Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 238 (2006) 348–367
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

# Transcriptional stochasticity in gene expression

Tomasz Lipniacki[a,b,c,*], Pawel Paszek[b], Anna Marciniak-Czochra[d], Allan R. Brasier[e], Marek Kimmel[a,f]

[a]Institute of Fundamental Technological Research, Swietokrzyska 21, 00-049 Warsaw, Poland
[b]Department of Statistics, Rice University, 6100 Main St. MS-138, Houston, TX 77005, USA
[c]Bioinformatics Program, University of Texas Medical Branch, Galveston, TX 77555-1060, USA
[d]Institute of Applied Mathematics, University of Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany
[e]Department of Internal Medicine, University of Texas Medical Branch, Galveston, TX 77555-1060, USA
[f]Institute of Automation, Silesian Technical University, 44-100 Gliwice, Poland

Received 19 March 2005; received in revised form 16 May 2005; accepted 23 May 2005
Available online 21 July 2005

**Abstract**

Due to the small number of copies of molecular species involved, such as DNA, mRNA and regulatory proteins, gene expression is a stochastic phenomenon. In eukaryotic cells, the stochastic effects primarily originate in regulation of gene activity. Transcription can be initiated by a single transcription factor binding to a specific regulatory site in the target gene. Stochasticity of transcription factor binding and dissociation is then amplified by transcription and translation, since target gene activation results in a burst of mRNA molecules, and each mRNA copy serves as a template for translating numerous protein molecules. In the present paper, we explore a mathematical approach to stochastic modeling. In this approach, the ordinary differential equations with a stochastic component for mRNA and protein levels in a single cells yield a system of first-order partial differential equations (PDEs) for two-dimensional probability density functions (pdf). We consider the following examples: Regulation of a single auto-repressing gene, and regulation of a system of two mutual repressors and of an activator–repressor system. The resulting PDEs are approximated by a system of many ordinary equations, which are then numerically solved.
© 2005 Elsevier Ltd. All rights reserved.

*Keywords:* Gene regulation; Transcription; Stochasticity; Probability density function; Transport-type equations

## 1. Introduction

Genes code for proteins, but the pathway between the code and the product involves several distinct processes. First, in eukaryotes, the majority of genes have to be activated. This typically happens by binding one or more transcription factors to the specific promoter regions. Then, RNA polymerase binds to the gene promoter, and an open complex is formed which initiates transcription. Resulting mRNA is spliced and polyadenylated. Next, mRNA is exported from the nucleus to the cytoplasm, where translation occurs. In many cases, the newly translated protein must be further processed to form its biologically active form. Some of the above processes are reversible like binding of transcription factors, recruitment of RNA polymerase and formation of an open complex. A simplified schematic diagram of gene expression is shown in Fig. 1. In eukaryotes, each gene has two homologous copies, which can be independently activated and inactivated. In some cases, one of these copies is transcriptionally inactive. In addition, transformed cells may have gene or chromosomal duplications producing a larger number of homologous gene copies. The

*Corresponding author. Department of Statistics, Rice University, 6100 Main St. MS-138, Houston, TX 77005, USA. Tel.: +1 713 348 36 85; fax: +1 713 348 5476.
*E-mail addresses:* tomek@rice.edu, tlipnia@ippt.gov.pl (T. Lipniacki).

0022-5193/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2005.05.032

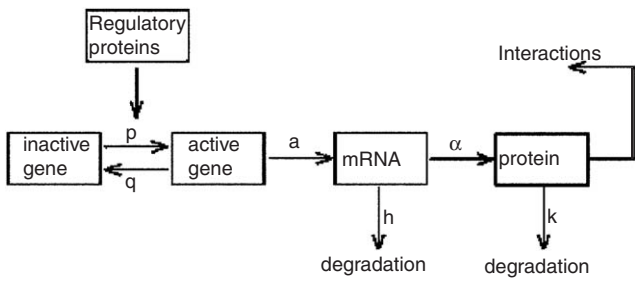


Fig. 1. Simplified schematic of gene expression.

number of copies of molecules involved grows as the process proceeds from DNA to protein. Tens or hundreds of mRNA molecules of a given type and tens of thousands of protein molecules are produced using two gene copies as templates. This implies that stochastic effects due to gene activation and inactivation followed by pulses of mRNA production (Femino et al., 1998; Blake et al., 2003), are much stronger than stochastic effects caused by production and degradation of single mRNA or protein molecules. In contrast, in prokaryotes, where the mRNA molecules are typically very unstable (half-life of the order of 1 min) and therefore much less abundant, the stochasticity of formation, degradation and translation of single mRNAs is of great importance (McAdams and Arkin, 1997; Ackers et al., 1982; Thattai and Oudenaarden, 2001; Kierzek et al., 2001; Swain et al., 2002). As a result, in prokaryotes there is a competition between stochastic effects caused by gene activation and mRNA processing. In this paper we focus on stochasticity in transcriptional regulation neglecting the mRNA and protein production/decay noise. We start from the approximation of single cell kinetics in which processes involving a large number molecules (i.e., transcription, translation and degradation of protein and mRNAs) are considered continuous, and are described by ordinary differential equations (ODEs). These equations, describing evolution of mRNA and protein levels in a single cell, contain a stochastic switch associated with gene activity. These equations yield a system of first-order partial differential equations for pdf's, from which two-dimensional mRNA-protein distributions, stationary and time-dependent, will be numerically calculated. Marginal distribution of protein is compared to the marginal distribution resulting from the Kepler–Elston model, in which mRNA is disregarded and direct translation of protein from the gene is assumed. This gives us the range of applicability of the Kepler–Elston model. Finally, we use this approximate model to analyse the two-gene systems for which we will calculate the two-dimensional protein–protein distributions.

## 2. Model

### 2.1. Preliminaries

Since gene activation and inactivation is due to binding and dissociation of regulatory factors to and from DNA, it is natural to assume that activation and inactivation rates depend on amounts (concentrations) of regulatory proteins. As a simplest example, let us consider regulation of a gene by an activating regulatory protein, the level ($y$) of which is constant in time. It is assumed that each gene copy may exist only in two states (Ko, 1991; Walters et al., 1995; Kepler and Elston, 2001; Pirone and Elston, 2004; Raser and O'Shea, 2004). The state of the $i$-th gene copy is denoted by $g_i \in \{0, 1\}$. We assume that the $i$-th gene copy is activated with rate $py$, and inactivated with rate $q$. The mRNA production efficiency from gene copy $i$-th is assumed to be equal to $a\, g_i$, where $a$ is transcription rate of the active gene copy. Thus the amount of transcript, $x$, follows the equation below:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -hx + aG(t, y(\cdot)), \tag{1}$$

where $h$ is the mRNA degradation rate and $G(t, y(\cdot)) = \sum_{i=1}^{n} g_i$, where $n$ is number of homologous gene copies. The approximation in which the amount of transcript is described by continuous variable $x$ is justified only when number of mRNA molecules is sufficiently large. For prokaryotes, in which number of mRNA is typically small, it is not justified. Eq. (1) is similar to a stochastic differential equation except that in our case the stochastic term is not an additive white noise, but a time continuous Markov process of gene activation and inactivation. First, let us calculate the mean (expected) mRNA level $E[x]$ in the population. If $y$ is assumed constant, the probability $P$ that the $i$-th gene copy is active equals to $P = py/(q + py)$, which is constant in time. The expected mRNA production rate is the product of $Pn$ and the transcription rate $a$, hence the expected amount of transcript equals

$$E[x] = \frac{napy}{h(q + py)}. \tag{2}$$

Although the expected level agrees with the classical result, the mRNA level in each cell, given by Eq. (1), oscillates, as shown in Fig. 2. The deterministic limit is attained under very frequent binding and dissociation of the regulatory factor, when $py \longrightarrow \infty$, $q \longrightarrow \infty$, with $py/q = \text{const}$. In this latter case, $P[y(t)] = py(t)/(q + py(t))$ determines the expected transcription rate $naP[y(t)]$. However, experimental data indicate that mRNA is produced in bursts, which suggests that $py$ and $q$ are small (Femino et al., 1998). There are two characteristic regimes; (1) $py \gg h, q \gg h$ and (2) $py \ll h, q \ll h$. In the first regime, the dynamics

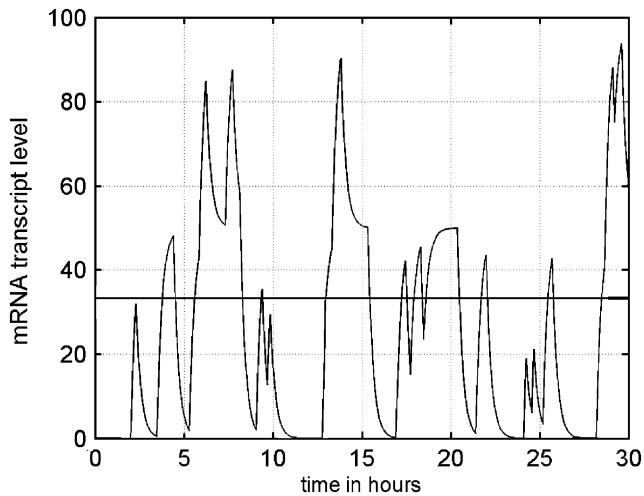


Fig. 2. A sample solution of Eq. (1), $n = 2$, $a = 0.1$, $py = 10^{-4}$, $q = 5 \cdot 10^{-4}$, $h = 10^{-3}$. Although each cell oscillates the averaged mRNA level in population remains constant at 33 mRNA/cell.

resembles that of the thermodynamic equilibrium limit, in the second regime, we have long bursts of mRNA from active gene copies and the level of mRNA varies significantly, shown in Fig. 2. The assumption that gene promoters are in thermodynamic equilibrium with regulatory protein molecules, was proposed over 20 years ago by Ackers et al. (1982) and Shea and Ackers (1985). It was used by Arkin et al. (1998) and Gilman and Arkin (2002) in analysis of prokaryotic gene expression, in which stochastic effects are due to a small number of mRNA molecules. In this work we will exploit the second regime, where changes in gene status are relatively infrequent.

### 2.2. Regulation of auto-repressive gene—single cell analysis

In this section we present a heuristic analysis of stochastic regulation of an auto-repressive gene in a single cell, whereas the corresponding probability distributions will be analysed in next section. Let $x$ and $y$ denote mRNA and protein levels, respectively. Since the gene is auto-repressive, we assume that the $i$-th gene copy is activated at a constant rate $p$, and inactivated at a rate $qy$. The resulting dynamics is given by the following system:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -hx + aG(t, y(\cdot)), \tag{3}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \alpha x - ky, \tag{4}$$

$$I \xrightarrow{p} A, \quad I \xleftarrow{qy(t)} A, \tag{5}$$

where $\alpha$ is the translation rate, $k$ the protein degradation rate, $I$ denotes the gene inactive state, $A$ denotes the active state, and $G(t, y(\cdot)) = \sum_{i=1}^{n} g_i$, where $g_i(I) = 0$, $g_i(A) = 1$. We introduce rescaled variables for problem (3)–(5):

$$x^* = \frac{h}{a}x, \quad y^* = \frac{kh}{a\alpha}y, \quad t^* = th. \tag{6}$$

Substituting new variables into Eqs. (3)–(5) and dropping the asterisks, we obtain the following system:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -x + G, \tag{7}$$

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = r(x - y), \tag{8}$$

$$I \xrightarrow{c} A, \quad I \xleftarrow{by(t)} A, \tag{9}$$

where

$$c = \frac{p}{h}, \quad b = \frac{qa\alpha}{kh^2}, \quad r = \frac{k}{h}. \tag{10}$$

For a given state of $n$ gene copies, which determines $G = \sum_{i=1}^{n} g_i$, the system is analytically solved as an initial value problem $x(0) = x_0$, $y(0) = y_0$, assuming that $G$ remains constant,

$$x(t) = x_0 \mathrm{e}^{-t} + G(1 - \mathrm{e}^{-t}), \tag{11}$$

$$y(t) = y_0 \mathrm{e}^{-rt} + G(1 - \mathrm{e}^{-rt}) + \frac{(G - x_0)r}{1 - r}(\mathrm{e}^{-t} - \mathrm{e}^{-rt}). \tag{12}$$

The risk that $G$ changes its value is equal to $\rho(t) = (n - G)c + Gby(t)$, where $n - G$, $G$ denote the number of inactive and active gene copies, respectively. This yields the cumulative distribution $F(t)$,

$$F(t) = 1 - \exp\left(-\int_0^t \rho(s)\,\mathrm{d}s\right). \tag{13}$$

We draw $z$ from the uniform distribution on $[0, 1]$ and calculate $t_f$, the time at which $G$ changes its value as $t_f = F^{-1}(z)$, where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. Finally, we decide whether $G$ switches to $G + 1$ or to $G - 1$. The conditional probability that it switches to $G + 1$, i.e. that one of homologous gene copies changes its status from inactive to active, is equal to

$$p^+ = \frac{\rho^+}{\rho(t_f)} = \frac{n_1 c}{n_1 c + n_2 by(t_f)}. \tag{14}$$

Finally, we evaluate $G$, and use $x(t_f), y(t_f)$ as the initial condition for next step. Inversion of $F(t)$ cannot be carried out analytically, since $\log(1 - F(t))$ is a transcendental function.

In the limit $c \to \infty$, $b \to \infty$, with $c/b = \text{const.}$, the promoter of each gene copy is in statistical equilibrium and the function $G = \sum_{i=1}^{n} g_i$ may be replaced by its expected value $E[G(y)] = nc/(c + by)$. As a result,

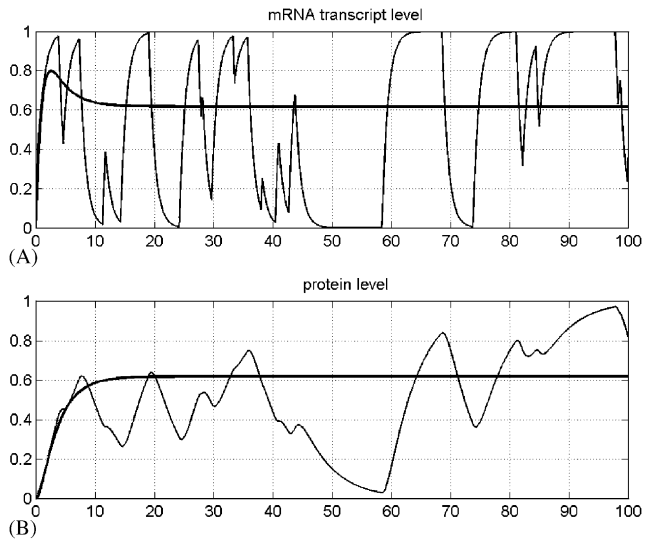


Fig. 3. A sample solution of system (7)–(9) for $n = 1$ (single gene copy), represented by a thin line. The initial condition is $x(0) = y(0) = G(0) = 0$ and parameters are $c = 0.5$, $b = 0.5$, $r = 0.2$. Since $c<1$, $b<1$ the solution is far from the solution of the limit system (15)–(16), represented by a bold line.

