

# Non-monotonicity of Fano factor in a stochastic model for protein expression with sequestration at decoy binding sites

Michal Hojcka, Pavol Bokes  
Department of Applied Mathematics and Statistics  
Comenius University  
Bratislava, Slovakia  
michal.hojcka@fmph.uniba.sk, pavol.bokes@fmph.uniba.sk

*Received: 11 July 2017, accepted: 21 October 2017, published: 2 November 2017*

**Abstract**—We present a stochastic model motivated by gene expression which incorporates unspecific interactions between proteins and binding sites. We focus on characterizing the distribution of free (i.e. unbound) protein molecules in a cell. Although it cannot be expressed in a closed form, we present three different approaches to obtain it: stochastic simulation algorithms, system of ODEs and quasi-steady-state solution. Additionally we use a large-system-size scaling to derive statistical measures of approximate distribution of the amount of free protein, such as the Fano factor. Intriguingly, we report that while in the absence of or in the excess of decoy binding sites the Fano factor is equal to one (suggestive of Poissonian fluctuations), in the intermediate regimes of moderate levels of binding sites the Fano factor is greater than one (suggestive of super-Poissonian fluctuations). We support and illustrate all of our results with numerical simulations.

**Keywords**—Gene Expression; Master Equation; Small Noise Approximation; Stochastic Simulation

## I. INTRODUCTION

The number of proteins and other species present in the biological processes inside the cells such as gene expression is usually small [6], [28]. Therefore, deterministic modeling of such reactions can be quite inaccurate and we often turn to stochastic methods [18]. They account for discrete number of molecules and can easily be simulated through stochastic simulation algorithms, in particular the Gillespie algorithm [10], [11]. Being a very timely topic, gene expression spurred a revival of interest in Markovian models of chemical kinetics, e.g. [25]. We assume that the protein is produced with a constant rate and that the rate of its decay is proportional to the number of proteins. We study the protein dynamics in presence of so-called decoy binding sites [17] on the DNA. Our model takes into account protein binding/unbinding reactions with these binding sites. Similar models have already been studied previ-

**Copyright:** © 2017 Hojcka et al. This article is distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Citation:** Michal Hojcka, Pavol Bokes, Non-monotonicity of Fano factor in a stochastic model for protein expression with sequestration at decoy binding sites, *Biomath* 6 (2017), 1710217, <http://dx.doi.org/10.11145/j.biomath.2017.10.217>

ously; in particular [9] investigated the model with protected complexes, i.e. the case when bound proteins were immune to degradation, showing that the steady-state distribution is Poissonian. Our model allows bounded proteins to degrade, which introduces additional noise into the model [2], [3]. For simplicity, we ignore effects of burst-like protein synthesis or transcriptional auto-regulation [2], [4], [26]. In section II we formulate the Master equation for the probability distribution of free protein; unfortunately, we are unable to find its solution in a closed form. However, biochemical reactions often operate on different timescales as was already thoroughly investigated in works such as [14], [12], [13]. Specifically, in the current context the interactions between the protein and its binding sites occur on a substantially faster timescale than the turnover of protein (by transcription and decay) does [1]. This allows us to successfully use singular perturbation methods [5], [22], [23] to obtain the quasi-steady-state solution to our problem. Obtaining the quasi-steady-state approximations in our model involves finding an equilibrium of binding/unbinding reaction, which is a specific case of a reversible bimolecular reaction studied by Laurenzi in [16]. In section III we expand the Master equation using the linear noise expansion as proposed in [29].

## II. STOCHASTIC APPROACH

The number of proteins expressed from a single gene can often be quite small [31]; it would therefore be inaccurate to use deterministic approach, which treats reactants as continuous variables. Instead we take a stochastic approach, in which we model each species as a discrete random variable and each reaction as a random event with a given probability to occur. Some protein species are present at even less than 10 copies per *E. coli* cell [28], therefore we also work with mean number of proteins below 100 in this paper. As we use discrete variables it is quite clear that we need to simulate these reactions through a simulation algorithm. Such algorithms gained a wider recognition after the work of Gillespie [10].

### A. Model Description

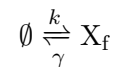
In this section we present a minimalistic model for gene expression. We neglect the effect of mRNA translation and production of proteins in bursts; instead we focus on the interaction between the proteins and decoy binding sites. Unlike in [26], we assume that bounded proteins are subject to degradation processes.

Let us use the following notation for our variables:

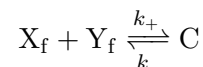
- $X$  - protein (free or bound),
- $X_f$  - free protein,
- $Y$  - binding site,
- $Y_f$  - free binding site,
- $C$  - complex (protein bound to the binding site).

For the sake of simplicity we omit 'decoy' from the binding site notation as we do not take into account any other binding sites. We assume that three reversible reactions can take place:

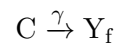
- 1) Protein production/decay.



- 2) Protein binding/unbinding reaction.



- 3) Decay of the complex (whereby a binding site is vacated).



We use upper-case letters in *italics* to represent a number of corresponding species throughout this paper. We reserve the corresponding lower-case letter as a notation for a concentration of a given species. In order to avoid confusion with  $X$ , we use  $N$  instead of  $X_f$  as the number of free protein. Although we mentioned five different variables, the problem is just two-dimensional, with the following straightforward conservation laws held between the variables:

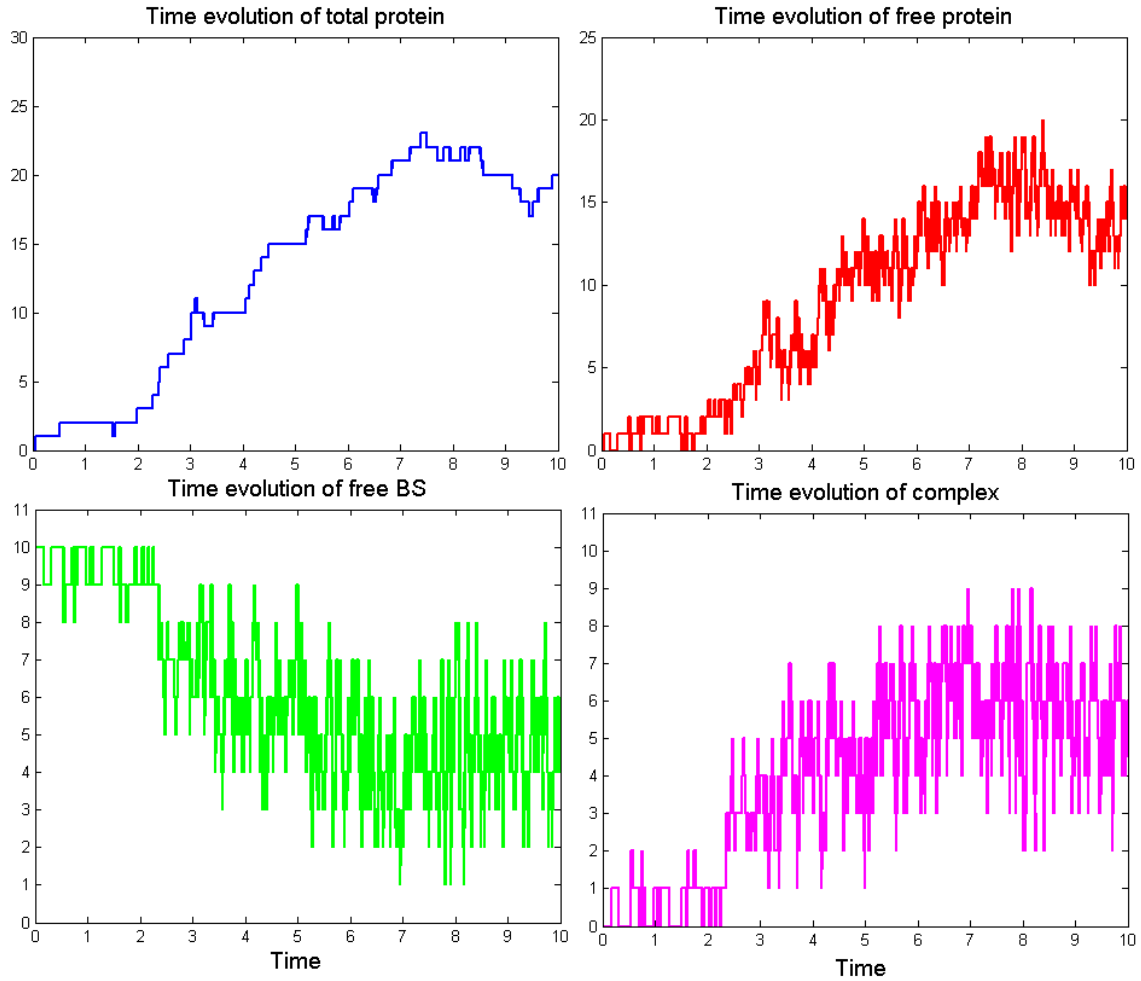


Fig. 1: Simulation using the Gillespie algorithm.

