Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Theoretical Biology 233 (2005) 423–433

Journal of Theoretical Biology

www.elsevier.com/locate/yjtbi

Stochastic effects of multiple regulators on expression profiles in eukaryotes

Pawel Paszek[a,*], Tomasz Lipniacki[a,b,d], Allan R. Brasier[c], Bing Tian[c], David E. Nowak[c], Marek Kimmel[a,1]

[a]Department of Statistics, Rice University, 6100 Main Street, MS-138, Houston, TX 77005, USA
[b]Institute of Fundamental Technological Research, Swietokrzyska 21, 00-049 Warsaw, Poland
[c]Department of Internal Medicine, University of Texas Medical Branch, Galveston, TX 77555-1060, USA
[d]Bioinformatics Program, University of Texas Medical Branch, Galveston, TX 77555-1060, USA

Received 25 June 2004; received in revised form 4 October 2004; accepted 19 October 2004
Available online 8 December 2004

Abstract

The stochastic nature of gene regulation still remains not fully understood. In eukaryotes, the stochastic effects are primarily attributable to the binary nature of genes, which are considered either switched "on" or "off" due to the action of the transcription factors binding to the promoter. In the time period when the gene is activated, bursts of mRNA transcript are produced. In the present paper, we investigate regulation of gene expression at the single cell level. We propose a mechanism of gene regulation, which is able to explain the observed distinct transcription profiles assuming the number of co-regulatory activities, without attempting to identify the specific proteins involved. The model is motivated by our experiments on NF-κB-dependent genes in HeLa cells. Our experimental data shows that NF-κB-dependent genes can be stratified into three characteristic groups according to their expression profiles: early, intermediate and late having maximum of expression at about 1, 3 and 6 h, respectively, from the beging of TNF stimulation. We provide a tractable analytical approach, not only in the terms of expected expression profiles and their moments, which corresponds to the measurements on the cell population, but also in the terms of single cell behavior. Comparison between these two modes of description reveals that single cells behave qualitatively different from the cell population. This analysis provides insights useful for understanding of microarray experiments.

Keywords: Stochastic gene regulation; Expression profiles; Single cell simulations; NF-κB

1. Introduction

Despite extensive studies carried out for the past several years the precise nature of transcriptional gene enhancement and regulation still remains not fully resolved. The fact that stochasticity plays an essential role in gene induction is well established, however there is still a discussion about major sources of randomness in the process of regulation. In prokaryotes, the stochasticity is attributed primarily to RNA polymerase binding, open complex formation, translation and degradation of mRNA and proteins (Johnson et al., 1979; Ackers et al., 1982; McAdams and Arkin, 1997). The stochastic effects are due to low abundance of mRNA transcript molecules (several molecules per cell on average), which means that production or degradation of just one mRNA transcript has a significant effect on cell behavior. On the other hand, in eukaryotic organisms, the average mRNA level is much higher (order of tens or hundreds of molecules) and the gene activity results in production of bursts of mRNA transcripts. This can be explained by the binary nature

[*] Corresponding author. Tel.: +1 713 348 6032; fax: +1 713 348 5476.
E-mail address: ppaszek@rice.edu (P. Paszek).
[1] Address for M. Kimmel: Institute of Automation, Silesian Technical University, Akademicka 16, 44-100 Gliwice, Poland.

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.10.023

of gene regulation; stochasticity is driven by binding and dissociation of transcription factors to and from the promoters of target genes. Thus, at a given time every gene copy is switched either "on" or "of", depending on whether the transcription factors are bound to the promoter or not (Ko et al., 1990; Ko, 1991; Kepler and Elston, 2001; Louis et al., 2003; Pirone and Elston, 2004). There is also another level of complexity in the gene regulation in eukaryotes corresponding to the chromatin–DNA interactions (reviewed in Wolfe and Pruss, 1996; Gregory et al., 2001; Eberharter and Becker, 2002) which should be taken into account. When the DNA is tightly bound by histones or other proteins, it may be too condensed to initiate transcription. In order for transcription to proceed, the DNA has to be accessible to the transcription factors and RNA polymerase. Chromatin structure has to be remodeled to expose the unfolded DNA for binding, which is accomplished by the process of histone acetylation. Acetylated histones relax chromatin structure, making the gene accessible to these regulatory proteins.

In the present paper, we address the problem of modeling stochastic regulatory mechanisms in eukaryotes. Available data provided by microarray experiments reveal that the genes can be clustered into distinct classes with respect to their expression profiles. We propose a model of gene regulation at the single cell level, which is able to generate distinct transcription profiles by involving different numbers of co-regulators. Without an attempt to identify these proteins, we hypothesize that they might be activating (e.g. histone acetylation) or repressing factors, not necessarily connected with DNA/protein binding. The model is motivated by the example of NF-κB-dependent genes in HeLa cells, which can be stratified into three characteristic classes according to their transcription profiles.

2. Motivation

The NF-κB- family of transcription factors plays an important role in pathogen or cytokine inflammation, immune response, cell proliferation and survival (Tian and Brasier, 2003). In mammals, this family contains five members, but the ubiquitously expressed NF-κB1/RelA heterodimer (referred herein as NF-κB) is responsible for the most common inducible NF-κB-binding activity. In resting cells NF-κB is sequestered in the cytoplasm by association with the members of another family of inhibitory proteins called IκB. In response to extracellular signals such as the tumor necrosis factor-α (TNF), IκB-inhibitory proteins are degraded, which exposes the NF-κB nuclear localization sequence and allows NF-κB to translocate into the nucleus, bind to κB motifs present in promoters of numerous genes and upregulate their transcription.

Because of its importance in inflammation and cell survival, the systematic identification of genetic networks downstream of NF-κB has been intensely pursued. Recently, cells engineered to have NF-κB activity controlled by exogenous doxycycline have been used to empirically identify the members of the NF-κB-dependent gene network by high-density microarrays (Tian et al., 2002; Tian and Brasier, 2003). In this system, the pattern of gene expression in wild-type cells in response to a stimulus is compared against the pattern produced by the same stimulus in the absence of NF-κB. In response to the cytokine TNF, 91 genes were identified to be NF-κB dependent by analysis of variance. Hierarchical clustering was used to group these genes into common expression profiles. As shown in Fig. 1, the NF-κB-responsive genes can be grouped into three characteristic classes: early (such as IκBα, A20 or IL8), for which the amount of mRNA transcript has its maximum at about 1 h, intermediate (such as NF-κB1 or TNFAIP2) with the maximum at 3 h, and late (such as NAF1 or NF-κB2) with the maximum at about 6 h.

