Revisiting demand rules for gene regulation†

Mahendra Kumar Prajapat,‡ Kirti Jain,‡ Debika Choudhury, Nikhil Raj and Supreet Saini*

Starting with Savageau's pioneering work regarding demand rules for gene regulation from the 1970s, here, we choose the simplest transcription network and ask: how does the cell choose a particular regulatory topology from all available possibilities? According to the demand rules, a cell chooses an activator based regulation of a target if the target protein is required for most of the time. On the other hand, if the target protein is only required sporadically, its control tends to be *via* a repressor-based regulatory topology. We study the natural distribution of topologies at genome, systems, and micro-levels in *E. coli* and observe deviations from demand rules. Analyzing the regulation of amino acid biosynthesis, transport, and carbon utilization in *E. coli* and *B. subtilis*, and comparing choice of topology with demand, we observe an alternate pattern emerging. Simulations of networks are used to help explain the natural distribution of topologies in nature. Overall, our results indicate that the choice of topology is drawn randomly from a pool of all networks which satisfy the dynamic requirements of the cell, as dictated by physiology. In short, our results suggest that the cell picks "whatever works".

Introduction

A critical feature of all living organisms is the ability to tune behavior in response to stimuli.[1–5] The most widespread and well-understood mode of this tuning is transcription, which enables cells to modulate gene expression in response to cues. Looking at the simplest transcription network, where a regulator R, in the presence or absence of a signal, controls the expression of a target T – different possibilities emerge. Control of the target might be *via* positive or negative regulation. When we consider the fact that most transcription factors in *E. coli* are also auto-regulators, six possible topologies emerge (Fig. 1).[3,4,6–10] In this study, we seek to answer the following question: among all the available regulatory designs, how does a cell pick one to control target expression?

In a series of papers in the 1970s, Savageau proposed "demand rules for gene regulation",[11–16] according to which, a target T is positively regulated (Fig. 1A–C) if, in the organism's natural habitat, T is required for a high fraction of time. On the other hand, if the target is only required sporadically, it tends to be regulated negatively (Fig. 1D–F).[12,13] Evidence for demand rules was provided as conformity in the regulation of sugar utilization enzymes in *E. coli* with the demand rules.[11,12]

In 2006, Alon *et al.* provided a functional explanation for demand rules.[17] They argued that positive regulation for a frequently needed target T ensured erroneous binding of other transcription factors to the promoter was minimized. Alon *et al.* demonstrate and propose, in a later report,[18] that such an approach for gene regulation acts as an insulator of the promoter regions, preventing erroneous transcription.

However, demand rules raise a few interesting questions. Active control, as proposed by the demand rules, will increase the demand of regulators in the cell. The cost associated with the production of additional regulators for control is likely detrimental for cellular growth.[19–21] In addition, demand rules seem in contrast to the concept of genetic robustness, which focuses on the loss of fitness due to mutations acquired by an individual.[22] How then do we reconcile these seemingly opposite logics? In a 2009 report, Hwa *et al.* have, *via* a theoretical framework, demonstrated that the choice of mode of gene regulation could be biased for or against demand rules, and is dictated by the population size and the time scale of environmental evolution.[23] Their framework remains to be experimentally tested though. An alternate approach can be to examine the response of different topologies to cues. The response can be quantified in terms of parameters like time of response, response to noise, and cost of control.[3,24–28] However, questions like whether, over physiologically relevant range of biochemical parameter values, there are inherent qualitative differences in the response that can be generated by different topologies remain unanswered.

In this work, we perform simulations of the simplest transcriptional network (Fig. 1), and compare our results with the natural

Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai - 400076, India. E-mail: saini@che.iitb.ac.in; Fax: +91 22 2572 6895; Tel: +91 22 25767216
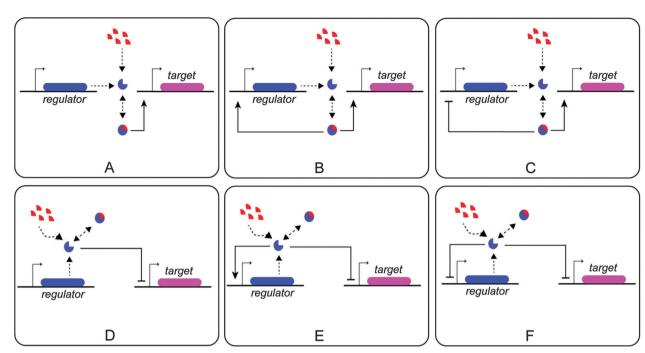
Fig. 1 Showcase of six topologies derived by interactions between a regulator R (blue) and a target, T (pink). In topologies A–C, in the presence of the appropriate environmental or cellular signal, the expression of target T is controlled positively by an induced regulator (R*). In contrast, target expression in topologies D–F is under repression by an un-induced regulator (R), and the repression is relieved under appropriate conditions.

distribution of regulatory interactions among topologies in *E. coli*. We revisit some of the results proposed by Savageau and study in detail the control of sugar utilization and amino acid biosynthesis and transport in *E. coli*. Finally, we characterize the role of control cost in dictating fitness of a cell. Put together, our results indicate several deviations from existing demand rules. Our results indicate that choice of a topology for gene expression control is likely chosen randomly from all available topologies which satisfy the dynamic demands of physiology associated with a particular cellular function.