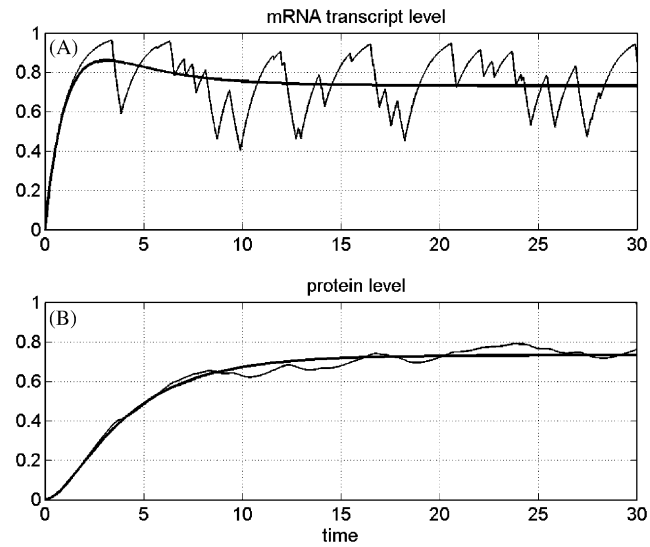


Fig. 4. A sample solution of system (7)–(9) for $n = 1$ (single gene copy), represented by a thin line. The initial condition is $x(0) = y(0) = G(0) = 0$ and the parameters are $c = 6$, $b = 3$, $r = 0.2$. Since $c>1$, $b>1$ the solution is relatively close to the solution of the limit system (15)–(16), represented by a bold line.

system (7)–(8) converges to the following,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -x + \frac{nc}{c + by}, \tag{15}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = r(x - y). \tag{16}$$

In Fig. 3 we show the evolution of system, for $c<1$, $b<1$, i.e., when system is far from its thermodynamic equilibrium limit. Large fluctuations in both mRNA and protein levels are present. In Fig. 4 we show the evolution of the system in proximity of the thermodynamic limit ($c = 6, b = 3$). Substantial fluctuations of the mRNA level still persist but the protein level exhibits relatively small fluctuations. Moreover, the solution is close to the solution of the limit system (15)–(16). However, we expect that the situation depicted in Fig. 3 is more common.

### 2.3. Equations for probability distributions of model variables

From now on, to simplify the analysis, we focus on the case, where the autoregulatory gene has only one potentially active copy. This implies that its state $G$ can be 0 or 1. Eqs. (7) and (8) generate stochastic trajectories, which can be described as a continuous-time Markov process. At each time $t$, the realizations of mRNA and protein levels $x(t)$ and $y(t)$, and of the transcription switch $G(t)$ are a triple of random variables, the first two of which are continuous and the third is binary. Therefore, their joint distribution can be described by a pair of pdf $f(x, y, t)$ and $g(x, y, t)$. The interpretation is that for given time $t$,

$$\begin{aligned}&\Pr[x(t) \in (x, x + \Delta x),\ y(t) \in (y, y + \Delta y) \text{ and } G(t) = 0]\\ &\quad = f(x, y, t)\Delta x \Delta y,\\ &\Pr[x(t) \in (x, x + \Delta x),\ y(t) \in (y, y + \Delta y) \text{ and } G(t) = 1]\\ &\quad = g(x, y, t)\Delta x \Delta y.\end{aligned}$$

One interpretation of these distributions is in the terms of frequencies of the mRNA and protein levels and of the state of the transcription switch in a large population of cells.

The equations describing evolution of densities $f$ and $g$ were first derived by us using an analogy between probability and compressible fluid. Using this approach, one writes the continuity equations with source terms following from change of gene status (transformation between $f$ and $g$), Eq. (9). Velocity fields $(\mathrm{d}x/\mathrm{d}t)_{|G=0}$ and $(\mathrm{d}x/\mathrm{d}t)_{|G=1}$ transforming $f$ and $g$, are given by Eqs. (7) and (8), for $G = 0$ and $G = 1$, respectively:

$$\frac{\partial f}{\partial t} + \mathrm{div}\left[\left(\frac{\mathrm{d}x}{\mathrm{d}t}_{|G=0}, \frac{\mathrm{d}y}{\mathrm{d}t}\right) f\right] = byg - cf, \tag{17}$$

$$\frac{\partial g}{\partial t} + \mathrm{div}\left[\left(\frac{\mathrm{d}x}{\mathrm{d}t}_{|G=1}, \frac{\mathrm{d}y}{\mathrm{d}t}\right) g\right] = -byg + cf. \tag{18}$$

The above system of first-order partial differential equations (PDE) is analogous to the Fokker–Planck equation which describes evolution of pdf in the process governed by the stochastic differential equation (Langevin equation), (Rao et al., 2002; Emch and Liu, 2002; Kepler and Elston, 2001). The difference is in

the right-hand term, which in the case of Fokker–Planck equation is a diffusion term resulting from the white noise term in Langevin equation. A derivation of a generalized version of Eqs. (17) and (18) is provided in Appendix A. Similar equations have been used in physics to describe noise-induced transitions (Horsthemke and Lefever, 1984) and in theoretical mechanics to describe dynamics of rigid bodies under random shocks (Iwankiewicz and Nielsen, 2000). A system analogous to our system (7)–(9) in which the transition intensities of the random forcing process $G(t)$ depend on state variables $x(t)$ and $y(t)$ was considered by Basak et al. (1999). Let us note that the solutions of system (7)–(8) for $G=0$ and $G=1$ are identical with the two characteristics of the PDE system. Using Eqs. (7) and (8) we obtain

$$\frac{\partial f}{\partial t}+\operatorname{div}[(-x, r(x-y))f]=byg-cf, \tag{19}$$

$$\frac{\partial g}{\partial t}+\operatorname{div}[(1-x, r(x-y))g]=-byg+cf, \tag{20}$$

and then

$$\frac{\partial f}{\partial t}-\frac{\partial}{\partial x}(xf)+r\frac{\partial}{\partial y}[(x-y)f]=byg-cf, \tag{21}$$

$$\frac{\partial g}{\partial t}+\frac{\partial}{\partial x}[(1-x)g]+r\frac{\partial}{\partial y}[(x-y)g]=-byg+cf. \tag{22}$$

Since Eqs. (21) and (22) are of first order and are coupled only by the free terms, the system is hyperbolic. Functions $f(x,y)$ and $g(x,y)$ may be considered on domain $D_0=[0,\infty)\times[0,\infty)$. However, we can restrict our considerations to a bounded domain $D\subset D_0$, defined below. Let us consider two following solutions of the system described by Eqs. (7) and (8), see Fig 5. The first $(x_1(t), y_1(t))$ assumes $G=0$ and the initial condition $(x_1(0), y_1(0))=(1,1)$, which results in

$$x_1=\mathrm{e}^{-t},\quad y_1=-\frac{r}{1-r}\,\mathrm{e}^{-t}+\frac{1}{1-r}\,\mathrm{e}^{-rt}. \tag{23}$$

The second $(x_2(t), y_2(t))$ assumes $G=1$ and initial condition $(x_2(0), y_2(0))=(0,0)$, which results in

$$x_2=1-\mathrm{e}^{-t},\quad y_2=1+\frac{r}{1-r}\,\mathrm{e}^{-t}-\frac{1}{1-r}\,\mathrm{e}^{-rt}. \tag{24}$$

These two solutions parametrically define two curves in the $x,y$ plane intersecting at points $(0,0)$ and $(1,1)$. Let us define the domain $D(r)$, as a subset of $D_0$ bounded by these two curves. If the kinetics of the cell is determined by system (7)–(8) then

1. if $(x(0), y(0)\in D(r)$ then $(x(t), y(t)\in D(r)$ for any $t>0$,
2. trajectories $x(t), y(t)$ of system (7)–(8) starting from an arbitrary point $(x(0), y(0))$ converge to the domain $D(r)$ as $t\to\infty$.

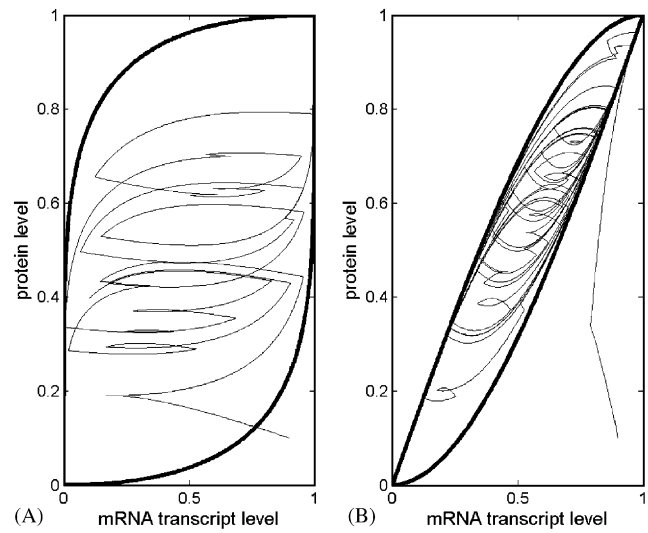


Fig. 5. The boundary of the domain $D(r)$—bold line, and a typical trajectory of system (7)–(9)—thin line. Panel A: $r=0.2$, $c=0.5$, $b=0.5$, Panel B: $r=3$, $c=2$, $b=2$.

In addition, $D(r)$ is the smallest domain satisfying these two conditions. In Fig. 5 we show domain $D(r)$ for $r=0.2$ and 3 together with the trajectories of system (7)–(9). After the trajectory enters domain $D(r)$ it remains there.

For stationary density functions we have $\partial f/\partial t=\partial g/\partial t=0$ and, consequently,

$$-\frac{\partial}{\partial x}(xf)+r\frac{\partial}{\partial y}[(x-y)f]=byg-cf, \tag{25}$$

$$\frac{\partial}{\partial x}[(1-x)g]+r\frac{\partial}{\partial y}[(x-y)g]=-byg+cf. \tag{26}$$

The above system, like (21)–(22), is hyperbolic. It has two families of characteristics, the first family (for function $f$) is determined by the solutions of system (7)–(8) for $G=0$, the second family (for function $g$) is determined by the solutions of system (7)–(8) for $G=1$. According to the definition, boundary of $D(r)$ consists of two characteristics of (25) and (26) given by Eqs. (23) and (24). Stationary distributions $f(x,y)$ and $g(x,y)$ must satisfy

$$\operatorname{supp}(f(x,y;c,b,r))=\operatorname{supp}(g(x,y;c,b,r))=D(r), \tag{27}$$

where the notation used underscores the parametric dependence of $f(\cdot)$ and $g(\cdot)$ on $c$, $b$ and $r$, and $\operatorname{supp}(f)$ is the closure of the set on which $f(\cdot)\neq 0$. The fact that the domain $D(r)$ is bounded by characteristics, makes the problem difficult for analytical and numerical analysis.

### 2.4. Numerical results

Numerical solutions of system (25)–(26) are calculated based on the discrete approximation introduced in Appendix B. This discretization technique results in a

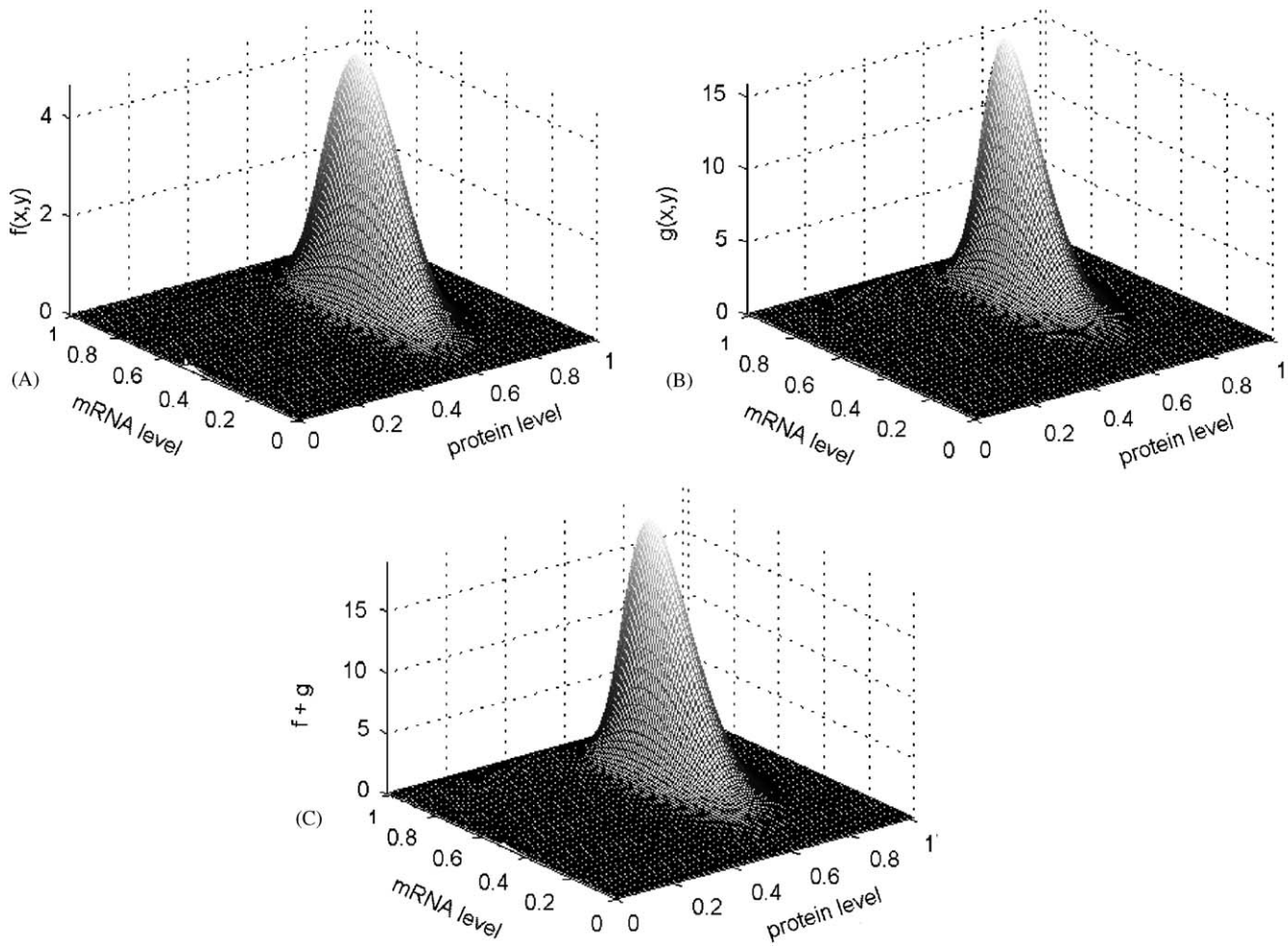


Fig. 6. Stationary distributions $f(x,y)$ (Panel A), $g(x,y)$ (Panel B) and $\rho(x,y) = f(x,y) + g(x,y)$ (Panel C) calculated on grid $100 \times 100$. The parameters are $c = 6$, $b = 3$, $r = 0.2$.

