

Functional organisation of *Escherichia coli* transcriptional regulatory network

Agustino Martínez-Antonio,^{1,□} Sarath Chandra Janga,² and Denis Thieffry^{3,4}

¹Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados, Instituto Politécnico Nacional, Campus Guanajuato, Irapuato 36500, México

²MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

³TAGC U928, INSERM and Université de La Méditerranée, Campus Scientifique de Luminy, Case 928, 13288 Marseille, France

⁴CONTRAIINTES Project, INRIA Paris-Rocquencourt Center, domaine de Voluceau, BP 105, 78153 Le Chesnay, France

Agustino Martínez-Antonio: amartinez@ira.cinvestav.mx

□ Corresponding author. Email: amartinez@ira.cinvestav.mx

Received 2008 Mar 7; Revised 2008 May 21; Accepted 2008 May 22.

Copyright © 2008 Elsevier Ltd.

This document may be redistributed and reused, subject to [certain conditions](#).

This document was posted here by permission of the publisher. At the time of the deposit, it included all changes made during peer review, copy editing, and publishing. The U. S. National Library of Medicine is responsible for all links within the document and for incorporating any publisher-supplied amendments or retractions issued subsequently. The published journal article, [guaranteed](#) to be such by Elsevier, is available for free, on ScienceDirect, at: <http://dx.doi.org/10.1016/j.jmb.2008.05.054>

Abstract

Taking advantage of available functional data associated with 115 transcription and 7 sigma factors, we have performed a structural analysis of the regulatory network of *Escherichia coli*. While the mode of regulatory interaction between transcription factors (TFs) is predominantly positive, TFs are frequently negatively autoregulated. Furthermore, feedback loops, regulatory motifs and regulatory pathways are unevenly distributed in this network. Short pathways, multiple feed-forward loops and negative autoregulatory interactions are particularly predominant in the subnetwork controlling metabolic functions such as the use of alternative carbon sources. In contrast, long hierarchical cascades and positive autoregulatory loops are overrepresented in the subnetworks controlling developmental processes for biofilm and chemotaxis. We propose that these long transcriptional cascades coupled with regulatory switches (positive loops) for external sensing enable the coexistence of multiple bacterial phenotypes. In contrast, short regulatory pathways and

variations. TFs at the core of the network coordinate the most basic endogenous processes by passing information onto multi-element circuits. Transcriptional expression data support broader and higher transcription of global TFs compared to specific ones. Global regulators are also more broadly conserved than specific regulators in bacteria, pointing to varying functional constraints.

Abbreviations: TF, transcription factor; CRP, cAMP receptor protein; cAMP, cyclic adenosine monophosphate; FFL, feed-forward loop; HNS, histone-like protein; FNR, fumarate and nitrate regulatory protein; IHF, integration host factor

Keywords: transcriptional regulation, network organisation, multiple phenotypes, homeostasis

Introduction

In bacteria, coupling of gene expression with external conditions is achieved through two molecular

functions: (i) binding of transcription factors (TFs) at specific sites in the genome and (ii) recognition of a relevant effector signal or metabolite.^{1,2} Typically, these functions are performed by different domains of a single polypeptide, but there are also cases where two interacting proteins are responsible for these functions, as in two-component systems.³

At the phenotypic level, there are evidences for the coexistence of multiple phenotypes in bacterial cultures, e.g., of cells with different morphological and physiological abilities such as motility, biofilm formation, drug resistance, etc.^{4,5} In particular, biofilm formation and chemotaxis are considered as multistage developmental processes, and in mature biofilms, a mixture of bacterial population from different developmental stages were found to coexist.^{6,7}

Here we take advantage of the wealth of experimental data on transcriptional regulation for the best-characterized bacterium, *Escherichia coli*, to analyse the structure of the transcriptional network in light of different functional constraints.

Results and Discussion

Topology of *E. coli* cross-regulatory transcriptional network

Available experimental data point to more than 3000 regulatory interactions between TFs and their regulated genes in *E. coli*. This information is integrated and documented in a specialised database called RegulonDB.⁸ Global analyses of this huge network have already been published, emphasising a hierarchical organisation and statistically overrepresented regulatory motifs.^{9–11} Here, our aim is to analyse the flow of regulatory information within the network of transcriptional interactions among TFs and sigmas (*E. coli* transcriptional cross-regulatory network). This network encompasses 115 TFs and 7 sigma factors, i.e., around one-third of the total predicted TF proteins in this bacterium (Fig. 1).^{12,13} On average, every TF is connected to two other TFs (i.e., more technically, the mean degree of the regulatory graph is 2.74). However, the connectivity distribution of TFs is not uniform, with a small fraction of global TFs with high out-degrees dominating the network.¹⁵ Seven global regulators were defined previously based on a collection of criteria:¹⁵ (i) number of regulated genes; (ii) number of regulated genes encoding for TFs; (iii) propensity of cooperative regulation of targets with the aid of other TFs; (iv) ability to directly affect the expression of a variety of promoters that use different sigma factors; (v) belonging to evolutionary families with few paralogs; and (vi) heterogeneity of the functional classes of the regulated genes.

In order to better visualise the informational flow through the network, the following graphical conventions have been used in Fig. 1 (see also legend): (i) the size of the nodes representing TFs is proportional to the number of genes they regulate [e.g., cAMP receptor protein (CRP) regulates 413 genes and is represented by the second biggest node, after the housekeeping sigma factor rpoD]; (ii) arrows and colours refer to the direction and sign of the regulatory interaction; (iii) arrow thickness is proportional to the impact of the interaction, computed as the number of genes thereby (in)directly regulated.

The majority of the TFs in this network are autoregulated (~ 70%), of which about two-thirds account for negative loops (see Table 1). This finding is consistent with the results of an analysis performed with a much smaller number of TFs about 10 years ago.¹⁷ This predominance of negative autoregulatory loops contrasts with the predominance of positive arcs between different TFs (about 54%, see Table 1). The dominance of positive regulatory interactions in the regulatory network of *E. coli* is not limited to those among TFs, as when we compute the regulation of all the target genes (3017 arcs) we found that about 54% (1630) are positively regulated, 40% (1206) are repressed, while about 6% (171) are dual regulated. This is especially interesting because a majority of the TFs in bacteria have been reported to act as repressors.^{12,18,19} The conventions used in Fig. 1 clearly display the hierarchical organisation of the network, with master regulators

such as CRP, fumarate and nitrate regulatory protein (FNR) or integration host factor (IHF) each (in)directly regulating a large number of other TFs. Furthermore, the layout emphasises important variations regarding the length of the transcriptional cascades.

