Published online 12 October 2007

Nucleic Acids Research, 2007, Vol. 35, No. 20 ***6963–6972***
doi:10.1093/nar/gkm743

Coordination logic of the sensing machinery in the transcriptional regulatory network of *Escherichia coli*

Sarath Chandra Janga[1,*], Heladia Salgado[1], Agustino Martínez-Antonio[2] and Julio Collado-Vides[1]

[1]Programa de Genómica Computacional, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, 62100, México and [2]Departmento de Ingenieria Genética, Centro de Investigacion y de Estudios Avanzados, Campus Guanajuato, Irapuato, 36500, México

Received July 10, 2007; Revised August 31, 2007; Accepted September 6, 2007

ABSTRACT

The active and inactive state of transcription factors in growing cells is usually directed by allosteric physicochemical signals or metabolites, which are in turn either produced in the cell or obtained from the environment by the activity of the products of effector genes. To understand the regulatory dynamics and to improve our knowledge about how transcription factors (TFs) respond to endogenous and exogenous signals in the bacterial model, *Escherichia coli*, we previously proposed to classify TFs into external, internal and hybrid sensing classes depending on the source of their allosteric or equivalent metabolite. Here we analyze how a cell uses its topological structures in the context of sensing machinery and show that, while feed forward loops (FFLs) tightly integrate internal and external sensing TFs connecting TFs from different layers of the hierarchical transcriptional regulatory network (TRN), bifan motifs frequently connect TFs belonging to the same sensing class and could act as a bridge between TFs originating from the same level in the hierarchy. We observe that modules identified in the regulatory network of *E. coli* are heterogeneous in sensing context with a clear combination of internal and external sensing categories depending on the physiological role played by the module. We also note that propensity of two-component response regulators increases at promoters, as the number of TFs regulating a target operon increases. Finally we show that evolutionary families of TFs do not show a tendency to preserve their sensing abilities. Our results provide a detailed panorama of the topological structures of *E. coli* TRN and the way TFs they compose off, sense their surroundings by coordinating responses.

INTRODUCTION

Organisms multiparty and re-associate their transcriptional regulatory network (TRN) orchestrating numerous transcriptional responses depending on the fluctuations in their internal and external conditions (1,2). The cellular components that sense these variations are linked to the transcriptional machinery through the activity of transcription factors (TFs). TFs can respond to specific signals resulting in allosteric modifications that change their affinities to specific DNA-binding sites (operators) or with the rest of the transcriptional machinery (3). These effector signals can be classified as exogenous or endogenous depending on their origin in the cellular context, i.e. whether the cell can take them from the milieu or produce them in the cytoplasm (4,5).

From a network perspective, TRNs have been studied in the following order of simplicity in terms of their topological organization: (i) at a global level TRNs have been shown to possess a multi-layer hierarchical modular structure using either a top-down or a bottoms-up approach for determining hierarchy (6,7), (ii) modules, these structures include the activity of several TFs sharing the regulation of related physiological functions (8–11) and (iii) motifs, which are composed of patterns constituting one or more TFs modulating the activity of a set of target genes (12–14). Three motif types were found to be dominant in TRNs, namely feed forward loops (FFLs) in which two TFs control the activity of the target gene with one of the TFs regulating the other, bifans in which two different TFs both control the expression of two target genes and single input modules (SIMs), where a single TF controls a group of target genes. On the other hand, from a genomic perspective, identification of operators in the non-coding regions of DNA responsible for the physical interaction between the DNA-recognition domain of TFs and the promoter zone has been an area of longstanding interest (15–17). However, from a signal integration perspective which involves our ability to understand how a simple organism like *Escherichia coli* can integrate

*To whom correspondence should be addressed. Tel: +52 777 313 2063; Fax: +52 777 317 5581; Email: sarath@ccg.unam.mx

signals from the exterior to the interior and coordinate its responses to changing environments, our knowledge is rather limited. Some recent studies tried to address this question, using publicly available expression data in *E. coli* and budding yeast, however they have mostly limited their study to understand global dynamics (1,18).

In this work, we study how the TFs that sense exogenous and/or endogenous signals [see Supplementary Data; (4,5)], constitute and define the behavior of topological structures in the *E. coli* regulatory network. In addition, we address how the regulatory code is interpreted by this set of TFs at the promoters they act upon. Although two recent studies have addressed certain aspects of combinatorial regulation in TRNs, no systematic in-depth large-scale comparison of the biological classification of TFs (based on the source of their biochemical signals) against different topological structures like motifs and modules has been performed (19,20). The results presented in this work enhance our understanding of the logic used by simple cells to sense and respond to environmental changes through the activity of TFs, which coordinate and partition the cell to respond to quotidian changing conditions.

MATERIALS AND METHODS

TFs and their evolutionary families in *E. coli*

Complete set of *E. coli* TFs analyzed in this study were obtained from RegulonDB (21), which is a manually curated database containing information on transcriptional regulation and operon organization in *E. coli*. Majority of the DNA-binding TFs in bacterial genomes can be classified in to a number of families based on structural homologies (22,23). TF families classified based on structural domains of the DNA-binding regions comprise of three folds, the helix–turn–helix, the winged helix and the beta ribbon with the most abundant among TFs being the classical helix–turn–helix domain (24). Evolutionary families for the complete repertoire of *E. coli* TFs having a helix–turn–helix domain, have been defined according to a previous study (22). Only those evolutionary families in which at least two TFs could be associated to a sensing class were considered. All calculations for each family were performed with respect to the total number of TFs which could be classified into a sensing category. A total of 13 families were analyzed for understanding preferences in sensing class distribution.

Data of transcriptional regulatory interactions and sensing classification in *E. coli*

The currently known network of transcriptional regulatory interactions in the complete genome of *E. coli* was obtained from RegulonDB (21). The network contained 1368 nodes and 2773 edges after removing autoregulatory and sigma-mediated interactions. TFs which act as dimers were also taken into account by manual curation to generate the final set of regulatory interactions. The basic unit of transcriptional sensing is composed of a TF and its corresponding effector genes; the former encode for a TF sensing the effector signal produced or obtained by the product of the second gene (4,5,25). The main character of the subclasses of the genetic sensing machinery in *E. coli* are shown in Supplementary Data and a more complete discussion is presented elsewhere (4). It was possible to obtain experimental or annotated information for 123 TFs and 324 effector genes. This set of TFs correspond to 41% of about 300 predicted TFs in *E. coli* (22,23).

