Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physics Letters A 339 (2005) 1–9
PHYSICS LETTERS A
www.elsevier.com/locate/pla

A stochastic approach to multi-gene expression dynamics

T. Ochiai [*,1], J.C. Nacher [1], T. Akutsu

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji 611-0011, Japan

Received 2 February 2005; received in revised form 24 February 2005; accepted 24 February 2005

Available online 16 March 2005

Communicated by V.M. Agranovich

Abstract

In the last years, tens of thousands gene expression profiles for cells of several organisms have been monitored. Gene expression is a complex transcriptional process where *mRNA* molecules are translated into proteins, which control most of the cell functions. In this process, the correlation among genes is crucial to determine the specific functions of genes. Here, we propose a novel multi-dimensional stochastic approach to deal with the gene correlation phenomena. Interestingly, our stochastic framework suggests that the study of the gene correlation requires only one theoretical assumption—*Markov property*—and the experimental transition probability, which characterizes the gene correlation system. Finally, a gene expression experiment is proposed for future applications of the model.

PACS: 89.75.-k; 87.14.Gg; 87.15.Aa; 87.15.Vv

Keywords: Gene expression; Markov property; Stochastic process

1. Introduction

In a living organism or cell the collective behavior of thousands of genes and their products (for example *mRNA* and proteins) are embedded in a complex architecture that creates the mystery of life. More than 10 years ago, methods in molecular biology worked on a "*one gene-one experiment*" framework, meaning that the throughput is constrained to only one gene and therefore, the whole image of gene function is difficult to visualize. An emerging technology called *DNA* Microarray/GeneChips [1,2] appeared in the recent years, attracting the interests among biologists, computer scientists, mathematicians and physicists. By using this technology, the whole transcribed genome can be monitored and the correlations among thousands of genes can be revealed.

In order to understand how the organism works, it is necessary to know *which* genes are expressed, *when* they are expressed and *how fast* they do. Gene expression is regulated by means of the gene regula-

* Corresponding author.
E-mail addresses: ochiai@kuicr.kyoto-u.ac.jp (T. Ochiai), nacher@kuicr.kyoto-u.ac.jp (J.C. Nacher).
[1] These authors contributed equally to this work.

0375-9601/$ – see front matter
doi:10.1016/j.physleta.2005.02.066

tion architecture system of the cells, which involves network of interactions among *DNA*, *mRNA*, proteins and hundreds of small ubiquitous molecules. These interactions involve many elements and different and complex mechanisms, therefore an intuitive understanding of the underlying dynamics is not easy to obtain. This is also true for the issue of gene correlation, which is the main aim of this Letter. In particular, although many approaches and techniques, as for example Boolean networks, graph theory and control theory, have been used successfully in many cases, they are still far to achieve a general description of the dynamics of the regulation and correlation among genes.

To shed light on this issue, here we propose a new theoretical model to deal with multi-gene correlation dynamics based on only one assumption: *Markov property*. Our approach will use the most general multi-variate stochastic process in order to obtain predictions about the correlations among genes.

In a previous work [3], we proposed a constructive approach to gene expression dynamics, which rebuilds the scale-free organization of genes (i.e., expression level k decays as a power-law $k^{-\gamma}$ [4,5]) observed in recent experiments [6–8]. There, we proposed a stochastic approach by assuming the Markov property and by using the observed experimental transition probability data, which characterize the gene expression system. Although our companion paper [3] succeeded to re-build the scale-free distribution, it may not provide much information about gene correlation phenomena because by construction it is one gene approach-like. Therefore, here our aim is to exploit the novelty of our previous constructive approach by extending that one-dimensional model, to a multi-dimensional analysis (i.e., multi-gene correlation).

Our approach is based on two fundamental aspects: Markov property and stochastic process.

1.1. Markov property

If we say that the system has a Markov property, we mean that the future is governed by the present and does not depend on the past. Our model assumes Markov property to describe the multi-gene expression dynamics. Certainly, the living organisms are systems with long term memory, and they are complex systems with large number of elements and interactions. However, from a physical point of view, although we may not know all the variables in real situations, it may be enough to find a reduced number of variables whose behavior in time can be described as a Markovian process. Therefore, in our study we may assume that the most relevant degree of freedom of the system is the gene expression level, and consequently, our gene system has the Markov property.

1.2. Stochastic process

Although many complex systems may be governed by non-stochastic processes, in the gene expression problem the random variation is reasonable, plays a relevant role in cellular process, and furthermore stochastic noise have recently been measured and studied theoretically [9–12]. For example, the expression level of thousands genes is very low, which creates intrinsic uncertainties in the number of expressed genes in the cells [6]. Furthermore, we can even distinguish between inherent stochasticity (*intrinsic noise*) and external stochasticity (*extrinsic noise*). While the origin of the first one are the biochemical processes, and motivates that two identical genes become uncorrelated due to that randomness, the second one represents sources of extrinsic noise, which change from cell to cell (i.e., fluctuations in elements among cells) [9,13]. Moreover, the number of molecules which are involved in signal transduction pathways fluctuates from 10^2 to 10^4. Therefore, the randomness connected with elementary molecular interactions and their amplification in the signaling cascade generates significant spatio-temporal noise. Therefore, the stochastic approach is justified, and it seems more appropriate and plausible than a deterministic approach. Finally, it is also worth reminding that the current experimental techniques also provide an additional source of fluctuation, which come from the ubiquitous instrumental noise (which may be around 30% or more) from chip to chip with the current GeneChips technologies.

On the other hand, one drawback of our approach is that the current existing experimental data of gene expression time series lacks of enough statistics. For example, in yeast [14] and human [15] organism experiments, they analyzed the fluctuations in time of many genes simultaneously (more than 30 000 in the case of human organism) by carrying out only *one* experiment.