Experimental procedure and mathematical analysis

Regulatory interactions in *E. coli*

The regulator–target (R–T) interaction dataset for the transcriptional regulatory networks was acquired from RegulonDB.[29] There are 197 transcription factors reported in RegulonDB, of which, seven are listed twice, individually, as well as in a dimeric form with another protein. We have considered 190 unique regulators for our study, and their interactions with targets have been classified into two ways; (i) between the regulator protein and the target gene and (ii) between the regulator protein and the promoter (all genes in an operon). This resulted in 4970 interactions in regulator–target gene classification (Sheet S1 in Excel, ESI†) and 2139 interactions in regulator–promoter classification (Sheet S2 in Excel, ESI†). Out of 190 transcription factors, seven are global regulators (CRP, H-NS, Lrp, IHF, ArcA, Fis, and FNR) and control around 51% of all genes in *E. coli*.[30] Excluding interactions of global

regulators, there are 2625 interactions in the regulator–target gene class and 1176 interactions in the regulator–promoter class. Multiple transcription factors feeding into a promoter were categorized into more than one topology, depending on the nature of interaction of the target gene with each interacting transcription factor.

Distribution of R–T interactions among six topologies

On the basis of the specific roles in cellular physiology, the interactions were further classified into six functional subgroups, as reported in *EcoCyc*.[31] For each functional sub-group, all the involved target genes were identified and distributed among six topologies. Biosynthesis pathways for amino acids and degradation pathways for carbohydrate were studied further in detail.

Biosynthesis and transport pathways of amino acids. For the biosynthesis of an amino acid, we considered the regulation of only target genes, which play a role in biosynthesis (Sheet S3 in Excel, ESI†) of that particular amino acid and its transport (Sheet S4 in Excel, ESI†) only.

Frequency of occurrence of amino acids. The frequency of each of the 20 amino acids from the coding region of *E. coli* DH10β genome was calculated to estimate the relative demand of all amino acids in *E. coli* (Sheet S5 in Excel, ESI†). The availability in the intestine of an amino acid was normalized by its demand (as per the relative expression of proteins in *E. coli*) to get estimate of the real availability of an amino acid.

Sugar utilization and transport. Genes encoding for enzymes involved in the metabolism of a particular carbon source until the metabolic branch merges with another in the network were

considered in our analysis (Sheet S6 in Excel). This was done to ensure that the regulation of only those enzymes which are exclusively involved in utilization of a particular carbon source are analyzed. The interactions between the identified genes and their regulators (R–T) have been distributed across the defined topologies. In addition, genes involved in the transport of sugars were also analysed in the same way (Sheet S7 in Excel, ESI†).

Mathematical analysis of the six topologies

A mathematical model for each topology was formulated by writing ordinary differential equations, and simulating deterministically. Three differential equations for each topology describe the rate of change in the inactive regulator, R, the rate of change in active regulator, R* *via* a signal and the rate of change in the target, T, as described in the ESI.†

Definition of the parameter space. To analyze the six topologies, networks were generated with different sets of biochemical parameters. To choose parameter values and range, physiologically observed values of all parameters were analyzed and the resultant space was called the parameter space (Fig. S1A, ESI†). From the parameter space, the red region represents commonly observed values reported in the literature,[32–38] biased towards exhibiting limited diversity in dynamics. In this work, we chose the "unbiased region" (blue) from the parameter space to explore all possible dynamics.[32–38]

Generation of networks from a topology. The model of topologies B, C, E, and F consists of nine parameters whereas of each from A and D consists of seven parameters. The inactive regulator, R, and the active regulator, R*, were assumed to be degraded at an equal rate. The ratio of the association constant to the dissociation constant of the signal to regulator was 100 and assumed to be constant in all topologies. The choice of this ratio was found to be close to experimental observations in many natural systems. Additionally, in our analysis, we found that changing this parameter had the least effect on dictating the performance of a network.

It was also assumed that all topologies regulate the same type of target T but just differ in the control mechanism. Hence, we did not vary the parameters of association and dissociation of signal to regulator and degradation of target T. We generated 14641 (11^4) networks for topology A and D by varying only four parameters and 15625 (5^6) networks for topologies B, C, E, and F by varying only six parameters. Such an approach was recently adopted by Ma and co-workers in the context of analysis of adaptation in biochemical networks.[39] We also performed simulations with log-normal parameter distributions across the range of values. However, different parameter ranges do not qualitatively affect our analysis. We also performed the simulations by defining parameter sets using the Latin Hypercube method. In addition, we performed simulations using the basal expression of the target T as non-zero. The results from all these simulations are as shown in Fig. S1 (ESI†).

Calculation of performance indicators. Each network was deterministically simulated using ODE45 in MATLAB.[40,41] Dynamic simulation of each network in both transitions from the OFF to ON and from the ON to OFF state was performed. To carry the simulation steps, we first simulated each network from an initial state of no regulator or target molecules, to the ON state. The steady state in the ON state was used as an initial condition to study transition from the ON to OFF state. The steady state from the OFF state was then used to study transition from the ON to OFF state. The dynamics of each network was recorded and performance indicators (as given below) for each network were calculated. The following five criteria were defined as indicators of performance of a network.

a. The steady state expression of T protein was calculated in the ON and OFF state.

b. Response time: time (t_{50}) to reach 50% of the total change in the steady state expression of T.

c. Cost of response: the amount of R to produce or remove unit T.

d. Switchability: steady state T in the ON condition divided by steady state T in the OFF condition.

e. Sensitivity: how sensitive is each topology to varying signal strength? The response level at the steady state and response time were calculated on switching the OFF condition to ON and *vice versa*.