Identification of motifs and motif subtypes in the TRN

Network motifs are defined as recurring regulation patterns which occur in the TRNs more often than expected by chance (12,26). In the regulatory network of *E. coli* three distinct types of motifs have been found to be predominant, namely (i) FFL, in which a TF regulates the expression of another transcription factor which together modulate the expression of the target gene; (ii) SIM, in which a single TF regulates several genes and is equivalent to a simple regulon (27); (iii) bifans in which two different TFs both regulate two target genes and are analogous to complex regulons (27). FFL appears to be the most abundant motif among the best studied transcriptional networks. To identify different kinds of motifs in the TRN we searched for the respective sub-graphs in the network with the specified topology. Note that for motif identification each gene of an operon would result in a regulatory interaction and hence, if an operon had three genes and was regulated by two TFs, this would lead to three bifans. However, our end results did not vary when an operon-based regulatory network was considered. We identified a total of 865 FFLs, 20480 bifans and 52 SIMs in the transcriptional network of *E. coli*. However, TFs in only 659 FFLs, 16759 bifans and 35 SIMs could be associated to a sensing class and hence were used for further analysis. Complete set of these network motifs along with the association of their TFs to sensing categories analyzed in this study can be obtained as Supplementary Data. FFLs can be sub-divided into eight different types of motifs depending on the mode of action of each of the two TFs involved in a FFL (28). Motif sub-types of FFLs were identified by using the mode of action of a TF on its target gene in the regulatory network. If a target gene is known to be both activated and repressed by a TF then each interaction was considered independently for identifying FFL motif-subtypes.

Identification of modules in the network of transcriptional regulatory interactions among TFs

Although there is no general consensus on the definition of a regulatory module (29), a transcriptional regulatory module is typically defined as a set of genes that are regulated by a common set of TFs. Under this definition, it is intuitive to expect that various cellular processes can be conveniently regulated by discrete and separable modules which can coordinate the activities of many genes and carry out complex functions. Therefore, identifying transcriptional modules is useful for understanding cellular responses to internal and external signals under different cellular conditions. Due to our interest in only TFs forming part of the modules we identified the modules in the regulatory network according to a previously

proposed approach (9). Briefly, we first constructed a distance matrix of the TFs, using the inverse square of the shortest path length between any two nodes and then hierarchically clustered the interactions to obtain modules. A total of eight modules were identified in the regulatory network. The obtained modules were then analyzed for the composition of different sensing classes.

Estimating the significance of enrichment for sensing categories in motifs

To calculate the significance of enrichment of sensing class combinations from different positions of FFL and bifan and to estimate the enrichment for a particular sensing category at a position in motifs, we compared the distribution of sensing classes with tendencies seen in 1000 randomly generated sets of motif structures same in count as the number of motifs seen in the regulatory network. Random motif structures for FFLs were generated by shuffling the labels of TFs between two motifs chosen at random while preserving their position in the motif. This was done to preserve the overall connectivity of the TF while still randomly associating it with a TF from another sensing class in the motif context. Similarly, random motif structures were created by shuffling the labels of TFs between pairs of bifan motifs to preserve the connectivity of a TF while changing its association with the sensing class of the other TF. The total number of random shuffles in the generation of each of the 1000 random sets was equal to two times the number of motifs detected of a given type. Shuffling in the order of the square of number of motifs of a given type did not vary our end results. For SIMs, 1000 sets of 35 TFs were sampled from those which could be classified into a sensing category and tested for enrichment of sensing class. For all observations reported in this study, statistical significance was assessed based on (i) z-score, calculated as the number of SDs the observed value is away from the randomly expected mean. This is obtained as the ratio between the differences of the observed, x, and random expected, μ, values to the standard deviation, σ i.e. $z = (x - \mu)/\sigma$) and (ii) P-values, defined as the fraction of the 1000 random trails which showed a value greater than equal to what was observed in the real dataset. Values of $|z\text{-scores}| \geq 3.3$ and $P\text{-value} \leq 10^{-3}$ (unless stated) were considered to show a significant difference in comparison to the null model. So in all the figures z-score was used as a parameter to assess significance.

RESULTS

Most FFL motifs co-ordinate their activity using a combination of internal and external sensing TFs

Motifs are sub-graphs which occur more frequently than expected by chance in networks. They have been first described in the TRN of *E. coli* and subsequently found in a variety of complex systems (12,26). FFL is a three-node subgraph and is one of the most abundantly found motif in all well characterized TRNs studied so far (14). This motif comprises of three genes: a regulator X, which regulates Y, and gene Z which is regulated by both X and Y (Figure 1a). Unlike bifan motifs (see below) FFLs are not symmetric for the positions of the two TFs comprising this motif as the first TF regulates two genes while the second regulates only one. To understand the organization of TFs in FFLs in the context of sensing classification we first identified the complete set of FFLs in the currently known TRN of *E. coli* (see Materials and Methods section). A total of 659 FFLs could be identified which could be associated to TFs with a classified category of sensing. To address the contributions and enrichment of different classes of sensing for the positions of the two TFs which comprise a FFL, we compared distributions seen to randomly generated FFLs (see Materials and Methods section). We found ISM (Internal Sensing Metabolites class) to be significantly enriched for the first position of the FFL while the sensing classes H (Hybrid; sensing transported and synthesized metabolites), ETC (External sensing Two-Components) and ETM (External sensing Transported Metabolites) are underrepresented in the first position (Table 1). On the other hand, the second position

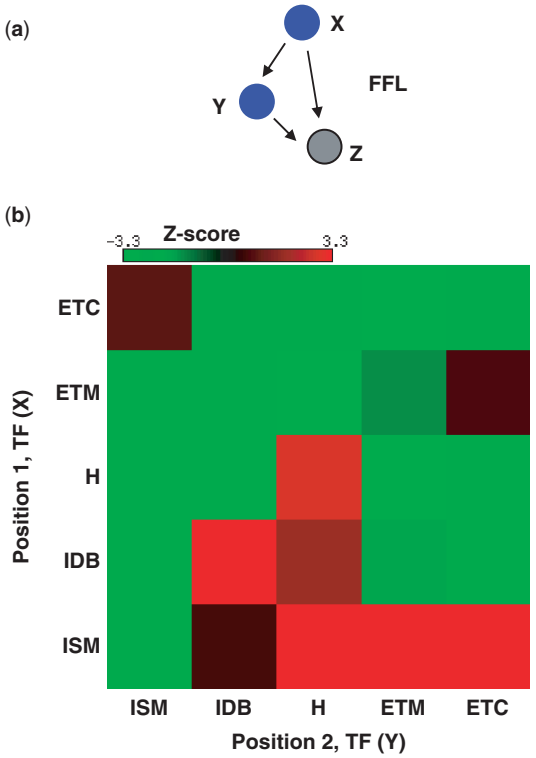

Figure 1. Statistical significance of co-occurrence of sensing classes in Feed Forward Loop (FFL) motifs observed in the transcriptional network of *E. coli*. (**a**) In a FFL motif, which is a commonly seen topological structure in transcriptional regulatory networks, TF X regulates another TF Y and both jointly modulate the expression of the target gene Z. Hence TF X can be considered to be in the first position while TF Y can be thought to be in the second position. (**b**) Matrix shows statistical significance for occurrence of different sensing category combinations in FFLs using the z-scores calculated by comparing against 1000 sets of randomly generated FFLs as described in Materials and Methods section. Positive z-scores correspond to favored combinations of sensing classes in FFLs and *vice versa*. $|z\text{-scores}| > 3.3$ were considered significant as they corresponded to P-values <0.001, unless otherwise stated.

