



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

C. R. Biologies 326 (2003) 171–188



Biological modelling / Biomodélisation

Genetic regulation networks: circuits, regulons and attractors

Réseaux de régulation génétique : circuits, régulons, attracteurs

Jacques Demongeot*, Julio Aracena, Florence Thuderoz, Thierry-Pascal Baum,
Olivier Cohen

Institut universitaire de France & CNRS TIMC-IMAG, Faculty of Medicine, 38700 La Tronche, France

Received 15 July 2002; accepted 4 December 2002

Presented by Jacques Ricard

Abstract

We deal in this paper with the concept of genetic regulation network. The genes expression observed through the bio-array imaging allows the geneticist to obtain the intergenic interaction matrix \mathbf{W} of the network. The interaction graph G associated to \mathbf{W} presents in general interesting features like connected components, gardens of Eden, positive and negative circuits (or loops), and minimal components having 1 positive and 1 negative loop called regulons. Depending on parameters values like the connectivity coefficient $K(\mathbf{W})$ and the mean inhibition weight $I(\mathbf{W})$, the genetic regulation network can present several dynamical behaviours (fixed configuration, limit cycle of configurations) called attractors, when the observation time increases. We give some examples of such genetic regulation networks and analyse their dynamical properties and their biological consequences. *To cite this article: J. Demongeot et al., C. R. Biologies 326 (2003).*

© 2003 Published by Académie des sciences/Éditions scientifiques et médicales Elsevier SAS.

Résumé

Cet article porte sur le concept de réseau de régulation génétique. L'expression génique, fournie par l'imagerie *bio-array*, permet d'obtenir la matrice \mathbf{W} d'interaction intergénique du réseau. Le graphe d'interaction G associé à \mathbf{W} présente en général des caractéristiques importantes telles que composantes connexes, jardins d'Eden, circuits (ou boucles) positifs et négatifs, ainsi que composants minimaux possédant une boucle négative et une boucle positive, appelés régulons. En fonction des valeurs de certains paramètres, tels que le coefficient de connectivité $K(\mathbf{W})$ et le poids moyen d'inhibition $I(\mathbf{W})$, le réseau de régulation génétique peut présenter différents comportements dynamiques (configuration fixe ou cycle limite de configurations) appelés attracteurs, lorsque le temps d'observation augmente. Nous donnerons des exemples de tels réseaux et analyserons leurs propriétés dynamiques, ainsi que leurs conséquences biologiques. Dans la partie consacrée à l'acquisition d'images *bio-array*, nous rappelons rapidement quelles sont leurs caractéristiques en termes de bruit et de signal et nous proposons une méthode (dite de l'emboutissage gaussien) permettant de les standardiser. Ensuite, nous donnons une méthode (dite des corrélations directionnelles) permettant d'extraire, à partir des images de co-expression des gènes, la matrice d'interaction inter-génique \mathbf{W} liée à l'activité du réseau étudié. Puis, après description des caractéristiques majeures de son graphe associé G , nous donnons

* Corresponding author.

E-mail address: Jacques.Demongeot@imag.fr (J. Demongeot).

une suite de propositions, lemmes et théorèmes permettant de faire le lien entre la phénotypie observée (configurations fixes ou cycles de configurations au cours du cycle cellulaire des tissus étudiés) et les contraintes qu'elle exerce sur la structure interne de \mathbf{W} et donc de G . Les deux résultats majeurs sont que l'existence d'au moins une boucle positive est une condition nécessaire de l'existence de plus d'un attracteur et que, si le nombre N de gènes est suffisamment grand et que $K(\mathbf{W}) = 2$, alors le nombre total $A(\mathbf{W})$ d'attracteurs est de l'ordre de \sqrt{N} et de toute manière inférieur à 2^m , où m est le nombre de boucles positives du réseau de régulation génique. Nous donnons ensuite les exemples des réseaux de régulation contrôlant la floraison d'*Arabidopsis thaliana*, la gastrulation, la fonction lytique du phage μ et la préséance des bourgeons axillaires de *Bidens pilosa*, dans lesquels nous retrouvons la mise en œuvre des principales notions introduites dans les parties précédentes de l'article. **Pour citer cet article : J. Demongeot et al., C. R. Biologies 326 (2003).**

© 2003 Published by Académie des sciences/Éditions scientifiques et médicales Elsevier SAS.

Keywords: genetic regulation network; intergenic interaction matrix; positive loops; regulons; attractors

Mots-clés : réseau de régulation génétique ; matrice d'interaction intergénique ; boucles positives ; régulons ; attracteurs

1. Introduction

During the recent years, the rapid development of the bio-arrays techniques [1] based on isotopic or fluorescent activity of hybridised DNA chips allowed the biologist to give to a grey level peak the signification of an expression rate for the genes studied in the bio-array. If we repeat this acquisition at different times of the cell cycle for different cells of a same tissue, we can calculate correlations between the genes expression rate and hence we are able to make explicit a matrix \mathbf{W} called the intergenic interaction matrix, representing the repression and induction influences a gene can exert on other genes.

1.1. The raw data from the bio-array imaging

The first encountered problem with the bio-array image is the noise and we have to low-pass it in order to suppress the high-frequency noise (see Fig. 1). The result of this pre-treatment is a better separation of

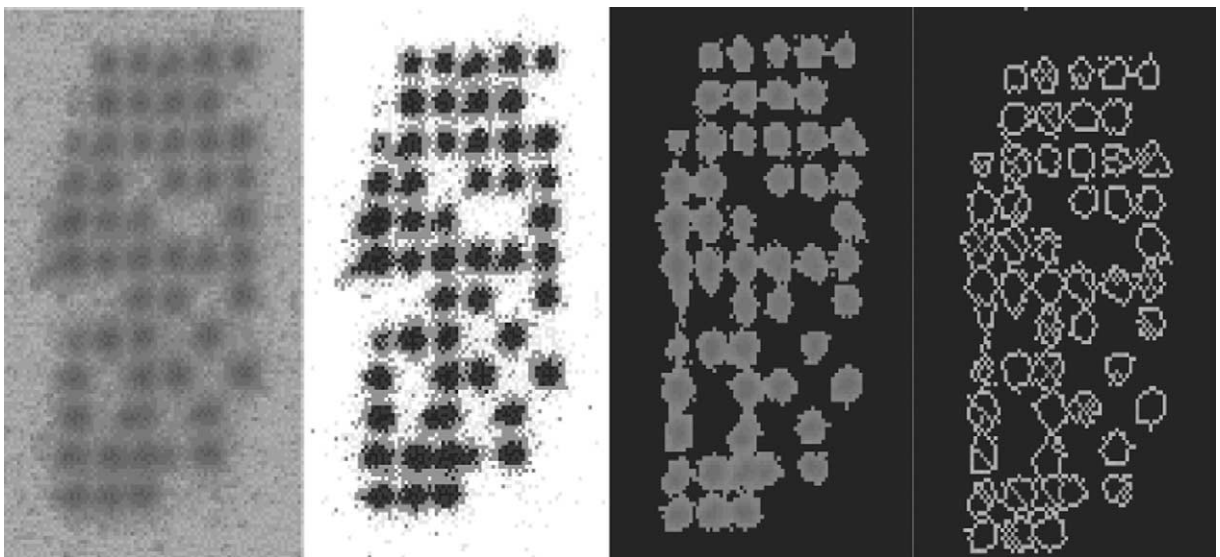


Fig. 1. (a) Raw data; (b) low-pass filtering; (c) watershed segmenting; (d) contouring.