reduction of system (21)–(22) to a system of linear ODEs, while system (25)–(26) for stationary distributions is reduced to a system of linear algebraic equations. (Fig. 6)

In Figs. 6 and 7, using the same set of parameters as in Fig. 4, we compare stationary distributions $f(x)$ and $g(x)$ given by system (25)–(26), with the distribution calculated directly from simulations of system (7)–(9). To draw Fig. 7, we simulate system (7)–(9) up to $t_f = 15$, $N = 500,000$ times, starting each simulation from the same initial condition $x(0) = y(0) = G(0) = 0$. The initial conditions for the simulation are $f(x,y,0) = \delta(x,y)$ (Dirac impulse at $(0,0)$), and $g(x,y,0) = 0$. As a result of simulation at any time $t \in [0,15]$, we obtain $N$ points $(x(t), y(t), G(t))$. Points $(x(t), y(t), 0)$ approximate the distribution $f(x,y,t)$, and points $(x(t), y(t), 1)$ approximate $g(x,y,t)$. In Fig. 7, we present numerical distributions for $t = t_f$, calculated using a relatively coarse grid ($N = 20$) to avoid noise. Since, for a typical trajectory $(x(t), y(t), G(t))$, the status of gene $G$ changes several times prior to $t_f$ (see Fig. 4) we may expect that distributions $f(x,y,t_f)$ and $g(x,y,t_f)$ are close to stationary. The marginal distribution $\rho(x,y) = f(x,y) + g(x,y)$ reflects the behavior of a single cell. The larger the single cell fluctuations are, the broader is the marginal distribution $\rho(x,y)$. Since in Fig. 4 the fluctuations in $y$ (protein level) are much smaller than those in $x$ (mRNA level) the corresponding distribution shown in Fig. 6 is much broader in the $x$ than in the $y$ direction.

In the case of Fig. 8, where the same parameters are used as in Fig. 3 the marginal distribution is much broader than in the previous case. Moreover, the profiles of distributions $f$ and $g$ are qualitatively different, the distribution $f$ having a maximum at $x$ close to 0, while $g$ having a maximum at $x$ close to 1. This property is in agreement with the single cell trajectory shown in Fig. 3. In Fig. 3 fluctuations in gene status are infrequent and therefore the mRNA transcript level is strongly correlated with gene status. As a result, the marginal distribution $\rho(x,y)$ has not one, but two, maxima. This property can be helpful in experimental verification of the model and in estimation of parameters $c$ and $b$.

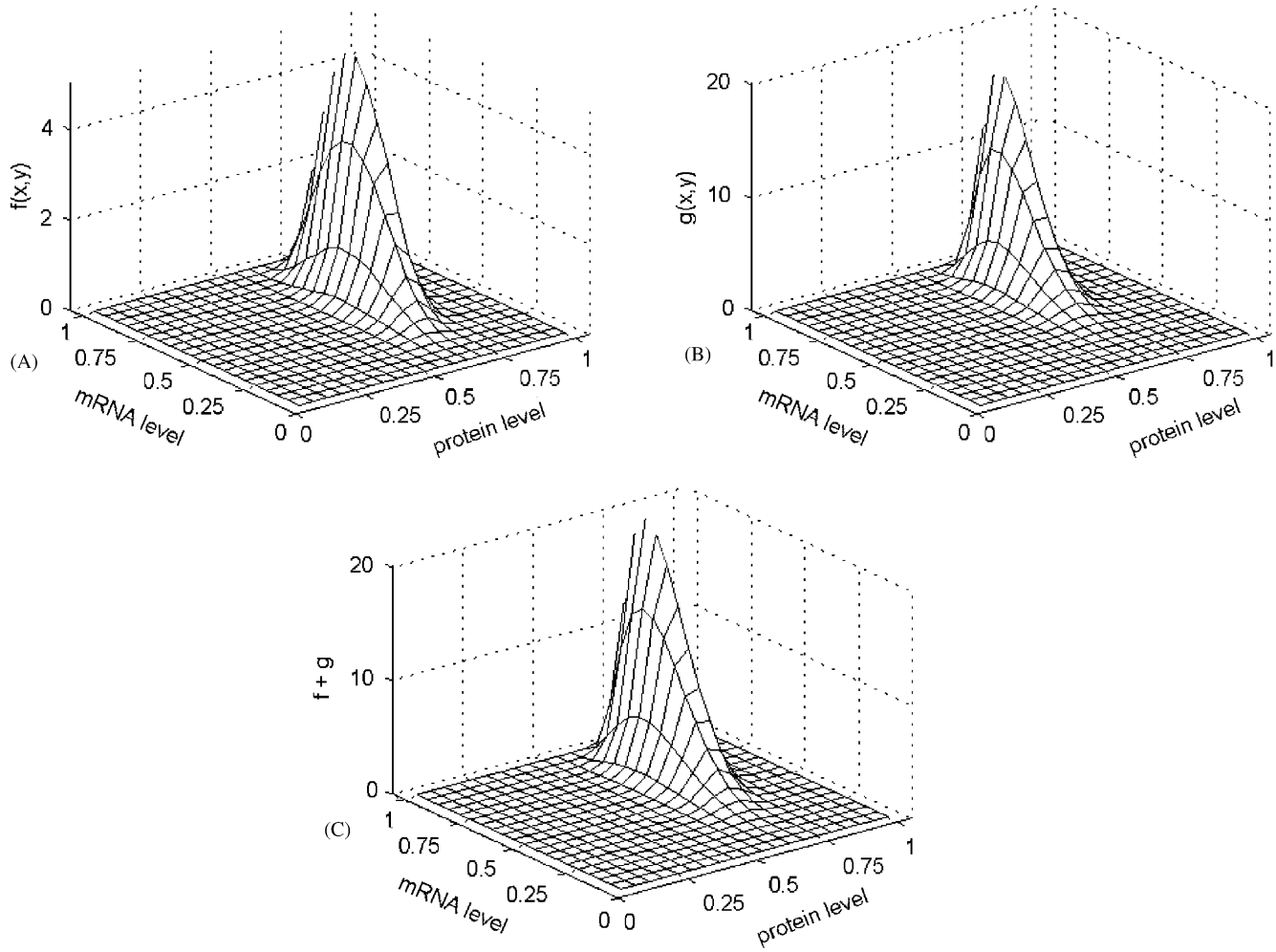


Fig. 7. Distributions $f(x,y,t)$ (Panel A), $g(x,y,t)$ (Panel B) and $\rho(x,y,t) = f(x,y) + g(x,y)$ (Panel C) calculated for $t = 15$ by running 500,000 single cell simulations of system (7)–(9). The same parameters $c = 6$, $b = 3$, $r = 0.2$ as in Fig. 6 are used. The data is shown using a relatively small grid $20 \times 20$ to avoid noise.

In Fig. 9 we analyse the case of large $r = 3$. As already said, for $r \to \infty$, $x - y \to 0$, and as a result large $r$ implies that $x$ and $y$ are close. This is visible in distributions presented in Fig. 9, which are concentrated close to the $x = y$ line.

Finally in Fig. 10 we show time evolution of distributions $f(x,y,t)$ and $g(x,y,t)$. Parameter values are the same as in Figs. 6 and 7, and the initial conditions are $f(x,y,0) = \delta(x,y)$, and $g(x,y,0) = 0$, same as for Fig. 7. Note that for $t = 10$, distributions are close to the stationary ones shown in Fig. 6. This could be expected from a single cell trajectory shown in Fig. 4, which reaches the stationary distribution (loses memory of the initial condition) about $t = 10$–15.

### 2.5. Kepler–Elston approximation

Analysing system (7)–(8), we may note that for $r \ll 1$, Eq. (7) is much faster than Eq. (8), which allows us to replace Eq. (7) by the equality $x = G$. As a result system (7)–(9) is transformed into

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = r(G - y), \tag{28}$$

$$I \xrightarrow{c} A, \quad I \xleftarrow{by(t)} A. \tag{29}$$

The above approximation is equivalent to the assumption made by Kepler and Elston (2001) that the protein is directly translated from the gene. Equations for the probability density functions $f(y,t)$, $g(y,t)$, corresponding to the simplified system (28)–(29) read

$$\frac{\partial f}{\partial t} - r\frac{\partial}{\partial y}(fy) = byg - cf, \tag{30}$$

$$\frac{\partial g}{\partial t} + r\frac{\partial}{\partial y}((1-y)g) = -byg + cf. \tag{31}$$

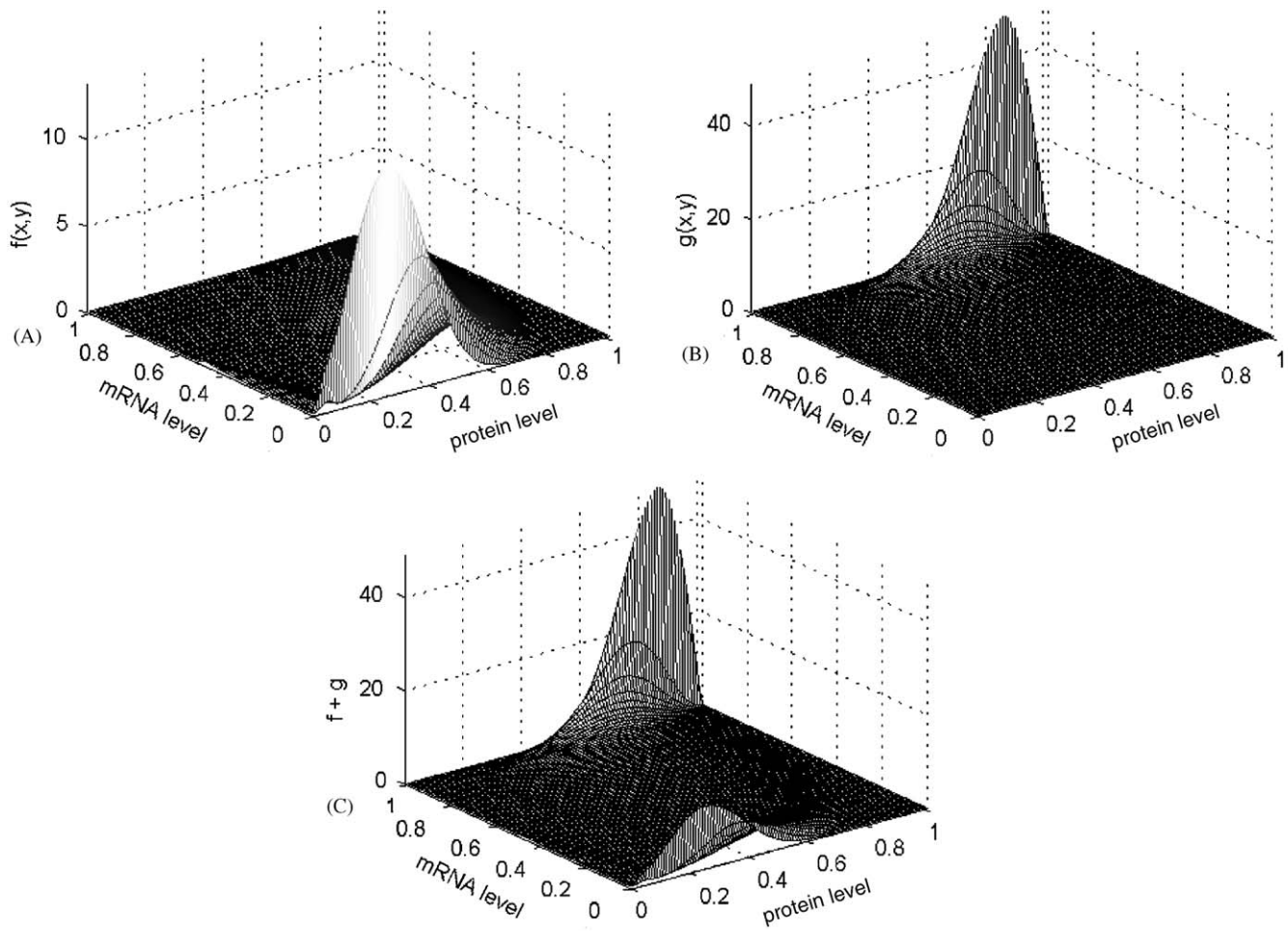


Fig. 8. Stationary distributions $f(x,y)$ (Panel A), $g(x,y)$ (Panel B) and $\rho(x,y,t) = f(x,y) + g(x,y)$ (Panel C) calculated on grid $100 \times 100$. The parameters are $c = 0.5$, $b = 0.5$, $r = 0.2$.