Table 1. Frequency distribution and statistical significance for TFs from different sensing classes to occupy the first and second positions of a FFL motif.

Sensing class	Position 1		Position 2	
	Proportion	*z*-score (*P*-value)	Proportion	*z*-score (*P*-value)
ISM	0.6464	14.394 (<0.001)	0.1533	−11.705 (<0.001)
IDB	0.1472	−1.353 (<0.076)	0.1730	0.502 (<0.282)
H	0.0440	−7.007 (<0.001)	0.2079	5.172 (<0.001)
ETC	0.1320	−5.821 (<0.001)	0.3126	5.083 (<0.001)
ETM	0.0303	−5.963 (<0.001)	0.1533	4.487 (<0.001)

First position corresponds to the TF with two outputs and second position corresponds to the TF with one output, both regulating the expression of a target gene.

is overrepresented by TFs from H, ETC and ETM and as expected suppressed for TFs from ISM. Interestingly, the IDB class, comprising of nucleoid-associated proteins responsible for controlling DNA topology and nucleoid organization, were not found to show any significant preference for the first or the second position however they were found to occur in a combinatorial fashion in FFLs (see below and Figure 1b). We then addressed the enrichment for combinatorial control by sensing classes in the entire pool of FFLs. To study this we compared against 1000 randomly generated FFL sets, same in size as the observed set, by shuffling TFs between pairs of FFLs as described in detail in Materials and Methods section. Figure 1b shows the pair wise combinations of sensing classes that were significantly enriched for positions 1 and 2 in FFLs. It should be noted that individual tendencies for either positions need not be the same as the pair-wise combinations in FFLs. We observed a clear and strong preference for the following series of combinations: ISM-H ($P < 0.001$), ISM-ETM ($P < 0.001$), ISM-ETC ($P < 0.001$), IDB-IDB ($P < 0.001$) and H-H ($P < 0.004$), suggesting that the first position of the FFL is preferentially occupied by ISM TFs when the second position is occupied by Hybrid (H) or one of the external class of TFs (ETM or ETC). These observations clearly suggest a strong coordination between the internal and external classes of TFs in FFLs. It is also interesting to note from this heatmap that IDB TFs strongly co-regulate their target promoters with only other of their kind. Similarly H TFs were also found to show this tendency indicating that these classes act independently and form self-consistent local structures. On the other hand, several other combinations did not show any preference for combinatorial control. In particular, we found that the combinations ISM-ISM, IDB-ISM, H-ISM, ETM-ISM, H-IDB, ETM-IDB, ETC-IDB, ETM-H, ETC-H, H-ETM, H-ETC, ETC-ETM, IDB-ETC and H-ETC were significantly underrepresented ($P < 0.001$) for the first and second TF positions, reinforcing that, although internal TFs dominantly occupy the first position of the FFL, they do not control their promoters independently but rather in coordination with the help of the external TFs. These observations also suggest that TFs sensing external signals almost never control FFLs i.e. they are not in the first position, but are mostly under the control of internal sensing TFs. Interestingly, neither IDB nor H, which can sense signals of internal origin, control the core internal ISM TFs when the later takes the second position, suggesting that neither nucleoid associated nor hybrid TFs start a FFL in coordination to responses from other kinds of sensing classes. However, IDB and H TFs tend to co-ordinate with TFs of the same class. It can also be noted that H and ETC or H and ETM TFs never work together in a FFL which is likely due to the fact that all H TFs, by definition, can sense signals of both internal and external origin and hence do not need any explicit co-ordination with TFs that only sense external signals.

It is interesting to note that in the composition of the FFLs, CRP (cAMP receptor protein) and FNR (fumarate and nitrate reduction regulatory protein) which are global regulators in *E. coli* (30), are the starting TFs in more than 60% of all the identified TF combinations. These TFs partition the regulatory network through FFLs in two different ways: FNR co-ordinates its activity almost exclusively with ArcA and NarL while CRP forms FFL motifs with TFs either sensing transported metabolites, hybrid or DNA-bending TFs. It is important to note how the external sensing TFs are differentially forming FFL with internal sensing TFs; FNR almost exclusively with two TFs of two-component systems (ArcA and NarL) and CRP forming FFLs mostly with TFs that use transported metabolites. Thus we can say that FNR almost exclusively coordinates the respiration mode through two TFs sensing external compounds for electron receptors while CRP coordinates particularly the uptake of biodegradable carbon sources. DNA-bending TFs and H TFs form FFL motifs entirely constituted by regulators of their own class; the former in a hierarchical order using the TFs IHF, FIS and HNS and the later using the TF combinations *tdcR-tdcA*, *galR-galS* and *gntR-idnR*. In line with these observations, a complex network of tightly co-ordinated interactions among nucleoid associated (IDB) TFs forming interdependent feedback loops is believed to play an important role in DNA supercoiling in enterobacterial genomes possibly explaining the reason for these TFs to form self-consistent local structures (31).

Congruent and incongruent FFLs show similar sensing class distributions

To unravel the function of the FFLs, one needs to understand how X and Y are integrated to regulate the promoter upstream of Z. Since each of the three regulatory interactions in the FFL can be either activation or repression, there are eight possible structural types of FFL. These eight types can in turn be divided into coherent or incoherent FFLs, depending on whether the sign of the direct path from TF X to output Z is the same as the overall sign of the indirect path through transcription factor Y or the two paths have opposite signs (28). Previous studies showed that two of these eight motif subtypes, namely coherent type-1 and incoherent type-1, occur much more predominantly in the TRNs of *E. coli* and yeast (28,32). While the former was shown to possess

