UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata
Dipartimento di Fisica

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata
Ciclo XXXI

Dynamical feedback models in cellular biophysics

Settore Scientifico Disciplinare FIS/03

Supervisore: Professor Guido TIANA

Coordinatore: Professor Francesco RAGUSA

Tesi di Dottorato di:
Filippo Cola

Anno Accademico 2017-2018

Commission of the final examination:

External Referees:
- **Francesco Ginelli**, ICSMB and Physics Department, University of Aberdeen, UK;
- **Mario Nicodemi**, Università degli Studi di Napoli Federico II, Dipartimento di Fisica.

External Members:
- **Michele Caselle**, Università degli Studi di Torino, INFN e Dipartimento di Fisica;
- **Raffaella Burioni**, Università degli Studi di Parma, Dipartimento di Scienze Matematiche, Fisiche e Informatiche.

Internal Member:
- **Stefano Zapperi**, Università degli Studi di Milano, Dipartimento di Fisica "Aldo Pontremoli".

Final examination:

20th November 2018

Università degli Studi di Milano, Dipartimento di Fisica "Aldo Pontremoli", Milano, Italy

To Giacomo Pietro Cola
(1928-2017)

Cover illustration:

Human Cervical Adenocarcinoma Cells (HeLa cell line)

MIUR subjects:

FIS/03

PACS:

87.18.Hf
87.19.xj

Contents

List of Figures

List of Tables

Introduction

The content of the present PhD thesis deals with computational models of cellular and subcellular phenomena, using as tools ordinary and delay differential equations and spatial stochastic simulations. In particular, the chapters of the thesis correspond to the two distinct works which have been the object of my PhD research: in the first we investigated the role of oscillations in a two-node genetic regulation network (GRN) with delay, while in the second we developed a model to explain the peculiar dynamics of a cancer stem cells population in a melanoma tissue as observed in a series of experiments in the wet lab.

The work of Chap. 1 ("How likely are oscillations in a genetic feedback loop with delay?") is an extension of the master thesis of F. Marchetti (1) and stems from the observation that genetic control networks display temporal oscillations as a result of delays in their homeostatic control. A relevant question about these systems is whether the oscillating regime is a rare feature, or it corresponds to a sizeable volume of the space of parameters. The importance of this question become clear if we transpose it in evolutionary terms: if we are in the first case it means that evolution has driven living systems to build specific configurations to obtain oscillatory behavior, through a fine-tuning of the network's parameters, while if oscillatory patterns are common features of genetic circuits (i.e. there is are a large volume in the space of parameters of the networks that corresponds to oscillatory states), it means that evolution has forced GRN to avoid them when not necessary. The answer to the question is not trivial mainly due to the large number of parameters controlling the rate equations which can be adopted to model the network (as the number of parameters scales quadratically with the number of nodes in the network) and the lack of sufficiently precise experimental estimations of their values.

To provide a first answer to the question, we considered a simple model of a two-node genetic regulation networks, consisting of two different molecular species (for example, proteins or RNAs) expressed by the cells and involved in a mutual feedback loop. Apart from being a common network motif that can constitute a functional sub–unit of more complex genetic networks, it is proved that a system of this kind is the most simple whose dynamics can exhibit oscillations. The evolution of the concentration of the two species has been modelled by a system of two differential (rate) equations characterised by a delay in the feedback, accounting for a set of molecular processes which take time to be finalised.

We have developed an efficient sampling scheme of the parameter space, based on a Monte Carlo algorithm, and applied it to the two-node system with delay, characterised by a 8-dimension parameter space. The result is that the volume fraction of the parameter space associated with oscillations is small but not negligible, and it is weakly de-

pendent on the duration of the delay. The most critical parameter to control oscillations is the coupling production rates, which must have opposite sign, giving rise to a negative feedback loop. The oscillating regions are connected except along the equilibrium constants between the two species, not allowing neutral evolution along this parameter. The content of this first chapter has resulted in the following research paper: Cola F., Marchetti F., Tiana G., "How likely are oscillations in a genetic feedback loop with delay?", Eur Phys J E Soft Matter. 2017 Aug; 40(8):74.

The work of Chap. 2 ("Control of phenotypic switching in cancer cells"), which is to be published soon, is concerned with the dynamics of cancer cells populations in an in-vitro experiment, with particular relevance accorded to cancer stem cells.

According to the so-called "cancer stem cell hyphotesis", which is gaining more and more experimental evidence in oncology, tumors are organized in an aberrant cell hierarchy, in which differentiated cancer cells (CC) have a limited capacity to proliferate and are produced by a tiny subpopulation of parent cancer stem cells (CSCs), that can replicate indefinitely. If validated, this model would represent a paradigm shift for the biology of carcinogenesis as well as having important clinical implications, possibly with the design of therapies that selectively target the subpopulation of CSC in a tumor and are able to effectively stop the dynamics of the whole tumor without the side effects provided by chemotherapy. However, the question has not yet been settled, and among the other issues there is debate about whether the tumor cell population is strictly hierarchically organized or whether stochastic transitions that mutate differentiated cancer cells to CSCs are also important. The degree of plasticity between the two compartments of cancer cells has been so far the subject of various experimental investigations, but without a definitive answer

In a significant series of experiments performed in the group of C. La Porta at the University of Milan, biologists sort human melanoma cells using cancer stem cell markers, i.e. surface proteins widely believed to be expressed only by CSC. A large number of CC (i.e. cells from cancer tissue negative to the CSC-biomarkers) are plated in a neutral culture medium with a tiny fraction (under the 1-3 % of unsorted cells) of positice cells (CSC). What is observed is that in a few days the fraction of CSC-marked cells first overshoots to a higher level and then returns to the level of unsorted cells. In successive experiments, biologists observed that a mass of negative cells plated in a special culture medium (one resulting from the growth of a mass of melanoma cells for several days in experiments like those in the previous series, and thus repleted with the molecular species expressed by positive and negative cells during their growth), the re-expression of positive markers is delayed, retarded with respect to the case of the neutral initial medium.

We developed a model to test whether the possibility of stochastic transitions from CC to CSC (called phenotypic switching), coupled with regulation performed by an underlying field of a (so far) unknown switching-inhibitor molecule produced by the CSCs themselves (possibly miRNA, according to the biologists), can qualitatively reproduce the homeostatic regulation of the CSC population as observed in the wet lab. The model consists in a system of differential equations which describe the evolution of the CSC an CC population. The model includes the process of aging of the CCs, the processes of cellular differentiation by division, the production of the switching-inhibitor molecule by the CSCs, the phenotypic switching of CC to CSC and makes the key hypothesis that switching is a delayed process, thus requires a certain time to complete. This simple system of delay differential equations is able to qualitatively reproduce the evolution of the concentration of CSCs in the different experiments performed, the timing of the overshoot and the peculiar dependence of the observed height of the overshoot peak on

the initial fraction of CSC.

Recently developed techniques of immunohistochemistry (among which immunofluorescence is very popular) allow biologists to visualize the spatial distribution of a target molecule through a tissue or a biological sample. In particular, cancer stem cells can be localized from the observed distribution of their specific biomarkers. In order to investigate the spatial organization of the cancer cell populations for a lab–grown tumor, to compare it with future experimental findings, we translated the rate equation model into a spatial stochastic model on a 2D regular square lattice. In this model the evolution of the different cellular processes is dictated by the stochastic Gillespie algorithm, coupled with the deterministic Fourier diffusion of the switching inhibitor-molecule. This model thus allow us to explore the role of diffusion on the spatial organization of the tumor and the fraction of stem cells on the total.

From the simulations it can be observed that the radial density of the fraction of CSC (starting from a single CSC in the centre of the lattice) always exhibits a peak at the exterior border of the growing tumor mass, and the higher the diffusion coefficient of the molecule, the higher the peak and the total fraction of CSCs. This peak, at least for lower values of the diffusion coefficient, grows in height for an initial transient time of several days, then the height becomes stationary, while the peak is transported exponentially in time in the radial direction. We then developed a simple model (in the form of a first-order transport P.D.E.) that is able to explain all these features of the CSC radial distribution, and also the stationarity of the total fraction of CSC for long times, as resultig from the simulations.

An interesting additional result from the spatial simulations is that lower values of the diffusion coefficient also favor anisotropy of the distribution of CSC and their localization in discrete high-density cluster at the border of the tumor mass, which is in accordance with the few experimental results available in literature.

CHAPTER 1

How likely are oscillations in a genetic feedback loop with delay?

1.1 Gene expression and regulation: introduction and modelling.

1.1.1 Gene expression and regulation.

All known forms of life – eukaryotes (including multicellular organisms), prokaryotes (such as bacteria), and also viruses – are in need of generating the macromolecules needed to perform the biological functions of life, starting from the information stored in the DNA.

The process by which information from a gene (a stretch of DNA that encodes information) is used in the synthesis of a functional gene product is called gene expression. It is a complex mechanism divided in several steps: DNA transcription, RNA processing and, in eukaryotes, transport, RNA translation and post-translational modification of proteins. Gene products are often proteins, but in non-protein coding genes such as transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA.

It is essential for all forms of biological life to be able to control the expression of gene products, since it increases the versatility and adaptability of an organism by allowing cells to express them when needed. Gene regulation gives the cell control over its structure and function, and is of supreme importance in the development of multicellular organisms, since it drives cellular differentiation and morphogenesis in the embryo, leading to the creation of different cell types that possess different gene expression profiles from the same genome sequence. It also makes organisms able to respond to environmental stimuli, and may also serve as a molecular basis for evolutionary change, since control of gene expression can have a profound effect on the functions of a gene in a cell or in a multicellular organism.

Since gene expression is made up of many molecular processes, there is a wide range of mechanisms that cells can use to regulate gene expression by modulating any such process. At the molecular level the easiest way to control gene expression is direct interaction with DNA: genes have in fact several binding sites i.e. molecules with a large affinity to bind a certain protein, near the coding region and in those sites proteins interact with the DNA (for example DNA polymerase starts to produce the RNA after is attached to the promoter region of the gene). In terms of regulation, binding sites may differ in function: an enhancer will encourage RNA transcription while a silencer prevents the binding of DNA polymerase to the promoter. Transcription factors are proteins specifically made to activate or inhibit some genes, by binding to an enhancer site of the gene and encouraging the production of the protein coded by the gene, or by binding to a silencer gene and thus stopping the production of the same protein. Transcription factors

action can also be modulated: other proteins which act as enzymes, i.e. micro-machines that that catalyse some reaction, can enhance or suppress the production and the action of transcription factors; a cellular signal can induce chemical transformation, such as phosphorylation, acetylation, or glycosylation of the initial factor. These chemical reactions, named post-translational modification, may change the ability of the protein to bind to DNA. Finally all the proteins and molecules undergo a degradation pathway in order to keep their activity under control, especially if the protein is damaged or overexpressed (2).

The molecules that interact with each other and with other substances in the cell to govern the expression levels of gene products, and thus constitute the physical substrate of the gene regulatory mechanism, are called gene regulators. In many cases, one gene regulator (which can be DNA, RNA, a protein or complexes of these) does not control only the expression level of one gene, but of many others. A genetic regulatory network is the collection of all the gene regulators involved in the control of expression levels of a set of genes, with a description of all of the possible interactions between each other and with the genes themselves.

The interaction of different kinds of molecular regulators can give birth to regulatory networks of growing complexity, obtained combining a huge variety of different network motifs (elementary irreducible sub-networks with specific topology and function); one of the most simple (and most common) example of a regulatory network motif is that of the negative feedback control loop, well known in electronics, which can be biologically implemented, for example, in the following way: a gene products its own protein, which in turn act as an enzyme that catalyses the production of a transcription factor which inhibit the production of the original protein, thus homeostatically regulating the amount of the original protein to be constant in time.

1.1.2 Graphical and quantitative models of genetic regulatory networks.

Gene regulatory networks (GRN) can be schematically represented with the tools of graph theory. A graph is an ordered pair $G = (V, E)$ comprising a set V of vertices, nodes or points together with a set E of edges, arcs or lines, which are 2-element subsets of V (i.e., an edge is associated with two vertices, and the association takes the form of the unordered pair of the vertices). The edges represents the links, the interactions between the nodes. If the interactions between the nodes are not symmetrical, and it is necessary to distinguish the edge outgoing from a node from the edge entering,, the arcs connecting two nodes can be represented by an oriented arrow (which is graphically very common), or by introducing a label s to the edge (typically $+$ for output and $-$ for input). The set $G = (V, E, s)$ is called a directed or oriented graph.

A GRN can be represented as an oriented graph, where each node represents either a gene whose activity is regulated or one of the molecular species that acts as a regulator. The edges connecting the nodes represent the molecular interactions between the objects represented by the nodes. They are oriented since the chemical reactions they represents have almost always a precise direction.

Graphs representing GRN are reconstructed collecting information about relationships between genes: information about these relationships is available in various biological database (for example EcoCyc (3), GeneNet (4), GeNet (5), KEGG (6) and RegulonDB (7)), which collect results from large-scale experiments on gene expression. Recently developed experimental techniques, like cDNA microarrays and oligonucleotidechips, can be employed to monitor the expression of thousands of genes at the same time, and have of course given a great impetus to the study of GRN: microarrays, in par-

ticular, are used for detecting difference in gene expression over different experiments, typically test vs control, helping in finding similarities in gene expression of multiple experiments, either following the changes in gene expression values in a time interval, or observing the response of the cell to an external stimulus.

The reconstruction of all the interactions between genes and regulators in a GRN allows us to make a first sketch of the network topology: it can be either a large-scale description (focusing only on the essential molecules) or as more fine-grained scheme, considering all the molecules involved in every intermediate process. This level of description, though giving some important information, cannot give us information about the actual overall expression of the genes involved (on the basis of the paradigm, stated by Philip Anderson, that "more is different": the overall behaviour of a complex system, like our GRN, is more than the sum of its constituent parts), nor about the temporal evolution of the level of gene expressions. In order to achieve this, we have to build a quantitative, rather than just a purely descriptive, model of the GNR. The benefits of reaching a quantitative setting are of undoubted importance: quantitative models of GRN could be used to simulate various scenarios, and even predict future behaviour of systems. The output of such simulations could provide hypotheses on the inner working of GRN to be tested in *ad hoc* designed experiments. At a further level, with increasingly large and refined models description may enlarge enough to elucidating the connection between gene regulation and phenotype, on a systemic scale, and complete understanding of the regulators, and their interactions, can help to identify molecular tar gets for specific drugs and consequently designing controlled systems with potential for producing disease-specific cures.

Several mathematical and computational models of GRN have been proposed in literature over time. They can be collected in different classes according to the specifications of the system and to the instruments employed. A first distinction to be made is between deterministic and stochastic models. In deterministic models, given the same initial conditions and the same parameters of the systems, the output of the model will be always the same. In stochastic models, the concentration or the number of regulator molecules are not defined with certainty, but are represented by a random variable which follows a certain probability distribution. Solving such a model implies finding the probability distribution for each molecule involved. The output of a specific realization of the systems (obtained almost always numerically) depends of course also on the random numbers employed in the stochastic simulation. The level of precision of the two kinds of models depends on the intensity and distribution of the intrinsic cellular noise: recent (as of 2007) experimental results (8; 9) have demonstrated that gene expression is an intrinsically stochastic process, and as such stochastic models have gained widespread diffusion and popularity. Deterministic models can however still provide a realistic description of genetic regulatory systems, especially if the numbers of molecules involved in the processes are sufficiently large and the microenvironment in which chemical and biological reactions take place is well-mixed.

Another distinction to be made is between statical and dynamical models. Statical models (such as Bayesian Networks) are only concerned with mutual interactions of genes and only gives information about equilibrium values. In dynamical models every quantity of interest is associated to a time-dependent function, so that it is possible to track the evolution in time of the system. Of course dynamical models are more accurate and more suitable to the aims set above, though they are of course more complex and dependent on more parameters.

An example of dynamical and stochastic model is that of the Boolean network (10). This stochastic technique defines the nodes of the graph as boolean variables (1 if the

gene is expressing and 0 if is dormant) and the interaction between the vertices are boolean function. The time-resolved expression is obtained integrating the sum of the input boolean functions for every node. Of course the information we get from this approach are only qualitative. In order to get quantitative details of the dynamics we have to build a more physical representation of the system. For a stochastic model, the most common choice is that of the Master Equation (coupled with computational algorithms to solve it efficiently on a calculator, such as the Gillespie algorithm (125)). Deterministic dynamical models are instead usually written in terms of differential equations (one of most common, powerful and versatile choice to analyse GRN).

1.1.3 Differential equations models of GRN.

Differential equations models for GRN are usually written in terms of ordinary differential equation (ODE). By doing this we make two key assumptions:

- the numbers or concentrations of molecules involved in the processes under examination are treated as continuous, rather than discrete, quantities: this is reasonable if the numbers of molecules involved in the processes are sufficiently large (so that the output is deterministic and noise is negligible);

- the reaction microenvironment is well-mixed (the dependence of the quantities of interest from the spatial variables is not relevant). Fluctuations in space and time are negligible and thus there are no stochastic or diffusion terms.

We assume that each node of the network correspond to a regulator molecule. We then associate to the node i of the network a function $x_i(t)$ representing the average number of molecules (or possibly the concentration).

To build the model we associate to every node of the genetic regulatory network a differential equation: thus the whole model will be made up of a system of ordinary differential equations, each for every node (that is, each for every dynamical variable). The equation is written in function of the concentration of the regulators involved in the biological process (those associated) and express the rate of production of the regulator associated to that node of the network. The general form of such equation will be the following

$$\frac{dx_i}{dt} = f_i(\underline{x}) \tag{1.1}$$

where x_i is the number (or concentration) of the regulator molecules associated with the node i, $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ is typically a non-linear function associated to the i-th node and $\underline{x} = [x_1, x_2, \ldots, x_n]$ are the quantities of the element of the network (generally in the function f_i only the terms corresponding to the regulators of the node connected to the i-th node will be present). The functions f_i are ultimately derived from basic principles of chemical kinetics or simple expressions derived from these, e.g. Michaelis-Menten enzymatic kinetics.

The most important principle of chemical kinetics that is regularly employed in determining th functional forms of the f_i is the *rate law*, which links the reaction rate of a chemical reaction (the speed at which reactants are converted into products) to the numbers or concentration of the reactants and the reaction rate constant (a specific, constant value for each process, usually associated with the edges of the graph representing the GRN).

