Predicting experimental designs leading to rewiring of transcription program and evolution of anticipatory regulation in *E. coli*.

Anjali Mahilkar, Akshat Mall, Supreet Saini*

Department of Chemical Engineering, Indian Institute of Technology Bombay, India 400076

* Corresponding author (saini@che.iitb.ac.in, 091 22 2576 7216)

Abstract.

Environmental cues in an ecological niche are temporal. In response to these temporal cues, bacteria have been known to exhibit learning or conditioning, whereby they trigger response to a yet to appear cue, anticipating its actual arrival in the near future. Such an anticipatory response in known to enhance Darwinian fitness, and hence, is likely an important feature in the regulatory networks in microorganisms. However, the conditions under which an anticipatory response optimizes cellular fitness are not known. Nor has evolution of anticipatory regulation in laboratory conditions been experimentally demonstrated. In this work, we develop a quantitative model to study response of a population to two temporal environmental cues, and present the key variables in cellular physiology associated with response to the cues whose modulation is likely to lead to evolution of anticipatory regulatory response. We predict experimental conditions, which are likely to lead to demonstration of rewiring of regulation, and evolution of anticipatory response in bacterial populations. Using inputs from the modeling results, we evolve *E. coli* in alternating environments of the pentose sugar rhamnose and paraquat, which induces oxidative stress. We demonstrate that growth in this cyclical environment leads to evolution of anticipatory regulation. Thus, we argue that in niches where environmental stimuli have a cyclical nature, conditioning evolves in a population as an adaptive response.

Keywords: Adaptation, conditioning, *E. coli*, laboratory evolution

Introduction.

Conditioning as a concept was first introduced by the classical experiment by Pavlov[1]. In this experiment, Pavlov conditioned dogs to associate condition 1 (whistle) with condition 2 (availability of food). Since then a number of examples of experimentally evolved conditioning in higher animals have been reported[2-4]. The logic associated with evolution of conditioning seems to be that associating a response with two signals (rather than one) increases the chances of launching an appropriate response. For example, instead of just responding to sight of food, if a dog responds to sound of the bell, the readiness to respond to food availability increases. This presumably increases fitness of an organism.

At the other level, limited work has been done with studying evolution of conditioning or anticipatory regulation in microorganisms. In the context of single-celled organisms, conditioning takes on a slightly different context. For bacteria like *E. coli*, conditioning means exhibition of behavior where, when the bacterium is repeatedly exposed to signals S1 and S2 in a temporal order, the introduction of S1 induces a response associated with S2. This is because, after repeated rounds of training, the bacterium "learns" that presence of S1 indicates an impending arrival of S2. The preemptive response to S2 enhances fitness.

Conditioning, at an experimental level, has been explicitly studied in specific ecological contexts. In two reports, *E. coli*, when entering a host is demonstrated to exhibit conditioned response. In the first such example, it was demonstrated, that S1, upon entering body, is an increase in temperature to (37 deg C), followed by non-availability of oxygen (S2)[5, 6]. This temporal order of S1 and S2, when experienced by the bacterium over an evolutionary timescale, causes exposure to S1 to launch a partial, presumably preemptive response to S2. This partial response to S2, upon exposure to S1, increases fitness when S2 is eventually introduced in the environment. The same pattern of response was observed in *E. coli*, when analyzed in the context of temporal hierarchy of sugars available to the bacterium in the small intestine. In this case, S1 was availability of sugar lactose, and S2 maltose[7, 8].

Conditioning, or preemptive gene regulation has also been observed in *Mycobacterium*, where the different two-component systems in the bacterium are so wired, so that activation of one, leads to partial activation of the downstream TCSs too[9]. This, again, the authors argue, help the bacterium anticipate and respond to upcoming environmental changes.

While these are the only studies which refer to conditioning directly, anticipatory regulation is a phenomenon much more ubiquitous than explicitly discussed in literature. For example, in *Salmonella enterica*, a successful infection of the host is contingent on hierarchical expression of a number of virulence factors[10]. This sequential activation is facilitated by conditioning where expression of virulence factor 1 (flagella) is known to partially activate expression of virulence factor 2 (a pathogenicity island-encoded type 3 secretion system). Similarly, in the phase when factor 2 is active, the genetic regulatory network is so wired that expression of virulence factor 3 (type I fimbriae) is partially on. This anticipatory activation is thought to help the bacterium transition from one phase of the infection cycle to the other faster. In a similar context, several other examples of conditioning in microorganisms exist[11, 12].

On the other hand, however, no experimental evolution study has addressed the question of the conditions under which conditioning can evolve in a microbial population. As a result, despite being a ubiquitous phenomenon, we have little understanding how conditioning can evolve. In this work, we address this outstanding question. Given two environmental stimuli, S1 and S2, which a microbial population experiences in a temporal fashion, when would the population “learn” to respond to S2, when only exposed to the signal S1? Towards this end, we develop a simple mathematical model, and identify the key variables in the experiment, which would lead to evolution of conditioning in an experimental set up.

Our results show that conditioning emerges as the optimal regulatory strategy as an interplay between three variables: the degradation rate associated with the target protein T2, the maximum benefit

conferred by protein T2, and the relative timing of signal S2 vs. that of S1. In such a context, we comment on the experimental conditions, which are likely to increase chances of emergence of conditioning as a regulatory response. We speculate that in a cyclical daily environment faced by microorganisms in their ecological niche, conditioning is likely a widely observed phenomenon. To test the predictions from our modeling exercise, we evolved *E. coli* in two stimuli (where S1 is the presence of the sugar rhamnose, and S2 is the presence of oxidative stress, paraquat). Evolving *E. coli* in these two alternating conditions for ~850 generations, we demonstrate that cells evolve to anticipate presence of paraquat, when grown in rhamnose. This confers an adaptive benefit to the cells, in view of the anticipated exposure to paraquat. These results demonstrate that conditioning between physiological response to two distinct stimuli can evolve in laboratory conditions in a fairly rapidly.

Methods.

Model system description.

To study the question of interest we simulate the following system. A microorganism experiences an environmental signal S1. Upon sensing this signal, it modulates gene expression. As part of this response, a transcription factor R1 gets activated and triggers expression of a target protein T1. The protein T1 is assumed to serve a physiological function, which helps the bacterium survive in S1. The signal S1 is assumed to be present for time *t1*. After this time, S1 is replaced by another signal S2. Cellular response to S2 is activation of a transcription factor R2, which then positively controls expression of the target protein T2, which helps the cell survive and replicate in R2. The signal S2 persists for time *t2* (**Figure 1**).

Initially, it is assumed that there is no regulatory crosstalk between the two signal responses. That is, activated R1 is responsible for response to S1 (controlling expression of T1), and activated R2 is responsible for response to S2 (controlling expression of T2). We call this design, a no-conditioning design. In such a setting, in order to maximize fitness, the bacterium will pick mutations, which change the biochemical interactions between the DNA, Proteins, etc. so that the fitness of the bacterium is enhanced. In such a setting, we solve for parameter values, which maximize fitness.

On the other hand, in the conditioning design, new transcriptional interactions could evolve. In such a scenario, the promoter of T2, or the regulator R1 could acquire mutations, which facilitate regulatory control of T2 by activated R1. In this case, expression of T2 is controlled by, activated R1 as well as activated R2. We solve for the values of these biochemical interaction that maximize fitness in this regulatory design.

In particular, we are interested in identification of parameter values such that the fitness conferred by design II (conditioning) exceeds that by design I.

