Harnessing the central dogma for stringent multi-level control of gene expression

F. Veronica Greco[1], Amir Pandi[2], Tobias J. Erb[2,3], Claire S. Grierson[1,4] and Thomas E. Gorochowski[1,4,*]

[1] School of Biological Sciences, University of Bristol, Tyndall Avenue, Bristol, UK

[2] Department of Biochemistry and Synthetic Metabolism, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

[3] SYNMIKRO Center of Synthetic Microbiology, Marburg, Germany

[4] BrisSynBio, University of Bristol, Tyndall Avenue, Bristol, UK

* Correspondence should be addressed to T.E.G. (thomas.gorochowski@bristol.ac.uk)

Keywords: gene expression; transcription; translation; genetic circuit; synthetic biology; regulatory motif; feed forward loop; control

Abstract

Strictly controlled inducible gene expression is crucial when engineering biological systems where even tiny amounts of a protein have a large impact on function or host cell viability. In these cases, leaky protein production must be avoided at all costs, but ideally without affecting the achievable range of expression. Here, we demonstrate how the central dogma offers a simple way to effectively address this challenge. By simultaneously regulating both transcription and translation, we show how relative basal expression of an inducible system can be greatly reduced, with minimal impact on the maximum induced expression rate. Using this approach, we create several stringent expression systems displaying >1000-fold change in their output after induction *in vivo* and up to a 350-fold change when used in a cell-free expression system. Furthermore, we find that multi-level regulation is able to suppress transcriptional noise and creates a digital-like switch when transitioning between ‘on’ and ‘off’ states. This work provides foundational knowledge and a genetic toolkit of parts to create multi-level gene expression controllers for those working with toxic genes or requiring precise regulation and propagation of cellular signals. It also demonstrates the value of exploring more complex and diverse regulatory designs for synthetic biology.

Introduction

Since the development of the first inducible systems in the early 1980s [1], the ability to dynamically control gene expression through the use of small molecules [2], light [3,4], and other signals [5] has revolutionized biotechnology. From controlling shifts between cell growth and protein production stages during large-scale fermentations [6], to the detailed characterization of genetic parts and circuitry [7], the control of gene expression underpins a huge variety of applications. However, while switching expression of a gene 'on' or 'off' is conceptually simple, it is rare for genes to have such discrete states or ever be completely silenced. Stochastic effects [8,9] and leaky expression are widespread and potentially important for adaptation in natural systems but can wreak havoc in engineered systems where genes are toxic to a host or responses are highly sensitive and easily triggered by unavoidable fluctuations [10,11].

Early systems for controlling gene expression relied on repurposing native regulatory components such as transcription factors. One of the most widely used is the P_{tac} system [1]. This consists of a constitutively expressed LacI repressor that can form dimers and tetramers to strongly bind operator sites within a P_{tac} promoter sequence and sterically block initiation of RNA polymerase (RNAP). LacI is sensitive to Isopropyl β-d-1-thiogalactopyranoside (IPTG) and at high concentrations, the DNA binding activity of LacI is abolished. This lifts repression of P_{tac} and leads to strong transcription of genes regulated by this promoter. While in most cases such systems offer strong repression, because such regulatory systems focus on a single step during protein synthesis (i.e. transcription), they are vulnerable to fluctuations in regulator production and the stochastic nature of biochemical reactions during gene expression [9].

Over the past decade, synthetic biologists have developed more advanced methods to control gene expression. These include engineered regulators based on DNA binding proteins such as zinc fingers [12], TALENs [13] and CRISPRi [14], RNA-RNA interactions [15–17], post-transcriptional/translational processes such as RNA and protein degradation [18], as well as using directed evolution to optimize existing inducible systems [19]. This offers a wealth of options to more strictly regulate gene expression through the coupling of multiple forms of regulation (e.g. affecting both transcription and translation of a gene) to reduce unwanted expression and improve the robustness of a system to component failure. However, few examples of such multi-level regulation have been implemented to date [20,21]. This has resulted in an unclear picture of how best stringent multi-level control can be achieved and the trade-offs that exist between performance, regulatory complexity, and cellular burden when designing these systems.

Here, we address this problem by systematically studying the combined use of transcriptional and translational regulators to stringently control protein expression. Using a

combination of mathematical modelling and combinatorial genetic assembly, we are able to design, build and test a variety of synthetic multi-level controllers (MLCs) and elucidate the relative performance of each. These controllers all implement a coherent type 1 feed-forward loop (C1-FFL) regulatory motif (**Figure 1A**) that is commonly found in natural genetic systems and is known to enable more stringent control of an output but is rarely used when designing new expression systems [22]. We show how MLCs offer advantages for many applications spanning the stringent control of protein expression to the accurate propagation of information in a cell [23,24] and demonstrate how applying modern synthetic biology tools to even simple regulatory systems can offer paths towards the precise and reliable control of biological systems.

Results

Stringent control of gene expression by harnessing the central dogma

In most synthetic genetic circuits, control of gene expression is achieved through the use of a single type of regulation (**Figure 1A**), with control of transcription predominantly used. While this type of single-level controller (SLC; **Figure 1B**) is often sufficient for many applications, the central dogma naturally lends itself to more stringent multi-level regulation where both transcription and translation are controlled simultaneously (e.g. via transcription factors and RNA-based translational switches). Such multi-level controllers (MLCs; **Figure 1C**) can be generalised by a genetic design that consists of an *L1* gene encoding a level 1 transcriptional regulator with cognate promoter P_{L1}, and an *L2* gene encoding a level 2 translational regulator. Both *L2* and the gene of interest (GOI) are separately transcribed by P_{L1} promoters and the product of *L2* activates translation of the GOI transcript. This MLC encapsulates a coherent type 1 feed forward loop (C1-FFL) in which both *L1* and *L2* are necessary for production of the GOI.

To explore the possible benefits of this regulatory motif, we developed mathematical models to capture how the rate of production of a GOI varied in response to differing concentrations of an input inducer for both the SLC and MLC designs (**Supplementary Note 1**; **Supplementary Data 1**). We generated steady state response functions by simulating the models using biologically realistic parameters (**Supplementary Table 1**) over a range of different input IPTG concentrations. As expected, the output production rate displayed a sigmoidal shape with both controllers reaching near identical maximum rates at high input IPTG concentrations (**Figure 1D**). The main difference was that the MLC design displayed a 50-fold lower output than the direct controller at low IPTG concentrations, leading to significantly reduced basal expression when no input was present (**Figure 1D**). This caused

the MLC design to have both an increased dynamic range and fold-change between 'off' and 'on' states when compared to the SLC design.

We also simulated the output protein production rate for both models when exposed to a range of dynamic inputs. These included delta functions, as well as pulse and step inputs (**Figure 1E**). Simulations showed that both types of controller displayed virtually identical output responses for both the pulse and step inputs, with only a small reduction in output expression rate for the MLC that matched its lower basal expression level. However, significant differences were observed in the responses to the delta function input. While the SLC led to moderate sized pulses in output, the MLC design fully suppressed all output activity with only tiny fluctuations in the output expression rate observed. The behaviour of the MLC arose from the need for both *L1* and *L2* to be expressed to sufficiently high levels for expression of the GOI to be triggered. The short pulses of expression caused by the delta function input were insufficient to cause this switch and allowed the MLC to effectively filter out these transient events in its input.

The ability to filter out rapid fluctuations is particularly important for stringent control in systems where input promoters exhibit high levels of intrinsic noise. In such scenarios, protein levels can vary significantly across a population of cells [8] due to the often bursty nature of gene transcription. This is commonly seen for weak promoters where intrinsic noise dominates. Rather than the activity of a weak promoter being uniformly low, it instead displays short bursts of strong activity separated by long periods of inactivity [8,25]. Across a population this averages out to a low overall expression level, but large variability is present between cells. As seen for the delta function inputs, such input profiles driving the SLC will lead to large fluctuations in the output. However, because intrinsic promoter noise is specific to an individual promoter and uncorrelated between multiple identical versions of a promoter within a construct, the MLC design which contains two copies of the input promoter P_{L1} should find that a burst of expression from one P_{L1} promoter is highly unlikely to occur at the same time as a burst from the other. Therefore, the MLC will suppress noise in the output.

To test this hypothesis, we generated accurate time-series promoter activity profiles based on a two-state model [25] where the mean length of time a promoter was in an 'on' active and 'off' silent state (Δt_{ON} and Δt_{OFF}, respectively) were $\langle \Delta t_{ON} \rangle$ = 6 min and $\langle \Delta t_{OFF} \rangle$ = 37 min. These values were taken from previous experimental measurements in *E. coli* [25]. We also set the activity of the P_{L1} promoter when in an 'on' state to a biologically realistic 0.25 RNAP/min. Independent time-series were generated for each P_{L1} promoter in the MLC and only one of these was used for the SLC where only a single P_{L1} promoter is present. These profiles were then fed into our existing dynamic models and the responses of the systems simulated. We found that the output production rate for the SLC saw large increases, especially where the input consisted of longer bursts of activity or several bursts in short succession (**Figure 1F**).