Although functional annotations on TFs are still limited, it is possible to classify the cross-regulating TFs into broad categories according to the physiological functions of the target structural genes: carbohydrate initial catabolism, respiration, biofilm formation and chemotaxis, etc. As shown in [Fig. 1](#), these broad classes correspond to different local network topologies. Due to their contrasting topologies, in what follows, we will focus our discussion on short regulatory cascades observed in the case of carbohydrate catabolism as opposed to long regulatory cascades seen in the case of biofilm and chemotaxis pathways ([Fig. 2a](#)). CRP resides at the top of both subnetworks. CRP is the only global TF acting hierarchically over local TFs for the usage of carbohydrates, whereas CRP's activity is comparable to the activity of other global regulators in the rest of the network. Note that the concentration of its effector metabolite, cyclic adenosine monophosphate (cAMP), is at par with that of adenosine triphosphate (ATP), which acts as the energetic currency of the cell.²⁰ This suggests that CRP not only regulates the use of these substrates for producing ATP, but also senses the energetic status of the cell to decide the execution of other cellular programs.

This study aims at understanding the network structure in relation to physiological roles played by the different modules, focusing on differences in the topologies of the subnetworks controlling metabolism *versus* motility and chemotaxis (cf. the following sections).

However, other subnetworks are also worth mentioning. In particular, all nine TFs controlling the expression of genes for amino acid biosynthesis seem to be expressed constitutively by sigma 70. Each TF regulates the transcription of the required genes for producing different amino acids. At high concentrations of the amino acids, allosteric modifications of TFs follow binding to their respective amino acids, resulting in TF autorepression as well as to the repression of the corresponding biosynthetic genes. Interestingly, the logic behind negative autoregulation in this case is different from that of the catabolism of carbohydrates. While in the latter case TFs are autorepressed until the substrate is available, in the case of amino acids, TFs are autorepressed only in the presence of an excess of the synthesized final product. Another interesting subnetwork is that for alleviating the stresses by drugs, solvents and weak organic acids. The regulatory logic in this complex subnetwork is peculiar, as their components form multi-element circuits (see [Fig. 3](#)) and their inputs are directed by Rob and SoxR, two small proteins constitutively expressed but with very short half-lives (1–2 min). Their stability/degradation depends on the presence/absence of their effector signals.^{21,22,23}

Multiple parallel feed-forward loops regulate the use of different carbon sources

Cellular feeding, which includes the uptake of carbon and energy sources and their metabolism, can be considered as one of the main physiological processes in bacterial systems. The regulation of these processes directly affects cellular fitness. The selection of carbon sources is regulated by CRP and about 20 more specific TFs ([Fig. 2a](#)). The hierarchical organisation of the corresponding subnetwork is characterized by a short average path length (cf. [Table 1](#)). Regulatory interactions between CRP and the specific TFs result in the occurrence of multiple feed-forward loops (FFLs) for the use of alternative sugar sources. FFL is a network motif recurrently found in transcriptional networks and is defined as a three-gene pattern composed of two input TFs, one of which regulates the other, both jointly regulating a target gene.^{9,24} Based on the mode of regulation of each TF, this motif is subdivided into eight different subtypes.²⁴ Coherent FFL type 1 corresponds to all the regulatory interactions in the motif being positive; in incoherent type 1 FFL, the first TF regulates positively both the targets, although the second TF represses the expression of the target gene thereby reversing the final effect. The majority of the FFLs present in the subnetwork for carbon catabolism belong to coherent and incoherent type 1 groups,²⁴ with both TFs working together, as a result of a persistent signal affecting the global TF (in this case, cAMP) and the presence of a signal affecting a TF

corresponding to a sugar alternative to glucose.^{24–26} This motif structure enables the filtering of short pulses of the signal affecting the global TF (cAMP) in case of transient glucose deprivation. Consequently, the target structural genes are activated only in the persistent absence of glucose and in the presence of an alternative carbon source.

The phosphotransferase system typically transports and phosphorylates certain sugars, including glucose, a preferred carbon source for *E. coli*, and this condition ultimately results in low levels of cAMP. Consequently, CRP does not activate the transcription of the genes responsible for the degradation of alternative sugars. Note that most structural genes involved in the transport and initial catabolism of alternative carbon sources are encoded in operons, each specifically repressed in the absence of the inducing sugar.

When glucose is lacking, cAMP level increases and CRP can activate the transcription of genes responsible for degrading alternative carbon sources.²⁷ Simultaneously, sugars (or a processed variant thereof) present in the cell bind their specific TF; allosteric interactions then result in TF unbinding from DNA, alleviating the repression and permitting the transcription of the corresponding target genes. This organisation involving multiple parallel FFLs coupled to phosphotransferase activity appears optimal for enabling rapid transcriptional responses to sudden lack of glucose in the presence of alternative carbon sources in the milieu.^{28,29}

Long hierarchical cascades regulate developmental processes

Biofilm formation and bacterial mobility can be seen as the outcome of specialised cell differentiation pathways. Biofilm formation involves subsequent cellular changes at the morphological and physiological levels resulting in bacterial populations with multiple phenotypes.^{4,30,31} Furthermore, bacteria living in biofilm communities are present in different developmental stages (at least four defined stages) as has been observed in *Cryptococcus*, *Pseudomonas*, *Staphylococcus*, *Xanthomonas*, etc.^{32–37}

The part of *E. coli* transcriptional cross-regulatory network involved in the control of biofilm formation and motility exhibits a relatively complex topology with several long cascades from CRP, IHF and FNR to downstream specialised TFs (see [Table 1](#) and [Fig. 2a](#)). Several of these cascades converge on the master regulators for motility and biofilm formation (FlhCD and CsgD, respectively).

Furthermore, a relatively high proportion of the downstream specialised TFs autoactivate themselves (see below), a feature that is rare at the level of the whole transcriptional network. In addition, we could identify 105 multi-element circuits in the transcriptional regulatory network involving 2 to 14 different elements (including transcription and sigma factors, see [Fig. 3](#) and supplementary material for the complete list of circuits). Interestingly, the information fluxes inside these circuits follow a frequent route in the network with the order CsgD > Cspa > HNS > GadX > RpoS > IHF. From IHF the regulatory flow diverge in two main directions: IHF > RpoH > RpoD and IHF > FIS > CRP. In line with this observation, CRP and RpoD are the major distributors, whereas CsgD and GadX are the main collectors of information. These multi-element circuits inside the cross-regulatory network are novel observations. The functional relevance of these regulatory structures remains to be assessed experimentally. Tentatively, these circuits may implement a feedback between the presence of different stresses and the basic machinery for replication and growth.

CsgD, the master regulator for biofilm formation, is directly involved in 28 of these long circuits, suggesting a particular tight coupling of CsgD activity with the intracellular status. In contrast, FlhCD, the master compound regulator for motility and chemotaxis, is known to be regulated by nine other TFs but has not yet been reported to regulate any other TF.