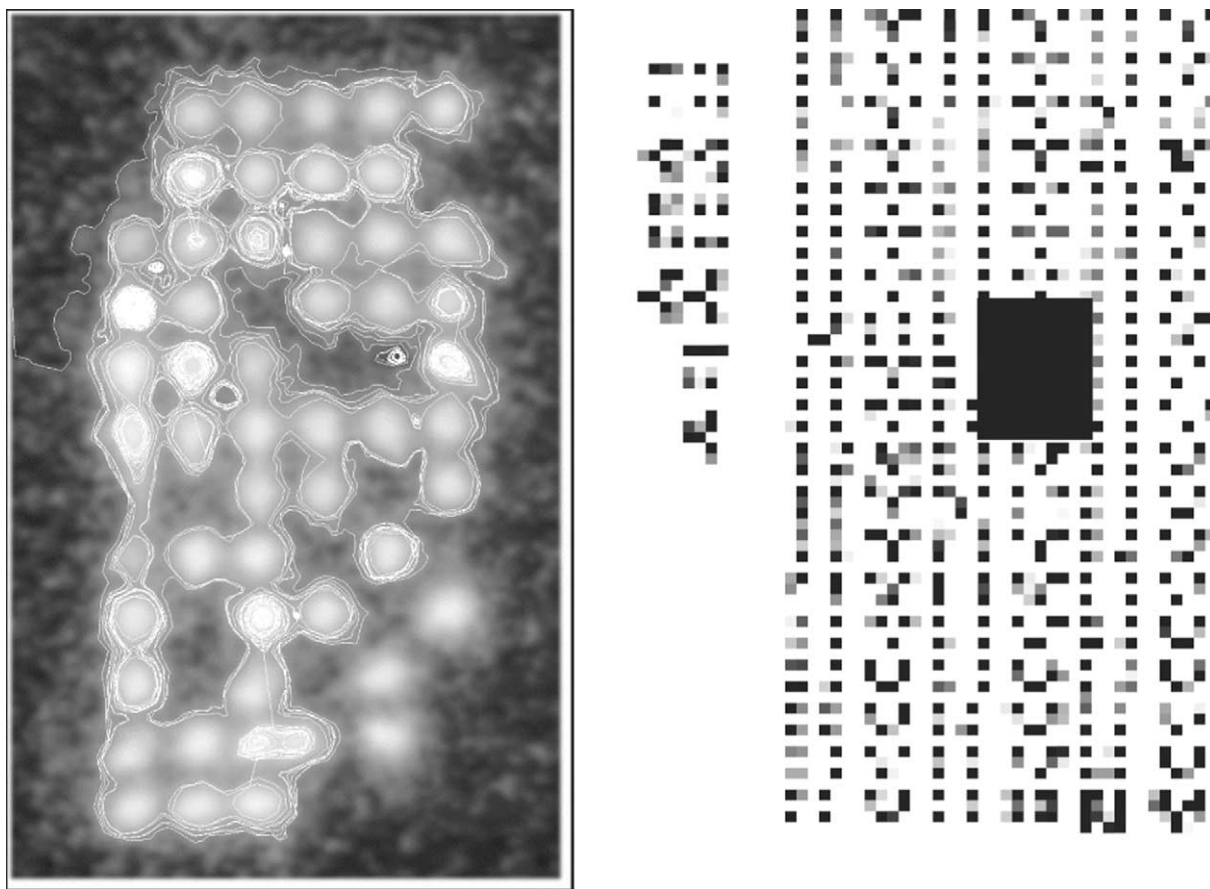


Fig. 2. 'Gaussian-stamping' contours (left) and standardized bio-array image (black rectangle right corresponds to the left image).

the isotopic activity peaks, allowing a watershed separation and contouring [2,3]. We can also apply a more accurate segmentation and contouring method called potential-Hamiltonian or 'Gaussian-stamping' method. Let us remark that the peaks are about Gaussian, with a weak kurtosis and skewness, allowing in particular the respect of the conservation 'law': 2/3 of the activity peak are concentrated into the set of points (x, y) where the Gaussian curvature $C(x, y)$ vanishes, i.e. inside the maximum gradient line, called the characteristic line of the peak. Then it is possible to neglect the part of the peak outside the projection of this line, whose equation is $C(x, y) = \partial^2 g / \partial x^2 \partial^2 g / \partial y^2 - (\partial^2 g / \partial x \partial y)^2 = 0$, $g(x, y)$ denoting the grey level function at the pixel of coordinates (x, y) .

We are thus led to consider the new height function $C(x, y)$ instead of the function $g(x, y)$ and its level line $C(x, y) = 0$. A new algorithm has been proposed in [1] to obtain automatically the characteristic line and, after, by integrating inside the projection of this line on the (x, y) plane and multiplying by 3/2 the obtained result, to standardize the estimated activity in terms of a bio-image with small squares symbolizing in grey levels the degree of hybridisation of the cDNA's expressing the regulation of a glioma tissue (Fig. 2). From such bio-array images acquired in cells of the same tissue at different times of the cell cycle, we can study the interactions between genes by estimating an interaction matrix.

2. Some rigorous results about the network attractors

2.1. The interaction matrix \mathbf{W}

The interaction matrix \mathbf{W} is similar to the synaptic weight matrix, which rules the relationships between neurons in a neural network. The general coefficient w_{ik} of such an interaction matrix \mathbf{W} is equal to +1 (resp -1, 0) if the gene G_k activates (resp inhibits, does not influence) the gene G_i , the state x_i of the gene G_i being equal to +1 (resp -1), if it is (resp is not) expressed. In the case of small regulatory genetic systems (called *operons*), the knowledge of such a matrix \mathbf{W} permits to make explicit all possible stationary behaviours of the organisms having the corresponding genome. The change of state of gene G_i between t and $t + 1$ obeys a threshold rule: $x_i(t + 1) = H(\sum_{k=1,n} w_{ik}x_k(t) - b_i)$ or $x(t + 1) = H(\mathbf{W}\mathbf{x}(t) - \mathbf{b})$, where H is the sign step function ($H(y) = 1$, if $y \geq 0$ and $H(y) = -1$, if $y < 0$) and the b_i s are threshold values. When t is increasing, the genes states reach a stable set of configurations (a fixed configuration or a cycle of configurations), called attractor of the genetic network dynamics. For example, in the regulatory network that regulates the *Arabidopsis thaliana* flower morphogenesis, the interaction matrix is a (11, 11)-matrix with only 22 non-zero coefficients (see Fig. 3). This matrix presents $P(\mathbf{W}) = 4$ positive loops and $A(\mathbf{W}) = 4$ attractors (see Section 3.1). Hence it is in general of a great biological interest and relevance to determine matrices having characteristic properties like (i) a minimal number of non-zero coefficients for a given set of attractors (fixed points or cycles) or (ii) a minimal number $P(\mathbf{W})$ of positive loops, controlling the number $A(\mathbf{W})$ of attractors and their stability (cf. [4–9] for the continuous case [10–12] and for the discrete one). In the following, we intend to partly solve the problems taken above by giving necessary and sufficient conditions to obtain properties (i) and (ii). In order to calculate the w_{ik} s, we can either determine the s -directional correlation $\rho_{ik}(s)$ between the state vector $\{x_k(t - s)\}_{t \in C}$ of gene j at time $t - s$ and the state vector $\{x_i(t)\}_{t \in C}$ of gene i at time t , t varying during the cell cycle C of length $M = |C|$ and corresponding to observation times of the bio-array images:

$$\rho_{ik}(s) = \left(\sum_{t \in C} x_k(t - s)x_i(t) - \sum_{t \in C} x_k(t - s) \sum_{t \in C} x_i(t)/M \right) / \sigma_k(s)\sigma_i(0)$$

where

$$\sigma_k(s) = \left(\sum_{t \in C} x_k(t - s)^2 - \left(\sum_{t \in C} x_k(t - s) \right)^2 / M \right)^{1/2}$$

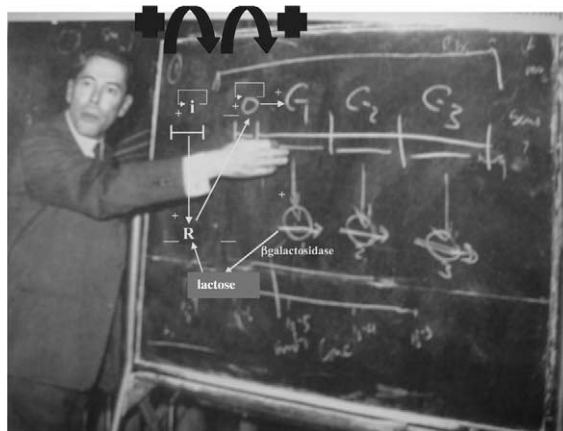


Fig. 3. The lactose operon exhibited by Jacques Monod.

and then take $w_{ik} = \text{sign}(\sum_{s=1, \dots, m} \rho_{ik}(s)/M)$, if $|w_{ik}| > \eta$, and $w_{ik} = 0$, if $|w_{ik}| < \eta$, where η is a de-correlation threshold, or identify the system with a Boolean neural network. When it is impossible to so obtain all the coefficients of \mathbf{W} (neither from the literature nor from such calculations), it is possible to complete \mathbf{W} : we choose randomly the missing coefficients by respecting the connectivity coefficient $K(\mathbf{W}) = I/N$, the ratio between the number I of interactions and the number N of genes, and the mean inhibition weight $I(\mathbf{W}) = R/I$, the ratio between the number of inhibitions (or repressions) R and $I \cdot K(\mathbf{W})$ is in general between 1.5 and 3 and $I(\mathbf{W})$ between 1/3 and 2/3 for many known operons or regulatory networks (lactose operon, Cro operon for the phage δ , lysogenic/lytic operon for the phage μ , gastrulation and *Arabidopsis thaliana* flowering regulatory networks...).

If G is the interaction graph associated to \mathbf{W} , then we call connected component C of G each set of genes C such that there is a path between every pair of genes of C along a sequence of arcs of G . A Garden of Eden is a gene receiving no arc, but influencing at least one other gene. A regulon is a connected component of G having exactly one positive (auto-catalysis) and one negative loop, these loops sharing the auto-catalysed node. In the lactose operon (see its G in Fig. 3), $K(\mathbf{W}) = 8/6$, $I(\mathbf{W}) = 3/8$, $P(\mathbf{W}) = 2$, $A(\mathbf{W}) = 2$ (β gal-activated and inactivated states), and G has one connected component and one regulon.