125 In comparison, the MLC fully suppressed all output production making it an excellent filter of
126 intrinsic promoter noise.
127
128 ***A genetic template to explore multi-level gene regulation***
129 There are many ways that an MLC could be implemented biologically. Furthermore, when
130 implementing such a controller it is often necessary to switch the input that is used and internal
131 regulators such that multiple controllers can be used simultaneously within the same cell. To
132 meet these requirements, we developed an 8-part genetic template and toolkit of parts to allow
133 for the rapid combinatorial assembly of MLCs (**Figure 2A**). The design enables both single
134 and multi-level regulation, has the option to introduce protein tags for further post-translational
135 control of the GOI (e.g. through signalled degradation) and is structured to minimise the
136 chance for transcriptional readthrough to cause unwanted expression of the component parts.
137 The toolkit comprises 8 types of part plasmid (pA–pH) and a backbone plasmid (pMLC-BB1)
138 in which the final MLC design is inserted (**Supplementary Data 2**). Assembly is performed
139 using a standard one-pot Golden Gate reaction with individual blocks designed to use 4 bp
140 overhangs with minimal cross reactivity to ensure the correct and efficient ligation of parts [26].
141 Furthermore, rapid screening of successful inserts is enabled by the drop-out of an orange
142 fluorescent protein (*ofp*) expression unit [27] (**Supplementary Note 2**).
143 Using this toolkit, we aimed to compare the *in vivo* behaviours of different SLC and
144 MLC designs with a focus on the different mechanisms that could be used for *L2* control and
145 the affect these might have on overall performance. For the SLC design we chose the widely
146 used P_{tac} system introduced earlier (**Figure 2C**). To simplify comparisons, we also used the
147 P_{tac} system for *L1* control in all the MLC designs and combined it with three different RNA-
148 based *L2* regulators. These included a toehold switch (THS; **Figure 2D**) [17,28], a small
149 transcription activating RNA (STAR; **Figure 2E**) [29], and a dual control system (DC; **Figure 2F**)
150 [21].
151 The THS regulator encodes a structural component followed by a ribosome binding
152 site (RBS) that is used to drive translation of the GOI (**Figure 2D**). The structural region is
153 designed to form a strong hairpin loop that when transcribed hinders the ability for ribosomes
154 to bind the RBS, and thus inhibits translation. Translation is activated by expression of a
155 complementary small RNA (sRNA) trigger that hybridizes to a short unstructured region of the
156 THS which causes a breakdown in its secondary structure. This conformational change allows
157 ribosomes to bind the RBS and translation of the GOI to proceed. THSs were selected
158 because they offer strong repression of translation, can be designed computationally, and
159 large libraries of designs exist with minimal crosstalk when used together [17,28].
160 Unlike the THS, the STAR regulator works at a transcriptional level. The STAR's target
161 is placed before an RBS in the 5' untranslated region (UTR) of the GOI (**Figure 2E**). This

162 forms an intrinsic terminator when transcribed and inhibits GOI expression. Activation is
163 achieved by expression of the STAR RNA, which interacts with the target, prevents terminator
164 formation and thus allows for expression of the downstream GOI. Similar to THSs, STARs
165 have been shown to offer strong repression and there exist large libraries of orthogonal
166 variants [16,29].

167 Finally, the DC regulator combines both transcriptional and translational control by
168 modifying the pT181 attenuator [21]. The DC target is placed in the 5' UTR of the GOI and
169 encodes an intrinsic terminator that includes the RBS (**Figure 2F**). When transcribed, the
170 intrinsic terminator not only halts transcription, but also represses translation by causing the
171 RBS to form a strong RNA secondary structure making it inaccessible to the ribosome.
172 Activation is achieved by expression of a STAR, which interacts with the target, both
173 preventing terminator formation and causing a conformational change in the RNA structure
174 that makes the RBS accessible for translation initiation. The DC regulator was chosen due to
175 this combined regulatory action which has been shown to produce strong repression [21].
176 However, to date, only a single of these regulators has been created, limiting future
177 applications.

178 DNA encoding parts for each of these regulatory systems was synthesised and our
179 toolkit used to assemble the SLC and three MLC designs. Superfolder green fluorescent
180 protein (*gfp*) was chosen as the GOI to allow for the measurement of output expression in
181 single cells using flow cytometry.

182

183 ***Performance comparison of the controllers***

184 To characterise the performances of the controllers, we transformed *Escherichia coli* cells with
185 each construct and measured GFP fluorescence using flow cytometry for 'off' and 'on' input
186 states. As P_{tac} was used as an input for all the designs, this corresponded to growing the cells
187 in either 0 or 1 mM IPTG, respectively (**Methods**). Data from these experiments was then
188 used to calculate the dynamic range and fold change in output GFP fluorescence (**Table 1**;
189 **Supplementary Figure 1**).

190 We found a clear separation between output states for all designs with little variation
191 between biological replicates (**Figures 3A**). All MLCs (THS, STAR and DC) reached higher
192 expression levels than the P_{tac} SLC, and the THS and DC designs achieved large >1000-fold
193 changes between output states. Notably, while the STAR design reached a much higher 'on'
194 state than the P_{tac} design, the STARs high levels of basal (leaky) expression when no input
195 was present resulted in a 43% lower fold change (**Table 1**).

196 A challenge when calculating these measures (especially fold change) is the ability to
197 accurately quantify very low level of output GFP fluorescence, which are near or identical to
198 the autofluorescence of the cells. To better understand this aspect, we measured the GFP

199 autofluorescence of untransformed *E. coli* cells, performing 11 biological replicates to estimate
200 a fluorescence distribution that could be used as an approximate detection limit. Overlaying
201 the average and standard deviation of the cell autofluorescence onto our results (**Figure 3A**,
202 dashed line and grey shaded region), we found that the 'off' states for the P_{tac}, THS and DC
203 designs all fell within this region and very close to the average suggesting they have virtually
204 no leaky expression at all.

205 Another difficulty when comparing the performance of the controllers is the need to
206 consider the large differences in the maximum expression rates (e.g. >60-fold difference
207 between the P_{tac} and THS designs for the 'on' state). It should be noted that the same P_{tac}
208 promoter is used as input to all our designs and that it includes a 15 bp upstream spacer
209 element to insulate its function from contextual effects arising from differing nearby sequences
210 that are present in each design [24,30]. It is therefore reasonable to expect the dynamic range of
211 the input promoter's transcriptional activity to be similar for each controller, with differences in
212 output protein expression rate related directly to the different strength ribosome binding sites
213 found in each *L2* regulator or the SLC design. Given these differences and to allow for an
214 unbiased comparison, we calculated the relative basal GFP expression level of each controller
215 as a percentage of its maximum output (**Table 1**). This showed that the THS performed best,
216 displaying a 25-fold decrease in relative basal expression compared to the P_{tac} SLC with
217 0.02% relative basal expression compared to 0.5%, respectively. The DC design also
218 performed well with 0.04% relative basal expression, while the STAR MLC saw the largest
219 relative basal expression of 1.45%, nearly 3 times that of the P_{tac} SLC.

220 While comparisons of average expression levels between 'on' and 'off' states are
221 useful, they are not able to capture the role of cell-to-cell variability inherent in all gene
222 expression [9]. Such variation is crucial when assessing the performance of stringent expression
223 systems because even though average output states might be sufficiently separated to be
224 distinguished, cell-to-cell variation across a population can lead to overlaps in the output
225 distributions. Cells falling in this overlap are impossible to classify resulting in some cells with
226 an undetermined state. Engineers have developed measures to help characterise the strength
227 and quality of a signal (i.e. the ability to distinguish 'on' and 'off' output states) with the Signal
228 to Noise Ratio (SNR) commonly used in other fields such as electronics. SNR has also
229 recently been adapted for use when studying engineered genetic systems making it easier to
230 understand how the quality of signals in a circuit are maintained or degraded as they pass
231 through various genetic devices [23].

