

RESEARCH ARTICLE

Inference of Quantitative Models of Bacterial Promoters from Time-Series Reporter Gene Data

Diana Stefan^{1,2}, Corinne Pinel^{1,2}, Stéphane Pinhal^{1,2}, Eugenio Cinquemani¹, Johannes Geiselmann^{1,2}, Hidde de Jong^{1*}

1 INRIA Grenoble – Rhône-Alpes, Grenoble, France, **2** Laboratoire Interdisciplinaire de Physique (LIPhy, CNRS UMR 5588), Université Joseph Fourier, Grenoble, France

* Hidde.de-Jong@inria.fr



 OPEN ACCESS

Citation: Stefan D, Pinel C, Pinhal S, Cinquemani E, Geiselmann J, de Jong H (2015) Inference of Quantitative Models of Bacterial Promoters from Time-Series Reporter Gene Data. *PLoS Comput Biol* 11(1): e1004028. doi:10.1371/journal.pcbi.1004028

Editor: Christos A. Ouzounis, Hellas, GREECE

Received: August 11, 2014

Accepted: November 5, 2014

Published: January 15, 2015

Copyright: © 2015 Stefan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supplementary Information files.

Funding: Rhône-Alpes region (cluster ISLE, PhD grant); DS Investissements d'Avenir Bio-informatique programme, project RESET (ANR-11-BINF-0005, <https://project.inria.fr/reset/>); DS, EC, JG, HDJ INRIA/INSERM project ColAge (<http://colage.saclay.inria.fr/>); JG, HDJ Agence Nationale de la Recherche, project GeMCo (ANR-2010-BLAN-0201-02, <http://www-sop.inria.fr/members/Madalena.Chaves/ANR-GeMCo/main.html>); EC, JG, HDJ The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The inference of regulatory interactions and quantitative models of gene regulation from time-series transcriptomics data has been extensively studied and applied to a range of problems in drug discovery, cancer research, and biotechnology. The application of existing methods is commonly based on implicit assumptions on the biological processes under study. First, the measurements of mRNA abundance obtained in transcriptomics experiments are taken to be representative of protein concentrations. Second, the observed changes in gene expression are assumed to be solely due to transcription factors and other specific regulators, while changes in the activity of the gene expression machinery and other global physiological effects are neglected. While convenient in practice, these assumptions are often not valid and bias the reverse engineering process. Here we systematically investigate, using a combination of models and experiments, the importance of this bias and possible corrections. We measure in real time and in vivo the activity of genes involved in the FliA-FlgM module of the *E. coli* motility network. From these data, we estimate protein concentrations and global physiological effects by means of kinetic models of gene expression. Our results indicate that correcting for the bias of commonly-made assumptions improves the quality of the models inferred from the data. Moreover, we show by simulation that these improvements are expected to be even stronger for systems in which protein concentrations have longer half-lives and the activity of the gene expression machinery varies more strongly across conditions than in the FliA-FlgM module. The approach proposed in this study is broadly applicable when using time-series transcriptome data to learn about the structure and dynamics of regulatory networks. In the case of the FliA-FlgM module, our results demonstrate the importance of global physiological effects and the active regulation of FliA and FlgM half-lives for the dynamics of FliA-dependent promoters.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

A wide variety of methods for the reverse engineering of regulatory networks and the identification of quantitative regulation functions are available. We investigate some common assumptions that are made in the application of these methods to time-series transcriptomics data, in the context of a central module in the motility network of *E. coli*. We show that these assumptions, which hypothesize that mRNA concentrations are good proxies for protein concentrations and that the gene expression machinery is equally active across different physiological conditions, are often not valid and may lead to biased inference results. We also show how models of gene expression can be used in combination with suitable experimental controls to correct for this bias and improve the inference process. The contribution of our work is thus not the addition of another method to the rich store of available reverse engineering algorithms, but lies in the critical examination of the information provided by the experimental data and new ways to exploit this information in the algorithms. The proposed approach is relevant for a wide range of applications using time-series transcriptomics data. For the motility system under study, it has underlined the importance of global physiological effects, the active degradation of the transcription factor FliA as well as the secretion of the anti-sigma factor FlgM for the network dynamics.

This is a *PLOS Computational Biology Methods* paper.

Introduction

DNA microarrays, RNA sequencing, and other high-throughput technologies yield huge amounts of data on the state of the transcriptional program in bacterial cells in different growth conditions and genetic backgrounds, at different time-points in an experiment. The information on the (relative) RNA abundances thus obtained, representative of the activity of the genes, have fueled the development of methods for inferring regulatory interactions among genes. In essence, these methods try to explain the observed variation in the activity of one gene in terms of the variation in the activity of other genes. A large number of inference methods have been proposed in the literature and have been successful in a variety of applications, although a number of difficult problems remain (see [1–7] for reviews).

A major problem with the use of transcriptome data for the inference of regulatory interactions is that often the active regulator is not mRNA but protein. Although protein and mRNA concentrations are moderately correlated at steady state [8, 9], this is generally not the case when the two are considered dynamically over time. Due to the fact that proteins and mRNAs have different half-lives, their concentrations evolve on different time-scales. For instance, mRNA half-lives are typically on the order of a few minutes in bacteria [10], whereas most proteins are quite stable [11, 12]. The effect of rapid responses in gene expression, within a single generation, may thus give rise to proteins persisting over several generations, endowing the cell with a memory of past events [9, 13]. As a consequence, inference of regulatory interactions from time-series transcriptome data alone may potentially lead to spurious results. Although quantitative proteomics techniques have much advanced recently [14, 15], it is not yet possible to directly measure protein concentrations *in vivo* and in real time.

A second problem derives from the fact that the dynamics of gene expression are not only controlled by transcription factors, small regulatory RNAs, and other specific regulators, but also by global physiological effects influencing the rates of transcription and translation of all

genes [16–19]. Large-scale differences in gene expression over time or across conditions may therefore not just derive from transcriptional regulatory interactions, but also reflect global changes in cellular physiology, notably the concentrations of (free) RNA polymerase and ribosome, gene copy number, and the size of amino acid and nucleotide pools. Ignoring such changes in the activity of the gene expression machinery, for example in experiments with important variations of the growth rate, may lead to the inference of spurious regulatory interactions [20, 21]. Unfortunately, concentrations of (active) RNA polymerase and ribosome, as well as many other global physiological parameters, are difficult to quantify in a direct way.

These problems for reverse engineering come from two basic, usually implicit assumptions on the biological processes under study: (i) mRNA abundance is a good proxy for protein concentrations and (ii) the gene expression machinery is equally active across different physiological conditions. Although the fact that these assumptions are often not valid has been broadly recognized, very little has been done to study the resulting bias in a systematic way. The aim of this paper is to propose a combined experimental and computational approach to show how these assumptions affect the inference of quantitative models of bacterial promoters from time-series gene expression data and to propose theoretically sound and practically useful procedures to correct for this bias and improve the inference process.

We will notably focus on the case of gene expression measurements obtained by means of fluorescent reporters [22]. These technologies, which have become widespread in recent years, allow the activity of genes to be monitored *in vivo* and in real time [23, 24]. Exploiting these data makes it possible to quantify the difference between mRNA and protein concentrations as well as global physiological effects. In short, if the half-lives of the proteins are available, the models used for deriving the activities of genes from fluorescence data can be integrated to yield estimates of protein concentrations [25]. The global physiological state of the cell can be estimated from the activity of a constitutively expressed gene [17, 18], that is, a gene whose expression is not controlled by any particular transcription factor, but only depends on the activity of the transcriptional and translational machinery [26]. To which extent does the integration of the above information into the inference procedure improve the identification results, both structurally and quantitatively?

In order to answer this question, we applied our methodology to a central module in the regulatory network controlling the synthesis of flagella and the chemotaxis sensing system in *Escherichia coli* [27–29]. This module comprises the FliA and FlgM transcription factors and their targets. FliA or σ^{28} is a sigma factor which directs RNA polymerase to operons coding for the flagellar filament and the chemotaxis sensing system controlling the flagellar motor. The effect of FliA is counteracted by the anti-sigma factor FlgM. As a typical example of a FliA-dependent gene we study *tar*. This gene encodes the aspartate chemoreceptor protein Tar, which responds to a decrease of the aspartate concentration in the medium. Tar stimulates the phosphorylation of downstream response regulators binding the flagellar motor component [30, 31]. The FliA-FlgM module forms a check-point in the temporally-organized expression cascade. It is particularly well-suited as a gold standard for our purpose, since the interactions in this network have been well-studied and protein stability has been found to play an important role in its functioning.

We experimentally excited the FliA-FlgM module in a variety of wild-type and mutant conditions, in different growth media, and measured the transcriptional response of the genes. These data were used to systematically test the information required for the reliable inference of the regulatory interactions and quantitatively predictive models of gene regulation. In a first step, we found that the use of *fliA* and *flgM* promoter activities, instead of their protein concentrations, did not allow the regulatory interactions to be recovered. Moreover, a quantitative model identified from the data fails to account for the observed dynamics of the *tar* promoter in most conditions considered here. The introduction, in a second step, of global regulatory effects, measured by means of a reporter gene driven by a constitutive promoter, results in the

expected structure of regulatory interactions. The fit of the quantitative model to the data, however, is only marginally improved. We therefore estimated in a third step the concentrations of FliA and FlgM from the observed promoter activities and physiologically plausible half-lives of the proteins. The model quantitatively reproduces the observed activity of *tar* across the different conditions much better and the estimated parameter values agree with the expected regulatory role of FliA. Additional simulation studies, in which we systematically varied the half-lives of the proteins and the importance of global physiological effects, show that these factors may be even more important in other regulatory networks, notably when involving transcription factors with half-lives longer than the (exceptionally short) half-lives of FliA and FlgM.

We conclude that for the reliable reconstruction of transcriptional regulatory networks in microorganisms it is important to monitor not only specific transcription factors, but also global effects imposed by the cellular physiology, and to take into account both short-term transcriptional responses as well as their long-term effects on protein abundance. We have presented and validated a practical approach to integrate information on protein concentrations and global regulatory effects into the network identification process. Since the proposed strategy does not depend on any specific network inference method, and can in principle be combined with data obtained from experimental techniques other than fluorescent reporter genes, the approach is applicable to a large variety of network inference problems in both prokaryotic and eukaryotic systems.

Results

Monitoring the transcriptional response of the FliA-FlgM module

The more than 60 genes responsible for motility in bacteria are structured in a transcriptional hierarchy of three operon classes that has been mapped in detail for *Escherichia coli* and *Salmonella enterica* [27–29, 32]. The single class 1 operon *flhDC* encodes the proteins FlhD and FlhC, which form a heteromultimeric complex activating σ^{70} -dependent transcription of the class 2 operons. The latter encode the proteins making up the flagellar motor structure as well as a major regulator of the class 3 operons, the sigma factor FliA (σ^{28}). When bound to core RNA polymerase, FliA directs the transcription of the class 3 operons [33] that code for the proteins forming the filament structure of the flagellum and the chemotaxis sensing system. The aspartate chemoreceptor Tar is an example of such a class 3 protein. The action of FliA is counteracted by the anti-sigma factor FlgM, which binds to FliA and thus prevents its association with RNA polymerase. FlgM is encoded by the gene *flgM*, which is transcribed from both a class 2 promoter and a class 3 promoter. FlgM can be excreted from the cell through the center of the basal-body structure (Fig. 1).

The transcriptional hierarchy underlies a temporally-arranged order of events during the assembly of the flagella and the chemotactic sensing system [27–29, 32]. On the highest level of the hierarchy, the transcription of the flagellar master regulator responds to a variety of signals [34, 35]. For instance, the expression of the *flhDC* operon is repressed when the bacteria are grown on minimal medium with glucose [36]. When glucose is depleted from the environment, however, the signalling molecule cyclic AMP (cAMP) accumulates in the cell, inducing *flhDC* transcription through the intermediary of the cAMP receptor protein Crp [37]. In the presence of FlhDC, the class 2 operons, and thus the genes encoding the hook basal-body (HBB) structure as well as FliA and FlgM, are actively transcribed. FlgM sequesters FliA, and prevents it from transcribing the class 3 operons [38]. When the HBB structures have been completed, however, FlgM is secreted from the cell, releasing FliA and relieving the repression of the class 3 operons [38]. The FliA-FlgM interactions thus form a check-point in flagella formation, ensuring that the filament proteins are produced only when the basal body and the hook, to which the flagellar filaments are attached, are in place.

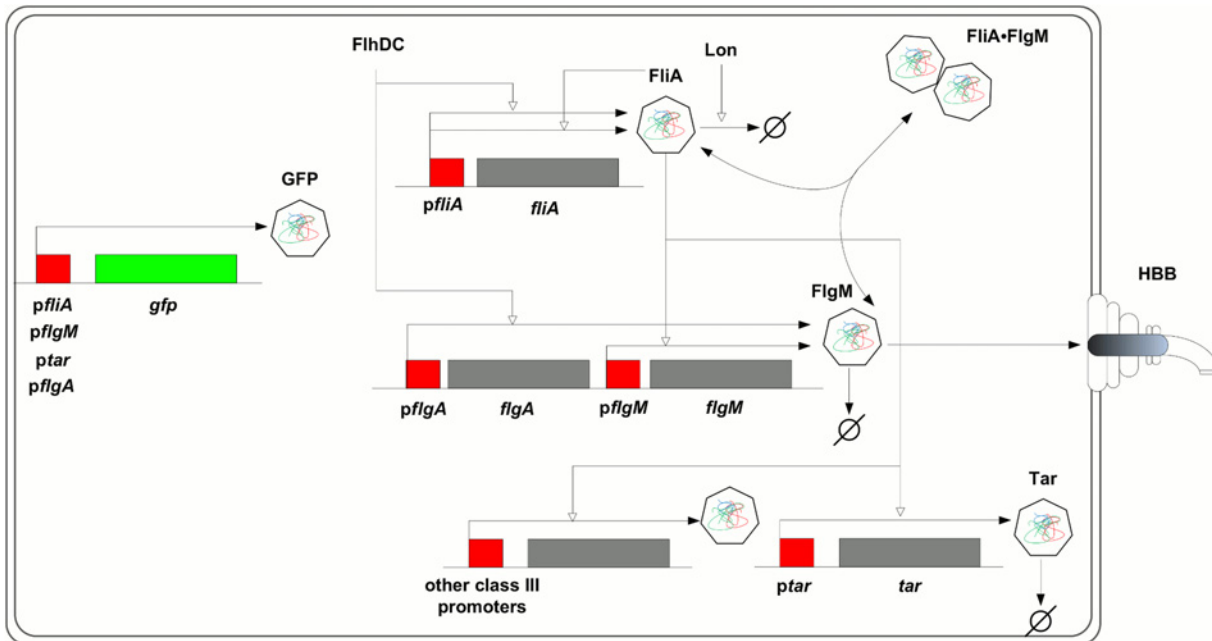


Figure 1. FliA-FlgM module. The regulatory circuit composed of the flagellar-specific transcription factor FliA, a sigma factor also known as σ^{28} , and the anti-sigma factor FlgM forms a check-point in the transcriptional hierarchy of the motility genes in *E. coli*. While *fliA* is transcribed from a single class 2 promoter (*pflIA*), *flgM* is transcribed from both a class 2 and a class 3 promoter (*pflgA* and *pflgM*, respectively). FliA binds to RNA polymerase core enzyme and directs transcription from a total of five class 3 promoters [33], including *ptar* and *pflgM*. When bound to FlgM, FliA cannot activate transcription. When the hook basal-body (HBB) structure is in place, however, FlgM is exported from the cell, thus releasing FliA from the inactive complex. FliA is subject to proteolysis by Lon, but FlgM-binding protects FliA from degradation. The *fliA* promoter is auto-regulated by FliA and by a number of other regulators, most importantly the motility master regulator FliDC. The expression of FliDC itself is under the control of a variety of regulatory factors, including RpoS, CpxR, and CsgD. The activity of the genes in the figure is measured by fusion of their promoters to a *gfp* reporter gene on a low-copy plasmid. Genes are shown in grey or green and their promoter regions in red. Regulatory interactions are represented by open arrows, association and dissociation of FliA and FlgM as well as degradation and export by filled arrows. The figure does not explicitly show that *fliA*, *flgM*, and *tar* are included in larger transcriptional units, the *fliAZY*, *flgAMN*, *flgMN* and *tar-tap-cheRBYZ* operons [33].

doi:10.1371/journal.pcbi.1004028.g001

In order to investigate the regulation of the genes involved in this check-point, we measured the time-varying transcription of *fliA*, *flgM*, and *tar* (as an example of a class 3 gene) in *E. coli*. This was accomplished by means of fluorescent reporter systems, consisting of transcriptional fusions of a *gfp* reporter gene to the promoters of the target genes, carried on a low-copy plasmid. The strains transformed with the reporter plasmids were grown in 96-well microplates, following a previously-established protocol [17, 39, 40]. After an overnight preculture, the bacteria were diluted into fresh medium in the microplate and the absorbance of the cultures and the emitted fluorescence were monitored at 37°C in a thermostated microplate reader for 7 to 16 h, until growth arrest occurred. These kinetic experiments were carried out in different growth media (minimal M9 medium with glucose, rich LB medium) and in different genetic backgrounds (wild-type and deletion mutants of the global transcription regulators RpoS, CsgD, and CpxR) [40]. The timing and the strength of the induction of the hierarchy of motility genes varies among conditions, leading to a different time-varying excitation of the FliA-FlgM module.

