky expression (i.e. protein production when in the presence of repressors).

Figure 3. Mean protein numbers of externally regulated and autoregulated genes as a function of $\tau_{\text{prior}}/\Delta t$ and inducer concentration. (a) Externally activated, (b) autoactivated, (c) externally repressed and (d) autorepressed genes. Note the different scales in the axis showing the mean protein levels at steady state.
3.3. Transcription initiation kinetics affects leaky expression in autorepressed genes

Repressed genes, especially autorepressed, exhibit leaky expression, which can be detrimental to cell growth rates [77–79] and facilitate fast state switching [80], etc. We investigate how leakiness can be tuned by $t_{\text{prior}}/\Delta t$ and autorepression strength. To change the latter, we alter the feedback strength, $f$ (equation (2.3)). In figure 4, we show the steady state expression levels of autorepressed genes as a function of inducer concentration and $t_{\text{prior}}/\Delta t$ for three levels of feedback strength. Weak, medium and strong FB (feedback strength) stand for $f = 1/50$, 1 and 50, respectively.

From figure 4, weak feedback strength causes the circuit to be nearly impervious to changes in inducer concentration and $t_{\text{prior}}/\Delta t$. Meanwhile, for strong feedback, protein numbers at steady state decrease with $t_{\text{prior}}/\Delta t$. Owing to the negative autoregulation, induction levels have little to no effects. Overall, this suggests that increasing $t_{\text{prior}}/\Delta t$ along with strengthening the feedback strength is the best strategy for reducing leaky expression in autorepressed circuits.

3.4. Transcription initiation kinetics affects biphasic behaviour in autoactivation genes

From figure 3b, autoactivated genes exhibit biphasic behaviour for higher values of $t_{\text{prior}}/\Delta t$. Next, we explore how to tune the threshold inducer concentration to reach biphasic behaviour as a function of $f$ (equation (2.3)) and $t_{\text{prior}}/\Delta t$. Results in figure 5 show the steady state expression levels of autoactivated genes as a function of these parameters.

From figure 5, stronger feedback allows the biphasic behaviour to occur at lower induction levels. Also, below a certain feedback strength, this phenomenon is no longer possible. The same occurs if
prior/\Delta t$ is too low. In this regard, figure 6a,b shows that having $k_1 > k_2$ allows higher expression levels as induction is increased (for high values of $\tau_{\text{prior}}/\Delta t$ alone), until a given threshold value, beyond which the opposite occurs, resulting in a biphasic behaviour. From figure 6c,d, this is not observed in autorepressed genes. Interestingly, the feedback strength determines the induction level at which the biphasic behaviour emerges. Overall, these results indicate that the kinetics of transcription initiation can play a key role in autoregulatory networks, even without affecting the mean transcription rate.

3.5. Transcription initiation kinetics affects cell-to-cell variability in protein numbers of autoregulated genes

Next, we study how the kinetics of transcription initiation and the regulatory mode can be combined to regulate cell-to-cell variability in protein numbers in single-gene circuits with feedback. To quantify this variability, we use $CV^2$ in protein numbers at steady state. Results are shown in figure 7.

From figure 7, in general, externally activated and autoactivated genes exhibit higher $CV^2$ in protein numbers at steady state than externally repressed and autorepressed ones. Also, they are more sensitive to inducer concentrations. Externally activated and autoactivated genes differ in that the latter has more variability in behaviour. Meanwhile, in all cases (except in figure 7c, i.e. for externally repressed genes), $\tau_{\text{prior}}/\Delta t$ does not tune this variability significantly. Overall, we find the inducer concentration and, secondly, $\Delta t$ to be the main regulators of cell-to-cell variability in protein numbers at steady state, confirming again the hypothesis that the kinetics of transcription initiation can play a key role in autoregulatory networks, even without affecting the mean transcription rate.
3.6. Autoregulation allows transcription initiation kinetics to tune activation time

Precise timing of events is essential in complex cellular processes [81,82]. Typically, the expression levels of specific genes need to reach a threshold level to trigger subsequent events [81,83,84]. We studied how the inducer concentration and $\tau_{\text{prior}}/\Delta t$ affect activation times (here assumed to be the time to reach half the steady state level). Results are shown in figure 8.

In general, externally activated and autoactivated genes respond slower than externally repressed and autorepressed ones, in agreement with [85–87] (figure 8). Meanwhile, in all cases, activation times are nearly independent of $\tau_{\text{prior}}/\Delta t$. Further, when and only when externally controlled, they are also nearly independent of $\Delta t$. Finally, all circuits are affected by the inducer concentration.