For many reactions $A + B \xrightarrow{k} C$ the rate r is given by a power law such as

$$r = k \cdot [A]^p \cdot [B]^q, \tag{1.2}$$

where $[A]$ and $[B]$ express the numbers or concentration of the reactant species A and B, the exponents p and q are the partial orders of reaction for A and B and k is the reaction rate constant. For $p = q = 1$ this is the usual law of mass action. The partial orders and the reaction rate constants are usually derived from simple phenomenological considerations of from known experimental results. In the case of complex biological processes, determining the reaction rate constants could be extremely difficult: we are in possession of numerical parameters only of an handful of well-know system, for example the bacteriophage λ lysis-lysogeny decision circuit (11), and this lack of in-vivo or in-vitro measurements is hindering the use of numerical techniques.

A different functional forms that is used when a reaction is triggered by molecular binding is known as *Hill's function*, which can be applied for example in modelling the rate at which a gene product is produced when its parent gene is being regulated by transcription factors (e.g., activators and/or repressors) (13). This function applies also ($h >> 1$) when a gene is regulated by multiple binding sites for transcription factors, in which case the transcription factors may bind the DNA in a cooperative fashion (14). The consequence of cooperative interaction is that the rate of expression of the gene's product is strongly dependent on the concentration of the transcription factor and the response tends to saturate, which is usually modelled with a step function. Hill's function $g(x)$ is one of the most simple and common functional form (15) to describe this interaction

$$g(x) = \frac{x^h}{x^h + K_A^h}, \tag{1.3}$$

where x is the concentration of the transcription factor, h is the Hill coefficient describing the degree of cooperation of the reaction (a small integer number, which defines the slope of the curve: an higher parameter determines an higher steepness of the curve), and K_A is a threshold value: it stands for the point at which the function reach $\frac{1}{2}$. Note that for large values of x the function (which represents a sigmoid curve) tends to saturate, i.e. come close in value to 1.

If the production of protein from gene X is up-regulated (activated) by a transcription factor Y, then the rate of production of protein $X_{produced}$ can be modelled as a differential equation in terms of the concentration of activated Y_{active} protein:

$$\frac{d[X_{produced}]}{dt} = k \cdot \frac{[Y_{active}]^h}{[Y_{active}]^h + (K_A)^h} = k \cdot g([Y_{active}]), \tag{1.4}$$

where k is the maximal transcription rate of gene X. Of course the function $1 - g(x)$ can model the production of a protein from a gene which is down-regulated (repressed) by a transcription factor.

Finally, we have consider the process of self-degradation of the gene product, which is represented as a self-interaction of the i-th node, and modelled with a decreasing linear function $f(x_i) = -\gamma_i \cdot x_i$, where the parameter γ_i is the self-degradation rate of the elements (usually measured as half-life).

At each node we apply the principle of superposition, assuming that different molecules control gene expression independently on each other, and we conclude that that the total

rate of production of the local regulator x_i is obtained summing separately the contributions $\overline{f}_{ij}$ from the connected nodes and the one from self regulation $\widetilde{f}_i$

$$\frac{dx_i}{dt} = f_i(\underline{x}) = \sum_{j=1}^{n} \overline{f}_i j(\underline{x}) + \widetilde{f}_i(x_i) \tag{1.5}$$

Thus the general ODE which models the temporal evolution of the i-th node of the network can be written in the following form

$$\frac{dx_i}{dt} = \sum_{j \neq i} \alpha_{ij} \frac{x_j^h}{x_j^h + K_{ij}^h} - \gamma_i x_i + s_i. \tag{1.6}$$

The parameters α_{ij} are the saturation values of the Hill's functions that quantify the interaction of the i-th node with the j-th node: they can be viewed as the network connection weight, that represent how much the node j stimulates the node i (as an enhancer if $\alpha_{ij} > 0$, as a repressor if $\alpha_{ij} < 0$). The parameter γ_i is the self-degradation rate of the regulator of the node i (usually measured as a half-life), while s_i is the rate at which the regulator of the node i is produced: a constant source term for the node i.

Two observations have to be made with reference to equations 1.6. First, the number of parameters needed to model a GRN with this kind of systems is quite large and moreover it grows rapidly with the number n of nodes in the network, exhibiting a "curse of dimensionality": in fact, as n grows, the total number of parameters of systems goes like $n \cdot (3n + 2)$. If n is sufficiently large, even with modern techniques for conducting large-scale experiments on genetic expression (gene-chips, microarrays), a huge numbers of experiments would be needed to obtain information about the parameters of the systems. Second, given the complexity of most biological networks, even with a large amount of data it can be a complex task to infer precise information about the system's parameters. As we mentioned earlier, we have accurate numerical parameters only of an handful of well-know system, while for most of the others only esteems of their orders of magnitude, at best. New methods would be needed if we want to study the general properties of such systems in absence of precise quantitative information about the parameters of the model.

1.1.4 The role of delay in GRN.

The model 1.6 we built to describe the most general GRN is not complete, since it leaves out the fact that many molecular processes in gene regulation exhibit time delays which follows from the obvious fact that chemical reactions takes some time to be completed (and time is also needed for reactants and produced molecule to diffuse in the cell). In table 1.1 are reported the typical time scales of a single-molecule process in eukaryotic cells, from the fastest to the slowest (16).

Process	Time scale
Translocation through nuclear pores	$10^{-4}\ s$
Molecular diffusion in the cell	$1\ s$
Translation	$30\ s$
Transcription	$3\ min$
mRNA degradation	$3\ min$
Protein degradation	$10\ min$ to $10\ h$
Cellular signals	$1\ h$

Table 1.1: Typical time scales of a single-molecule process in eukaryotic cells, presented in order from the fastest to the slowest (16).

It is evident that in a GRN, which often involves tens (or hundreds) of different chemical reactions, some of which organized hierarchically (the products of one are the reactants of another, so the second cannot start until the first has been completed), and which take different times to be completed, it is reasonable to assume that some process will display a net time delay τ, that can be conveniently modelled adding the time delay τ in the interactions among the graph nodes involved in the delayed process. A typical example is given by transcription and translation mechanisms, which are composed of a large number of in-between stages: immediately after the polymerase is binded to the DNA, the newly-formed closed complex performs a transition and modifies its shape in an elongated one; the polymerase then runs along the unrolled DNA, synthesizing the mRNA polymer until it reaches the end of the gene.

We should also consider that, when making a coarse-grained model of a GRN, we are not interested in all the different intermediate reactions that makes up a complex process: the intermediate nodes (and the corresponding reaction products) can be replaced by a time delay between the nodes corresponding to the initial and the final regulators of the process.

Coming to our general model 1.6, to make it more realistic we added a time delay in the Hill's function that models the interactions between different nodes. This is to all intents and purposes a coarse-graining of all the intermediate steps and reactants that make up the regulation process (diffusion of the regulator, binding to the promoter, enzymatic reactions, formation of a protein complex, and so on).

The final and complete model we propose for the general GRN is thus the following

$$\frac{dx_i}{dt} = \sum_{j \neq i} \alpha_{ij} \frac{x_j^h(t-\tau)}{x_j^h(t-\tau) + K_{ij}^h} - \gamma_i x_i + s_i. \tag{1.7}$$

A system of n equations like (1.7) is in general impossible to solve analytically and have to be solved numerically. Such a system can usually exhibit a complex dynamics, due to nonlinearity and the presence of time delay. In principle, for generic n

($i = 1, \ldots, n$), the dynamics of the variable x_i could be even chaotic. For $n = 2$, if the possibility of chaotic dynamics is ruled out, the dynamics of each variable then, however complicated, can display only three kind of behaviour: the dynamics could converge to one point (a constant steady state), diverge to infinity or have sustained periodic oscillations. The last case is of particular interest for us, given our aim to establish whether the oscillating regime is a rare feature of GRN, or it corresponds to a sizeable volume of the space of parameters. In the next section we shall therefore provide a short introduction to the role of oscillation in GRN, and then, in a subsequent section, we shall focus on a more specific and simpler case of GRN: a two-node GRN with negative feedback. This kind of GNR, though simple and computationally tractable (given the relatively low number of parameters), has nonetheless a sufficiently rich dynamics, so that it is worth to employ it to investigate the occurrence of oscillations in GNR.

1.2 The role of oscillations in GRN.

All organism have to deal with situations that require exact timing or an efficient response to environmental variations (think of the circadian rhythms and cell division). For this reason biological system have developed extraordinary spatial and temporal patterns, which reflects the fact the underlying cellular processes are highly dynamic and operate on a wide range of timescales and lengthscales, with complex mechanisms of cooperation and synchronization. From the great amount of measurement of the temporal dynamics of protein concentrations (made possible by recent techniques such as pulse-labelling, β-galactosidase measurements and immunoblotting, fluorescent techniques, chromatin immunoprecipitation assays and microarrays (16)) it is now evident that regulatory and signal transduction networks do not depend merely on shifting the relevant protein concentrations from one steady state level to another. Rather, the signals often have a significant temporal variation that is able to carry much more information, as we shall see.

In particular, several genetic control systems in cells have been observed to display regular temporal patterns in the form of stable periodic oscillations: this is the case, for example, of p53 ((17; 18)), Hes1 (19), NfκB (20), and Wnt (21). As a rule, these oscillations have a period of the order of tens of minutes to hours, are associated to some specific mechanism and involve two main species linked in a feedback loop: P53 activates the transcription of Mdm2 which, in turn, promotes the degradation of p53; the resulting 20 minute oscillations are thought to coordinate the cellular response to DNA damage, causing growth arrest and eventually apoptosis (22). Also hormones, such as the human growth factor, also show such intermittently periodic behaviour and pulsatile secretion (23). The oscillatory dynamics is not only limited to the temporal domain, but it is also fundamental for spatial organisation: the oscillations of Hes1, which inhibits the transcription of its own mRNA, are turned into spatially repetitive units such as vertebrae, ribs and skeletal muscles during embryogenesis (24); another example is parallel root branching in plants (25).

We have seen that there are many examples of periodic oscillations in the dynamical behaviour of GRN and that they are of great importance for cellular life, but regulatory system could display another behaviour: multistationarity (26). Instead of oscillate between two extreme values, in the case of multistationarity the dynamics is forced to settle, at least after a long evolution time, on one of the fixed points of the system, according to their basins of attraction: the system thus in any case reaches a steady state. This kind of control is required to account for epigenetic differences, including cell differentiation; a sort of trigger circuit able to switch from "gene on" to "gene off" and vice versa. Once is verified that there are multiple possibilities is interesting to investigate how difficult it is for the cell to produce temporal oscillations against stable state configuration, and what are the benefits of choosing one dynamical behaviour over the other.

In order to better investigate the nature of the oscillatory regime it is useful to draw a comparison between biological and electronic circuits. The aims and functions of the two kinds of systems are in fact remarkably similar: the genetic regulation network duty is to process input signals coming from the cellular environment to create an appropriate response. From a theoretical and systemic point of view, then, one can think of genetic control elements as biological computers whose goal is to elaborate input stimuli to obtain an appropriate output. For example, transcriptional inhibition can be regarded as a logical NOT node in the cellular computing network, whereas promoters can build out logical AND nodes. Cellular life can thus be summarized in a continuous elaboration and diffusion and transport (i.e. transmission) of signals (besides the simpler produc-

tion of energy, required for the former goal).

Having established this plausible comparison, we can borrow theoretical techniques from information theory to study how much oscillations (and multistationarity) are important for GNR. That is, we can ask whether a cell get some benefit from the stable oscillatory dynamical regime behaviour, with respect to multistationarity. To answer to this question we need to find a value that quantifies the efficiency in the transmission of information by a circuit, and such a value can be the channel capacity (27), i.e. the maximum of the rate at which a communication channel transfers information. A higher transmission rate implies a better ability in elaborate and propagate signals in reaction to external stimuli, and hence a finer efficiency and an evolutionary advantage.

In the case of a two-stable response system the associate capacity I, in term of Shannon's information entropy, is $I = 1$ even in presence of noise (28), while if we consider a periodic signal it has been proved by Shannon himself (29) that the channel capacity is given by

$$I = T \cdot W \cdot \log_2\Big(1 + \frac{P}{N}\Big), \tag{1.8}$$

where P and N are, respectively, the spectrum and noise power, W is the highest value of the spectrum and T is the duration of oscillations. To provide a quantitative esteem of the capacity I in the case of our system, let'as assume that the duration is several hours long, so $T \sim 10h$ (16); concerning the spectrum, we do not expect any frequency in it higher than the inverse of the fastest process taking place in the generating genetic control elements, so for the maximum of the spectrum is reasonable to take the inverse of the delay of the fastest process involved, and then we have $W \sim 10^{-1}\text{min}^{-1}$. To evaluate $\frac{P}{N}$ is usually a complicated task but even if we choose $P \sim N$, that is the worse possible case , the value of the capacity will be $I \sim 40$ (30).

Finally, we have to stress the fact that a continuous rate equation (which is the model we are considering for our system) implies that for every x variables in input a signal y is obtained in output, therefore the amount of information accountable to the system seems to be infinite, but our differential equation is actually an approximation, a coarse-graining of the real situation: the number of particles in the cell have to be finite and also the rate equation don't envision the presence of noise that appear in every molecular process.

We can conclude that the amount of information transmitted in a genetic control circuit that exhibit an oscillatory response is greater than the that transmitted in the multistationary regime, even though the system is affected by noise. Thus the oscillatory regime presents an evident advantage in terms of efficiency and adaptability for living systems. What our present analysis leaves out is how difficult it is for the cell to produce temporal oscillations against stable state configuration. To establish the amount trade-off between efficiency and ease of implementation could help us to assess the plausibility of one of the two possible scenarios: oscillations are a common feature of biological control systems or instead evolution has driven organisms to build specific configurations to obtain oscillatory behaviour. Some recent work (30) suggests that periodic patterns naturally arise in a regulatory system whereas steady state are an insulated incident. If that scenario be true, evolution forced the hand at the molecular regulation mechanisms stressing the pursuit of multiple constant steady states.

1.3 Two-nodes systems with feedback.

In our work we take a reductionist approach, choosing to analyse a simple network motif, which can constitute a recurrent and statistically significant sub-graphs of a more complex regulatory network. The motif of interest is a simple two-node network with feedback.

In the case $n = 2$ (two nodes, i.e. two regulator molecules) the GNR has the graphical representation displayed in Fig. 1.1.

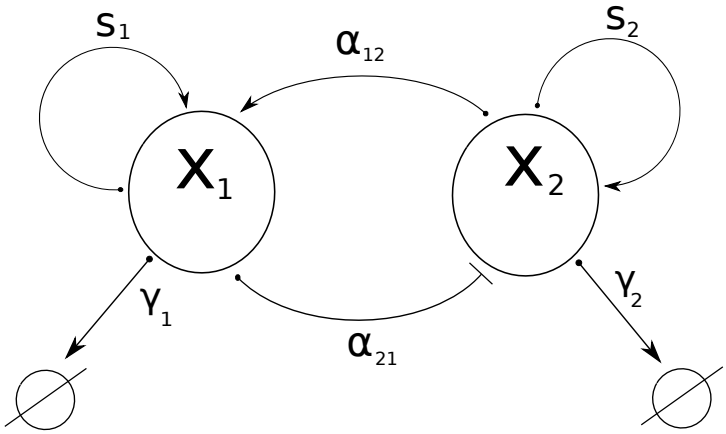

Figure 1.1: Simple graphical scheme of a GRN with two regulators molecule: the 1st node is a repressor and the 2nd an activator. x_1 and x_2 are the concentration of the two regulator species, α_{ij} are the connections weights, γ_i the self-degradation rates, s_i the local source rates.

The system of equations modelling the GNR (according to the general model of the previous section) is the following

$$\begin{aligned}\frac{dx_1}{dt} &= s_1 + \alpha_{12} \cdot \frac{x_2^h(t-\tau)}{k_{12}^h + x_2^h(t-\tau)} - \gamma_1 x_1 \\ \frac{dx_2}{dt} &= s_2 + \alpha_{21} \cdot \frac{x_1^h(t-\tau)}{k_{21}^h + x_1^h(t-\tau)} - \gamma_2 x_2,\end{aligned} \tag{1.9}$$

where x_1 and x_2 are the concentration of the two regulator species. Note that we assumed for simplicity that the processes of coupling between the two nodes are similar, so that they share the same time delay τ and the same Hill's coefficient h.

For this kind of system, the possibility of chaotic dynamics is ruled out by a generalized version of the Poincaré-Bendixson theorem for monotone cyclic feedback systems of ODEs with delay (31). The dynamics of each variable then, however complicated, can display only three kind of behaviour: the dynamics could converge to one point (ba steady state), diverge to infinity or have sustained periodic oscillations. In the first two cases if the equilibrium (or infinity) is reached directly the system is called stable "node" or fixed point (unstable "node") but if oscillates towards the steady state (or towards infinity) we talk about stable spiral (unstable). The oscillatory case is called limit cycle. The introduction of time delays can induce instability of the fixed point, and hence causing

the transition to a a limit cycle. In a large number of systems it has been verified that sufficiently large delay produce oscillations ((32; 33)).

1.3.1 Feedback loops.

With the graphical convention employed in Fig. 1.1, the node 2 is an *activator* for the node 1 ($\alpha_{12} > 0$), while the node 1 is a *repressor* for the node 2 ($\alpha_{21} < 0$). In other terms, node 2 has a positive feedback on node 1, while node 1 has a negative feedback on node 2. Positive feedback of regulator A on regulator B means that if the concentration of a regulator A goes up it makes the regulator B concentration growth, while in negative feedback if the concentration of regulator A goes up it makes the protein B concentration decrease.

The kind of system depicted in Fig. 1.1 is an example of a *negative feedback loop*, i.e. a control loop network with an even number of repressors (while of course a *positive feedback loop* is control loop network with an odd number of repressors). Feedback loops are, in general, the most common network motifs in cellular organization, especially when one considers the regulation of small molecules (34). Of course the simplest negative feedback loop (single-node) is a protein which represses itself; there are many examples of such proteins: the main repressor of the SOS regulon in *E. coli*, LexA, also represses its own production (35); Hes1 represses transcription of its own gene ((16; 35)).

