

Emergent Principles in Gene Expression Dynamics

J.C. Nacher^{*,a} and T. Ochiai^b

^aDepartment of Complex and Intelligent Systems, Future University-Hakodate, 116-2 Kamedanakano-cho Hakodate, Hokkaido 041-8655, Japan

^bFaculty of Engineering, Toyama Prefectural University, 5180 Kurokawa Imizu-shi, Toyama 939-0398, Japan

Abstract: Rapid advances in data processing of genome-wide gene expression have allowed us to get a first glimpse of some fundamental laws and principles involved in the intra-cellular organization as well as to investigate its complex regulatory architecture. However, the identification of commonalities in dynamical processes involved in networks has not followed the same development. In particular, the coupling between dynamics and structural features remains largely uncovered. Here, we review several works that have addressed the issue of uncovering the gene expression dynamics and principles using micro-array time series data at different environmental conditions and disease states as well as the emergence of criticality in gene expression systems by using information theory. Moreover, we also describe the efforts done to explore the question of characterizing gene networks by using transcriptional dynamics information. The combination of the emergent principles uncovered in the transcriptional organization with dynamic information, may lead to reconstruct, characterize and complete gene networks. We also discuss several methods based on simulations of a series of enzyme-catalyzed reaction routes and Markov processes as well as combination of complex network properties with stochastic theory.

Keywords: Systems biology, dynamical modeling, complex networks, gene expression.

INTRODUCTION

High-throughput data acquisition with DNA microarray technologies have led to drastic changes in molecular biology, a field historically characterized by reductionism for decades. The simultaneous measure of the concentration of tens of thousand of RNA transcripts from a diverse sample of cells, including normal and cancer tissues, offers a new challenge in biology in the twenty-first century [1-3]. Instead of focusing on individual bio-molecules, scientists investigate biological functionalities that emerge from complex interactions of fundamental molecules in distinct temporal states, making it possible to envision, for the first time, a model of the cell's underlying topological structure and its transcriptional dynamics [4-9]. Rapid advances in the field have contributed to develop multi-disciplinary based reverse-engineering techniques to infer models of transcriptional regulatory networks and predict the cellular response under stresses and therapy treatments [9-16].

Bio-molecular regulatory networks aim to describe the causal, functional and physical relationships between functional molecules like genes and proteins. These bio-molecules are the nodes in the network and two nodes can be connected by edges representing the relationships between them. Several engineering-oriented strategies and methods have been suggested to reconstruct these networks and analyze its properties. Boolean networks, differential equation models, Bayesian network models and association networks

are some representative examples of the rapid expansion of the field [10, 12, 16-19]. However, complex systems, and the cell is not an exception, are composed of a large number of elements. A human cell is composed of tens of thousand unique transcripts, each of them having up to hundreds of copies. Furthermore, the interacting rules between cellular components may change over time in response to environmental stresses. On the other hand, a more systemic-oriented approach that combines previous ideas seeks to identify and unveil fundamental organizing principles that govern the formation and evolution of transcription networks. This research based on complex systems science focus on determining the dynamics of a real system by proposing simple models that capture its main features. In this systemic view, the whole is much more than the sum of its parts, and consequently, the functionality of the system is given by the interactions between its constituents [4, 6-7]. In this context, qualitative approaches can identify the existence of universal laws and generic principles that lead to a greater insight into the cell's functional organization and biological phenomenon in general.

Rapid advances in bio-knowledge and data processing of genome-wide gene expression have allowed us to get a first glimpse of some fundamental laws and organizing principles that govern the cellular functions. One of these findings has shown that the transcriptional organization is very heterogeneous, suggesting that the probability that a gene has an amount of expression k decays as a power-law [20-22]. This organization seems to be universal and conserved across species, cell types and disease states suggesting a robust property in gene expression dynamics. Furthermore, recent research suggests that the gene expression dynamics exhibits criticality in the macrophase [23]. These findings, the exist-

*Address correspondence to this author at the Department of Complex and Intelligent Systems, Future University-Hakodate, 116-2 Kamedanakano-cho Hakodate, Hokkaido 041-8655, Japan; Tel: +81-138-34-6123; Fax: +81-138-34-6124; E-mail: nacher@fun.ac.jp

tence of a power-law as stationary state and the emergence of criticality in gene expression are extremely important to understand the cell's functional organization and to unify its study of structure and dynamics.

Here, we review several works that have addressed the issue of uncovering the principles of gene expression dynamics using micro-array time series data at different environmental conditions and disease states as well as the emergence of criticality in gene expression systems by using information theory. Moreover, we also describe the efforts done to explore the challenging question of predicting gene networks by only using transcriptional dynamics information. The combination of the emergent principles uncovered in the transcriptional organization with dynamic information, may lead to construct, characterize and complete gene networks. In particular, we explore the possibility of identifying structural features of networks by only computing dynamic information obtained from gene expression time series data. We also discuss several methods based on simulations of a series of enzyme-catalyzed reaction routes and Markov processes as well as combination of complex network properties with stochastic theory.

