From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*

Denis Thieffry,* Araceli M. Huerta, Ernesto Pérez-Rueda, and Julio Collado-Vides

Summary

Because a large number of molecular mechanisms involved in gene regulation have been described during the last decades, it is now becoming possible to address questions about the global structure of gene regulatory networks, at least in the case of some of the best-characterized organisms.This paper presents a global characterization of the transcriptional regulation in *Escherichia coli* on the basis of the current data. The connectivity of the corresponding network was evaluated by analyzing the distribution of the number of genes regulated by a given regulatory protein, and the distribution of the number of regulatory genes regulating a given regulated gene. The mean connectivity found (between 2 and 3) shows a rather loosely interconnected structure. Special emphasis is given to circular sequences of interactions (''circuits'') because of their critical dynamical properties. Only one-element circuits were found, in which negative autoregulation is the dominant architecture. These global properties are discussed in light of several pertinent theoretical approaches, as well as in terms of physiological and evolutionary considerations. *BioEssays* **20**:433–440, 1998.

Introduction

Knowledge in molecular genetics has largely focused on the analysis of specific genetic systems. However, at least in some privileged cases, the corpus already accumulated about individual interactions now permits to get a glimpse at the global regulatory structures involved. In this study, focused on transcriptional regulation, *Escherichia coli* was chosen because it is no doubt the best-characterized free-living organism. Indeed, given its role as a model organism in molecular biology, information is available about its metabolism,[1–3] the organization of its molecular components into different classes,[4,5] and its transcription machinery and regulation.[6] The classic books of the American Society of Microbiology on *E. coli* and *Salmonella* illustrate the large quantity of information available compared with almost any other organism, and certainly to any other bacterium.[7,8] More recently, the group of F. Blattner completed the sequencing of the whole genome of *E. coli*.[9]

Our group has been collecting and analyzing all available information about transcriptional regulation in *E. coli*. This information has been compiled into a dedicated relational database christened *RegulonDB*.[10] The current version contains information about 500 regulatory mechanisms, 100 regulatory proteins and close to 300 promoters or operons. On the basis of this information, we derived a first characterization of the transcriptional network of *E. coli* (defined here as the complete set of transcriptional regulatory interactions),

Centro de Investigación sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México, A.P. 565-A, Cuernavaca, Morelos 62100, México.
Contract grant sponsor: DGAPA-UNAM; Contract grant sponsor: Conacyt.
*Correspondence to: Denis Thieffry, Centro de Investigación sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México, A.P. 565-A, Cuernavaca, Morelos 62100, México. E-mail: denis@cifn.unam.mx

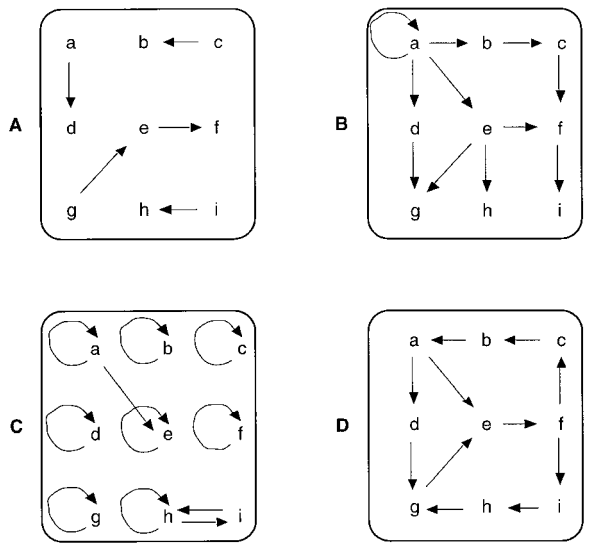

Figure 1. Four didactic examples of gene networks are given, each involving 9 genes and corresponding to one hypothetical organism. The first one (A) includes 5 regulatory interactions, whereas the others (B–D) each includes 11 interactions. Thus, the first network has a average connectivity of 5/9, whereas the others have an identical average connectivity of 11/9. However, networks B–D have strikingly different architectures. Indeed, the network B constitutes a hierarchical regulatory tree, network C can be seen as a set of small independent regulatory modules, whereas network D contains a set of intertwined feedback circuits. Interaction signs have been omitted here for sake of simplicity. Strictly speaking, a network is a fully connected set of genes where there is always a path to navigate from one gene to any other gene through regulatory interactions. Here, we use a more flexible definition that encompasses sets of genes where not all genes are fully connected, leaving the possibility for isolated subnetworks or even isolated genes.

estimating its connectivity (i.e., the mean number of regulatory interaction per gene in the network), together with an estimation of the numbers and types of regulatory feedback circuits (i.e., circular cascades of interactions).