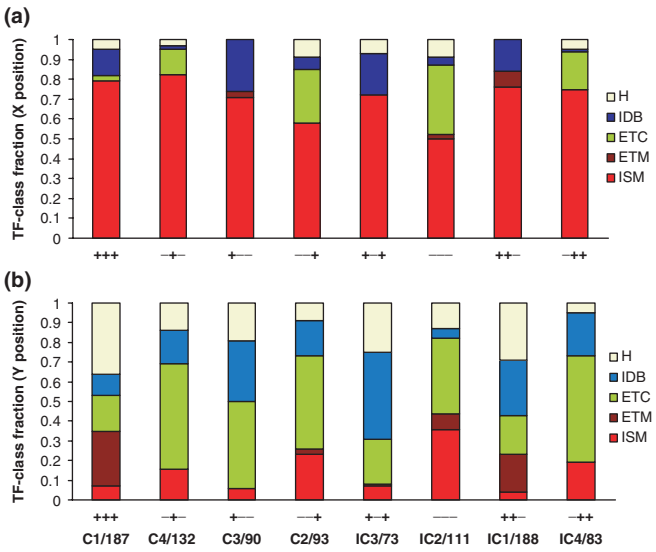

Figure 2. Distribution of different sensing classes in coherent and incoherent FFL motif sub-types identified in *E. coli*. (**a**) Proportion of different sensing categories among the TFs occupying the first position of the different FFL sub-types. (**b**) Proportion of different sensing categories among the TFs occupying the second position for all the FFL sub-types. FFL motif sub-types are named according to Mangan and Alon (28). C1-4 correspond to the coherent FFL types while IC1-4 correspond to the incoherent motif types, described earlier. For instance in a FFL if TF X regulates the activity of the genes Y and Z, while TF Y regulates Z, the sign $-+$ in the figure corresponds to the repression of Y and Z by X while activation of Z by Y thus representing the regulatory links between the gene pairs XY, XZ and YZ, respectively. Numbers below each motif type show the abundance of motifs of that kind identified in the TRN.

the property of producing a delay in the initial response when the input function at the Z promoter is AND, the later was demonstrated to function as response accelerator and pulse generator of the Z promoter (32,33). Given these observations and understanding of the motif dynamics we sought to address the composition of the sensing classes for the first and second position of different motif subtypes. Figure 2 shows the distribution of different sensing classes for the first and second position of the eight FFL motif subtypes identified in the TRN of *E. coli* (see Materials and Methods section). We found that apart from the coherent and incoherent type-1 subtypes of FFLs, coherent type-4 and incoherent type-2 motif subtypes are also dominant in the currently known TRN of *E. coli*. Although previous theoretical works have shown that increased effective cooperativity of the coherent type-1 FFL could be evolutionarily advantageous and selected for due to its capability to reduce noise propagation associated with the input signal, no strong theoretical rational could be arrived at for the prevalence of incoherent FFLs (28,34,35). Therefore, it is possible that other motif subtypes which are also found to be prevalent in the TRN of *E. coli* have important functions which are yet to be explored in detail both theoretically and experimentally. It is interesting to note that four motif subtypes, namely coherent and incoherent types 1 and 3 clearly show a preference for IDB TFs in the first position and about 50% of the TFs in their respective second positions are occupied by H and IDB classes put together, consistent with previous observation that IDB TFs frequently coregulate their targets in conjunction with either IDB or H TFs. It is also evident that in only coherent and incoherent types 2 and 4, ETC TFs are mostly found in the first position. Curiously, the same sensing class is also enriched in the second position for these motif subtypes. From the perspective of the second position, it is worth noting that IDB TFs show a preference to occur in coherent and incoherent type-3 FFLs, given their number of instances while ISM TFs appear more commonly in the coherent and incoherent type-2 motif subtypes. ETM TFs, which sense external transported metabolites and were found to be significantly co-occurring with the ISM TFs in FFLs show a strong tendency to occur in the second position of coherent and incoherent type-1 FFLs. Despite the sensing classification of the TFs, which is based on literature evidence about the physiological role of the TFs and the motif structures, which are based on their non-random occurrence in the TRNs, being very different they still show tendencies for similar distributions in the corresponding coherent and incoherent motif subtypes. For instance, in several cases discussed above similar coherent and incoherent subtypes show very similar preferences for sensing classes in both TF positions, suggesting that the mode of action of the TF (activation or repression) in the second position has little influence in their sensing class distribution. This is especially curious to note given that most TFs occupying the second position of the FFL are not dual regulators but rather one of the other two kinds of regulators. A possible explanation for the observed tendencies is that the second TF (Y) of the FFLs in coherent and incoherent types varies from condition to condition, depending on the available metabolites exterior to the cell. For instance, alternative metabolites to the cell like galactose could be degraded rapidly in the absence of core metabolites using an incoherent system while coherent system with the help of its initial response delay can wait for a persistent external signal to degrade the available metabolites, as has been demonstrated in the case of the arabinose system in *E. coli*, thereby providing a defined order for the degradation of different substrates (32,33). This could also imply that coherent type could be used for uptake of metabolites which are most commonly available in the cell's natural environment as they could have been tuned for low cellular noise due to a persistent signal, while incoherent types could be used for optional metabolites that need to be degraded by the cell under starvation conditions.

Bi-fan motifs

A Bifan motif is composed of two TFs, A and B, which both control the expression of two different target genes C and D (see Figure 3a). Thus, unlike a FFL motif, the Bifan motif is not hierarchical but rather forms a horizontal layer of interactions. In fact, bifan motifs are a particular subset of complex regulons. A simple regulon being a group of genes regulated by one regulator, and complex regulons, groups of genes regulated by the same set of two

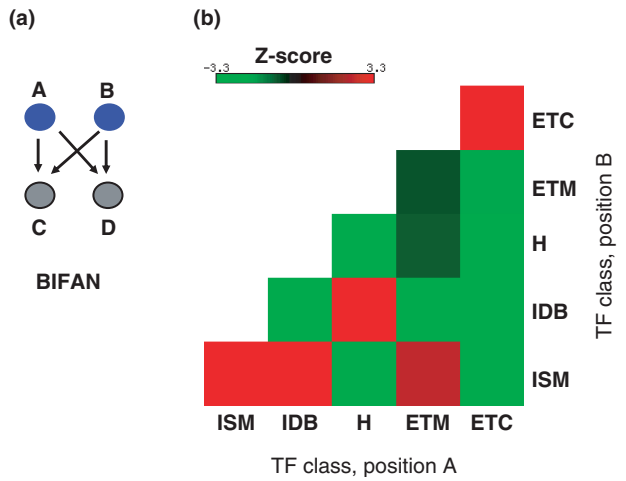

Figure 3. Statistical significance of combination of sensing classes in bifan motifs observed in the transcriptional network of *E. coli*. (**a**) In a bifan motif, two TFs A and B, both regulate the expression of two different target genes C and D. Thus making the positions of the TFs to be symmetric, unlike in FFLs. (**b**) Matrix shows statistical significance for occurrence of different sensing category combinations in bifans using the *z*-scores calculated by comparing against 1000 sets of randomly generated bifans as described in Materials and Methods section. Notice that since the positions of the TFs are not relevant only a lower-triangular matrix is shown. Positive *z*-scores correspond to favored combinations of sensing classes in bifan motifs and *vice versa*. |*z*-scores| >3.3 were considered significant as they corresponded to *P*-values <0.001, unless otherwise stated.

