Differential Equations and Cellular Automata Models of the Growth of Cell Cultures and Transformation Foci

Roberto Serra[*]
Marco Villani

Montecatini Environmental Research Center,
Edison Group,
v. Ciro Menotti 48,
I-48023 Marina di Ravenna (RA)

Annamaria Colacci

National Institute for Cancer Research,
Biotechnology Satellite Unit,
Viale Filopanti, 20/22,
I-40126 Bologna (BO)

Two different modeling approaches are discussed in the study of *in vitro* cell cultures which, after exposure to a carcinogen, may develop transformation foci that may be considered the *in vitro* analogue of tumors. The most important variables that are measured in these tests are the number of foci found at the end of the experiment, starting from a different number of initial cells. It is shown that an approach based upon ordinary differential equations (ODEs) may fit the data, but in a fragile way, while a cellular automata (CA) approach provides a robust agreement. However, the story told here is not that of a conflict, but rather of a cooperation between the two modeling approaches: the results of the ODE study guided our exploration of the different alternatives in CA simulations, and provided checks during model development and testing.

The CA model led us to consider the importance of the initial seeds, a point which has not been stressed in the previous literature, and to reinterpret published experimental data. It is shown that CA models, which retain cell individuality, can handle this aspect in a straightforward way, which would have been very difficult to introduce in methods based upon partial differential equations. It is also shown that quantitative modeling provides useful insights for the interpretation of experimental data as well as suggestions for further experiments.

1. Introduction

Different approaches have been used to model the dynamical behavior of biological systems, including ordinary differential equations (ODEs),

[*]Electronic mail address: rserra@cramont.it.

discrete difference equations (DDEs), partial differential equations (PDEs) and cellular automata (CA). CA are well suited to describe several interesting biological phenomena (see [1] and further references quoted therein) including immune system response [2], biodegradation of organic compounds by soil microbes [3], and tumor growth [4, 5].

ODEs and DDEs can be usefully applied whenever spatial homogeneity (on a suitable scale) can be assumed, while PDEs and CA are able to describe spatially heterogeneous systems. In dealing with reproduction phenomena, DDEs are more naturally suited whenever reproduction timing is synchronized within the population (e.g., all the newborns are born in spring) while ODEs represent a favorite choice when no such synchronization can be assumed and when the time scale of the observation is large enough to justify a limiting operation $\Delta t \to 0$.

We describe here an interesting biological system of cell cultures which are exposed to a suspect carcinogen, that appeared at first very well suited for an ODE approach. However, as we shall show, the ODE description was partly unsatisfactory, so we developed a CA model that can take into account some features that proved to be very important. The CA model has provided robust agreement with experimental data and has led to a re-interpretation of some literature data, pointing to a possible role of a variable which had so far been unnoticed, to the best of our knowledge. However, the previous work was not lost because the results of the ODE models not only led us to introduce the new model, they also provided guidance in our exploration of the "simulation space" of the CA. We will see below that using different modeling tools to study a complex system, in particular ODEs and CA, can be fruitful.

The cell cultures we studied may develop, a few weeks after exposure, so-called "transformation foci" whose number provides an indication about the carcinogenic affect of the substance under scrutiny. These systems are very useful in cancer risk assessment and are gaining widespread interest and acceptance. They are described in section 2.

The variables that are measured in typical cancer risk tests are the numbers of foci, which are found at the end of the experiment, under different conditions. "The end of the experiment" is defined as a fixed elapsed time from the beginning of the experiment, and is dictated by intrinsic features of the cells, which after some time lack their capacity to stick to the walls of the culture plate. A global variable is usually measured; that is, the average number of foci per culture plate, while no attention is paid to the location of foci in the plate itself. Therefore an approach based upon ODEs or DDEs might seem appropriate. Indeed, models of this kind have already been applied [6].

However, during this process the cell population increase takes place through the growth of clusters of cells surrounding those which had been plated on the bottom of the culture plate at the beginning of the experiment (initial seeds). During this replication process, some

cells become "transformed," and each of them is able to give rise to an observable transformation focus. The process of cell growth and transformation is local and a mean-field approach, which ignores spatial heterogeneities, might miss some important aspects.

The results of the ODE approach are described in section 3. Here we will only synthesize the reasons why it was necessary to move to a different modeling framework.

The ODE model can reach a reasonable agreement with the experiment in regards to the most important variable which can be actually measured, that is, the number of foci F_{fin} found at the end of the experiment, in correspondence of different values of the number of initially seeded cells M_0. Experimantally [6] one finds a power-law relationship:

$$F_{fin} = cM_0^q \tag{1}$$

with c independent of M_0 and $q < 1$ (typically, for the frequently used C3H10T1/2 cell lines, $q \approx 0.4$).

Upon closer examination of the model behavior, such agreement can be proved to be fragile, as it is heavily dependent upon the value of a parameter which, in order to provide the agreement, should be constrained to a narrow, rather unrealistic interval. This parameter is the exponent which relates the rate of formation of new cells $(dM/dt)_+$ to the number of cells which are already present: $(dM/dt)_+ \propto M^\nu$.

In the beginning of the test each cell is surrounded by empty space, so every existing cell can replicate and it is therefore reasonable to assume linear growth, that is, $(dM/dt)_+ \propto M$, but when clusters become large the internal cells cannot replicate due to contact inhibition and replication takes place mainly on the borders so that $(dM/dt)_+ \propto M^{1/2}$ (in the case of circular clusters). For reasonable values of the other model parameters, the agreement between the theoretical and the measured shape of $F_{fin}(M_0)$ is reached only when the exponent ν is very close to one (see section 3 for a quantitative analysis). But, as we have seen, the actual exponent is likely to be smaller than that for a large part of the cell growth process.

Moreover, it is becoming increasingly apparent that power laws like equation (1) can be obtained by a number of different models, so the match between model and experiment does not *per se* guarantee adequate modeling.

We tried to model the growth of the clusters while still adhering to an ODE scheme by varying the effective growth exponent as a function of the number of cells per cluster. But this model requires the introduction of *ad hoc* hypotheses and geometrical parameters and its dependence upon M_0 is heavily dependent upon other parameter values.

We then developed a model where the spatial constraints to cell growth are taken into account from the very beginning. A natural framework to accomplish this task is that of CA [7–10], which is sup-

posed known to the reader, because it allows the introduction of discrete units (cells) in a straightforward way, and because it requires that inter- actions be local, involving only a cell and its neighbors, as it happens in cell reproduction. For a recent review on CA see [11] and further references quoted therein.

A possible alternative would have been that of resorting to PDEs, but this would have required considering a cluster as a continuum, which in turn implies the choice of a length scale much wider than that of the single cell. Moreover, as we shall see, this would have prevented us from discovering a very interesting feature of these systems.

The CA model is described in section 4. Here we just mention that in our initial attempts we found that for reasonable parameter values our simulation yielded a function $F_{fin}(M_0)$ which was approximately flat, at least in some intervals; that is, $(dF_{fin}/dM_0) \approx 0$, in disagreement with the empirical behavior of equation (1).

The reason for the disagreement turned out to be subtle: according to our model (and to most commonly shared views of the process as in [6]) the change from a normal cell (B-type, B for "buono") to a transformed cell (T-cell) does not take place all of a sudden, but requires an intermediate step, where cells become activated (A-type). The role of the carcinogen is that of activating a certain number of cells, while the transformation A → T takes place spontaneously during the subsequent phase of cell growth without carcinogen. In the following, we refer to A-type cells simply as "A cells," to B-type cells as "B cells," and so on.

Our simulations start ($t = 0$) at the moment when the carcinogen is washed away, and we must therefore assume that in the initial popula- tion a certain fraction of the cells are type A. This is the initial condition used for ODE simulations. In the CA simulations we placed at random locations either a B or an A cell (taking into account the overall ratio A/B). B cells gave rise to clusters of B cells, A cells gave rise to clusters of A cells, some of which might possibly turn into type T. After some time, clusters started to encounter each other, and B cells started to compete with A cells on the colliding borders. For reasons described in section 3, B cells are favored in this competition with respect to A cells.

In simulations carried out with the above initial conditions, we found the already mentioned result that $(dF_{fin}/dM_0) \approx 0$ (except for rather extreme and unreasonable values of some model parameters). But a closer examination of the replication process described in section 4 showed that, whenever a B cell is turned into an A cell by the carcinogen, there is very often at least one other cell close to it, which is still of type B. This is due to the fact that activation takes place during cell reproduction and it is highly unlikely that the same process involves both daughter cells. Therefore the initial conditions had to be changed. We used as initial seeds pairs of cells composed of either one A and one B, or two B cells (but never by two A cells). The simulations with these

initial conditions provided a close match with the observed behavior described by the empirical equation (1). The behavior was robust for a wide range of parameter values and also with respect to some different model alternatives; such as, for example, endowing the cells with a limited capability to move, looking for free space to reproduce.

A very interesting observation is that some researchers had performed reseeding experiments, which are described in section 4, where the initial seeds are likely to be actually composed of separate A and B clusters. In these cases most researchers report that the number of foci seems largely insensitive to the number of initial seeds, as predicted by our CA model with initial conditions composed of all type A and all type B clusters.

While the experimental confirmation of this finding requires further tests it is already clear that the CA model provides a convenient framework to deal with such systems, and that its study allows one to reinterpret experimental data in an interesting and meaningful way allowing the design of further tests.

Moreover, the interplay between the ODE study of the system and the CA simulations has been very fruitful. The results of the former have allowed us to better understand the behavior of the simulations, guided our experimentation in the space of possible parameter values, and provided checks for the validity of the model assumptions. We will come back to the issue of the comparison and cooperation of the different modeling techniques in section 5.

