Reconstructing repressor protein levels from expression of gene targets in *Escherichia coli*

R. Khanin*†, V. Vinciotti‡, and E. Wit§

*Department of Statistics, University of Glasgow, Glasgow G12 8QW, United Kingdom; ‡School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge UB8 3PH, United Kingdom; and §Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, United Kingdom

Edited by Philip P. Green, University of Washington School of Medicine, Seattle, WA, and approved September 28, 2006 (received for review May 5, 2006)

The basic underlying problem in reverse engineering of gene regulatory networks from gene expression data is that the expression of a gene encoding the regulator provides only limited information about its protein activity. The proteins, which result from translation, are subject to stringent posttranscriptional control and modification. Often, it is only the modified version of the protein that is capable of activating or repressing its regulatory targets. At present there exists no reliable high-throughput technology to measure the protein activity levels in real-time, and therefore they are, so-to-say, lost in translation. However, these activity levels can be recovered by studying the gene expression of their targets. Here, we describe a computational approach to predict temporal regulator activity levels from the gene expression of its transcriptional targets in a network motif with one regulator and many targets. We consider an example of an SOS repair system, and computationally infer the regulator activity of its master repressor, LexA. The reconstructed activity profile of LexA exhibits a behavior that is similar to the experimentally measured profile of this repressor: after UV irradiation, the amount of LexA substantially decreases within a few minutes, followed by a recovery to its normal level. Our approach can easily be applied to known single-input motifs in other organisms.

Michaelis–Menten kinetics | statistical reconstruction | transcrtiption factor activity

The changes in expression of *Escherichia coli* genes as a result of DNA damage, collectively named the SOS response, have been the subject of numerous experimental and theoretical studies (1–5). The general picture of the SOS response has slowly started to crystallize, but the precise kinetics and complex regulation of the processes involved can still generate surprises after 30 years of intensive study (6).

The SOS system includes >30 genes controlled at the transcriptional level by the transcriptional repressor protein LexA (2). Under normal conditions, the level of LexA is high and the SOS genes are suppressed. Upon DNA damage, one of the SOS proteins, RecA, binds to regions of single-stranded DNA that are produced as a consequence of this damage and becomes conformationally active. The active form of RecA facilitates the inactivation of the LexA repressor. As a result, the level of LexA diminishes, thereby causing the up-regulation of the genes that are suppressed by LexA under normal conditions. Because the damage has been repaired, the level of the activated RecA drops, LexA accumulates, and a decrease in the activation of SOS genes is observed.

It has recently been reported that the SOS response is "digital" (5): the number of pulses but not their amplitude increases with the level of DNA damage. This finding indicates an additional level of regulation of the master repressor, LexA, possibly executed by a product of another target gene *umuD* (5). In fact, it has been suggested that the proteins and regulatory systems involved in sensing the damage, transducing the signal, and implementing and relieving the checkpoint are heavily intertwined (7).

An interesting feature of the LexA/RecA regulatory module is that the timing, duration, and level of activation varies for each of the LexA-regulated genes (2, 3, 5). Ronen *et al.* (3) developed a combined experimental and theoretical approach based on accurate high temporal-resolution measurements of promoter activities using GFP technology. They applied this approach to several of the SOS genes and found a strikingly detailed temporal program of expression. Based on the accurate measurements of the promoter activities of several SOS genes, these authors computed the temporal profile of relative levels of the transcriptional regulator, LexA.

In this work, we reconstruct the activity level of the transcriptional repressor, LexA, from the expression time profiles of its target genes. The LexA profile, reconstructed by our method from microarray data, shows good correspondence with the LexA profiles measured by immunoblots technology (8) and computed from the measurements of promoter activities (3).

Results

Model Description. In this work, we aim to reconstruct the activity of the master regulator in a single input motif (SIM) from gene expression data of its targets. The computational approach involves embedding a set of differential equations that describe kinetics of gene regulation, within a statistical noise model and to recover the kinetic parameters by a maximum likelihood approach. The model equations used in the present paper are given in *Materials and Methods*. Here, we describe the main assumptions of our modeling approach.