Note that the motility module has its own sigma factor, FliA, regulated by FlhCD. FliA is required for the

transcription of the genes required for the last part of flagella development and for chemotaxis machinery.^{38,39} In contrast, the genes for biofilm development are transcribed by the housekeeping sigma 70 and RpoS, the sigma factor expressed in response to general stress.⁴⁰

The execution of such long regulatory cascades requires time. Indeed, complete flagella assembly may take a generation time or longer.^{41–43} The occurrence of positively autoregulated TFs at several intermediate steps enables informed decisions about the cellular/environmental condition. In some conditions, cellular duplication might be faster than the conclusion of a long regulatory cascade. This implies that bacterial populations likely consist of mixtures of bacteria with transcriptional programs at different levels in long regulatory cascades.

Multiple conditions regulate the first steps of developmental processes in *E. coli*

Both biofilm and motility are regulated by multiple, long regulatory cascades (Fig. 2a). This organisation is congruent with the observation that these developmental processes are regulated by multiple environmental conditions, ranging from nutritional deprivation to environmental stresses.^{6,7,43} In *E. coli*, several global TFs sit at the top in these regulatory cascades, namely, CRP, IHF and FNR, all sensing endogenous signals.² Additional regulatory inputs are provided by more specialised TFs such as ArcA (aerobic respiration regulatory protein), EvgA (environmentally responsive activator) and TorR (trimethylamine N-oxide reductase regulator) involved in exogenous sensing. Remarkably, downstream players are enriched in two-component system partners: RcsAB (capsule biosynthesis regulator), OmpR (outer membrane protein regulator), CpxR (regulator of cell envelope proteins folding and degradation) and QseB (quorum sensing regulator) (see the next section). Consequently, the first developmental steps likely occur in most of the cellular population in response to nutritional (CRP) or respiration (FNR-ArcA) stresses, yet depending on nucleoid structure status (sensed through IHF), while the final steps critically depend on exogenous conditions, as observed in other bacteria.^{7,44} For instance, in *Pseudomonas aeruginosa*, the initial steps for biofilm development (motility and reversible attachment to solid surfaces) are independent of quorum sensing, whereas the final steps (irreversible attachment and biofilm maturation) are strongly dependent on quorum-sensing regulation, an exogenous condition.⁴⁵ Both processes (motility and biofilm) must be tightly interconnected, since motility is important for the concerted movement of groups of bacteria over solid surfaces, followed by cell membrane modification, excretion of polysaccharides, causing them to be less motile and smaller in mature biofilms.^{46–48} Interestingly, in *E. coli*, the activation of nucleoid-associated proteins such as histone-like protein (HNS) occurs relatively early along both biofilm and chemotaxis developmental pathways. The expression of these nucleoid-associated proteins is known to be growth-phase dependent (HNS is mostly expressed in the exponential phase, whereas IHF and factor for inversion stimulation (FIS) are expressed in arrested growing cells and during the early exponential phase, respectively), pointing to an influence of nucleoid structure and specific bacterial growth phases on the first steps of these developmental processes.^{42,49,50}

External input information defines the last steps of the developmental processes

Downstream of biofilm and chemotaxis developmental cascades, transcriptional autoactivations and cross-inhibitions become more frequent (Figs. 1 and 2a). This part of the network comprises several two-component systems, enabling the sensing of external information. In particular, OmpR is responsible for switching the phenotype between biofilm and motility (activating the first and repressing the latter) depending on external stresses. During stressing conditions as well as in mature biofilms, cells tend to be smaller compared to rapidly growing cells. This phenotypic transition depends on the expression of BofA (cell morphogenetic regulator), which is transcribed by RpoS and repressed by OmpR.⁵¹

The activation of the master gene for motility, FlhCD, is further controlled by the quorum-sensing detector

QseB, while the activity of the master regulator for biofilm formation, CsgD, is repressed in the presence of extracellular stressing conditions, detected by CpxR. In addition, the BaeSR system controls the expression of export complexes, conferring multidrug resistance phenotypes,⁵² while RcsB mediates the glutamate-dependent acid resistance.^{53,54} QseB, CpxR, BaeR, OmpR and RcsAB are all members of two-component systems that are modified by protein–protein phosphorylation, enabling quick response in comparison to *de novo* protein production along transcriptional cascades. Strikingly, almost all of these two-component TFs (except OmpR) autoactivate their expression, an observation fitting the contention that positive feedback circuits are necessarily found at the core of all differentiation switches. Indeed, such TF autoactivations likely enable persistent TF expression in two stable states: “on” or “off”. Thus, they constitute a potentially robust machinery for all-or-nothing output response depending on transient exogenous signals.^{55–62}

Expression of master regulators

In the preceding sections, we have emphasised the hierarchical organisation as well as important local structural variations in *E. coli* transcriptional network. In the subsequent section, we consider the expression of a TF as a function of its position in the regulatory hierarchy. Although it is difficult to address this at the protein level because of the current relative scarcity of proteomic data, this question can be readily addressed at the level of transcription. Therefore, we have analysed the levels of mRNA reported by two independent experiments on *E. coli* cells grown on minimal medium + glucose.^{63–65} In summary, the mean mRNA levels of global TFs such as CRP, FNR, ArcA and IHF are significantly higher than those of more specialised regulators (Fig. 2b). This result supports the contention that the master regulators are continuously required at high levels to bind to numerous DNA sites across the genome, whereas downstream, specialised TFs can be expressed or not depending on the environmental conditions. Indeed, in conditions favourable to planktonic growth (as in the condition for mRNA quantification mentioned above), TFs controlling the final stages of biofilm/motility development are lowly expressed. Similarly, the TFs controlling specific sugar catabolism regulons show much lower mean concentration compared to CRP, suggesting that these TFs are sporadically required, depending on the metabolic state of the cell. A similar tendency has been recently observed by measuring the fluorescence of 15 different TF–green fluorescent protein gene fusions.⁶⁶

Conservation of the transcriptional cross-regulatory network in bacteria

To what extent can our observations on *E. coli* transcriptional network be generalised to other bacteria? To address this question, we have compiled orthologs of *E. coli* TFs across 216 non-redundant bacterial genomes (using the approach described in Ref. ⁶⁷) and analysed their conservation. As a general observation, global regulators (and negatively autoregulated TFs) of *E. coli* tend to be more conserved than more specialised regulators (see Fig. 4). In spite of this observation, previous studies indicated that even the TFs found at the top of regulatory hierarchies vary across bacterial phyla.^{67–69} Intriguingly, orthologs of CsgD and FlhCD are found in less than 25 genomes, whereas orthologs for OmpR and FliA are found in more than 75 genomes. More precisely, FliA appears to be conserved in many organisms lacking FlhCD orthologs, although FlhCD is involved in the transcriptional regulation of FliA in *E. coli*. Tentatively, in these organisms, a different TF might regulate FliA. Alternatively, FliA alone might control the transcription of genes required in the final stages of motility and chemotaxis, or FliA might be directing the transcription of genes for unrelated functions in these bacteria. This set of regulatory components and the corresponding network structure are likely to be conserved in a relatively small group of bacteria (less than 25 bacteria closely related to *E. coli*). However, the general network organisation (long regulatory cascades *versus* multiple FFLs, and endogenous *versus* exogenous sensing) apparently constitute a common theme for the regulation of physiological and developmental processes in bacteria.