The equations dictating the synthesis and degradation rate of proteins in the cell, upon sensing signal S1 or S2 can be written as below.

$$R1 + [s1] \leftrightarrow R1^*$$

and,

$$R2 + [s2] \leftrightarrow R2^*$$

is represented as:

$$k_{f1}[R1][s1] \leftrightarrow k_{r1}[R1^*]$$

and

$$k_{f2}[R2][s2] \leftrightarrow k_{r2}[R2^*]$$

(1)

The dynamics of target proteins T1 and T2 is represented as the following equations. *R1** is the transcription factor responsible for expression of T1, *Km* is the Michaelis-Menten kinetic constant for T1 production, *b* is the maximum production rate of T1, and k_d is the degradation rate of T1 in the cell. In our analysis, we assume that expression of R1 is not contingent on the presence or absence of signal S1 (but its activation is).

$$\frac{dT1}{dt} = b_0 \frac{R1^*}{R1^* + K_{m0}} - k_{d1} T1$$

(2)

Similarly, we write the equation for T2 in the cell, in response to signal S2. In this equation, we assume that expression of T2 is under the control of R2 exclusively. Hence, there is no conditioning in the system.

$$\frac{dT2}{dt} = b_2 \frac{R2^*}{R2^* + K_{m2}} - k_{d2}\, T2 \tag{3}$$

However, in the case when conditioning controls gene expression, the amount of T2 in the cell can be quantified as:

$$\frac{dT2}{dt} = b_1 \frac{R1^*}{R1^* + K_{m12}} + b_2 \frac{R2^*}{R2^* + K_{m2}} - k_{d2}\, T2 \tag{4}$$

Where,

$$K_m = \frac{k_off}{k_on}$$

for each transcription factor-DNA interaction. In this manner, given the biochemical parameters of a network, we can quantify the dynamics of protein expression in a cell.

Cost-benefit framework.

From the time-course data of protein expression, we use a cost-benefit model to understand the cellular fitness, as a result of the gene expression dynamics [13]. Such a framework has been used in a number of studies to better understand cellular behavior and physiology [14-16]. In this formulation, the benefit conferred to the cell by the target protein T2 is given by the following expression:

$$Benefit, b = b_{max} \frac{[T2]}{[T2] + K_{mT2}} \tag{4}$$

In addition, protein production is known to have an energetic cost associated with it [17]. This cost is a linear function of the amount of protein being synthesized. Hence, the cost component associated with synthesis of protein T2 can be represented as:

$$Cost, c = c_o\ [T2] \tag{5}$$

The overall fitness of a cell is simply the difference of the benefit and cost associated with making T2. Hence, fitness is given by:

$$Fitness, f = b - c \tag{6}$$

Given a set of network topology and the biochemical parameters associated with it, this framework allows us to compute the benefit of production of T2 and track this quantity with time. In the context of this work, we compute the fitness conferred by T2 in the time window [t1, t2] and the total cost associated with synthesis of protein T2 in the window [*t1*, *t2*]. Note that synthesis of T2 in the regulatory design with conditioning starts before time *t1*.

Solving for optimal fitness.

The equations represented in equations (1-6) have biochemical constants appear as equation parameters. Each parameter represents specific cellular processes. For instance, Km represents the interaction dynamics between the transcription factor and DNA. The parameter b represents the maximal rate of transcription from the promoter of gene *t2*.

In this work, we assume that the parameters associated with changing the regulatory patterns of gene expression (like, Km and b) are evolvable on a time scale much faster than the time needed to change parameters associated with protein property (like, enzyme activity, degradation rate) [18]. This has been

shown to be experimentally true in several studies [19-21]. As a result, for the purpose of our work, we let the parameters, which correspond to regulatory timing and level evolve, while keeping the others constant. Moreover, the parameters associated with these biochemical functions cannot take arbitrary values. Each of these parameters is constrained by the thermodynamics associated with biological processes. Hence, we let the parameters change within a defined range (given in **Table 1**). The range chosen for each parameter represents the biologically permissible window in which they can take a value.

Given the above, we introduce "mutations" (change parameters) in the network and for each of the two regulatory designs (with and without conditioning, as defined above), find the parameter set which optimizes fitness. In doing so, if the optimal fitness in the regulatory topology associated with the case of conditioning is more than the optimal fitness in the regulatory topology where conditioning is not taken into account, we note that conditioning is a better adaptive response than non-conditioning. Moreover, in such a scenario, should we design an experiment with appropriate parameter values, we are likely to observe evolution of conditioning. We note that in the evolution scheme, when conditioning is not the optimal response, $b_{12} = 0$ should be optimal design in the regulatory topology which permits conditioning to evolve.

All simulations were performed in Matlab 7.0.

Media and reagents

M9 glycerol medium contained 0.2% glycerol unless otherwise stated. M9 salts, trace elements and casamino acids were prepared in concentrated stocks. Stock of 5X M9 salts solution consisted of 64 g/L $Na_2HPO_4.7H_2O$, 15 g/L KH_2PO_4, 2.5 g/L NaCl, and 5 g/L NH_4Cl dissolved in Milli-Q filtered water. Casamino acids were prepared as 10x solution. $MgSO_4$ and $CaCl_2$ were prepared at 1M stock solutions each. All stock solutions were sterilized by autoclaving. The stocks of rhamnose and paraquat was prepared at 20% and 5mM respectively and sterilized by filtration through 0.22 μm filters.

Laboratory evolution experiment

The ancestral *E. coli* strain used for this study is K12 MG1655 (ATCC 47076). Adaptive evolution experiments were performed by serial dilutions every 12 hours in M9 medium with appropriate stimulus. All cultures were grown at 37°C and 250rpm unless otherwise stated. Every 12 hours, growing cultures of bacteria in triplicates were diluted 1:100 in fresh M9 medium with appropriate stimulus yielding ~ 6-7 generations every 12 hours. The evolution experiment was carried out for a total of 850 generations.

Conditions for evolution of anticipatory regulation

We used rhamnose and paraquat at a concentration of 0.2% and 40 μM respectively, as the two stimuli for evolution of anticipatory regulation. The evolution for anticipatory regulation was carried out for three replicate lines serially diluted in alternating conditions of rhamnose and paraquat every 12 hours as described above. For controls, we evolved lines in triplicates in either 0.2% rhamnose or 40uM paraquat. The control experiment were both the stimuli were added together would not grow beyond four such serial dilutions, hence was dropped out of the experiment. Control lines where neither of the two stimuli were present were also evolved.

Analysis of evolved lines for anticipatory regulation

All evolved lines and the ancestor were revived from freezer stocks from the deep freezer into 2ml LB and incubated for 12-14 hours at 37ºC. The saturated cultures were then sub-cultured 1:100 in 2ml M9 glucose media for 12 hours. All lines were then sub-cultured into fresh M9 medium with 0.2% rhamnose as the first inducer. As a control, all the lines were also sub-cultured into M9 medium without rhamnose. These were grown for another 12 hours before we test them for their performance in paraquat. To test this, all lines were sub-cultured 1:100 into 2ml M9 with PQ (40uM) and measured OD every 60 minutes to plot the growth curves.

Results.

A framework to identify when conditioning evolves in transcription networks.

In this work, we ask the following question: in what conditions does conditioning become an adaptive response? The question is best tested in bacteria, where evolutionary experiments can be carried out in relatively short times, and rewiring of regulatory networks can occur in a few hundred generations. To test this question, we simulate a scenario where a bacterium is responding to two cues encountered in temporal succession. The response to cue A is expression of target protein T1 and the response to cue B is expression of target protein T2 (**Figure 1A**). As a starting point, we assume that expression of T1 and T2 (under appropriate conditions) is triggered by transcription factors R1 and R2, respectively. Under what conditions, however, is this regulatory design rewired such that transcription factor R1, in addition to activating expression of T1, also controls expression of T2 (**Figure 1B**)? In other words, under what conditions associated with the environments or properties of the proteins involved does the design represented by Figure 1B confer a greater fitness as compared to one in Figure 1A?