2. Cell cultures and transformation foci

While many studies exist concerning *in vivo* cancer growth (see [12] for a review), the mathematical analysis of *in vitro* assays is less developed. However, in order to study the main features of tumor formation, *in vitro* tests provide very useful information and reduce the need for animal experimentation [13–15]. Moreover, the recent developments of molecular biology allow for a careful comparison, at the level of patterns of gene expression, between *in vivo* and *in vitro* systems, and it is therefore expected that the importance of these latter methods will further increase in the near future.

These tests [13] are based upon the use of well-defined cell clones, some of which are plated on a Petri dish and exposed to a chemical (e.g., a suspected carcinogen) for a short period of time. After that the chemical is washed away and the cells are cultured for a longer period. They reach confluence (i.e., they cover the bottom of the plate) in some days but the test continues for some more weeks. While the growth of the number of normal cells is inhibited, transformed cells which are not affected by "contact inhibition," undergo further growth giving rise to macroscopic structures (transformation malignant foci), each one composed of many transformed cells.

Macroscopic foci are counted at the end of the experiment. In this way it is possible to evaluate the carcinogenic effect of the chemical under study by comparison with the results of other known substances. Indeed, some of these tests; for example, those using Balb/c 3T3 clones, show good correlation with *in vivo* tests [16]. Moreover, the use of *in vitro* systems with well-defined cell clones makes the molecular characterization of the steps leading to transformation easier than their *in vivo* analogues.

Even these *in vitro* systems are complex biological systems and are subject to a high level of variability among different tests. The following variables are usually determined.

- The number of transformation foci per dish which are found at the end of the experiment F_{fin} in different experimental conditions (e.g., by using different concentrations of a suspected carcinogen).

- The transformation frequency $T_f \equiv F_{fin}/M_0$, where M_0 is the number of cells which were initially plated (more precisely, those which survive after initial plating and exposure to the carcinogen, which may well have a citotoxic effect).

However, by focusing only upon the number of foci at different carcinogen concentrations one is likely to ignore some useful information that may be provided by these methods, perhaps introducing some further tests. Mathematical modeling might therefore improve our capability to extract meaningful information from *in vitro* tests and to suggest further experiments by providing a framework to interpret the time development of cell cultures.

Existing dynamical models of the birth of transformation foci are usually of the population dynamics type [17], and they treat cell growth as spatially homogeneous [6]. This would be appropriate if the cells were free to wonder in the plate, but they are actually bound to the bottom of the dish, so they cannot move and interact freely with each other and, as observed in section 1, grow in approximately circular clusters.

A different model was proposed by Mordan in [18], which takes into account the local features related to the development of cell clusters. However, this model assumes that the number of final foci is a logistic function of the average size of the cell clusters (microcolonies) at the time of confluence, and no attempt is made to describe the dynamics of growth and transformation. Therefore it represents at most a phenomenological relationship between different variables and not a dynamical model.

As several phenomena take place during cell growth it is appropriate to try to develop a minimal model which aims at describing the most relevant aspects, leaving aside many details (until comparison with ex-

perimental data does not compel us to take into account what had been previously considered as a "detail").

The models proposed in sections 3 and 4 are based upon the well known fact that the cell transformation process involves more than one step [6, 19, 20]; a minimal model therefore requires two steps, so we will suppose that normal cells (called "B" cells in section 1) can become active (type "A") under the effect of the carcinogen. As the cell lines which are used for *in vitro* tests have already undergone some of the mutations which finally lead to the transformed state, they are not really "normal" cells like those found *in vivo*, but are normal from the viewpoint of the test. When no activation event occurs B cells grow from their initial number to a full monolayer which covers the bottom of the culture plate.

After the carcinogen has been removed, the A cells can spontaneously undergo a further change, leading them to transformed (i.e., "T") cells. Type A cells can possibly be detected by the cell defense system and induced to death, so their probability of dying may exceed that of the type B cells. Activation represents the intermediate step between the original cell and the transformed one: we will suppose that it is a property that is inherited by daughter cells; like, for example, a mutation. A and B cells, which are supposed to be indistinguishable under the microscope, feel contact inhibition, but T cells do not and continue to grow. A single T cell can give rise to a full macroscopic transformation focus.

The growth of transformation foci could be dealt with by a further model, which can be built on top of the previous one, which should describe how a single transformed cell may give rise to a focus. This process is not modeled in the present work wherein we assume that each newborn T cell gives rise to a full focus (unless it is too close to another T seed; in this case, as coalescence between nearby foci may occur, we count the two as a single focus when comparing the model with experimental data). The choice of separating the description of foci growth from that of the monolayer of type B and A cells is based upon the assumption that the perturbations due to infiltration of transformed cells among the others do not play a significant role in the formation of new foci.

Following [6] we will assume that during the initial exposure to the carcinogen some B cells may become activated. The models presented here do not describe the events which take place during this initial phase (which lasts typically 1 or 2 days), but do focus upon the phenomena which take place after the carcinogen has been washed away. During this 3 to 5 week period the culture medium is periodically changed so that cells are provided with fresh nutrients and metabolic wastes are removed.

Let us also remark that, in negative controls without a carcinogen, transformation foci may occasionally be observed. However, as the

number of foci in negative controls is typically much lower than in the case of exposure to a carcinogen, this phenomenon is neglected in the following models.

3. The ordinary differential equation model

ODEs represent a "natural language" to describe dynamical systems which are homogeneous in space, or whenever we do not care about their spatial heterogeneity. A possible alternative could be that of the DDE, which has already been used in the literature [6]. However, the lack of a self-maintaining synchronization among the cells makes the former choice a more natural one in this case [17]. Moreover, in the interesting regions of parameter space in the models considered, the two classes of equations yield similar results (although it is well known that in general DDEs can lead to instabilities even when their ODE counterparts do not).

Let us first consider the growth of normal cells until confluence, when they completely cover the bottom of the Petri dish. We can describe the time evolution of a population of cells that feel contact inhibition by the following equation

$$\frac{dM}{dt} = [Gs(M) - \omega]M \tag{2}$$

where $M(t)$ is the density of cells at time t (as the culture plates are of a fixed size, density is proportional to the total number of cells), ω is the spontaneous cell death rate, and $s(M) \in [0, 1]$ is a growth limiting function. Two very well known empirical models of population growth that have also been applied to tumor growth, are those of Verhulst and Gomperz in [12]. They are special cases of equation (2), the former corresponding to the choice $s(M) = 1 - bM$, the latter to $s(M) = (\omega/G) - \ln(M(t)/M_\infty)$. Here M_∞ is a constant equal to the asymptotic limit $\lim_{t\to\infty} M(t)$.

Comparison with experimental data on the growth of cells in the culture plate without carcinogen indicate comparable and not fully satisfactory agreement for the models in [12] (data not reproduced here), but a closer match is obtained by using a modified version of the Verhulst equation, which takes into account the facts that (*i*) contact inhibition starts to be felt only after a certain population density has been reached and (*ii*) at very low levels of cell density the growth rate may be an increasing function of cell density. In this case $s(M)$ can be given by the following piecewise linear expression (Figure 1):

$$
\begin{aligned}
s(M) &= \alpha_0 + c_0 M = 1 - c_0(s_1 - M) & 0 < M < s_1 \\
s(M) &= 1 & s_1 \leq M \leq s_2 \\
s(M) &= 1 - b(M - s_2) = b(s_3 - M) & s_2 < M \leq s_3 \\
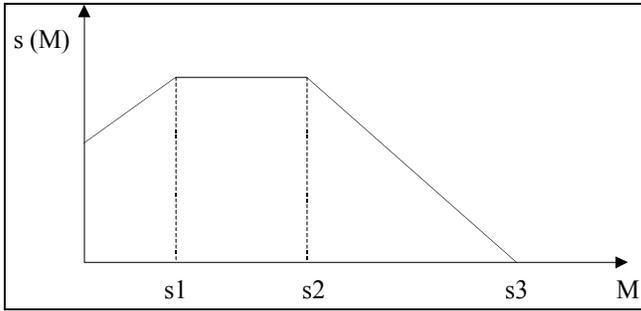s(M) &= 0 & s_3 < M
\end{aligned}
\tag{3}
$$

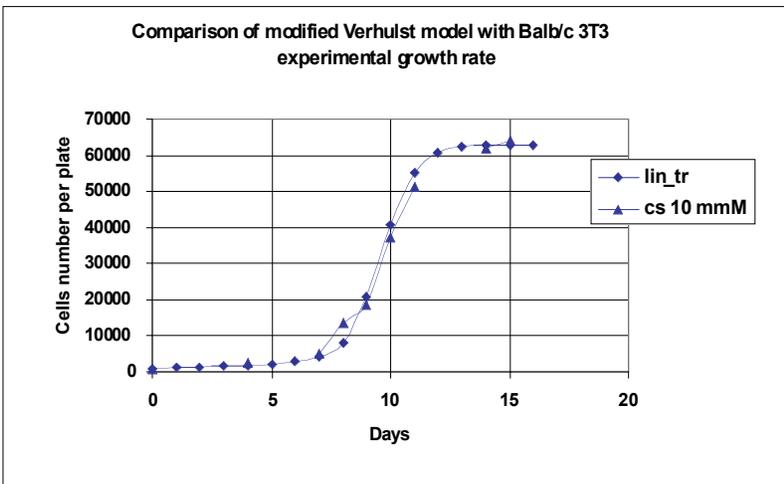**Figure 1**. Shape of the function $s(M)$ from equation (3).

