Reducing Gene Regulatory Networks by Decomposition

Luonan Chen
Osaka Sangyo University
Nakagaito 3-1-1, Daito, Osaka 574-8530, Japan

Ruiqi Wang Kazuyuki Aihara
The University of Tokyo
Komaba 4-6-1, Meguro, Tokyo 153-8505, Japan

1 Introduction

Explicitly considering all variables and chemical reactions in a cell is unrealistic for a gene regulatory network from modeling, analyzing and computing viewpoint. However, in a cell, many different time scales characterize the gene regulatory processes, which can be exploited to reduce the complexity of the mathematical models. For instance, the transcription and translation processes generally evolve on a time scale that is much slower than that of phosphorylation, dimerization or binding reactions. Moreover, in biological systems, there are many subsystems, such as gene regulatory network, protein network or metabolic network, which dynamically interact with each other but also are relatively independent. In this work, we exploit such properties to simplify the gene-protein network, which can be applied to the quantitative simulation of a large cellular system.

2 Gene-Protein Network

Consider a system containing m chemical reactions with n molecular species. Let $z = (z_1, ..., z_n)$ be state of the species, i.e. z_i is the number of the $i-$th species at t, which is a non-negative integer. Define $p(z)$ to be the probability for the state z at t. Then the dynamics of the system is described by a master equation with initial state z_0 at $t = 0$

$$\frac{\partial p(z;t)}{\partial t} = \sum_{k=1}^{m}[w_k(z - r_k)p(z - r_k;t) - w_k(z)p(z;t)] \quad (1)$$

where $r_k = (r_{k,1}, ..., r_{k,n})$ is an integer vector for the change of state, i.e. $r_{k,j}$ is the change in the number of the $j-$th molecule by the $k-$th reaction. $w_k(z)$ is a transition rate (≥ 0) from state z to state $z + r_k$ by the $k-$th reaction, which depends on the cell volume v (scalar) and gene number vector u. u and v are generally all periodic functions, synchronized with cell cycles. Notice that $z - r_k$ should be non-negative although $r_{k,j}$ can be negative.

(1) theoretically provides full information of system performances, but only a few simple cases are amenable to exact solution due to its complexity for a large number of variables. Next we exploit the fast-slow reactions to simplify the master equation.

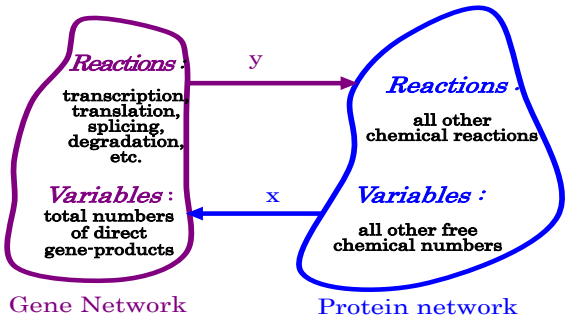

Figure 1. Gene-protein network

3 Decomposition

Although dynamics are intertwined between gene network and protein network, interactions among each network are generally more active than those between them. Such property can be explored to decompose the gene-protein network. As schematically shown in figure 1, we define the gene and protein networks in the following manner.

- Gene network is mainly composed of transcription, translation, splicing, and degradation reactions, with total numbers of direct gene-products, such as protein and RNA numbers as variables.
- Protein network is comprised of all other chemical reactions, such as, phosphorylation, dimerization, binding, enzyme reactions and chemical modifications, with free chemical numbers as variables.

Gene network involves the gene regulatory, and its dynamics are generally slow in contrast to the relatively fast reactions in protein network. In gene network, rather than the free monomers, we adopt total numbers of gene products as variables, which include not only free monomers but also those among all complexes, such as dimers and other multimers. Note that the degradation or depletion reactions of direct gene-products (chemical monomers) are also included in gene network. With such definition, we will show that the gene-protein network can generally be decomposed into gene network and protein network.

Without loss of generality, assume that the first m_0 reactions are in protein network. Rearrange the state variables by $z = (x, y)$ and $r_k = (\phi_k, \theta_k)$, where $x = (x_1, ..., x_{n_x})$,

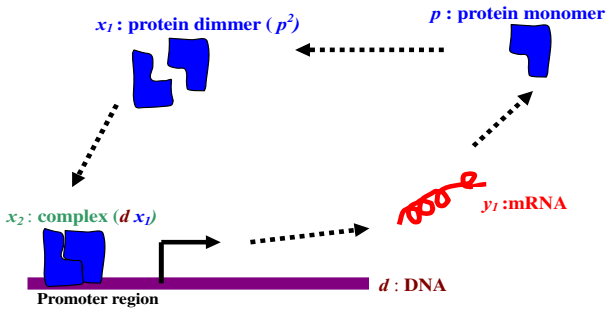

Figure 2. A simple network with one gene.

$y = (y_1, ..., y_{n_y})$. x_i is the number of a molecule synthesized in a fast chemical reaction, and y_i is the total number of a mRNA produced by transcription reaction, or the total number of a protein produced by the translation reaction in gene network. Notice that y_i for a protein is the total number including dimers and other complexes.