For previously studied human fibroblast, the nuclear activity of NF-κB is terminated by the newly synthesized IκBα, which enters the nucleus, binds to NF-κB and takes it out into cytoplasm (Tian et al., 2002; Nelson et al., 2002; Lipniacki et al., 2004). Our experiments reveal that in HeLa cells, in contrast, NF-κB is not effectively lead out of the nucleus by the IκBα, but rather, after

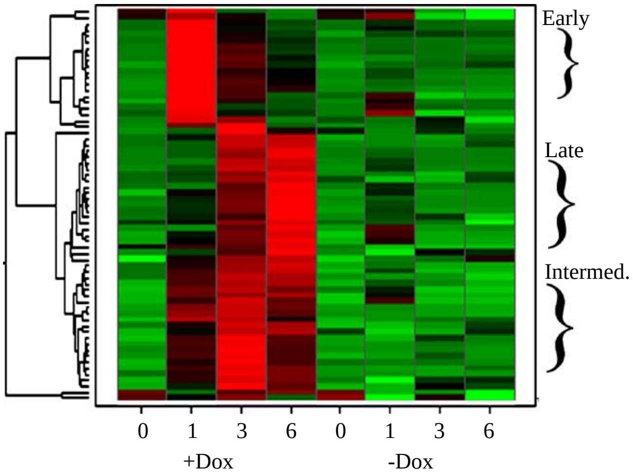

Fig. 1. Kinetics of NF-κB-dependent gene expression in HeLa cells (high expression depicted with red color). Data represent the mean of three independent time course experiments analysed by high-density microarrays. In this experiment the NF-κB nuclear translocation is enabled by culturing cells in the presence of doxycycline (Dox). The expression profiles for selected genes of each group where confirmed by Northern blots. The data reveal three characteristic classes of genes: early, intermediate and late, stratified by the time of 1, 3 and 6 h at which the level of the mRNA transcript reaches the maximum value.

entering the nucleus at 15 min from the beginning of TNF stimulation, it remains there for at least 6 h at a steady level. The Electrophoretic Mobility Shift Assay (EMSA) of the NF-κB-binding activity in TNF-stimulated HeLa cells was performed. In this assay, nuclear extract was prepared from a time series of stimulated cells and tested for binding to a radiolabeled high-affinity NF-κB-binding site (containing the sequence 5′-GGGATTTCCC-3′). After separation by non-denaturing gel electrophoresis, the relative change in NF-κB binding is observed (Fig. 2A). Based on their relative migration rate, the characteristic NF-κB1/RelA heterodimers are identified and found strongly induced by stimulation.

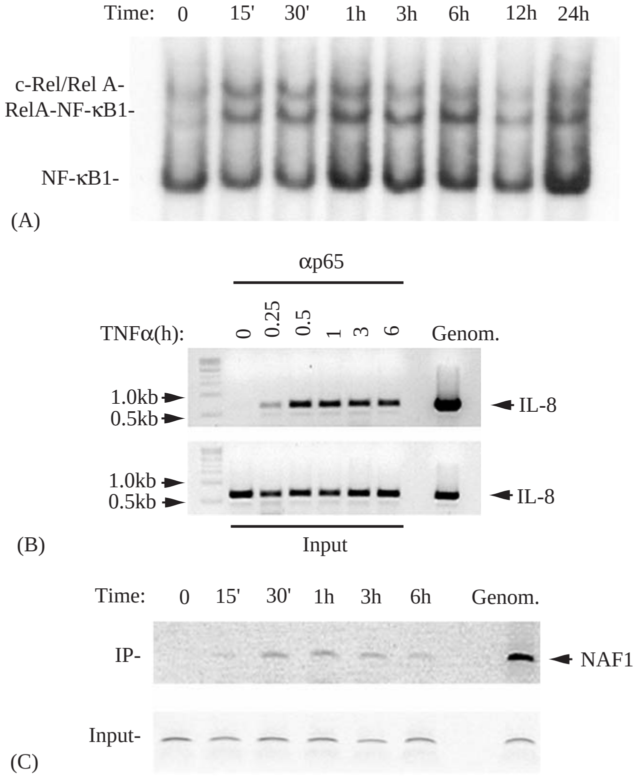
Time: 0 15' 30' 1h 3h 6h 12h 24h
c-Rel/Rel A-
RelA-NF-κB1-
NF-κB1-
(A)
αp65
TNFα(h): 0 0.25 0.5 1 3 6 Genom.
1.0kb
0.5kb
IL-8
1.0kb
0.5kb
IL-8
Input
(B)
Time: 0 15' 30' 1h 3h 6h Genom.
IP-
NAF1
Input-
(C)

Fig. 2. [A] The electrophoretic mobility shift assay (EMSA) of the NF-κB binding activity. HeLa S3 cells were stimulated with TNFα (30 ng/ml) for the indicated times prior to nuclear extraction and analysis of NF-κB binding by EMSA. Shown is an autoradiogram of the protein–DNA complexes. The relative migration of the specific NF-κB heterodimers is labeled. Rel A/NF-κB1 and c-Rel–RelA complexes are rapidly induced by TNF treatment within 15 min and persist in the nucleus for 6 h. A later peak at 24 h is also seen. [B] ChIP analysis of NF-κB association with IL-8 promoter (an early gene). HeLa cells were stimulated for various times with TNFα prior to formaldehyde fixation. Rel A was used as the immuno-precipitating antibody. After reversal of the crosslinks, qualitative PCR was performed using DNA from the immunoprecipitates (top panel) and input DNA as control (bottom panel). Far right lane is genomic DNA control. Rel A binds to the IL-8 promoter within 30 min of stimulation and persists for 6 h. [C] ChIP analysis of NF-κB association with NAF1 promoter (a late gene). The experiment is carried out on A549 respiratory (alveolar) epithelial cells, which have the same NF-κB kinetics as HeLa cells. Rel A was used as the immuno-precipitating antibody. After reversal of the crosslinks, qualitative PCR was performed using DNA from the immunoprecipitates (top panel) and input DNA as control (bottom panel). Far right lane is genomic DNA control. Rel A binds to the NAF1 promoter within 30 min of stimulation and persists for 6 h.

Next, we analysed NF-κB binding in the nucleus to endogenous target gene promoters. The experiments were carried out using Chromatin Immunoprecipitation (ChIP) Assay. In brief, NF-κB is crosslinked to its target promoter by addition of DNA–protein crosslinking agent, formaldehyde. After extraction of the DNA–-protein complexes, genes associated with RelA are purified by immunoprecipitation, and relative changes in NF-κB binding to its target gene are identified by the polymerase chain reaction. This assay provides an accurate index of the relative changes in NF-κB binding to endogenous gene promoters within its native chromatin context. We studied NF-κB association with promoters of two highly inducible genes: IL8, which is an early gene and NAF1, an example of a late gene. For both genes we observe similar kinetics, within 30 min from TNF treatment, transcriptionally active RelA binds to IL8 and NAF1 promoters and persists bound for at least 6 h at a steady level (Fig. 2B and C).

The experimental finding that NF-κB remains associated with IL-8 throughout the 6 h time course, even though its expression is actively being terminated strongly suggests the presence of a “repressor” that terminates NF-κB activity. Additionally, the fact that the same NF-κB-binding kinetics results in three different transcription profiles among the dependent genes, reveals that NF-κB by itself is not able to explain the observed phenomena. We hypothesize that some other factors are needed to initiate and terminate gene expression.

