Hierarchical dynamics of a transcription factors network in *E. coli*†

Agustino Martínez-Antonio,*^a David A. Velázquez-Ramírez,^a
Joel Sánchez-Mondragón^a and Moisés Santillán^b

Received 12th April 2012, Accepted 31st July 2012

DOI: 10.1039/c2mb25236h

In the present work we study the transcription-factor regulatory network that controls the synthesis of flagella in *E. coli*. Our objective is to address how the transcription-factor dynamics (in terms of their promoter activities and associated rates) correlate with their positions in the hierarchical organization of this regulatory network. Our results suggest that global-regulator promoters express at higher rates than those of local regulators, particularly when the bacterial populations are actively growing. Furthermore, promoter activity decreases together with the rate of cellular division. And finally, local-regulator promoters reach their maximal activity later than global-regulator promoters do. In summary, our results suggest a strong correlation between promoter activities and their hierarchical organization in this particular regulatory network.

Introduction

Interactions among the molecular entities of a living organism are essential to understand its capacity to maintain an optimal performance in a constantly changing environment. These interactions usually take place at either the transcriptional—by tuning the expression level of specific genes—or the post-transductional—by modifying the existing molecular machinery—levels. In the case of *E. coli*, there exist a number of databases compiling the interactions at the transcriptional level. Arguably, the most complete of them is RegulonDB.¹ The *E. coli* transcriptional regulatory network has been reconstructed from the information in this database.² Some other works have been devoted to studying the network structure and to inferring biological functions,^{2–9} while their theoretical predictions have been experimentally validated.^{10–13} In a particularly interesting paper,¹⁴ Zaslaver *et al.* studied the sequential expression of the genes coding for the enzymes that participate in the biochemical pathways leading to the synthesis of some amino acids. They identified a temporal expression program and expression hierarchy that matches the enzyme order in unbranched pathways. In fact, they found a couple of design principles: the closer the enzyme is to the beginning of the pathway, the shorter the response time for the activation of its promoter and the higher its maximal promoter activity. They suggest from these results that metabolic regulation networks are designed to generate precision promoter timing and activity. Having this in mind, here we test whether these design

principles can be identified in complex transcriptional regulatory networks among transcription factors in which self-regulation is present, the genes coding for transcription factors are hierarchically distributed in the network, and several genes are regulated by more than one transcription factor.

In the present work we analyze the transcription-factor regulatory network that controls the synthesis of flagella in *E. coli*. In particular, we investigate how the activity of the genes coding for eight transcription factors (TFs) depends on their hierarchical localization in the network, as reconstructed from literature. Our results suggest that the genes coding for global TFs express at higher rates, and that they reach their maximal expression level earlier than the genes coding for local TFs.

The network of transcription factors considered in this study

Previously, we reported a literature-reconstructed transcriptional regulatory network for *E. coli*, consisting of 115 transcription and 7 sigma factors.² In the present work we study the sub-network responsible for controlling the synthesis of flagella. This sub-network comprises nine regulatory genes hierarchically organized (Fig. 1), and has the following characteristics: (a) it is the largest sub-network in the *E. coli* transcriptional regulatory pathway;² (b) it includes both global (nucleoid-associated proteins, NAPs) as well as local regulators;¹⁵ and (c) given the growing conditions here employed, none of these regulators is directly involved in activities that would imply a high demand like: carbon source uptake, respiration mode, a defined stress, *etc.*

The hypothesis of hierarchical activity of TFs

The hypothesis behind this work is that, given the different qualitative roles of the TFs in a hierarchical network, the activities of the corresponding promoters should be correlated

^a Departamento de Ingeniería Genética, Cinvestav, Km. 9.6 Libramiento Norte Carr. Irapuato-León 36821 Irapuato Guanajuato, México

^b Centro de Investigación y de Estudios Avanzados del IPN, Unidad Monterrey, Vía del Conocimiento 201, 66600 Apodaca NL, México

