CORRELATED FLUCTUATIONS PROBE DYNAMICS OF TRANSCRIPTIONAL REGULATION

Vijayanarasimha H. Pakka[1], *Adam Prügel-Bennett*[2], *Srinandan Dasmahapatra*[1]

[1]SENSe Group, [2]ISIS Group,
School of ECS, University of Southampton
Southampton SO17 1BJ, UK
{vhp05r,apb,sd}@ecs.soton.ac.uk

ABSTRACT

The dynamics of transcriptional control involves small numbers of molecules and result in significant fluctuations in protein and mRNA concentrations. The correlations between these intrinsic fluctuations then offer, via the fluctuation dissipation relation, the possibility of capturing the system's response to external perturbations, and hence the nature of the regulatory activity itself. We study time-dependent noise correlations in simple networks of activators and repressors, varying the topology of causal influence, and using different mechanisms or parameter choices. The distinct correlated fluctuations could be used as signatures for mechanism identification. To that end, we present analytical and numerical results on peaks and delays in correlations between proteins within networks, and the dependence of these features on parameter and mechanism.

1. INTRODUCTION

A gene regulatory network (GRN) consists of nodes representing genes and edges representing regulation of gene expression by proteins encoded by the genes. Since the set of processes of transcription and translation are inherently stochastic within a cellular environment, the levels of mRNA and protein for each species is a fluctuating quantity, and referred to as noise in gene expression(1). The rate of production of the average levels of protein expression of a gene as a function of its regulatory factors is used to describe the kinetics of each node in a GRN, from which the dynamics of the network, represented as a coupled set of ordinary differential equations (ODE), forms a model for the functional properties of the network. Experimental access to distributions of protein levels by flow cytometry (1) have demonstrated the observability of noisy genes, and there is further evidence that noise drives biologically meaningful behaviours in living systems (2). Given the stochasticity of the very processes that constitute regulatory responses to stimuli, it is likely that correlated fluctuations will illuminate the dynamics of regulatory interactions. Indeed, the intuition behind Onsager's regression hypothesis (3) is that the regression to an equilibrium state after a short-lived external perturbation or an intrinsic fluctuation are identical, which has been developed further in several fluctuation-dissipation theorems even away from equilibrium (3). Regulation of expression can be viewed as a set of responses to endogenous signals or exogenous stimuli organised by the cell to form functional pathways. In this paper, we consider correlated fluctuations in elementary fragments of GRN to illustrate what they can tell us about the nature of the regulatory function enacted by the network.

Experimental protocols circumscribe the nature of the observations of gene noise, which then determine the nature of the theoretical tools used to describe or model the data. Time-independent single-variable statistics characterise histograms of levels of fluorescent proteins in a population of cells accumulated using flow cytometric techniques (1), which can also be described using the fluctuation-dissipation relations (4; 5). Single molecule methods (6) quantify bursts of protein production, fitted to geometric distributions.

Instead of flow cytometry which leaves out the temporal aspects, time-lapse fluorescence measurements(7) are needed to track the causal dynamics as revealed by noise, and thereby facilitate our understanding of the link between regulatory implementation and correlative effects. Indeed, recent reports suggest that the presence of regulatory activity shows up in the time dependent correlations of fluctuations in protein levels(8; 7). That is the object of our study.

The analytical framework of this paper builds upon relations between the macroscopic dynamics and the fluctuation properties of the system, such as the fluctuation-dissipation relations(3). In the context of biochemical reaction systems, it emerges naturally out of an expansion of the chemical master equation in terms of the inverse of large system size(9; 10) assuming gaussian deviates from average concentrations of molecules. This linear noise approximation (LNA) is equivalent to the Fokker-Planck equation for these process(9).

Our main focus in this work is to derive the time-correlation functions in a sum-of-exponentials form (eq. (1)) to study the dynamic correlations between two molecular species in a GRN, such as proteins of the regulator and regulated genes. We demonstrate the distinct be-

havioural patterns in the protein correlations for changes in the regulatory networks. For such a study, we considered simple two-gene networks where, the changes in these networks were introduced by employing three strategies: (1) varying the values of the reaction rate constants (2) varying the regulatory mechanisms, say, activation with or without co-operative mechanisms, and repression (3) introducing additional player *i.e.,* a gene into the model, thus varying the parameters, regulatory mechanisms and the network structure respectively.

