Published in IET Systems Biology
Received on 22nd January 2008
Revised on 12th May 2008
doi: 10.1049/iet-syb:20080097

Special Issue – Selected papers from the First q-bio Conference on Cellular Information Processing

ISSN 1751-8849

Serially regulated biological networks fully realise a constrained set of functions

A. Mugler[1] *E. Ziv*[2] *I. Nemenman*[3] *C.H. Wiggins*[4]

[1]*Department of Physics, Columbia University, New York, NY 10027, USA*
[2]*College of Physicians and Surgeons, Columbia University, New York, NY 10027, USA*
[3]*Computer, Computational and Statistical Sciences Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[4]*Department of Applied Physics and Applied Mathematics, Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10027, USA*
E-mail: ajm2121@columbia.edu

Abstract: It is shown that biological networks with serial regulation (each node regulated by at most one other node) are constrained to direct functionality, in which the sign of the effect of an environmental input on a target species depends only on the direct path from the input to the target, even when there is a feedback loop allowing for multiple interaction pathways. Using a stochastic model for a set of small transcriptional regulatory networks that have been studied experimentally, it is further found that all networks can achieve all functions permitted by this constraint under reasonable settings of biochemical parameters. This underscores the functional versatility of the networks.

1 Introduction

A driving question in systems biology in recent years has been the extent to which the topology of a biological network determines or constrains its function. Early works have suggested that the function follows the topology [1–4], and this continues as a prevailing view even though later analyses (at least in a small corner of biology) have questioned the paradigm [5, 6]. It remains unknown if a small biochemical or regulatory network can perform multiple functions, and whether the function set is limited by the network's topological structure. To this extent, in this paper, we develop a mathematical description of the functionality of a certain type of biological networks and show that the answer to both questions is 'yes': the networks can perform many, but not all possible functions, and the set of attainable functions is constrained by the topology. We illustrate these results in the context of an experimentally realised system [1].

Following [1] and our earlier work [6], we focus on the steady-state functionality of transcriptional regulatory networks. In this case, the input is the 'chemical environment,' that is a binary vector of presence/absence of small molecules that affect the regulation abilities of the transcription factors; and the output is the steady-state expression of a particular gene, hereafter called the reporter. Different functions of the network correspond then to different ways to map the small molecule concentrations into the reporter expression.

In our setup, the effect of introducing a small molecule S_j specific to a transcription factor X_j is to modify the affinity of X_j to its binding site. Equivalently, one can think of S_j as modulating or renormalising the transcription factor concentration X_j by some factor s_j, making the effective concentration $\chi_j = \chi_j(X_j, s_j)$. A simple example of such a modulation function is

$$\chi_j(X_j, s_j) \equiv \frac{X_j}{s_j} \tag{1}$$

in which the presence of the small molecule reduces the effective concentration of the transcription factor by the factor s_j.

© The Institution of Engineering and Technology 2008

The function of the circuit will depend on how the steady-state expression G^* of the reporter gene G changes as the modulation factor s_j is varied from some 'off' value s_j^- to some 'on' value s_j^+

$$\frac{\Delta G^*}{\Delta s_j} = \frac{G^*(s_j^+) - G^*(s_j^-)}{\Delta s_j} = \frac{1}{\Delta s_j}\int_{s_j^-}^{s_j^+} \frac{\mathrm{d}G^*}{\mathrm{d}s_j}\,\mathrm{d}s_j \qquad (2)$$

where $\Delta s_j = s_j^+ - s_j^-$. For example, if $\chi_j = X_j/s_j$, then $s_j^- = 1$, indicating that the small molecule is absent, and $s_j^+ > 1$ is the factor by which effective concentration is reduced when the small molecule is present.

If the sign of $\mathrm{d}G^*/\mathrm{d}s_j$ does not change for $s_j \in [s_j^-, s_j^+]$, then the sign of ΔG^* is fixed. For networks with only serial regulation, that is, each gene is regulated by at most one other gene, we will show that the sign of $\mathrm{d}G^*/\mathrm{d}s_j$ is unique and in accord with the direct path from S_j to G, a property we term direct functionality. This constrains the possible responses and hence the functionality of serial networks. Importantly, we will then show that all admissible functions indeed can be attained by all the networks we studied operating at different parameter values. Although throughout this work we focus on the setup pioneered experimentally by Guet *et al.* [1], we also show that the constraint to direct functionality holds for any network with serial regulation.

2 Direct functionality in small networks

As in Guet *et al.* [1], we consider networks with $N = 4$ genes (three transcription factors plus a reporter G), in which each gene is regulated by exactly one other gene. This admits three topologies and a total of 24 networks, as described in Fig. 1. All three topologies consist of a cycle and a cascade that begins in the cycle and ends at the reporter gene G. Once outside the cycle, there is only one path to G, so it suffices to study a topology consisting of an n-gene cycle with a gene G immediately outside (Fig. 1*c* is an example with $n = 3$), and extensions to topologies where the cycle is connected to the reporter by a linear cascade are trivial.

In this section, we will perform the steady-state analysis of such single-cycle networks to lay the groundwork for understanding the effect of topology on allowed functionality.

The process of protein expression has been modelled with remarkable success by combining transcription and translation into one step and directly coupling genes by a deterministic dynamics [7–9]. Accordingly, we model mean expressions $\bar{X}_i$ with the system of ordinary differential equations [we later distinguish between entire probability distributions $P(X_i)$ and the means of these

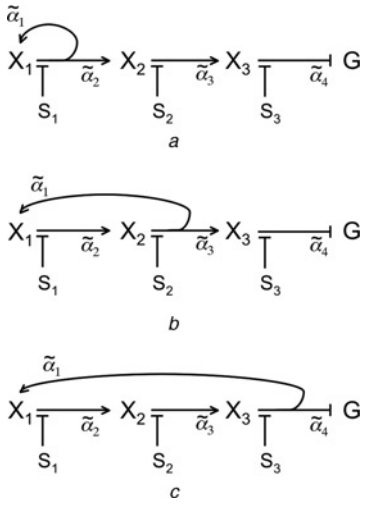

Figure 1 *Four-gene networks (three transcription factors X_i plus one reporter gene G) in which each gene is regulated by one other gene, as studied in [1, 6]*

