View Article Online / Journal Homepage / Table of Contents for this issue
Dynamic Article Links

Molecular BioSystems

Cite this: *Mol. BioSyst.*, 2012, **8**, 511–530

www.rsc.org/molecularbiosystems

PAPER

Optimality and thermodynamics determine the evolution of transcriptional regulatory networks†

Marco Avila-Elchiver,‡[a] Deepak Nagrath‡[b] and Martin L. Yarmush*[a]

Received 10th May 2011, Accepted 23rd September 2011
DOI: 10.1039/c1mb05177f

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

Transcriptional motifs are small regulatory interaction patterns that regulate biological functions in highly-interacting cellular networks. Recently, attempts have been made to explain the significance of transcriptional motifs through dynamic function. However, fundamental questions remain unanswered. Why are certain transcriptional motifs with similar dynamic function abundant while others occur rarely? What are the criteria for topological generalization of these motifs into complex networks? Here, we present a novel paradigm that combines non-equilibrium thermodynamics with multiobjective-optimality for network analysis. We found that energetic cost, defined herein as specific dissipation energy, is minimal at the optimal environmental conditions and it correlates inversely with the abundance of the network motifs obtained experimentally for *E. coli* and *S. cerevisiae*. This yields evidence that dissipative energetics is the underlying criteria used during evolution for motif selection and that biological systems during transcription tend towards evolutionary selection of subgraphs which produces minimum specific heat dissipation under optimal conditions, thereby explaining the abundance/rare occurrence of some motifs. We show that although certain motifs had similar dynamical functionality, they had significantly different energetic cost, thus explaining the abundance/rare occurrence of these motifs. The presented insights may establish global thermodynamic analysis as a backbone in designing and understanding complex networks systems, such as metabolic and protein interaction networks.

1. Introduction

Network motifs are the basic building blocks of complex networks and are the smallest overrepresented repeated subgraphs occurring commonly in both man-made large-scale networks (such as the world-wide web) and complex natural networks (such as cellular networks).[1–4] Motifs in transcriptional regulatory networks (TRNs) have numerous functions that help maintain phenotypes. The three-node feed-forward loop (FFL) motifs: (1) are among the most abundant and conserved TRNs; (2) are the smallest repeated interacting unit between genes/operon and transcription factors, and maintain gene regulation; and (3) have dynamical functions such as pulse generation, response delays, and noise filtering.[5] TRN motifs of various types including FFL, bifan, and single input module (SIM) (Fig. 1) have been found to occur in real networks in organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*.

There has been a continued struggle in understanding the common basis of occurrence of these motifs in network biology. Insights about the wiring of these network motifs could explain their evolutionary selection criteria, uncover the mechanism behind TRNs evolution, and decipher the basis behind the coordination of regulatory processes. Intense interest in explaining the selection of one motif over another and the frequency of occurrence of these network motifs in TRNs has focused attention on the structural and dynamical basis of these motifs.[5–7] Despite the advances in the identification of the mechanisms for the natural occurrence of these motifs, structural and dynamical functional bases have failed to provide an understanding of the properties and the density of occurrence of these motifs. Thus, there has been the lack of a universal basis which can describe the underlying mechanisms behind the occurrence of these motifs, as well as the way in which these motifs encode functional information and the way that both the dynamical function and the topological generalization may evolve.[6]

Cellular systems perform an array of regulatory,[8–11] homeostatic,[12–16] and phenotypic functions thus they exhibit a tradeoff between proliferation and differentiation,[17] cellular functions and growth,[18,19] and cellular functions and robustness.[20–22] Therefore, a multiobjective-optimal approach (where tradeoff

[a] *Massachusetts General Hospital and the Harvard Medical School, Shriners Hospitals for Children, 51 Blossom Street, Boston, MA 02114. E-mail: ireis@sbi.org; Tel: (617) 371-4882*
[b] *Chemical and Biomolecular Engineering Department, Rice University, Houston, TX 77005. E-mail: dn7@rice.edu*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c1mb05177f
‡ Equal Contribution.

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

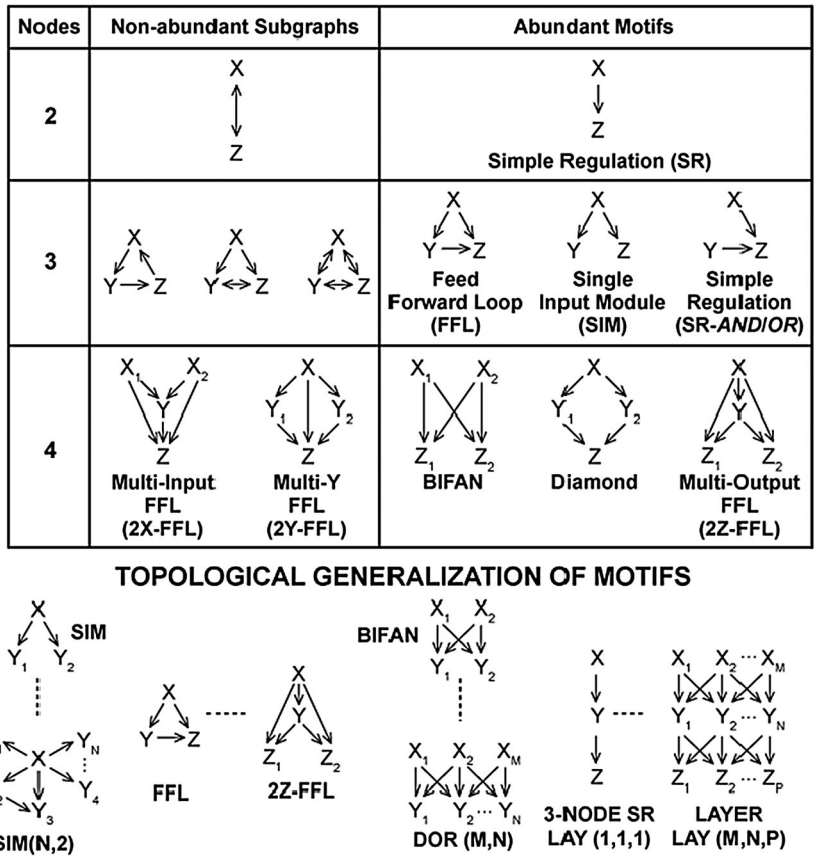


Fig. 1 Description of Network Motifs and non-equilibrium TRNs. Patterns of subgraphs and network motifs found in *E. coli* and *S. cerevisiae* networks. FFL is the common three node subgraph, bifan is the common four node subgraph, and SR is the common simple regulation motif. Higher order FFLs (multi-output, 2*Z*-FFL for four node TRNs) and diamond motif are common four node subgraphs in these organisms. Also shown are the topologically generalized networks from simple network motifs.

between several objectives has to be attained simultaneously) is necessary when seeking optimal functional analysis of these systems.[17,23–24] A multiobjective solution is said to be Pareto-optimal if there is no other feasible solution that will yield an improvement in an objective without causing degradation in at least one other objective.[25–27] Similar to cellular processes, transcription also works using the Pareto principle, *i.e.* higher transcription occurs at the cost of robustness or a cellular function.[28] We have developed a Pareto thermodynamic criterion that couples the non-equilibrium thermodynamic analysis for TRNs with Pareto-optimal solutions for attaining biological functions of a motif.

Here, we postulate that the abundance of certain motifs in a network can be predicted based on a conceptual framework that integrates non-equilibrium thermodynamics with multiobjective-optimality of the biological functions to be carried out by the motif (such as transcriptional rate, and/or robustness). We present an energetic-cost (defined herein as specific dissipation energy) theory that can explain which network motif has a higher probability of selection under a given environment, as well as explain the topological generalization of the subgraphs compared to other circuit designs. Through the developed framework, we have tried to answer the questions of why evolution converges to the same network motifs in TRNs and what advantage these selected network motifs offer as compared to other subgraphs for both steady state and dynamic analyses.

2. Systems and methods

2.1. Kinetic modeling of activation of *Z* by *X* using cycle concept

Our goal here is to develop a mathematical treatise which allows estimation of energetic cost during transcriptional regulation. We first present the formulation for the simplest case $X \rightarrow Z$ (TF X activates Z).

2.1.1 Mechanism. The mechanism by which the protein Z is transcribed by the transcription factor X is divided in a three-step process as shown below and in Fig. 2.

Step 1: The transcription factor X binds a free DNA site of the promoter region of protein Z (D_Z) to form a complex (D_{ZX}): $D_Z + X \leftrightarrow D_{ZX}$. If k_1^{ZX} [s^{-1} nM^{-1}] and k_{-1}^{ZX} [s^{-1}] are the forward and backward kinetic constants, respectively, then the net flux for this reaction is given by

$$J_1^{ZX} = k_1^{ZX}[D_Z][X] - k_{-1}^{ZX}[D_{ZX}] \quad (1)$$

and the reaction chemical potential can be estimated as:

$$\Delta\mu_1^{ZX} = \Re T \ln\left(\frac{k_{-1}^{ZX}[D_{ZX}]}{k_1^{ZX}[D_Z][X]}\right) = \Re T \ln\left(\frac{[D_{ZX}]}{K_1^{ZX}[D_Z][X]}\right) \quad (2)$$

where $\Re$[J K^{-1} mol^{-1}] is the gas constant, T[K] is the absolute temperature and $K_1^{ZX} = k_1^{ZX}/k_{-1}^{ZX}$ [nM^{-1}] is the association equilibrium constant.

This journal is © The Royal Society of Chemistry 2012

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

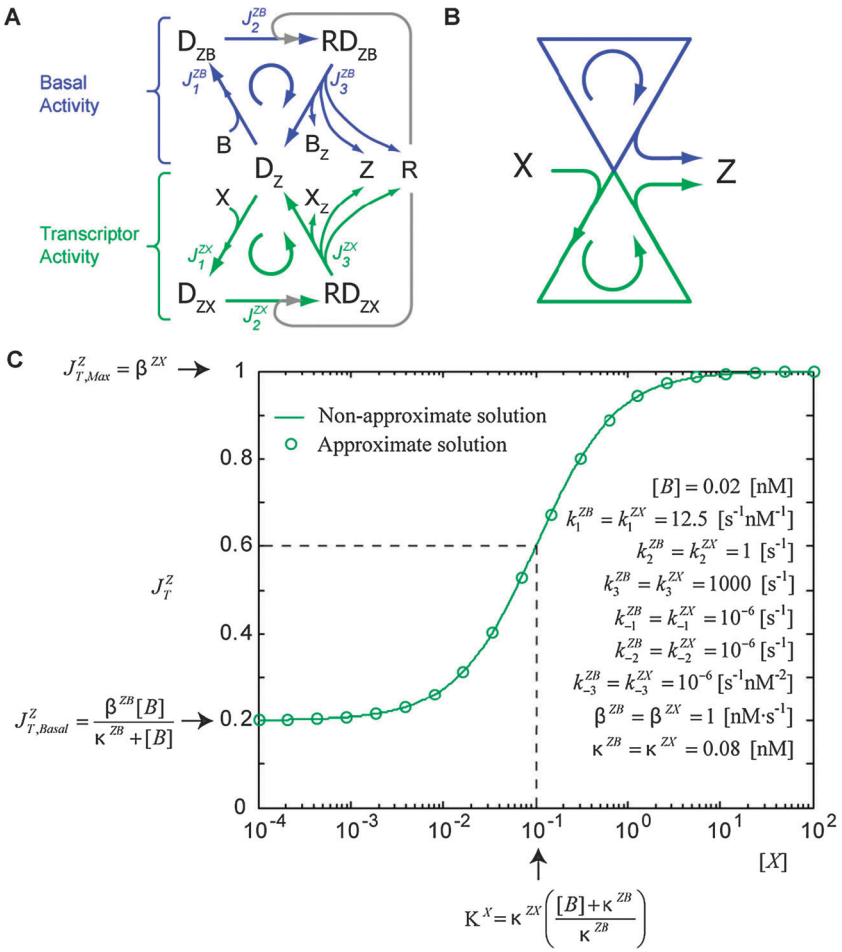


Fig. 2 Cyclic representation of the activation pathway of protein *Z* by transcription factor *X*. (A) Cyclic transcriptional activation schematic for protein synthesis which includes the transcriptional activation of gene *Z* by TF *X* (green cycle) and basal activators *B* (blue cycle), and subsequent translation steps for protein synthesis with amino acid incorporation. (B) Simplified representation of the cyclic activation model. (C) Transcription rate of *Z* as a function of its activator *X*. Curve follows a first order Hill's function, with maximal transcription rate of 1 [nM s^{-1}] and activation coefficient of 0.1 [nM].

Step 2: RNA polymerase (R) binds D_{ZX} and forms a complex (RD_{ZX}): $R + D_{ZX} \leftrightarrow RD_{ZX}$. If $k_2^{0,ZX}$ [s^{-1} nM^{-1}] and k_{-2}^{ZX} [s^{-1}] are the forward and backward kinetic constants for this step, respectively then net flux for this reaction can be written as:

$$J_2^{ZX} = k_2^{0,ZX}[R][D_{ZX}] - k_{-2}^{ZX}[RD_{ZX}] = k_2^{ZX}[D_{ZX}] - k_{-2}^{ZX}[RD_{ZX}] \quad (3)$$

where $k_2^{ZX} = k_2^{0,ZX}[R]$ [s^{-1}] is a pseudo-first order kinetic constant. If $K_2^{ZX} = k_2^{ZX}/k_{-2}^{ZX}$ is the association equilibrium constant for this reaction, then the reaction chemical potential is given by

$$\Delta\mu_2^{ZX} = \Re T \ln\left(\frac{k_{-2}^{ZX}[RD_{ZX}]}{k_2^{ZX}[D_{ZX}]}\right) = \Re T \ln\left(\frac{[RD_{ZX}]}{K_2^{ZX}[D_{ZX}]}\right) \quad (4)$$

Step 3: Once RNA polymerase is bound, there is a transcription step to form mRNA and then a translation step which requires recruitment of the amino acids (AA_Z) required to form the protein Z. In our model, for the sake of simplicity, we have combined both steps: $RD_{ZX} + AA_Z \leftrightarrow R + D_Z + X_Z + Z$. Note that as a result of transcription of Z, RNA polymerase molecules are released, thus freeing the DNA site and releasing the bound co-activator (X_Z). If $k_3^{ZX} = k_3^{0,ZX}[AA_Z]$[$s^{-1}$] and $k_{-3}^{ZX} = k_{-3}^{0,ZX}[R]$[$s^{-1}$ nM^{-2}] are pseudo-first order kinetic constants, and $K_3^{ZX} = k_3^{ZX}/k_{-3}^{ZX}$ [nM^2] is the association equilibrium constant for this reaction, then the net reaction flux and reaction chemical potential are given by:

$$J_3^{ZX} = k_3^{ZX}[RD_{ZX}] - k_{-3}^{ZX}[D_Z][X_Z][Z] \quad (5)$$

$$\Delta\mu_3^{ZX} = \Re T \ln\left(\frac{[D_Z][X_Z][Z]}{K_3^{ZX}[RD_{ZX}]}\right)$$

The described steps 1 to 3 are represented as a green triangular cycle in Fig. 2A and B. An analogous mechanism describes the basal activity (blue cycle in Fig. 2A and B) in which a basal transcription factor B binds the free promoter region of Z, initiating the cycle and basal level transcription of Z.