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb25236h

Fig. 1 Schematic representation of the transcription-factor regulatory network controlling the synthesis of flagella in *E. coli*. The nodes represent the transcription factor genes, while the edges denote regulatory interactions: green, activation; and red, repression.

to their hierarchy within the network. This should be evident in at least two variables: the maximal promoter activities and the promoter temporal expression patterns. The first point is supported by the fact that global and local classes of regulatory proteins differ in their abundance of mRNAs and protein products per cell; *i.e.* higher levels for NAPs and global regulators and lower levels for the local ones.^{16–18} On the other hand, a temporal expression pattern of promoter activities might be also expected due to the following observations:

(1) The most common mode of regulation between every two genes in the *E. coli* transcriptional regulatory network is activation. This further indicates the existence of regulatory pathways that convey transcriptional programs inside the genetic regulatory network of *E. coli*, even if self-regulation is found at each step in the pathway;²

(2) Global TFs respond to endogenous conditions¹⁹ directing global processes for cellular maintenance and growth, whereas local regulators sense mostly exogenous signals and work under stressing conditions when bacteria decrease their growth rate.

In the following sections we shall present experimental evidence supporting this hypothesis.

Material and methods

Bacterial strains and media

In all our experiments we employ the *E. coli* K12 MG1655 strain. We use the LB and M9 media (the latter supplemented with 0.4% glucose) following the protocols in ref. 10. When indicated, we use kanamycin (Km) 50 µg ml⁻¹. The pre-inoculates were grown in 5 ml of LB medium with agitation (200 rpm). Next, the cultures were diluted 1 : 100 in 150 µl of fresh M9 media in micro-titer plates of 96 wells and incubated for 11 h with agitation (250 rpm) and at 37 °C. We follow bacterial growth, by measuring OD_{595nm}, and fluorescence (535 nm) every hour in a Perkin Elmer Victor X3 plate multi-lector.

Construction of transcriptional fusions

Plasmids pUA66 and pUA139 were used as backbones to clone promoters on direct or inverted orientations of the nine TFs considered. Both plasmids contain the *gfpmut2* gene, which codes for a GFP (green fluorescent protein) protein of rapid maturation.¹⁰ The primers used to amplify PCR products were designed to cover all the TF regulatory sites annotated in RegulonDB.¹ A list of primers for each promoter sequence is available in the ESI.† With these genetic constructions, harboring the different transcriptional fusions, we transformed the K12 MG1655 strain using standard protocols. The only transcriptional fusion we were unable to achieve was that corresponding to the FlhDC regulator (a master regulator of motility and flagellar genes), and so it was ruled out in this study.

Data treatment

The numerical data obtained from the Victor X3 plate multi-lector (OD and GFP) were processed using the QtiPlot software (QtiPlot 0.9.7.10 Copyright © 2004–2009, Ion Vasilief, released: 2009/10/08). The raw data consisted of discrete measurements of growth curves (optical density *vs.* time) and GFP profiles (fluorescence intensity *vs.* time). We found that, in all cases, the sigmoidal Boltzmann function fits both the growth curves and the GFP profiles, with correlation factors higher than 0.99. We limited our analysis to the first 8 hours after culture because this time period precedes the bacterial death phase.

The functions used to fit the OD and GFP profiles are:

$$\text{OD}(t) = \frac{a_1 + b_1 e^{(t-t_1)/z_1}}{1 + e^{(t-t_1)/z_1}}, \quad (1)$$

$$\text{GFP}(t) = \frac{a_2 + b_2 e^{(t-t_2)/z_2}}{1 + e^{(t-t_2)/z_2}}, \quad (2)$$

In eqn (1), a_1 and b_1 , respectively, stand for the minimum and maximum possible values of the function. Moreover, t_1 is the time at which the function attains the midpoint between a_1 and b_1 : $\text{OD}(t_1) = (a_1 + b_1)/2$. This point is also the inflection point of the growth curve: for $t < t_1$ ($t > t_1$) the population growth speed is an increasing (decreasing) function of time. Finally, z_1 determines the function slope at the inflection point: larger z_1

values imply smaller slopes. Regarding eqn (2), the meaning of parameters a_2 , b_2 , t_2 and z_2 is equivalent to that of the corresponding parameters in eqn (1).