While *fliA* and *tar* have a single promoter, this is not the case for *flgM*, which is transcribed from both a class 2 and a class 3 promoter, as discussed above. The fluorescence signal from the class 2 promoter, however, was found to be almost indistinguishable from background levels in all conditions (Text S7), consistent with the observation that most FlgM in the cell derives from the FliA-dependent promoter [29, 41]. In the analysis that follows, we therefore neglected *flgM* transcription from the class 2 promoter.

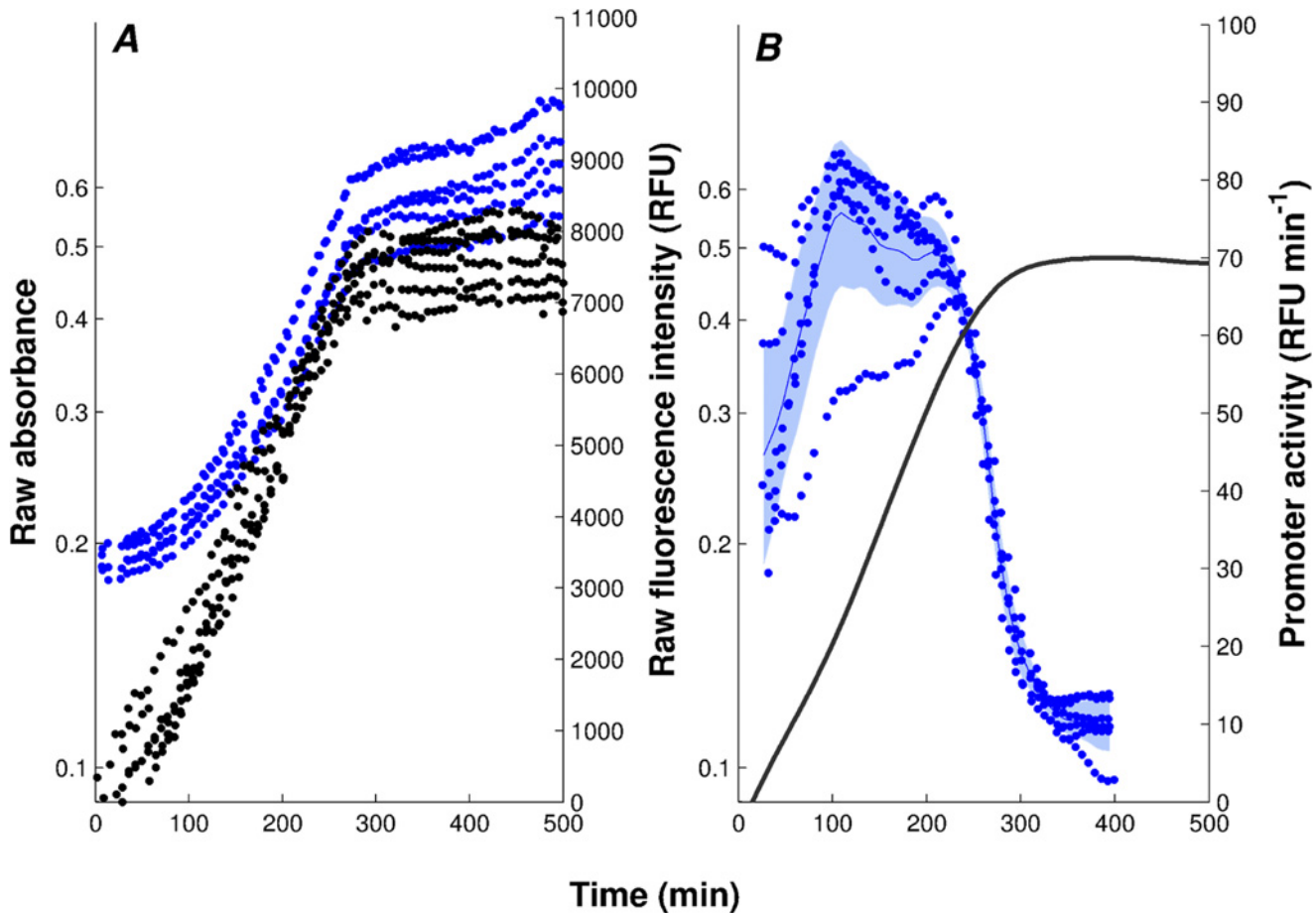


Figure 2. Primary data and promoter activities. *A*: Absorbance (•, black) and fluorescence (•, blue) data, corrected for background intensities, obtained with the $\Delta cpxR$ strain transformed with the *ptar-gfp* reporter plasmid and grown in M9 with glucose. *B*: Activity of the *tar* promoter, computed from the primary data as described in the [Methods and materials](#) and in [Text S3](#). The solid black line corresponds to the mean of 6 replicate absorbance measurements and the shaded blue region to the mean of the promoter activities \pm twice the standard error of the mean.

doi:10.1371/journal.pcbi.1004028.g002

As illustrated in [Fig. 2](#), and explained in more detail in the [Methods and materials](#) and [Text S3](#), the primary absorbance and fluorescence signals can be transformed into promoter activities using kinetic models of gene expression. More precisely, the reporter gene data can be used to deduce protein synthesis rates [[23](#), [25](#)]. Under certain conditions, the latter are proportional to mRNA concentrations and promoter activities, and thus reflect the transcriptional activity of the gene ([Text S2](#)). Following established terminology, we will refer to the measured protein synthesis rates as promoter activities, or more generally, activities of genes.

In each of the experimental conditions, we have acquired 5 to 8 replicate measurements, which makes it possible to estimate the uncertainty in the derived promoter activities. [Fig. 3](#) shows the results for the five conditions considered here: (i) $\Delta rpoS$ strain grown in M9 ($\Delta rpoS$ -M9), (ii) $\Delta cpxR$ strain grown in M9 ($\Delta cpxR$ -M9), (iii) $\Delta csgD$ strain grown in M9 ($\Delta csgD$ -M9), (iv) $\Delta csgD$ strain grown in LB ($\Delta csgD$ -LB), and (v) wild-type strain grown in LB (WT-LB). As expected [[36](#)], the fluorescence signals in the wild-type strain grown in minimal M9 medium with glucose were mostly indistinguishable from the background fluorescence and therefore this condition was not further considered. In one condition (WT-LB), the activities measured by means of reporter genes were validated using RT-qPCR ([Text S6](#)).

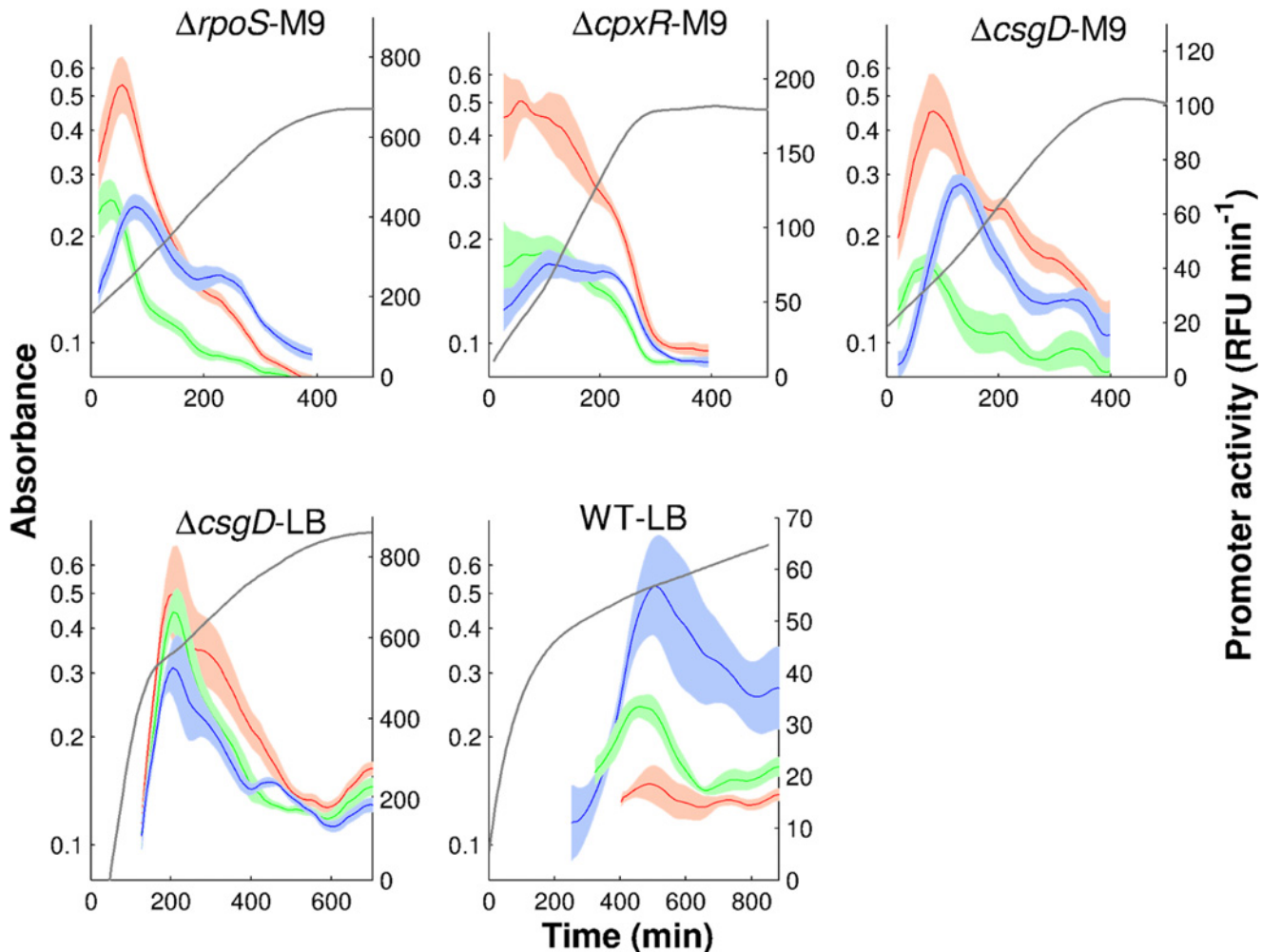


Figure 3. Promoter activities of genes in the FliA-FlgM module. The promoter activities of *fliA* (green), *flgM* (red), and *tar* (blue) measured by means of fluorescent reporter genes in the following experimental conditions: $\Delta rpoS$ strain grown in M9 ($\Delta rpoS$ -M9), $\Delta cpxR$ strain grown in M9 ($\Delta cpxR$ -M9), $\Delta csgD$ strain grown in M9 ($\Delta csgD$ -M9), $\Delta csgD$ strain grown in LB ($\Delta csgD$ -LB), and wild-type strain grown in LB (WT-LB). Grey lines report mean absorbance measurements in the various conditions. The promoter activities and absorbance profiles have been derived from the primary data as illustrated in Fig. 2.

doi:10.1371/journal.pcbi.1004028.g003

The measured promoter activities in Fig. 3 show some common features, such as a transient activity peak of the genes during exponential growth, followed by stabilization at a low level after growth arrest. The induction of the individual promoters has a distinct temporal order, corresponding to the level of the promoters in the transcriptional hierarchy [39]: *fliA*, *flgM*, *tar*. There are also clearly visible differences between the profiles across the conditions though. In M9 medium with glucose the motility genes in the mutant strains are transcribed right from the start, whereas in LB induction occurs only after a number of generations, consistent with previous reports [27, 36]. Moreover, the strength of induction and the duration of the activity peak varies from one condition to the other. For instance, the maximal activity of *tar* varies 10-fold between the WT-LB and $\Delta csgD$ -LB conditions.

Identification of gene regulation functions from promoter activities

The circuit in Fig. 1 has been well-studied over several decades and its regulatory structure is well-known [27–29, 32, 33]. This makes it an excellent test case for investigating which

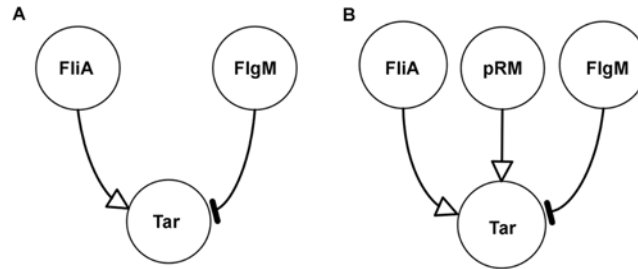


Figure 4. Pattern of regulatory interactions for *tar* and other class 3 genes. A: FliA activates and FlgM inhibits *tar*. B: Idem, but with global physiological effects, measured by the activity of the pRM promoter.

doi:10.1371/journal.pcbi.1004028.g004

information is required for the reliable inference of regulatory interactions and quantitative regulation models from time-series expression data. In a first step, we tested if we could account for measured time-varying promoter activities while ignoring the distinction between mRNA and protein concentrations as well as the activity of the gene expression machinery and other global physiological effects. This corresponds to the usual assumptions made in the analysis of transcriptome data.

We expect FliA to be an activator and FlgM an inhibitor of target genes like *tar* and other class 3 genes (Fig. 4A). In order to check if this regulatory pattern is consistent with the reporter gene data, we used minimal sign pattern analysis [42]. This approach exploits time-series data to invalidate patterns of regulatory interactions, based on the assumption that the activity of a gene is a monotonic function of its regulators. For the patterns that remain after the invalidation step, so-called minimal sign patterns are computed, equivalent to the regulatory structures in Fig. 4. These patterns are minimal in the sense that removing any of the regulators results in an inconsistency with the data, while adding other regulators preserves consistency (see *Methods and materials* and *Text S5* for details on the method).

We applied minimal sign pattern analysis to the reporter gene data in Fig. 3. In particular, we tested if the expected regulatory pattern in Fig. 4A is conserved when replacing the concentrations of FliA and FlgM by the measured promoter activities. We found that the activator role of FliA and the inhibitor role of FlgM are not consistent with the data. This is due to the fact that, over some interval of time in the condition $\Delta rpoS$, a decrease of the promoter activity of *fliA* and an increase of the promoter activity of *flgM* coincide with an increase of the activity of *tar*. As a consequence, the sign pattern corresponding to the expected structure in Fig. 4A is rejected in the analysis.

