CONDITIONS FOR EQUIVALENCE OF STOCHASTIC AND DETERMINISTIC GENE TRANSCRIPTION MODELS

Jaroslaw Smieja
Institute of Automatic Control
Silesian University of Technology
Akademicka 16, 44-100 Gliwice, Poland
email: Jaroslaw.Smieja@polsl.pl

ABSTRACT
The paper deals with two different approaches of modeling gene transcription in eukaryotic cells: stochastic and deterministic. The analysis is based on the assumption that dynamics of the transcription factor concentration is known from experimental work, first increasing and then decreasing, without oscillations. The expected value of transcription rate in a stochastic case is compared to deterministic transcription rate. Additionally, the dynamics of mRNA resulting from both models is analyzed. It appears that, despite claims in broad literature devoted to the topic, the deterministic description is a good approximation of an actual stochastic process.

KEY WORDS
Systems biology, transcription modeling

1 Introduction
Following rapid developments in new experimental techniques, mathematical modeling of regulatory pathways that control intracellular biological and chemical processes is gaining increasing interest in the biomedical research [1, 2]. Analysis of biological data has led to much better understanding of the nature of intracellular processes.

Contrary to standard approaches for identification of processes and their parameters, used e.g. in control theory and its applications, models of the pathways cannot be built as input-output models. They must relate directly to biochemical processes involved in the regulatory networks. There are two main reasons for such approach. First, to uncover new mechanisms regulating intracellular processes, they must be directly described. Second, these mechanisms vary from one cell type to another. If one chose an input-output representation, it had to be built from scratch for each cell type and it would be impossible to use knowledge gained in other experiments.

Various processes involved in any signaling pathway exhibit clearly stochastic character, since most of them require collision and subsequent interaction of at least two molecules that move randomly in intracellular environment. Therefore, stochastic approach is very often recommended as the right tool for pathway modeling and analysis (see e.g., [3, 4, 5] and references therein). On the other hand, the number of just one species of interacting molecules in a single cell can reach hundreds of thousands [6], justifying the deterministic approach, based on the law of mass action. There are two basic exceptions: 1) ligand binding to the receptor on a cell surface (either small concentration of ligands or small number of the receptors could be the case) and 2) individual interactions of transcription factors and their corresponding promoter regions in DNA. This paper is focused on the latter problem.

In most cases any gene under consideration in a given pathway is present only in a few copies (usually two) in a cell. Therefore, the assumption underlying the law of mass action is not satisfied. However, as proved in the following sections, the deterministic approach based on this law can still be used, at least in some cases. It has been shown that a single event of binding of an active transcription factor (TF) to the promoter region can result in a burst of mRNA level and, consequently, newly synthesized proteins, transcription is a process where stochastic effects are the most apparent [7].

2 Transcription factors (TFs) and gene transcription

There are basically two ways in which transcription process can be explained in terms of TF binding (the complex process of assembling polymerase complex is not the subject of this discussion). In the first one, when the TF is bound to a promoter region of a given gene, it initiates transcription that proceeds until a repressor is bound to a respective regulatory sequence (or actively unbinds the TF from the promoter region). Alternatively, induced gene transcription might consist in frequent, successive binding and unbinding of the TF to the promoter region. Both cases are briefly analyzed in the following sections.

In the deterministic approach the rate of transcription is a function of TF concentration. Stochastic modeling, in turn, considers moments of activator (and/or repressor) binding random variables determining how transcription is switched on and off. The results of these two approaches are compared in the subsequent section, leading to assumptions about applicability of deterministic modeling.