Conclusions

Our structural analysis of the transcriptional cross-regulatory network in *E. coli* suggests that regulatory interactions between TFs are predominantly positive, while autoregulatory interactions are mostly negative. However, this general trend appears to be reversed in the case of most downstream TFs involved in the regulation of biofilm/chemotaxis modules.

We also note that there are striking topological differences between the subnetwork controlling metabolic activities, such as carbon metabolism, and that controlling developmental processes; the former encompasses many parallel short transcriptional cascades and multiple FFLs, each enabling the use of one alternative carbon source, while the latter involves long and intertwined regulatory cascades. These long transcriptional cascades typically include multiple autoactivated intermediate TFs, as well as regulatory circuits between TFs and sigma factors in the case of biofilm formation.

We further observe that TFs acting at the end of these regulatory cascades often belong to two-component systems. This topology suggests that cell homeostasis is maintained through multiple regulatory cascades with commonly autorepressed TFs, while the regulatory memory within the network is preserved by the sequential activation of TFs and by multi-element circuits at the core of the network. Downstream of the hierarchical network, two-component systems can memorise transient external signals through autoactivation loops, thus acting as molecular switches enabling the coexistence of alternative phenotypes.

As shown in a recent study, the *E. coli* cross-regulatory network appears to be robust enough to tolerate the rewiring between members high and low in the network hierarchy.⁶⁶ This study also indicates that the allosteric signals are the mandatory input elements for network function. Thus, TFs present in a condition different from the natural one(s) would have limited activity due to the absence of their effector signals.

In this respect, a proper global understanding of the organisation of the *E. coli* transcriptional network (combining sigma and TFs) could contribute to the interpretation of network-rewiring experiments as well as foster more efficient design of synthetic regulatory circuits.

The general significance of the observed organisation of the *E. coli* transcriptional cross-regulatory network remains to be assessed. A more comprehensive picture of the network organisation in bacteria will progressively be drawn as additional regulatory elements such as small RNAs, anti-sigma factors and riboswitches are integrated.⁷⁰ In addition, the combination of transcriptional and metabolic networks should provide important insights by linking effector metabolites and regulatory elements. Clearly, variations in regulatory network topology might be expected in the case of bacteria with asymmetric cell division (mostly α -proteobacteria), where the offspring asymmetric cells cause a transient genetic asymmetry that triggers different developmental processes, such as the formation of stalked and swarmer cells in *Caulobacter* or vegetative and spore-forming cells in *Bacillus*.^{71–75} Future comparisons between network topologies for different model systems should further enhance our understanding of regulatory network organisation and its conservation or variations among different bacterial phyla.

Acknowledgements

We thank Heladia Salgado and the RegulonDB team for compiling and providing *E. coli* TF network data. Part of this work is a result of discussions among the three authors at the workshop “Dynamics of Regulatory Networks II” held in 2007 at the Centro Internacional de Ciencias A.C, Cuernavaca, Mexico.

Appendix A. Supplementary data

Supplementary material:

Transcriptional cross-regulatory network analysed in this study, FFL motifs identified in the network, mRNA abundances and conservation of transcription factors are available in a supplementary excel file.

References

1. Jacob F. Une Histoire de L'Hérédité; Gallimard, Paris: 1970. La Logique du Vivant.
2. Martinez-Antonio A., Janga S.C., Salgado H., Collado-Vides J. Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. Trends Microbiol. 2006;14:22–27. [PubMed: 16311037]
3. Ulrich L.E., Koonin E.V., Zhulin I.B. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. 2005;13:52–56. [PMCID: PMC2756188] [PubMed: 15680762]
4. Ehrlich G.D., Hu F.Z., Shen K., Stoodley P., Post J.C. Bacterial plurality as a general mechanism driving persistence in chronic infections. Clin. Orthop. Relat. 2005;20–24. [PMCID: PMC1351326]
5. Balaban N.Q., Merrin J., Chait R., Kowalik L., Leibler S. Bacterial persistence as a phenotypic switch. Science. 2004;305:1622–1625. [PubMed: 15308767]
6. O'Toole G., Kaplan H.B., Kolter R. Biofilm formation as microbial development. Annu. Rev. Microbiol. 2000;54:49–79. [PubMed: 11018124]
7. Stoodley P., Sauer K., Davies D.G., Costerton J.W. Biofilms as complex differentiated communities. Annu. Rev. Microbiol. 2002;56:187–209. [PubMed: 12142477]
8. Salgado H., Gama-Castro S., Peralta-Gil M., Diaz-Peredo E., Sanchez-Solano F., Santos-Zavaleta A. RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. Nucleic Acids Res. 2006;34:D394–D397. [PMCID: PMC1347518] [PubMed: 16381895]
9. Shen-Orr S.S., Milo R., Mangan S., Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat. Genet. 2002;31:64–68. [PubMed: 11967538]
10. Dobrin R., Beg Q.K., Barabasi A.L., Oltvai Z.N. Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network. BMC Bioinf. 2004;5:10. [PMCID: PMC357809]
11. Ma H.W., Kumar B., Ditges U., Gunzer F., Buer J., Zeng A.P. An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. Nucleic Acids Res. 2004;32:6643–6649. [PMCID: PMC545451] [PubMed: 15604458]
12. Perez-Rueda E., Collado-Vides J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res. 2000;28:1838–1847. [PMCID: PMC102813] [PubMed: 10734204]
13. Madan Babu M., Teichmann S.A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. Nucleic Acids Res. 2003;31:1234–1244. [PMCID: PMC150228] [PubMed: 12582243]
14. Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–2504. [PMCID: PMC403769] [PubMed: 14597658]
15. Martinez-Antonio A., Collado-Vides J. Identifying global regulators in transcriptional regulatory networks in bacteria. Curr. Opin. Microbiol. 2003;6:482–489. [PubMed: 14572541]
16. Hu Z., Mellor J., Wu J., DeLisi C. VisANT: an online visualization and analysis tool for biological interaction data. BMC Bioinf. 2004;5:17. [PMCID: PMC368431]

