Inferring the connectivity of a regulatory network from mRNA quantification in *Synechocystis* PCC6803

Sylvain Lemeille, Amel Latifi[1] and Johannes Geiselmann*

Laboratoire Adaptation et Pathogénie des Micro-Organismes, CNRS UMR5163 Université Joseph Fourier Bâtiment Jean Roget, Faculté Médecine-Pharmacie, Domaine de la Merci, 38700 La Tronche, France and [1]Laboratoire de chimie bactérienne, IBSM-CNRS, 31 chemin Joseph Aiguier, 13402 Marseille cedex 20, France

ABSTRACT

A major task of contemporary biology is to understand and predict the functioning of regulatory networks. We use expression data to deduce the regulation network connecting the sigma factors of *Synechocystis* PCC6803, the most global regulators in bacteria. *Synechocystis* contains one group 1 (SigA) and four group 2 (SigB, SigC, SigD and SigE) sigma factors. From the relative abundance of the *sig* mRNA measured in the wild-type and the four group 2 sigma mutants, we derive a network of the influences of each sigma factor on the transcription of all other sigma factors. Internal or external stimuli acting on only one of the sigma factors will thus indirectly modify the expression of most of the others. From this model, we predict the control points through which the circadian time modulates the expression of the sigma factors. Our results show that the cross regulation between the group 1 and group 2 sigma factors is very important for the adaptation of the bacterium to different environmental and physiological conditions.

INTRODUCTION

All living organisms must finely control the expression of their genes in order to adapt most effectively to the changes of their environment. Genes that encode regulators of many other genes are called global regulators and play the essential role in this process.

Bacterial sigma subunits of RNA polymerase are global regulators of gene expression. They confer specificity to the recognition of promoters by the core enzyme. Two broad families of sigma factors have been identified: the σ^{70} type and the σ^{54} type factors (1). The σ^{54} family regulates a variety of genes, such as those involved in chemotaxis, synthesis of structural components of flagella and enzymes involved in the response to nitrogen starvation (2). The σ^{70} family is subdivided into three groups (1). Group 1 comprises the primary sigma factors that control the transcription of housekeeping genes, and these sigma factors are therefore essential for cell viability. Groups 2 and 3 include the so-called alternative sigma factors that coordinate the regulation of gene expression in bacteria on a global level. They direct the transcription of specific genetic programs that allow bacteria to cope with particular environmental changes and stress conditions. Group 2 sigma factors are similar in sequence to the primary sigma factors and include proteins, such as the stationary-phase-specific sigma factor, RpoS (3). Group 3 sigma factors show less sequence similarity with those of group 1 and include proteins required for the heat-shock response (4) and motility (5).

The inactivation of a gene encoding a particular group 2 or group 3 sigma factor usually produces growth defects or other phenotypes under specific physiological or environmental conditions. For example, an *Escherichia coli rpoS* mutant has a pleitropic phenotype: it shows a loss of viability in stationary phase and a decreased resistance to some stresses, such as the osmotic stress (6). In *Synechocystis* PCC6803, inactivation of the *sigF* gene, encoding a group 3 sigma factor, leads to the loss of motility and pilus formation (7). In *Synechococcus elongatus* PCC7942 mutants of *rpoD* (*rpoD* genes correspond to *sig* genes) show defects in the circadian expression of the *psbAI* gene, encoding the protein D1 of the photosystem II reaction center (8).

The unicellular cyanobacterium *Synechocystis* sp. strain PCC6803 possesses one group 1 sigma factor, *sigA* (slr0653), four group 2 sigma factors, *sigB* to *sigE* (sll0306, sll0184, sll2012 and sll1689) and four group 3 sigma factors (sll0687, sll0856, slr1545 and slr1564) (9). For example, SigE is involved in response to nitrogen stress (10), and the SigB/SigD factors participate in regulating dark/light adaptation (11). SigC contributes to the transcription of *glnB*, a key regulatory gene involved in nitrogen metabolism during stationary phase (12). The synthesis of the other alternative sigma factors is also modulated in response to particular stresses (13,14).

*To whom correspondence should be addressed. Tel: +33 4 76 63 74 96; Fax: +33 4 76 63 74 97; Email: hans.geiselmann@ujf-grenoble.fr

Since the *sig* genes are transcribed by an RNA polymerase holoenzyme, they necessarily regulate each other's transcription. Such crossregulation has been well documented in several bacterial and cyanobacterial systems (15–17). In order to investigate this regulatory network in *Synechocystis*, we have analyzed the transcription of groups 1 and 2 *sig* genes in the *Synechocystis* PCC6803 wild-type strain and in mutants lacking the group 2 sigma genes. Our results demonstrate that these sigma factors regulate each other's transcription. We derive a first-order predictive model of the regulatory network based only on steady-state measurements of mRNA (18). To our knowledge, this constitutes the first quantitative model of an unknown mutual regulation network of five global regulators. This network suggests that sigma factors act in concert in the global transcriptional control of this bacterium.

The bacterium adjusts the relative proportion of each sigma factor in response to environmental changes. We have studied more closely a recurrent environmental change: the daily cycle of day and night imposed by terrestrial rotation, in others words circadian rhythm (19). The regulatory network controlling the circadian cycle consists of three components [reviewed in (20)]: (i) a central oscillator that generates rhythmicity (the clock), (ii) an input pathway that receives and transduces the environmental signal to the oscillator in order to synchronize it with the environment (21,22) and (iii) an output components that connects the clock information to the expression of target genes.

A major characteristic of the circadian regulation in cyanobacteria is that it affects the transcription of the majority of genes. In *S.elongatus* PCC7942, >90% of the genome shows circadian expression and the genes can be grouped into five classes according to the waveforms of their transcriptional profiles (19). These observations suggest that several output pathways coexist and that signal transduction passes through global regulators of gene expression. Such widespread circadian transcriptional activity could be mediated by a circadian expression of the sigma factors. Indeed, four group 2 sigma factors were described as being important for the circadian expression of the *psbAI* gene in *Synechococcus* (8). However, very few components of the output pathways are identified. At present, only one sensory histidine kinase has been shown to interact with the clock proteins (23). We know very little about how the signal generated by the clock is transmitted to the entire genome.