2. DYNAMIC CORRELATIONS

The Chemical Master equation,

$$\partial_t P(\mathbf{X},t) = \sum_{j=1}^{M} \{P(\mathbf{X}-\boldsymbol{\nu}_j,t)a_j(\mathbf{X}-\boldsymbol{\nu}_j) - P(\mathbf{X},t)a_j(\mathbf{x})\}$$

represents the updates in the probability distributions of a collection of numbers of molecules $\mathbf{X}$ due to reactions R_j $(j = 1\ldots,M)$ occurring with probability $a_j(\mathbf{x})dt$ in an infinitesimal time interval$(t, t+dt)$ in a volume Ω, resulting in changes ν_{ji} in the number of S_i molecules. In this paper, the molecules of interest are mRNAs and proteins encoded by genes. The corresponding reactions studied are given in Table 1.

By multiplying with X_i and X_iX_j and taking expectation values, it is easy to show that the means $\langle X_i\rangle$ and covariances $\mathrm{Cov}(X_i,X_j) = \langle X_iX_j\rangle - \langle X_i\rangle\langle X_j\rangle$ satisfy

$$\begin{aligned}\frac{d}{dt}\langle X_i\rangle &= \sum_{j=1}^{M}\nu_{ji}\langle a_j(\mathbf{X})\rangle, \text{ and}\\ \frac{d}{dt}\mathrm{Cov}(X_i,X_j) &= \sum_{k=1}^{M}\langle X_i\nu_{kj}a_k(\mathbf{X})\rangle + \langle X_j\nu_{ki}a_k(\mathbf{X})\rangle\\ &+ \langle\nu_{ki}\nu_{kj}a_k(\mathbf{X}-\boldsymbol{\nu}_k)\rangle.\end{aligned}$$

In terms of concentrations $x_i = X_i/\Omega$, the expectation values yield mass action kinetics description of the chemical system:

$$\frac{d}{dt}\langle x_i\rangle = \sum_{j=1}^{M}\nu_{ji}R_j(\mathbf{x}), \text{ where } R_j(\mathbf{x}) := \lim_{\Omega\to\infty}\frac{1}{\Omega}\langle a_j(\frac{\mathbf{X}}{\Omega})\rangle. \quad (1)$$

Only when the propensity functions are linear in its arguments do the covariance equations form a closed set, else higher-order statistics are necessary for finding the covariances. By linearising that the kinetic rates $R_j(\mathbf{x})$ around the steady state solutions $\langle\dot{x_i}\rangle = 0$, where the means are time independent, we get

$$\frac{\partial \boldsymbol{C}}{\partial t} = \boldsymbol{A}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{B}^T \quad (2)$$

where $\boldsymbol{C}$ stands for the matrix of covariances C_{ij} =Cov(x_i,x_j), $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined as :

$$A_{ik} = \frac{\partial}{\partial x_k}(\sum_{j=1}^{M}\nu_{ij}R_j(\mathbf{x})),\ \boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{\nu}\ diag(\mathbf{R}(\mathbf{x}))\ \boldsymbol{\nu}^T.$$