3. Model

To explain the expression profiles among three classes of genes we propose a model involving three activators and one repressor. This approach can be easily generalized to a problem of explaining N distinct expression profiles by N activators and one repressor. In the model we assume the following:

- Each gene has n potentially active homologous copies, and the activation and repression of these copies proceed independently. Typically each gene has two homologous copies, however sometimes only one copy is transcriptionally active. In cancer cells and in transfected cells the number of gene copies can be larger.
- Expression of an early or first class gene is initiated by the first activator, while to initiate expression of a gene from the second class one additional co-activator is needed, and to initiate expression of a gene from the

third class two co-activators are needed, in addition to the first one.

- Transcription of each gene is terminated by the repressor.
- Co-activators act according to a conditional mechanism, so there exists an order of events.
- All genes within each class have the same mRNA degradation half-time equal to δ_i, $i = 1, 2, 3$.
- Binding of the activators and of the repressor occurs in a stochastic way, with binding rates for the activators equal to λ_1, λ_2 and λ_3, respectively, and with the rate equal to λ_0 for the repressor.

Amount of mRNA transcript produced by the single gene copy j of a gene from class i, is modeled by the following ordinary differential equation:

$$\frac{\mathrm{d}x_i^j(t)}{\mathrm{d}t} = G_i^j(t) \cdot k_{prod} - \delta_i \cdot x_i^j(t), \tag{1}$$

where

- $i = 1, 2, 3$ corresponds to first, second and third class of genes, respectively,
- x_i is the amount of mRNA transcript,
- k_{prod}, δ_i are transcription and degradation rates, respectively,
- G_i^j is the status of the j homologous gene copy from the i-th class, $G_i^j = 1$ whenever all activators considered occupy promotory region, but the repressor is not bound, and $G_i^j = 0$ otherwise,
- initial conditions are $x(0) = 0$ and $G(0) = 0$.

The amount of mRNA transcript in a cell is a sum over amounts of transcript produced by each of homologous gene copies, given in Eq. (1).

We assume that co-activators act according to a conditional mechanism, so there exists an order of events. Let us define t_1 to be the time of the first activator binding, from the beginning of stimulation. Then, let t_2 be the time of the second activator binding, counted from the time of binding of the first activator. Similarly, let t_3 be equal to the duration of the period between second and third activator binding. Finally, let t_0 be the time between binding of last activator and repression event. In the model, the t_i's, $i = 0, 1, 2, 3$ are assumed to be independent and distributed exponentially with parameters λ_i, $i = 0, 1, 2, 3$, respectively. This is a consequence of the assumption that binding rates (or more precisely risk functions) of the regulatory factors are constant, which is equivalent to their constant amount (concentration) in the cell. In such scheme, the binding times are exponentially distributed random variables.

In addition, let us define a_i and r_i to be activation and repression times of genes from the i-th class. For the first class the activation time is equal to $a_1 = t_1$, for the second $a_2 = t_1 + t_2$, and for the third $a_3 = t_1 + t_2 + t_3$. The repression time among genes from i-th class is equal to $r_i = a_i + t_0$. Each of these characteristic times is a sum of the corresponding exponentially distributed random variables and their distributions can be analytically derived. The schematic representation of the model is depicted in Fig. 3 for the case of the genes from the second class.

The solution to Eq. (1) depends on the function $G_i^j(t)$ which is determined by the underlying stochastic process. The status of each homologous gene copy j from the i-th gene class $G_i^j(t) = 1$ if $t_i \in [a_i, r_i)$ and zero otherwise. Thus, to obtain the amount of transcript in the single cell produced by the gene from the i-th class, first, the function $G_i^j(t)$ is simulated by drawing a_i and r_i from the underlying probability densities for each gene copy j, the Eq. (1) is solved, and then the amount of transcript is summed over all gene copies.

It is possible to obtain the expected value of the analytical solution of Eq. (1), which describes behavior of a cell population. Let us notice that from the Eq. (1), the expected number of mRNA molecules $E[x_i^j(t)]$ over time (the average expression profile of a gene copy j from the i-th class) is described by the following differential equation:

$$\frac{\mathrm{d}E[x_i^j(t)]}{\mathrm{d}t} = g_i^j(t) \cdot k_{prod} - \delta_i \cdot E[x_i^j(t)], \tag{2}$$

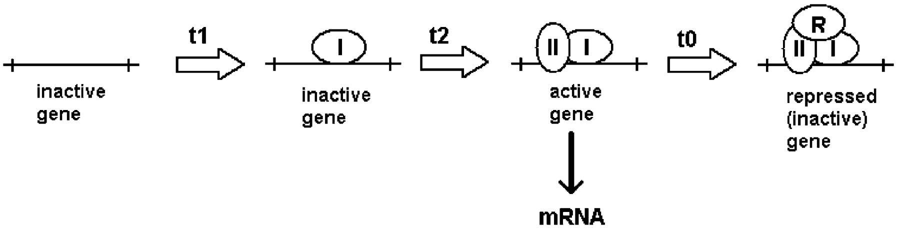

Fig. 3. Schematic representation of the model for the genes from the second (intermediate) class. Two activators, I and II, required for gene to start the mRNA transcription bind at the exponentially distributed times, t_1 and t_2, respectively. Gene activity is terminated by repressor, R, binding at exponentially distributed time t_0 following activation. The representation for other gene classes would differ only in number of activators needed for initiation of gene expression.

where $g_i^j(t) = E[G_i^j(t)]$ is the expected gene activity. The expected gene activity $g_i^j(t)$ can be evaluated as

$$g_i^j(t) = \int_0^\infty \int_0^\infty 1_{([a_i, a_i + t_0))}(t) f_{a_i}(a_i) f_0(t_0)\, \mathrm{d}t_0\, \mathrm{d}a_i$$
$$= \int_0^t \int_{t-a_i}^\infty f_{a_i}(a_i) f_0(t_0)\, \mathrm{d}t_0\, \mathrm{d}a_i, \qquad (3)$$

where $1_{(A)}(t)$ is an indicator function equal to 1 if $t \in A$ and zero otherwise, $f_0\,(t_0)$ is the probability density function of repressor binding and $f_{a_i}\,(a_i)$, $i = 1, 2, 3$, are activation distributions for consecutive gene classes. In general, in the case of N gene classes the expected gene activity $g_i^j(t)$, $1 \leqslant i \leqslant N$, can be written as (see the Appendix A for the detailed derivations in the case of $N = 3$):

$$g_i^j(t) = \sum_{k=0}^{i} \left[\left(\frac{\prod_{l=1}^{i} \lambda_l}{\prod_{l=0, l \neq k}^{i} (\lambda_l - \lambda_k)} \right) \mathrm{e}^{-\lambda_k t} \right]. \qquad (4)$$

Then the average expression profile of a gene copy j from the i-th class, can be derived by substituting $g_i^j(t)$ from Eq. (4) into Eq. (2) and solving it for $E[x_i^j(t)]$

$$E[x_i^j(t)] = \sum_{k=0}^{i} \left[\frac{k_{prod}}{\delta_i - \lambda_k} \left(\frac{\prod_{l=1}^{i} \lambda_l}{\prod_{l=0, l \neq k}^{i} (\lambda_l - \lambda_k)} \right) \times (\mathrm{e}^{-\lambda_k t} - \mathrm{e}^{-\delta_i t}) \right]. \qquad (5)$$