Transcription factor efficacies are influenced by small molecules S_i. Regulation functions $\tilde{\alpha}_i$ are assigned to the edges. The three edges $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$ can be up-regulating or down-regulating, giving $3 \times 2^3 = 24$ possibilities; the reporter gene is repressed in all cases

distributions $\bar{X}_i$; cf. appendix]

$$\frac{\mathrm{d}\bar{X}_1}{\mathrm{d}t} = \tilde{\alpha}_1(\bar{\chi}_n) - r_1\bar{X}_1 \qquad (3)$$

$$\frac{\mathrm{d}\bar{X}_i}{\mathrm{d}t} = \tilde{\alpha}_i(\bar{\chi}_{i-1}) - r_i\bar{X}_i \quad (2 \le i \le n) \qquad (4)$$

$$\frac{\mathrm{d}\bar{G}}{\mathrm{d}t} = \tilde{\alpha}_{n+1}(\bar{\chi}_n) - r_{n+1}\bar{G} \qquad (5)$$

where $\tilde{\alpha}_i$ are the creation rates for the species X_i (and $X_{n+1} \equiv G$), each monotonically regulated by the effective concentration $\bar{\chi}_{\pi_i}$ of its parent π_i and the r_i are the decay rates. Note that we have set

$$\pi_1 = n \qquad (6)$$

$$\pi_i = i - 1 \quad (2 \le i \le n+1) \qquad (7)$$

to create the n-gene cycle with one gene immediately outside. The regulation functions $\tilde{\alpha}_i$ will be up- or down-regulating according to the network topology. A common example is the familiar Hill functions

$$\tilde{\alpha}(\bar{\chi}) = a_0 + a\frac{\bar{\chi}^h}{K^h + \bar{\chi}^h} \text{ (up-regulating)} \qquad (8)$$

$$\tilde{\alpha}(\bar{\chi}) = a_0 + a\frac{K^h}{K^h + \bar{\chi}^h} \text{ (down-regulating)} \qquad (9)$$

with basal and maximal expression levels a_0 and $a_0 + a$, respectively, Michaelis–Menten constants K and cooperativities h. Although we use the functional forms in

© The Institution of Engineering and Technology 2008

doi: 10.1049/iet-syb:20080097

(8) and (9), as well as the functional form for the modulation function in (1), for our numerical experiment (cf. Section 3), the analytic result derived in this section will be valid for any monotonic functions $\bar{\alpha}(\bar{\chi})$ and any function $\bar{\chi}(\bar{X}, s)$.

Fixed points of the dynamical system in (3)–(5) satisfy

$$\bar{X}_1^* = \alpha_1(\bar{\chi}_n^*) \tag{10}$$

$$\bar{X}_i^* = \alpha_i(\bar{\chi}_{i-1}^*) \quad (2 \le i \le n) \tag{11}$$

$$\bar{G}^* = \alpha_{n+1}(\bar{\chi}_n^*) \tag{12}$$

where we define

$$\alpha_i \equiv \frac{\tilde{\alpha}_i}{r_i} \tag{13}$$

We may now, as in [10, 11], use the chain rule to calculate the derivative of $\bar{G}^*$ with respect to a particular input factor s_j. For illustration, we will do so first for the concrete example in Fig. 1*c*, in which $n = 3$. Let us consider the derivative of $\bar{G}^*$ with respect to s_1

$$\frac{\mathrm{d}\bar{G}^*}{\mathrm{d}s_1} = \frac{\partial\alpha_4}{\partial\bar{X}_3}\frac{\partial\alpha_3}{\partial\bar{X}_2}\left[\frac{\partial\alpha_2}{\partial s_1} + \frac{\partial\alpha_2}{\partial\bar{X}_1}\frac{\mathrm{d}\bar{X}_1^*}{\mathrm{d}s_1}\right] \tag{14}$$

where all derivatives are evaluated at the fixed point, and it is understood that α_i depends on either $\bar{X}_{\pi_i}$ or s_{π_i} through $\bar{\chi}_{\pi_i}$, that is, that

$$\frac{\partial\alpha_i}{\partial\bar{X}_{\pi_i}} = \frac{\partial\alpha_i}{\partial\bar{\chi}_{\pi_i}}\frac{\partial\bar{\chi}_{\pi_i}}{\partial\bar{X}_{\pi_i}} \quad \text{and} \quad \frac{\partial\alpha_i}{\partial s_{\pi_i}} = \frac{\partial\alpha_i}{\partial\bar{\chi}_{\pi_i}}\frac{\partial\bar{\chi}_{\pi_i}}{\partial s_{\pi_i}} \tag{15}$$

If we introduce the notation

$$\alpha_i' \equiv \frac{\partial\alpha_i}{\partial\bar{X}_{\pi_i}} \tag{16}$$

$$\dot{\alpha}_i \equiv \frac{\partial\alpha_i}{\partial s_{\pi_i}} \tag{17}$$

then (14) becomes

$$\frac{\mathrm{d}\bar{G}^*}{\mathrm{d}s_1} = \alpha_4'\alpha_3'\left[\dot{\alpha}_2 + \alpha_2'\frac{\mathrm{d}\bar{X}_1^*}{\mathrm{d}s_1}\right] \tag{18}$$

The first term reflects the direct chain to G from S_1, and the second term incorporates further contributions around the cycle and will need to be evaluated self-consistently.

For a cycle of arbitrary length n and for an arbitrary input factor s_j $(1 \le j \le n)$, (18) generalises to

$$\frac{\mathrm{d}\bar{G}^*}{\mathrm{d}s_j} = \left[\dot{\alpha}_{j+1} + \alpha_{j+1}'\frac{\mathrm{d}\bar{X}_j^*}{\mathrm{d}s_j}\right]\prod_{k=j+2}^{n+1}\alpha_k' \tag{19}$$

where we use the convention that

$$\prod_{k=a}^{b}[\,\cdot\,] = 1 \quad \text{if} \quad a > b \tag{20}$$

We may also use the chain rule for $\mathrm{d}\bar{X}_j^*/\mathrm{d}s_j$

$$\frac{\mathrm{d}\bar{X}_j^*}{\mathrm{d}s_j} = \left[\dot{\alpha}_{(j \bmod n)+1} + \alpha_{(j \bmod n)+1}'\frac{\mathrm{d}\bar{X}_j^*}{\mathrm{d}s_j}\right]\frac{\prod_{k=1}^{n}\alpha_k'}{\alpha_{(j \bmod n)+1}'} \tag{21}$$

and now we may solve for $\mathrm{d}\bar{X}_j^*/\mathrm{d}s_j$ self-consistently

$$\frac{\mathrm{d}\bar{X}_j^*}{\mathrm{d}s_j} = \frac{\dot{\alpha}_{(j \bmod n)+1}}{\alpha_{(j \bmod n)+1}'}\frac{\prod_{k=1}^{n}\alpha_k'}{1 - \prod_{l=1}^{n}\alpha_l'} \tag{22}$$