Despite this structural problem, we also tested to which extent it is possible to quantitatively predict the activity of *tar* from the activities of their regulators. To this end, we developed a mechanistic model of the regulation of this promoter by FliA and FlgM. The model takes into account the titration of FliA by FlgM and the activation of transcription by (free) FliA. We made a quasi-equilibrium assumption for FliA-FlgM association and dissociation, justified by the fast time-scale on which these reactions occur in comparison with transcription and translation processes [43, 44]. Moreover, we chose a Hill function to describe promoter activation and included a basal synthesis rate. The resulting model is:

$$f(t) = k_0 + k_1 \frac{p_{A,free}(t)^n}{\theta^n + p_{A,free}(t)^n}, \quad (1)$$

$$p_{A,free}(t) = \frac{1}{2} \left(-(K + p_M(t) - p_A(t)) + \sqrt{(K + p_M(t) - p_A(t))^2 + 4Kp_A(t)} \right), \quad (2)$$

where $f(t)$ is the time-varying promoter activity, $p_{A,free}(t)$ is the concentration of free FliA, θ is a threshold constant for promoter activation, k_0 and $k_0 + k_1$ are the basal and maximal synthesis rates, respectively, and n is a Hill constant. The concentration of free FliA is computed from the concentrations $p_A(t)$ and $p_M(t)$ of total FliA and FlgM, respectively, and the FliA-FlgM dissociation constant K . All variables and parameters are non-negative and $n \geq 1$. The concentration variables, as well as θ and K , have (arbitrary) units RFU, while the promoter activity and the rate constants have units RFU min⁻¹. The derivation of the model is described in detail in [Text S4](#). Notice that the model is in agreement with the expected pattern of regulatory interactions ([Fig. 4A](#)).

How well does this model fit the data when the total concentrations of FliA and FlgM in [Eq. 2](#), p_A and p_M , are replaced by the measured activities of *fliA* and *flgM*, respectively? We estimated the values of the kinetic parameters $c = (k_0, k_1, n, \theta, K)$ in the regulation model from the data obtained in all five conditions, using a multistart global optimization algorithm [45] to minimize the fitting error $Q(c)$ ([Methods and materials](#)). The algorithm minimizes the mean-square error between the observed promoter activities and the predictions of the model of [Eqs. 1–2](#), while taking into account differences in absolute promoter activity across conditions as well as the time-varying size of confidence intervals ([Methods and materials](#)). The parameters are chosen within physiologically plausible intervals. We notably require that the threshold θ lies within the range of observed FliA concentrations, which corresponds to making the assumption that within the conditions considered here, *tar* varies between its minimal and maximal activity. This is consistent with the observation that motility is low during exponential growth in LB medium [27, 36] and high in a $\Delta csgD$ strain [40].

The predictions of the identified regulation function for *tar* as well as the estimated parameter values are shown in [Fig. 5](#). We computed confidence intervals for the parameter estimates by means of a bootstrap-like procedure resampling the measured promoter activities at each time-point from an experimentally-determined distribution (see [Text S10](#) for details). In this case, and for all parameter values reported in later figures, the confidence intervals are small (< 2-fold). This indicates that there are no identifiability issues, that is, the parameter values can be unambiguously inferred from the data [2, 7, 46].

When analyzing the estimated parameter values, we observe that the cooperativity parameter n equals 1 and that the equilibrium constant K has a value such that the regulator is fully active over the duration of the experiment ([Text S9](#)). As a consequence, the regulation function of the *tar* promoter is essentially a linear transformation of *fliA* activity. While the fit with the experimental data is quite good for the $\Delta csgD$ -LB and WT-LB conditions, the model is not able to account for the peak in *tar* activity in the M9 conditions. The model either predicts no peak or a peak occurring more than an hour before it is observed. In conclusion, replacing protein concentrations by promoter activities in the FliA-FlgM module is inappropriate for obtaining reliable models of the promoter activities, both structurally and quantitatively.

Identification of gene regulation functions from promoter activities including global physiological effects

A possible explanation for the difficulty to identify quantitative regulation functions from information on promoter activities alone may be that, in addition to transcription regulators and other specific regulators, the activity of the transcriptional and translational machinery also affects gene expression [16, 47–49]. Contrary to FliA and FlgM, which affect specific genes, all motility genes are affected by the activity of the gene expression machinery and other global physiological effects. [Fig. 6](#) shows the network structure of the FliA-FlgM module when such global physiological effects are taken into account.

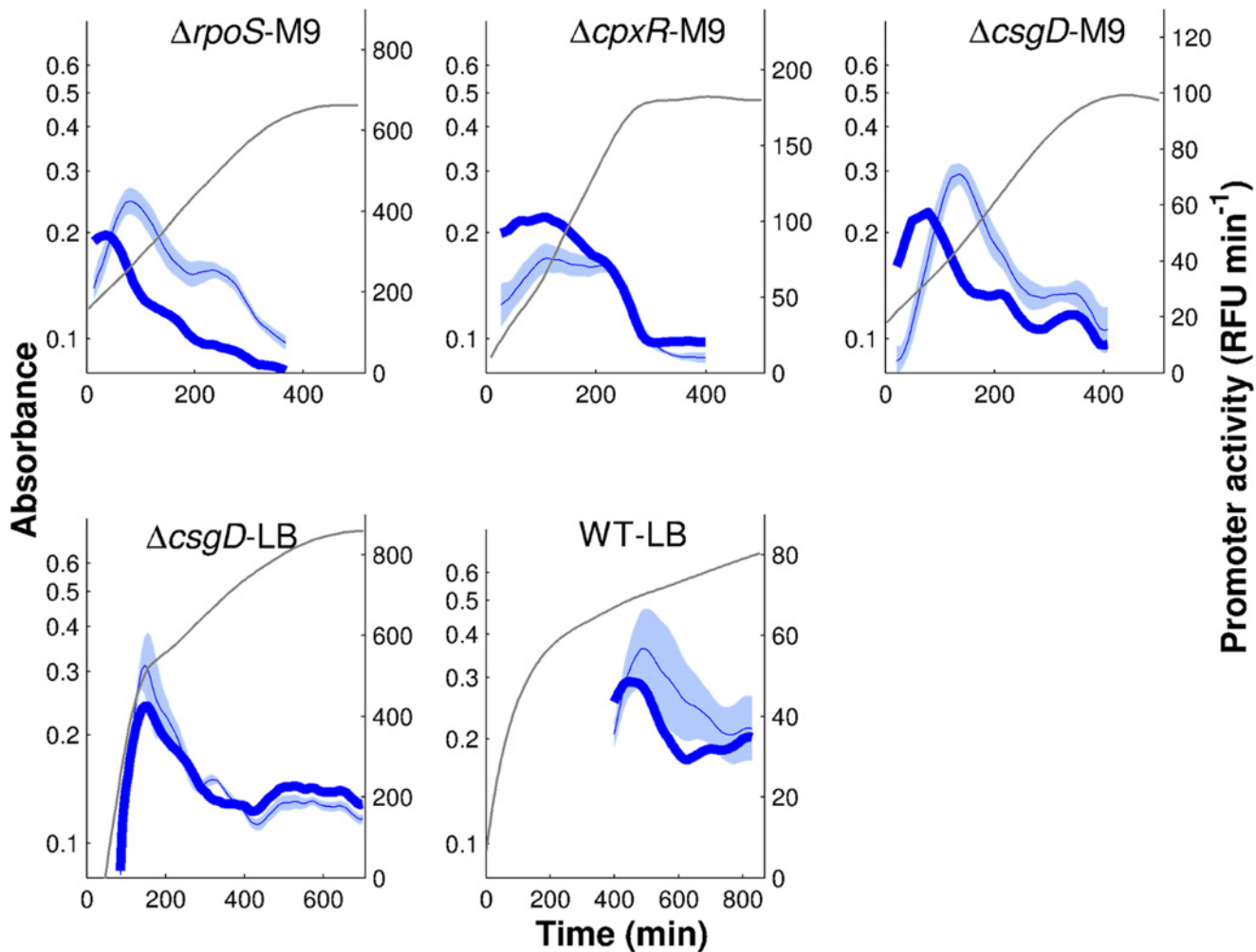


Figure 5. Regulation function of *tar* fitted to reporter gene data when replacing protein concentrations by promoter activities. The regulation function of Eqs. 1–2 was fitted using the promoter activities for *tar*, *flhA*, and *flgM* shown in Fig. 3, where the latter two replace the concentrations of FlhA and FlgM, respectively. Model predictions are in dark blue (thick solid line), *tar* reporter data are in light blue (thin solid line and shaded area). The parameters were estimated using a multistart global optimization algorithm (see [Methods and materials](#) for details). The best fit returns the value $Q = 33.4$ for the objective function, for the parameter vector $(k_0, k_1, n, \theta, K) = (7.6, 853, 1, 663, 14615)$. Confidence intervals for the parameter values are reported in [Text S10](#).

doi:10.1371/journal.pcbi.1004028.g005

The activity of the gene expression machinery includes the abundance and activity of RNA polymerase and ribosome, as well as pools of metabolic precursors, and is therefore difficult to quantify in a direct way. This has motivated the use of the growth rate or the activity of constitutive genes, whose expression is in principle not controlled by any specific regulators, as an indirect read-out of the global physiological state of the cell [16–17, 50]. In this study, following [17], we used the activity of the pRM promoter of phage λ , which is constitutive in non-infected *E. coli* cells, as a quantitative measure of the activity of the gene expression machinery and the global physiological state more generally. In Fig. 7 the time-varying activity of the constitutively-expressed reporter gene is shown, together with the activity of *tar*. Similar to the latter, in almost all conditions, the activity of the constitutive promoter shows a peak, though occurring somewhat later (in WT-LB, the detection of a peak is obscured by the analysis of fluorescence data that were extremely close to background, as witnessed by larger confidence bands).

Does the inclusion of global physiological effects enable the identification of quantitatively predictive gene regulation functions? In order to answer this question, we again applied

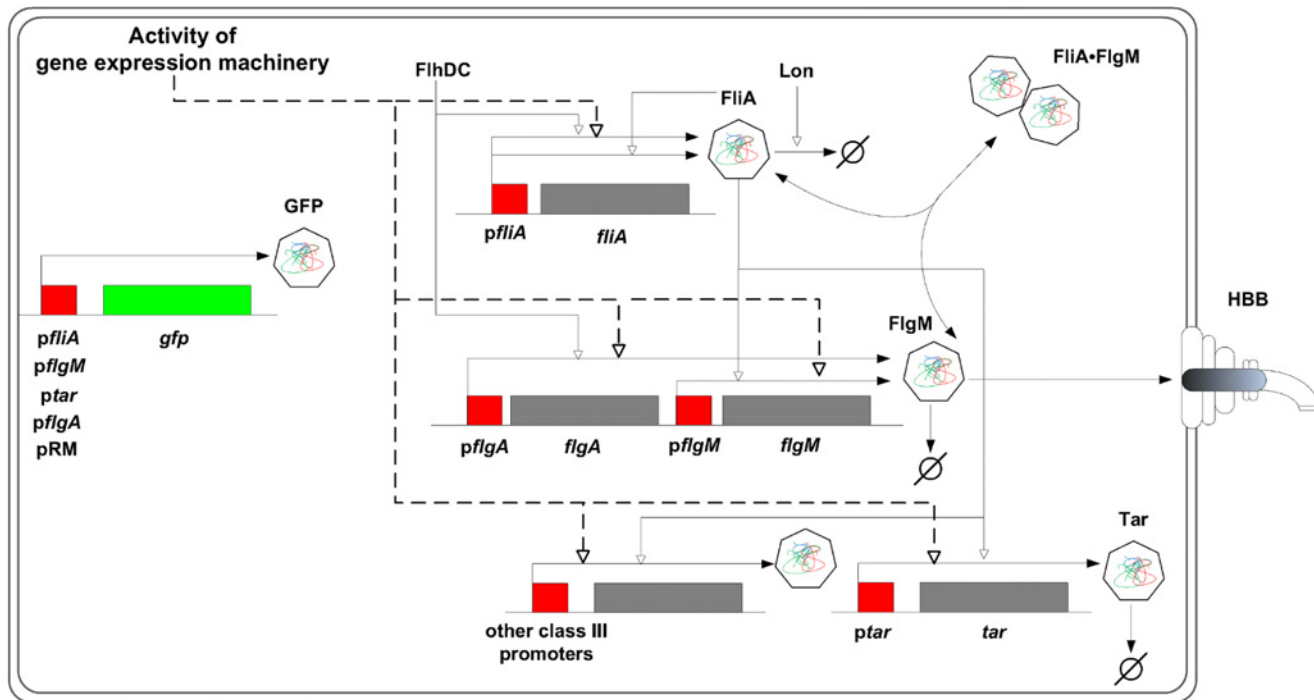


Figure 6. FliA-FlgM module extended with activity of the gene expression machinery. The network is the same as in Fig. 1, but the regulation of the motility genes by global physiological effects, in particular the activity of the gene expression machinery, has been included. These regulatory interactions are shown by bold, dashed lines.

doi:10.1371/journal.pcbi.1004028.g006

minimal sign pattern analysis to the reporter gene data, this time including the activity of the constitutive phage promoter as a proxy for the activity of the gene expression machinery. Like in the previous section, the FliA and FlgM concentrations were replaced by the activities of their genes. The expected pattern of regulatory interactions (activation of the promoter by the gene expression machinery and FliA, inhibition by FlgM) was indeed found to be consistent with the data for *tar* (Fig. 4B).

We also checked if the proposed extension improves the capability of the model to quantitatively account for the time-varying activity of a FliA-controlled promoter. To this end, we multiplied Eq. 1 with $f_{const}(t)$, the measured activity of a constitutive promoter:

$$f(t) = f_{const}(t) \left[k_0 + k_1 \frac{P_{A,free}(t)^n}{\theta^n + P_{A,free}(t)^n} \right]. \quad (3)$$

The fits shown in Fig. 8, obtained with the parameter estimation approach outlined in the previous section, are somewhat better than those obtained with a model accounting for the effects of FliA and FlgM only, especially for the $\Delta rpoS$ -M9 and $\Delta cpXR$ -M9 conditions. The better fit is also reflected in a lower value of the fitting error ($Q = 30.9$ vs $Q = 33.4$). Notice that the extended model has the same number of parameters as the model without global physiological effects in Eqs. 1–2, so that the improvement is not simply due to an increase in the degree of freedom of the model. The parameter estimates are quite similar to those of the previous model.

Although taking into account the activity of the gene expression machinery improves the results, the quantitative predictions of FliA-dependent regulation functions are still unsatisfactory for some conditions, notably $\Delta rpoS$ -M9 and WT-LB. As explained in the Introduction, this may be due to the use of promoter activities as proxies for protein concentrations.