In the case of three gene classes we obtain the following expected expression profiles:

$$E[x_1^j(t)] = \frac{\lambda_1 k_{prod}}{(\lambda_1 - \lambda_0)(\delta_1 - \lambda_0)} (\mathrm{e}^{-\lambda_0 t} - \mathrm{e}^{-\delta_1 t}) + \frac{\lambda_1 k_{prod}}{(\lambda_0 - \lambda_1)(\delta_1 - \lambda_1)} (\mathrm{e}^{-\lambda_1 t} - \mathrm{e}^{-\delta_1 t}), \qquad (6)$$

$$E[x_2^j(t)] = \frac{\lambda_1 \lambda_2 k_{prod}}{(\lambda_1 - \lambda_0)(\lambda_2 - \lambda_0)(\delta_2 - \lambda_0)} (\mathrm{e}^{-\lambda_0 t} - \mathrm{e}^{-\delta_2 t}) + \frac{\lambda_1 \lambda_2 k_{prod}}{(\lambda_0 - \lambda_1)(\lambda_2 - \lambda_1)(\delta_2 - \lambda_1)} (\mathrm{e}^{-\lambda_1 t} - \mathrm{e}^{-\delta_2 t}) + \frac{\lambda_1 \lambda_2 k_{prod}}{(\lambda_0 - \lambda_2)(\lambda_1 - \lambda_2)(\delta_2 - \lambda_2)} \times (\mathrm{e}^{-\lambda_2 t} - \mathrm{e}^{-\delta_2 t}), \qquad (7)$$

$$E[x_3^j(t)] = \frac{\lambda_1 \lambda_2 \lambda_3 k_{prod}}{(\lambda_1 - \lambda_0)(\lambda_2 - \lambda_0)(\lambda_3 - \lambda_0)(\delta_3 - \lambda_0)} \times (\mathrm{e}^{-\lambda_0 t} - \mathrm{e}^{-\delta_3 t}) + \frac{\lambda_1 \lambda_2 \lambda_3 k_{prod}}{(\lambda_0 - \lambda_1)(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\delta_3 - \lambda_1)} \times (\mathrm{e}^{-\lambda_1 t} - \mathrm{e}^{-\delta_3 t}) + \frac{\lambda_1 \lambda_2 \lambda_3 k_{prod}}{(\lambda_0 - \lambda_2)(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2)(\delta_3 - \lambda_2)} \times (\mathrm{e}^{-\lambda_2 t} - \mathrm{e}^{-\delta_3 t}) + \frac{\lambda_1 \lambda_2 \lambda_3 k_{prod}}{(\lambda_0 - \lambda_3)(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)(\delta_3 - \lambda_3)} \times (\mathrm{e}^{-\lambda_3 t} - \mathrm{e}^{-\delta_3 t}). \qquad (8)$$

In addition, let us define a gene from the i-th class to be active at given time t, if mRNA transcript is produced by at least one of its copies at time t. Assuming that n copies of a gene act independently, the proportion of active cells (i.e. cells with an active gene) in the population at time t can be evaluated using Eq. (4):

$$E[g_i(t)] = 1 - (1 - g_i^j(t))^n. \qquad (9)$$

Proposed here, relatively simple mathematical formulation of the problem allows finding an analytical solution. We derive it for the case of N gene classes corresponding to N distinct expression profiles [Eq. (5)]. Unfortunately, even in the case of three gene classes the final expressions are fairly complicated. Thus, we propose another, more tractable description of the problem in the terms of moments of the expression profiles. We use the Laplace transform to derive first three moments of $x_i^j(t)$ describing the properties of expression profiles (see the Appendix B for detailed derivations). We obtain the following zeroth, first and second absolute moments of the expression profiles of j homologous gene copy from the i-th class, $M_i^{j(0)}$, $M_i^{j(1)}$ and $M_i^{j(2)}$, respectively:

$$M_i^{j(0)} = \frac{k_{prod}}{\delta_i} t_0, \qquad (10)$$

$$M_i^{j(1)} = \frac{1}{\delta_i} + a_i + \frac{t_0}{2}, \qquad (11)$$

$$M_i^{j(2)} = \frac{2}{\delta_i^2} + \frac{t_0^2}{3} + \frac{t_0}{\delta_i} + t_0 a_i + \frac{2a_i}{\delta_i} + a_i^2. \qquad (12)$$

The first moment describes center of gravity of the expression profiles, while the second corresponds to their variation. The advantage of this approach is the simplicity of the derived mathematical expressions, which are still capable of giving insights into the analysis. Evaluating the expectations of the moments with respect to distribution of a_i and t_0 corresponds to the moments of expression profiles averaged over the cell population. The first moments of the average profile are given then by the following:

$$\bar{M}_1^{(1)} = \frac{1}{\delta_1} + \frac{1}{2\lambda_0} + \frac{1}{\lambda_1}, \qquad (13)$$

$$\bar{M}_2^{(1)} = \frac{1}{\delta_2} + \frac{1}{2\lambda_0} + \frac{1}{\lambda_1} + \frac{1}{\lambda_2}, \qquad (14)$$

$$\bar{M}_3^{(1)} = \frac{1}{\delta_3} + \frac{1}{2\lambda_0} + \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3}, \qquad (15)$$

where the subscript i in $\bar{M}_i^{(1)}$ corresponds to the i-th gene class. Similarly, the second average moments are the following:

$$\bar{M}_1^{(2)} = \frac{2}{\delta_1^2} + \frac{2}{3\lambda_0^2} + \frac{1}{\delta_i\lambda_0} + \left(\frac{1}{\lambda_0} + \frac{2}{\delta_1}\right)\frac{1}{\lambda_1} + \frac{2}{\lambda_1^2}, \tag{16}$$

$$\begin{aligned}\bar{M}_2^{(2)} = {} & \frac{2}{\delta_2^2} + \frac{2}{3\lambda_0^2} + \frac{1}{\delta_i\lambda_0} + \left(\frac{1}{\lambda_0} + \frac{2}{\delta_2}\right)\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right) \\ & + \frac{2}{\lambda_1^2} + \frac{2}{\lambda_2^2} + \frac{2}{\lambda_1\lambda_2},\end{aligned} \tag{17}$$

$$\begin{aligned}\bar{M}_3^{(2)} = {} & \frac{2}{\delta_3^2} + \frac{2}{3\lambda_0^2} + \frac{1}{\delta_i\lambda_0} + \left(\frac{1}{\lambda_0} + \frac{2}{\delta_3}\right)\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3}\right) \\ & + \frac{2}{\lambda_1^2} + \frac{2}{\lambda_2^2} + \frac{2}{\lambda_3^2} + \frac{2}{\lambda_1\lambda_2} + \frac{2}{\lambda_1\lambda_3} + \frac{2}{\lambda_2\lambda_3}.\end{aligned} \tag{18}$$

It is possible to express the second absolute moments as a function of the corresponding first moment, which leads to the following formula:

$$M_i^{j(2)} = (M_i^{j(1)})^2 + Var(a_i) + \frac{1}{\delta_i^2} + \frac{1}{6\lambda_0^2}, \tag{19}$$

hence the second central moments (i.e. the variation in the expression profile) is given by

$$V_i^j = M_i^{j(2)} - (M_i^{j(1)})^2 = Var(a_i) + \frac{1}{\delta_i^2} + \frac{1}{6\lambda_0^2}, \tag{20}$$

It is noteworthy that the differences in the second central moment among the gene classes are driven mostly by the variance of the corresponding activation time a_i, which can be further decomposed into sum of variances of respective random variables $t_i^\prime$s. Thus, according to the model, the variation among the profiles increases while considering more activators, thus the transcription profiles of later genes are characterized by broader distributions and smaller number of mRNA at the peak of transcriptional activity.