Here, we examine the possibility that the circadian time is transmitted to the organism via the most global transcriptional regulators, the sigma factors. We show that the circadian time cyclically influences the expression of the sigma factors. These changes can be explained by supposing that the circadian pacemaker controls the expression of only two of the sigmas. The mutual connections between the sigmas transmit this perturbation to the other sigmas, and hence to the entire organism.

MATERIALS AND METHODS

Bacterial strains and growth conditions

All cyanobacterial strains were obtained and grown as described previously (24).

For the study of the circadian cycle, the samples were subjected to an entraining period of 12 h incubation in the dark and then returned to LL (continuous light) conditions. All cyanobacterial strains were grown on BG11 medium (25) Petri plates containing 1.5% Difco Bacto Agar. When needed, chloramphenicol was added to a concentration of 10 μg/ml. Growth rates of mutants were compared with a *Synechocystis* strain carrying the same antibiotic resistance cassette inserted into an inessential gene, *ureA*. We call this strain the wild-type for our experiments.

Reverse transcription and real-time quantitative PCR

The reactions and their quantitative analysis were carried out as described previously (24).

Formal description of the regulation network

We assume that transcription by a particular sigma factor is proportional to its concentration and we can therefore describe the expression of the sigma factors by the following linear differential equations:

$$\frac{dsig_i}{dt} = \sum_j m_{ij} \cdot Sig_j - k_i \cdot sig_i$$

$$\frac{dSig_i}{dt} = r_i \cdot \delta_i \cdot sig_i - \lambda_i \cdot Sig_i, \qquad \textbf{1}$$

where sig_i and Sig_i are the mRNA and protein concentrations of the sigma factors, respectively. The m_{ij} are the influences (in biological terms, the promoter strengths) of protein Sig_j on the transcription of gene sig_i. The r_i are the efficiencies with which the mRNA sig_i is translated into the protein Sig_i. The δ_i is equal to 1 for all sig_i except in the sig_i mutant where δ_i is equal to zero. The λ_i are the degradation rates of the proteins, k_i the degradation rates of the mRNAs. At steady state, the net production rates are zero. Because the degradation rates have to be greater than zero, we can solve the system of Equation 1 as follows:

$$sig_i = \frac{1}{k_i} \cdot \sum_j \frac{m_{ij} \cdot r_j}{\lambda_j} \cdot \delta_j \cdot sig_j. \qquad \textbf{2}$$

The coefficients k_i, m_{ij}, λ_j and r_j can be combined into a single constant n_{ij}, representing the overall effect of transcription, translation and degradation.

$$n_{ij} = \frac{m_{ij} \cdot r_j}{k_i \cdot \lambda_j}. \qquad \textbf{3}$$

Equation 2 becomes

$$sig_i = \sum_j n_{ij} \cdot \delta_j \cdot sig_j. \qquad \textbf{4}$$

Equation 4 can be rewritten in more compact form using matrix notation:

$$\vec{x}_l = N \cdot \vec{y}_l, \qquad \textbf{5}$$

where $\vec{x}_l$ is a 5×1 vector of mRNA concentrations of the five *sig* genes in strain l, N is a 5×5 connectivity matrix, composed of elements n_{ij}, and $\vec{y}_l$ is a 5×1 vector proportional to protein concentrations of the five sigma factors in strain l, composed of elements $\delta_j sig_j$. In a mutant strain, one of the elements of this vector will be zero. Inferring the network in this context means to retrieve matrix N.

This can be accomplished by measuring the mRNA concentrations of all five *sig* genes at steady state in the wild-type and in the four group 2 sigma mutants, and solving the system of equations:

$$X = N \cdot Y, \qquad\qquad 6$$

where X is a 5×5 matrix composed of columns $\vec{x}_l$, Y is a 5×5 matrix composed of columns $\vec{y}_l$.

Optimal model

Although we can solve Equation 6 exactly, this solution is not biologically relevant. Consistent with biological knowledge about transcription in cyanobacteria, we assume that each *sig* gene is only transcribed from a small number of promoters. We search for the minimal number of promoters necessary to explain the observed expression profile. Since a *sigA* mutant is not viable, we cannot estimate the influence of SigA on itself and therefore set this effect to zero. We calculate the error μ of a particular network comprising a certain number of promoters as follows:

$$\mu = \sum \left[\log\left(N \cdot Y\right) - \log X \right]^2. \qquad\qquad 7$$

Using the logarithm assures that an x-fold overestimate is penalized the same way as an x-fold underestimate. A perfect model would yield an error of zero. We use standard 'non-linear minimization' to obtain the optimal parameters for each variant of the network, similar to the method described by Tibshirani (26).

In order to obtain the optimal network (the smallest number of promoters sufficient to explain the observations), we proceed by systematically eliminating each influence, i.e. promoter, and we calculate the coefficients that minimize the prediction error for this particular network geometry. We repeat this procedure for networks with all possible numbers of promoters. The optimal network is the one with the least number of promoters, but nevertheless a reasonably small error of the prediction (see below).

Sigma gene expression during the circadian cycle

The expression of the sigma genes in synchronized cells follows the same procedure as the one described above. In order to synchronize the cells, the samples were placed into the dark for 12 h and then returned to constant light conditions. The return to light is called time 0 in our experiment. The expression level of each sigma factor in each strain was then quantified every 3 h for 1.5 periods of the circadian cycle, in others words for 36 h. The concentration of each mRNA species was calculated as before.

We eliminated outliers from the data and corrected the efficiency of the quantitative PCR at each time point by a small correction factor (<20%) in such a way as to optimize the fit of the expression profiles to a sin curve. We used these fitted curves to calculate the influence of the circadian clock on the expression of each sigma factor during the circadian cycle (see below).