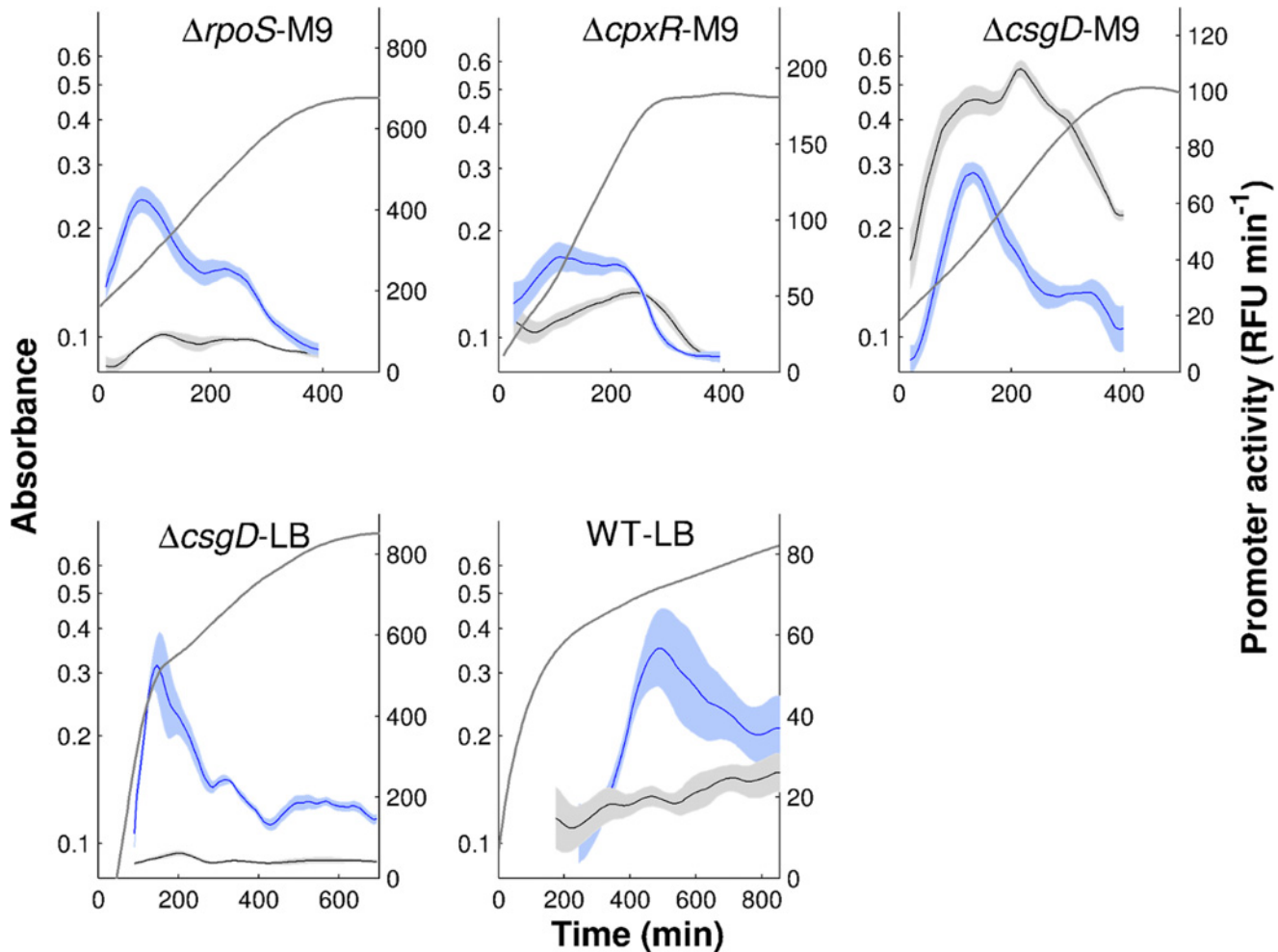


Figure 7. Activities of constitutive phage promoter. Activity of the phage λ promoter pRM (grey) and activity of *tar* (blue) measured in all experimental conditions considered in this study. The *tar* promoter activities (and the mean absorbance profiles) are the same as shown in Fig. 3.

doi:10.1371/journal.pcbi.1004028.g007

We therefore investigated how information on protein concentrations can be integrated into the inference process and if this improves the identification results.

Identification of gene regulation functions from estimates of protein concentrations

It is straightforward to provide an estimate of the GFP concentration, by dividing the fluorescence intensity by the absorbance (*Methods and materials*). The results are shown in Fig. 9. As can be seen, the transcriptional pulse in exponential phase (Fig. 3), leading to a transient accumulation of mRNA, is seen to be followed by the prolonged presence of stable protein, indicating the temporal decorrelation of the promoter activity and the protein concentration. Unfortunately, reporter concentrations are not always representative of the concentrations of proteins of interest, that is, proteins naturally expressed from a promoter. Post-transcriptional regulation and coding bias may cause divergent synthesis rates. The main bias, however, comes from the fact that the two proteins may have different half-lives and thus different degradation rates [25].

Available data in the literature indicate that the half-lives of FliA and FlgM are much shorter than the 19 h of the GFP reporter. The measured half-lives of FliA and FlgM in *Salmonella*

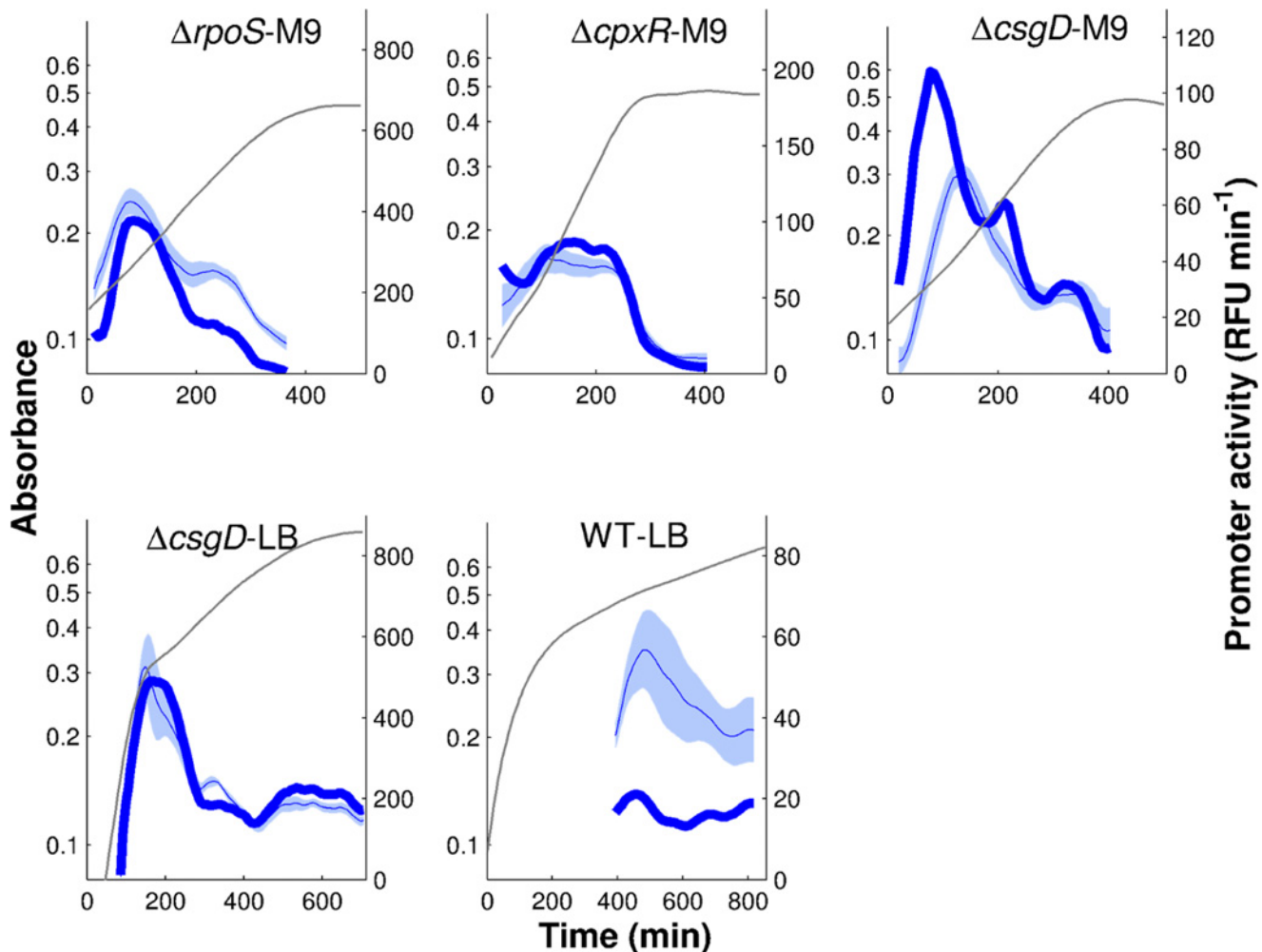


Figure 8. Regulation function of *tar* fitted to reporter gene data when replacing protein concentrations by promoter activities and including global physiological effects. The regulation function of Eqs. 2-3 was fitted using the promoter activities for *tar*, *flhA*, and *flgM* shown in Fig. 3, where the latter two replace the concentrations of FliA and FlgM, respectively. Moreover, global physiological effects are quantified by the activity of the constitutively expressed pRM promoter (Fig. 7). Model predictions are in dark blue (thick solid line), *tar* reporter data are in light blue (thin blue line and shaded area). The parameters were estimated using a multistart global optimization algorithm (see *Methods and materials* for details). The best fit returns the value $Q = 30.9$ for the objective function, for the parameter vector $(k_0, k_1, n, \theta, K) = (0.24, 13.9, 1.2, 353, 14615)$. Confidence intervals for the parameter values are reported in Text S10.

doi:10.1371/journal.pcbi.1004028.g008

enterica wild-type strains growing in LB were found to be 30 min and 18 min, respectively [51]. These half-lives are much shorter than those commonly found in *E. coli*. This can be explained by the fact that, in addition to being physically degraded, FlgM is secreted from the cell (Fig. 1). Moreover, FliA is subject to active degradation by Lon [52].

How can we exploit this information to reconstruct the protein concentration from the promoter activity? As shown in [25] and the *Methods and materials* section, if the half-live of the protein of interest is known, then an estimate of its concentration can be reconstructed from the observed promoter activity using a simple kinetic model integrating the effects of protein synthesis and degradation as well as growth dilution of the protein. Fig. 9 shows the result that is obtained for the FliA concentration, using the above-mentioned half-life. Although the difference with the promoter activities is less striking than for the GFP concentrations, the computation of the concentration via integration of the corresponding activity smoothens out the rapid variations of the activities and changes the time-varying profile of the regulators.

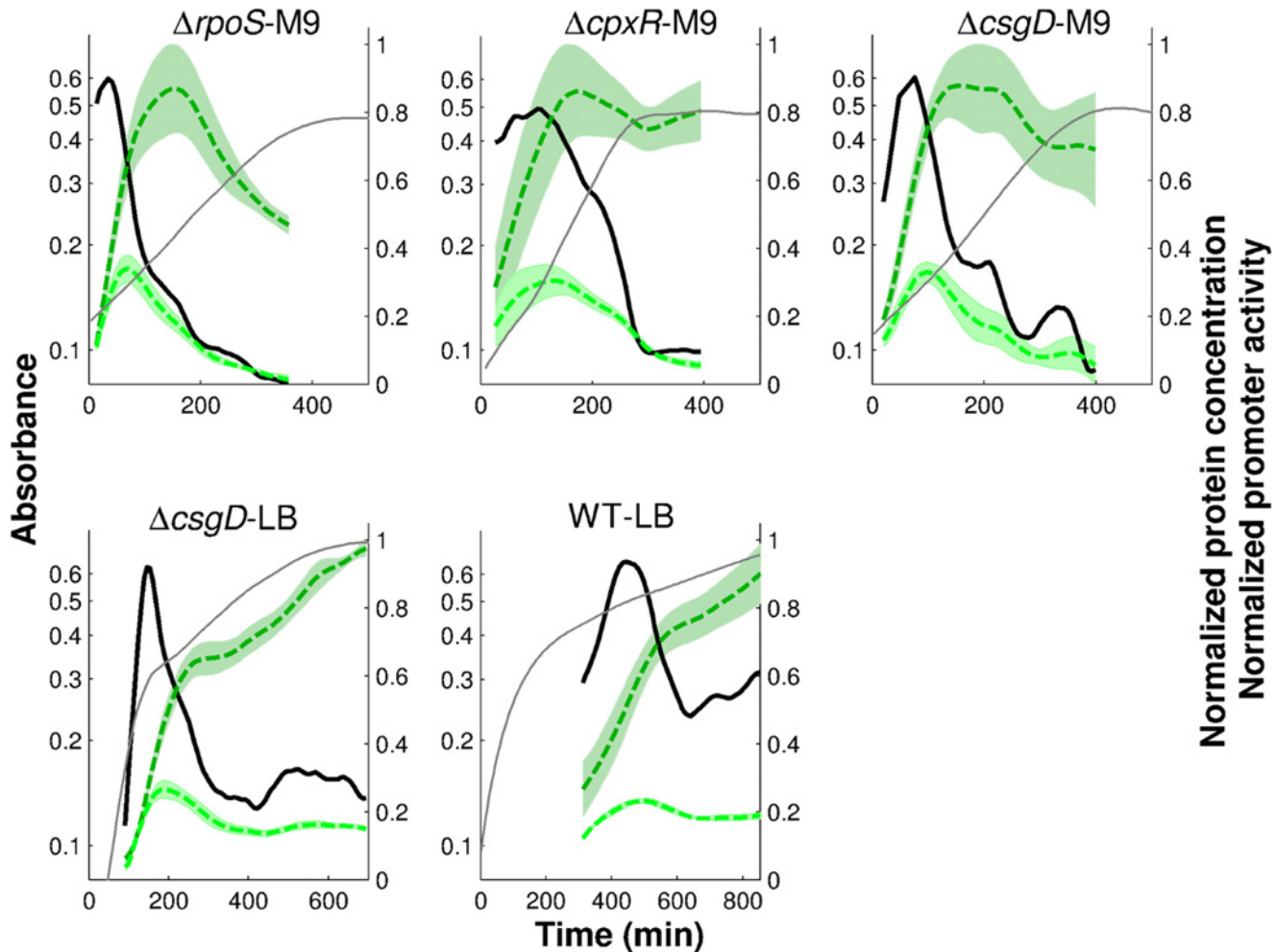


Figure 9. Estimates of FliA concentrations from reporter gene data. Concentrations of FliA computed from *fliA* promoter activity (thick, black solid line) in all experimental conditions considered in this study. The *fliA* activities are the same as shown in Fig. 3. The dashed, dark green line represents the concentration of the reporter protein, while the dashed, light green line represents the reconstructed FliA concentration for the measured half-life of 30 min. In each condition, promoter activity has been normalized with respect to its maximum value. The protein concentrations have been normalized with respect to the maximum of the upper limit of the confidence interval of the reporter concentration. The shaded regions correspond to the mean of the protein concentrations \pm twice the standard error of the mean. For clarity, the confidence intervals of the promoter activities have been omitted.

doi:10.1371/journal.pcbi.1004028.g009

A tacit assumption in the computation of protein concentrations from promoter activities is that the half-lives of the proteins are constant over the duration of the experiment. This may not be true in the system considered here, since the apparent half-lives of FliA and FlgM are regulated. In particular, the secretion rate of FlgM varies with the synthesis of HBB structures. Data from the literature indicate that the first FlgM molecules appear in the extracellular medium shortly after the induction of *fliA* [52, 53]. Once the cell population stops growing, the rate of assembling new flagella and thus the secretion of FlgM come to a halt as well. This increases the apparent half-lives of FliA and FlgM to 2 h and 3 h, respectively [52, 53]. Since our kinetic experiments focused on the exponential growth phase, and the analysis is limited to the time frame in which *fliA* and *flgM* are expressed, it is justified here to assume that the half-lives of FliA and FlgM are constant.

