Article

Escherichia coli transcriptional regulatory network

Agustino Martínez-Antonio
Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. 36500 Irapuato, Guanajuato, México
E-mail: amartinez@ira.cinvestav.mx

Received 27 February 2011; *Accepted 28 March 2011*; *Published online 10 May 2011*
IAEES

Abstract
Escherichia coli is the most well-know bacterial model about the function of its molecular components. In this review are presented several structural and functional aspects of their transcriptional regulatory network constituted by transcription factors and target genes. The network discussed here represent to 1531 genes and 3421 regulatory interactions. This network shows a power-law distribution with a few global regulators and most of genes poorly connected. 176 of genes in the network correspond to transcription factors, which form a sub-network of seven hierarchical layers where global regulators tend to be set in superior layers while local regulators are located in the lower ones. There is a small set of proteins know as nucleoid-associated proteins, which are in a high cellular concentrations and reshape the nucleoid structure to influence the running of global transcriptional programs, to this mode of regulation is named analog regulation. Specific signal effectors assist the activity of most of transcription factors in *E. coli*. These effectors switch and tune the activity of transcription factors. To this type of regulation, depending of environmental signals is named the digital–precise-regulation. The integration of regulatory programs have place in the promoter region of transcription units where it is common to observe co-regulation among global and local TFs as well as of TFs sensing exogenous and endogenous conditions. The mechanistic logic to understand the harmonious operation of regulatory programs in the network should consider the globalism of TFs, their signal perceived, co-regulation, genome position, and cellular concentration. Finally, duplicated TFs and their horizontal transfer influence the evolvability of members of the network. The most duplicated and transferred TFs are located in the network periphery.

Keywords *Escherichia coli*; transcriptional network; network hierarchy.

1 Introduction to the Basic Biology of *Escherichia coli*

Escherichia coli (*E. coli*) is a Gram-negative, facultative anaerobic and non-sporulating bacterium. Cells are typically rod-shaped and are about 2 micrometers (μm) long and 0.5 μm in diameter, with a cell volume of 0.6 - 0.7 μm^3. *E. coli* was discovered by German pediatrician and bacteriologist Theodor Escherich in 1885 and is now classified as part of the Enterobacteriaceae family of gamma-proteobacteria (Blattner et al., 1997). Optimal growth of *E. coli* occurs at 37°C but some laboratory strains can multiply at temperatures of up to 49°C. This bacterium is commonly found in the lower intestine of warm-blooded organisms. *E. coli* normally colonizes an infant's gastrointestinal tract within 40 hours of birth, arriving with food or water or with the

individuals handling the child. Most *E. coli* strains are harmless, but some, such as serotype O157:H7, can cause serious food poisoning in humans (Ohnishi et al., 1999). The harmless strains are part of the normal flora of the gut, and can benefit their hosts by helping with assimilation of food, by providing some vitamins, and by preventing the establishment of pathogenic bacteria within the intestine (Steed et al., 2008).

Cultivated strains (e.g. *E. coli* K12) are well adapted to the laboratory environment, and, unlike wild type strains, have lost their ability to thrive in the intestine. In 1946, Joshua Lederberg and Edward Tatum first described the phenomenon known as bacterial conjugation (Lederberg and Tatum, 1953), and was an integral part of the first experiments to understand phage genetics, and early researchers, such as Seymour Benzer, used *E. coli* and phage T4 to understand the topography of gene structure (Miller et al., 2003). After extensive studies in bacteria and their phages, molecular biologists centered their investigations in bacteria, being *E. coli* the first-choice model. Studies in bacteria like *E. coli* set many of the modern fundaments of molecular biology. Today, we know the DNA sequence of about 40 genomes of *E. coli* strains, most of them pertaining to pathogens (Wirth et al., 2006). The *E. coli* pangenome (the total repertoire of genes in all the sequenced genomes of the genus) is thought to be as large as those of a human being (Rasko et al., 2008; Touchon et al., 2009). Nowadays *E. coli* continues to be the choice model organism driving pioneering studies as is the case where their functional genetic-elements –along with those of their phages- are the biological parts list –biobricks- used to construct new arrangements or forms of life by researches in the new field of synthetic biology (Anderson et al., 2010; Arkin, 2001; Benner, 2003).

2 Transcriptional Regulatory Machinery of *E. coli*

In spite that bacteria could be initially considered as a biochemical reactor inside a bag, conform we gain knowledge about its molecular components we realize that there is an order behind its functioning. *E. coli* K-12 (the most well-studied strain and to which I will refer in the rest of this review) have around 4,605 open reading frames (ORF) in a 4.6 Mpb genome long (Blattner et al., 1997). Its genes are located in 3,386 transcription units (TUs), (see Box 1 for additional description of *E. coli* K-12 bio-molecular components), (Gama-Castro et al., 2008). From a transcriptional point of view, the TU is the operative unit for gene expression as is delimited by promoter and terminator regions, which correspond to the start and terminus of transcription respectively. A TU could contain encoded one (monocistronic) or more (polycistronic) genes. Of the total genes in *E. coli* around 300 are predicted for encoding for transcription factors (TFs) and 7 for σ factors (all these type of genes correspond to ~7% of total genes), (Madan Babu and Teichmann, 2003; Perez-Rueda and Collado-Vides, 2000). Of the total of ~300 TF, 176 (60%) have experimental evidences for DNA-binding sites on TUs regulatory regions, while the seven σ have identified promoter on TUs as documented in RegulonDB (Box 1 and Table 1). The joining activity of TFs and σ are directly responsible for the *turn on* or *off* of genes as assist or obstructs the action of RNA polymerase on the promoters regions of TUs, (Browning and Busby, 2004).