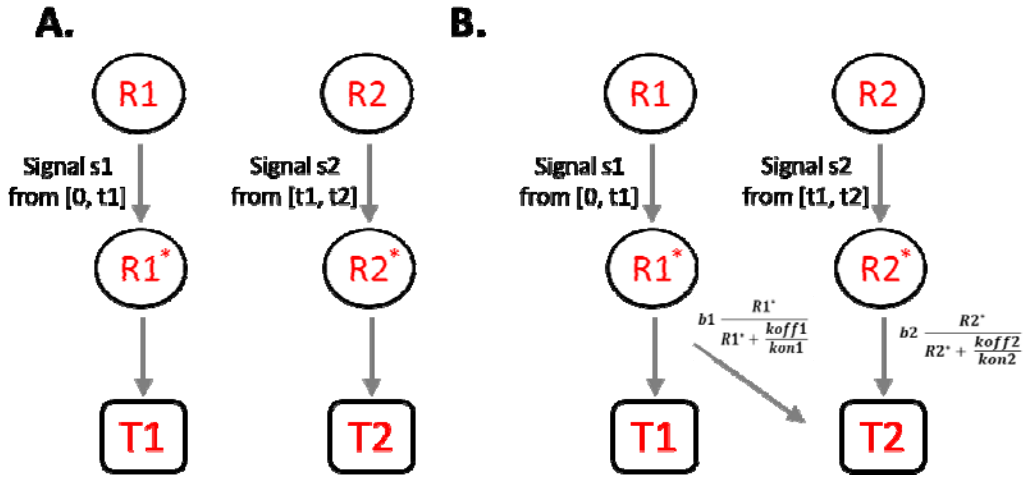


Figure 1. Regulatory designs without (A) and with (B) conditioning. In the prevailing environment, signal s1 is present from [0, t1] and signal s2 is present from [t1, t2]. In the no-conditioning design, activated regulator R1* controls expression of target protein T1, and R2* controls expression of T2. In the regulatory design with conditioning, R1*, in addition to controlling expression of T1, also partially/fully controls expression of target protein T2.

Conditioning appears as an adaptive response for intermediate values of degradation rates of the protein T2.

We simulate the networks in Figure 1A and 1B for a set of parameter values, chosen from a window typically representing each biochemical interaction in a cell [22-28]. We focus on control of transcription as the most flexible parameters of genetic rewiring of networks, as promoter mutations and evolution is thought to be accomplished much quicker than protein property changes (like, enzyme activity) [18]. For each of the two designs, we identify, for a given value of *kd* (degradation rate constant for protein T2), the design which leads to a maximal fitness.

Towards this end, we note that for a given set of fixed parameters, maximal fitness is obtained via three distinct regulatory logics, depending on the degradation rate of the protein T2 (**Figure 2**). For small values of the *kd* of T2, the maximal fitness is achieved when control of T2 is entirely under the regulator R1. That is, R2 does not play any role in controlling expression of T2. This can be understood as follows. Expression of T2 by R1 leads to T2 production even before signal S2 is present in the environment. As a result, the T2 so produced does not confer any advantage to the cell. Thus, this preemptive production of T2 does not give any benefit to the cell, but does come at an additional cost associated with protein production in the cell. This cost, however, is offset by the enhanced T2 levels inside the cell (and consequently the benefit conferred to the cell), from the time when signal B is introduced to the system. When T2 is stable, the steady state T2 in the cell is high, and as a result, the fitness of the cell at the moment signal B is introduced to the system is also high.

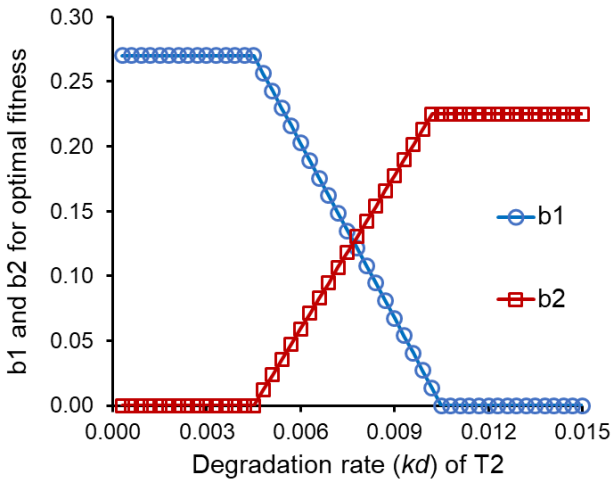


Figure 2. For intermediate values of the degradation rate *kD* for target protein, T2, conditioning is the optimal response. For low values of the degradation constant, expression of T2 controlled entirely by T1 (i.e. b2 = 0) is the optimal response. For high degradation rates, optimal fitness is obtained when expression of T2 is controlled by R2 (i.e. b1 = 0). [b1 is the maximal promoter strength when R1* triggers expression from promoter of T2; b2 is the maximal promoter strength when R2* triggers expression from promoter of T2].

At the other extreme, when the degradation rate constant associated with T2 is high, the regulatory design that optimizes cellular fitness is one where control of expression of T2 is not by R1, but is instead entirely by R2. In such a context, the energetic cost of T2 production in the "futile" time-period [0, *t1*] prior to introduction of signal B, because of the high protein turnover, is high. In such a scenario, the cellular adaptive strategy is control of each target protein by its cognate regulatory partner.

However, at intermediate values of the protein T2 degradation rate constant, for maximal cellular fitness, both b1 and b2 are non-zero. This window of the values of the degradation rate constant of T2 represents the scenario where conditioning is the adaptive response in cellular functioning. Such a set up identifies conditions for a distributed regulatory design for control of gene expression. Interestingly, distributed control strategies are known to be more robust in control systems, and have also been shown to be present in biological regulatory designs [10].

Among the three regimes, we note that conditioning is only observed when b1 and b2 are both non-zero. The other two regimes are solutions to the optimization problem, which lead to different regulatory designs. Thus, we note that one of the key variables which leads to conditioning as the optimal solution is the degradation rate of the target protein.

In the results above, we only compare the optimal solutions of the two regulatory designs. That is, we locate the parameter values of the two regulatory designs (with and without conditioning), which optimize fitness. In such a context, we note that conditioning can outperform the non-conditioning design, under certain parameter regimes. We do not, however, comment on the accessibility of these parameter values via an adaptive path on the fitness landscape. Walks on adaptive landscapes is a multidimensional problem and little is known about structures of actual fitness landscape [29]. In the context of the problem we discuss, how a mutation changes the parameter values is an open problem [30]. Therefore, from a starting parameter set, to what adaptive solution does a population evolve to is an open problem, and has been observed to depend on the details of the problem. Little is known about this facet of evolutionary dynamics [31].

Anticipatory regulation as the optimal strategy at intermediate protein degradation rates can also be arrived at by changing protein-DNA interaction.

The parameter b1 represents the strength of the promoter associated with the gene T1. Essentially, in our setup, expression of T2 is governed by two promoters – one is driven by activated R1 and the other by activated R2. The maximal strengths of these promoters is represented by b1 and b2, respectively. By allowing the parameters b1 and b2 to evolve, we allow the nature of the promoter to change. This change, essentially, is independent of the transcription factor-promoter binding. Another scheme to incorporate changes in regulatory design would be to not change the maximum strength of the promoter, but change the binding affinity of the transcription factor with the promoter region of the DNA. In the context of the model, we change the parameter k_on, which represents the affinity of the DNA for the transcription factor.