2.1.2 Mass balance equations. Here J_i and α_i are defined as the external intake flux in [nM s^{-1}] and the degradation rate in [s^{-1}] of the species i, respectively. Flux is defined positive for

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

incoming flux and negative otherwise. The mass balance of all species in Fig. 2A can be described by the following differential equations:

$$\frac{d[X]}{dt} = J_X - J_1^{ZX} - \alpha_X[X] \quad (6)$$

$$\frac{d[B]}{dt} = J_B - J_1^{ZB} - \alpha_B[B] \quad (7)$$

$$\frac{d[Z]}{dt} = J_3^{ZX} + J_3^{ZB} + J_Z - \alpha_Z[Z] \quad (8)$$

$$\frac{d[X_Z]}{dt} = J_3^{ZX} + J_{X_Z} - \alpha_{X_Z}[X_Z] \quad (9)$$

$$\frac{d[B_Z]}{dt} = J_3^{ZB} + J_{B_Z} - \alpha_{B_Z}[B_Z] \quad (10)$$

$$\frac{d[D_Z]}{dt} = J_3^{ZX} + J_3^{ZB} - J_1^{ZX} - J_1^{ZB} \quad (11)$$

$$\frac{d[D_{ZX}]}{dt} = J_1^{ZX} - J_2^{ZX} \quad (12)$$

$$\frac{d[D_{ZB}]}{dt} = J_1^{ZB} - J_2^{ZB} \quad (13)$$

$$\frac{d[RD_{ZX}]}{dt} = J_2^{ZX} - J_3^{ZX} \quad (14)$$

$$\frac{d[RD_{ZB}]}{dt} = J_2^{ZB} - J_3^{ZB} \quad (15)$$

Additionally, mass balance has to satisfy DNA site balance equality constraint:

$$[D_Z]_{\text{tot}} = [D_Z] + [D_{ZX}] + [D_{ZB}] + [RD_{ZX}] + [RD_{ZB}] \quad (16)$$

In eqn (8), the mass balance is between generated transcription factor Z through main cycle (J_3^{ZX}) and through basal cycle flux (J_3^{ZB}), incoming transcription factor Z flux (through different TF cycles, can be assumed negligible if Z is not generated through any other cycles) and the degradation of product Z. Here, X_Z and B_z denotes the modified transcription factors released from the cycle after the transcription and translation process.

2.1.3 Steady state solution. At steady state, $J_1^{ZX} = J_2^{ZX} = J_3^{ZX} = J_C^{ZX}$, where J_C^{ZX} [nM s^{-1}] is the cyclic flux of cycle ZX. Analytical solution of the mass balance leads to:

$$J_C^{ZX} = \frac{[D_Z]_{tot}(k_1^{ZX}k_2^{ZX}k_3^{ZX}[X] + k_{-1}^{ZX}k_{-2}^{ZX}k_{-3}^{ZX}[X_Z][Z])}{k_{-1}^{ZX}k_{-2}^{ZX} + k_{-1}^{ZX}k_3^{ZX} + k_2^{ZX}k_3^{ZX} + (k_2^{ZX} + k_{-2}^{ZX} + k_3^{ZX})k_1^{ZX}[X] + (k_{-1}^{ZX} + k_2^{ZX} + k_{-2}^{ZX})k_{-3}^{ZX}[X_Z][Z]} \quad (17)$$

A similar equation can be found for the basal activity cyclic flux J_C^{ZB}. Since protein transcription is a highly irreversible process, one can expect that the rate of the forward reaction in each of the three steps, to be much higher than the rate of the corresponding backward reaction. This assumption shifts the entire thermodynamic analysis away from equilibrium. This contrasts many other postulated models where the forward and backward rates are assumed to be equal, leading to thermodynamic equilibrium and zero net flux. If we assume that the order of magnitude for concentration of the species is the same, then by assuming $k_i^{ZX} \gg k_{-i}^{ZX}$ for $i = \{1,2,3\}$, the cyclic flux ZX can be written as:

$$J_C^{ZX} \approx \frac{k_1^{ZX}k_2^{ZX}k_3^{ZX}k_2^{ZB}k_3^{ZB}[X][D_Z]_{tot}}{k_2^{ZX}k_3^{ZX}k_2^{ZB}k_3^{ZB} + k_1^{ZX}k_2^{ZB}k_3^{ZB}(k_2^{ZX} + k_3^{ZX})[X] + k_1^{ZB}k_2^{ZX}k_3^{ZX}(k_2^{ZB} + k_3^{ZB})[B]} \quad (18)$$

Let J_T^Z be the total transcriptional rate of protein Z in [nM s^{-1}], which is the sum of the transcription rate of Z due to the activity of transcription factor (TF) X, J_T^{ZX}, and the basal activity of basal TF B, J_T^{ZB}. Since protein Z is being transcribed in the third reaction of the proposed mechanism, $J_T^{ZX} = J_3^{ZX}$ and $J_T^{ZB} = J_3^{ZB}$. Thus, $J_T^Z = J_T^{ZX} + J_T^{ZB} = J_3^{ZX} + J_3^{ZB} = J_C^{ZX} + J_C^{ZB}$. For $I = X, B$ we define:

$$\beta^{ZI} = \frac{k_2^{ZI}k_3^{ZI}[D_Z]_{\text{tot}}}{(k_2^{ZI} + k_3^{ZI})}, \quad [\text{nM s}^{-1}], \quad (19)$$

and

$$\kappa^{ZI} = \frac{k_2^{ZI}k_3^{ZI}}{k_1^{ZI}(k_2^{ZI} + k_3^{ZI})}, \quad [\text{nM}], \quad (20)$$

Then, the cyclic flux of ZX (eqn (18)) and the corresponding flux for ZB can be written as follows:

$$J_C^{ZX} = \frac{\beta^{ZX}[X]}{\kappa^{ZX} + [X] + \frac{\kappa^{ZX}}{\kappa^{ZB}}[B]} \quad (21)$$

$$J_C^{ZB} = \frac{\beta^{ZB}[B]}{\kappa^{ZB} + [B] + \frac{\kappa^{ZB}}{\kappa^{ZX}}[X]} \quad (22)$$

In principle, J_T^Z is a function of both activator $[X]$ and $[B]$. However, if the concentration of the basal activator does not change in the system, then the transcription rate of Z is only a function of its main activator X. The basal and maximal transcription rates, $J_{T,\text{Basal}}^Z$ and $J_{T,\text{Max}}^Z$, are defined as the transcription rates when $[X] \to 0$ and $[X] \to \infty$, respectively:

$$J_{T,\text{Basal}}^Z = \frac{\beta^{ZB}[B]}{\kappa^{ZB} + [B]} \quad (23)$$

$$J_{T,\text{Max}}^Z = \beta^{ZX} \quad (24)$$

Notice that $J_{T,\text{Basal}}^Z$ is constant for a specified $[B]$. In addition, from eqn (24), β^{ZX} can be conceptually defined as the ***maximal transcription rate*** $J_{T,\text{Max}}^Z$ of protein Z due to the activity of its transcription factor X. This parameter includes a cooperativity term, ϖ^{ZX} which deals with various input logics. The cooperativity term (ϖ^{ZX}) quantifies the interaction between two proteins bound on two sites. As indicated earlier, each

 This journal is © The Royal Society of Chemistry 2012

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

cycle represents the binding of a transcription factor on a DNA site to induce transcription. Hence, cooperativity term in our model explicitly allows binding interactions to be taken into account. In principle, $\beta^{ZX} = \varpi^{ZX}\,\beta^{Z}$ where β^{Z} is a common maximal transcription rate for all the cycles involved in the transcription of Z. Essentially, different β^{ZI} values are obtained by incorporation of different cooperativity terms in the cycles, represented by ϖ^{ZI}. This effect was previously described by Buchler *et al.*, 2003.[29] In general, ϖ^{ZI} is low for *OR* input logic and high for *AND* input logics. Moreover, the concentration of X required to reach half of $(J^{Z}_{T,\mathrm{Max}} - J^{Z}_{T,\mathrm{Basal}})$ known as the ***activation coefficient*** K^{X} is determined by the following relationship between κ^{ZX} and κ^{ZB}:

$$K^{X} = \kappa^{ZX}\left(\frac{[B] + \kappa^{ZB}}{\kappa^{ZB}}\right) \tag{25}$$

Fig. 2C shows the approximate solution for highly irreversible reactions far away from thermodynamic equilibrium. ***Note that the transcription rate of Z follows a first order Hill's function with basal transcriptional activity***:

$$J^{Z}_{T} = J^{Z}_{T,\mathrm{Basal}} + \frac{(J^{Z}_{T,\mathrm{Max}} - J^{Z}_{T,\mathrm{Basal}})[X]}{K^{X} + [X]} \tag{26}$$

When X transcribes Z in absence of the basal activity ($[B] = 0$, $J^{Z}_{T,\mathrm{Basal}} = 0$), then $K^{X} = \kappa^{ZX}$ and

$$J^{Z}_{T} = \frac{\beta^{ZX}[X]}{\kappa^{ZX} + [X]} \tag{27}$$

which is same as the first order Hill's equation.

2.1.4 Energy of activation. Because these reactions are highly irreversible and thus are far apart from chemical equilibrium, the energy dissipated is high. As the reaction progresses, the rate at which energy is dissipated is expressed by the product between the flux and the chemical potential of reaction. Whereas, the total heat dissipation rate HDR^{Z} [J s^{-1}] is given by the addition of the heat dissipated by each cycle:

$$\begin{aligned}\mathrm{HDR}^{Z} &= \mathrm{HDR}^{ZX} + \mathrm{HDR}^{ZB}\\ &= -\left(\sum_{i=1}^{3} J_i^{ZX}\Delta\mu_i^{ZX} + \sum_{i=1}^{3} J_i^{ZB}\Delta\mu_i^{ZB}\right)\\ &= -\sum_{I=X,B}\sum_{i=1}^{3} J_i^{ZI}\Delta\mu_i^{ZI}\end{aligned} \tag{28}$$

At steady state:

$$\begin{aligned}\mathrm{HDR}^{ZI} &= -(J_1^{ZI}\Delta\mu_1^{ZI} + J_2^{ZI}\Delta\mu_2^{ZI} + J_3^{ZI}\Delta\mu_3^{ZI})\\ &= -J_C^{ZI}(\Delta\mu_1^{ZI} + \Delta\mu_2^{ZI} + \Delta\mu_3^{ZI})\\ &= -J_C^{ZI}\Delta\mu_C^{ZI}\end{aligned} \tag{29}$$

where

$$\begin{aligned}\Delta\mu_C^{ZI} &= \Delta\mu_1^{ZI} + \Delta\mu_2^{ZI} + \Delta\mu_3^{ZI}\\ &= \Re T\ln\left(\frac{[I_Z][Z]}{K_1^{ZI}K_2^{ZI}K_3^{ZI}[I]}\right)\end{aligned} \tag{30}$$

is defined as the chemical potential of the cycle for I = TFs X, B.

Here, we introduce the concept of specific dissipation energy (SDE), defined by us as the ratio between the heat dissipation rate to the input mass flux required to keep the system under non-equilibrium steady state (NESS) conditions:

$$\mathrm{SDE}^{Z} = \mathrm{HDR}^{Z}/J_X \tag{31}$$

2.2 Repression of *Z* by *X*

2.2.1 Mechanism. Similar to the activation case, the mechanism of repression by repressor X can be divided in a three step process (Fig. 3):

Step 1: As in the activation case, the transcription factor X (in this case repressor) binds a free DNA site of the promoter region of protein Z (D_Z) to form an occupied DNA site (D_{ZX}): $D_Z + X \leftrightarrow D_{ZX}$. If K_1^{ZX} [s^{-1} nM^{-1}] and K_{-1}^{ZX} [s^{-1}] are the forward and backward kinetic constants, respectively, and $K_1^{ZX} = k_1^{ZX}/k_{-1}^{ZX}$ [nM^{-1}] is the association equilibrium constant, then the net flux and the chemical potential for this reaction are:

$$J_1^{ZX} = k_1^{ZX}[D_Z][X] - k_{-1}^{ZX}[D_{ZX}] \tag{32}$$

$$\Delta\mu_1^{ZX} = \Re T\ln\left(\frac{k_{-1}^{ZX}[D_{ZX}]}{k_1^{ZX}[D_Z][X]}\right) = \Re T\ln\left(\frac{[D_{ZX}]}{K_1^{ZX}[D_Z][X]}\right) \tag{33}$$

Step 2: Since X is a repressor, RNA polymerase (R) cannot bind to the occupied DNA site (D_{ZX}). In turn, the X-bound site changes its configuration into another energetic state (D^{*}_{ZX}) : $D_{ZX} \leftrightarrow D^{*}_{ZX}$. Let k_2^{ZX} [s^{-1}] and k_{-2}^{ZX} [s^{-1}] be the forward and backward kinetic constants, respectively, and $K_2^{ZX} = k_2^{ZX}/\,k_{-2}^{ZX}$ the association equilibrium constant. Then, the net flux and chemical potential for this step are:

$$J_2^{ZX} = k_2^{ZX}[D_{ZX}] - k_{-2}^{ZX}[D^{*}_{ZX}] \tag{34}$$

$$\Delta\mu_2^{ZX} = \Re T\ln\left(\frac{k_{-2}^{ZX}[D^{*}_{ZX}]}{k_2^{ZX}[D_{ZX}]}\right) = \Re T\ln\left(\frac{[D^{*}_{ZX}]}{K_2^{ZX}[D_{ZX}]}\right) \tag{35}$$

Step 3: In this last step, the activated X-bound site releases the free DNA site and the co-repressor necessary for this process (X_Z). Because RNA polymerase is not bound, there is no recruitment of amino acids (AA_Z) and transcription and translation of Z does not proceed: $D^{*}_{ZX} \leftrightarrow D_Z + X_Z$. If k_3^{ZX} [s^{-1}] and k_{-3}^{ZX} [s^{-1} nM^{-1}] are the forward and backward kinetic constants, respectively, and $K_3^{ZX} = k_3^{ZX}/k_{-3}^{ZX}$ [nM^{1}] is the association equilibrium constant for this reaction, then the reaction flux and reaction chemical potential are given by:

$$J_3^{ZX} = k_3^{ZX}[D^{*}_{ZX}] - k_{-3}^{ZX}[D_Z][X_Z] \tag{36}$$

$$\Delta\mu_3^{ZX} = \Re T\ln\left(\frac{[D_Z][X_Z]}{K_3^{ZX}[D^{*}_{ZX}]}\right) \tag{37}$$

The described steps 1, 2 and 3 are represented as a red triangular cycle in Fig. 3A and B. The basal activity (blue cycle in Fig. 3A and B), in which a basal transcription factor B binds the free promoter region of Z, initiating the cycle and further transcription of Z, follows the activation mechanism shown earlier.