3 Topological Properties of the *E. coli* Transcriptional Network

Here we will consider the regulatory network formed by transcription factors and their target genes (TGs) (Fig. 1). This indicate that for each interaction in the network there are experimental evidences about the existence of functional DNA-binding site(s) for the respective TF -where the arc start- onto the regulatory region of the TU encoding for the target gene that is draw in the network –where the arc finish-. In this way, the network in Fig. 1 contains 1531 genes (nodes) and 3421 regulatory interactions (arcs). Of the total of nodes, 176 correspond to TFs and the rest for structural and mRNAs genes. Taking into account the total number of genes, this entire network involves around one third of the total genes in *E. coli.* However it is not easy to calculate

which proportion of the total of regulatory interactions in *E. coli* represents the 3421 interactions between TFs and TGs. Since this network is formed by regulatory interactions from TF to TG it result in a directed graph with topological properties common to other biological networks (see Fig. 2 for the more common topological properties of this network). All biological networks described until date, have a node-degree distribution nearest to a power-law topology, which means there are few genes highly connected while most of them are poorly connected. The most highly connected nodes (HUBs in the network) correspond to

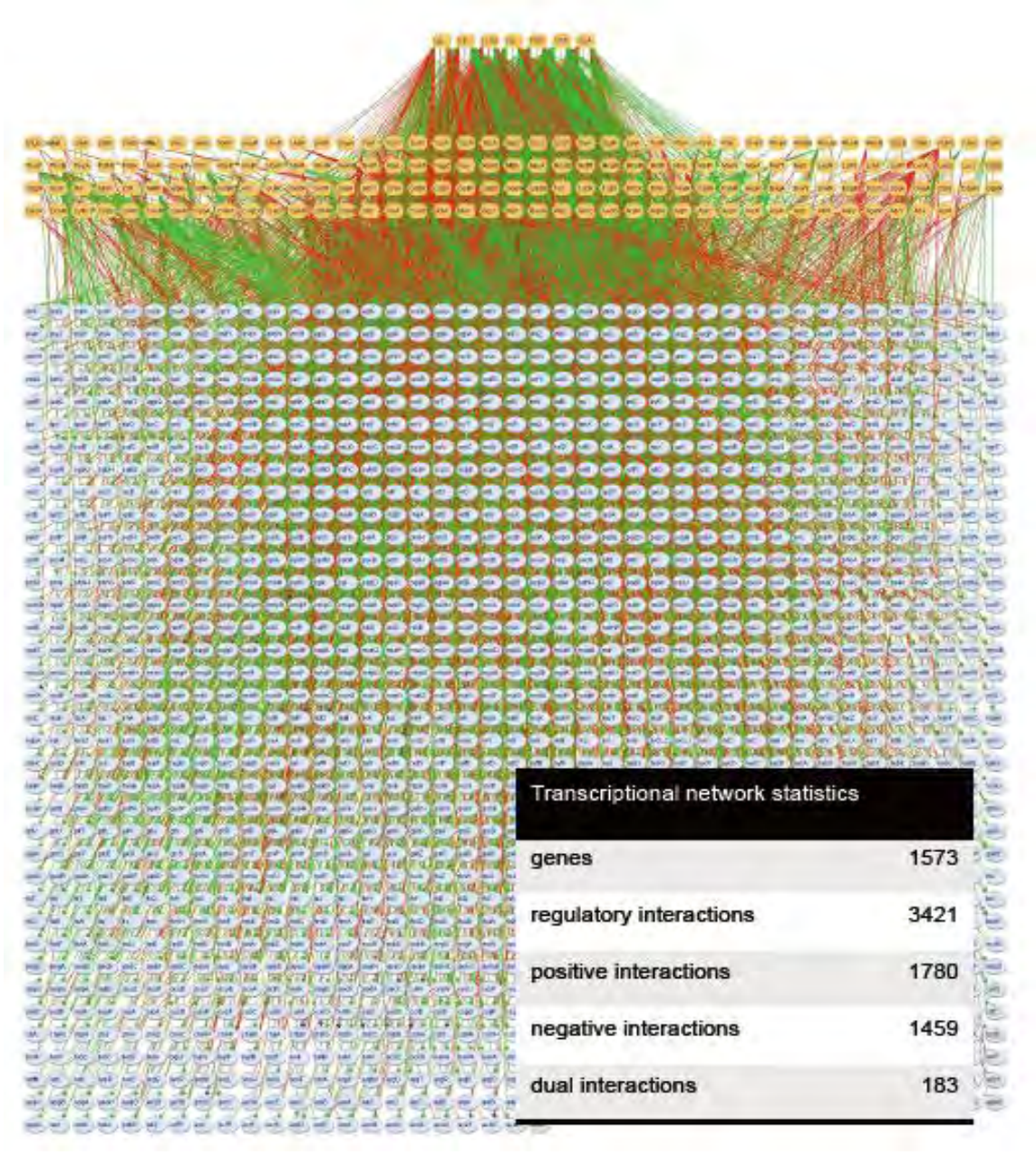

Fig. 1 *Escherichia coli* transcriptional regulatory network. This network corresponds to the regulonDB 6.7 version. Nodes represent genes and arcs the regulatory interactions among them respectively. This figure is as an updated version of Figure 1 in (Martinez-Antonio and Collado-Vides, 2003). The table inside this figure shows the statistics of this network. Green arcs represent activation; red, repression and blue, dual auto-regulation (updated versions of this figure is available on RegulonDB http://regulondb.ccg.unam.mx/images/RegulationNetworks/NetWorkTFGene.jpeg).