- $Y$  is a known constant (the number of binding sites)
- $C = X - N$  (the number of complexes is the same as the number of bound protein)
- $Y_f = Y - C = Y - X + N$  (the number of free binding sites is the same as the number of all binding sites without the complexes)

$$\begin{aligned}
 \dot{P}_{X,N} = & kP_{X-1,N-1} - kP_{X,N} \\
 & + \gamma(N+1)P_{X+1,N+1} - \gamma NP_{X,N} \\
 & + k_+(N+1)(Y-X+N+1)P_{X,N+1} \\
 & - k_+N(Y-X+N)P_{X,N} \\
 & + k_-(X-N+1)P_{X,N-1} - k_-(X-N)P_{X,N} \\
 & + \gamma(X-N+1)P_{X+1,N} - \gamma(X-N)P_{X,N},
 \end{aligned} \tag{1}$$

Therefore we can express the Master Equation of the system in terms of  $X$  and  $N$  (with a constant total number of binding sites  $Y$ ):

where  $P_{X,N}$  is the abbreviation of  $P(X(t) = X, X_f(t) = N)$ . We can use the Gillespie algorithm to simulate this process (parameters of the reactions are set to  $Y = 10$ ,  $k = 3$ ,  $\gamma = 0.1$ ,  $k_+ = 1$ ,  $k_- = 10$  with no proteins at the

beginning:  $P_{X,N}(0) = \delta_{X,0}\delta_{N,0}$ . We illustrate the individual components of the process in Figure 1.

**B. Total Protein distribution**

Our first goal is to obtain the marginal distribution of the total protein number. It can be obtained as the sum of  $P_{X,N}$  through all possible  $N$ 's:  $P_X = \sum_{N=-\infty}^{\infty} P_{X,N}$ . Let us use an abbreviation  $\sum_N$  for the sum through all integers. In order to derive an equation  $P_X$  from (1) let us sum both sides of the equation through all  $N$ 's. As the sum goes through all integers (if  $N < 0$  is the probability is naturally equal to zero), we can use the fact that  $\sum_N f(N) = \sum_N f(N \pm 1)$ , obtaining

$$\begin{aligned} \dot{P}_X = & k \sum_N (P_{X-1,N} - P_{X,N}) \\ & + \gamma \sum_N N (P_{X+1,N} - P_{X,N}) \\ & + \gamma \sum_N ((X - N + 1)P_{X+1,N} - (X - N)P_{X,N}), \end{aligned}$$

which simplifies to

$$\dot{P}_X = k (P_{X-1} - P_X) + \gamma ((X + 1)P_{X+1} - X P_X). \tag{2}$$

A system of differential equations in this form can be solved using the method of generating functions [15]. First we multiply both sides by  $s^X$  and sum them through all integer  $X$ 's.

$$\begin{aligned} \frac{\partial}{\partial t} \sum_X s^X P(X,t) = & k \left( \sum_X s^X P_{X-1} - \sum_X s^X P_X \right) \\ & + \gamma \left( (X + 1) \sum_X s^X P_{X+1} - X \sum_X s^X P_X \right). \end{aligned} \tag{3}$$

Now we can use the definition of the generating function  $G(s,t) = \sum_X s^X P_X$  and its derivative  $\frac{\partial G}{\partial s} = \sum_X X s^{X-1} P_X$  in order to transform (3) into

a partial differential equation:

$$\begin{aligned} \frac{\partial G}{\partial t} &= k(s - 1)G + \gamma(1 - s) \frac{\partial G}{\partial s} \\ &= (s - 1) \left( kG - \gamma \frac{\partial G}{\partial s} \right). \end{aligned} \tag{4}$$

In the steady state we can omit the left-hand side of the equation (as  $\frac{\partial G}{\partial t} = 0$ ) and easily solve it using the separation of variables. We come to the solution  $G(s, \infty) = e^{\frac{k}{\gamma}(s-1)}$ , which can be recognized as the probability generating function of the Poisson distribution parametrized by  $\lambda = \frac{k}{\gamma}$ . Using the notation  $\langle X \rangle = \lambda$  for the distribution's mean we can write  $P_X(\infty) = \frac{\langle X \rangle^X e^{-\langle X \rangle}}{X!}$ . Away from the steady state, (4) is an example of a non-homogeneous first-order linear partial differential equation, which can be solved for suitable initial conditions. Using a common extension of the method of characteristics for (quasi-)linear PDE's (see e.g. [8]) we introduce an auxiliary function  $u = u(s, t, G)$ , which satisfies

$$\frac{\partial u}{\partial t} + \gamma(s - 1) \frac{\partial u}{\partial s} + k(s - 1)G \frac{\partial u}{\partial G} = 0.$$

The characteristic system of this equation has the form:

$$\begin{aligned} \dot{t} &= 1, & t(\tau) &= \tau + C_1 \\ \dot{s} &= \gamma(s - 1), & s(\tau) &= C_2 e^{\gamma\tau} + 1 \\ \dot{G} &= k(s - 1)G, & \frac{dG}{G} &= k(C_2 e^{\gamma\tau}) d\tau \end{aligned} \tag{5}$$

By finding the functions which are constant on the characteristics, we obtain solutions in the form  $G(s, t) = \psi \left( t - \frac{\ln(s-1)}{\gamma} \right) \cdot e^{\frac{ks}{\gamma}}$ . We can specify the results for any particular initial conditions. Let us investigate the case when there is no protein at the beginning, i.e.  $P(x, 0) = \delta_{x,0}$ , which gives us an initial condition  $G(s, 0) = 1$  also for the generating function. Using the initial condition we come to the solution  $G(s, t) = e^{-\frac{k}{\gamma}(e^{-\gamma t}(s-1)+1) + \frac{ks}{\gamma}} = e^{-\frac{k}{\gamma}(s-1)(e^{-\gamma t}-1)}$ , which is again the Poisson distribution, just with different, time-dependent, parameter meaning that  $\langle X(t) \rangle = \frac{k}{\gamma} \cdot (1 - e^{-\gamma t})$ . Taking  $t \rightarrow \infty$  we can easily see that the time-dependent solution converges to the steady-state one.

C. Free Protein distribution

Searching for an exact formula for the free protein distribution yields no apparent results: the master equation (1) does not have solution in the closed form (unless  $Y = 0$ ). Therefore we have to look for alternative ways to obtain it.

1) *Stochastic Simulation Algorithms*: The first approach consists of using the techniques of stochastic simulation algorithms, in particular the Gillespie algorithms (see [7] for a practical guide). With their help we can simulate the time evolution of all species involved in a system of reactions. The simulation technique is based upon the following scheme: we generate two uniformly distributed random numbers, the first of which determines the time of the next reaction and the second one determines which reaction will occur. When we use a sufficient number of sample trajectories, we obtain a robust estimate for the distribution of any variable. The main problem of this approach is its limited computational capacity. In order to obtain a robust distribution of variables we have to run a substantial number of simulations; this can be very time-consuming and even unrealistic when calculating results for many different parameter sets.

2) *System of ODEs*: The second option is to transform the Master equation, which is an infinite system of equations, into a finite system of ordinary differential equations. The most straightforward way to obtain this is to set threshold values of the discrete variables, replacing the unknown probability by zero whenever these thresholds are exceeded (see [20], in our case we set all  $P_{X,N}$  to zero, whenever  $X > 100$ ). In this way, the overall sum of probability distribution is no longer conserved at one, but in return we get a finite system of differential equations. We use the MATLAB *ode15s* solver for stiff problems to calculate the solution. But this system is also very time-consuming, as the number of equations and the computational difficulty grows rapidly with raising the thresholds for the non-zero probabilities of  $P_{X,N}$ .