For the special case of $j = n$, where $(j \bmod n) + 1 = 1$, substituting (22) into (19) obtains

$$\frac{\mathrm{d}\bar{G}^*}{\mathrm{d}s_n} = \left[\frac{1}{1 - \prod_{l=1}^{n}\alpha_l'}\right]\left[\dot{\alpha}_{n+1} + (\dot{\alpha}_1\alpha_{n+1}' - \dot{\alpha}_{n+1}\alpha_1')\prod_{k=2}^{n}\alpha_k'\right] \tag{23}$$

$$= \left[\frac{1}{1 - \prod_{l=1}^{n}\alpha_l'}\right]\dot{\alpha}_{n+1} \tag{24}$$

where the second step follows from

$$\dot{\alpha}_1\alpha_{n+1}' = \left(\frac{\mathrm{d}\alpha_1}{\mathrm{d}\bar{\chi}_n}\frac{\partial\bar{\chi}_n}{\partial s_n}\right)\left(\frac{\mathrm{d}\alpha_{n+1}}{\mathrm{d}\bar{\chi}_n}\frac{\partial\bar{\chi}_n}{\partial\bar{X}_n}\right) \tag{25}$$

$$= \left(\frac{\mathrm{d}\alpha_1}{\mathrm{d}\bar{\chi}_n}\frac{\partial\bar{\chi}_n}{\partial X_n}\right)\left(\frac{\mathrm{d}\alpha_{n+1}}{\mathrm{d}\bar{\chi}_n}\frac{\partial\bar{\chi}_n}{\partial s_n}\right) \tag{26}$$

$$= \alpha_1'\dot{\alpha}_{n+1} \tag{27}$$

in which the first step recalls (15). For $1 \le j \le n-1$, where $(j \bmod n) + 1 = j + 1$, substituting (22) into (19) obtains

$$\frac{\mathrm{d}\bar{G}^*}{\mathrm{d}s_j} = \left[\frac{1}{1 - \prod_{l=1}^{n}\alpha_l'}\right]\dot{\alpha}_{j+1}\prod_{k=j+2}^{n+1}\alpha_k' \tag{28}$$

which, upon inspection of (24), is valid for $j = n$ as well.

Stability of the fixed point $\bar{X}_j^*$ requires that the Jacobian of (3) and (4)

$$J = \begin{bmatrix} -r_1 & & & & & \tilde{\alpha}_1' \\ \tilde{\alpha}_2' & -r_2 & & & & \\ & \tilde{\alpha}_3' & -r_3 & & & \\ & & \ddots & \ddots & & \\ & & & \tilde{\alpha}_{n-1}' & -r_{n-1} & \\ & & & & \tilde{\alpha}_n' & -r_n \end{bmatrix} \tag{29}$$

be negative definite or, since the determinant is the product

© The Institution of Engineering and Technology 2008

of the eigenvalues, that

$$0 < (-1)^n \det(J) \tag{30}$$

$$= \prod_{k=1}^{n} r_k - \prod_{l=1}^{n} \bar{\alpha}'_l \tag{31}$$

$$= \prod_{k=1}^{n} r_k \left(1 - \prod_{l=1}^{n} \alpha'_l\right) \tag{32}$$

Since the decay rates r_k are positive, (32) states that the term inside the brackets in (28) is positive for stable fixed points.

For the networks in Fig. 1, where in the one- and two-cycles, the reporter is attached by means of intermediates, the analogue of (28) is calculated similarly to be

$$\frac{\mathrm{d}\bar{G}^*}{\mathrm{d}s_j} = \left[\frac{1}{1-\theta(n-j)\prod_{l=1}^{n}\alpha'_l}\right]\dot{\alpha}_{j+1}\prod_{k=j+2}^{N}\alpha'_k \tag{33}$$

where $N = 4$ is the number of genes, $1 \le j \le N-1$ for each of the three possible small molecule inputs and n is the length of the cycle ($1 \le n \le N-1$). Here, θ is the Heaviside function, for which we use the convention $\theta(0) = 1$. Its presence reduces the bracketed term to 1 when the input S_j is outside the cycle, leaving only the contribution corresponding to the cascade from S_j to G, as must be the case.

In (33), the term outside the brackets represents the direct (i.e. the shortest) path from S_j to G and fixes the sign of $\mathrm{d}\bar{G}^*/\mathrm{d}s_j$ (since the term inside the brackets is positive at a stable fixed point). If the creation rates are monotonic (which is the usual model for transcriptional regulation, but may be violated in protein signalling due to competitive inhibition and other effects), this sign is unique and fixes the sign of $\Delta\bar{G}^*/\Delta s_j$ via (2). Importantly, this shows that the feedback in each of the topologies in Fig. 1 is irrelevant in determining the sign of $\Delta\bar{G}^*/\Delta s_j$ for a steady-state analysis. As an example, for the network in Fig. 2*a* (inset), $\bar{G}^*$ changes with increasing s_1 according to $\dot{\alpha}_2\alpha'_3\alpha'_4$, which, since S_1 inhibits the activation, is negative × positive × negative = positive, just as one would expect if the feedback was ignored.

2.1 Direct functionality corresponds to specific orderings of output states

Consider the case in which there are only two small molecule inputs, S_1 and S_2, as in Fig. 2*a* (inset). Since each input can be absent or present, $S_1, S_2 \in \{-, +\}$, there are four chemical input states $c = S_1S_2 \in \{--, -+, +-, ++\}$. Direct functionality admits only two orderings of the four output states $\bar{G}^*_c$, and hence the functionality of the network is severely limited by its serial topology. To see this, note that for Fig. 2*a* (inset) we have

$$\Delta\bar{G}^*/\Delta s_1 \ge 0 \Rightarrow \bar{G}^*_{+-} \ge \bar{G}^*_{--} \quad \text{and} \quad \bar{G}^*_{++} \ge \bar{G}^*_{-+} \tag{34}$$

$$\Delta\bar{G}^*/\Delta s_2 \ge 0 \Rightarrow \bar{G}^*_{-+} \ge \bar{G}^*_{--} \quad \text{and} \quad \bar{G}^*_{++} \ge \bar{G}^*_{+-} \tag{35}$$

These conditions permit only the following output orderings, irrespective of biochemical parameters

$$\begin{aligned} &\bar{G}^*_{--} \le \bar{G}^*_{-+} \le \bar{G}^*_{+-} \le \bar{G}^*_{++} \quad \text{or} \\ &\bar{G}^*_{--} \le \bar{G}^*_{+-} \le \bar{G}^*_{-+} \le \bar{G}^*_{++} \end{aligned} \tag{36}$$

These two orderings nevertheless allow the realisation of a significant subset of all possible logical functions that one can build with two binary inputs, depending on the distinguishability of the four output states, as described in the next section. Quantifying the distinguishability demands careful treatment of the noise with a stochastic equivalent of our deterministic dynamical system, as described in the appendix.

3 Numerical results

We numerically solved the system in (3)–(5) [with stochastic effects given by (61)] with many parameter settings for all 24 networks represented in Fig. 1. In addition to verifying the restriction to direct functionality, we find that all networks can achieve all possible direct functions, suggesting that the networks are still quite versatile within the functional constraint.