UNIVERSALITY IN SYSTEMS BIOLOGY

The molecular interactions within a cell are very complex and their direct study poses enormous difficulties from experimental and theoretical view point. However, the cell is not the only example of complexity. We are surrounded by many disparate complex systems like, for example, financial systems, social networks, fluid dynamics and Internet evolution. In these cases, it is simply impracticable to solve and predict the behavior of single stock prices, individuals, water atoms and web pages, respectively. In spite of that, these systems often show a remarkably simple behavior and commonalities. While the understanding of these systems using a microscopic approach seems a hard and intractable task, its study from a macroscopic view leads to an understanding of the observed behavior. This coarse-grained approach is central in statistical mechanics, whose concepts and techniques have permeated different fields like economics, engineering, computer science and life sciences [24-28]. In short, the great success of statistical mechanics and its predictive power in so many disparate fields has been possible because macroscopic phenomena, and their relative simple observed laws, do not depend on the microscopic particulars in a large extent. Well-known examples are the Newton's laws or the Navier-Stokes equations. These laws emerge from underlying complex interactions and properly reproduce the effective observed behavior of the system. However, the water motion and gravity effect do not largely depend on details of the elementary molecular interactions of water molecules and matter forces, respectively. This fact has definitely contributed to identify continuum laws and principles in physics that apply to many fields. With the advent of high-throughput technologies in the post-genome era, there has been a necessity to process this new and massive information and, at the same time, to provide answers to fundamental questions on the cell's functional organization and its elemental interacting constituents. This huge flow of data has led to hypothesize about the emergence of simple laws for dynamic molecular processes in life sciences. Bioinformaticians and computer scientists have dealt with these issues by

combining algorithms and computational techniques, based on data processing, with leading statistical mechanics concepts to unravel the principles of cell biology. In particular, gene expression dynamics has been one of the central topics of interests in the recent years.

LINEAR RESPONSE IN GENE EXPRESSION

As we have mentioned above, an understanding of the gene expression organization and its dynamics is largely lacking. An attempt to uncover fundamental dynamic principles that generate complex gene expression fluctuations was carried out in [22]. A series of genome-wide experiments were conducted using high-density oligonucleotide arrays to investigate the large-scale organization of gene expression dynamics and elucidate its fundamental principles. Gene expression data was collected from experiments performed using six organisms (*E. coli*, *S. cerevisiae*, *A. thaliana*, *D. melanogaster*, *M. musculus* and *H. sapiens*) under several environmental conditions (for example, constant light/dark cycles, normal or symbiotic conditions). First, the global RNA expression features were investigated by analyzing the distribution of gene expression. The results showed that the probability to find a gene with an amount of expression k decays as a power-law with the analytical form $P(k) \sim k^{-\gamma}$. Interestingly, this behavior is conserved from *E. coli* to *H. sapiens* with degree exponents in the vicinity of the value -2, suggesting that prokaryotes and eukaryotes share a similar genome-wide transcriptional organization.

However, a question arises about whether this observed power-law will be conserved when different temporal states or tissues are investigated. The analysis of RNA expression using data at different circadian time (CT) or various tissues showed that the distribution of gene expression displayed a power-law in spite of the fact that many genes changed their expression values. For example, while 3% and 16 % of genes in *S. cerevisiae* and *A. thaliana*, respectively, changed their expression values, this proportion increased up to 95 % and 88 % of genes in *H. sapiens* and *M. musculus*, respectively. Therefore, it means that even though a large proportion of genes changed the individual expression level, in average the changes compensated each other and the power-law remained invariant with the same exponent. These fractions of differentially expressed genes were observed in 47 tissues of *H. sapiens* and 45 tissues of *M. musculus*. This finding raised a new question about what kind of dynamics generated this invariant transcriptional organization across time and space (i.e., tissues).

The transition probability $T(k_1, k_2)$ of expression change for individual genes with expression level k_1 to another expression level k_2 , along two different conditions was computed for several organisms. This expression was calculated by counting the number of genes with expression level k_1 that changes to another expression k_2 . The results showed that the transition probability does not show a random pattern. In contrast, it largely depends on the before-transition expression level k_1 . It means that while highly expressed genes have a higher probability to largely change its expression level, low expressed genes have low probability to show large changes. To precisely determine the nature of the dy-

namics involved in these transitions, the absolute change $|\Delta k| = |k_2 - k_1|$ was computed. The results indicated that this absolute change was proportional to the initial before transition value, suggesting a proportional expression dynamics in agreement with the results shown by the transition probability. To understand this finding in a theoretical context, the authors proposed a model where the standard deviation of gene expression change increases proportionally to the before-transition gene expression level k . In this approach, the gene expression dynamics is represented by a continuous Markov process. In this context, the forward Kolmogorov equation, or equivalently the Fokker-Plank equation in one dimension can be written as:

$$\frac{\partial P(k,t)}{\partial t} = -\frac{\partial}{\partial k} [\langle dk \rangle P(k,t)] + \frac{1}{2} \frac{\partial^2}{\partial k^2} [\langle |dk|^2 \rangle P(k,t)] \quad (1)$$

where $\langle dk \rangle$ and $\langle |dk|^2 \rangle$ are the drift and diffusion terms, respectively. The drift term $\langle dk \rangle$ indicates the average of instantaneous expression changes and can be computed using the following expression:

$$\langle dk \rangle = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int T(k + \Delta k, k) \Delta k d(\Delta k) \quad (2)$$

where Δk is an expression change during time interval Δt . It was assumed that there was not specific tendency in the average value of gene expression level. In other words, for each gene, the expression level could increase or decrease, so the value of the drift term averaged over time is taken as $\langle dk \rangle = 0$.