Definition of "High Performance Box". A box in the performance indicator space defining networks with robust response, low cost, and fast dynamics was named as a "High Performance Box". Networks with the steady state expression of $T \geq 1$ Arbitrary Units (A.U.) in the ON state, with a minimum switchability factor of 1.4; activation time (t_{50}) $\leq$ 694 (A.U.) and deactivation time $\leq$ 788 (A.U.); and cost of activation and cost of deactivation ≤ 2 (A.U.) were considered to define the outer edges of the "High Performance Box". Networks outside of the "Performance Box" were assumed to be more costly, exhibiting minimal expression of the target protein T, or/and slow responding networks, and hence excluded from our analysis. The precise definition of the "High Performance Box" does not qualitatively change our results. As shown later in the paper, we note that on relaxing (making the "High Performance Box" bigger) or constraining (making the "High Performance Box" smaller), the relative number of parameter sets in the "High Performance Box" do not change qualitatively.

Evolutionary experiments

E. coli grown overnight in LB at 37 °C with shaking was subcultured (1 : 100) in tubes containing 1 ml M9 media, 1% casamino acids and a sugar source. The tubes either contained 0.4% arabinose or 0.35% glucose with 0.05% arabinose. The cultures were grown for 24 hours at 37 °C, and propagated daily by sub-culturing 1 : 100 into fresh M9 media containing respective sugars for 3000 generations. The last strains from each lineage was transformed with plasmid based promoter fusions of arabinose metabolic genes (*araB* (PEC3876-98156236)), from Thermo Scientific *E. coli* promoter collection (PEC3877). Fluorescence (488/525 nm) and absorbance (600 nm) values were measured using a Tecan microplate reader (Infinite M200 PRO).

Results

At the genome scale, *E. coli* chooses topologies differentially for the control of gene expression

To understand the ''logic'' behind the choice of topology for gene expression control, we enumerated all regulatory interactions in *E. coli*, and classified them into one of the six topologies as shown in Fig. 1.[29] We note that there is no qualitative difference in the number of interactions which are controlled *via* positive ($\sim$49.6%) or negative regulation ($\sim$50.4%) (Fig. 2A). The number of interactions belonging to topologies [A–C] or [D–F] again remain qualitatively equal to each other if we consider the regulator–operon as one inter-action (instead of regulator–gene) (Fig. 2B). Including global regulators (and their regulons) in the enumeration yields similar results (Fig. S2A and B, ESI†).

However, the R–T frequency distribution in *E. coli* changes qualitatively when we analyze the number of interactions in each of the six topologies in Fig. 1. As represented in Fig. 2C, among the six topologies, F is over-represented. This is followed by topologies A–D, with no statistically significant difference between them. Lastly, topology E is the least represented ($\sim$5% of all interactions). We observe the same general trend when we define one interaction as regulator R controlling a promoter, instead of a gene (Fig. 2D). On including the global regulators in the analysis, a slightly different picture emerges, where topologies C and F are the most represented (as most global regulators auto-regulate themselves), followed by topologies B, D, A, and E (which is again under represented) (Fig. S2C and D, ESI†).

Overall, our analysis suggests that *E. coli* prefers certain regulatory arrangements over others. What are the factors that dictate this choice? Various possibilities exist, including, demand rules,[11–16] error-minimization,[17] or minimizing cost of control.[19–21] To understand the differences between the frequencies of the six to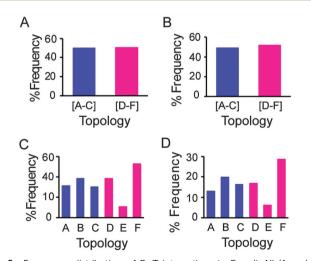pologies, we study and analyze their distribution at two different scales. At the first level, we analyze the frequency distribution of regulatory arrangements at a systems level, where a system is defined as the sum of all interactions which serve the cell towards a broad common function (such as carbohydrate utilization, or stress response). For example, all regulatory interactions which control amino acid biosynthesis were classified as a system. At the second level, we analyze in detail the demand and corresponding regulatory design at a micro level. Here by the micro level, we mean, for example, analysing regulatory interactions which dictate biosynthesis of a particular amino acid (as against a general analysis of control of all topologies involved in amino acid biosynthesis). As another example, we place all metabo-lism related regulatory interactions at the systems level, and regulatory interactions which define the control of catabolism of a particular sugar as an example of a micro level.

Differential choice of topology at a systems scale

To analyze the frequency distribution of topologies in further detail, we separated R and T interactions in *E. coli* into six functional sub-groups, namely: (a) amino acid transport and metabolism, (b) sugar metabolism and energy production, (c) coenzyme transport and metabolism, (d) inorganic transport and metabolism, (e) cell division and nucleotide metabolism, and (f) stress response (Fig. 3). In almost all groups, the distribution of the number of interactions among A–C and D–F is statistically identical to the distribution observed at the genome scale in *E. coli*. Moreover, when analyzed individually, we note that in each of the six classifications, topology E is under-represented, and topology F is over-represented. In addition, the frequency distribution in all six groups is statistically similar to the one observed globally in *E. coli* (Fig. 3A–F). We revisit the reason and nature of this distribution later in this manuscript. It is not wholly surprising that topology F is over-represented in nature. Negative auto-regulation is known to speed up the response in cellular systems,[9,25] and both topologies C and F possess that architecture. However, subtle differences exist. While topology C speeds up the response when the system transitions from an OFF to an ON state, topology F speeds up cellular response on transition from the ON to OFF state. Could other similar dynamic criteria explain the differential use of topologies in *E. coli*?