Figure 2. Comparison of modified Verhulst growth rate (indicated as **lin_tr**) with experimental data on Balb/c 3T3 data (indicated as **cs 10 mmM**). Parameter values are $s_1 = 9500$, $s_2 = 15000$, $s_3 = 63000$, and $\alpha_0 = 0.01$.

where α_0 is the value of the growth limiting function when $M = 0$, s_1 and s_2 are thresholds which separate the different regimes, s_3 is the maximum cell density, c_0 is the slope of the growing region, and b that of the decreasing region. Note that the parameters are not independent since $\alpha_0 + c_0 s_1 = 1$ and $1 - b(s_3 - s_2) = 0$. The original Verhulst model corresponds to the decreasing region only, that is, $s_1 = s_2 = 0$.

We refer to the growth equation described by equations (2) and (3) as the "modified Verhulst equation" in the following. It is worth observing that it describes experimental data concerning cell growth well (an example is shown in Figure 2) while both the Gomperz and the original Verhulst models typically overestimate the initial growth rates.

In the following, for the sake of simplicity we will often neglect the effects of the initial growing portion of the curve, and will therefore limit ourselves to the case where $s_1 = 0$.

We now come to transformation tests. Let $B(t)$, $A(t)$, $T(t)$ be the numbers at time t of cells of type B, A, and T respectively. Let $t = 0$ be the moment when the cells are fed a new culture medium after removal of the carcinogen. Following [6] we assume that all the cells are born B, that some become A during exposure to the carcinogen, and that some A can spontaneously turn into T during the subsequent culture. Recall also that A and B cells, which are indistinguishable by visual inspection, feel contact inhibition: each A cell feels the effect of neighboring A and B cells, each B cell feels the effect of neighboring A and B cells, both A and B cells do not feel the T cells. This picture is supported by experimental data concerning cell to cell communications, which indicate the existence of two subpopulations that communicate within themselves [21]. We will not follow the time evolution of the population of T cells, which are mainly composed of growing foci.

The equation for the B cells is given by a modified Verhulst-type equation, where the limitation to growth comes from the sum of the A and B cells. The equation for the A cells contains similar terms, and in addition two loss terms describing the transformation from A to T and the increased probability that an A cell dies, with respect to a B cell.

As an alternative, one may think that the repair mechanisms are able to transform an A into a B cell, as was done in [6]: however, as it happens that $A(t) \ll B(t)$ in almost every case, the addition of some more B cells does not lead to any appreciable effects, so the two hypotheses (increased A cell death rate or back-transformation of some A to B) lead to similar results.

It is appropriate to discuss further the hypothesis that A cells are more likely to die out than B. While this appears highly likely, due to the cell repair mechanisms, it may also seem odd. In the B $\rightarrow$ A $\rightarrow$ T sequence the net growth rate increases from B to T, while we suppose that it decreases from B to A (due to the increased death rate). Indeed, let us suppose that it were not so; that is, that the growth rate of A is greater than that of B, and let us compare two different experiments, one with an initial seed of, say, 90 B and 10 A cells, and another with 900 B and 100 A cells. In experiment 1 cells grow and when they number 1000, the relative proportion of A cells would be higher than the initial 10%. From that time on, the number of A cells would constantly exceed that of experiment 2, and so also the number of T cells and therefore of transformation foci. We would then expect that the number of final foci decreases as M_0 increases, while the opposite has been observed. So we rule out the possibility that the net growth rate of A cells exceeds that of B cells. Indeed, while it is likely that a single surviving A cell has a faster reproductive cycle than a B cell, the

net growth rate of the population of the A cells is smaller than that of the B cells.

The model equations according to the above hypotheses are therefore

$$\frac{dA}{dt} = [Gs(A + B) - \omega - p]A$$

$$\frac{dB}{dt} = [Gs(A + B) - \omega]B \tag{4}$$

where it has been assumed that the growth and death rates (G and ω) are the same for both cell types and p is the sum of the extra death term for activated cells and of the loss term due to transformation from type A to T.

If we let $M = A + B$ be the total number of nontransformed cells, it is straightforward to see that

$$\frac{dM}{dt} = [Gs(M) - \omega]M - pA. \tag{5}$$

Let us recall that in our hypotheses the number of foci is equal to the number of A cells which become T cells. If we assume that transformation from A to T takes place during cell replication, then the number of A cells which become T cells in a given time interval is proportional to the number of A cells which undergo reproduction in that interval. Let us denote the rate of generating new A cells by $(dA/dt)_+$, and the number of foci at time t by $F(t)$, so $dF/dt \propto (dA/dt)_+$. The rate of generation of A cells is the nonnegative term in equation (4), so, if we let p' be the probability that a reproducing A gives birth to a T cell, then

$$\frac{dF(t)}{dt} = p'\left(\frac{dA}{dt}\right)_+ = p'Gs(M(t))A(t). \tag{6}$$

Therefore, $(dA/dt)_+ \propto A$, but the proportionality coefficient changes in time. In the following, we will suppose that $s_1 = 0$ and therefore that the growing portion of the curve of Figure 1 plays no role. In this case, in the beginning of the test $(dA/dt)_+ = GA$; afterwards, when inhibition starts to be felt, $s(M)$ decreases. Summarizing:

$$M \le s_2 : \qquad \frac{dF(t)}{dt} = p'GA(t) \tag{7}$$

$$s_2 < M \le s_3 : \qquad \frac{dF(t)}{dt} = p'G[1 - b(M(t) - s_2)]A(t)$$

$$= p'Gb(s_3 - M(t))A(t). \tag{8}$$

We will now summarize the main features of the system described by equations (4) through (8). The details of the calculations are given in appendix A.

As far as the asymptotic behavior is concerned, in the interesting case where $A(0)$ and $B(0)$ are both nonvanishing, the final state is one where

all the A cells become extinct. It can be proven that the ratio y between A and B cells

$$y(t) = \frac{A(t)}{B(t)} \tag{9}$$

decreases exponentially in time

$$y(t) = y_0 e^{-pt}. \tag{10}$$

From equation (5) one finds that, if pA is much smaller than $[Gs(M) - \omega]M$, then an approximate equation for the total number of cells can be obtained:

$$\frac{dM}{dt} \simeq [Gs(M) - \omega]M. \tag{11}$$

Its solution is given by

$$t \leq t^* : M(t) = M_0 e^{(G-\omega)t}$$
$$t > t^* : M(t) = s_2 \frac{bs_3 - \frac{\omega}{G}}{bs_2 + \left(1 - \frac{\omega}{G}\right)e^{-(Gbs_3 - \omega)(t-t^*)}} \tag{12}$$

where

$$t^* = \frac{\ln \frac{s_2}{M_0}}{G - \omega}. \tag{13}$$

Equations (9), (10), and (11) allow an approximate analytical evaluation for the growth of the population of A cells. By neglecting the number of initial foci (a reasonable assumption), the number of foci F_{fin} is the sum of two terms: $\Delta F(t^*, 0)$, those which are formed before contact inhibition starts to be felt (i.e., before t^*); and $\Delta F(t_{\text{fin}}, t^*)$, those which are formed after t^* and before the end of the experiment, t_{fin}. The dependence of $\Delta F(t^*, 0)$ and $\Delta F(t_{\text{fin}}, t^*)$ upon the number of initial seeds M_0 is studied in appendix A. Summarizing, both terms approximately follow a power law dependence upon M_0 with the same exponent, therefore

$$F_{\text{fin}} = \text{const} \cdot M_0^{p/(G-\omega)}. \tag{14}$$

Note that this law compares well with the empirical relationship (equation (1)) and that the exponent of the power law has a clear physical meaning, as it is the ratio between the increased death rate of the A cells and the net cell growth rate far from saturation. As typically $G \gg \omega$, from the slope experimentally observed one is tempted to infer that the cell repair mechanisms are able to "kill" 30 to 40% of activated cells at each reproduction.

These results generalize those of [6], which were obtained using a crude approximation where an exponential growth is supposed to take place until saturation, when any further growth would stop.

The agreement with experimental data might appear satisfactory, but a closer examination reveals that cell growth takes place through the formation of clusters surrounding each initial seed cell. The effective cell density which is "felt" by cells in a cluster is higher than the average cell density in the plate, and contact inhibition is therefore felt in the interior of the cluster at a density value which would give no effect if the cells were scattered on the whole culture plate. Growth of the cell population then takes place on the borders of the clusters and the growth rate term can no longer be assumed proportional to the total number of cells, but only to a fraction. When clusters become large, the number of cells involved in replication scales roughly with the square root of the total number of cells.

In order to deal with the aspects related to cluster growth it is possible to resort to a different model wherein the growth term scales with some power v of the number of cells, like

$$\frac{dA}{dt} = [Gs(A + B) - \omega - p]A^v$$
$$\frac{dB}{dt} = [Gs(A + B) - \omega]B^v. \tag{15}$$

Numerical simulations show that for a wide range of parameter values the number of foci turns out to be a growing function of the number of initial cells if the growth exponent v is very close to 1, but it becomes a decreasing function of M_0 if the exponent becomes slightly smaller.

This model behavior can be understood by considering the time evolution of $y = A/B$. Qualitatively, let us recall that in order to obtain more foci by seeding more cells it is necessary (although not sufficient) that the ratio A/B decreases in time. Otherwise, when comparing two experiments with different M_0, the one starting with the smaller number would show at the time when the number of its cells equals the initial number of the other experiment a higher fraction of A cells, and therefore a higher number of final foci, so that one would observe $dF_{fin}/dM_0 < 0$. Therefore, in order for the model to provide a value of F_{fin} growing with M_0 it is necessary that, at least for small t, $dy/dt < 0$.