For the stationary solutions $f(y)$, $g(y)$ we have

$$-\frac{\mathrm{d}}{\mathrm{d}y}(fy) = b_r y g - c_r f, \tag{32}$$

$$\frac{\mathrm{d}}{\mathrm{d}y}((1-y)g) = -b_r y g + c_r f, \tag{33}$$

where $c_r = c/r$ and $b_r = b/r$. The above system can be solved analytically for $y \in (0,1)$. Adding Eqs. (32) and (33), we obtain the first integral

$$\frac{\mathrm{d}}{\mathrm{d}y}[-yf + (1-y)g] = 0. \tag{34}$$

This implies

$$-yf + (1-y)g = -f(1) = g(0). \tag{35}$$

Since we require that $f(\cdot)$ and $g(\cdot)$ are nonnegative, the condition $-f(1) = g(0)$ implies $f(1) = g(0) = 0$. Therefore from Eq. (35) we have $g = yf/(1-y)$. Inserting this into Eq. (32) we obtain

$$b_r \frac{y^2}{(1-y)} f - c_r f = -\frac{\mathrm{d}}{\mathrm{d}y}(fy), \tag{36}$$

which implies

$$f(y) = A\mathrm{e}^{b_r y} y^{c_r - 1}(1-y)^{b_r}, \tag{37}$$

and further

$$g(y) = A\mathrm{e}^{b_r y} y^{c_r}(1-y)^{b_r - 1}, \tag{38}$$

where $A = [\int_0^1 \mathrm{e}^{b_r y} y^{c_r - 1}(1-y)^{b_r - 1}\,\mathrm{d}y]^{-1}$, since for $c_r > 0$, $b_r > 0$ both $f(y)$ and $g(y)$ are integrable on $(0,1)$. For $c_r < 1$, $\lim_{y\to 0} f(y) = \infty$ and for $b_r < 1$, $\lim_{y\to 1} g(y) = \infty$, while for $c_r > 1, b_r > 1$ we have $f(0) = g(0) = f(1) = g(1) = 0$. The marginal distribution $\rho(y) := f(y) + g(y)$, has the form of

$$\rho(y) = A\mathrm{e}^{b_r y} y^{c_r - 1}(1-y)^{b_r - 1}. \tag{39}$$

It describes the protein level with no regard to gene status, and may be more adequate to compare the theory with experimental data based on flow cytometry.

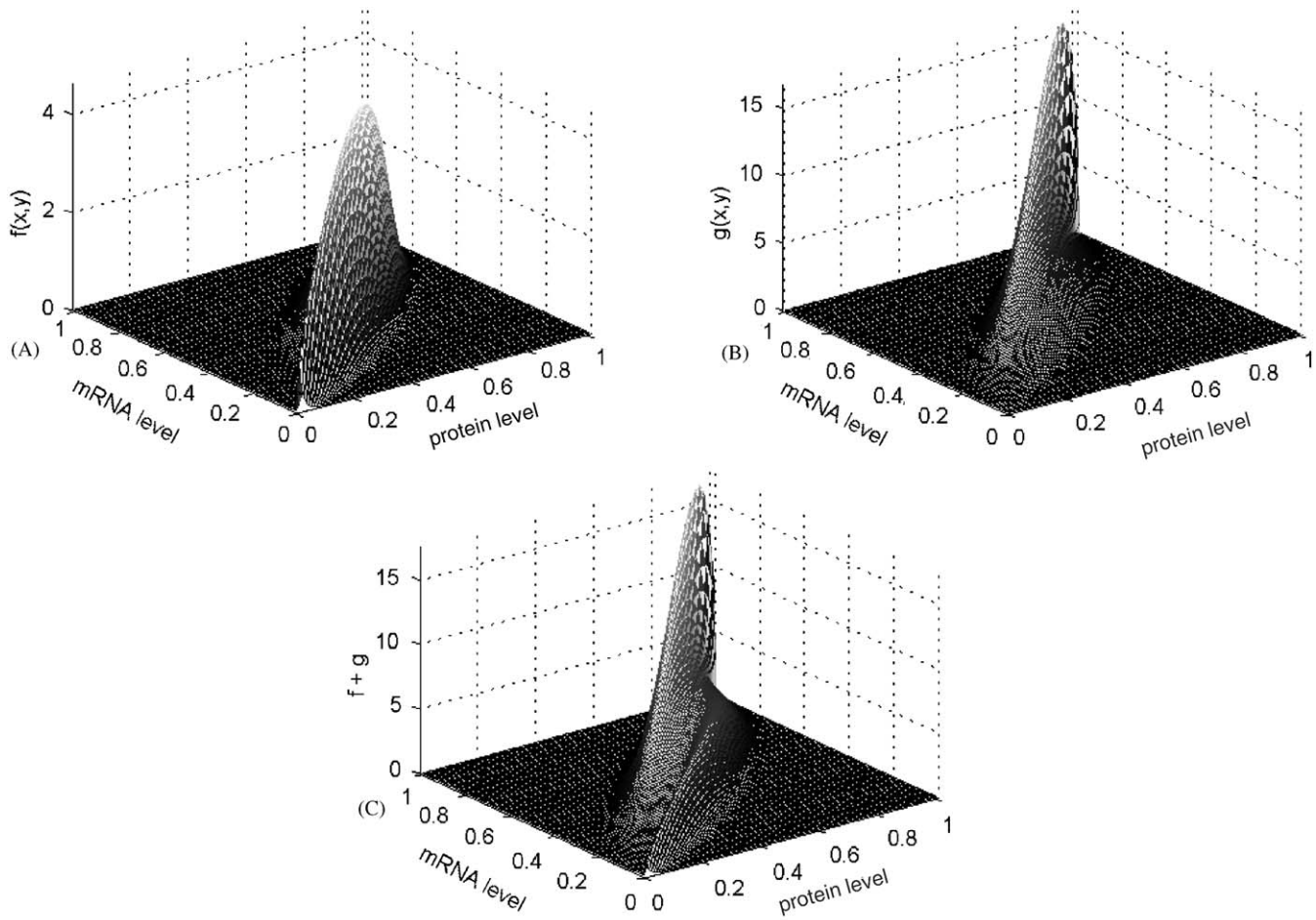


Fig. 9. Stationary distributions $f(x,y)$ (Panel A), $g(x,y)$ (Panel B) and $\rho(x,y,t) = f(x,y) + g(x,y)$ (Panel C) calculated on grid $100 \times 100$. The parameters are $c = 2$, $b = 2$, $r = 3$.

For $c_r < 1$ and $b_r < 1$, which corresponds to a low frequency of gene status switching, function $\rho(y)$ has a minimum between 0 and 1, whereas $\lim_{y\to 0}\rho(y) = \infty$ and $\lim_{y\to 1}\rho(y) = \infty$. For $c_r > 1$ and $b_r > 1$, $\rho(y)$ has one maximum. The larger $c_r$ and $b_r$ are, the more concentrated is the distribution $\rho(y)$. This is not surprising, since large $c_r$ and $b_r$ imply that gene status is frequently changed, and cell to cell fluctuations in the mRNA level are small. To verify the Kepler–Elston approximation, distributions $f(y)$, $g(y)$ and $\rho(y)$ given by Eqs. (37)–(39) should be compared with the marginal distributions $\int f(x,y)\mathrm{d}x$, $\int g(x,y)\,\mathrm{d}x$, $\int \rho(x,y)\,\mathrm{d}x$ calculated numerically from the two-dimensional distribution. As one may expect, the Kepler–Elston approximation is accurate for small $r$ (Fig. 11A–F, where $r = 0.2$), especially for large $c$ and $d$. However, it is unacceptable for $r > 1$: In Fig. 11I we show that for $c = 2$, $b = 2$ and $r = 3$, the stationary distribution $\rho(y)$ calculated from Eq. (39) has two maxima, while the marginal distribution $\int \rho(x,y)\,\mathrm{d}x$ has one. This is due to the fact that $c > 1$, $b > 1$ but $c_r < 1$, $b_r < 1$. In this latter case, assumption that Eq. (8) is fast, i.e., $y = x$ and thus $\rho(y) = A\mathrm{e}^{by}y^{c-1}(1-y)^{b-1}$, is more appropriate. The case $r \ll 1$, where the Kepler–Elston approximation is accurate, corresponds to the situation, where the protein is much more stable than the mRNA. Typically, this is the case, however in some situations the protein is actively degraded and its half-life can be shorter than that of the mRNA. We encountered this situation analysing the NF-$\kappa$B regulatory module (Lipniacki et al., 2004); the NF-$\kappa$B inhibitor I$\kappa$B$\alpha$ is catalytically degraded with a half-life of about 10 min, while its mRNA has a half-life on the order of 30 min. In this case, the amount of I$\kappa$B$\alpha$ protein is not proportional to the amount of mRNA. To the contrary, we found that the higher level of the I$\kappa$B$\alpha$ protein produces stronger inhibition of NF-$\kappa$B, and thus a lower level of I$\kappa$B$\alpha$ mRNA which is under NF-$\kappa$B control (Lipniacki et al., 2004, 2005). This implies that the two-dimensional I$\kappa$B$\alpha$ mRNA-protein distribution is important for understanding the NF-$\kappa$B regulatory module. Another important example of regulation by rapid proteolysis

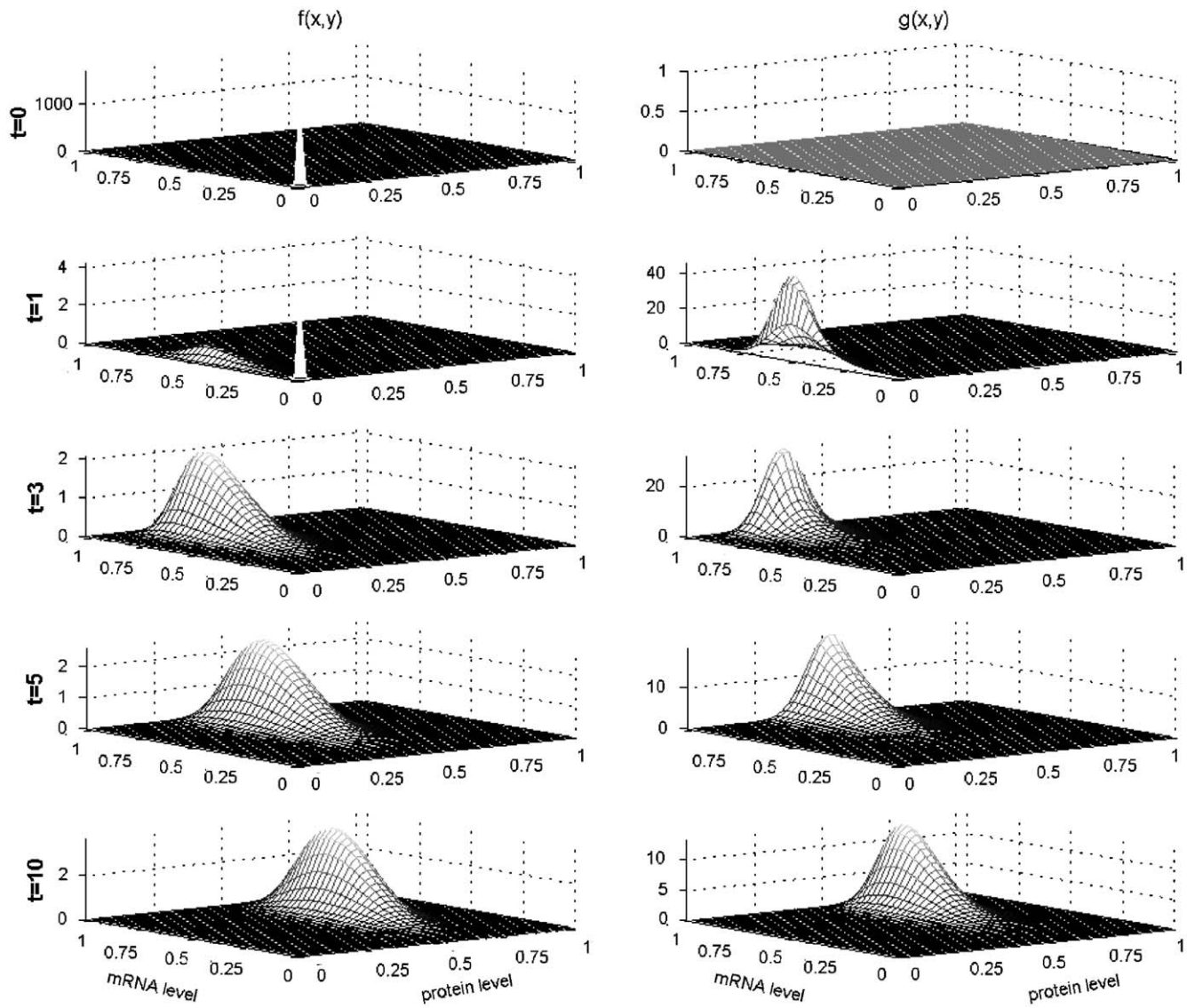


Fig. 10. Time evolution of distributions $f(x, y, t)$-left, and $g(x, y, t)$-right calculated on grid $40 \times 40$. The parameters are $c = 6$, $b = 3$, $r = 0.2$, as in Figs. 6 and 7. The initial condition is $f(x, y, 0) =$ Dirac function, $g(x, y, 0) = 0$.

we want to mention is cell cycle regulation in *Caulobacter* (McAdams and Shapiro, 2003). In a narrow window of the cell cycle, just prior to cell division, most of the master regulatory protein called CtrA is degraded in the stalked compartment, while in the swarmer compartment, its level remains unchanged. This time and space-specific protein degradation is crucial for the fates of the two daughter cell fates since CtrA controls 26% of the *Caulobacter* cell-cycle-regulated genes.

Nevertheless, since the Kepler–Elston approximation is, for $r \ll 1$, well justified and quite accurate, and constitutes a great simplification to the analysis, we use it to analyze the system of two interacting genes and calculate two-dimensional protein–protein distributions. Without this assumption, the analysis of two-gene system would require calculation of four-dimensional distributions.