Network Structure	Regulatory Mechanism	Elementary Reactions
$X \to Y$	Monomers	$P_x + G_y \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathcal{C}$ $\mathcal{C} \xrightarrow{k_1} \mathcal{C} + M_y$
	Dimers	$2P_x \underset{k_b}{\overset{k_a}{\rightleftharpoons}} P_{x_2}$ $P_{x_2} + G_y \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathcal{C}$ $\mathcal{C} \xrightarrow{k_1} \mathcal{C} + M_y$
Direct activation		
$X \to Z \to Y$	Monomers	$P_x + G_z \underset{k_{r_1}}{\overset{k_{f_1}}{\rightleftharpoons}} \mathcal{C}_1$ $\mathcal{C}_1 \xrightarrow{k_1} \mathcal{C}_1 + M_z$ $M_z \xrightarrow{t_z} M_z + P_z$ $P_z + G_y \underset{k_{r_2}}{\overset{k_{f_2}}{\rightleftharpoons}} \mathcal{C}_2$ $\mathcal{C}_2 \xrightarrow{k_2} \mathcal{C}_2 + M_y$
Activation via intermediary		
$X \to Y \leftarrow Z$	Monomers	$P_x + G_y \underset{k_{r_1}}{\overset{k_{f_1}}{\rightleftharpoons}} c$ $P_z + c \underset{k_{r_2}}{\overset{k_{f_2}}{\rightleftharpoons}} \mathcal{C}$ $\mathcal{C} \xrightarrow{k_1} \mathcal{C} + M_y$
	Processive activation	
	Monomers	$P_x + G_y \underset{k_{r_1}}{\overset{k_{f_1}}{\rightleftharpoons}} \mathcal{C}_1$ $P_z + G_y \underset{k_{r_1}}{\overset{k_{f_1}}{\rightleftharpoons}} \mathcal{C}_2$ $\mathcal{C}_1 \xrightarrow{k_1} \mathcal{C}_1 + M_y$ $\mathcal{C}_2 \xrightarrow{k_2} \mathcal{C}_2 + M_y$
Combined activation	Parallel activation	
$X \dashv Y$	Monomers	$P_x + G_y \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathcal{C}$ $G_y \xrightarrow{k_1} G_y + P_y$
Direct repression		

Table 1. Table showing Network Topologies of simple activator and repressor systems and their regulatory mechanisms. Only those elementary reactions that *distinguish* the regulatory network are shown in the table above. Reactions such as production and decay of upstream mRNAs and their translation into proteins are not shown, but are included in the analysis.

The time evolution of the averages in the linearised equations is

$$\langle\mathbf{x}(t)\rangle = \boldsymbol{V}e^{\boldsymbol{\Lambda}(t-t_0)}\boldsymbol{U}^T\mathbf{x}(t_0),$$

with $\boldsymbol{V}$ and $\boldsymbol{U}^T$ being the matrices comprising of the left and right eigenvectors of the Jacobian matrix $\boldsymbol{A}$ respectively, and $\boldsymbol{\Lambda}$ the eigenvalues.

To solve for the stationary covariance matrix $\dot{\boldsymbol{C}} = 0$, in eq.(2) we can derive

$$\langle\ \delta x_i(t+\tau)\ \delta x_j(t)\ \rangle = \sum_{l=1}^{N} e^{\lambda_l\tau}V_{il}\left(\sum_{k=1}^{N}U^T_{lk}C_{kj}\right), \quad (3)$$

where $\delta x_i(t) := x_i(t) - x_i^s$ and x_i^s satisfies $\dot{x_i} = 0$ in eq. (1). We observe that the relaxation to a steady state, governed by the Jacobian matrix $\boldsymbol{A}$ is related to the fluctuation terms $\boldsymbol{B}\boldsymbol{B}^T$ which makes it possible to compute the covariances; this is the power of the fluctuation dissipation relation. For effective comparison of results from various networks below, we use the dynamic correlations $Corr\ [\delta x_i(t+\tau), \delta x_j(t)]$ by normalising the covariances by the stationary standard deviations.

3. RESULTS

3.1. Elementary Activator

The effect of a transcription factor on its target protein is studied. Here we have used the *cha4* $\longrightarrow$ *cha1* regulatory link (11) and parameters for the various decay rates and half-lives are obtained from the data compiled in (12). The proteins of the regulated and regulating genes are positively correlated, but with a delay τ^* in the attainment of the peak $Corr^*$, indicating the time scales for signal propagation through the transcriptional dynamical system to perturbations introduced. Note, positive corelations indicate activation. For the mRNA correlations, the peak occurs at $\tau^* = 36$ mins, while the peak is at $\tau^* = 49$ mins for protein correlations.