In order to understand the behavior of our equations, let us then consider the initial value of dy/dt. In the case of equation (4), y was indeed a growing function of time, but this is no longer guaranteed in the sublinear growth of equation (15). Let us first consider the initial period of growth, when $M < s_2$ and therefore $s = $ constant $= 1$:

$$\frac{dy}{dt} = \frac{A^v B(G - \omega - p) - AB^v(G - \omega)}{B^2}$$
$$= \frac{G - \omega}{B^2}[A^v B - AB^v] - p\frac{A^v}{B}$$

recalling that $y = A/B$, this can be rewritten as

$$\frac{dy}{dt} = \frac{G - \omega}{B^2} A^\nu B[1 - y^{1-\nu}] - p\frac{A^\nu}{B}$$
$$= \frac{A^\nu}{B}[(G - \omega)(1 - y^{1-\nu}) - p]. \tag{16}$$

Since $G \gg \omega$, $G > p$, dy/dt may well be positive if $y < 1$ and $\nu < 1$.

Even with a very high (perhaps unrealistic) number of initial seeds $M_0 > s_2$ the initial value of dy/dt may be positive; in this case, $s(M) = bs_3 - b(A + B)$ so

$$\frac{dA}{dt} = (Gbs_3 - Gb(A + B) - \omega - p)A^\nu$$
$$\frac{dB}{dt} = (Gbs_3 - Gb(A + B) - \omega)B^\nu.$$

It is straightforward to obtain

$$\frac{dy}{dt} = \frac{A^\nu[1 - y^{1-\nu}]}{B}(Gbs_3 - \omega - GbM) - p\frac{A^\nu}{B} \tag{17}$$

since $s_3 > M$ and $G > p$, dy/dt may have also in this case the same sign as $1 - y^{1-\nu}$, that is, positive.

Note also that equations (16) and (17) allow us to roughly estimate the threshold value for the exponent ν above which one may expect that $dy/dt < 0$, and therefore that $dF_{\text{fin}}/dM_0 > 0$. As shown in appendix B, if $p/G = 0.3$ and $y_0 = 0.1$, the threshold is 0.87 for the case of equation (16) (few initial seeds) and 0.90 for the case of equation (17) (many initial seeds). Numerical simulations support the above conclusions. Moreover, similar behaviors are also found by using slightly different versions of the model which include the following features.

- The use of a linear death term (cells may die even in the interior of a growing cluster).

- A power law dependence upon M of the term responsible for slowing down cell growth; like, for example, $dA/dt = (G - \omega - p - \beta M^\mu)A^\nu$ and a similar equation for dB/dt.

The observation that $dF_{\text{fin}}/dM_0 < 0$ with sublinear growth exponents demonstrates that the agreement between the ODE model and experiments is fragile and does not survive a parameter change which is biologically well founded.

In order to analyze the problem we also considered models wherein the effective growth exponent gradually decreases as the average cluster size increases. These models aim at describing the growth of clusters of cells by supposing that the total growth rate of the cells of a given kind (say dB/dt) is proportional to the number of initial clusters multiplied

by a function which describes the growth of a single cluster. The effective exponent of cluster growth varies as the dimension of the cluster changes, ranging between 1 (single seed) and 1/2 (large cluster). The death terms are linear, since cells may die in the interior of a cluster as well as on its boundary.

We call these models variable effective exponent (VEE). They have the form

$$\frac{dA}{dt} = [G - \beta M]A_0 f\left(\frac{A}{A_0}\right) - (\omega + p)A$$

$$\frac{dB}{dt} = [G - \beta M]B_0 f\left(\frac{B}{B_0}\right) - \omega B \qquad (18)$$

where $f(x)$ accounts for the growth of a single cluster. A specific form which was tested is

$$f(x) = Q(x)x^{\nu(x)}$$

$$\nu(x) = \frac{1}{2} + \frac{e^{-kx}}{1 + e^{-kx}}$$

$$Q(x) = 4 - 6\frac{e^{-Kx}}{1 + e^{-Kx}}.$$

The shape of $\nu(x)$ is similar to that of the well known Fermi–Dirac distribution in quantum statistical mechanics, but here it is used in a phenomenological way to assure a smooth transition of the exponent from $\nu = 1$ when isolated seeds are present, to $\nu = 1/2$ when big clusters have developed.[1] $Q(x)$ is a form factor which ranges from 1 to 4 in the case of square clusters, the precise shape of the cluster does not modify the form factor too much (e.g., for a circular cluster it would range from 1 to π).

We assume as usual that the number of foci is proportional to the number of newly created A cells, that is,

$$\frac{dF(t)}{dt} = p'\left(\frac{dA}{dt}\right)_+ = p'[G - \beta M]A_0 f\left(\frac{A}{A_0}\right). \qquad (19)$$

In these cases one finds that, for a reasonable set of parameter values,[2] $dy/dt < 0$; however, the total number of new A cells (and therefore of the final foci F_{fin}) upon M_0 (y_0 fixed) depends upon several factors.

A frequently found behavior of the number of foci is shown in Figure 3. If M_0 is small enough, F_{fin} grows with M_0 but, after a certain

---

[1]Note that x ranges from 1 upwards; however, if k and K are much smaller than 1, the smallest value of the expression is very close to that which can be found by putting $x = 0$.

[2]The parameters used in most simulations are the following: kinetic coefficients $G = 1$, $b = 6.10^{-6}$, $p = 0.3$, $w = 10^{-3}$; Fermi function $k = 0.1$, $K = 0.02$; initial conditions: $A_0/B_0 = 0.1$, B_0 ranging from 500 to 100,000; duration of the simulation is 29 days.
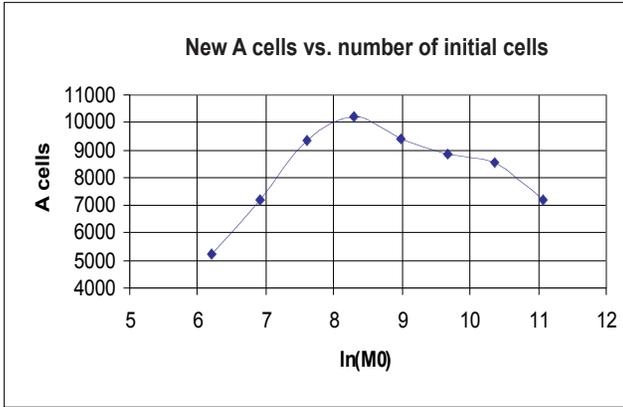
New A cells vs. number of initial cells

Figure 3. Dependence of the number of new A cells (proportional to the number of foci) as a function of M_0 for the VEE model of equation (18).

value, it decreases with M_0. By changing the parameter values even more complicated dependencies can be observed.

However, a robust feature observed is that the number of foci shows a very weak dependence upon the number of initial cells. While varying M_0 between 500 and 64,000, as shown in Figure 3, we find that the number of new A cells is always in the interval [5000, 10,000]. Simulations performed with different sets of parameters support this conclusion.

So, the observed "fragility" of the agreement between observed data and the ODE model has not been robustly eliminated by the choice of more elaborate models of the same kind. Therefore we turned to a different model that would deal with locality in a straightforward way. As we shall see, the models developed in this section are however useful, as they allow us to better understand the behavior of the CA model.

4. The cellular automata model

We have developed a model using the framework of CA which is well suited for this task [7–10]. The available space is divided into pieces of equal size, which are called *cells* in CA jargon. Each of these cells may be either occupied by a biological cell or not. In order to avoid confusion, we will limit the use of the term "cell" to biological cells, and will explicitly refer to the "CA cell," "lattice site," or simply "site" in the other case.

The model is described in detail in [22] and is briefly summarized here. A formal synthetic description is given in appendix C.

We divide the two-dimensional space (which physically corresponds to the bottom of the culture plate) into a fixed number N of CA cells,

whose size is chosen to match the average size of a biological cell. A cells are supposed to be macroscopically indistinguishable from B cells, so their average size is the same. T cells are smaller than B cells but we neglect this aspect here. Therefore, in any lattice site there may be either a B cell, an A cell, a T cell, or no cell at all. The state of site i will be called X_i; $X_i \in \{B, A, T, E\}$, where the first three symbols refer to the kind of (biological) cell which may occupy the site, and the last refers to the case where the site is empty.

For simplicity a square topology is used with the nine-membered Moore neighborhood [8]. Updating is synchronous, as is typical in CA.

At each time step, a cell may either do nothing, duplicate itself, give birth to a different cell, or die. Note that it is supposed that the major events in cell life, like the transition from type A to T, take place only when the cell enters its reproductive cycle and its DNA double strand is open. Therefore, at each time step, a biological cell located at site i may try to reproduce according to a stochastic rule if there is at least one empty neighbor of lattice point i.

The probabilities associated to the different "reaction channels" are $p_{B \to A}$ (probability that a B cell gives birth to an A cell), $p_{A \to T}$ (probability that an A cell gives birth to a T cell), and p_{RB}, p_{RA} (probability that a B or an A cell duplicates itself) respectively.

A reproducing cell will try to place its daughter cell in one of the empty neighbors at random. Actually, the algorithm for updating the state of the lattice sites proceeds in several steps. First, the CA sites occupied by type B or A cells are considered and each reproducing cell identifies the empty site which is to be occupied by its daughter cell. In a second step, all the empty CA cells are considered, and those which have been selected by a neighbor for reproduction become occupied. If there are conflicts (i.e., two or more neighbors trying to occupy the same empty space) then a stochastic choice is performed.

The search for empty neighbors is iterated. Indeed, without iterations an unrealistically high slow down of growth would be observed due to the fact that some empty space would be left unoccupied even if it were available. It has been verified that three iterations suffice to make the artificial slow down negligible.