3) *Quasi-steady-state solution*: In order to obtain the solution in a closed form, which is as close to exact solution as possible, we proceed to obtain the so-called quasi-steady-state solution [24], [27]. We utilize the fact from biological background that binding/unbinding reactions are fast compared to protein production/degradation ( $k_- \gg \gamma$ ). In order to use this fact we perform the singular perturbation reduction [9] using the ratio  $\gamma/k_-$  as a small parameter. Let us introduce new dimensionless time and the parameters:

$$\varepsilon = \frac{\gamma}{k_-}, \quad k_b = \frac{k_-}{k_+}, \quad t = \frac{\tau}{\gamma}.$$

The master equation then reads

$$\begin{aligned} \varepsilon \frac{d}{d\tau} P_{X,N} = & \varepsilon \frac{k}{\gamma} (P_{X-1,N-1} - P_{X,N}) \\ & + \varepsilon ((N+1)P_{X+1,N+1} - \gamma N P_{X,N}) \\ & + \frac{1}{k_b} ((N+1)(Y-X+N+1)P_{X,N+1} \\ & - N(Y-X+N)P_{X,N}) \\ & + (X-N+1)P_{X,N-1} - (X-N)P_{X,N} \\ & + \varepsilon ((X-N+1)P_{X+1,N} - (X-N)P_{X,N}). \end{aligned} \tag{6}$$

As  $\varepsilon$  tends to zero, we obtain an equation for the leading-order approximation of  $P_{X,N}$ . The number of total protein  $X$  is constant after this approximation, so that we can abbreviate  $P_{X,N}$  by  $P_N$ , writing

$$\begin{aligned} (N+1)(Y-X+N+1)P_{N+1} = \\ (N(Y-X+N) + k_b(X-N))P_N - \\ - k_b(X-N+1)P_{N-1}. \end{aligned} \tag{7}$$

Solving (7) subject to boundary conditions (see e.g. [16])  $P_N(N < 0) = 0$  and  $P_N(N > X \vee N < X - Y) = 0$  yields

$$P_N = P_{X,N} = \frac{k_b^N C(X)}{N!(X-N)!(Y-X+N)!}, \tag{8}$$

where  $C(X)$  is a constant with respect to  $N$ , dependent only on the value of  $X$ .

The multiplicative constant  $C(X)$  can be deter-

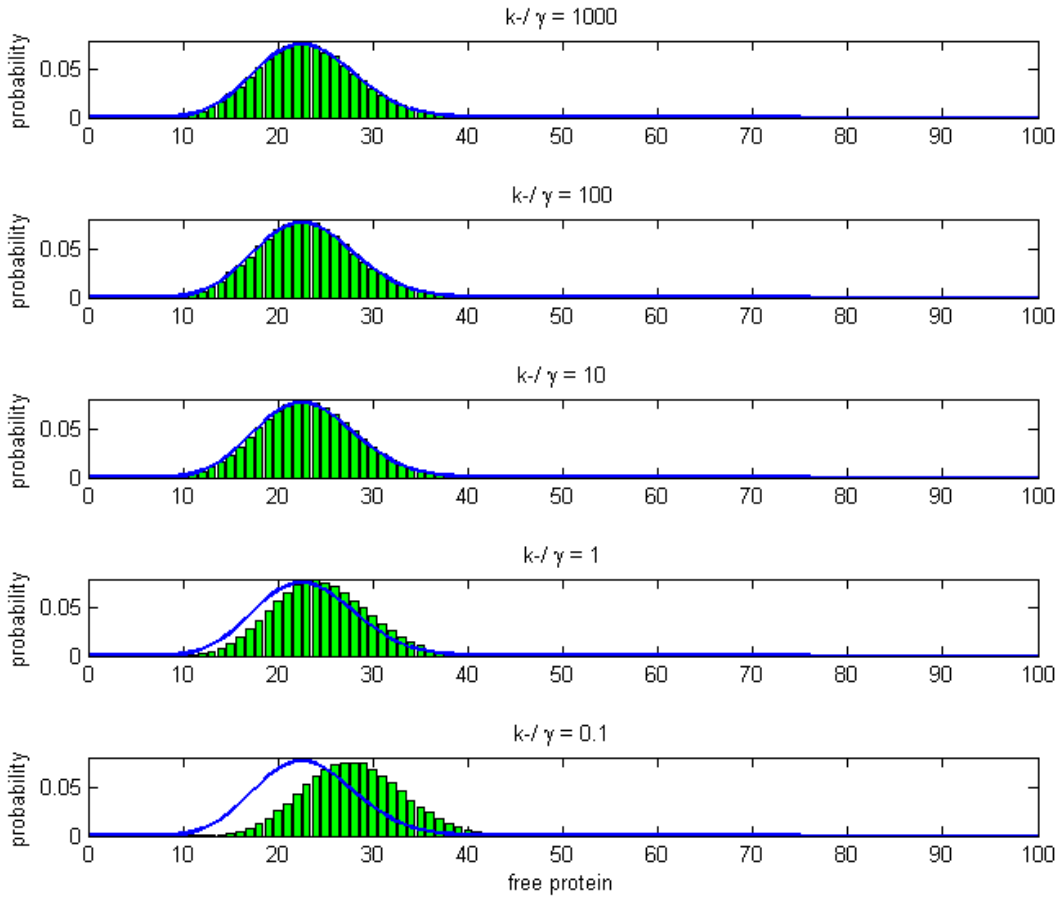


Fig. 2: Quality of approximation by the quasi-steady-state solution as function of  $k_-/\gamma$

mined from the relation

$$P_X = \sum_{N=\max\{X-Y,0\}}^X P_{X,N},$$

since

$$\begin{aligned} P_X &= \sum_{i=\max\{0,X-Y\}}^X P_{X,i} \\ &= C(X) \sum_{i=\max\{0,X-Y\}}^X \frac{k_b^i}{(X-i)!(Y-X+i)!i!}, \end{aligned}$$

and thus

$$C(X) = P_X \left( \sum_{i=\max\{0,X-Y\}}^X \frac{k_b^i}{(X-i)!(Y-X+i)!i!} \right)^{-1}.$$

Since we already know about the Poisson character of steady-state solution for  $P_X$ , we are ready to calculate the number of free protein in quasi-steady state:

$$P_N = \sum_{X=N}^{N+Y} \frac{P_X \cdot \frac{k_b^N}{N!(X-N)!(Y-X+N)!}}{\sum_{i=\max\{0,X-Y\}}^X \frac{k_b^i}{(X-i)!(Y-X+i)!i!}} \quad (9)$$

This represents a major improvement in terms of computational requirements necessary to get the distribution of free protein amount.

In order to illustrate the quality of this approximate solution, let us investigate the differences between the simulated distribution (by Gillespie algorithm) and the approximate quasi steady state

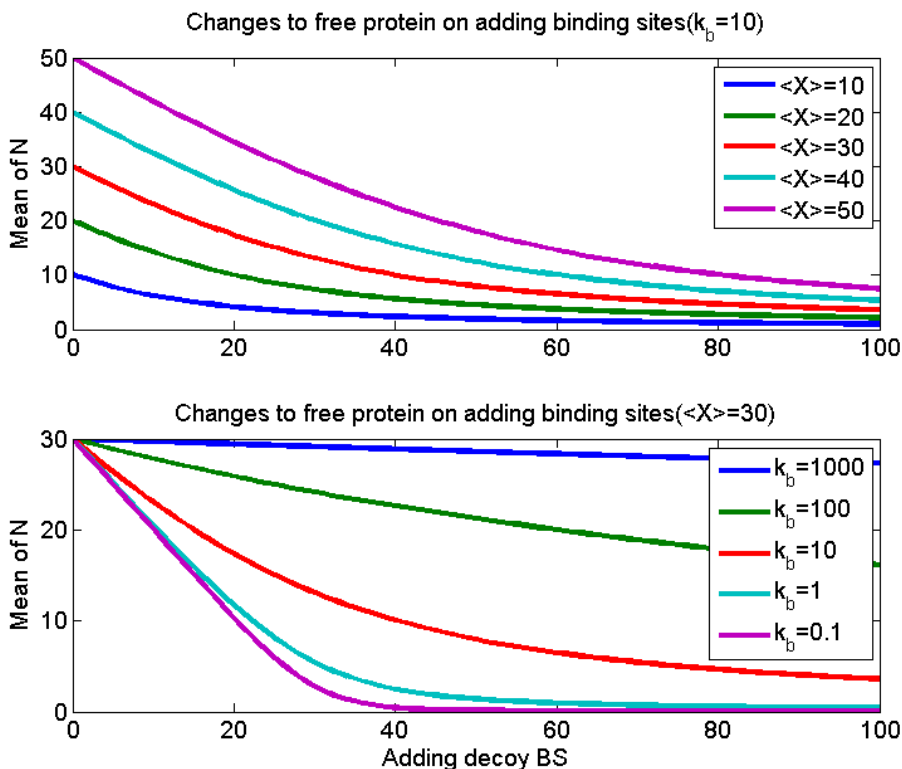


Fig. 3: Mean of free protein probability distribution

solution (9) for selected values of  $\varepsilon^{-1} = \frac{k_-}{\gamma}$ , which is assumed to be large as justified above. We assumed the following values of the parameters:  $Y = 10; k = 3, \gamma = 0.1 \Rightarrow \langle X \rangle = 30; k_+ = a, k_- = 10a \Rightarrow k_b = 10$ . We vary the value of  $a$  in order to change the value of  $(\varepsilon^{-1})$  while keeping  $k_b$  fixed; thus the distribution obtained from (9) remains the same. The results are shown in Figure 2, where the green histogram is estimated from repeated simulations ( $10^5$  trajectories) of the Gillespie algorithm on a sufficiently long timescale. The blue line is the probability distribution of free protein in quasi-steady state (9). It is easy to see that with the lowering of  $\frac{k_-}{\gamma}$ , we diverge from exact distribution results estimated by the Gillespie algorithm histogram. To evaluate the goodness of the fit, we calculate the  $\ell^2$  distance between these two realizations of free protein distribution for all  $N$ 's from 0 to 100 (see TABLE I).