Revisiting demand rules for gene regulation

Next, we analyse demand rules at a micro scale in *E. coli* and *B. subtilis*. Application of demand rules require that we quantify demand for a protein in the environment that the bacterium has evolved in. We do this using two different examples, as discussed below.

In the first example, we study amino acid biosynthesis and transport in *E. coli*, primarily present in mammalian intestine. While *E. coli* has the ability to synthesize all 20 amino acids, because of their unequal presence in the intestine, not all amino acids are required equally by the bacterium.[47,48] The demand for amino acids is further biased by the number of codons for each amino acid in the *E. coli* coding region, and
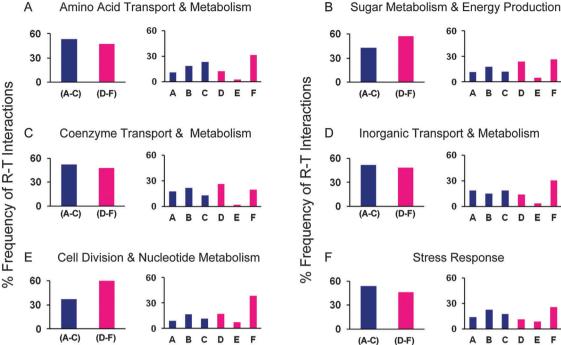

Fig. 2 Frequency distribution of R–T interactions in *E. coli*. All (A and C) regulator–gene interactions (B and D) regulator–promoter interactions in *E. coli*, without including global regulators, were calculated from RegulonDB and then distributed (A and B) among activator- [A–C] and repressor-based [D–F] topologies and (C and D) among the six distinct topologies.

Fig. 3 Division of R–T interactions among six functional groups. All regulator–promoter interactions in *E. coli* were classified into six functional groups. For each functional group without including global regulators, the frequency of activator-based topologies [A–C] roughly equals repressor-based topologies [D–F]. Among individual topologies, F is over-represented and E is under-represented in all six functional groups.

the relative expression of each protein in the *E. coli* genome. To get a better estimate of the relative demand for each amino acid, we use data published by Sunney Xie and co-workers in 2009.[52] In this study, the authors report absolute levels of expression of more than 1000 proteins in *E. coli*, thus providing data for demand for amino acids. In our analysis, we identified the biosynthesis pathway(s) which are uniquely dedicated to the synthesis of a particular amino acid only,[31] analysed regulation of each gene in the pathway(s), and classified regulation as positively or negatively regulated topologies.

Amino acid biosynthesis and transport are cellular functions with inversely related demand. For instance, if an amino acid is not present in the surroundings (resulting in low demand for transporters), the biosynthetic demand would be high, and *vice versa*. In Fig. 4A, the *x*-axis represents amino acids in increasing availability in *E. coli* habitat, while the *y*-axis gives the fraction of all regulatory interactions, controlling the biosynthesis and transport of that amino acid, belonging to topologies A–C. Our analysis shows that the regulation of both biosynthesis and transport exhibit a statistically insignificant correlation
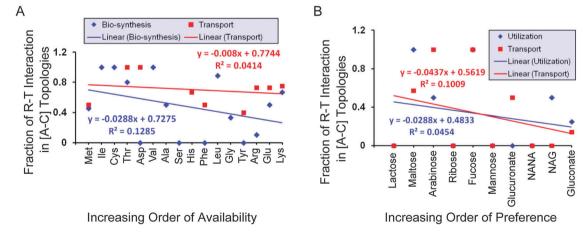

Fig. 4 Correlation between demand and activator-based control. (A) *X*-axis represents amino acids in increasing availability to *E. coli*, and *y*-axis represents the fraction of regulatory interactions in the activator-based topologies for each amino acid biosynthesis regulon (blue) and its transport (red). (B) *X*-axis represents carbon sources in increasing order of preference to *E. coli* in the intestine, and *y*-axis, the fraction of regulatory interactions in the activator-based topologies controlling expression of metabolic genes involved in utilization (blue) of each sugar and its transport (red).

with increasing availability. This is in contrast to the demand rules. Adherence to the demand rules would have meant that transporters of abundant amino acids are regulated by A–C topologies, and biosynthetic genes for such amino acids are primarily regulated by D–F topologies. The reverse would have held true for scarcely available amino acids. The same results hold on including interactions involving global regulators or when we normalize demand by including a number of codons for an amino acid in our analysis (Fig. S3A–C, ESI†). All the results show a statistically insignificant relationship against demand rules. Additionally, we performed a similar analysis for the soil bacterium *B. subtilis*, and found no correlation between the choice of topology and availability of an amino acid in the surroundings (Fig. S4A and C, ESI†).[49,50]