E. coli transcriptional regulatory network

Connectivity

Several theoretical studies have already questioned the relation between gene network architecture and dynamical behavior.[11] In this respect, a first important architectural criterion is certainly the average *connectivity*, which gives an estimate of the number of interactions per gene. Figure 1 gives four didactic examples of cellular regulatory networks, each comprising nine genes.

Network A includes 5 interactions, whereas networks B, C, and D each contains 10 interactions. It is easy to compute that network A has a connectivity slightly greater than 1/2, whereas the other networks have a connectivity slightly greater than 1. However, even though they have similar connectivity, networks B, C, and D have very different architectures (see next section).

In order to characterize the *E. coli* transcriptional regulatory network, we first addressed the two following questions: (1) What is the connectivity distribution in the *E. coli* regulatory network? and (2) What is the mean connectivity of the entire network?

A genetic regulatory network can be conceived as a set of interconnected elements, where the elements are regulatory and regulated genes, and whose connections are interactions directed from regulatory genes towards regulated genes. Given the data available, we were limited to only those interactions describing regulation at the level of the initiation of transcription. These are the direct regulatory interactions. However, depending on the connectivity of the network, one can also think of indirect interactions by means of a cascade of direct interactions.

To obtain an estimate of the connectivity of the whole network, we distinguished the number of *interactions leaving* a regulatory gene on the one hand, and the number of *interactions arriving* at regulated genes on the other hand. Figure 2a shows the number of regulators regulating a given number of genes, i.e., *interactions leaving* regulatory genes. The second histogram (Fig. 2b) gives the distribution of genes regulated by a given number of regulators, that is to say, the distribution of *arriving interactions.* For example, the last bar of Figure 2a indicates that there is a single protein (CRP) regulating 133 genes, whereas the last bar of Figure 2b indicates that there is only one gene (sodA) regulated by as many as six different proteins.

On the basis of these histograms, one can evaluate that the mean connectivity lies somewhere between two and three. More precisely, a transcription factor regulates on average three genes, and an *E. coli* gene is under direct control of two transcription factors.

Figure 2 gives an idea of the complexity of transcriptional regulation in *E. coli*, which reflects how flexible the expression of a given gene can be. This fits with the observed prominence of promoters regulated by a single regulator in *E. coli* (see Table 1 in reference 6). Two transcriptional factors on average are sufficient in *E. coli* to determine the patterns of gene regulation.

It should be taken into account, however, that *E. coli* genes often lie within multigenic operons. Figure 2c shows the number of proteins regulating a given number of promoters, and Figure 2d indicates the number of promoters regulated by a given number of proteins. It can be seen that, when