A SIM is a simple network architecture that frequently occurs in gene regulatory networks of different organisms (9). In a SIM, there is one transcription factor, either an activator or repressor, that regulates (either activates or represses) the transcription of several target genes. The targets are not controlled by other regulators. Therefore, the transcription of each target gene depends on the regulator profile that is common to all genes and on the kinetic parameters of regulation, that are presumably gene specific. The SOS repair system regulated by the repressor LexA is an example of a SIM (Fig. 1).

The dynamics of changes in expression of a target gene are determined by the gene's transcription rate, which depends on the promoter activity, and by the mRNA degradation. Previous approaches to infer the activity of a transcription factor assumed that the kinetics of gene transcription is adequately described by linear or log-linear models (10–14). However, it has been noted (15) that gene transcription regulated by a transcription factor resembles the process of enzyme-mediated reactions. The latter process has been intensively studied and is known to exhibit

Author contributions: R.K., V.V., and E.W. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: SIM, single input motif; MM, Michaelis–Mentin; ODE, ordinary differential equation.

†To whom correspondence should be addressed. E-mail: raya@stats.gla.ac.uk.

www.pnas.org/cgi/doi/10.1073/pnas.0603390103

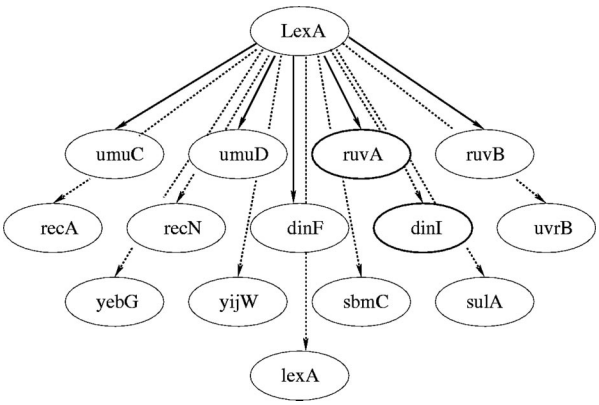

Fig. 1. Schematic drawing of the SOS single input motif, regulated by the LexA repressor protein. Fourteen genes are differentially expressed between wild type and mutant samples at a *P* value cutoff of 0.01. These genes are therefore targets of LexA repressor, and together they constitute single input motif.

saturation effects, which are captured by a so-called Michaelis–Menten (MM) ordinary differential equation (ODE).

Based on the analogy between enzymatic mediated reactions and regulator mediated transcription, gene regulation has been modeled previously by a MM kinetics model (3, 15, 16). Qualitatively, the MM model implies that in the case of the activator, higher levels of transcription factor yield a higher transcription rate of a regulated gene, until the transcription rate saturates. Raising level of transcription factor beyond a certain threshold value will no longer result in a higher production of mRNA. For the repressor, higher levels of transcription factor result in a lower transcription rate, until the gene is completely suppressed, i.e., mRNA production is at its basal level. Linear models of regulation cannot account for such saturation effects. Nor can linear models account for the case of repressor regulator.

It is well known that microarray data are noisy. To account for the noise in the observed gene expression data, we embed the MM model in a statistical framework. Because the terms of the MM model can be traced back to the actual biological processes of protein binding and mRNA degradation, we choose to use the expression data on its original scale, rather than the log-scale that is commonly used in microarray analysis. We expect the noise to depend on the amount of transcription. This means that low levels of gene expression have less noise and higher levels of gene expression have more noise. A log-normal distribution accommodates variance inflation at higher expression levels.

The profile of the master regulator that is common to all genes in the SIM regulatory module, and the gene-specific kinetic parameters of regulation, are inferred from the gene expression data by a maximum likelihood approach. The maximum likelihood estimates of the model parameters are computed by a conjugate gradient method. For further details, see supporting information, which is published on the PNAS web site.