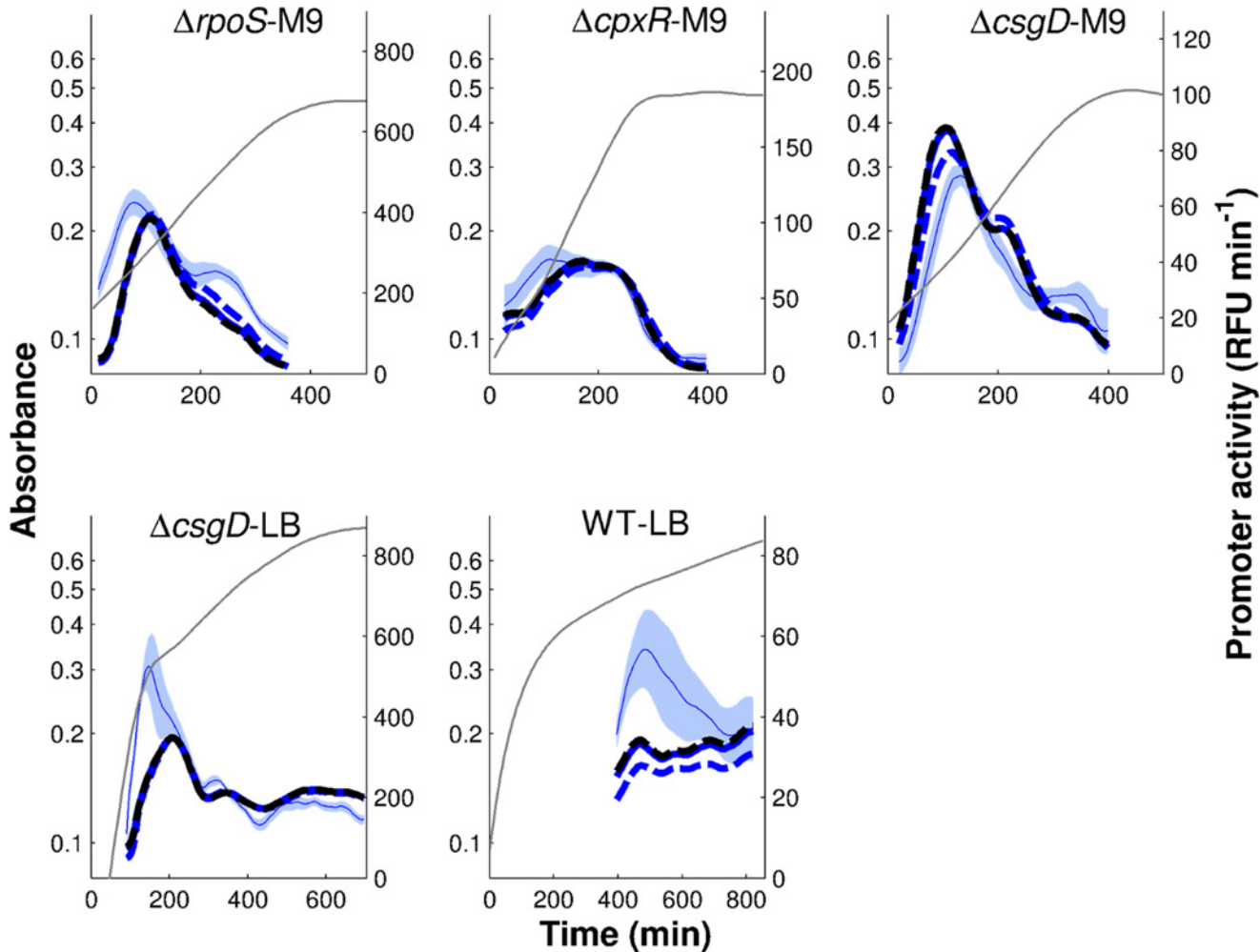


Figure 10. Regulation function of *tar* fitted to reporter gene data when reconstructing protein concentrations from the reporter gene data and including global physiological effects. The regulation function of Eqs. 2–3 was fit to the data using the promoter activity for *tar* (Fig. 3), concentrations of FliA and FlgM reconstructed from the activities of their promoters for physiologically realistic half-lives (Fig. 9 and Text S7), and the activity of the constitutively expressed pRM promoter quantifying global physiological effects (Fig. 7). Model predictions are in thick black and blue lines, *tar* reporter data are in light blue (thin line and shaded area). Three fits are shown: the best fit for measured half-lives of FliA and FlgM of 30 min and 18 min, respectively (thick, blue solid line, $Q = 25.5$, $(k_0, k_1, n, \theta, K) = (0.26, 5.0, 1.99, 3542, 447499)$) and two other fits for comparable half-lives (blue and black dashed lines). Parameter values were estimated using a multistart global optimization algorithm (see [Methods and materials](#) for details). Their confidence intervals are reported in [Text S10](#).

doi:10.1371/journal.pcbi.1004028.g010

Does the estimation of time-varying protein concentrations from promoter activities, by means of a kinetic model and physiologically realistic half-lives, improve the inference of regulatory interactions and gene regulation functions?

We first verified that a model using the reconstructed FliA and FlgM concentrations as regulators of *tar*, in addition to the activity of the gene expression machinery, is structurally compatible with the data. Minimal sign pattern analysis accepted the expected pattern of regulatory interactions. Second, we identified the gene regulation model of Eqs. 2–3 from the data, with the estimated FliA and FlgM concentrations for p_A and p_M , respectively. As shown in [Fig. 10](#), the model better captures the quantitative trend in the data, including in WT-LB, where the improvement was moderate though, and the resulting fit still improvable ($Q = 25.5$).

Since the half-lives were taken to be those measured for a different species in growth conditions that are similar but not identical to ours, and measurement errors were not reported, we slightly relaxed the reported values. This did not much change the quality of the fit (Fig. 10). We conclude that even approximately correct half-live values may allow the results of the inference process to be improved.

The above analysis ignores a particularity of the FliA-FlgM module, namely that although the half-lives are constant in the time-window of the experiment, they may be different across growth conditions. Generally speaking, in environmental conditions favoring a larger number of flagella, and thus completed HBB structures, the secretion rate of FlgM is higher and therefore the apparent half-life shorter. For example, during growth of a wild-type strain in LB medium, the apparent half-live of FlgM is 18 min [51], but in conditions of strong induction of the flagellar hierarchy half-lives up to 7 min were measured [54]. The half-life of FliA, the flagellar sigma factor, is also variable. FliA is subject to active degradation by the Lon protease, but stabilized when bound to FlgM (Fig. 1). This makes its apparent half-life dependent on the concentration of its anti-sigma factor [52].

The observation that the half-lives of FliA and FlgM are not identical across all growth conditions considered suggests a final extension of the analysis to improve the inference results. We allowed the FliA and FlgM half-lives to vary between physiologically possible bounds in each of the conditions and estimated not only the parameters of the regulation functions, but also the half-lives. In order to reduce the computational complexity of this procedure, we discretized the possible half-live values for FliA and FlgM (27 values, between 7 min and 4 h), and we precomputed the protein concentration profiles for each half-life in each of the experimental conditions. The resulting time-course patterns were used for the same analyses as above.

Fig. 11 shows the results for the structural inference of *tar* regulators. As can be seen, almost all combinations of half-lives are compatible with activation of *tar* by FliA and the gene expression machinery as well as with inhibition by FlgM. This means that the returned structure of interactions is robust over the range of half-lives, a desirable property for network inference. Fig. 12 illustrates that the obtained quantitative regulation function of *tar* activity fits the data better than in all other previously considered situations ($Q = 21.0$), while the parameter values are similar to those obtained in the previous sections. Although we substantially relaxed the possible half-live values of FliA and FlgM, it is remarkable that the optimal values are close to the reported values for LB medium (Fig. 12). This emphasizes the importance of active degradation of FliA and secretion of FlgM for the dynamics of the motility network. Moreover, while the proportion of FliA released by FlgM varies across conditions, most FliA is predicted to be free over the duration of the experiment (Text S9). This is also intuitively expected, as FlgM is actively exported in the exponential growth phase considered.

We conclude that the reconstruction of protein concentrations from reporter gene data results in much better inference results for the FliA-FlgM module, for physiologically plausible values. The computation of the protein concentrations requires a simple kinetic model, accounting for protein synthesis and degradation, as well as estimates of the protein half-lives. While this increases the complexity of the data analysis procedures, it reflects the actual dynamics of gene expression and is thus critical for exploiting time-series measurements. Moreover, the availability of information on protein half-lives may not be constraining in practice, since even rough half-live estimates from the literature were seen to preserve the expected interaction pattern and provide a significant improvement of the ability of the models to quantitatively describe the time-varying promoter activity. It is important to remark, however, that adding information on protein half-lives is not enough. When repeating the identification process with the measured half-lives, but ignoring global physiological effects, the results are far worse ($Q = 36.3$, Text S7).

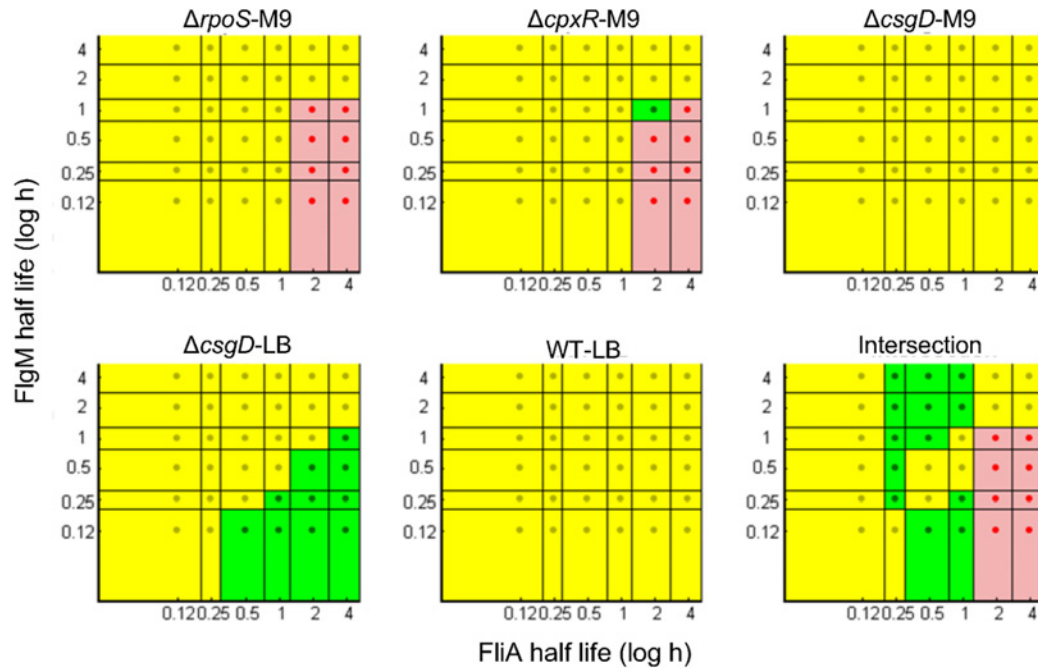


Figure 11. Minimal patterns of regulatory interactions for *tar* over a range of physiologically realistic half-lives. The minimal regulatory patterns for the gene *tar* in the motility network of Fig. 6 as a function of the half-lives of FliA and FigM. The plots correspond to the five experimental conditions considered ($\Delta rpoS$ -M9, $\Delta cpxR$ -M9, $\Delta csgD$ -M9, $\Delta csgD$ -LB, and WT-LB) as well as the pooling of the data sets from all five conditions. The dot in the center of each region in the plots corresponds to a tested combination of half-lives of FliA and FigM, and thus to specific protein concentration profiles computed from the kinetic model of gene expression (Methods and materials). The minimal regulatory patterns were obtained by applying the minimal sign pattern algorithm [42]. The color codes represent the different categories of minimal signal patterns inferred. A region is colored green if the expected regulatory pattern is among the minimal sign patterns returned by the algorithm, and yellow if it is compatible with the returned sign patterns. A region is colored red if none of the returned sign patterns is consistent with the data only. Two examples of inconsistent sign patterns are shown. Note that, for every combination of half-lives, the analysis of the pooled data (results reported as “Intersection”) is generally more constraining than the pooling of the results from individual analyses: The expected pattern may be consistent (yellow) with all individual datasets but not minimal (green) for any of them, and turn out to be consistent and minimal (green) when all datasets are analyzed at once (see also Text S5).

doi:10.1371/journal.pcbi.1004028.g011

Determination of conditions in which protein half-lives and global physiological effects are important

The importance of accounting for global physiological effects and protein half-lives was demonstrated above for the regulation of the expression of *tar*. The same analysis was repeated for the regulation of the *flgM* promoter. Results are reported in Text S7. We found that, for this promoter, the improvement in the fit to the experimental data obtained by including global physiological effects and protein kinetics is not as pronounced as for *tar*. One possible explanation is that the *flgM* activity profile happens to be already well explained using the promoter activities of *fliA* and *flgM* as proxies for the corresponding protein concentrations (Text S7), thus leaving little space for improvement. In addition, from a mathematical viewpoint, we notice

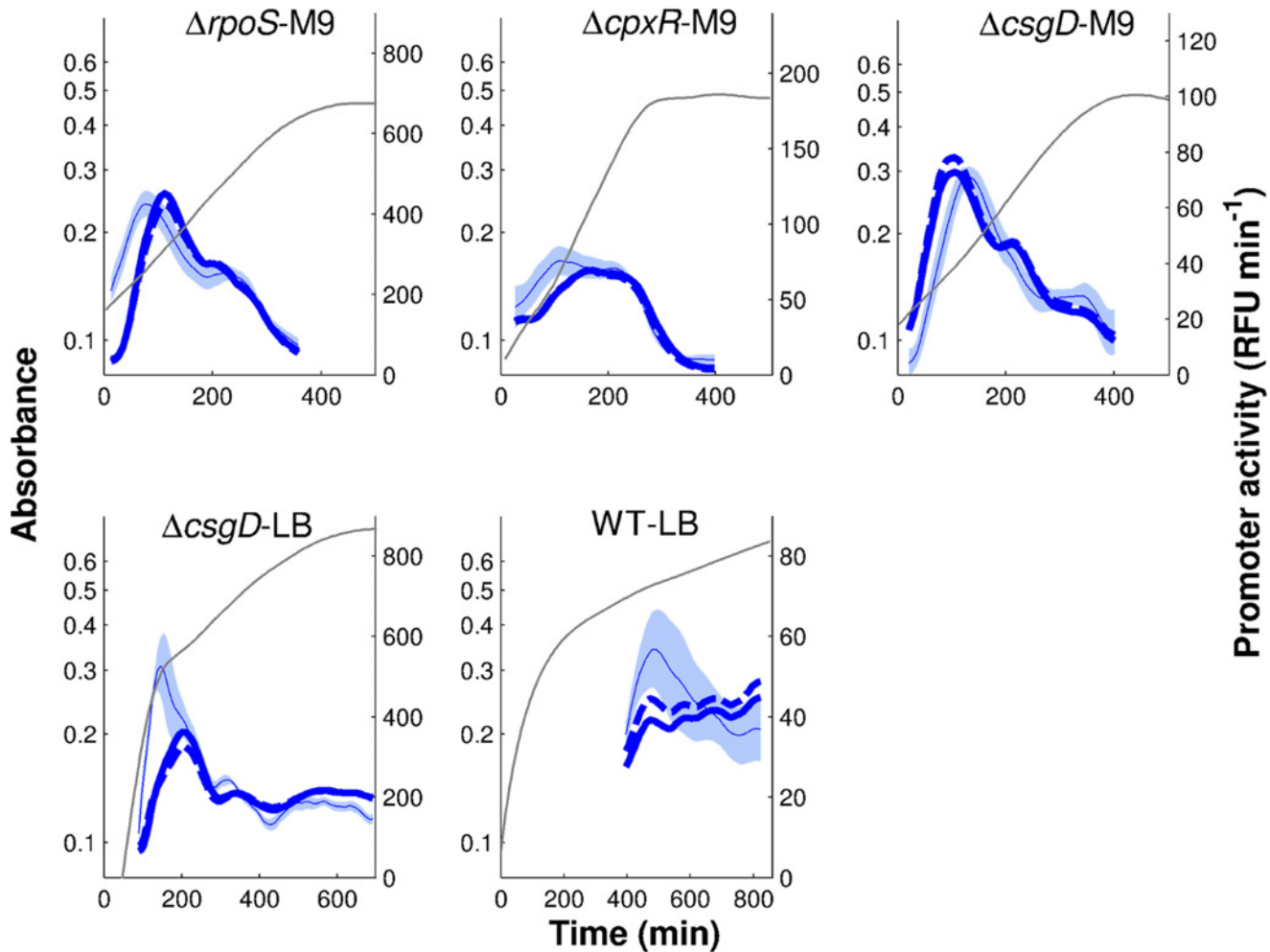


Figure 12. Regulation function of *tar* fitted to reporter gene data when reconstructing protein concentrations from the reporter gene data for physiologically realistic half-lives and including global physiological effects. As in Fig. 10, but the half-lives have now also been estimated from the data, within a physiologically plausible range. Model predictions are in thick solid and dashed blue lines, *tar* reporter data are in light blue (thin line and shaded area). Two example fits are shown, namely the best fit for estimated half-lives of FliA and FlgM (solid line, $Q = 21.0$, $(k_0, k_1, n, \theta, K) = (0.22, 6.6, 1.38, 6252, 47467)$) and another example of a high-ranking fit (dashed line). In the case of the best fit, the half-lives of FliA are equal to (60,30,24,30,60) min in the ($\Delta rpoS$, $\Delta cpxR$, $\Delta csgD$ -M9, $\Delta csgD$ -LB, WT-LB) conditions, respectively, while the half-lives of FlgM are equal to (45,7,24,11,9) min. Confidence intervals for the parameter values are reported in Text S10.

doi:10.1371/journal.pcbi.1004028.g012

that using the promoter activity of *flgM* for the fitting of the same quantity may render the regression problem degenerate. Still, these results raise a more general question: When is it important to take into account protein half-lives and global physiological effects?