Similar to the results obtained above, we note that conditioning could also outperform for intermediate values of the degradation rate of protein T2, if k_on1 or k_on2 is the parameter which is allowed to vary. Also similar with the previous results, while conditioning only outperforms the non-conditioning designs for intermediate values of the degradation rate of protein T2, positive fitness in regulatory designs with conditioning exist for all values of kd. This again suggests that conditioning regulatory design can emerge as the regulatory design which maximizes fitness at intermediate values of degradation rate of protein T2 (**Figure 3**).

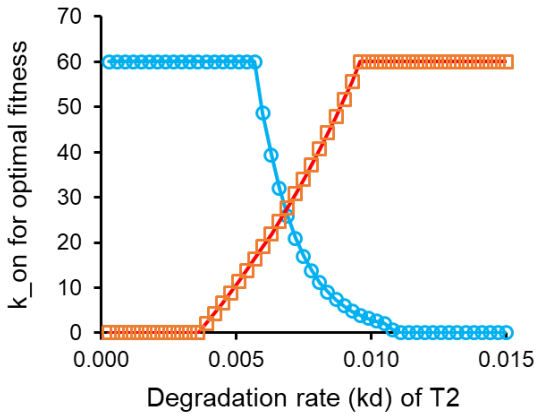


Figure 3. For intermediate values of the degradation rate *kD* for target protein, T2, conditioning is the optimal response. For low values of the degradation constant, expression of T2 controlled entirely by T1 is the optimal response. For high degradation rates, optimal fitness is obtained when expression of T2 is controlled by R2 (i.e. k_on for b1 = 0). Blue circles represent the k_on for interaction between R1* and promoter of T2. Orange squares represent k_on for interaction between R2* and promoter of T2.

This suggests that regulatory structures are extremely flexible in terms of incorporating newer connections (or changing/eliminating the ones existing) to enhance cellular fitness in terms of growth rate. From our work, we aim to show that this statement is true for the simplest regulatory network – one containing a single regulator R and a target T – too. It is likely that as we move towards regulatory topologies of greater complexities, the ability to incorporate newer features in the dynamical responses of networks also increases. These responses could include cell-cell

heterogeneity, memory, speed of response among others. All these properties have shown to have fitness consequences for a cell, in different contexts [16, 28, 32, 33]. In addition, the number of ways in which these responses could be incorporated into the cellular manifestations also increases as the size of the network increases. This facet of networks makes tracking evolutionary dynamics of a system quite daunting – not only in terms of how a design could incorporate a specific feature into cellular response, but also in terms of what additional features (and also future possibilities) are going to get incorporated in the cellular physiology as a fallout of the adaptive response. Predicting the long-term implications of this remains a major challenge for evolutionary biology.

Conditioning, though not optimal, confers a positive fitness to the cell.

Our results in Figure 2 and Figure 3 show that conditioning is not the optimal response for very low and very high values of the degradation rate constant associated with the target protein T2. This is further illustrated in **Figure 4**. Here, ΔF corresponds to the difference between the fitness when conditioning is permitted (for a corresponding (b1, b2) or (kon1, kon2) pair) and the best value of fitness when conditioning is not permitted. Thus, for a high degradation rate (which lies in the region where regulatory design in Figure 1A maximizes fitness), $\Delta F < 0$ for all combinations of (b12, b2) (**Figure 4A**) and (k_on12, k_on2) (**Figure 4B**).

However, for a large parts of the parameter scanned in these heat plots, the fitness when conditioning is permitted, is greater than zero (Figures 4C and 4D). Hence, even though, conditioning might not be the best possible response in these parameter regions, it may still be observed as the prevailing regulatory design. This is particularly so when strength of selection on the target protein is not very high, or in population regimes where drift has a significant role to play.

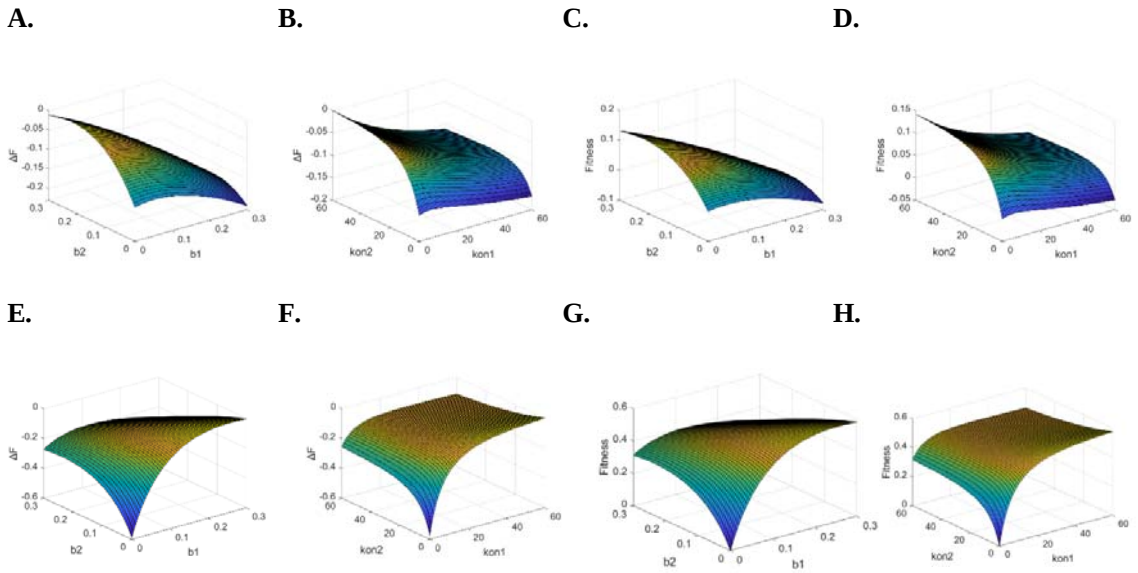


Figure 4. Conditioning yields positive fitness at high protein degradation rates (A-D). When the promoter strength (A) or the DNA-protein interaction (B) is allowed to change, the fitness of conditioning regulatory logic is always less than of a non-conditioning regulatory design in Figure 1A. However, for large parts of the range of parameter values, the absolute fitness conferred in the regulatory design corresponding to conditioning is greater than zero (C-D). **Conditioning yields positive fitness at low protein degradation rates (E-H).** When the promoter strength (E) or the DNA-protein interaction (F) is allowed to change, the fitness of conditioning regulatory logic is always less than the design when R1* controls both T1 and T2 gene expression. However, for large parts of the range of parameter values, the absolute fitness conferred in the regulatory design corresponding to conditioning is greater than zero (G-H). (A-D) are for *kd* (for T2) equal to 0.0144. (E-H) are for *kd* (for T2) equal to 0.0025.

Similar results are also observed when parallel heat plots are analyzed for small values of degradation rate constant of T2. We note that while conditioning might not be the best possible response, it does correspond to non-zero fitness values of the system (**Figure 4E-4H**).

Transcription factor-target data from *E. coli*.

Previous studies have established a link between conditioning and the ecological niche of the bacterium [5-8]. These studies naturally raise the question of whether conditioning as a phenomenon is wide-spread in the genome of the bacterium, or are these examples to be found only rarely in the genome. To explore this at the genome wide scale, we examine if there exists a correlation between

the degradation rates of proteins (targets) and the number of transcription factors controlling their expression. One of the predictions from our study is that should the degradation rate of proteins being in an intermediate region, the target is likely to be controlled by more than one transcription factor. On the other hand, if the degradation rate is high, the target is likely to be controlled by an exclusive transcription factor, which exhibits exclusive control of its expression. Last, if the degradation rate of the target protein is very low, its expression is likely to be controlled by the same transcription factor, which controls the target protein of the previous environment (in the context of conditioning).