General feedback loops with an arbitrary number of components N can be modelled by equations like the following

$$\frac{dx_i}{dt} = g_i(x_i, x_{i-1}) \quad i = 1, \ldots, N, \tag{1.10}$$

where the rate of change of the variable x_i at the node i of the feedback loop depends only on itself and the previous variable x_{i-1}. In modelling feedback loops, one usually assumes the function g_i of (1.10) to be monotonic in x_{i-1} (under this assumption, we shall call (1.10) a *monotonic system*). This reflects the hypothesis that a protein that activates a particular process cannot change to repress it at some other concentration, and vice versa, which is the case for most transcription factors (and implies that there is no ambiguity about whether the loop implements negative or positive feedback,).

For monotone systems with a negative feedback loop it can be proved rigorously (16) that there can be only one fixed point. If this point is stable, the system at infinity tends to reach the only available steady state, while if it is unstable, and the system's trajectories are bounded (again, a reasonable assumption for a biological system) then the system must display periodic oscillations, as it follows from the standard Poincaré-Bendixson theorem (36).

For positive feedback loops, on the other hand, it can be proved (16) that they can have multiple fixed points. Thus in this case the system tend to reach at infinity one of the available stable points,according to their respective basins of attraction.

Positive and negative feedback loops, in conclusions, represents two dynamical and biological paradigm: negative feedback loops represent stable periodical oscillations and homeostasis, while positive feedback loops stand for multistability, which can be seen as a scheme for the differentiation processes of developmental biology. In fact it can be proved (37; 38) that for a wide class of systems the presence of at least one negative feedback loop is a necessary condition for oscillations.

The condition is necessary, but not sufficient: it can be shown using Dulac's criterion that oscillations are not possible in the case of two-variable monotonic systems (16; 39). What is needed is the possibility for the system's only fixed point to become unstable,

thus generating sustained periodic oscillations with the mechanism cited above. To generate the instability of the fixed point, and hence oscillations, a time delay in the system is required : if a perturbation in the concentration of one variable instantaneously affects the concentration of the next one, and so on, then for a negative feedback loop any perturbation will be immediately cancelled and the steady state will be stable. A sufficiently large time delay on the other hand will produce oscillations: the mechanism involved is that of the Hopf bifurcation (16).

For a feedback loop system to have the possibility of generating periodic oscillations we have shown that the feedback must be negative, the trajectories of the system bounded, and if the system is monotone and two-component it must have an explicit time delay. In the next subsection we analyse the two-node system (1.9) and prove that it is the smallest realistic system of ODE which can generate sustained, periodic, oscillations.

1.3.2 Analysis of the two-node monotone systems with delay.

It can be easily seen that equations (1.9) of the two-node system are a particular case of monotone system (1.10) (with the additional features of the delayed variables). (1.9) are of the type

$$\frac{dx_i}{dt} = F_i(x_i(t), x_{i-1}(t-\tau)) \quad i = 1, \dots, 1, \tag{1.11}$$

where the functions F_i are monotonic in the second variable. We note that their trajectories, appropriately for the model of a biological system, are bounded; it is in fact

$$\frac{dx_i}{dt} \leq s_i + \max\{\alpha_{ij}, 0\} - \gamma_i x_i, \tag{1.12}$$

and the trajectories of the system

$$\frac{dy_i}{dt} = s_i + \max\{\alpha_{ij}, 0\} - \gamma_i y_i \tag{1.13}$$

are certainly boundend (incidentally, from this follows that the domain of existence of the solution x_i is the whole $\mathbb{R}$, since growth is sublinear).

We can quite easily show that if the system (1.9) is a negative feedback loop (i.e. if $\alpha_{12} \cdot \alpha_{21} < 0$ it has just one fixed point. Some of the variables in the system are delayed delayed but the presence of the delay do not affect the value of the fixed point, since when the equilibrium is reached $x_i(t) = x_i(t-\tau)$, therefore we can just analyse the following system

$$\begin{aligned} \frac{dx_1}{dt} &= s_1 + \alpha_{12} \cdot \frac{x_2^h(t)}{k_{12}^h + x_2^h(t)} - \gamma_1 x_1 = F_1(x_1, x_2) \\ \frac{dx_2}{dt} &= s_2 + \alpha_{21} \cdot \frac{x_1^h(t)}{k_{21}^h + x_1^h(t)} - \gamma_2 x_2 = F_2(x_1, x_2). \end{aligned} \tag{1.14}$$

We introduce the equilibrium values $\widetilde{x}_1$ and $\widetilde{x}_2$ (for which $\frac{dx_1}{dt}\Big|_{\widetilde{x}_1, \widetilde{x}_2} = 0$ and $\frac{dx_2}{dt}\Big|_{\widetilde{x}_1, \widetilde{x}_2} = 0$)

$$\widetilde{x}_1 = \frac{\alpha_{12}}{\gamma_1} \cdot \frac{\widetilde{x}_2^h}{k_{12}^h + \widetilde{x}_2^h} + \frac{s_1}{\gamma_1}$$
$$\widetilde{x}_2 = \frac{\alpha_{21}}{\gamma_1} \cdot \frac{\widetilde{x}_1^h}{k_{21}^h + \widetilde{x}_1^h} + \frac{s_2}{\gamma_2}. \tag{1.15}$$

Thus we have two coupled equations

$$\widetilde{x}_1 = f_1(\widetilde{x}_2)$$
$$\widetilde{x}_2 = f_2(\widetilde{x}_1). \tag{1.16}$$

Making a simple substitution we obtain $\widetilde{x}_1 = f_1(\widetilde{x}_2) = f_1(f_2(\widetilde{x}_1)) = f_1 \circ f_2(\widetilde{x}_1) = \Gamma(\widetilde{x}_1)$, where Γ is the composed function $f_1 \circ f_2$.

The number of fixed points of 1.14 is the number of fixed points of the map Γ, so we have to calculate the number of intersections of the curves $y = \Gamma(x)$ and $y = x$. The functions f_i are monotonic, so Γ, as a composition of monotonic function, is monotonic. If $\Gamma' < 0$ the two curves can intersect in just one point, since $y = \Gamma(x)$ is a decreasing curve, while $y = x$ is an increasing curve. On the other hand, if $\Gamma' < 0$ the two curves may cross in more than one point. By the chain rule for derivatives we have $\Gamma'(x) = f_1' \cdot f_2'$ and it is $\text{sign}(\Gamma'(x)) = \text{sign}(\alpha_{12} \cdot \alpha_{21})$.

Thus if $\alpha_{12} \cdot \alpha_{21} < 0$ (negative feedback: one repressor and one activator), the system (1.9) has just one fixed point, while if $\alpha_{12} \cdot \alpha_{21} > 0$ (positive feedback: two repressors or two activators) can have multiple fixed points.

In conclusion, if $\alpha_{12} \cdot \alpha_{21} < 0$ the system has only one fixed point: since its trajectories are always bounded, then if that fixed point is unstable, for the standard Poincaré-Bendixson theorem, the system dynamics exhibits a limit cycle, that is oscillates periodically. Since the system has an explicit delay in time, it can undergo a Hopf bifurcation and thus have an unstable fixed point (16).

We have proved then that our two-node delay system of ODE is the simplest non trivial system that can display a periodic stable oscillatory dynamics. Since it is of low dimension and thus depends on a relatively low number of parameters (in total 8, if we fix Hill's coefficients), it is a computationally tractable system to test our methods for the detection of stable periodic oscillatory states.

1.4 Sampling oscillatory states in the phase space: a Metropolis algorithm.

Our aim in this work is to investigate whether the the oscillating regime is a rare feature of GRN, or it corresponds to a sizeable volume of the space of parameters. The answer to this question is of importance to understand network motifs of gene regulatory systems in evolutionary terms: if oscillatory patterns are common features of genetic circuits (i.e. there is are a large volume in the space of parameters of the networks that corresponds to oscillatory states), it means that evolution has forced GRN to avoid them when not necessary, while if the size of oscillatory volume in the phase space is tiny, and thus oscillations are a rare feature, it means that evolution has driven living systems to acquire them through a fine-tuning of the network's parameters.

As we have already stated before, the main problem in evaluating the size of the phase space corresponding to oscillatory states is that the dimensionality of the space of parameters needed to model a GRN rapidly increases with the number N of nodes, scaling as N^2, thus exhibiting a "curse of dimensionality".

In order to make the analysis more manageable, our first choice has been to concentrate on a model which is simple, but not too simple, i.e. that combines a relatively low number of parameters and a sufficiently rich dynamics. As we argued in the previous sections, such a model (which shall be our model of reference) is that of a two-species feedback loop with delay, modelled via the following system of rate equation

$$\begin{aligned}\frac{dx_1}{dt} &= s_1 + \alpha_{12} \cdot \frac{x_2^h(t-\tau)}{k_{12}^h + x_2^h(t-\tau)} - \gamma_1 x_1 \\ \frac{dx_2}{dt} &= s_2 + \alpha_{21} \cdot \frac{x_1^h(t-\tau)}{k_{21}^h + x_1^h(t-\tau)} - \gamma_2 x_2,\end{aligned} \tag{1.17}$$

The model is specified by the following parameters

- s_1 and s_2 are the unconditioned production rates of the two molecular species, whose concentrations are x_1 and x_2;
- α_{12} and α_{21} are the production rates of each molecular regulator dependent on the other species;
- γ_1 and γ_2 are the the natural degradation rates of the molecular species;
- k_{12} and k_{21} are the thresholds of the Hill's functions that provide the regulatory coupling between the two species;
- h is the HIll's coefficient, which is assumed to be the same for both processes of coupling, for simplicity;

We have that the dynamics of the system (1.17) is described by 9 numerical parameters. Assuming that $h = 2$ (a common biological value for Hill's coefficient, see (40)), we further reduce the dimensionality of the phase space to 8.

The next step in our analysis is to devise a serial procedure that assign at each of the parameters of the system a numerical value (among those permitted), then run a numerical simulation of the system of rate equations (1.17). Once obtained the solution, the procedure evaluates if it displays an oscillatory dynamics or not: if it oscillates, the set of chosen parameters, which can be seen as a point in the 8-dimensional phase space,

is assigned to the volume of the phase space corresponding to oscillations of the system. At the end of the procedure, the ratio of points of the phase space belonging to the oscillatory volume to the total number of considered points will provide an esteem of the ratio of the oscillatory volume to the total volume of the phase space, if the sample is ergodic.

Despite the relatively low number of parameters, the systematic exploration of a 8–dimensional parameter space is rather cumbersome, so the point is to develop a procedure for sampling efficiently the phase space: in our work we developed a sampling scheme based on an optimisation principle. The general idea behind our method is to define a function of the trajectory of the system which is minimised as oscillations appear, and to bias the exploration of parameter space with a Metropolis algorithm (41) using this function as an energy. After the sampling is finished, one estimates the volume of parameter space associated with oscillations by subtracting the effect of the bias, much similarly to what one does to obtain the density of states from a thermodynamic sampling (49). The data can be analysed with some advanced tools developed in connection with molecular simulations, like the weighted–histogram method (48). In this way one can quantify the fraction of the space of parameters associated with oscillations, and also estimate the effect of the different parameters in inducing them.

The main advantage of employing the Metropolis algorithm is that it can make the sampling faster and more efficient than if we were doing a random, uniform sampling of the phase space, since with this method we can choose to explore configurations in the phase space with a given probability distribution. In the implementation of the model, we chose of course to explore more frequently the configurations associated with an oscillatory dynamics: this is necessary under the working hypothesis that the fraction of the space of parameters corresponding to an oscillating behaviour is small, and therefore a random sampling of the oscillating regions is not expected to be efficient (it would miss out most of the points, resulting in a severe underestimation of the oscillatory volume). The Metropolis algorithm is precisely what is needed to generate a given arbitrary probability distribution, which in our choice was a Boltzmann-like distribution.

The algorithm we devised starts with a randomly chosen network configuration (i.e. set of parameters of the system (1.17)) and then simulates the dynamics of the system by integrating the ODEs for every node. To determine whether the resulting dynamics is oscillatory or not, we use a technique of symbolic dynamics, consisting in counting the peaks (i.e. the maximum points) of the solution for one variable of the system. The number of the peaks detected in the dynamics is defined to be the "energy" Ω of the system. To explore the oscillatory state with a likelihood given by the Boltzmann distribution, we have to employ a Markov Chain Monte Carlo algorithm with weights $w_i = \mathrm{e}^{\frac{\Omega}{T}}$ being the temperature of the simulation): with this kind of settings the network is pushed to reach the maximum number of peaks available. At every successive step we change the network configuration moving the parameters randomly in a bounded interval. Here we can provide a schematic description of the algorithm:

1. an arbitrary set of parameters is chosen;
2. we move the system, i.e. the parameters are changed multiplying them by a random number;
3. we integrate the system of equations with this new set of parameters, obtaining the dynamics;
4. the "energy" function Ω is evaluated counting the peaks of the solution with a symbolic dynamics algorithm;

5. the new parameter set is confirmed with probability $\min(1, \mathrm{e}^{\frac{\Omega_{new}-\Omega_{old}}{T}})$;

6. the process is repeated from point (2);

In the next subsections we provide details about the implementation of some of these steps of the algorithm.

1.4.1 Solving the system of delay equations.

At each step of our Metropolis-Monte Carlo algorithm we need to solve the system of delay differential equations (1.17). Before solving it it is necessary to define its 8 numerical parameters (and the time delay τ). The 6 rates and the 2 threshold constants (k_{ij}) were allowed to vary in a range of values which we assume to be biologically relevant (42). In particular, we set the constraints (expressing all times in seconds and all concentrations in mol)

- $10^{-10} \leq s_i \leq 10$ mol/s;
- $-10 \leq \alpha_{ij} \leq 10^{-10}$ and $10^{-10} \leq \alpha_{ij} \leq 10$ mol/s;
- $10^{-9} \leq k_{ij} \leq 10$ mol;
- $10^{-6} \leq \gamma_i \leq 10\ \mathrm{s}^{-1}$.

These constraints on the variables define a hyper–parallelepiped in the parameter space, which we shall consider our phase space to be sampled to quantify the fraction of it which is associated with oscillatory dynamics.

Once defined the parameters of the system, we must address the delicate issue of the initial conditions for our system of ordinary delay differential equations (DDE). In the case of DDE we have in fact to define an entire set of initial conditions (since we need the value of the delayed function at the start and at every following integration step), providing the values on the whole interval $[-\tau, 0]$ (if the system has a single time delay τ) in the form of an initial history function $\phi(s) = x(s)$, $s \in [-\tau, 0]$. DDEs thus need *functional initial condition*, and are an example of an *infinite dimensional dynamical system*: this is why they are in general more complicated than simple ODEs, and as such their dynamics is much more rich.

The set of initial conditions $\{x_1(t), x_2(t)\}_{-\tau \leq t \leq 0}$ was set to the constant values $\{x_1^0, x_2^0\}$, obtaining resolving the fixed-point system of algebraic equations (1.15), which defines the steady state of the system (1.17), with a small perturbation ϵ in the source term

$$\begin{aligned} x_1^0 &= \frac{\alpha_{12}}{\gamma_1} \cdot \frac{{x_2^0}^h}{k_{12}^h + {x_2^0}^h} + (1+\epsilon) \cdot \frac{s_1}{\gamma_1} \\ x_2^0 &= \frac{\alpha_{21}}{\gamma_1} \cdot \frac{{x_1^0}^h}{k_{21}^h + {x_1^0}^h} + (1+\epsilon) \cdot \frac{s_2}{\gamma_2}. \end{aligned} \tag{1.18}$$

If either x_1^0 or x_2^0 is negative, it is set to zero. In the present calculation we chose $\epsilon = 0.2$.

The flat initial condition is both computationally tractable and biologically plausible, since it is based on the assumption that the initial condition is due to a perturbation of the steady state (if the system has one), which put the dynamics in motion. We note that

the choice of the initial conditions do not afflict the simulation results, i.e. if the dynamics is oscillating or not. In fact the stability of a fixed point depends only on the parameters of the equation (and on the delay τ) and, in the case of a negative feedback loop, which exhibit only a possible fixed point, if that fixed point is unstable the oscillations necessary occur.

The system of equations (1.17) was solved with the adaptive-step Runge-Kutta method implemented in the gsl libraries (43), with an absolute integration error set to 10^{-4} and the relative error to 10^{-3}; the time length of each simulation is 10^5 s, corresponding in order of magnitude to the length of the cell cycle.

The variables x_i of our system represent the concentration of the molecular species in the cell and is unreasonable for a concentration, which is a normalized count on the molecules number, to be negative. Therefore we add an approximation in the solution of our system: when the x_i value goes below zero we check if the derivative if time derivative $\frac{dx_i}{dt}$ is also negative: if so, the variable x_i is automatically set to zero (and also its time derivative), while if $\frac{dx_i}{dt} > 0$ only the variable x_i is set to zero. We thus interrupt the dynamics and force it to be positive only in the case if the variable is actually expected to go negative.

1.4.2 Detection of oscillations and symbolic dynamics.

The Metropolis algorithm provides a method to sample the phase space according to a given probability distribution. In the case of thermodynamic sampling, the probability distribution is the Boltzmann distribution $p(E) \sim \mathrm{e}^{-\frac{E}{kT}}$, so that the points in the phase space which are selected with more probability are the ones which corresponds to the minimum of the energy E, which are the states most relevant for the statistics and corresponding to physical conformations. In our simulations we would like to select with more probability the states corresponding to oscillations, thus we define as our "energy" $E = -\Omega$, where Ω is a function of our state (i.e. of the set of parameters of the network) whose main requirement is to distinguish between an oscillating state and one which settle on a constant steady state. Our definition $E = -\Omega$ indicates that the function will have its minimum if the dynamics is not oscillatory.

The most simple function Ω which forces the Monte Carlo algorithm to sample with more probability oscillatory states is the one which counts the number of peaks in the dynamics of x_i. The choice of Ω as the number of peaks reflects the obvious fact that an oscillatory dynamics will display an interchanging succession of maxima and minima in the solution x_i, and a solution could be considered to be "more oscillating" if in the time given by the simulation length display an higher number of peaks, with respect to another solution.