2.2. Some definitions and notations

In the following, we give some definitions about the rigorous mathematical description of the discrete Boolean networks used to describe the genes interaction dynamics, and their associated graph G and matrix \mathbf{W} . Then we will present some theoretical results with rigorous proofs only for the first and last results in order to show what kind of reasoning we have to perform and we will refer to [13–16] for complete demonstrations. Let us consider a graph $G = (V, E)$, where $V = \{1, \dots, n\}$ is the set of nodes and E is the set of arcs. Let $\mathbf{W} = (w_{ik})$ be a real (n, n) -matrix. We call G the incidence graph of \mathbf{W} , if (k, i) , the arc going from k to i , belongs to E , then $w_{ik} \neq 0$. By extension, \mathbf{W} will also be called the incidence matrix of G . We will define the sign of an arc (k, i) , denoted by $\text{sign}((k, i))$, as the sign of w_{ik} . Let us denote by $\Gamma^-(i)$ (resp $\Gamma^+(i)$) the set of nodes $\{i_1, i_2, \dots, i_{k(i)}\}$ such that (i_j, i) (resp (i, i_j)) belongs to E , for each $j = 1, \dots, k(i)$. We will say that a set of arcs $C = \{e_1, e_2, \dots, e_r\}$ is a chain if each e_k in C has a node belonging to e_{k-1} and the other one belonging to e_{k+1} . We will say that C is a simple (resp elementary) chain if the arcs (resp nodes) are different. In the sequel we will understand by chain a simple and elementary chain. In the same way we will call C a path if $e_k = (i_k, i_{k+1})$ implies $e_{k+1} = (i_{k+1}, i_{k+2})$, for all $k = 1, \dots, r$, that is to say the final node of each arc is the beginning node of the next arc in C . The sign of a path or a chain C (denoted by $\text{sign}(C)$) is positive if the number of negative arcs of C is even and negative if this number is odd. A cycle (resp circuit or loop) is defined as a chain (resp path) where each of the two extremities of any arc belongs to two and only two arcs. For simplicity of notation, we will say that a node i belongs to a cycle C if there exists a node j such that (i, j) or (j, i) belongs to C . Every other definition of graph theory used here will be consistent with that in [17,18]. We will call a circuit or cycle C negative (resp positive) if $\text{sign}(C)$ is negative (resp positive). Define now a discrete state regulatory network, acting on the set of states $\{-1, 1\}$, here and subsequently denoted by N , as the 4-uple $N = (G, \mathbf{W}, b, \text{sign})$, where G is the incidence graph of \mathbf{W} , b is a threshold real vector and the local transition function is given by $x_i(t + 1) = \text{sign}(\sum_{k=1, \dots, n} w_{ik}x_k(t) - b_i)$, $\forall \mathbf{x} \in \{-1, 1\}^n$, where $\text{sign}(u) = 1$ if $u \geq 0$ and $\text{sign}(u) = -1$ otherwise. The sequential iteration consists to update the nodes one by one in a prescribed periodic update $I = (i(1), i(2), \dots, i(n))$, where $\{i(1), i(2), \dots, i(n)\} = \{1, 2, \dots, n\}$, that is to say, starting with a given $\mathbf{x}(0) = (x_1(0), \dots, x_n(0))$ in $\{-1, 1\}^n$, we generate a sequence of iterates:

$$x_{i(k)}(t) = \text{sign}\left(\sum_{j < k} w_{i(k)i(j)}x_{i(j)}(t) - b_{i(k)} + \sum_{j \geq k} w_{i(k)i(j)}x_{i(j)}(t - 1) - b_{i(k)}\right), \quad \forall k \in \{1, \dots, n\}$$

Now, the parallel iteration consists in updating all the nodes synchronously:

$$x_i(t + 1) = \text{sign}\left(\sum_{k=1, \dots, n} w_{ik}x_k(t) - b_i\right), \quad \forall i \in \{1, \dots, n\}, \quad \text{with } \mathbf{x}(0) \in \{-1, 1\}^n$$

We shall say that \mathbf{x} is a fixed point if it is invariant under the application of the complete sequence of updates. Observe that the kind of iteration does not change the set of fixed points, but only change their attraction basins. In the following we will use systematically the parallel iteration.

2.3. Relations between positive and negative cycles, and fixed points

In the sequel, we will assume that the graph G is connected, since otherwise one can apply the results to each of connected components of G . In addition, we will suppose with no loss of generality that $|\Gamma^-(i)| > 0$, for all $i \in V$. Since otherwise, if there exists a node $i \in V$ such that $\Gamma^-(i)$ is empty, then we can assume that the arc (i, i) exists in E ; in this way the dynamics of both networks are the same. It evidently follows from this property that there exists at least one circuit C in G (possibly a circuit of the form (i, i)). Finally, we suppose that the graph G and the matrix \mathbf{W} have a quasi-minimal structure, that is to say, all (j, i) , such as $i \neq j$, belong to E (or equivalently $w_{ij} \neq 0$, if $i \neq j$), if there exists $\mathbf{x} \in \{-1, 1\}^n$, such that:

$$\text{sign}\left(\sum_k w_{ik}x_k - b_i\right) \neq \text{sign}\left(\sum_{k \neq j} w_{ik}x_k - b_i\right)$$

Hence, we have the following necessary condition to have a quasi-minimal structure:

$$-\sum_k |w_{ik}| < b_i \leq \sum_k |w_{ik}|, \quad \forall i \in 1, \dots, n$$

The following property will be very useful in the following for characterizing a cycle.

Proposition 1. *A cycle C is positive if and only if there exists a vector $\mathbf{x} \in \{-1, 1\}^n$ such that for all $(k, i) \in C$, $\text{sign}(w_{ik}) = x_i x_k$ or equivalently, for all $(k, i) \in C$, $x_i = \text{sign}(w_{ik})x_k$ (1).*

Proof. Let C be a positive cycle and $i(0)$ a fixed node belonging to C . Let us enumerate the nodes belonging to C by $i(0), i(1), \dots, i(j)$, such that $\forall k = 0, \dots, j$, $(i(k-1), i(k))$ or $(i(k), i(k-1)) \in C$ (by identifying j and -1). Finally, let us define the vector \mathbf{x} as follows:

$$\begin{aligned} x_{i(0)} &= 1 \quad \text{and} \quad x_{i(k)} = \text{sign}(w_{i(k)i(k-1)})x_{i(k-1)} \quad \text{if } (i(k-1), i(k)) \in C \quad \text{or} \\ x_{i(k)} &= \text{sign}(w_{i(k-1)i(k)})x_{i(k-1)} \quad \text{if } (i(k), i(k-1)) \in C, \quad \forall k = 1, \dots, j \end{aligned}$$

Obviously \mathbf{x} is satisfying equation (1). Hence $-\mathbf{x}$ satisfies equation (1) too. Finally, it is direct that there does not exist another vector $\mathbf{y} \notin \{\mathbf{x}, -\mathbf{x}\}$ that satisfies equation (1).

Let C be now a negative cycle, and let us suppose that equation (1) is true, then $\prod_{(j,i) \in C} \text{sign}(w_{ij}) = \prod_j x_j \prod_i x_i = (\prod_j x_j)^2$, but $\text{sign}(C) = \prod_{(j,i) \in C} \text{sign}(w_{ij}) < 0$, which is contradictory. \square

Theorem 1. *Given N , if all cycles of the incidence graph G are positive, then there exists a vector $\mathbf{x} = (x_1, \dots, x_n) \in \{-1, 1\}^n$ such that \mathbf{x} and $-\mathbf{x} = (-x_1, \dots, -x_n)$ are fixed points of N .*

Remark. There are two remarkable fixed points having by construction a non-frustration property, that is on each cycle the sign changes of x_i s are identical to the sign changes of the arcs. For other possible fixed points, there is at least one cycle for which sign changes are frustrated.

Theorem 2. *If all circuits of the incidence graph G are negative, then N has no fixed points.*

2.4. Minimal regulatory networks

The previous results allow us to characterize some minimal regulatory networks. The following propositions constitute examples of minimal regulatory networks. They solve in part the inverse problem consisting in the description of \mathbf{W} only from the knowledge of a phenotypic \mathbf{x} observed from bio-array images.

Proposition 2. *Let N having n nodes and n connections, a necessary and sufficient condition of existence of a fixed point \mathbf{x} is the existence of a positive circuit. In this case, \mathbf{x} and $-\mathbf{x}$ are both fixed points. Hence we can characterize the set of minimal N s having \mathbf{x} as fixed point.*

Proposition 3. *Given a state vector \mathbf{x} , the set of minimal networks $N = (G, \mathbf{W}, b, \text{sign})$ having \mathbf{x} as fixed point is given by the following conditions:*

- (1) $w_{ik} = \alpha_{ik} x_i x_k$, where $\alpha_{ik} \geq 0$ and, for all i , there exists $k(i)$ such that $\alpha_{ik(i)} \neq 0$ and
- (2) $-|\alpha_{ik(i)}| < b_i \leq |\alpha_{ik(i)}|$.

Proposition 4. *Let N with n nodes and $n + 1$ connections, a necessary and sufficient condition for existence of an attractor of all points parallel iterated, is a negative circuit and a positive circuit intersecting.*

2.5. Fixed points bounds in regulatory networks

Given N , let C be a positive circuit of N , then by Proposition 1, there exists $\mathbf{x} \in \{-1, 1\}^{|\mathbf{V}(C)|}$, such that \mathbf{x} and $-\mathbf{x}$ satisfy the equation: $\forall (k, i) \in C, \text{sign}(w_{ik}) = x_i x_k$. We denote $\mathbf{u}(C) \in \{-1, 1\}^{|\mathbf{V}(C)|}$ the vector defined by: $\mathbf{u}(C) = \mathbf{x}$ (resp $-\mathbf{x}$), if $x_{i(0)} = 1$ (resp -1), where $i(0) = \min\{i \mid i \in C\}$.

Lemma 1. *Given N and \mathbf{y} a fixed vector of N , then for all $i \in \mathbf{V}$, there exists a positive circuit $C(i)$ in G such that for all k in $C(i)$, $y_k = u(C(i))_k$ or for all k in $C(i)$, $y_k = -u(C(i))_k$.*

Theorem 3. *If m is the total number of positive circuits of N , then the number of fixed points of N is $\leq 2^m$, and this upper bound is reached if and only if for all circuits C of N there does not exist an arc (k, i) in C^C ending in C (there is no garden of Eden k pending to C).*

Remark. We have to notice that the condition concerns the number of circuits and not of cycles, these last being in general very more numerous.