Data. In the *E. coli* SOS repair system, the master repressor LexA regulates a SIM, which consists of several targets shown in Fig. 1. The temporal changes in expression of *E. coli* genes, caused by irradiation on a genome-wide level, have been studied in ref. 2. These authors examined the changes in gene expression after UV exposure (40 J/m^2) in both wild-type cells and *lexA1* mutants, which are unable to induce genes under LexA control. Their data are publicly available at http://genome-www.stanford.edu/Uvirradiation.

The data contain two times series at six time points (0, 5, 10, 20, 40, and 60 min), one for wild-type and one for mutant. For the use in further analysis, raw data have been normalized to account for spatial, dye, and across-array effects using the allnorm function from the smida R package (17), which is available for download at www.stats.gla.ac.uk/~microarray/book/smida.html. The data were normalized on the log-scale (see supporting information).

LexA Single Input Motif Module. Among >30 genes that are known to contain the LexA-binding boxes (ref. 2; full list of genes that we used is available), 20 genes are present on all arrays in the current data set. To find which of those 20 genes constitute the targets of LexA repressor, we looked for differentially expressed genes between wild and mutant types. We used an additive factorial model for estimating the time and UV exposure effects. The model has 4 degrees of freedom (2×6 observations minus 1+1+5 main effect parameters) to test for a significant UV exposure effect. The number of differentially expressed genes identified depends on the stringency of the comparisons. We select a cutoff *P* value of 0.01, which results in discovering 14 genes as potential SIM targets. This cutoff corresponds to a kink in the *P* value plot (supporting information) and to a corresponding false discovery rate of 1.4%, which on average would almost certainly result at most 1 false discovery. Because it is not known which of the genes might have been falsely discovered, we have performed a sensitivity analysis by iteratively leaving one of the putative targets out. The LexA profiles reconstructed by using each of the subsets of 13 genes shows a very good agreement with the profile reconstructed using all 14 targets (Fig. 2*b*).

Further computations are performed with 14 target genes. Their expression profiles have now been returned to original scale for further kinetic modeling and optimization.

Our analysis confirms the conclusion of ref. 2 that the expressions of some of the genes that are documented to be LexA-regulated genes do not rise significantly following UV irradiation. These genes include *dinG*, *molR*, *uvrD*, *uvrA*, *hokE*, and *ssb*. For at least two of these genes, *uvrA* and *uvrD*, a rise in the promoter activities after lower levels of UV irradiation has been recorded (3). Possible reasons for this discrepancy between known behavior of some genes confirmed by promoter activity measurements and gene expression data can be due to differences in the experimental set-ups, as discussed in ref. 2. Alternative reasons may include the fact that the measurements performed over a population of cells might be limited in their ability to accurately describe the network responses in the case of a nonhomogenous population or an unsynchronized response (5). Genes that were not differentially expressed after UV irradiation were not included in the LexA SIM (Fig. 1) for the present study.

Most of the differentially expressed genes (e.g., *dinF*, *dinI*, *lexA*, *ruvA*, *ruvB*, *yebG*) in the LexA regulatory module show significant up-regulation after UV radiation and subsequent decline in the gene expression levels. This is very much in line with the measured promoter activities of the target genes (3). The expression profiles of other differentially expressed genes exhibit a different behavior pattern: some genes (*recN*, *uvrB*, *yijW*) do not decline after reaching their respective peaks after UV radiation, whereas others (*umuC*) reach their maximum levels rather slowly (Fig. 3).

It has recently been reported that the promoter activities of three genes (*lexA*, *recA*, and *umuD*) exhibit more than one peak after irradiation of 20 J/m^2 and 50 J/m^2 (5). At 20 J/m^2, the second peak of *recA* promoter activity occurs at 57 ± 3 min. Observations on single cells for lower levels of irradiation show that 60% of cells show a single peak at a UV dose of 10 J/m^2 (5). These authors concluded that the number of peaks increases with irradiation. Because the microarray studies of Courcelle *et al.* (2) were conducted at a UV dose of 40 J/m^2 and the last