The standard deviation of instantaneous expression change (in the context of Fokker-Planck equation, this expression is called diffusion term) $\langle |dk|^2 \rangle$ was computed using the following expression:

$$\langle |dk|^2 \rangle = \left[\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int T(k + \Delta k, k) (\Delta k)^2 d(\Delta k) \right]^{1/2} \quad (3)$$

Furthermore, it was hypothesized that this standard deviation of instantaneous expression change was proportional to the before-transition expression level k . This leads to the following expression:

$$\langle |dk|^2 \rangle = a(k + b)^s, \quad (4)$$

where $a > 0$, $k \gg b > 0$, and $s \sim 1$. By considering these assumptions and by imposing the stationary state condition, a stationary distribution of gene expression level characterized by a power-law with exponent -2 was obtained, in good agreement with experimental observations. This proportional or “rich-travel-more” dynamics could generate a power-law distribution, in a similar way that “rich-get-richer” mechanism (also known as preferential attachment) could derive the power-law distribution in complex growing networks. The described gene expression dynamics represents a generic principle in the regulation of transcriptome based on large-scale statistics, for both highly or lower expressed genes, that co-exists with more rich and specialized dynamics, in a lower-level view, that characterizes specific modules like feed-back loops and cycles [29]. However, although this

“rich-travel-more” mechanism has an intrinsic beauty and attractiveness, its linear form is surprising. It could be expected that a complex regulatory network, that is able to adapt environmental or rapid developmental stages, might be governed by a highly complex non-linear response dynamics. Perhaps, a new avenue for the investigation of gene expression dynamics could be the simultaneous investigation of structure and dynamic properties [23]. In the next sections, we will see that, in spite of the fact that a complete framework is still lacking, recent and promising studies suggest that the information theory could significantly contribute to this issue.

CRITICALITY

If we think in terms of atomic matter, fluids or even larger-scales like social networks, populations, cities or ecosystems, we observe that these systems are composed of multiple fundamental elements or individuals that interplay by means of physical forces, social relationships or information exchange. While these interactions are originated by intrinsic features of the systems, external forces, like electromagnetic and gravitational fields, social rules as well as drastic and severe climate changes, may also drive the evolution of the system. An intriguing phenomena is that even though intrinsic and extrinsic forces co-exist, it seems that systems share a high degree of commonality and behavior, which seems to be independent of the nature and details of the system itself. In [30, 31], the Bak, Tang and Wiesenfeld (BTW) model shows that, in fact, any system consisting of many interacting components could display general and common behavior by organizing itself into a state where there is no a homogeneous or specific size for characteristic event. Instead of that, it is suggested that there is a large-range of possible time or length scales, where events can occur in a heterogeneous way (i.e., the observed distributions exhibits a power-law decay). A simple example is the earthquake activity characterized by many small quakes but few large ones far in the tail of the distribution. Systems that obey this kind of power-law dynamics exhibit the same structure (self-similar) over all scales. Moreover, the exponents of the power-laws could be relatively similar between systems that show a disparate microscopic nature. This description certainly resembles the characterization of real networks carried out by the new science of networks [7, 32]. The main statement in BTW is that the system can self-organize itself without needing specific external constraints. Their global organization is the result from the interactions between the components. Moreover, these interactions are independent of the physical nature of the individual components. Interestingly, in the self-organized states, different systems show similar emergent behavior, characterized by scale-free or power-laws distributions. Several works have suggested that life has precisely emerged and developed at this self-organized state, a critical point between order and chaos [33].

A very fresh and novel approach was recently conducted based on information theory. As discussed above, the cell is a highly regulated system that is able to respond to external signal, adapt environmental changes and conserve stability. In this process, a fundamental and recurrent feature in self-organized systems is the existence of criticality. In addition, it is well-known that criticality also represents a trade off

between adaptability and stability. It indicates a state of the system where signals and perturbations can be efficiently propagated over longer temporal and spatial states, without being repressed nor amplified. While it has been suggested since several decades ago, that biological systems are critical and live at the edge of the criticality, Schmulevich group conducts experiments on gene expression and addresses this issue with biological experiments. In particular, they examine the large-scale organization of gene expression dynamics in the macrophage [23]. The striking aspect of their approach is that, based on existing measures of information theory, they propose a novel method that determines the existence of criticality in macrophage cell types. These cells were stimulated by a variety of stresses and the results suggested that macrophage cells operate on a critical state-based dynamics.