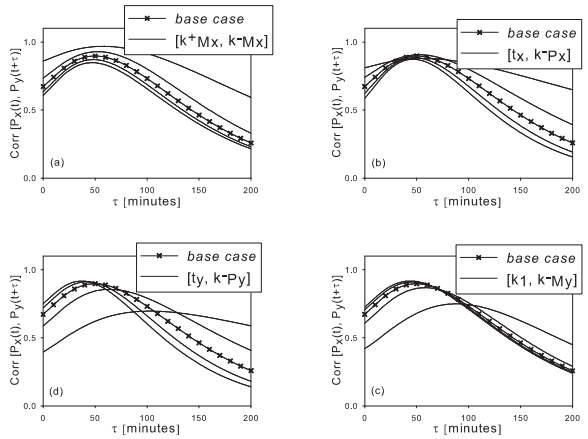

Figure 1. In all these figures, the dashed line is the covariance plot in the *base case*, the *cha4* $\longrightarrow$ *cha1* parameters, while solid lines are due to the step changes by 20% in the rate constants. (a)-(d) are labelled clockwise from top left. (a) $k^+_{M_x}$ and $k^-_{M_x}$, the production and decay rates of M_x varied. While $Corr^*$ decreases rapidly for increase in $k^+_{M_x}$ and $k^-_{M_x}$, τ^* experiences slight reduction from 56 to 46 minutes. (b) Translation and protein decay rates t_x and $k^-_{P_x}$ are varied. τ^* decreases from 55 to 45 minutes. (d) t_y and $k^-_{P_y}$ varied. τ^* drops from 102 to 36 minutes. (c) Transcription rate constant k_1, and decay rate $k^-_{M_y}$ are varied. τ^* drops from 87 to 43 minutes.

We notice that for the same mean steady state values for the proteins and mRNAs, the dynamic correlations vary for changes in the reaction rate constants. For example, the pair of parameters $[k^+_{M_x}, k^-_{M_x}]$ are varied from their *base* values in equal measures so that the mean value of $\langle M_x \rangle = k^+_{M_x}/k^-_{M_x}$ remains unaffected. This is illustrated with the help of the figure 3.1. The observations can be explained with the help of the eigenvalues of the Jacobian – 3 out of 5 eigenvalues are the decay rates $-k^-_{M_x}$, $-k^-_{M_y}$, $-k^-_{P_y}$, and in the limit $K_d \gg [G_y]$ (K_d is the dissociation constant for protein-DNA binding, $[G_y]$ concentration of gene), the remaining two are $-k^-_{P_x}$ and $-k_r/[G_y]$. It is this dependence on decay constants that accounts for the variability in τ^* observed.

3.2. Co-operative activation

In this model, G_y is activated by the action of dimers P_{x_2} that are formed by a dimerization process whose dissociation constant (for protein-protein interactions) is taken to be $K_{dim} = k_b/k_a = 200\ nM$. Here, we shall see the effect that dimerization and its dissociation constant has on the activator system. The effective transcription rate for species y (using the quasi-steady state assumption) is proportional to $P_x^2/(K_d K_{dim} + P_x^2)$, K_d being the protein-DNA dissociation constant. The effects of dimerization on the correlation functions are threefold: firstly, the peak correlation is larger than for monomers; secondly, τ^* is advanced by a few minutes; and thirdly, the shape is flatter. For both monomers and dimers, only the dissociation constants (the ratio, not the forward and reverse rates) affect the covariances. Also, while the covariances of P_y with P_x and P_{x_2} differ, their normalized dynamic correlations do not.

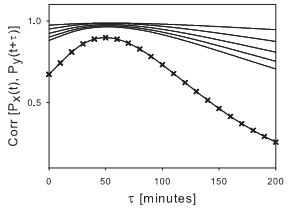

Figure 2. Activation by dimers: The dimer dissociation constant $k_{d_{ba}}$ is varied [by varying k_b, while k_a is held constant] in steps of $\pm 20\%$ from its base value of 200 nM corresponding to the stationary mean dimer concentration of $P_{x_2} = 783\ nM$. As the dissociation constant is altered, $P_{x_2} = P_x^2/k_{d_{ba}}$ and consequently the transcription rate $k_1 = k^-_{M_y} M_y/C$ varies. τ^*_{new} shows negligible variation from 53 to 51 minutes.