17. Thieffry D., Huerta A.M., Perez-Rueda E., Collado-Vides J. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays*. 1998;20:433–440. [PubMed: 9670816]
18. Moreno-Campuzano S., Janga S.C., Perez-Rueda E. Identification and analysis of DNA-binding transcription factors in *Bacillus subtilis* and other Firmicutes—a genomic approach. *BMC Genomics*. 2006;7:147. [PMCID: PMC1524751] [PubMed: 16772031]
19. Struhl K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell*. 1999;98:1–4. [PubMed: 10412974]
20. Bettenbrock K., Sauter T., Jahreis K., Kremling A., Lengeler J.W., Gilles E.D. Correlation between growth rates, EIICrr phosphorylation, and intracellular cyclic AMP levels in *Escherichia coli* K-12. *J. Bacteriol.* 2007;189:6891–6900. [PMCID: PMC2045212] [PubMed: 17675376]
21. Shah I.M., Wolf R.E., Jr Novel protein–protein interaction between *Escherichia coli* SoxS and the DNA binding determinant of the RNA polymerase alpha subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. *J. Mol. Biol.* 2004;343:513–532. [PubMed: 15465042]
22. Martin R.G., Gillette W.K., Rosner J.L. Promoter discrimination by the related transcriptional activators MarA and SoxS: differential regulation by differential binding. *Mol. Microbiol.* 2000;35:623–634. [PubMed: 10672184]
23. Griffith K.L., Wolf R.E., Jr Genetic evidence for pre-recruitment as the mechanism of transcription activation by SoxS of *Escherichia coli*: the dominance of DNA binding mutations of SoxS. *J. Mol. Biol.* 2004;344:1–10. [PubMed: 15504398]
24. Mangan S., Alon U. Structure and function of the feed-forward loop network motif. *Proc. Natl Acad. Sci. USA.* 2003;100:11980–11985. [PMCID: PMC218699] [PubMed: 14530388]
25. Janga S.C., Salgado H., Martinez-Antonio A., Collado-Vides J. Coordination logic of the sensing machinery in the transcriptional regulatory network of *Escherichia coli*. *Nucleic Acids Res.* 2007;35:6963–6972. [PMCID: PMC2175315] [PubMed: 17933780]
26. Alon U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 2007;8:450–461. [PubMed: 17510665]
27. Deutscher J., Francke C., Postma P.W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* 2006;70:939–1031. [PMCID: PMC1698508] [PubMed: 17158705]
28. Mangan S., Itzkovitz S., Zaslaver A., Alon U. The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *J. Mol. Biol.* 2006;356:1073–1081. [PubMed: 16406067]
29. Dekel E., Mangan S., Alon U. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys. Biol.* 2005;2:81–88. [PubMed: 16204860]
30. Shapiro J.A. Thinking about bacterial populations as multicellular organisms. *Annu. Rev. Microbiol.* 1998;52:81–104. [PubMed: 9891794]
31. Hooshangi S., Thiberge S., Weiss R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA.* 2005;102:3581–3586. [PMCID: PMC552778] [PubMed: 15738412]

32. Aldridge P., Hughes K.T. Regulation of flagellar assembly. *Curr. Opin. Microbiol.* 2002;5:160–165. [PubMed: 11934612]
33. Guerrero A., Jain N., Goldman D.L., Fries B.C. Phenotypic switching in *Cryptococcus neoformans*. *Microbiology.* 2006;152:3–9. [PMCID: PMC2721797] [PubMed: 16385110]
34. Chantratita N., Wuthiekanun V., Boonbumrung K., Tiyawisutrisri R., Vesaratchavest M., Limmathurotsakul D. Biological relevance of colony morphology and phenotypic switching by *Burkholderia pseudomallei*. *J. Bacteriol.* 2007;189:807–817. [PMCID: PMC1797308] [PubMed: 17114252]
35. Massey R.C., Buckling A., Peacock S.J. Phenotypic switching of antibiotic resistance circumvents permanent costs in *Staphylococcus aureus*. *Curr. Biol.* 2001;11:1810–1814. [PubMed: 11719226]
36. Kamoun S., Kado C.I. Phenotypic switching affecting chemotaxis, xanthan production, and virulence in *Xanthomonas campestris*. *Appl. Environ. Microbiol.* 1990;56:3855–3860. [PMCID: PMC185079] [PubMed: 16348384]
37. Handke L.D., Conlon K.M., Slater S.R., Elbaruni S., Fitzpatrick F., Humphreys H. Genetic and phenotypic analysis of biofilm phenotypic variation in multiple *Staphylococcus epidermidis* isolates. *J. Med. Microbiol.* 2004;53:367–374. [PubMed: 15096544]
38. Kalir S., McClure J., Pabbaraju K., Southward C., Ronen M., Leibler S. Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science.* 2001;292:2080–2083. [PubMed: 11408658]
39. Kalir S., Alon U. Using a quantitative blueprint to reprogram the dynamics of the flagella gene network. *Cell.* 2004;117:713–720. [PubMed: 15186773]
40. Hengge-Aronis R. Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* 2002;66:373–395. table of contents. [PMCID: PMC120795] [PubMed: 12208995]
41. Aizawa S.I., Kubori T. Bacterial flagellation and cell division. *Genes Cells.* 1998;3:625–634. [PubMed: 9893020]
42. Pruss B.M., Matsumura P. Cell cycle regulation of flagellar genes. *J. Bacteriol.* 1997;179:5602–5604. [PMCID: PMC179437] [PubMed: 9287021]
43. Macnab R.M. How bacteria assemble flagella. *Annu. Rev. Microbiol.* 2003;57:77–100. [PubMed: 12730325]
44. Crawford E.W., Jr, Shimkets L.J. The stringent response in *Myxococcus xanthus* is regulated by SocE and the CsgA C-signaling protein. *Genes Dev.* 2000;14:483–492. [PMCID: PMC316387] [PubMed: 10691740]
45. Sauer K., Camper A.K., Ehrlich G.D., Costerton J.W., Davies D.G. *Pseudomonas aeruginosa* displays multiple phenotypes during development as a biofilm. *J. Bacteriol.* 2002;184:1140–1154. [PMCID: PMC134825] [PubMed: 11807075]
46. Beloin C., Valle J., Latour-Lambert P., Faure P., Kzreminski M., Balestrino D. Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. *Mol. Microbiol.* 2004;51:659–674. [PubMed: 14731270]
47. Danese P.N., Pratt L.A., Kolter R. Biofilm formation as a developmental process. *Methods Enzymol.* 2001;336:19–26. [PubMed: 11398398]