### 2.6. The systems of two interacting genes

Let us consider the system of two genes. We use the Kepler–Elston approximation, assuming a direct translation from DNA into protein. Let $x$ and $y$ now denote the amounts of protein related to the first and second genes, respectively. The system has the form

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -x + G_x, \tag{40}$$

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = -r\,y - G_y, \tag{41}$$

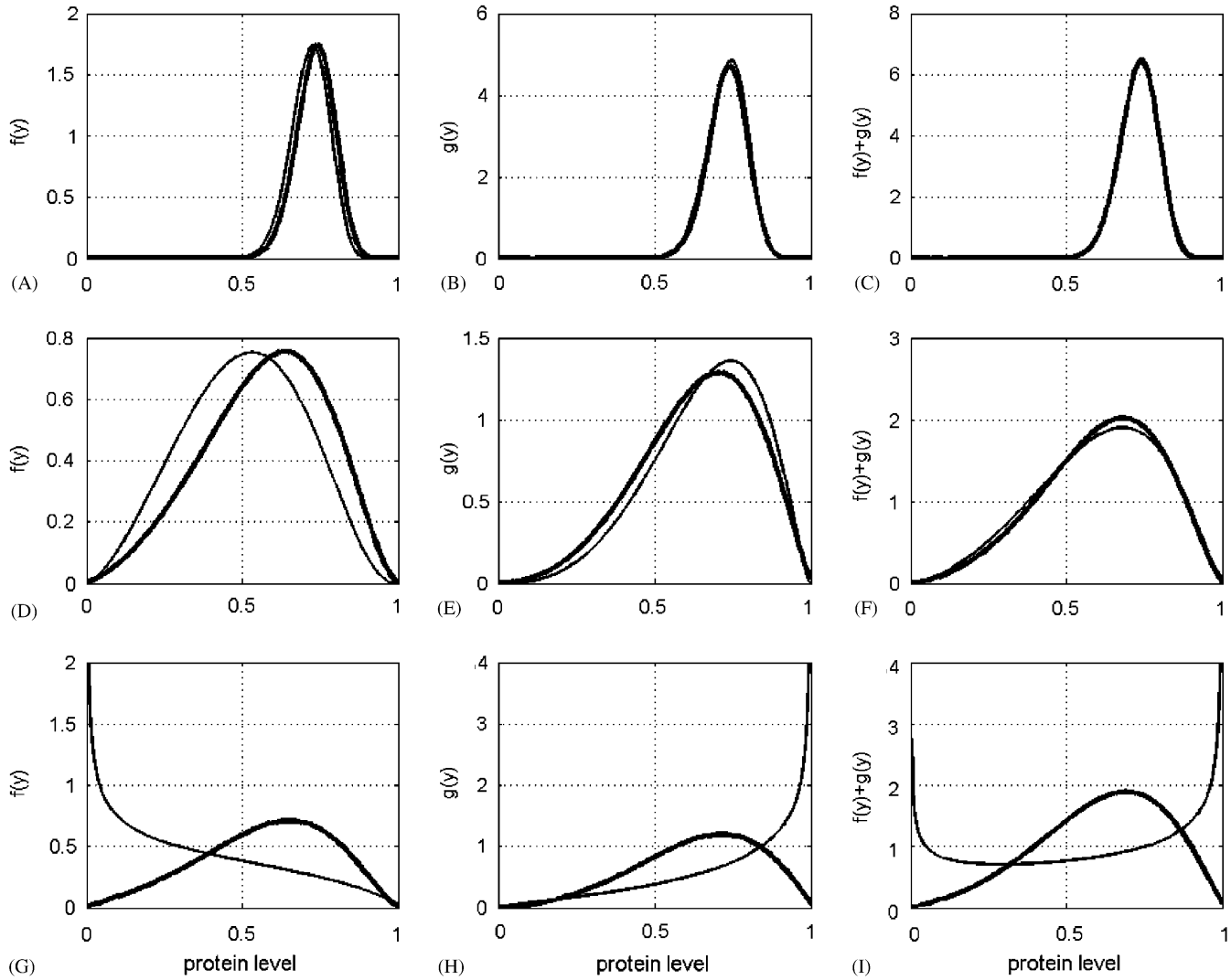


Fig. 11. Marginal distributions $f(y)$, $g(y)$, $\rho(y)$ calculated based on Eq. (39) — and numerically from two-dimensional distributions presented in Figs. 6, 8 and 9 —. Panels A, B and C correspond to $c = 6$, $b = 3$, $r = 0.2$; Panels D, E and F correspond to $c = 0.5$, $b = 0.5$, $r = 0.2$; Panels G, H and I correspond to $c = 2, b = 2, r = 3$.

where $G_x$, $G_y$ are the transcription switches of the first and second genes, respectively. We denote

$$f_{ij}(x,y,t)\Delta x\Delta y = \Pr[x(t) \in (x, x+\Delta x),\ y(t) \in (y, y+\Delta y),\ G_x = i,\ G_y = j], \quad i,j = 0,1. \tag{42}$$

### 2.6.1. Activator–repressor system

In the case of an activator $x$–repressor $y$ system, we assume the following transition rules:

$$I_x \xrightarrow{p} A_x, \quad I_x \xleftarrow{qy(t)} A_x, \quad G_x(I_x) = 0, \quad G_x(A_x) = 1, \tag{43}$$

$$I_y \xrightarrow{kx(t)} A_y, \quad I_y \xleftarrow{h} A_y, \quad G_y(I_y) = 0, \quad G_y(A_y) = 1. \tag{44}$$

The above relations imply that inactivation of the activator $x$ is proportional to the amount of the repressor $y$, and that the activation of the repressor is proportional to the amount of activator. Using Eqs. (43) and (44), we transform Eqs. (40) and (41) into a system of 4 PDEs for $f_{ij}(x,y,t)$:

$$\frac{\partial f_{00}}{\partial t} + \operatorname{div}\left[f_{00}\left(\frac{\mathrm{d}x}{\mathrm{d}t}_{|G_x=0}, \frac{\mathrm{d}y}{\mathrm{d}t}_{|G_y=0}\right)\right] = -(p+kx)f_{00} + f_{01}h + f_{10}qy, \tag{45}$$

$$\frac{\partial f_{10}}{\partial t} + \operatorname{div}\left[f_{10}\left(\frac{\mathrm{d}x}{\mathrm{d}t}_{|G_x=1}, \frac{\mathrm{d}y}{\mathrm{d}t}_{|G_y=0}\right)\right] = -(qy+kx)f_{10} + pf_{00} + hf_{11}, \tag{46}$$

$$\frac{\partial f_{01}}{\partial t} + \operatorname{div}\left[f_{01}\left(\frac{\mathrm{d}x}{\mathrm{d}t}_{|G_x=0}, \frac{\mathrm{d}y}{\mathrm{d}t}_{|G_y=1}\right)\right] = -(p+h)f_{01} + f_{00}kx + f_{11}qy, \tag{47}$$

$$\frac{\partial f_{11}}{\partial t} + \operatorname{div}\left[f_{11}\left(\frac{\mathrm{d}x}{\mathrm{d}t}_{|G_x=1}, \frac{\mathrm{d}y}{\mathrm{d}t}_{|G_y=1}\right)\right] = -(qy+h)f_{11} + kxf_{10} + pf_{01}. \tag{48}$$

In the steady state this assumes the form:

$$\frac{\partial}{\partial x}(-xf_{00}) + \frac{\partial}{\partial y}(-ryf_{00}) = -(p+kx)f_{00} + f_{01}h + f_{10}qy, \tag{49}$$

$$\frac{\partial}{\partial x}[(1-x)f_{10}] + \frac{\partial}{\partial y}(-ryf_{10}) = -(qy+kx)f_{10} + pf_{00} + hf_{11}, \tag{50}$$

$$\frac{\partial}{\partial x}(-xf_{01}) + \frac{\partial}{\partial y}[(1-ry)f_{01}] = -(p+h)f_{01} + f_{00}kx + f_{11}qy, \tag{51}$$

$$\frac{\partial}{\partial x}[(1-x)f_{11}] + \frac{\partial}{\partial y}[(1-ry)f_{11}] = -(qy+h)f_{11} + kxf_{10} + pf_{01}. \tag{52}$$

In Figs. 12 and 13 we show the solutions of the activator–repressor system for two sets of parameters.

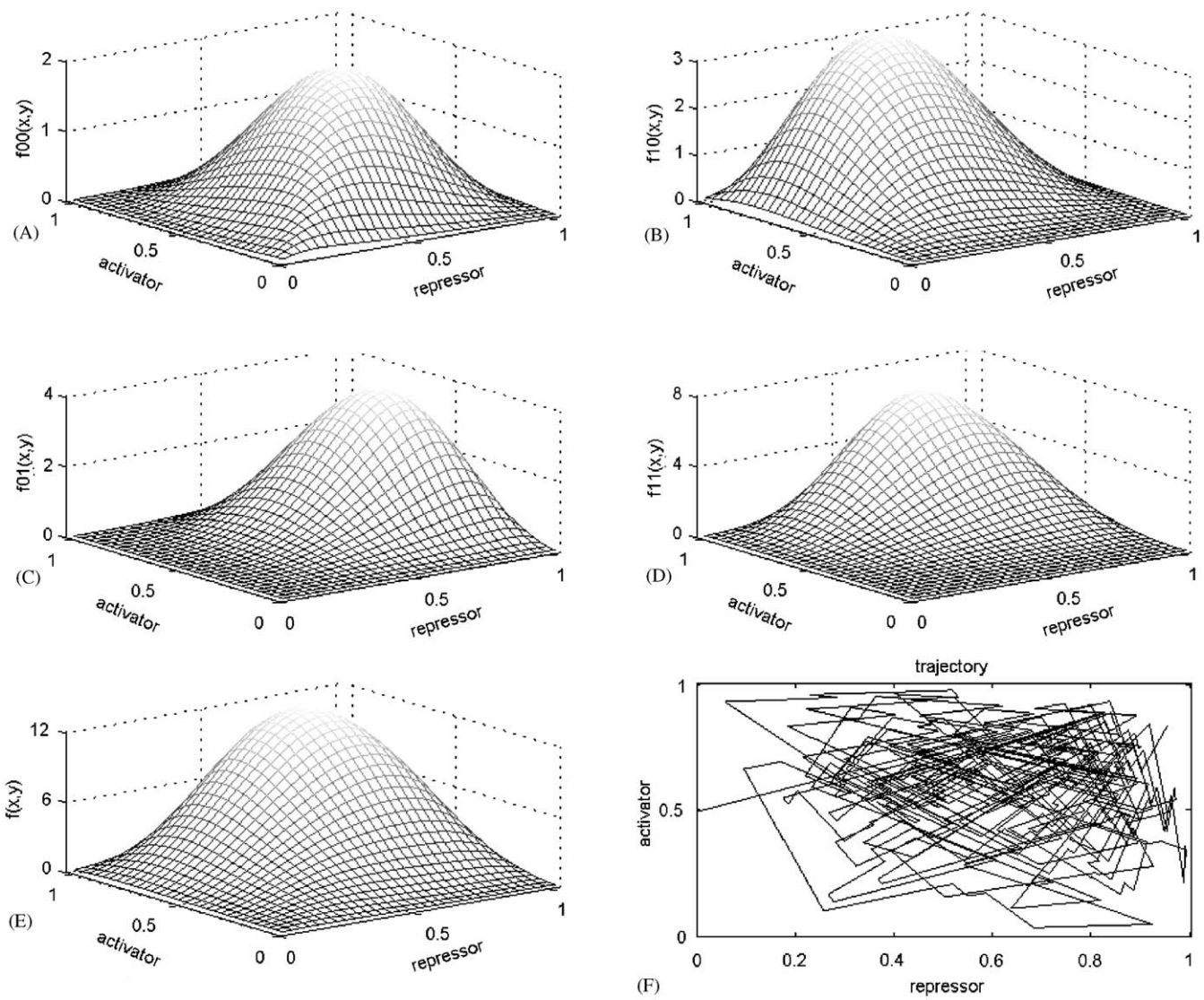


Fig. 12. Protein–protein distributions for the activator–repressor system for $p=3$, $q=3$, $k=5$, $h=2$, $r=1$. Panels A, B, C, D and E show functions $f_{0,0}, f_{1,0}, f_{0,1}, f_{1,1}$ and the marginal distribution $f = f_{0,0} + f_{1,0} + f_{0,1} + f_{1,1}$. In Panel F we show the example single cell trajectory for the same set of parameters.

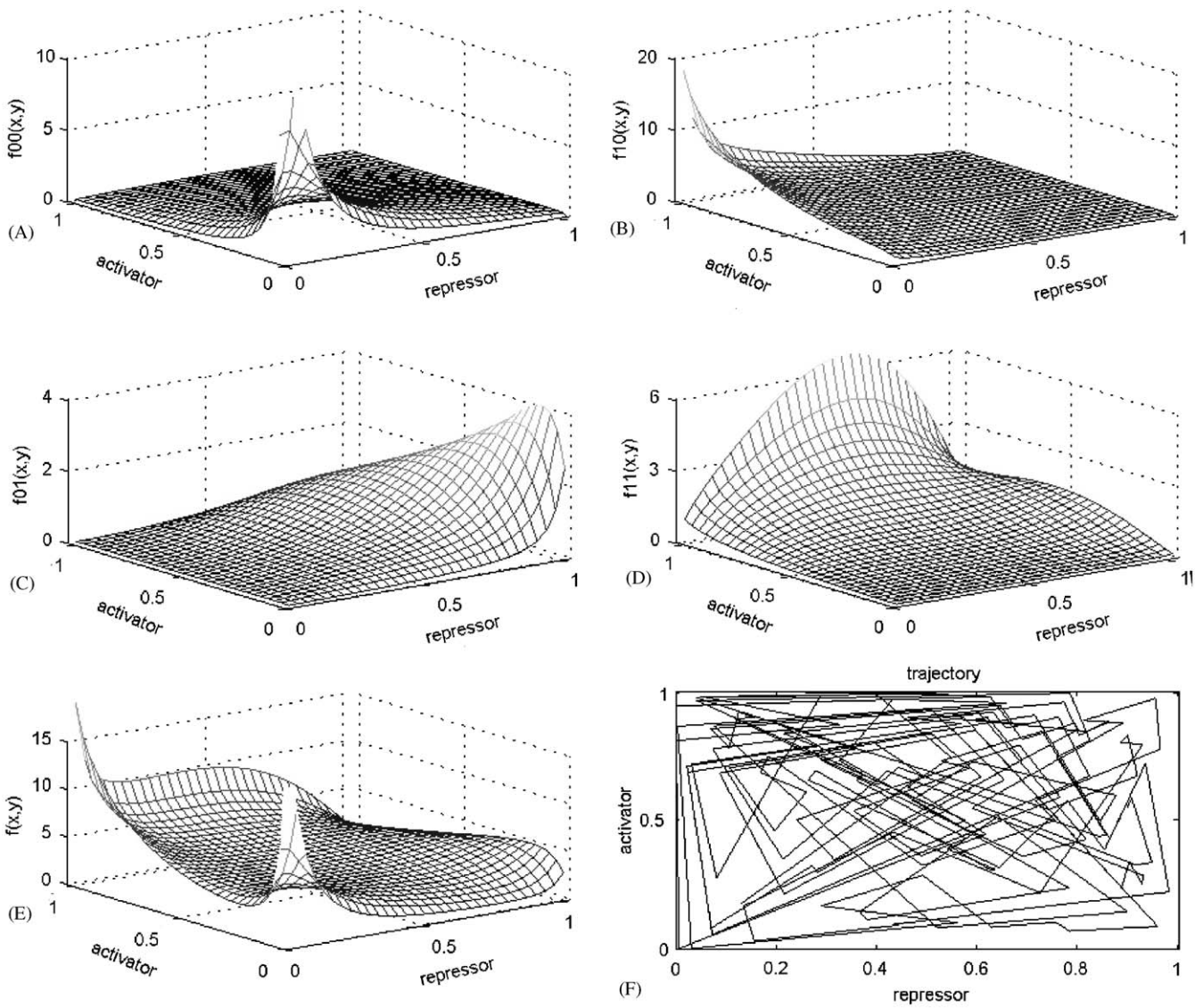


Fig. 13. Protein–protein distributions for the activator-repressor system for $p = 1$, $q = 2$, $k = 1.5$, $h = 1$, $r = 1$. Panels A, B, C, D and E show functions $f_{0,0}, f_{1,0}, f_{0,1}, f_{1,1}$ and the marginal distribution $f = f_{0,0} + f_{1,0} + f_{0,1} + f_{1,1}$. In Panel F we show the example single cell trajectory for the same set of parameters.