In the second example, we focus on the metabolism of sugars preferred by *E. coli* in its natural habitat.[51] Based on their abundance, we obtained the relative demand for carbohydrates in the intestine.[51] For our analysis, we only considered part of the metabolism which exclusively deals with the catabolism of that particular carbon source only. The genes encoding the respective enzymes and their regulators were studied, and classified into activator- or repressor-based topologies. Our results indicate that the regulation of enzymes involved in carbon utilization is independent of the availability of the sugar (Fig. 4B). A similar statistically insignificant result was obtained on including global regulators (Fig. S3D, ESI†). In case of carbon utilization, the expression of transporter genes should be positively correlated with the expression of genes involved in the catabolism. However, our analysis shows that the choice of topology does not seem linked with the availability of the carbon source (Fig. 4B and Fig. S3D, ESI†). A similar analysis was performed for carbon utilization in *B. subtilis* and no statistical correlation was observed between the choice of topology and demand for the product of gene of interest (Fig. S4B and D, ESI†).[49,50] As shown in Table 1, the Spearman correlation coefficients between the demand and fraction of [A–C] topologies controlling expression show weak statistical significance, at best.

Overall, our results indicate deviations from Savageau's demand rules. Rather, looking at our results from the micro level analysis, it seems that the choice of topology is made independent of demand. A major difference in our and Savageau's analysis is that we consider all regulatory interactions controlling cellular functions, while Savageau's work only accounted for the key regulator involved in a particular cellular process, for example, AraC for arabinose catabolism.[11,13]

Speculative evolutionary experiment to note switch in mode of regulation

As a speculative test of the demand rules, we performed long-term experiments where we fed 0.2% arabinose as a sole carbon source to *E. coli* for 3000 generations. In parallel, we also grew *E. coli* in 0.2% glucose and 0.005% arabinose. The culture grown on arabinose had high demand for arabinose utilization genes, whereas the culture with small amounts of arabinose would only express from the *araBAD* operon when out of glucose – thus creating differential demand for the *araBAD* gene products. AraC, when bound to arabinose, is known to be an activator of the *araBAD* operon.[42–46] In the absence of arabinose, AraC is known to be a repressor of the *araBAD* operon. This dual regulation can be observed experimentally in wild type *E. coli*. On altering demand for the *araBAD* gene products, the dual regulation can still be observed in both (with high and low demand for *araBAD* gene products) the strains (Fig. 5). In a relatively short span of 3000 generations, no switch in mode of
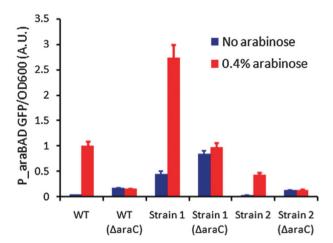

Fig. 5 Long-term experiment to track changes, if any, in mode of regulation. P_{araBAD} expression in strain grown in 0.4% arabinose for 3000 generations (called Strain 1) and strain grown in 0.35% glucose and 0.05% arabinose for 3000 generations (called Strain 2). WT refers to MG1655. Δ*araC* strain refers to the mutant created by knocking out *araC* from the parent evolved strain.

Table 1 Spearman's rank correlation between demand and fraction of regulatory interactions belonging to topologies [A–C]

	Sugar		Amino acid	
	Utilization	Transport	Biosynthesis	Transport
E. coli				
Without global regulator	−0.097	−0.212	−0.304	−0.046
With global regulator	−0.02	−0.206	−0.145	−0.527
Without global regulator (normalized with genome demand)	—	—	−0.323	0.0
With global regulator (normalized with genome demand)	—	—	−0.334	0.085
B. subtilis				
Without global regulator	0	—	−0.183	—
With global regulator	0	—	−0.217	—

regulation was observed, though the absolute levels of expression were different in the two strains, and had evolved from the parent wild-type *E. coli*. Our efforts in searching in literature for experimental evidence for switch in mode of regulation in response to demand were unsuccessful.

If demand rules do not explain the choice of topology, then how can we understand the differences in frequency distribution of the six topologies we are analysing? To answer this, we perform simulations of the six topologies and characterize their performance as described in the next section.

Simulations to quantify performance of networks across topologies

We define a list of factors that best define the performance of a regulatory circuit. These include (a) steady state target expression, (b) time of response, (c) control cost, (d) ability to be effectively switched ON and OFF, and (e) sensitivity towards environmental signal (see Methods for more details). We hypothesize that these five indicators define the performance of a genetic network, and hence simulate all six topologies and quantify performance. However, network dynamics are dictated by the values of the associated biochemical parameters. To account for biases introduced by parameters, we simulated about 90 000 networks (~15 000 networks in each topology). The choice of parameters for these networks was taken from a range. Each network was simulated using ODE45 in MATLAB, and transition from OFF to ON and ON to OFF tracked.

Because of our choice of parameters from a parameter space, many networks are "dead" (steady state target expression less than one); have infinite cost; or have physiologically unviable dynamics. Among the six topologies, topology E has the greatest percent of these "dead" networks (data not shown). For analysis, we considered only those networks which express target T and are able to effectively switch ON/OFF. In addition, we impose limits on the time of activation (and deactivation) and control cost. Placing these constraints allows us to define a "High Performance Box". In rest of the article, we only consider networks which lie within this "High Performance Box", unless otherwise stated. A frequency count of the networks with positive and negative control of T shows that the two are identical (Fig. 6A), consistent with the global distribution in *E. coli* (Fig. 2A and B).