2.2.2 Mass balance equations. The differential equations describing the repression system are the same as those presented

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

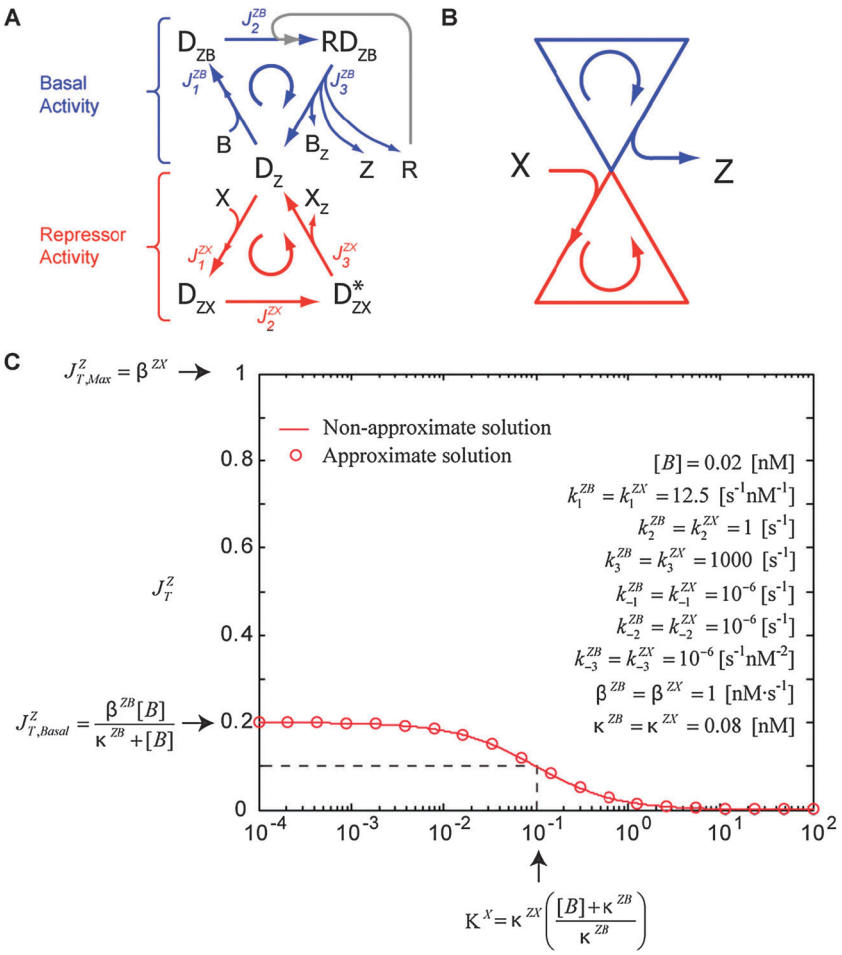


Fig. 3 Cyclic representation of the repression pathway of protein *Z* by transcription factor *X*. (A) Cyclic transcriptional repression schematic of gene *Z* (red cycle). Detailed mechanism showing both basal and repressor activity. (B) Simplified representation of the cyclic repression model. (C) Transcription rate of *Z* as a function of the repressor concentration *X*. Curve follows a first order Hill's function, with maximal transcription rate given by the basal activity of 0.2 [nM s^{-1}] and activation coefficient of 0.1 [nM].

in the activation mechanism with only two differences. First the mass balance equation of $[Z]$ does not include J_3^{ZX}, because the transcription of Z is only due to the basal activity. Second, RD_{ZX} is conceptually replaced by D^*_{ZX}. For the repression process, the total concentration of DNA sites is now written as:

$$[D_Z]_{\text{tot}} = [D_Z] + [D_{ZX}] + [D_{ZB}] + [D^*_{ZX}] + [RD_{ZB}] \quad (38)$$

2.2.3 Steady state solution. At non-equilibrium steady state (NESS), it can be shown that eqn (19)–(23) are also valid in the repression mechanism. However, when X is a repressor, the total transcription rate of Z equals only the basal cyclic flux: $J_T^Z = J_T^{ZB} = J_3^{ZB} = J_C^{ZB}$. As $[X] \to 0$, the basal transcription rate $J^Z_{T,\text{Basal}}$ (eqn (23)) follows the same relationship as in the activation case: the basal transcription rate is independent of the activating or repressing nature of X. However, as expected, the basal transcription rate corresponds to the maximal transcription rate. As $[X] \to \infty$, the transcription rate of Z decreases and approaches to zero: $J^Z_{T,\text{Min}} = 0$. Moreover, the concentration of X required to reach half of $(J^Z_{T,\text{Basal}} - J^Z_{T,\text{Min}}) = J^Z_{T,\text{Basal}}$, K^X, is determined by the same relationship between κ^{ZX} and κ^{ZB} shown in eqn (25). Fig. 3C shows the solution for irreversible reactions that are far away from thermodynamic equilibrium. Finally, it is observed that the transcription rate of Z follows a Hill's first order equation for repressor activity:

$$J_T^Z = \frac{J^Z_{T,\text{Basal}}}{1 + \frac{[X]}{K^X}} \quad (39)$$

2.2.4 Energy of repression. Eqn (28)–(29) express the general concept of the HDR^Z [J s^{-1}] associated with any transcription process, and thus are valid for representing the repression scenario. However, the cyclic chemical potential equation given by eqn (30) must be changed since the overall reactions are different:

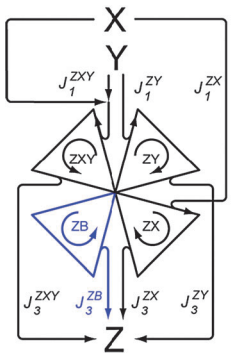
$$\Delta\mu_C^{ZI} = \Delta\mu_1^{ZI} + \Delta\mu_2^{ZI} + \Delta\mu_3^{ZI} = \Re T \ln\left(\frac{[I_Z][P_Z]}{K_1^{ZI} K_2^{ZI} K_3^{ZI} [I]}\right) \quad (40)$$

where

$$[P_Z] = \begin{cases} [Z], & \text{if } I = B \text{ (or } I \text{ is an activator)} \\ 1, & \text{if } I = X \text{ (or } I \text{ is a repressor)} \end{cases}$$

 This journal is © The Royal Society of Chemistry 2012

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013



OR	Cycle			J_1			J_3		
	ZX	ZY	ZXY	ZX	ZY	ZXY	ZX	ZY	ZXY
X→ Z ←Y	↻	↻	↻	↓	↓	↓	↓	↓	↓
X→ Z ⊣Y	↻	↻	↻	↓	↓	↓	↓		↓
X⊣ Z ←Y	↻	↻	↻	↓	↓	↓		↓	↓
X⊣ Z ⊣Y	↻	↻	↻	↓	↓	↓			

AND	Cycle			J_1			J_3		
	ZX	ZY	ZXY	ZX	ZY	ZXY	ZX	ZY	ZXY
X→ Z ←Y	↻	↻	↻	↓	↓	↓			↓
X→ Z ⊣Y	↻	↻	↻	↓	↓	↓	↓		
X⊣ Z ←Y	↻	↻	↻	↓	↓	↓		↓	
X⊣ Z ⊣Y	↻	↻	↻	↓	↓	↓	↓	↓	

Fig. 4 Multigene Regulation: Cyclic model of gene Z transcription by two TF inputs X and Y. The schematic shows three-step transcription process of Z by X and Y. Black arrows become green (activator behavior) or red (repressor behavior) based on the individual activation or repressor nature of X and Y, and the input logic AND or OR. When a transcription factor acts as a repressor, the corresponding output flux of Z, J_3, is not present and not shown in the right-side tables. The adjoining table shows various combinatorial Boolean transcription logics of AND and OR. Activation is shown by green arrows and repression by red.

Solution of the mass balance will allow the computation of J_X for the final determination of SDE^Z, as defined by eqn (31).

2.3 Transcriptional regulation of Z by two transcription factors X and Y

In this section, we will analyze multi-gene regulation. Specifically, we will present formulation for the transcription (activation and repression) of the protein Z by two transcription factors X and Y (Fig. 4).

2.3.1 Mechanism. The basic mechanism by which two transcription factors activate or repress protein transcription is the three step process described previously for activation and for repression in sections 2.1 and 2.2. Here, we study multi-gene regulation and allow both X and Y (and basal transcription factor B) to compete for a free DNA site in the promoter region of Z. The competitive binding is irrespective of whether X (or Y) is an activator or a repressor. However, in addition to the individual TF binding to D_Z, there is also a joint interaction between X and Y, so in principle the complex XY may be considered as an additional transcription factor whose concentration is given by the product between the concentration of X and Y. In other words, $[XY] = [X][Y]$. The activator or repressor nature of this complex depends on two factors:

1. The independent activation-repression nature of X and Y.
2. The Boolean input logic used to transcribe Z which can be AND or OR.

In order to explain the Boolean input logic, let us assume that both X and Y are activators of Z. In principle, protein Z should be transcribed at least from the individual binding of X and Y. However, this is not completely true. If both TF X *and* Y are required to transcribe Z (Boolean input logic AND), then transcription cannot proceed with individual binding of X or Y. Therefore when individually present, TF X or Y will bind to the promoter site without transcribing Z, and thus acts as repressor contrary to their activating nature. Therefore, when both X and Y are present for AND input logic, transcription of Z can only proceed from the XY cycle (and from the basal cycle ZB which by definition is always transcribing Z). This XY complex will bind the promoter region, recruit RNA polymerase and finally transcribe (and translate) Z. On the other hand, if only one of the two transcription factors is required (Boolean input logic OR) for transcription of Z, then the individual binding as well as the complex XY binding, will lead to formation of Z, as shown in Fig. 4, which gives a simplified representation of the four cycles is presented. The cycles, entering and leaving fluxes for ZX, ZY and ZXY are in black color. The activator and repressor behaviors are presented as green and red arrows, respectively, according to the input logic and independent nature of X and Y as shown in the tables in Fig. 4. When the cycle ZI behaves as a repressor (for $I = X, Y, XY$), the transcription flux J_3^{ZI} is not present (as shown in Fig. 3), and thus the corresponding arrow is not shown in Fig. 4.

2.3.2 Mass balance solution. If the irreversibility condition is satisfied, then the cyclic fluxes can be approximated as follows:

$$J_C^{ZI} = \frac{\beta^{ZI}[I]}{\kappa^{ZI} + \sum_{J=B,X,Y,XY} \frac{\kappa^{ZI}}{\kappa^{ZJ}}[J]}, \tag{41}$$

where β^{ZI} and κ^{ZI} are defined by eqn (19) and (20), respectively. Fig. S1 presents the transcription rate surface for both input logics OR and AND, using the approximated solution.

For Boolean input logic OR, $J_T^Z = J_C^{ZB} + J_C^{ZX} + J_C^{ZY} + J_C^{ZXY}$. If both basal activity and the transcription rate from the combined activity of X and Y are zero (*i.e.*, β^{ZXY} is very low and κ^{ZXY} is very high), the expression for the total transcription rate of Z, given by eqn (41) equals the relationship given by Mangan and Alon (2003):[5]

$$J_T^Z = \frac{\beta^{ZX}\frac{[X]}{\kappa^{ZX}}}{1 + \frac{[X]}{\kappa^{ZX}} + \frac{[Y]}{\kappa^{ZY}}} + \frac{\beta^{ZY}\frac{[Y]}{\kappa^{ZY}}}{1 + \frac{[Y]}{\kappa^{ZY}} + \frac{[X]}{\kappa^{ZX}}} \tag{42}$$

For Boolean input logic AND *case*, transcription rate of Z is given by the basal activity and the combined interaction of X and Y: $J_T^Z = J_C^{ZB} + J_C^{ZXY}$. If basal activity is zero and $\kappa^{ZXY} = \kappa^{ZX}\kappa^{ZY}$, then the expression for J_T^Z (41) also equals the relationship given by Mangan and Alon (2003):[5]

$$J_T^Z = \frac{\beta^{ZX}\frac{[X]}{\kappa^{ZX}}}{1 + \frac{[X]}{\kappa^{ZX}}} \cdot \frac{\beta^{ZY}\frac{[Y]}{\kappa^{ZY}}}{1 + \frac{[Y]}{\kappa^{ZY}}} \tag{43}$$

2.3.3 Energy of transcription. At steady state, the total HDR for this four-cycle system can be written as follows:

$$\mathrm{HDR}^Z = -\sum_{I=B,X,Y,XY} J_C^{ZI}\Delta\mu_C^{ZI} \tag{44}$$

where

$$\Delta\mu_C^{ZI} = \Re T \ln\left(\frac{[I_Z][P_Z]}{K_1^{ZI}K_2^{ZI}K_3^{ZI}[I]}\right) \tag{45}$$

Analytical solution for HDR^Z is highly complex and input logic dependent in this scenario, thus numerical solutions are required.

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

The involved steps and generalized formulation for cyclic transcriptional network model for FFL is presented in Table 1.

3. Multiobjective optimality of transcriptional networks

3.1 Multiobjective optimization

A multiobjective optimization is a problem involving several competing objectives and constraints. The best solution of multiobjective optimization is the one that satisfies the conflicting objectives. A Pareto solution is one where any improvement in one objective can only take place at the cost of another objective. A Pareto set is a set of Pareto-optimal solutions. The "anchor value" is obtained for a particular objective function if that function alone is optimized, given the bounds on the design parameters.

3.2 Pareto concept

Fig. 5A presents a general scheme of a Pareto set for a bi-objective maximization and minimization problem. As seen in the figure, generally there is a tradeoff between cellular objectives of robustness (f_1) and transcription rate (f_2). If objective function f_1 alone is optimized (maximized), then the optimal value is $f_1^{\max}$ (shown as point P_1). Similarly, if objective function f_2 alone is optimized then the optimal value is $f_2^{\max}$ (shown as point P_2). Here, $f_1^{\max}$ and $f_2^{\max}$ are the anchor values for objective functions f_1 and f_2, respectively. The ideal or Utopian optimal solution ($f_1^{\max}$, $f_2^{\max}$) obtained by the individual maximization of the objective functions is not a feasible solution of the multiobjective optimization problem. As seen in Fig. 5A, the line joining points P_1 and P_2 defines the boundary of the feasible space and is termed as the Pareto frontier. That is, for every optimal solution on arc P_1–P_2, it is not possible to improve both objectives simultaneously. If one objective is improved, it must be at the expense of the other one. These optimal points on the arc are often referred to as extreme Pareto points. In view of their stated characteristics, Pareto points are usually the candidates of choice in the process of multi-objective optimization.