To answer this question we performed an *in-silico* analysis where the regulation model of Eqs. 2–3 is simulated for different protein half-lives and varying strength of the global physiological contribution, using the *pflia*, *pflgM*, and *pRM* activity profiles reported in Figs. 3 and 7. Identification is then attempted from the simulated data with models ignoring protein half-lives and global physiology. This enables us to quantify the relevance of the analysis in the previous sections for a variety of realistic scenarios, starting from experimentally measured activities of bacterial promoters.

To evaluate the importance of protein half-lives, we simulated FliA and FlgM concentration profiles for half-lives ranging between 7 minutes and 16 hours. The other relevant parameters

in the model (k_0, k_1, n, θ, K) were fixed in agreement with the best fit obtained for the reference half-lives of 30 min for FliA and 18 min for FlgM, shown in Fig. 10. More precisely, the relative position of the parameter values within the interval of physiologically plausible values, which may depend on the FliA and FlgM concentrations, as explained in the [Methods and materials](#), was conserved across conditions. Activity profiles of *tar* were then generated in accordance with Eqs. 2–3 based on the experimentally measured pRM activities. We then attempted to identify from these simulated data a gene regulation model accounting for the global physiological effects, but using promoter activities in place of FliA and FlgM concentrations. The results are reported in Fig. 13.

As can be seen, the quality of the fit decreases with longer half-lives of FliA, but is rather insensitive to the half-life of FlgM. The strong dependency on the half-life of FliA shows that, in general, accounting for slow protein kinetics is important, but that promoter activities can be safely used in place of protein concentrations for very fast-degrading proteins. This is intuitively explained by the fact that fast-degrading protein concentration profiles reproduce promoter activity profiles quite closely, while this is not true in case of slow degradation (Fig. 9). The relative insensitivity to FlgM half-lives can be explained by the fact that, in the time window considered in our experimental set-up, a good fit requires most FliA to be free (Text S9). Longer half-lives, and therefore higher concentrations of FlgM, favor lower free FliA concentrations, but this tendency is compensated in the parameter optimization process by higher values for the equilibrium constant K . The actually measured reference half-lives of 18 min for FlgM and 30 min for FliA are located in the upper left corner of Fig. 13A, where fitting residuals are comparably small. Therefore, for networks involving regulators with longer half-lives than the exceptionally short half-lives observed for FliA and FlgM, it will be even more critical to account for protein kinetics than for the genes considered here.

To evaluate the importance of global physiological effects, starting from the experimentally measured pRM activity profiles, we simulated global physiological effects of different strength. In particular, we rescaled the variations of $f_{const}(t)$ around its temporal mean across all conditions, \bar{f}_{const} , by a factor α ranging from 0 (no variability, no regulatory effect) to 1 (measured variability, moderate regulatory effect) and 1.25 (increased variability, strong regulatory effect). That is, synthetic activity profiles of FliA-dependent promoters were generated in accordance with the model

$$f(t) = [\alpha \cdot (f_{const}(t) - \bar{f}_{const}) + \bar{f}_{const}] \cdot \left[k_0 + k_1 \frac{p_{A,free}(t)^n}{\theta^n + p_{A,free}(t)^n} \right], \quad (4)$$

with $p_{A,free}(t)$ computed from the FliA and FlgM concentration profiles according to Eq. 2. The upper bound of 1.25 for α was chosen so as to avoid negative values of the promoter activity $f(t)$.

Identification results using FliA and FlgM concentrations computed for the reference half-lives of 30 min and 18 min, respectively, but ignoring global physiological effects are reported in Fig. 13B. It is clear that the misfit of the *tar* promoter activity data increases with the strength of the ignored physiological effects. In particular, with the experimentally observed pRM activity ($\alpha = 1$), the discrepancy between the data and the best model fit is quite significant. This is in agreement with the results of previous sections and especially Text S7, where it is shown that ignoring global physiological effects, even when computing protein concentrations from promoter activities, leads to poor model fits. While neglecting small variations of global physiological state ($\alpha \ll 1$) may be safe, ignoring highly varying global physiological effects ($\alpha > 1$) may have even more severe repercussions on the inference results than those observed here.

In summary, the simulation study shows that, as expected, the importance of accounting for protein kinetics and global physiological effects depends on the strength of these effects,

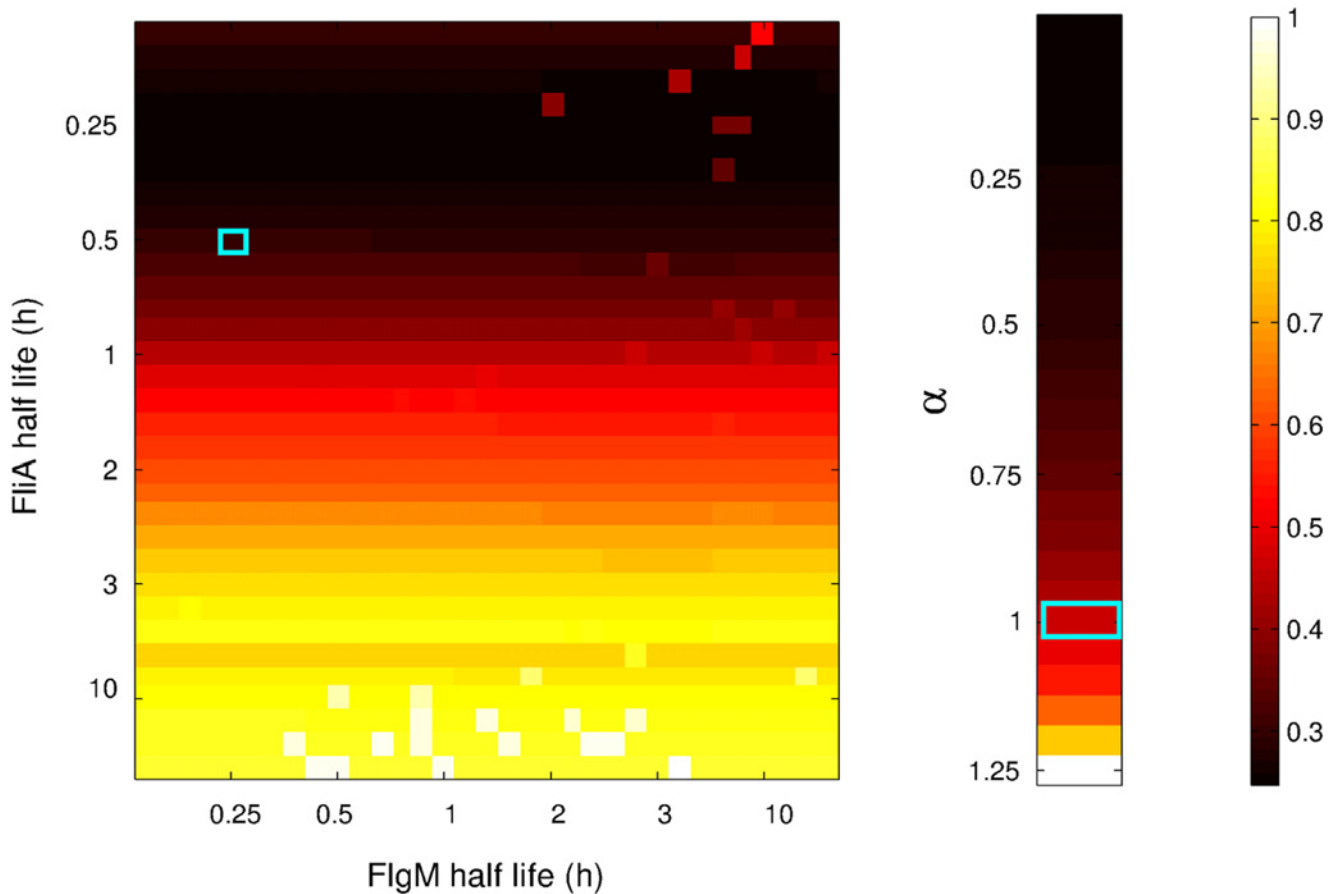


Figure 13. Heatmap of the fitting residuals for simulated data generated for different protein half-lives and for different strengths of global physiological effects. A: For all different combinations of 33 half-lives of FlgM (horizontal axis) and FliA (vertical axis), the residual of the fit for a model ignoring protein kinetics is represented by the color code reported in the right bar. For clarity of presentation, the residual values Q have been normalized with respect to the maximum value of Q over all half-life combinations. The combination corresponding to the measured half-lives in LB medium is marked with a light blue square (18 min for FlgM, 30 min for FliA). B: For 26 different values of the strength parameter α , defined in Eq. 4, the residual of the fit by a model ignoring global physiological effects is represented by the color code. The residual values Q have been normalized with respect to the maximum value of Q over the different strengths of physiological effects. The value corresponding to the real data is marked with a light blue rectangle ($\alpha = 1$).

doi:10.1371/journal.pcbi.1004028.g013

although the structure of the system itself may also play a role, as illustrated by the differences in the dependency of the fit quality on FliA and FlgM concentrations (Fig. 13A). As a general rule, ignoring significant fluctuations of the global physiology or large differences between mRNA and protein half-lives is very likely to result in modelling bias and hence poor inference results. Interestingly, in the previous sections a substantial improvement of the fit of a quantitative regulation function to *tar* activity was already obtained when taking into account concentrations of short-lived proteins and moderately-variable global physiological effects. In the light of the analysis of this section, the contribution of our approach becomes even more fundamental in other systems, bearing in mind that the vast majority of bacterial proteins are much more stable than FliA and FlgM, which are actively degraded and exported from the cell (Fig. 1).

Discussion

Experimental techniques developed over the past two decades have made it possible to monitor gene expression with high precision and temporal resolution. The interpretation of these data requires reliable mathematical and computational tools for the inference of regulatory

interactions as well as the identification of quantitative gene regulation functions. While enormous progress has been made on such inference methods, many problems remain. We believe that the solution of these problems should not only be sought in technical improvements of the algorithms themselves, but should also come from a better understanding of the precise information on gene expression provided by the experimental data. The relation between the primary data and physiological quantities like the cellular concentrations of mRNA and protein is usually indirect and obscured by simplifications and assumptions that do not generalize beyond the specific situations for which they were designed.

In this paper we have made explicit the relation between experimental data and physiological quantities by means of mathematical models of gene expression, calling into question two basic assumptions that are commonly made in the inference of regulatory interactions and quantitative gene regulation functions from time-series data.

The first assumption is that transcriptome data alone are sufficient to capture the time-varying state of gene expression. Often, the regulators of gene expression are proteins and, whereas mRNA and protein concentrations are correlated at steady state, this is generally not the case when the two are considered dynamically over time. As a consequence, neglecting the distinction between mRNA and protein may hamper the full and correct exploitation of the information contained in time-series transcriptome data. This might explain why the comprehensive evaluation of network inference methods carried out in the DREAM initiative concluded that steady-state transcriptome data comparing wild-type and mutant strains are usually more informative for network inference than time-series data [55]. The temporal decorrelation of the mRNA and protein concentrations makes the former generally an unreliable proxy of the latter.

A second implicit assumption in the analysis of transcriptome data is that gene regulation can be reduced to the action of transcription factors and other specific regulators. This ignores the fact that the activity of the transcriptional and translational machinery, as well as other global physiological effects such as gene copy number and DNA supercoiling, may drastically change over the course of an experiment, a fact that has been well-documented for microorganisms [56–58]. As Lovén *et al.* demonstrate, a global increase or decrease of transcriptional activity across conditions may lead to erroneous interpretations and the inference of spurious regulatory interactions [21].

The main contribution of this paper is an integrated experimental and computational approach for addressing the above two problems, in the context of time-series measurements of gene expression by means of fluorescent reporter genes. We propose new controls for transcriptome experiments, in particular the use of constitutively-expressed genes, as well as mathematical models and computational procedures for reconstructing protein concentrations and for integrating global physiological effects into the network inference process. The reconstruction of protein concentrations from real-time promoter activities by means of kinetic models as well as the quantification of global physiological effects by means of reporter genes have been proposed before [17–18, 25, 59]. For instance, Gerosa *et al.* have developed quantitative models to dissect global and specific regulation of *E. coli* genes involved in arginine biosynthesis [18]. To our knowledge, however, the work presented here is the first systematic study of how the integration of information on both global physiological effects and protein concentrations can improve the inference of regulatory interactions and the identification of regulation functions from time-series gene expression data.

It is important to emphasize that the proposed approach is orthogonal to existing inference methods and that the models and analysis procedures proposed in this study can be directly combined with many of the methods described in the literature [1, 3–7]. The models and analysis procedures we have used are explicitly detailed and can be easily integrated into available methods, as illustrated for the MATLAB implementation of the minimal sign pattern

algorithm [42]. While reporter gene data were used in this study, other experimental techniques may also yield time-series transcriptome data suitable for our purpose. The main requirement is that an estimate of the expression of a constitutive gene can be obtained and sampling times are sufficiently dense and precise to allow time-varying mRNA concentrations to be reliably measured.

We have validated our approach by means of a central module of the motility network in *E. coli*. The FliA-FlgM module has been very well-studied and has characteristics that make it atypical but particularly suitable for our purpose. FlgM is secreted from the cell and FliA is a target for proteolysis, which causes these regulators to have apparent half-lives that are quite short in comparison with typical *E. coli* proteins. Moreover, the secretion and degradation rates may change across conditions, depending on the strength of induction of the flagella synthesis network. This yields a rich and challenging data set for testing how accounting for the distinction between cellular responses on the level of mRNA and protein influences the results of the inference process.

We investigated the capability to infer from reporter gene data both the regulatory structure and the quantitative regulation function of a FliA-dependent motility gene, not known to be regulated by any other transcription factors. When progressively solving the problems mentioned above, by integrating information on the activity of the gene expression machinery and computing estimates of protein concentrations from promoter activities, both the structure and the dynamics of the regulation of the *tar* promoter could be identified successfully. We emphasize that, when using available measurements of FliA and FlgM half-lives, this was achieved without increasing the number of parameters in the models and is therefore not simply a consequence of increasing the degrees of freedom. Moreover, *a-posteriori* analysis of the confidence intervals of the parameter estimates (Text S10) confirmed that there are no identifiability issues, that is, the models are fully determined by the available data.

The results underline the important roles played by global physiological effects and the active regulation of FliA and FlgM half-lives in shaping the dynamics of FliA-dependent promoters. When global physiological effects were ignored, or the FliA and FlgM half-lives were set to typical values of *E. coli* proteins, a sharp drop in the quantitative predictivity of the gene regulation models was observed (Text S7). In other words, both the inclusion of global physiological effects and realistic half-lives were necessary to improve the inference results in our example network.

More generally, under which conditions does the inclusion of the above factors lead to better results and when can they be ignored? We performed a simulation study in which we systematically varied the relative contribution of global physiological effects to cross-condition variations in the expression of a target gene and the half-lives of the regulators. These results showed that longer half-lives of the activating transcription factor and stronger variations of global physiological effects make it more difficult to obtain good fits when using promoter activities and data on specific regulators only, respectively. While these conclusions are not surprising, it is important to emphasize that in the system studied here, where FliA and FlgM have half-lives that are exceptionally short for bacterial proteins, a considerable improvement of the fit could be obtained. For regulatory proteins with more typical half-lives, the gain may therefore be even more important than observed here.