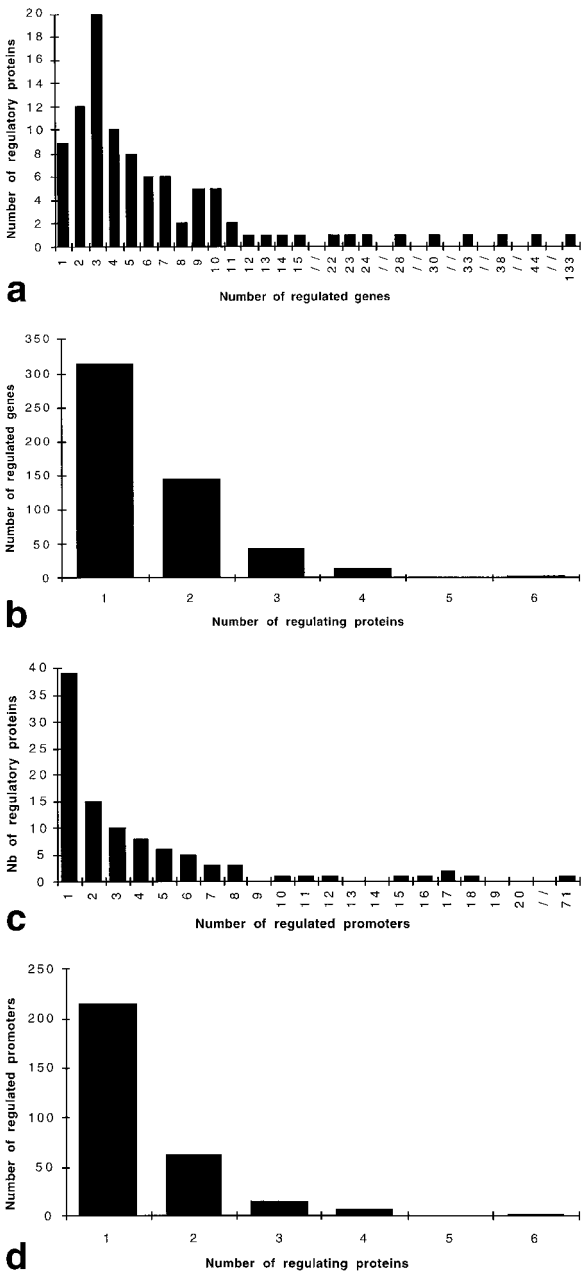

Figure 2. Histograms giving the distributions of characterized transcriptional regulations in *E. coli*. a: Numbers of regulatory proteins regulating 1, 2, 3, . . . , 133 genes. b: Numbers of genes regulated by 1, 2, . . . , 6 regulatory proteins. c: Numbers of regulatory proteins regulating 1, 2, 3, . . . , 71 promoters. d: Numbers of promoters regulated by 1, 2, . . . , 6 regulatory proteins. From these histograms, one can estimate that the mean connectivity lies somewhere between two and three.

TABLE 1. Main Dynamic and Biological Properties of the Two Families of Feedback Regulatory Circuits

Characteristics	Positive circuits	Negative circuits
No. of negative interactions	Even	Odd
Dynamic property	Multistationarity	Homeostasis
Biological property	Differentiation	Homeostatic regulation

focusing on operons or promoters, the connectivity estimation is significantly lowered.

Feedback circuits

The connectivity of the *E. coli* regulatory network provides interesting insights into the structure of the entire regulatory system. However, this is not sufficient for understanding the system as a whole. Indeed, very different types of regulatory structures might be characterized by roughly an identical mean connectivity of departing and arriving interactions. In Figure 1, for example, network B constitutes a hierarchical regulatory tree, network C a set of independent, small regulatory modules, and network D contains a set of intertwined feedback circuits.

A complementary description of the *E. coli* transcriptional regulatory network could be focused on the regulatory interactions forming closed circuits, as well as on the number of elements involved in such circuits. These circuits are crucial with respect to the dynamical behavior of the network, as shown by the work of R. Thomas and his collaborators.[12–16] Following the definition of Thomas, a regulatory circuit simply consists of a cascade of regulatory interactions closing on itself. Table 1 summarizes the properties of positive and negative regulatory circuits.* The importance of regulatory circuits leads us to address the following questions:

1. How many regulatory feedback circuits are found in the *E. coli* transcriptional network?
2. What is their average length?

*It is crucial to distinguish between the notion of "interaction" or simple "feedback" vs. the notion of "regulatory feedback circuit." For example, the *lac* system involves a positive "feedback circuit," but its mode of regulation is negative (repressor). In fact, the positive circuit involves the inducing metabolite (lactose), the specific permease, and the repressor. Thus, in addition to the transcriptional control, this circuit involves several metabolic events. For an interesting description of this system, including an account of some of the early experiments, see ref. 14 (pp. 201–202)