For all networks, we consider the case of two small molecule inputs S_1 and S_2, as in the experimental setup of Guet *et al.* [1], and as shown for an example network in Fig. 2*a* (inset). We take s_j to be a multiplicative factor by which the transcription factor concentration $\bar{X}_j$ is effectively scaled, that is

$$\bar{\chi}_j(\bar{X}_j, s_j) \equiv \frac{\bar{X}_j}{s_j} \tag{37}$$

Then, $s_j^- \equiv 1$ for the 'off' settings and $s_j^+ > 1$ are the free parameters for the 'on' settings.

We model the regulation using the familiar Hill form (which is monotonic and thus satisfies the direct functionality conditions)

$$\tilde{\alpha}(\bar{\chi}) = a_0 + a\frac{\bar{\chi}^b}{K^b + \bar{\chi}^b} \text{ (up-regulating)} \tag{38}$$

$$\tilde{\alpha}(\bar{\chi}) = a_0 + a\frac{K^b}{K^b + \bar{\chi}^b} \text{ (down-regulating)} \tag{39}$$

with basal and maximal expression levels a_0 and $a_0 + a$, respectively, Michaelis–Menten constants K and

© The Institution of Engineering and Technology 2008

doi: 10.1049/iet-syb:20080097

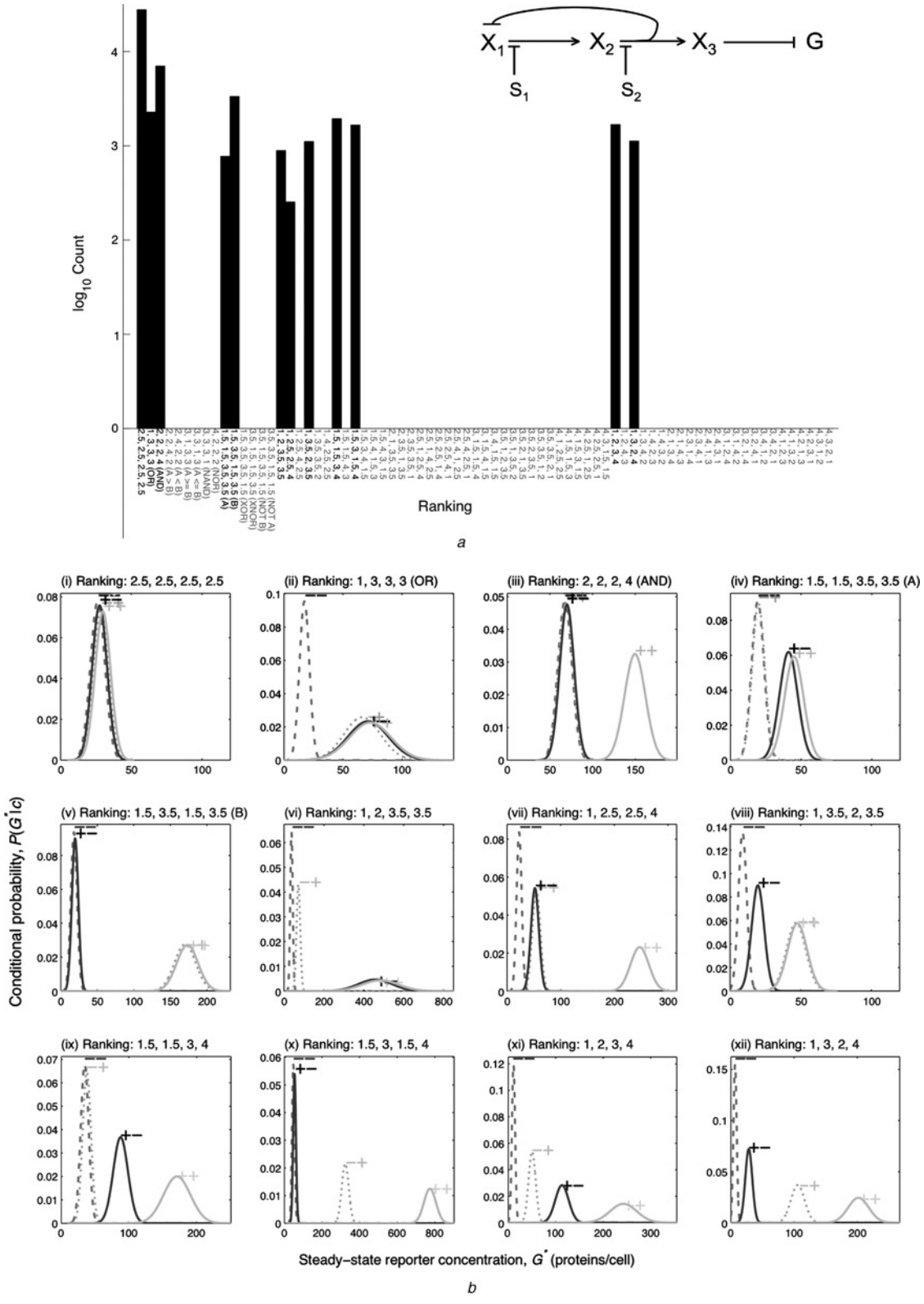

Figure 2 *Direct functionality in a representative network with serial regulation [network shown in the inset in (a)]*

a Histogram of logical functions, as defined by the ranking of the output distributions $P(G^*|c)$ (cf. Section 3)
Binary logic names are included after rankings when applicable, with 'A' and 'B' corresponding to inputs S_1 and S_2, respectively. Direct functions are labelled in black and indirect functions in grey. Note that only direct functions are observed, and that all direct functions can be attained by the network

b An example of each direct function
Two distributions are considered indistinguishable in rank when their means are separated by less than the smaller of their standard deviations. For example, in (ix), the distributions for the first and second states ($--$ and $-+$, respectively) are indistinguishable, so they are tied in rank at 1.5 (the average of ranks 1 and 2). Note that all functions satisfy the ordering constraints in (36)

© The Institution of Engineering and Technology 2008

Table 1 Parameters and ranges from which each is randomly drawn, with $1 \leq i \leq 4$ for the four genes, and $1 \leq j \leq 2$ for the two small molecule inputs

Parameters	Range
decay rates, r_i	10^{-4}–10^{-3}
Michaelis–Menten constants, K_i	10^{0}–10^{3}
basal expression levels, $a_{0,i}$	10^{-3}–10^{-2}
expression level ranges, a_i	10^{0}–10^{2}
cooperativities, h_i	10^{0}–10^{1}
'on' input factors, s_j^+	10^{2}–10^{3}

Ranges are representative of typical cell conditions [7, 12].

cooperativities h. For the four-gene networks in Fig. 1, with only two small molecule inputs S_1 and S_2, this gives 22 parameters in total (cf. Table 1).