or more TFs. The structure of the bifan motif makes the two positions of the TFs symmetric in contrast to the organization of FFLs. Bifan motifs are essential to maintain the network backbone and link it in a horizontal way by connecting across transcriptional regulatory modules (9,36). Figure 3b shows the *z*-score matrix for co-occurrence of TFs from different sensing classes to appear in bifan motifs identified in the TRN of *E. coli* (see Materials and Methods section). A clear and strong tendency of coregulation was found between the following pairs of sensing classes: ISM-ISM, ISM-IDB, IDB-H and ETC-ETC ($P < 0.001$ in each case) suggesting that unlike in FFLs, there is a preference for internal sensing classes to coregulate their targets with other internal TFs and external sensing ETC TFs to control their targets with other TFs of their own class. The enrichment seen in two-component systems to frequently occur together in bifans, i.e regulating the same set of targets, might be a means of feeding multiple external signals, each corresponding to different environmental conditions, as inputs to the interior of the cell. This observation implies that bifans are more homogenous in their sensing class composition compared to FFLs and do not link the external signals using the two component systems with internal TFs, but rather could link to internal machinery with signal transduction cascades. However, bifans do link the ETM with ISM TFs ($P < 0.005$). On the other hand, the following sensing class combinations showed no preference to coregulate their targets: ISM-H, ISM-ETC, IDB-IDB, IDB-ETM, IDB-ETC, H-H and H-ETC ($P < 0.001$). IDB and H TFs, which showed strong preference to occur together with other of their kind in FFLs were found to show avoidance to appear together in bifans. This observation implies that bifans are used by IDB and H TFs to connect to each other, while FFLs are used by these TF classes to link among themselves in a hierarchical fashion. It is possible to speculate from these observations that DNA-bending TFs (IDB class) and Hybrid (H) class of TFs combinatorially regulate their targets with the help of TFs of their own class in FFLs, but at the same time also coordinate and integrate their signal responses in a horizontal way with the help of bifan motifs. Taken together, our results suggest that FFLs and bifans distribute the sensing classes of TFs in very distinct ways. While FFLs link internal and external sensing machinery in a hierarchical fashion, bifans show a tendency to connect internal or external sensing classes among themselves in a horizontal manner.

SIMs are enriched for IDB class

SIMs form yet another class of local network structures which appear more frequently than expected by chance in TRNs (12). A SIM consists of a very simple pattern in which a transcriptional regulator X regulates a series of target genes Z_1 to Z_n. In a strict sense, no other TF regulates any of its target genes in a SIM, however the TF often autoregulates its own activity. The major function of this motif is to allow coordinated expression of a group of genes with related functions. This typically happens by generating a temporal expression program, with a defined order of activation of each of the target promoters. X often has different activation thresholds for each gene, owing to variations in the sequence and context of its binding site in each promoter. So, when X activity rises gradually with time, it crosses these thresholds in a defined order, resulting in a temporal order of expression. This has been demonstrated in the arginine-biosynthesis system in which the repressor ArgR regulates several operons that encode enzymes in the arginine-biosynthesis pathway. The order of activation was found to match the position of the enzymes in the arginine pathway (37). We found that the TFs from internal sensing classes especially those of IDB showed a preference to form part of SIMs, possibly suggesting that IDB TFs besides coregulating with H TFs in bifans also regulate their targets independently as has been previously observed in FFL organization (Table 2). However, this observation should be taken with caution, as the network could be under-represented by interactions from local and external class of TFs due to the fact that it is far from complete. Nevertheless, this result is interesting as each protein of the nucleoid associated (IDB) class tends to be preferentially present in some phase of bacterial population growth; e.g. FIS at early exponential, HNS during exponential and IHF during early stationary growth phase (38). So the sequential expression of genes assisted by these nucleoid-associated proteins in various stages of growth might indicate their temporal expression depending only on nucleoid organization in parallel to those genes depending on the presence of exogenous and endogenous signals.

Table 2. Abundance and statistical significance of occurrence of TFs from different sensing classes in SIMs

Sensing class	Abundance	*z*-score (*P*-value)
ISM	12	2.021 (<0.016)
IDB	4	2.766 (<0.003)
H	7	−1.203 (<0.071)
ETC	8	−0.079 (<0.388)
ETM	4	−1.818 (<0.014)

A total of 35 SIMs of those identified in the regulatory network of *E. coli* could be associated to sensing classes.

Modules are heterogeneous in sensing context

Motifs, topological units in regulatory networks, which control the local dynamic behavior of transcriptional regulation, do not exist in isolation but aggregate into motif clusters or modules (9,36). However, little is known about the components which form modules, subsets of genes that integrate cellular responses in a given condition. In several recent works it has been reported that the TRN of *E. coli* is a scale-free hierarchial network organized into modules with no feedback regulation at the level of transcription, from the bottom to the top of the hierarchy (7,36,39). So we sought to ask if modules show any preferences in their distribution of sensing classes. Figure 4 shows the distribution of sensing classes among the eight modules identified in the regulatory network according to the shortest path distance metric among the TFs (9) (see Materials and Methods section). Although all the modules are shown for completeness, only the modules M1, M2, M3, M4.1 and M4.6 could be associated with at least three TFs of known sensing class. Therefore, we based our observations and discussion based on them. In all these modules we found that there is a strong coordination between internal and external classes of sensing although the specific classes which dominate and coordinate to work together in each module seem to depend on the physiological role of the module. For instance, the modules M4.1 and M4.6 which were found to be related to 'carbon sources assimilation' showed an integration of ISM and ETM classes with H or IDB TFs, suggesting that in *E. coli*, carbon sources are mostly transported and degraded with the integration of signals from transported metabolites and internally synthesized metabolites, thereby linking extracellular changing environment with intracellular conditions. Similarly, in stress response conditions which is the physiological function associated with the module M2, there is a coordination between internal TFs, which can sense metabolites synthesized in the interior of the cell with those of TFs, which can sense metabolites transported into the cell. On the other hand, module M3, which is known to have a functional role in chemotaxis, motility and biofilm formation was found to be composed of DNA-bending TFs of internal class and two-component systems which sense external conditions. Module M1, which comprises of TFs playing role in aerobic and/or anaerobic states of the cell was found to show a clear dominance for TFs from external sensing classes, ETM and ETC, implying that