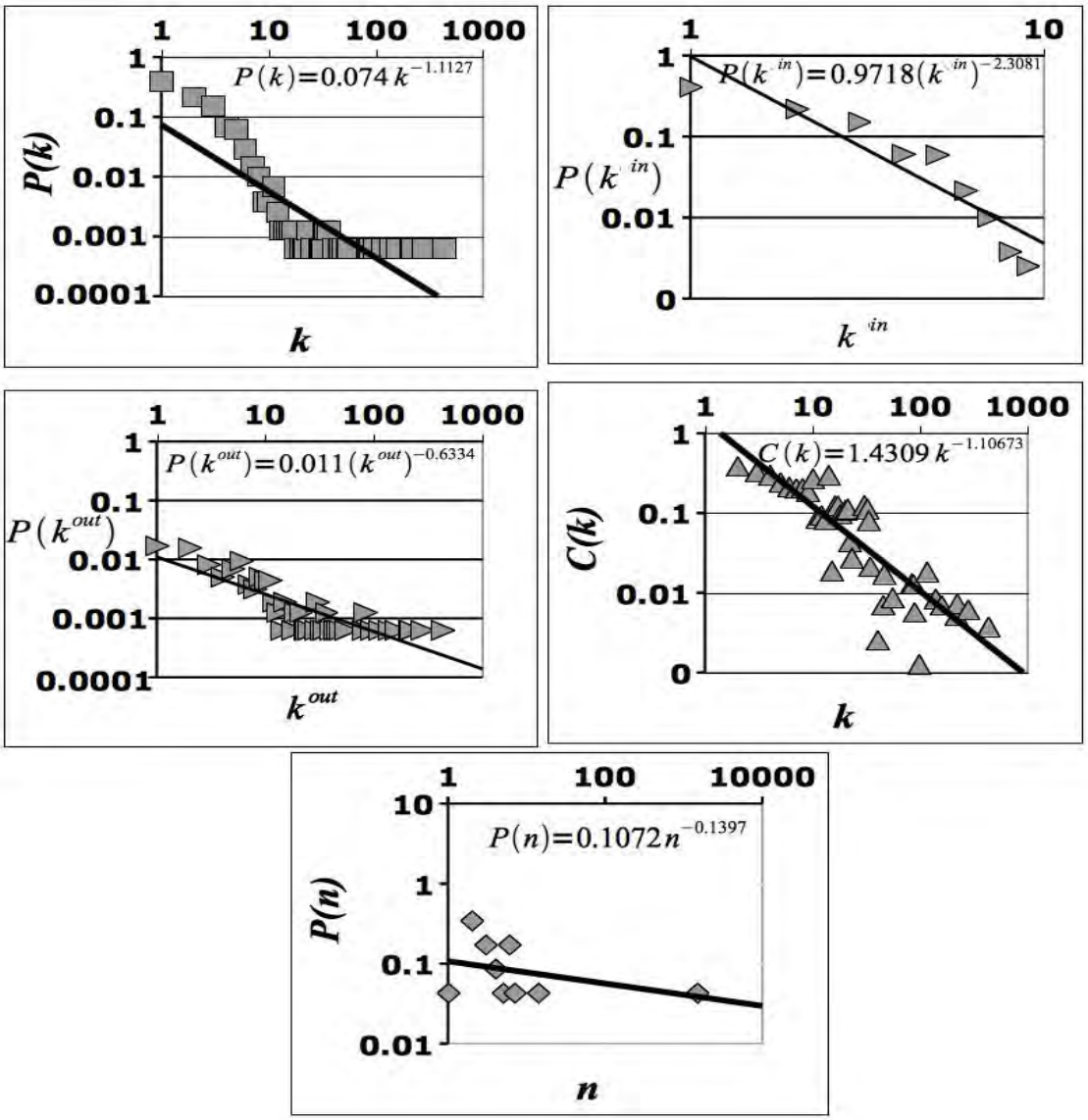

Fig. 2 Topological properties of the *E. coli* transcriptional regulatory network (as of figure 1). (a) Degree distribution *P(k)*. If we consider no directions in the interactions the network contains a high diversity in node degrees. However it better fit a power-law more than a normal distribution. (b) Input degree distribution $P(k^{in})$. Considering the incoming interactions in each node (analysis centered on target genes); their distribution fit nicely a scale-free system. It is because there are few nodes with 9 TFs regulating them while most of nodes have few TFs regulating them. (c) Output degree distribution $P(k^{out})$. Considering the outgoing interactions for each TF in the network the scale-free distribution is biased by few global TFs (having most of the interactions). Output interactions in the whole network dominate their entire distribution in the undirected network as this network distribution is similar to those of panel a. (d) Clustering coefficient *C(k)*. This analysis indicates that in nodes with few neighbors these are connected among them (as in a small well-known community) while in nodes connected to many other nodes these tend to be unconnected between them (as in a big city where is difficult that each person know each other). (e) The size distribution of connected components in the network *P(n)*. We can observe a giant connected component of 1476 nodes (on the right) and 23 isolated clusters with less of 15 nodes each.

Box 1 RegulonDB

RegulonDB (http://regulondb.ccg.unam.mx/) is currently the major electronically-encoded regulatory network of any free-living organism. This database is based on continuously updated curated knowledge of original scientific literature complemented with comprehensive computational predictions on transcriptional regulation of *E. coli* K-12 MG1655. RegulonDB have information of the complex regulation of transcription initiation, and also on the organization of the genes in transcription units, operons and simple and complex regulons. You can find graph and text integrated information by three major navigation streams: genes, operons and regulons. Some statistics of information you can find in RegulonDB is show below:

Bio-Objects (biobricks)	With experimental evidences	Total predicted
Genes		4605
Transcription units	2385	3386
Promoters	1798	4010
Terminators	228	
TF DNA-binding sites	1852	2123
TF regulatory interactions	2543	
sRNA interactions	81	
Shine-Dalgarno (RBS)	179	
Sigma Factors	7	7
Transcription Factors	175	300
Simple regulons	93	
Complex regulons	294	
Small molecule effectors	77	
Attenuators	246	
Riboswitches	20	

If you are interested to known additional information of *E. coli*, one good choice should be to visit the international *E. coli* Alliance database portal (http://www.uni-giessen.de/ecoli/IECA) where you can find links to around 40 additional databases with information on molecular components and their functional interactions. For those with interest in modeling can visit CyberCell or an additional resource on quantitative information of cellular component and some rates of reactions is Bionumbers (http://bionumbers.hms.harvard.edu/).

Table 1 Interesting statistics on operons organization and their regulation in *E. coli*.