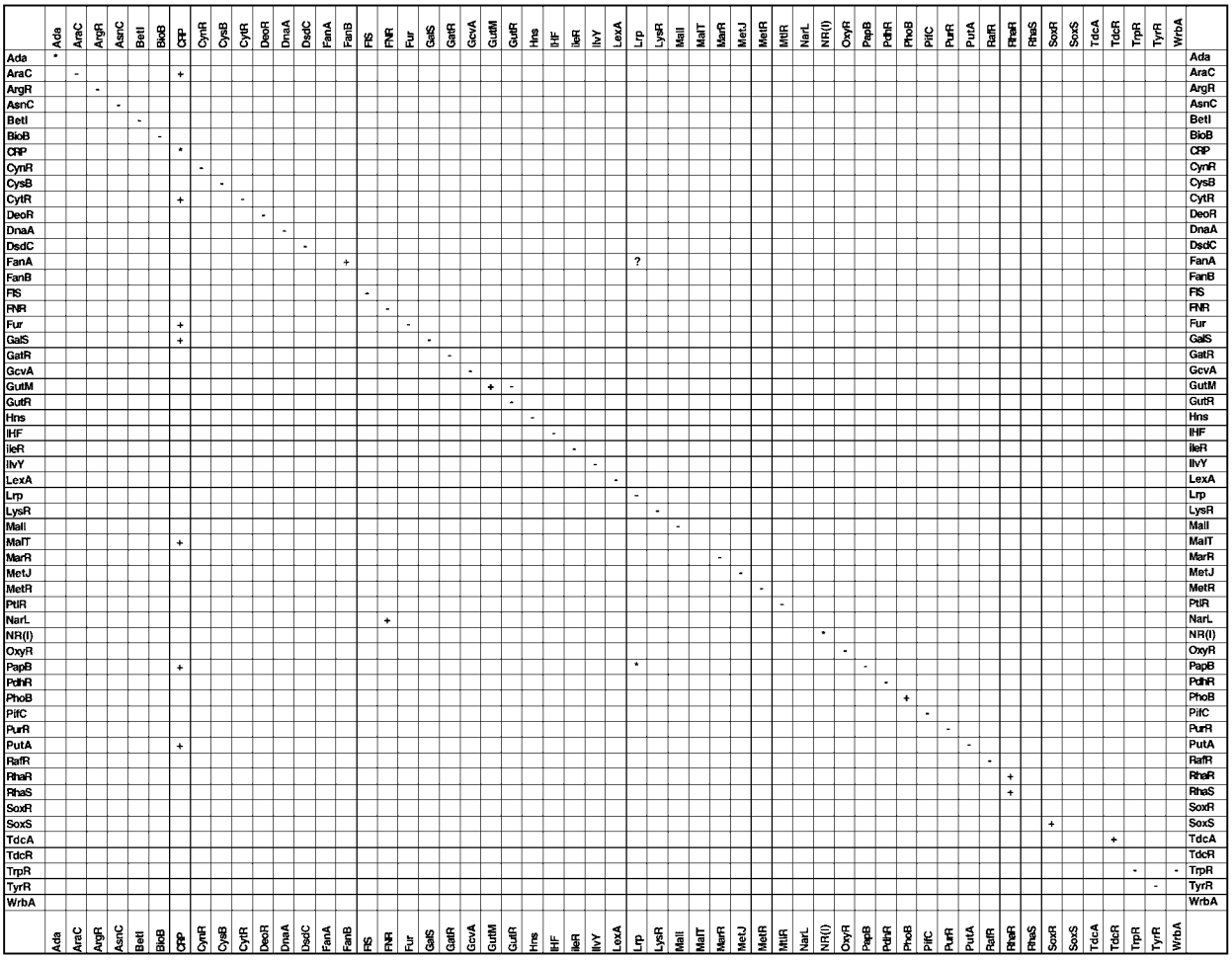

Figure 3. Matrix summarising all characterised cross-interactions occurring among *E. coli* transcriptional regulators. When filled, a box of this matrix gives the sign of the effect of the regulator of the corresponding column on the expression of the regulator of the corresponding row; "+", "−", and "*" stand for positive, negative, and dual interactions, respectively; "?" stands for an interaction whose mode remains uncharacterised.

3. What fraction of feedback circuits is positive, and what fraction is negative?
4. How are these circuits interconnected?

Even though it is known that many regulatory circuits involve both genetic and metabolic elements, here we limited ourselves to the analysis of the sole transcriptional regulatory network. To address our four questions, we built a matrix containing the 96 well-characterized genes involved in the regulation of transcription in *E. coli*. Out of these 96 transcriptional regulators, we found that 55 are involved in interregulation. Figure 3 gives the 55 × 55 interaction matrix obtained after discarding all regulatory genes that do not exert or receive interactions involving other regulatory genes.

In this matrix, all the genes controlled by the same protein are located in the same column, whereas all regulators of a given gene are found in the same row. One-element circuits are easily identified on the diagonal of the matrix. Two-element circuits would be found as pairs of filled boxes symmetrically located with respect to the diagonal. More generally, any circuit would consist of a set of boxes whose row and column coordinates form a circular permutation.

The structure of this matrix came as a surprise. Indeed, this matrix contains mainly diagonal terms (with the notable exception of CRP that regulates the transcription of several other regulators). This indicates that many regulators autoregulate themselves, but that there is little cross-transcriptional regulation among *E. coli* regulators. In addition, most of

the autoregulations are in fact auto-inhibitions; only six regulators are positively autoregulated—Ada, CRP, GutM, NR(I), PhoB, and RhaR—including three regulators that exert both positive and negative regulation on their own expression, i.e., Ada, CRP, and NR(I).

There are several interesting properties to be observed in this matrix. The total number of feedback circuits is 45, which primarily includes negatively regulated circuits of one element (negative autoregulation), plus 3 positive and 3 dual (positive and negative) circuits of one element. This gives 87% negative circuits, 6.5% positive circuits, and 6.5% dual circuits. The low number of elements within these closed circuits is quite surprising. We discuss plausible interpretations of these observations below.