One may think that to model expression profiles among the N gene classes we do not need N activators. At first glimpse, it seems that all profiles can be fitted using even one activator by changing degradation rate δ_i (i.e. fitting the first moment described by Eq. (13) to all gene classes). Later in the manuscript, we show that to fit desired profiles very unrealistic values of this parameter have to be taken and the fit is still not accurate. Therefore, we assume that the variation among expression profiles between different gene groups is due to the number of activators and not to the degradation rate of mRNA transcript. This leads to simplification of the problem, when assuming that all genes have the same mRNA degradation half-time δ. The moments corresponding to the i-th gene class can be expressed as the functions of moments of the corresponding $i-1$ gene class, which may be of use while fitting the desired experimental profiles. In the case of the first moment we have the following:

$$\bar{M}_1^{(1)} = \frac{1}{\delta} + \frac{1}{2\lambda_0} + \frac{1}{\lambda_1}, \tag{21}$$

$$\bar{M}_2^{(1)} = \bar{M}_1^{(1)} + \frac{1}{\lambda_2}, \tag{22}$$

$$\bar{M}_3^{(1)} = \bar{M}_2^{(1)} + \frac{1}{\lambda_3}. \tag{23}$$

Thus, having the moments calculated from the experimental data it is possible to uniquely determine values of parameters λ_2 and λ_3, while relationship between λ_0 and λ_1 is given by the Eq. (21). In case of estimating these parameters from the expected expression profiles [Eq. (6)–(8)] relationships between model parameters are much more complicated.

4. Results

In this section, we fit the model to reproduce 3 classes of expression profiles of NF-κB-dependent genes presented in Fig. 1. Having not enough data from the microarray experiments to calculate the first moments we decided to fit the characteristic times of the maximum mRNA transcript abundance given at 1, 3 and 6 h after the stimulation for early, intermediate and late genes, respectively. It is assumed that every gene has two potentially active homologous copies. In addition, it is assumed that all genes have the same mRNA degradation half-time equal to 20 min, which corresponds to degradation rate $k_{dgr} = 0.00057\,\mathrm{s}^{-1}$. This is in agreement with experimental data from Blattner et al. (2000), who estimated the degradation half-time for early gene IκBα to be within 15–30 min range. In fact, as we show later, the two-fold increase of the degradation half-time results in relatively small change in expression profiles. Also we assume a common transcription rate among genes equal to 4 mRNA molecules per minute per gene copy ($k_{prod} = 0.0667\,\mathrm{s}^{-1}$), which was confirmed for β-actin by single RNA transcript visualization (Femino et al., 1998). Given this, the profiles are determined by the set of four parameters λ_i, $i = 0, 1, 2, 3$ describing binding rates of three activators and one repressor.

The fitting procedure was carried out analytically by taking the time derivative of the profiles given by Eqs. (6)–(8) and equating it to zero. Given the three characteristic times of peak transcription at 1, 3, and 6 h the parameters can be derived by solving obtained equations. Unfortunately the problem is indeterminate, since there are four unknown parameters and only three expressions describing them. Thus, taking into account the fast dynamics of NF-κB, we assume that the expected binding time of the first activator is equal to

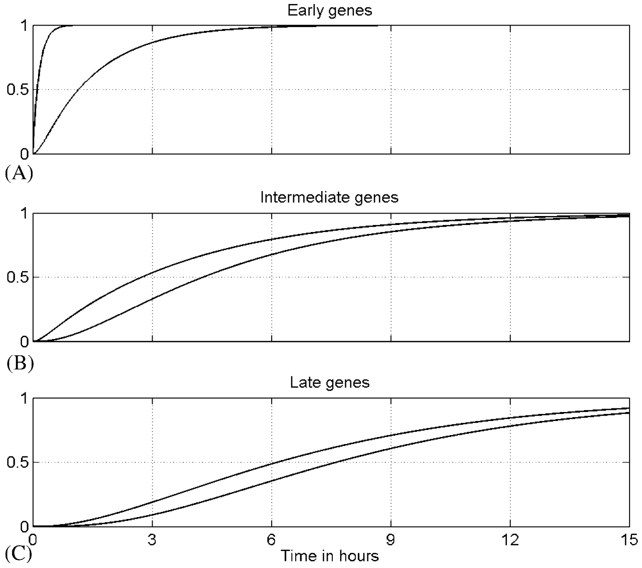

Fig. 4. Cumulative distribution functions of activation, a_i, and repression times, r_i, for three hypothetical genes belonging to each of the early, panel [A], intermediate, panel [B], and late, panel [C], classes (activation distribution function is located on the top of its corresponding repression distribution function). Cumulative distribution functions were derived from their probability density functions obtained similarly to derivations in Eq. (A.1) and Eq. (A.2).

10 min, which corresponds to the binding rate $\lambda_1 = 1.667 \times 10^{-3}\,\mathrm{s}^{-1}$. Then, the remaining parameters are determined by numerically solving given equations (this cannot be done analytically, since the expressions are transcendental). As a result, the following binding rates were fitted: $\lambda_2 = 7.4 \times 10^{-5}$, $\lambda_3 = 8.1 \times 10^{-5}$ and $\lambda_0 = 1.96 \times 10^{-4}\,\mathrm{s}^{-1}$which corresponds to the expected binding times of 225, 205 and 85 min for the two remaining activators and one repressor, respectively.