$k_-/\gamma$	Diff
1000	0.0031
100	0.0036
10	0.0071
1	0.048
0.1	0.1483

TABLE I: The  $\ell^2$  distance between the quasi-steady state approximation and the exact distribution estimated by stochastic simulation.

#### D. Moments of free protein distribution

In this section we focus on the analysis of the free protein probability distribution in the quasi-steady state. We focus on four basic statistical characteristics: mean ((Figure 3)), variance, the Fano factor ( $F = \frac{\sigma^2}{\mu}$ ) (Figure 4) and the squared coefficient of variation ( $CV^2 = \frac{\sigma^2}{\mu^2}$ ). We use the number of binding sites as a free parameter, which we vary on the horizontal axis. We sometimes

use an abbreviation BS instead of binding site. We study the change in these characteristics for selected choices of the total protein production  $\langle X \rangle$  and the dissociation constant  $k_b$ . The values  $\gamma = 0.1$  and  $k_- = 10$  are fixed. Different choices of  $k$  and  $k_+$  are used to modify  $\langle X \rangle$  and  $k_b$ .

Without any binding sites ( $Y = 0$ ) the Poisson distribution follows, so that  $\mu = \sigma^2$  and thus  $F = 1$ . Very large dissociation constants imply that proteins bind weakly to the binding sites and thus the system exhibits characteristics that are consistent with the Poisson distribution. It is apparent that for large values of  $Y$ , the mean and variance tend to 0, Fano factor tends to 1 and  $CV^2$  diverges to infinity. The convergence of the Fano factor to one suggests an approximate Poisson distribution in the excess of decoy binding sites. Below, we show with asymptotics that this is indeed the case.

Provided that  $Y \gg X$ , we can use the approximation for the factorial  $(Y - X + N)! \approx Y! Y^{N-X}$ . Substituting this into (9) and using binomial theorem and basic algebraic manipulation, we obtain a simplified equation for the probability distribution of the free protein:

$$P_N = \sum_{X=N}^{N+Y} P_X \cdot \binom{X}{N} \left( \frac{k_b}{k_b + Y} \right)^N \left( \frac{Y}{k_b + Y} \right)^{X-N}.$$

This formula can also be rewritten as  $\sum P_X \cdot P_{N|X}$ , where  $P_X$  has a Poisson distribution with parameter  $\langle X \rangle$ ; we see that  $P_{N|X}$  has a binomial distribution with probability of a Bernoulli trial given by  $\frac{k_b}{k_b + Y}$  and the number of trials given by  $X$ . Therefore, we can express the free protein distribution to be that of a random  $N = \sum_{i=0}^X \xi_i$ ,

where  $P(\xi_i = 1) = \frac{k_b}{k_b + Y}$ . After some algebraic manipulation with probability generating functions (such as the rule for calculating probability generating function of the sum of independent random variables) we arrive at:

$$G_N(s) = G_X(G_\xi(s)). \tag{10}$$

It is clear now that probability generating function for free protein is given as a composition of

generating functions for  $G_X$  (Poisson distribution) and  $G_\xi$  (Bernoulli distribution). Substituting the well-known formulae for generating functions of these distributions into (10), we obtain

$$G_N(s) = \exp(\langle X \rangle (G_\xi(s) - 1)) = \exp\left(\left(\frac{k_b \langle X \rangle}{k_b + Y}\right) (s - 1)\right), \tag{11}$$

which again indicates the Poisson distribution, but with a different parameter

$$\langle N \rangle = \langle (N - \langle N \rangle)^2 \rangle = \frac{k_b \langle X \rangle}{k_b + Y}. \tag{12}$$

Computing the probability distribution in quasi-steady state for large values of binding sites is infeasible as the formula works with the terms of the order  $Y!$ ; therefore, we use the Gillespie algorithm (with  $10^5$  repetitions) to estimate the mean and the variance of the distribution. The results are provided in TABLE II. Other parameters used are  $\langle X \rangle = 10$  and  $k_b = 10$ ). As the distribution is Poissonian, the Fano factor tends to one and  $CV^2 = \frac{1}{\langle N \rangle}$ , which tends to infinity.

$Y$	approximate	$\langle N \rangle$	$\langle (N - \langle N \rangle)^2 \rangle$
100	0.9091	1.0027	1.0124
500	0.1961	0.1999	0.2009
1000	0.099	0.0999	0.0998
5000	0.02	0.02	0.02
10000	0.01	0.0101	0.01

TABLE II: The approximate value (12) against simulated estimates of the mean and variance in the large  $Y$  regime.

### E. Comparison of the methods

In this section we compare the probability distributions generated by the three methods mentioned earlier: stochastic simulation by Gillespie algorithm (referred to as Gill in the TABLE III), transforming master equation into the finite system of ODEs (ODE in TABLE III) and calculating the solution in quasi-steady state (Quasi in TABLE III). Obviously we use the same set of parameters in all three approaches:  $Y = 10$ ,  $k = 3$ ,  $k_+ =$



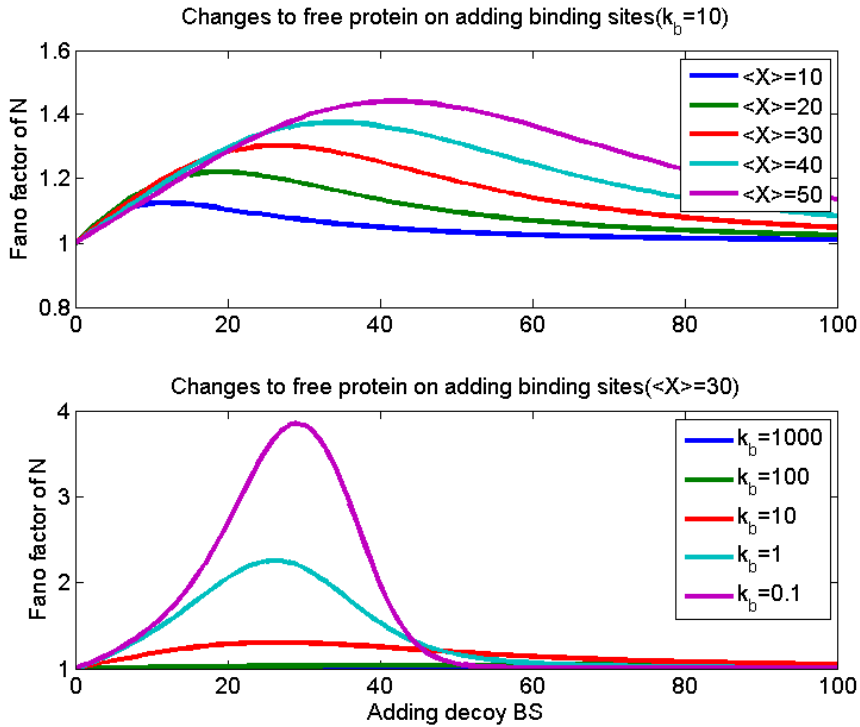


Fig. 4: Fano factor of free protein probability distribution

1,  $k_- = 10$  and  $\gamma = 0.1$ . Initially, the system is set to contain no proteins:  $P_{X,N} = \delta_{(0,0)}$ . The output of each method is the probability surface  $P_{X,N}$  in all given time points. In TABLE III we record the differences in the probability surfaces obtained by the three methods. We define the distance between the two methods as the  $\ell^2$ -matrix norm, i.e.:

$$\|A\|_2 = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2},$$

which we apply to the differences between the two probability surfaces.

### III. SMALL NOISE APPROXIMATION

In the second part of this paper we focus on the limit case when the size  $\Omega$  of the system is large enough and try to obtain the statistical moments of free protein distribution in the closed form for this case. The noise in free protein distribution has two sources; the first one is the reversible

Time	Gill-ODE	Gill-Quasi	ODE-Quasi
0	0	0	0
1	0.02	0.0111	0.0219
2	0.0058	0.0055	0.0071
3	0.0027	0.0035	0.0031
4	0.0022	0.0034	0.0027
5	0.0066	0.0024	0.0064
10	0.0044	0.0018	0.0043
20	0.0016	0.0019	0.00061
50	0.0014	0.0016	0.00061
100	0.0018	0.0019	0.00055

TABLE III: Distances between various numerical implementations of the probability surface.

association of protein and binding site and the second one is new protein production/degradation. Let us first concentrate on the reversible association  $X_f + Y_f \xrightleftharpoons[k_-]{k_+} C$ . In the course of this reaction, we can treat the values of  $x$  (total protein concentration) and  $y$  (total binding sites concentration)

as constants. We would like to underscore the fact that lowercase letters denote the concentration of a given reactant denoted by corresponding uppercase letter.