Biological significance

These results indicate a rather low connectivity for the *E. coli* transcriptional regulatory network. At the level of operon regulation, the average connectivity is even lower. From our point of view, however, the most intriguing and robust result of our analysis of the *E. coli* transcriptional regulatory matrix consists of the clear prominence of one-element regulatory circuits, especially of auto-inhibitions.

Escherichia coli thus appears to be composed essentially of small regulatory subnetworks that are loosely interconnected. This seems reasonable when we recall that most perturbations of the expression of a given transcriptional regulator typically leads to a change of expression of a limited number of coordinated genes, usually grouped into one operon or within a regulon. This modular organization of the complete network should help the cell to evolve efficiently.

As suggested by R. Thomas, each negative circuit might result in a *homeostatic* expression of the repressor gene and of the genes it controls. Indeed, while a simple negative control tends to maintain the regulated gene ''off'' under proper external conditions, a negative autoregulation tends to keep the regulated gene ''half-on,'' ''half-off.'' Strikingly, even most of the ''positive'' transcriptional regulators exert negative control on themselves. This could indicate that homeostatic control might be selected to preserve bacteria from overexpression of regulatory genes and their associated toxic effects (classic examples include genes cII and N of λ phage.[17] In addition, it is presumably less expensive to control the expression of a gene directly by its product than via another protein. Finally, such negative autoregulation could buffer stochastic variations of gene expression (for a discussion of the importance of these effects, see ref. 18.

On the other hand, each of the six positive circuits found in *E. coli* could constitute a switch between two stable regimes of gene expression, directly or indirectly affecting an important number of genes. Moreover, these positive circuits can typically be switched *stably* from one expression regime to another one, even under the action of a *transient* signal (see ref. 14, pp 173–179). Clearly, four of these factors could constitute such switching devices: *CRP*, the regulator of catabolic repression, which is the regulator with the highest number of regulated genes[19]; *NR(I)*, the key regulator of nitrogen metabolism[20–22]; and *PhoB*, which regulates the expression of a series of genes involved in the transport and intracellular regulation of Pi.[23,24] Each of these genes appears to play a decisive role in the irreversible triggering, at least on a short time scale, of a series of other (structural or regulatory) genes. This could also be the case of *Ada*, which is involved in the induction of the DNA-repair system that protects the bacterium against methylating and alkylating agents.[25–27] However, the other two positively autoregulated genes (*GutM* and *RhaR*) seem to be involved in more local regulatory systems.

The data just described suggest that there are few positive feedback circuits in *E. coli*. It should be recalled, however, that these results refer only to transcriptional regulation. In fact, other positive regulatory feedback circuits have already been characterized, although these circuits usually include at least one metabolite as a key feedback element, such as in the *lac* system (see the footnote above). Thus, the low number of positive circuits found could be the consequence of a systematic involvement of metabolic signals in positive circuits. Because such mixed metabolic/genetic positive circuits need the continuous presence of the involved metabolites to remain active, they allow the cell to monitor the presence of such metabolites continuously.

Discussion

Robustness of the data

The total number of genes in *E. coli* is roughly estimated at 4,000, from which about 1,000 are now functionally characterized. Recalling that the database we used in this analysis contains around 500 genes and assuming 20–50% of all *E. coli* genes to be constitutively expressed, we estimate that the set used in this analysis represents between 15–25% of the whole transcriptional regulatory network. Alternatively, if one assumes a 1 : 10 ratio of regulatory to regulated genes,[3,28] one would expect about 400 regulatory genes. Since our data set contains 100 regulatory proteins, this produces again an estimate of a quarter of the entire transcriptional network.

A different question is to determine how representative are the data with respect to the number and type of transcriptional interactions in the whole network. This is a much more difficult matter. Indeed, the experimental techniques and concepts available for the study of gene regulation may well produce a bias toward what we currently know about the complete regulatory network.

Because we limit ourselves here to interactions at the level of transcription, we are well aware that the interactions

considered in this analysis represent a small fraction of the biological regulatory interactions. Furthermore, even though we took into account all recognized specific (e.g., MalT, LacI) and global (e.g., CRP, FNR, NR(I)) transcriptional regulators, we did not include the elements involved in the RNA polymerase holoenzyme, nor the σ factors. However, in *E. coli*, these transcriptional proteins seem to be involved in transcriptional regulatory cascades rather than in regulatory circuits and therefore should not affect the estimate on regulatory circuits. Nonetheless, important integrative regulatory responses of the cell occur at the level of protein–protein interactions, not considered in this initial global analysis.