The proposed approach to better exploit the information contained in time-series data of the transcriptional response of bacterial cells depends on kinetic models of gene expression, relating the primary fluorescence and absorbance data to promoter activities and protein concentrations. The models used in this study could be further refined, by taking into account delays that are due to the maturation of GFP and the time for rounds of transcription and translation to complete [25, 60–62]. These refinements were neglected here, since the GFP reporter used in this study is fast-folding and the transcription and translation delays are short on the

time-scale of the experiments. The computation of protein concentrations by means of these measurement models depends on the availability of approximate values of the protein half-lives. While genome-wide studies of the stability of individual proteins exist, *e.g.*, for yeast [63], quantitative information on the stability of individual proteins in microorganisms is seldom available. Still, it is known that most proteins in *E. coli* are stable, with half-lives > 10 h. In most experimental scenarios in the laboratory, bacterial growth occurs at a much higher pace, *i.e.*, $\mu \gg \gamma_p$, which is sufficient to ensure correct applicability of our measurement models regardless of the specific (poorly known) value of γ_p . Turned another way, to apply our measurement models in cases of poorly known half-lives of stable proteins, it suffices to perform experiments with cell doubling times well below 10 h.

In conclusion, the applicability of our principled approach to account for protein degradation and global effects in network reconstruction from reporter gene data goes well beyond the simple and well-understood biological system on which it was illustrated and the specific network analysis and identification methods utilized. In fact, the use of the proposed approach becomes even more important in problems involving networks that are less known and/or of greater complexity, in that the identification problem becomes intrinsically more difficult, and therefore the biases introduced by common though weakly justified hypotheses or approximations become even more difficult to discern. Due to the generality of both the problem and the proposed solutions, we believe that the methodology presented in this paper has broad practical applicability for analyzing time-series transcriptome data and improving network inference in a variety of organisms.

Methods and materials

Strains and growth conditions

The *E. coli* strains used for this study are the wild-type strain BW25113 and isogenic deletion mutants $\Delta rpoS$, $\Delta csgD$ and $\Delta cpxR$. The strains were taken from the Keio collection [64] and the kanamycin resistance cassette was removed [40]. The wild-type and mutant strains were transformed with low-copy plasmids bearing a fusion of a *gfpmut2* reporter gene with the promoter regions of the genes *tar*, *fliA*, and *flgM*. These plasmids were selected from the plasmid library constructed at the Weizmann Institute [65]. A reporter for the pRM promoter of phage λ was constructed in the same plasmid vector to provide information on the physiological state of the bacteria, following the approach in [17]. The pRM promoter fused with the *gfp* reporter gene was also inserted into the chromosome of the BW25113 wild-type strain as reference for the qRT-PCR assays. All plasmids carry the kanamycin resistance gene. All the strains and plasmids were verified by PCR. More details on the strains and plasmids used in this study can be found in [Text S1](#).

The strains were recovered from glycerol stock and grown overnight (16 h) at 37°C in LB rich medium and M9 minimal medium [66] supplemented with 0.3% glucose and mineral trace elements. For the preculture of strains containing plasmids, kanamycin (50 $\mu\text{g/ml}$) was added. The overnight cultures were diluted (10- to 100-fold) into a 96-well microplate, so as to obtain an adjusted initial OD₆₀₀ of 0.2. The wells of the microplate contain 150 μl of the above medium, to which was added 1.2% of the buffering agent HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) for maintaining a constant external pH. The wells were covered with 60 μl of mineral oil to avoid evaporation. The microplate cultures were then grown for about 16 h at 37°C, with agitation at regular intervals, in a microplate reader (Fusion Alpha, Perkin-Elmer).

Experimental monitoring of gene expression in real time

The expression of the fluorescent reporter genes in different genetic backgrounds and different growth media was monitored *in vivo* and in real time. About 150 readings each of absorbance

(600 nm) and fluorescence (485/520 nm) were obtained during a typical experiment using the Perkin-Elmer microplate reader. In order to compute promoter activities and protein concentrations from these data, data analysis procedures were designed and implemented in MATLAB, completing earlier work [17, 25]. These data procedures account for the specific half-life of the fluorescent reporter protein and take special care in the subtraction of the autofluorescence background (see [Text S3](#) for details on the data analysis procedures).

Denoting by $A(t)$ and $I(t)$ the (background-corrected) time-varying absorbance and fluorescence signals, we computed the reporter concentration $r(t)$ and the promoter activity $f(t)$ by means of the following formulas:

$$r(t) = \frac{I(t)}{A(t)}, \tag{5}$$

$$f(t) = \frac{d}{dt}r(t) + (\gamma_r + \mu(t))r(t) = \frac{\frac{d}{dt}I(t)}{A(t)} + \gamma_r \frac{I(t)}{A(t)}, \tag{6}$$

where γ_r [min^{-1}] is the degradation constant of the reporter and $\mu(t)$ [min^{-1}] the growth rate of the bacteria. The half-life of the protein is defined as $t_{1/2} = \ln 2 / \gamma_r$. The reporter concentration is expressed in units RFU and the promoter activity in units RFU min^{-1} , as is usual for this kind of measurements (see [17] and [Text S3](#)). The growth rate is easily estimated from the time-varying absorbance, using the standard relation $\mu(t) = d \ln A(t) / dt$. The above equations rely on the use of a kinetic model of the expression of the reporter gene, as explained in [Text S2](#). We used cubic smoothing splines (`csaps` function in MATLAB) to fit the fluorescence and absorbance data and obtain estimates of $A(t)$, $I(t)$, $dA(t)/dt$, and $dI(t)/dt$. The half-life of the GFPmut2 reporter is 19 h ($\gamma_r = 0.0006 \pm 0.0001$). The maturation time of GFPmut2 is so short (4 min, [65]) that it can be safely ignored.

A similar measurement model was used for the expression of the actual gene of interest, encoding a protein with concentration $p(t)$ [RFU min^{-1}]:

$$\frac{d}{dt}p(t) = f(t) - (\gamma_p + \mu(t))p(t), \quad p(0) = p_0, \tag{7}$$

where γ_p [min^{-1}] is the degradation constant of the protein. Notice that in the case of FlgM, protein degradation includes both physical degradation of the protein and secretion through the cell membrane. When the degradation constant is known, we can compute the protein concentration by numerical integration, starting from the initial concentration p_0 . This initial concentration is obtained from the reporter gene data, by realizing that the bacterial cells at the beginning of the experiment are rediluted cells from a preculture grown in the same medium. In particular, assuming that gene expression in the preculture is at steady-state, it follows from [Eq. 7](#) that

$$p(0) = p(T) = \frac{\mu(T) + \gamma_r}{\mu(T) + \gamma_p} r(T), \tag{8}$$

where $\mu(T)$ is the growth rate at the end of the preculture (at time T), $p(T)$ and $r(T)$ are the corresponding concentrations of the protein of interest and reporter protein, respectively. Usually, the bacteria in the preculture are in stationary phase, so $\mu(T) = 0$. [Eq. 8](#) was solved by numerical integration using the `quad` function in MATLAB.

In the case of the motility network there are two complications that slightly modify this general scheme. First, the half-lives of FliA and FlgM are variable over the time-course of the experiment. During exponential growth, when the motility genes are expressed, FliA and FlgM have short half-lives, due to proteolysis and secretion, respectively. During stationary phase, at

the end of the preculture, this is no longer the case and FliA and FlgM have longer half-lives (2 h for FliA and 3 h for FlgM [52, 53]). As a consequence, when computing the initial protein concentrations from the reporter concentrations at time T , we need to take protein degradation constants γ'_p corresponding to these longer half-lives. Second, in some experimental conditions, notably in rich medium like LB, the activity of the *fliA*, *flgM*, and *tar* promoters is negligible in the first few hours of the experiment [27]. As a consequence, the fluorescence intensity in the corresponding reporter strains is indistinguishable from the background fluorescence. We assume the promoter activity of the genes to be 0 in this case and back-extrapolate the observed promoter activities at earlier times towards 0. In Text S8 we illustrate the effects of variable half-lives and extrapolation of promoter activities on the computation of FliA and FlgM concentrations in a WT strain grown in LB.

For each of the derived quantities $r(t)$, $f(t)$, and $p(t)$, confidence intervals (defined as ± 2 standard errors of the mean) were computed from 6–7 experimental replicates.

Relative quantification of gene expression using real-time qRT-PCR

We verified the reporter gene measurements by means of qRT-PCR in the WT-LB condition, following a previously validated protocol [67]. Details of the experimental procedure can be found in Text S6.

Inference of minimal patterns of regulatory interactions

We use the method introduced in [42] to infer minimal pattern of regulatory interactions from time-series reporter gene data. The assumption of the method is that a regulator (*e.g.*, a transcription factor, but also the gene expression machinery) cannot operate both as a repressor and as an activator of a target gene, while it is allowed to operate as a repressor for one gene and as an activator for another gene. This corresponds to assuming that the activity of a gene is a monotone nondecreasing function of activators and a monotone nonincreasing function of repressors. Any such regulatory pattern can be encoded in terms of a sign pattern, *i.e.*, a vector containing one entry per regulator, taking value +1 for activators, -1 for repressors, and 0 for factors that do not affect the expression of the gene under consideration.

For every target gene, the method scans the measured promoter activities and concentrations of putative regulators. A sign pattern, *i.e.* a hypothesis on the regulatory structure, is rejected if it is found to be inconsistent with the data, *i.e.*, if measurements violate the monotonicity properties corresponding to that sign pattern. For instance, in the network module considered in this paper, the assumption that both FlgM and FliA activate *tar* can be rejected if any two measurement times are found such that, for higher concentrations of FlgM and FliA, the promoter activity of *tar* is lower. The algorithm makes the above verifications in a computationally efficient way and returns, for every target gene, a set of minimal sign patterns. The minimal sign patterns are regulatory patterns consistent with the data, having the properties that removal of any interaction results in an inconsistent pattern, whereas addition of a regulator (activator or repressor) preserves the consistency. In order to check the robustness of the minimal patterns thus obtained, we verified that no sign patterns were dismissed because of a single pair of measurements in the time-series. Mathematical details on the minimal sign pattern method can be found in Text S5 and [42].

Parameter estimation

The promoter activity models we considered in the main text have the form $f(t) = f(x(t), c)$, where c is a vector of unknown parameters and x is a vector of regressors. The specific form of

$f(x(t), c)$ is given in Eq. 1 and 3. The regressors take different forms in consecutive sections of this paper, consisting either of the activities f_A and f_M of the *fliA* and *flgM* promoters ($x = (f_A, f_M)$) or the reconstructed concentrations p_A and p_M of *FliA* and *FlgM* ($x = (p_A, p_M)$). In all sections, $c = (k_0, k_1, n, \theta, K)$, as defined in the section *Identification of gene regulation functions from promoter activities*. The superscript symbol s indicates the experimental condition, where $s \in \mathcal{S} = \{\Delta rpoS-M9, \Delta cpxR-M9, \Delta csgD-M9, \Delta csgD-LB, WT-LB\}$. Given measurements $(\bar{x}^s(t), \bar{f}^s(t))$ of $(x(t), f(t))$ (averages of 6–7 experimental replicates) at times $t \in \mathcal{T}^s$ along with confidence intervals $(\bar{f}^s(t) \pm \varepsilon^s(t))$ (computed from the same experimental replicates with ε^s equal to twice the standard error of the mean \bar{f}^s), we estimate c by solving the optimization problem

$$\hat{c} = \min_{c \in C} Q(c), \quad Q(c) = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}^s} \frac{1}{2\varepsilon^s(t)} |\bar{f}^s(t) - f(\bar{x}^s(t), c)|.$$

The solution is found in MATLAB using the multistart global search function `gs` with standard settings (interior-point method, `fmincon` for local minimizations). We tried several other global optimization function available in the MATLAB global search toolbox, but the `gs` function was found to perform best. The parameter search space C is given by the constraints

$$0 \leq k_0 \leq \max\{\bar{f}^s(t) : t \in \mathcal{T}^s, s \in \mathcal{S}\},$$

$$0 \leq k_1 \leq 10 \cdot \max\{\bar{f}^s(t) : t \in \mathcal{T}^s, s \in \mathcal{S}\},$$

$$1 \leq n \leq 4,$$

$$0 \leq \theta \leq \max\{\bar{x}_1^s(t) : t \in \mathcal{T}^s, s \in \mathcal{S}\},$$

$$0 \leq K \leq 20 \cdot \max\{\bar{x}_2^s(t) : t \in \mathcal{T}^s, s \in \mathcal{S}\}.$$

The above procedure applies to the estimation of the regulation function of both *tar* and *flgM*. However, for the estimation of the regulation function of *flgM*, the condition WT-LB is not available and hence excluded from the computation of $Q(c)$. Moreover, in the latter case, K is fixed for biological consistency to the value inferred from the fitting of *tar* promoter activity. For all parameter fits shown in the main text, we performed *a-posteriori* identifiability analysis to ensure that no structural or practical identifiability issue affects our results (Text S10). Overfitting issues were also excluded based on the results of this analysis and visual inspection of the fits.

Supporting Information

S1 Text. Strains and plasmids.

(PDF)

S2 Text. Measurement models for reporter gene data.

(PDF)

S3 Text. Data analysis.

(PDF)

S4 Text. Derivation of regulation function of motility genes.

(PDF)

S5 Text. Computation of minimal consistent sign patterns.
(PDF)

S6 Text. Validation of reporter gene data using qRT-PCR.
(PDF)

S7 Text. Additional data, analysis results, and fits.
(PDF)

S8 Text. Initial conditions for computing protein concentrations.
(PDF)

S9 Text. Parameter analysis: computation of active FliA.
(PDF)

S10 Text. Computation of parameter confidence intervals.
(PDF)

Acknowledgments

The authors would like to thank Julien Demol, Omayya Dudin, Nils Giordano, Stephan Lacour, and Delphine Ropers for providing bacterial strains, technical assistance, and discussions.

Author Contributions

Conceived and designed the experiments: DS EC JG HDJ. Performed the experiments: DS CP HDJ. Analyzed the data: DS EC JG HDJ. Contributed reagents/materials/analysis tools: CP SP. Wrote the paper: DS EC JG HDJ.