LINKING INFORMATION THEORY AND CRITICALITY IN GENE EXPRESSION DYNAMICS

Macrophages, also known as white blood cells, are good examples of how molecular entities are able to respond to environmental signals [23]. This type of cell is involved in the immune systems and as well as involved in the inflammatory response programs. Cell surface receptors detect molecular patterns that pose perils to the cell, and transmit the information through signal transduction pathways where macrophages detect it, and initiate complex and adapted gene expression programs that lead to modify their functional state, start producing specific chemical compounds, and release them to deal with the pathogen. It is particularly interesting and intriguing how the information that reaches the cell is transmitted through the molecular pathways and activates these complex responses [34]. To deal with this issue, it has been suggested since long, that living organisms and cells operate at the critical boundary between a perfect ordered state (like a lattice, for example) and a completely disorganized state (chaotic system, for example). Several studies have analyzed these features in biological systems [35, 36]. The interplay between these two very different states seems to be an essential feature of emergence of life and adaptability. If the system becomes too chaotic, it self-evolves into a more ordered state and vice-versa. Criticality represents the key link that connects two drastically different worlds and where cell behavior emerges in a compromise between stability and adaptability [33]. It could explain why so disparate dynamic systems share universal features.

The approach is based on information theory, concretely in the information dynamics and how to measure the information distance between different states. The authors consider two different external signals that make evolve the cell into two well-defined states. Each state can be determined by measuring the time-course transcriptional data. By measuring the information difference between two states and how this difference changes over time, we can characterize the dynamics of the system. An informational convergence would suggest a world where the uniqueness of both signals are progressively vanished, suggesting a stable and robust behavior. In contrast, an informational divergence scenario would indicate that even a small differential signal could be amplified, which drastically compromises the transmission of information over time. A critical behavior lies in the middle, in the most effective balance point, showing an equilib-

rium between stability and adaptability and supporting the “*life on the edge of chaos*” hypothesis.

In Li *et al.* an information-based measure was developed to measure the universal information distance between two objects [37, 38]. The minimal information distance between discrete objects x and y is the length of the shortest program that transforms x into y as well as computing y from x . It can also be defined in terms of the Kolmogorov complexity. The Kolmogorov complexity or algorithmic entropy $K(x)$ of a string x is the length of a shortest binary program to compute x on an appropriate universal computer. In other words, it represents the minimal amount of information required to generate x by any algorithm. It can also be interpreted as the compressed form of x . As this concept is not computable, the authors approximate its computation by using real-world data compressors (i.e., gzip) and used the normalized compression distance NCD :

$$NCD(x, y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}} \quad (5)$$

This distance has been normalized because, for example, when it is required to express similarity between two strings, it is more appropriate to show the relative distances among them. Two strings of 10,000 bits differing by only 100 bits seem to be similar. However, two strings of 100 bits differing by 100 bits are totally different [38]. In this expression, $C(x)$ represents the compressed size of x and xy means the concatenation of the strings x and y . Schmulevich group performed gene expression experiments and measured the course of the time series. Data was obtained from murine bone marrow-derived macrophages under six different treatments or conditions. For each pair of distinct experiments ($S1, S2$), NCD was computed for every pair of time-point measurements. Then, by selecting consecutive time points ($t1, t2$) in distinct experiments ($S1, S2$), $NCD(S1(t1), S2(t1))$ was plotted versus the next time step $NCD(S1(t2), S2(t2))$. Here, $S1(t)$ and $S2(t)$ refers to the expression states for two micro-array experiments. Therefore, the NCD uses the compressed size of each micro-array data at a given time t , $C(S1(t))$, $C(S2(t))$ and the compressed size (i.e., binary information corresponding to expression values compressed using a real-world data compressor like gzip) of the corresponding concatenated string $C(S1(t)S2(t))$. The result showed that the scatter plot follows the main diagonal, with slope 1. This finding suggests that the transcriptional network dynamics is governing the macrophage functions in the critical regime. As showed above, this finding is consistent with the results showing that gene expression is organized following power-law distribution. In critical regime, the systems tend to self-organize to be scale-free. Data above the diagonal would imply a chaotic or divergent behavior, indicating a system hypersensitive to external signals or constraints and with difficulties to carry out a coherent response. In contrast, a distribution below the diagonal would suggest an informational convergence. The system would progressively become more insensitive to signals [23]. Furthermore, these results, showing that that gene expression dynamics in the macrophage exhibits criticality, were validated by using Boolean networks simulations. This method of computation is very useful because it can potentially applied to genomic sequences, networked structures, time series data, and in general to any objects that can be represented using comput-