The frequency distribution of the six topologies shows that, just as in *E. coli* (Fig. 2C and D), topology E is under-represented, and topology F, most represented in the "Performance Box" (Fig. 6B). Hence, our simulations suggest, and we speculate that the distribution of a topology in a cell is proportional to the frequency of the topology in the "High Performance Box" (*p*-value = 0.276, thus rejecting the null hypothesis that the natural and "High Performance Box" distributions are statistically different). Changing the dimensions of the "High Performance Box" does not alter the frequency distribution of the topologies (Fig. 7). We note, however, that there are subtle differences in the distribution of the networks among the six topologies between our computational results and the *E. coli* distribution. We hypothesize that these differences in distributions are due to the inherent differences in the dynamic features of the six topologies (Fig. S5–S9, ESI†).

One interesting qualitative difference in the performance of the six topologies is shown in Fig. S7 (ESI†). For all networks in the "High Performance Box", change in steady state target response (ΔTarget) in both switching ON (blue) and OFF (red) against change in the amount of total regulators (ΔRegulator) at steady state were plotted. Total regulators include both the active and inactive form of regulators. The positive value of ΔTarget indicates the switching of a network from the OFF to

Fig. 7 Frequency distribution of R–T interactions from simulations. A fraction of networks belonging to each of six topologies in the "Performance Box". (A and B) represent the fraction of networks in transient performance boxes when the boundaries of nominal performance box have been tightened up, (C) represents the most constrained performance box, (D and E) represent the fraction of networks in transient performance boxes when the boundaries of nominal performance box have been relaxed, (F) represents the most relaxed performance box. In each transient performance box, the fractional distribution changes by the number but not by the pattern.
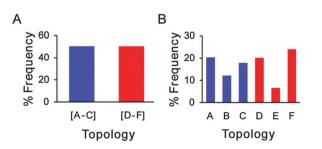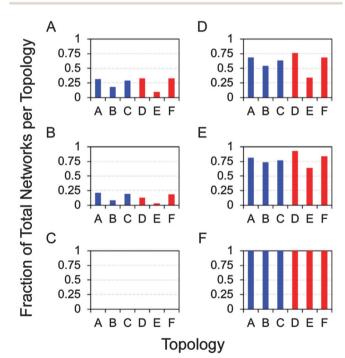

Fig. 6 Frequency distribution of R–T interactions from simulations. (A) Frequency distributions of networks belonging to the activator-based [A–C] and repressor-based [D–F] topologies in the "Performance Box" and (B) percent frequency of networks belonging to each of six topologies in the "Performance Box".

ON condition whereas the negative value indicates switching from the ON to OFF condition. The positive value of ΔRegulator indicates that a particular switching needs production of more regulators and the negative value of ΔRegulator indicates regulators need to be degraded for that particular switching.

Intuitively, networks in topologies A–C should offer qualitatively similar ΔTarget–ΔRegulator profile. Likewise, networks in topologies D–F should have similar ΔTarget–ΔRegulator profile qualitatively. However, as shown in Fig. S7, ESI,† we note that topologies A and D, topologies B and F, and topologies C and E offer qualitatively similar ΔTarget–ΔRegulator profile. In topologies A and D, due to the absence of feedback, ΔTarget values for all networks are plotted at almost zero change in total regulators. The ΔTarget–ΔRegulator plane shows that topologies C and E offer dynamics where the production of regulators is required to switch the network OFF and degradation of regulators is required to switch the network ON. On other hand, topologies B and F offer a dynamics where they require the production of regulators to switch the network ON and degradation of regulators during switching network OFF. In other words, it suggests that topologies C and E are expensive in switching OFF whereas topologies B and F are expensive in switching ON. In Fig. S7, ESI,† we show that how economic a control in a particular topology is to maintain a certain response level at the steady state. Our analysis (Fig. S7, ESI†) suggests that topologies A–C and topologies D–F are qualitatively similar in showing maintenance cost dynamic. However, topology B and C, for some networks, require a high number of regulators to maintain target response in ON and OFF conditions respectively. Again, only topology E, in both ON and OFF conditions, remains expensive by requiring a high number of regulators to maintain the response level at the steady state. In terms of control of switching and maintaining the response level, topology F is suited for most of the cellular functions in the cell whereas topology E stands the most expensive.

Discussion

Transcription networks are of interest from a number of perspectives like structure, topology, dynamics, and evolution.[3,24,37,53–58] Despite significant effort in trying to understand dynamic features of topologies – an open question remains. Why does a cell choose a particular topology over the others?

Demand rules provide an insight into this question. However, our analysis reveals that the agreement with demand rules is rather limited. What then could be the additional determinants? Dynamically, as our analysis shows, there are several subtle differences across topologies. Our simulations and comparison with the natural distribution of topologies in *E. coli* suggests that, given a performance criteria, the frequency distribution among the six topologies in *E. coli*, and our simulations is identical. The simplest explanation from this analysis regarding the choice of topology could be that a network is perhaps randomly picked out from a group that

satisfies the demands of performance. Or simply, we speculate that the cell picks "whatever works".

In a 2009 study, Hwa *et al.* demonstrated that different modes of regulation lead to qualitatively different patterns of protein levels when cells are grown under conditions supporting different growth rates.[59] They demonstrate that constitutively and positively controlled genes exhibit a decrease in the steady state with growth rate, negatively regulated genes can exhibit a weakly negative or a strongly positive correlation between protein levels and growth rate. Could additional considerations like maintenance of protein levels at a constant level, independent of growth rate, be a selective force for certain physiological roles?