3.3. Elementary Repressor

We model repression as follows: the protein X blocks the incorporation of the RNA polymerase molecule onto the promoter of gene Y. The average steady state value of $M_y = k_1 G_y/k^-_{M_y}$. The basal transcription rate k_1 is taken to be $1.7\ min^{-1}$ and the dissociation constant $K_d = 100\ nM$. Other rate constants are the same as in the case of the elementary activator. G_y then is equal to $0.2\ nM$, simulating low probability of occupancy by the TF in the presence of a repressor. By doing so, we retain the mean values of the mRNAs and proteins.

If we artificially hold the mean values $P_y > P_x$, the repressor covariances can be viewed as a reflection, by the τ-axis, of the activator correlations. For $P_x > P_y$, by assuming concentrations of $\langle M_y \rangle = 0.1\ nM$ and $\langle P_y \rangle = 100\ nM$ (by reducing k_1, t_y), the correlation plots in Figure 3 are similar except with reduced magnitudes (compared to the activator case), $Corr^* = -0.47$.

3.4. Activation via intermediaries

Here we consider the transcriptional chain $X \rightarrow Z \rightarrow Y$, where we explore the effect of the intermediary TF Z on

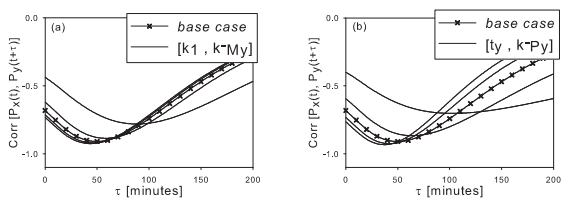

Figure 3. (a) $[k^-_{M_y}, k_1]$ are varied in pair in steps of $\pm 20\%$ from their base values. While $Corr^*$ increases in magnitude, τ^* decreases from 87 to 43 minutes for step-wise increase in these parameters. (b) $[k^-_{P_y}, t_y]$ are varied in pair in steps of $\pm 20\%$. While $Corr^*$ increases in magnitude, τ^* once again reduces from 102 to 36 minutes for step-wise increase in these parameters.

the $\langle P_x P_y \rangle$ correlations. In Figure 4, the dynamic correlation between P_x and P_y is shown to be sensitive in the decay rates of the intermediary element Z, $k^-_{M_z}$ and $k^-_{P_z}$. This is mainly due to the fact that four of the eight eigenvalues of the system are $[-k^-_{M_x}, -k^-_{M_z}, -k^-_{M_y}, -k^-_{P_y}]$ with the other four being the roots of two quadratic equations that are approximated as $[-k^-_{P_x}, -k^-_{P_z}, -\frac{k_{r_1}}{[G_z]}, -\frac{k_{r_2}}{[G_y]}]$. Thus, the effect of adding an intermediary regulatory element to the original two-gene activator network is two-fold: firstly, there is a large shift in the correlation along the time-axis. $\tau^*_{new} = 97$ minutes. Secondly, this τ^*_{new} is sensitive in the decay rates of Z. Note that, for mRNA and protein half-lives of around 1 minute each, of the intermediate element Z, $\tau^* \approx 49$. Hence, for rapidly decaying mRNA/proteins the effect of the Z on the correlations between X and Y are negligible.

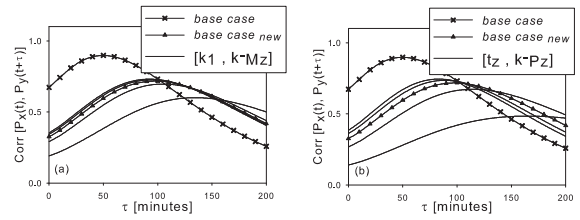

Figure 4. $X \rightarrow Z \rightarrow Y$ (a) The decay rate of the *intermediary* mRNA M_z is varied in steps of $\pm 20\%$ from an initial half-life of 8 minutes. The new τ^* and $Corr^*$ are sensitive to variations in $k^-_{M_z}$. τ^*_{new} decreases through 134 to 92 minutes, for increase in $k^-_{M_z}$. (b) $k^-_{P_z}$ is varied in steps of $\pm 20\%$ from an initial half-life of 50 minutes. Once again, for increasing $k^-_{P_z}$, τ^*_{new} decreases from 163 to 82 minutes. $Corr^*_{new}$ increases as well.