232 Using the flow cytometry distributions, we calculated the SNR for each controller in
233 decibel (dB) units (**Table 1**; **Methods**). We found that the P_{tac} SLC performed worst with a low
234 SNR of 0.2 dB, corresponding to a signal barely larger than the noise. This was evident for
235 the flow cytometry distributions where a sizable overlap in the 'on' and 'off' states was seen

236 (**Figure 3B**). All MLCs performed better with the THS achieving an SNR >10 dB. This
237 improved performance was also evident from the flow cytometry data with clear gaps of
238 varying sizes between the 'on' and 'off' output distributions (**Figure 3B**). This clearer
239 separation between cells in an 'on' and 'off' state would make these parts ideal for genetic
240 logic circuits, ensuring signals are cleanly propagated.
241
242 ***Burden of controllers on the host cell***
243 There is a growing awareness of the importance of considering the burden that engineered
244 genetic parts and circuits place on their host cell [31]. The introduction of a genetic construct
245 that sequesters large quantities of shared cellular resources like ribosomes or heavily impacts
246 core metabolic fluxes can lead to reduced growth rates and trigger stress responses that
247 impair the function of engineered genetic parts [32–37]. When designing the MLCs, we
248 purposefully selected RNA-based regulators as previous results suggest that they impose a
249 small metabolic burden on the cell [38]. However, to experimentally verify this in our cells, we
250 generated growth curves for all SLC and MLC designs (**Supplementary Figure 2**). Because
251 the metabolic demands of the controllers would vary based on the concentration of inducer
252 present (due to the varying levels of sRNA or STAR produced), cells were exposed to 4
253 different concentrations of IPTG (0, 0.1, 1, 10 mM) spanning the 'off' and 'on' states of the
254 controllers.
255 From these growth curves, we estimated the doubling time during the exponential
256 growth phase (**Methods**). We found that the SLC and all MLCs displayed similar doubling
257 times of ~70 min (**Figure 3C**). Furthermore, we saw a slight decrease in the doubling times of
258 all controllers as the IPTG concentration increased. This trend is counterintuitive given that an
259 increasing IPTG concentrations will cause expression of the GOI and any *L2* regulators,
260 increasing the burden on the cell. However, it is known that IPTG can have unexpected effects
261 on cell physiology [39] and cause changes in plasmid stability [40], which could lead to reduced
262 overall burden due to fewer copies of the controller plasmid or more efficient utilisation of
263 available nutrients by the cell.
264 We also measured the lag time after inoculation into fresh media before the cells
265 entered exponential growth (**Methods**). We found differences between many of the controllers
266 with a lag time of ~165 min for the P_{tac} and THS designs, a shorter lag time of 88 min for the
267 DC design, and a significantly longer lag time of 373 min for the STAR design (**Figure 3D**).
268 Closer inspection of the growth curves showed that the DC design had a consistently higher
269 initial cell density (optical density at 600 nm of 0.07 compared to 0.04 for the THS design),
270 which could account for the shorter lag phase (**Supplementary Figure 2**). For the STAR
271 design the elongated lag phase coincided with a consistently longer additional time of ~100
272 min to reach saturation of the culture.

To better understand if the extended lag phase of the STAR-based MLC was a general feature to be expected when using this type of regulator, we rebuilt this construct using a different STAR ($STAR_2$) that had an identical initial 72 bp sequence, but unique 10 bp sequence at its 3'-end (**Supplementary Table 2**). As we would expect for such a similar design, testing of the $STAR_2$ construct showed similar performance to the initial STAR design with a good dynamic range and similar leaky expression in its output (**Table 1**; **Supplementary Figure 3**; **Methods**). However, unlike the original, the $STAR_2$ design displayed a lag phase (161 min) and doubling time (72 min) that closely matched the other MLCs. This suggests that long lag times observed for the original STAR design were likely due to some highly specific and uncharacterised off-target interactions with endogenous cellular processes and not due to a general feature of the STAR's regulatory mechanism.

Digital-like transitions and suppression of weak input signals

Our previous modelling of the MLCs showed that in addition to improved performance in 'on' and 'off' states, the addition of the *L2* regulator also altered the response function, causing a sharper transition from an 'off' to 'on' state due to the lower basal expression, and an ability to suppress low level noise in the input (**Figure 1D, E**).

To assess if these features were present, we generated response functions of the controllers by growing the cells in varying concentrations of input inducer and measuring steady state output GFP fluorescence. The sharpness of the transition is captured by the cooperativity of Hill function fits to this data. We found that in comparison to the P_{tac} SLC, both the THS and STAR MLCs saw more than a doubling in this value from 3.4 to more than 7, while the DC design maintained an identical value (**Table 1**). High cooperativities correspond to a very sharp step-like transition between 'on' and 'off' states that is clearly evident from the response function curves (**Figure 4A**). The high non-linearity in the response functions of the THS and STAR MLCs is potentially useful for information processing tasks. In particular, implementing digital logic within cells requires clear 'on' and 'off' states and limited chance for signals to reside at intermediate states. Sharp transitions in the response function ensure that there is less room for an input to fall at an intermediate point during the transition, ensuring an 'on' or 'off' state is always given. Furthermore, a high non-linearity can also be exploited to generate bimodality. For example, if a noisy input is positioned to span the transition point in the response function, a population of cells will have large groups of cells in 'on' and 'off' states, with much fewer in intermediate states because of the sharp transition and small probability of falling in this small region.

To quantify the ability of each MLC to suppress low level input noise, we further analysed the response functions. As mentioned earlier, the large differences in dynamic range make comparisons between designs difficult. Given that the promoter driving transcription for

310 each MLC is identical (P_{tac}), the discrepancies arise from differing *gfp* translation rates
311 controlled by the associated ribosome binding sites. These do differ in sequence and strength
312 for each design and in some cases are specific and integral to the RNA regulator's function.
313 Therefore, to allow for comparisons, we normalised the output of each MLC to its maximum
314 output and used data from the P_{tac} SLC to estimate the input activity of the P_{tac} promoter used
315 in each controller. If no secondary regulation was present (as in the SLC), then we would
316 expect the normalised input and output to follow a straight line where one equals the other
317 (see P_{tac} design in **Figure 4B**). However, if the secondary regulation suppresses the input P_{tac}
318 activity then a lower normalised output to input will be seen, and conversely, an amplification
319 of the input will lead to a higher normalised output to input.

320 Using this approach, we assessed the responses of each MLC and found that all
321 caused a suppression of low levels of input promoter activity and an amplification of higher
322 input activities. This effect was most prominent for the THS and STAR designs, with both able
323 to ensure controller output is maintained below 1% even when the input promoter reaches
324 3.5% activity (**Figure 4B**, insert). These results confirm the findings of our modelling and
325 demonstrate the potential for using MLCs to filter out unwanted input activity in noisy
326 environments.

327

328 ***Controller performance in a cell-free expression system***

329 There has been growing interest in the use of cell-free protein synthesis (CFPS) systems [41]
330 as a means to prototype synthetic genetic circuits [42], enable the rapid characterisation of
331 genetic parts and metabolic pathways [43], and more recently as a novel bioproduction
332 platform[44]. While great progress has been made in expanding the applications of CFPS
333 systems [45–48], strategies to stringently control protein expression have yet to be developed.

334 To assess the performance of our controllers in a cell-free context, we used a CFPS
335 system created from crude *E. coli* cell lysate and performed simple batch reactions (**Methods**)
336 that we continuously monitored so that output expression rate could be inferred from changes
337 in GFP fluorescence over time. These experiments showed that all controllers were also
338 functional in the CFPS system and behaved qualitatively similar to the *in vivo* situation (**Figure**
339 **5A**; **Supplementary Table 3**). Overall, MLC designs performed better than the SLC design,
340 by showing a lower percentage of basal expression, a larger dynamic range and fold changes
341 between 'off' and 'on' states with sharper, more digital-like, transitions (i.e. higher co-
342 operativity in Hill function fits).

343 However, compared to the *in vivo* situation, we observed distinct differences in
344 performance. The largest drop in performance was observed for the P_{tac} SLC design, for which
345 basal expression reached 10% of the maximal output and only a 10-fold dynamic range (i.e.
346 between 'off' and 'on' output states). Performance losses were also observed for the other

347 MLC designs; however, the THS design displayed <1% basal expression and a ~350-fold
348 change between 'off' and 'on' output states. These performance losses were likely caused by
349 the relatively low effective concentrations of regulators that can be achieved in a CFPS system
350 compared to the highly crowded cytoplasm of a living cell [49]. Especially low concentrations of
351 the LacI repressor will likely limit the maximal repression that can be achieved in the CFPS
352 system, accounting for the higher basal expression observed in performance observed.

353 Notably, the STAR MLC showed a similar performance in respect to percentage of
354 basal expression, fold-change and cooperativity compared to the *in vivo* setting (**Table 1** and
355 **Supplementary Table S3**). This robustness may stem from the fact that the STAR design
356 exploits secondary control at a transcriptional level, specifically, through premature
357 termination of transcription in the 5' UTR of the output gene, and therefore regulation limits
358 the potentially active transcripts that are present within the reaction (**Figure 2E**). In contrast,
359 the THS design produces full length transcripts and relies on continuous suppression of
360 translation initiation by RNA secondary structures (**Figure 2D**). Our data suggests that the
361 STAR regulator is less affected by the differing environment of the CFPS system than the
362 THS, enabling the STAR design to maintain virtually identical performance across these
363 contexts.