Here, we describe simulations of the simplest regulatory topology. In terms of increasing the complexity of the network structures we can explore, our simulation approach very rapidly approaches saturation. In addition, combinatorial inputs of multiple regulators into one promoter remain unanswered and unexplored. These additional interactions would make the possible range of dynamic behaviour much more complex and richer, but at the same time computationally intractable.

Author contributions

KJ: All RegulonDB analysis. MKP: All simulation work. DC: Evolutionary experiment. NR: Analysis on relative amino acid demand data. MKP, KJ, SS: Conceived the study. MKP, KJ, SS: Wrote the paper.

Conflicts of interest

The authors declare that no competing interests exist.

Acknowledgements

References

1 F. C. Neidhardt, *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology*, American Society for Microbiology, 1996.

2 M. Ptashne and A. Gann, *Genes and Signals*, Cold Spring Harbor Laboratory Press, 2002.

3 S. S. Shen-Orr, R. Milo, S. Mangan and U. Alon, *Nat. Genet.*, 2002, **31**, 64–68, Epub 2002 Apr 2022.

4 D. Thieffry, A. M. Huerta, E. Perez-Rueda and J. Collado-Vides, *BioEssays*, 1998, **20**, 433–440.

5 J. D. van Elsas, A. V. Semenov, R. Costa and J. T. Trevors, *ISME J.*, 2011, **5**, 173–183, DOI: 110.1038/ismej.2010.1080, Epub 2010 Jun 1024.

6 S. T. Crews and J. C. Pearson, *Curr. Biol.*, 2009, **19**, R241–R246, DOI: 210.1016/j.cub.2009.1001.1015.

7 M. Madan Babu and S. A. Teichmann, *Nucleic Acids Res.*, 2003, **31**, 1234–1244.

8 R. P. Ngondo and P. Carbon, *Nucleic Acids Res.*, 2014, **42**, 2171–2184, DOI: 2110.1093/nar/gkt1136, Epub 2013 Nov 2114.

9 N. Rosenfeld, M. B. Elowitz and U. Alon, *J. Mol. Biol.*, 2002, **323**, 785–793.

10 M. A. Savageau, *Nature*, 1974, **252**, 546–549.

11 M. A. Savageau, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 2453–2455.

12 M. A. Savageau, *Proc. Natl. Acad. Sci. U. S. A.*, 1977, **74**, 5647–5651.

13 M. A. Savageau, *Proc. Natl. Acad. Sci. U. S. A.*, 1983, **80**, 1411–1415.

14 M. A. Savageau, *Genetics*, 1998, **149**, 1665–1676.

15 M. A. Savageau, *Genetics*, 1998, **149**, 1677–1691.

16 C. Sybesma, *Biophysics*, Springer Netherlands, 1989, ch. 12, pp. 281–290, , DOI: 10.1007/978-94-009-2239-6_12.

17 G. Shinar, E. Dekel, T. Tlusty and U. Alon, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 3999–4004, Epub 2006 Mar 3996.

18 V. Sasson, I. Shachrai, A. Bren, E. Dekel and U. Alon, *Mol. Cell*, 2012, **46**, 399–407, DOI: 310.1016/j.molcel.2012.1004.1032.

19 A. L. Koch, *J. Mol. Evol.*, 1983, **19**, 455–462.

20 C. G. Kurland and H. Dong, *Mol. Microbiol.*, 1996, **21**, 1–4.

21 I. Shachrai, A. Zaslaver, U. Alon and E. Dekel, *Mol. Cell*, 2010, **38**, 758–767, DOI: 710.1016/j.molcel.2010.1004.1015, Epub 2010 Apr 1029.

22 J. A. de Visser, J. Hermisson, G. P. Wagner, L. Ancel Meyers, H. Bagheri-Chaichian, J. L. Blanchard, L. Chao, J. M. Cheverud, S. F. Elena, W. Fontana, G. Gibson, T. F. Hansen, D. Krakauer, R. C. Lewontin, C. Ofria, S. H. Rice, G. von Dassow, A. Wagner and M. C. Whitlock, *Evolution*, 2003, **57**, 1959–1972.

23 U. Gerland and T. Hwa, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 8841–8846, DOI: 8810.1073/pnas.0808500106, Epub 0808502009 May 0808500122.

24 U. Alon, *Nat. Rev. Genet.*, 2007, **8**, 450–461.

25 D. Madar, E. Dekel, A. Bren and U. Alon, *BMC Syst. Biol.*, 2011, **5**, 111, DOI: 10.1186/1752-0509-1185-1111.

26 S. Mangan, S. Itzkovitz, A. Zaslaver and U. Alon, *J. Mol. Biol.*, 2006, **356**, 1073–1081, Epub 2005 Dec 1019.

27 S. Mangan, A. Zaslaver and U. Alon, *J. Mol. Biol.*, 2003, **334**, 197–204.

28 K. Wu and C. V. Rao, *Mol. Microbiol.*, 2010, **75**, 513–527, DOI: 510.1111/j.1365-2958.2009.07011.x, Epub 02009 Dec 07016.