A cell may also die. Let p_{DA} and p_{DB} denote the probability that at each time step an A or B cell dies off (note that cells may die at every time step, not only when trying to reproduce). For reasons discussed in section 3, it will be assumed that $p_{DA} > p_{DB}$. The possibility that a T cell dies off will be neglected here as the dynamics of T cells are not described in detail. In the simulations we also assume that $p_{B \to A}$; that is, the probability of "spontaneous activation," is negligible during cell growth. Cells only become activated in the initial phase of exposure to the carcinogen.

The time origin is set at the end of this exposure period. We also assume that there is no nutrient limitation during the culture period,

only the crowding of other cells bounds the growth of the cell population and no lack of essential nutrients is experienced.

In order to estimate the number of foci for comparison with experimental data it should be noted that if two activated cells which are close to each other become transformed, the two nearby foci may coalesce so that a single focus will be observed by the experimenter. Therefore, in foci counting we consider that two transformed cells which are very close to each other actually give rise to a single focus.

The model as described introduces, besides others, two simplifications that might be relaxed, giving rise to two major variants.

1. It has been assumed in the original model that an A cell can become fully transformed in a single generation. Looking at the phenomenon at a more microscopic level, it may be supposed that transformation from A to T corresponds to a genetic change, which is likely to initially take place only on one of the two DNA strands of one of the two daughter cells of the original A parent. If transformation were dominant, it would show up immediately, otherwise it would need another generation to appear at a phenotypic level. In this latter case competition for available space from other cells and contact inhibition might actually change the transformation frequency and it is not obvious *a priori* that the model features remain unaltered. Therefore, we have also tested a modified version of the model wherein the final transformation from A to T takes place in two generations: $A \rightarrow A' \rightarrow T$.

2. Some form of taxis (cell movement) in the dish cannot be *a priori* excluded. In this case newborn cells could migrate far from their parents so that the effects of contact inhibition would be initially mitigated. We have therefore also tested a model wherein newborn biological cells are allowed to move one CA cell away from their parents, therefore limiting crowding effects in the initial growth phases.

However, the main features and results of the model described below are not deeply affected by these modifications.

A series of experiments has been performed to study the dynamical properties of the CA model on a grid of $400 \times 400 = 160{,}000$ CA cells. The graphs and figures in this section were obtained with the set of parameters shown in Table 1 (where N_p is the number of simulations run for each set of parameter values and initial conditions). In the

Parameter	Value	Parameter	Value	Parameter	Value
N_p	10	M_0	160-32000	$p_{B \rightarrow A}$	0
p_{RB}	1	y_0	1/9	p_{DB}	0
p_{RA}	0.1-0.7	$p_{A \rightarrow T}$	0.001	p_{DA}	0

Table 1. List of parameters of the CA model. The meaning of symbols is given in the main text.

following $A(t)$, $B(t)$, and $T(t)$ will denote the number of cells of type A, B, and T at time step t; $M(t) = A(t) + B(t)$ is the total number of nontransformed cells; $y(t) = A(t)/B(t)$; subscript "0" denotes initial values, for example, $M_0 = M(0)$. As already observed, the origin of time is set at the end of exposure to the carcinogen (which lasts 1 to 2 days).

Note that in order to use this model to simulate actual *in vitro* carcinogenesis tests one must stop the replication after a certain number of generations (t_{fin}) instead of studying the limit $t \rightarrow \infty$.

The cells which have survived seeding and the initial treatment may then have already undergone some reproduction during the exposure period, so we investigate two classes of cases, differing for the configuration of initial cells ("seeds").

1. Sparse seeds, where at $t = 0$ there are M_0 cells, either of type A or B, placed at random in the spatial grid.

2. Grouped seeds, where the initial clusters are composed of two cells. In this case, whenever there is an A cell in the initial seed, there is also a B cell in the same seed (due to the fact that A cells originate from B cells and to the hypothesis that activation initially affects only one strand of DNA). In some cases initial clusters formed by four cells (with at most one A cell) have been tested.

As previously mentioned, the major results are similar for the different variants that have been tested. Figure 4 refers to the variant with moving cells.

The expected growth of the clusters of cells is observed, until confluence. The growth of the number of cells in time follows a familiar S-shaped curve. In a log-log plot of M *versus* time one easily verifies that in the initial part of the curve M grows approximately with t^2, as is to be expected if growth takes place on the borders of a two-dimensional cluster ($dM/dt \propto M^{1/2}$). Later growth is slowed by collision between clusters.

Let us consider the dependence of the final number of transformation foci upon the number of initial seed cells. As coalescence is a rare event, we can actually consider T_{fin}, the number of new cells of type T which are found, instead of F_{fin}. Transformation frequency is low so typical diagrams which show how the number of T cells scales with M_0 are rather noisy [23]. In this model transformation occurs with a fixed probability every time a new A cell is generated and we can study directly the number of newly generated A cells during the whole experiment (let it be N_A). This is much larger and less affected by noise and should be proportional, in the limit of a very large number of simulations, to the number of transformations from A to T cells. A diagram of this variable is shown in Figure 5.

With the exception of very small M_0 values the dependence of N_A (and therefore of F_{fin}) upon M_0 is rather flat. Note that this feature is
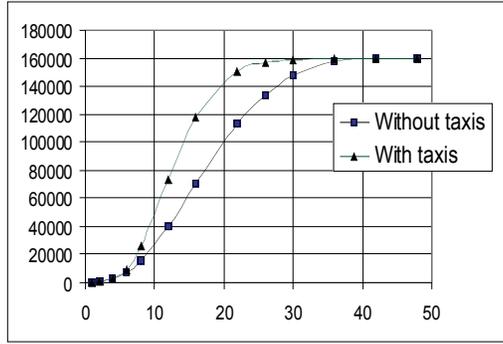
(a)
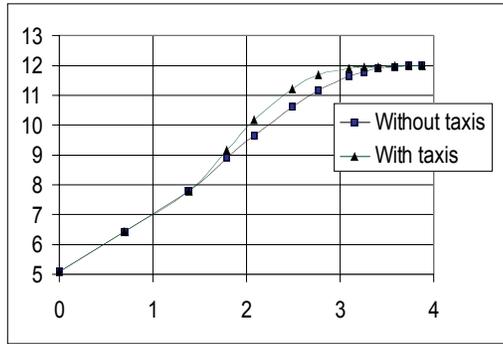
(b)

Figure 4. The observed growth of the cells until confluence (a) normal scale and (b) log-log scale. $M_0 = 160$ cells.
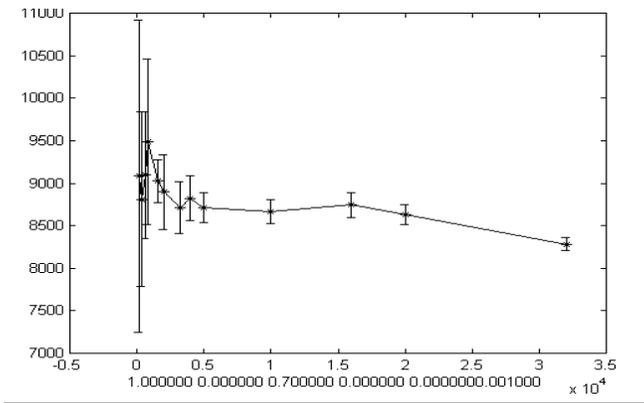

Figure 5. Number of A cells which have been generated during the experiment (N_A) versus the number of initially seeded cells $M_0 \times 10^{-4}$ ($p_{RA} = 0.7, p_{DA} = 0$).
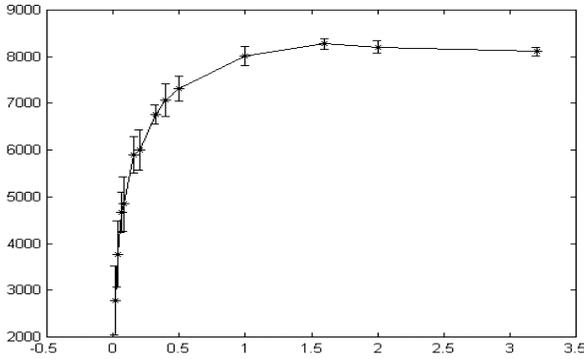
Figure 6. Number of A cells which have been generated during the experiment N_A *versus* the number of initial seed cells $M_0 \times 10^{-4}$ ($p_{RA} = 0.7$, $p_{DA} = 0$).
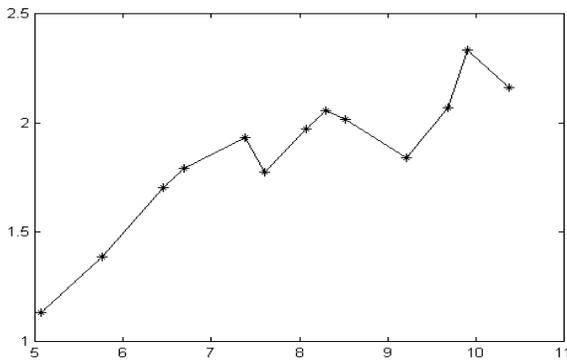

Figure 7. Number of transformed cells found at the end of the experiment T_{fin} *versus* $M_0 \times 10^{-4}$ (data on a log-log scale, $p_{RA} = 0.7$, $p_{DA} = 0$).

similar to that of the previously described EEV models. So the mean-field model described in section 3 allows us to interpret the results of the CA model. However, experimental data mentioned in section 1 do not appear to behave according to the model.

So let us now consider the case where the seeds are formed by pairs of cells, recalling that the initial clusters are composed of either two B cells or one A and one B cell. In this case each initial A cell immediately faces competition with the faster growing B cells, and the overall behavior is such that the number of A cell reproductions, and therefore the number of foci, is a growing function of M_0 (Figures 6 and 7).