Sigma network during the circadian cycle

An external stress may increase or decrease the expression of one or several of the sigma factors. The circadian time can be considered as such an external factor. We model this influence by adding the appropriate term to Equation 4. At each time, t, of the circadian cycle, we can therefore write

$$sig_i^t = \sum_j n_{ij} \cdot \delta_j \cdot sig_j^t + u_i^t. \qquad\qquad 8$$

Equation 8 can be rewritten at each time, t, using matrix notation:

$$X_t = M_t \cdot Z_t, \qquad\qquad 9$$

where X_t is a 5×5 matrix composed of columns $\vec{x}_{lt}$, M_t is a 5×6 matrix whose first five columns contain the parameters n_{ij} obtained for the optimal network. The last column is made up of the vector $\vec{u}_t$ comprising the parameters u_i^t, and Z_t is a 6×5 matrix identical to Y_t, but with an additional sixth line composed of ones.

We then calculated the parameters u_i^t for each time point of the circadian cycle by minimizing the difference between observed and measured values of the concentrations of the sigma mRNAs as before. The other parameters were kept constant at their values obtained for the optimal network. In biological terms, u_i^t represent the varying influence of the circadian clock on the expression of each sigma gene.

Assignment of an index

For a more succinct description, we assign an index between 1 and 25 to each coefficient n_{ij} of the matrix N in the following way.

$$n_{ij} \Leftrightarrow (i - 1) \cdot 5 + j. \qquad\qquad 10$$

RESULTS

Transcription of *sig* genes in σ mutants

We have shown previously (24) that sigma factors regulate each other's transcription. Results are compiled in Table 1. The multiple effects of inactivation of a sigma factor confirm the existence of complex regulatory connections between the different sigma factors: (i) mutation of the *sigB* gene leads to a 3–4-fold decrease of the expression of the *sigA*, *sigC* and *sigE* genes; (ii) in the *sigD* mutant, transcription of the *sigA* and *sigB* genes decrease ~2–3-fold and the expression of the *sigD*

Table 1. Transcription of *sig* genes in sigma mutants

	sigA	*sigB*	*sigC*	*sigD*	*sigE*
wt	8.2	0.9	9.3	4.3	2.2
Δ*sigB*	2.6	0.4	2.5	4.7	0.6
Δ*sigC*	6.4	0.8	11.5	4.5	1.6
Δ*sigD*	4.1	0.5	7.8	12.2	2.1
Δ*sigE*	0.6	0.05	3.5	3.6	1.2

sigABCDE genes were quantified in the wild-type and the four sigma mutants (*sigB*, *sigC*, *sigD* and *sigE*). The gene quantified is indicated at the top of each column. The genotype is indicated to the left of each line. All measurements were carried out under identical culture conditions for the *wt* and mutants strains. Our method can detect transcripts of all *sig* genes in all mutants because the cDNA synthesized during the RT–PCR used primers that anneal upstream of the inactivating chloramphenicol cassette. The absolute values of the RT–PCR were rescaled to coincide with the protein concentrations measured by Imamura *et al.* (13).

gene increases ∼3-fold; (iii) the *sigE* mutation leads to a strong decrease (∼14–18-fold) of the transcription of the *sigA* and *sigB* genes and to a 2–3-fold decrease of the expression of the *sigC* and *sigE* genes; (iv) mutation of the *sigC* gene does not strongly affect the transcription of any of the other four sigma genes.

SigE seems to be a particularly important sigma factor because it controls directly or indirectly the expression of three other *sig* genes. The mutation of the *sigE* gene had the strongest effects among all mutants inactivating sigma genes: its inactivation particularly affected the housekeeping genes *sigA* and *sigB*. The role of the housekeeping sigma factor, SigA, remains less well defined because a deletion mutant is not viable.

Regulation network connecting the sigma genes

Since all sigma genes are transcribed by an RNA polymerase containing one of the sigma factors, we should be able to calculate the expression of the sigma genes as a function of the concentration of all other sigmas. To a first approximation, the effect of a sigma factor is proportional to its concentration. In other words, a promoter will be twice as active when the sigma concentration is increased 2-fold.

If all sigmas would transcribe all others, i.e. five different promoters for each sigma gene, a linear model comprising five equations would exactly predict the observations. From our data, we can calculate the influence of each factor on the transcription of all the others in such a completely connected network. However, it is biologically unreasonable to suppose that all sigma factors are directly regulated by all others.

In order to obtain a biologically reasonable vision of the mutual connections between the *sig* genes, we successively eliminate interactions and adjust parameters, such as to best fit the observation. We systematically continued by canceling further interactions until the discrepancy between prediction and observation became too large. There are ∼33.5 million

possible combinations of networks describing the mutual regulation of the sigma genes. For each combination, we calculated the error between the predicted and observed expression level of each sigma factor. A parameter n_{ij} was set to zero (meaning that such a promoter does not exist) only if its cancellation did not significantly increase the error calculated by Equation 7, and represented in Figure 1a. This Figure shows the error of the 500 best solutions for a network with a specified number of promoters.

Figure 1a clearly shows that the error increases considerably when an additional promoter is removed from the 10-promoters network. We therefore consider the regulation network comprising 10 promoters to be the optimal solution. In addition, of all possible connections of the five sigma factors with 10 promoters, the best solution clearly stands out, i.e. the lowest point of the cluster is separated from the others. Figure 1b shows the errors of all possible 10-promoters networks. While there are very many possible connections with 10 promoters, only very few predict reasonably well the measured quantities of the *sig* mRNAs. Below we further analyze this set of best solutions with 10 promoters.