364 Time course measurements from these experiments also allowed us to quantify the
365 output GFP production rate as the reaction proceeded. This data revealed a key difference
366 between the *in vivo* and CFPS system that was observed for all controller designs. For the
367 first 2 hours the expression rate for 'off' and 'on' states for each design were virtually identical,
368 with regulation only being observed after this point and strongly affecting output GFP
369 production rate after 4 hours (**Figure 5B**). The initial constant output GFP production rates in
370 the CFPS system matched the order of different RBS strengths measured *in vivo*. The more
371 rapid decrease in expression observed for the MLC designs versus the SLC for the regulated
372 'off' state is expected because of the additional regulatory layer (*L2* regulator) of these
373 designs. The observed 'lag' phase of the regulation reflects very likely the time required for
374 each controller to express sufficient LacI to interact with the P_{tac} promoters that act as the input
375 in all our designs. In contrast to the CFPS system, in the *in vivo* experiments the cells had
376 reached exponential growth and the systems were at steady state equal with LacI degradation
377 and dilution rates equalling production rate to keep repressor concentration constant.
378 Therefore, while multi-level regulation offers greatly improved control over gene expression in
379 CFPS systems, for batch reactions, it is crucial that necessary regulatory components (e.g.
380 repressor proteins) are present at sufficient concentrations from the start of an experiment to
381 enable stringent regulation. This could be achieved by generating the CFPS system from cells
382 that already express the regulators at high concentrations, by separately adding these
383 components into the reaction mix before an experiment starts, or by making use of

microreactors to enable the CFPS system to maintain steady state concentrations of regulators through continual dilution of the reaction products [50].

Discussion

In this work we have shown how multi-level control of gene expression offers a means to more stringently regulate gene expression both *in vivo* and *in vitro*. By harnessing the multi-step process of transcription and translation that underpins the central dogma of biology and simultaneously regulating both processes in response to an input signal, we demonstrate through modelling (**Figure 1**) and experiments (**Figures 3–5**) how inducible expression systems can be created with greatly reduced leaky expression when in an 'off' state, while also maintaining high expression rates once induced. Furthermore, we have shown that multi-level regulation creates a more digital-like switch when transitioning between 'off' and 'on' states and suppresses low-level transcriptional noise (**Figure 4**), both of which are valuable properties when developing genetic systems for information processing or when highly toxic products or excitable systems act as downstream products.

Our top MLC design, which makes use of a THS for *L2* regulation, achieved >2000-fold change in output upon induction *in vivo* and displayed a 10 dB SNR (**Table 1**) making it one of the most tightly controlled and high-performance induction systems built to date. Furthermore, the flexibility of our modular genetic toolkit for assembling new multi-level controllers (**Figure 2**), and the availability of many other THSs, makes it easy to develop additional orthogonal MLCs that could be used in parallel within the same cell. It is worth noting that the underlying P_{tac} promoter that the THS MLC uses, achieved only a 93-fold change and 0.2 dB SNR when used alone as an SLC. Therefore, employing the multi-level regulatory approach outlined in this work could offer a means to greatly improve the performance of many existing low-performance transcriptional sensors, without any need to modify the transcription factors or promoter sequences making up these devices.

With the improvements we see when employing multi-level regulation, it is likely no coincidence that small interfering RNAs (siRNAs) are also widely used by bacteria to refine the regulation of many endogenous processes [51–53]. RNAs are perfectly tailored for this task, imposing a small metabolic burden and offering a fast response. In this work, we selected synthetic RNA-based regulators that function through RNA-RNA hybridisation alone. While this reduces our dependencies on other cellular machinery and makes them easier to transfer between strains/organisms, it is known that many endogenous siRNA regulators employ protein chaperones such as Hfq to increase their binding affinity to targets and strengthen their regulatory effect [54]. It would be interesting to explore the use of synthetic regulators that

make use of these chaperones [38] or exploit recent advances in the RNA part design [16] to see whether further improvements in performance are possible.

The stringent regulation of our controllers is achieved by incorporating a C1-FFL regulatory motif that is known to be evolutionarily selected in many natural and engineered systems [55] and can be used to implement many useful functionalities [22]. More recent work has also demonstrated the importance of interconnections and clustering of many motifs in coordinating more complex behaviours [56,57]. While this work focused on demonstrating that transcriptional and translational regulation can fit neatly into a C1-FFL structure, an intriguing future direction would be to explore how these higher-level structures (e.g. motif clusters or higher-level network structures) might be implemented using the approaches outlined in this work to aid the coordination of multiple interrelated processes in parallel.

This study started with the goal of more stringently controlling gene expression. However, through the design of our MLCs it became evident that the more intricate regulatory designs we built had many other benefits. Synthetic biology to date has often focused on simplifying complexity and reducing systems to their minimal parts. Our findings indicate that complementary studies exploring the complexification of synthetic regulatory systems might also reap rewards allowing us to more efficiently exploit the capabilities of biology by combining many diverse processes and parts in unison. The genetic toolkit presented here offers a starting point for such studies focused on the fundamental processes of transcription and translation.

Methods

Strains, media and chemicals

All cloning and characterization of genetic constructs was performed using *Escherichia coli* strain DH10-β (Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC) (New England Biolabs, C3019I). Cells were grown in DH10-β outgrowth medium (New England Biolabs, B9035S) for transformation, LB broth (Sigma-Aldrich, L3522) for general propagation, and M9 minimal media supplemented with glucose (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl (Sigma-Aldrich, M6030), 0.34 g/L thiamine hydrochloride (Sigma T4625), 0.4% D-glucose (Sigma-Aldrich, G7528), 0.2% casamino acids (Acros, AC61204-5000), 2 mM $MgSO_4$ (Acros, 213115000), and 0.1 mM $CaCl_2$ (Sigma-Aldrich, C8106)) for characterization experiments. Antibiotic selection was performed using 100 µg/mL ampicillin (Sigma-Aldrich, A9518) and 50 µg/mL kanamycin (Sigma-Aldrich, K1637). Induction of the expression systems was performed using varying concentrations of isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich, I6758).

455

456 ***Assembly of controllers***

457 All part plasmids were either directly synthesised (GeneArt, Thermo Fisher Scientific) or
458 assembled as complementary single-stranded DNA oligos annealed together. Controllers
459 consisting of 8-parts (pA–pH) plus a backbone (pMLC-BB1) were assembled using a standard
460 Golden Gate cloning method (**Figure 2B**) [27]. Briefly, for each assembly, we started from the
461 18.5 ng of required part plasmids (pA–pH) and 18.5 ng of the backbone (pMLC-BB1) to be
462 added to a 5 µL Golden Gate reaction. The standard manufacturer's reaction conditions were
463 used, but at a quarter of their normal volume (New England Biolabs, E1601). 2 µL of this
464 reaction mix was then used to transform 12.5 µL of chemically competent DH10-β cells (New
465 England Biolabs, C3019) for further experiments. All assembled constructs were sequence
466 verified by Sanger sequencing (Eurofins Genomics). Annotated sequences of all part and
467 backbone plasmids and assembled controllers are provided in GenBank format in
468 **Supplementary Data 2**. Plasmid maps are shown in **Supplementary Figures 4** and **5**.

469

470 ***Characterisation experiments***

471 Single colonies of cells transformed with an appropriate genetic construct were inoculated in
472 200 µL M9 media supplemented with glucose and kanamycin for selection in a 96-well
473 microtiter plate (Thermo Fisher Scientific, 249952). Cultures were grown for 14 hours in a
474 shaking incubator (Stuart, S1505) at 37 °C and 1250 rpm. Following this, the cultures were
475 diluted 3:40 (15 µL in 185 µL) in M9 media supplemented with glucose, kanamycin for
476 selection and IPTG for induction in a new 96-well microtiter plate and grown for a further 4
477 hours under the same conditions. Finally, the cultures were further diluted 1:10 (10 µL into 90
478 µL) in phosphate-buffered saline (PBS) (Gibco,18912-014) containing 2 mg/mL kanamycin to
479 halt protein translation. These samples were incubated at room temperature for 1 hour to allow
480 for full maturation of GFP before flow cytometry was performed.

481

482 ***Flow cytometry***

483 Measurements of GFP fluorescence in single cells was performed using an Acea Biosciences
484 NovoCyte 3000 flow cytometer equipped with a NovoSampler to allow for automated collection
485 of samples from a 96-well microtiter plate. Data collection was performed using the
486 NovoExpress software. Cells were excited using a 488 nm laser and GFP fluorescence
487 measurements taken using a 530 nm detector. At least 10^6 events were captured per sample.
488 In addition, to enable conversion of GFP fluorescence into calibrated MEFL units [58] a single
489 well per plate contained 15 µL of 8-peak Rainbow Calibration Particles (Spherotech, RCP-30-
490 5A) diluted into 200 µL PBS. Automated gating of events and conversion of GFP fluorescence
491 into MEFL units was performed using the forward (FSC) and side scatter (SSC) channels and