Thus, the precise numbers and distributions here presented should be considered as preliminary estimations of the connectivity and structure of the *E. coli* transcriptional network. As additional molecular data become available, the description of the complete network might be attainable. Nevertheless, several features that arise from the analysis presented in this paper are likely to be sufficiently robust to remain true. Indeed, several arguments based upon physiological considerations and computational studies might be advanced.

Concerning the low connectivity obtained, one could think that new studies will progressively uncover additional interactions, leading to an increase of the connectivity of the *E. coli* transcriptional network. Nonetheless, out of the 300 promoters encompassed by this study, about 130 have regulatory regions that have been well characterized experimentally.[29] This set includes some promoters that have been studied for years (e.g., the *lac, ara, gln*, and *deo* systems). It seems quite improbable that new regulatory proteins affecting these promoters will be discovered.

Another piece of evidence comes from recent computational studies of the DNA sequences of these regulatory regions. When searching for binding sites of the 40 regulators that are known to regulate these 130 promoters, we found little convincing evidence for strong new sites, supporting the idea that our knowledge of the regulation of this set is close to complete.[30,31]

Finally, the low connectivity observed might also be understood in terms of the structural properties of the transcription machinery in *E. coli*. Certainly, regulation of transcription in σ^{70} promoters is known to require sites that are located close to the transcription initiation to enable direct contact of the regulator and the RNA polymerase. This restriction will make it difficult for a promoter to be independently regulated by a large number of transcriptional factors, as compared with regulation in eukaryotic promoters, where regulation occurs from remote distances. Given that the σ^{70} factor defines an evolutionary family of σ factors in bacteria,[32] one can assume that (perhaps with the exception of the evolutionarly distant σ^{54} promoters) this structural restriction will prevail.

Finally, the present study has been performed twice: first with the limited collection of 130 promoters already mentioned, and a second time with an extended collection including all the information contained in various databases (e.g., MedLine, GenBank). Even though the number of regulatory genes and interactions considered was roughly doubled in the process, the mean connectivity, as well as the distribution of the different types of feedback circuits was largely conserved.

Comparison with other theoretical approaches

Since the late 1960s, Stuart Kauffman has modeled genetic networks with Boolean equations (i.e., equations whose variables can take only two values, 0 and 1).[11,33,34] On the basis of both biological case studies and theoretical analyses considering any set of interactions that might occur in random Boolean networks, Kauffman predicts a low connectivity and the occurrence of intertwined feedback circuits. The first prediction is clearly compatible with the structure found for *E. coli* transcriptional network; the prediction of intertwined feedback circuits, however, is clearly challenged by our observations, at least for *E. coli*.

Beginning in the 1970s, another important theoretical approach was initiated by Michael Savageau.[35,36] Savageau initiated a systematic comparative analysis of regulatory mechanisms in relation to the natural environment of organisms. According to the *"demand theory of gene regulation,"* Savageau predicted that regulation by a repressor is selected when there is a low demand for the expression of the regulated gene(s) in the natural environment of the cell, whereas regulation by an activator is selected when there is a high demand for their expression. Moreover, when the demand for the expression of a given structural gene changes as a consequence of changes in the environment, then a switch in the regulatory mode is predicted.[37,38]

As far as we know, the properties of the *E. coli* network reported here are consistent with Savageau's predictions. In particular, in the case of autonomous systems, in which a protein directly regulates its own expression, Savageau predicted the prevalence of the negative mode.[39,40] In addition, in the case of moderate coupling between regulator and regulated genes, which is the most common case, Savageau predicted on functional grounds that the regulator should affect its own transcription negatively regardless of whether it affects other transcriptional units negatively or positively. Detailed experimental data, as well as our global analysis, unambiguously support these predictions. In addition, we suggest that other constraints might play a role in establishing the regulatory pattern of *E. coli*, i.e., an additional evolutionary advantage related to transcriptional auto-inhibition, together with a preferred occurrence of mixed metabolic/genetic, positive regulatory circuits.

Conclusions and perspectives

This paper presents a global picture of the *E. coli* network based on available information about transcriptional regulation. Some features that begin to emerge are a low connectivity, a large number of autoregulatory interactions with a predominance of negative autoregulation, and very few feedback circuits involving more than one gene.

Even though we still have limited information on transcriptional regulation in eukaryotes, we already have indications that the regulatory structures involved are quite different (e.g., wider occurrence of positive autoregulations, as well as at least some multielement feedback circuits). It will be interesting to learn if the structural features of the *E. coli* transcriptional regulatory network are common to other unicellular organisms and/or prokaryotes, as opposed to higher organisms with richer developmental processes. These questions will be amenable to evaluation as our knowledge of other organisms progresses.

