RESEARCH ON BEHAVIOR OF GOVERNING GENE/EPIGENE NETWORKS AS A PROBLEM OF CELLULAR AUTOMATA IDENTIFICATION

Tchuraev R.N.*
Department of Physicochemical Biology and Epigenetics, Ufa Research Center, RAS, Ufa, Russia
* Corresponding author: e-mail: tchuraev@anrb.ru

Key words: governing gene networks; cellular automata and cell ensembles; metastability

SUMMARY

Motivation: A body of mathematics used in automata and graphs theories is adequate for revealing the general dynamic properties of governing gene and epigene networks and provides a basis for efficient analytical algorithms.

Results: The paper presents the results of research on the general properties of cellular automata characteristic functions as models for intracellular networks that govern gene expression.

INTRODUCTION

By now mathematical methods of different efficiency and purpose have been worked out for formalized description of gene network dynamics allowing the models be constructed for a number of real eukaryotic gene networks (Tchuraev, Galimzyanov, 2001; Ratushny et al., 2003). Previously, as applied to intracellular governing gene networks of multicellular organisms, a simple formal system has been put forward, with cell governing networks represented as the inner structures of cellular automata $A^C$, and cell aggregates, being formed during ontogenesis, as the ensembles of cellular automata $A_\alpha$. In the context of this theoretical model, we have stated the general principles of organization and dynamics laws of functioning in governing gene networks on the whole (Tchuraev, 2006).

For the purpose of investigating more thoroughly into the behavior of governing gene networks, the present paper deals with the general properties of a cellular automaton as a part of self-reproduced cell ensemble corresponded to a multicellular organism in the process of its individual development. These properties reveal the molecular and genetic mechanisms of storing, passing and transforming hereditary information during ontogenesis and phylogensis.

MODEL

In order to reveal the general properties of cellular automata, the cellular automaton $A^C$, being an element of the cellular ensemble $A_\alpha$, will be treated separately from specific elements of this ensemble. Then cells of a multicellular organism cultured in vitro may be taken as prototypes of such automata.

Let us introduce the premises necessary in constructing the models.
1. Let us assume that cells cultured in vitro and all their clonal derivatives formed under successive mitotic divisions are provided with substance and energy in amounts sufficient for reproduction, i.e., we believe that resources of the outer medium impose no restrictions on intracellular metabolism (postulate A).

2. Let us also suppose that all external effects performed by an experimenter are only signal ones (postulate B).

By doing so, any fluctuations of the nutrient medium composition are excluded, along with other external stochastic effects. External (input) signals may be carried by specific proteins and metabolites and/or temperature and other pulses.

3. Let us introduce one more rather evident premise (postulate C).

Every cytotype (molecular phenotype) in any cell of an organism \( x \) evolving in a neutral environment is determined by gene differentiated activity of a given genome. Specific differentiated activity of the genome may be expressed, at each instant of time, by the \( \gamma \)-vector value of the activities \( \Gamma(t) = \langle \gamma_1(t), \ldots, \gamma_j(t), \ldots, \gamma_h(t) \rangle \) of genetic blocks \( G_j \), elements of the intracellular governing gene network \( S^e_x(G) \), which are considered to be the internal structure of cellular automaton \( A^e_x \). The governing gene network \( S^e_x(G) \) may be transformed into the epigen network \( \bar{S}^e_x(G) \) that involves the following modules: genetic triggers (bistable memory modules), oscillators and delay logical combinators (Tchuraev, 2006).

A canonical description has already been given for the cellular automaton \( A^e_x \), whose internal structure is represented by the governing gene/epigen network \( S^e_x \).

In the general form the cellular automaton \( A^e_x \) is described with five symbols \( (\Sigma, \nabla, \Omega, \Phi, \Psi) \), where \( \Sigma \) and \( \nabla \) are the input and output alphabets, \( \Omega \) is the set of internal memory states \( \Xi \), \( \Phi \) and \( \Psi \) denote the transition and output functions, respectively.