492 the FlowCal Python package version 1.2.2 with default parameters [58]. To correct for the GFP
493 autofluorescence of cells, *E. coli* DH10-β cells containing no genetic construct were grown in
494 identical conditions. An average measurement of GFP fluorescence in MEFL units from three
495 biological replicates of these cells was then subtracted from fluorescence measurements of
496 cells containing our genetic constructs to correct for cell autofluorescence.
497
498 ***Plate reader measurements of construct performance in vivo***
499 Single colonies of cells transformed with an appropriate genetic construct were inoculated in
500 200 µL M9 media supplemented with glucose and kanamycin for selection in a 96-well
501 microtiter plate (Thermo Fisher Scientific, 249952). Cultures were grown for 4 hours in a
502 shaking incubator (Stuart, S1505) at 37 °C and 1250 rpm. Following this, the cultures were
503 diluted 3:40 (15 µL in 185 µL) in M9 media supplemented with glucose, kanamycin for
504 selection and IPTG for induction in a 96-well 190 µm clear base imaging microplate (4titude,
505 Vision Plate™, 4ti-0223). This spectrophotometric assay was performed using a BioTek
506 Synergy Neo2 plate reader at 37°C. Optical density at 600 nm (OD_{600}) was measured every
507 10 min over a 16-hour period. OD_{600} measurements were also taken from samples of M9
508 medium supplemented with glucose containing no cells to allow for quantification of media
509 autofluorescence. Shaking was automated for the all the length of the experiment. Data
510 collection was performed using the Gen5 version 3.04 software. For each time point, media
511 autofluorescence was subtracted from the sample measurement.
512
513 ***Cell-free expression***
514 The *E. coli* cell lysate for CFPS was prepared using an autolysis protocol [59]. In this protocol,
515 *E. coli* BL21-Gold (DE3) cells harboring a pAS-LyseR plasmid give a high-quality cell lysate
516 by freeze-thawing. Specifically, these cells were grown overnight at 37 °C in LB broth
517 supplemented with ampicillin. On the following day, cells were sub-cultured in 2 L of 2X YTPG
518 medium supplemented with ampicillin and grown at 37 °C to an OD_{600} of 1.5. Cells were then
519 harvested at 2000*g* for 15 min at room temperature in four centrifuge bottles and 45 mL of
520 cold S30A buffer (50 mM Tris-HCl at pH 7.7, 60 mM potassium glutamate, 14 mM magnesium
521 glutamate, final pH 7.7) was added to each. Cells were then resuspended by vigorous vortex
522 mixing and poured into a pre-weighted 50 mL falcon tube and centrifuged as in the previous
523 step. The supernatants were completely removed, and the falcon tubes weighted again. The
524 net weight of each pellet was calculated and relative to its weight two volumes of cold S30A
525 supplied with 2 mM DTT was added (3 mL for 1.5 g of pellet). After vigorous vortex mixing,
526 the samples were stored at –80 °C. The next day, frozen cells were placed in a room
527 temperature water bath to thaw, vigorously vortexed, incubated at 37 °C on a shaker for 45
528 min, vortex mixed again, and then incubated at 37 °C for 45 min. The samples were then

529 centrifuged at 30000*g* for 60 min at 4 °C. The supernatants were carefully pipetted out and
530 aliquoted in 1.5 µL tubes, and then finally centrifuged at 20000*g* using a tabletop centrifuge
531 for 5 min to remove residual cell debris. Aliquots of the lysate were stored at –80 after flash-
532 freezing in liquid nitrogen.

533 For the prepared lysate, Mg-glutamate and K-glutamate were titrated in with all
534 components of the cell-free reaction based on the protocol of Sun *et al.* [60] resulting in
535 concentrations of 10 nM and 60 mM, respectively, for optimal GFP production. Each reaction
536 was prepared at a final volume of 10.5 µL containing 33% lysate, Mg-glutamate and K-
537 glutamate as titrated, and amino acids mix, energy mix, and PEG 8000. For cell-free
538 experiments of the SLC and MLC constructs, maxi-prepped plasmids (using Machery-Nagel
539 NucleoBond Xtra Maxi kit) were added at a final concentration of 10 nM along with varying
540 concentrations of IPTG (0, 0.002, 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 5, 10 mM). While gently
541 mixing by pipette, 10 µL of reactions were transferred to a 384-well plate (Greiner Bio-One,
542 784076) and GFP fluorescence was monitored (excitation/emission wavelengths of 485/528
543 nM and gain = 100) every 10 min in a plate reader (Tecan Infinite 200 PRO).

544

545 ***Signal to noise ratio***

546 The signal to noise ratio (SNR) in decibel (dB) units was calculated from the flow cytometry
547 GFP fluorescence distributions using the equation [23]

548 $$\mathrm{SNR_{dB}} = 20 \cdot \log_{10} \frac{|\log_{10}(\mu_{\mathrm{ON}}/\mu_{\mathrm{OFF}})|}{2 \cdot \log_{10}(\sigma)}. \quad (1)$$

549 Here, μ_{ON} and μ_{OFF} are the geometric means of distributions for the 'on' and 'off' states,
550 respectively, and σ is the geometric standard deviation of the distribution for the 'off' state. 'off'
551 and 'on' states correspond to cells grown in 0 and 1 mM IPTG, respectively.

552

553 ***Data analysis and numerical simulation***

554 Data analysis was performed using Python version 3.7.4 and the NumPy version 1.17.4, SciPy
555 version 1.3.1, Pandas version 1.0.3, FlowCal version 1.2.2, and Matplotlib version 3.1.1
556 libraries. ODE models were simulated using the odeint function of the SciPy.integrate Python
557 package version 1.1 with default parameters. Steady-state response functions of the
558 controllers were calculated by fitting median GFP fluorescence values from the flow cytometry
559 distributions for a range of input IPTG concentrations to the following Hill function

560 $$y = y_{\min} + (y_{\max} - y_{\min}) \frac{x^n}{K^n + x^n}. \quad (2)$$

561 Here, y is the output GFP fluorescence in MEFL units, $y_{\min}$ and $y_{\max}$ are the minimum and
562 maximum output GFP fluorescence in MEFL units, respectively, K is the input IPTG
563 concentration at which the output is half-maximal, n is the Hill coefficient, and x is the input
564 IPTG concentration. Fitting of the experimental data was performed using non-linear least

565 squares and the curve_fit function from the SciPy.integrate package version 1.1. Genetic
566 diagrams were generated using DNAplotlib version 1.0 [61,62] and figures were composed using
567 Omnigraffle version 7.15 and Affinity Designer version 1.8.3.
568

569 ## Data availability

570 Python scripts simulating the ODE models of the direct and multi-level controllers be found in
571 **Supplementary Data 1**. Annotated sequences for all plasmids in GenBank format are
572 available in **Supplementary Data 2**. All plasmids used in this study are available from
573 Addgene.
574

575 ## Acknowledgements

576 This work was supported by BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research
577 Centre grant BB/L01386X/1 (T.E.G., C.S.G.), a Royal Society PhD Studentship (F.V.G.), a
578 Royal Society University Research Fellowship grant UF160357 (T.E.G.), the Max Planck
579 Society (T.J.E.), and a European Molecular Biology Organization (EMBO) long-term
580 postdoctoral fellowship (A.P.)
581

582 ## Author Contributions

583 T.E.G. conceived the study. V.G. designed the genetic toolkit, assembled all controllers and
584 performed the *in vivo* experiments. A.P. performed the *in vitro* cell-free experiments. T.E.G.
585 developed and simulated the mathematical models. V.G. and T.E.G. analysed the data.
586 T.E.G., C.S.G., T.J.E. supervised the work. V.G., T.E.G. and C.S.G. wrote the manuscript with
587 input from all other authors.
588

589 ## Competing Interests

590 The authors declare no competing interests.

591 ## References

592 1. de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived
593 from the trp and lac promoters. *Proc. Natl. Acad. Sci.* **80**, 21 (1983).
594 2. Gallivan, J. P. Toward reprogramming bacteria with small molecules and RNA. *Curr. Opin.*
595 *Chem. Biol.* **11**, 612–619 (2007).
596 3. Baumschlager, A., Aoki, S. K. & Khammash, M. Dynamic Blue Light-Inducible T7 RNA
597 Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control. *ACS*
598 *Synth. Biol.* **6**, 2157–2167 (2017).
599 4. Castillo-Hair, S. M., Baerman, E. A., Fujita, M., Igoshin, O. A. & Tabor, J. J. Optogenetic
600 control of Bacillus subtilis gene expression. *Nat. Commun.* **10**, 3099 (2019).
601 5. Sen, S., Apurva, D., Satija, R., Siegal, D. & Murray, R. M. Design of a Toolbox of RNA
602 Thermometers. *ACS Synth. Biol.* **6**, 1461–1470 (2017).
603 6. Sivashanmugam, A. *et al.* Practical protocols for production of very high yields of
604 recombinant proteins using Escherichia coli. *Protein Sci.* **18**, 936–948 (2009).
605 7. Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R. & Tabor, J. J. Characterizing
606 bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat.*
607 *Methods* **11**, 449–455 (2014).
608 8. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a
609 Single Cell. *Science* **297**, 1183 (2002).
610 9. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression
611 and Its Consequences. *Cell* **135**, 216–226 (2008).
612 10. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli:
613 advances and challenges. *Front. Microbiol.* **5**, 172 (2014).
614 11. Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene
615 regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545–550 (2006).
616 12. Khalil, A. S. *et al.* A Synthetic Biology Framework for Programming Eukaryotic
617 Transcription Functions. *Cell* **150**, 647–658 (2012).
618 13. Deng, D., Yan, C., Wu, J., Pan, X. & Yan, N. Revisiting the TALE repeat. *Protein Cell* **5**,
619 297–306 (2014).
620 14. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and
621 Activation. *Cell* **159**, 647–661 (2014).
622 15. Bartoli, V., Meaker, G. A., di Bernardo, M. & Gorochowski, T. E. Tunable genetic devices
623 through simultaneous control of transcription and translation. *Nat. Commun.* **11**, 2095
624 (2020).