Analytical results of the fitted model for three hypothetical genes belonging to the early, intermediate and late classes, respectively, are depicted in Figs. 4 and 5. Fig. 4 presents cumulative distribution functions of activation and repression times a_i and r_i, $i = 1, 2, 3$, derived as sums of independent exponential random variables. One may notice that probability that an early gene is activated within first hour is close to 1, Fig. 4A, but the corresponding probability for intermediate and late genes is much smaller. In fact, the cumulative probability that a late gene was activated within first hour is 0.025, and about 0.5 at 6 h from beginning of stimulation, Fig. 4C. The less steep distribution functions of activation and repression for intermediate and late genes than for the early genes result from larger variability in binding times of the former. This

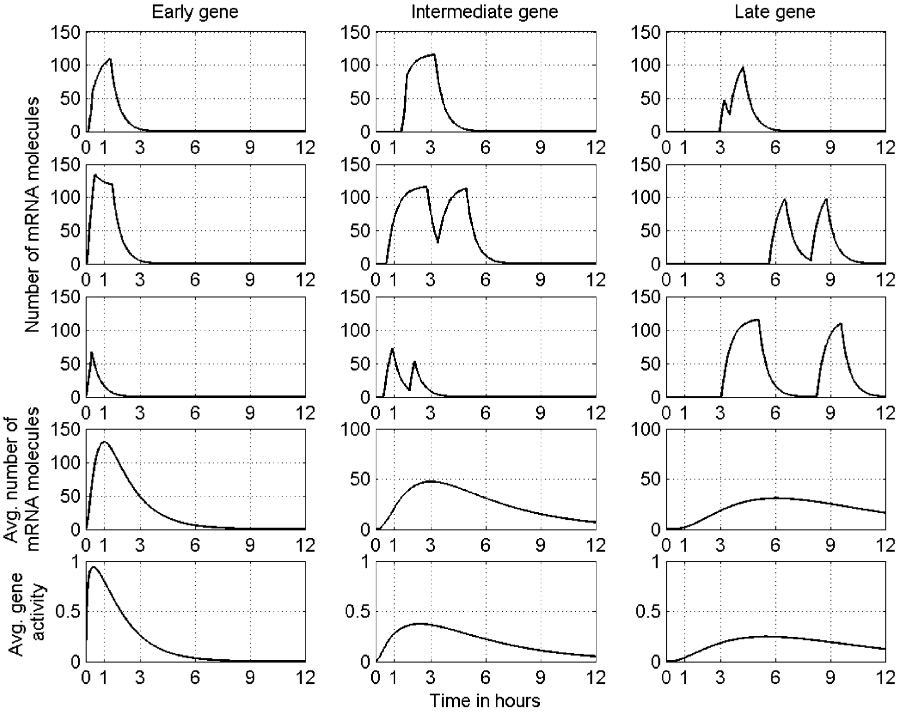

Fig. 5. The mRNA profiles for early, intermediate and late genes. First 3 rows show mRNA profiles in single cells, while the profiles in fourth row depict the expected expression in the cell culture derived in Eq. (6)–(8). These latter profiles should be compared to experimental data in Fig. 1. Finally, in the last row, gene activity over time is shown, defined in Eq. (9) as the proportion of cells in the culture with at least one of two gene copies active, given in Eq. (A.3)–(A.5). The kinks visible on single cell profiles correspond to initiation or termination of expression in any of the two homologous copies of the gene. The difference among single cell profiles is larger for the late genes and, as a result, the averaged expression profile is broader, what is well confirmed by Northern blot data (Nowak, 2004).

immediately affects the expression profiles of genes among different classes. Corresponding results are depicted in Fig. 5. First three rows correspond to simulations of the single cells. These profiles have characteristic kinks, which reflect initiation or termination of expression in any of two homologous copies of the gene. The fourth row includes mRNA profiles analytically derived in Eqs. (6)–(8), corresponding to expression profiles averaged over the population of cells. These should be compared to experimental data in Fig. 1. Finally, in the last row, the gene activity over time is shown, defined in Eq. (9) as the proportion of the cells in the culture with at least one of the two gene copies active, derived in Eqs. (A.3)–(A.5). Single cell expression profiles are significantly different from the profiles obtained for the population of cells. In fact, no individual cell behaves like an "average" one. This is especially visible for the late genes, where the variability among single cell profiles is much larger than for early and intermediate genes. Another interesting observation is that not all cells (genes) are active in the population even at their peak transcriptional activity. In fact, the proportion of active cells at their maximum activity is about 0.93 for the early class, but significantly decreases to 0.37 and 0.25 for intermediate and late genes, respectively.

Obtained average expression profiles fit our microarray observations on NF-κB-dependent genes very well. Also, the fact that variability among single cell profiles is larger for late genes and, as a result, the averaged expression profile is broader, was well confirmed by Northern blot data (Nowak, 2004).

5. Discussion

In the present paper, we propose the mechanism of gene regulation at a single cell level, which is able to generate N characteristic classes of expression profiles obtained by microarray experiments, involving N activators and 1 repressor. We developed an analytical description of the model in the terms of expected profiles corresponding to the measurements in the cell culture and applied it to the three characteristic classes of profiles of NF-κB-dependent genes in HeLa cells. It was shown that the simulated results are in a strong agreement with our experimental observations from microarray study. The analysis was carried out based on the assumption that degradation rates among gene classes are equal. Such constrained model is less flexible and more difficult to fit the experimental data. The full model, which allows to fit separate degradation rates will be of use when more time course data is available. One can expect the expression profiles to be more diverse in that case, hence the unconstrained model with its flexibility will be more suitable. Moreover, we developed another analytical description of the model in terms of moments of expression profiles, which is also of potential use when more data is available. The latter description might be more straightforward and robust, since the estimates of the moments are much more stable than the estimates of the maximum intensity of the expression profiles.

We are still missing some information about parameters in the model. Based on the experimental data from Blattner et al. (2000), the mRNA degradation rate was set to 20 min for all genes (results depicted in Fig. 5). The two-fold increase of the degradation half-time results in relatively small change in expression profiles, Fig. 6. To avoid further uncertainty in the model, 20 min degradation rate was set as a common value for all genes. Another assumed parameter is the maximum transcription rate, which has been measured for β-actin by single RNA transcript visualization, and found to be equal to 4 mRNA molecules per minute per gene copy (Femino et al., 1998). We followed this estimate, however changing it does not affect the shape of expression profiles, it only changes levels of mRNA transcript. Having the degradation rate defined, the set of parameters which determines the expression profiles in the model is given by four binding probabilities of three activators and one repressor. The fitting problem is still not fully determined, since there are four unknown parameters and three degrees of freedom, corresponding to three characteristic times of peak transcription. As a result there is some freedom in

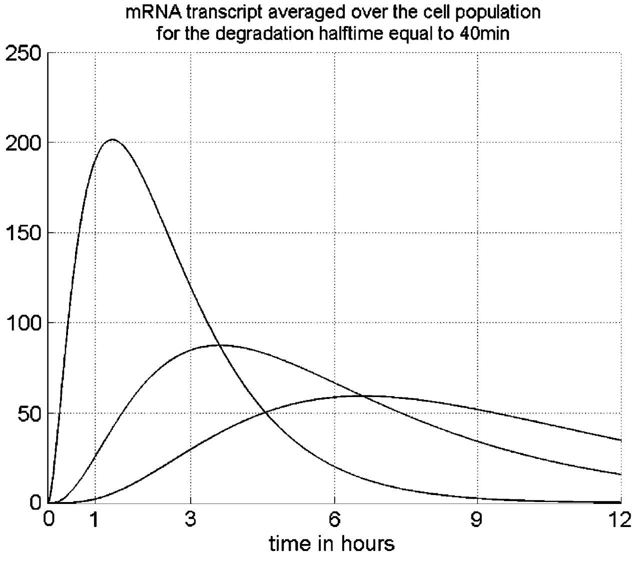

Fig. 6. Expected mRNA expression profiles for early, intermediate and late genes (characterized by decreasing maximum, respectively). Degradation half-time is set to 40 min for all gene classes. Other parameter values are assumed as in Fig. 5. The two-fold increase of the degradation half-time results in a relatively small change in expression profiles.