The optimal network is shown in Figure 2. This model is the simplest reasonable network of transcriptional interactions between the sigma genes of *Synechocystis*. The thickness of the arrows is proportional to the effect of a given mutation on sigma gene expression. Activation of transcription is represented by arrows; repression is represented by a line ending in a cross-bar. The most important effects are (i) a strong influence of SigE on the transcription of *sigA* and *sigD*, (ii) *sigB* seems to be only transcribed by SigA and (iii) the protein SigB influences the transcription of *sigC* and *sigE*. *sigA* is mainly transcribed by SigE, and the protein SigA is entirely responsible for the transcription of *sigB*. The interconnections between SigA, SigB and SigE can explain not only the observed effect on the transcription of *sigB* in a *sigE* null mutant, but also the decrease of *sigC* in this same mutant. The protein SigC does not strongly affect the transcription of others sigma genes,
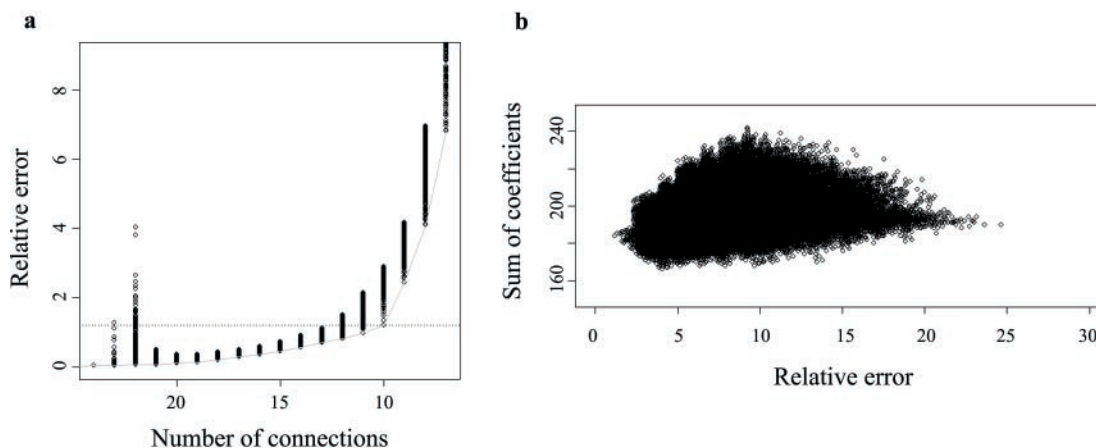

Figure 1. The mutual regulation network of the sigmas. (**a**) Error of the 500 best solutions for each number of promoters. In order to obtain the optimal network connecting the sigma factors, we calculated the quality of all networks with a given number of connections, i.e. promoters. The error of the prediction is shown for the 500 best solutions for each number of promoters. The optimal network is the one providing a good prediction with a minimal number of connections. This is the case for a network with 10 connections; removal of any of these connections considerably increases the error of the prediction. The horizontal line shows the error of the optimal 10-promoter network. (**b**) Exhaustive search for 10-promoters networks. For each network with 10 promoters, the error of the prediction is reported on the *x*-axis. The *y*-axis presents the sum of the indices of the remaining connection as defined by Equation 10. The best solution corresponding to the optimal network is obtained with an error of ∼1.19.
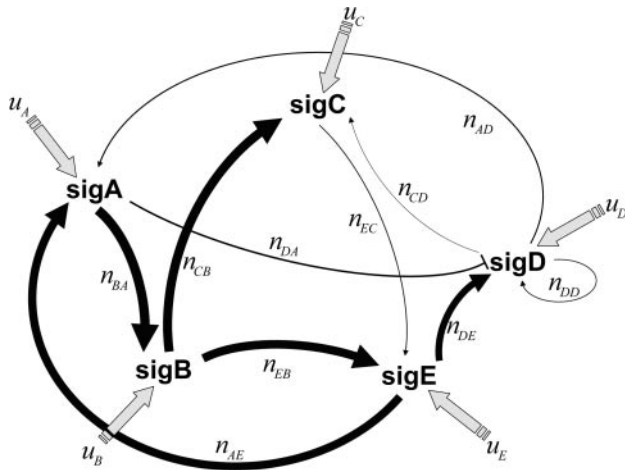
Figure 2. Optimal network of transcriptional interactions between sigma genes in *Synechocystis*. The thickness of the arrows represents the relative importance of a sigma factor for the expression of the target gene. Activation of transcription is represented by arrows; repression is represented by a line ending in a perpendicular bar. The indices of Equation 8 are shown next to the arrows.

a result that may have been intuitively predicted from the measurements obtained in the *sigC* mutant strain. SigA negatively regulates *sigD* transcription. The negative effect of SigA on *sigD* is most likely indirect, e.g. SigA could transcribe a repressor of *sigD*.

The first prediction of our optimal network is that sigma genes are transcribed from multiple promoters. The network of cross regulations of the sigma factors was obtained by the measurement of the respective concentrations of each factor in exponential phase of growth, and the data of Imamura *et al.* (13) concerning the number of promoters of each sigma factor in this same phase of growth agree with our predictions about the number of promoters upstream of each sigma gene. However, it is necessary to keep in mind that our predictions cannot be directly compared with the results of primer extensions because the same promoter can be recognized by several sigma factors and the same sigma factor can recognize several different promoters.

In order to estimate the robustness of our optimal model, we further analyzed the 50 best solutions comprising 10 promoters. Inspection of these best 50 solutions shows one aspect of the robustness of certain network connections. We calculate two parameters: (i) the fraction of the best solutions that retain a certain connection and (ii) the variation of the strength (numerical value of the coefficient) of a particular connection within these best solutions.

Each connection is numbered as described in Materials and Methods (see Equation 10) and its importance is measured as the fraction of the best 50 solutions that contain this particular promoter. As shown in Figure 3a, 5 of the 10 connections composing the minimal network between the sigma factors are extremely robust since they are found in >80% of the group of the 50 best solutions. The five remaining promoters, even though less highly represented, are nevertheless largely more often observed than any of the other connections. This robustness of connections is further reinforced by analyzing the best networks with >10 promoters whose error of prediction is lower than the one of the optimal network (Figure 1a).

As shown in Figure 3b–d; the 10 optimal connections are present in almost all good networks with more promoters, while, at same time, no other connection between the sigma factors is consistently represented within the group of best networks. These 10 connections are thus essential since all of the best networks have them. In other words, removing any one of them produces very much worse predictions. The minimal network of interconnections between the sigma factors is therefore optimal in the sense that only essential connections remain, i.e. those which will considerably increase the error of the prediction when they are removed. This first parameter shows that the geometry of the optimal network is robust.