ers. The interplay between networked structures and dynamics using Boolean networks as well as the phylogenetic analysis of the metabolism of 107 organisms were conducted using NCD information [39]. Basically the information distance can be applied to any object. Although it was applied to gene expression data, it can also be applied to networks, graphs or any other type of data. The pairwise distances between the metabolic networks of more than hundred organisms were computed, and a phylogenetic tree, generated by means of the complete linkage method was constructed. This approach was able to clearly group all the organisms into the corresponding three domains of life and subclasses of species. It shows that fundamental structural differences in networks are highlighted by using this method. On the other hand, the relationship between dynamics and structure was investigated in an experiment using Boolean networks. Six ensembles of random Boolean networks were prepared with $N=1000$ nodes with average network connectivity $K=1,2,3$. For random wiring of edges, by appropriately setting the free parameter in the critical phase transition curve, $K=1,2,3$ will correspond to ordered, critical and chaotic states, respectively. Moreover, each ensemble consisted of 150 networks. Then, NCD measures were computed between all pairs of network structures and between the corresponding dynamic state trajectories. By plotting the NCD computed using structural information in x axis and NCD calculated using the dynamics state trajectories in y axis, the six ensembles were clearly separated and visible in the figure [39]. Interestingly, only the ensemble corresponding to random topology with $K=2$ exhibited an elongated pattern along the vertical axis at a fixed point in the horizontal axis, showing a more rich and diverse dynamical behavior. This supports the idea that systems in critical regime operate and exhibit a maximal dynamic diversity. It is precisely on the edge of the critical regime, where complexity arises and where the interplay between structure and dynamic becomes more relevant. These findings showed the power of the approach based on NC measure in order to combine networked structures and dynamic information into the same framework. It would certainly be very interesting if this analysis could be carried out using real biological systems, like gene expression data, instead of using computer simulation models.

ZIPF'S LAW IN GENE EXPRESSION

An extensive analysis on gene expression data using a large variety of organisms and tissues revealed that the abundances of expressed genes follow a power-law distribution [21] with exponent close to -1 (i.e., Zipf's law [40]). The analyzed tissues were publicly available from SAGE (serial analysis of gene expression) database and included not only human normal cells but also different types of cancer tissues as well as embryonic stem cells. Furusawa and Kaneko investigated the possible universal features of cell's reaction dynamics, and in particular, what kind of molecular dynamics can generate the observed power-law behavior. It is worth noticing that there is a relationship between the exponent of the power-law derived from a Zipf's law and the exponent from a power-law probability distribution. The functional form of the Zipf's law can be written as

$$Z(r) \sim r^{-b} \quad (6)$$

It indicates the frequency of an occurrence of an event relative to its rank r . Zipf's law states that the frequency of the r th largest occurrence of the event (in this case gene expression) is inversely proportional to its rank. Therefore, Furusawa and Kaneko computed the rank-ordered frequency distributions of expressed genes. In contrast, Ueda *et al.* focused on the probability distribution to find genes with expression level k

$$P(k) \sim k^{-\gamma} \quad (7)$$

It can be shown that both exponents are related by means of the relationship:

$$\gamma = 1 + (1/b) \quad (8)$$

which shows the agreement between both studies since Ueda *et al.* reported an exponent close to 2.

Their model defines a variety of chemicals (n_1, \dots, n_k) , where n_i denotes the number of molecules of the chemical species i . The chemical reaction dynamics is determined by considering a catalytic network composed of individual chemical reactions. This reaction network transforms nutrient chemicals into proteins. Each reaction is catalyzed by a chemical l and transforms a chemical i to some other chemical j (i.e., $i+l \rightarrow j+l$). An interesting feature of the model is that the results are robust to a uniform selection of reaction coefficients as well as the internal network structure (i.e., random or scale-free topologies) of the catalytic networks, which denotes the flexibility of the approach.

To simulate chemical reactions, it was necessary to consider fundamental compounds that play the role of resources or nutrients supplied from the environment and that permeated by diffusion through the membrane. Although these nutrient chemicals did not have a catalytic capability, they could be transformed into other chemicals by means of chemical reactions. Due to that some of the chemicals could not permeate through the membrane, the number of chemicals inside the cell could grow and reach a threshold beyond which the cell divides into two and the number of molecules also splits among the two new cells. The mathematical expression for the reaction dynamics is described using the following rate equation:

$$\frac{dn_i}{dt} = \sum_{j,l} \text{Con}(j,i,l) \mathcal{E} n_j n_l / N^2 - \sum_{j,l} \text{Con}(i,j,l) \mathcal{E} n_i n_l / N^2 + D \sigma_i (\bar{n}_i / V - n_i / N) \quad (9)$$

where $\text{Con}(i,j,l)$ takes the value 1 if there is a reaction $i+l \rightarrow j+l$ and 0 otherwise. Next, \mathcal{E} indicates the coefficient for each chemical reaction and σ_i is 1 if the chemical i is penetrable and 0 otherwise. The last term shows the diffusion of chemicals through the membrane, where \bar{n}_i denotes the number of the i th chemical species in the environment and V shows the volume of the environment in units of initial cell size. The analysis and simulation results indicated that the reaction dynamics exhibited a critical value for the diffusion coefficient $D=D_c$. For values above this critical point, the flow from external environment is too rapid and the cells stops growing. Continuous cellular growth, including cell

divisions, are only maintained for values $D \leq D_c$, where the internal reaction dynamics is faster than the external flow of nutrients. The results show that the slope in the rank-ordered number distributions of chemical species increases with the diffusion coefficient and only at the critical point $D=D_c$, it becomes a power-law with exponent -1.