Once again, the problem is to find an efficient algorithm to computer the function Ω, since this calculation has to be performed at every step of the Monte Carlo algorithm. Since we are looking for stable periodic solutions, we can calculate the discrete Fourier transform of the solution with a Fast Fourier Transform (FFT) algorithm: the basic idea is that once the FFT is done and we can detect a peak in it (or possibly more than one), the peak reveals that the function is oscillating at the frequency corresponding to the peak. However, this solution is computationally expensive and the results are not so easy to interpret (because the dynamics is generally not purely sinusoidal and thus the FFT exhibits different peaks of various height for the different frequencies of oscillations, and it is difficult to find the right threshold to select just the relevant peaks).

The method we used in the end was simpler and based on *symbolic dynamics*, which is the collective name employed for every mathematical technique that maps a contin-

uous dynamical expression into a discrete space. The mapping consists in associating a symbol to every state the system can produce. The basic idea is to follow the dynamics and every time the system changes state a symbol (correspondent to the new state) is noted. In the end a sequence of symbol is obtained and at this point is possible to study the patterns formed by the symbolic expressions, easier to handle than the continuous case.

We are analysing a continuous expression but we are only interested in the general behaviour of the dynamics: if is oscillating or not. An oscillating dynamic shows an interchanging succession of maxima and minima, and our goal is to choose a suitable labelling that still reproduce this periodic pattern. The procedure, developed by Pigolotti (44), consists in labelling our dynamics with symbols representing the monotonicity of our curve. Since we are considering continuous functions, in the interval between a maximum and a successive minimum the sign of the slope of the function remains constant. The dynamics is then investigated dividing the space in intervals with a defined monotonicity, to which is associated a symbol, typically $+$ if is growing and $-$ if is decreasing. Doing so we passed from a continuous variable to discrete one: an example of this method applied in reconstructing the dynamics of the $kaiA - kaiC1 - kaiB3$ regulation model is showed in the figure 1.2 taken from the work of Pigolotti (45) (see also (16)).

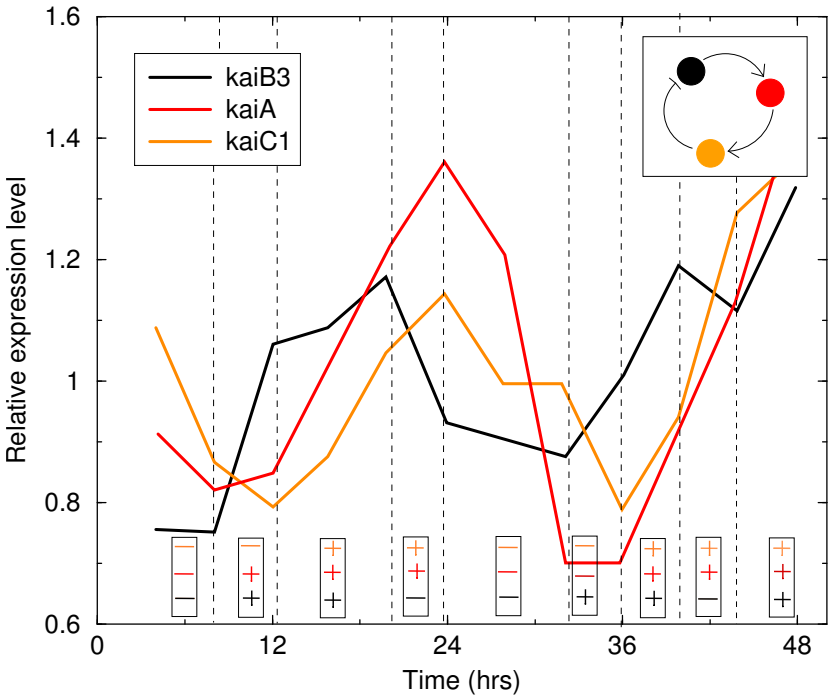

Figure 1.2: Symbolic dynamics algorithm applied to reconstructing the dynamics of circadian rhythms in cyanobacteria from the empirical dynamics of concentration of three protein, from the work of Pigolotti (45). The algorithm predicts that protein $kaiA$ activates $kaiC1$, which represses $kaiB3$, which, in turn, activates $kaiA$. The final sequence of symbols has allowed to know that the dynamics is oscillating and also which variable is a repressor. In fact when the system pass from a $(+,+)$ zone to a $(-,+)$ it means that the activity of the second variable has repressed the first variable.

We note that for a two-dimensional system we can just observe a single symbolic sequence (if one variable in a feedback loop oscillates, so will the other one): the number of maxima in the solution is equal to the number of the times the symbols change sign.

The procedure we have adopted takes place in parallel with the integration of the dynamics. While solving the differential equations, when we print the value x_j of the solution x at the discrete time j (which happens every 5000 numerical time step, since the time step to ensure stability and accuracy of the numerical scheme is quite small), we also calculate the increment $\Delta x_j = x_j - x_{j-1}$ and we compare it to the previous increment Δx_{j-1}. If the increment is positive, we associate the value 1 to that discrete time in a specific array, while if it is negative the value -1. In order to prevent considering increments, both positive and negative, which are not real, but only due to numerical approximations, we consider the increment to be real if $|\Delta x_j| \geq \frac{xmax_j}{c}$, where c is a constant much greater than 1 (in our case $c = 10^5$) and $xmax_j$ the last maximum of the function found at the discrete time j.

It is important to note that the initial dynamics of the system is strongly affected by the chosen initial conditions (while the dynamics for large evolution time, as we stated earlier, is minimally affected), and anyway it requires a certain time to reach stability (in our case, a stable periodic oscillatory regime). Therefore, it is necessary to exclude a certain transient time at the beginning of the dynamics for each simulation. It is obvious to exclude the interval $[0, \tau]$, since this time will be strongly affected by the initial conditions. Concerning the time needed to reach a stable regime, it cannot be precisely quantified, since it depends of many parameters, but we can make a reasonable approximation of its timescale, saying it should be equal to the timescale of the simulation, given by the inverses of the autoregulation parameters γ_i. Thus the cut-off time at the beginning of the dynamics is calculated at every step of our algorithm (i.e. for each change of parameters), and is chosen as the maximum between the time delay and the inverse of the γ_i of the system, so in the case of our two-nodes system the cut-off interval is $\left[0, \max\left(\tau, \frac{1}{\gamma_1}, \frac{1}{\gamma_2}\right)\right]$. Since the terms $\frac{1}{\gamma_i}$ are generally not so negligible, for our choices of "realistic" values of γ_i, with respect to the duration of the simulation ($T = 10^5$ s), there is no risk, in most cases, of including the transient dynamics in the symbolic analysis of the extremal points. Actually, the risk is quite the opposite. To avoid missing the analysis of a large part of the dynamics of the solution, a further condition could be imposed to the cut-off time (for example, not to exceed half the total time duration of the simulation), or we could extend suitably the time duration of the simulation, if the cut-off time happens to be too long.

In the course of simulation we also save the time when the symbols change sign, i.e. the time where maxima are present, so that we can compute compute the time between two neighbouring maxima (which, in case of stable periodic oscillations, would be equal to the double of the period). This set of times could be used to calculate the frequencies of the oscillations in line of principle, taking the average of the distribution, or to distinguish between stable oscillations and other kinds of dynamics. In case of stable oscillations, the distributions of times between two successive maxima will have in fact a low standard deviation. In our successive data analysis, we used as a criterion to distinguish the oscillatory regime from other kinds of dynamics the presence of at least two maxima outside of the cut-off interval defined before. A further question was to distinguish between stable oscillations, whose amplitude (i.e. the difference between a maximum and its successive minimum) remains stable with time, and damped oscillations, whose amplitude goes to zero as time advances (thus the solution effectively settles on a fixed point). For doing so we decided in the course of the dynamics

of each simulation to compute the amplitudes between every maximum of the solution and its successive minimum, and to discard those solutions where the amplitude decreased with time, more than a certain tolerance threshold (set in our case at 1%). This guarantees that the oscillating solution is not damped.

1.4.3 Thermodynamic-like sampling of the space of parameters.

As stated above, our preliminary working hypothesis that the fraction of the space of parameters corresponding to an oscillating behaviour is small, and thus a random sampling of the oscillating regions is not expected to be efficient. This problem is analogous to that of thermodynamic sampling, that is looking for the conformations of a complex physical system which have such a low energy to be statistically relevant according to Boltzmann distribution (where the density of states ρ of the system is distributed, in function of the energy E of the system, as $\rho(E) \sim \mathrm{e}^{-\frac{E}{k_b T}}$, k_b being the Boltzmann constant and T the absolute temperature). In the thermodynamic problem, a purely random sampling is inefficient because the density of states is a non-decreasing function of energy, and consequently a random sampling would visit mainly high-energy states which are irrelevant according to Boltzmann statistics.

To solve this problem, Monte Carlo methods have been introduced (41), which generate a Markov chain that samples the space according to the very probability distribution one wants to recover, usually the Boltzmann distribution. In this way relevant states are visited more often than poorly-relevant ones. In our case we are not interested in sampling the states according to the Boltzmann distribution, but in selecting with more probability the states corresponding to oscillating solutions, and thus our choice of the function to be maximised is the number Ω of stable oscillations of the solution.

Our strategy consists in sampling the parameter space of the model making use of a Markov chain much similar to that used by Monte Carlo methods. Given a set of parameters $\pi \equiv \{s_1, s_2, \alpha_{12}, \alpha_{21}, k_{12}, k_{21}, \gamma_1, \gamma_2\}$ we generated the solution trajectories, as described in 1.4.1, and associated to it the quantity Ω, which counts how many steady oscillations the trajectory displayed (calculated as described in 1.4.2, as the number of maxima of $x_1(t)$ which assume, within a relative error of 1%, the same value of the first relevant maximum).

To favour the sampling of the high-Ω regions of the space of parameters, we started from a random set π chosen within the ranges defined in 1.4.1 and changed randomly the parameters, multiplying them by 10^r, where r is a random number such that $-1 < r < 1$ (uniformly distributed), with a probability of 0.8 ("small move"), or multiplying them by 10^r, where r is a random number such that $-2 < r < 2$, with a probability of 0.2 ("big move"). In any case the parameters are moved globally, i.e. each of them is multiplied for a random number in the described ranges (not the same for each parameter, of course). The possibilities of a single move (moving only one parameter in the set for each Monte Carlo step) or of an addictive move (adding a random number, rather than multiplying by one) have been considered but then rejected, because of the amplitude of the phase space: the multiplicative change of the parameters allows the system to explore efficiently all the orders of magnitude in the space of parameters (i.e., effectively exploring the logarithm of the rates).

In the case of α_{ij} , when the random move brings it above -10^{-12} or below 10^{-12}, the sign is changed: in this way the algorithm could sample continuously the two semi–regions in which the α_{ij} are allowed to move.

After moving the parameters, solving the system and calculating Ω, the move was

then accepted with a rate given by the Metropolis-like expression

$$w(\pi \to \pi') = \min\Big[1, \exp\Big(\frac{\Omega(\pi) - \Omega(\pi')}{T}\Big)\Big] \tag{1.19}$$

Here the parameter T is analogous to the temperature in a standard Metropolis simulation, and controls the bias towards high values of Ω (essentially, the higher the temperature, the more the Metropolis algorithm resembles a random sampling). This procedure is essentially equivalent to a Metropolis simulation controlled by the energy $-\Omega$, where the minus causes the algorithm to maximise Ω instead of minimising it, as usually done with the energy.

Given Eq. (1.19), one can then use all the armoury of Monte Carlo algorithms to improve the sampling of the space of parameters and to analyse the data. For example, one can perform a parallel tempering sampling, in which different replica are sampled simultaneously at different temperatures, to help the system to jump through disconnected regions of non-zero values of Ω (46). Actually, one can use all the advanced sampling techniques which were developed for enhancing the exploration of molecular conformations, such as simulated tempering, umbrella sampling, metadynamics, etc. (47).

Simulations of the kind described here were conducted for different choices of the delay ($\tau = 500, 1000, 1500, 2000, 2500$ s), using each time three temperature parameters ($T = 1, 2, 3$), with each single Monte Carlo simulation running for 150000 steps. The simulations are computationally quite demanding, and took approximately 6 months of CPU time on an Intel i7 3.4Ghz processor.

In Fig. 1.3, in the upper panel, it is shown an example of an oscillating solution of the system, found in a step of the Monte Carlo simulation, displaying a number of maxima $\Omega = 54$. It is evident from the picture the presence of a short transient time at the beginning, after which the dynamics settles on a stable oscillatory regime. In the lower panel we can see an example of a whole Monte Carlo sampling of the space of parameters, where for each Monte Carlo step of the simulation is plotted the corresponding value of Ω for the solution of the corresponding system.

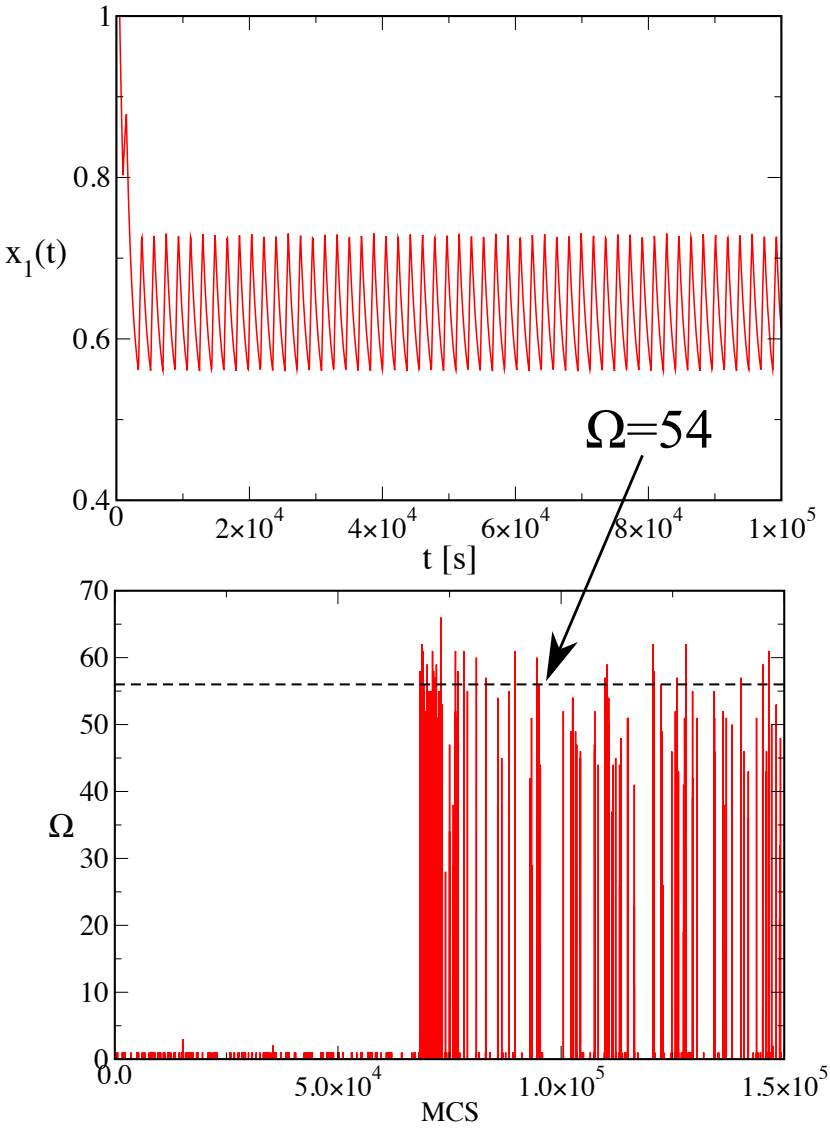

Figure 1.3: (Upper panel) An example of oscillating solution $x_1(t)$ of the feedback loop, displaying $\Omega = 54$. (Lower panel) An example of Monte Carlo sampling of the space of parameters, where the system first visits a region corresponding to stationary solutions ($\Omega \leq 1$) and later finds oscillating solutions.

1.5 How likely are oscillations?

Once all the samplings reached convergence, to quantify the density of oscillating state we employed a technique once again borrowed from the field of thermodynamic sampling, the *weighted-histogram* method (48). We note in fact that we have, for each value of the time delay τ, multiple Monte Carlo sampling of the function Ω (the corresponding of the energy for thermodynamic sampling) over the phase space: these multiple

samplings are conducted with different biases (i.e. the temperature T of the Metropolis algorithm) in order to make the sampling more efficient and provide more statistics for the esteem. The Metropolis bias was introduced to make the sampling more effective that a pure random sampling, and this bias in the sampling must be removed if we want to calculate an estimation of the real fraction of oscillating volume. Thus a technique is needed to provide an esteem of the fraction of oscillating volume that is independent on the bias used and that uses all the data from the samplings conducted at different biases: the weighted-histogram method provides exactly such a tool.

The fraction of the space of parameters associated with oscillations is the density

$$g(\Omega) = \int d^8\pi \, \delta(\Omega(\pi) - \Omega), \tag{1.20}$$

which is analogous to the density of states of thermodynamic systems. A multiple–histogram algorithm extracts the density of states $g(E)$ from the histograms of sampled energy (the outputs of our simulations) (49).

From the histograms $N_T(\Omega)$ collected from the simulations at different temperatures, one can obtain (maximising the likelihood) the most likely form of $g(\Omega)$ compatible with them, which is the following (48)

$$g(\Omega) = \frac{\sum_T N_T(\Omega)}{\sum_T N_T \exp\left(\frac{\Omega}{T}\right)/Z_T}, \tag{1.21}$$

where $N_T = \sum_T N_T(\Omega)$, and Z_T is obtained solving the set of implicit equations

$$\sum_\Omega \frac{\sum_{T'} N_{T'}(\Omega)}{\sum_{T'} N_{T'} \exp\left(\frac{\Omega}{T'} - \frac{\Omega}{T}\right) \cdot Z_T/Z_{T'}} = 1. \tag{1.22}$$

In this way, the function $g(\Omega)$ is proportional to the volume of the parameter space characterised by a given oscillation number Ω, independent on the bias used to make the sampling efficient.

The functions $g(\Omega)$ resulting from the simulations for different delays are displayed in Fig 1.4 in semi-logarithmic scale.