We see that the CA model can lead to two different scaling behaviors. In the case where the clusters develop from isolated cells the dependence of the number of final foci upon the initial number of cells is rather flat for a wide range of parameters and for the different model versions

which we have tested (including moving and nonmoving cells, as well as one-generation and two-generation models). On the other hand, if the initial seeds are formed by groups of cells, F_{fin} is a growing function of M_0. This is again a robust result that continues to hold, for example, if we suppose that the initial seeds are composed by quadruples of cells instead of couples.

It should be remarked that this difference in behavior is not due to the actual number of initial cells (as can easily be checked) but to the fact that in the second case each A cell is born close to at least one B cell, and it is therefore subject to competition pressure from the very beginning. In the former case nuclei composed of all A cells have some generations available for reproduction before colliding with other nuclei composed of B cells.

It has been observed in reported experimental data that the growth of cell cultures exposed to a chemical carcinogen show a power-law increase of F_{fin} with M_0 like equation (1). The data by Fernandez in [6] concerning C3H10T1/2 cells exposed to methylcholanthrene, show an initial slope of about 0.4, in agreement with those of Figure 7. Note that no adjustable parameters have been used here, although we have chosen a value for the ratio of the parameters $p_{\text{RA}}/p_{\text{RB}} = 0.7$, which agrees with the estimate in [6] of the analogous parameter in their model (see section 3). Also the data by Haber in [24] concerning C3H10T1/2 cells exposed to benzo(a)pyrene and those of Reznikoff in [25] concerning C3H10T1/2 cells exposed to 3-methylcholanthrene show an increase in the number of foci per dish, with increasing seeding density, with a similar slope.

This is the kind of behavior which we would expect in our model, if the initial event of activation would affect only one of the two strands of the DNA of a replicating cell—so that A cells always appear close to their parent B cells. There are however other experiments (reseeding experiments) where some cells are taken from a dish after reaching confluence and then replated to a new dish. In this case, replating would lead to initially sparse cells, so we would expect the dependence of F_{fin} upon M_0 to be rather low. Actually, interesting data by Kennedy in [19, 20] and Mordan in [18] show that in replating experiments there exist large intervals of M_0 values where no appreciable difference in foci per dish is observed. Contrasting with this observation, there is indeed one data set [24] where F_{fin} seems to be a growing function of M_0 in replating experiments, but the values of the variables are such that in several cases no foci are actually observed.

Note also that some interesting experiments have been performed using x-ray irradiation instead of a chemical carcinogen [19, 20, 26, 27]. It has been suggested that in this case the intermediate activated state of the cell corresponds to a large genome-wide damage to the DNA [27] which is supposed to be inheritable and prone to develop

later in a fully transformed state. Therefore, our model should also be applicable to this case. One could speculate that in such a case the wide damage could affect both DNA strands, and therefore there would exist initial clusters formed by activated cells only. Since these experiments have been performed by reseeding, it is not possible to provide an independent test of this hypothesis.

5. Conclusions

We wish to comment here on the major lessons learned in the development and testing of the models and to address some issues that need further investigation.

The capability of mean-field ordinary differential equation (ODE) models to describe the major experimental data was shown to be fragile. Models which are similar to those of section 3 (although making use of cruder approximations) had been proposed in the past [6] and this feature had not been noticed. Moreover, the relevant variables are the number of foci, no attention being paid to their position or dimension, so ODE or discrete difference equation (DDE) approaches might seem perfectly fit. An interesting aspect is that the locality of the foci generating interactions cannot be ignored even if we are not concerned with the space dependence of the variables.

It should then be recalled that one of the major drawbacks of cellullar automata (CA) modeling is the lack of "analytical" methods, similar to those that are so useful in dealing with (some) differential equations. Although the body of theoretical knowledge about CA is growing [28] there is so far no analytical treatment of models like those discussed here (see also [11] for a recent review).

In this situation an approach based upon a mixed use of differential equations and CA can be very powerful. While this statement may appear obvious in the case of partial differential equations (PDEs), it is not trivial in the case of ODEs. The example discussed in this paper shows that indeed, at least in this case, the combined use of the two modeling methods may provide useful insights.

The ODEs by themselves are not able to provide a satisfactory account of the phenomena. Although the model fits the data it does so in a fragile way which embodies unrealistic values of a key parameter and when the unrealistic value is changed the agreement is lost.

CA provide the required, robust agreement, when equipped with the appropriate initial conditions. They also allow one to ascertain that the initial conditions have a crucial role in determining the shape of the curve $F_{\text{fin}}(M_0)$ and to interpret old experimental data in a new way, pointing out a major (so far unnoticed) difference between standard tests and reseeding experiments. These observations came out of a number of simulations performed with different parameter values, and

the analytical results described in section 3, concerning the dependence of the shape of the curve upon the effective exponent of the growth of the population of cells, have been of great help in interpreting the results of the simulation tests.

In particular, the fact that $dF_{\text{fin}}/dM_0 < 0$ for effective growth exponents smaller than 0.8 to 0.9 (obtained by studying ODEs) indicated that the "cure" of the observed discrepancy between data and the CA model was not to be found by slightly modifying the system parameters, but that a different attack was needed, and this led to the analysis of the initial conditions.

The analytical results were also useful to provide consistency checks of the model validity. For example, the fact that the number of non-transformed cells M grows approximately with t^2, as it is to be expected if the growth takes place on the borders of a two-dimensional cluster ($dM/dt \propto M^{1/2}$), gave support to the kind of replication dynamics we have developed.

The overall idea is then the following: CA models of natural phenomena require the exploration of a large parameter space, a procedure that is long and sometimes misleading. A fundamental caveat in order to avoid trivial mistakes is that of looking for robust agreement between model and data. The exploration of simplified ODE descriptions may provide hints for better modeling as well as checks of the models and, even more important, it may help in understanding the behavior of the CA model.

One further point is worth stressing: the possible crucial role of the initial conditions has been relatively easy to identify within the CA framework, where individual biological cells are modeled, so that they are born, compete, and die. If we had adopted a PDE approach, which looks at the cell population as a continuum, we could not have outlined the role of the initial conditions (unless perhaps in a very cumbersome way). Needless to say, this would have been impossible or very cumbersome and *ad hoc* also within an ODE framework. So the use of agent-based models like the CA described here, where single agents (biological cells, in our case) retain their individuality, is very well suited to deal with phenomena of this kind.

While the present work is concerned with *in vitro* tumors, it is tempting to speculate about the usefulness of adopting a similar approach to the study of *in vivo* tumors (for a review see [12, 29] and further references given therein). While the feasibility and usefulness of a CA description has already been proven for some solid tumors [4, 5], it would be interesting to try to combine differential equation methods with CA simulations, in an approach similar to the one of sections 3 and 4.

As far as the indications for further investigations are concerned, there are different aspects that are worth deepening. The following

indications for further work are not meant to cover every detail, but rather to point out some aspects which we feel particularly important.

From the experimental viewpoint, it would be important to assess the role of initial conditions by careful comparison of tests performed with and without reseeding. Moreover, it would be interesting to gain further information about cluster growth by performing a set of experiments that should be stopped at intermediate times between 0 and t_{fin}. Further information—although perhaps indirect—might be provided by the knowledge of the distribution of dimensions of the foci which are found at the end of the experiment.

The model developed here, following the lines of previous researchers, including Fernandez [6] and Little [26], is based upon the hypothesis that cells which appear "normal" may exist in at least two different states, B and A. Molecular biology might perhaps provide methods and techniques to distinguish between activated and normal cells in order to directly verify this key hypothesis. In this case, it could become possible to study the change in time of the corresponding populations.

From the theoretical and modeling viewpoint some issues which need further investigation include the developments of models which can describe the following.

- The growth of foci in time.

- The effects of chemicals (nutrients, promoters, etc.) and the actual change of culture medium.

- The effects of the carcinogen, that is, the birth of activated cells during the initial exposure period.

Moreover, there are also more technical aspects related to the CA modeling methods which may need to be addressed (e.g., testing different neighborhoods, using asynchronous updating, etc.).

Finally, we wish to stress that the present work provides a further confirmation of the usefulness of the framework of CA, invented more than 50 years ago by von Neumann, in describing the unfolding in time of complex biological processes.

Acknowledgments

Appendix

A. Calculations for the ordinary differential equation model

We now study the asymptotic behavior of equation (4) (section 3). We show that, in the interesting case where there are initially both A and B cells, the asymptotic state is one where all the A cells become extinct.

Let for simplicity $\mathbf{X} = (A, B)$: then $\mathbf{X} = \mathbf{0}$ (i.e., $A = B = 0$) is a fixed point of equation (4) but it is an unstable fixed point.

In order to demonstrate this statement we perform a linear stability analysis by letting $\mathbf{X} = \mathbf{0} + \delta\mathbf{X}$. Neglecting second and higher order terms we obtain:

$$\frac{d\delta A}{dt} \approx [Gs(0) - \omega - p]\delta A$$

$$\frac{d\delta B}{dt} \approx [Gs(0) - \omega]\delta B$$

which can be written as $d\delta\mathbf{X}/dt = P\delta\mathbf{X}$. The matrix P is diagonal and its eigenvalues are the solutions of

$$\det(P - \lambda I) = [Gs(0) - \omega - p - \lambda][Gs(0) - \omega - \lambda] = 0.$$

Provided that $Gs(0) > \omega + p$ (which is the interesting case where cells do actually grow in the plate) both eigenvalues are positive and the instability is therefore proven. Of course if there are no cells, this state persists, but the state of complete absence of cells cannot be reached from any state with nonzero cells (according to the deterministic dynamics used here, if we were to adopt a more realistic stochastic dynamics, then small initial populations might become extinct).

