From Protein Interaction to Cell Population Dynamics

A theoretical framework combining Boolean networks with Markov processes illustrated in an ovarian carcinoma

Frédéric Fourré[1], Denis Baurain[2] and Hervé Barbason[3]

[1] Département de Mathématiques Appliquées, Institut d'Électricité Montefiore B28, ULg, B-4000 Liège.
[2] Département des Sciences de la Vie, Institut de Botanique B22, ULg, B-4000 Liège.
[3] Cancérologie Expérimentale, Radiobiologie et Chronobiologie B23, ULg, B-4000 Liège.
fourre@montefiore.ulg.ac.be, denis.baurain@ulg.ac.be

Abstract

At the single-cell level, cellular processes are commonly viewed as intricate biochemical networks the dynamics of which is regulated by interactions between genes and proteins. Building a theoretical framework for studying cellular processes is a hard task, mainly because of the huge number of components involved and the high complexity of the associated regulatory mechanisms. From a qualitative point of view, Boolean networks are among the most popular approaches. In this formulation, the activity level of a network component is represented by a binary variable (taking the value 0 or 1) while interactions between the components are described by Boolean switching rules. A particular feature of Boolean networks is their ability to self-organize into so-called basins of attraction that partition the state space of the network. Combining Boolean networks with Markov processes, we develop a simple stochastic model describing the time evolution of a protein-protein interaction network. To account for the inherent uncertainty of biological processes, we introduce the parameter p as a measure of the degree of intracellular disorder. This parameter is defined as the probability that one and only one component switches to its opposite activity level within a given time interval, therefore allowing transitions between basins to occur. We then propose a new generic property of the dynamic organization of these networks which concerns the probability distribution of the time spent in a basin of attraction. Using this property, we are led to a system of ordinary differential equations as a model of cell population dynamics. Furthermore, the general structure of the system is such that most macroscopic models encountered in cell population dynamics are retrieved. These findings are illustrated in an ovarian carcinoma by the simulation of a desynchronization curve in the S phase of the cell cycle that shows good agreement with BrdU labeling experimental data.

Background

CURRENT cell biology still lacks a mathematical framework for both unravelling the laws that govern cellular processes and bridging the gap between single-cell and population levels (see Figures 1 and 5). Examples of such processes include cell proliferation, differentiation and migration. Our modelling approach relies on the principle of *modularity*, where a module is defined as a protein-protein interaction network. A cellular process is viewed as a collection of modules connected through gene expression and working together to execute a function.

1. A stochastic model to depict protein interaction at the single-cell level

A powerful idealization of protein-protein interaction networks well-suited for investigating their qualitative properties is the Boolean network approach.

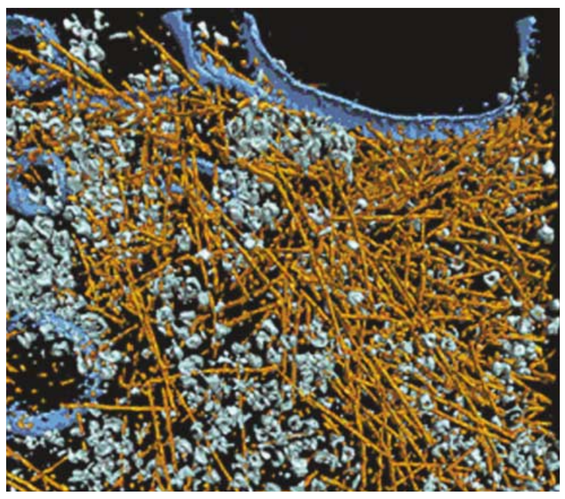

Figure 1: ***Illustration of the single-cell level.*** *Cytoplasm of an intact motile* Dictyostelium discoideum *cell. The orange linear complexes are actin filaments; ribosomes and other macromolecular complexes are in grey; membranes are in blue.*

Modelling aspects

1. m protein species (including complexes) $\rightarrow$ m nodes;
2. the activity state of a node is represented by a Boolean variable (0 or 1) $\rightarrow$ the state of the network is described by a bit vector of length m and the size of the state space is 2^m;
3. interactions among the nodes are modelled using Boolean switching rules that determine the output of the nodes at time $n+1$ according to their inputs at time n, $n = 0, 1, 2, \ldots$

Important features

1. the time evolution of the network activity is deterministic;
2. state space self-organizes into subsets of states called basins of attraction, each containing an attractor which captures the long-term behaviour of the system;
3. transiently flipping the activity of single nodes (unit perturbations) allows transitions between basins.