29 H. Salgado, M. Peralta-Gil, S. Gama-Castro, A. Santos-Zavaleta, L. Muniz-Rascado, J. S. Garcia-Sotelo, V. Weiss, H. Solano-Lira, I. Martinez-Flores, A. Medina-Rivera, G. Salgado-Osorio, S. Alquicira-Hernandez, K. Alquicira-Hernandez, A. Lopez-Fuentes, L. Porron-Sotelo, A. M. Huerta, C. Bonavides-Martinez, Y. I. Balderas-Martinez, L. Pannier, M. Olvera, A. Labastida, V. Jimenez-Jacinto, L. Vega-Alvarado, V. Del Moral-Chavez, A. Hernandez-Alvarez,

E. Morett and J. Collado-Vides, *Nucleic Acids Res.*, 2013, **41**, D203–D213, DOI: 210.1093/nar/gks1201, Epub 2012 Nov 1029.

30 A. Martinez-Antonio and J. Collado-Vides, *Curr. Opin. Microbiol.*, 2003, **6**, 482–489.

31 I. M. Keseler, A. Mackie, M. Peralta-Gil, A. Santos-Zavaleta, S. Gama-Castro, C. Bonavides-Martinez, C. Fulcher, A. M. Huerta, A. Kothari, M. Krummenacker, M. Latendresse, L. Muniz-Rascado, Q. Ong, S. Paley, I. Schroder, A. G. Shearer, P. Subhraveti, M. Travers, D. Weerasinghe, V. Weiss, J. Collado-Vides, R. P. Gunsalus, I. Paulsen and P. D. Karp, *Nucleic Acids Res.*, 2013, **41**, D605–D612, DOI: 610.1093/nar/gks1027, Epub 2012 Nov 1099.

32 A. Y. Mitrophanov and E. A. Groisman, *BioEssays*, 2008, **30**, 542–555, DOI: 510.1002/bies.20769.

33 E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman and A. van Oudenaarden, *Nat. Genet.*, 2002, **31**, 69–73, Epub 2002 Apr 2022.

34 N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain and M. B. Elowitz, *Science*, 2005, **307**, 1962–1965.

35 N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain and M. B. Elowitz, *Mol. Syst. Biol.*, 2007, **3**, 143, Epub 2007 Nov 2013.

36 M. Santillan and M. C. Mackey, *Chaos*, 2001, **11**, 261–268.

37 K. Sneppen, S. Krishna and S. Semsey, *Annu. Rev. Biophys.*, 2010, **39**, 43–59.

38 G. M. Suel, J. Garcia-Ojalvo, L. M. Liberman and M. B. Elowitz, *Nature*, 2006, **440**, 545–550.

39 W. Ma, A. Trusina, H. El-Samad, W. A. Lim and C. Tang, *Cell*, 2009, **138**, 760–773, DOI: 710.1016/j.cell.2009.1006.1013.

40 D. T. Gillespie, *J. Comput. Phys.*, 1976, **22**, 403–434.

41 D. T. Gillespie, *J. Phys. Chem.*, 1977, **81**, 2340–2361.

42 A. Haldimann and B. L. Wanner, *J. Bacteriol*, 2001, **183**, 6384–6393.

43 W. G. Miller, J. H. Leveau and S. E. Lindow, *Mol. Plant-Microbe Interact.*, 2000, **13**, 1243–1250.

44 K. Skarstad, B. Thony, D. S. Hwang and A. Kornberg, *J. Biol. Chem.*, 1993, **268**, 5365–5370.

45 S. P. Cohen, S. B. Levy, J. Foulds and J. L. Rosner, *J. Bacteriol.*, 1993, **175**, 7856–7862.

46 N. Lee, C. Francklyn and E. P. Hamilton, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 8814–8818.

47 S. E. Nixon and G. E. Mawer, *Br. J. Nutr.*, 1970, **24**, 241–258.

48 S. E. Nixon and G. E. Mawer, *Br. J. Nutr.*, 1970, **24**, 227–240.

49 R. Caspi, T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang and P. D. Karp, *Nucleic Acids Res.*, 2014, **42**, D459–D471, DOI: 410.1093/nar/gkt1103, Epub 2013 Nov 1012.

50 H. Fischer, A. Meyer, K. Fischer and Y. Kuzyakov, *Soil Biol. Biochem.*, 2007, **39**, 2926–2935.

51 D. E. Chang, D. J. Smalley, D. L. Tucker, M. P. Leatham, W. E. Norris, S. J. Stevenson, A. B. Anderson, J. E. Grissom, D. C. Laux, P. S. Cohen and T. Conway, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 7427–7432, Epub 2004 May 7423.

52 Y. Taniguchi, P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili and X. S. Xie, *Science*, 2010, **329**, 533–538.

53 H. Dong, L. Nilsson and C. G. Kurland, *J. Bacteriol.*, 1995, **177**, 1497–1504.

54 E. Dekel and U. Alon, *Nature*, 2005, **436**, 588–592.

55 D. M. Stoebel, A. M. Dean and D. E. Dykhuizen, *Genetics*, 2008, **178**, 1653–1660, DOI: 1610.1534/genetics.1107. 085399, Epub 082008 Feb 085393.

56 M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein and S. A. Teichmann, *Curr. Opin. Struct. Biol.*, 2004, **14**, 283–291.

57 L. H. Hartwell, J. J. Hopfield, S. Leibler and A. W. Murray, *Nature*, 1999, **402**, C47–C52.

58 R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, *Science*, 2002, **298**, 824–827.

59 S. Klumpp, Z. Zhang and T. Hwa, *Cell*, 2009, **139**, 1366–1375, DOI: 1310.1016/j.cell.2009.1312.1001.

This journal is © The Royal Society of Chemistry 2016