Transition parameters $p, q, k$ and $h$ used to obtain the distribution shown in Fig. 12 are relatively large and thus the resulting protein–protein distribution $f = f_{00} + f_{10} + f_{01} + f_{11}$ has one maximum. For smaller values of transition parameters, as shown in Fig. 13, partial distributions $f_{i,j}$ are much different and, as a result, their sum $f = f_{00} + f_{10} + f_{01} + f_{11}$ has three maxima located close or at the points (0,1), (0,0) and (1,0), and a quite complicated profile.

### 2.6.2. Repressor–repressor system

A simplified system of two repressors has been analysed by Kepler and Elston (2001). To reduce the number of possible states these authors assume that the two genes coding for $x$ and $y$ share the same operator, thus the state $f_{11}$ is excluded. Moreover, they assume the same parameter values describing the kinetics of the two genes and their products. With these simplifications they calculate the histogram for the marginal distribution $f(x) = f(y)$ using the Monte Carlo method.

As in the previous case, our approach allows us to calculate two-dimensional (protein–protein) distributions. We assume the following transition rules:

$$I_x \xrightarrow{p} A_x, \quad I_x \xleftarrow{qy(t)} A_x, \quad G_x(I_x) = 0, \quad G_x(A_x) = 1, \tag{53}$$

$$I_y \xrightarrow{k} A_y, \quad I_y \xleftarrow{hx(t)} A_y, \quad G_y(I_y) = 0, \quad G_y(A_y) = 1. \tag{54}$$

The resulting system of PDE's describing the stationary distribution is analogous to system (49)–(52) and has the

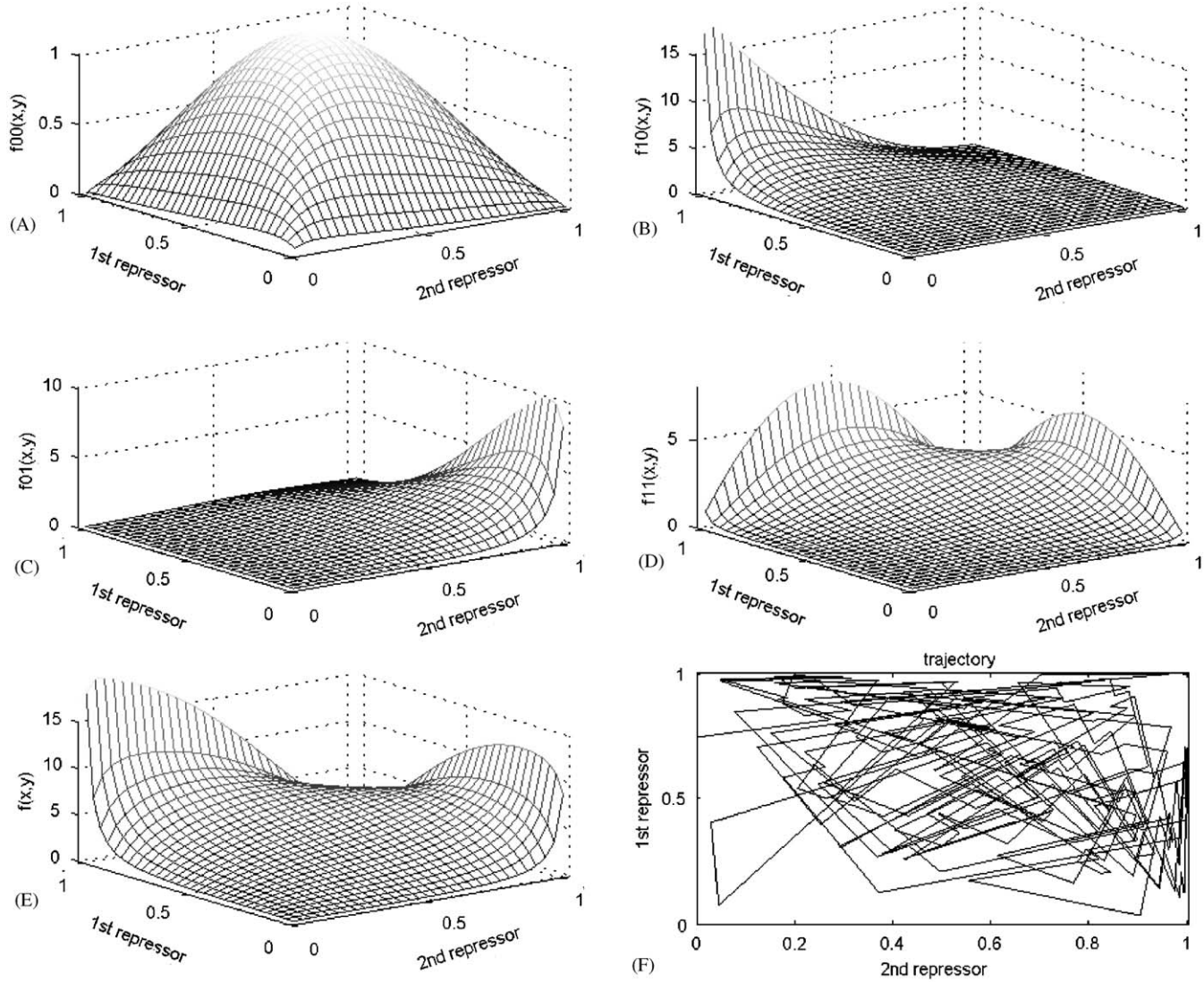


Fig. 14. Protein–protein distributions for the repressor–repressor system for $p = 1.7$, $q = 2$, $k = 1.5$, $h = 2$, $r = 2$. Panels A, B, C, D and E show functions $f_{0,0}, f_{1,0}, f_{0,1}, f_{1,1}$ and the marginal distribution $f = f_{0,0} + f_{1,0} + f_{0,1} + f_{1,1}$. In Panel F we show the example single cell trajectory for the same set of parameters.

form of

$$\frac{\partial}{\partial x}(-xf_{00}) + \frac{\partial}{\partial y}(-ryf_{00}) = -(p+k)f_{00} + f_{01}h(x) + f_{10}qy, \tag{55}$$

$$\frac{\partial}{\partial x}[(1-x)f_{10}] + \frac{\partial}{\partial y}(-ryf_{10}) = -(qy+k)f_{10} + pf_{00} + h(x)f_{11}, \tag{56}$$

$$\frac{\partial}{\partial x}(-xf_{01}) + \frac{\partial}{\partial y}[(1-ry)f_{01}] = -[p+h(x)]f_{01} + f_{00}k + f_{11}qy, \tag{57}$$

$$\frac{\partial}{\partial x}[(1-x)f_{11}] + \frac{\partial}{\partial y}[(1-ry)f_{11}] = -[qy+h(x)]f_{11} + kf_{10} + pf_{01}. \tag{58}$$

Figs. 14 and 15 depict the solutions of the repressor–repressor system for two sets of parameters. In Fig. 14, we observe that the repressor–repressor system is unstable in the sense that relatively small differences in the activation constants of two repressors ($p = 1.7, k = 1.5$) lead to a substantial asymmetry in the resulting protein–protein distribution. In Fig. 15 the values of activation and inactivation coefficients are the same, and the asymmetry between the two repressors results from the assumption that the first repressor has half the degradation rate of the second one, and thus its level is higher.

## 3. Discussion and conclusions

The intrinsic stochasticity in gene expression may result from small number of mRNA and protein

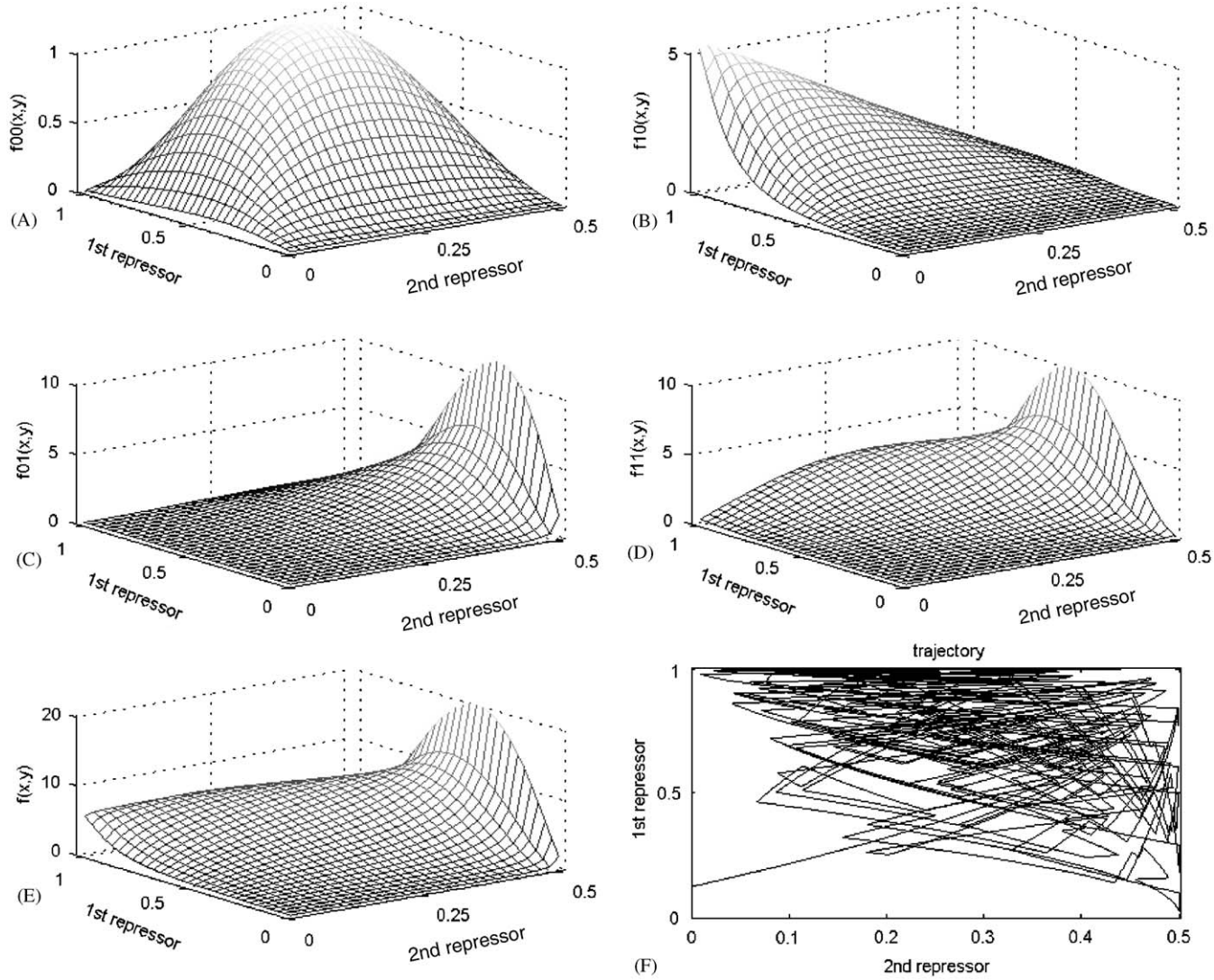


Fig. 15. Protein–protein distributions for the repressor–repressor system for $p = 3$, $q = 3$, $k = 3$, $h = 3$, $r = 2$. Panels A, B, C, D and E show functions $f_{0,0}, f_{1,0}, f_{0,1}, f_{1,1}$ and the marginal distribution $f = f_{0,0} + f_{1,0} + f_{0,1} + f_{1,1}$. In Panel F we show the example single cell trajectory for the same set of parameters.

molecules, and from intermittent gene activity. It is expected that the first source is the most important in prokaryotes, in which the number of mRNA and even protein molecules per cell is very small. In eukaryotes, and especially in higher eukaryotes, where the number of mRNAs is fairly large, the main source of stochasticity is intermittent gene activity. Typically, to activate the eukaryotic gene, several transcription factors are needed together with chromatin remodeling, and therefore longer periods of gene inactivity and activity resulting in large bursts of mRNA molecules are expected.

Depending on the biology of the phenomena, various methods for the intrinsic noise analysis in genetic regulatory networks have been proposed. The approach of McAdams and Arkin (1997), Arkin et al. (1998) and Gilman and Arkin (2002) was designed to explore the effects of small number of mRNA and protein molecules in bacteria. McAdams and Arkin (1997) follow the assumption made by Ackers et al. (1982), and Shea and Ackers (1985) that there is a rapid equilibrium between regulatory proteins and corresponding gene promoters. The same assumption was made by Arkin et al. (1998) who applied the stochastic formulation of chemical kinetics proposed by Gillespie (1977) to analyse the phage $\lambda$ lysis-lysogeny decision circuit in *Esherichia coli*. Recently, Tao (2004a) analysed a single auto-regulatory gene, considering both negative and positive feedback. He employs the Chapman–Kolmogorov equation for the probability distribution $P(x, y, t)$, where $x$ and $y$ denote numbers of mRNA and protein molecules, respectively, to calculate the first two moments of$P(x, y, t)$. Assuming that the transcription rate depends on the amount of protein, he also neglects

stochasticity caused by switching of the gene status. When analysing a two-gene network Tao (2004b) disregards mRNA, assuming that the rate of protein synthesis does not depend explicitly on the amount of the corresponding mRNA. The noise in one- and two-gene regulatory networks also has been analysed by Tomioka et al. (2004). These authors assumed that the regulatory network is close to a deterministic stable equilibrium state and applied linear noise approximation of a chemical master equation in order to evaluate the system fluctuations around this state.