Next, we suppress the explicit time dependence of $p(z)$ for readability. Define a marginal function $p(y) = \int p(x, y)dx$, where the integration is simply a summation over all discrete x. $p(x)$ is similarly defined as $p(y)$. Then, the joint probability function is written by the marginal probability and conditional probability as

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x). \quad (2)$$

Example 3.1 *Figure 2 shows a simple gene regulatory network. A gene transcribes into mRNAs, which are further translated into protein monomers. Then the protein monomers are dimerized and act as transcriptional factors to regulate the gene activity by binding to the promoter site. Let numbers of free protein monomer and free DNA are p and d respectively. Define numbers of mRNA, protein dimers, and [protein:DNA] complex to be y_1, x_1 and x_2 respectively. Then the total numbers of protein and DNA are $y_2 = p + 2x_1 + 2x_2$ and $u = d + x_2$.*

The protein dimerization and DNA binding reactions are fast dynamics.

$$p + p \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} x_1 \; ; \; d + x_1 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} x_2. \quad (3)$$

On the other hand, the transcription, translation and degradation reactions are considered as slow dynamics.

$$d \overset{\alpha k_3}{\rightharpoonup} y_1 + d \; ; \; x_2 \overset{k_3}{\rightharpoonup} y_1 + x_2 \quad (4)$$

$$y_1 \overset{k_4}{\rightharpoonup} p + y_1 \; ; \; y_1 \overset{d_1}{\rightharpoonup} \emptyset \; ; \; p \overset{d_2}{\rightharpoonup} \emptyset \quad (5)$$

The transcription reaction occurs with or without x_2 but at different rates. There are five slow reactions.

Let $z = (x, y) = (x_1, x_2, y_1, y_2)$ where $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Therefore, $n = 4$, $m = 9, m_0 = 4$, and $n_x = 2, n_y = 2$. The terms for $k = 1, ..., 4$ in Table 1 correspond to (3), whereas the terms for $k = 5, ..., 9$ are derived from (4)-(5). Then, the master equation (1) is obtained straightforward from Table 1 for Example 3.1.

Table 1. r_k and w_k for Example 3.1

k	r_k = $(\phi_k,$	$\theta_k)$	$w_k(x, y)$
1	1,0,	0,0	$k_1p^2/v = k_1(y_2 - 2x_1 - 2x_2)^2/v$
2	-1,0,	0,0	$k_{-1}x_1$
3	-1,1,	0,0	$k_2x_1d/v = k_2x_1(u - x_2)/v$
4	1,-1,	0,0	$k_{-2}x_2$
5	0,0,	1,0	$\alpha k_3 d = \alpha k_3(u - x_2)$
6	0,0,	1,0	k_3x_2
7	0,0,	0,1	k_4y_1
8	0,0,	-1,0	d_1y_1
9	0,0,	0,-1	$d_2p = d_2(y_2 - 2x_1 - 2x_2)$

where $r_k = (\phi_k, \theta_k) = (\phi_{k,1}, \phi_{k,2}, \theta_{k,1}, \theta_{k,2})$. Exactly $w_1 = k_1p(p-1)/v$. θ_k for all $k = 1, ..., 4$ and ϕ_k for all $k = 5, ..., 9$ are zero vectors.

Clearly, ϕ_k for $k = m_0 + 1, ..., m$ and θ_k for $k = 1, ..., m_0$ are zero vectors. Actually the total numbers of direct gene-products are affected only by transcription, translation and degradation reactions which are all in gene network, whereas other chemical numbers vary only in the protein network although there exist interactions between gene and protein networks. Then the protein network can be described as

$$\frac{\partial p(x|y)}{\partial t} = \sum_{k=1}^{m_0} [w_k(x - \phi_k, y)p(x - \phi_k|y) - w_k(x, y)p(x|y)] \quad (6)$$

Substituting (2) into (1) and summing over all x, we have the evolution equations of the marginal functions for gene network

$$\frac{\partial p(y)}{\partial t} = \sum_{k=m_0+1}^{m} [\bar{w}_k(y - \theta_k)p(y - \theta_k) - \bar{w}_k(y)p(y)] \quad (7)$$

where

$$\bar{w}_k(y) = \sum_x w_k(x, y)p(x|y) \quad (8)$$

is average value conditional to y. $\bar{w}_k(y)$ can be expressed by conditional moments or cumulants of x because $w_k(x, y)$ is generally a polynomial of x and y.

According to (2), clearly we can obtain the dynamics of gene-protein network by (6) and (7), which is much simpler than the original (1).

4 Conclusion

We theoretically provided a general framework to derive gene regulatory networks with stochasticity. We exploit the fast-slow dynamics of biological systems to reduce the dimensionality, and take advantage of special interaction structure of fast-slow variables to simplify the mathematical model, which significantly reduce the complexity of gene networks. The numerical simulation also confirmed the effectiveness of our method, which can be applied to a large-scale quantitative simulation of cellular dynamics.