Operons		
Biggest operon (with larger number of genes)	16 genes	*mra*Z-*rsm*H-*fts*LI-*mur*EF-*mra*Y-*mur*D-*fts*W-*mur*GC-*ddl*B-*fts*QAZ-*lpx*C
Operon with larger number of promoters (and therefore transcription units)	12 promoters	*mra*Z-*rsm*H-*fts*LI-*mur*EF-*mra*Y-*mur*D-*fts*W-*mur*GC-*ddl*B-*fts*QAZ-*lpx*C
Operon with larger number of terminators	5 terminators	*rho*L-*rho*
Operon regulated by the larger number of TFs	12 transcription factors	*gad*AXW
Operon with larger number of DNA-binding sites for TFs	21 DNA-binding sites	*glp*TQ
Operon with more promoters for different sigma factors	4 different sigma factors	*rpo*H and *clp*PX-*lon*
Transcription Factors		
Larger DNA-binding site for a TF	46 bp	ZraR
TF with the larger number of DNA-binding sites	231 DNA-binding sites	CRP

global regulators, which have the highest output degree that means they are regulating to many TG (Shen-Orr et al., 2002; Yu and Gerstein, 2006). Although the high number of regulated genes is a common criterion to consider a TF as global they should have additional properties to be considered as true global regulators (Martinez-Antonio and Collado-Vides, 2003): i) should regulate to a larger number of other TFs; ii) co-regulate with them; iii) their target genes should be transcribed by more than one kind of σ factor; iv) the products of their TG should fall in different functional classes; v) the TF should be active in different growth conditions, and; vi) these TFs commonly pertain to a protein-families with few paralogous.

Global TFs are commonly set at the highest positions in the regulatory network and they normally auto-regulates in a dual way (i.e. they auto-activate and auto-repress) which guarantee the protein level of these important regulators fluctuate between certain levels but never fall to zero (Savageau, 1977; Thomas, 1973). In agreement with this prediction, quantification of mRNA and protein levels for global regulators are found higher compared with the rest of TF in the network, which make biological sense since they should be functioning most of the time (see below). In *E. coli* these global regulators might be divided in those controlling the global metabolism and those know as nucleoid-associated proteins (NAPs). In the first group we found regulators for controlling carbon uptake (CRP), respiration mode (FNR and ArcA) and stringed response to the lack of important amino acids (Lrp), (first part of Table 2). In the second group we find to FIS, H-NS and IHF whose maximal production had been associated to a different points in a growing population curve (Ali Azam et al., 1999). Since NAPs have properties of DNA-bending and bridging, these kinds of proteins are considered as analogs regulators of gene expression (Marr et al., 2008), exerting their regulation by structuring the bacterial nucleoid in different forms (see below). Physically, it is difficult to test this hypothesis however as the nucleoid seems to be extremely dynamic, which make almost impossible to define the nucleoid structures, and even to identify the physical limits of the proposed chromosomal loops (Postow et al., 2004; Travers, 2006).

On the contrary, the most highly regulated TU (nodes with the highest input degree) are show in the second part of Table 2. As a general description we can see that most of these genes are encoding for products that define metabolic and adaptive capabilities of this bacterium: motility, response to stresses and nitrogen metabolism. It is interesting that two TUs encoding for TFs result among the most highly regulated ones. These genes are not the most highly conserved in bacteria but probably are more specific to *E. coli* genus and their tight regulation might be responsible to give the fitness required for the life-style distinctive of this bacterium.

4 Hierarchical Organization of the *E. coli* Transcriptional Network

Since a TF could regulate to genes that encode for TF (including itself) there is possible to extract from network in Fig. 1 the connected sub-network that represent one fraction of the regulatory hardware of *E. coli*, represented in Fig. 3. This network among 97 TFs was draw considering the operons where each TF-gene is encoded. From this graph, it is easy to distinguish that global regulators are distributed in the higher levels (left on the figure) of this hierarchical arrangement. This connected network (with 222 interactions and 69 TFs autoregulating) gives us a visual image from where a TF co-regulates with other TF to whom the first is regulating (as in the case of CRP). A close analysis of this network gives insights about the gene regulatory strategies in this bacterium and probably is representative of a more common strategy in bacteria: metabolic activities, whose clearest example is carbon utilization, is regulated by short transcriptional cascades (1 or 2 steps) where a global regulator -CRP- is normally co-regulating with one of the numerous –specific- catabolic TFs. In absence of glucose in the milieu, the coupled activity of CRP and one of the local regulators activates the machinery that direct the uptake and the first catabolic steps of one of the alternative carbon sources until

their derivative products fall over the general streams of glycolysis or pentose pathways. Other way, the longest transcriptional cascades are regulating physiological processes being considered as for development in bacteria, that is, those regulatory cascades that direct flagella and biofilm formation (Martinez-Antonio et al., 2008). If we consider the time the transcriptional program run along this network it is natural to expect that their execution in short cascades is quicker and therefore could result in a more uniform response of the bacterial population. On the contrary, the execution of regulatory programs along large transcriptional cascades should take more time the cell to accomplish it. As in addition each TF in the cascade autoregulates, we expect these longer transcriptional programs give place to a more noise-responding population. This might result in heterogeneity on bacterial phenotypes, as certainly have been seen in populations during flagella and biofilm formation (Gaynor et al., 2004; Sauer et al., 2002).