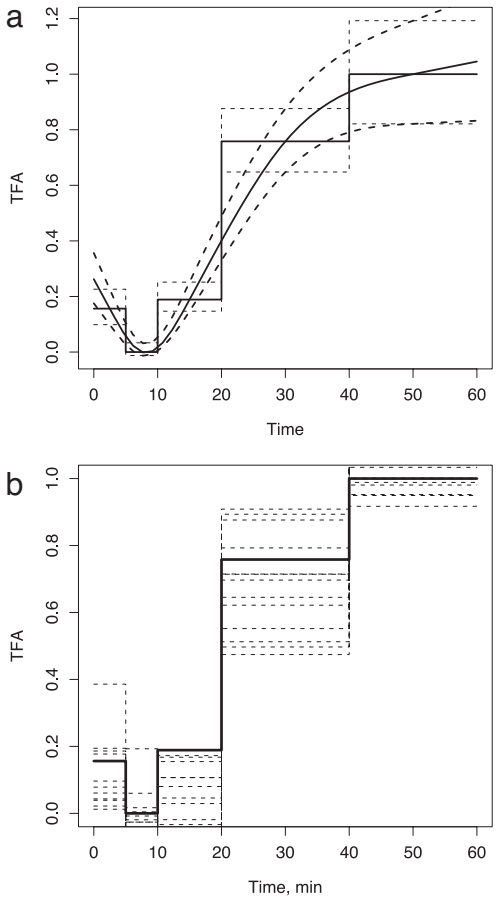

Fig. 2. Reconstructed activity level of master repressor LexA, following a UV dose of 40 J/m^2. (*a*) The activity profile of LexA approximated by piece-wise constant function (bold solid line), is reconstructed from the MM kinetic model (Eqs. **1**–**4**, *Methods*) using microarray data of 14 target genes in the LexA SIM (Fig. 1). The LexA profile is estimated by the maximum likelihood method. 95% confidence bounds (dashed lines) are computed independently for each component of the LexA profile, while keeping all other parameters fixed. LexA profile is rescaled between 0 and 1. Confidence bounds are rescaled accordingly. Time is given in minutes as in the experiment. The smoothed LexA profile (solid) is obtained by cubic spline interpolation from the reconstructed piece-wise constant profile and its 95% confidence bounds, using the R function spline. TFA, transcription factor activity. (*b*) LexA approximated by piece-wise constant function (solid line) is the same as in *a*. LexA profiles estimated by the maximum likelihood method by iteratively leaving each one of the putative targets out (dashed lines). The mean correlation between the LexA profile reconstructed from expression of 14 genes and profiles reconstructed from each subset of 13 target genes is 0.84.

measurement was taken at 60 min, the observations that some of the gene profiles stayed constant or went up at ≈60 min after 40 J/m^2 are consistent with the appearance of the second peak. The fact that only some genes show a build-up to the second peak might indicate an additional level of regulation of these genes (3, 5) as well as different kinetic constants of regulation. Different decay constants could also be partly responsible for the fact that some genes reach their normal level after the first peak, whereas others stay activated for longer times.

Reconstructed LexA Activity Profile. The reconstruction procedure is based on finding the maximum likelihood for the whole SIM (Fig. 1) using the kinetic model (Eqs. **1**–**4**) and gene expression data of the 14 targets (taken on original scale). Reconstruction yields the profile of the master regulator (LexA) that is common to all target genes and maximum likelihood estimates for the gene-specific kinetic parameters of regulation. The 14 targets are the genes that have been identified as differentially expressed SOS genes between the wild and mutant types. Reconstruction is based on R function optim() that performs the maximization of the likelihood with constraints, using a conjugate gradient method. Our search procedure uses multiple starts for locating the maximum with respect to all unknown parameters of interest, that is transcription factor activity profile and gene-specific kinetic parameters.