Let us now assume that the culture optical density, $OD(t)$, is proportional to the bacterial count, $N_b(t)$. That is: $OD(t) = \alpha N_b(t)$. Let $\mu(t)$ be the bacterial population growth rate. Hence:

$$\frac{dN_b(t)}{dt} = \mu(t)N_b(t).$$

On the other hand, assume that the GFP fluorescence intensity is proportional to both the bacterial count and the average protein count per bacterium, $n_p(t)$: $GPF(t) = \beta N_b(t)n_p(t)$.

Finally, if the protein degradation rate is negligible as compared to the bacterial growth rate:

$$\frac{dn_p(t)}{dt} = \kappa(t) - \mu(t)n_p(t),$$

where $\kappa(t)$ is the protein production rate per bacterium. It follows from the above considerations that

$$\frac{dGFP(t)/dt}{OD(t)} = \frac{\beta}{\alpha} \kappa(t).$$

From the above considerations we define the promoter activity, $PA(t)$, as

$$PA(t) = \frac{dGFP(t)/dt}{OD(t)}.$$

As we have seen, this function estimates the average protein production rate per cell.

Although the same growth conditions were employed in all the experiments, the growth curves differed in the t_1 values. We believe this variability is due to intrinsic biochemical noise and/or the metabolic burden due to GFP expression.^{20–26,28} Thus, to make all promoter activity curves comparable we translate them along the temporal axis so the zero time point is at the inflection point of the corresponding growth curve: $PA(t) = PA(t + t_1)$. After substitution of eqn (1) and (2) into the definition of $PA(t)$ we obtain

$$PA(t) = -\frac{(a_2 - b_2)(1 + e^{t/z_1})e^{(t+t_1-t_2)/z_2}}{z_2(a_1 + b_1e^{t/z_1})(1 + e^{(t+t_1-t_2)/z_2})^2} \quad (3)$$

Results and discussions

Growth and promoter-activity profiles

As described in the *Materials and methods* section, we engineered nine bacterial strains, each one containing a reporter *gfp* gene, whose expression is controlled by the promoter of one of nine different transcription factors in the gene network responsible for the synthesis of flagellar proteins. We cultured all the strains in a glucose-enriched minimal medium (see *Materials and methods*), and recorded the optical density and GFP-fluorescence profiles every hour, for eleven hours. We repeated these experiments three times for each bacterial strain. Next, we took the raw data and fitted the optical density and fluorescence profiles to the sigmoidal Boltzmann

Fig. 2 Representative plots of the experimental data OD vs. t (A) and GFP vs. t (B), as well as of the corresponding best fitting sigmoidal Boltzmann functions, for each of the nine different engineered bacterial strains. In both plots, AU stands for arbitrary units. Each plot corresponds to a different strain, with the color code as follows: blue, *hns::gfp*; green, *ihfB::gfp*; brown, *ihfA::gfp*; purple, *fis::gfp*; orange, *ompR::gfp*; light blue, *hdfR::gfp*; red, *lrhA::gfp*; gray, *fliA::gfp*; and pink, *qseB::gfp*.

functions (1) and (2), see Fig. 2. Finally, we took the resulting fitness parameters to compute the promoter activity vs. time curves for every experiment by means of eqn (3). In Fig. 3 we show representative PA vs. t curves, one for every bacterial strain. Observe that all the promoter activity curves are concave functions of time with a single maximum. Based on this observation, we calculated the maximum PA values ($PA_{\max}$) and the temporal position of these maximal values ($t_{\max}$). In Table 1 we tabulate the $PA_{\max}$ and $t_{\max}$ average values, as well as the corresponding standard deviations, computed for each one of the nine engineered bacterial strains.