2.6. Asymptotic mean value for the number of fixed configurations (fixed vectors or limit cycles of vectors) in the case $K(\mathbf{W}) = 2$

Let us consider now a network N having n nodes and $2n$ connections such as $K(\mathbf{W}) = 2$. We search for a mean value of the number of fixed configurations, when n is growing to infinity.

Lemma 2. *For any graph G having m non-oriented edges, the mean number of oriented edges we can define on G from the non-oriented configuration is equal to $4m/3$.*

Proof. Let us note $\langle o \rangle$ the mean number of oriented edges we can construct from a configuration of m non-oriented edges; then, if exactly k from the m non-oriented edges are decomposed into two oriented opposite connections, we have $C_m^{m-k} 2^{m-k}$ different ways to dispatch the not double connections into the $(m - k)$ other non-oriented edges; hence we can write: $\langle o \rangle = \sum_{k=0}^m (k + m) C_m^{m-k} 2^{m-k} / \sum_{k=0}^m C_m^{m-k} 2^{m-k} = 4m/3$. \square

Theorem 4. *If the network N has n nodes and K(W)n connections, with K(W) = 2, then the expectation of the number of fixed configurations of N is n^{1/2}, if n is sufficiently large.*

Proof. Following [19], if the connections of N are random, and if the mean number *c* of non-oriented edges per node is equal to 3/2, then the random variables *X_i* equal to the number of disjoint cycles of length *i* of N are independent and Poissonian with parameter $\lambda(i) = 2^{i-1}/i$, if *n* is sufficiently large. From Lemma 2, we are just in this case, because we have $2n = 4m/3$ connections, hence $m = 3n/2$ and $c = m/n = 3/2$. Then we have, for the mean number $\langle f \rangle$ of fixed configurations of N: $\langle f \rangle = \sum_{s=0}^n \sum_{k=s}^n \sum_{\sigma \in \Omega(s,k)} A(\sigma) \Pi_{\sigma}$, where $\Omega(s, k) = \{\sigma = (s(1), \dots, s(n)) / s(i) \geq 0, \sum_{i=1}^n s(i) = s, \sum_{i=1}^n i s(i) = k\}$, $\Pi_{\sigma} = P(\{X_i = s(i), s(i) \geq 0, \sum_{i=1}^n s(i) = s, \sum_{i=1}^n i s(i) = k\}) = e^{-\sum \lambda(i)} \prod_{i=1}^n \lambda(i)^{s(i)} / s(i)!$ is the probability to have the *X_i*s equal each to *s(i)*, and *A(σ)* is the mean number of fixed configurations, when each *X_i*s is equal to *s(i)*.

We will now evaluate the expectation *A(σ)*. Each disjoint positive circuit bringing two fixed points (Theorem 3 above), an isolated positive non-circuit cycle bringing also two fixed points and an isolated negative circuit bringing one limit cycle, we can first calculate *A(0, σ)*, the expected number of fixed configurations in the case where we have only disjoint positive circuits, the rest of the nodes depending on these circuits (and hence their states being fixed by the states of the circuit): $A(0, \sigma) = B(0, \sigma) / D(\sigma)$, where $B(0, \sigma) = 2^s$ (number of fixed points of $s = \sum_{i=1}^n s(i)$ disjoint positive circuits, from Theorem 3 above) $\times 2^k$ (number of different signs for each of the $k = \sum_{i=1}^n i s(i) \leq n$ nodes involved in the *s* circuits) $\times [2^s$ (number of different directions – left or right – for each of the *s* circuits) / 2^s (reduction factor for having only positive circuits)] $\times N(\sigma)$, $D(\sigma) = 2^k$ (number of different directions for each of the *k* connections) $\times 2^k$ (number of different signs for each of the *k* connections) $\times N(\sigma)$, where *N(σ)* is the number of choices for the *s* disjoint cycles: $N(\sigma) = C_k^{s(1), \dots, s(n)} \prod_{i=1}^n (i-1)!^{s(i)} / 2^s$. *N(σ)* just equals the number of choices of *k* nodes in *s(1)* subsets of size 1, ..., *s(n)* of size *n* times the number of choices of different loops (without multiple points) connecting the vertices inside each of these subsets.

In the same way, we can calculate *A(1, σ)* (resp *A(j, σ)*) the expected number of attractors of N in the case where we have among the *s* disjoint cycles 1 (resp *j*) isolated positive non-circuit cycles (bringing two fixed vectors) or isolated negative circuits (bringing 1 fixed cycle).

We have also:

$$A(1, \sigma) = B(1, \sigma) / D(\sigma)$$

where

$$B(1, \sigma) = 2^{s-1} (2^{s-1} / 2^{s-1}) N(\sigma) [2^{k-1} s(1) (2^1 2^1 2^1 + 2^1 2^1 - 2^1 2^1 2^1) / 2^1 + \dots + 2^{k-i} s(i) (2^1 2^i 2^i + 2^i 2^1 - 2^1 2^i 2^1) / 2^1 + \dots + 2^{k-n} s(n) (2^1 2^n 2^n + 2^n 2^1 - 2^1 2^n 2^1) / 2^1]$$

where $s(i) 2^1 2^i 2^i / 2^1$ is just the number (2^1) of fixed points of 1 positive cycle (circuit or not) of length *i* times the number of such configurations $s(i) (2^i 2^i / 2^1)$, $s(i) 2^i 2^1 / 2^1$ is the number (1) of attractors of 1 isolated negative circuit of length *i* times the number of such configurations $s(i) (2^i 2^1 / 2^1)$ and $-s(i) 2^1 2^i 2^1 / 2^1$ is the number (2^1) of fixed points of 1 positive circuit (already counted in *B(0, σ)*) times the number of such configurations $s(i) (2^i 2^1 / 2^1)$; $2^{s-1} (2^{s-1} / 2^{s-1}) N(\sigma) 2^{k-i}$ is equal to the number of configurations of *s* – 1 positive circuits with *s(1)* of length 1, ..., *s(i) – 1* of length *i*, ..., *s(n)* of length *n*. Then we have: $A(2, \sigma) = B(2, \sigma) / D(\sigma)$, where

$$B(2, \sigma) = 2^{s-2} (2^{s-2} / 2^{s-2}) N(\sigma) [2^{k-2} s(1)^2 (2^2 2^2 2^2 - 2(2^2 2^2 2^1 - 2^1 2^2 2^1) + 2^2 2^1) / 2^2 + \dots + 2^{k-i-j} s(i) s(j) (2^2 2^{i+j} 2^{i+j} - (2^2 2^{i+j} 2^i 2^1 - 2^1 2^{i+j} 2^i 2^1) - (2^2 2^{i+j} 2^j 2^1 - 2^1 2^{i+j} 2^j 2^1) + 2^{i+j} 2^2) / 2^2 + \dots + 2^{k-2n} s(n)^2 (2^2 2^{2n} 2^{2n} - 2(2^2 2^{2n} 2^n 2^2 - 2^1 2^{2n} 2^n 2^2) + 2^{2n} 2^2) / 2^2]$$

where $s(i) s(j) (2^2 2^{i+j} 2^{i+j} - (2^2 2^{i+j} 2^i 2^2 - 2^1 2^{i+j} 2^i 2^2) - (2^2 2^{i+j} 2^j 2^2 - 2^1 2^{i+j} 2^j 2^2) + 2^{i+j} 2^2) / 2^2$ is just the number of the fixed configurations of a couple made of positive not circuit cycles or negative circuits, the (*s* – 2) remaining cycles being positive circuits, by paying attention to the fact that the fixed configurations of the couple

of a positive not circuit cycle combined with a positive circuit $(s(i)s(j)2^{2i+j}(2^i + 2^j)2^2)$ have been already counted both in $B(1, \sigma)$ and in $B(0, \sigma)$ and hence have to be taken away (by using sign $-$) from the sum $s(i)s(j)(2^{2i+j}2^{i+j} + 2^1 2^{i+j} 2^i 2^1 + 2^1 2^{i+j} 2^j 2^1 + 2^{i+j} 2^2)/2^2$.

Finally, more generally, we have $A(j, \sigma) = B(j, \sigma)/D(\sigma)$, where

$$\begin{aligned}
 B(j, \sigma) &= 2^{s-j} (2^{s-j} / 2^{s-j}) N(\sigma) \\
 &\times \left[\sum_{\zeta \in I} 2^{k-r(\zeta)} \left[2^j 2^{r(\zeta)} 2^{r(\zeta)} + \sum_{m=1}^j 2^{j-1} 2^{r(\zeta)-i(m)} 2^{r(\zeta)-i(m)} (2^{i(m)} 2^1 - 2^1 2^{i(m)} 2^1) + \dots \right. \right. \\
 &+ \sum_{\xi=(m(1), \dots, m(v)) \in \{1, \dots, n\}} v 2^{j-v} 2^{r(\zeta)-r(\xi)} 2^{r(\zeta)-r(\xi)} (2^{r(\xi)} 2^v - v(2^2 2^{r(\xi)} 2^v - 2^1 2^{r(\xi)} 2^v) \\
 &+ v(v-1)(2^2 2^{r(\xi)} 2^v - 2 \cdot 2^3 2^{r(\xi)} 2^v + 2^4 2^{r(\xi)} 2^v)/2 + \dots \\
 &\left. \left. + (-1)^v 2^v 2^{r(\xi)} 2^v + \dots + (-1)^j 2^j 2^{r(\zeta)} \right] / 2^j \right] \\
 &= 2^{s-j} (2^{s-j} / 2^{s-j}) N(\sigma) \sum_I \prod_{t=1}^j (2^{i(t)-1} - 1/2)^{u(i(t))}
 \end{aligned}$$