Fig. 2 shows the maximum likelihood reconstructed profile of the LexA repressor, approximated by the piece-wise constant function on each sampling interval that is determined by available microarray measurements as discussed in *Materials and Methods* (Eq. **2**). The LexA profile reconstructed for the SIM with 14 targets is shown on Fig. 2*a* (bold solid line) with 95% confidence bounds (dashed lines; computed via a classical Wilks procedure). The activity of LexA, here referred to as transcription factor activity, is expressed in arbitrary units and can be interpreted as relative "levels." The activity of a protein is generally determined by its relative level and the appropriate kinetic rate constant. Fig. 2*a* also shows the smoothed LexA profile obtained by a cubic spline interpolation (R function spline) of the reconstructed piece-wise constant function.

Crucially, the reconstructed profile of LexA exhibits behavior that is similar to the experimentally observed profiles in refs. 3 and 8. After UV irradiation, the amount of LexA substantially decreases (>10-fold) within a few minutes (5–10 min) (6). This reduction is followed by a recovery phase, wherein LexA goes back to its levels under normal conditions. The reconstructed LexA profile in Fig. 2 exhibits a faster recovery phase compared with the experimentally predicted rise from measurements of promoter activities (3). The reconstructed LexA profile recovers in 40 min compared with ≈60 min predicted from experiments (3, 8). In addition, reconstructed LexA profile goes to the level that is higher than its initial level. This is, perhaps, a reflection of the experimental data, wherein the expression levels of some target genes (e.g., *recN* and *umuC*, see Fig. 3) do not show any sign of a decline of their expressions 60 min after UV irradiation. It might also suggest that a standard assumption of linear degradation has to be reevaluated.

Given the limitations of the current microarray data set (2), which include (*i*) averaging gene expressions from nonhomogenous cells, (*ii*) using potentially different experimental conditions and protocols from promoter activity measurements in ref. 3, and (*iii*) the limited available data (only six time points available), the LexA profile reconstructed by our method exhibits a very good similarity with the profile computed from precise measurements of the promoter activities.

Interpretation of Kinetic Parameters. A very structured temporal order of activation of SOS genes has been reported both in microarray studies (2) and promoter activity measurements (3). Based on available microarray data, we have reconstructed the kinetic profiles of all target genes in the SIM regulatory module in Fig. 1. The kinetic profiles of four representative genes (*lexA*, *ruvB*, *recN*, and *umuC*) are shown in Fig. 3. The reconstructed profiles (solid lines) of these four genes show excellent fit with the data (points). It is interesting to note that the noise tends to be higher at higher levels of expression: compare profiles of *recN* and *umuC* genes with the profiles of *lexA* and *ruvB* genes in Fig. 3. This finding is in good agreement with the log-normal assumption (Eq. **3**). Kinetic profiles of other target genes in the LexA module can be found in supporting information.

Four genes, represented in Fig. 3, exhibit very different behavior. Genes *lexA* and *ruvB* show up-regulation at ≈10 min followed by a subsequent decline, whereas profiles of two other genes (*recN* and *umuC*) do not decline during 60 min after UV radiation. Estimated kinetic rate constants for the latter two