The points at $\Omega = 0$ and 1 indicate sets of parameters not associated with oscillations. At all values of the delay τ, $g(\Omega)$ displays an overall decreasing trend, compatibly with the idea that the stationary behaviour is the most common outcome, while oscillations are somewhat disfavoured. However, the strongest decrease occurs at $\Omega \lesssim 10$, corresponding to a oscillation frequency of less that $10^{-4}\,\mathrm{s}^{-1}$, while the decrease is only marginal at larger values of Ω. From the simulations it is difficult to assess if the wavy behaviour at large Ω is a real physical effect or noise associated with the finite statistics.

The volume fraction of the space of parameters associated with oscillating solutions of Eq. (1.17), defined by

$$f = \frac{\sum_{\Omega>1} g(\Omega)}{\sum_{\Omega\geq 0} g(\Omega)} \tag{1.23}$$

is displayed in Fig. 1.5 as a function of the delay τ . The volume fraction displays a spike at small delays ($\tau = 500$ s), reaching a fraction of $f = 0.027$, and then converges to an almost constant value of $f = 0.004$. Perturbing the definition of f, for example defining as oscillating also the solutions displaying $\Omega = 1$ does not change the qualitative behaviour of this plot, suggesting that it is quite robust.

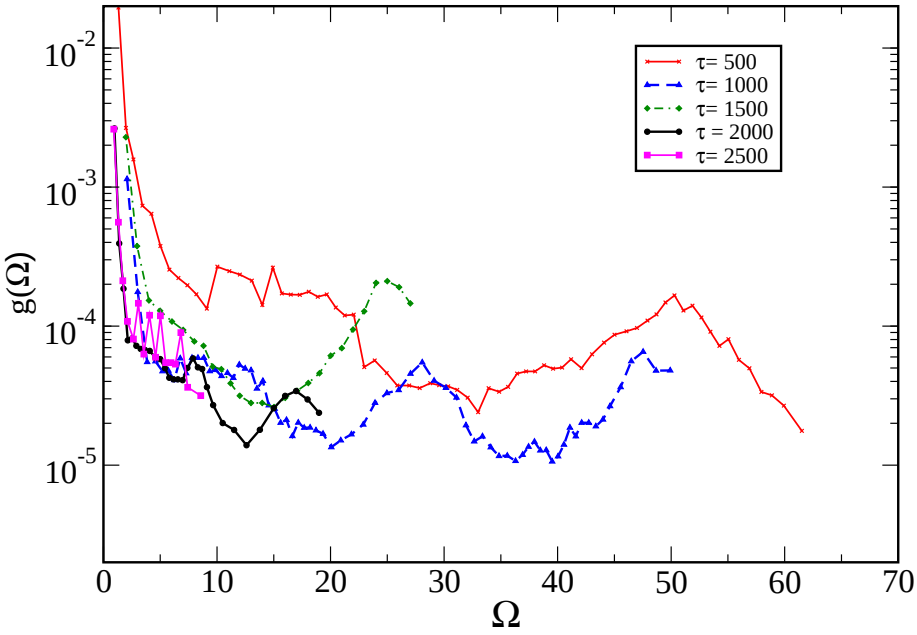

Figure 1.4: The density $g(\Omega)$ calculated for different values of the delay τ.

Once provided an esteem of the fraction f of the oscillating volume, what is left is to investigate the effect of the 8 parameters of the model in inducing oscillations. In Fig. 1.6 the value of Ω sampled at the different values of the parameters of the rate equations are plotted for the Monte Carlo simulation generated at $\tau = 500$ s. The simulations with the other delays are qualitatively similar (data shown in App. A, Figs. A.1 - A.5). Since the system of rate equations is symmetric, we collected in the same plots the rates corresponding to the two species.

It must be noted that what Fig. 1.6 only consider the dependence of Ω from the single parameters separately, and so it is a very coarse analysis of the possible patterns in the combination of the parameters that induce an oscillatory dynamics. However, some useful information could anyway be drawn.

The dependence of oscillations on the mutual couplings α_{ij} $(i = 1, 2)$ is bimodal, displaying large values of Ω at small (in absolute value) negative and at large positive values of α_{ij} $(i = 1, 2)$. To further investigate this point, we plotted in Fig. 1.7 the dependence of Ω on the product $\alpha_{12} \cdot \alpha_{21}$. Oscillations ($\Omega > 1$) appears only if the product is negative and small: in other words, the equations must describe a negative feedback loop, in which one species activate the other, which, in turn, inhibits the former. Moreover, in the inset of Fig. 1.7 it is shown the maximum value of Ω as a function of the larger between α_{12} and α_{21}. The curve is approximately straight, indicating that

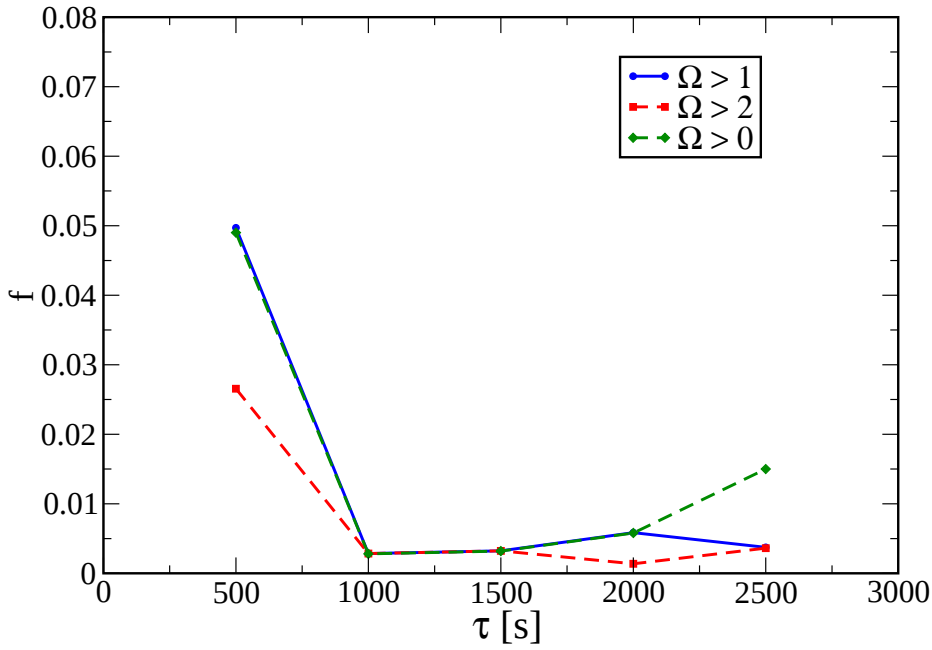

Figure 1.5: The fraction f of space of parameters associated with fluctuating trajectories, defined as $\Omega > 1$, as a function of the delay τ (circles). To test the robustness of the result, we perturbed the definition of fluctuating trajectories, calculating the behaviour of f at $\Omega > 2$ (squares) and $\Omega > 0$ (diamonds).

to obtain oscillations with high frequency, the larger α_{ij} must be large and that such a relation is linear.

As a function of the equilibrium constants k_{ij}, the oscillation parameter Ω is rather variable, filling densely the scatter plot except for a region between 10^{-7} and 10^{-5} mol. Although one cannot rule out that the lack of points in this region is due to a bad sampling, the otherwise dense appearance of the plot, and the straight boundary between the populated and non–populated region may suggest the this region of parameter space is actually depleted of oscillating trajectories.

The scatter plot of Ω as a function of the degradation rates γ_i gathers all points with $\Omega > 1$ at $\gamma_i > 10^{-1}\text{s}^{-1}$. Note that the values of γ_i determine the time scale of the dynamics of $x_1(t)$ and, in absence of fixed ranges of variability of all the parameters, its value would be irrelevant, setting the time units of all the rates. The accumulation of the points with $\Omega > 1$ at the higher border of γ_i indicates that if it were allowed to be increased beyond this border, the system could find new oscillating regions by decreasing the other rates; this would anyway results in rates assuming unrealistic values when compared with the underlying molecular processes.

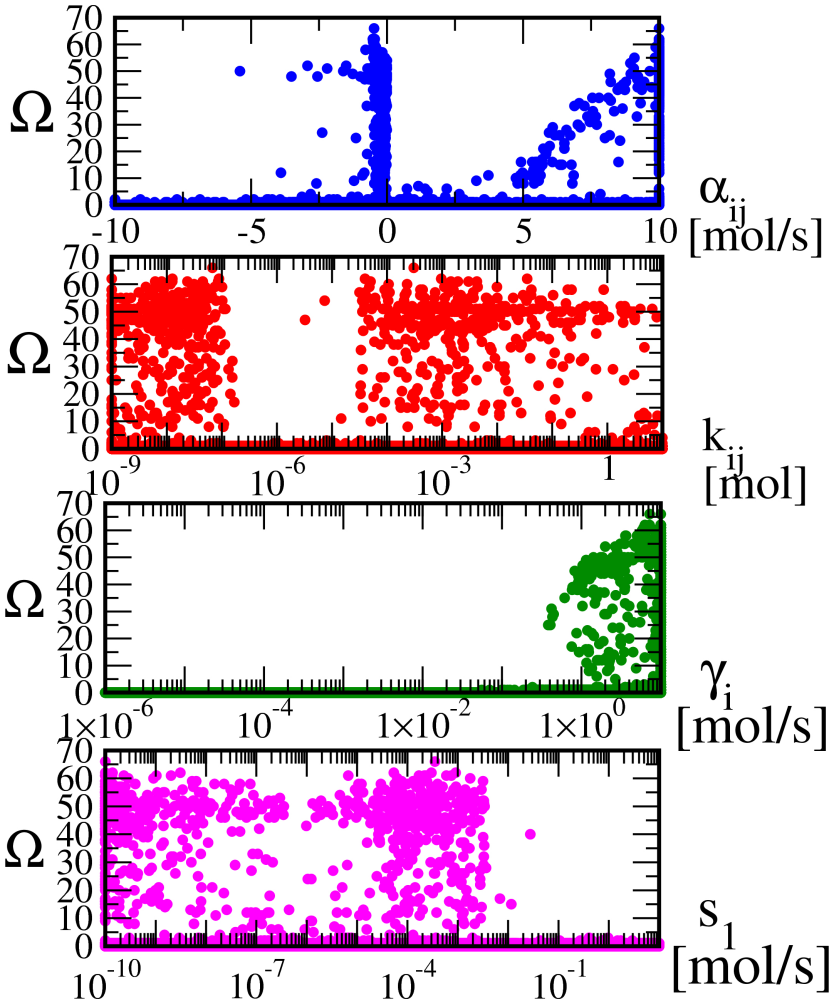

Figure 1.6: The values of the number of oscillations Ω associated with the points of parameter space sampled by the simulation carried out at $\tau = 500$ s and $T = 1$ projected, respectively, onto the parameters α_{ij} , k_{ij} , γ_i and s_i . To make the results more readable, α_{ij} is plotted in a linear scale, while the other parameters in logarithmic scale. The density of points in each region of the plots reflects the probability that the delay equations display a specific value of Ω for the value of the parameters that define that region.

The dependence of Ω on the constant sources s_i is straightforward, the system displaying oscillations only when both s_i are negligible. When this is not the case, the two equations are essentially decoupled, and the process becomes approximately a birth-death process, converging to a stationary solution.

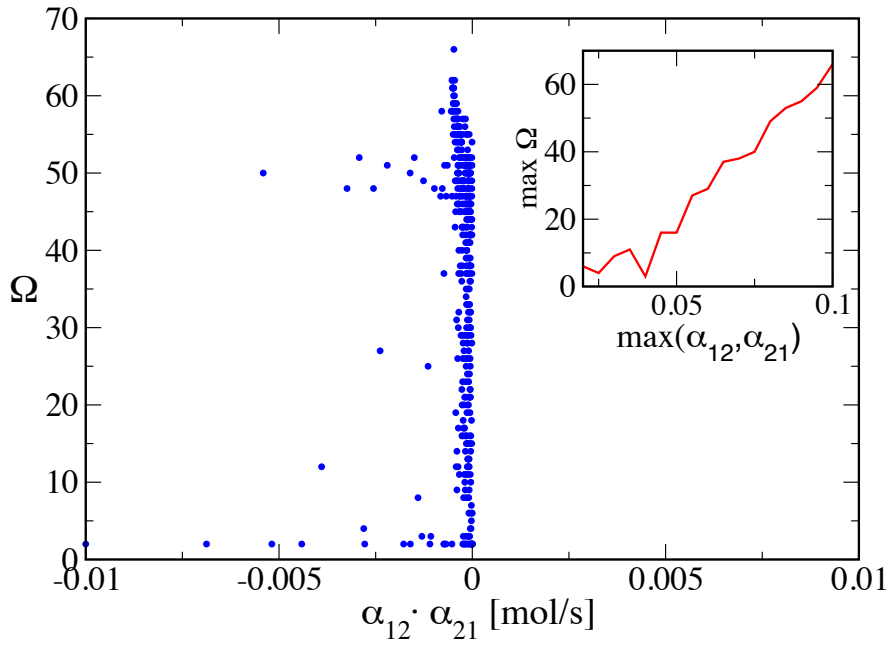

Figure 1.7: The values of Ω sampled by the simulation carried out at $\tau = 500$ s and $T = 1$ as a function of the product between the α_{ij} of the two species. In the inset, the maximum value of Ω as a function of the larger between α_{12} and α_{21}.

1.6 Discussion and conclusions.

Oscillations in the concentrations of proteins are a response of the cell to external stimuli that has been observed in several cellular contexts (50; 51; 52). Even one of the simplest system displaying oscillations, that is a two-node delayed feedback loop, can be difficult to study because one has little control over the numerical values of the parameters which control the model. For most systems, only a qualitative accounting of the relation between genes is available (53), sometimes with the specification that the relation is excitatory or inhibitory, but the associated numbers are unknown (54). When these are available, they are often extracted from biochemical or model experiments, and not directly related to the in vivo situation. Consequently, one is compelled to study the model by varying its parameters in a realistic, usually wide, range of values. However, this procedure is not straightforward even in the simple case of a two-node system, because this is defined by eight numerical parameters, and their exhaustive scan is infeasible.

To face this problem, we have applied the sampling techniques and the analysis strategies originally developed to study molecular systems in the canonical ensemble. Analogously to the sampling of the space of parameters of out rate equations, molecular systems typically display a very high dimensionality, often reaching $10^5 - 10^6$ degrees of freedom. This suggests that these techniques can be effective also to study larger regulatory networks, controlled by many more parameters. In the case of molecular systems, sampling techniques build a Markov chain in conformational space to exploit energy correlations and thus find the low-energy conformations, which are the most rare but at

the same time the most significant according to Boltzmann distribution. Similarly, we generated Markov chains in the space of the parameters of the delay equations, biasing the search for highly-oscillating solutions through a temperature-like parameter. Once we explored the available parameter space, using multiple temperatures to fasten the exploration, we removed the bias to obtain the actual density of parameters associated with oscillations at various frequencies.

The result is that oscillations occupy a small, but not negligible volume (fraction of $\sim 1\%$) of the entire parameter space, as is can be seen, for example, in Fig. 1.8. This fraction is quite independent on the delay, and increases only at small delay values (cf. Fig. 1.5). Oscillations appear only for negative feedback loops, within a rather well-defined range of the parameter which couples the two species, and only if their spontaneous production rate is negligible. All these data point to a scenario in which evolution had to search actively for oscillations. Although we already know that α_{12} and α_{21} must have opposite sign for the system to display oscillations (16), we let the system explore the whole parameter space to prove the generality of the approach and to constrain the sampling as little as possible; the constrain on the sign of this parameter eventually emerges as a result.

The projections of the oscillating regions of the space of parameters along the direction of the different rates are usually connected. Different seems to be the case along the direction of the equilibrium constants k_{ij} , which displays a gap spanning more than two orders of magnitude. This means that once evolution tunes the rates to produce oscillations, it is quite easy to adjust them, for example, to change their frequency. On the other hand, the model predicts an evolutionary barrier on the equilibrium constant which defines the interaction properties between the two species.

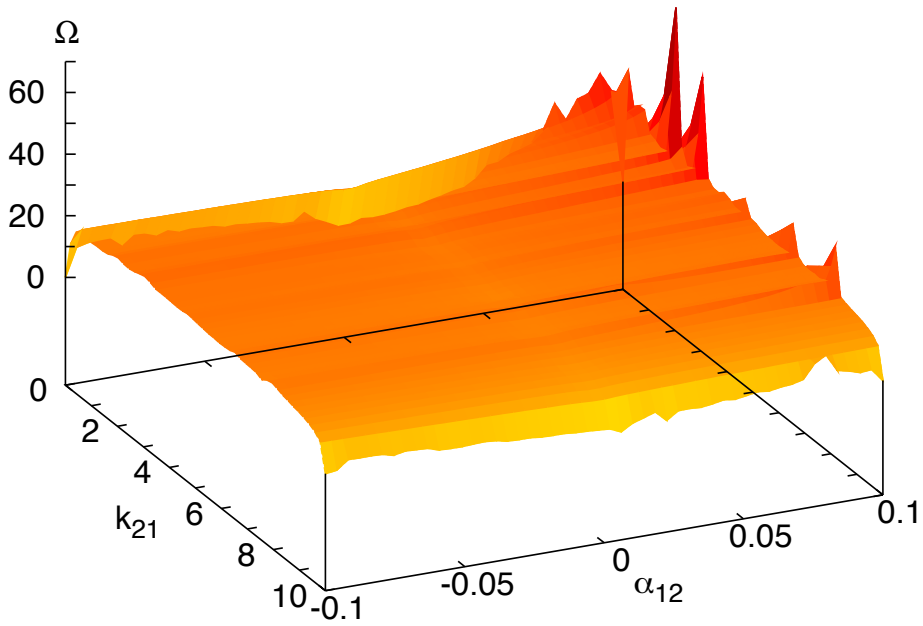

Figure 1.8: The values of Ω sampled as a function of α_{12} and k_{21} suggest that the oscillating region is a small but not negligible part of the parameter space.

CHAPTER 2

Control of phenotypic switching in cancer cells

2.1 The cancer stem cell hypothesis.

Cancer is a group of about 200 known diseases involving abnormal cell growth with the potential to invade and spread to other parts of the body: it is fundamentally a disease associated with tissue growth regulation. The incidence of cancer and cancer types are influenced by many factors such as age, gender, race, local environmental factors, diet, and genetics (55). It is a leading cause of death worldwide: it accounted for 8.2 million deaths (around 22% of all deaths not related to communicable diseases, according to the World Health Organization (WHO) (56).