where $I = \{\zeta = (i(1), \dots, i(j)) \in \{1, \dots, n\}^j \mid \forall t = 1, \dots, j, \text{ the number } u(i(t)) \text{ of cycles of size } i(t) \text{ satisfies: } 0 < u(i(t)) \leq s(i(t))\}$, $j = \sum_{t=1}^j u(i(t))$, $r(\zeta) = \sum_{t=1}^j i(t)$, and $2^j 2^{r(\zeta)} 2^{r(\zeta)} / 2^j$ is just the number (2^j) of fixed points of j positive cycles (circuits or not) of lengths $i(1), \dots, i(j)$ multiplied by the number of such configurations of j positive cycles $(2^{r(\zeta)} 2^{i(1)+\dots+i(j)} / 2^j)$, $\sum_{m=1}^j 2^{j-1} 2^{r(\zeta)-i(m)} 2^{r(\zeta)-i(m)} \cdot (2^{i(m)} 2^1 - 2^1 2^{i(m)} 2^1) / 2^j$ being the number of attractors in a configuration where we have 1 negative circuit of length $i(m)$ among the $(j-1)$ other positive cycles (circuits or not) diminished by the number of the configurations having $(j-1)$ positive non-circuit cycles and $(k-j+1)$ positive circuits (already counted in $B(j-1, \sigma)$). The other terms of $B(j, \sigma)$ correspond to the number of fixed configurations of sub-graphs having $(k-j)$ positive circuits and j either positive non-circuit cycles or negative circuits, diminished by the number of already counted fixed configurations in the $B(m, \sigma)$ s, for $m < j-1$, and not yet taken away. We remark to end the calculation of $B(j, \sigma)$ that:

$$\begin{aligned}
 &2^{k-r(\zeta)} \left[\sum_{\xi=(m(1), \dots, m(v)) \in \{1, \dots, n\}} v 2^{j-v} 2^{r(\zeta)-r(\xi)} 2^{r(\zeta)-r(\xi)} (2^{r(\xi)} 2^v - v(2^2 2^{r(\xi)} 2^v - 2^1 2^{r(\xi)} 2^v) \right. \\
 &\left. + v(v-1)(2^2 2^{r(\xi)} 2^v - 2 \cdot 2^3 2^{r(\xi)} 2^v + 2^4 2^{r(\xi)} 2^v)/2 + \dots + (-1)^v 2^v 2^{r(\xi)} 2^v \right] / 2^j \\
 &= \sum_{\xi=(m(1), \dots, m(v)) \in \{1, \dots, n\}} v 2^v 2^k 2^{r(\zeta)-r(\xi)} (1/2 - 1)^v = \sum_{\xi=(m(1), \dots, m(v)) \in \{1, \dots, n\}} v 2^k 2^{r(\zeta)-r(\xi)} (-1)^v
 \end{aligned}$$

By summing the $A(j, \sigma)$ s and after the $A(\sigma) \Pi_{\sigma}$ s, $\langle f \rangle$ is clearly of the order of $n^{1/2}$:

$$\begin{aligned}
 \langle f \rangle &= \sum_{s=0}^n \sum_{k=s}^n \sum_{\sigma \in \Omega(s,k)} e^{-\sum \lambda(i)} \prod_{i=1}^n \lambda(i)^{s(i)} / s(i)! \sum_{j=0}^n A(j, \sigma) \\
 &= \sum_{s=0}^n \sum_{k=s}^n \sum_{\sigma \in \Omega(s,k)} (e^{-\sum \lambda(i)} / s!) \left(s! / \prod_{i=1}^n s(i)! \right) K_{\sigma}
 \end{aligned}$$

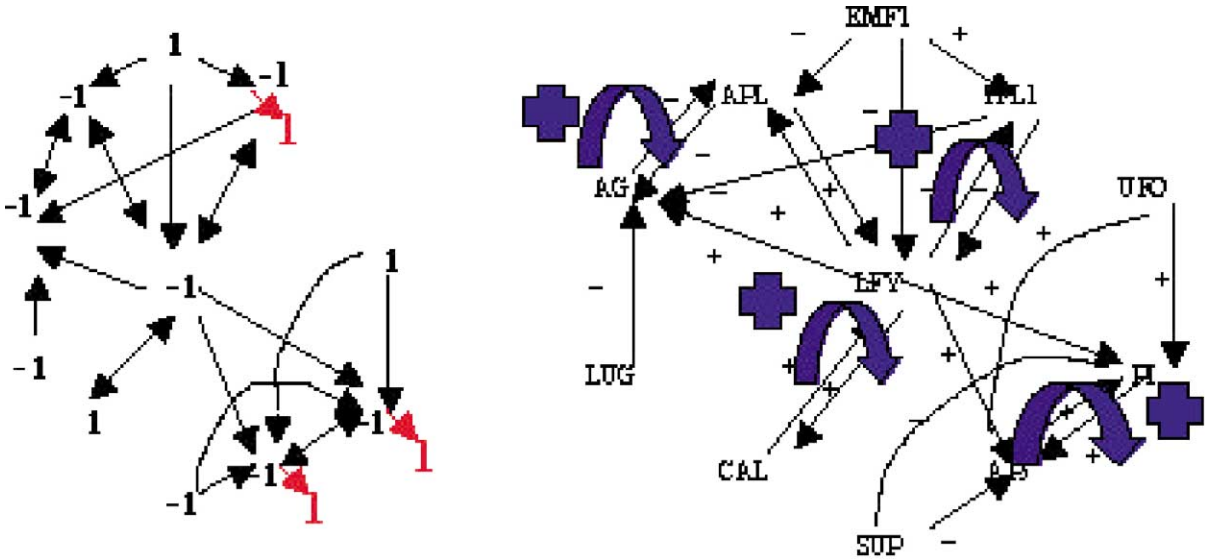


Fig. 4. Interaction graph of the flowering regulatory network of *Arabidopsis thaliana* (right) and an attractor of its Boolean dynamics (left).

where

$$K_\sigma = 2^{s-k} \prod_{i=1}^n \lambda(i)^{s(i)} (2^{i-1} - 1/2 + 1)^{s(i)} = \prod_{i=1}^n (2^i/i)^{s(i)} (1 + 1/2^i)^{s(i)} \quad \text{and}$$

$$\sum_{i=1}^n \lambda(i) = \left(\sum_{i=1}^n 2^i/i \right) / 2 = \sum_{i=1}^n \int_0^2 x^{i-1} dx / 2 = \int_0^2 \sum_{i=1}^n x^{i-1} dx / 2$$

$$= \int_0^2 (x^n - 1)/(x - 1) dx / 2 < 2^{n-1}$$

Then we have: $\langle f \rangle \sim \sum_{s=0}^n e^{-\sum \lambda(i)} (\sum_{i=1}^n \lambda(i) + \ln n/2)^s / s! = O(\sqrt{n})$. \square

Remark. Theorem 4 corresponds to the Kaufmann’s conjecture [20].

3. Examples of genetic regulation networks

3.1. The flowering regulatory network of *Arabidopsis thaliana*

If we consider the interaction graph of the flowering regulatory network of *Arabidopsis thaliana* (Fig. 4, right) [10], then we can easily define from it a Boolean dynamics, Fig. 4 (left) giving an example of attractor with final states (in bold red) different from the initial conditions. The characteristics of the associated interaction matrix \mathbf{W} are: $K(\mathbf{W}) = 22/11 = 2$, $I(\mathbf{W}) = 10/22$, $P(\mathbf{W}) = 4$, $A(\mathbf{W}) = 4$ (corresponding to the 4 differentiated tissues of the flower, i.e. sepals, petals, stamens and carpels). \mathbf{W} has two connected components and four gardens of Eden. $A(\mathbf{W})$ is well $\leq 2^4$ and in fact exactly equal to 2^2 , where 2 is the number of connected components having at least one positive loop. Then we can recall that:



Fig. 5. The Watt regulator, the prototype of negative regulatory loop in cybernetics.

- in 1948, Delbrück [21] conjectured that positive loops in the interaction graph of a regulatory network was a necessary condition for cell differentiation, i.e. for the existence of multiple attractors of the genes expression; this conjecture has been written in a good mathematical context by Thomas in 1980 [22]; we have proved above that the positive loops were related to the observation of multiple attractors, which definitively gives to the positive loops another signification than to the negative ones, more related to the stability of the system (like in the classical Watt regulator, well known in cybernetics, cf. Fig. 5);
- in 1992, Kauffman [20] conjectured that the mean number of attractors for a Boolean genetic network with n genes and with connectivity 2 was of the order of \sqrt{n} (see Theorem 4 above). This conjecture is now supported by real observations: we have about 35 000 genes in the human genome and about 200 different tissues, which can be considered as different attractors of the same dynamics. For *Arabidopsis thaliana*, there is $A(\mathbf{W}) = 4 \approx \sqrt{11}$ different tissues [10] and for the Cro operon of the phage λ , $K(\mathbf{W}) = 14/5 = 2.8$, $I(\mathbf{W}) = 9/14$, $A(\mathbf{W}) = 2 \approx \sqrt{5}$ (lytic and lysogenic attractors) [23,24], with Boolean [25] or discrete multi-level [26] models.