48. Prigent-Combaret C., Brombacher E., Vidal O., Ambert A., Lejeune P., Landini P., Dorel C. Complex regulatory network controls initial adhesion and biofilm formation in *Escherichia coli* via regulation of the *csgD* gene. *J. Bacteriol.* 2001;183:7213–7223. [PMCID: PMC95571] [PubMed: 11717281]
49. Luijsterburg M.S., Noom M.C., Wuite G.J., Dame R.T. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.* 2006;156:262–272. [PubMed: 16879983]
50. Grainger D.C., Hurd D., Goldberg M.D., Busby S.J. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.* 2006;34:4642–4652. [PMCID: PMC1636352] [PubMed: 16963779]
51. Vieira H.L., Freire P., Arraiano C.M. Effect of *Escherichia coli* morphogene *bolA* on biofilms. *Appl. Environ. Microbiol.* 2004;70:5682–5684. [PMCID: PMC520865] [PubMed: 15345459]
52. Nishino K., Honda T., Yamaguchi A. Genome-wide analyses of *Escherichia coli* gene expression responsive to the BaeSR two-component regulatory system. *J. Bacteriol.* 2005;187:1763–1772. [PMCID: PMC1063996] [PubMed: 15716448]
53. Fredericks C.E., Shibata S., Aizawa S., Reimann S.A., Wolfe A.J. Acetyl phosphate-sensitive regulation of flagellar biogenesis and capsular biosynthesis depends on the Rcs phosphorelay. *Mol. Microbiol.* 2006;61:734–747. [PubMed: 16776655]
54. Castanie-Cornet M.P., Treffandier H., Francez-Charlot A., Gutierrez C., Cam K. The glutamate-dependent acid resistance system in *Escherichia coli*: essential and dual role of the His–Asp phosphorelay RcsCDB/AF. *Microbiology.* 2007;153:238–246. [PubMed: 17185552]
55. Brandman O., Ferrell J.E., Jr, Li R., Meyer T. Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science.* 2005;310:496–498. [PMCID: PMC3175767] [PubMed: 16239477]
56. Suel G.M., Garcia-Ojalvo J., Liberman L.M., Elowitz M.B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature.* 2006;440:545–550. [PubMed: 16554821]
57. Acar M., Becskei A., van Oudenaarden A. Enhancement of cellular memory by reducing stochastic transitions. *Nature.* 2005;435:228–232. [PubMed: 15889097]
58. Ferrell J.E., Jr Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell. Biol.* 2002;14:140–148. [PubMed: 11891111]
59. Goulian M. Robust control in bacterial regulatory circuits. *Curr. Opin. Microbiol.* 2004;7:198–202. [PubMed: 15063859]
60. Demongeot J., Kaufman M., Thomas R. Positive feedback circuits and memory. *C. R. Acad. Sci. III.* 2000;323:69–79. [PubMed: 10742912]
61. Thomas R. Laws for the dynamics of regulatory networks. *Int. J. Dev. Biol.* 1998;42:479–485. [PubMed: 9654035]
62. Wall M.E., Hlavacek W.S., Savageau M.A. Design of gene circuits: lessons from bacteria. *Nat. Rev. Genet.* 2004;5:34–42. [PubMed: 14708014]
63. Covert M.W., Knight E.M., Reed J.L., Herrgard M.J., Palsson B.O. Integrating high-throughput and computational data elucidates bacterial networks. *Nature.* 2004;429:92–96. [PubMed: 15129285]
64. Lu P., Vogel C., Wang R., Yao X., Marcotte E.M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* 2007;25:117–124.

[PubMed: 17187058]

65. Liu M., Durfee T., Cabrera J.E., Zhao K., Jin D.J., Blattner F.R. Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. *J. Biol. Chem.* 2005;280:15921–15927.

[PubMed: 15705577]

66. Isalan M., Lemerle C., Michalodimitrakis K., Horn C., Beltrao P., Raineri E. Evolvability and hierarchy in rewired bacterial gene networks. *Nature.* 2008;452:840–845. [PMCID: PMC2666274]

[PubMed: 18421347]

67. Lozada-Chavez I., Janga S.C., Collado-Vides J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res.* 2006;34:3434–3445. [PMCID: PMC1524901] [PubMed: 16840530]

68. Madan Babu M., Teichmann S.A., Aravind L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J. Mol. Biol.* 2006;358:614–633. [PubMed: 16530225]

69. Salgado H., Martinez-Antonio A., Janga S.C. Conservation of transcriptional sensing systems in prokaryotes: a perspective from *Escherichia coli*. *FEBS Lett.* 2007;581:3499–3506.

[PMCID: PMC2238691] [PubMed: 17617412]

70. Gama-Castro S., Jimenez-Jacinto V., Peralta-Gil M., Santos-Zavaleta A., Penaloza-Spinola M.I., Contreras-Moreira B. RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.* 2008;36:D120–D124. [PMCID: PMC2238961] [PubMed: 18158297]

71. Dworkin J. Transient genetic asymmetry and cell fate in a bacterium. *Trends Genet.* 2003;19:107–112.

[PubMed: 12547520]

72. Dworkin J., Losick R. Differential gene expression governed by chromosomal spatial asymmetry. *Cell.* 2001;107:339–346. [PubMed: 11701124]

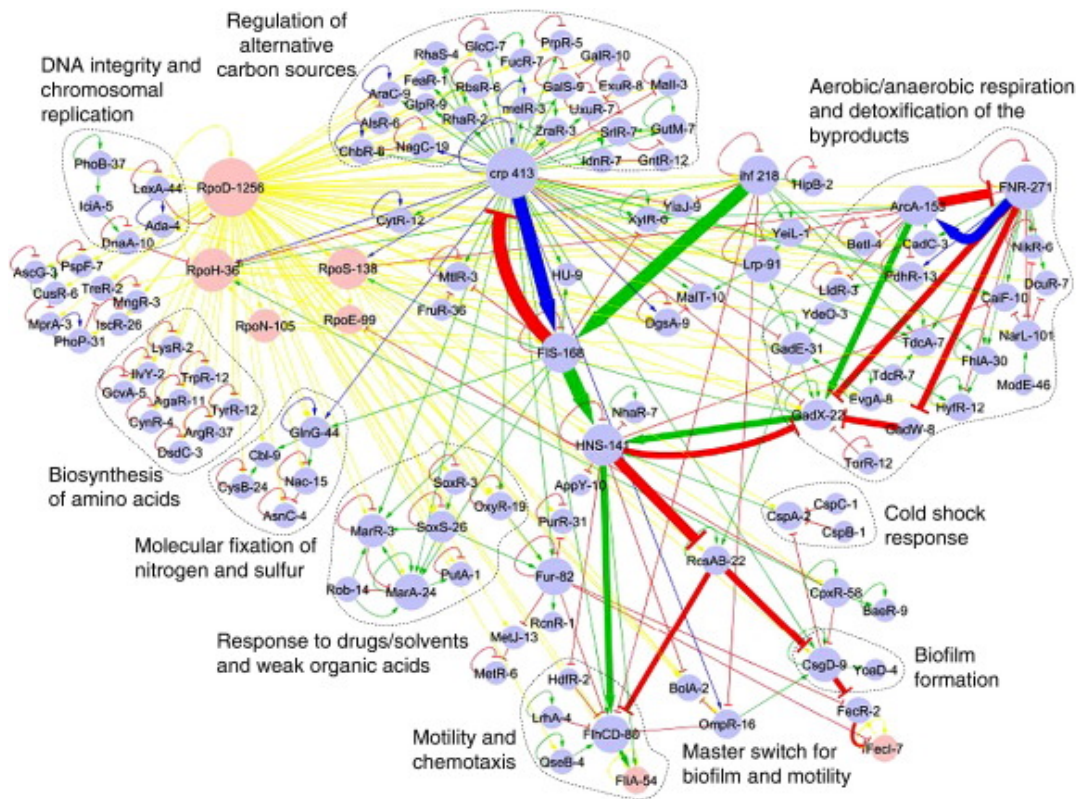
73. Yudkin M.D., Clarkson J. Differential gene expression in genetically identical sister cells: the initiation of sporulation in *Bacillus subtilis*. *Mol. Microbiol.* 2005;56:578–589. [PubMed: 15819616]