The commonly accepted model for carcinogenesis, as developed by Nowell and Vogelstein (57; 58), is the so-called "stochastic model" of carcinogenesis. This model considers cancer as a purely genetic disorder, which develops by sequential accumulation of mutations in oncogenes and tumour suppressor genes. In this view, cancer cells continue to evolve by acquiring new mutations, thus leading to an heterogeneous population of cells competing for resources in a highly dynamic process. Note that, in this model, any cell may be the target of carcinogenesis.

Several other models for the formation of tumours have been proposed over the years: in particular, the hypothesis that cancer might arise from a rare population of cells with stem cell properties was proposed about 150 years ago (59; 60; 61), but it was only recently, owing to advances in experimental tools and studies, that this hypothesis has gained attention in the scientific community.

Stem cells are characterized by two specific features: the ability to differentiate into all range of specialized cell types and the ability to renew themselves [1]. According to this analogy, therefore, properties of the supposed "cancer stem cells" (CSC) include unlimited self-renewal, which drives tumorigenesis, and differentiation, albeit aberrant, into "normal" cancer cells (CC), that contributes to cellular heterogeneity.

The first evidence for the existence of CSCs came from acute myeloid leukaemia (62; 63), a tumour of the blood, in which a rare subset of cells, comprising 0.01–1% of the total population, could induce leukaemia when transplanted into immunodeficient mice. This first suggested, on the basis of experimental findings, that only a small subset of the total tumour population has the capacity for self-renewal, and thus could actively drive tumour growth and recurrence. Thus initially, cancer stem cells could be determined only operationally by their ability to seed new tumours in animal models, and, for this reason, they have also been termed ”tumour-initiating cells” (64; 65). More

[1] There are two distinct types of mammalian stem cells: embryonic stem cells that are totipotent, e.g. having the possibility to differentiate into all specialized tissues of the developing organism, and the adult stem cells, which are multipotent, meaning the possibility to substitute specialized cells of the corresponding tissues, thus maintaining the normal turnover of it.

recently, for some cancer cell lines specific cancer stem cell biomarkers were reported, which allow biochemical determination of this population and its further analysis (65).

Populations of cancer cells that match the described criteria for CSCs, that is, self-renewal and the ability to give rise to a variety of differentiated cells, have since been identified in a diversity of human malignant solid tumours (65). The first solid tumour type from which a CSC population was isolated using cell surface markers was breast cancer. In 2003, Al-Hajj et al. (66) showed that one hundred $CD44^{+}\ CD24^{low/-}\ ESA^{+}$ cells from breast carcinomas were able to grow a differentiated mammary carcinoma in SCID mice. This is in contrast to the $CD44^{-}$ and $CD24^{+}$ tumour cells that were not able to grow a tumour even when 10^5 cells were injected. Currently, populations with CSC characteristics have been identified in brain tumours, including medulloblastoma and glioblastoma, head and neck squamous cell, colon, prostate, lung, pancreas, ovarian and hepatic carcinoma (65). The CSCs in solid tumours are directly selected after dissociation based on the expression of cell surface proteins frequently associated with a primitive, non-mature cell type. These protein biomarkers can in some cases be associated with stem cells or progenitor cells in the specific tissue from which the cancer arose, like $CD133$ in medulloblastomas (67) , while in other instances the marker protein used to enrich the CSC population was not associated with the stem cells in the normal tissue (69).

The picture that seems to emerge is that a small subpopulation of CSC have an unlimited capacity to proliferate and gives birth by division to the population of differentiated cancer cells (CC), which can divide for a certain number of generations and then die (see ref. (63)). Thus, according to this "aberrant hierarchy theory" (70; 65), the dynamics of tumour growth is driven ultimately by the tiny subpopulation of CSC, in analogy with the dynamics of normal tissues. It is evident that his model represents a paradigm shift in our thinking and has fundamental consequences for understanding the biology of carcinogenesis as well as important clinical implications for early detection, prevention, and therapy of human malignancies (65). One of the most important clinical implications, obtained in (68) by a simple ordinary differential equation model (considering normal and abnormal stem and differentiated cell populations of the hematopoietic system), showed that successful therapy must eradicate the CSCs, while any therapy that targets mature CCs or partially induces differentiation of stem cells is unable to provide tumour control. In this direction, the design of therapies that selectively target the subpopulation of CSC in a tumour and are able to effectively stop the dynamics of the whole tumour without the side effects provided by chemotherapy would be a major achievement.

Despite recent experiments in vivo have confirmed the presence of an aggressive CSC-like subpopulation in benign and malignant tumours [2] (71; 72; 73)), and despite cancer stem cell research has provided us with important new insights for future oncological research, definitive proof of the CSC model is still lacking in most malignancies (65; 74). In particular, there is lack on consensus even about the proportion of CSC in the whole mass of the tumour. In many cases it has been assumed that CSCs necessarily constitute only a small minority of the cells in a cancer, at most few percents of the total (76; 69; 75). However, more recent work on melanomas (77) has shown that, while earlier techniques could identify only one in $8.37 \cdot 10^5$ cells in a melanoma as being cancer stem cells,newer assays suggested that as many as one in four cells from melanomas could form tumours in more stringently immuno-deprived mice and observed over a longer period of time. Previous studies on mouse leukaemias have also shown that at

[2]Experiments confirming the presence of CSC in vivo are of particular importance, since animal tumour xenografts in which CSCs were first identified miss the physiological environment in which the tumour grows.

least some malignancies may be maintained by more than 10% of the cells in a tumour (78). In (79) it is shown that the proportion of cancer stem cells that can be identified in colorectal cancer derived cell lines can vary quite widely, from perhaps 20% of the total to virtually all cells. It is therefore possible that the theory is faulted, or that different individual cancers may contain widely differing proportions of CSCs.

About this, it has been shown recently (80), through computational models, corroborated by in vitro studies, that tumours modelled in a CSC context more faithfully resemble human malignancies and show invasive behaviour, whereas tumours without a CSC hierarchy do not. The same model also predicts that CSC paradigm allows for significantly higher tumour heterogeneity, which may affect therapy resistance to the point that therapy which fails to target the CSC population is not only unsuccessful in curing the patient, but also promotes malignant features in the recurring tumour. Another mathematical model, presented in (75), and supported by experimental evidence, argues that CSCs may comprise any possible proportion of the tumour, and that the higher the proportion the more aggressive the tumour is likely to be.

It is evident from the examples reported that, at our present level of knowledge, the question of whether most tumours present a hierarchical organization in their cellular population, as that described by the theory of the cancer stem cells and their "aberrant hierarchy", is a complex one. While it is obvious that the subdivision in compartments of the cells of the tumours is necessarily schematic (for example, phenotypic heterogeneity within CSC subpopulations is likely to exist (81)), we may ask whether the hierarchical model with the CSC subpopulation is too simple and some important biological mechanism (at the macroscopic, cellular level) is missed out in this picture. In the next section we shall report some recent important experimental results which can suggest a new way to integrate the CSC model; before doing so, in the next subsection we collect the essential known facts about cellular kinetics in the CSC model, which shall be our standard reference throughout our work.

2.1.1 Cellular kinetics in the CSC hypothesis and related computational models.

According to the model of the aberrant hierarchy with a CSC subpopulation, the total cell population of a tumour can be divided in two compartment: cancer stem cells (CSC), a subpopulation of cells with an unlimited capacity of self-renewal and also able to differentiate by division, and the population of "normal" differentiated cancer cells (CC), which are produced by CSC as they differentiate and have a limited capacity of division, just like normal cell in the body. We could also consider the compartment of dead, or quiescent cells (DC), which is the state that CC reach after exhaust their proliferative potential, having undergone the maximum number of available cell divisions.

The different kinds of cells in the tumour undergo different biological processes. CSC cells can undergo *symmetric division* ((82; 83; 84)), whereby one CSC originates two CSCs

$$CSC \rightarrow CSC + CSC. \tag{2.1}$$

CSC can also undergo symmetric division producing two CCs (85; 86) (which is a process of cell differentiation)

$$CSC \rightarrow CC + CC. \tag{2.2}$$

Alternatively, CSC can undergo *asymmetric division* (118; 87), producing one CSC and one CC

$$CSC \rightarrow CSC + CC. \tag{2.3}$$

This is done in analogy with normal stem cells, which generally undergo asymmetric division in order to preserve their numbers constant, and thus maintain general tissue homeostasis. However, cancer stem cells are generally believed to undergo asymmetric divisions with relatively slow rates (88).

Concerning CCs, they can undergo *symmetric division* (90), giving birth to two CCs

$$CC \rightarrow CC + CC. \tag{2.4}$$

This cannot be done for a virtually unlimited number of times, as in the case of CSC, but only for a limited number of times. CC are subjected to senescence as they divide (89), and after they reach the maximum number of divisions, cells go quiescent, losing their proliferative potential, and die

$$CC \rightarrow DC. \tag{2.5}$$

In Fig. 2.1 we summarize the whole theory of the aberrant hierarchy with CSC, collecting the different kinds of cells and the processes they can undergo

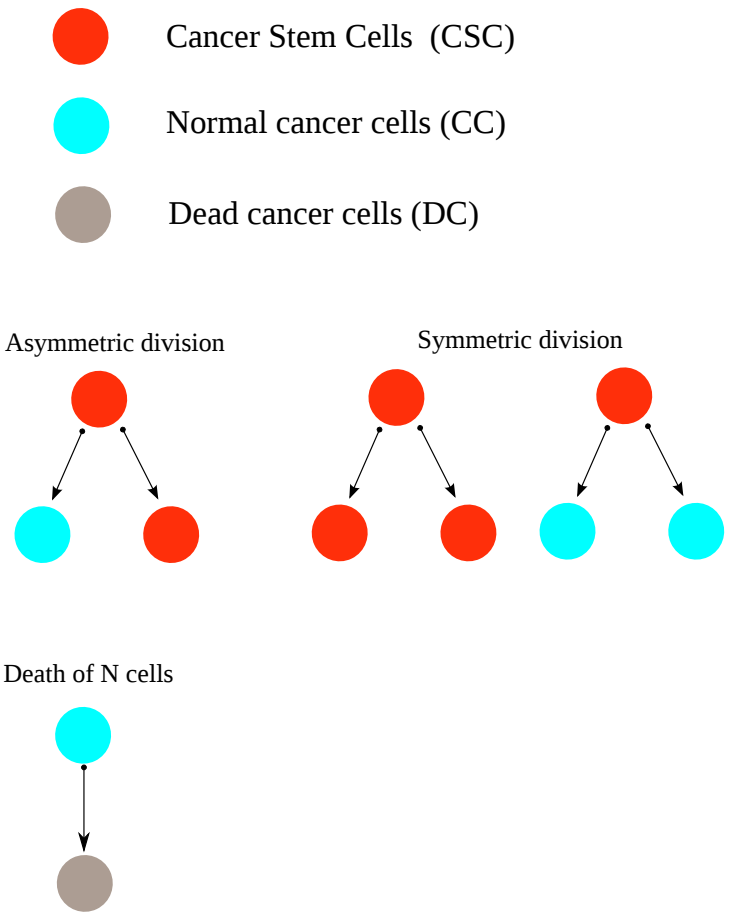

Figure 2.1: The theory of the aberrant hierarchy with CSC: cells and processes.

One factor responsible for establishing the CSC fraction within a tumour is the relative frequency with which CSCs either create another CSC (by symmetric division) or a non-CSCs (by asymmetric division) (91) (92). Mechanisms known to directly affect the symmetric division probability, in turn, include availability of certain host growth factors such as EGF, and growth-factor-rich niches, which can skew division modes in favour of symmetric production of CSC up to 85% (93).

The quantification of the fraction of CSC in a tumour is a matter hotly debated (75; 81). In fact, the biology of CSCs has not been fully elucidated and many questions still remain unresolved, most of which related to the dynamics of tumour growth (94; 81). However, both experimental and computational modelling results give strong evidence to the following facts, that we shall take as working hypotheses in the rest of our work:

- CSCs are generally a small subpopulation within tumours. The fraction of CSC in a tumour can go, according to the most realistic estimates, from fractions of percent to about 5%. In addition to identification of CSCs through surface proteins in vitro and in vivo mouse xenograft transplantation assays, novel approaches emerge that trace tumour hierarchy and help estimate CSC kinetics and frequency in tumours. One approach is to monitor the division kinetics of stem and progenitor cells in normal epithelial tissues, skin papilloma, and invasive squamous cell carcinoma during unperturbed growth emerged from clonal analysis using genetic lineage tracing in mice (91; 81). In (95) an integrated experimental and cellular Potts model approach to simulate glioblastoma population growth and response to irradiation, which identified the (a)symmetric division kinetics of glioblastoma stem cells necessary to reproduce the observed ratio of 2-3% of such cells. Another integrated approach (96) of single-molecule genomic data, spatial agent-based modelling, and statistical inference was recently introduced to derive tumour ancestral trees in patient-specific colorectal cancer samples that lead to the observation of a CSC fraction of 0.5-4%;

- the percentage of cancer stem cell population is maintained at the same level during several years of passages, as results from works on several cancer cell lines (88). Another important observation is that when a cancer stem cell population is isolated, purified and maintained in culture, the percentage of stem cells rapidly decreased and finally stabilized at the initial level, characteristic for this given cell culture (88). Thus there seems to be a homeostatic control mechanicsm that aims at keeping the percentage of CSC constant in the tissue. One of the hypotheses about the nature of this tissue regulation mechanism involves the existence of ”instructive signals” for cell-cell communication, in the form of an underlying field, that is a diffusive molecular species (88).

Our last observations are about the translation of information about the hierarchical model as summarized in Fig. 2.1 into mathematical and computational models which have proved extremely useful in testing hypotheses about cellular kinetics and investigating the dynamics of tumour growth. Rate equations models, as systems of ordinary differential equations, are particularly suited to simulate compartment models like our hierarchical model, which simplify the cell heterogeneity found in cancer considering only few specific cell subpopulations. In this models, events related to CSC self-renewal, to differentiation of CSCs into CC cells, to ageing of CC cells, and to death of all cell subtypes, are represented as "chemical reactions" and are mediated by specific rate constants. These reactions occur in a system that has no nutrient or spatial limitations, typically during the phase of exponential tumour growth. Such kind of models have

been employed for investigating how different values of the proliferation rates of different cells affect the steady-state fraction of CSC in the tumour (82), the role of various feedbacks and control mechanisms in the general dynamics of cell populations (97), the effects of the combination of different therapies in the evolution of a tumour (98), and the estimation of cell proliferation potential (99). When it is necessary to incorporate the inherent stochasticity of biological phenomena, especially in the case of small numbers of cells, deterministic models such as the rate equation fail, and stochastic models are needed, for example in the mathematical form of master equations or computationally as stochastic simulations with the Gillespie algorithm (94). Simulation at a further level of complexity, which include the role of the spatial dimension, can be performed with the introduction of *agent-based cellular automaton models*, where single cells are considered to be individual objects whose behaviour is determined by a fixed set of basic intrinsic cell kinetics and their local environment. Cell are located in some point of the space, which in many cases is a discrete regular lattice, bidimensional or tridimensional. At each discrete simulation time point, the cell behaviour for all cells is updated and, over time, complex population dynamics emerge. This approach is particularly useful for studying carcinogenesis, natural selection, genetic instability and interactions of individual cells with each other and the microenvironment, and its limitations in describing very large systems can be overcome by carefully designing hybrid continuous-discrete and multiscale model (100), (101). Agent based model have been employed, for example, to study tissue homeostasis in the context of CSC (100), to investigate tumour morphology at the variation of the fraction of CSC (80), the paradoxical dependencies of tumour dormancy and progression on cellular proliferation and death rates (102), and the role of spatial distribution of stem cells in determining tumour response to radiotherapy (103).

2.2 Phenotypic switching in the CSC context: theory and experiments.

We have seen in the previous section that the variations in the observed fractions of CSC in tumours pose some challenge to the CSC hypothesis. One possibility is that the model, as described in 2.1.1 is incomplete, and that some key cellular mechanism is missing, and that the population dynamics of CSCs is more complex than the strict hierarchy originally proposed.

In particular, some recent experimental findings points in the direction that CC can revert to a stem-cell-like state even in the absence of mutations, undergoing a process what has been called *phenotypic switching*. The first experimental evidence of transition from CC to CSC state has been observed in human breast cancer cell lines (104): in this work it has been shown that subpopulations of cells purified for a given phenotypic state (i.e. sorted and collected with proportions of the different subpopulations different from the equilibrium ones) return towards equilibrium proportions over time, and this could be explained by a simple Markov model in which cells transition stochastically between states. From this model it can be inferred that, given certain conditions, any subpopulation of cells will return to equilibrium phenotypic proportions over time (in accordance with the working hypothesis of the previous section) and, crucially, that breast cancer stem-like cells arise de novo from non-stem-like cells. Similarly, in melanoma a small population of CSC-like JARID1B positive cells has been shown to be dynamically regulated in a way that differs from the standard hierarchical CSC model (105), in accordance with earlier findings ((106; 106)).

The biological factors regulating the process of dedifferentiation of non-CSCs into CSCs are still unclear: phenotypic switching may arise due to stochastically acquired genetic or epigenetic changes in genes governing the CSC state (107): for instance, it has been found that overexpression of BMI1 (polycomb ring finger oncogene) in transformed epithelial cells can promote their conversion into a CSC-like state with increasing tumorigenicity and metastatic potential (108). Microenvironmental factors, such as TGFβ, are also found to enhance the rate of switching from non-CSC cells to the CSC state (109). This is in accordance with results showing that ABCG2 negative cells isolated from human melanoma biopsies re-express this marker after a few generations in vitro (110; 71).

Despite increasing experimental evidence supports the hypothesis of switching from CC to CSC (71), the question is still debated, and some papers do not support this hypothesis: for example, in (111) it is shown that in melanoma, ABCB5- cells are not able to generate ABCB51 cells, while in (112) CD341Cd271/Ngfr/p75- cells formed tumours CD271- restricted, whereas CD34CD271/Ngfr/p75- cells formed tumours containing both CD2711 and CD271- cells [3]. The identification of cancer stem cells in vivo and in vitro still relies mainly on detecting specific surface markers, characteristic for the subpopulation of cells, and thus the hypothesis of phenotypic switching mainly originated from experiments reporting that sorted cancer cell populations after some time tend to express again all the original markers (like in (104)): mathematical modelling for cancer cell proliferation has been employed (113) to test with statistical significance if the reappearance of original marker could be explained only by phenotypic switching, or alternatively also by imperfect markers and error in the sorting process of cellular populations. The conclusion was that the observation of reversible expression of surface markers after sorting does not provide sufficient evidence in support of phenotypic switching.