Here it is interesting to discuss why this power-law is maintained in so complex and fluctuating medium. It was suggested that a hierarchical organization of catalytic reactions is behind this scale-freeness behavior at the critical point. In this view, the generation of higher ranking chemicals (those that are present in very high number ($n_i > 1000$)) is catalyzed by lower ranking chemicals (those that are present in lower number). Next, the latter chemicals are again catalyzed by chemicals that show a lower abundance, and this hierarchically organized sequence of reactions continues until the chemical species with lower number of molecules are reached ($n_i < 5$). The presence of hierarchy in the reaction dynamics leads to explain the observed exponent in the power-law distribution by means of a mean-field approximation (see Ref. [21] for details).

An important finding of the present model is that the observed power-law distribution does not depend on the details on the model, such a kinetic ratio, reaction network configuration or specific functional molecules. It rather means a universal property of the cell that is able to grow using environmental nutrients and operate complex biochemical reactions at a precise critical state. The fact that a critical point exists in reaction dynamics raises the question about its origin and why it has been assimilated by the cell system. This could be investigated by computing simulations of the cell growth at different diffusion values. The results showed that the growth ratio of a cell is maximized at $D=D_c$, indicating that a critical state is one of the best possible candidates for a probabilistic natural selection. Furthermore, and interestingly, the computation of the degree of similarity, measured as the scalar product of k -dimensional vectors of chemical composition, between chemicals of the original cell and split cells, revealed that the critical state also maximizes this value. These findings suggest that evolutionary dynamics supports a critical state based on maximization principles of cell growth and chemical homogeneity and similarity. This research is interesting since it provides new insights on principles closely related to dynamic process, instead of using only topological information.

FLUCTUATIONS IN GENE EXPRESSION

Another recent finding on gene expression dynamics has shown that the distribution of gene expression fluctuations (i.e., standard deviation) in gene expression time series follows a power-law distribution [41]. This scaling indicates that while most genes exhibit a relatively low variation in expression level, a few genes show a highly variable pattern. The analyzed time series experiments included human cell data corresponding to cell-cycle-regulated transcripts measured using high-density oligonucleotide arrays and the cell cycle of the budding yeast. The observed fat-tailed distribution was observed in both human and yeast organisms. Then, the relationship between the dynamic information of gene expression time series and the underlying network structure was investigated. In order to elucidate a possible relationship

between the fluctuation of genes and its topological characteristic (i.e., node degree) and explain the power-law of fluctuations, we propose a simple mathematical model based on stochastic formulation [42-44].

First, a set of nodes that correspond to genes is considered. Each node has an intrinsic value that reflects its expression level. Next, it is assumed these nodes may be interconnected and that the node degree distribution follows a power-law with exponent γ . Then, a general stochastic partial differential equation is considered with the multi-dimensional stochastic variable $X_t = (X_t^1, \dots, X_t^M)$ denoting the gene expression level and the multi-dimensional Wiener process represented by $W(t) = (W_1(t), \dots, W_M(t))$, where all the processes are assumed to be independent to each other. The model assumption is that each node i fluctuates as described by $dW_i(t)$. Moreover, fluctuations can be induced between adjacent nodes. We then can formulate the dynamic stochastic partial differential equation (SPDE) for gene expression dynamics as follows:

$$dX_t^i = \mu^i (m^i - X_t^i) dt + C dW_i(t) + D \sum_{j=1, j \neq i}^M a^{ij} dW_j(t) \quad (10)$$

where μ^i is the strength of the mean reverting mechanism, a^{ij} denotes the adjacency matrix and C and D are proportional coefficients. In previous expression, the first term on the right hand side represents the mean-reverting mechanism. This mechanism means that the gene expression level of a given gene i tends to fluctuate around its mean value m . The second term denotes a self-regulatory fluctuation (or loops). The last one takes into account the regulatory interaction transmitted from the other nodes connected to node i .

By using Ito formula, the SPDE can be solved and the dynamic solution for gene expression system can be derived. Moreover, by computing the variance $V[X_t^i]$ when the system reaches the steady state, it is possible to derive the following relationship between standard deviation and node degree for steady state:

$$\sigma^i = \sqrt{V^i} = \sqrt{\frac{C^2 + D^2 k_i}{2\mu^i}} \quad (11)$$

Then, by combining this equation with the power-law distribution of the underlying regulatory network $k^{-\gamma}$, we find:

$$P(\sigma) \propto \frac{2D^{2(\gamma-1)}}{\sigma^{2\gamma-1} (2\mu)^{\gamma-1} (1 - C^2 / (2\mu\sigma^2))^{\gamma}} \sim \frac{1}{\sigma^{2\gamma-1}} \quad (12)$$

This relationship shows a power-law distribution for the distribution of fluctuations, similar to that found in the gene expression data. Furthermore, it allows us to establish a link between the degree exponent of the network and the exponent of the node fluctuation α . To be precise, it shows that the underlying network exponent can be computed using

only dynamic information by means of the expression $\gamma = (\alpha + 1) / 2$.