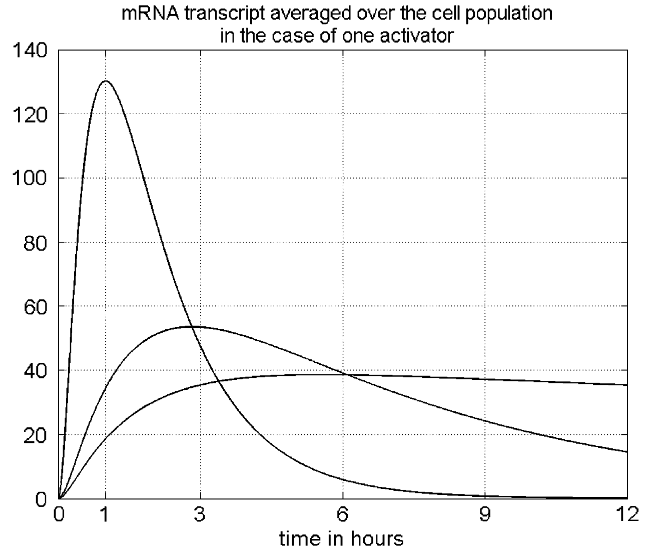

Fig. 7. Expected mRNA expression profiles for early, intermediate and late genes (characterized by decreasing maximum, respectively) considering one activator (NF-κB). To obtain the desired maximum times, the degradation half-times were set to 20 min, 3 and 40 h, respectively. Corresponding transcription rates were set to 4, 0.8 and 0.25 mRNA per minute per gene copy.

choosing λ_0 and λ_1, while other parameters can be determined uniquely.

It is still possible to fit the three characteristic profiles of gene expression when only one or two activators (including NF-κB) are considered. The case of one activator is depicted in Fig. 7. To obtain the desired maximum times the degradation half-times were set to 20 min, 4 and 40 h for the early, intermediate and late genes, respectively. Then, to preserve the general property that intermediate and late genes are characterized by lower expression, the transcription rate was chosen separately for each gene class and set to 4, 0.5 and 0.25 mRNA per minute per gene copy. The fit obtained is not fully satisfactory. For example, for the late genes the amount of the transcript increases rapidly and stabilizes for several hours. Moreover, the 40 h mRNA degradation rate, necessary to obtain the maximum at 6 h, is very unrealistic. Similarly, the transcription rate of 0.25 mRNA per minute is 16 times smaller than the rate measured for β-actin in situ (Femino et al., 1998). For the case of two activators it is also possible to obtain the desired expression profiles, but we still have problems with the fit (data not shown).

The model proposed assumes a conditional mechanism of activator binding, i.e. there is an order of binding events (NF-κB binding is followed by other events). It is intuitively justified to think that for the intermediate and late genes NF-κB binds first, and then, after this information is acquired by the cell, other co-activators bind. The question arises, whether this latter binding occurs in a certain order or not. In the model we consider a conditional mechanism of co-activators binding, however this assumption can be easily relaxed to the unconditional mechanism. In terms of fitting the desired expression profiles, both models are equally good. The model with unconditional mechanism requires slightly different parametrization to obtain the fit.

The proposed method of modeling gene expression at a single cell level provides some interesting insights. Single cell expression profiles are significantly different from the profiles constructed by averaging over the population of cells. No individual cell behaves like an "average" one. This is especially visible in the example of the late NF-κB-dependent genes, where the variability among single cell profiles is much larger than for early and intermediate genes. This observation has a strong implication in the terms of understanding the microarray experiments. The time course microarray experiments provide us with measurements of gene expression averaged over the cell population. These measurements have continuous values, but in a single cell, at a given time moment, the targeted gene is either "on" or "off". It is also shown that not all cells (genes) are always active in the culture, in fact, their number may be unexpectedly small. Thus it is misleading to think that every cell in the tissue responds gradually in the terms of the expression level, as the microarray measurements might seem to suggest. One should rather think about a proportion of the transcriptionally active cells, at the given time in the population. This behavior of single cells was confirmed in Ko et al. (1990) and Ko (1992), when the microarray experiments were not available yet.

Acknowledgements

We would like to thank Dr. Bruce Luxon for discussion and support. This work was supported by NHLBI Contract N01-HV-28184, Proteomic technologies in airway inflammation (A. Kurosky, P.I.) and by KBN (Polish Committee for Scientific Research) Grant No. 8T07A 045 20.

Appendix A

A.1. Derivation of expected gene activity

The expected gene activity $g_i^j(t)$ given by Eq. (3) can be evaluated in straightforward manner using model assumptions and convolution formula. The probability density functions of repressor binding and activation in the first gene class, $f_0\ (t_0)$ and $f_{a_1}\ (a_1)$, are given by the assumption:

$$f_0(t_0) = \lambda_0 \mathrm{e}^{-\lambda_0 t_0}.$$

$$f_{a_1}(a_1) = \lambda_1 \mathrm{e}^{-\lambda_1 a_1}.$$

To calculate f_{a_2} (a_2), let us notice that $a_2 = t_1 + t_2$ is a sum of two exponentially distributed independent random variables, thus its distribution is given by the convolution formula:

$$f_{a_2}(a_2) = \int_0^{a_2} \lambda_1\lambda_2 \mathrm{e}^{-\lambda_1 t}\mathrm{e}^{-\lambda_2(a_2-t)}\,\mathrm{d}t = \frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}(\mathrm{e}^{-\lambda_1 a_2} - \mathrm{e}^{-\lambda_2 a_2}). \quad (A.1)$$

Similarly, the distribution f_{a_3} (a_3) can be calculated by noticing that $a_3 = a_2 + t_3$ and using the previous result,

$$f_{a_3}(a_3) = \frac{\lambda_1\lambda_2\lambda_3}{\lambda_2-\lambda_1}\left[\frac{1}{\lambda_3-\lambda_1}(\mathrm{e}^{-\lambda_1 a_3} - \mathrm{e}^{-\lambda_3 a_3}) - \frac{1}{\lambda_3-\lambda_2}(\mathrm{e}^{-\lambda_2 a_3} - \mathrm{e}^{-\lambda_3 a_3})\right]. \quad (A.2)$$