Fig. 3 Representative plots of promoter activity PA vs. t , for each of the nine different engineered bacterial strains. AU stands for arbitrary units. The color code is the same as in Fig. 2.

Table 1 Average and standard deviation values for the promoter maximal activity, $PA_{\max}$, and the time at which this maximal value is achieved, $t_{\max}$, for the nine engineered bacterial strains. AU stands for arbitrary units

Promoter fusion	$PA_{\max} \times 10^4$ (AU)		$t_{\max}$ (h)	
	Average	Std. Dev.	Average	Std. Dev.
<i>hns::gfp</i>	43.52	0.36	-1.08	0.03
<i>ihfB::gfp</i>	8.44	0.15	-0.76	0.09
<i>ihfA::gfp</i>	7.27	0.38	-1.02	0.11
<i>fis::gfp</i>	5.00	0.11	-0.98	0.03
<i>ompR::gfp</i>	3.65	0.18	-0.53	0.28
<i>hdfR::gfp</i>	2.38	0.08	-0.29	0.02
<i>lrhA::gfp</i>	2.33	0.04	-0.09	0.10
<i>fliA::gfp</i>	2.11	0.02	0.17	0.15
<i>qseB::gfp</i>	1.82	0.08	0.28	0.11

Global-promoter activities are higher than those of local regulators

Previous studies have demonstrated the existence a differential abundance of mRNA and protein molecules, correlated with the activity of global and local—or dedicated—regulators.^{29–31} Global regulators regulate the expression of many genes, and their proteins bind to hundreds or thousands of sites in their target genes. Thus a large amount of proteins is required. In contrast, local-regulator proteins are not as demanded because they control the expression of a reduced number of genes. In agreement with the above referred findings, the results in Table 1 show that promoters with the highest activities are those corresponding to global regulators; in particular those known as nucleoid-associated proteins (NAPs). H-NS (Histone-like Nucleoid Structuring protein) has the most active promoter, followed by those expressing the two subunits of IHF (*ihfA* and *ihfB*, Integration Host Factor), and that of FIS (Factor for Inversion Stimulation). At an intermediate level of activity we observe the promoter of *OmpR* (Outer Membrane Protein Regulator). The rest of promoters (*lrhA* - LysR homologue A; *hdfR* - *hns*-dependent *flhDC* regulator; *fliA* - sigma factor for late flagellar genes; and *qseB* - quorum sensing regulator) have the lowest levels of activity, in agreement with their roles as local regulators.

There is a timing of maximal rates of promoter activities among the different regulators

Observe in Table 1 that, except for the *ihfB* promoter, the higher the maximal activity of a given promoter, the sooner the peak of activity is achieved. To better visualize this assertion, we plot in Fig. 4 the average maximal promoter activity ($PA_{\max}$) vs. the average time at which the maximal promoter activity is attained $t_{\max}$, for each of the nine engineered bacterial strains. Notice that, as expected, the $PA_{\max}$ and $t_{\max}$ values are negatively correlated. Another feature worth noticing in Fig. 4 is the fact that, even if we take into consideration the error bars, all the promoters corresponding to global regulators have higher maximal activities and reach their peaks sooner than the local-regulator promoters. All the promoters corresponding to global regulators reach their maximal activity well before the inflection point of the growth curve. That is, while the bacterial population is exponentially growing. On the other hand the maximum activity of local promoters occurs around

Fig. 4 Plot of the average $PA_{\max}$ vs. average $t_{\max}$ values. The error bars denote the standard error of the corresponding data set. AU stands for arbitrary units. The color code is the same as in Fig. 2.

Fig. 5 Linear regression plot of all the $(t_{\max}, PA_{\max}^{-1})$ points, directly calculated from the experimental data. AU stands for arbitrary units.

the inflection point, when the exponential growth phase is about to cease or has already ceased, and the cell-division rate has started to decrease due to overpopulation-originated stress.