It should be noted that the description given above is...
a simplified view of a real process. In many cases a complex of several TFs must be created on the promoter region to start transcription or stabilize the required spatial structure of DNA. Such complexes are formed by subsequent binding of cofactors in a determined order. Quite often only the first of them is a known TF, while the others are non-specific members of transcriptional machinery and their concentration can be assumed constant. Then, the simplest deterministic approach, tying transcription rate to one known TF, is not sufficient to model transcription process, even if the model is used to estimate average levels of mRNA in a homogeneous population of cells. Once again, one could take a stochastic approach to solve such modeling problem, since the process involves small number of molecules if the construction of the complex at the particular place (promoter region) is considered. Another way to deal with that would be to introduce a pure time delay into the system of equations, the value of delay being the only additional parameter to be estimated. However, it seems to be an unacceptable simplification. First, such time delay should, in fact, be a random variable rendering the deterministic approach inappropriate. Second, the most plausible model for the process of transcription should allow for changes in transcription rate other than simple on/off switch, imitating a situation in which forming the complex on the promoter region increases the transcription rate that reaches its maximum when the complex is completed. Therefore, introduction of additional dynamical elements seems to be a reasonable solution in this case.

Importance of stochastic effects on gene transcription depends on two factors:

1. Stability of DNA binding by TFs
2. The number of cofactors needed to start the transcription.

3 Modeling gene transcription

3.1 A deterministic model

Let us denote TF (activator) concentration, transcript concentration, and transcript degradation constant by \( (TF) \), \((mRNA)\) and \(k_{\text{deg}}\), respectively. Then, the equation describing transcript dynamics in the deterministic model is as follows:

\[
\frac{d(mRNA)}{dt} = -k_{\text{deg}}(mRNA) + f(t) \tag{1}
\]

where the first term represents transcript degradation, and the second - mRNA production (called from this point a transcription rate) for a gene activated in a given signaling pathway.

Gene transcription considered here can be either constitutive, or induced. Constitutive transcription is incorporated in the deterministic models simply by adding a positive constant to the right hand side of equation (1). It contributes to non-zero initial level of mRNA in the model and provides a return to a normal steady state after disappearance of active TF or turning on repressor activity. For the sake of simplicity, and easier comparison with the stochastic model, in the analysis that follows it will be assumed that it is equal to zero. Such assumption implies that the genes are silent in unstimulated cells. However, the conclusions drawn from this analysis are relevant also in the case when there is a low level of constitutive transcription, and much higher level of induced transcription. In fact, many deterministic models include terms describing both types of mRNA production. Although this leads to an error in calculations - one cannot add both transcription rates - it is assumed that due to much larger value of the induced type, it is small. In the case of stochastic modeling, these two types of transcription are distinguished by different probability distribution functions of binding/unbinding events.

Following the explanation given above, when comparing stochastic and deterministic models, only induced transcription will be taken into account. Then, the function \( f(t) \) is usually assumed to be of Michaelis-Menten type:

\[
f(t) = \frac{k_1 \cdot (TF)}{k_{\text{MM}} + (TF)}, \tag{2}
\]

where \(k_1\) and \(k_{\text{MM}}\) are constant parameters, or it is a linear function of the concentration of an active transcription factor

\[
f(t) = k_2 \cdot (TF) \tag{3}
\]

with \(k_2\) a constant parameter. In both cases, the parameters should be identified basing on experimental measurements of \( (TF) \). However, it should be noted that there is a biochemical constraint imposed on the maximum transcription rate, determined by how fast the polymerase can move along DNA and what is the minimum distance between two subsequent polymerase complexes transcribing the same gene copy. According to [8] the maximum transcription speed is approximately 40 nucleotides per second. Taking into account that the single polymerase covers about 2030 nucleotides of the DNA strand, it is reasonable to assume that the minimum distance between two polymerases should be approximately 250 nucleotides [9]. Then, the transcription rate is limited by the value of 0.16 mRNA molecule per second, that can be transformed into appropriate \(v_{\text{max}}\) given in molar concentration units for a cell of a given volume. For (2) it is sufficient to set \( k_1 \leq v_{\text{max}}\) to satisfy this constraint. In the case when (3) is applied, one should check if \( k_2 \cdot (TF)_{\text{max}} \leq v_{\text{max}}\), where \((TF)_{\text{max}}\) denotes the maximum concentration of a given TF. Clearly, (3) cannot be used when the model is to describe perturbed processes in which \((TF)\) reaches very high levels. However, for a constant \((TF)\), both formulae are basically identical. If \(k_{\text{MM}} \gg (TF)_{\text{max}}\), (3) is a good approximation of (2). In the following analysis both of them are considered.
3.2 A stochastic model