3.5. Activation with co-operation from other TFs

Figure 3.5 shows the sensitivity of correlations *w.r.t* decay rates of M_z and P_z. The covariance peak Cov^* is insensitive to variations in these parameters but not the correlations. Also, though the decay rates $k^-_{M_z}$ and $k^-_{P_z}$ show up as the eigenvalues, they do not influence the time-characteristics. Thus, the method of *parallel* activation desensitizes τ^* and brings about variation only in $Corr^*$.

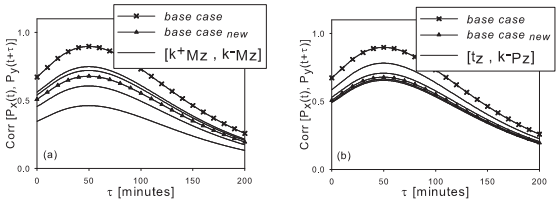

Figure 5. The presence of Z reduces the dynamic correlations, leaving τ^* unaltered at 49 minutes. (a) and (b) show the variation in $Corr^*$ for changes in $k^-_{M_z}$ and $k^-_{P_z}$ for the processive case. Parallel activation produces similar results, except that $Corr^*_{new}$ is reduced to 0.4. Base values of $k^-_{M_z}$ and $k^-_{P_z}$ correspond to half lives of 8 and 40 minutes respectively. $\langle M_z \rangle = 0.3\ nM$, $\langle P_z \rangle = 240\ nM$.

4. CONCLUSION

We have identified trends in changes in correlated fluctuations in protein levels as a function of rates, mechanism and structure. Thus the nature of the regulatory function leaves its signature via internal noise and can be used to quantitatively characterise speeds and magnitudes of transcriptional responses. While there are many networks with the same correlation, the variations in the correlations under modification could suggest putative structure and mechanism of transcritional control.

References

[1] M. B. Elowitz *et al*, "Stochastic gene expression in a single cell.," *Science*, vol. 297, pp. 1183-6, 2002.

[2] A. Raj and A. van Oudenaarden, "Nature, nurture, or chance: Stochastic gene expression and its consequences," *Cell*, vol. 135, no. 2, pp. 216-26, 2008.

[3] J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes*, Springer-Verlag, 1987.

[4] M. Thattai and A. van Oudenaarden, "Intrinsic noise in gene regulatory networks," *PNAS*, vol. 98, no. 15, pp. 8614-9, 2001.

[5] J. Paulsson, "Summing up the noise in gene networks.," *Nature*, vol. 427, no. 415-8, 2004.

[6] L. Cai *et al*, "Stochastic protein expression in individual cells at the single molecule level," *Nature*, vol. 440, pp. 358-62, 2006.

[7] A. Sigal *et al*, "Variability and memory of protein levels in human cells," *Nature*, vol. 444, pp. 643-6, 2006.

[8] M. J. Dunlop *et al*, "Regulatory activity revealed by dynamic correlations in gene expression noise," *Nat. Genetics*, vol. 40, pp. 1493-8, 2008.

[9] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, third edition, 2007.

[10] J. Elf and M. Ehrenberg, "Fast evaluation of fluctuations in biochemical networks with the linear noise approximation," *Genome Research*, vol. 13, pp. 2475-84, 2003.

[11] M. C. Teixeira *et al*, "The YEASTRACT database: a tool for the analysis of transcription regulatory associations in S. cerevisiae," *Nucl. Acids Res.*, vol. 34, pp. D446-51, 2006.

[12] C. D. Cox *et al*, "Using noise to probe and characterize gene circuits," *PNAS*, vol. 105, no. 31, pp. 10809-14, 2008.