However, no database exists for protein degradation rates in *E. coli*, the most well-studied model organism. Moreover, it is known that protein degradation rates are a strong function of cellular and surrounding environments. We do know, however, that cellular protein degradation rates are quite small under normal physiological conditions, often of the order of a day [34]. This protein degradation rate however increases significantly if an environmental stress is placed on the cell, for example, lack of resources [35].

One of the likely consequence of this fact is that cellular regulatory networks are more centralized in certain few regulatory nodes – when *kd* is low. This corresponds to the design where T2 is entirely under the control of R1. When the regulatory targets of all transcription factors is studied, we note that >78% of transcription factors have more than one promoter target. That is, less than a fourth of all transcription factors have an exclusive promoter target (**Figure 5**).

Figure 5. Transcription factors in *E. coli*. Blue represents the number of transcription factors, which control one promoter only. This number is 22% of all transcription factors in *E. coli*. The orange

represents transcription factors, which control more than one promoter (~78%). The mean number of promoters controlled by transcription factors in this category is approx. 10.

Can conditioning be evolved in a laboratory set-up?

Given the above, we speculate the requirements of an experimental design to demonstrate evolution of conditioning in laboratory conditions. Till date, only limited experimental and theoretical studies have focused on examining presence of conditioning in living organisms. The context of conditioning in complex organisms is quite different from how this problem has been approached in microorganisms. While evolution of conditioning in complex organisms/animals is almost trivial, evolution of conditioning in microorganisms (bacteria or yeast) has not been attempted, to the best of our knowledge. Our model lays down conditions, which might facilitate design of experiments towards this goal.

We predict that conditioning is most likely to evolve if the target protein has an intermediate degradation rate. This is perhaps the single most important criteria for conditioning to emerge as an outcome of an evolutionary experiment. In addition to this, the target protein responding to condition/environment 2 should enable a physiological function which enhances cellular fitness considerably. Alternatively, the absence of target protein T2 should cause a significant fitness cost to the cell (**Figure 6A**). In this context, for experimental evolution of conditioning, it makes sense if the design of experiment is such that the signal S2 is a stress, which, in the absence of the appropriate cellular response, has a significant fitness penalty to the cell.

The other variable which dictates the likelihood of conditioning being the optimal response is the relative timing of the two environmental signals. In this context, we note that the greater the relative time for which the signal S2 prevails in the environment, the greater the likelihood of conditioning evolving as the optimal fitness solution (**Figure 6B**). Thus, in the context of design of experimental evolution, the relative timing of the two signals is another key variable which is likely to dictate the evolutionary fate of the population. From the context of bacteria in their ecological niches, cyclical

nature of the environment is not too stringent a requirement to be fulfilled for conditioning to emerge. Organisms in ecological niches are exposed to cyclical daily environments, and hence, selection is oscillates with time.

A. **B.**

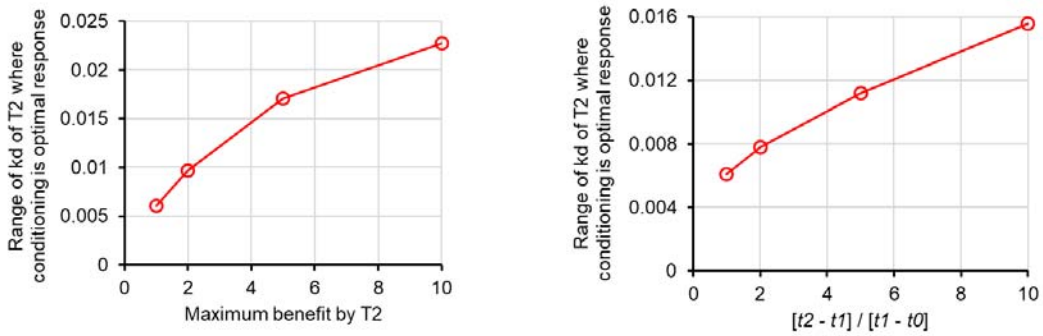


Figure 6. Anticipatory regulation emerges as optimal solution as (A) Maximum benefit conferred by T2 increases, and (B) Relative time of S2 with respect to S1 increases. The y-axis on the plots represents the range of the values of *kd* for protein T2, for which anticipatory regulation is the optimal regulatory design.

However, we realize that these considerations are unlikely to be sufficient for evolution of conditioning with certainty. While several evolutionary experiments have reported convergent evolution at a molecular level (at the gene/operon level), this is hardly a universal theme in experimental evolution [36, 37]. In fact, multiplicity of solutions, in response to a selection pressure, remains one of the lessons from the evolutionary experiments from the last few decades. In such a context, convergent evolution is likely still only reported in a limited number of cases, as compared to widely different solutions that a cell undertakes in parallel experiments, in response to an environmental stress. Hence, our predictions from the model regarding evolution of conditioning (via rewiring the genetic regulatory networks) remain to be experimentally tested. It is possible that even for the same environmental signals S1 and S2, the adaptive solutions in multiple parallel setups could involve different molecular routes towards adaptation. A fraction of these routes, we argue, would involve evolution of conditioning.

Moreover, while we only analyze incorporation of anticipatory regulation via transcription, several other possibilities remain. Gene regulation can be incorporated at via post-transcription, translation, and/or post-translation mechanisms. These regulatory possibilities only extend the number of ways that anticipatory regulation can be built into a regulatory design of cellular networks, should it confer adaptive benefit to the cell.

The other question is that how long is this evolution of conditioning going to take? Does one need to set up an evolutionary experiment of geological timescale, or would it be a rapid process. It depends on the selection pressure we put, and the choice of S1 and S2. There is no reason to doubt that should S1 and S2 be chosen appropriately, conditioning should evolve fairly rapidly. In fact, theory has previously suggested that conditioning in asexual populations could likely evolve in a timescale of the order of a hundred generations [7].

Evolving conditioning in *E. coli*.

From our modeling exercise we note that two key parameters for evolution of conditioning are (a) an intermediate degradation rate of the target T2, and (b) a large cost associated with not launching a rapid response to signal S2. In view of these inputs, we chose a pair of environmental signals, which satisfy these conditions. We chose the pentose sugar rhamnose as signal 1, S1 and an oxidative stress molecule, paraquat as signal 2, S2. Rhamnose utilization is controlled by two transcription factors, RhaS and RhaR[38, 39]. In the presence of rhamnose, RhaR activates the expression from the *rhaSR* promoter[40]. RhaS, on the other hand, activates expression of the rhamnose transporter protein, RhaT, and the catabolic operon, *rhaBAD*[41]. Both, RhaR and RhaS, belong to the AraC/XylS family of transcriptional regulators, and share sequence similarity in the DNA binding regions of the proteins[42-44].

On the other hand, oxidative stress response is largely controlled by SoxRS system in *E. coli*[45]. To the best of our knowledge, the two systems are not linked by any direct transcriptional regulation. A

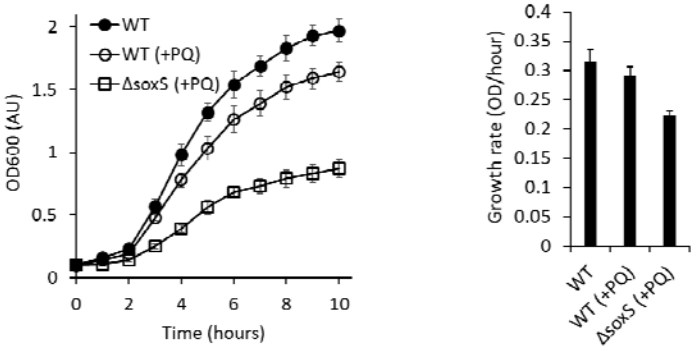
stress of 40 uM paraquat is quite significant for *E. coli*, and introduces a severe growth defect. In a strain lacking *soxRS* the same stress hinders growth considerably (**Figure 7A**). Thus, any anticipated expression of the SoxRS regulon is likely to provide a large adaptive benefit to the individual. With this premise, we performed an evolutionary experiment where the cells were exposed to S1 and S2 alternatively for a total of about 850 generations (See experimental design in **Figure 7B**, and methods section). Alongside, we also performed experiments where populations were exposed to only S1, only S2, or neither S1 or S2.