625 16. Chappell, J., Westbrook, A., Verosloff, M. & Lucks, J. B. Computational design of small
626 transcription activating RNAs for versatile and dynamic gene regulation. *Nat. Commun.* **8**,
627 1051 (2017).
628 17. Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold Switches: De-Novo-Designed
629 Regulators of Gene Expression. *Cell* **159**, 925–939 (2014).
630 18. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nat. Biotechnol.*
631 **32**, 1276–1281 (2014).
632 19. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. Escherichia coli
633 "Marionette" strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.* **15**,
634 196–204 (2019).
635 20. Calles, B., Goñi-Moreno, Á. & de Lorenzo, V. Digitalizing heterologous gene expression
636 in Gram-negative bacteria with a portable ON/OFF module. *Mol. Syst. Biol.* **15**, e8777
637 (2019).
638 21. Westbrook, A. M. & Lucks, J. B. Achieving large dynamic range control of gene expression
639 with a compact RNA transcription–translation regulator. *Nucleic Acids Res.* **45**, 5614–
640 5624 (2017).
641 22. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc.*
642 *Natl. Acad. Sci.* **100**, 11980 (2003).
643 23. Beal, J. Signal-to-Noise Ratio Measures Efficacy of Biological Computing Devices and
644 Circuits. *Front. Bioeng. Biotechnol.* **3**, 93 (2015).
645 24. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352**, aac7341 (2016).
646 25. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-Time Kinetics of Gene Activity
647 in Individual Bacteria. *Cell* **123**, 1025–1036 (2005).
648 26. Woodruff, L. B. A. *et al.* Registry in a tube: multiplexed pools of retrievable parts for genetic
649 design space exploration. *Nucleic Acids Res.* **45**, 1553–1565 (2016).
650 27. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot
651 DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE* **4**, e5553
652 (2009).
653 28. Green, A. A. *et al.* Complex cellular logic computation using ribocomputing devices.
654 *Nature* **548**, 117 (2017).
655 29. Chappell, J., Takahashi, M. K. & Lucks, J. B. Creating small transcription activating RNAs.
656 *Nat. Chem. Biol.* **11**, 214–220 (2015).
657 30. Zong, Y. *et al.* Insulated transcriptional elements enable precise design of genetic circuits.
658 *Nat. Commun.* **8**, 52 (2017).
659 31. Boo, A., Ellis, T. & Stan, G.-B. Host-aware synthetic biology. *Synth. Biol.* **14**, 66–72 (2019).
660 32. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene
661 expression designs with reduced burden. *Nat. Methods* **12**, 415–418 (2015).

662 33. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat. Methods* **15**,
663 387–393 (2018).
664 34. Gorochowski, T. E., van den Berg, E., Kerkman, R., Roubos, J. A. & Bovenberg, R. A. L.
665 Using Synthetic Biological Parts and Microbioreactors to Explore the Protein Expression
666 Characteristics of Escherichia coli. *ACS Synth. Biol.* **3**, 129–139 (2014).
667 35. Gorochowski, T. E., Avcilar-Kucukgoze, I., Bovenberg, R. A. L., Roubos, J. A. & Ignatova,
668 Z. A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression
669 between Endogenous and Synthetic Genes. *ACS Synth. Biol.* **5**, 710–720 (2016).
670 36. Gyorgy, A. *et al.* Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophys.*
671 *J.* **109**, 639–646 (2015).
672 37. Weiße, A. Y., Oyarzún, D. A., Danos, V. & Swain, P. S. Mechanistic links between cellular
673 trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci.* **112**, E1038 (2015).
674 38. Kelly, C. L. *et al.* Synthetic negative feedback circuits using engineered small RNAs.
675 *Nucleic Acids Res.* **46**, 9875–9889 (2018).
676 39. Malakar, P. & Venkatesh, K. V. Effect of substrate and IPTG concentrations on the burden
677 to growth of Escherichia coli on glycerol due to the expression of Lac proteins. *Appl.*
678 *Microbiol. Biotechnol.* **93**, 2543–2549 (2012).
679 40. Gomes, L., Monteiro, G. & Mergulhão, F. The Impact of IPTG Induction on Plasmid
680 Stability and Heterologous Protein Expression by Escherichia coli Biofilms. *Int. J. Mol. Sci.*
681 **21**, (2020).
682 41. Perez, J. G., Stark, J. C. & Jewett, M. C. Cell-Free Synthetic Biology: Engineering Beyond
683 the Cell. *Cold Spring Harb. Perspect. Biol.* **8**, (2016).
684 42. Niederholtmeyer, H. *et al.* Rapid cell-free forward engineering of novel genetic ring
685 oscillators. *eLife* **4**, e09771 (2015).
686 43. Karim, A. S. & Jewett, M. C. A cell-free framework for rapid biosynthetic pathway
687 prototyping and enzyme discovery. *Metab. Eng.* **36**, 116–126 (2016).
688 44. Kelwick, R. J. R., Webb, A. J. & Freemont, P. S. Biological Materials: The Next Frontier
689 for Cell-Free Synthetic Biology. *Front. Bioeng. Biotechnol.* **8**, 399 (2020).
690 45. Laohakunakorn, N. *et al.* Bottom-Up Construction of Complex Biomolecular Systems With
691 Cell-Free Synthetic Biology. *Front. Bioeng. Biotechnol.* **8**, 213 (2020).
692 46. Borkowski, O. *et al.* Cell-free prediction of protein expression costs for growing cells. *Nat.*
693 *Commun.* **9**, 1457 (2018).
694 47. Swank, Z., Laohakunakorn, N. & Maerkl, S. J. Cell-free gene-regulatory network
695 engineering with synthetic transcription factors. *Proc. Natl. Acad. Sci.* **116**, 5892 (2019).
696 48. Pandi, A., Grigoras, I., Borkowski, O. & Faulon, J.-L. Optimizing Cell-Free Biosensors to
697 Monitor Enzymatic Production. *ACS Synth. Biol.* **8**, 1952–1957 (2019).

49. McGuffee, S. R. & Elcock, A. H. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLOS Comput. Biol.* **6**, e1000694 (2010).
50. Niederholtmeyer, H., Stepanova, V. & Maerkl, S. J. Implementation of cell-free biological networks at steady state. *Proc. Natl. Acad. Sci.* **110**, 15985 (2013).
51. Gottesman, S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet.* **21**, 399–404 (2005).
52. Storz, G., Opdyke, J. A. & Zhang, A. Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol.* **7**, 140–144 (2004).
53. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628 (2009).
54. Soper, T., Mandin, P., Majdalani, N., Gottesman, S. & Woodson, S. A. Positive regulation by small RNAs and the role of Hfq. *Proc. Natl. Acad. Sci.* **107**, 9602 (2010).
55. Milo, R. *et al.* Network Motifs: Simple Building Blocks of Complex Networks. *Science* **298**, 824 (2002).
56. Gorochowski, T. E., Grierson, C. S. & di Bernardo, M. Organization of feed-forward loop motifs reveals architectural principles in natural and engineered networks. *Sci. Adv.* **4**, eaap9751 (2018).
57. S. Dunn, H. Kugler & B. Yordanov. Formal Analysis of Network Motifs Links Structure to Function in Biological Programs. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1–1 (2019) doi:10.1109/TCBB.2019.2948157.
58. Castillo-Hair, S. M. *et al.* FlowCal: A User-Friendly, Open Source Software Tool for Automatically Converting Flow Cytometry Data from Arbitrary to Calibrated Units. *ACS Synth. Biol.* **5**, 774–780 (2016).
59. Didovyk, A., Tonooka, T., Tsimring, L. & Hasty, J. Rapid and Scalable Preparation of Bacterial Lysates for Cell-Free Gene Expression. *ACS Synth. Biol.* **6**, 2198–2208 (2017).
60. AU - Sun, Z. Z. *et al.* Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *J. Vis. Exp.* e50762 (2013) doi:10.3791/50762.
61. Bartoli, V., Dixon, D. O. R. & Gorochowski, T. E. Automated Visualization of Genetic Designs Using DNAplotlib. in *Synthetic Biology: Methods and Protocols* (ed. Braman, J. C.) 399–409 (Springer New York, 2018). doi:10.1007/978-1-4939-7795-6_22.
62. Der, B. S. *et al.* DNAplotlib: Programmable Visualization of Genetic Designs and Associated Data. *ACS Synth. Biol.* **6**, 1115–1119 (2017).