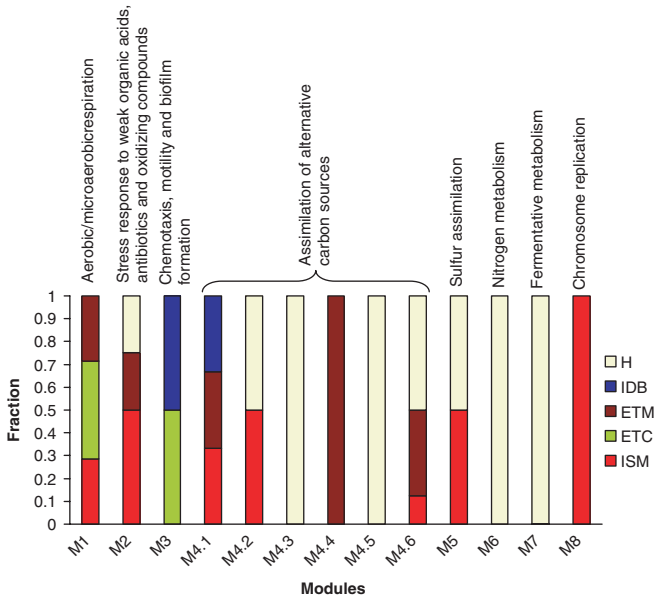

Figure 4. Distribution of sensing classes in the modules identified in the transcriptional network of *E. coli*. Modules were identified as described earlier (9), using the regulatory interactions among TFs and clustering the resulting network using the shortest path distance metric. A total of eight modules were identified although not all TFs could be associated to sensing classes. After mapping to the sensing classification modules M4.2, M4.3, M4.5, M5 and M8 comprised of two TFs each, modules M4.1 had 3, M2 and M3 were left with four, while M1 and M4.6 were found to consist of seven and eight TFs, respectively. Other modules (M4.4, M6 and M7) contained only one TF each and hence were bound to be homogenous in sensing context. However most of the modules with atleast four TFs can be seen to be heterogenous in sensing context with a clear combination of internal and external classes.

aerobic and anaerobic respiration response in *E. coli* is mostly controlled by TFs which can sense external signals. Based on our results one can speculate that regulatory modules in *E. coli* are not homogenous in their sensing categories but rather composed of a mix of TFs which can sense signals from both internal and external origin independent of their physiological role.

Distribution of sensing classes in promoter regions which are regulated by more than two TFs

Understanding network motifs, which are topological structures controlling the local expression dynamics in the cell, although is of great interest, studies on them is often limited to combinatorial regulation of a promoter by no more than two TFs. However, to appreciate the effect of the action of multiple TFs on a single promoter in the context of sensing classes, one has to study the frequency distribution of TFs from different sensing classes when promoters are regulated by one or more TFs. Indeed control of transcription by multiple TFs at a promoter has been an area of immense interest in itself due to a variety of possible mechanisms by which TFs can combinatorially control the expression of a gene (3,19). As seen in Figure 5a, very few promoters are regulated by four or more TFs suggesting limitations on the number of different binding sites possible in the already short

intergenic regions in prokaryotic organisms. A possible explanation for this limitation could be the structural and functional constraints imposed in the intergenic regions of bacterial genomes due to restrictions on the sizes of the protein complexes formed during transcription.

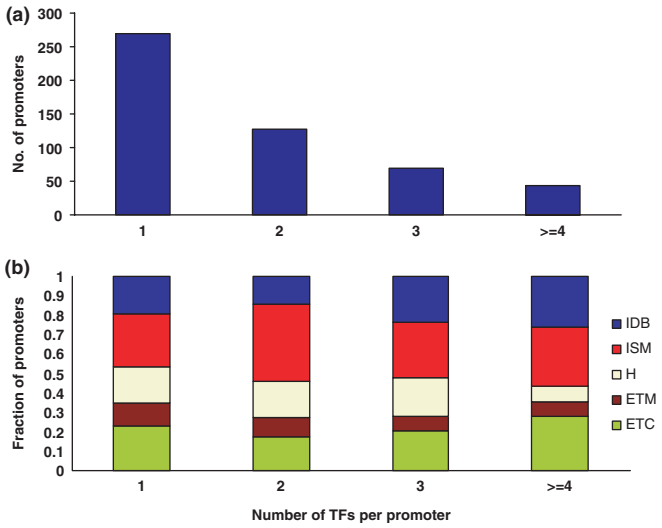

Figure 5. Distribution of sensing classes of TFs regulating the promoter regions upstream of experimentally known operons in *E. coli*. (**a**) Number of promoters regulated by a given number of transcription factors. (**b**) Proportion of different sensing classes transcriptionally controlling the expression of operons in *E. coli* at various thresholds of the number of TFs controlling an operon. To represent, for each bin, first a vector showing the proportion of sensing classes for each operon was calculated and then an average was obtained over the total number of operons present in a given bin, thus obtaining the occurrence of a sensing class. As the number of TFs regulating an operon increased proportion of ISM and hybrid (H) TFs showed a tendency to decrease, while the external TFs from two-component systems increased slightly. Interestingly, IDB TFs were also found to increase their propensity as the number of TFs increased.

Figure 5b shows the distribution of sensing classes when the promoters are regulated by only one, two, three and four or more TFs. It is evident from this figure that as the number of TFs regulating a promoter increases the proportion of TFs of internal origin remain constant (ISM and IDB taken together form about 50% of the TFs regulating a promoter) while the ETC TFs show an inclination to increase suggesting that promoters with several TFs regulating them could show a propensity for external TFs after a saturation threshold of the number of internal TFs controlling them. This observation implies that certain transcription units with multiple inputs can be used under a variety of exogenous conditions depending on the external TFs which modulate their activity particular to a condition.

Distribution of sensing classes in evolutionary families

Most prokaryotic TFs are multi-domain proteins, typically composed of a DNA-binding domain along with a signaling small molecule-binding domain. Since the majority of the bacterial TFs can be classified into evolutionary families based on their helix–turn–helix DNA-binding domain, we identified a total of 13 evolutionary families of TFs for which sensing classes could be assigned (see Materials and Methods section). One could expect that TFs belonging to a common evolutionary family might be composed of the same sensing class as they might have evolved from a common DNA-binding domain. However, as can be seen from Figure 6, this does not appear to be the case. Most families which contain at least three TFs are composed of an extensive mix of TFs from different sensing classes indicating that TFs of the same evolutionary family need not belong to the same class of sensing. For example, four major families LacI, LuxR, AraC and GntR in this figure do not clearly show enrichment for one or the other kind of sensing category.