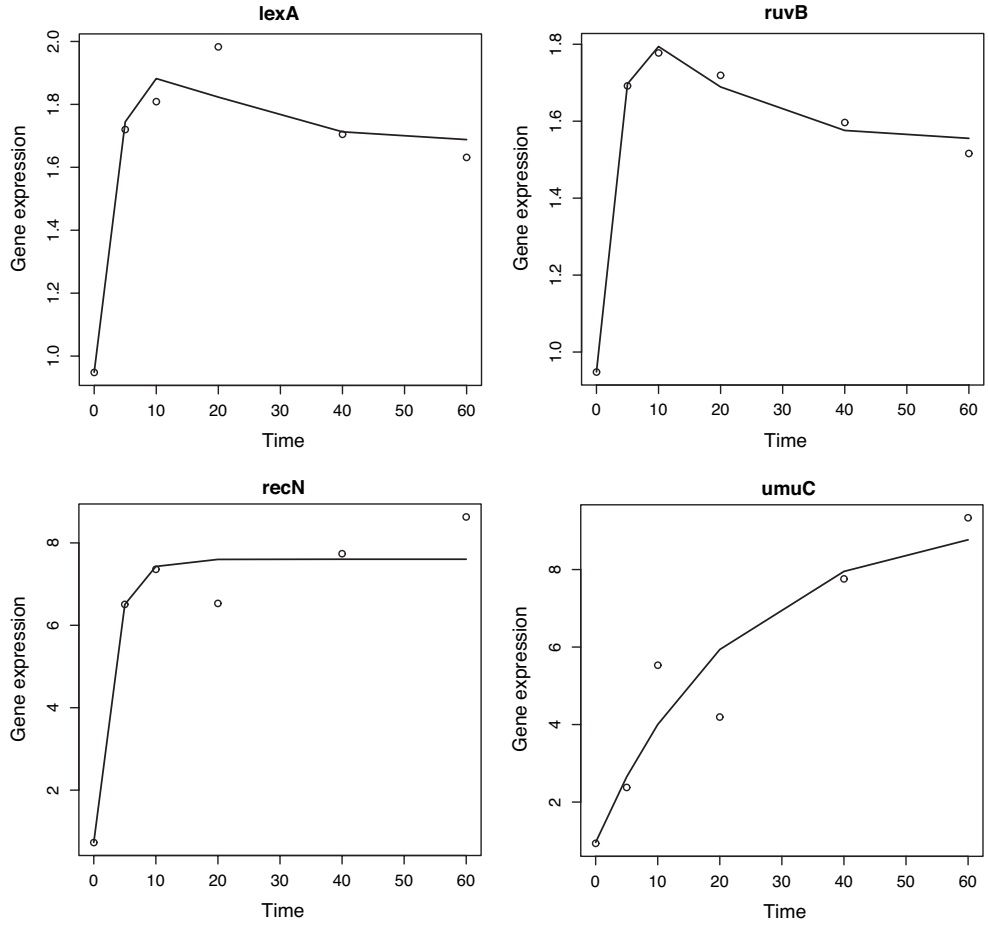

Fig. 3. Reconstructed profiles for four genes in the LexA SIM. Points stand for the data values. Data and reconstructed profiles are presented on the original scale. Time is given in minutes as in the experiment. Profiles of other target genes in the SIM are available on request.

genes show that this type of behavior can actually be achieved by linear degradation alone: $\dot{\mu} \approx \alpha + \delta\mu$. This finding clearly indicates that, within the first 60 min after radiation, the basal transcription rate, α, and decay play important roles in the mRNA levels of the *recN* and *umuC* genes.

From the MM kinetic reconstruction, it is possible to evaluate qualitatively the relative numerical values of the gene specific kinetic parameters. Ranking the other 12 genes by their effective production rate, $r = \beta/(\gamma + \bar{\eta})$, results in *recA* being the fastest gene ($r \approx 18.2$) and *uvrB* being the slowest one ($r \approx 1.4$). The other 10 genes can be ranked in the following order (from the fastest to the slowest): *sbmC*, *umuD*, *yebG*, *ruvB*, *sulA*, *ruvA*, *yjiW*, *lexA*, *dinI*, *dinF*.

For three genes (*dinF*, *uvrB*, and *ruvA*) the estimates of γ are very small, resulting in a transcription rate inversely proportional to the LexA protein levels. For five genes (*recA*, *lexA*, *sbmC*, *dinI*, and *yebG*), estimates of γ are comparable with the $\bar{\eta}$, indicating that saturation in production rate occurs for already moderate levels of LexA. Degradations rates, δ, for target genes vary by two orders of magnitude, with *umuC* having the slowest rate and *ruvA* being the fastest one to degrade. It is also interesting to note that estimates for basal level of production, α, were found to be negligible for many genes with an exception of *recN*, *uvrB*, and *umuC*.

Reconstructed profiles of most of the target-genes show excellent fit with the data (see Fig. 3 and supporting information). However, the fit for four genes (*dinI*, *sulA*, *umuD*, *yjiW*) is not particularly good, indicating that either the data are too noisy or there might be nonlinear degradation and cooperativity in the production term. Although it might be tempting to try to interpret the absolute values of the kinetic parameters directly, this is typically not possible because there are no explicit units in the microarray measurements on which all of the estimates are based. However, relative comparisons of such values, as we have done above, are meaningful.