Let us consider other fixed points of equation (4). They are solutions of

$$[Gs(A_\infty + B_\infty) - \omega - p]A_\infty = 0 \tag{A.1}$$

$$[Gs(A_\infty + B_\infty) - \omega]B_\infty = 0. \tag{A.2}$$

Let us suppose that $A_\infty \neq 0$, $B_\infty \neq 0$: dividing equation (A.1) by A_∞ and equation (A.2) by B_∞, we obtain that $[Gs(A_\infty + B_\infty) - \omega - p] = 0$ and $[Gs(A_\infty + B_\infty) - \omega] = 0$, which cannot both be satisfied if $p \neq 0$. Therefore asymptotic states where both A and B take nonvanishing values are not allowed in this system.

The only possible asymptotic states are therefore either of the form $(A_\infty, 0)$ or $(0, B_\infty)$. Note that, if $A(0) \neq 0$ and $B(0) = 0$, no B cell will ever appear, so the final state will be of the form $(A_\infty, 0)$ (and of course if $A(0) = 0$ and $B(0) \neq 0$, no A cell will ever appear, so the final state will be of the form $(0, B_\infty)$). Let us consider what happens in the really interesting case where $A(0) \neq 0$ and $B(0) \neq 0$ by letting

$$y(t) = \frac{A(t)}{B(t)} \tag{A.3}$$

from equation (4) one directly finds that

$$\frac{d \ln(y)}{dt} = -p < 0. \tag{A.4}$$

$\ln y$ is a decreasing function of t, and so is y. The asymptotic state is therefore $(0, B_\infty)$ where B_∞ is determined by $Gs(B_\infty) - \omega = 0$.

Let us also observe that by integrating equation (A.4) one finds

$$y(t) = y_0 e^{-pt} \tag{A.5}$$

which provides a quantitative description of the evolution towards the asymptotic state.

Let us now determine the time history $\{M(t)\}$. From equation (5) one finds that if pA is much smaller than $[Gs(M) - \omega]M$, then

$$\frac{dM}{dt} \approx [Gs(M) - \omega]M. \tag{A.6}$$

The approximation holds well at initial times, when saturation effects are not apparent. At later times, when $Gs(M) - \omega$ is small, the A cells have almost disappeared from the system so the approximation used here continues to hold.

Let t^* be the moment when M equals s_2 (we suppose that $s_1 = 0$):

$$M(t^*) = s_2. \tag{A.7}$$

M grows with pure first order kinetics in the interval $[0, t^*]$ so

$$M(t) = M_0 e^{(G-\omega)t}$$

therefore:

$$t^* = \frac{1}{G - \omega} \ln\left(\frac{s_2}{M_0}\right). \tag{A.8}$$

According to equation (A.6), M is ruled by a Verhulst equation after t^*:

$$\frac{dM(t)}{dt} \simeq [Gbs_3 - \omega]M - GbM^2$$

whose analytic solution is known [17]. Therefore, if we adopt the

approximation of equation (A.6), $M(t)$ is given by

$$t \leq t^* : \quad M(t) = M_0 e^{(G-\omega)t}$$

$$t > t^* : \quad M(t) = s_2 \frac{bs_3 - \frac{\omega}{G}}{bs_2 + \left(1 - \frac{\omega}{G}\right) e^{-(Gbs_3 - \omega)(t-t^*)}}. \tag{A.9}$$

$A(t)$ can then be determined from the relationship $A = zM$, where $z = y/(1 + y)$. In the following we consider cases where the number of initially activated cells is always smaller than the number of B cells. Note also that if y_0 is small, $y(t)$ will be even smaller. The number A can therefore be estimated by the approximate relationship $A = yB \approx yM$ (which is used here as it leads to analytical results, see below).

Therefore, within this approximation

$$t \leq t^* : \quad A(t) = y_0 M_0 e^{(G-\omega-p)t}$$

$$t > t^* : \quad A(t) = s_2 y_0 \frac{\left(bs_3 - \frac{\omega}{G}\right) e^{-pt}}{bs_2 + \left(1 - \frac{\omega}{G}\right) e^{-(Gbs_3 - \omega)(t-t^*)}}. \tag{A.10}$$

Let us now consider the number of foci $F(t)$. From equation (6) one gets (for $t > t^*$):

$$F(t) = F_0 + p'G \int_0^{t^*} A(s)ds + p'Gb \int_{t^*}^t (s_3 - M(s))A(s)ds$$
$$= F_0 + \Delta F(t^*, 0) + \Delta F(t, t^*). \tag{A.11}$$

The total number of foci is expressed as the sum of the following three terms.

- F_0, those which were already there at $t = 0$ (and are neglected in the following).

- $\Delta F(t^*, 0)$, those which are formed before contact inhibition starts to be felt, that is, before t^*.

- $\Delta F(t, t^*)$, those which are formed after t^*.

We first estimate $\Delta F(t^*, 0)$. From equations (A.10) and (A.11) one gets

$$\Delta F(t^*, 0) = p'G \int_0^{t^*} A(s)ds = p'GM_0 y_0 \int_0^{t^*} e^{(G-\omega-p)s} ds$$
$$= \frac{p'GM_0 y_0}{G - \omega - p} \left\{ \left(\frac{s_2}{M_0}\right)^{1 - \frac{p}{G-\omega}} - 1 \right\}. \tag{A.12}$$

If $M_0 \ll s_2$, the last term on the righthand-side of equation (A.12) can be neglected and

$$\Delta F(t^*, 0) \simeq \frac{p'GM_0 y_0}{G - \omega - p} \left(\frac{s_2}{M_0}\right)^{1 - \frac{p}{G-\omega}} = \frac{p'Gy_0}{G - \omega - p} s_2^{1 - \frac{p}{G-\omega}} M_0^{\frac{p}{G-\omega}}, \tag{A.13}$$

therefore $\Delta F(t^*, 0)$ grows with M_0 according to a power law.

We now consider the third term in equation (A.11), that is,

$$\Delta F(t, t^*) = p'Gb \int_{t^*}^{t} (s_3 - M(s))A(s)ds$$

$$= p'Gby_0 \int_{t^*}^{t} e^{-ps}(s_3 M(s) - M^2(s))ds.$$

Note that, according to equation (A.10), M depends upon $s - t^*$. We can make this explicit by posing $s' = s - t^*$ and $M'(s') = M(s)$, that is,

$$M'(s') = s_2 \frac{bs_3 - \frac{\omega}{G}}{bs_2 + \left(1 - \frac{\omega}{G}\right)e^{-(Gbs_3 - \omega)s'}}.$$

Since

$$e^{-ps} = e^{-p(s' + t^*)} = e^{-pt^*}e^{-ps'} = \left(\frac{s_2}{M_0}\right)^{-\frac{p}{G-\omega}}e^{-ps'},$$

we obtain

$$\Delta F(t, t^*) = p'Gby_0 e^{-pt^*} \int_0^{t-t^*} e^{-ps'}(s_3 M'(s') - M'^2(s'))ds'$$

$$= p'Gby_0 \left(\frac{M_0}{s_2}\right)^{\frac{p}{G-\omega}} \int_0^{t-t^*} e^{-ps'}(s_3 M'(s') - M'^2(s'))ds'.$$

Now let

$$\phi(x) \equiv e^{-px}(s_3 M'(x) - M'^2(x))$$

(note that M' remains bounded at $\lim_{x\to\infty} \phi(x) = 0$), then

$$\Delta F(t, t^*) = p'Gby_0 \left(\frac{M_0}{s_2}\right)^{\frac{p}{G-\omega}} I(t - t^*)$$

$$I(t - t^*) \equiv \int_0^{t-t^*} \phi(s)ds.$$

$\Delta F(t, t^*)$ shows an explicit dependence upon M_0 in the first equation. $I(t - t^*)$ might depend upon M_0 through the dependence of the upper integration limit upon t^*, which in turns depends upon M_0. By applying the Leibniz rule one finds

$$\frac{dI(t - t^*)}{dM_0} = \phi(t - t^*)\frac{d(t - t^*)}{dM_0} = -\frac{\phi(t - t^*)}{G - \omega}\frac{d}{M_0}\ln\frac{s_2}{M_0} = \frac{\phi(t - t^*)}{M_0(G - \omega)}.$$

However, since $\phi(t - t^*)$ decays exponentially with time constant $1/p$, the derivative approximately vanishes for sufficiently long times, leaving only the previously determined dependence of $\Delta F(t, t^*)$ upon M_0:

$$\Delta F(t, t^*) \simeq p'Gby_0 s_2^{-\frac{p}{G-\omega}} I M_0^{\frac{p}{G-\omega}}, \tag{A.14}$$

where I depends upon t but is approximately independent of M_0.

The dependence of $\Delta F(t^*, 0)$ and $\Delta F(t_{\text{fin}}, t^*)$ upon the number of initial seeds M_0 is in both cases (equations (A.13) and (A.14)) a power law with the same exponent. Therefore

$$F_{\text{fin}} = \text{const} \cdot M_0^{\frac{p}{G-\omega}}. \tag{A.15}$$

B. Threshold values

Equations (16) and (17) from section 3 allow us to roughly estimate the threshold value for the exponent v below which one may expect that $dy/dt < 0$, and therefore that $dF/dM_0 > 0$.