74. Ausmees N., Jacobs-Wagner C. Spatial and temporal control of differentiation and cell cycle progression in *Caulobacter crescentus*. *Annu. Rev. Microbiol.* 2003;57:225–247. [PubMed: 14527278]

75. Hilbert D.W., Piggot P.J. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol. Mol. Biol. Rev.* 2004;68:234–262. [PMCID: PMC419919] [PubMed: 15187183]

Figures and Tables

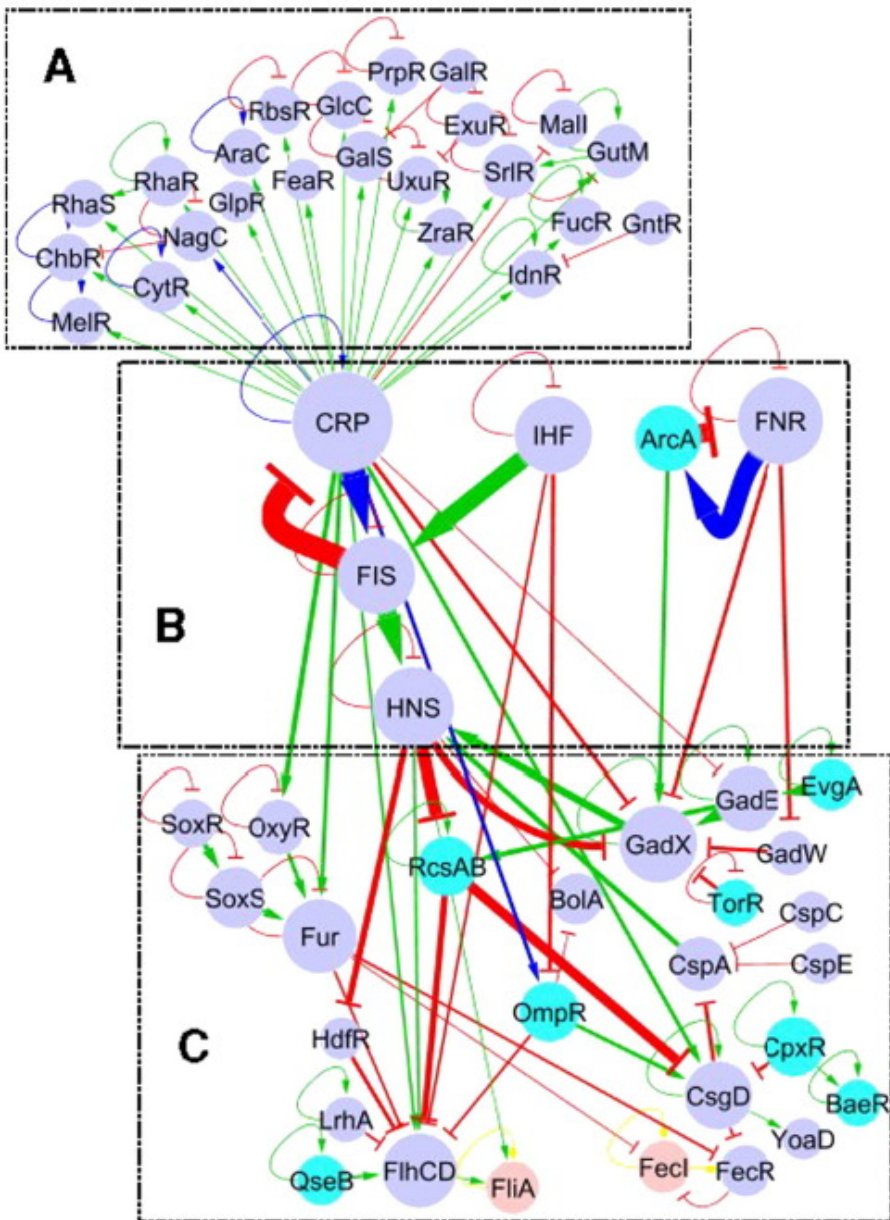
Fig. 1



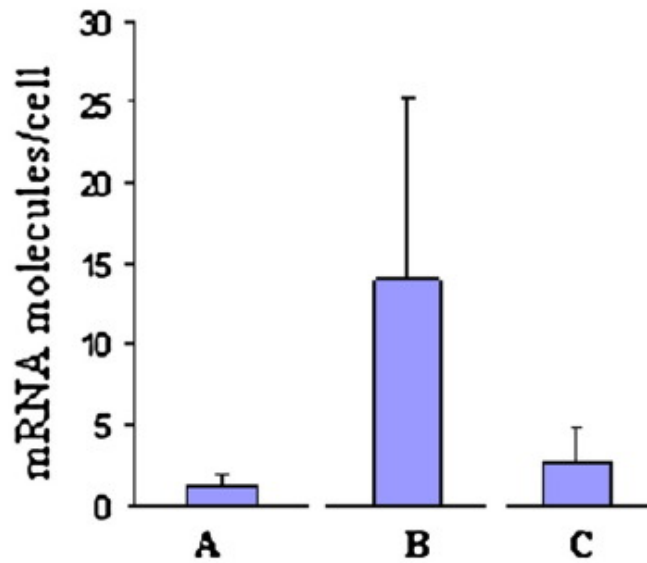
Core transcriptional regulatory network of *E. coli*. Blue and pink nodes represent genes encoding for TFs and sigma factors, respectively; each node label is accompanied with its connectivity showing the number of regulatory targets. Edges represent cross-regulatory interactions (green for activation, red for repression, blue for dual interactions and yellow for sigma transcription), whereas loops represent transcriptional autoregulations. Specific subnetworks, such as the one associated with the regulation of carbon sources, are delineated with dashed lines to distinguish different regulatory modules. This figure was generated using Cytoscape.¹⁴

Fig. 2

(a)