3.3 Normalized constraint (NC) method

The NC method is based on the design space reductions using reduction constraints. The reduction constraint is constructed by ensuring the orthogonality by constructing the dot product between the normal w and r_0 an arbitrary point on a plane. The vector equation of a plane is expressed as $w\cdot(r - r_0) = 0$. To solve for multiobjective solutions, a reduced feasible space is constructed using the previous equation as $w\cdot(r - g) \leq 0$ where g is any point in the feasible space. In Fig. 5B (1) the non-normalized design space and the Pareto frontier of a bi-objective problem is shown. Fig. 5B (3) represents the normalized Pareto frontier in the normalized design space. In the normalized objective space, while the utopia point is at the origin, all anchor points are one unit away from the utopia point. A bar over a variable implies that it is normalized. The two anchor points denoted by g_1^* and g_2^*, are obtained by successively minimizing the first and second design metrics. The line joining these two points is the *utopia line*. The actual optimization takes place in the normalized objective space. Let $\bar{g}$ be the normalized form of g and g^u, the utopia point is defined as $g^u = [g_1(x^{1^*})\ g_2(x^{2^*})]^T$ where ℓ_1 and ℓ_2 be the distances between g^{2^*} and g^{1^*}, and the Utopia point, g^u, respectively (Fig. 5B). Then $\ell_1 = g_1(x^{2^*}) - g_1(x^{1^*})$ and $\ell_2 = g_2(x^{1^*}) - g_2(x^{2^*})$. The normalized design objectives can then be evaluated as:

$$\bar{g}^T = \left[\frac{g_1(x) - g_1(x^{1^*})}{\ell_1}\ \frac{g_2(x) - g_2(x^{2^*})}{\ell_2}\right]$$

$\bar{N}_1$ is defined as the direction from $\bar{g}^{1^*}$ to $\bar{g}^{2^*}$, yielding $\bar{N}_1 = \bar{g}^{2^*} - \bar{g}^{1^*}$. Next, the utopia line is divided into $m_1 - 1$ segments, resulting in m_1 points. A normalized increment, δ_1 along the direction $\bar{N}_1$ for a prescribed number of solutions, m_1, is obtained as:

$$\delta_1 = \frac{1}{m_1 - 1}$$

As seen in Fig. 5B, the next step involves generating a set of evenly distributed points on the utopia line as $\bar{X}_{Pj} = \alpha_{1j}\bar{g}^{1^*} + \alpha_{2j}\bar{g}^{2^*}$ where $0 \leq \alpha_{1j} \leq 1$, $\alpha_{1j} + \alpha_{2j} = 1$ and α_{1j} is incremented by δ_1 between 0 and 1 (Fig. 5B), with values of j as $j \in \{1, 2, \ldots, m_1\}$.

Fig. 5B (4) shows one of the generic points intersecting the segments used to define a normal to the utopia line. This normal line is used to reduce the feasible space as indicated in the figure. As observed in the figure, if we minimize $\bar{g}_2$, the resulting optimum point is $\bar{g}^{2^*}$. By translating the normal line, a corresponding set of solutions are generated. This is done by generating a corresponding set of Pareto points by solving a succession of optimization runs of Problem P2 (see Text S2). Each optimization run corresponds to a point on the utopia line. For each generated point on the utopia line, solve for the jth point:

Problem P2 (for jth point):

$$\begin{aligned} & \min_x\{\bar{g}_2(x)\} \\ \text{subject to:}\quad & h_k(x) = 0, 1 \leq k \leq r \\ & f_j(x) = 0, 1 \leq j \leq s \\ & x_l \leq x_i \leq x_u 1 \leq i \leq n_x \\ & \bar{N}_1(\bar{g} - \bar{X}_{Pj})^T \leq 0 \end{aligned}$$

This results in a set of vectors for the design parameters. Each Pareto point gives one vector x. Then, cellular objective functions are computed by evaluating the non-normalized design metrics that correspond to each Pareto point. The non-normalized design objectives can be obtained by using the relation $g = [\bar{g}_1\ell_1 + g_1(x^{1^*})\ \bar{g}_2\ell_2 + g_2(x^{2^*})]^T$.

Importantly, the generation of the set of Pareto points is performed in the normalized objective space, resulting in critically beneficial scaling properties. The steps involved and the essential mathematical formulation for the NC method are presented in Table S1 for the n-objective case.

3.4 Description of non-equilibrium TRNs

Non-equilibrium processes require an external signal, flux or driving force to maintain the system far away from equilibrium. The role of non-equilibrium thermodynamics in small-scale systems such as biological molecular machines and RNA folding/unfolding has sparked a surge of interest in the field.[30–33] Equilibrium systems

 This journal is © The Royal Society of Chemistry 2012

View Article Online

Table 1 Cyclic transcriptional network mathematical forumlation for FFL

A. Steps for estimating specific dissipation energy (SDE)

1. Set global parameters:
Concentration of basal activator of Y, $[A]$
Concentration of basal activator of Z, $[B]$
Concentration of total available sites for transcription of Y, $[D_Y]_{tot}$
Concentration of total available sites for transcription of Z, $[D_Z]_{tot}$
Degradation rates for each species, α
Temperature, T

2. For each of the 8 cycles forming the FFL motif YA, YX, ZB, ZX, ZY, ZXY (in general, cycle WI corresponds to transcription or repression of W due to transcription factor I):

For step 1:
Set the kinetic parameters k_1^{WI} and k_{-1}^{WI}
State the net reaction flux as $J_1^{WI} = k_1^{WI}[D_W][I] - k_{-1}^{WI}[D_{WI}]$
State the net change in chemical potential as $\Delta\mu_1^{WI} = \Re T \ln\left(\frac{k_{-1}^{WI}(D_{WI})}{k_1^{WI}[D_W][I]}\right)$

For step 2:
Set the kinetic parameters k_2^{WI} and k_{-2}^{WI}
State the net reaction flux as $J_2^{WI} = k_2^{WI}[D_{WI}] - k_{-2}^{WI}[RD_{WI}]$
State the net change in chemical potential as $\Delta\mu_2^{WI} = \Re T \ln\left(\frac{k_{-2}^{WI}[RD_{WI}]}{k_2^{WI}[D_{WI}]}\right)$

For step 3:
Set the kinetic parameters k_3^{WI} and k_{-3}^{WI}
State the net reaction flux as $J_3^{WI} = k_3^{WI}[RD_{WI}] - k_{-3}^{WI}[D_W][I_W][P_W]$
State the net change in chemical potential as $\Delta\mu_3^{WI} = \Re T \ln\left(\frac{k_{-3}^{WI}[D_W][I_W][P_W]}{k_3^{WI}[RD_{WI}]}\right)$

If $l^{WI} = 1$ (*i.e.*, I is an activator of W), $[PW] = [W]$. **If** $l^{WI} = 0$ (*i.e.*, I is a repressor of W), $[P_W] = 1$. The attached table shows the logic parameter l^{WI} for each of the 8 FFLs using AND and OR logics.

3. For each of the involved species state the mass conservation equations. This includes transcription factors (A, B, X, Y, Z), released cofactors (A_Y, X_Y, B_Z, X_Z, Y_Z, XY_Z, or in general, I_W), and occupied DNA sites (in general, $[D_W]$, $[D_{WI}]$ and $[RD_{WI}]$), resulting in a set of 24 ODE's

4. Use the conservation equation of the total available sites:

$$[D_Y]_{tot} = D_Y + \sum_{I=\{A,X\}}[D_{YI}] + \sum_{I=\{A,X\}}[RD_{YI}]$$

$$[D_Z]_{tot} = D_Z + \sum_{I=\{B,X,Y,XY\}}[D_{ZI}] + \sum_{I=\{B,X,Y,XY\}}[RD_{ZI}]$$

5. Solve the steady state equations to find the concentrations of the involved species

6. Calculate the net change in chemical potential and the net flux for each step in each cycle

7. Calculate the total Heat Dissipation Rate (HDR)**:**

$$\text{HDR} = \sum_{\text{cycle}}\sum_{\text{step}}(J\Delta\mu)$$

8. Calculate the external consumption flux of X**:**

$$J_X = J_1^{YX} + J_1^{ZX} + J_1^{ZXY} + \alpha[X]$$

9. Calculate the Specific Dissipation Energy (SDE)**:**

$$\text{SDE} = \text{HDR}/J_X$$

B. Logic parameter l^{WI} as a function of the FFL motif with AND or OR logic.

	OR Logic								AND Logic							
	C1	C2	C3	C4	I1	I2	I3	I4	C1	C2	C3	C4	I1	I2	I3	I4
l^{YA}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
l^{YX}	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
l^{ZB}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
l^{ZX}	1	0	0	1	1	0	0	1	0	0	1	1	1	1	0	0
l^{ZY}	1	1	0	0	0	0	1	1	0	1	1	0	0	1	1	0
l^{ZXY}	1	1	0	1	1	0	1	1	1	0	0	0	0	0	0	1

are generally governed by classical thermodynamics and have reversible work equal to the Gibbs free energy change at constant temperature and pressure. Although the combination of non-equilibrium thermodynamics with modern network theory dates back more than 40 years,[34] it was not until the recent formulation of non-equilibrium steady-state (NESS) analysis that it was utilized extensively in biological systems.[32,35] In NESS analysis, Gibbs free energy is unequal to the work

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

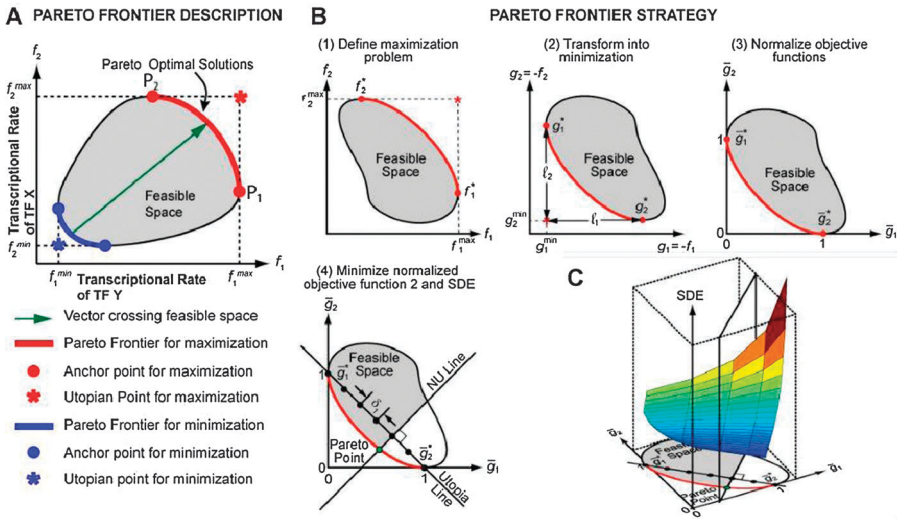


Fig. 5 Optimal SDE serves as the basis for efficient resource utilization in metabolic and transcriptional networks. (*A*) Pareto frontiers for a bi-objective maximization and minimization problem.[17] If objective functions f_1 (for instance, robustness) and f_2 (for example, transcriptional rate) alone are individually maximized, then the optimal values are f_1^{max} (point P_1) and f_2^{max} (point P_2), respectively. Here g_1^* and g_2^* are the anchor values for design objectives g_1 and g_2, respectively. The ideal or Utopian solution (g_1^*,g_2^*) obtained by the individual maximization of the objective functions is an infeasible solution of the multiobjective optimization problem. The arc P_1P_2 is defined as the Pareto frontier containing multiobjective-optimal or tradeoff solutions. Pareto-optimal and non-Pareto utilization of available energy and resources are shown as the corresponding circular area in the triangle for two objectives of a cellular TRN. (B) Steps involved for obtaining Pareto frontiers. First, maximization problem between cellular objectives is defined. The maximization problem is then converted into minimization and normalized in a feasible space using anchor points. The utopian line is joined between anchor points and the normal to utopian line is moved along the utopian line to expand or restrict the feasible space. Optimization of objectives in reduced feasible space (*i.e.* shaded region) eventually gives Pareto solutions on the maximal surface of feasible space. (C) Using bi-level optimization SDE was found to be minimal at the Pareto frontier compared to any other point in the feasible space.

done on the system because of heat/work dissipation. Moreover, chemical potential is equivalent to heat dissipated by the system (or work done on the system) which in turn is equivalent to entropy generated by the system.[32,35,36] It has been shown earlier that the average work dissipated along any trajectory between different non-equilibrium states is always positive.[37] This result led to our two point hypothesis. First, the heat dissipation normalized to the input mass flux, defined by us as specific dissipation energy (SDE), should be minimal for biological systems under optimal nutrient conditions. Second, systems with lowest SDE should win out during evolutionary selection. Specifically, the dissipated work will be directly proportional to the input mass flux or raw materials utilized in moving from one state to another, hence, normalized heat dissipation may serve as a criterion for choosing a trajectory to move from between different states. Interestingly, a recent study[38] observed that robustness is inversely related to dissipation cost in signal transduction networks.

To evaluate the energetic cost involved during transcription, we utilized previous NESS analysis and developed a novel non-equilibrium thermodynamic kinetic formulation for gene transcription. Since the developed modeling framework for feed-forward loop network motifs is further used for computationally expensive nonlinear multiobjective Pareto optimal solutions, we made several assumptions to obtain simplified steady state transcriptional rate solutions. Although these assumptions were warranted to reduce the complexity in our modeling framework, the solutions obtained using our model is similar to the current paradigm of using Hill's approximation for modeling transcription. Another advantage of simplifying our model in Hill's form is that it allows easy generalization of smaller network motifs into complex large-scale networks and straightforward integration of well-developed transcriptional network theories based on Hill's framework in our model. The developed non-equilibrium cyclic TRN model (a) explicitly deals with as many interactions as required with no limit on interactions (activation and repression) (Fig. 2–4), (b) uses a competitive binding scheme for Boolean input logics (Fig. 2–4), (c) can be easily generalized to complex networks, and (d) provides estimation of the SDE for a gene-transcription factor (TF) combination during a transcription-translation process. As mentioned above, in order to simplify our model we made several assumptions. We utilized a lumped parameter approach where energetic terms involved during various molecular steps are lumped into a single reaction. This was done not only to obtain a simplified mass flux but also because of complexity involved during TF-DNA interactions. Since our model clearly provides similar mass flux solutions when compared to Hill's model in terms of lumped reaction, we thus used the lumped reaction for estimating energetic in all of our analysis. Various mechanisms involved during TF-based protein synthesis (TF-DNA binding energetics,[39,40] mRNA binding effecting translation,[33,41] structural changes,[42] chromatin conformations,[43] and post-translational changes[44]) may contribute towards energetic-cost of cellular regulation. However, for the sake of simplicity, heat dissipation obtained from the cyclic TRN model developed here lumps these steps into reactions having pseudo rate constants. We integrated these steps in various reaction fluxes by explicitly taking into account appropriate concentrations as indicated earlier in Section 2.