Stochasticity due to switching of the gene status was first recognized by Ko (1991) and then was analyzed by Kepler and Elston (2001). Their approach involves the Chapman–Kolmogorov equation for probability distribution defined on discrete states, which is then approximated by the Fokker–Planck equation. The authors consider synthesis of protein oligomers, but assume direct translation of proteins from the DNA. In the case of a single self-promotory gene, the Fokker–Planck equation is further simplified by neglecting the diffusion term, which leads to the first-order system of PDEs analogous to our system (32)–(33). Kepler and Elston (2001) also consider a system of two mutual repressors, and assuming that they are identical they compute the marginal distribution of the protein using Monte Carlo simulations. The main difference between the Kepler–Elston (2001) approach and the approach we propose is the manner in which we pass from the discrete to a continuous description. We do this at the level of a single cell description, approximating the Gillespie scheme by a system of stochastic ODEs, while Kepler and Elston pass from the Chapman–Kolmogorov equation to Fokker–Planck equation. The advantage of our approach lies in the possibility to validate it by single cell simulations, which generally is much simpler than comparing a solution of Chapman–Kolmogorov equation and its Fokker–Planck approximation. In their recent work, Pirone and Elston (2004) use the Fokker–Planck equation to calculate all first and second moments of the probability density function. In the latter paper more attention is focused on oligomerization reactions leading to the formation of dimers and tetramers.

Transcriptional regulation involving switching between discrete high and low transcriptional rates was also considered in a frequency domain by Simpson et al. (2004). Their approach provides the frequency distribution of noise associated both with mRNA synthesis/degradation and noise resulting from the operator binding events that cause bursts of transcription.

Following Ko (1991) and others, here we focus on stochasticity in eukaryotic gene expression, which is introduced at the level of transcriptional regulation. The approach combines the stochastic switch description of kinetics of reactants present in a small number of copies (in this case gene copies) with ODE description for processes involving larger number of reacting molecules (i.e., mRNAs and proteins). The model we explore is based on the assumption that the gene promoters, in the time-scale on the order of the mRNA half-life, are not in a statistical equilibrium. This assumption is supported by a growing number of experiments on single cell gene expression, showing cell-to-cell heterogeneity in mRNA levels, fluctuations of which are too large to be explained only by effects of the finite number of mRNAs (Takasuka et al., 1998; Stirland et al., 2003). The experiments also show strong time-dependent fluctuations of single cell mRNA levels, in response to a steady stimulation. Moreover, these fluctuations have a tendency towards desynchronization (Takasuka et al., 1998; Stirland el al., 2003; Elowitz et al., 2002).

Low frequency of gene switching causes activation of a single gene, leading to the production of a flux in mRNA transcript abundance, on the order of the total mRNA for the gene being considered. Since the relative strength of stochastic effects grows as the number of reacting molecules decreases, we expect that the stochasticity due to switching of the gene status is the most important, at least for eukaryotes. Accordingly, we neglect the mRNA/protein production/decay randomness. For simplicity, we assume that gene activation or repression is due to a single molecule; however in many cases the gene is turned on (or off) due to a collective action of several different regulatory factors. This problem has been analysed recently by Paszek et al. (2005).

The resulting stochastic ODEs for mRNA and protein in a single cell yield a system of linear first-order PDEs for pdf's. Using the discretization, we reduced the problem of finding time-dependent pdf's to solving a large system of linear differential equations. Stationary pdf's are solutions of a system of linear algebraic equations. This approach allows us to calculate the two-dimensional (mRNA and protein) time-dependent and stationary pdf. Having the two dimensional pdf, we calculate numerically the marginal distribution for protein in order to compare it with protein distribution in the Kepler–Elston model (2001), in which a direct translation of the protein is assumed without mediation of mRNA. We found the Kepler–Elston approximation to be satisfactory in the case when protein half-life is much longer than mRNA half-life. However, the approximation fails when the protein is degraded faster than mRNA, what may happen in the case of active protein degradation (e.g. Lipniacki et al., 2004; McAdams and Shapiro, 2003). In this case, the Kepler–Elston approximation may produce artificially bimodal protein distributions. Based on the introduced approach, and using the Kepler–Elston approximation, we considered the following two-gene systems: activator–repressor and repressor–repressor, and have calculated their two-dimensional protein–protein probability distribution functions.

In summary, we derive the first-order PDEs for probability distribution function from stochastic ODEs describing approximate kinetics of a single cell. Resulting equations enable us to calculate, for the first time, the two-dimensional distributions, i.e., the mRNA-protein distribution in the case of single gene regulation and the protein–protein distribution in the case of two-gene regulatory systems.

**Acknowledgments**

We 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. Derivation of PDEs for the pdf's—general case

We consider the following general set-up:

A system of $K$ nonlinear autonomous ODEs of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \varphi(x,\gamma),\ \ t \geqslant 0, \tag{59}$$

$$x(0) = x_0, \tag{60}$$

where $x = x(t)$, $x_0$ and $\varphi$ are column $K$-vectors with components $x_k$, $x_{0k}$, and $\varphi_k$, respectively. $\gamma = \gamma(t)$ is a right-continuous function assuming values from a finite set equivalent to $\{1,2,\ldots,M\}$. We assume that functions $\varphi$ are continuous and bounded

$$\varphi : D \times \{1,2,\ldots,M\} = [0,a_1] \times [0,a_2] \times \cdots \times [0,a_K] \times \{1,2,\ldots,M\} \to \mathbb{R}^K \tag{61}$$

and such that $D$ is invariant for the system, i.e., for initial conditions $x_0 \in D$ and any $\gamma$, $x(t) \in D$. A solution of the ODE system (59)–(60), extended from value $x_s$ at time $s$, is frequently denoted

$$x(t) = x(t; x_s, s), \tag{62}$$

which for given $\gamma(t)$ defines a mapping $X(\cdot; t, s) = x(t; \cdot, s) : \mathbb{R}^K \to \mathbb{R}^K$, a translation from $s$ to $t$ along the solution of the ODE (59).

Furthermore, we let $\gamma$ be a random function, following the rules of a continuous-time non-autonomous finite Markov chain. Specifically, if at time $t \geqslant 0$, $\gamma(t) = m$, then the probability that at time $t + \Delta t$ it will be in a different state $l$, is equal to

$$\Pr[\gamma(t+\Delta t) = l \,|\, \gamma(t) = m] = q_{lm}(x(t))\Delta t + o(\Delta t), \tag{63}$$

where $q_{lm}(x)$ are bounded continuous on $D$, and $\lim_{\Delta t \to 0} o(\Delta t)/\Delta t = 0$. In other words, the transition intensities of process $\gamma$ are functions of the state $x$ of the ODE system (59)–(60).

We will derive expressions for the joint distributions of the random variables $x(t)$ and $\gamma(t)$, at a given time $t$. Let us denote by $f_\gamma(x,t)$, the joint function of probability density (in $x(t)$) and probability (in $\gamma(t)$), so that

$$\begin{aligned} P_l(\xi, \Delta\xi, t) &= \Pr[x_k(t) \in (\xi_k, \xi_k + \Delta\xi_k), k = 1,\ldots,K; \\ &\qquad \gamma(t) = l] \\ &= f_l(\xi, t)\Delta\xi_1 \cdots \Delta\xi_K + o(\Delta\xi). \end{aligned} \tag{64}$$

Furthermore, let us assume $f_\gamma(x,0)$ is given. For $t \geqslant 0$, let us consider $P_l(\xi, \Delta\xi, t)$ and $P_l(\xi, \Delta\xi, t, \Delta t)$, this latter being the probability that $x(t+\Delta t)$ falls into the region containing solutions of system (59)–(60), which at time $t$ were enclosed in the rectangle $(\xi, \xi + \Delta\xi)$, and that $\gamma(t+\Delta t) = l$,

$$\begin{aligned} P_l(\xi, \Delta\xi, t, \Delta t) &= \Pr[\{x(t+\Delta t; x_t, t), x_t \in (\xi, \xi+\Delta\xi)\}, \\ &\qquad \gamma(t+\Delta t) = l]. \end{aligned} \tag{65}$$

If there is no jump of $\gamma(t)$ in the interval $(t, t+\Delta t)$ then $P_l(\xi, \Delta\xi, t, \Delta t) = P_l(\xi, \Delta\xi, t)$. However, in general, transitions to $l$ may occur from all other states $m \in \{1,2,\ldots,M\}$, $m \neq l$, in the interval $(t, t+\Delta t)$, so that

$$P_l(\xi, \Delta\xi, t, \Delta t) = P_l(\xi, \Delta\xi, t)\left[1 - \sum_{m \neq l} q_{lm}(x(t))\Delta t - o(\Delta t)\right] \tag{66}$$

$$+ \sum_{m \neq l} P_m(\xi, \Delta\xi, t)[q_{ml}(x(t))\Delta t + o(\Delta t)] \tag{67}$$

and consequently

$$\begin{aligned} \lim_{\Delta t \downarrow 0} \frac{P_l(\xi, \Delta\xi, t, \Delta t) - P_l(\xi, \Delta\xi, t)}{\Delta t} &= P_l(\xi, \Delta\xi, t) q_{ll}(x(t)) \\ &\quad + \sum_{m \neq l} P_m(\xi, \Delta\xi, t) q_{ml}(x(t)), \end{aligned} \tag{68}$$

where $q_{ll}(t) = -\sum_{m \neq l} q_{lm}(t)$. Let us notice that since $P_l(\xi, \Delta\xi, t, 0) = P_l(\xi, \Delta\xi, t)$, the limit at the left-hand side of (67) is the derivative $\partial P_l(\xi, \Delta\xi, t, \Delta t)/\partial \Delta t$ at $\Delta t = 0$. The following expression involving a change of variables:

$$\begin{aligned} P_l(\xi, \Delta\xi, t, \Delta t) &= \int \cdots \int_{u \in (\xi, \xi+\Delta\xi)} \left| \det\left[\frac{\partial X(u; t, t+\Delta t)}{\partial u}\right] \right| \\ &\quad \times f_l[X(u; t, t+\Delta t), t+\Delta t]\,\mathrm{d}u,\ \ l = 1,\ldots,M \end{aligned} \tag{69}$$

is employed, since it would be difficult to express the region being the image of the rectangle $(\xi, \xi+\Delta\xi)$ through mapping $X(\cdot; t, t+\Delta t)$. Let us also note that

$$P_l(\xi, \Delta\xi, t) = \int \cdots \int_{u \in (\xi, \xi+\Delta\xi)} f_l(u,t)\,\mathrm{d}u, \quad l = 1,\ldots,M. \tag{70}$$

As it is known, the sensitivity matrix $W(u;s,t) = \partial X(u;s,t)/\partial u$ is the solution of a matrix ODE

$$\frac{\mathrm{d}}{\mathrm{d}t} W(u;s,t) = \frac{\partial \varphi(u,\gamma)}{\partial u} W(u;s,t), \quad W(u;s,s) = I, \tag{71}$$

and therefore, for small $\Delta t$,

$$W(u;t,t+\Delta t) = I + \frac{\partial \varphi(u,\gamma)}{\partial u}\Delta t + o(\Delta t),$$
$$\frac{\mathrm{d}}{\mathrm{d}t} W(u;t,t+\Delta t) = \frac{\partial \varphi(u,\gamma)}{\partial u} + O(\Delta t) \tag{72}$$

which implies

$$|\det[W(u;t,t)]| = 1, \tag{73}$$

$$\mathrm{d}\{|\det[W(u;t,t+\Delta t)]|\}/\mathrm{d}t_{|\Delta t=0} = \sum_{k=1}^{K} \frac{\partial \varphi_k(u,\gamma)}{\partial u_k}. \tag{74}$$

Substituting into Eq. (69), and carrying out the required differentiations under the integral, we obtain

$$\partial P_l(\xi,\Delta\xi,t,\Delta t)/\partial \Delta t_{|\Delta t=0} = \int \cdots \int_{u\in(\xi,\xi+\Delta\xi)} \left\{ \sum_{k=1}^{K} \frac{\partial \varphi_k(u,l)}{\partial u_k} f_l(u,t) + \sum_{k=1}^{K} \varphi_k(u,l) \frac{\partial f_l(u,t)}{\partial u_k} + \frac{\partial f_l(u,t)}{\partial t} \right\} \mathrm{d}u. \tag{75}$$

Comparing expressions under the integrals in Eqs. (68) and (75) (note Eq. (70)), we obtain the following system of first-order linear PDEs:

$$\frac{\partial f_l(u,t)}{\partial t} + \sum_{k=1}^{K} \frac{\partial \varphi_k(u,l)}{\partial u_k} f_l(u,t) + \sum_{k=1}^{K} \varphi_k(u,l) \frac{\partial f_l(u,t)}{\partial u_k} = f_l(u,t) q_{ll}(x(t)) + \sum_{m\neq l} f_m(u,t) q_{ml}(x(t)), \quad l = 1,\ldots,m$$

or

$$\frac{\partial f_l(u,t)}{\partial t} + \operatorname{div}[\varphi(u,l) f_l(u,t)] = f_l(u,t) q_{ll}(x(t)) + \sum_{m\neq l} f_m(u,t) q_{ml}(x(t)), \quad l = 1,\ldots,m. \tag{76}$$

### *A. 1. Preservation of probability norming*

Adding Eqs. (76) together and remembering that $q_{ll}(t) = -\sum_{m\neq l} q_{lm}(t)$ we obtain

$$\sum_{l=1}^{M} \frac{\partial f_l(u,t)}{\partial t} + \operatorname{div}[\varphi(u,l) f_l(u,t)] = 0. \tag{77}$$

The assumption that there exists a domain $D$ invariant for system (59)–(60) ( i.e., that for initial conditions $x_0 \in D$, and any $\gamma$, $x(t) \in D$) implies that for all $l$ and $t$ supp $f_l(\cdot,t) \subseteq D$. Thus we can show applying Gauss–Green theorem that the integral of the marginal distribution function $\varrho(u,t)$,

$$\varrho(u,t) = \sum_{l=1}^{M} f_l(u,t) \tag{78}$$

is preserved in time by system (76). Let $D_0$ be a domain in $\mathbb{R}^K$such that closure $D \subset$ interior $D_0$. Thus we have $f_{l|\partial D_0} = 0$, $l = 1,\ldots,M$. The Gauss–Green Theorem (Evans, 2002) states that, provided $f(x) \in C^1(D_0)$ is a vector function bounded on $\partial D_0$,

$$\int_{D_0} \frac{\partial f}{\partial x_i}\,\mathrm{d}x = \int_{\partial D_0} \mathbf{n}\cdot f(x)\,\mathrm{d}S, \tag{79}$$

where $\mathbf{n}$ is a unit vector normal to $\partial D_0$ directed outwards. Thus, since div $f = \sum_i \partial f/\partial x_i$, from Eq. (77) and Gauss–Green theorem we obtain,

$$\frac{\partial}{\partial t}\int_{D_0} \varrho(u,t)\,\mathrm{d}u = -\sum_{l=1}^{M} \int_{D_0} \operatorname{div}[\varphi(u,l) f_l(u,t)]\,\mathrm{d}u = -\sum_{l=1}^{M} \int_{\partial D_0} f_l(u,t)\mathbf{n}\cdot\varphi(u,l)\,\mathrm{d}S = 0, \tag{80}$$

The last equality follows from the fact that $f_l(u,t) \equiv 0$ on $\partial D_0$, for each $l = 1,\ldots,M$. Let us note that since system (76) is linear, the value of marginal distribution norming $\int_{D_0} \varrho(u,t)\,\mathrm{d}u$ is not determined by the system itself, thus we have freedom to impose $\int_{D_0} \varrho(u,t) = 1$. System (59)–(60) can be interpreted as system describing the motion of a particle in $\mathbb{R}^K \times \{1,2,\ldots,M\}$ space. Intuitively, the conservation of marginal probability norming $\varrho$ is a direct consequence conservation of particles, i.e., a particle, kinetics of which is given by Eqs. (59) and (60), remains in the $D \times \{1,2,\ldots,M\}$ subdomain.