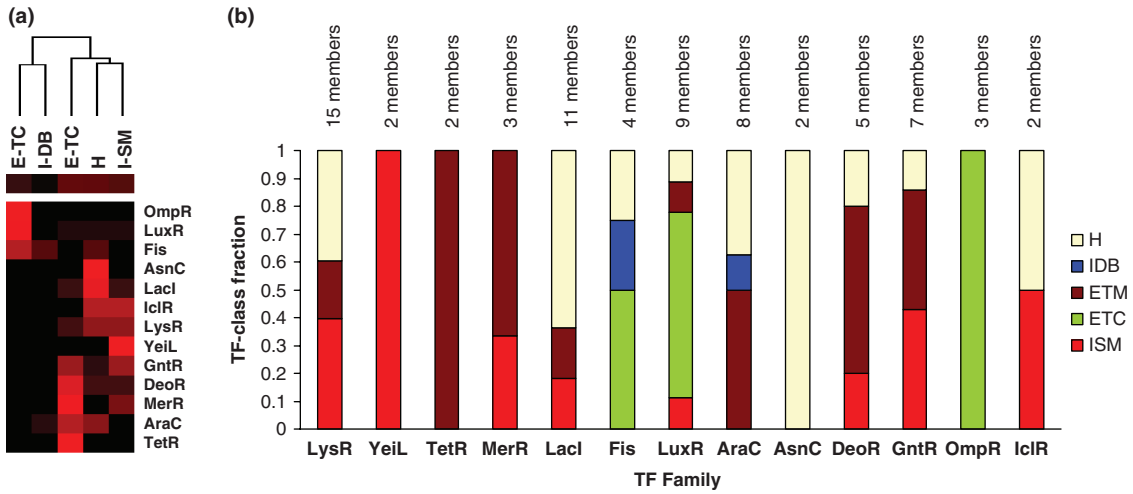

Figure 6. Distribution of different sensing categories seen in 13 evolutionary families of transcription factors. Only those families in which atleast two of the transcription factors could be associated with a sensing class are shown. (**a**) Heatmap showing the fraction of TFs in a given family belonging to different sensing classes. A dark red corresponds to 1.0 while a black color cell represents no TF associated to a class. Families and sensing classes are clustered hierarchically to view similarities in distributions. (**b**) Histogram showing the proportion of different sensing classes in each evolutionary family calculated against the total number of TFs of a family assigned to sensing classes. Certain families like IclR, AsnC, TetR were found to contain only two TFs which could be associated to sensing classes, while LacI, LuxR, AraC and GntR comprised of 11, 9, 8 and 7 TFs, respectively. The number of TFs in all other families was between 2 and 6.

These observations can be explained under the premise that the sensing class of a TF could actually depend on its signaling domain rather than its DNA-binding domain and hence, although two TFs can belong to the same evolutionary family they could still correspond to different classes as their ability to respond to the signals will depend on their signaling domain. In addition, it is now well accepted that there are extensive variations and frequent recombinations and rearrangements occurring in the signaling domains of TFs having the same DNA-binding domain, suggesting that unless TFs of an evolutionary family are recent duplicates, in which case they might conserve their signaling domain, it is unlikely that they still preserve their sensing mechanism (23,24,40). It can also be seen from the hierarchical clustering of sensing classes in Figure 6a that the majority of the TF families show a tendency to come from one of the following two combinations of sensing classes: ISM and H appearing with ETM or IDB appearing with ETC in a given family. This suggests a higher order relationship in the evolution of sensing mechanisms in TFs—the former indicating a link among those sensing endogenous and endogenous metabolites and the later linking nucleoid-associated TFs with two-component systems.

DISCUSSION AND CONCLUSIONS

The study presented here provides an understanding of how a simple unicellular model bacterium like *E. coli* partitions its regulatory network components like motifs and modules in response to changing exogenous and endogenous conditions based on a previously proposed classification of TFs, relying on the location of the origin of the signals, which control the activity of the TFs. We found that although the transcriptional response is mainly coordinated by internal TFs, their participation in various network structures is clearly dependent on their activity; internal global TFs sensing cAMP (CRP) and redox potential (FNR) were found to be initiating most of the FFL motifs while those sensing external signals frequently occupied the second position of the FFLs, indicating a clear partitioning and integration between internal and external sensing TFs. In the most frequently occurring form of FFLs, both TFs and their effector signals should be present to turn on the system avoiding response to transient changes, however the system would quickly shut off when one of the signals is absent or in low levels (33). The first TF might be responding to persistent or general signals to be turned on and therefore this motif might be designed to respond to a particular or transient conditions sensed by the second TF. Our observations suggest that while FFL motifs integrate internal and external signals, bifan motifs maintain connectivity and cohesion in the network by integrating signals in a horizontal manner across transcriptional modules. It is possible to envisage based on our observations that bifans connect TFs across different members of a layer in the hierarchical structure of the TRN, linking global regulators with other global regulators and local regulators with other local regulators while FFLs connect down the hierarchy connecting globals or other highly connected TFs with local regulators.

Our results hint that long-time cellular memory by means of mainly intracellular global TFs working in concert with short-time memory systems (local TFs sensing environmental conditions) could be a common theme of bacterial TRNs. In other words, one main program might always be working depending on the internal status of the cell while the sporadic ones follow required subroutines (41). In such an organization, it is possible to envisage the advantages of signal transduction between the signal and the TF gene products by an integrated mechanism with a quick response time, which is feasible in prokaryotes by keeping the gene products of the signal gene and the corresponding TF physically close on the genome (5,42). This may contribute to the possibility of having the two gene products spatially close in the cellular compartment which would require lesser concentrations for driving local dynamics. A second major advantage of such a hierarchical organization with local modules, is an ability of a cell to respond quickly to fluctuant extracellular responses using these local modules, while persistent global signals control TFs sensing intracellular conditions.

We found that transcriptional modules and evolutionary families of TFs in *E. coli* are not homogenous in terms of sensing classes, instead are often comprised of a mix of internal and external sensing TFs, suggesting that most physiological roles of the cell use TFs which sense both internal and external conditions and that the DNA-binding domain of the TF does not govern its sensing class. The later observation indicates the need to understand the sensing ability of TFs when their signaling domains are conserved across TFs, as this would not only aid the prediction of the sensing mode of a TF but also better our understanding on the principles that direct the evolution of sensing mechanisms in TFs.

SUPPLEMENTARY DATA

Supplementary Data available at http://www.ccg.unam.mx/Computational_Genomics/TRNS/TopologyvsCellsensing/

The regulatory sub-systems classified according to the location of signal metabolites, used for the entire analysis and additional figures can be obtained from:

http://www.ccg.unam.mx/Computational_Genomics/regulondb/CellSensing/

ACKNOWLEDGEMENTS

We thank Arthur Wuster and Gabriel Moreno-Hagelsieb for providing helpful suggestions and comments on previous versions of this manuscript. This work was partially supported by NIH grant RO1 GM 071962. S.C.J. has been supported by grants given to J.C.V. Funding to pay the Open Access publication charges for this article was provided by NIH grant GM071962.

Conflict of interest statement. None declared.