 This journal is © The Royal Society of Chemistry 2012

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

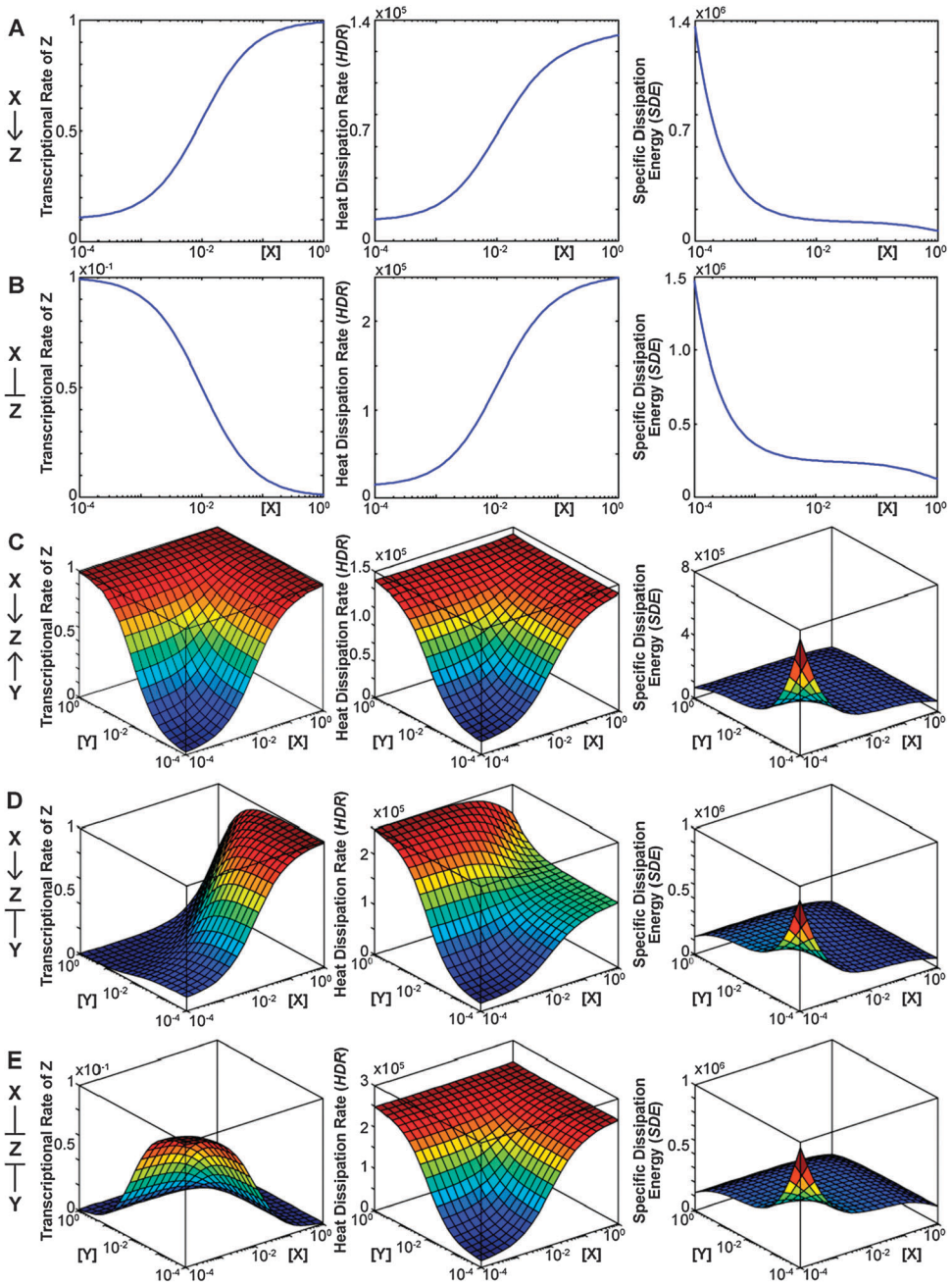


Fig. 6 Energetic cost analysis of TRNs using kinetic model. (A) Transcription of Z as a function of activator X using non-equilibrium cyclic activation model. Curve follows a first order Hill's function. The corresponding heat dissipation rate (HDR) increases with an increase in transcription. Conversely, normalized energetic cost *i.e.* specific dissipation energy (SDE) decreases with increase in activator concentration and is minimal at highest transcription. (B) Transcriptional repression of Z as a function of repressor X using kinetic model. HDR is highest at maximal repression and SDE is minimal at the maximal repressor concentration. (C–E) Transcription rate of Z with two factor regulation with OR input logic using the developed kinetic model for multifactor regulation cases: both X and Y are activators (C), X is activator and Y is repressor (D), X and Y are repressors (E). HDR and SDE for multi factor cases follow individual activator and repressor behavior.

4. Results and discussion

4.1 Analysis of energetic cost for transcriptional activation and repression

We first analyzed total heat dissipation rate and specific dissipation energy for various transcriptional regulatory processes. During transcriptional activation (Fig. 6A), with an increase in TF activator concentration there is a corresponding increase in transcription and heat dissipation rate. Interestingly, energetic cost (SDE) was found to be minimal at the highest activator concentration. During transcriptional repression (Fig. 6B), with an increase in repressor concentration there is a corresponding decrease in the rate of transcription and an increase in heat dissipation rate. Similar behavior was found for multigene regulation as shown in Fig. 6C–E. Our analysis shows that heat dissipation rate (HDR) increases with increase

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

in cycle flux when driven by activator TF during transcriptional activation. Interestingly, HDR increases even with reduced cycle flux when driven by repressor TF during transcriptional repression. This shows that dissipation rate is dependent only on input flux which may be either a transcriptional activator flux or a transcription repression flux. Notably, SDE is minimal during high activation and repression, thus it could be used as a correlate in either activation and/or repressive linkages in TRN motifs.

4.2 Specific dissipation energy is minimal for biological systems under optimal conditions

Numerous studies abound in which biological systems have been shown to exhibit the property of optimality. However, these studies have utilized either performance or mass resources-based objectives to show that biological systems operate under optimal conditions. Resultantly, the role of energetic under function or mass resources-based optimal conditions has remained unknown for biological systems. We first wanted to assess whether there is any correlation between Pareto-optimal transcriptional rates and SDE, where SDE is the ratio of total heat dissipated by the system to the input mass flux. To achieve this we used a bi-level optimization. In the first step we obtained Pareto optimal transcription fluxes using the Normalized Constraint method (Fig. 5B) and in the second level we obtain transcription fluxes with minimal specific dissipation energy (Fig. 5C).

As seen in Fig. 5B, the normalized constraint Pareto frontier method is based on the design space reductions using reduction constraints. The reduction constraint is constructed by ensuring orthogonality by constructing the dot product between the normal $\vec{w}$ and r_0 an arbitrary point on a plane. The vector equation of a plane is expressed as

$$\vec{w}\cdot(r - r_0) = 0$$

To solve for multiobjective solutions in Fig. 5B, a reduced feasible space is constructed using the above equation as $\vec{w}\cdot(r - g) \leq 0$

$$\begin{aligned} \underset{\vec{x}}{\text{Maximize}} \quad & g = (J_Y, J_Z) \\ \text{subject to:} \quad & f(\vec{x}) = 0 \\ & \text{HDR}_{lb} \leq \text{HDR} \leq \text{HDR}_{ub} \\ & \vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub} \\ & \bar{N}_1(\bar{g} - \bar{X}_{Pj})^T \leq 0 \end{aligned}$$

where N_1 is the normal to the utopia line, $f(\vec{x})$ represents the nonlinear mass balance equations at steady state and $\vec{x} = [[X]\ J_Y\ J_Z]^T$. In the second level of optimization we solved an optimization problem in the reduced feasible space to obtain minimal specific dissipation energy.

$$\begin{aligned} \underset{\vec{x}}{\text{Minimize}} \quad & \text{SDE} \\ \text{subject to:} \quad & f(\vec{x}) = 0 \\ & \text{HDR}_{lb} \leq \text{HDR} \leq \text{HDR}_{ub} \\ & \vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub} \\ & \bar{N}_1(\bar{g} - \bar{X}_{Pj})^T \leq 0 \end{aligned}$$

Remarkably, as we moved normal along the utopian line as shown in Fig. 5B, we found that the corresponding Pareto optimal transcriptional rates were the same as the optimal transcriptional fluxes with minimal SDE. This essentially shows that Pareto optimal transcriptional fluxes indeed are the fluxes that have minimal SDE (Fig. 5C).

To explain the variation of SDE in the transcription space and on Pareto frontier, below we evaluate the energetic cost along the vector from minimal transcription fluxes to the Pareto optimal transcriptional fluxes for both FFL and feedback TRNs.

4.2.1 Correlating pareto optimal surface with dissipative energetics for FFL TRN motifs. The FFL TRN corresponds to a coherent type-1 FFL (C1-FFL) with an *OR* input logic. Essentially, the FFL network is composed of six cycles as shown in Fig. 7A: two of them are involved in the transcription of Y by X, and four in the transcription of Z by both X and Y. The first two cycles correspond to cycle YA (transcription of Y from its basal activator A) and cycle YX (transcription of Y from the activator X), and the last four cycles are cycle ZB (transcription of Z from its basal activator B), cycle ZX (transcription of Z by the activator X), cycle ZY (transcription of Z by activator Y) and cycle ZXY (transcription of Z by both X and Y together). The kinetic constants used in Fig. 7B are shown in Fig. 7A. The basal cycles are colored blue and the other cycles are green, indicating that they are all transcriptionally active. In fact, since this network follows an *OR* input logic for transcription of Z (notice that transcription of Y does not need any input logic, because only one transcription factor is present), transcription of Z can be seen from all of the six cycles. As both external fluxes of Y and Z are desired to be maximized when X is externally provided, there is a tradeoff region between these two objective functions: when more Y is taken out of the system (higher J_Y), the concentration of Y goes down and its contribution towards further transcription of Z decreases, leading to lower J_Z, and *vice versa*. This Pareto frontier is shown in Fig. 7B and the optimization problem can be expressed as follows:

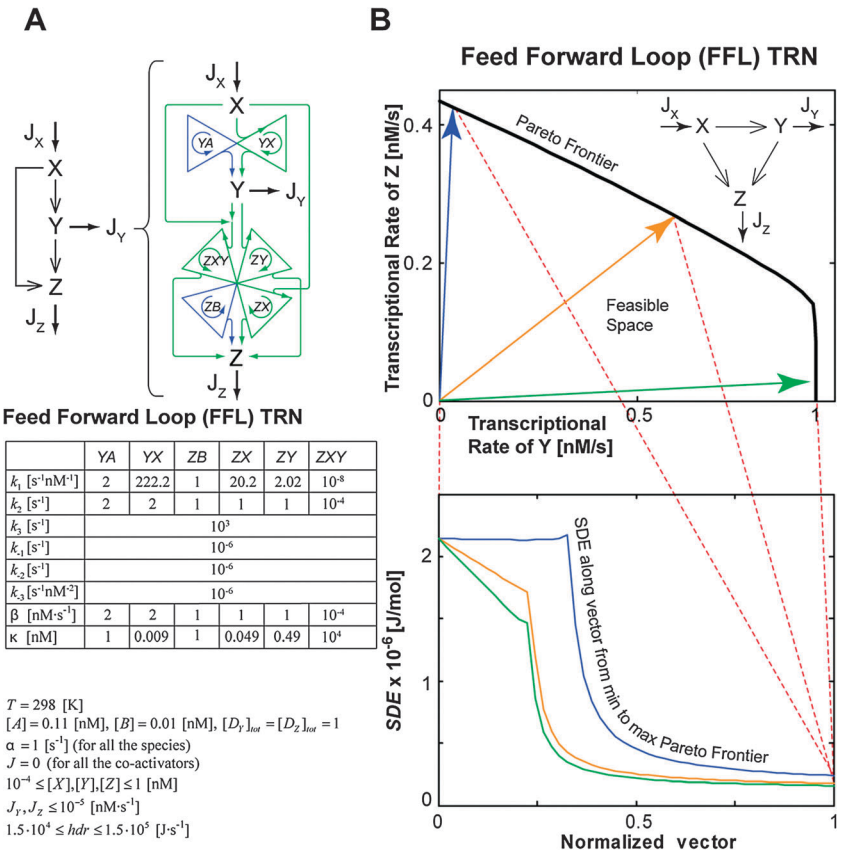
$$\begin{aligned} \underset{\vec{x}}{\text{Maximize}} \quad & (J_Y, J_Z) \\ \text{subject to:} \quad & f(\vec{x}) = 0 \\ & \text{HDR}_{lb} \leq \text{HDR} \leq \text{HDR}_{ub} \\ & \vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub} \end{aligned}$$

where $f(\vec{x})$ represents the nonlinear mass balance equations at steady state and $\vec{x} = [[X]\ J_Y\ J_Z]^T$. Recall that both J_Y and J_Z are negative since these fluxes are leaving the system. It is important to note that at the Pareto frontier there is optimal utilization of nutrient resources. SDE being minimal at the optimal production indicates optimal production with efficient energy utilization. This is in concurrence with the recent experiments where evolutionary adaptation of *E. coli* towards optimal metabolic fluxes and the key role of fitness criteria in evolutionary selection of certain mutations has been shown.[45]

The minimal Pareto frontier can be obtained by simultaneously minimizing J_Y and J_Z for this transcription network. This region is confined to a single point characterized by the vector $\vec{x} = [10^{-4}, -10^{-5}, -10^{-5}]^T$. A vector in the space $J_Y J_Z$ can be traced from the minimal point to any point on the Pareto frontier for maximization (as shown in blue, green and

 This journal is © The Royal Society of Chemistry 2012

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013



	YA	YX	ZB	ZX	ZY	ZXY
k_1 [$s^{-1}nM^{-1}$]	2	222.2	1	20.2	2.02	10^{-8}
k_2 [s^{-1}]	2	2	1	1	1	10^{-4}
k_3 [s^{-1}]	10^3					
k_{-1} [s^{-1}]	10^{-6}					
k_{-2} [s^{-1}]	10^{-6}					
k_{-3} [$s^{-1}nM^{-2}$]	10^{-6}					
β [$nM{\cdot}s^{-1}$]	2	2	1	1	1	10^{-4}
κ [nM]	1	0.009	1	0.049	0.49	10^4