For a given parameter set, we numerically solve (3)–(5) (using MATLAB's ode15s) for each input state $c \in \{--, -+, +-, ++\}$ to find mean steady-state concentrations $\bar{G}_c^*$. We then solve (61) to find fluctuations around these means, giving probability distributions $P(G^*|c)$ (cf. appendix). The function is defined by the ranking of the conditional distributions $P(G^*|c)$. That is, if two distributions are distinguishable, then the one with the larger mean is ranked higher. We consider two distributions to be indistinguishable when their means are separated by less than the smaller of their standard deviations (alternative definitions do not change our results qualitatively), in which case they both take on the average of their two ranks. When there are only two distinguishable output states, this rank-based classification reduces to that defining the familiar binary logical functions AND, OR, XOR and so on [see, e.g. Fig. 2*b*, (ii–v)]. More generally, for one, two, three and four distinguishable responses, there are 75 total rankings (as listed on the horizontal axis of Fig. 2*a*). However, only 12 of these satisfy the ordering constraints for each network analogous to those in (36) and therefore correspond to direct functions (for the network in Fig. 2*a* these 12 are shown in black on the horizontal axis).

We ran 50 000 trials for each of the 24 networks, in which the parameters were randomly selected (using a distribution uniform in log-space) from the ranges in Table 1. We found the steady-state reporter expression distributions $P(G^*|c)$ and classified the responses by ranking. All 24 networks displayed only direct functions. However, every network was able to achieve all 12 of its direct functions with parameters selected via Table 1, meaning that the networks fully realised all the functionality allowed by the constraint. This suggests that the networks studied are both constrained and versatile, and that a cell may still use a serial network to perform multiple logical functions by varying biochemical parameters, despite the restriction to direct functionality. Fig. 2 shows a histogram of functions and an example of each type of direct function for a representative network.

We note that Guet *et al.* experimentally observed both direct and indirect functions [1]. However, they explicitly call the indirect functions into question, citing several possible unanticipated effects including RNA polymerase read-through. We have not incorporated such effects into the current model.

4 Multiple fixed points

For the 12 networks in which the overall sign of the feedback cycle is positive, there are parameter settings that support multiple stable fixed points. In this section, we evaluate the extent to which the presence of multiple fixed points affects the constraint to direct functionality, and we find that violation of the constraint is possible but unlikely.

Although the function of a network has been defined in terms of $P(G^*|c)$, the linear noise approximation (cf. appendix) only gives us access to $P(G^*|c, \bar{\boldsymbol{X}}_m^*)$, the distribution expanded around a particular fixed point $\bar{\boldsymbol{X}}_m^*$. The two are related by a weighted sum

$$P(G^*|c) = \sum_m \pi_m P(G^*|c, \bar{\boldsymbol{X}}_m^*) \tag{40}$$

where the probabilities π_m of being near the mth fixed point will depend on the basins of attraction and curvatures near the fixed points. Numerical solution for $P(G^*|c)$ directly is possible in principle, although computationally difficult. Whether the statistical steady state distribution is calculated numerically or is approximated as in this paper, if we continue to define the function of the network by the ranking of the means of the $P(G^*|c)$, we have

$$\frac{\mathrm{d}\bar{G}^*}{\mathrm{d}s_j} = \frac{\mathrm{d}}{\mathrm{d}s_j}\int \mathrm{d}G^* G^* P(G^*|c) \tag{41}$$

$$= \sum_m \frac{\mathrm{d}}{\mathrm{d}s_j}\pi_m \int \mathrm{d}G^* G^* P(G^*|c, \bar{\mathbf{X}}_m^*) \tag{42}$$

$$= \sum_m \left(\pi_m \frac{\mathrm{d}\bar{G}_m^*}{\mathrm{d}s_j} + \frac{\mathrm{d}\pi_m}{\mathrm{d}s_j}\bar{G}_m^* \right) \tag{43}$$

The expressions for the individual $\mathrm{d}\bar{G}_m^*/\mathrm{d}s_j$ are given by (33), so the first term in (43) exhibits direct functionality. If the weights π_m do not depend appreciably on the s_j, the second term will be small, and the restriction to direct functionality will be maintained. If, on the other hand, the weights do change appreciably (an obvious case might be the presence of a bifurcation at a particular value of s_j), then the second term may overpower the first enough to change the sign of $\Delta\bar{G}^*/\Delta s$ and violate the restriction to direct functionality.

© The Institution of Engineering and Technology 2008

doi: 10.1049/iet-syb:20080097

We investigate this effect in two ways. First, we show analytically that, in the case of a one-cycle, crossing a bifurcation does not violate direct functionality. Second, we subject all positive-feedback networks to a numerical test to estimate the dependence of the weights π_m on the s_j. The results of both techniques suggest that the likelihood of a violation of direct functionality due to the presence of multiple fixed points is low.

4.1 Bifurcations do not violate direct functionality (1D)

Consider the case of a positive one-cycle with a gene G immediately outside, as shown in Fig. 3*a* (inset). For $n = 1$, (10)–(12) become

$$\bar{X}^* = \alpha_1(\bar{\chi}^*) \tag{44}$$

$$\bar{G}^* = \alpha_2(\bar{\chi}^*) \tag{45}$$

where unnecessary subscripts are dropped and $\bar{\chi} = \bar{X}/s$ as in (37). With α_1 of the form in (38), there are at most two stable fixed points $\bar{X}_1^*$ and $\bar{X}_2^*$, with $\bar{X}_1^* < \bar{X}_2^*$, as illustrated by an example in Fig. 3*a*. As shown in Fig. 3*b*, bifurcations occur at s_1 and s_2 such that only $\bar{X}_2^*$ exists when $s < s_1$, only $\bar{X}_1^*$ exists when $s > s_2$, and $\bar{X}_1^*$ and $\bar{X}_2^*$ are found with (unknown) probabilities $\tilde{\pi}_1(s)$ and $\tilde{\pi}_2(s) = 1 - \tilde{\pi}_1(s)$, respectively, when $s_1 < s < s_2$. These statements can be combined such that

$$\pi_1(s) = \theta(s - s_1)\theta(s_2 - s)\tilde{\pi}_1(s) + \theta(s - s_2) \tag{46}$$

and $\pi_2(s) = 1 - \pi_1(s)$ define the probabilities of approaching $\bar{X}_1^*$ and $\bar{X}_2^*$, respectively, for any s. Here, θ is the Heaviside function. As we go from an 'off' value s^- to an 'on' value s^+, let us assume that we hit both bifurcations, such that $s^- < s_1 < s_2 < s^+$. To test for direct functionality, we investigate the sign of

$$\frac{\Delta\bar{G}^*}{\Delta s} = \frac{1}{\Delta s}\int_{s^-}^{s^+}\sum_m \pi_m \frac{\mathrm{d}\bar{G}_m^*}{\mathrm{d}s}\mathrm{d}s + \frac{1}{\Delta s}\int_{s^-}^{s^+}\sum_m \frac{\mathrm{d}\pi_m}{\mathrm{d}s}\bar{G}_m^*\,\mathrm{d}s \tag{47}$$

$$\equiv T_1 + T_2 \tag{48}$$

obtained using (2) and (43). The first term T_1 depends on

$$\frac{\mathrm{d}\bar{G}_m^*}{\mathrm{d}s} = \frac{\dot{\alpha}_2}{1 - \alpha_1'} \tag{49}$$

[from (28); $\alpha' \equiv \partial\alpha/\partial\bar{X}$ and $\dot{\alpha} \equiv \partial\alpha/\partial s$ as before, both evaluated at the mth fixed point], which, as previously discussed, is always of the sign of $\dot{\alpha}_2$, consistent with direct functionality.