730 ## Tables

731 **Table 1: Performance summary of the single- and multi-level controllers *in vivo*[a]**

Controller	Type[b]	Basal[c] (%)	Dynamic range[d] (10^3 MEFL)	Fold change[d]	Cooperativity[e], *n*	SNR[f] (dB)
P_{tac}	SLC	0.5	0.9	93	3.4	0.2
THS	MLC	0.02	65.5	2166	7.3	10.1
STAR	MLC	1.45	4.6	53	9.3	4.7
$STAR_2$	MLC	2.0	3.4	37	7.7	4.5
DC	MLC	0.04	10.4	1030	3.4	7.1

732 a. All values are averages calculated from three biological replicates. Key performance features of
733 the controllers are visually shown in **Supplementary Figure 1**.
734 b. SLC refers to 'single-level controller' and MLC refers to 'multi-level controller'.
735 c. Relative basal expression calculated when no IPTG is present and as a percentage of the
736 expression level for the 'on' state (1 mM IPTG).
737 d. Calculated between 'on' and 'off' states for cells grown in 0 and 1 mM IPTG, respectively, and
738 given in calibrated molecules of equivalent fluorescein (MEFL) units.
739 e. From the Hill function fitting of the steady state response functions (**Figure 4A**).
740 f. SNR refers to 'Signal to Noise Ratio'.

Figures and captions

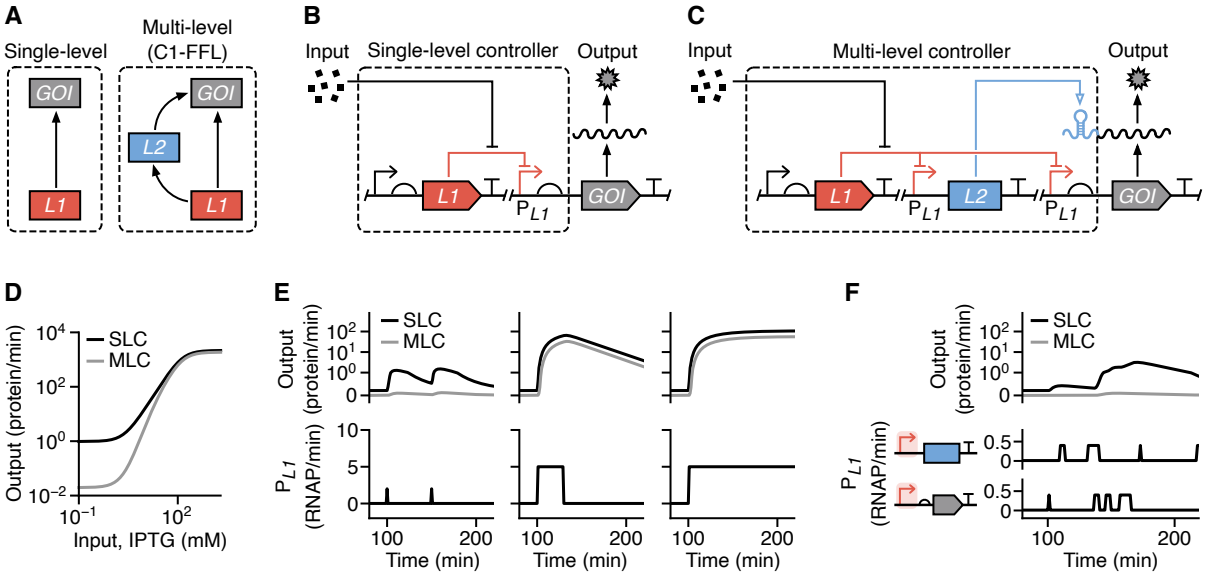


Figure 1: Stringent control of protein expression through multi-level gene regulation. (**A**) Two possible regulatory schemes to control the expression of a gene of interest (GOI): 1. control using a single regulator (*L1*), and 2. multi-level control using two separate regulators (*L1* and *L2*) connected in the form of a coherent type 1 feed-forward loop (C1-FFL). (**B**) Schematic of a genetic implementation of a single-level controller (SLC) that uses only transcriptional (red lines) regulation. An input (e.g. small molecule) modulates activity of the P_{L1} promoter and production of the GOI. (**C**) Schematic of a genetic implementation of a multi-level controller (MLC) that uses both transcriptional (red lines) and translational (blue line) regulation. An input (e.g. small molecule) modulates activity of the two P_{L1} promoters and an internal *L2* regulator activates the translation of GOI transcripts to finally produce the output protein. (**D**) Steady state response functions from mathematical models of the SLC and MLC. (**E**) Dynamic model simulations of the SLC and MLC and their response to different forms of temporal input (left to right): delta functions (P_{L1} activity = 2 RNAP/min for 1 min at 100 min and 150 min), a pulse (P_{L1} activity = 5 RNAP/min from 100–130 min), and a step function (P_{L1} activity = 5 RNAP/min from 100 min onwards). The activity of both P_{L1} promoters in the MLC is considered identical. (**F**) Dynamic model simulations of the SLC and MLC showing suppression of intrinsic promoter noise by the MLC. The two identical P_{L1} promoters for the *L2* regulator and GOI are separately driven by independent and biologically realistic bursty transcriptional activity profiles (**Methods**).

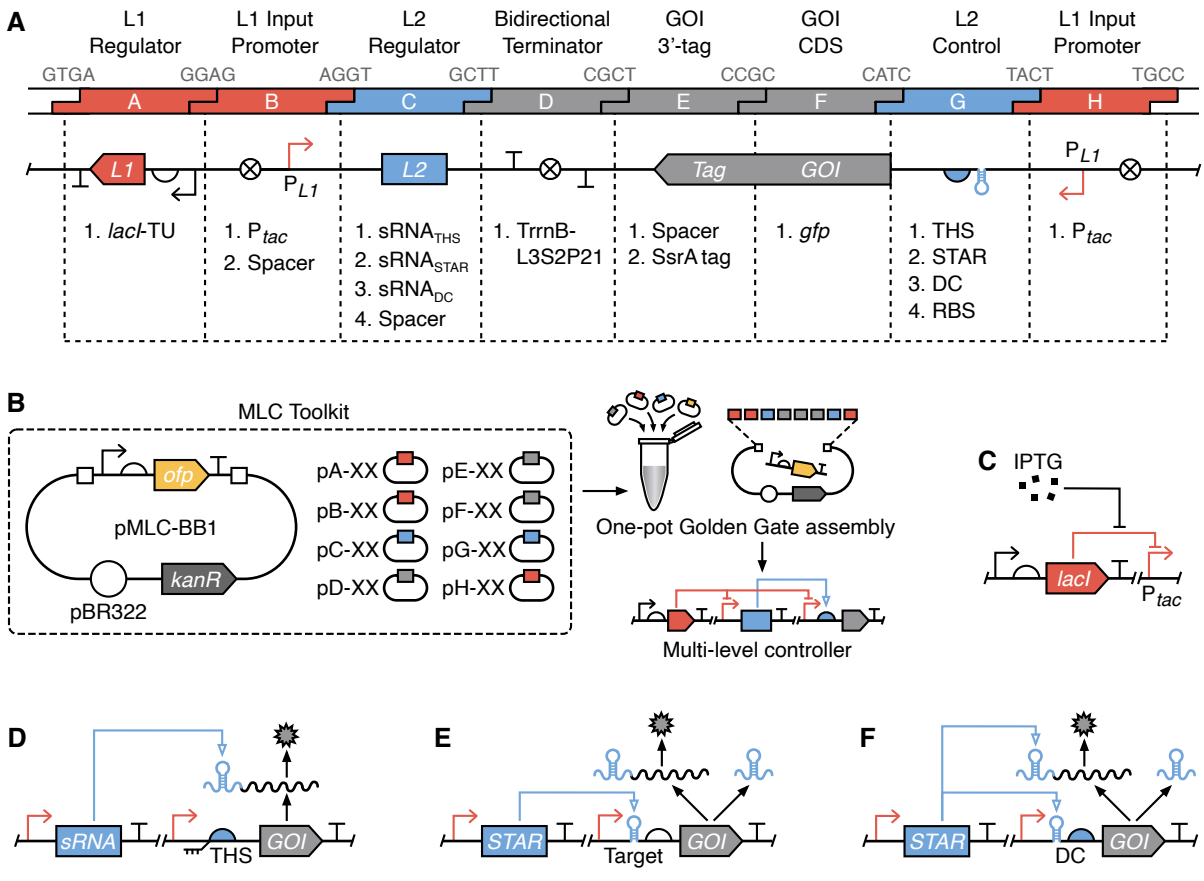
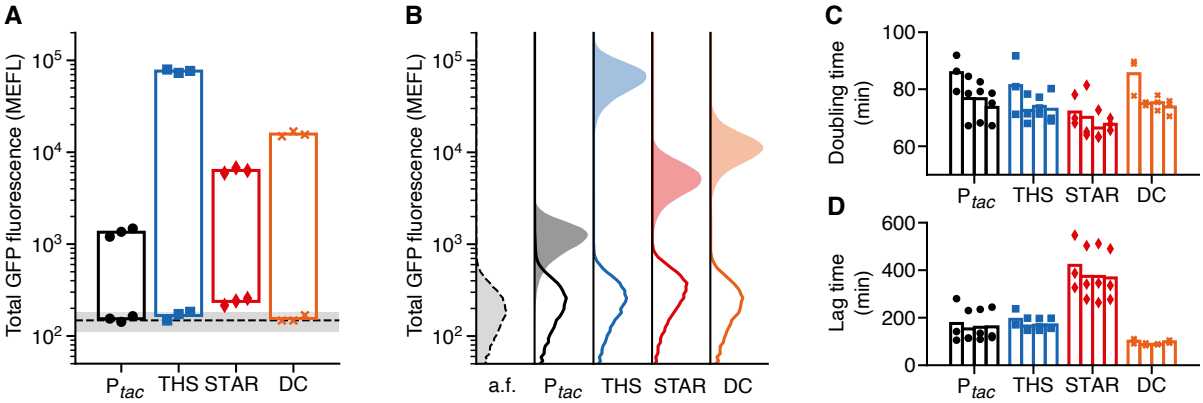