Fig. 7 (A) The Coherent Type-1 FFL with *OR* input logic. Parameters and optimization constraints are shown. Basal TF refers to the basal activator (blue cycle), individual TF refers to only one transcription factor (for instance, X or Y in the transcription of Z), and combined TF refers to the joint interaction between the individual activators (for instance, XY in the transcription of Z). (B) Pareto frontiers for transcriptional regulatory networks (TRN) were obtained by considering the maximization of protein production rates of Y and Z as the appropriate TRN objectives. Three node feedforward TRNs follow the Pareto dominance rule. SDE was found to be minimal at the Pareto frontier compared to the non-Pareto feasible solutions. Following the solutions on the line vector from Pareto frontier of minimization to the Pareto frontier of maximization there is a constant decrease in the SDE and it is minimal at the Pareto frontier optimal transcriptional rates. Here, basal kinetic parameters and the basal maximal transcription rates and activation coefficients were selected in order to have $J^Y_{T,\mathrm{Basal}} = 0.1{\cdot}J^Y_{T,\mathrm{Max}}$ and $J^Z_{T,\mathrm{Basal}} = 0.001{\cdot}J^Z_{T,\mathrm{Max}}$. The kinetic parameters of the other cycles were estimated in order to get the following activation coefficients: $K^{YX} = 0.01$, $K^{ZX} = 0.05$, $K^{ZY} = 0.5$, $K^{ZXY} = 10^4$ [nM][5]. The combined contribution of XY is reduced in order to observe the individual effect of X and Y. By making β^{ZXY} low and K^{ZXY} high, the transcription rate from the ZXY cycle is negligible. The maximal transcription rate of Y ($\beta^{YX} = 2$) is twice the maximal transcription rate of Z ($\beta^{ZX} = 1$ and $\beta^{ZY} = 1$). This was done in order to have concentrations of Y of the same order of magnitude of Z since Y is being consumed when transcribes Z. The maximal consumption rate of Y corresponds to the maximal transcription rate of that cycle ($\beta^{ZY} = 1$), and the net transcription rate of Y when fully transcribes Z is 1 [nM s^{-1}].

orange in Fig. 7B). As the coordinates of this vector are known, it can be equally divided in n points where point i has the coordinates (J^i_Y, J^i_Z). Each of these points can then be fully determined by minimizing the input flux J_X in order to maximize the fixed "benefit" (given by the values of J^i_Y and J^i_Z) at the minimal "cost" (given by J_X). This optimization problem can be described as follows:

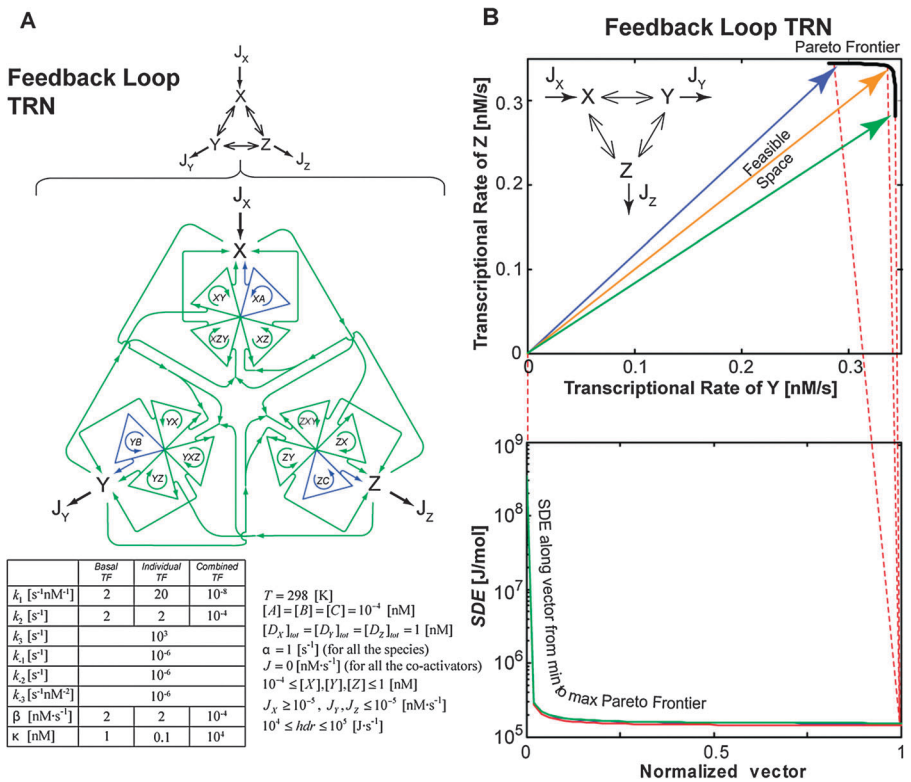
$$\begin{aligned} &\underset{\vec{x}}{\text{Minimize}} \quad (J_X) \ \ (\text{for } i = 1:n) \\ &\text{subject to:} \quad f(\vec{x}) = 0 \\ &\qquad\qquad \mathrm{HDR}_{lb} \le \mathrm{HDR} \le \mathrm{HDR}_{ub} \\ &\qquad\qquad J_Y = J^i_Y, \ \ J_Z = J^i_Z \\ &\qquad\qquad \vec{x}_{lb} \le \vec{x} \le \vec{x}_{ub} \end{aligned}$$

Interestingly, we found that at Pareto-optimal environmental surface, SDE is always the minimum for TRNs as seen in Fig. 7B. Importantly, SDE steadily decreased when moving throughout the feasible space along the vector from minimum transcriptional fluxes to the Pareto-optimal maximal transcriptional fluxes for FFL TRN systems.

4.2.2 Correlating pareto optimal surface with dissipative energetics for feedback TRN loop subgraphs. The Feedback TRN loop subgraph corresponds to a three-node system with complete feedback and an *OR* input logic as shown in Fig. 8A. This network is composed of twelve cycles: four of them are involved in the transcription of X by both Y and Z, another four cycles in the transcription of Y by X and Z, and four in the transcription of Z by both X and Y. The first four cycles correspond to cycle XA (transcription of X from its basal activator A), cycle XY (transcription of X from the activator Y), cycle XZ (transcription of X by Z), and cycle XYZ (transcription of X by both Y and Z). The next four cycles involve cycle YB (transcription of Y from its basal activator B), cycle YX

View Article Online

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013



	Basal TF	Individual TF	Combined TF
k_1 [$s^{-1}nM^{-1}$]	2	20	10^{-8}
k_2 [s^{-1}]	2	2	10^{-4}
k_3 [s^{-1}]	10^3		
k_{-1} [s^{-1}]	10^{-6}		
k_{-2} [s^{-1}]	10^{-6}		
k_{-3} [$s^{-1}nM^{-2}$]	10^{-6}		
β [$nM{\cdot}s^{-1}$]	2	2	10^{-4}
κ [nM]	1	0.1	10^4

Fig. 8 (A) The three-node transcription network with complete feedback and *OR* input logic. Parameters and optimization constraints are shown. Basal TF refers to the basal activator (blue cycle), individual TF refers to only one transcription factor (for instance, *X* or *Y* in the transcription of *Z*), and combined TF refers to the joint interaction between the individual activators (for instance, *XY* in the transcription of *Z*). (B) Pareto frontiers for transcriptional regulatory networks (TRN) were obtained by considering the maximization of protein production rates of *Y* and *Z* as the appropriate TRN objectives. Three node feedback TRNs follow the Pareto dominance rule. SDE was found to be minimal at the Pareto frontier compared to the non-Pareto feasible solutions. Following the solutions on the line vector from Pareto frontier of minimization to the Pareto frontier of maximization there is a constant decrease in the SDE and it is minimal at the Pareto frontier optimal transcriptional rates. Notice that the basal kinetic parameters and thus the basal maximal transcription rates and activation coefficients were selected in order to have $J^I_{T,\mathrm{Basal}} = 10^{-4}.J^I_{T,\mathrm{Max}}$, for $I = X,Y,Z$. In addition, the kinetic parameters of the other cycles were found in order to get the following activation coefficients: $K^{\mathrm{individual}} = 0.1$ and $K^{\mathrm{combined}} = 10^4$ [nM]. $K^{\mathrm{individual}}$ refers to the individual action of a single transcription factor (cycles *XY*, *XZ*, *YX*, *YZ*, *ZX* and *ZY*) and K^{combined} refers to the combined action of two transcription factors (cycles *XYZ*, *YXZ*, and *ZXY*).[5] As in the FFL TRN, $\beta^{\mathrm{combined}}$ and K^{combined} are low and high, respectively and $\beta^{\mathrm{individual}} = 2$.

(transcription of *Y* from the activator *X*), cycle *YZ* (transcription of *Y* by *Z*), and cycle *YXZ* (transcription of *Y* by both *X* and *Z*). The last four cycles are cycle *ZC* (transcription of *Z* from its basal activator *C*), cycle *ZX* (transcription of *Z* by the activator *X*), cycle *ZY* (transcription of *Z* by activator *Y*) and cycle *ZXY* (transcription of *Z* by both *X* and *Y* together). As being depicted throughout this work, the basal cycles are colored blue and the other cycles are green, meaning that there is active transcription from all of them. Transcription of *Z* occurs from all of the nine cycles because the *OR* input logic. For this TRN, the minimal Pareto frontier is confined to a single point characterized by the vector. $x = [1.961 \times 10^{-4}, -10^{-5}, -10^{-5}]^T$.

Notably, as seen in Fig. 8B we found that at Pareto-optimal environmental surface, SDE is always the minimum for feedback subgraphs. Importantly, similar to FFL TRNs, SDE steadily decreased when moving throughout the feasible space along the vector from minimum transcriptional fluxes to the Pareto-optimal maximal transcriptional fluxes for feedback subgraphs.

The behavior of energetic cost at optimal conditions is along the lines of existing paradigms which indicate that biological systems operate under optimal environmental conditions with optimal selection and utilization of existing mass resources in *E. coli*.[18,45] Extending the previous theory, our analysis concludes that (a) maximally optimal transcriptional rates are directly related with optimal utilization of available energetic resources, and (b) at the maximal Pareto-optimal condition the SDE is minimal. Therefore, natural systems operating under Pareto-optimal conditions will always have lowest specific energetic cost. We tried various architectures of TRN motifs and found this phenomenon to be true in all cases. Importantly, our optimality framework provides a rationale for how cells integrate optimal selection and utilization of resources with energetic-cost. Paradoxically, our optimality results elicit that energetic cost is minimal at the maximal/optimal resource utilization conditions although it was not part of the cellular objectives being optimized. Thus, rather than operating at optimal nutrient and functional conditions, cellular regulatory systems simultaneously use energetic

 This journal is © The Royal Society of Chemistry 2012

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

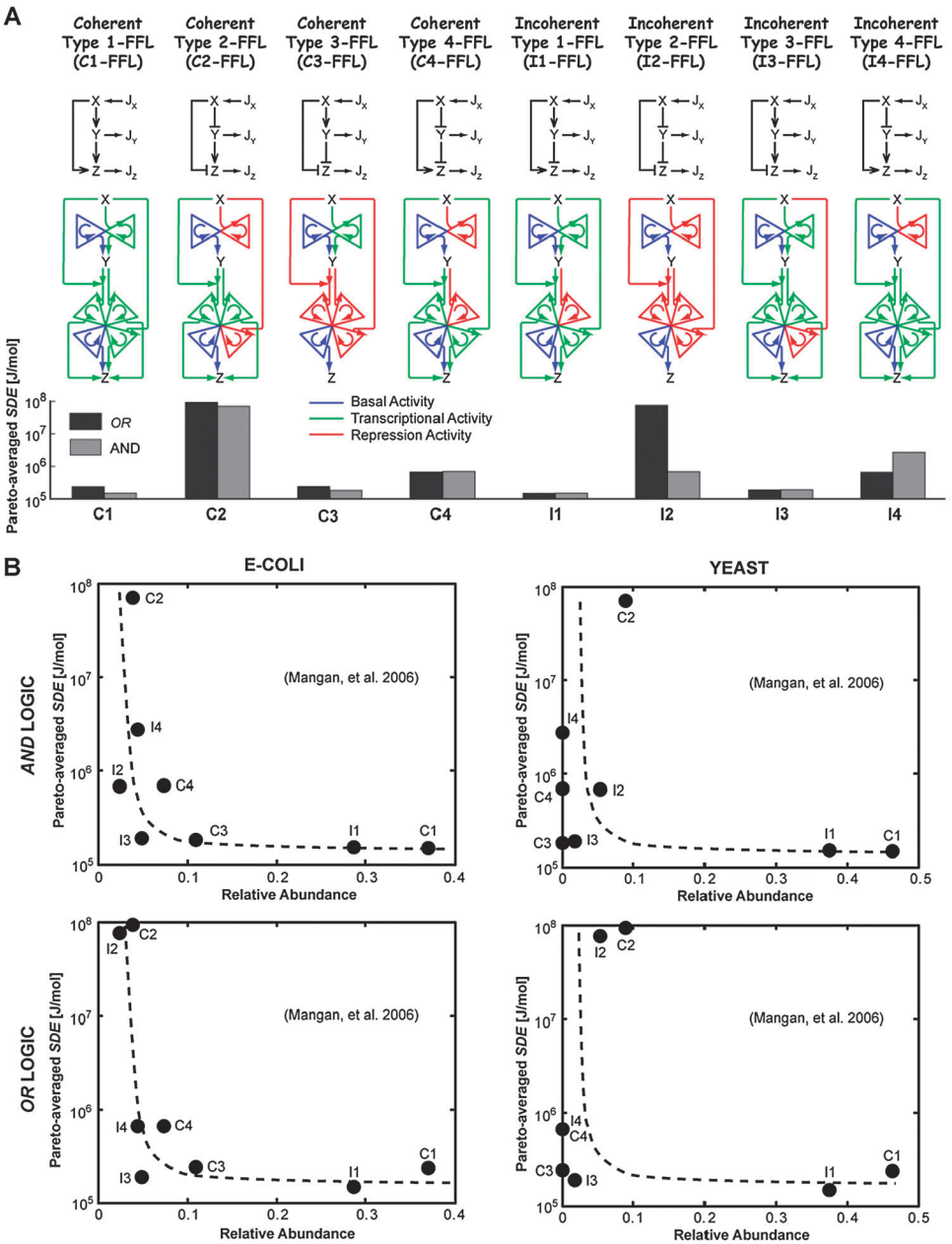
cost minimization as the underlying basis, to operate under globally optimal conditions.

4.3 Specific dissipation energy under pareto optimal conditions predicts frequency of occurrence of motifs

We further studied the implications of energetic-cost on network structure and motifs. We investigated whether biological systems utilize this as a rule of selection, contributing to high abundance of certain subgraphs (also defined as density or occurrence frequency). FFLs can be classified as coherent (Type-1 to 4) or incoherent (Type-1 to 4). Fig. 9A shows that with either *AND* or *OR* input logics, FFLs can be divided into Coherent-Type 1 to 4 (C1-FFL to C4-FFL) and Incoherent-Type 1 to 4 (I1-FFL to I4-FFL). Independently of the Boolean input logic or the Coherent or Incoherent structure, these networks are composed of six cycles as shown in Fig. 7A:

Since both external fluxes of Y and Z are desired to be maximized when X is externally provided, there is a tradeoff region between these two objective functions. The details of cyclic representation and parameters used for energetic estimation in these FFLs are shown in Fig. S2.