3.2. The gastrulation regulatory network

If we consider the regulatory network ruling the gastrulation in *Drosophila* (cf. Fig. 6 and [28]), it is easy to check that $K(\mathbf{W}) = 25/15$, $I(\mathbf{W}) = 5/15$, $P(\mathbf{W}) = 4$ and $A(\mathbf{W}) = 2$ (the corresponding cells being the ordinary ectoderm cell and the trapezoidal invagination cell called bottle cell). The regulation graph contains five connected components (among which three are singletons). In this case, the classical Kolmogorov–Rashevski–Turing models of reaction–diffusion [29–31] are well explaining the epigenetic part of the invagination at the start of the gastrulation, but only after the apparition of a new bottle cell presenting an apical constriction due to the change of intracellular balances ATP/ADP and GTP/GDP (whose ratios increase), due to the expression of the kinase (ADK and GDK) genes. This is due to a change of attractor basin by the bottle cell starting the gastrulation

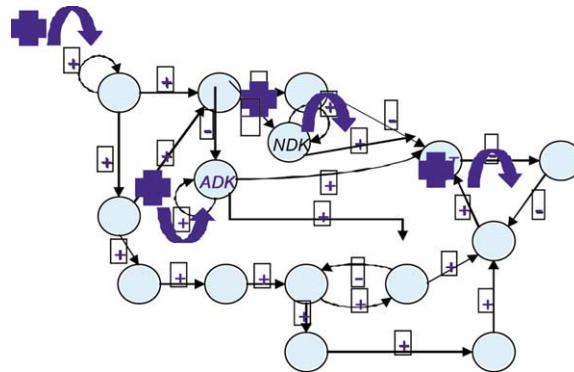


Fig. 6. The gastrulation regulatory network (after [27]) (ADK and NDK are respectively the adenylate kinase and the nucleotide diphosphate kinase).

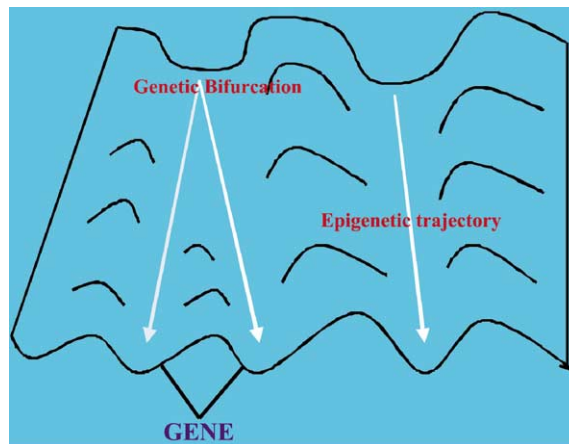


Fig. 7. The Waddington chreode or morphogenetic landscape.

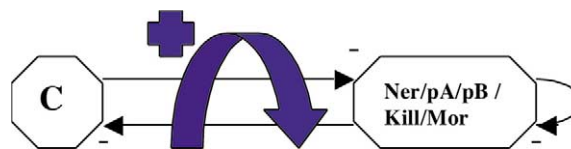


Fig. 8. The phage μ operon.

process due to the genetic regulation pathway of Fig. 6. This context describing the morphogenesis from genetic and epigenetic forces was called by Waddington a chreode or a morphogenetic landscape (cf. Fig. 7).

3.3. The phage μ lytic-lysogenic attractor [32]

If we consider the operon governing the expression of the phage μ , we obtain the graph given in Fig. 8. It is interesting to notice that $K(\mathbf{W}) = 3/2$, $I(\mathbf{W}) = 1$, $P(\mathbf{W}) = 1$, $A(\mathbf{W}) = 2$ (the two corresponding states in the host cell being the lytic and lysogenic ones, like for the Cro operon of the phage λ [24]). There is only one connected component.

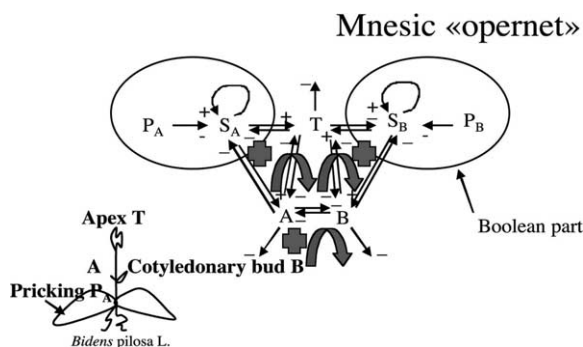


Fig. 9. The mnesic operon.

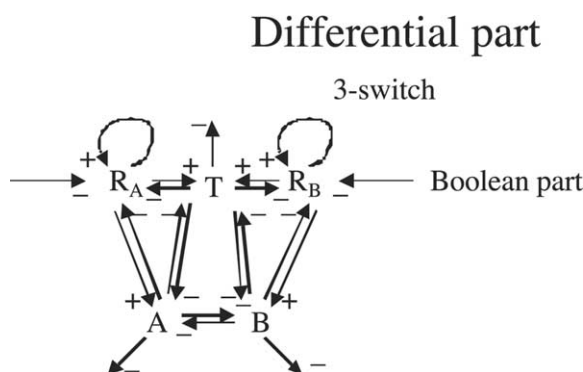


Fig. 10. Interaction graph of the epigenetic part as a continuous differential system.

3.4. The mnesic 'operon'

We will call in the following mnesic 'operon' the system obtained by merging the genetic (*operon*) and epigenetic parts (*metabolic net*) of the system ruling the cotyledonary buds growth (cf. [33,34] and Fig. 9).

The genetic part of the mnesic operon has been modelled by a Boolean system [34]:

- variable P_A (resp P_B) represents pricking treatment (or any other stress action) on the side A (resp B) of the plant and its value is 1 if treatment has been done, and 0 if not;
- variable S_A (resp S_B) represents the discrete part of the operon; we suppose that it contributes to mobilize a morphogenetic cotyledonary material R_A (resp R_B) responsible for the growth of the apex and of the cotyledonary bud A (resp B).

We will suppose in the following that the variable R_A (resp R_B) representing the concentration of R on side A (resp B) is continuous and that its velocity dR_A/dt (resp dR_B/dt) is ruled by a differential system containing a three-switch between the continuous variables T (apex growth metabolites concentration), A and B (cotyledonary buds A and B growth metabolites concentrations on respectively A and B side) (see Fig. 10).

Graph G of Fig. 9 is such that $K(\mathbf{W}) = 16/5$, $I(\mathbf{W}) = 10/16$, $P(\mathbf{W}) = 3$, $A(\mathbf{W}) = 3$; G is connected and contains two regulons.

Then we can write the differential system governing the continuous variables T , A , B , R_A and R_B as follows:

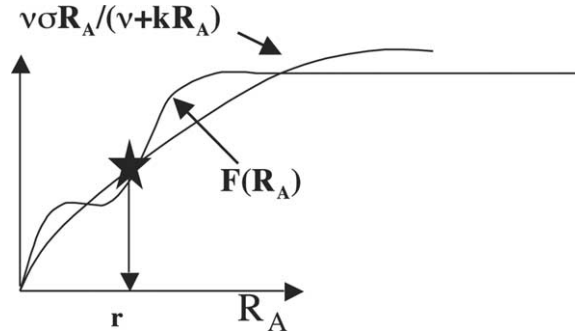


Fig. 11. Allosteric function F , which presents two successive inflection points.

$$\begin{aligned}
 dR_A/dt &= (\sigma - kT - 4kA/5 - kB/5)R_A - F(R_A)(T + 4A/5 + B/5)/(T + A + B) \\
 &\quad - wP_A - wP_B/2 \\
 dR_B/dt &= (\sigma - kT - 4kB/5 - kA/5)R_B - F(R_B)(T + A/5 + 4B/5)/(T + A + B) \\
 &\quad - wP_B - wP_A/2 \\
 dT/dt &= (F(R_A) + F(R_B))T/(T + A + B) - \nu T \\
 dA/dt &= F(R_A)4A/5(T + A + B) + F(R_B)B/5(T + A + B) - \nu A \\
 dB/dt &= F(R_B)4B/5(T + A + B) + F(R_A)A/5(T + A + B) - \nu B
 \end{aligned}$$

The two first equations correspond to the dynamics of R_A (resp R_B) whose concentration derivative at time t $dR_A(t)/dt$ (resp $dR_B(t)/dt$) results from an auto-catalytic term $\sigma R_A(t)$ (resp $\sigma R_B(t)$) diminished by the term $(-kT(t) - 4kA(t)/5 - kB(t)/5) R_A(t)$ (resp $(-kT(t) - 4kB(t)/5 - kA(t)/5) R_B(t)$) expressing the inhibition by T , A and B , plus a production of growth metabolites term denoted by $F(R_A(t)) (T(t) + 4A(t)/5 + B(t))/5 / (T(t) + A(t) + B(t))$ (resp $F(R_B(t))(T(t) + A(t)/5 + 4B(t)/5) / (T(t) + A(t) + B(t))$), by supposing that the R_A (resp R_B) consumption is competitively inhibited by the bud growth on its side A (resp B) and by the bud growth on the other side and by considering that K_{ms} and $K_{min hib}$ s are equal to 1, plus the instantaneous perturbation $wP_A(t)$ from its side A (resp B) and $wP_B(t)/2$ from the other side B (resp A), the value of $P_A(t)$ (resp $P_B(t)$) being 1 if the pricking treatment occurs on A (resp B) at time $t = t_p$, and 0 elsewhere.

The equations for the apex and cotyledonary buds growth metabolites concentrations just express that their production comes from R_A and R_B , with a competitive inhibition by the other sources of growth, plus a linear degradation term.