The disproportionate number of highly fluctuating genes shown by the power-law of fluctuations resembles the existence of hub nodes observed in many biological networks [7, 27, 45]. This result suggests a possible link between structure and dynamics to explore real system characterized by many fluctuating and networked elements, and supports previous findings on the power-law nature of gene regulatory networks [46-49].

CONCLUSION

Systems whose structural information is difficult to collect represent a challenge to reconstruct its underlying topology. While metabolic networks are being reconstructed rapidly for hundreds of organisms, gene networks have not followed the same development. Large-scale genomic technologies have made it possible to measure thousands of mRNA levels in parallel and collect their fluctuation patterns by means of time series experiments. However, it has not yet implied an accurate and extensive knowledge about the gene regulatory network structure for a significant large number of organisms. This is even more evident when the question is to elucidate the dynamical processes that occur in these networks [50]. Furthermore, the links between structural features and dynamical processes remain largely uncovered. In the same way, as we have shown here, recent qualitative approaches focused on identifying the existence of universal laws and generic principles, have provided a novel and remarkable insight into the cell's functional organization and its dynamics. It is expected that more complex models and new methodologies that include and combine fundamental properties observed across species, may lead to understand the coupling between dynamic and structural characteristics and help to reconstruct and systematically identify large fragments of the cellular maps still hidden to the eyes of the world.

REFERENCES

- [1] P.O. Brown, and D. Botstein, "Exploring the new world of the genome with DNA microarrays," *Nat. Genet.*, vol. 21, pp. 33-37, 1999.
- [2] R.J. Lipshutz, S.P. Fodor, T.R. Gingeras, and D.J. Lockhart, "High density synthetic oligonucleotide arrays," *Nat. Genet.*, vol. 21, pp. 20-24, 1999.
- [3] J.B. Fan, X. Chen, M.K. Halushka, A.H. Berno, X. Huang, T. Ryder, R.J. Lipshutz, D.J. Lockhart, and A. Chakravarti, "A parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays," *Genome Res.*, vol. 10, pp. 853-860, 2000.
- [4] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray, "From molecular to modular cell biology," *Nature*, vol. 402, pp. C47-C51, 1999.
- [5] J. Hasty, D. McMillen, and J. J. Collins, "Engineered gene circuits," *Nature*, vol. 420, pp. 224-230, 2002.
- [6] H. Kitano, "Computational systems biology," *Nature*, vol. 420, pp. 206-210, 2002.
- [7] A.-L. Barabasi, and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nat. Rev. Genet.*, vol. 5, pp. 101-103, 2004.
- [8] J. Tegnér, M. S. Yeung, J. Hasty, and J. J. Collins, "Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 5944-5949, 2003.
- [9] M. Bansal, G.D. Gatta, and D. di Bernardo, "Inference of gene regulatory networks and compound mode of action from time course gene expression profiles," *Bioinformatics*, vol. 22, pp. 815-822, 2006.
- [10] H. de Jong, "Modeling and simulation of genetic regulatory systems: a literature review," *J. Comput. Biol.*, vol. 9, pp. 67-103, 2002.
- [11] E. Davidson, *Genomic Regulatory Systems: Development and Evolution*. San Diego: Academic Press, 2001.
- [12] T. S. Gardner, and J. J. Faith, "Reverse-engineering transcription control networks," *Phys. Life. Rev.*, vol. 2, pp. 65-88, 2005.
- [13] J. Hasty, D. McMillen, F. Isaacs, and J.J. Collins, "Computational studies of gene regulatory networks: In molecular biology," *Nat. Rev. Genet.*, vol. 2, pp. 268-279, 2001.
- [14] M. Kaern, W.J. Blake, and J.J. Collins, "The engineering of gene regulatory networks," *Annu. Rev. Biomed. Eng.*, vol. 5, pp. 179-188, 2003.
- [15] M. Kaern, T.C. Elston, W.J. Blake, and J.J. Collins, "Stochasticity in gene expression: from theories to phenotypes." *Nat. Rev. Genet.*, vol. 6, pp. 451-464, 2005.
- [16] P. Brazhnik, A. de la Fuente, and P. Mendes, "Gene networks: how to put the function in genomics," *Trends Biotechnol.*, vol. 20, pp. 467-472, 2002.
- [17] A de la Fuente, P. Brazhnik P, and P. Mendes P. "Linking the genes: inferring quantitative gene networks from microarray data," *Trends Genet.*, vol. 18, pp. 395-398, 2002.
- [18] de la Fuente, N. Bing, I. Hoeschele, and P. Mendes, "Discovery of meaningful associations in genomic data using partial correlation coefficients," *Bioinformatics*, vol. 20, pp. 3565-3574, 2004.
- [19] M. A. Savageau, "Design principles for elementary gene circuits: elements, methods, and examples," *Chaos*, vol. 11, pp. 142-159, 2001.
- [20] V. A. Kuznetsov, G. D. Knott, and R. F. Bonner, "General statistics of stochastic process of gene expression in eukaryotic cells genetics," *Genetics*, vol. 161, pp. 1321-1332, 2002.
- [21] Furusawa, and K. Kaneko, "Zipf's law in gene expression," *Phys. Rev. Lett.*, vol. 90, pp. 088102, 2003.
- [22] H. Ueda, S. Hayashi, S. Matsuyama, T. Yomo, S. Hashimoto, S. A. Kay, J.B. Hogenesch, and M. Iino, "Universality and flexibility in gene expression from bacteria to human," *Proc. Natl. Acad. Sci. USA*, vol. 101, pp. 3765-3769, 2004.
- [23] M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, "Gene expression dynamics in the macrophage exhibit criticality," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, pp. 1897-1900, 2008.
- [24] A.-L. Barabasi, R. Albert, and H. Jeong "Mean-field theory for scale-free random networks", *Physica A*, vol. 272, pp. 173-187, 1999.
- [25] J. Park, and M. E. J. Newman, "The statistical mechanics of networks", *Phys. Rev. E*, vol. 70, pp. 066117, 2004.
- [26] J. Sethna, *Statistical Mechanics: Entropy, Order and Complexity*. Oxford: Oxford University Press, 2006.
- [27] M.C. Gonzalez, C.A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns", *Nature*, vol. 453, pp. 779-782, 2008.
- [28] P. Wang, M. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding the spreading patterns of mobile phone viruses", *Science*, vol. 324, pp. 1071-1076, 2009.
- [29] S. Shen-Orr, S. Milo, R. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of *E. coli*," *Nat. Genet.*, vol. 31, pp. 64-68, 2002.
- [30] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality: an explanation of 1/f noise," *Phys. Rev. Lett.*, vol. 59, pp. 381-384, 1987.
- [31] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality", *Phys. Rev. A*, vol. 38, pp. 364-374, 1988.
- [32] S. N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks: From biological networks to the Internet and WWW*. Oxford: Oxford University Press, 2003.
- [33] S.A. Kauffman, *The origins of order: Self-Organization and Selection in Evolution*. New York: Oxford University Press, 1993.
- [34] M. Underhill, and A. Ozinsky, "Phagocytosis of microbes: complexity in action", *Annu. Rev. Immunol.*, vol. 20, pp. 825-852, 2002.
- [35] Shmulevich, S.A. Kauffman, and M. Aldana, "Eukaryotic cells are dynamically ordered or critical but not chaotic," *Proc. Natl. Acad. Sci. USA*, vol. 101, pp. 13439-13444, 2005.
- [36] P. Ramo, J. Kesseli, and O. Yli-Harja, "Perturbation avalanches and criticality in gene regulatory networks," *J. Theor. Biol.*, vol. 242, pp. 164-170, 2006.