A. Deterministic case

If we assume large protein numbers, we can obtain the mean of the distribution by finding the stationary state of deterministic reaction kinetics. This approach is used since the inception of chemical kinetics [19]. We know that in the stationary state (see e.g. [21]), we have  $n \cdot y_f = c$ , where the concentrations  $x_f$  of free protein,  $y_f$  of free binding site, and  $c$  of their complex are measured in units of the dissociation constant. If we combine this equation with conservation laws,  $n + c = x$  for the total protein and  $y_f + c = y$  for the total binding-site concentration, we will obtain a system of three equations with three unknown variables. This system has a single non-negative solution, which we refer to as  $\bar{n}, \bar{y}_f, \bar{c}$ . We can obtain it by solving the quadratic equation

$$\bar{c}^2 + \bar{c}(-x - y - 1) + xy = 0,$$

which gives us the solution in the form

$$\begin{aligned} \bar{c}(x, y) &= \frac{x+y+1 - \sqrt{x^2+y^2+1+2x+2y-2xy}}{2}, \\ \bar{n}(x, y) &= x - \bar{c}, \\ \bar{y}_f(x, y) &= y - \bar{c}. \end{aligned} \tag{13}$$

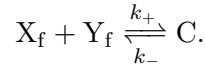
For the sake of parsimony, we omit the explicit notation of dependence of  $\bar{c}, \bar{n}, \bar{y}_f$  on the total concentration of  $x$  and  $y$  in all non-ambiguous cases.

B. Stochastic component

As we mentioned before, we were not able to obtain the distribution of free protein in a closed form; therefore we have to choose an alternative approach. The number of free protein is affected by binding and unbinding reactions as well as by the production/degradation of new proteins. As the timescales of these two processes are diametrically different, we can treat them separately. Let us

divide the stochastic component into the part corresponding to binding/unbinding and the part corresponding to total protein production/degradation.

1) *Constant X (total protein count)*: First we assume that the number of total protein remains constant, focusing only on the binding/unbinding reactions:



Using the shift-operator notation from [29] ( $\mathbb{E}^i f(x) = f(x+i)$ ), we can write down the Master equation in the steady state for this reaction as

$$0 = (\mathbb{E} - 1)k_- C P_C + (\mathbb{E}^{-1} - 1)k_+ N Y_f P_C, \tag{14}$$

where  $P_C$  denotes probability mass function of  $C$ ,  $X_f = X - C$  and  $Y_f = Y - C$ , and the shifting is executed with respect to  $C$ . We further assume that size  $\Omega$  of the system is large, whereby we identify the system size with the dissociation constant  $\Omega = \frac{k_-}{k_+}$ . This will naturally lead to the macroscopic dissociation constant being equal to unity after the system-size reduction is performed. Now we can write down the relation between the total species count (uppercase letters in italics) and their corresponding concentrations (lowercase letters):  $N = \Omega \cdot n$ ,  $Y_f = \Omega \cdot y_f$  and  $C = \Omega \cdot c$ . Inserting this scaling into (14) yields

$$0 = (\mathbb{E} - 1)c P_c + (\mathbb{E}^{-1} - 1)ny_f P_c, \tag{15}$$

Using the Taylor expansion together with the fact that  $C$  is large (i.e.  $\Omega^{-1}$  is small), the shift operators can formally be expanded as follows:

$$\begin{aligned} \mathbb{E} = e^{\partial_c} &= e^{\Omega^{-1}\partial_c} = 1 + \Omega^{-1}\partial_c + \frac{\Omega^{-2}}{2}\partial_c^2 + \dots \\ \mathbb{E}^{-1} = e^{-\partial_c} &= e^{-\Omega^{-1}\partial_c} = 1 - \Omega^{-1}\partial_c + \frac{\Omega^{-2}}{2}\partial_c^2 - \dots \end{aligned} \tag{16}$$

Inserting (16) into (15) and neglecting terms of order of  $\Omega^{-1}$  we obtain

$$\partial_c \left[ (c - ny_f)P_c + \frac{\Omega^{-1}}{2}\partial_c((c + ny_f)P_c) \right] = 0. \tag{17}$$

If we now integrate this equation with respect to  $c$  (with zero-flux conditions) we come to

$$(c - ny_f)P_c + \frac{\Omega^{-1}}{2} \partial_c ((c + ny_f)P_c) = 0.$$

We denote by terms  $A = (c - ny_f)$ ,  $B = c + ny_f$  the terms which appear in the above equation. Since  $\Omega^{-1} \ll 1$ , we can use a small-noise approximation (variance of  $c$  is of order  $\Omega^{-1}$ ), which is based on Taylor-expanding  $A$  and  $B$  around the deterministic mean value:

$$\begin{aligned} A(c) &= c - (x - c)(y - c) \\ A'(c) &= 1 + x + y - 2c = 1 + n + y_f \\ A(c) &\simeq A(\bar{c}) + A'(\bar{c})(c - \bar{c}) = A'(\bar{c})(c - \bar{c}) \\ &= (1 + \bar{n} + \bar{y}_f)(c - \bar{c}) \\ B(c) &\simeq B(\bar{c}) \\ &= \bar{c} + \bar{n}\bar{y}_f = 2\bar{n}\bar{y}_f. \end{aligned} \tag{18}$$

Substituting approximations (18) into (17) yields

$$(1 + \bar{n} + \bar{y}_f)(c - \bar{c})P_c + \Omega^{-1}\bar{n}\bar{y}_f\partial_c P_c = 0,$$

i.e.

$$\partial_c P_c = -\frac{1 + \bar{n} + \bar{y}_f}{\Omega^{-1}\bar{n}\bar{y}_f} \cdot (c - \bar{c})P_c \tag{19}$$

This equation can be easily solved by separation of variables, which yields

$$P_c \propto \exp\left(\frac{1}{2} \cdot \frac{(1 + \bar{n} + \bar{y}_f)}{\Omega^{-1}\bar{n}\bar{y}_f} \cdot (c - \bar{c})^2\right).$$

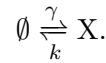
We have found an approximate distribution of  $c$ ; to put it in other words, we can say that

$$c \sim \mathcal{N}\left(\bar{c}, \frac{\Omega^{-1}\bar{n}\bar{y}_f}{1 + \bar{n} + \bar{y}_f}\right) \tag{20}$$

as  $\Omega \rightarrow \infty$ . As  $n$  and  $y_f$  can be computed directly from  $c$  as  $x - c$ , resp.  $y - c$ , they have the same variance as  $c$ .

2) *Fluctuating  $X$  (total protein count)*: In the second part, instead of taking  $X$  to be a constant, we assume that total protein count fluctuates due to creation of new protein molecules and decay of old ones, which is denoted by a reversible pair of

chemical reactions



We already know that these processes operate on a much slower timescale than the association/dissociation processes. Furthermore, we have already found the steady-state distribution of  $X$  (from II-B); the result is that it is Poissonian with  $\langle X \rangle = \text{Var}(X) = \frac{k}{\gamma}$ . We use the system-size scaling  $\frac{k}{\gamma} = \langle x \rangle \cdot \Omega$ . As  $\Omega \rightarrow \infty$ , we can approximate the Poissonian distribution by a small-noise Gaussian distribution

$$x = \frac{X}{\Omega} \sim \mathcal{N}(\langle x \rangle, \Omega^{-1} \cdot \langle x \rangle). \tag{21}$$

In order to calculate the statistics of  $n$  (free protein concentration) we have to combine the results (20) and (21) given that  $n$  is expressed in terms of slowly fluctuating  $x$  (total protein concentration). The free protein concentration also naturally depends on  $y$  (total concentration of binding sites), but as it is a constant through the whole process, we can neglect it from our notation for the sake of simplicity. We can use the variance decomposition theorem in order to solve this problem (see [30]). It expresses the total (unconditional) variance as the sum of the expectation of conditional variances and the variances of conditional expectations, in our case, we can write

$$\text{Var}(n) = E(\text{Var}(n|x)) + \text{Var}(E(n|x)). \tag{22}$$

We already know (from (20)) the solution for  $E(n|x)$  and  $\text{Var}(n|x)$ ; we utilize the fact that  $x$  fluctuates much more slowly than  $n$  and thus we can obtain results for  $n$  subject to constant  $x$ . Using the large  $\Omega$  approximation, we can write

$$\begin{aligned} E(n|x) &= \bar{n} = \bar{n}(x) \\ \text{Var}(n|x) &= \frac{\Omega^{-1}\bar{n}(x)\bar{y}_f(x)}{1 + \bar{n}(x) + \bar{y}_f(x)}. \end{aligned} \tag{23}$$