Illustration of modularity in budding yeast

Let us consider a simple module in budding yeast (*Saccharomyces cerevisiae*) that concerns Start events controlling the G_1/S transition of the cell cycle. The associated Boolean network consists of three nodes: Cln2/Cdc28, Clb5/Cdc28 and Sic1, where Cln2 and Clb5 are cyclins, Cdc28 a cyclin-dependent kinase, and Sic1 a cyclin-dependent kinase inhibitor. Update rules are chosen to be respectively:

$$\xi_1, \quad \neg\xi_3, \quad \neg\xi_1, \tag{1}$$

where $\neg$ represents the NOT operator (see Figures 2 and 3).

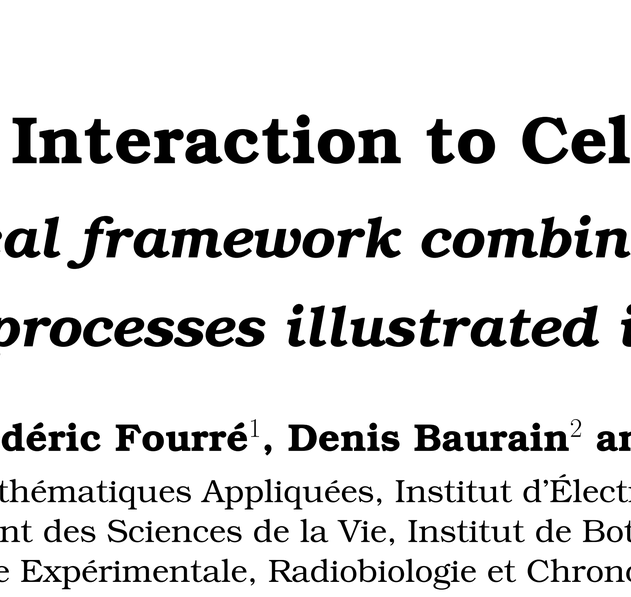

Figure 2: ***Illustration of modularity in budding yeast (Saccharomyces cerevisiae) using Boolean networks.*** *Start events controlling the G_1/S transition of the cell cycle. Active and inactive states are in green and red, respectively.*

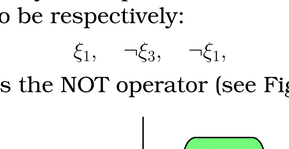

Figure 3: ***State diagram corresponding to rules (1).*** *The 2^3 states in state space are organized into 2 basins of attraction, namely B_1 and B_2. Dashed arrows indicate transitions between basins.*

Accounting for stochasticity of biological systems

We introduce the parameter $0 < p < 1$ which is a measure of the degree of *intracellular disorder*. This is the probability that one and only one of the m bits switches transiently to its opposite value within Δt, where Δt is the time interval between any two successive updates. Each bit is supposed to switch with the same probability (p/m). The following assumptions are made:

1. p is time independent and depends only on the length of Δt;
2. the probability of more than one unit perturbation in time interval Δt is negligibly small compared to p;
3. unit perturbations occur independently of each other.