Indeed, we are interested in values at times close to $t = 0$, so we substitute y_0 for y in the two equations. Since $G \gg \omega$, from equation (16) we obtain

$$\frac{dy}{dt} < 0 \Leftrightarrow G(1 - y_0^{1-v}) < p$$

which in turns implies that

$$v < v_{\text{th}} = 1 - \frac{\ln\left[1 - \frac{p}{G}\right]}{\ln y_0} \simeq 1 + \frac{p}{G \ln y_0}.$$

If $p/G = 0.3$ and $y_0 = 0.1$, this provides a threshold of about 0.87. Numerical simulations support the above conclusions.

If the initial number of cells is very high equation (17) should be used. Since $s_3 \gg M$, $G \gg \omega$, and bs_3 is of the order of unity, then $Gbs_3 - \omega - GbM \simeq Gbs_3$ and equation (17) leads to

$$\dot{y} \simeq \frac{A^v}{B}[Gbs_3(1 - y^{1-v}) - p]$$

therefore $dy/dt > 0$ implies

$$1 - y^{1-v} > \frac{p}{G}\left(1 - \frac{s_2}{s_3}\right)$$

which in turn leads to the following condition for a positive derivative

$$v < v_{\text{th}} = 1 - \frac{\ln\left[1 - \frac{p}{G}\left(1 - \frac{s_2}{s_3}\right)\right]}{\ln y_0} \simeq 1 + \frac{p}{G \ln y_0}\left(1 - \frac{s_2}{s_3}\right)$$

which provides a slightly higher threshold than the previous case (i.e., $v_{\text{th}} \simeq 0.9$ if $s_2/s_3 \simeq 1/4$).

C. The cellular automata model

The CA is formally defined as a quadruple $A = \langle \mathbf{G}, \mathbf{V}, \mathbf{Q}, f \rangle$ with the terms defined as follows.

- **G** is the cellular space, in our case $\mathbf{G} \subset \mathbf{Z}^2$ is a finite subset of the set of points with integer coordinates in two-dimensional euclidean space.

- **V** defines the neighborhood; using relative coordinates
  $\mathbf{V} = \{(0,0), (0,1), (1,1), (1,0), (1,-1), (0,-1), (-1,-1), (-1,0), (-1,1)\}$.

- **Q** is the state space: it is the cartesian product of the state space of (*i*) a variable describing the kind of biological cell, (*ii*) a variable describing the reproductive state, and (*iii*) a variable describing the relative direction in which reproduction could take place: $\mathbf{Q} = \mathbf{X} \times \mathbf{Y} \times \mathbf{D}$, where $\mathbf{X} = \{B, A, T, E\}$; $\mathbf{Y} = \{0, 1\}$; $\mathbf{D} = \mathbf{V} - \{0, 0\}$.

- *f* is the transition function which describes how the state of a CA cell is determined from the knowledge of the previous states of its neighbors. The basic steps are the following.

  for every *t*
  {
  for every site *i* such that $X_i \neq E$

  - determine whether reproduction will be attempted (by comparing a stochastic variable with a threshold which determines the reproduction rate); let $\mathbf{G}_1$ be this set of sites

  - for every site in $\mathbf{G}_1$, verify whether there is at least one empty site in the neighborhood; let $\mathbf{G}_2$ be the set of CA cells with at least one empty neighbor

  - repeat

  - {

  - for every site in $\mathbf{G}_2$, determine the direction of reproduction; if there are more available sites, choose at random among them

  - for every site *i* such that $X_i = E$

    * determine whether at least one of the neighbors has a direction of reproduction pointing to *i*; let $\mathbf{G}_i$ be the set of these CA cells
    * if all the sites in $\mathbf{G}_i$ have a common state, then set $X_i(t + 1)$ equal to that state; otherwise, choose at random among the states
    * assign at random the newborn biological cell in $\mathbf{G}_i$ to one of its parents (let it be P_i)
    * set $X_i(t + 1)$ equal to the state of P_i
    * if the new state is A, then change it to T according to a fixed probability

  - }

  - remove P_i from $\mathbf{G}_2$

  }

References

[1] G. B. Ermentrout, L. Edelstein-Keshet, "Cellular Automata Approaches to Biological Modelling," *Journal of Theoretical Biology*, **160** (1993) 97–133.

[2] P. E. Seiden and F. Celada, "A Model for Simulating Cognate Recognition and Response in the Immune System," *Journal of Theoretical Biology*, **158** (1992) 329–357.

[3] S. Di Gregorio, R. Serra, and M. Villani, "A Cellular Automata Model of Soil Bioremediation," *Complex Systems*, **11**(1) (1998) 31–54.

[4] An-Shen Qi, Xiang Zheng, Chan-Ying Du, and Bao-Sheng An, "A Cellular Automaton Model of Cancerous Growth," *Journal of Theoretical Biology*, **161** (1993) 1–12.

[5] A. R. Kansal, S. Torquato, G. R. Harsh IV, E. A. Chiocca, and T. S. Deisboeck, "Simulated Brain Tumor Growth Dynamics Using a Three Dimensional Cellular Automaton," *Journal of Theoretical Biology*, **203** (2000) 367–382.

[6] A. Fernandez, S. Mondal, and C. Heidelberger, "Probabilistic View of the Transformation of C3H10T1/2 Mouse Embryo Fibroblasts by 3-methylcholanthrene," *Proceedings of National Academy of Sciences USA*, **77**(12) (1980) 7272–7276.

[7] A. W. Burks, *Essays on Cellular Automata* (University of Illinois Press, Urbana, 1970).

[8] T. Toffoli and N. Margolus, *Cellular Automata Machines* (MIT Press, Cambridge, MA, 1987).

[9] S. Wolfram, *Theory and Application of Cellular Automata* (World Scientific, Singapore, 1986).

[10] R. Serra and G. Zanarini, *Complex Systems and Cognitive Processes* (Springer-Verlag, 1990).

[11] S. Bandini, G. Mauri, and R. Serra, "Cellular Automata: From a Theoretical Computational Model to its Application to Complex Systems," *Parallel Computing*, **27** (2001) 539–553.

[12] J. A. Adam and N. Bellomo, *A Survey of Models for Tumor-immune System Dynamics* (Birkhauser, Boston, 1997).

[13] T. Kakunaga and H. Yamasaki, "Transformation Assay of Established Cell Lines: Mechanisms and Application," *IARC Scientific Publication n.67* (Oxford University Press, 1985).

[14] H. Yamasaki, *et al.*, "Nongenotoxic Carcinogens. Development of Detection Methods Based on Mechanisms: A European Project," *Mutat. Res.*, **353** (1996) 47–63.

[15] M. Vaccari, *et al.*, "Effects of the Protease Inhibitor Antipain on Cell Malignant Transformation," *Anticancer Research*, **19** (1999) 589–596.

[16] E. J. Matthews, J. W. Spalding, and R. W. Tennant, "Transformation of Balb/c 3T3 Cells: V. Transformation Responses of 168 Chemicals Compared with Mutagenicity in Salmonella and Carcinogenicity in Rodent Bioassay," *Environmental Health Perspectives*, **101**(Supplement 2) (1993) 347–482.

[17] L. Edelstein-Keshet, *Mathematical Models in Biology* (McGraw-Hill, New York, 1998).

[18] L. J. Mordan, J. E. Martner, and J. S. Bertram, "Quantitative Neoplastic Transformation of C3H10T1/2 Fibroblasts: Dependence upon the Size of the Initiated Colony at Confluence," *Cancer Research*, **43** (1983) 4062–4067.

[19] A. R. Kennedy, M. Fox, G. Murphy, and J. B. Little, "Relationship between X-ray Exposure and Malignant Transformation in C3H10T1/2 Cells," *Proceedings of National Academy of Sciences USA*, **77**(12) (1980) 7262–7266.

[20] A. R. Kennedy, J. Cairns, and J. B. Little, "Timing of the Steps in Transformation of C3H10T1/2 Cells by X-irradiation," *Nature*, **307**(5946) (2001) 85–86.

[21] M. Mesnil and H. Yamasaki, "Cell-cell Communication and Growth Control of Normal and Cancer Cells: Evidence and Hypothesis," *Molecular Carcinogenesis*, **7** (1993) 14–17.

[22] R. Serra, M. Villani, L. Castelli, and A. Colacci, "Modelling the Birth of Transformation Foci in Cell Cultures," *International Journal of Computing Anticipatory Systems*, **8** (2001) 371–385.

[23] R. Serra and M. Villani, "Describing *in vitro* Cell Proliferation and Transformation with Cellular Automata," in *Artificial Intelligence and Heuristic Methods in Bioinformatics*, edited by R. Shamir and P. Frasconi (Springer (NATO ASI Series), Heidelberg, in press).

[24] D. A. Haber, D. A. Fox, W. S. Dynan, and W. G. Thilly, "Cell Density Dependence of Focus Formation in the C3H10T1/2 Transformation Assay," *Cancer Research*, **37** (1977) 1644–1648.

[25] C. A. Reznikoff, J. S. Bertram, D. W. Brankow, and C. Heidelberger, *Cancer Research*, **33** (1973) 3239–3249.

[26] J. B. Little, "Cellular Mechanisms of Oncogenic Transformation *in vitro*," in edited by Kakunaga and Yamasaki, 1985.

[27] J. B. Little, "Radiation-induced Genomic Instability," *International Journal of Radiation Biology*, **74**(6) (1998) 663–671.

[28] A. Wuensche, "Basins of Attraction in Cellular Automata; Order-Complexity-Chaos in Small Universes," *Complexity* **5**(6) (2000) 19–25.

[29] N. Bellomo and L. Preziosi, "Modelling and Mathematical Problems Related to Tumor Evolution and its Interaction with the Immune System," *Mathematical and Computer Modelling*, **32** (2000) 413–452.