We suppose now for interpreting a minima the experimental results given in Tables 1 and 2 of [33] that F is an allosteric function of order 4 having two successive inflection points (involving that the protein catalysing the production of apex and buds growth metabolites has four catalytic subunits) as allowed by the Monod–Wyman–Changeux equation (see [35] and Fig. 11).

If the value of F' verifies $F(r) - rF'(0) < 0$ and $\nu < F(r)/r - F'(r) < 2(krF'(r))^{1/2} - \nu$, then the differential system above possesses at most 16 stationary states, whose 0 is a stable focus, two (respectively $(r, 0, \sigma r / (\nu + kr))$ and $(0, r, \sigma r / (\nu + kr))$) are unstable focuses surrounded by limit cycles α and β , and $C(r, r, 2\sigma r / (\nu + kr))$ is an attractor (either stable focus, or limit cycle) as shown in Fig. 12, by supposing known the dynamics of the inhibitory three-switch [36] between T , A and B .

More generally, if we have an inhibitory n -switch between the A_i 's verifying:

$$dA_i/dt = KA_i / \sum_{j=1, \dots, n} A_j - \nu A_i$$

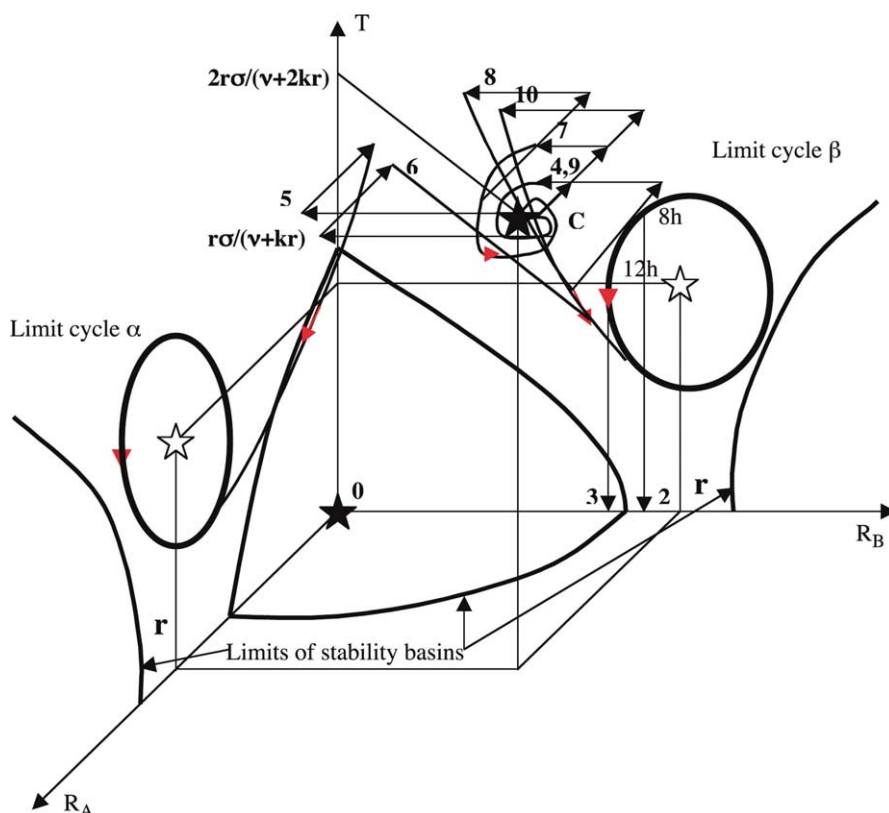


Fig. 12. Experimental perturbations in the state space (R_A, R_B, T) .

then the stationary states $A_i = k = K/v, A_j = 0$, for $j \neq i$ are stable and the stationary states having $m > 1$ metabolites equal to $k = K/mv > 0$ and the other vanishing are unstable. It is easily to check this property on the Jacobian matrix of the differential system above [37] whose unique eigenvalue $\lambda = -v < 0$ in the first case and, in the second case, whose spectrum has the general eigenvalue $\lambda = -v + v(m-1)/m - rv/m - r^2v/m - \dots - r^{m-1}v/m$, where r is one of the m th root of the unity. Then $\lambda = -(v/m)(1 + r + r^2 + \dots + r^{m-1}) = 0$, if $r = e^{i2\pi/m}$, which implies the non-stability.

The possible configurations of experimental perturbations as reported in Table 1 below and in Tables 1 and 2 of [33] give different trajectories after perturbations, as explained in the following. If n_A (resp n_B) represents the number of seedlings beginning to elongate on the side of bud A (resp B), then $g = (n_A - n_B)/n$, where $n = n_A + n_B$, is an asymmetry growth index.

We can make in Fig. 12 above the following observations.

- (1) If the initial conditions are inside the basin of stability of the attractor C, near C, then the whole trajectory without perturbation lies in this basin and tends to the attractor C (where $R_A \approx R_B$) as time t is tending to infinity. After decapitation without primitive pricking (non-pricked control), A and B buds have the same chance to begin to grow up, then to inhibit the bud growth on the opposite side, hence $g \approx 0$.
- (2–3) If we do four pricking treatments on A, then from initial conditions near C we observe a shift to the basin of the limit cycle β and following the decapitation time, we get a B domination if decapitation is done at the onset daylight (dod) and a symmetry in buds growth if it occurs at midday (dm) (see Fig. 12): it is due to the

Table 1
Experimental results about domination growth after decapitation

Pricking treatment	g	Domination
(1) Non-pricked control	0.02	$A = B$
(2) 4A with decapitation at onset daylight (dod)	0.35 ± 0.02	$A < B$
(3) 4A with decapitation at midday (dm)	0.08 ± 0.15	$A = B$
(4) 1A (dod)	0.01	$A = B$
(5) 4B (dod)	-0.35	$A > B$
(6) 1A (1 h) 4B (dod)	0.39	$A < B$
(7) 2A (dod)	0.06	$A = B$
(8) 2A (1 h) 2A/2B (dod)	0.32	$A < B$
(9) 2A (1 h) 2A/2B (3 h) 2A/2B (dod)	0.05	$A = B$
(10) 2A (1 h) 2A/2B (3 h) 2A/2B (5 h) 2A/2B (dod)	0.34	$A < B$

fact that R_A and R_B after decapitation lies in the basin of the limit cycle β where $R_B > R_A$ in the first case and are in the basin of the stable focus 0 in the second case.

- (4) If the system is starting near C and if one pricking 1A is done on cotyledon A (resp B), then we suppose that the perturbation results in a decrease of intensity w of R_A (resp R_B) and in a decrease of intensity $w/2$ of R_B (resp R_A) at time t_P of pricking; then the variables R_A and R_B remain in the basin of C.
- (5) If we are doing four pricking treatments on side B, then the point representing the system leaves the basin of C with a decrease of $4w$ in R_A and $2w$ in R_B to go to the basin of the limit cycle β .
- (6) If we do one pricking on side A and 1 h after four prickings on side B, then the system goes in the basin of the limit cycle α and for opposite reasons than for (5) above, $R_B < R_A$ and A dominates B after decapitation.
- (7) If we do two prickings on side A, then the system remains in the basin of C.
- (8) If we do two prickings on side A and 1 h after two prickings on each side, then the system goes in the basin of the limit cycle β .
- (9) If we do two prickings on side A, 1 h after two prickings on each side and 3 h after two prickings on each side, then the system returns in the basin of C due to the special shape of the trajectory (8) going to the limit cycle β passing near the frontiers of the basin of C.
- (10) If we do two prickings on side A, 1 h after two prickings on each side, 3 h after two prickings on each side and 5 h after anew two prickings on each side, the system goes in the basin of C like in (9) after 4 h, then 5 h after it goes from C to the basin of the limit cycle β .

We have then shown only by using the qualitative description of both the genetic (after pricking or any stress) and epigenetic forces exerted on the opernet that all the simulated behaviours described above qualitatively fitted the observed phenomenology (Table 1 above and Tables 1 and 2 of [33]).

4. Conclusion

An important first conclusion we have to make explicit in this paper concerns the relationship between the number F of fixed points and the number S of interaction circuits of the interaction matrix \mathbf{W} : the problem is in fact to find the best upper bound for F for a given interaction matrix \mathbf{W} . This question is related to the famous 16th Hilbert's problem, whose one of the aim is to give an efficient upper bound for the number of limit cycles of a polynomial differential system. Let us summarize the role of the architecture of positive and negative circuits of \mathbf{W} on the occurrence of multiple stationary behaviours, as obtained above: if the number of nodes and the number of arcs are the same, there is only one isolated interaction circuit ($S = 1$) in \mathbf{W} and either this circuit is negative and the lowest bound (0) for F is reached, or this circuit is positive and the upper bound (2^1) for F is reached. If the number of nodes is n and the number of arcs is $n + 1$, there is two interaction circuits ($S = 2$) with the

following structure: if both circuits are negative, $F = 0$; if there is a positive circuit and a negative circuit disjoint, $F = 0$; if there is a positive circuit intersecting a negative circuit, $F = 1$; if there is a positive circuit intersecting a positive circuit, $F = 1$; if there is two disjoint positive circuits, $F = 2^2$. If, more generally, the number S of interaction circuits of \mathbf{W} is m , then: if all circuits are negative, $F = 0$; if all circuits are positive, $2 \leq F \leq 2^m$ and if all circuits are positive and disjoint, $F = 2^m$. An interesting open problem is now to make exhaustive the determination of F and S and in particular to find the circumstances (related to the circuits structure) in which we can relate the number of intersecting and isolated circuits to F . A conjecture we could make is that $F = 2^c$, where c is the number of not-singleton connected components of the interaction graph G having at least one positive loop: it holds for the lactose, *Arabidopsis*, phage μ and gastrulation regulatory systems. The approach for solving this open problem could consist in finding coherent relationships between analogous properties discovered independently for continuous and Boolean versions of regulatory networks [4–9].