The probability that the first unit perturbation occurs on the kth time step is:

$$g_k = p(1-p)^{k-1}, \quad k = 1, 2, \ldots, \tag{2}$$

which is the geometric distribution with mean $\mu = 1/p$. The resulting semi-random walk of the cell in state space is a discrete-time Markov chain $\{X_n, n = 0, 1, 2, \ldots\}$ with state space $\{1, 2, 3, \ldots, 2^m\}$. The time-dependent probabilities $z_i^{(n)}$ of a cell being in state i at time n can be obtained through the following vector equation:

$$\mathbf{z}^{(n+1)} = \mathbf{z}^{(n)}\Pi, \quad n = 0, 1, 2, \ldots, \tag{3}$$

with $\mathbf{z}^{(n)}$ the vector of probabilities $z_i^{(n)}$, $\mathbf{z}^{(0)}$ an initial state vector, and Π the matrix of transition probabilities.

Computing ψ_s sojourn time distributions

As a new generic property of random Boolean networks, the time spent in a basin of attraction closely follows a geometric distribution provided $p \ll 1$ (see Figure 4).

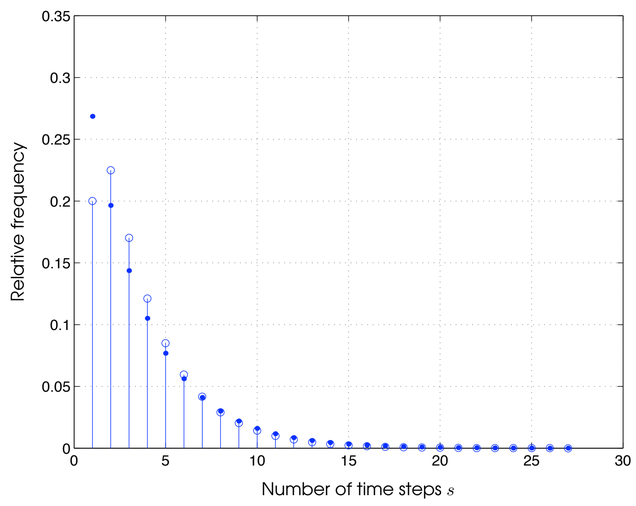

Figure 4: ***Sojourn time probability distributions in a basin of attraction.*** *The initial state vector $\mathbf{z}_\beta^{(0)}$ is uniform and $p = 0.3$. Dots: exact $\psi_{s,u}$ Markov distribution with $\mu_{e,u} = 3.72$; open circles: approximate $g_{s,u}$ geometric distribution with parameter $p_{e,u} = 0.27$. The maximum discrepancy between $\hat{\psi}_{s,u}$ and $\hat{g}_{s,u}$ is 6.86%.*

2. Bridging the gap to the population level

Moving from one basin to another with $\Delta t \rightarrow 0$

Since sojourn time distributions are quasi-geometric, the process which consists of moving from one basin to another is quasi-Markovian. The chain $\{X'_n, n = 0, 1, 2, \ldots\}$ with state space $\{1, 2, 3, \ldots, \beta\}$, where β is the number of basins that compose the network, is a subchain of $\{X_n\}$ ($\beta < 2^m$). Let us denote by $w_j^{(n)}$ the state probability of being in basin j at time n. If $p \ll 1$ then $w_j^{(n)} \rightarrow w_j(t)$ as $\Delta t \rightarrow 0$ and $w_j(t)$ satisfies the Kolmogorov differential equation:

$$\frac{d}{dt}w_j(t) = -k_{jj}w_j(t) + \sum_{j' \neq j} k_{j'j}w_{j'}(t), \tag{4}$$

where the k's are the transition rates.
Therefore, protein-protein interaction networks can be depicted using ordinary differential equations provided cell number be greater than 10^2, so that fluctuations can be ignored.

Coming back to cellular processes...

Below, we propose a system of ordinary differential equations to describe the progression of cells in their cycle:

$$\begin{aligned}
\frac{dy_1}{dt} &= 2k_v y_v - k_1 y_1 + k'_1 y_2 \\
\frac{dy_l}{dt} &= k_{l-1}y_{l-1} + k'_l y_{l+1} - (k'_{l-1} + k_l)y_l, \quad l = 2, 3, \ldots, v-1 \\
\frac{dy_v}{dt} &= k_{v-1}y_{v-1} - k'_{v-1}y_v - k_v y_v,
\end{aligned} \tag{5}$$

where the cell cycle is partitioned into v compartments and $y_l(t)$ is the number of cells in compartment l at time t.