- [37] M. Li, C. Xin, L. Xin, M. Bin, and P.M.B. Vitanyi, "The similarity metric," *IEEE T. Inform. Theory*, vol. 50, pp. 3250-3264, 2004.
- [38] M. Li, and P.M.B. Vintanyi, *An introduction to Kolmogorov Complexity and its Applications*, New York: Springer, 1997.
- [39] M. Nykter, P.D. Nathan, A. Larjo, A.T. Antti, A.S. Kauffman, O. Yli-Harja, I. Shmulevich, "Critical network exhibit maximal information diversity in structure-dynamics relationship," *Phys. Rev. Lett.*, vol. 100, no. 05, pp. 058702, 2008.
- [40] G. K. Zipf, *Human Behavior and the Principle of Least Effort*. Cambridge: Addison-Wesley, 1949.
- [41] J.C. Nacher, and T. Ochiai, "Power-law of gene expression distribution," *Phys. Lett. A*, vol. 327, pp. 6202-6206, 2008.
- [42] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier Science B.V., 1992.
- [43] Wong, *Stochastic Processes in Information and Dynamical Systems*. New York: McGraw-Hill, 1971.
- [44] T. Mikosch, *Elementary Stochastic Calculus with Finance in View*, Singapore: World Scientific Publishing Co. Pte. Ltd, 1998.
- [45] A.-L. Barabasi, and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, 1999.
- [46] M. Gustafsson, M. Hornquist, A. Lombardi, "Constructing and analyzing a large-scale gene-to-gene regulatory network-lasso-constrained inference and biological validation," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 2, pp. 254-262, 2005.
- [47] H. Agrawal, "Extreme self-organization in networks constructed from gene expression data," *Phys. Rev. Lett.*, vol. 89, pp. 268702, 2002.
- [48] E. Feather, and K. Broadie, "Wrestling with pleiotropy: Genomic and topological analysis of the yeast gene expression network," *Bioessays*, vol. 24, pp. 267-272, 2004.
- [49] S. Shaw, "Evidence of scale-free topology and dynamics in gene regulatory networks," Proc. of the ISCA 12th international conference on intelligent and adaptative systems and software engineering, pp. 37-40, 2003.
- [50] A.-L. Barabási, "Scale-free networks: A decade and beyond," *Science*, vol. 325, pp. 412-4123, 2009.

Received: January 18, 2010

Revised: August 06, 2010

Accepted: August 06, 2010

© Nacher and Ochiai; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.