Conclusions

We have demonstrated that the protein level of a regulator can be inferred from microarray data measured on its target genes. Using a microarray time course experiment on wild and mutant types of *E. coli* after UV irradiation, we have successfully reconstructed the activity of the repressor LexA from its target genes. The reconstructed profile is a piece-wise constant function on each time interval, determined by available microarray measurements. The piece-wise nature of the reconstructed profile might seem highly approximative; however, our method is highly flexible because it does not make any parametric assumptions and it permits a closed-form solution of the integration of the kinetic equations (see Eq. **2**). At the same time, it is a parsimonious way to deal with the limited amount of data. The reconstructed profile and its confidence bands can be smoothed with an interpolation spline, yielding a profile without discontinuities at the sampling time points.

Our method enables the reconstruction of the temporal profile of a regulatory protein given the gene expressions of its targets. It has wide applicability because SIMs frequently occur

in many organisms. Once all or a subset of the target genes in the SIM are known, and such information is rapidly becoming available from ChIP–Chip experiments (18), the activity profile of the master regulator can be reconstructed. This approach can be used to reconstruct the activity of an activator, as demonstrated in an example of the cdaR activator in *Streptomyces* (19), or of a repressor, as in the present paper.

Materials and Methods

The SOS regulatory module is an example of a SIM (9) where one regulator, a repressor in our case, controls several target genes. The activity profile of the regulator, LexA, is denoted by $\eta(t)$. The expression of a target gene k, $\mu_k(t)$, changes due to mRNA production, which depends on the level of the regulator $\eta(t)$ and mRNA degradation. The rate of the transcription production of a gene k is modeled by the MM kinetics, $\beta_k/(\gamma_k + \eta(t))$, where β_k and γ_k are gene-specific kinetic parameters for gene k. This equation is widely used for modeling processes in enzyme mediated kinetics. The kinetics of expression of gene k is described by

$$\dot{\mu}_k(t) = \alpha_k + \beta_k \frac{1}{\gamma_k + \eta(t)} - \delta_k \mu_k(t), \qquad [1]$$

where δ_k is the rate of linear mRNA degradation. The additive constant α_k accounts for the basal level of transcription as well as for nuisance effects from microarrays.

Given measurements of gene expression at N time points (t_0, $t_1, \ldots, t_{N-1}$), the temporal profile of a gene k, $\mu_k(t)$, that solves the ODE in Eq. **1** can be approximated by

$$\mu_k(t) = \mu_k^0 e^{-\delta_k t} + \frac{\alpha_k}{\delta_k} + \beta_k e^{-\delta_k t} \frac{1}{\delta_k} \sum_{j=0}^{N-2} (e^{\delta_k t_{j+1}} - e^{\delta_k t_j}) \frac{1}{\gamma_k + \bar{\eta}_j}, \qquad [2]$$

where $\bar{\eta}_j = (\eta(t_j) + \eta(t_{j+1}))/2$ on each subinterval (t_j, t_{j+1}), $j = 0, \ldots, N-2$. This is under the simplifying assumption that $\eta(t)$ is a piece-wise constant function on each subinterval (t_j, t_{j+1}). One can come up with linear (or higher order) $\eta(t)$ approximations on each subinterval. This will introduce additional parameters, which will be impossible to infer with any certainty given limited amount of data.

The observed gene expression of a target gene k, taken on its original scale, is assumed to be log-normally distributed

$$g_k(t) \approx \text{lognorm}(m_k(t), \sigma_k^2), E(g_k(t)) = \mu_k(t), \qquad [3]$$

where the location parameter is $m_k(t) = \log[\mu_k(t)] - 1/2\sigma_k^2$.