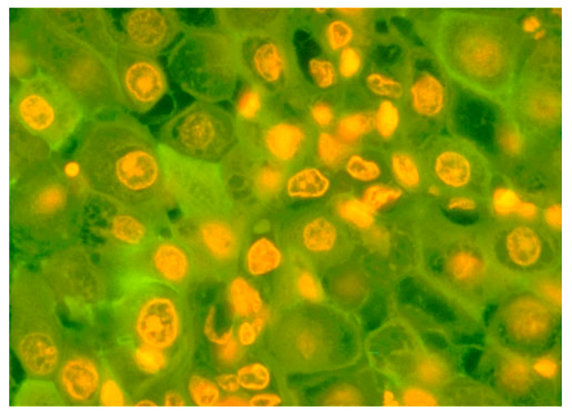

Figure 5: ***Illustration of the population level.*** *Adenocarcinoma of the human salivary gland. Nuclei appear orange on a green background of cellular components when stained with propidium diiodide and fluorescein isothiocyanate.*

Illustration of cell-to-cell variability

We compare system (5) with Lotka-McKendrick equation for age-structured populations:

$$\frac{\partial}{\partial t}z(a,t) + \frac{\partial}{\partial a}z(a,t) = -\eta(a)z(a,t), \quad 0 < a < \infty, \quad t > 0, \tag{6}$$

with $z(a,t)$ the age density at time t. From (6), Chiorino et al. (2001) have established the formula:

$$I_s(t) = \bar{I}_s + e^{-Rt}C\cos\left(\frac{2\pi}{T}t - \phi\right), \quad t \geq 0, \tag{7}$$

where $I_s(t)$ is the labeling index at time t.

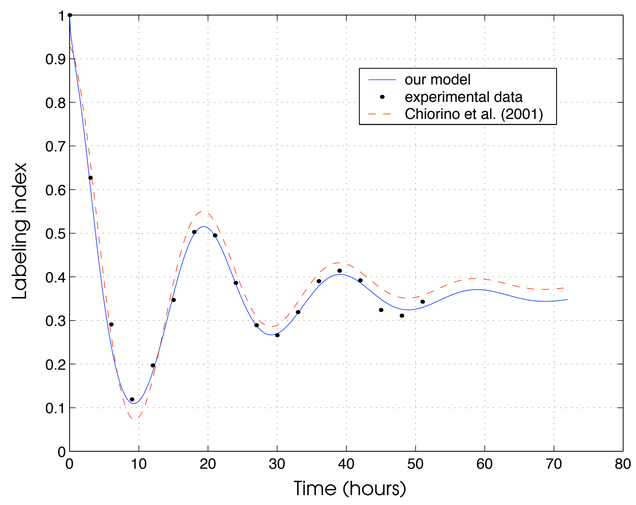

Figure 6: ***Desynchronization of cells in their cycle.*** *BrdU labeling index of an ovarian carcinoma* in vitro *under the assumption of asynchronous exponential growth. Both curves are fitted by the least-squares method. (– –), Chiorino et al. (2001) J. Theor. Biol. 208, 185-99: $\bar{I}_s = 0.364$, $R = 0.074$, $C = 0.597$, $T = 19.73$ h, $\phi = 0.0546$ and $r_2 = 1.30 \times 10^{-2}$; (—), system (5): $v = 201$, $v_1 = 90$, $v_2 = 80$, $k_l = 63.5$ h^{-1}, $k'_l = 53.7$ h^{-1} and $r_2 = 4.57 \times 10^{-3}$.*

Conclusion

1. intracellular disorder p is interpreted as the primary cause of cell-to-cell variability;
2. our model can also accounts for distribution of cell cycle times, fraction labeled mitosis analysis and rhythms in cell populations (data not shown).

BIOFORUM 2004, Université de Liège, Liège, 7 December 2004

D. Baurain is a Postdoctoral Researcher of the FNRS.