A second measure of robustness can be derived from analyzing the numerical value of the coefficient associated with each connection between the sigmas. As before, we looked at the variation of the coefficients of each connection in the best networks predicted with 10 connections, but also when a greater number of connections were allowed. Remarkably, the coefficients of the 10 major connections hardly vary even when >10 connections are allowed. This second parameter shows that not only the geometry is important, but also the absolute value of each connection.

These two parameters together attest to the robustness and the quality of the minimal network that we have obtained. Indeed, not only are all these connections essential, but also the intensity of these connections does not vary even when they could have been assisted by other interactions.

The σ-network during the circadian cycle

External or internal stimuli may affect the strength of these 10 promoters, but it is unlikely that a stimulus would modify the network geometry or act at many of the promoters simultaneously. In order to test this prediction, we measured the expression of each one of the sigma factors when the cells were subjected to a regularly changing stimulus: the circadian time.

We propose that the circadian clock exerts a cyclic influence on one or more of the sigma factors. Because all sigmas are connected to each other, this would result in a circadian expression of all sigmas even if only one or two of them were cyclically expressed. Such a model is consistent with the nature of the circadian clock in cyanobacteria. The *kai* genes provide the clock mechanism, which involves many non-transcriptional regulatory mechanisms [reviewed in (20)]. The circadian time then has to be transmitted to the entire organism, which could be accomplished by the sigma factors. We therefore measured the expression of the sigma factors during the circadian cycle in the various contexts (*wt* and *sig* mutants) and tried to explain the observed measurements by a cyclic influence of the circadian clock on the network of the sigma factors. Such a signal of the clock regulating the network of the sigma is already partially characterized in other cyanobacteria and is commonly called output signal.

We therefore modeled the transcription of the sigma factors during the circadian cycle by a changing influence of the circadian signal on each sigma factor. We measured the expression level of each sigma factor in each strain (*wt* and *sig* mutants) for 36 h (Figure 4). All sigma genes are cyclically expressed with a maximum expression level during the subjective day, as are almost all genes in cyanobacteria (19).
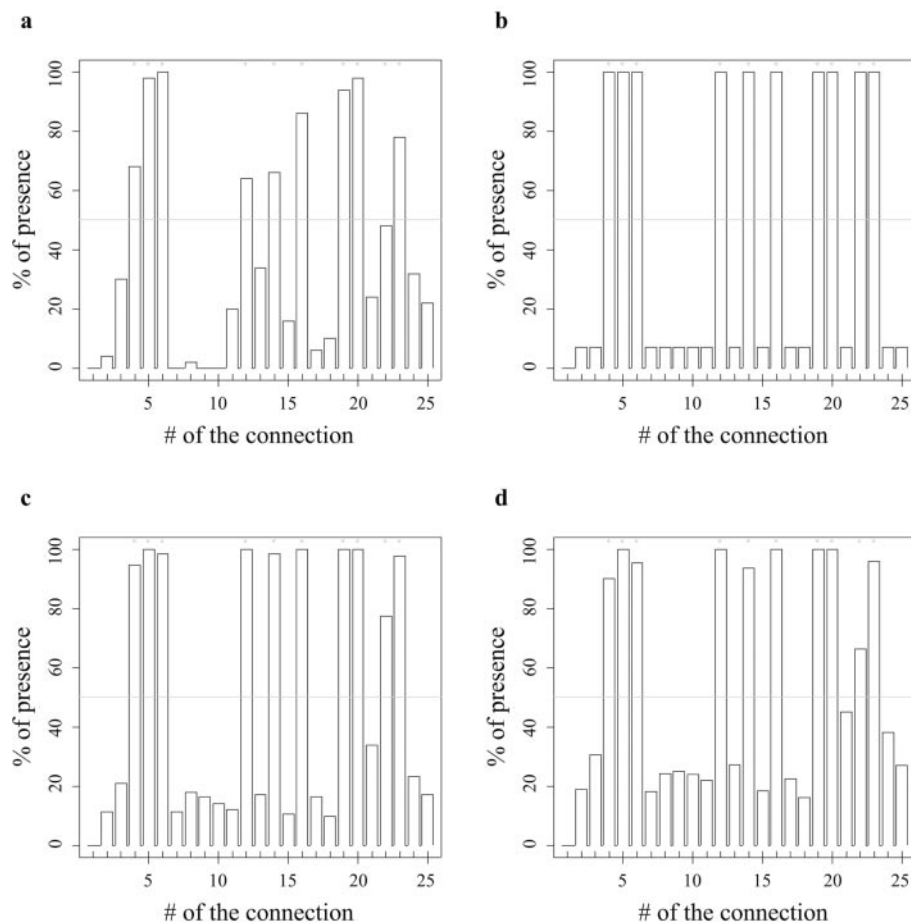
Figure 3. Importance of each connection in the best networks. As a measure of the robustness of the network geometry, we calculated the proportion of network models that contain a particular connection. For the 10-promoters network, we only plot the 50 best solutions because the others have a very much higher error. For networks with more promoters, we include only those that perform better than the optimal 10-promoters network (the number of such networks is in parentheses). The connections present in the optimal network are marked with a star. (**a**) Relative abundance of each connection in the 50 best networks with 10 promoters. (**b**) Relative abundance of each connection in the 11-promoters networks (14) performing better than the optimal 10-promoters network. (**c**) Relative abundance of each connection in the 12-promoters networks (133) performing better than the optimal 10-promoters network. (**d**) Relative abundance of each connection in the 13-promoters networks (802) performing better than the optimal 10-promoters network.

We then compared the predictions of the model with the measured expression level. The same minimization procedure used to obtain the coefficients of the σ-network was now used to determine the impact of circadian time on the expression of each sigma factor at each time point of the cycle (Figure 5). Surprisingly, only two sigma factors are predicted to undergo a circadian regulation: SigB and SigD. This Figure shows the influence of the circadian signal on each sigma factor. The signal modulating the expression of SigB and SigD follows cyclic curves with a period of 24 h. Moreover, the circadian cycle seems to have a greater influence on these sigma factors during the subjective day than during the subjective night. This result is consistent with observations by Imamura *et al*. (11), who demonstrated that *sigD* expression is modulated by the redox state of the cell in the light phase. Moreover, as shown in Figure 4, when the *sigB* gene is interrupted, the circadian expressions of the *sig* genes have reduced amplitude: this observation is consistent with the predicted influence of the clock on the *sigB* gene expression since the clock seems to exert it strongest influence on this sigma factor.