3.7. Transcription initiation kinetics affects the memory storage capacity of autoactivated circuits

We explore how the memory storage capacity of autoactivated circuits can be jointly tuned by the feedback strength ($f$) and the transcription initiation kinetics ($\tau_{\text{prior}}/\Delta t$ by varying $k_1$, equation (2.21)). For this, for each case, cells were simulated under different induction levels until reaching their respective steady states (ON’ states). We studied the transition from the ON to the OFF state, by simulating cells whose initial condition corresponds to the steady state in protein numbers under full induction. For each model, two curves were generated, from OFF to ON, and from ON to OFF. Results in figure 9 show that, in all cases, the two curves do not overlap, demonstrating storage capacity for memory from past states.

From figure 9, increasing the feedback strength enhances the memory storage. Decreasing $k_1$ reduces it. Namely, cells almost failed to store any memory when combining weak feedback with the smaller value of $k_1$ tested (figure 9a). Surprisingly, the higher value of $k_1$ caused the ON state to remain, even after removing the inducer (figure 9c). Finally, by changing $k_1$ and the feedback strength, a wide range of values of $\tau_{\text{prior}}/\Delta t$ was covered (0.2 to 0.8). This range is reduced by weakening the feedback strength and/or $k_1$ (figure 9, Insets).

Figure 7. Cell-to-cell variability in the steady state protein numbers. (a) Externally activated, (b) autoactivated, (c) externally repressed and (d) autorepressed genes for three different values of $\Delta t$ and various inducer concentrations and values of $\tau_{\text{prior}}/\Delta t$. 

\[
\Delta t = \begin{cases} 
1 \text{ min} & \Delta t = 10 \text{ min} & \Delta t = 20 \text{ min} 
\end{cases}
\]
3.8. Transcription initiation kinetics regulates the modality of cell populations with autoactivated genes

We study whether the behaviour of positively regulated genes can be jointly tuned by the transcription initiation kinetics ($k_1$) and feedback strength ($f$). From the simulations, we obtained histograms of the fraction of cells with a given number of proteins at different time points (figure 10).

Figure 10 shows that the feedback strength can tune the probability of emergence of two ‘sub-populations’ from ‘one initial population’, as well as the fraction of cells in each sub-population. These are also affected by $\tau_{\text{prior}}/\Delta t$ (as regulated by $k_1$). Meanwhile, the feedback strength affects the ranges of $\tau_{\text{prior}}/\Delta t$ that can be reached by tuning $k_1$. 

Figure 8. Activation time of externally regulated and autoregulated genes. The time to reach half the mean steady state expression levels of (a) activated, (b) autoactivated, (c) repressed and (d) autorepressed genes with different $\Delta t$ (represented in different colours) was estimated as a function of inducer concentration and $\tau_{\text{prior}}/\Delta t$.

Figure 9. Memory storage capacity of autoactivated genes as a function of feedback strength and $\tau_{\text{prior}}/\Delta t$. (a) $k_1 = 0.4 \text{ min}^{-1}$, (b) $k_1 = 4 \text{ min}^{-1}$ and (c) $k_1 = 40 \text{ min}^{-1}$. In each, three circuits differing in feedback strength were simulated (green, blue and red colours represent weak, medium and strong feedback strength, respectively). The primary $y$-axis shows the normalized mean protein numbers at steady state and the secondary $y$-axis shows the estimated value of $\tau_{\text{prior}}/\Delta t$ (using equation (2.21)). Weak, medium and strong FB (feedback strength) stand for $f = 1/50$, 1 and 50, respectively.
3.9. Transcription initiation kinetics affects oscillatory behaviour in autorepressed genes

Since \( \tau_{\text{prior}}/\Delta t \) affects the time for transcript production to initiate, we hypothesize that it can be used to tune the dynamics of oscillations in the protein numbers resulting from autorepressed genes. We simulated these models for various values of \( k_1 \) and feedback strength and extracted the mean frequency and spread (equation (2.19)) of the oscillations (figure 11). We also calculated the ranges of values of \( \tau_{\text{prior}}/\Delta t \) (equation (2.22)) reached when changing the feedback strength and \( k_1 \).

Results in figure 11 suggest an increase in both the mean frequency and spread of the oscillations for increasing feedback strength. This increase is sensitive to the value of \( k_1 \). Overall, longer-lasting transcription initiation results in faster, more spread oscillations. Finally, in the regime of weak induction, \( \tau_{\text{prior}}/\Delta t \) is mainly controlled by the feedback strength while, overall, the ranges of \( \tau_{\text{prior}}/\Delta t \) decrease for increasing \( k_1 \).

4. Discussion

The dissection of the dynamics of transcription initiation using \textit{E. coli} as a model organism (first \textit{in vitro} [14] and, more recently, \textit{in vivo} [12,23]) has subsequently allowed showing that the kinetics of the rate-limiting steps in transcription initiation is sequence dependent, thus evolvable, and subject to external regulation, and thus adaptive. There is also much evidence that the kinetics of the two main rate-limiting steps can be tuned independently [11,18]. This has several consequences, e.g. two genes with similar rates of mRNA and protein production in one condition can differ widely in other conditions if the new conditions cause the relative durations of the rate-limiting steps of the two genes to differ (e.g. [13]). In recent works, it was also suggested that the effects of this phenomenon may have multi-scale effects, i.e. are tangible not only at the single-gene level but also at the level of small and large-scale genetic circuits [88–90].