Since the variance of  $x$  is of order  $\Omega^{-1}$  and we neglect all terms of order higher than  $\Omega^{-1}$ , we can use the approximations  $\bar{n}(x) \simeq \bar{n}(\langle x \rangle)$  and  $\bar{y}_f(x) \simeq \bar{y}_f(\langle x \rangle)$  in the formula for the conditional

variance, which yields

$$E(\text{Var}(n|x)) \simeq \text{Var}(n|\langle x \rangle) = \frac{\Omega^{-1} \cdot \bar{n}\bar{y}_f}{1 + \bar{n} + \bar{y}_f}, \quad (24)$$

evaluated at  $\langle x \rangle$  (mean of total protein concentration) and  $y$  (constant concentration of binding sites). The final term left to calculate is  $\text{Var}(\bar{n}(x))$ . If we used the approximation  $\bar{n}(x) = \bar{n}(\langle x \rangle)$ , we would end with zero, which would incorrectly neglect all the variance. Therefore we also include terms of next order by using the linear variance approximation, i.e. by writing

$$\bar{n}(x) \simeq \bar{n}(\langle x \rangle) + \frac{d\bar{n}}{dx} \cdot (x - \langle x \rangle).$$

From this expression we can find that

$$\begin{aligned} \text{Var}(\bar{n}(x)) &\simeq \left( \frac{d\bar{n}}{dx} \right)_{|x=\langle x \rangle}^2 \cdot \text{Var}(x) \\ &= \left( \frac{d\bar{n}}{dx} \right)_{|x=\langle x \rangle}^2 \cdot \Omega^{-1} \langle x \rangle \end{aligned} \quad (25)$$

Hence, the last item to calculate is the derivation of  $\bar{n}$  with respect to  $x$ . In order to find it, let us differentiate the equations

$$\bar{n}\bar{y}_f = \bar{c}, \quad \bar{n} + \bar{c} = x, \quad \bar{y}_f + \bar{c} = y$$

that define the dependence of  $\bar{n}$ , among others, on  $x$ , with respect to variable  $x$ ; this yields

$$\frac{d\bar{n}}{dx} \cdot \bar{y}_f + \frac{d\bar{y}_f}{dx} \cdot \bar{n} = \frac{d\bar{c}}{dx}, \quad (26)$$

$$\frac{d\bar{n}}{dx} + \frac{d\bar{c}}{dx} = 1, \quad (27)$$

$$\frac{d\bar{y}_f}{dx} + \frac{d\bar{c}}{dx} = 0. \quad (28)$$

Substituting (28) into (26), we eliminate  $\frac{d\bar{y}_f}{dx}$  and obtain

$$\frac{d\bar{n}}{dx} \cdot \bar{y}_f = \frac{d\bar{c}}{dx} \cdot (1 + \bar{n}).$$

Now using equation (27) we get

$$\frac{d\bar{n}}{dx} = \frac{1 + \bar{n}}{1 + \bar{n} + \bar{y}_f}. \quad (29)$$

Thus we are now in a position to find the moments

of  $n$ . We get the mean by substituting  $\langle x \rangle$  into the deterministic results (13) and we get the variance by substituting the partial results (24), (25) and (29) into the variance decomposition theorem (22). For mean we obtained the formula

$$\begin{aligned} E(n) &= \bar{n}(\langle x \rangle) \\ &= \frac{\langle x \rangle - y - 1 + \sqrt{\langle x \rangle^2 + y^2 + 1 + 2\langle x \rangle + 2y - 2\langle x \rangle y}}{2}. \end{aligned} \quad (30)$$

Variance can be expressed in the form

$$\text{Var}(n) = \Omega^{-1} \left( \frac{\bar{n}\bar{y}_f}{1 + \bar{n} + \bar{y}_f} + \left( \frac{1 + \bar{n}}{1 + \bar{n} + \bar{y}_f} \right)^2 \langle x \rangle \right). \quad (31)$$

Combining these two results gives us the possibility to obtain the Fano factor as

$$F = \frac{\text{Var}(N)}{E(N)} \simeq \frac{\text{Var}(\Omega \cdot n)}{E(\Omega \cdot n)} = \Omega \cdot \frac{\text{Var}(n)}{E(n)}. \quad (32)$$

After some simplifying steps we come to the expression

$$F = 1 + \frac{\bar{n}\bar{y}_f(1 + \bar{n})}{(1 + \bar{n} + \bar{y}_f)^2}, \quad (33)$$

where 1 can be interpreted as the Poissonian noise and the residual fraction as an additional non-Poissonian noise present due to the interaction with decoy binding sites.

### C. Numerical simulations

In the current section we perform numerical simulations in order to visualize of our results and to confirm the validity of the approximation scheme presented above.

1) *Fano factor based on system-size approach:* In the first simulation we investigate the relation between  $\langle x \rangle$  and the Fano factor. In the first figure (Figure 5) we plot the dependance of Fano factor on the number of binding sites. We calculate the Fano factor for different values of  $\langle x \rangle$ . All values are meant as the concentrations; therefore to obtain the corresponding number of molecules we have to multiply them by  $\Omega$ . We clearly see that for larger values of  $\langle x \rangle$  we are able to reach larger values of  $F$ . An interesting point to observe in

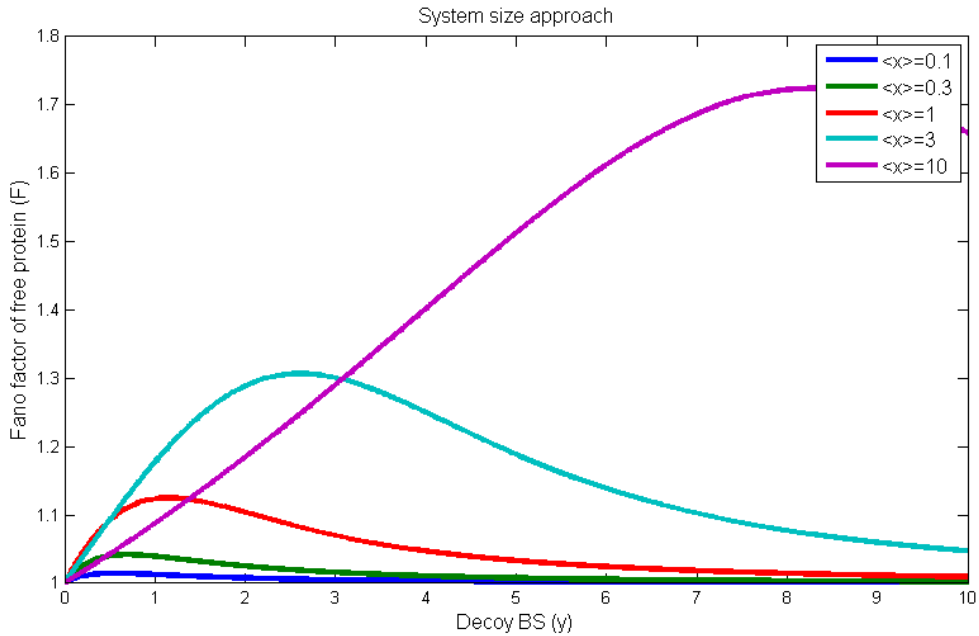


Fig. 5: Fano factor for large system size.

the graphs is the slope near  $y = 0$ . We see that for small values of  $\langle x \rangle$  the slope increases with increasing  $\langle x \rangle$ , but for larger values of  $\langle x \rangle$  the slope starts to decrease. In order to investigate this phenomenon into further depth, let us calculate Taylor expansion near  $F(x, 0)$ :

$$F(x, \Delta y) = 1 + \Delta y \cdot \frac{x}{(1+x)^2}.$$

We can find the maximum of this value by differentiating

$$\frac{\partial F(x, \Delta y)}{\partial x} = \Delta y \cdot \frac{1-x}{(1+x)^3},$$

which confirms that the maximal slope is obtained for  $\langle x \rangle = 1$ .

In Figure 6 we take the ratio  $y/\langle x \rangle$  (number of BS divided by mean number of total protein) as the independent variable and plot Fano factor again for several different choices of  $\langle x \rangle$ .