In such a context, conditioning can be said to have evolved when pre-exposure to rhamnose confers a growth advantage in paraquat. Moreover, this advantage should only be present in the lines which were evolved in an environment with alternating presence of rhamnose and paraquat. As shown in the **Figure 7C**, the populations grown in rhamnose and paraquat alternatively demonstrated a significant growth advantage when transferred to paraquat from rhamnose as compared to when the same lines were transferred to paraquat from a glucose environment. On the other hand, this advantage was absent in the lines evolved in paraquat only environments. This growth advantage is contingent on the populations "seeing" the two stimuli alternatively. Thus, a pre-exposure to rhamnose confers an adaptive advantage in the paraquat environment, to lines which were evolved in the two alternating environments.

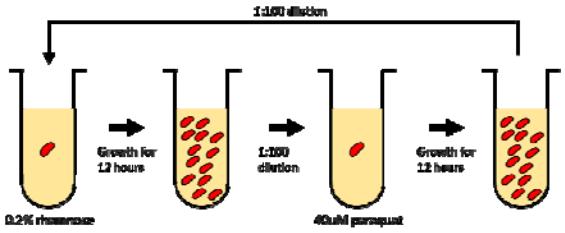
Our data shows that when conditioning-evolved lines were exposed to paraquat after growth in glucose (where rhamnose utilization genes are in the OFF state, the adaptive benefit was no longer present. Thus, demonstrating that the growth advantage is seen exclusively in lines which were exposed to rhamnose and paraquat alternatively, and only when the cells are exposed to rhamnose. This demonstrates evolution of conditioning. While the molecular basis of this adaptive response is not known, we demonstrate phenomenology associated with evolution of conditioning. The growth advantage on transition from rhamnose to paraquat environments is absent in the ancestor and the rhamnose-evolved lines (**Figure 7D**).

559

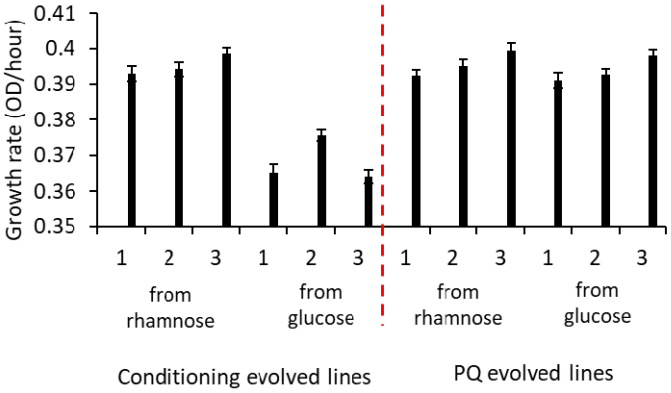
A.


WT
WT (+PQ)
ΔsoxS (+PQ)
OD600 (AU)
Time (hours)
Growth rate (OD/hour)
WT
WT (+PQ)
ΔsoxS (+PQ)

B.


1:100 dilution
Growth for 12 hours
1:100 dilution
Growth for 12 hours
0.2% rhamnose
40uM paraquat

C.


Growth rate (OD/hour)
from rhamnose
from glucose
from rhamnose
from glucose
Conditioning evolved lines
PQ evolved lines

D.

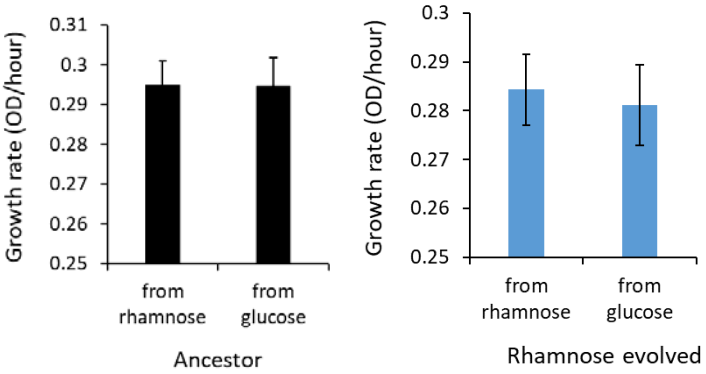

Growth rate (OD/hour)
from rhamnose
from glucose
Ancestor
Growth rate (OD/hour)
from rhamnose
from glucose
Rhamnose evolved

Figure 7. (A) Growth in stressor in WT and soxSR mutant. Wild type *E. coli* and *soxRS* mutant was grown in presence of paraquat (left). On right, the growth rate of the wild type and the mutant in each condition. **(B) Experimental Design for evolution of conditioning.** *E. coli* was grown in 0.2% rhamnose, and then transferred to 40 μM paraquat. Growth in each environment was for 12 hours. The cycle was repeated for a total growth duration of ~850 generations. **Evolution of conditioning. (C) Conditioning improves growth advantage.** When conditioning-evolved lines are transferred to M9 media with paraquat, they exhibit a higher growth when coming from a rhamnose environment, as compared to when they are introduced from a glucose environment. This advantage is absent in the paraquat-evolved lines. The numbers 1, 2, 3 refer to three independent lines for the conditioning-evolved and paraquat-evolved environments. **(D) There is no growth advantage when the ancestor or the rhamnose-evolved lines are transitioned to a paraquat environment, after growth in rhamnose.**

Discussion.

Transcription networks evolve faster than other biological networks[46]. This happens via various such as gene duplication, horizontal gene transfer, and via point mutations in the transcription factor and/or target promoter[47-49]. Changes in the transcription program/control of a cell which are adaptive in nature get selected. In this sense, appropriate control of gene expression becomes an important challenge for the cell.

How promoters tune accuracy of gene regulation is an open question. Theory suggests that gene regulation evolved to minimize noise[50]. However, noise is also an integral part of the gene regulation process[51]. However, how this noise changes in the face of selection is largely unknown[52]. In a recent work, it was demonstrated that promoters in *E. coli* exhibit low noise by default, and suggested that noise in promoter dynamics must have been selected for[53]. This noise in the transcriptional process, the authors argue, generates phenotypic heterogeneity in the population, allowing for evolution of transcriptional networks in the cell. In this sense, noise in transcription aids adaptation.

Moreover, promoter regions are extremely malleable. In another recent study, it was demonstrated that a majority of randomly selected DNA sequences (which were not promoters) could evolve into a promoter via acquisition of only a single nucleotide mutation[54]. It was hypothesized that, similarly, regulation could evolve by acquisition of a few mutations in a promoter region. Laboratory experiments have further demonstrated that transcription factors, particularly global regulators are particular targets of modulating gene expression patterns, and hence physiology[55, 56]. How gene regulatory networks impact physiology has also been studied using ecological isolates[57-59]. Thus, transcriptional reprogramming is evolutionarily easily achieved, provided selection is appropriately strong. Sequence evolution of promoters, in this context, is easily permitted by the cellular machinery[60]. Thus via small adjustments in the cis-acting regulatory regions, large changes in the gene expression pattern can be brought about. The mechanistic and theoretical aspects of how these changes are brought about has also been investigated recently[61, 62].