In a purely stochastic model, the variables correspond to the number of molecules and it is assumed that in a small time interval $dt$ only one event can take place, changing the number of molecules by $+1$ or $-1$ depending on the type of the event (production or degradation, respectively). Most often a mixed models are used, where degradation is assumed to follow the law of mass action and the transcription process is of a stochastic nature. Then, the change of mRNA amount is given by:

$$\frac{d(mRNA)}{dt} = -k_{deg}(mRNA) + v_s \cdot g(t)$$

(4)

where $v_s$ is a transcription rate, $g(t)$ is a binary function equal to 1 when the gene is transcribed and 0 otherwise.

Though it has been postulated that $v_s = v_{max}$ [7], it seems more reasonable to assume that $v_s \leq v_{max}$, particularly in eukaryotic cells. Otherwise, one would assume always reaching maximum transcription rate. However, if long genes were involved, it would mean unfolding them along their whole length, which seems unlikely.

To prove applicability of a deterministic approach, one has to show that the results obtained using a deterministic model correspond to the expected value of mRNA concentration (calculated directly from expected number of mRNA molecules) in a stochastic model. This can be done in one of the following ways:

1. numerically - generating random moments of time corresponding to activator and repressor binding, numerical simulation for large number of cells and subsequent averaging of results; or

2. analytically or semi-analytically (wherever possible), where the expected value for gene activity $E[g(t)]$ is considered.

Below, the second approach will be employed, utilizing a-priori assumed concentration profiles of an active TF. From (4), $E[g(t)]$ is a solution to the following differential equation:

$$\frac{dE[(mRNA)]}{dt} = -k_{deg}E[(mRNA)] + v_s \cdot E[g(t)]$$

(5)

Since the degradation term is the same in both (1) and (5), it is sufficient to compare the terms describing transcript production - functions $f(t)$ and $E[g(t)]$. As the same qualitative behavior is required for both of them, the functions will be normalized when showing their plots.

4 Analysis of a transcription rate for a typical TF dynamics

Due to regulatory processes, the concentration of a TF in a stimulated cell often increases at first and then decreases.

Such pattern can be expressed as a sum of exponential functions. For example, the time profile of one of the TFs in the signaling pathway analyzed in [10] can be approximated by

$$f(t) = k_0(\exp(-\mu_1 t) - \exp(-\mu_2 t))$$

(6)

where $\mu_1 > \mu_2$ are constant parameters (see Fig. 1)

As shown in the Fig. 1, the level of TF for a gene activated in a given pathway is often characterized by an increasing-decreasing pattern. However, the assumption about a single binding event, taken in the preceding sections, is more appropriate for modeling processes in prokaryotic cells, where repressors of transcription are also more common [11]. In eukaryotic cells, the TFs can bind and unbind to their respective promoter regions, thus regulating the transcriptional process. While this does not change the way the transcription is modeled in the deterministic approach, the calculations in stochastic modeling must be done in a different way.

Let us assume that the TF concentration is defined by
is the solution to the following differential equation:

\[ \frac{d}{dt} P(t) = \frac{1}{\mu} (e^{-\mu t} - e^{-\mu t}) P(t) + \lambda_0 P_1(t) \]

The solution for the average time spent in the bound state is given by the exponential distribution with mean \(1/\mu\). The value of \(\mu\) depends, among others on the specificity of the TF and its affinity to the promoter region. If \(P_0\) and \(P_1\) denote probabilities of being in free and bound states, respectively, the system can be described by the following set of equations:

\[
\begin{align*}
\dot{P}_0(t) &= -k \cdot (e^{\mu_1 t} - e^{\mu_2 t}) P_0(t) + \lambda_0 P_1(t) \\
\dot{P}_1(t) &= k \cdot (e^{\mu_1 t} - e^{\mu_2 t}) P_0(t) - \lambda_0 P_1(t)
\end{align*}
\]

(7)