In order to corroborate whether the values of $PA_{\max}$ and $t_{\max}$ are somehow correlated, we performed a linear correlation analysis of all $(t_{\max}, PA_{\max}^{-1})$ points, directly calculated from the experimental data. The results are shown in Fig. 5. As seen there, a straight line fits well the experimental data, with a correlation coefficient of 0.93, proving a linear correlation between $t_{\max}$ and $1/PA_{\max}$.

Conclusions

In this work we present experimental evidence in favor of a just-in-time activation of the promoters controlling the expression of the genes that code for the transcription factors in a bacterial hierarchical transcriptional regulatory network: that responsible for the synthesis of the flagellum proteins in *E. coli*. Our results suggest a correlation between the rate and timing of promoter activities and the hierarchical position of the corresponding genes in the network. These results agree with the fact that the abundance of mRNA and protein molecule counts corresponding to global regulators is, in general, higher than those of local regulators.

We found that the global-regulator promoters maximize their activity prior to the inflection point of the growth curve; that is, before the cell division rate starts to decline. This suggests that

global regulators are in charge of the maintenance and division tasks. On the other hand, the activity of the local-regulator promoters reaches its maximal value around the inflection point of the growth curve. Hence, it seems that local regulators are expressed to respond to specific stressing conditions appearing after the exponential growth phase. Since the global and local regulators are hierarchically organized in the transcriptional regulatory network, the above results further indicate that the flow of information starts at the network core, where global regulators respond to endogenous conditions, and ends at the network periphery, with the local regulators responding to specific stimuli.

The notion of a correlation between promoter timing and strength, on the one hand, and the hierarchical position in a biochemical pathway of the enzymes the corresponding genes code for, on the other, was put forward for the enzymes responsible for the synthesis of branched amino acids.¹⁴ In the present paper we prove that the same design principles are valid for a set of promoters organized in a functional, hierarchical, transcriptional, regulatory network. This is observed even when the corresponding coding genes are dispersed in the chromosome, present self-regulation, and are subject to multiple regulatory interactions from other transcription factors.

It might be interesting to test the generality of these results in other sub-networks of the *E. coli* transcriptional regulatory network; and even in other organisms' transcriptional-regulatory networks. If proved, it would indicate that the execution of the regulatory software is designed to have order, timing and direction. It would also be interesting to evaluate the impact that mutations on the regulator's genes have on the network global dynamic response.

Contributions

AM-A designed the study, DAV-R carried out the experiments, JS-M and MS performed the mathematical analysis, AM-A and MS drafted the ms and all the authors read and approved the manuscript.

Abbreviations

TF transcription factor
NAPs nucleoid-associated proteins

Acknowledgements

We thank Alejandro Hernández-Morales and Susana Ruiz for their assistance in the construction of transcriptional fusions, and Denis Thieffry, Julio Collado-Vides and Roberto Olivares-Hernández for critical comments to the manuscript. This work was supported by CONACYT grants 103686 and 102854. DAV-R has a bachelor fellowship through the 102854

grant whereas JS-M has a postdoctoral fellowship from the 103686 grant.