DISCUSSION

Hypotheses used to deduce the regulatory network

The assumptions used to establish the network connections between the *sig* genes are as follows: (i) a proportionality between the expression level of the genes and the protein level of each sigma factor, (ii) the activity of each protein in transcription is linearly related to its concentration and (iii) the network of the sigma factors is a network at steady state.

(i) Imamura *et al*. (13) quantified the protein levels of each sigma factor and of the alpha-subunit of RNA polymerase. By comparing our measurements of mRNA with the concentrations of the σ proteins measured by Imamura *et al*., we conclude that the mRNA concentration of each factor is indeed roughly proportional to the protein concentration of each sigma factor.
(ii) The data of Imamura *et al*. (13) also suggest a very simple mechanism by which varying proportions of sigma factors shift the transcriptional program of the bacterium.
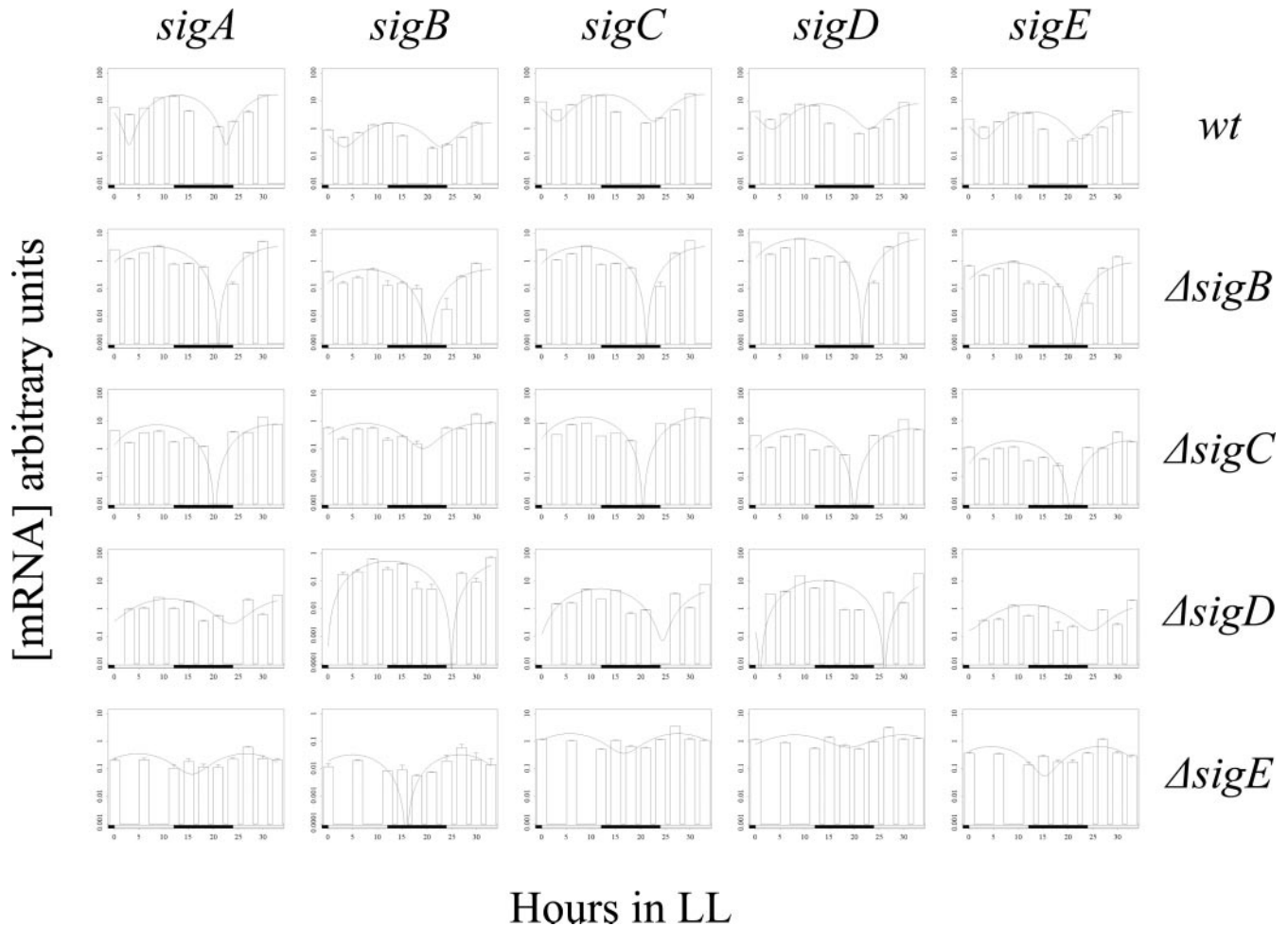
Figure 4. Circadian expression of *sig* genes in the wild-type and the four sigma mutants. Circadian expression of the *sigA*, *sigB*, *sigC*, *sigD* and *sigE* genes in *wt* and mutant strains. The gene that has been inactivated in each reporter strain is indicated at the end of each row. The *x*-axis shows time in hours after the cells were released into LL. The *y*-axis indicates mRNA concentration on a logarithmic scale. Expression levels were measured every 3 h for 1.5 periods of the circadian cycle and the curve is a sin fit of the circadian expression. The patterns of the sin curves are unusual since we use a logarithmic scale. The black rectangle below the graph indicates the subjective night; the white rectangle corresponds to the subjective day.

According to their quantitative western blots, *Synechocystis* always produces more core RNA polymerase than the sum of all five sigma subunits. The measurement is based on the quantification of RpoA, and the assumption that all core subunits are produced in equimolar amounts. According to this simple model, there would be little competition between the sigma subunits for association with the core enzyme (always in excess), and the activity of promoters recognized by a particular sigma factor would be roughly proportional to the concentration of the respective sigma factor.