Transcription is an inherently flexible and dynamic process. This is illustrated by existence of multiple DNA-Transcription Factor variants in a population or within a single individual[63-65]. But perhaps this should not be surprising. Massive transcriptional reprogramming was observed in *E. coli* strains evolved for 20,000 generations[66].

In the context of our study, rhamnose utilization is facilitated by two transcription factors in *E. coli*, RhaS and RhaR[67-69]. Both bind upstream of the coding sequence of the genes essential for rhamnose utilization as dimers[67]. This fact and that the coding regions of one of the transcription factors (RhaR) and the transporter (RhaT) overlap make the rhamnose utilization regulon unique in *E. coli*[41, 67, 70]. How and in what context this transcriptional interference, and presence of two regulators facilitates survival is not known. In view of our work, the presence of two transcription factors provides a greater opportunity for evolution of transcriptional reprogramming.

While the molecular basis of the evolution of conditioning in our experimental setup is as yet unknown, we identify and demonstrate that under appropriate conditions rewiring of evolutionary networks takes place readily. The rhamnose utilization regulators, RhaR and RhaS, both are known to bind ~30 base pair regions in the promoter sequence of *rhaSR*, *rhaT*, and *rhaBAD* promoters[67, 71]. No other regulatory targets are known for these two transcription factors. As a result, the consensus binding site for them is unknown. However, experimental evolution with bacteria has shown that promoter regions can evolve from random DNA sequences in an extremely short period of time[54]. Thus, it is very likely that transcription factor binding sites are also introduced in promoter sequences, should selection for their presence be strong. In the context of the stress response, oxidative stress is primarily handled by the *soxRS* regulon in *E. coli*. This regulon comprises of more than a 100 downstream genes, including those related to transport, metabolism, pH control[72-75]. The control of this regulon is primarily under the control of *soxRS* locus, although the control of their expression is distributed with the *marRAB* and *rob* regulators in the cell[45, 76]. The precise mutations which enable transcriptional rewiring leading to evolution of conditioning can be investigated via whole genome

sequencing. Moreover, there is likely to be multiple mutations which help the cell confer a response to the challenge posed in our experimental set up[77]. In a broader perspective, from the context of ecological niches of bacteria, where several factors of the environment are periodic in nature (e.g 24 hour Temperature, light cycles), conditioning is extremely likely to have shaped the genetic regulatory patterns in the organisms.

Acknowledgements.

The authors thank the anonymous reviewers for their comments.

Conflict of interest.

The authors declare that they have no conflict of interest.

Table 1. Range of values of the parameters used in this study.

Parameter	Value/Range
b (for all promoters)	0 to 0.3
k_on (for all DNA-protein interactions)	0 to 60
k_off (for all DNA-protein interactions)	0.3
kd (degradation rate constant)	0 to 0.015
Signal (S1 or S2)	0 or 1
b_max (maximum benefit)	1.5
Km (half-maximum benefit)	15
co (cost of synthesis of one protein molecule)	0.01

References.

1. Dennis Coon, J.M. Introduction to Psychology: Gateways to Mind and Behavior. (Cengage Learning, 2008).
2. Watson, J.B. & Rayner, R. Conditioned emotional reactions. 1920. *Am Psychol* **55**, 313-317 (2000).
3. Glanzman, D.L. The cellular basis of classical conditioning in Aplysia californica--it's less simple than you think. *Trends Neurosci* **18**, 30-36 (1995).
4. Woods, S.C. & Strubbe, J.H. The psychobiology of meals. *Psychon Bull Rev* **1**, 141-155 (1994).
5. Cooper, T.F. Evolutionary biology: Microbes exploit groundhog day. *Nature* **460**, 181 (2009).
6. Mitchell, A. et al. Adaptive prediction of environmental changes by microorganisms. *Nature* **460**, 220-224 (2009).
7. Tagkopoulos, I., Liu, Y.C. & Tavazoie, S. Predictive behavior within microbial genetic networks. *Science* **320**, 1313-1317 (2008).
8. Baliga, N.S. Systems biology. The scale of prediction. *Science* **320**, 1297-1298 (2008).
9. Agrawal, R., Pandey, A., Rajankar, M.P., Dixit, N.M. & Saini, D.K. The two-component signalling networks of Mycobacterium tuberculosis display extensive cross-talk in vitro. *Biochem J* **469**, 121-134 (2015).
10. Saini, S., Slauch, J.M., Aldridge, P.D. & Rao, C.V. Role of cross talk in regulating the dynamic expression of the flagellar Salmonella pathogenicity island 1 and type 1 fimbrial genes. *J Bacteriol* **192**, 5767-5777 (2010).
11. Huang, H. et al. An integrated genomic regulatory network of virulence-related transcriptional factors in Pseudomonas aeruginosa. *Nat Commun* **10**, 2931 (2019).
12. Erhardt, M. & Dersch, P. Regulatory principles governing Salmonella and Yersinia virulence. *Front Microbiol* **6**, 949 (2015).
13. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588-592 (2005).
14. Fritz, G., Walker, N. & Gerland, U. Heterogeneous Timing of Gene Induction as a Regulation Strategy. *J Mol Biol* **431**, 4760-4774 (2019).
15. Roselius, L. et al. Modelling and analysis of a gene-regulatory feed-forward loop with basal expression of the second regulator. *J Theor Biol* **363**, 290-299 (2014).
16. Zaslaver, A., Mayo, A., Ronen, M. & Alon, U. Optimal gene partition into operons correlates with gene functional order. *Phys Biol* **3**, 183-189 (2006).
17. Kafri, M., Metzl-Raz, E., Jona, G. & Barkai, N. The Cost of Protein Production. *Cell Rep* **14**, 22-31 (2016).

18. Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. & Teichmann, S.A. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14**, 283-291 (2004).
19. Rubinstein, M. & de Souza, F.S. Evolution of transcriptional enhancers and animal diversity. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130017 (2013).
20. Wittkopp, P.J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* **13**, 59-69 (2011).
21. Wray, G.A. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**, 206-216 (2007).
22. Mitrophanov, A.Y. & Groisman, E.A. Positive feedback in cellular control systems. *Bioessays* **30**, 542-555 (2008).
23. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat Genet* **31**, 69-73 (2002).
24. Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. & Elowitz, M.B. Gene regulation at the single-cell level. *Science* **307**, 1962-1965 (2005).
25. Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. & Elowitz, M.B. Accurate prediction of gene feedback circuit behavior from component properties. *Mol Syst Biol* **3**, 143 (2007).
26. Sneppen, K., Krishna, S. & Semsey, S. Simplified models of biological networks. *Annu Rev Biophys* **39**, 43-59 (2010).
27. Suel, G.M., Garcia-Ojalvo, J., Liberman, L.M. & Elowitz, M.B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545-550 (2006).
28. Prajapat, M.K., Jain, K., Choudhury, D., Raj, N. & Saini, S. Revisiting demand rules for gene regulation. *Mol Biosyst* **12**, 421-430 (2016).
29. Fragata, I., Blanckaert, A., Dias Louro, M.A., Liberles, D.A. & Bank, C. Evolution in the light of fitness landscape theory. *Trends Ecol Evol* **34**, 69-82 (2019).
30. Brajesh, R.G., Dutta, D. & Saini, S. Distribution of fitness effects of mutations obtained from a simple genetic regulatory network model. *Sci Rep* **9**, 9842 (2019).
31. Blount, Z.D., Lenski, R.E. & Losos, J.B. Contingency and determinism in evolution: Replaying life's tape. *Science* **362** (2018).
32. Lan, G. & Tu, Y. Information processing in bacteria: memory, computation, and statistical physics: a key issues review. *Rep Prog Phys* **79**, 052601 (2016).
33. van Hoek, M. & Hogeweg, P. The effect of stochasticity on the lac operon: an evolutionary perspective. *PLoS Comput Biol* **3**, e111 (2007).
34. Maurizi, M.R. Proteases and protein degradation in Escherichia coli. *Experientia* **48**, 178-201 (1992).
35. Flick, K. & Kaiser, P. Protein degradation and the stress response. *Semin Cell Dev Biol* **23**, 515-522 (2012).