We can see that for larger values of  $\langle x \rangle$  the maximum of Fano factor is achieved near  $y/\langle x \rangle = 1$ . In order to investigate this phenomenon further, let us express the Fano factor in terms of new variables  $a = \frac{y}{\langle x \rangle}$  and  $b = \frac{1}{x}$ . This substitution

yields

$$F(a, b) = 1 + \frac{1}{2} \cdot \frac{a(1-a-b+\sqrt{\star})}{\star}, \quad (34)$$

$$\star = a^2 + b^2 + 1 + 2b + 2a - 2ab.$$

The problem of finding the maximum of Fano factor with respect to  $a$  is equivalent to finding the solution of  $\frac{\partial F(a,b)}{\partial a} = 0$ , with  $F(a, b)$  from (34). This yields a very complex implicit function. As we are interested in cases with large values of  $\langle x \rangle$ , we want to investigate its solution for small values of  $b$ . As we are unable to express the value of  $b$  in terms of  $a$  from the implicit function in reasonable way, we have to settle for numerical and graphical solution for this equation, which is provided in Figure 7. This verifies our hypothesis that for large  $\langle x \rangle$  the maximal Fano factor is achieved near the point where  $y = \langle x \rangle$ .

2) *Quality of system size approach:* In this section we investigate whether the linear noise approximation is consistent with the stochastic

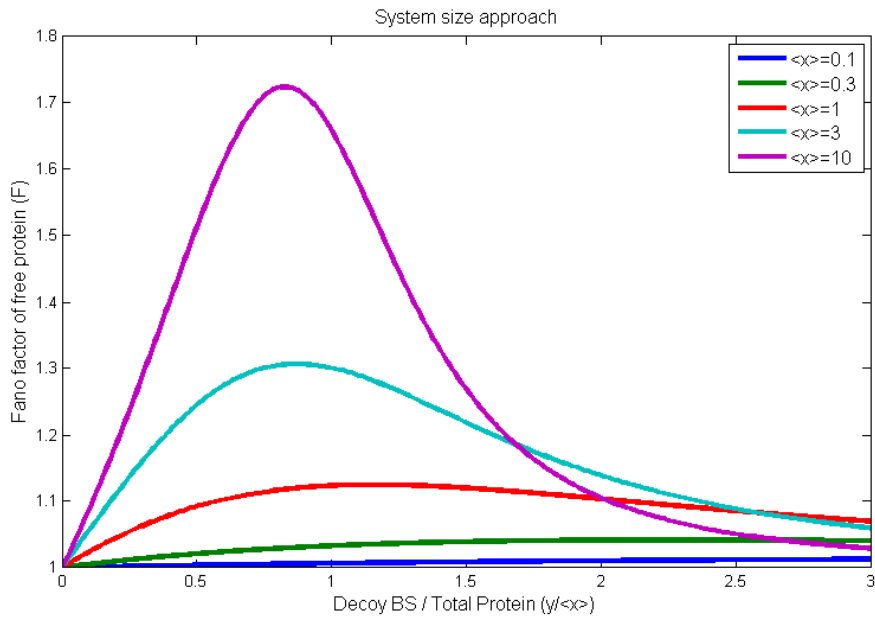


Fig. 6: Fano factor for large system size.

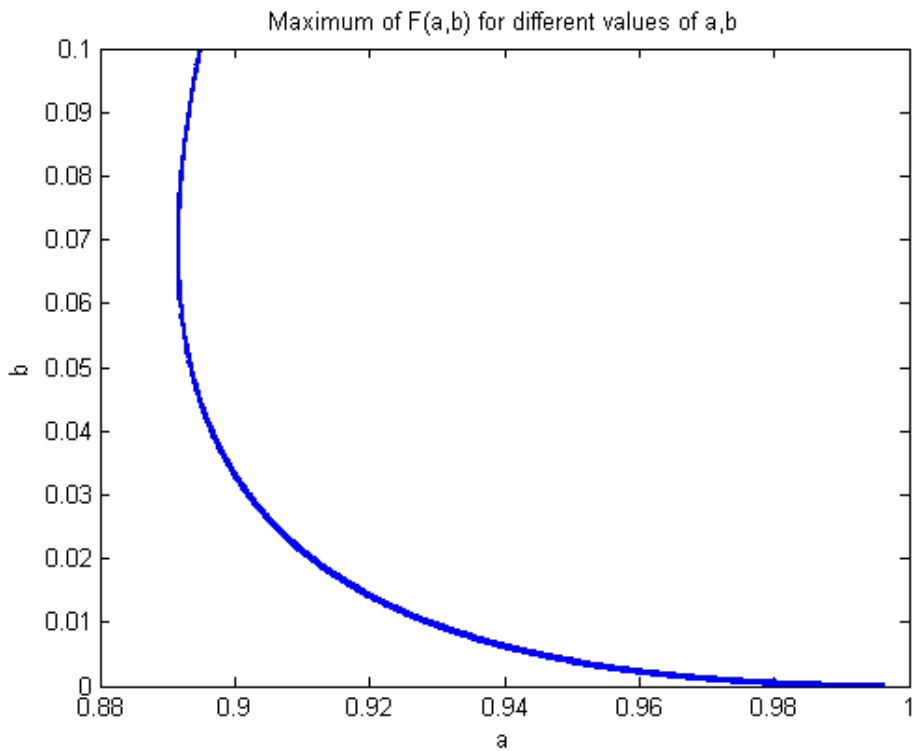


Fig. 7: Maximum of Fano factor.

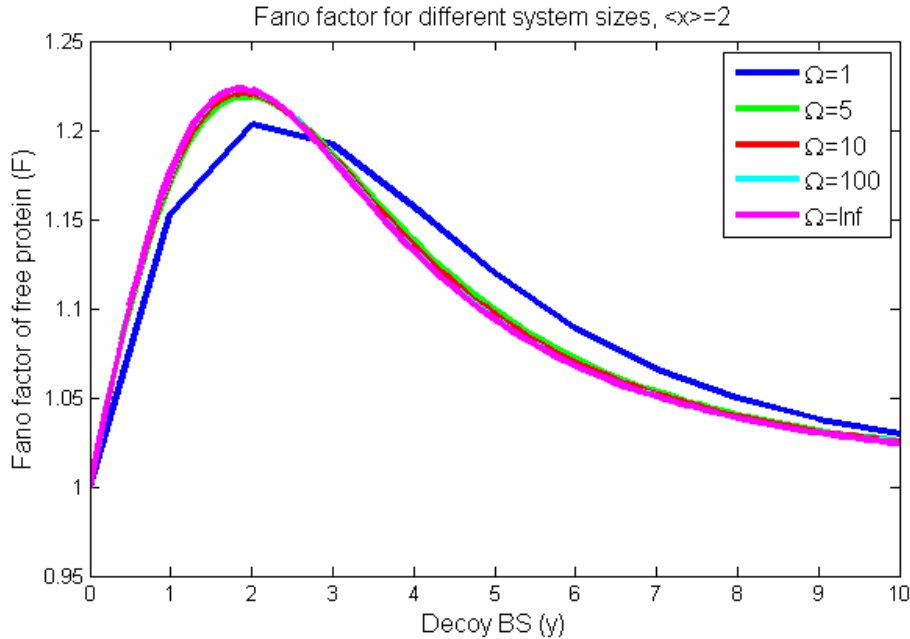


Fig. 8: Fano factor for different system sizes.

results for quasi-steady state:

$$P_N = \sum_{X=N}^{N+Y} P_X \cdot \frac{k_b^N}{N!(X-N)!(Y-X+N)!} \cdot \left( \sum_{i=\max\{0, X-Y\}}^X \frac{k_b^i}{(X-i)!(Y-X+i)!i!} \right)^{-1} \quad (35)$$

Using this formula we can obtain probability distribution of free protein and therefore calculate its Fano factor. But it is not that straightforward. In order to get the results for big values of  $N$ , we have to rewrite the sum in a way that we get around the problems of calculating big factorials and multiplying number close to 0 with an extremely large number. For this purpose, (9) can be rewritten as:

$$P_N = \sum_{X=N}^{N+Y} P_X \cdot \left( \sum_i k_b^{i-N} \cdot \frac{N!(X-N)!(Y-X+N)!}{i!(X-i)!(Y-X+i)!} \right)^{-1} \quad (36)$$

Now we can proceed with the calculations:

we use 1, 5, 10 and 100 as the value for  $\Omega$  and compare the numerically calculated exact Fano factor with the LNA expression for Fano factor (33) for values of  $y$  between 0 and 10. In the case where  $\Omega = 1$  we can only use integer values of  $y$ , for  $\Omega = 5$  we can use multiples of 0.2 for  $y$ , in other cases we use multiples of 0.1 for  $y$  in order to plot the graphs. We use the same setup for the different values of  $\langle x \rangle$ , in particular 0.1, 0.5, 1 and 2. Results for scenario with  $\langle x \rangle = 2$  is shown in the Figure 8; results corresponding to other values of  $\langle x \rangle$  are analogous. Furthermore, in Table IV we present the sum of squared residuals in the points  $y = 1, \dots, 10$ . (If  $y = 0$ , then  $F = 1$  always, so we can omit this point.)