Hence, we get the following description of the discrete finite automaton \( A^e_x \). The input alphabet \( \Sigma \) of the automaton \( A^e_x \) with \( n_1 \) number of input channels constitutes a set of cortege (words of length \( n_1 \)): \( \Sigma = \{ \epsilon \} \), where \( \epsilon = \langle \epsilon_1(t), \epsilon_2(t), \ldots, \epsilon_j(t), \ldots, \epsilon_{n_1}(t) \rangle \), \( l = 1, n_1 \), and the elements \( \epsilon \) of the cortege \( \epsilon \) are the binary values. The output alphabet \( \nabla \) of the automaton \( A^e_x \) is best represented as a set of \( \gamma \)-vectors of the activities \( \Gamma(t) = \langle \gamma_1(t), \gamma_2(t), \ldots, \gamma_j(t), \ldots, \gamma_N(t) \rangle \), where \( \gamma_j = \gamma_j(t) \) denoting the activity of the gene block \( G_j \) are the elements of governing gene network \( S^e_x(G) \). In other words, at each discrete instant of time \( t \) it is possible to record the observable values, or the activities of all genes in the governing gene network \( S^e_x(G) \), i.e., for the output channels of the automaton \( A^e_x \) we consider the output channels of all its elements, not only those unconnected to the other elements of the network. Such a representation of output symbols in the cellular automaton is motivated by a possibility to have experimentally observable patterns of gene activities in the governing gene network judging, for example, by the presence (or absence) of primary transcripts. Thus, the output alphabet \( \nabla = \{ \Gamma \} \) of the cellular automaton \( A^e_x \) represents a set of all possible words \( \Gamma \) of length \( N \):

\[
\nabla = \{ \Gamma_1, \Gamma_2, \ldots, \Gamma_2 \}, \quad \text{where } \Gamma = \Gamma(t) = \langle \gamma_1(t), \gamma_2(t), \ldots, \gamma_j(t), \ldots, \gamma_N(t) \rangle
\]

is the \( \gamma \)-vector of the gene activities in the control gene network. A set of states \( \Omega \) in the memory \( \Xi \) of the automaton \( A^e_x \) is:

\[
\Omega = \{ \omega_1, \ldots, \omega_m \}, \quad m = (1, M)
\]
GENERAL PROPERTIES OF CELLULAR AUTOMATA

Since only living systems are treated as prototypes, one may exclude the previously introduced subnetwork $S^G_e(G)$ governing the processes of apoptosis (Tchuraev, 2006) from the network $S^G_e(G)$. The following statement about strong connection (1) has been derived for this case: Any cellular automaton $A^e_ζ ∈ A_ε$ (involved in cell ensemble $A_ζ$) is strongly connected.

This statement allows consequence (2): There are successions of input signals that take the cellular automaton $A^e_ζ$ from any given state $ω_ι$ to any other given state $ω_j$.

According to Harary (1969), from statement (1) it follows that the transition graph of the automaton $A^e_ζ$ relates to Euler oriented graphs and permits the application of the theorem BEST (de Bruijn, van Aardene-Ehrenfest, Smith and Tutte), in which a formula is given for the number of Euler contours in Euler graphs. Thus, we get statement (3): In any transition graph of the cellular automaton $A^e_ζ$, number $n_e$ of Euler contours is expressed by the formula:

$$n_e = c \prod_{j=1}^{M} (d_j-1)!,$$

where $d_j = id(ω_j)$ is the number of vertices (states) adjacent to $ω_ι$ (semidegree of entry), and $c$ is the common value of all cofactors of the matrix $M_{cd}$ obtained by substituting $od(ω_ι)$ for main diagonal. Here $od(ω_ι)$ is the number of states (vertices) adjacent from $ω_ι$, $M$ is the number of states of the automaton $A^e_ζ$. As seen from this formula, number $n_e$ of Euler contours in Moor’s diagrams is strongly dependent on the number of states of the cellular automaton $A^e_ζ$ and value $d_ι$ (semidegree of entry), i.e., the number of vertices (states) adjacent to $ω_ι$. According to statement (2), the semidegree of entry $d_ι = id(ω_ι) > 2$. Thus, the number of Euler contours in the transition graph of an arbitrary cellular automaton $A^e_ζ$ is likely to be sufficiently high.

What is the way of interpreting this result? Let two automata $A^e_ζ/ω_α$ and $A^e_ζ/ω_β$ ($ω_α ≠ ω_β$) enter one and the same cellular ensemble $A_ζ$. If the output symbol (signal) of one automaton, $A^e_ζ/ω_α$, is the input signal for another automaton $A^e_ζ/ω_β$, i.e., $Γ_α = ω_β$, in consequence of Euler contours occurred in Moor’s diagrams the automaton $A^e_ζ/ω_β$ can transit to one of the previous states. It should be noted that in this case the copies of automata $A^e_ζ/ω_α$ transit to another state $ω_ι$ and the development of the cell ensemble will continue with conservation of the automata subpopulation of previous (non-specialized) states – "stem cells of different tissues" and “ontogenetic variability” reserve.