36. Buller, A.R. & Townsend, C.A. Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc Natl Acad Sci U S A* **110**, E653-661 (2013).
37. Dodson, G. & Wlodawer, A. Catalytic triads and their relatives. *Trends Biochem Sci* **23**, 347-352 (1998).
38. Egan, S.M. & Schleif, R.F. DNA-dependent renaturation of an insoluble DNA binding protein. Identification of the RhaS binding site at rhaBAD. *J Mol Biol* **243**, 821-829 (1994).
39. Egan, S.M. & Schleif, R.F. A regulatory cascade in the induction of rhaBAD. *J Mol Biol* **234**, 87-98 (1993).
40. Tobin, J.F. & Schleif, R.F. Transcription from the rha operon psr promoter. *J Mol Biol* **211**, 1-4 (1990).
41. Via, P., Badia, J., Baldoma, L., Obradors, N. & Aguilar, J. Transcriptional regulation of the Escherichia coli rhaT gene. *Microbiology* **142 (Pt 7)**, 1833-1840 (1996).
42. Egan, S.M. Growing repertoire of AraC/XylS activators. *J Bacteriol* **184**, 5529-5532 (2002).
43. Gallegos, M.T., Michan, C. & Ramos, J.L. The XylS/AraC family of regulators. *Nucleic Acids Res* **21**, 807-810 (1993).
44. Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K. & Ramos, J.L. Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev* **61**, 393-410 (1997).
45. Jain, K. & Saini, S. MarRA, SoxSR, and Rob encode a signal dependent regulatory network in Escherichia coli. *Mol Biosyst* **12**, 1901-1912 (2016).
46. Shou, C. et al. Measuring the evolutionary rewiring of biological networks. *PLoS Comput Biol* **7**, e1001050 (2011).
47. Teichmann, S.A. & Babu, M.M. Gene regulatory network growth by duplication. *Nat Genet* **36**, 492-496 (2004).
48. Price, M.N., Dehal, P.S. & Arkin, A.P. Horizontal gene transfer and the evolution of transcriptional regulation in Escherichia coli. *Genome Biol* **9**, R4 (2008).
49. Price, M.N., Dehal, P.S. & Arkin, A.P. Orthologous transcription factors in bacteria have different functions and regulate different genes. *PLoS Comput Biol* **3**, 1739-1750 (2007).
50. Shinar, G., Dekel, E., Tlusty, T. & Alon, U. Rules for biological regulation based on error minimization. *Proc Natl Acad Sci U S A* **103**, 3999-4004 (2006).
51. Golding, I., Paulsson, J., Zawilski, S.M. & Cox, E.C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025-1036 (2005).
52. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216-226 (2008).
53. Wolf, L., Silander, O.K. & van Nimwegen, E. Expression noise facilitates the evolution of gene regulation. *Elife* **4** (2015).

54. Yona, A.H., Alm, E.J. & Gore, J. Random sequences rapidly evolve into de novo promoters. *Nat Commun* **9**, 1530 (2018).
55. Conrad, T.M., Lewis, N.E. & Palsson, B.O. Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol* **7**, 509 (2011).
56. Philippe, N., Crozat, E., Lenski, R.E. & Schneider, D. Evolution of global regulatory networks during a long-term experiment with Escherichia coli. *Bioessays* **29**, 846-860 (2007).
57. King, T., Ishihama, A., Kori, A. & Ferenci, T. A regulatory trade-off as a source of strain variation in the species Escherichia coli. *J Bacteriol* **186**, 5614-5620 (2004).
58. Le Gall, T., Darlu, P., Escobar-Paramo, P., Picard, B. & Denamur, E. Selection-driven transcriptome polymorphism in Escherichia coli/Shigella species. *Genome Res* **15**, 260-268 (2005).
59. Vital, M. et al. Gene expression analysis of E. coli strains provides insights into the role of gene regulation in diversification. *ISME J* **9**, 1130-1140 (2015).
60. Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**, 1-4 (1999).
61. Lagator, M., Paixao, T., Barton, N.H., Bollback, J.P. & Guet, C.C. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *Elife* **6** (2017).
62. Igler, C., Lagator, M., Tkacik, G., Bollback, J.P. & Guet, C.C. Evolutionary potential of transcription factors for gene regulatory rewiring. *Nat Ecol Evol* **2**, 1633-1643 (2018).
63. Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M. & Snyder, M. Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464**, 1187-1191 (2010).
64. Kasowski, M. et al. Variation in transcription factor binding among humans. *Science* **328**, 232-235 (2010).
65. McDaniell, R. et al. Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328**, 235-239 (2010).
66. Cooper, T.F., Remold, S.K., Lenski, R.E. & Schneider, D. Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in Escherichia coli. *PLoS Genet* **4**, e35 (2008).
67. Wickstrum, J.R. & Egan, S.M. Amino acid contacts between sigma 70 domain 4 and the transcription activators RhaS and RhaR. *J Bacteriol* **186**, 6277-6285 (2004).
68. Holcroft, C.C. & Egan, S.M. Interdependence of activation at rhaSR by cyclic AMP receptor protein, the RNA polymerase alpha subunit C-terminal domain, and rhaR. *J Bacteriol* **182**, 6774-6782 (2000).
69. Wickstrum, J.R. et al. Transcription activation by the DNA-binding domain of the AraC family protein RhaS in the absence of its effector-binding domain. *J Bacteriol* **189**, 4984-4993 (2007).

70. Baldoma, L., Badia, J., Sweet, G. & Aguilar, J. Cloning, mapping and gene product identification of rhaT from Escherichia coli K12. *FEMS Microbiol Lett* **60**, 103-107 (1990).

71. Wickstrum, J.R., Skredenske, J.M., Balasubramaniam, V., Jones, K. & Egan, S.M. The AraC/XylS family activator RhaS negatively autoregulates rhaSR expression by preventing cyclic AMP receptor protein activation. *J Bacteriol* **192**, 225-232 (2010).

72. Wu, J. & Weiss, B. Two divergently transcribed genes, soxR and soxS, control a superoxide response regulon of Escherichia coli. *J Bacteriol* **173**, 2864-2871 (1991).

73. Wu, J. & Weiss, B. Two-stage induction of the soxRS (superoxide response) regulon of Escherichia coli. *J Bacteriol* **174**, 3915-3920 (1992).

74. Martin, R.G. & Rosner, J.L. Genomics of the marA/soxS/rob regulon of Escherichia coli: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. *Mol Microbiol* **44**, 1611-1624 (2002).

75. Martin, R.G., Bartlett, E.S., Rosner, J.L. & Wall, M.E. Activation of the Escherichia coli marA/soxS/rob regulon in response to transcriptional activator concentration. *J Mol Biol* **380**, 278-284 (2008).

76. Prajapat, M.K., Jain, K. & Saini, S. Control of MarRAB Operon in Escherichia coli via Autoactivation and Autorepression. *Biophys J* **109**, 1497-1508 (2015).

77. Brajesh, R.G., Raj, N. & Saini, S. Optimal parameter values for the control of gene regulation. *Mol Biosyst* **13**, 796-803 (2017).