## Appendix B. Discretization techniques

Here we discuss the numerical method applied to calculate the stationary distributions $f(x,y), g(x,y)$ of system (21)–(22). To illustrate the technique let us begin with the simplified model (28)–(29). Let us consider the spatially discretized problem on the grid corresponding to the interval $[0,1]$. The continuous variable $x$ is replaced by $i/N$, $0 \leqslant i \leqslant N$. Let $f_i$ and $g_i$ denote probability distribution functions $f$ and $g$ at point $i$ of the grid. The discretized analog of system (30)–(31) now reads

$$\frac{\mathrm{d}f_i}{\mathrm{d}t} = b\frac{i}{N} g_i - c f_i + r((i+1)f_{i+1} - i f_i), \tag{81}$$

$$\frac{\mathrm{d}g_i}{\mathrm{d}t} = -b\frac{i}{N} g_i + c f_i + r((N+1-i)g_{i-1} - (N-i)g_i). \tag{82}$$

The resulting system consists of $2 \times (N+1)$ linear ODEs. In each of the equations the first two right-hand side terms correspond to exchange between distributions $f$ and $g$. The last two terms describe transport into grid point $i$, and from the grid point $i$. Note that in the case of distribution $f$ transport to grid point $i$ proceeds from grid point $i+1$, while for the distribution $g$ transport proceeds to grid point $i$ from the grid point $i-1$. Assumption that $f$ and $g$ vanish out of interval $[0,1]$ may be replaced by setting $f_{N+1} = g_{-1} = 0$, which closes the system.

The stationary distributions is calculated by setting $\mathrm{d}f_i/\mathrm{d}t = \mathrm{d}g_i/\mathrm{d}t = 0$. This simplifies system (81)–(82) to a system of linear algebraic equations

$$b_r \frac{i}{N} g_i - c_r f_i + (i+1) f_{i+1} - i f_i = 0, \tag{83}$$

$$-b_r \frac{i}{N} g_i + c_r f_i + (N+1-i) g_{i-1} - (N-i) g_i = 0, \tag{84}$$

where, recall, $b_r = b/r$ and $c_r = c/r$. The resulting system has no free terms and therefore it may not have unique solution. To make the system unique, we replace one of its $2 \times (N+1)$ equations by the normalization equation $[1/(N+1)\sum_i (f_i + g_i) = 1$. In Fig. 16 we compare the two solutions of system (83)–(84) calculated for $N = 50$ and 200 with the analytic result (37)–(38). As one may expect, the accuracy of the applied method grows with the size of the grid. For $N = 500$ the difference would be indistinguishable by eye.

The full system (21)–(22) is considered on the grid $i,j$, where $0 \leqslant i \leqslant N, 0 \leqslant j \leqslant N$. The continuous variables $x$ and $y$ are replaced by $i/N$ and $j/N$, respectively. Let $f_{i,j}$ and $g_{i,j}$ denote distributions $f$ and $g$ at point $i,j$ of the grid. This way we replace $f(x,y)$ and $g(x,y)$ by $2 \times (N+1)^2$ variables.

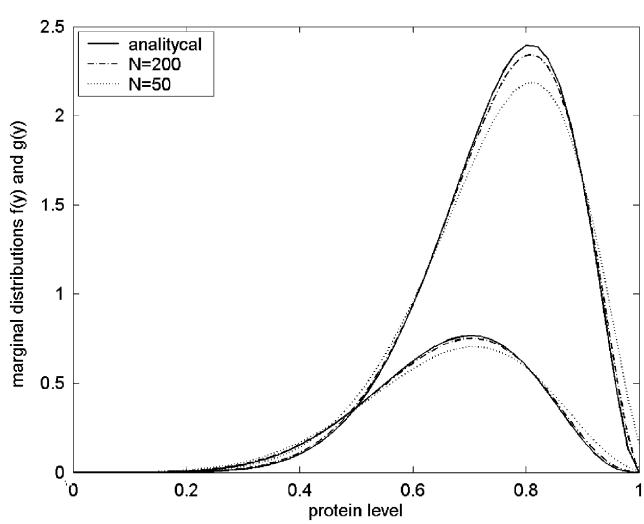


Fig. 16. Numerical solutions for distributions $f(y)$ and $g(y)$ on the grid of size $N = 50$ (dotted lines) and $N = 200$ (dashed-dotted lines) are compared with analytical solutions (37,38), (continuous line). The parameters are $c_r = 6$, $b_r = 3$.

The discretized analog of system (21)–(22) now reads

$$\begin{aligned}\frac{\mathrm{d}f_{i,j}}{\mathrm{d}t} = {} & b\frac{j}{N} g_{i,j} - c f_{i,j} - i f_{i,j} + (i+1) f_{i+1,j} \\ & - r|i-j| f_{i,j} + r(i+1-j) f_{i,j-1} L_1 \\ & + r(j+1-i) f_{i,j+1} L_2,\end{aligned} \tag{85}$$

$$\begin{aligned}\frac{\mathrm{d}g_{i,j}}{\mathrm{d}t} = {} & -b\frac{j}{N} g_{i,j} + c f_{i,j} - (N-i) g_{i,j} + (N+1-i) g_{i-1,j} \\ & - r|i-j| g_{i,j} + r(i+1-j) g_{i,j-1} L_1 \\ & + r(j+1-i) g_{i,j+1} L_2,\end{aligned} \tag{86}$$

where $L_1$ and $L_2$ are the logical variables,

$$L_1 = 1 \text{ for } i > j-1 \quad \text{and} \quad L_1 = 0 \text{ for } i \leqslant j-1 \tag{87}$$

and

$$L_2 = 1 \text{ for } i < j+1 \quad \text{and} \quad L_2 = 0 \text{ for } i \geqslant j+1. \tag{88}$$

As in the previous case, the stationary distributions are calculated by setting $\mathrm{d}f_{ij}/\mathrm{d}t = \mathrm{d}g_{ij}/\mathrm{d}t = 0$. As a result we obtain system of $2 \times (N+1)^2$ algebraic linear equations. To make the solution unique we replace one of the equations by the normalization $[1/(N+1)^2]\sum_{ij}(f_{i,j} + g_{i,j}) = 1$. We note that the matrix of the resulting system is relatively sparse, and the number of non-zero entries grows as $N^2$, not as $N^4$. The random access memory (RAM) shortage is the main problem here. We used the MATLAB 7.0 solver and to save RAM we declare matrix as *sparse* and then we invert it using matrix left division function to solve the system. The problem can be solved on a coarse grid, $N = 50$, in less than a minute, on an average PC, but to solve it on a finer grid (up $N = 300$) we use server with 8 GB RAM. In a same way we discretize systems (49)–(52) and (55)–(58) for protein–protein stationary distributions. In the case of activator–repressor (49)–(52) we obtain

$$\begin{aligned}& -i f_{ij}^{0,0} + (i+1) f_{i+1,j}^{0,0} - rj f_{ij}^{0,0} + r(j+1) f_{i,j+1}^{0,0} \\ & \quad - \left(p + k\frac{i}{N}\right) f_{ij}^{0,0} + f_{ij}^{0,1} h + f_{ij}^{1,0} q\frac{j}{N} = 0,\end{aligned} \tag{89}$$

$$\begin{aligned}& -(N-i) f_{ij}^{1,0} + (N+1-i) f_{i-1,j}^{1,0} - rj f_{ij}^{1,0} + r(j+1) f_{i,j+1}^{1,0} \\ & \quad - \left(q\frac{j}{N} + k\frac{i}{N}\right) f_{ij}^{1,0} + p f_{ij}^{0,0} + h f_{ij}^{1,1} = 0,\end{aligned} \tag{90}$$

$$\begin{aligned}& -i f_{ij}^{0,1} + (i+1) f_{i+1,j}^{0,1} - r(N-j) f_{ij}^{0,1} + r(N+1-j) f_{i,j-1}^{0,1} \\ & \quad - (p+h) f_{ij}^{0,1} + f_{ij}^{0,0} k\frac{i}{N} + f_{ij}^{1,1} q\frac{j}{N} = 0,\end{aligned} \tag{91}$$

$$\begin{aligned}& -(N-i) f_{ij}^{1,1} + (N+1-i) f_{i-1,j}^{1,1} - r(N-j) f_{ij}^{1,1} \\ & \quad + r(N+1-j) f_{i,j-1}^{1,1} - \left(q\frac{j}{N} + h\right) f_{ij}^{1,1} \\ & \quad + k\frac{i}{N} f_{ij}^{1,0} + p f_{ij}^{0,1} = 0.\end{aligned} \tag{92}$$

System (85)–(86) of linear ODEs can be used to simulate time evolution of densities $f(x, y, t)$ and $g(x, y, t)$. We use functions *lsim* or *initial* designated for systems of linear ODEs. The simulation needs however even larger RAM and is also much more time consuming. Using our solver we can simulate evolutions on grids with up to $N = 50$. On such a grid it takes several hours until the proximity of steady state is reached. On a grid with $N = 15$, the problem can be solved on PC in less than a minute.

## References


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

Arkin, A., Ross, J., McAdams, H.H., 1998. Stochastic kinetics analysis of developmental pathway bifurcation in $\lambda$-phage infected *Escherichia Coli* cells. Genetics 149, 1633–1648.

Basak, G.K., Bisi, A., Ghosh, M.K., 1999. Stability of degenerate diffusions with state-dependent switching. J. Math. Anal. Appl. 240, 219–248.

Blake, W.J., Kaern, M., Cantor, C.R., Collins, J.J., 2003. Noise in eucaryotic gene expression. Nature 422, 633–637.

Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S., 2002. Stochastic gene expression in a single cell. Science 297, 1183–1186.

Emch, G.G., Liu, C., 2002. The Logic of Thermostatistical Physics, Springer, Berlin, p. 494.

Evans, L.C., 2002. Partial Differential Equations. AMS, Providence, RI, p. 627.

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

Gillespie, D.T., 1977. Exact stochastic simulations of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361.

Gilman, A., Arkin, A.P., 2002. Genetic code: representation and dynamics models of genetic components and networks. Annu. Rev. Genomics Hum. Genet. 3, 341–369.

Horsthemke, W., Lefever, R., 1984. Noise Induced Transitions. Theory and Applications in Physics, Chemistry and Biology. Springer, Berlin.

Iwankiewicz, R., Nielsen, S.R.K., 2000. Solution techniques for pulse problems in non-linear stochastic dynamics. Prob. Eng. Mech. 15, 25–36.

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

Kierzek, A.M., Zaim, J., Zielenkiewicz, P., 2001. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. J. Biol. Chem. 276, 8165–8172.

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

Lipniacki, T., Paszek, P., Brasier, A.R., Luxon, B., Kimmel, M., 2004. Mathematical model of NF-kB regulatory module. J. Theor. Biol. 228, 195–215.

Lipniacki, T., Paszek, P., Brasier, A.R., Tian, B., Wang, H-Q., Luxon B., Kimmel, M., 2005. Stochastic regulation in early immune response. Biophys. J., submitted.

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

McAdams, H.H., Shapiro, L., 2003. A bacterial cell-cycle regulatory network operating in time and space. Science 301, 1874–1877.

Paszek, P., Lipniacki, T., Brasier, A.R., Tian, B., Nowak D. E., 2005. Stochastic effects of multiple regulators on expression profiles in Eukaryotes. J. Theor. Biol. 233, 423–433.

Pirone, J.R., Elston, T.C., 2004. Fluctuations in transcription factor binding can be explain the graded and binary responses observed in inducible gene expression. J. Theor. Biol. 226, 111–121.

Rao, Ch.V., Wolf, D.M., Arkin, A.P., 2002. Control, exploitation and tolerance of intracellular noise. Nature 420, 231–237.

Raser, J.M., O'Shea, E.K., 2004. Control of stochasticity in eukaryotic gene expression. Science 304, 1811–1814.

Shea, M.A., Ackers, G.K., 1985. The Or control system of bacteriophage lambda: a physical–chemical model for gene regulation. J. Mol. Biol. 181, 211–230.

Simpson, M.L., Cox, Ch.D., Sayer, G.S., 2004. Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. J. Theor. Biol. 229, 383–394.

Stirland, J.A., Seymour, Z.C., Windeatt, S., Norris, A.J., Stanley, P., Castro, M.G., Loudon, A.S.I., White, M.R.H., Davis, J.R.E., 2003. Real-time imaging of gene promoter activity using an adenoviral reporter construct demonstrates transcriptional dynamics in normal anterior pituary cells. J. Endocrinol. 178, 61–69.

Swain, P.S., Elowitz, M.B., Siggia, E.D., 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc. Natl Acad. Sci. USA 99, 12795–12800.

Tao, Y., 2004a. Intrinsic and external noise in an auto-regulatory genetic network. J. Theor. Biol. 229, 147–156.

Tao, Y., 2004b. Intrinsic noise gene regulation and steady-state statistics in a two-gene network. J. Theor. Biol. 231, 563–568.

Takasuka, N., White, M.R.H., Wood, C.D., Robertson, W.R., Davis, J.R.E., 1998. Dynamic changes in prolactin promoter activation in individual living lactotrophic cells. Endocrinology 139, 1361–1368.

Thattai, M., Oudenaarden, A., 2001. Intrinsic noise in gene regulatory networks. Proc. Natl Acad. Sci. USA 98, 8614–8619.

Tomioka, R., Kimura, H., Kobayashi, T.J., Aihara, K., 2004. Multivariate analysis of noise in genetic regulatory networks. J. Theor. Biol. 229, 501–521.

Walters, M.C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M., Martin, D.I.K., 1995. Enhancers increase the probability but not the level of gene expression. Proc. Natl Acad. Sci. USA 92, 7125–7129.