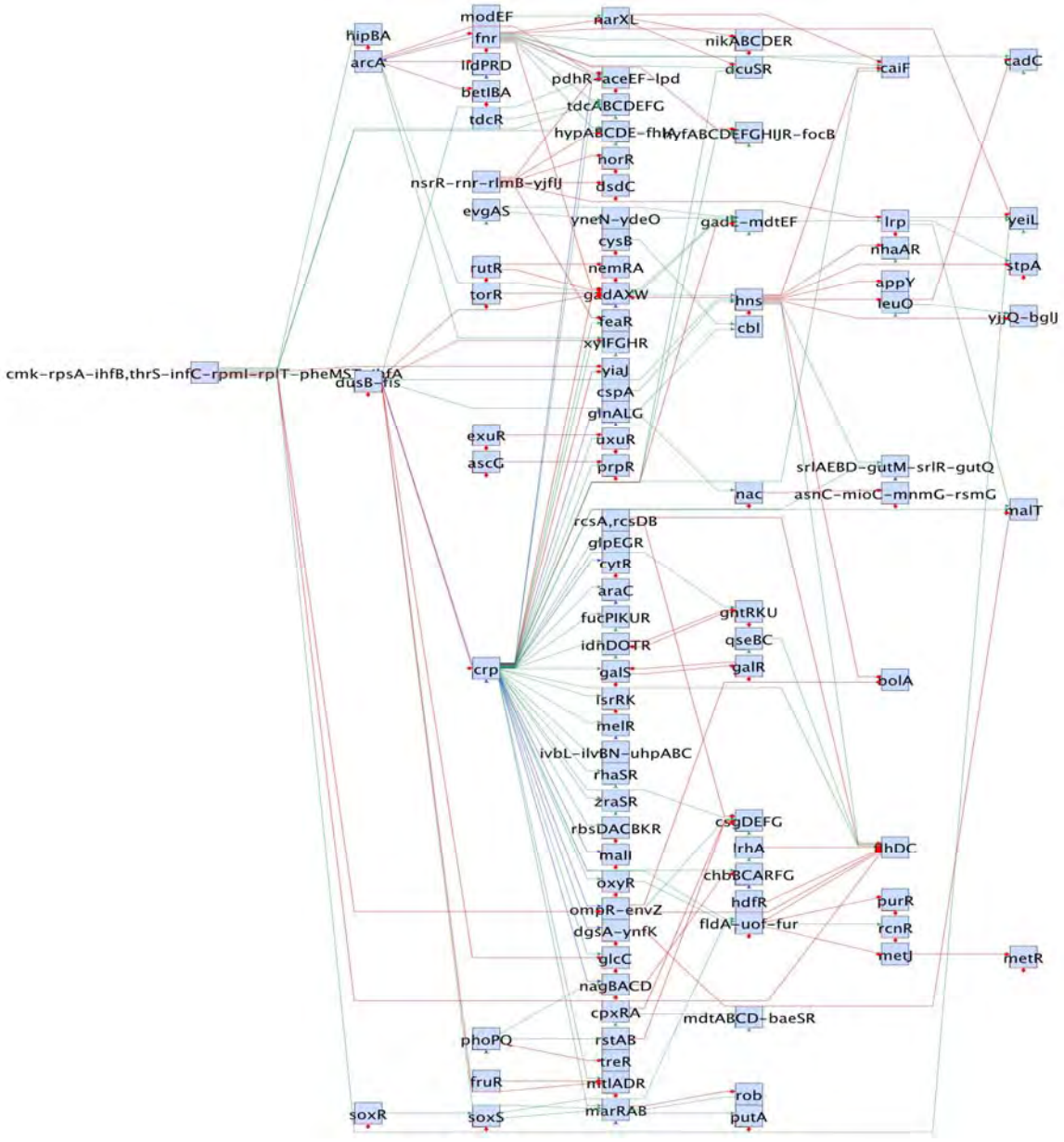

Fig 3 Hierarchical network organization of transcription factors in *E. coli*. The sub-network represents the connected regulatory hardware of *E. coli*. Nodes represent operons where each TF is encoded. Color of arcs is the same as in Figure 1. This figure was draw with the cytoscape software (Shannon et al., 2003).

Other interesting observation in the regulatory program of *E. coli* is that in spite that most to TF are negatively auto-regulated, the regulation of one TF to the following in the network is mostly of activation. Thus, it is possible to consider that the sequential order in which the regulatory software, encoded in the genome, should be executed influence the way the network is structured. That is to say, the regulatory network topology might be an evolutionary physical structure that facility the functional interactions of their molecular components following the scripts encoded in the genome. This structure of regulatory interactions is then the responsible of the transitions from genotype to phenotype in this bacterium (Martinez-Antonio et al., 2008). Similar topological regulatory networks could be working in other bacteria.

Table 2 Hubs and highly regulated target genes in the regulatory network of *E. coli*.

Global regulators (HUB TFs)	Name	Number of regulated genes
CRP	Cyclic AMP receptor protein, also known as catabolite activator protein (CAP)	440 genes in 128 regulons
H-NS	Histone-like nucleoid-associated protein	286 genes in 44 regulons
FNR	Fumarate and nitrate reductase regulatory protein	284 genes in 69 regulons
FIS	Factor for inversion stimulation	225 genes in 48 regulons
IHF	Integration host factor	223 genes in 60 regulons
ArcA	Aerobic respiration regulatory protein	160 genes in 48 regulons
Lrp	Leucine-responsive regulatory protein	97 genes in 26 regulons
Most highly regulated transcription units	Transcription factors regulating	Encoded gene(s)
*flh*DC	9: CRP, Fur, H-NS, HdfR, IHF, LrhA, OmpR, QseB, RcsAB	Heterodimer master regulator of flagella synthesis
*sod*A	8: ArcA, CRP, FNR, Fur, IHF, MarA, Rob, SoxS	Superoxide dismutase, Alleviate oxidative stress
*nir*BCD	8: CRP, FNR, Fis, FruR, H-NS, IHF, NarL, NarP	Large and small subunits of nitrite reductase and nitrite transporter
*mic*F	8: H-NS, HU, IHF, Lrp, MarA, OmpR, Rob, SoxS	Antisense negative regulator of OmpF abundance
*gad*AX	8: ArcA, CRP, FNR, GadE, GadW, GadX, H-NS, TorR	Regulators of glutamate decarboxylase synthesis
*omp*F	7: CRP, CpxR, EnvY, Fur, IHF, Lrp, OmpR	Outer membrane porin for secretion of toxic compounds
*nrf*ABCDEFG	7: FNR, Fis, FlhDC, IHF, NarL, NarP, NsrR	Nitrite reductase, formate-dependent, cytochrome c
*glt*BDF	7: ArgR, CRP, FNR, GadE, IHF, Lrp, Nac	Large and small subunits of glutamate synthase
*dcu*B-*fum*B	7: ArcA, CRP, DcuR, FNR, Fis, Fur, NarL	C4-dicarboxylate antiporter and fumarase B
*nap*FDAGHBC-*ccm*ABCDEFGH	6: FNR, FlhDC, IscR, ModE, NarL, NarP	Proteins with predicted roles in electron transfer to periplasmic nitrate reductase
*mar*RAB	6: CRP, Fis, MarA, MarR, Rob, SoxS	Regulators of weak acids and antibiotics resistance systems