The second term T_2 can be written

$$T_2 = \frac{1}{\Delta s}\int_{s^-}^{s^+}\left(\frac{\mathrm{d}\pi_1}{\mathrm{d}s}\bar{G}_1^* + \frac{\mathrm{d}\pi_2}{\mathrm{d}s}\bar{G}_2^*\right)\mathrm{d}s \tag{50}$$

$$= \frac{1}{\Delta s}\int_{s^-}^{s^+} -\frac{\mathrm{d}\pi_1}{\mathrm{d}s}(\bar{G}_2^* - \bar{G}_1^*)\mathrm{d}s \tag{51}$$

and since

$$\frac{\mathrm{d}\pi_1}{\mathrm{d}s} = \theta(s_2 - s)\tilde{\pi}_1(s)\delta(s - s_1) + \left[1 - \theta(s - s_1)\tilde{\pi}_1(s)\right]\delta(s - s_2) + \theta(s - s_1)\theta(s_2 - s)\frac{\mathrm{d}\tilde{\pi}_1}{\mathrm{d}s} \tag{52}$$

(where δ is the Dirac delta function) we have

$$T_2 = \frac{1}{\Delta s}\left\{ -\left[\tilde{\pi}_1(\bar{G}_2^* - \bar{G}_1^*)\right]_{s_1} - \left[\tilde{\pi}_2(\bar{G}_2^* - \bar{G}_1^*)\right]_{s_2} - \int_{s_1}^{s_2}\frac{\mathrm{d}\tilde{\pi}_1}{\mathrm{d}s}(\bar{G}_2^* - \bar{G}_1^*)\mathrm{d}s\right\} \tag{53}$$

The first two terms in (53) represent the contributions from crossing the bifurcations at s_1 and s_2, respectively. Using (45), we may write them as

$$T_2 = \frac{1}{\Delta s}\left\{\sum_{m=1}^{2}\left[\tilde{\pi}_m\left(-\frac{\Delta\alpha_2}{\Delta\bar{\chi}^*}\right)\Delta\bar{\chi}^*\right]_{s_m} - \int_{s_1}^{s_2}\frac{\mathrm{d}\tilde{\pi}_1}{\mathrm{d}s}(\bar{G}_2^* - \bar{G}_1^*)\mathrm{d}s\right\} \tag{54}$$

where $\Delta\alpha_2 = \alpha_2(\bar{\chi}_2^*) - \alpha_2(\bar{\chi}_1^*)$ and $\Delta\bar{\chi}^* = \bar{\chi}_2^* - \bar{\chi}_1^* = (\bar{X}_2^* - \bar{X}_1^*)/s > 0$. Since α_2 is monotonic in $\bar{X}$, $-\Delta\alpha_2/\Delta\bar{\chi}^*$ at fixed s is of the same sign as $-\alpha_2'$, which is of the same sign as $\dot{\alpha}_2$ since s effectively reduces X [see (37)]. Therefore the contributions to $\Delta\bar{G}^*/\Delta s$ from crossing the bifurcations do not violate direct functionality. A violation, at least in the case of a one cycle, can only come from variations in the probabilities $\tilde{\pi}_m$ within the region $s_1 < s < s_2$, as described by the last term in (54). Next, we describe a numerical test that suggests such violations are rare.

4.2 Numerics suggest violations from multiple fixed points are rare

For each of the 12 positive-feedback networks, we numerically found the steady state of the dynamical system with randomly sampled parameters as before (cf. Section 3). However, now for each parameter set, we solved the system many times with randomly selected initial conditions. When multiple fixed points were found, the fraction of trials approaching the mth fixed point was used for the weight π_m. This assumes that the π_m are determined only by the basins of attraction of each fixed point and by the distribution of the initial conditions.