Figure 2: Combinatorial assembly of gene expression controllers. (**A**) Summary of the 8-part genetic template used to allow for systematic exploration of direct and multi-level gene regulation. The 4 bp overhangs used for Golden Gate assembly are shown in grey at their respective junctions. Available genetic elements are listed below each corresponding part type (A–H). (**B**) The MLC toolkit contains a set of plasmids that can be combined using Golden Gate assembly to create a variety of direct and multi-level controllers (**Supplementary Figure 3**). (**C**) The *lacI* transcription factor responsive to IPTG used for level 1 (*L1*) transcriptional regulatory control. (**D**) Toehold switch (THS) translational regulator used for level 2 (*L2*) control. (**E**) Small transcription activating RNA (STAR) transcriptional regulator used for *L2* control. (**F**) Dual control (DC) transcriptional and translational regulator used for *L2* control.



773

774 **Figure 3: Performance comparison of single- and multi-level controllers *in vivo*.** (**A**)
775 Total GFP fluorescence for 'off' and 'on' input states (0 and 1 mM IPTG, respectively). Points
776 show the three biological replicates for each controller and condition (black circles, P_{tac}; blue
777 squares, THS; red diamonds, STAR; orange crosses, DC). Black dashed line denotes the
778 mean fluorescence of cell autofluorescence (a.f.) controls containing no plasmid with grey
779 shaded region showing ± 1 standard deviation of 11 biological replicates. Fluorescence given
780 in calibrated molecules of equivalent fluorescein (MEFL) units. (**B**) Flow cytometry
781 distributions of total GFP fluorescence for 'off' (line) and 'on' (shaded) input states. Cell
782 autofluorescence (a.f.) controls containing no controller are shown by black dashed line and
783 light grey filled distributions. (**C**) Doubling time of cells harbouring direct and multi-level
784 controllers for varying concentrations of IPTG (bars left to right for each design: 0, 0.1, 1, 10
785 mM IPTG). (**D**) Lag time calculated as the time to reach an OD_{600} = 0.15 after inoculation of
786 cells harbouring controllers for varying concentrations of IPTG (bars left to right for each
787 design: 0, 0.1, 1, 10 mM IPTG).

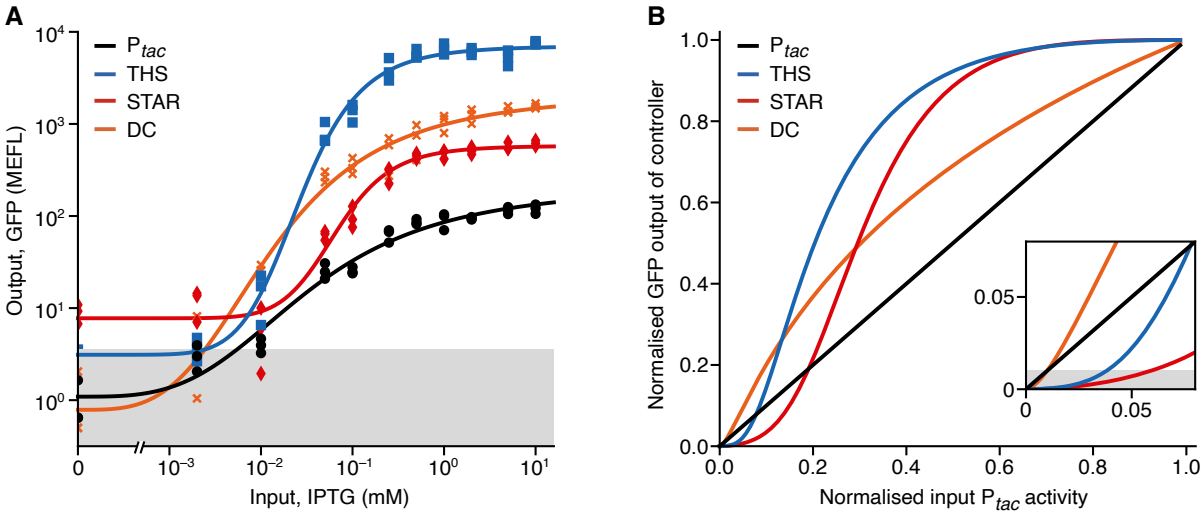


Figure 4: Response functions of single- and multi-level controllers *in vivo*. (**A**) Steady state response functions of the controllers showing output GFP fluorescence (corrected for cell autofluorescence) for varying input IPTG concentrations (0, 0.002, 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 5, 10 mM). Points show the three biological replicates for each controller and condition (black circles, P_{tac}; blue squares, THS; red diamonds, STAR; orange crosses, DC). Grey shaded region shows the standard deviation of cellular GFP autofluorescence from 11 biological replicates. (**B**) Comparison of how normalised GFP output (as a fraction of the maximum GFP fluorescence) varies in response to changes in the normalised transcriptional activity of P_{tac} (as a fraction of its maximum activity). Multi-level regulation can lead to the suppression or amplification of the output GFP production rate compared to direct transcriptional regulation (i.e. a specific multi-level controller's line falls below or above the diagonal, respectively). Insert shows zoomed area and grey shaded region denotes a GFP output level of 1% for the controller.

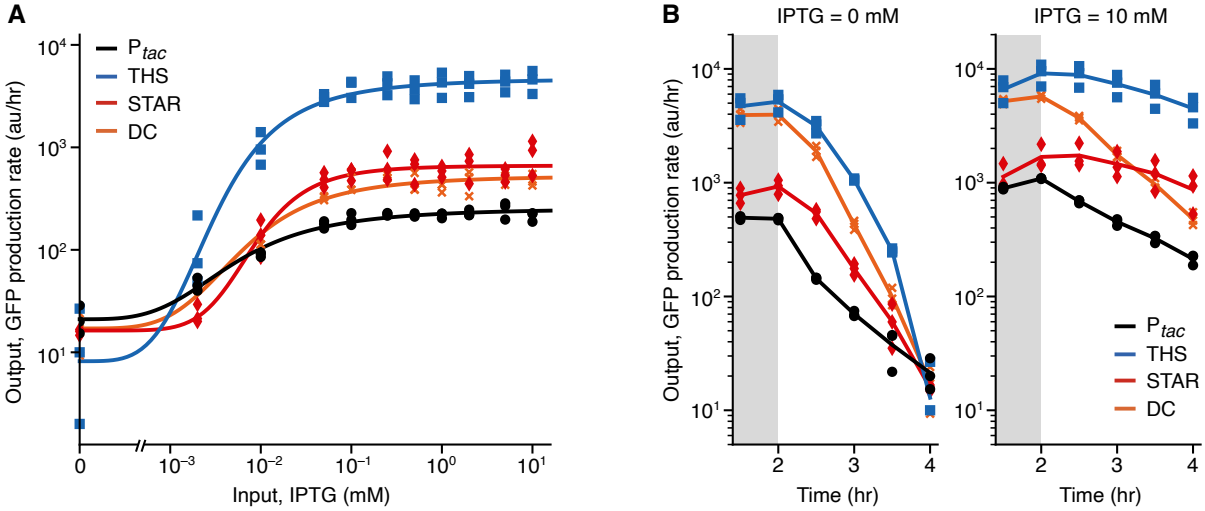


Figure 5: Performance of single- and multi-level controllers in a cell-free expression system. (**A**) Response functions of the controllers showing output GFP production rates in arbitrary fluorescence units per hour (au/hr) at 4 hours after the start of the cell-free reaction for varying input IPTG concentrations (0, 0.002, 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 5, 10 mM). Points show the three biological replicates for each controller and condition (black circles, P_{tac}; blue squares, THS; red diamonds, STAR; orange crosses, DC) (**B**) Output GFP production rate of the controllers over time since the start of the reaction. Time courses shown for controllers in an 'off' (0 mM IPTG; left) and 'on' (10 mM IPTG; right) state. GFP production rates at each time point calculated as an average GFP production rate over the previous 1.5 hours (**Methods**).