Multiple lines of evidence[5,46–49] suggest that coherent and incoherent Type-1 FFL(I1-FFL) motifs are abundant in both prokaryotic *E. coli* and the eukaryotic *S. cerevisiae* TRN systems, while other FFL motifs rarely occur in these small organisms. However, no explanation has yet been provided for the rare occurrence of certain FFL types and the abundance of a few motifs.[5,48,49] To answer this question, we hypothesized that SDE itself is the universal correlate used by biological systems for selection during evolution. To evaluate this, we first obtained the Pareto-averaged SDE (the average SDE over the maximal Pareto frontier between transcription rates of



Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

proteins Y and Z) for all eight FFL motifs for both *AND* and *OR* logics (Fig. 9A). Pareto-averaged SDE is found to be the lowest for highly abundant FFL motifs and highest for rarely occurring network motifs (Fig. 9A). Significantly, we observed that Pareto-averaged SDE for both input logics (*AND* and *OR*) correlates inversely with the abundance of the network motif as experimentally determined[5] for microorganisms (*E. coli* and *S. cerevisiae*) as seen in Fig. 9B (Spearman's correlation coefficient $= -0.43$; $p = 0.007$). Notably, similar SDE-network motif abundance correlation is obtained when genes were used as nodes[47] (Fig. S3). The preceding results show that during evolution minimal SDE may have been an important criteria used for development of highly organized complex TRNs. The success of predicting abundance or the regulatory role of network motifs depends on various factors which may be limiting cells in different stages of their life cycle, as well as their evolutionary history. Our results elicit that although energetics necessarily may not be the rate limiting factor, at some period of evolution it might have been the "bottleneck".

4.4 SDE links network topology with dynamic functionality and evolutionary adaptations

The dynamical function has been used to explain the occurrence frequency of motifs;[1,5,7] however, it fails to do so for functionally similar motifs. Type 1 coherent FFL as shown earlier has higher occurrence frequency than Type 4 coherent FFL. Similarly, Type 1 incoherent FFL has higher occurrence frequency than Type 4 incoherent FFL. We hypothesized that SDE can be used as the mapping function between network topology and dynamic output.

We first estimated SDE for a dynamic scenario where a phenotype exhibited by an incoherent FFL motif in the galactose system of *E. coli* has been shown to be of pulse generation and response acceleration.[5,48,49] The dynamic responses of incoherent FFLs with *AND* input logic (I1-FFL–I4-FFL) are compared in Fig. 10 for an ON-OFF step change in the concentration of X from 10^{-4} to 1 [nM]. At times $t < 0$, the system is allowed to reach the initial steady state (ss_i) for $[X] = 10^{-4}$ [nM]. At $t = 0$ the ON step change in the concentration of X occurs and is maintained at $[X] = 1$ [nM] for 10 [s]. This time is enough for the system to reach a new steady state (ss_f). At $t \geq 10$ [s], the OFF step change is induced and the concentration of X is again brought to 10^{-4} [nM] and kept at this value for another 10 [s]. After this, the system attains the initial steady state (ss_i). Notice that in these FFLs, the transcription factor X always activates Z in the direct transcriptional pathway. In the indirect pathway (X activates or represses Z by transcribing the intermediate Y first), different combinations between activation and repression are considered. The dynamic response of FFLs is compared to the dynamic response of simple regulation (SR) $X \rightarrow Z$. The parameters used in Fig. 10 for the FFLs are the same as used in Fig. 9. The SR case is formed by only two cycles: cycle ZB and ZX. The parameters used for these cycles are the same for their corresponding cycles in the FFLs. In addition, the lower and upper limits in the step change in the concentration of X correspond to the lower and upper constraints in the Pareto frontiers used in Fig. 9. In all of the responses, the concentrations were normalized from 0 to 1 as shown in the equation below where C represents any concentration, and $[C]_{ssi}$ and $[C]_{ssf}$ are the concentrations of C at ss_i and ss_f, respectively.

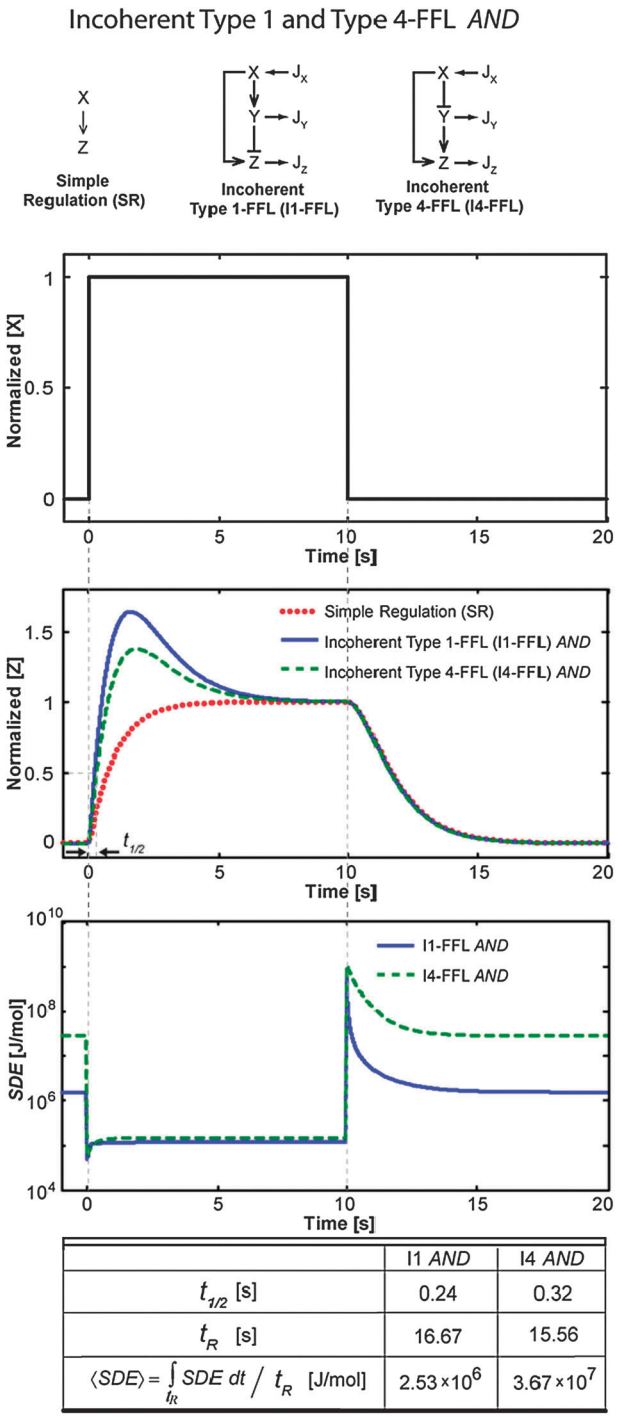
$$[C]_{\text{normalized}} = \frac{[C] - [C]_{ss_i}}{[C]_{ss_f} - [C]_{ss_i}}$$

As theoretically explained[5] and experimentally demonstrated in the galactose system of *E. coli*,[5,48–51] I1-FFL with *AND* input logic generates a pulse and speeds up the response time only in the ON phase. I4-FFL with *AND* input logic also accelerates the response in this sign sensitive manner and

Fig. 9 Optimal SDE predicts the network motif abundance. (A) Optimal SDE of eight coherent and incoherent FFL motifs at the Pareto frontiers with two input functions *OR* and *AND*. The input functions integrate the incoming signals at the promoter of gene Z. Here, arrow denotes activation and symbol $\perp$ denotes repression. The bar graph shows the averaged SDE over the Pareto frontier of a FFL motif. The Pareto frontier was obtained between transcription fluxes J_Y and J_Z. The symbolic representation of the corresponding cycles involved is shown below each FFL in the figure and FFL types are marked C and I for coherent and incoherent. (B) Correlation of averaged Pareto-optimal SDE with relative abundance of FFL motifs for *E. coli* and *S. cerevisiae* TRNs. Correlation is done for network motif relative abundance data obtained from experimentally verified *E. coli*[52] and *S. cerevisiae* 3 databases by considering gene as nodes. A statistically significant negative correlation (Spearman correlation coefficient $= -0.43$ and p value $= 0.007$) between the Pareto-averaged SDE and the published data of relative abundance of FFL motifs was observed. Spearman's correlation coefficient was used to study the relationship between Pareto-averaged SDE values and relative abundance of the 8 FFL motifs in yeast and E. coli based on the data published.[5] SDE data from the AND and OR configurations were pooled to increase the power of the test. The basal cycles (YA and ZB) are shown in blue, and the other cycles are either green if the transcription factor behaves as an activator, or red if the transcription factor acts as a repressor. In the latter case, no transcription occurs and when there is no arrow leaving the cycle. The basal kinetic parameters and thus the basal maximal transcription rates and activation coefficients were selected in order to have $J^Y_{T,\text{Basal}} = 0.1 \cdot J^Y_{T,\text{Max}}$, $J^Z_{T,\text{Basal}} = 0.001 \cdot J^Z_{T,\text{Max}}$. The kinetic parameters of the other cycles were found in order to get the following activation coefficients: **For activation**: $K^{YX} = 0.01$, $K^{ZX} = 0.05$, $K^{ZY} = 0.5$, $K^{ZXY} = 10^4$ (*OR* input logic) and $K^{ZXY} = 0.025$ (*AND* input logic) [nM]. **For repression**: $K^{YX} = 1$, $K^{ZX} = 1$, $K^{ZY} = 10$, $K^{ZXY} = 10^7$ (*OR* input logic) and $K^{ZXY} = 0.025$ (*AND* input logic) [nM]. For this analysis, the combined contribution of XY for *OR* input logic is desired to be minimal in order to observe the individual effect of X and Y. By making β^{ZXY} low and K^{ZXY} high, the transcription rate from the ZXY cycle is negligible. On the other hand, when the input logic is *AND*, the contribution from the ZXY cycle has to be as important as the one from the individual transcription factors. In this case, $\beta^{ZXY} = 1$ is same as $\beta^{ZX} = 1$ and $\beta^{ZY} = 1$ and for the activation case $K^{ZXY} = K^{ZX}K^{ZY}$, as previously suggested.[5] The maximal transcription rate of Y ($\beta^{YX} = 2$) is twice the maximal transcription rate of Z. This was done in order to have concentrations of Y of the same order of magnitude of Z since Y is being consumed when Z is being transcribed. The maximal consumption rate of Y corresponds to the maximal transcription rate of that cycle ($\beta^{ZY} = 1$), and the net transcription rate of Y when Z is fully transcribed is 1 [nM s^{-1}].

 This journal is © The Royal Society of Chemistry 2012

Downloaded by Rice University on 30 January 2013
Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F

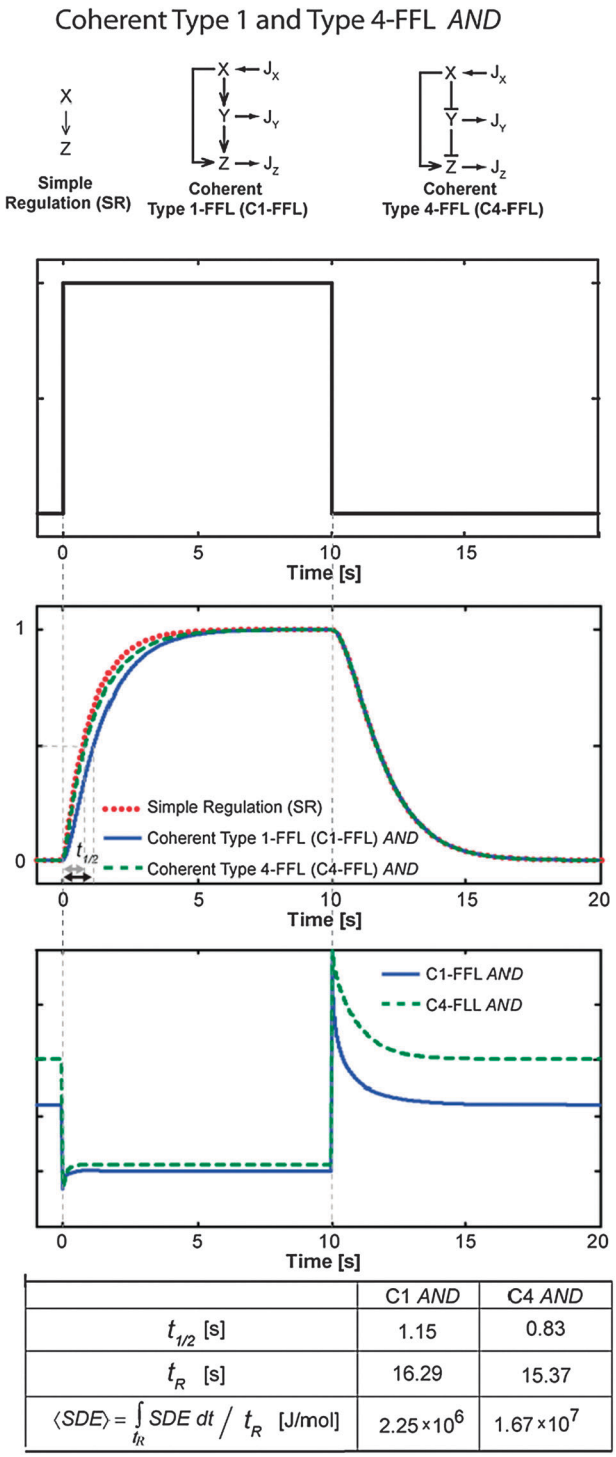


	I1 *AND*	I4 *AND*
$t_{1/2}$ [s]	0.24	0.32
t_R [s]	16.67	15.56
$\langle SDE \rangle = \int_{t_R} SDE\, dt \,/\, t_R$ [J/mol]	2.53×10^6	3.67×10^7

Fig. 10 Selection of network motifs for dynamic functionality utilizes energetics as the underlying basis. Comparison between I1-FFL and I4-FFL with *AND* logic for an input pulse in input *X*. I1-FFL and I4-FFL with *AND* logic motifs have pulse generation and response acceleration in ON step as the dynamic function. $t_{1/2}$ is the time to reach 50% of the steady state and response time, t_R is the time to attain the final steady state. $t_{1/2}$ is slightly lower for I1-FFL than I4-FFL but SDE of I1-FFL is significantly lower (10×) than I4-FFL. The dynamic responses are obtained for an ON-OFF step change in the concentration of *X*. The step pulse duration is chosen to be long enough for the system to reach a new steady state. The response speed ($t_{\frac{1}{2}}$) is quantified as the time required to reach 50% of the final steady state during the ON response. The response time (t_R) is the time required for the system to return to the initial steady state within a 1% of error. Using the cyclic kinetic TRN model we estimate the specific dissipated energy required for attaining the dynamic function. The time-averaged SDE ($\langle SDE \rangle$) represents the mean SDE during the response time: $\langle SDE \rangle = \frac{\int_0^{t_R} SDE(t)\, dt}{t_R}$.

generates a pulse response.[5] Both dynamical behaviors are predicted by our transcriptional model. As shown in Fig. 10 during the ON response, I1-FFL and I4-FFL speeds up the response by generating a pulse that overshoots the final steady state (ss_f) when compared to the SR. However, during the OFF response, I1-FFL, I4-FFL, and SR have similar trajectory.