There is no doubt that additional feedback circuits that were not taken into account in this analysis play an important role in *E. coli* gene regulation, in particular circuits involving post-transcriptional regulation or various metabolic events (e.g., proteolysis, phosphorylation). Thus, a full characterization of the *E. coli* regulatory network will imply not only the inclusion of post-transcriptional regulations, but also a further integration of the genetic and metabolic knowledge of *E. coli*.

In addition to new insights and global perspectives on the *E. coli* regulatory network, our analysis also leads to specific experimental suggestions. For example, in the case of the positive autoregulatory circuits, one could look for experimental conditions in which a transient signal leads to a durable and coordinated switching of controlled genes. On the other hand, it would be interesting to examine the physiological role that we proposed for the ubiquitous negative autoregulations, i.e., buffering of gene expression and protection against toxic effects.

Clearly, as attested by several recent publications,[41–43] we are still far from understanding the internal organization of a whole cell, even in the case of *E. coli.* Nonetheless, the work presented here constitutes a preliminary step in that direction. As a result of the multiplication of genome projects and annoted databases, describing, analyzing, and understanding the internal network of regulatory interactions in a cell is now becoming attainable. Such analyses are likely to play an important role in the search for a deeper understanding of the physiology of whole organisms, as well as in the development of comparative genomics.

Acknowledgments

We acknowledge Roberto Kolter, Boris Magasanik, John Reinitz, Michael Savageau, Judy Schloegel, Jim Shapiro, René Thomas, and Ariane Toussaint for their critical reading, fruitful discussions and many useful suggestions. We also acknowledge Heladia Salgado for her contribution to the collection of regulatory information. This work was supported by grants from DGAPA-UNAM and Conacyt (to J. C.-V.).

References

1 Médigue C, Viari, A, Henaut A, and Danchin A (1993) Colibri: A functional data base for the *Escherichia coli* genome. *Microbiol Rev* 57:623–654.
2 Rouxel T, Danchin A, and Henaut A (1993) METALGEN.DB: Metabolism linked to the genome of *Escherichia coli*, a graphics-oriented database. *Comput Appl Biosci* 9:315–324.
3 Karp PD, Riley M, Paley SM, and Pelligrini-Toole A (1996) EcoCyc: An encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res* 24:32–39.
4 Labedan B and Riley M (1995) Widespread protein sequence similarities: Origins of *Escherichia coli* genes. *J Bacteriol* 177:1585–1588.
5 Labedan B and Riley M (1995) Gene products of *Escherichia coli*: Sequence comparisons and common ancestries. *Mol Biol Evol* 12:980–987.
6 Gralla JD, and Collado-Vides J (1996) Organization and Function of Transcription Regulatory Elements. In F.C. Neidhardt, R. Curtiss III, J. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W. Reznikoff, M. Schaechter, H.E. Umbarger, and M. Riley (eds): Cellular and Molecular Biology: *Escherichia coli* and *Salmonella typhimurium.* 2nd Ed. Washington, DC: American Society for Microbiology, pp 1232–1245.
7 Neidhardt FC, et al (ed.) (1987) *Escherichia coli* and *Salmonella typhimurium. Cellular* and *Molecular Biology.* Washington, DC: American Society for Microbiology.
8 Neidhardt FC, et al (ed.) (1996). *Escherichia coli* and *Salmonella typhimurium. Cellular* and *Molecular Biology.* 2nd Ed. Washington, DC: American Society for Microbiology.
9 Blattner F, et al (1997) The complete sequence of *Escherichia coli* K-12. *Science* 277:1453–1461.
10 Huerta AM, Salgado H, Thieffry D, and Collado-Vides J (1998) RegulonDB: A database on transcriptional regulation in *Escherichia coli*. Nucleic Acids Res. 26:55–59.
11 Kauffman SA (1993) *The Origins of Order. Self-Organization and Selection in Evolution.* Oxford: Oxford University Press.
12 Thomas R (1973) Boolean formalization of genetic control circuits. *J Theor Biol* 42:563–585.
13 Thomas R (1978) Logical analysis of systems comprising feedback loops. *J Theor Biol* 73:631–656.
14 Thomas R and D'Ari R (1990) *Biological Feedback.* Boca Raton, FL: CRC Press.
15 Thomas R (1991) Regulatory networks seen as asynchronous automata: A logical description. *J Theor Biol* 153:1–23.
16 Thomas R, Thieffry D, and Kaufman M (1995) Dynamical behaviour of biological regulatory networks. I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol* 57:247–276.
17 Shimatake H and Rosenberg M (1981) Purified λ regulatory protein CII positively activates promoters for lysogenic development, *Nature* 292:128–132.
18 McAdams HH and Arkin A (1997) Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA* 94:814–819.
19 Ishizuka H, Hanamura A, Inada T, and Aiba H (1994) Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: Role of autoregulation of the crp gene. *EMBO J* 13:3077–3082.
20 Reitzer LJ and Magasanik B (1986) Transcription of glnA in *E. coli* is stimulated by activator bound to sites far from the promoter. Cell 45:785–792.
21 Popham DL, Szeto D, Keener J, and Kustu S (1989) Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* 243:629–635.
22 Fiedler U and Weiss V (1995) A common switch in activation of the response regulators NtrC and PhoB: Phosphorylation induces dimerization of the receiver modules. *EMBO J* 14:3696–3705.
23 Torriani A (1990) From cell membrane to nucleotides: The phosphate regulon in *Escherichia coli. Bioessays* 12:371–376.
24 Wanner BL (1993) Gene regulation by phosphate in enteric bacteria. *J Cell Biochem* 51:47–54.
25 Nakabeppu Y and Sekiguchi M (1986) Regulatory mechanisms for induction of synthesis of repair enzymes in response to alkylating agents: Ada acts as a transcriptional regulator. *Proc Natl Acad Sci USA* 83:6297–6301.
26 Landini P and Volkert MR (1995) RNA polymerase alpha subunit binding site in positively controlled promoters: A new model for RNA polymerase–promoter interaction and transcriptional activation in the *Escherichia coli* ada and aidB genes. *EMBO J* 14:4329–4335.
27 Saget BM and Walker GC (1994) The ada protein acts as both a positive and a negative modulator of *Escherichia coli*'s response to methylating agents. *Proc Natl Acad Sci USA* 91:9730–9734.
28 Riley M (1993) Functions of the gene products of *Escherichia coli. Microbiol Rev* 57:862–952.
29 Collado-Vides J, Magasanik B, and Gralla JD (1991) Control site location and transcriptional regulation in *Escherichia coli. Microbiol Rev* 55:371–394.