References

1. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. *Mol Syst Biol* 3. doi: [10.1038/msb4100120](https://doi.org/10.1038/msb4100120) PMID: [17299415](https://pubmed.ncbi.nlm.nih.gov/17299415/)
2. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp J, Blom J (2009) Systems biology: parameter estimation for biochemical models. *FEBS J* 276: 886–902. doi: [10.1111/j.1742-4658.2008.06844.x](https://doi.org/10.1111/j.1742-4658.2008.06844.x) PMID: [19215296](https://pubmed.ncbi.nlm.nih.gov/19215296/)
3. Bar-Joseph Z, Gitter A, Simon I (2012) Studying and modelling dynamic biological processes using time-series gene expression data. *Nat Rev Genet* 13: 552–64. doi: [10.1038/nrg3244](https://doi.org/10.1038/nrg3244) PMID: [22805708](https://pubmed.ncbi.nlm.nih.gov/22805708/)
4. de Smet R, Marchal K (2010) Advantages and limitations of current network inference methods. *Nat Rev Microbiol* 8: 717–29. doi: [10.1038/nrmicro2419](https://doi.org/10.1038/nrmicro2419) PMID: [20805835](https://pubmed.ncbi.nlm.nih.gov/20805835/)
5. Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R (2009) Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Biosystems* 96: 86–103. doi: [10.1016/j.biosystems.2008.12.004](https://doi.org/10.1016/j.biosystems.2008.12.004) PMID: [19150482](https://pubmed.ncbi.nlm.nih.gov/19150482/)
6. Marbach D, Costell J, Küffner R, Vega N, Prill R, et al. (2012) Wisdom of crowds for robust gene network inference. *Nat Methods* 9: 796–804. doi: [10.1038/nmeth.2016](https://doi.org/10.1038/nmeth.2016) PMID: [22796662](https://pubmed.ncbi.nlm.nih.gov/22796662/)
7. Villaverde A, Banga J (2014) Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *J R Soc Interface* 11: 20130505. doi: [10.1098/rsif.2013.0505](https://doi.org/10.1098/rsif.2013.0505) PMID: [24307566](https://pubmed.ncbi.nlm.nih.gov/24307566/)
8. Lu P, Vogel C, Wang R, Yao X, Marcotte E (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 25: 117–24. doi: [10.1038/nbt1270](https://doi.org/10.1038/nbt1270) PMID: [17187058](https://pubmed.ncbi.nlm.nih.gov/17187058/)
9. Taniguchi Y, Choi P, Li GW, Chen H, Babu M, et al. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329: 533–9. doi: [10.1126/science.1188308](https://doi.org/10.1126/science.1188308) PMID: [20671182](https://pubmed.ncbi.nlm.nih.gov/20671182/)
10. Bernstein J, Khodursky A, Lin PH, Lin-Chao S, Cohen S (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* 99: 9697–702. doi: [10.1073/pnas.112318199](https://doi.org/10.1073/pnas.112318199) PMID: [12119387](https://pubmed.ncbi.nlm.nih.gov/12119387/)

11. Larrabee K, Phillips J, Williams G, Larrabee A (1980) The relative rates of protein synthesis and degradation in a growing culture of *Escherichia coli*. *J Biol Chem* 255: 4125–30. PMID: [6989832](#)
12. Mosteller R, Goldstein R, Nishimoto K (1980) Metabolism of individual proteins in exponentially growing *Escherichia coli*. *J Biol Chem* 255: 2524–32. PMID: [6987224](#)
13. Maier T, Guell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett* 583: 3966–73. doi: [10.1016/j.febslet.2009.10.036](#) PMID: [19850042](#)
14. Picotti P, Clément-Ziza M, Lam H, Campbell D, Schmidt A, et al. (2013) A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* 494: 266–70. doi: [10.1038/nature11835](#) PMID: [23334424](#)
15. Trauchessec M, Jaquinod M, Bonvalot A, Brun V, Bruley C, et al. (2014) Mass spectrometry-based workflow for accurate quantification of *E. coli* enzymes: how proteomics can play a key role in metabolic engineering. *Mol Cell Proteomics* 13: 954–68. doi: [10.1074/mcp.M113.032672](#) PMID: [24482123](#)
16. Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139: 1366–75. doi: [10.1016/j.cell.2009.12.001](#) PMID: [20064380](#)
17. Berthoumieux S, de Jong H, Baptist G, Pinel C, Ranquet C, et al. (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol Syst Biol* 9: 634. doi: [10.1038/msb.2012.70](#) PMID: [23340840](#)
18. Gerosa L, Kochanowski K, Heinemann M, Sauer U (2013) Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol Syst Biol* 9: 658. doi: [10.1038/msb.2013.14](#) PMID: [23591774](#)
19. Keren L, Zackay O, Lotan-Pompan M, Barenholz U, Dekel E, et al. (2013) Promoters maintain their relative activity levels under different growth conditions. *Mol Syst Biol* 9: 701. doi: [10.1038/msb.2013.59](#) PMID: [24169404](#)
20. Regenberg B, Grotkjaer T, Winther O, Fausbøll A, Akesson M, et al. (2006) Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biol* 7: R107. PMID: [17105650](#)
21. Lovén J, Orlando D, Sigova A, Lin C, Rahl P, et al. (2012) Revisiting global gene expression analysis. *Cell* 151: 476–82. doi: [10.1016/j.cell.2012.10.012](#) PMID: [23101621](#)
22. Chudakov D, Matz M, Lukyanov S, Lukyanov K (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev* 90: 1103–63. doi: [10.1152/physrev.00038.2009](#) PMID: [20664080](#)
23. Ronen M, Rosenberg R, Shraiman B, Alon U (2002) Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci USA* 99: 10555–60. doi: [10.1073/pnas.152046799](#) PMID: [12145321](#)
24. Elowitz M, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403: 335–8. doi: [10.1038/35002125](#) PMID: [10659856](#)
25. de Jong H, Ranquet C, Ropers D, Pinel C, Geiselman J (2010) Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst Biol* 4: 55. doi: [10.1186/1752-0509-4-55](#) PMID: [20429918](#)
26. Liang S, Bipatnath M, Xu Y, Chen S, Dennis P, et al. (1999) Activities of constitutive promoters in *Escherichia coli*. *J Mol Biol* 292: 19–37. doi: [10.1006/jmbi.1999.3056](#) PMID: [10493854](#)
27. Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, et al. (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* 292: 2080–3. doi: [10.1126/science.1058758](#) PMID: [11408658](#)
28. Macnab R (1996) Flagella and motility. In: Neidhardt F, Curtiss R III, Ingraham J, Lin E, Low K, et al., editors, *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Washington, DC: ASM Press. 2nd edition, p. 12345.
29. Chevance F, Hughes K (2008) Coordinating assembly of a bacterial macromolecular machine. *Nat Rev Microbiol* 6: 455–465. doi: [10.1038/nrmicro1887](#) PMID: [18483484](#)
30. Macnab R (1996) Chemotaxis. In: Neidhardt F, Curtiss R III, Ingraham J, Lin E, Low K, et al., editors, *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Washington, DC: ASM Press. 2nd edition, p. 110329.
31. Wadhams G, Armitage J (2004) Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* 5: 1024–37. doi: [10.1038/nrm1524](#) PMID: [15573139](#)
32. Kutsukake K, Ohya Y, Iino T (1990) Transcriptional analysis of the flagellar regulon of *Salmonella typhimurium*. *J Bacteriol* 172: 741–7. PMID: [2404955](#)

33. Keseler I, Collado-Vides J, Santos-Zavaleta A, Peralta-Gi M, Gama-Castro S, et al. (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res* 39: D583–90. doi: [10.1093/nar/gkq1143](https://doi.org/10.1093/nar/gkq1143) PMID: [21097882](https://pubmed.ncbi.nlm.nih.gov/21097882/)
34. Mika F, Hengge R (2013) Small regulatory RNAs in the control of motility and biofilm formation in *E. coli* and *Salmonella*. *Int J Mol Sci* 14: 4560–79. doi: [10.3390/ijms14034560](https://doi.org/10.3390/ijms14034560) PMID: [23443158](https://pubmed.ncbi.nlm.nih.gov/23443158/)
35. Girgis H, Liu Y, Ryu W, Tavazoie S (2007) A comprehensive genetic characterization of bacterial motility. *PLoS Genet* 3: e154. doi: [10.1371/journal.pgen.0030154](https://doi.org/10.1371/journal.pgen.0030154) PMID: [17941710](https://pubmed.ncbi.nlm.nih.gov/17941710/)
36. Adler J, Templeton B (1967) The effect of environmental conditions on the motility of *Escherichia coli*. *J Gen Microbiol* 46: 175–84. doi: [10.1099/00221287-46-2-175](https://doi.org/10.1099/00221287-46-2-175) PMID: [4961758](https://pubmed.ncbi.nlm.nih.gov/4961758/)
37. Soutourina O, Kolb A, Krin E, Laurent-Winter C, Rimsky S, et al. (1999) Multiple control of flagellum biosynthesis in *Escherichia coli*: Role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. *J Bacteriol* 181: 7500–8. PMID: [10601207](https://pubmed.ncbi.nlm.nih.gov/10601207/)
38. Chilcott G, Hughes K (2000) Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar typhimurium and *Escherichia coli*. *Microbiol Mol Biol Rev* 64: 694–708. doi: [10.1128/MMBR.64.4.694-708.2000](https://doi.org/10.1128/MMBR.64.4.694-708.2000) PMID: [11104815](https://pubmed.ncbi.nlm.nih.gov/11104815/)
39. Kalir S, Alon U (2004) Using a quantitative blueprint to reprogram the dynamics of the flagella gene network. *Cell* 117: 713–20. doi: [10.1016/j.cell.2004.05.010](https://doi.org/10.1016/j.cell.2004.05.010) PMID: [15186773](https://pubmed.ncbi.nlm.nih.gov/15186773/)
40. Dudin O, Geiselmann J, Ogasawara H, Ishihama A, Lacour S (2014) Repression of flagellar genes in exponential phase by CsgD and CpxR, two crucial modulators of *Escherichia coli* biofilm formation. *J Bacteriol* 196: 707–15. doi: [10.1128/JB.00938-13](https://doi.org/10.1128/JB.00938-13) PMID: [24272779](https://pubmed.ncbi.nlm.nih.gov/24272779/)
41. Gillen K, Hughes K (1993) Transcription from two promoters and autoregulation contribute to the control of expression of the *Salmonella typhimurium* flagellar regulatory gene *flgM*. *J Bacteriol* 175: 7006–15. PMID: [7693654](https://pubmed.ncbi.nlm.nih.gov/7693654/)
42. Porreca R, Cinquemani E, Lygeros J, Ferrari-Trecate G (2010) Identification of genetic network dynamics with unate structure. *Bioinformatics* 26: 1239–45. doi: [10.1093/bioinformatics/btq120](https://doi.org/10.1093/bioinformatics/btq120) PMID: [20305266](https://pubmed.ncbi.nlm.nih.gov/20305266/)
43. Bundschuh R, Hayot F, Jayaprakash C (2003) Fluctuations and slow variables in genetic networks. *Biophys J* 84: 1606–15. doi: [10.1016/S0006-3495\(03\)74970-4](https://doi.org/10.1016/S0006-3495(03)74970-4) PMID: [12609864](https://pubmed.ncbi.nlm.nih.gov/12609864/)
44. Buchler N, Louis M (2008) Molecular titration and ultrasensitivity in regulatory networks. *J Mol Biol* 384: 1106–19. doi: [10.1016/j.jmb.2008.09.079](https://doi.org/10.1016/j.jmb.2008.09.079) PMID: [18938177](https://pubmed.ncbi.nlm.nih.gov/18938177/)
45. Ugray Z, Lasdon L, Plummer J, Glover F, Kelly J, et al. (2007) Scatter search and local NLP solvers: A multistart framework for global optimization. *INFORMS J Comput* 19: 32840. doi: [10.1287/ijoc.1060.0175](https://doi.org/10.1287/ijoc.1060.0175)
46. Berthoumieux S, Brilli M, Kahn D, de Jong H, Cinquemani E (2013) On the identifiability of metabolic network models. *J Math Biol* 67: 1795–832. doi: [10.1007/s00285-012-0614-x](https://doi.org/10.1007/s00285-012-0614-x) PMID: [23229063](https://pubmed.ncbi.nlm.nih.gov/23229063/)
47. Bremer H, Dennis P (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt F, Curtiss R III, Ingraham J, Lin E, Low K, et al., editors. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Washington, DC: ASM Press. 2nd edition, pp. 1553–69.
48. Maaløe O, Kjeldgaard N (1966) Control of Macromolecular Synthesis: A Study of DNA, RNA and Protein Synthesis in Bacteria. New York: W.A. Benjamin.
49. Klumpp S, Hwa T (2008) Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci USA* 105: 20245–50. doi: [10.1073/pnas.0804953105](https://doi.org/10.1073/pnas.0804953105) PMID: [19073937](https://pubmed.ncbi.nlm.nih.gov/19073937/)
50. Wanner B, Kodaira R, Neidhardt F (1977) Physiological regulation of a decontrolled *lac* operon. *J Bacteriol* 130: 212–22. PMID: [323228](https://pubmed.ncbi.nlm.nih.gov/323228/)
51. Aldridge P, Karlinsey J, Aldridge C, Birchall C, Thompson D, et al. (2006) The flagellar-specific transcription factor, σ^{28} , is the type III secretion chaperone for the flagellar-specific anti- σ^{28} factor FlgM. *Genes Dev* 20: 2315–26. doi: [10.1101/gad.380406](https://doi.org/10.1101/gad.380406) PMID: [16912280](https://pubmed.ncbi.nlm.nih.gov/16912280/)
52. Barembruch C, Hengge R (2007) Cellular levels and activity of the flagellar sigma factor FlhA of *Escherichia coli* are controlled by FlgM-modulated proteolysis. *Mol Microbiol* 65: 76–89. doi: [10.1111/j.1365-2958.2007.05770.x](https://doi.org/10.1111/j.1365-2958.2007.05770.x) PMID: [17537210](https://pubmed.ncbi.nlm.nih.gov/17537210/)
53. Karlinsey J, Tanaka S, Bettenworth V, Yamaguchi S, Boosa W, et al. (2000) Completion of the hook-basal body complex of the *Salmonella typhimurium* flagellum is coupled to FlgM secretion and *flhC* transcription. *Mol Microbiol* 37: 1220–31. doi: [10.1046/j.1365-2958.2000.02081.x](https://doi.org/10.1046/j.1365-2958.2000.02081.x) PMID: [10972838](https://pubmed.ncbi.nlm.nih.gov/10972838/)
54. Karlinsey J, Tsui H, Winkler M, Hughes K (1998) Flk couples *flgM* translation to flagellar ring assembly in *Salmonella typhimurium*. *J Bacteriol* 180: 5384–97. PMID: [9765570](https://pubmed.ncbi.nlm.nih.gov/9765570/)
55. Marbach D, Prill R, Schaffter T, Mattiussi C, D DF, et al. (2010) Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci USA* 107: 6286–91. doi: [10.1073/pnas.0913357107](https://doi.org/10.1073/pnas.0913357107) PMID: [20308593](https://pubmed.ncbi.nlm.nih.gov/20308593/)

56. Scott M, Hwa T (2011) Bacterial growth laws and their applications. *Curr Opin Biotechnol* 22: 559–65. doi: [10.1016/j.copbio.2011.04.014](https://doi.org/10.1016/j.copbio.2011.04.014) PMID: [21592775](https://pubmed.ncbi.nlm.nih.gov/21592775/)
57. Dennis P, Ehrenberg M, Bremer H (2004) Control of rRNA synthesis in *Escherichia coli*: a systems biology approach. *Microbiol Mol Biol Rev* 68: 639–68. doi: [10.1128/MMBR.68.4.639-668.2004](https://doi.org/10.1128/MMBR.68.4.639-668.2004) PMID: [15590778](https://pubmed.ncbi.nlm.nih.gov/15590778/)
58. Travers A, Muskhelishvili G (2005) DNA supercoiling—a global transcriptional regulator for enterobacterial growth? *Nat Rev Microbiol* 3: 157–69. doi: [10.1038/nrmicro1088](https://doi.org/10.1038/nrmicro1088) PMID: [15685225](https://pubmed.ncbi.nlm.nih.gov/15685225/)
59. Huang Z, Senocak F, Jayaraman A, Hahn J (2008) Integrated modeling and experimental approach for determining transcription factor profiles from fluorescent reporter data. *BMC Syst Biol* 2: 64. doi: [10.1186/1752-0509-2-64](https://doi.org/10.1186/1752-0509-2-64) PMID: [18637177](https://pubmed.ncbi.nlm.nih.gov/18637177/)
60. Leveau J, Lindow S (2001) Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J Bacteriol* 183: 6752–62. doi: [10.1128/JB.183.23.6752-6762.2001](https://doi.org/10.1128/JB.183.23.6752-6762.2001) PMID: [11698362](https://pubmed.ncbi.nlm.nih.gov/11698362/)
61. Wang X, Errede B, Elston T (2008) Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters. *Biophys J* 94: 2017–26. doi: [10.1529/biophysj.107.122200](https://doi.org/10.1529/biophysj.107.122200) PMID: [18065460](https://pubmed.ncbi.nlm.nih.gov/18065460/)
62. Subramanian S, Srien F (1996) Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J Bacteriol* 49: 137–51.
63. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci USA* 103: 13004–9. doi: [10.1073/pnas.0605420103](https://doi.org/10.1073/pnas.0605420103) PMID: [16916930](https://pubmed.ncbi.nlm.nih.gov/16916930/)
64. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2: 2006.0008. doi: [10.1038/msb4100050](https://doi.org/10.1038/msb4100050) PMID: [16738554](https://pubmed.ncbi.nlm.nih.gov/16738554/)
65. Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, et al. (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* 3: 623–8. doi: [10.1038/nmeth895](https://doi.org/10.1038/nmeth895) PMID: [16862137](https://pubmed.ncbi.nlm.nih.gov/16862137/)
66. Miller J (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: CSHL Press.
67. Lee C, Kim J, Shin S, Hwang S (2004) Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. *J Biotechnol* 123: 273–80. doi: [10.1016/j.jbiotec.2005.11.014](https://doi.org/10.1016/j.jbiotec.2005.11.014)