We see that the differences between approximations increase with  $\langle x \rangle$ , which is no surprise as Fano factor achieves higher values in such cases.

#### IV. CONCLUSION

We introduced three methods to obtain free protein distribution of our gene expression model. In addition to stochastic simulation and numerical simulation of the Master equation, we also employed singular-perturbation reduction tech-

$\Omega$	$\langle x \rangle = 0.1$	$\langle x \rangle = 0.5$	$\langle x \rangle = 1$	$\langle x \rangle = 2$
1	$1.8 \cdot 10^{-5}$	$2.2 \cdot 10^{-4}$	$8.0 \cdot 10^{-4}$	0.0032
5	$5.6 \cdot 10^{-7}$	$8.9 \cdot 10^{-6}$	$3.3 \cdot 10^{-5}$	$1.4 \cdot 10^{-4}$
10	$1.4 \cdot 10^{-7}$	$2.2 \cdot 10^{-6}$	$8.4 \cdot 10^{-6}$	$3.6 \cdot 10^{-5}$
100	$1.3 \cdot 10^{-9}$	$2.2 \cdot 10^{-8}$	$8.5 \cdot 10^{-8}$	$3.6 \cdot 10^{-7}$

TABLE IV: Sum of squares from LNA.

niques to obtain a quasi-steady-state approximation, which was helpful in finding a relatively simple explicit formula for the free protein distribution. Using this formula we were able to observe its statistical moments for many different input parameters. Very interesting results were obtained for the Fano factor, which substantially differed from Poissonian case. In the second part of the paper we considered the case of large system size, using the dissociation constant as the measure of size. This approach yielded a tractable expression for the Fano factor of free protein distribution. Through numerical simulations we showed that this solution is consistent with results from the first part of the paper. While we have employed our methodologies to explore the properties of a relatively simple model, we expect that the same approaches can yield valuable insights in stochastic gene-expression models in particular as well as problems of mathematical biology more widely.

#### ACKNOWLEDGMENT

PB and MH are supported by the Slovak Research and Development Agency grant APVV-14-0378. PB acknowledges the support from the VEGA grant agency (grant no. 1/0319/15). PB would like to thank A. Singh for helpful conversation about questions related to the research presented herein.

#### REFERENCES

[1] Alon, U.: *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall/CRC, 2007.  
dx.doi.org/10.1063/1.2754606

[2] Bokes, P., Singh, A.: *Protein copy number distributions for a self-regulating gene in the presence of decoy binding sites*, PLoS ONE Vol. 10, No. 3: 1-19, 2015.  
dx.doi.org/10.1371/journal.pone.0120555

[3] Burger, A., Walczak, A.M., Wolynes, P.G.: *Abduction and asylum in the lives of transcription factors*, P. Natl. Acad. Sci. USA Vol. 107: 4016-4021, 2010.  
dx.doi.org/10.1073/pnas.0915138107

[4] Burger, A., Walczak, A.M., Wolynes, P.G.: *Influence of decoys on the noise and dynamics of gene expression*, Physical Review E Vol. 86(4): 041920, 2012.  
dx.doi.org/10.1103/physreve.86.041920

[5] Cao, Y., Gillespie, D.T., Petzold, L.R.: *The slow-scale stochastic simulation algorithm*, Journal of Chemical Physics, 122(1): 014116, 2005.  
dx.doi.org/10.1063/1.1824902

[6] Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S.: *Stochastic gene expression in a single cell*, Science Vol. 297: 1183-1186, 2002.  
dx.doi.org/10.1126/science.1070919

[7] Erban, R., Chapman, S.J., Maini, P.K.: *A practical guide to stochastic simulations of reaction-diffusion processes*, arXiv:0704.1908, 2007.

[8] Fritz, J.: *Partial differential equations (4th ed.)*, Springer, 1991.

[9] Ghaemi, R., Del Vecchio, D.: *Stochastic analysis of retroactivity in transcriptional networks through singular perturbation*, American Control Conference: 2731-6, 2012.  
dx.doi.org/10.1109/acc.2012.6315645

[10] Gillespie, D.T.: *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions*, Journal of Computational Physics 22 (4): 403434, 1976.  
dx.doi.org/10.1016/0021-9991(76)90041-3

[11] Gillespie, D.T.: *Exact Stochastic Simulation of Coupled Chemical Reaction*, The Journal of Physical Chemistry 81 (25): 23402361, 1977.  
dx.doi.org/10.1021/j100540a008

[12] Kang, H-W., Kurtz, T.G.: *Separation of time-scales and model reduction for stochastic reaction networks*, The Annals of Applied Probability Vol. 23, No. 2, 529-583, 2013.  
dx.doi.org/10.1214/12-AAP841

[13] Kang, H-W., Kurtz, T.G., Popovic, L.: *Central limit theorems and diffusion approximations for multiscale Markov chain models*, The Annals of Applied Probab-



- bility Vol. 24, No. 2, 721-759, 2014.  
dx.doi.org/10.1214/13-AAP934
- [14] Kim, J.K., Sontag, E.D.: *Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation*, PLOS Computational Biology Vol.13(6): e1005571, 2017.  
dx.doi.org/10.1371/journal.pcbi.1005571
- [15] Knuth, D.E., Graham R.L., Patashnik O.: *Concrete mathematics*, Addison Wesley, 1989.
- [16] Laurenzi, I.J.: *An analytical solution of the stochastic master equation for the reversible bimolecular reaction kinetics*, The Journal of Chemical Physics 113: 3315-3322, 2000.  
dx.doi.org/10.1063/1.1287273
- [17] Lee, T.H., Maheshri, N.: *A regulatory role for repeated decoy transcription factor binding sites in target gene expression*, Mol. Syst. Biol. Vol. 8: 576, 2012.  
dx.doi.org/10.1038/msb.2012.7
- [18] McAdams, H.H., Arkin, A.: *Stochastic mechanisms in gene expression*, P. Natl. Acad. Sci. USA Vol. 94: 814-819, 1997.  
dx.doi.org/10.1073/pnas.94.3.814
- [19] Michaelis, L., Menten, M.L.: *Die Kinetik der Invertinwirkung*, Biochem Z 49: 333369, 1913.
- [20] Munsky, B., Khammash, M.: *The finite state projection algorithm for the solution of the chemical master equation*, The Journal of Chemical Physics 124: 044104, 2006.  
dx.doi.org/10.1063/1.2145882
- [21] Murray, J.D.: *Mathematical biology: An introduction*, Springer, 2002.
- [22] Feigelman, J., Marr, C., Popovic, N.: *A Case Study on the Use of Scale Separation-Based Analytic Propagators for Parameter Inference in Stochastic Gene Regulation*, J. Coupled Syst. Multiscale Dyn. 3(2): 164-173, 2015.  
dx.doi.org/10.1166/jcsmd.2015.1074
- [23] Popovic, N., Marr, C., Swain, P.S.: *A geometric analysis of fast-slow models for stochastic gene expression*, Journal of Mathematical Biology 72(1): 87-122, 2016.  
dx.doi.org/10.1007/s00285-015-0876-1
- [24] Schauer, M., Heinrich, R.: *Quasi-steady-state approximation in the Mathematical Modeling of Biochemical Reaction Networks*, Mathematical Biosciences Vol. 65: 155-170, 1983.  
dx.doi.org/10.1016/0025-5564(83)90058-5
- [25] Shahrezaei, V., Swain, P.S.: *Analytical distributions for stochastic gene expression*, P. Natl. Acad. Sci. USA Vol. 105: 1725617261, 2008.  
dx.doi.org/10.1073/pnas.0803850105
- [26] Soltani, M., et al.: *Nonspecific transcription factor binding can reduce noise in the expression of downstream proteins*, Physical Biology 12: 055002, 2015.  
dx.doi.org/10.1088/1478-3975/12/5/055002
- [27] Stiefenhofer, M.: *Quasi-steady-state approximation for chemical reactional networks*, Journal of Mathematical Biology 36: 593-609, 1998.  
dx.doi.org/10.1007/s002850050116
- [28] Taniguchi, Y., et al.: *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells*, Science Vol. 329: 533-538, 2010.  
dx.doi.org/10.2142/biophys.51.136
- [29] Van Kampen N.G.: *Stochastic processes in physics and chemistry*, Elsevier Science B.V., 1992.  
dx.doi.org/10.1016/b978-0-444-52965-7.x5000-4
- [30] Weiss, N.A.: *A Course in Probability*, AddisonWesley, 2005.
- [31] Yu, J., Xiao, J., Ren, X., Lao, K., Xie, X.S.: *Probing gene expression in live cells, one protein molecule at a time*, Science Vol. 311: 1600-1603, 2006.  
dx.doi.org/10.1126/science.1119623