From statement (1) about strong connection it also follows that any cellular automaton $A^e_ζ$ is reversible, i.e., always capable of being set in its initial state. Moreover, it can always be set in any predetermined state.
CELLULAR AUTOMATA IDENTIFICATION

The identification problem is that when an automaton is seen as a “black box” and one needs to find transition and output functions by means of measurements at its output channels. An automaton is identifiable if it can be identified irrespective of its initial state. It is obvious that if there is no sufficient information on the automaton, the general problem of its identification cannot be solved. By and large four laws of functioning and principles of organization have been stated for an arbitrary cellular automaton \( A^c \) (Tchuraev, 2006). That is why, strictly speaking, the cellular automaton \( A^c \) cannot be thought of as a “black box”, where only output responses of the automaton to input effects are observable by an experimenter. Hence, it may be believed that an arbitrary cellular automaton \( A^c \) is identifiable.

Let us next invoke general results given in (Moor, 1956) and (Gill, 1961; 1964). In keeping with the cited studies, the process of applying input sequences to automata, observation of resultant output sequences and deduction of conclusions based on these observations are said to be an experiment.

According to statement 1, any cellular automaton \( A^c \) is strongly connected. Since for automaton \( A^c \) we have a predetermined number of states \( M \), input symbols \( n_1 \) and output symbols \( 2^N \), in compliance with (Gill, 1964) we then get the following statement (4): Isolated cellular automaton \( A^c \) may always be identified by a simple unconditional experiment of length \( r \), where

\[
 r \leq \frac{(2M-1)(M 2^N)^{nM}}{(M-1)!} \exp \left[ -\frac{M(M-1)}{2(M 2^N)^n} \right]
\]

From statement (1), according to the theorem in (Gill, 1964) statement (5) also follows: Each cellular automaton \( A^c \in \mathbb{A} \) may be set in any predetermined state by a simple conditional experiment of length \( l \) and order \( d \), where

\[
 l \leq \frac{1}{2}(M+2)(M-1), \\
 d \leq M,
\]

where \( M \) is the number of states of cellular automaton inner memory.

In order to set an arbitrarily chosen automaton \( A^c \) to a predetermined state, particularly to \( z_{\omega} \) (“zygotic state”), the given automaton should be isolated from adjacent automata (“to have it cultured in vitro”).

INTERPRETATION OF CELLULAR AUTOMATON AS A FUNCTIONAL CONSTRUCTION

States \( \omega \in \Omega \) of any cellular automaton \( A^c \) encode functional information processed in the cell governing gene/epigene network, the state \( z_{\omega} \) therewith encodes the inherited functional information. Functional information is encoded in a cell by the pattern (CM-pattern) of different regulatory molecules and their complexes varying both in “age” and quantity as well as by spatial coordinates, when it has coding sense. As, according to the
description of the automaton $A_c^e$, the $\gamma$-vector value of activities $\Gamma(t)$ in neutral environment (envastat) is uniquely determined by its state, it may be stated that the $\gamma$-vector value of activities $\Gamma(t) = \langle \gamma_1(t), \gamma_2(t), \ldots, \gamma_j(t), \ldots, \gamma_h(t) \rangle$ in genetic blocks of the network $S^g_e(G)$ under envastat conditions is uniquely determined by the CM-pattern. From this statement and postulate C statement (5) follows: Every cytotype (molecular phenotype) in any cell of a given organism $x$ under envastat conditions is uniquely determined by the CM-pattern at previous moment of time. There are two consequences emerging from this state.

**Consequence (6).** A zygote cytotype of a given organism $x$ under envastat conditions is uniquely determined by the CM-pattern formed in the previous generation.

**Consequence (7).** Any somatic cell cytotype of a given organism $x$ is determined by the CM-pattern formed during ontogenesis of this generation.

**CONCLUSION**

Thus, while complete identification algorithms of cellular automaton $A_c^e$ are yet to be found, i.e., detection of its characteristic functions $\Phi$ (transitions) and $\Psi$ (outputs), it is still possible to reveal their properties, particularly by means of graph theory.

**REFERENCES**