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013



	C1 *AND*	C4 *AND*
$t_{1/2}$ [s]	1.15	0.83
t_R [s]	16.29	15.37
$\langle SDE \rangle = \int_{t_R} SDE\, dt \,/\, t_R$ [J/mol]	2.25×10^6	1.67×10^7

Fig. 11 Selection of network motifs for dynamic functionality utilizes energetics as the underlying basis. Comparison between C1-FFL and C4-FFL with *AND* logic for an input pulse in input *X*. C1-FFL and C4-FFL with *AND* logic motifs have delay in ON step with marginally higher delay for C1-FFL. SDE is significantly lower (10×) for C1-FFL than for C4-FFL. The dynamic responses are obtained for an ON-OFF step change in the concentration of *X*. The step pulse duration is chosen to be long enough for the system to reach a new steady state. The response speed ($t_{\frac{1}{2}}$) is quantified as the time required to reach 50% of the final steady state during the ON response and the response time (t_R) is the time required for the system to return to the initial steady state within a 1% of error. Using the cyclic kinetic TRN model we estimate the specific dissipated energy required for attaining the dynamic function. The time-averaged SDE ($\langle$SDE$\rangle$) represents the mean SDE during the response time: $\langle \text{SDE} \rangle = \frac{\int_0^{t_R} SDE(t)\, dt}{t_R}$.

We quantified the accelerated response as the time required to reach 50% of the final steady state during the ON response ($t_{1/2}$) as seen in Fig. 10. For I1-FFL and I4-FFL, these times are $t_{1/2}^{\text{I1-FFL}} = 0.24$ and $t_{1/2}^{\text{I4-FFL}} = 0.32$ [s], respectively. It is to be noted that the desired dynamic functionality (accelerated ON response compared to SR) is present in both I1-FFL

 This journal is © The Royal Society of Chemistry 2012

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

and I4-FFL, and is only marginally better for I1-FFL. Our kinetic TRN model allowed us to calculate the dissipated energy required to fulfill the dynamic requirement during the entire response time (t_R) or time required for the system to return to the initial steady state within a 1% of error ($t_R^{\text{I1-FFL}} = 16.7$ and $t_{1/2}^{\text{I4-FFL}} = 15.6$ [s]). The table in Fig. 10 shows the quantified values of dynamic responses of SDE for I1-FFL and I4-FFL. When SDE is time averaged during the response time, we found that the total averaged specific dissipated energy $\langle \text{SDE} \rangle$ for both systems as: $\langle \text{SDE} \rangle^{\text{I1-FFL}} = 2.53 \times 10^6$ and $\langle \text{SDE} \rangle^{\text{I4-FFL}} = 3.67 \times 10^7$ [J mol^{-1}].

Both Type-1 (I1-FFL) and Type-4 (I4-FFL) incoherent FFLs with *AND* input logic were found to have similar dynamic phenotype for an ON-OFF step change in the input signal X (Fig. 10). Thus, the current paradigm of associating dynamic functionality with frequency of occurrence of motifs does not explain the variations in abundance density since both FFL types had pulse generation and response acceleration as a function with similar magnitude. However, energetic-cost SDE indicates large magnitude differences (at least one order of magnitude) between time-averaged SDE ($\langle \text{SDE} \rangle$) of I4-FFL and I1-FFL (Fig. 10). The energetic hypothesis seems to perform better than dynamical function in explaining abundance/rare occurrence of incoherent FFLs.

Coherent FFLs have been shown in the arabinose system of *E. coli*[5] to have the sign-sensitive delay as the major dynamic function. Sign-sensitive delay is defined as the response delay when compared to simple regulation (SR) and it depends on the sign of the input step (ON or OFF). As theoretically explained[5] and experimentally demonstrated in the arabinose system of *E. coli*,[49] the C1-FFL with *AND* logic presents a sign sensitive delay for ON steps of the external signal X compared to the SR. Similarly, C4-FFL with *AND* logic exhibits this delay during the ON response but not during the OFF step change.[5] Both dynamical behaviors are predicted by our kinetic model of TRNs, as shown in Fig. 11: during the ON response. C1-FFL and C4-FFL initially is delayed compared to the simple regulation, but during the OFF response, C1-FFL, C4-FFL, and SR follow the same trajectory. For C1-FFL and C4-FFL, the delay times are $t_{1/2}^{\text{C1-FFL}} = 1.15$ and $t_{1/2}^{\text{C4-FFL}} = 0.83$ [s], respectively. Observe that the desired functionality of "delayed response" (delayed ON response compared to SR) is present in both C1 *AND* and C4 *AND*, but is slightly better for C1 *AND*. We next calculated the average SDE over the response time ($t_R^{\text{C1-AND}} = 16.3$ and $t_{1/2}^{\text{I4-AND}} = 15.4$ [s]). The table in Fig. 11 shows that averaged SDE for both FFLS are $\langle \text{SDE} \rangle^{\text{C1-FFL}} = 2.25 \times 10^6$ and $\langle \text{SDE} \rangle^{\text{C4-FFL}} = 1.67 \times 10^7$ [J mol^{-1}].

Both, coherent Type-1 (C1-FFL) and coherent Type-4 (C4-FFL) FFLs were seen to have similar sign-sensitive delay functionality (Fig. 11); hence, it cannot clearly establish the rare occurrence of C4-FFL compared to C1-FFL. For a step input in activator X, the delay with respect to SR was marginally pronounced in C1-FFL than C4-FFL (Fig. 11). Strikingly, we found that $\langle \text{SDE} \rangle$ was significantly lower for C1-FFL than C4-FFL with *AND* logic thus consistent with their abundance.

5. Conclusion

The energetic-cost theory presented here clearly indicates that SDE may be a suitable basis for evolutionary selection of one motif over another and could provide an explanation for the rare occurrence of various network motifs. Here, we have used a simplified non-equilibrium cyclic network modeling approach to estimate heat dissipation involved during transcription. Our analysis indicates that the Pareto-optimality principle, when combined with NESS analysis, leads to energetically efficient solutions for transcription. We showed that the underlying energetic-cost criterion, SDE, for Pareto-optimal conditions is a measure that reflects maximal transcription at the lowest energetic demand. Beyond its application as a functional basis in TRN motifs, the Pareto-optimal SDE concept may also lead to an optimal and energetically efficient design of synthetic gene circuits. Further validation of this concept for protein and metabolic networks is required to confirm its generality; however, the corresponding abundance data for these networks is unavailable. The finding that energetic cost may be used as an underlying basis for evolutionary selection of between motifs having similar dynamic functionality is of major significance. In our future work, we are addressing the limitations imposed by using lumped model and simplified approach for estimating energetics. The new developments in high-performance computing will allow explicit modeling of various processes involved during transcription and translation such as confirmation changes, structural modifications, DNA-protein interactions and post translational changes in transcriptional network models. The overwhelming diversity of possible dynamical functions with highly-interactive biological networks limits effective learning from experimental data alone. Network analyses using knowledge of the often ignored energetics may greatly reduce the hypothesis space, enabling identification of new functionalities of dynamically perturbed large-scale networks. Appropriate identifications of cellular objectives involved in evolutionary decision making may provide a potentially novel approach to identify optimal environmental conditions and therefore, as a stand-alone strategy, may provide a more efficacious simultaneous prediction and validation strategy for biological networks.

Acknowledgements

We are grateful to Arno Tilles, Syed Asad Hussain, Ron Milo, Marcos Sotomayor, Sachet Shukla, Lisa Pabst and Roberto Rondanelli for discussions, comments and suggestions for this article.

References

1 U. Alon, *Nat. Rev. Genet.*, 2007, **8**, 450–461.
2 R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer and U. Alon, *Science*, 2004, **303**, 1538–1542.
3 R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, *Science*, 2002, **298**, 824–827.
4 G. Balazsi, A. L. Barabasi and Z. N. Oltvai, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7841–7846.

View Article Online

5 S. Mangan and U. Alon, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 11980–11985.
6 A. Vazquez, R. Dobrin, D. Sergi, J. P. Eckmann, Z. N. Oltvai and A. L. Barabasi, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17940–17945.
7 R. J. Prill, P. A. Iglesias and A. Levchenko, *PLoS Biol.*, 2005, **3**, e343.
8 E. H. Yang, R. R. Almon, D. C. Dubois, W. J. Jusko and I. P. Androulakis, *PLoS One*, 2009, **4**, e5992.
9 E. Yang, M. L. Yarmush and I. P. Androulakis, *J. Theor. Biol.*, 2009, **256**, 393–407.
10 J. F. Moxley, M. C. Jewett, M. R. Antoniewicz, S. G. Villas-Boas, H. Alper, R. T. Wheeler, L. Tong, A. G. Hinnebusch, T. Ideker, J. Nielsen and G. Stephanopoulos, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 6477–6482.
11 G. Stephanopoulos, H. Alper and J. Moxley, *Nat. Biotechnol.*, 2004, **22**, 1261–1267.
12 C. Chan, D. Hwang, G. N. Stephanopoulos, M. L. Yarmush and G. Stephanopoulos, *Biotechnol. Prog.*, 2003, **19**, 580–598.
13 C. Chan, F. Berthiaume, K. Lee and M. L. Yarmush, *Biotechnol. Bioeng.*, 2003, **81**, 33–49.
14 A. P. Burgard, P. Pharkya and C. D. Maranas, *Biotechnol. Bioeng.*, 2003, **84**, 647–657.
15 A. P. Burgard and C. D. Maranas, *Biotechnol. Bioeng.*, 2003, **82**, 670–677.
16 W. W. Wong, T. Y. Tsai and J. C. Liao, *Mol. Syst. Biol.*, 2007, **3**, 130.
17 D. Nagrath, M. Avila-Elchiver, F. Berthiaume, A. W. Tilles, A. Messac and M. L. Yarmush, *Ann. Biomed. Eng.*, 2007, **35**, 863–885.
18 E. Dekel and U. Alon, *Nature*, 2005, **436**, 588–592.
19 K. S. Varedi, A. C. Ventura, S. D. Merajver and X. N. Lin, *PLoS One*, 2010, **5**, e14029.
20 M. A. Savageau and R. R. Freter, *Biochemistry*, 1979, **18**, 3486–3493.
21 R. K. Dash, Y. Li, J. Kim, D. A. Beard, G. M. Saidel and M. E. Cabrera, *PLoS One*, 2008, **3**, e3168.
22 H. Zhou, N. Lai, G. M. Saidel and M. E. Cabrera, *Ann. N. Y. Acad. Sci.*, 2008, **1123**, 178–186.
23 M. S. Dasika, A. Burgard and C. D. Maranas, *Biophys. J.*, 2006, **91**, 382–398.
24 S. S. Fong, A. P. Burgard, C. D. Herring, E. M. Knight, F. R. Blattner, C. D. Maranas and B. O. Palsson, *Biotechnol. Bioeng.*, 2005, **91**, 643–648.
25 D. Nagrath, M. Avila-Elchiver, F. Berthiaume, A. W. Tilles, A. Messac and M. L. Yarmush, *Metab. Eng.*, 2010, **12**, 429–445.
26 D. Nagrath, B. W. Bequette, S. M. Cramer and A. Messac, *Aiche J.*, 2005, **51**, 511–525.
27 S. Ghose, D. Nagrath, B. Hubbard, C. Brooks and S. M. Cramer, *Biotechnol. Prog.*, 2004, **20**, 830–840.
28 H. Kitano, *Mol. Syst. Biol.*, 2007, **3**, 137.
29 N. E. Buchler, U. Gerland and T. Hwa, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5136–5141.
30 J. M. Vilar and J. M. Rubi, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11081–11084.
31 C. Bustamante, J. Liphardt and F. Ritort, *Phys. Today*, 2005, **58**, 43–48.
32 H. Qian and D. A. Beard, *Biophys. Chem.*, 2005, **114**, 213–220.
33 J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr. and C. Bustamante, *Science*, 2002, **296**, 1832–1835.
34 G. Oster, A. Perelson and A. Katchals, *Nature*, 1971, **234**, 393.
35 K. J. Rothschild, S. A. Ellias, A. Essig and H. E. Stanley, *Biophys. J.*, 1980, **30**, 209–230.
36 H. Qian, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **69**, 012901.
37 V. Chernyak, M. Chertkov and C. Jarzynski, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2005, **71**, 025102.
38 S. Lapidus, B. Han and J. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 6039–6044.
39 M. A. Shea and G. K. Ackers, *J. Mol. Biol.*, 1985, **181**, 211–230.
40 N. K. Jana, S. Deb, B. Bhattacharyya, N. C. Mandal and S. Roy, *Protein Eng., Des. Sel.*, 2000, **13**, 629–633.
41 S. P. Walton, G. N. Stephanopoulos, M. L. Yarmush and C. M. Roth, *Biophys. J.*, 2002, **82**, 366–377.
42 T. Xia, A. Frankel, T. T. Takahashi, J. Ren and R. W. Roberts, *Nat. Struct. Biol.*, 2003, **10**, 812–819.
43 I. Russo, P. Barboro, I. Alberti, S. Parodi, C. Balbi, C. Allera, G. Lazzarini and E. Patrone, *Biochemistry*, 1995, **34**, 301–311.
44 P. P. Dzeja and A. Terzic, *J. Exp. Biol.*, 2003, **206**, 2039–2047.
45 R. U. Ibarra, J. S. Edwards and B. O. Palsson, *Nature*, 2002, **420**, 186–189.
46 N. Kashtan, S. Itzkovitz, R. Milo and U. Alon, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **70**, 031909.
47 H. W. Ma, B. Kumar, U. Ditges, F. Gunzer, J. Buer and A. P. Zeng, *Nucleic Acids Res.*, 2004, **32**, 6643–6649.
48 S. Mangan, S. Itzkovitz, A. Zaslaver and U. Alon, *J. Mol. Biol.*, 2006, **356**, 1073–1081.
49 S. Mangan, A. Zaslaver and U. Alon, *J. Mol. Biol.*, 2003, **334**, 197–204.
50 A. Zaslaver, A. Mayo, M. Ronen and U. Alon, *Phys. Biol.*, 2006, **3**, 183–189.
51 A. Zaslaver, A. E. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M. G. Surette and U. Alon, *Nat. Genet.*, 2004, **36**, 486–491.
52 S. S. Shen-Orr, R. Milo, S. Mangan and U. Alon, *Nat. Genet.*, 2002, **31**, 64–68.

Published on 10 November 2011 on http://pubs.rsc.org | doi:10.1039/C1MB05177F
Downloaded by Rice University on 30 January 2013

 This journal is © The Royal Society of Chemistry 2012