Given that, $g_i^j(t)$, the expected gene activity of a single copy j from the i-th class, $i = 1, 2, 3$, can be obtained by solving Eq. (3). We obtain the following:

$$g_1^j(t) = \frac{\lambda_1}{\lambda_1-\lambda_0}\,\mathrm{e}^{-\lambda_0 t} + \frac{\lambda_1}{\lambda_0-\lambda_1}\,\mathrm{e}^{-\lambda_1 t}, \quad (A.3)$$

$$g_2^j(t) = \frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_0)(\lambda_2-\lambda_0)}\,\mathrm{e}^{-\lambda_0 t} + \frac{\lambda_1\lambda_2}{(\lambda_0-\lambda_1)(\lambda_2-\lambda_1)}\,\mathrm{e}^{-\lambda_1 t} + \frac{\lambda_1\lambda_2}{(\lambda_0-\lambda_2)(\lambda_1-\lambda_2)}\,\mathrm{e}^{-\lambda_2 t}, \quad (A.4)$$

$$g_3^j(t) = \frac{\lambda_1\lambda_2\lambda_3}{(\lambda_1-\lambda_0)(\lambda_2-\lambda_0)(\lambda_3-\lambda_0)}\,\mathrm{e}^{-\lambda_0 t} + \frac{\lambda_1\lambda_2\lambda_3}{(\lambda_0-\lambda_1)(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}\,\mathrm{e}^{-\lambda_1 t} + \frac{\lambda_1\lambda_2\lambda_3}{(\lambda_0-\lambda_2)(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)}\,\mathrm{e}^{-\lambda_2 t} + \frac{\lambda_1\lambda_2\lambda_3}{(\lambda_0-\lambda_3)(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}\,\mathrm{e}^{-\lambda_3 t}. \quad (A.5)$$

This result can be further generalized to the case of N gene classes (see Eq. (4) in the Model section).

Appendix B

B.1. Derivation of moments of expression profiles using Laplace transform

We transform Eq. (1) by multiplying both sides by e^{-st} and integrating out with respect to dt:

$$\int_0^\infty \frac{\mathrm{d}x_i^j(t)}{\mathrm{d}t}\,\mathrm{e}^{-st}\,\mathrm{d}t = \int_0^\infty (k_{prod} \cdot G_i^j(t) - \delta_i \cdot x_i^j(t))\mathrm{e}^{-st}\,\mathrm{d}t = \int_0^\infty (k_{prod} \cdot 1_{([a_i, a_i+t_0))}(t) - \delta_i \cdot x_i^j(t))\mathrm{e}^{-st}\,\mathrm{d}t,$$

this gives

$$X_i^j(s) = \frac{k_{prod}}{s(\delta_i + s)}\,\mathrm{e}^{-a_i s}(1 - \mathrm{e}^{-t_0 s}),$$

where

$$X_i^j(s) = \int_0^\infty x_i^j(t)\mathrm{e}^{-st}\,\mathrm{d}t \quad (B.1)$$

is the Laplace transform of $x_i^j(t)$.

Now we calculate the absolute moments of $x_i^j(t)$ from Eq. (B.1). The zeroth moment is

$$M_i^{j(0)} = \int_0^\infty x_i^j(t)\,\mathrm{d}t = X_i^j(s)|_{s=0}. \quad (B.2)$$

The higher order absolute moments $M_i^{j(k)}$, $k>0$, (normalized by $M_i^{j(0)}$) are obtained by differentiating Eq. (B.1) k times with respect to s:

$$M_i^{j(k)} = \frac{1}{M_i^{j(0)}}\int_0^\infty t^k x_i^j(t)\,\mathrm{d}t = \frac{1}{M_i^{j(0)}}(-1)^k \left.\frac{d^k X_i^j(s)}{\mathrm{d}s^k}\right|_{s=0}. \quad (B.3)$$

References

Ackers, G.K., Johnson, A.D., Shea, M.A., 1982. Quantitative model for gene regulation by λ phage repressor. Proc. Natl Acad. Sci. USA 79, 1129–1133.

Blattner, C., Kannouche, P., Litfin, M., Bender, K., Rahmsdorf, H.J., Angulo, J.F., Herrlich, P., 2000. UV-induced stabilization of c-fos and other short-lived mRNAs. Mol. Cell. Biol. 20, 3616–3625.

Eberharter, A., Becker, P.B., 2002. Histone acetylation: a switch between repressive and permissive chromatin. EMBO Rep. 3, 224–229.

Femino, A.M., Fay, F.S., Fogarty, K., Singer, R.H., 1998. Visualization of single RNA transcripts in situ. Science 280, 585–590.

Gregory, P.D., Wagner, K., Hörz, W., 2001. Histone acetylation and chromatin remodeling. Exp. Cell Res. 265, 195–202.

Johnson, A.D., Meyer, B.J., Ptashne, M., 1979. Interactions between DNA-bound repressors govern regulation by the λ phage repressor. Proc. Natl Acad. Sci. USA 76, 5061–5065.

Kepler, T.B., Elston, T.C., 2001. Stochasticity in transcriptional regulation: orgins, consequences, and mathematical representations. Biophys. J. 81, 3116–3136.

Ko, M.S.H., 1991. A stochastic model for gene induction. J. Theor. Biol. 153, 181–194.

Ko, M.S.H., 1992. Induction mechanism of a single gene molecule: stochastic or deterministic? Bioassays 14 (5), 341–346.

Ko, M.S.H., Nakauchi, H., Takahashi, N., 1990. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. EMBO J. 9 (9), 2835–2842.

Lipniacki, T., Paszek, P., Brasier, A.R., Luxon, B., Kimmel, M., 2004. Mathematical model of NF-κB module. J. Theor. Biol. 228, 195–215 (doi:10.1016/j.jtbi.2004.01.001).

Louis, M., Holm, L., Sanchez, L., Kaufman, M., 2003. A theoretical model for the regulation of *sex-lethal*, a gene that controls sex determination and dosage compensation in *Drosophila melanogaster*. Genetics 165, 1355–1384.

McAdams, H.H., Arkin, A., 1997. Stochastic mechanisms in gene expression. Proc. Natl Acad. Sci. USA 94, 814–819.

Nelson, G., Paraoan, L., Spiller, D.G., Wilde, G.J.C., Browne, A.M., Djali, P.K., Unitt, J.F., Sullivan, E., Floettmann, E., White, M.R.H., 2002. Multi-parameter analysis of the kinetics of NF-κB signaling and transcription in single living cells. J. Cell Sci. 115, 1137–1148.

Nowak, D., E., 2004. Unpublished data.

Pirone, J.R., Elston, T.C., 2004. Fluctuations in transcription factor binding can explain the graded and binary responses observed in inducible gene expression. J. Theor. Biol. 226, 111–121 (doi:10.1016/j.jtbi.2003.08.008).

Tian, B., Brasier, A.R., 2003. Identification of a nuclear factor kappa B-dependent gene network. Recent Progr. Hormone Res. 58, 95–130.

Tian, B., Zhang, Y., Luxon, B.A., Garofalo, R.P., Casola, A., Sinha, M., Brasier, A.R., 2002. Identification of NF-κB-dependent gene networks in respiratory syncytial virus-infected cells. J. Virol. 76, 6800–6814.

Wolfe, A.P., Pruss, D., 1996. Targeting chromatin disruption: transcription regulators that acetylate histones. Cell 84, 817–819.