© The Institution of Engineering and Technology 2008

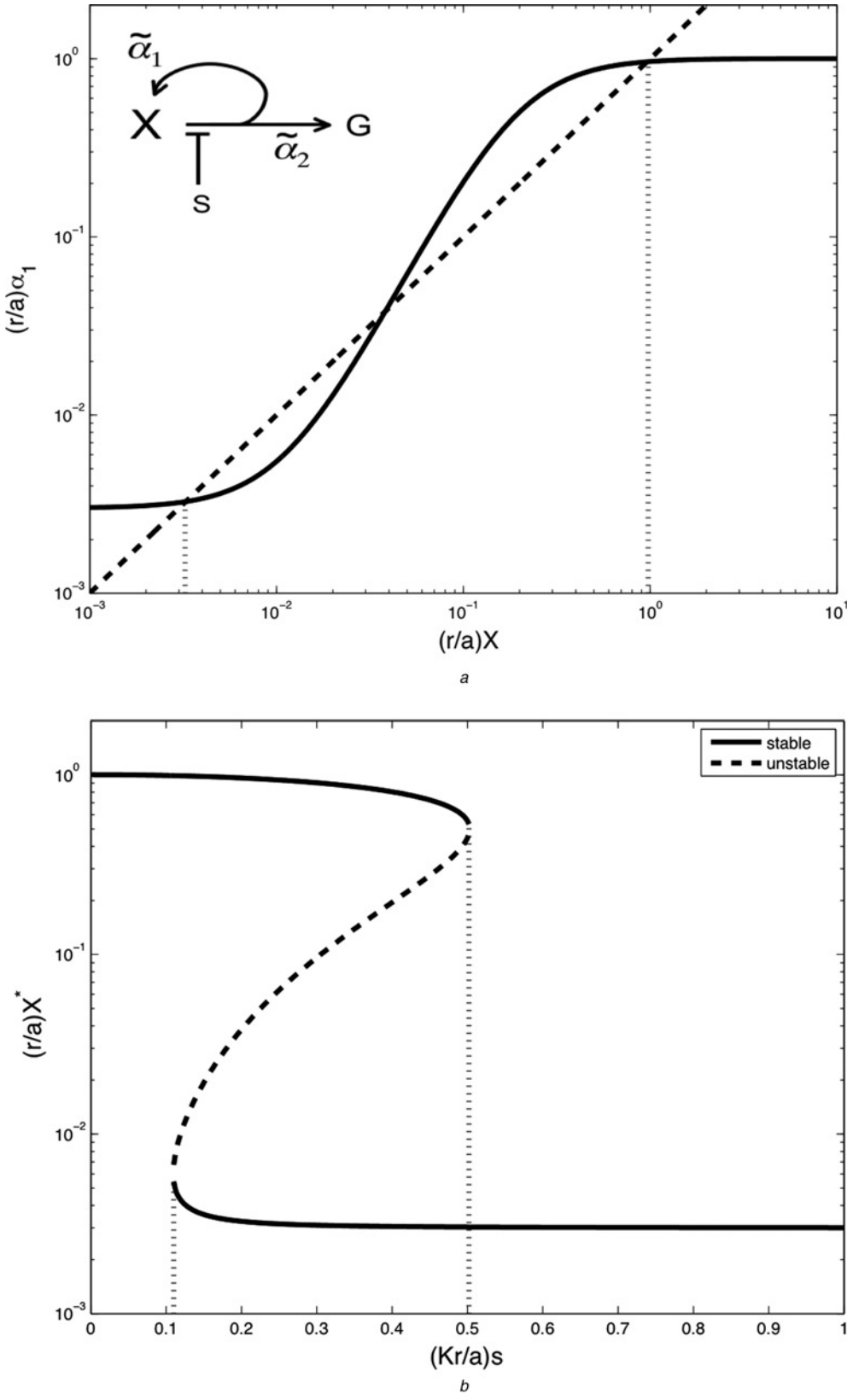

Figure 3 *Multiple fixed points in a one-cycle*

a Solid: plot of regulation function $\alpha_1 = \tilde{\alpha}_1/r$ (refer to inset network), as defined in (38) and (37), with parameters $h = 2$, $a_0 = 0.03$, $a = 10$, $K = 2$ and $r = 1$
Dashed line shows $\alpha_1 = \bar{X}$ such that dotted lines indicate locations of stable fixed points $\bar{X}_1^*$ and $\bar{X}_2^*$ (take $\bar{X}_2^* > \bar{X}_1^*$)
b Stable fixed points X_m^* (solid) and unstable fixed points (dashed) as a function of s, with $a_0/a = 0.003$ as in (a)
Dotted lines indicate locations of bifurcation points s_1 and s_2 such that only $\bar{X}_1^*$ exists when $s < s_1$, only $\bar{X}_1^*$ exists when $s > s_2$, and both $\bar{X}_1^*$ and $\bar{X}_2^*$ exist for $s_1 < s < s_2$

© The Institution of Engineering and Technology 2008

doi: 10.1049/iet-syb:20080097

However, different distributions of initial conditions do not give qualitatively different results.

For each network, 2000 parameter sets were selected (uniform randomly in log-space), at which the system was solved 100 times with initial protein counts selected uniform randomly from 0 to 1000 proteins per cell. Over all positive-feedback networks, 37% of the parameter sets supported multiple fixed points for at least one of the settings of S_1 and S_2. However, only 0.46% of parameter sets produced violations of direct functionality. Moreover, this number is likely an overestimate, as no distinguishability criterion was imposed as was done in the single-fixed point case (cf. section 2). It is likely that this fraction would remain low if the estimation of the π_m was refined to incorporate the curvatures of the fixed points, or if alternative distributions were used for the sampling.

5 All networks with serial regulation exhibit only direct functionality

In this section, we extend our analytic constraint as derived in the context of the system studied experimentally by Guet *et al.* [1] to show that any network with only serial regulation – each node having 0 or 1 parent – exhibits only direct functionality, that is, any target node X_i changes with any input S_j according to the direct path between them.

We first consider a connected directed graph in which every node has in-degree 1, called a contrafunctional graph [13]. One can show that a contrafunctional graph has exactly one cycle, each of whose nodes is the root of a tree if the cycle edges are ignored [13]. Now consider changing one node's in-degree to 0, or equivalently, removing an edge. If the edge is in the cycle, the graph remains connected and becomes a tree. If the edge is not in the cycle, the graph is cut into two components: a contrafunctional graph and a tree.

A tree exhibits only direct functionality since there is at most one path from an input S_j to a gene S_i, which is therefore the direct path.

In a contrafunctional graph, we first consider the case where the target node X_i is inside the cycle. Only inputs S_j that are inside the cycle can affect X_i because the rest of the graph consists of trees that all point away from the cycle. Since we can start labelling nodes at any point in the cycle, we may take $i \leq j$ without loss of generality. Then, using the chain rule

$$\frac{d\bar{X}_i^*}{ds_j} = \left[\dot{\alpha}_{(j \bmod n)+1} + \alpha'_{(j \bmod n)+1} \frac{d\bar{X}_j^*}{ds_j}\right] \times \frac{\prod_{k=1}^{n} \alpha'_k}{\prod_{l=i}^{j} \alpha'_{(l \bmod n)+1}} \quad (55)$$

$$= \left[\frac{1}{1 - \prod_{m=1}^{n} \alpha'_m}\right] \dot{\alpha}_{(j \bmod n)+1} \frac{\prod_{k=1}^{n} \alpha'_k}{\prod_{l=i}^{j} \alpha'_{(l \bmod n)+1}} \quad (56)$$

where the second step follows from (22).

We next consider the case where the target node is outside the cycle. An input S_j can only affect the node if it is either in the cycle or above the node in its tree. The portion of the path in the tree will exhibit direct functionality. Therefore in looking for possible indirect functionality we may, without loss of generality, take the node to be immediately outside the cycle, as we did for G in the previous section. $d\bar{G}^*/ds_j$ is then given by (28).

In both (56) and (28), the term outside the brackets represents the direct path from S_j to the target node, and the term inside the brackets is positive for stable fixed points. Therefore a contrafunctional graph exhibits only direct functionality. Since each connected component of a network in which every node has in-degree 0 or 1 is either a contrafunctional graph or a tree, such networks exhibit only direct functionality. Thus, in general, the possible logical functions of topologies with at most one regulator per node are severely constrained.

6 Acknowledgment

We are grateful to the organisers and participants of The First q-bio Conference, where a preliminary version of this work was presented. This work was partially supported by NSF Grant No. ECS-0425850 to CW and IN. IN was further supported by LANL LDRD program under DOE Contract No. DE-AC52-06NA25396.