(iii) Finally, the establishment of the network was carried out by measuring the level of expression of each sigma factor in non-synchronized cells during balanced growth. By definition, we therefore measure the steady state concentration of the cellular components. Even for the measurements carried out during the circadian cycle we can assume the existence of a steady state. Transcriptional adaptations are generally rapid, in the order of minutes, whereas the circadian cycle imposes modifications in the scale of several hours. Because of these different time scales, we can assume that the fast process, transcription, is always at steady state relative to the slow process, the gradual change of a circadian signal.

The basic assumptions necessary for establishing this network between the sigma factors thus appear justified.

Obtaining the network

The optimal network of connections between the sigma factors was obtained by searching for the minimal number of connections (promoters) between the sigmas still capable of predicting the observed behavior. Our optimal solution comprises only 10 connections. None of these 10 connections can be removed without considerably deteriorating the predictions of this model. Indeed, the curve showing the error of the predictions presents two quite distinct parts; with a clear cut at 10 promoters. The minimal network comprising 10 connections is at the border between these two parts and any further reduction of connections drastically increases the error of the prediction.
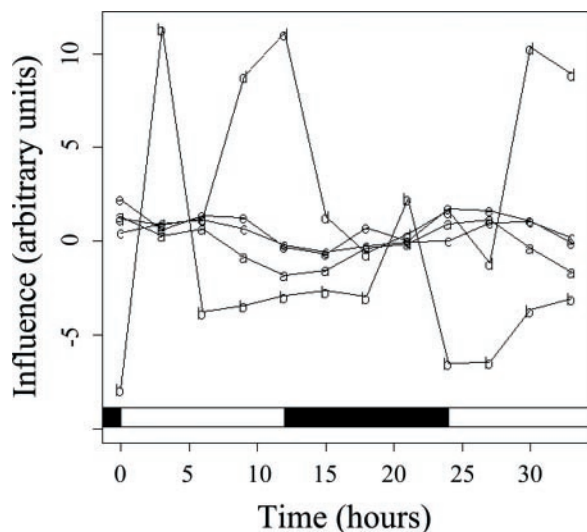
Figure 5. Impact of the circadian clock on the sigma network during the circadian cycle. The external influence of the circadian clock on each sigma promoter is plotted for 1.5 periods of the circadian cycle. The value represents the fraction of this circadian influence with respect to all influences on this gene. The curve of each sigma promoter is marked by the letter of the corresponding *sig* gene. The black rectangle below the graph indicates the subjective night; the white rectangle corresponds to the subjective day.

The validity of a minimal network with 10 connections is reinforced by the fact that the connections present in this minimal network are also found in most of the best solutions, and that their intensities are conserved even when additional promoters are added. Furthermore, the predicted network is also robust with respect to experimental errors. We obtain the same connections and only slight (10%) variations of the parameters when we derive the network from the average of the circadian expression data instead of using the data from the non-synchronized cultures.

The basic method used for obtaining a network from steady-state expression data has already been validated by Gardner *et al.* (18). They have perturbed their system by the overexpression of certain components and measured the resulting change in the expression of the other components. Since their perturbations are relatively small, they can justify the linearity assumption. In our case, we deal with very particular transcription factors for which the linearity assumption should hold for a large range of concentrations, from zero up to the normal physiological concentration of the sigma factors.

CONCLUSION

Synechocystis possesses five genes encoding four group 2 sigma factors and one principal sigma factor. The transcriptional program of this bacterium is largely determined by the activity of these multiple sigma factors. All sigma factors are expressed in all environmental conditions. Similar results had been obtained previously in *S.elongatus* PCC7942, where all sigma factors were found to be active under many growth conditions (16). Imamura *et al.* (13) have measured the concentration of all five sigma factors during normal growth of
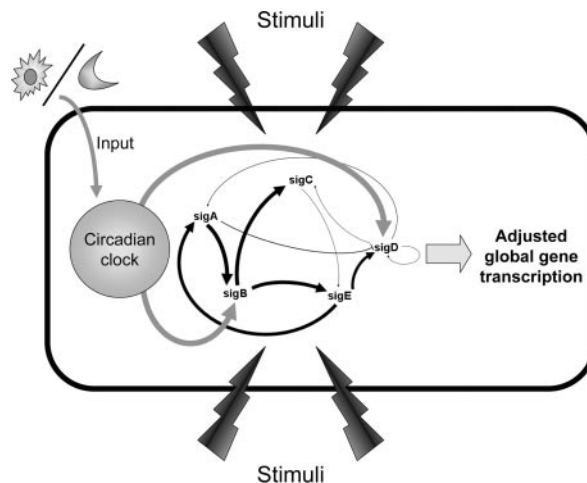

Figure 6. Points of control of gene transcription in *Synechocystis*. This model shows the role of the σ-network in the global regulation of gene transcription in the transmission of the circadian signal and others environmental stimuli (see text).

Synechocystis PCC6803. They found amounts of proteins between 1 and 10 fmol/μg of total protein. These data suggest that all five factors are important for the cellular physiology of *Synechocystis* under standard conditions.

By quantifying the sigma transcripts in different sigma mutants, we have shown that the transcription of the *sig* genes is controlled by a network of mutual connections between the sigmas. Previous studies in related organisms have also shown a mutual transcriptional regulation of sigma factors: in *Synechococcus* PCC7942 the *rpoD1* gene is transcribed by RpoD3 and RpoD4 factors (16), and SigC factor has a negative effect on SigB expression (11). In *Borrelia burgdorferi*, RpoN regulates the expression of *rpoS* (17). However, for all these systems, we crucially lack an understanding of the complete mutual interdependence of all sigma factors.

To our knowledge, this is the first study that explores biologically and mathematically the quantitative relationships between all members of one family of sigma factors in eubacteria. The robustness of this network permits us to consider a model of global gene regulation in *Synechocystis* (Figure 6). We have already explored how external stimuli could impinge on this network. In the case of circadian time, we show that modulating the expression of only two sigma factors can explain the circadian regulation of all five factors. Other stimuli representing environmental changes, such as the osmotic stress, starvation, etc., could act via the same mechanism.