30 Rosenblueth DA, Thieffry D, Huerta AM, Salgado H, and Collado-Vides J (1996) Syntactic recognition of regulatory regions in *Escherichia coli. Comput Appl Biosci* 12:415–422.
31 Thieffry D, Rosenblueth DA, Huerta AM, Salgado H, and Collado-Vides J (1996) Definite-Clause Grammars for the analysis of cis-regulatory regions in *E. coli.* In R. Altman, K. Dunker, L. Hunter, and T. Klein (eds): *Proceedings of the Pacific Symposium on Biocomputing '97.* Singapore: World Scientific, pp 441–452.
32 Lonetto M, Gribskov M, and Gross CA (1992) The sigma 70 family: Sequence conservation and evolutionary relationships. *J Bacteriol* 174:3843–3849.
33 Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 22:437–467.
34 Kauffman SA (1973) Control circuits for determination and transdetermination. *Science* 181:310–318.
35 Savageau MA (1977) Design of molecular mechanisms and the demand for gene expression. *Proc Natl Acad Sci USA* 74:5647–5651.
36 Savageau MA (1983) Regulation of differentiated cell-specific functions. *Proc Natl Acad Sci USA* 80:1411–1415.
37 Hlavacek WS and Savageau MA (1995) Subunit structure of regulator proteins influences the design of gene circuitry: Analysis of perfectly coupled and completely uncoupled circuits. *J Mol Biol* 248:739–755.
38 Savageau MA (1976) Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology. Reading, MA: Addison-Wesley.
39 Savageau MA (1979) Autogenous and classical regulation of gene expression: A general theory and experimental evidence. In R.F. Goldberger, P. Berg, R.T. Schimke, K. Moldave, P. Leder, and L.E. Hood (eds): *Biological Regulation and Development,* Vol. I. New York: Plenum, pp 57–108.
40 Hlavacek WS and Savageau MA (1996) Rules for coupled expression of regulator and effector genes in inducible circuits. *J Mol Biol* 255:121–139.
41 Palsson BO (1997) What lies beyond bioinformatics? Nature Biotechnol 15:3–4.
42 Strohman RC (1997) The coming Kuhnian revolution in biology. *Nature Biotechnol* 15:194–200.
43 Moxon ER, and Higgins CF (1997) A blueprint for life. *Nature* 389:120–121.