[3] the strings of letters are the name of the marker for CSCs, and the sign $-$ indicates that they are not expressing the markers, i.e. that they are negative cells.

Advances in proving or disproving the phenotypic switching hypothesis could be of decisive importance for laying the aberrant hierarchy theory on more solid ground (with a possible modification to account for a certain degree of plasticity between the non-CSC and CSC compartments, as shown in Fig. 2.2), and even more so for its clinical implications, in the planning of appropriate and efficacious treatment strategies. In fact, under the hypothesis of phenotypic switching, therapies that target primarily CSCs may ultimately be unsuccessful if non-CSCs are able to dedifferentiate into a CSC state. That would make the CSC population extremely difficult to eradicate completely (even not considering the fact that, as several recent studies have indicated, that CSCs are resistant to many of the standard therapeutic regimes, including chemotherapy and radiation therapy (114; 107)). Hence, it is of critical importance to develop a deeper understanding of the plasticity and heterogeneity of cancer cells to overcome these therapeutic limitations, and for this the combination of appropriate mathematical models and sufficient experimental data could be decisive. It is for this purpose that in the next subsection we illustrate a particularly meaningful series of experimental results, performed by La Porta's group in the Department of Biosciences at the University of Milan, which provide crucial new information about phenotypic switching and shall be the main reference of our subsequent work.

2.2.1 Overshoot of CSCs in sorted cancer cells: experiments and models.

In (71) further evidence is provided about the existence of phenotypic switching in a population of human melanoma cells cultivated in vitro and, most interestingly, it is shown that the transition from the CC state to the CSC state in melanoma cells is regulated by an internal underdamped homeostatic mechanism controlling the fractional population, not dependent on external microenvironment.

In the experiments, human melanoma cells (a minimum of $5 \cdot 10^5$ cells for each experiment) are sorted using three distinct cancer stem cell (CSC) markers (CXCR6, CD271 and ABCG2); cells are then collected and plated in a culture medium, in such a way that the percentage of cells positive to the CSC markers is very tiny, generally inferior to 1% of the total number of cells (in any case inferior to the homeostatic fraction of CSC for that specific kind of cancer). Then cells are let grow and cell populations dynamics for longer times than earlier studies (such as (104)), because the experimenters recognized that negative cells may require time to revert their phenotype (71). What is observed in the following days is that, if the initial percentage of CSCs is below a certain threshold (about the 1% mentioned before), the population of negative cells switch en mass substantially overshooting the fraction seen in unsorted cells. The maximum in the percentage of CSC is reached in the course of 6-10 days, and then the fraction of CSC slowly returns to the level of unsorted cells. This clearly indicates that the CSC population is homeostatically regulated. in Fig. 2.3

The initial concentration of sorted CSC cells (on the total population) is in percentage ranging from 0.2% to 1%, while it is usually ranging from one to few percents for unsorted cells (i.e. in physiological conditions in tissue). We see that the percentage of CSC can reach very high values (up to 25% of the total), which would be very difficult to explain without the hypothesis of phenotypic switching; we see also a peculiar dependence of the height of the peak on the initial concentration of CSC: the lower the initial percentage of positive cells, the higher the maximum peak observed during the dynamics.

During the experiments, the percentage of CSC cells is obtained, as mentioned before, measuring the expression levels of three biomarkers (CXCR615, ABCG212, and

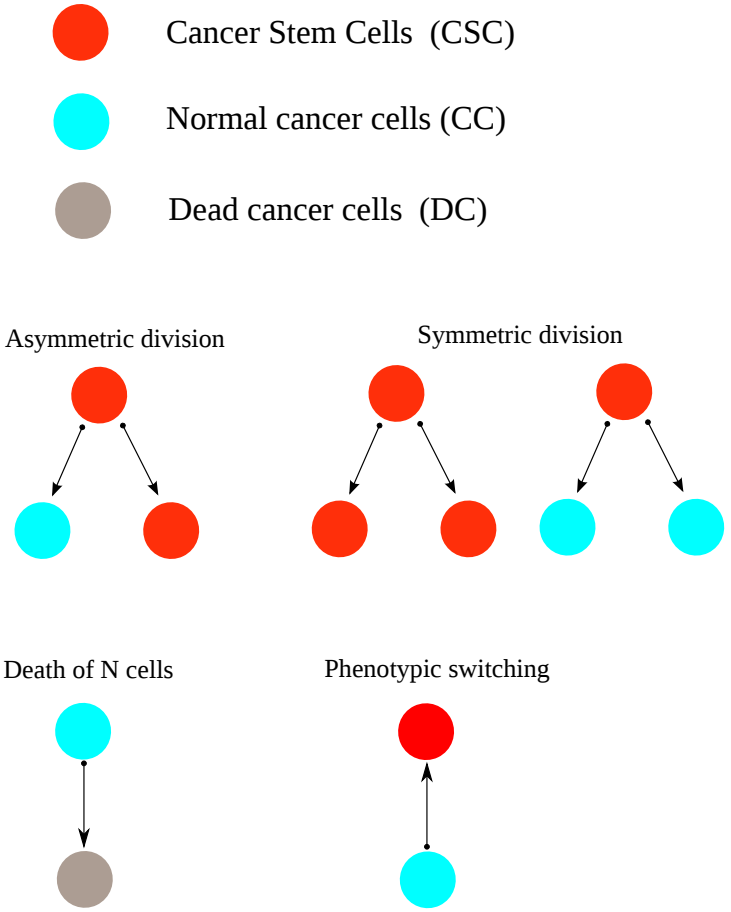

Figure 2.2: The theory of the aberrant hierarchy with CSC with the inclusion of the phenotypic switching hypothesis: cells and processes.

CD27110). For all three markers, positive cells are known to give rise to a bigger tumour in immunodeficient mice than the negative ones (the small tumour obtained with the negative population could be due to the use of an imperfect marker, to sorting errors or to phenotypic switching). Cells that are negative for CXCR6, ABCG2 or CD271 markers are sorted by flow cytometry, mixed with the negative ones in a fixed percentage ranging from 0.2% to 1% (see Fig. 2.3) and then plated and grown under standard conditions. Considering that the positive cells represent only a very small fraction of the entire population, the possible error due to sorting is extremely low. Cells are then collected 3, 10 and 20 days after sorting and analysed for each marker by flow cytometry or immunofluorescence (for the complete set of results, see (71)). In all cases, the marker-negative cells re-express their marker in a time-dependent manner. In particular, the marker-positive population overshoots its initial level at 10 days post-sorting and then returns to the steady-state level at 20 days. With a procedure of remixing cells sorted using the level of expression of one biomarker and analysed the levels of expression of one of the other biomarkers during the growth of the tumoral mass, the authors prove that there is a strong overlap between the populations of cell positive to each biomarker, and

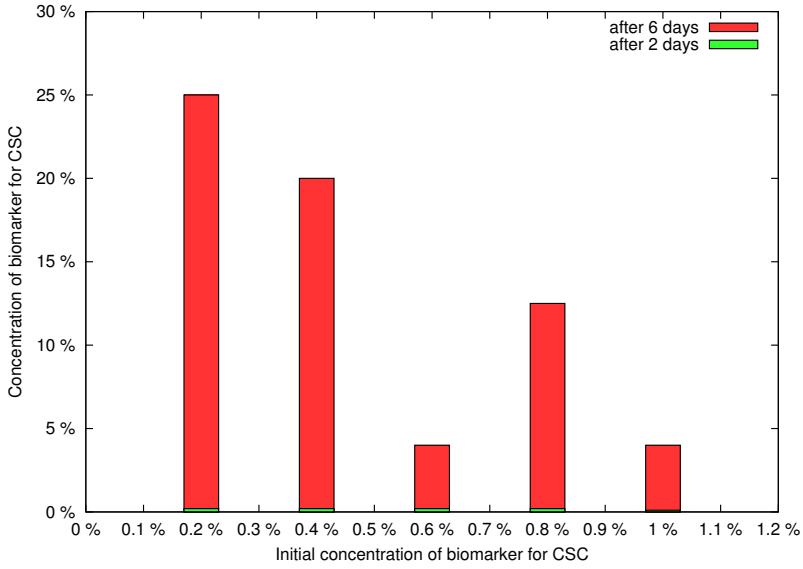

Figure 2.3: The percentage of positive cells obtained in the different experimental conditions (i.e. with different initial CSC concentration), taken from (71).

thus that the overshoot is independent of the particular marker chosen to characterize and sort CSCs. Indeed, the overshoot is observed for each biomarker.

The results of the experiments (the overshoot and the lack of significant mutations) seems to indicate quite clearly that the fraction of CSCs in cancer tissue is subject to a control mechanism of homeostatic regulation, aimed at maintaining the level of CSC at about 1% of the total cancer cell population, and that switching is not a statistically random event, but is regulated by the CSC population. What is missing is the exact nature of this biological control mechanism. Concerning this, the authors of (71) focus on the role of miRNA expressed by melanoma cells. miRNAs influence most fundamental biological processes by ultimately altering the expression levels of proteins either through interference with mRNA translation or by reducing the stability of the mRNA in the cytoplasm, and human melanoma cells express many miRNAs. The authors identified all the miRNAs that were significantly altered with respect to the unsorted condition in the subsequent cellular dynamics of the experiments and, using bioinformatics tools and the Diana-MirPath biological database, were able to identify the signalling pathways (Wnt and PI3K-AkT) most likely to be targeted by the miRNAs. The same miRNAs control network also regulates factors related to the epithelial–mesenchymal transition (EMT), a profound event for large-scale cell movement during morphogenesis at the time of embryonic development, which, because a similar process has been observed at the invasive front of metastatic cancer, is considered to be a hallmark of neoplastic transformation and seems to be related to CSC switching (71).

The action of the miRNAs regulatory network, as described in (71), on biological processes related to the control of CSC switching is very complex and not understood in detail. To understand the observed CSC population dynamics in which the CSC-like cell population overshoots after sorting, the authors propose a simple mathematical model for the dynamics of hierarchical cell populations with CSCs, cancer cells and senescent

cells. Previous models for phenotypic switching, consisting in the introduction of a small probability for cancer cells to switch back to the CSC state (104; 113), predicted that the CSC population would reapproach its steady-state concentration after sorting, but monotonically, without an overshoot. In the model proposed in (71), the switching probability is assumed to be usually zero, but increases drastically after the activation of a miRNA network which responds to the lack of CSCs in the population. The switch probability is thus dependent on the expression level of a set of relevant miRNAs which is in turn controlled by the population of CSCs.

Here we provide a brief description of the model, written in terms of rate equations. Let's indicate with S the number of CSC, C the number of CC, D the number of senescent cells. CSC duplicates symmetrically into two CSC with probability p_2 or into two cancer cells with probability p_0, or asymmetrically into one CSC and one cancer cell with probability p_1, so that $p_0+p_1+p_2=1$. Cancer cells can switch to CSCs with a probability $p(\mu)$ depending on the concentration of relevant miRNAs μ, or duplicate and give rise to two non-duplicating cancer cells with probability $1-p$. Defining $\epsilon = p_2 - p_0$ and fixing for simplicity the rates of duplication and switching equal to R_d for all the cells, the model assumes the following form

$$\frac{dS}{dt} = R_d(\epsilon S + p(\mu))(C+D) \tag{2.6}$$

$$\frac{dC}{dt} = R_d(1-\epsilon)S - R_d C(1-p(\mu)) \tag{2.7}$$

$$\frac{dD}{dt} = R_d(2C - p(\mu)D). \tag{2.8}$$

In the model the switching probability is not constant but is controlled by the level of relevant miRNAs, which responds to the depletion of the CSC population, as observed experimentally. The authors assume that the miRNAs regulating phenotypic switching are produced with a rate that rapidly vanishes when the fraction of CSCs $f_S = \frac{S}{S+C+D}$ is sufficiently large, and decay with a constant rate γ, which can be modelled as follows

$$\frac{d\mu}{dt} = \beta \exp\Big(-\frac{f_S}{s_0}\Big) - \gamma\mu. \tag{2.9}$$

The authors suppose switching probability is activated when the level of miRNA is above a threshold μ_0. The probability of switching is then modelled with a smooth sigmoid function

$$p(\mu) = 1 + \tanh\Big(\frac{\mu-\mu_0}{\sigma}\Big). \tag{2.10}$$

According to this model the switching probability is usually zero but increases drastically after the activation of a miRNA network which responds to the lack of CSCs in the population. This is an intuitive explanation for the presence of the overshoot in the evolution of f_S: when the CSC population is depleted, the miRNA expression level shoots up, triggering a large switching probability, which in turn induces an overshoot in the fraction of CSCs in the population that then falls back to the steady state level as the level of miRNAs decreases. As can be seen from the output of numerical simulations of the model (71) (performed choosing realistic parameters for (2.8)), the amplitude of the overshoot whose decreases as the initial fraction of CSCs increases, in qualitative accordance with results reported in Fig. 2.3.

There are well-grounded reasons to be dissatisfied with the model presented in (71) and summarized in Eqs. (2.8) – (2.10): in particular, it does not give a clear picture of the molecular mechanism which regulates the switching of cancer cells; in particular, it is not clear how the exponential function in the equation (2.9) can be justified, nor what kind of molecular regulation processes (even at a coarse-grained level) are needed to regulate in such a way the production rate of molecule μ.

Having acknowledged the said shortcomings of the presently available model, in the next section we shall build an alternative mathematical model for the overshoot dynamics of cancer cells as observed in the experiments. This model will rely on the hypothesis (specular to that of the previous model) that in the culture medium of the grown tumor mass a molecular species (probably a miRNA) is present that acts as an inhibitor o the process of phenotypic switching. As we shall see in the next section, one of the key assumption we make is that such molecule is a product of the CSCs themselves: it is therefore part of a system of homeostatic control by which CSCs autoregulate their own fraction (on the total number of cells) in the whole cancer tissue. This hypothesis can be easily subjected to an experimental test: the procedures of the experiments reported in (71) should substantially be repeated, but plating only negative cells in two different culture media, one of which is the neutral, standard one, also employed in the first series of experiments. The second culture medium employed in this kind of experiments should be the one resulting from the growth of a mass of the same kinds of melanoma cells for several days in experiments like those in the previous series (71), and so, according to our working hypothesis, repleted with the molecular species expressed by cancer cells during their growth. If our proposed control mechanism is correct, after some days the experimenters would observe a massive re-expression of biomarkers for positive cells, due to phenotypic switching, in both media, but in the case of the non-neutral initial medium this effect should be delayed, retarded with respect to the case of the neutral initial medium.

2.3 A rate equations model with delay for phenotypic switching.

In this section we describe an alternative model of the control mechanism phenotypic switching that is simple, can be easily interpreted in terms of molecular interactions and can reproduce the experimental data. At variance with other models, we suggest that phenotypic switching is a spontaneous tendency of all cancer cells, and that, under standard conditions of tumour growth, switching is inhibited by a specific (and so far unknown) molecular species produced by CSC. If CSC are sorted out, the environment experiences a depletion of this molecular species and its inhibition effect is abolished, promoting phenotypic switching and allowing CSC to reach their steady fraction.

Other ingredients that characterize the model are that CSCs are capable of unlimited proliferation and can undergo different kinds of cellular division, into two CSCs, into two normal, non–stem cancer cells (CC), or into one CSC and one CC: these are the biological processes of the classical theory of CSC aberrant hierarchy, as described in Fig. 2.1. On the other hand, CC can only divide into two CCs for a limited number g of generations, after which they become senescent and die; we note that the process of senescence of tumours cells is well documented, and has been investigated even in the context of the CSC hierarchy (89). Thus we do not have just one compartment of CC, like in 2.1, but g compartments of CC, one for each generation of CC. Cells of generation g cannot divide anymore and die at a constant rate (see Fig. 2.5).

We also assume that CC can switch stochastically to CSC at constant rate (in absence of inhibition), regardless of their age, while dead cells cannot switch anymore. Moreover, CSC produce a molecular species that inhibits switching according to a Michaelis–Menten molecular mechanism.

A further ingredient that allowed us to improve the agreement with the experimental data is the introduction of a delay in the action of CCs, that account for the time needed for cellular replication and switch, and have the effect of avoiding an immediate overshoot of CSC, that would not agree with what observed (see below). We shall further discuss later the role and implications of time delay in switching.

Under well-mixed microenvironment hypothesis, and considering we are dealing with a number of cells of the order of 10^6, so that we can neglect (at least for the time being) fluctuations in space, the model can be expressed by a system of delay rate equations in the concentration S of CSC, the number N_j of CC of generation j (with $0 \leq j \leq g$) and the number m of the inhibiting species. The overall mechanism, with the complete reaction networks, is sketched in Fig. 2.5 and its dynamics can be described by the system of equations (2.11)

$$\begin{cases} \dot{m} & = k_{mp}S - k_m m \\ \dot{S} & = (k_{2S} - k_{2N})S + \sigma \sum_{i=0}^{g-1} N_i(t-\tau)\Big(1 - \frac{m^h(t-\tau)}{K^h + m^h(t-\tau)}\Big) \\ \dot{N}_0 & = (2k_{2N} + k_{1N1S})S - k_d N_0 - \sigma\Big(1 - \frac{m^h(t-\tau)}{K^h + m^h(t-\tau)}\Big)N_0(t-\tau) \\ \cdots & \\ \dot{N}_i & = 2k_d N_{i-1} - k_d N_i - \sigma\Big(1 - \frac{m^h(t-\tau)}{K^h + m^h(t-\tau)}\Big)N_i(t-\tau) \\ \dot{N}_{i+1} & = 2k_d N_i - k_d N_{i+1} - \sigma\Big(1 - \frac{m^h(t-\tau)}{K^h + m^h(t-\tau)}\Big)N_{i+1}(t-\tau) \\ \cdots & \\ \dot{N}_g & = 2k_d N_{g-1} - 2k_{dfin} N_g \end{cases} \tag{2.11}$$

Table 2.1 show the list of all the coefficient expressing the temporal rates of the cellu-

lar and molecular processes included in our model and the other parameters needed to describe our system.