REFERENCES

1. Lonetto,M., Gribskov,M. and Gross,C.A. (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *J. Bacteriol.*, **174**, 3843–3849.
2. Kustu,S., Santero,E., Keener,J., Popham,D. and Weiss,D. (1989) Expression of sigma 54 (*ntrA*)-dependent genes is probably united by a common mechanism. *Microbiol. Rev.*, **53**, 367–376.
3. Mulvey,M.R. and Loewen,P.C. (1989) Nucleotide sequence of *katF* of *Escherichia coli* suggests *KatF* protein is a novel sigma transcription factor. *Nucleic Acids Res.*, **17**, 9979–9991.
4. Hecker,M., Schumann,W. and Volker,U. (1996) Heat-shock and general stress response in *Bacillus subtilis*. *Mol. Microbiol.*, **19**, 417–428.
5. Arora,S.K., Ritchings,B.W., Almira,E.C., Lory,S. and Ramphal,R. (1997) A transcriptional activator, FleQ, regulates mucin adhesion and flagellar gene expression in *Pseudomonas aeruginosa* in a cascade manner. *J. Bacteriol.*, **179**, 5574–5581.
6. Hengge-Aronis,R. (2002) Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.*, **66**, 373–395.
7. Bhaya,D., Watanabe,N., Ogawa,T. and Grossman,A.R. (1999) The role of an alternative sigma factor in motility and pilus formation in the cyanobacterium *Synechocystis* sp. strain PCC6803. *Proc. Natl Acad. Sci. USA*, **96**, 3188–3193.
8. Nair,U., Ditty,J.L., Min,H. and Golden,S.S. (2002) Roles for sigma factors in global circadian regulation of the cyanobacterial genome. *J. Bacteriol.*, **184**, 3530–3538.
9. Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S. *et al.* (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.*, **3**, 109–136.
10. Muro-Pastor,A.M., Herrero,A. and Flores,E. (2001) Nitrogen-regulated group 2 sigma factor from *Synechocystis* sp. strain PCC 6803 involved in survival under nitrogen stress. *J. Bacteriol.*, **183**, 1090–1095.
11. Imamura,S., Asayama,M., Takahashi,H., Tanaka,K., Takahashi,H. and Shirai,M. (2003) Antagonistic dark/light-induced SigB/SigD, group 2 sigma factors, expression through redox potential and their roles in cyanobacteria. *FEBS Lett.*, **554**, 357–362.
12. Asayama,M., Imamura,S., Yoshihara,S., Miyazaki,A., Yoshida,N., Sazuka,T., Kaneko,T., Ohara,O., Tabata,S., Osanai,T. *et al.* (2004) SigC, the group 2 sigma factor of RNA polymerase, contributes to the late-stage gene expression and nitrogen promoter recognition in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Biosci. Biotechnol. Biochem.*, **68**, 477–487.
13. Imamura,S., Yoshihara,S., Nakano,S., Shiozaki,N., Yamada,A., Tanaka,K., Takahashi,H., Asayama,M. and Shirai,M. (2003) Purification, characterization, and gene expression of all sigma factors of RNA polymerase in a cyanobacterium. *J. Mol. Biol.*, **325**, 857–872.
14. Tuominen,I., Tyystjarvi,E. and Tyystjarvi,T. (2003) Expression of primary sigma factor (PSF) and PSF-like sigma factors in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.*, **185**, 1116–1119.
15. Caslake,L.F., Gruber,T.M. and Bryant,D.A. (1997) Expression of two alternative sigma factors of *Synechococcus* sp. strain PCC 7002 is modulated by carbon and nitrogen stress. *Microbiology*, **143**, 3807–3818.
16. Goto-Seki,A., Shirokane,M., Masuda,S., Tanaka,K. and Takahashi,H. (1999) Specificity crosstalk among group 1 and group 2 sigma factors in the cyanobacterium *Synechococcus* sp. PCC7942: *In vitro* specificity and a phylogenetic analysis. *Mol. Microbiol.*, **34**, 473–484.
17. Hubner,A., Yang,X., Nolen,D.M., Popova,T.G., Cabello,F.C. and Norgard,M.V. (2001) Expression of *Borrelia burgdorferi* OspC and DbpA is controlled by a RpoN-RpoS regulatory pathway. *Proc. Natl Acad. Sci. USA*, **98**, 12724–12729.
18. Gardner,T.S., di Bernardo,D., Lorenz,D. and Collins,J.J. (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, **301**, 102–105.
19. Liu,Y., Tsinoremas,N.F., Johnson,C.H., Lebedeva,N.V., Golden,S.S., Ishiura,M. and Kondo,T. (1995) Circadian orchestration of gene expression in cyanobacteria. *Genes Dev.*, **9**, 1469–1478.
20. Kondo,T. and Ishiura,M. (2000) The circadian clock of cyanobacteria. *Bioessays*, **22**, 10–15.
21. Mutsuda,M., Michel,K.P., Zhang,X., Montgomery,B.L. and Golden,S.S. (2003) Biochemical properties of CikA, an unusual phytochrome-like histidine protein kinase that resets the circadian clock in *Synechococcus elongatus* PCC 7942. *J. Biol. Chem.*, **278**, 19102–19110.
22. Schmitz,O., Katayama,M., Williams,S.B., Kondo,T. and Golden,S.S. (2000) CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. *Science*, **289**, 765–768.
23. Iwasaki,H., Williams,S.B., Kitayama,Y., Ishiura,M., Golden,S.S. and Kondo,T. (2000) A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria. *Cell*, **101**, 223–233.
24. Lemeille,S., Geiselmann,J. and Latifi,A. (2005) Crosstalk regulation among group 2-Sigma factors in *Synechocystis* PCC6803. *BMC Microbiol.*, **5**, 18.
25. Rippka,R., Deruelles,J., Waterbury,J.B., Herman,M. and Stanier,R.Y. (1979) Genetics assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.*, **111**, 1–61.
26. Tibshirani,R. (1996) Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. B.*, **58**, 267–288.