References

- 1 S. Gama-Castro, V. Jimenez-Jacinto, M. Peralta-Gil, A. Santos-Zavaleta and M. I. Penaloza-Spinola, *et al.*, *Nucleic Acids Res.*, 2008, **36**, D120–D124.
- 2 A. Martínez-Antonio, S. C. Janga and D. Thieffry, *J. Mol. Biol.*, 2008, **381**, 238–247.
- 3 S. S. Shen-Orr, R. Milo, S. Mangan and U. Alon, *Nat. Genet.*, 2002, **31**, 64–68.
- 4 O. Resendis-Antonio, J. A. Freyre-González, R. Menchaca-Méndez, R. M. Gutiérrez-Ríos, A. Martínez-Antonio, C. Avila-Sánchez and J. Collado-Vides, *Trends Genet.*, 2005, **21**, 16–20.
- 5 H. W. Ma, J. Buer and A. P. Zeng, *BMC Bioinf.*, 2004, **5**, 199.
- 6 N. Bhardwaj, P. M. Kim and M. B. Gerstein, *Sci. Signaling*, 2010, **3**, 79.
- 7 M. Kittisopikul and G. M. Süel, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13300–13305.
- 8 I. Thiele, N. Jamshidi, R. M. Fleming and B. Ø. Palsson, *PLoS Comput. Biol.*, 2009, **5**, e1000312.
- 9 C. Lagomarsino, P. Jona, B. Bassetti and H. Isambert, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 5516–5520.
- 10 J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins and T. S. Gardner, *PLoS Biol.*, 2007, **5**, 54–66.
- 11 M. Isalan, C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao and E. Raineri, *et al.*, *Nature*, 2008, **452**, 840–845.
- 12 N. Sonnenschein, M. Geertz, G. Muskhelishvili and M. T. Hütt, *BMC Syst. Biol.*, 2011, **5**, 40.
- 13 S. Kaplan, A. Bren, A. Zaslaver, E. Dekel and U. Alon, *Mol. Cells*, 2008, **29**, 786–792.
- 14 A. Zaslaver, A. Mayo, R. Rosenberg, P. Bashkin, H. Sberro and M. Tsalyuk, *et al.*, *Nat. Genet.*, 2004, **5**, 486–491.
- 15 A. Martínez-Antonio, *Network Biol.*, 2011, **1**, 21–33.
- 16 P. Lu, C. Vogel, R. Wang, X. Yao and E. M. Marcotte, *Nat. Biotechnol.*, 2007, **25**, 117–124.
- 17 S. C. Janga, H. Salgado and A. Martínez-Antonio, *Nucleic Acids Res.*, 2009, **37**, 3680–3688.
- 18 J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski and G. Cottarel, *et al.*, *PLoS Biol.*, 2007, **5**, e8.
- 19 A. Martínez-Antonio, S. C. Janga, H. Salgado and J. Collado-Vides, *Trends Microbiol.*, 2006, **1**, 22–27.
- 20 S. Kalir, J. McClure, K. Pabbaraju, C. Southward, M. Ronen and S. Leibler, *et al.*, *Science*, 2001, **292**, 2080–2083.
- 21 S. V. Avery, *Nat. Rev. Microbiol.*, 2006, **4**, 577–587.
- 22 B. Snijder and L. Pelkmans, *Nat. Rev. Mol. Cell Biol.*, 2011, **12**, 119–125.
- 23 K. Kaneko, *J. Biosci.*, 2009, **34**, 529–542.
- 24 M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain, *Science*, 2002, **297**, 1183–1186.
- 25 J. M. Raser and E. K. O'Shea, *Science*, 2010, **309**, 2010–2013.
- 26 E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman and A. van Oudenaarden, *Nat. Genet.*, 2002, **31**, 69–73.
- 27 N. E. Freed, O. K. Silander, B. Stecher, A. Böhm, W. D. Hardt and M. Ackermann, *PLoS Genet.*, 2008, **4**, e1000307.
- 28 S. Tsuru, J. Ichinose, A. Kashiwagi, B. W. Ying, K. Kaneko and T. Yomo, *Phys. Biol.*, 2009, **6**, 036015.
- 29 B. Munsky, B. Trinh and M. Khammash, *Mol. Syst. Biol.*, 2009, **5**, 318.
- 30 T. A. Azam, A. Iwata, A. Nishimura, S. Ueda and A. Ishihama, *J. Bacteriol.*, 1999, **181**, 6361–6370.
- 31 A. Martínez-Antonio and J. Collado-Vides, *Curr. Opin. Microbiol.*, 2003, **6**, 482–489.