Previous studies have thoroughly investigated how negative and positive regulation affects noise (e.g. [91]) and response times (e.g. [80]) in gene expression. Here, we focused on observing the state space of these complex small genetic circuits’ models that combine autoregulatory mechanisms with rate-limiting steps in transcription. To assess the added value of autoregulation, the effects of a combined modification of the parameters of the multi-step transcription and the autoregulation on these circuits were compared to those in constitutive (used as null-models) and externally regulated genes.
Overall, we found that the efficiency with which the models exhibited complex dynamics regulation, such as minimization of leaky expression, biphasic behaviour, regulation of cell-to-cell variability in protein numbers, tuning of activation times, memory storage capacity and bimodal and oscillatory behaviour, was achieved by combining the tuning of the autoregulatory mechanism parameter values with the tuning of the rate-limiting steps in transcription. This suggests that the strategy here used could be of assistance to improve the efficiency of presently existing synthetic genetic circuits. Relevantly, most predictions regarding the changes in kinetics obtained from the simulations could be tested by using such already engineered circuits (e.g. [46]), by changing their original promoters for others with different initiation kinetics (strength and, in particular, relative duration of the rate-limiting steps in transcription initiation [13,18,23]). Similar tests could be performed by changing the binding affinities of TFs to the promoters, whose original values can be found in [15,92–94], as these changes are expected to also allow changes in the transcription initiation kinetics of the genes composing the circuits.

While too extensive to introduce in the present work, in the future, it will be of interest to focus on specific models and further analyse how the various parameter values combine to generate the complex behaviours here reported, such as biphasic response and behavioural transitions.

Evidence suggests that prokaryotic cells evolved several autoregulated genes for time tracking, memory storage and decision making [29]. Given the results above, we hypothesize that this may have been made possible by co-evolving the transcription initiation kinetics of the component genes and the rate constants controlling the autoregulation.

In conclusion, our results may assist the engineering of single-gene synthetic circuits with predefined dynamics using the combined tuning of the feedback strength of the proteins and the kinetics of the rate-limiting steps in transcription initiation of the component promoter in order to maximize the circuit’s efficiency. Such circuits, if efficient, may become of wide use due to their relative simplicity.

Figure 11. Oscillatory dynamics of autorepressed genes as a function of feedback strength and $\tau_{prior}/\Delta t$. Autorepressed genes with different values of $k_1$ ($k_1 = 0.4 \text{ min}^{-1}$, $k_1 = 4 \text{ min}^{-1}$ and $k_1 = 40 \text{ min}^{-1}$) and feedback strengths were simulated, and the mean frequency of oscillations and the spread of amplitudes estimated from protein numbers over time. For each value of $k_1$ and feedback strength, the values of $\tau_{prior}/\Delta t$ were calculated (using equation (2.22)). ‘FB’ stands for feedback strength, which is varied between $f = 1/50$ and 50.

Overall, we found that the efficiency with which the models exhibited complex dynamics regulation, such as minimization of leaky expression, biphasic behaviour, regulation of cell-to-cell variability in protein numbers, tuning of activation times, memory storage capacity and bimodal and oscillatory behaviour, was achieved by combining the tuning of the autoregulatory mechanism parameter values with the tuning of the rate-limiting steps in transcription.

This suggests that the strategy here used could be of assistance to improve the efficiency of presently existing synthetic genetic circuits. Relevantly, most predictions regarding the changes in kinetics obtained from the simulations could be tested by using such already engineered circuits (e.g. [46]), by changing their original promoters for others with different initiation kinetics (strength and, in particular, relative duration of the rate-limiting steps in transcription initiation [13,18,23]). Similar tests could be performed by changing the binding affinities of TFs to the promoters, whose original values can be found in [15,92–94], as these changes are expected to also allow changes in the transcription initiation kinetics of the genes composing the circuits.

While too extensive to introduce in the present work, in the future, it will be of interest to focus on specific models and further analyse how the various parameter values combine to generate the complex behaviours here reported, such as biphasic response and behavioural transitions.

Evidence suggests that prokaryotic cells evolved several autoregulated genes for time tracking, memory storage and decision making [29]. Given the results above, we hypothesize that this may have been made possible by co-evolving the transcription initiation kinetics of the component genes and the rate constants controlling the autoregulation.

In conclusion, our results may assist the engineering of single-gene synthetic circuits with predefined dynamics using the combined tuning of the feedback strength of the proteins and the kinetics of the rate-limiting steps in transcription initiation of the component promoter in order to maximize the circuit’s efficiency. Such circuits, if efficient, may become of wide use due to their relative simplicity.

Data accessibility. All relevant data, general methods and figures are within the main article. The Matlab codes are available on request from the authors.