5 Sensing and Condition Dependent Activity of the *E. coli* Transcriptional Network

As of the first studies with TFs, they were recognized as two-headed molecules given that normally they have two functional domains: one for DNA-binding and the other one for the sensing of small molecule effectors or for the interaction with additional proteins. The existence of this additional domain, to the DNA-binding, is very important in TFs as it gives the "switch" character to these protein regulators, (Jacob, 1970). About three-quarters of the *E. coli* transcription factors were identified as two-domain proteins (Madan Babu and Teichmann, 2003). The bindings of specific effectors, for each TF, induce conformational changes on them that switch their regulatory activity (from active to inactive state or vice verse), (Jacob and Monod, 1961; Wall et al., 2004). The signal effectors for TFs could be of diverse nature: osmotic pressure, light, temperature, organic compounds, waste products, metal ions, etc. These effectors could be sensed –by the TFs- inside or in the periphery of the cell (*i.e.* in the periplasm). In a gross description, small-molecule metabolites inside the cell are products of enzymatic reactions, respiration or waste products of the whole metabolism. On the other hand, signals sensed in the cell periphery are mainly of two types: organic molecules transported into the cell, which serve to supply energy, or as precursors to construct the cellular building blocks and; the other type correspond to physicochemical conditions in the milieu (heat, osmotic stress, etc.). As an overview, all the signal effectors produced inside the cell and transported from the milieu are sensed by the so-called "one-component" sensory systems (Ulrich et al., 2005). Other way, physicochemical conditions in the milieu are sensed by the "two-component" sensory systems, (Alex and Simon, 1994). TFs sensing effectors produced inside the cell might be considered as sensing "internal or endogenous" conditions while those sensing transported molecules and milieu conditions are considered as sensing "external o exogenous conditions" (Martinez-Antonio et al., 2006), some TFs however can sense signal effectors produced either inside or outside the bacteria, these are considered as "hybrid" sensory systems, these systems typically correspond to those TFs sensing amino acids, which may be synthesized in the cell or transported from the milieu.

Each these sensory systems are represented by TFs, the transcriptional network could be also considered as a network of transcriptional sensory systems. The logic behind the coordinated function of sensory systems seems to be simple to understand in some cases and complex in other ones. For example, it make biological sense that would be a co-regulatory activity among the TFs that control the transport of molecules from the milieu with those TFs that control the expression of enzymes that should degrade further this metabolites inside the cell, as it really happen. On the other hand, it is more complicated to understand the co-regulatory activity of two-component systems (for environment perception) with the activity of nucleoid-associated proteins as it also happen in *E. coli*; it means that environment conditions and nucleoid structure (i.e. their regulatory actors) are influencing together the expression of genes (Martinez-Antonio et al., 2006).

6 The Operation Mechanism of the Regulatory Program

The main actors in a transcriptional regulatory network are the nodes representing to TFs (and sometimes sigma factors) since they correspond to the regulatory machinery whose transcriptional switching outcome on the promoter region of TU, for activating or repressing genes. However each TF into the network repertoire has their own distinctiveness: i) a wide heterogeneity in the number of target genes, ii) the type of signals perceived (as from endogenous or exogenous origin) and, iii) their co-regulatory ability to work with different sigma and other TFs. Taking into account all these features in the catalog of TFs, the question then is: what is the operative logic which harmonize the running of the different cellular regulatory programs in agreement with the environment conditions and the cellular necessities? A careful analysis of data generated by experimentalists might offer us clues about this operative logic. First, it is empirically know that TFs have a wide rank of target genes, which permit classify them as from global to local regulators conform they regulates

as of hundreds to a less of a dozen of genes respectively. This decreasing globality of TFs correlates with their diminishing on cellular concentrations (Isalan et al., 2008) but with a rising specificity for their target DNA-binding sites (Lozada-Chavez et al., 2008). Second, even if there is observed a co-regulatory activity among global regulators, there is more common to observe co-regulation among global and local regulators, and hardly is it observed a local-local co-regulation. Third, local regulators tend to be encoded proximal to their target genes in the linear molecule of DNA. In fact, many times local TFs are encoded divergent or in the same operon with their regulated genes, which in addition to facilitate their regulation might put up functional the horizontal transfer of these genetic modules in bacteria. Fourth, there is known that NAPs proteins could wholly shape the bacterial nucleoid and that these proteins are preferentially expressed at different points in a growing population. NAPs are small proteins of less than 20 KDa although present in several dozen of thousands by cell. These small proteins bind cooperatively over the DNA bending or bridging it, reshaping their whole structure. The hypothesis states that these different conformations of the nucleoid influence the running of global transcriptional programs, as an analog regulation. This analog regulation is complemented with the more precise –digital- regulation exerted by local TFs whose switching activity are drive by specific signal effectors and thus fine-tuning the expression of genes in response to precise changing conditions (Marr et al., 2008). Functionally, this hypothesis has sustention as in most of cases co-regulation is observed between NAPs and some effector-using TF. Concerning the physicochemist of regulation at the transcription initiation it make sense that local regulators are produced very close respect their DNA-binding sites as they are produced in very low quantities (Kolesov et al., 2007), (one or two dozen of molecules per cell and many of them expressed less of one time per cell generation). Other way, global regulators are not impeded to diffuse and reach their target DNA-binding sites as they are normally at some few of thousands of molecules per cell. In this way we can get an idea about the complementing activity of NAPs, global and local regulators into the network hierarchy (Janga et al., 2009).