REFERENCES

1. Luscombe,N.M., Babu,M.M., Yu,H., Snyder,M., Teichmann,S.A. and Gerstein,M. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, **431**, 308–312.
2. Balazsi,G. and Oltvai,Z.N. (2005) Sensing your surroundings: how transcription-regulatory networks of the cell discern environmental signals. *Sci. STKE*, **2005**, e20.
3. Browning,D.F. and Busby,S.J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, **2**, 57–65.
4. Martinez-Antonio,A., Janga,S.C., Salgado,H. and Collado-Vides,J. (2006) Internal-sensing machinery directs the activity of the regulatory network in Escherichia coli. *Trends Microbiol.*, **14**, 22–27.
5. Janga,S.C., Salgado,H., Collado-Vides,J. and Martinez-Antonio,A. (2007) Internal versus external effector and transcription factor gene pairs differ in their relative chromosomal position in Escherichia coli. *J. Mol. Biol.*, **368**, 263–272.
6. Ma,H.W., Buer,J. and Zeng,A.P. (2004) Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics*, **5**, 199.
7. Yu,H. and Gerstein,M. (2006) Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl Acad. Sci. USA*, **103**, 14724–14731.
8. Ihmels,J., Friedlander,G., Bergmann,S., Sarig,O., Ziv,Y. and Barkai,N. (2002) Revealing modular organization in the yeast transcriptional network. *Nat. Genet.*, **31**, 370–377.
9. Resendis-Antonio,O., Freyre-Gonzalez,J.A., Menchaca-Mendez,R., Gutierrez-Rios,R.M., Martinez-Antonio,A., Avila-Sanchez,C. and Collado-Vides,J. (2005) Modular analysis of the transcriptional regulatory network of E. coli. *Trends Genet.*, **21**, 16–20.
10. Segal,E., Shapira,M., Regev,A., Pe'er,D., Botstein,D., Koller,D. and Friedman,N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, **34**, 166–176.
11. Bar-Joseph,Z., Gerber,G.K., Lee,T.I., Rinaldi,N.J., Yoo,J.Y., Robert,F., Gordon,D.B., Fraenkel,E., Jaakkola,T.S. *et al.* (2003) Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.*, **21**, 1337–1342.
12. Shen-Orr,S.S., Milo,R., Mangan,S. and Alon,U. (2002) Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.*, **31**, 64–68.
13. Ma,H.W., Kumar,B., Ditges,U., Gunzer,F., Buer,J. and Zeng,A.P. (2004) An extended transcriptional regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res.*, **32**, 6643–6649.
14. Alon,U. (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, **8**, 450–461.
15. Collado-Vides,J., Magasanik,B. and Gralla,J.D. (1991) Control site location and transcriptional regulation in Escherichia coli. *Microbiol. Rev.*, **55**, 371–394.
16. Harbison,C.T., Gordon,D.B., Lee,T.I., Rinaldi,N.J., Macisaac,K.D., Danford,T.W., Hannett,N.M., Tagne,J.B., Reynolds,D.B. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.
17. Wang,T. and Stormo,G.D. (2005) Identifying the conserved network of cis-regulatory sites of a eukaryotic genome. *Proc. Natl Acad. Sci. USA*, **102**, 17400–17405.
18. Balazsi,G., Barabasi,A.L. and Oltvai,Z.N. (2005) Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli. *Proc. Natl Acad. Sci. USA*, **102**, 7841–7846.
19. Hermsen,R., Tans,S. and Wolde,P.R. (2006) Transcriptional Regulation by Competing Transcription Factor Modules. *PLoS Comput. Biol.*, **2**, e164.
20. Balaji,S., Babu,M.M. and Aravind,L. (2007) Interplay between network structures, regulatory modes and sensing mechanisms of transcription factors in the transcriptional regulatory network of E. coli. *J. Mol. Biol.* doi: 10.1016/j.jmb.2007.06.084.
21. Salgado,H., Gama-Castro,S., Peralta-Gil,M., Diaz-Peredo,E., Sanchez-Solano,F., Santos-Zavaleta,A., Martinez-Flores,I., Jimenez-Jacinto,V., Bonavides-Martinez,C. *et al.* (2006) RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.*, **34**, D394–397.
22. Perez-Rueda,E. and Collado-Vides,J. (2000) The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Res.*, **28**, 1838–1847.
23. Madan Babu,M. and Teichmann,S.A. (2003) Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Res.*, **31**, 1234–1244.
24. Aravind,L., Anantharaman,V., Balaji,S., Babu,M.M. and Iyer,L.M. (2005) The many faces of the helix–turn–helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231–262.
25. Wall,M.E., Hlavacek,W.S. and Savageau,M.A. (2004) Design of gene circuits: lessons from bacteria. *Nat. Rev. Genet.*, **5**, 34–42.
26. Milo,R., Shen-Orr,S., Itzkovitz,S., Kashtan,N., Chklovskii,D. and Alon,U. (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.
27. Gutierrez-Rios,R.M., Rosenblueth,D.A., Loza,J.A., Huerta,A.M., Glasner,J.D., Blattner,F.R. and Collado-Vides,J. (2003) Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles. *Genome Res.*, **13**, 2435–2443.
28. Mangan,S. and Alon,U. (2003) Structure and function of the feed-forward loop network motif. *Proc. Natl Acad. Sci. USA*, **100**, 11980–11985.
29. Wolf,D.M. and Arkin,A.P. (2003) Motifs, modules and games in bacteria. *Curr. Opin. Microbiol.*, **6**, 125–134.
30. Martinez-Antonio,A. and Collado-Vides,J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.*, **6**, 482–489.
31. Travers,A. and Muskhelishvili,G. (2005) DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.*, **3**, 157–169.
32. Mangan,S., Itzkovitz,S., Zaslaver,A. and Alon,U. (2006) The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli. *J. Mol. Biol.*, **356**, 1073–1081.
33. Mangan,S., Zaslaver,A. and Alon,U. (2003) The coherent feed-forward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.*, **334**, 197–204.
34. Thattai,M. and van Oudenaarden,A. (2002) Attenuation of noise in ultrasensitive signaling cascades. *Biophys. J.*, **82**, 2943–2950.
35. Ghosh,B., Karmakar,R. and Bose,I. (2005) Noise characteristics of feed forward loops. *Phys. Biol.*, **2**, 36–45.
36. Dobrin,R., Beg,Q.K., Barabasi,A.L. and Oltvai,Z.N. (2004) Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network. *BMC Bioinformatics*, **5**, 10.
37. Zaslaver,A., Mayo,A.E., Rosenberg,R., Bashkin,P., Sberro,H., Tsalyuk,M., Surette,M.G. and Alon,U. (2004) Just-in-time transcription program in metabolic pathways. *Nat. Genet.*, **36**, 486–491.
38. Luijsterburg,M.S., Noom,M.C., Wuite,G.J. and Dame,R.T. (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.*, **156**, 262–272.
39. Lagomarsino,M.C., Jona,P., Bassetti,B. and Isambert,H. (2007) Hierarchy and feedback in the evolution of the Escherichia coli transcription network. *Proc. Natl Acad. Sci. USA*, **104**, 5516–5520.
40. Anantharaman,V., Koonin,E.V. and Aravind,L. (2001) Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *J. Mol. Biol.*, **307**, 1271–1292.
41. McAdams,H.H. and Arkin,A. (1998) Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 199–224.
42. Menchaca-Mendez,R., Janga,S.C. and Collado-Vides,J. (2005) The network of transcriptional interactions imposes linear constrains in the genome. *Omics*, **9**, 139–145.