The second conclusion concerns the practical use of the presented results; a geneticist can for example exploit the minimality results in the following sense: we have shown in the paper that it would be possible to characterize the minimal interaction matrices having certain state vectors as fixed points. The determination of these matrices is not unique, but permits to focus on a certain important equivalence class to which the expected matrix has to belong. This considerably restricts the choice of the possible interaction matrices compatible with observed fixed points, when it is impossible to directly get from experiments all interaction coefficients, but when it is only possible to observe the phenomenology of fixed points or limit cycles. This corresponds in genetics to the observation of stationary expression behaviours (for example from bio-array imaging) without experimental measure of the inhibitory and activatory coefficients of repressors and promoters. The possibility to obtain (even in an equivalence class) a sketch of the interaction matrix permits to construct (by randomising in a Bayesian way the unknown coefficients of \mathbf{W}) more complicated interaction matrices, then to test if they still have the observed states as fixed points and finally keep or reject definitively the so-tested matrices and propose further experimental strategies refining the knowledge about the interaction structure of a genetic regulatory network and then answer crucial biological questions like the relationship between genetic expression and recombination [16,38–40] (the crossing-over and translocations break points seeming correlated with the ubiquity genes expression sites) or the relative parts taken by genetic and epigenetic forces in morphogenesis (embryogenesis or tumorigenesis). The last (but not the least) application of the interaction matrices introduced above is the ability to calculate the barycentre between two matrices by using classical (spectral or L_2) distances between matrices, then we could build phylogenetic trees among a set of species avoiding the complex problems coming from the non-unicity of L_1 (Hamming or Manhattan) barycentres met in the sequence based phylogenetic trees. The interaction based phylogenetic trees could reflect more the genomic function than the genomic anatomy hence could explain more deeply the evolution trends.

Acknowledgements

We have done this work thanks to the support of the National Network for Technology Research RNTS ‘Technologies for Health’ from the French Ministry of Research.

References

- [1] J. Demongeot, J.-P. Françoise, M. Richard, F. Senegas, T.P. Baum, A differential geometry approach for biomedical image processing, *C. R. Biologies* 325 (2002) 367–374.
- [2] J. Demongeot, M. Richard, New segmenting and matching algorithms as tools for modeling and comparing medical images, *Imacs 2000*, EPFL, Lausanne, 2000, CD 127.
- [3] J. Mattes, M. Richard, J. Demongeot, Tree representation for image matching and object recognition, *Lect. Notes Comput. Sci.* 1568 (1999) 298–309.
- [4] E. Plahte, T. Mestl, S.W. Omholt, Feedback loops, stability and multi-stationarity in dynamical systems, *J. Biol. Syst.* 3 (1995) 409–414.

- [5] J. Demongeot, Multi-stationarity and cell differentiation, *J. Biol. Syst.* 6 (1998) 1–2.
- [6] E.H. Snoussi, Necessary condition for multi-stationarity and stable periodicity, *J. Biol. Syst.* 6 (1998) 3–10.
- [7] J.-L. Gouzé, Positive and negative circuits in dynamical systems, *J. Biol. Syst.* 6 (1998) 11–16.
- [8] O. Cinquin, J. Demongeot, Positive and negative feedback: striking a balance between necessary antagonists, *J. Theoret. Biol.* 216 (2002) 239–246.
- [9] O. Cinquin, J. Demongeot, Positive and negative feedback: mending the ways of sloppy systems, *C. R. Biologies* 325 (2002) 1085–1095.
- [10] L. Mendoza, E.R. Alvarez-Buylla, Dynamics of the genetic regulatory network for *Arabidopsis thaliana* flower morphogenesis, *J. Theoret. Biol.* 193 (1998) 307–319.
- [11] E.H. Snoussi, R. Thomas, Logical identification of all steady states: the concept of feedback loop characteristic states, *Bull. Math. Biol.* 55 (1993) 973–991.
- [12] J. Demongeot, M. Kaufman, R. Thomas, Interaction matrices, regulation circuits and memory, *C. R. Acad. Sci. Paris, Ser. III* 323 (2000) 69–80.
- [13] J. Demongeot, J. Aracena, S. Ben Lamine, M.-A. Mermet, O. Cohen, Hot spots in chromosomal breakage: from description to etiology, in: D. Sankoff, J.H. Nadeau (Eds.), *Comparative Genomics*, Kluwer, Amsterdam, 2000, pp. 71–85.
- [14] J. Demongeot, J. Aracena, S. Ben Lamine, S. Meignen, A. Tonnelier, R. Thomas, Dynamical systems and biological regulations, in: E. Goles, S. Martinez (Eds.), *Complex Systems*, Kluwer, Amsterdam, 2000, pp. 107–151.
- [15] J. Aracena, J. Demongeot, E. Goles, Fixed points and maximal independent sets on AND-OR networks, *Discrete Appl. Math.* (in press).
- [16] J. Aracena, S. Ben Lamine, M.-A. Mermet, O. Cohen, J. Demongeot, Mathematical modelling in genetic networks: relationships between the genetic expression and both chromosomal breakage and positive circuits, in: N. Bourbakis (Ed.), *BIBE 2000, IEEE, Piscataway*, 2000, pp. 141–149.
- [17] C. Berge, *Graphes et Hypergraphes*, Dunod, Paris, 1974.
- [18] E. Goles, S. Martinez, *Neural and Automata Networks*, in: *Maths. Appl. Ser.*, Vol. 58, Kluwer, Amsterdam, 1991.
- [19] B. Bollobas, *Random Graphs*, Academic Press, London, 1985.
- [20] S. Kauffman, *The Origins of Order*, Oxford University Press, Oxford, UK, 1993.
- [21] R. Thomas, On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations, *Springer Ser. Synerget.* 9 (1980) 1–23.
- [22] M. Delbrück, Discussion, *Unités biologiques douées de continuité génétique*, *Colloques internationaux CNRS* 8 (1949) 33–35.
- [23] R. Thomas, D. Thieffry, M. Kaufman, Dynamical behavior of biological regulatory networks. I. Biological role and logical analysis of feedback loops, *Bull. Math. Biol.* 57 (1995) 328–339.
- [24] D. Thieffry, M. Colet, R. Thomas, Formalization of regulatory networks: a logical method and its automatization, *Math. Model. Sci. Comput.* 2 (1993) 144–151.
- [25] R. Thomas, R. D’Ari, *Biological Feedback*, CRC Press, Boca Raton, 1990.
- [26] F. Plouraboué, H. Atlan, G. Weisbuch, J.-P. Nadal, A network model of the coupling of ion channels with secondary messenger in cell signaling, *Network Computation in Neural Networks Systems* 3 (1992) 393–406.
- [27] J. Aracena, *Modèles mathématiques discrets associés à des systèmes biologiques. Application aux réseaux de régulation génétique*, PhD thesis, U. Chile & UJF, Santiago, Chile, & Grenoble, France, 2001.
- [28] M. Leptin, Gastrulation in *Drosophila*: the logic and the cellular mechanisms, *EMBO J.* 18 (1999) 3187–3192.
- [29] N. Rashevsky, *Mathematical Biophysics*, Cambridge United Press, London, 1948.
- [30] A. Turing, The mathematical basis of morphogenesis, *Phil. Trans. Ro. Soc. B* 237 (1952) 37–47.
- [31] A.N. Kolmogorov, I. Petrowski, N. Piscounov, Étude de l’équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique, *Mosc. Univ. Bull. Math.* 1 (1937) 1–25.
- [32] F. Thuderoz, DEA Report, UJF, Grenoble, France, 2000.
- [33] M. Thellier, L. Le Sceller, V. Norris, M.C. Verdus, C. Ripoll, Long-distance transport, storage and recall of morphogenetic information in plants. The existence of a sort of primitive plant ‘memory’, *C. R. Acad. Sci. Paris, Ser. III* 323 (2000) 81–91.
- [34] J. Demongeot, M. Thellier, R. Thomas, A mathematical model for storage and recall functions in plants, *C. R. Acad. Sci. Paris, Ser. III* 323 (2000) 93–97.
- [35] J. Demongeot, M. Laurent, Sigmoidicity in allosteric models, *Math. Biosci.* 67 (1983) 1–17.
- [36] R. Thomas, La logique des systèmes vivants, *Bull. Cl. Sci. Acad. R. Belg.* 74 (1988) 432–442.
- [37] O. Cinquin, J. Demongeot, Inhibitory n-switch dynamics and applications, *Math. Biosci.* (in preparation).
- [38] O. Cohen, M.A. Mermet, J. Demongeot, HC Forum[®]: a web site based on an international human cytogenetic data base, *Nuclear Acids Research* 29 (2001) 305–307.
- [39] O. Cohen, C. Cans, M. Cuillel, J.-L. Gilardi, H. Roth, M.A. Mermet, P. Jalbert, J. Demongeot, Cartographic study: breakpoints in 1574 families carrying human reciprocal translocations, *Hum. Genet.* 97 (1996) 659–667.
- [40] O. Cohen, C. Cans, M.-A. Mermet, J. Demongeot, P. Jalbert, Viability thresholds for partial trisomies and monosomies. A study of 1159 viable unbalanced reciprocal translocations, *Hum. Genet.* 93 (1994) 188–194.