7 Evolutionary Dynamics of the *E. coli* Transcriptional Network

Bacterial genomes size from 120 to 10,000 ORFs and in this context *E. coli* has a medium-size genome. It is natural to assume that conform genomes increase not all the type of genes should increase in the same proportion. In larger genomes the genes that normally are enriched correspond to those encoding for regulatory and secondary metabolism functions (van Nimwegen, 2003). Respect the content of transcriptional machinery; larger genomes harbor more transcription factors per gene than smaller ones and this trend apply also for σ factors. Even more, not all the evolutionary families of TFs and σ increases uniformly as genomes sizes, but only in some families burst their members in larger genomes. In the case of TFs these families gift to bacteria abilities to contend with environment changes: cell-cell communication (LuxR), sensing of exogenous signals (OmpR), use of exogenous metabolites (GntR and AraC), and response to toxics and antibiotics (TetR). In the same logic, the family of σ that is more expanded correspond to those contributing with extracytoplasmic functions (ECF), which normally endow bacteria to compete with exogenous conditions also (Perez-Rueda et al., 2009). All this suggests that conform there are more variant environmental conditions, signal integration and regulation of gene expression require a more complex coordination for to enable rapid response and adaptation of bacteria (Cases et al., 2003; Molina and van Nimwegen, 2008). This implies that the repertoire of interactions by gene increases conform increases the number of genes in larger genomes. To understand the mechanisms how regulatory networks flexibly in each bacterium we should consider three main biological mechanisms driven it: (i) gene duplication, (ii) rewiring of edges by mutation/selection of TF/DNA interactions and (iii) horizontal gene transfer. Gene duplication is the driving force for creating new genes in genomes: it had been estimated that at least 50% of prokaryotic genes are result of this process. Many

transcription factors have identical domain architectures in each bacterium, and this implies that roughly three-quarters of the transcription factors might arisen as a consequence of gene duplication. In contrast, there is little evidence of duplication of regulatory regions together with regulated genes or of transcription factors together with regulated genes. It had been estimate that just one-third of known regulatory interactions were inherited from the ancestral transcription factor or target gene after duplication, and roughly one-half of the interactions were gained during divergence after duplication (Teichmann and Babu, 2004).

In studies in *E. coli, S. cerevisiae* and at least extent in *B. subtilis* it had been found that duplication of TF result in a conservation of auto-regulation in the network, and duplicates of TFs tend to populate the same layer with their ancestors, indicating conservation of hierarchy. In addition, layer populated by duplicated TFs tend to be lower (or in the periphery) of the network (Sellerio et al., 2009). On the contrary, other study indicates that even when most of transcription factors have paralogous, these usually arose by horizontal gene transfer rather than by duplication within the *E. coli* lineage. In general, horizontal gene transfer acquired most neighbor regulators -regulators that are adjacent to genes that they regulate-, whereas most global regulators evolved vertically within the γ-proteobacteria (Price et al., 2008). Consistent with these observations the transcriptional sensory machinery for exogenous signals in *E. coli* is less conserved respect those sensory systems for endogenous conditions (Salgado et al., 2007).

Take-home messages

- The present version of the regulatory network of *E. coli* includes 1531 genes and 3421 regulatory interactions (around one third of their total genes). This network shows a power-law distribution with a few global regulators and most of genes poorly connected.
- There is an operons sub-network of seven hierarchical layers formed by the connected activity of TFs encoded on them. Global regulators tend to be set in superior layers while local regulators are located in the lower ones.
- There is a small set of proteins (<8) know as NAPs, which are in a high cellular concentrations. These proteins reshape the nucleoid structure and influence the running of global transcriptional programs, this mode of regulation is proposed as an analog regulation.
- The binding of specific signal effectors assists the activity of most of TFs in *E. coli*. These effectors switch and tune the activity of TFs. To this type of regulation depending of environmental signals is named the digital –more precise- regulation.
- The integration of regulatory programs have place in the promoter region of TUs where it is common to observe co-regulation among global and local TFs as well as of TFs sensing exogenous and endogenous conditions.
- The mechanistic logic to understand the harmonious operation of regulatory programs in the network should consider the globality of TFs, their signal perceived, co-regulation, genome position, and cellular concentration.
- Duplicated TFs and their horizontal transfer influence the evolvability of the network. The most duplicated and transferred TFs are located in the network periphery.

Acknowledgements

The author thank to John Délano and Gabriela Olmedo for the critical reading of the manuscript. I also thank to Beatriz C. Luna for their help with the topological analyzes in Fig. 2, and to Heladia Salgado for providing some data from RegulonDB and with Figures 1 and 3. This work was partially supported by CONACYT Young-researcher grant 102854.