The kinetic parameters $\theta_k = \{\alpha_k, \beta_k, \gamma_k, \delta_k, \mu_0^k\}$ and the variance of the log-normal distribution, σ_k^2, are assumed to be gene-specific. The activity profile $\eta = \{\eta_1, \ldots, \eta_N\}$ of the repressor LexA, and the parameters of regulation for each target gene are sought by maximizing the overall likelihood of the SIM

$$L_{\text{overall}}(\Theta, \Sigma^2, \eta) = \prod_{k=1}^{K} L_k(g_k(t); \theta_k, \sigma_k^2, \eta). \qquad [4]$$

Here, Θ represents all of the gene-specific kinetic parameters of the kinetic model, θ_k, for all targets $k = 1 \ldots K$ in the regulatory module; Σ^2 stands for all of the scale-parameters of the log-normal distribution, σ_k^2; and $L_k(g_k(t); \theta_k, \sigma_k^2, \eta)$ is the gene-specific likelihood of a gene k given the observed data $g_k(t)$ and the transcription factor activity η. The computational method that maximizes the log-likelihood with respect to η, Θ, and Σ for all genes in the SIM is based on a conjugate gradient method. For a SIM, we have six kinetic parameters per gene to estimate and additional $N - 1$ parameters for the regulator profile $\bar{\eta}$.

Confidence intervals for the maximum likelihood estimate of η ($\hat{\eta}$) are calculated via a classical Wilks method. Each component of $\hat{\eta}$ is perturbed, while keeping all other parameters fixed, such that the marginal confidence decreases by 95%, which corresponds approximately to a decrease in the likelihood by $2\chi^2_{1,095}$. The smoothed η -profile (bold solid line on Fig. 2*A*) is obtained by the cubic spline interpolation method (R function spline).

All of the code for estimating the parameters, calculating the confidence intervals and plotting the figures has been implemented in the statistical language R (www.r-project.org) are available upon request and can be obtained from authors.

1. Aksenov S, Krasavin E, Litvin A (1997) *J Theor Biol* 186:251–260.
2. Courcelle J, Khodursky A, Peter B, Brown P, Hanawalt P (2001) *Genetics* 158:41–64.
3. Ronen M, Rosenberg R, Shraiman B, Alon U (2002) *Proc Natl Acad Sci USA* 99:10555–10560.
4. Gardner T, deBernardo D, Lorentz D, Collins JJ (2003) *Science* 301: 102–105.
5. Friedman N, Vardi S, Ronen M, Alon U, Stavans J (2005) *PLoS Biol* 7:e238.
6. Michel B (2005) *PLOS* 3:1174–1176.
7. Opperman T, Murli S, Smith B, Walker G (1999) *Proc Natl Acad Sci USA* 96:9218–9223.
8. Sassanfar M, Roberts J (1990) *J MolBiol* 212:79–96.
9. Shen-Orr S, Milo R, Mangan S, Alon U (2002) *Nat Genet* 31:64–68.
10. Kao K, Yang Y, Boscolo R, Sabatti C, Roychowdhury V, Liao J (2004) *Proc Natl Acad Sci USA* 101:641–646.
11. Boulesteix A-L, Strimmer K (2005) *Theor Biol Med Model* 2:23–35.
12. Tran L, Brynildsen M, Kao K, Suen J, Liao J (2005) *MetabEng* 7:128–141.
13. Yang Y, Suen J, Brynildsen M, Galbraith S, Liao J (2005) *BMC Genomics* 6:90.
14. Yang Y, Liao J (2005) *Metab Eng* 7:280–290.
15. Bolouri H, Davidson E (2002) *BioEssays* 24:1118–1129.
16. Nachman I, Regev A, Friedman N (2004) *Bioinformatics* 20:I248–I256.
17. Wit E, McClure J (2004) *Statistics for Microarrays: Design, Analysis and Inference* (Wiley, Chichester, UK).
18. Lee T, Rinaldi N, Robert F, Odom D, Bar-Joseph Z, Gerber G, Hannett N, Harbison C, Thompson C, Simon I, *et al.* (2002) *Science* 298:799–804.
19. Khanin R, Vinciotti V, Mersinias V, Smith C, Wit E (2006) *Biometrics*, in press.