The expected value for gene activity \(E[g(t)]\) in 5 is the solution to the following differential equation:

\[ \dot{P}_1(t) = k \cdot (e^{\mu_1 t} - e^{\mu_2 t}) (1 - P_1(t)) - \lambda_0 P_1(t) \]  

(8)

The solution is given by

\[ P_1(t) = C(t) \cdot \exp \left\{ k \left( -\frac{1}{\mu_1} e^{-\mu_1 t} + \frac{1}{\mu_2} e^{-\mu_2 t} \right) + \lambda_0 t \right\} \]

(9)

where

\[ C(t) = \int_{-\infty}^{t} \left\{ -\lambda_0 \exp \left( k \left( \frac{1}{\mu_1} e^{-\mu_1 s} - \frac{1}{\mu_2} e^{-\mu_2 s} \right) - \lambda_0 s \right) \right\} ds \]

(10)

It is known that, in a general case, the average gene activity in a stochastic model can be proved to correspond to the transcription rate in the deterministic approach only for \(\lambda_0 \rightarrow \infty\) [12]. To check this convergence in the case discussed in this section, (8) was solved numerically for various values of \(\lambda_0\). Moreover, respective mRNA levels, stemming from (1) and (5) were compared. The results are shown in Figs. 2 - 5. As expected, if the TFs bind and free the promoter region with a relatively high frequency, determined by high probability of unbinding, the deterministic model yields similar results as the stochastic one. Moreover, the maximum value of \(P_1(t)\) (not shown in the plot, as the results are normalized there) does not exceed 0.1. In the simulation of a a pathway presented in the next chapter, the maximum transcription rate was also more than 10 times lower than the maximum biological value. It can be concluded, therefore, that the linear approximation of the transcription rate is also a reasonable approximation of the average value obtained in the stochastic approach, as long as the mean value of the distribution of unbinding moment is much smaller than the parameters of (6) (less than 1 minute vs. 30 and 60 minutes, respectively). An additional, implicit condition is that in the transcription process, the promoter remains mostly free.

If the transcription rate observed in experiments is approaching its maximum possible value, the deterministic
Figure 4. Influence of the parameters $k$ and $\mu$ on the probability of the promoter region being occupied by a TF - (a) and (b), respectively, and on the transcription rate (normalized value) - (a) and (b), respectively. The arrows show increasing value of a given parameter.

Figure 5. Comparison of (a) transcription rate and (b) resulting mRNA level (for an assumed half-life time of 30 minutes) in deterministic and stochastic models. In the stochastic model, the average time of the promoter region being occupied was 1 hour, as indicated in the legend.

linear model is no longer appropriate. One of the main reasons is that if the level of a given TF was increased, the model would yield biologically unrealistic transcription rate. Therefore, the Michaelis-Menten relation should be used instead of the linear one. Such model, however, leads to qualitatively different results (Fig. 3). As a result the time course of the mRNA level, indicating the time course of the transcription rate, can provide hints about the right type of a deterministic model. Moreover, high transcription rate implies high probability of the promoter region being occupied by a TF in a stochastic model. Here, in (7), higher values of $P_1(t)$ can result from larger $k$ and/or smaller $\mu$ (Fig. 4). One should also remember that the case when $k$ (7) is increased, is indistinguishable from increasing the value of $k_0$ in (6). It can be noted that if the transcription in the deterministic model is given by Michaelis-Menten kinetics (2), once again it is similar to the results of stochastic model. However, the differences here are larger than in the previous case (Fig. 5).

5 Conclusions

Models of signaling pathways and regulatory networks that control intracellular processes are usually high-dimensional and nonlinear. It is known that in nonlinear systems a deterministic approach usually yields different results than a stochastic one (in terms of an average values of variables). This paper shows that in a particular case of a biological system, when gene transcription is the stochastic
process, a deterministic description provides a reasonable approximation of a transcription process, allowing for subsequent application of control theory methods to analyze dynamics of a whole regulatory network.

Acknowledgements

This work was partially supported by the grant N N514 4119 36 for years 2009-2012.

References