References

Alex LA, Simon MI. 1994. Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. Trends in Genetics, 10: 133-138

Ali Azam T, Iwata A, Nishimura A, et al. 1999. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. The Journal of Bacteriology, 181: 6361-6370

Anderson JC, Dueber JE, Leguia M, et al. 2010. BglBricks: A flexible standard for biological part assembly. Journal of Biological Engineering, 4(1): 1-12

Arkin AP. 2001. Synthetic cell biology. Current Opinion in Biotechnology, 12: 638-644

Benner SA. 2003. Synthetic biology: Act natural. Nature, 421: 118

Blattner FR, Plunkett G, Bloch CA, et al. 1997. The complete genome sequence of Escherichia coli K-12. Science, 277: 1453-1474

Browning DF, Busby SJ. 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol, 2: 57-65

Cases I, de Lorenzo V, Ouzounis CA. 2003. Transcription regulation and environmental adaptation in bacteria. Trends in Microbiology, 11: 248-253

Gama-Castro S, Jimenez-Jacinto V, Peralta-Gil M, et al. 2008. RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. Nucleic Acids Research, 36: D120-124

Gaynor EC, Cawthraw S, Manning G, et al. 2004. The genome-sequenced variant of Campylobacter jejuni NCTC 11168 and the original clonal clinical isolate differ markedly in colonization, gene expression, and virulence-associated phenotypes.The Journal of Bacteriology, 186: 503-517

Isalan M, Lemerle C, Michalodimitrakis K, et al. 2008. Evolvability and hierarchy in rewired bacterial gene networks. Nature, 452: 840-845

Jacob F. 1970. La Logique du Vivant, Une Histoire de L'Hérédité. Gallimard, Paris, France

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol, 3: 318-356

Janga SC, Salgado H, Martinez-Antonio A. 2009. Transcriptional regulation shapes the organization of genes on bacterial chromosomes. Nucleic Acids Research, 37: 3680-3688

Kolesov G, Wunderlich Z, Laikova ON, et al. 2007. How gene order is influenced by the biophysics of transcription regulation. Proceedings of the National Academy of Sciences of the USA, 104: 13948-13953

Lederberg J, Tatum EL. 1953. Sex in bacteria; genetic studies, 1945-1952. Science, 118: 169-175

Lozada-Chavez I, Angarica VE, Collado-Vides J, et al. 2008. The role of DNA-binding specificity in the evolution of bacterial regulatory networks. Journal of Molecular Biology, 379: 627-643

Madan Babu M, Teichmann SA 2003. Evolution of transcription factors and the gene regulatory network in Escherichia coli. Nucleic Acids Research, 31: 1234-1244

Marr C, Geertz M, Hutt MT, et al. 2008. Dissecting the logical types of network control in gene expression profiles. BMC Systems Biology, 2: 18

Martinez-Antonio A, Collado-Vides J. 2003. Identifying global regulators in transcriptional regulatory networks in bacteria. Current Opinion in Microbiology, 6: 482-489

Martinez-Antonio A, Janga SC, Salgado H, et al. 2006. Internal-sensing machinery directs the activity of the regulatory network in Escherichia coli. Trends in Microbiology, 14: 22-27

Martinez-Antonio A, Janga SC, Thieffry D. 2008. Functional organisation of Escherichia coli transcriptional regulatory network. Journal of Molecular Biology, 381: 238-247

Miller ES, Kutter E, Mosig G, et al. 2003. Bacteriophage T4 genome. Microbiology and Molecular Biology Reviews, 67: 86-156

Molina N, van Nimwegen E. 2008. Universal patterns of purifying selection at noncoding positions in bacteria. Genome Research, 18: 148-160

Ohnishi M, Tanaka C, Kuhara S, et al. 1999. Chromosome of the enterohemorrhagic Escherichia coli O157:H7; comparative analysis with K-12 MG1655 revealed the acquisition of a large amount of foreign DNAs. DNA Research, 6: 361-368

Perez-Rueda E, Collado-Vides J. 2000. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. Nucleic Acids Research, 28: 1838-1847

Perez-Rueda E, Janga SC, Martinez-Antonio A. 2009. Scaling relationship in the gene content of transcriptional machinery in bacteria. Molecular Biosystems, 5: 1494-1501

Postow L, Hardy CD, Arsuaga J, et al. 2004. Topological domain structure of the Escherichia coli chromosome. Genes and Development, 18: 1766-1779

Price MN, Dehal PS, Arkin AP. 2008. Horizontal gene transfer and the evolution of transcriptional regulation in Escherichia coli. Genome Biology, 9: R4

Rasko DA, Rosovitz MJ, Myers GS, et al. 2008. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. The Journal of Bacteriology, 190: 6881-6893

Salgado H, Martinez-Antonio A, Janga SC. 2007. Conservation of transcriptional sensing systems in prokaryotes: A perspective from Escherichia coli. FEBS Letters, 581: 3499-3506

Sauer K, Camper AK, Ehrlich GD, et al. 2002. Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm. The Journal of Bacteriology, 184: 1140-1154

Savageau MA. 1977. Design of molecular control mechanisms and the demand for gene expression. Proceedings of the National Academy of Sciences of the USA, 74: 5647-5651

Sellerio AL, Bassetti B, Isambert H, et al. 2009. A comparative evolutionary study of transcription networks. The global role of feedback and hierachical structures. Molecular Biosystems, 5: 170-179

Shannon P, Markiel A, Ozier O, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13: 2498-2504

Shen-Orr SS, Milo R, Mangan S, et al. 2002. Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics, 31: 64-68

Steed H, Macfarlane GT, Macfarlane S. 2008. Prebiotics, synbiotics and inflammatory bowel disease. Molecular Nutrtion and Food Research, 52: 898-905

Teichmann SA, Babu MM. 2004. Gene regulatory network growth by duplication. Nat Genet, 36: 492-496

Thomas R. 1973. Boolean formalization of genetic control circuits. Journal of Theoretical Biology, 42: 563-585

Touchon M, Hoede C, Tenaillon O, et al. 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genetics, 5: e1000344

Travers A. 2006. DNA topology: dynamic DNA looping. Current Biology, 16: R838-840

Ulrich LE, Koonin EV, Zhulin IB. 2005. One-component systems dominate signal transduction in prokaryotes. Trends in Microbiology, 13: 52-56

van Nimwegen E. 2003. Scaling laws in the functional content of genomes. Trends in Genetics, 19: 479-484

Wall ME, Hlavacek WS, Savageau MA. 2004. Design of gene circuits: lessons from bacteria. Nature Reviews Genetics, 5: 34-42

Wirth T, Falush D, Lan R, et al. 2006. Genomic analysis of the hierarchical structure of regulatory networks. Proceedings of the National Academy of Sciences of the USA, 103: 14724-14731