7 References

[1] GUET C.C., ELOWITZ M.B., HSING W., LEIBLER S.: 'Combinatorial synthesis of genetic networks', *Science*, 2002, **296**, (5572), pp. 1466–1470

[2] SHEN-ORR S.S., MILO R., MANGAN S., ALON U.: 'Network motifs in the transcriptional regulation network of *Escherichia coli*', *Nat. Genet.*, 2002, **31**, (1), pp. 64–68

[3] MANGAN S., ALON U.: 'Structure and function of the feed-forward loop network motif', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, (21), pp. 11980–11985

[4] KOLLMANN M., LØVDOK L.K., BARTHOLOMÉ TIMMER J., SOURJIK V.: 'Design principles of a bacterial signalling network', *Nature*, 2005, **438**, (7067), pp. 504–507

[5] WALL M.E., DUNLOP M.J., HLAVACEK W.S.: 'Multiple functions of a feedforward-loop gene circuit', *J. Mol. Biol.*, 2005, **349**, (3), pp. 501–514

© The Institution of Engineering and Technology 2008

[6] ZIV E., NEMENMAN I., WIGGINS C.H.: 'Optimal signal processing in small stochastic biochemical networks', *PLoS ONE*, 2007, **2**, (10), article id: e1077

[7] ELOWITZ M.B., LEIBLER S.: 'A synthetic oscillatory network of transcriptional regulators', *Nature*, 2000, **403**, (6767), pp. 335–338

[8] GARDNER T.S., CANTOR C.R., COLLINS J.J.: 'Construction of a genetic toggle switch in escherichia coli', *Nature*, 2000, **403**, (6767), pp. 339–342

[9] HASTY J., MCMILLEN D., ISAACS F., COLLINS J.J.: 'Computational studies of gene regulatory networks: in numero molecular biology', *Nat. Rev. Genet.*, 2001, **2**, (4), pp. 268–279

[10] KHOLODENKO B.N., HOEK J.B., WESTERHOFF H.V., BROWN G.C.: 'Quantification of information transfer via cellular signal transduction pathways', *FEBS Lett.*, 1997, **414**, (2), pp. 430–434

[11] KHOLODENKO B.N., KIYATKIN A., BRUGGEMAN F.J., SONTAG E., WESTERHOFF H.V., HOEK J.B.: 'Untangling the wires: a strategy to trace functional interactions in signaling and gene networks', *Proc. Natl. Acad. Sci. USA*, 2002, **99**, (20), pp. 12841–12846

[12] BRAUN D., BASU S., WEISS R.: 'Parameter estimation for two synthetic gene networks: a case study'. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2005, vol. 5, pp. v769–v772

[13] HARARY F.: 'Structural models: an introduction to the theory of directed graphs' (Wiley, New York, 1965)

[14] ELF J., PAULSSON J., BERG O.G., EHRENBERG M.: 'Near-critical phenomena in intracellular metabolite pools', *Biophys. J.*, 2003, **84**, (1), pp. 154–170

[15] PAULSSON J.: 'Summing up the noise in gene networks', *Nature*, 2004, **427**, (6973), pp. 415–418

[16] ELF J., EHRENBERG M.: 'Fast evaluation of fluctuations in biochemical networks with the linear noise approximation', *Genome Res.*, 2003, **13**, (11), pp. 2475–2484

[17] VAN KAMPEN N.G.: 'Stochastic processes in physics and chemistry' (North-Holland, Amsterdam, 1992)

[18] GILLESPIE D.T.: 'Exact stochastic simulation of coupled chemical reactions', *J. Phys. Chem.*, 1977, **81**, (25), pp. 2340–2361

8 Appendix: the stochastic model

The dynamical system in (3)–(5) provides a deterministic description of mean expression levels but fails to capture fluctuations around these means. A full stochastic description is given by the chemical master equation. For N species participating in R elementary reactions in a system with volume Ω, the master equation reads

$$\frac{\mathrm{d}P(\boldsymbol{n}, t)}{\mathrm{d}t} = \Omega \sum_{j=1}^{R} \left(\prod_{i=1}^{N} E^{-Z_{ij}} - 1 \right) f_j(\boldsymbol{X}, \Omega) P(\boldsymbol{n}, t) \quad (57)$$

where $P(\boldsymbol{n}, t)$ is the probability of having the copy number vector $\boldsymbol{n} = \Omega \boldsymbol{X} = \Omega(X_1, \ldots, X_N)$ at time t, Z_{ij} is the $N \times R$ stochiometric matrix, $E^{-Z_{ij}}$ is the step operator which acts by removing Z_{ij} molecules from n_i and f_j is the rate for reaction j. The f_j are the $\bar{\alpha}_j$ and $r_j X_j$ of (3)–(5) in the macroscopic limit.

As in previous work [6], we employ the much-used linear noise approximation [14–17] to make (57) tractable by expanding in orders of $\Omega^{-1/2}$. Introducing $\boldsymbol{\xi}$ such that $n_i = \Omega X_i + \Omega^{1/2} \xi_i$ and treating $\boldsymbol{\xi}$ as continuous, the first two terms in the expansion yield the macroscopic rate equations [e.g. (3)–(5) in our case] and the linear Fokker–Plank equation, respectively

$$\sum_{i=1}^{N} \frac{\partial \bar{X}_i}{\partial t} \frac{\partial P(\boldsymbol{\xi}, t)}{\partial \xi_i} = \sum_{i=1}^{N} \sum_{j=1}^{R} Z_{ij} f_j(\bar{\mathbf{X}}) \frac{\partial P(\boldsymbol{\xi}, t)}{\partial \xi_i} \quad (58)$$

$$\frac{\partial P(\boldsymbol{\xi}, t)}{\partial t} = -\sum_{i,k} J_{ik} \frac{\partial (\xi_k P)}{\partial \xi_i} + \frac{1}{2} \sum_{i,k} D_{ik} \frac{\partial^2 P}{\partial \xi_i \partial \xi_k} \quad (59)$$

where $J_{ik} = \sum_{j=1}^{R} Z_{ij} (\partial f_j / \partial X_k)$ is the Jacobian matrix [e.g. (29)] and $D_{ik} = \sum_{j=1}^{R} Z_{ij} Z_{kj} f_j(\mathbf{X})$ is a diffusion-like matrix. The steady-state solution to (59) is the multivariate Gaussian

$$P(\boldsymbol{\xi}) = \left[(2\pi)^N \det \boldsymbol{\Xi} \right]^{-1/2} \exp\left(-\frac{\boldsymbol{\xi}^{\mathrm{T}} \boldsymbol{\Xi} \boldsymbol{\xi}}{2} \right) \quad (60)$$

where the covariance matrix $\boldsymbol{\Xi}$ satisfies

$$\boldsymbol{J}\boldsymbol{\Xi} + \boldsymbol{\Xi}\boldsymbol{J}^{\mathrm{T}} + \boldsymbol{D} = 0 \quad (61)$$

We solve for $\boldsymbol{\Xi}$ using standard matrix Lyapunov equation solvers (e.g. MATLAB's *lyap*). Thus, fluctuations are captured to leading order by Gaussian distributions with means $\bar{X}_i$ given by the macroscopic equation and variances given by the diagonal entries of $\boldsymbol{\Xi}$. For example, Gaussian distributions $P(G^*|c)$ are shown in Fig. 2*b* for the steady-state concentration of the reporter gene G under chemical input states c. In [6], we have compared the distributions $P(G^*|c)$ obtained using the linear noise approximation with those obtained via direct stochastic simulations [18] and found the results almost indistinguishable for molecular copy number above 10–20.

© The Institution of Engineering and Technology 2008

doi: 10.1049/iet-syb:20080097