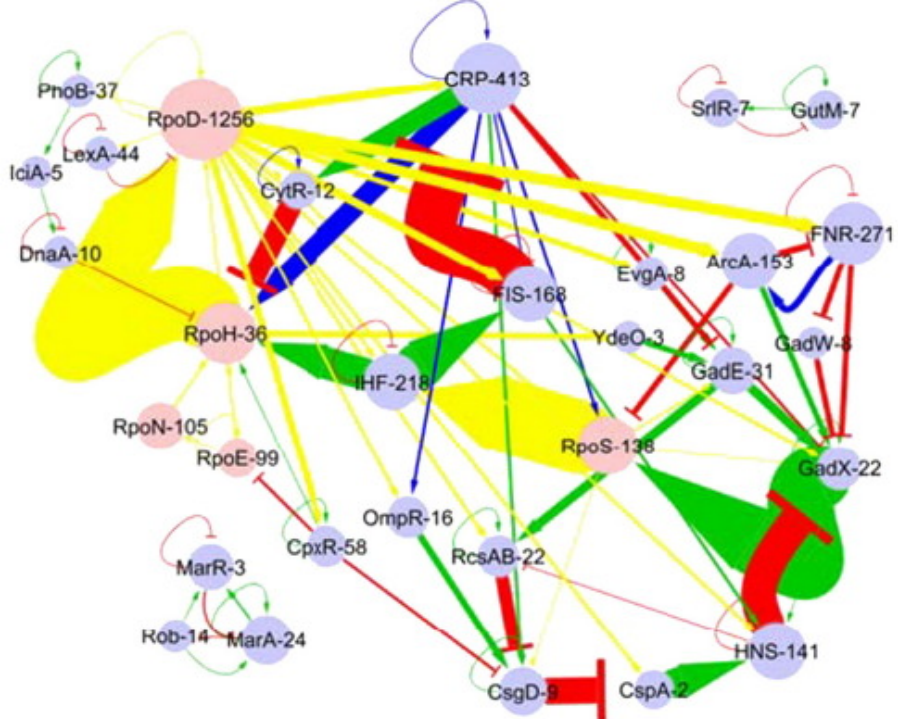


(b)



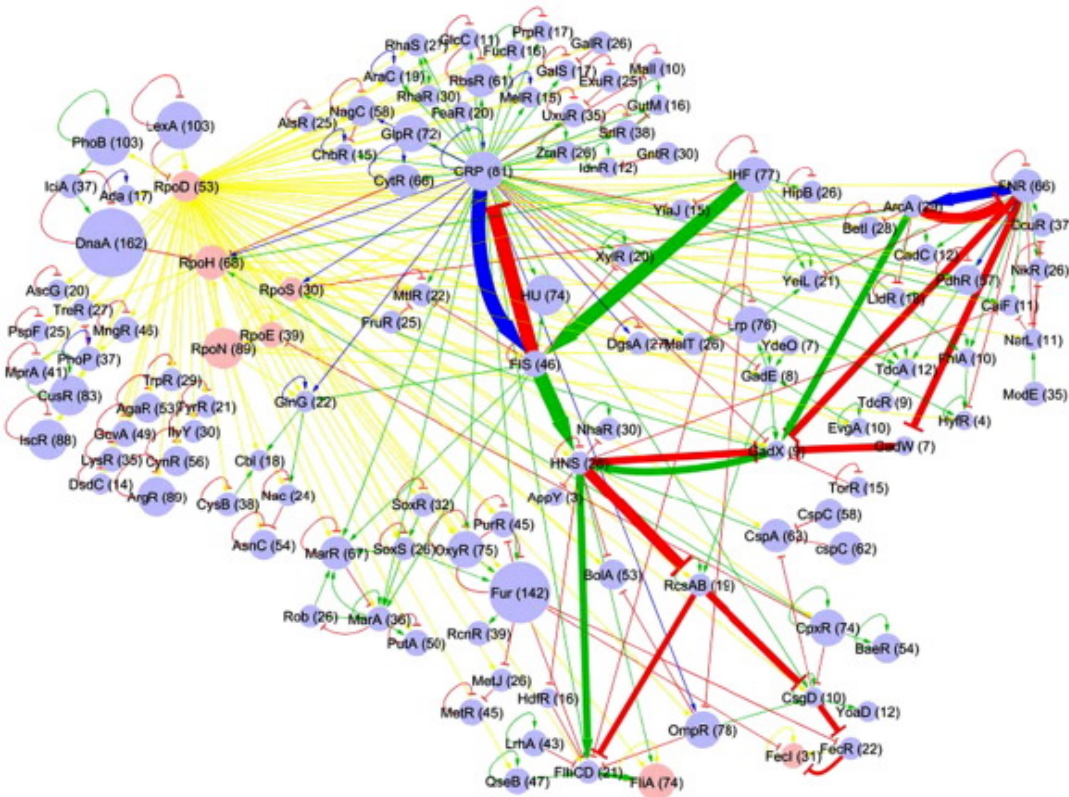
Functional organisation of *E. coli* core transcriptional network. (a) selection of carbon source (group A), global regulation (group B), and regulation of developmental processes (biofilm and chemotaxis, group C). (b) Average mRNA levels per cell for each TF group defined in (a), together with standard deviations. Levels of mRNA were recovered for 13 TFs (56%) of members of group A, all TFs of group B, and 13 TFs (54%) of group C. Nodes in light blue represent members of two-component systems.

Fig. 3



Multi-element regulatory circuits found in the cross-regulatory network of *E. coli*. Extent of informational flux between two nodes is denoted by the thickness of the edges.

Fig. 4



Conservation of *E. coli* sigma and transcription factors across 216 non-redundant bacterial genomes. Node sizes are proportional to the corresponding conservation interval and to the number of genomes in which an ortholog was found (shown in parentheses).

Table 1

Distributions of positive, negative and dual (auto)regulatory interactions, mean path length¹⁶ and observed maximum out- and in-degrees for different sections of the *E. coli* transcriptional cross-regulatory network shown in [Fig. 1](#)

| Network section | No. of TFs | No. of Autoregulations | Positive autoregulations | Negative autoregulations | Dual autoregulation | Regulatory arcs ^a | Positive arcs | Negative arcs |
|-----------------------------------|------------|------------------------|--------------------------|--------------------------|---------------------|------------------------------|---------------|---------------|
| All | 115 | 80 (70) | 24 (30) | 48 (60) | 8 (10) | 166 | 90 (54) | 67 (40) |
| Carbon sources | 24 | 19 (79) | 5 (26) | 9 (48) | 5 (26) | 27 | 20 (74) | 6 (22) |
| Biofilm and motility ^d | 32 | 18 (56) | 9 (50) | 9 (50) | 0 | 52 | 22 (42) | 27 (52) |

Values in parentheses are percentages. Subnetworks for alternative carbon sources and for biofilm and chemotaxis development processes are defined in [Fig. 2a](#). Note that while the carbon sources module has a relatively high maximum out-degree compared to the biofilm/motility module, the latter has a higher maximum in-degree, clearly suggesting that the motility TF, flhCD (with nine inputs), is directed by several TFs to control its regulation.

^aRegulatory interactions from TFs to others TFs or towards sigma factors.

^bAverage path lengths in the (sub)network(s) were calculated with the ViSANT program.¹⁶

^cExcluding sigmas.

^dOnly the TFs forming cascades ending on biofilm and chemotaxis modules were computed, the autoregulation of CRP was computed in the carbon sources module ([Fig. 2a](#)).