Coefficient	Cellular/molecular process
$\mathbf{k}_{mp}$	rate of production of signal molecule concentration m by CSC
$\mathbf{k}_m$	rate of decay of signal molecules concentration m
$\boldsymbol{\sigma}$	rate of switching of CC cells
$\mathbf{k}_{2S}$	rate of symmetric division of CSC cells into 2 CSC cells
$\mathbf{k}_{2N}$	rate of symmetric division of CSC cells into 2 CC cells
$\mathbf{k}_{1N1S}$	rate of asymmetric division of CSC cells into 1 CC cell and 1 CSC cell;
$\mathbf{k}_d$	rate of symmetric division of CC cells into 2 CC cells
$\mathbf{k}_{dfin}$	death rate of senescent (last generation) CC cells
$\mathbf{K}$	threshold of Hill's inhibition function
$\mathbf{h}$	exponent of Hill's inhibition function
g	number of generations to senescence of CC cells

Table 2.1: Parameters of the model and corresponding molecular/cellular processes.

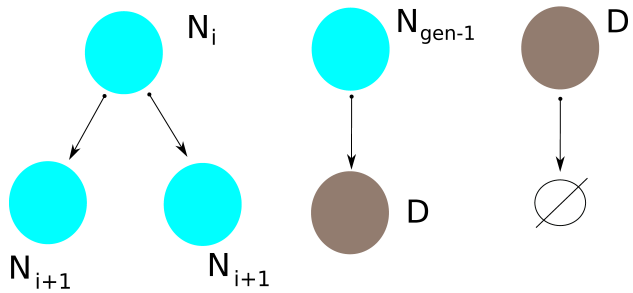

Figure 2.4: Ageing of normal cancer cells (CCs): a CC of generation N_i ages undergoing division, producing two CCs of generation $N_i + 1$ (left). After a maximum number g of cellular divisions, CCs get senescent (dead CCs) and cannot divide (or switch) anymore (centre). Dead CCs are cleared at a certain rate (right).

We note that the Michaelis–Menten molecular mechanism of switching inhibition results in the functional form of a Hill's function (with Hill's coefficient h and threshold K) that models the level of inhibition depending on the number m of the inhibiting species. The other terms are simply derived from the law of mass action.

Of course the quantity resulting from the solution of (2.11) which is of greatest interest and that can be compared with experiments is the fraction $c = \frac{S}{S+N}$ of CSC over the total number of cells, with $N = \sum_{i=0}^{g} N_i$. being the total number of CC.

The rate equations are to be solved numerically using a custom made C code. The numerical method used to solve the system of differential equations is a variable-coefficient linear multistep Adams method in Nordsieck form (from the GNU Scientific Library),

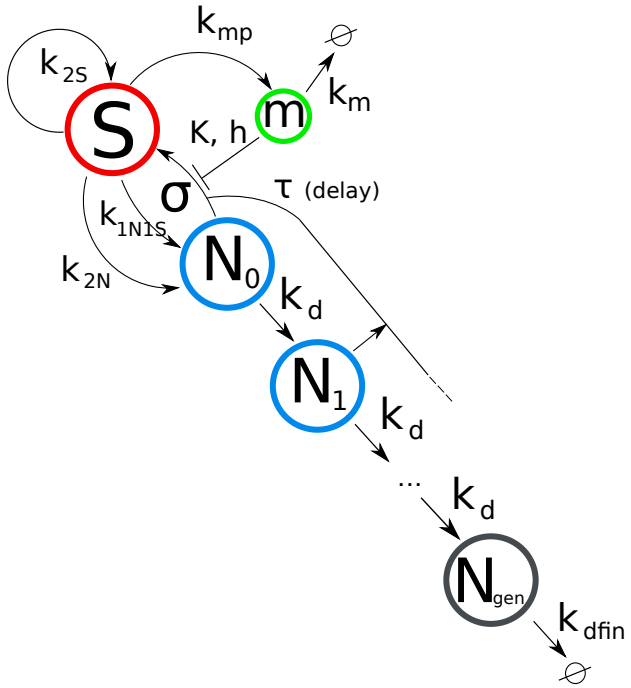

Figure 2.5: Reaction network for the mean–field model, involving cancer stem cells (S), normal cancer cells (N_j) and inhibitor molecule m.

which employs a variable time-step. In order to solve the system (2.11), one has to define the numerical values of the parameters which appear in it (see the next subsection) and a set of initial conditions compatible with the experimental settings reported in (71). Initial populations of CC is typically $N^0 \sim 10^6$,while the initial fraction of sorted CSC cells ranges from 0.001 (0.1%) to 0.01 (1%) of the total number of cells. We thus take $N^0 = N(0) = 10^6$ and $S_0 = S(0) = f_0 \cdot N^0$, with f_0 ranging from 0.001 (0.1%) to 0.01 (1%). Since our system (2.11) is made up of delay ordinary differential equations, initial conditions have to be defined in the whole interval $[-\tau, 0]$. It is natural to choose $S(t) = 0$, $N(t) = 0$ and $m(t) = 0$ for $t \in [-\tau, 0)$, since cancer cells are plated at time $t = 0$; we assume that $m_t = 0$ or $t \in [-\tau, 0)$, since it is evident that the technique of flow cytometry to sort cells in a fluid flow should be quite efficient at removing most traces of the previous environment and selecting the cells to be plated apart only (see (71) for a description of the technique employed). Even if there is strong evidence that during the sorting procedure the inhibitor molecule is mostly washed out, a small amount of $m(0)$ left does not change qualitatively the results; cf. Fig. D.2 in Appendix D .

Finally, we initially assume for simplicity that CC cells at time $t = 0$ are distributed equally among the g generations. Another possibility, is that CC cells at time $t = 0$ are distributed reflecting the steady-state distribution among generations associated with our model. This alternative possibility is discussed with more details in Appendix B.

2.3.1 Overshoot in the mean–field model.

As we have extensively discussed in Chap. 1, the main problem with models of complex biological networks is that the number of independent parameters increases rapidly (super-linearly) with the number of nodes of the networks. Making things worse, given the complexity of biological systems, it is extremely difficult to have reliable and precise measures of quantitative parameters such as the rates of biological processes: in fact we have accurate numerical parameters only of an handful of well-know system, while for most of the others only esteems of their orders of magnitude. Thus to infer meaningful information about the dynamics of systems from models like (2.11) one has to make some hypothesis about the values of the parameters appearing in them or at least define the most stringent constraints on their range of variations which is compatible with experimental results. Only few of the numerical values that define Eq. (2.11) are known, at least as order of magnitude. The division rate of melanoma cells is known from literature to be 0.5-0.6/day (71).

As for the constraints to be imposed on the parameters, since we assume phenotypic switching is naturally always active, it must be $\sigma > 0$. Beside this, an important experimental fact that the model should reproduce is that, describing tumoral cells, their total number $S + N$ must be an increasing and unbounded function of time. A necessary and sufficient condition for this is that (see Appendix C)

$$\epsilon \equiv k_{2S} - k_{2N} \geq 0 \tag{2.12}$$

As is explained in Appendix C, the parameter ϵ is of particular importance, since it is the main driver of the long-term dynamics of the system. If $\epsilon > 0$ s and N grow exponentially-like in time, thus reproducing the uncontrolled cellular growth of cancer tissues. Incidentally, one can argue that the opposite case, that of $\epsilon < 0$, can describe the case of a benign tumour, that does not grow indefinitely invading neighbouring tissues. This case is discussed in Appendix C.

The parameters relative to the processes of cellular division should be chosen so that they satisfy the order of magnitude relative to experimental data about cellular division (see (71) (88)) and to reproduce the experimental values of the steady-state concentration of CSC c_{staz} (which according to (71) are of the order of at most few percent, the same as in homeostatic conditions within the tumour). From the mathematical analysis of the system (2.11), it follows that while both S and N, in the case $\epsilon > 0$, grows exponentially-like, the concentration of CSC $c = \frac{S}{S+N}$ always reaches a constant steady state, for which we can provide a closed-form expression in function of some of the parameters of (2.11) (see App. B and App. C).

At the conclusion of this process of estimation, a set of numerical parameters was chosen so that the parameters reflected the constraints provided by biology (known numerical values or orders of magnitude) and were able to provide the best fit with experimental results (height and timing of the peak of CSC concentration, steady-state value of CSC concentration) when simulating the model (2.11). The set of parameters associated with an overshoot which is most similar to the experimental data is given in Table 2.2.

The dynamics of the fraction $c = \frac{S}{S+N}$ of CSC in time, obtained solving numerically the system (2.11) with the previously defined initial conditions and the parameters in Table 2.2, are shown in Fig. 2.6, for different values of the initial concentration c_0.

In Fig. 2.7 are reported the fraction of CSC at 2 and 6 days (red bars), in function of the initial concentration $c(0)$, as observed in the experiments (71), while the continuous lines represents the values of the concentration at the same time obtained solving the system (2.11).

Coefficient	value	Cellular/molecular process
$\mathbf{k}_{mp}$	300.0 day^{-1}	rate of production of signal molecule concentration m by CSC
$\mathbf{k}_m$	4.0 day^{-1}	rate of decay of signal molecules concentration m
$\boldsymbol{\sigma}$	0.22 day^{-1}	rate of switching of CC cells
$\mathbf{k}_{2S}$	0.4 day^{-1}	rate of symmetric division of CSC cells into 2 CSC cells
$\mathbf{k}_{2N}$	0.1 day^{-1}	rate of symmetric division of CSC cells into 2 CC cells
$\mathbf{k}_{1N1S}$	0.2 day^{-1}	rate of asymmetric division of CSC cells into 1 CC cell and 1 CSC cell
$\mathbf{k}_d$	0.4 day^{-1}	rate of symmetric division of CC cells into 2 CC cells
$\mathbf{k}_{dfin}$	0.4 day^{-1}	death rate of senescent (last generation) CC cells
$\mathbf{K}$	600000	threshold of Hill's inhibition function
$\mathbf{h}$	4	exponent of Hill's inhibition function
g	20	number of generations to senescence of CC cells
τ	4 (days)	time delay associated to the switching-regulation process

Table 2.2: Parameters of the model, their values associated to best fit with overshoot results, and corresponding molecular/cellular processes.

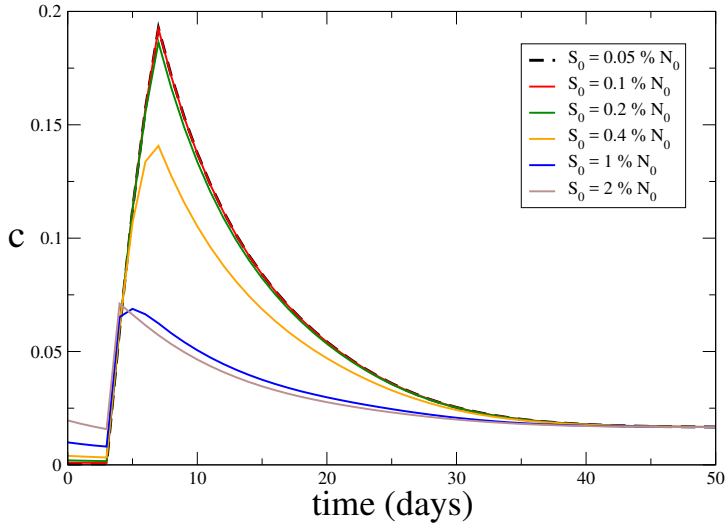

Figure 2.6: The dynamics of the concentration $c(t)$ of CSC with different choices of the initial value $c(0)$ ($N_0 = 10^6$, $m_0 = 0$).

Consequently, there is at least a set of parameters for which the model reproduces qualitatively the experimental data. Being a thorough exploration of the whole parameter space of the model computationally infeasible, we started from those of Table 2.2 and studied the dynamics of the system varying one parameter at a time.

The value of the maximum c_{max} concentration of CSC during the overshoot and of its stationary value c^* are displayed in Fig. 2.8 as a function of some of the parameters of the system. The size of the overshoot increases smoothly but considerably as a function of the switching rate σ and of the rate difference ϵ, and decreases as a function of

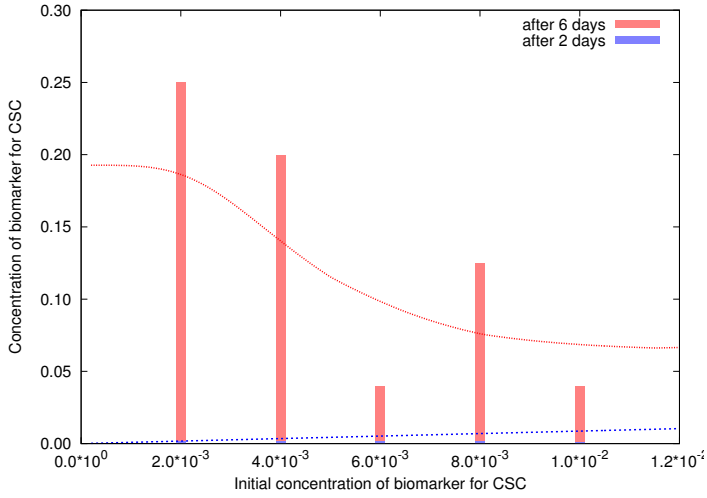

Figure 2.7: The experimental fraction $c(t)$ of CSC after 2 days (blue bars) and after 6 days (red bars) from their sort out, as a function of their remaining fraction $c(0)$. The curves indicate the simulated results.

the ageing rate k_d of CC. The dependence on the other parameters is much weaker (cf. Figs. D.5 – D.18 in Appendix D). These results suggest that the overshoot is overall a robust property of the model. The strong dependence on the switching rate σ indicates that phenotypic switch is an important ingredient to reproduce the experimental observations. Two important parameters seem to be k_d and g, that control together the total amount of CC. For large values of k_d or g, the overall number of CC is large and the overshoot effect disappears. A simple qualitative explanation for this is that if we increase the values of k_d or g, then the total number of CC cells increases exponentially, while the number of CSC increases only as there are more CC cells able to switch. Therefore we expect the increase in those values to reduce the amount of the overshoot.

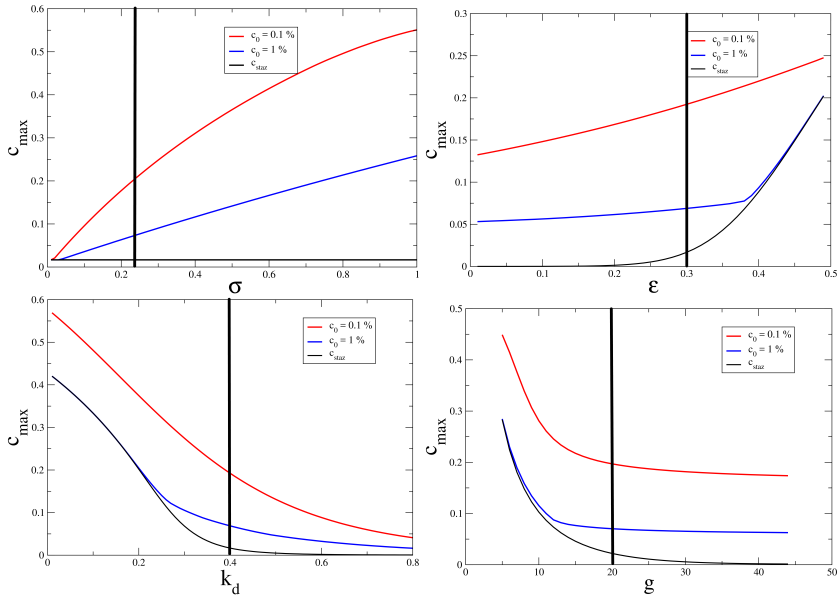

Figure 2.8: The size c_{max} of the overshoot as a function of the switching rate σ, of the rate difference ϵ, of the rate k_d and of the number g of generations of CC. The black vertical lines indicate the values of the parameters employed in our simulations.

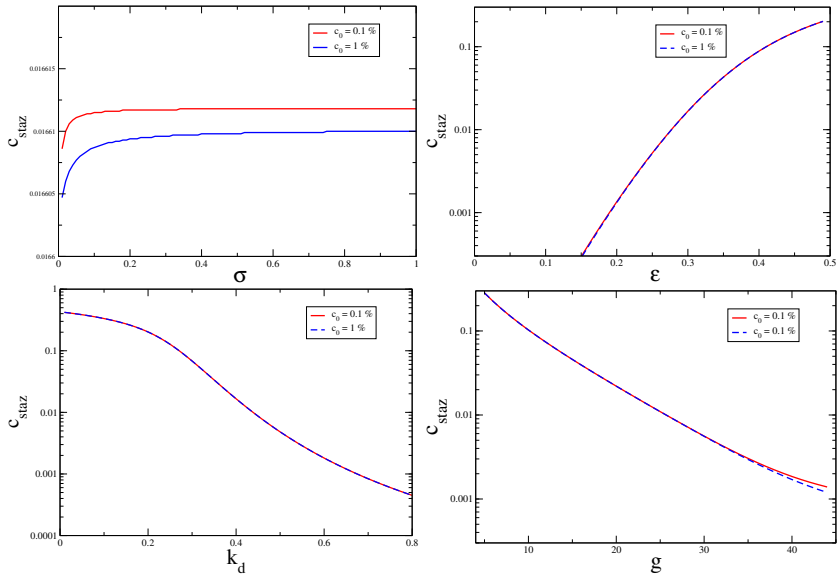

Figure 2.9: The size c_{staz} of the steady state CSC concentration as a function of the switching rate σ, of the rate difference ϵ, of the rate k_d and of the number g of generations of CC.

2.3.2 Is this a minimal model?

A relevant question about the model is whether all the terms in Eqs. (2.11) are really necessary to describe the system. If the time delay τ is not introduced in the switch-related terms of the equations 2.11, the overshoot peak always occurs within the few hours after sorting (see Fig. D.18 in Appendix D), while it is observed that some days are needed for the increase of $c(t)$ (71). Moreover, removing the time delay, the dependence of the maximum of CSC concentration c_{max} on the initial concentration (as observed in (71): see 2.3 and 2.7) is lost. To see an example of this, we simulate the system 2.11, with time delay $\tau = 0$ and different initial concentrations c_0. The output is shown in Fig.2.10.

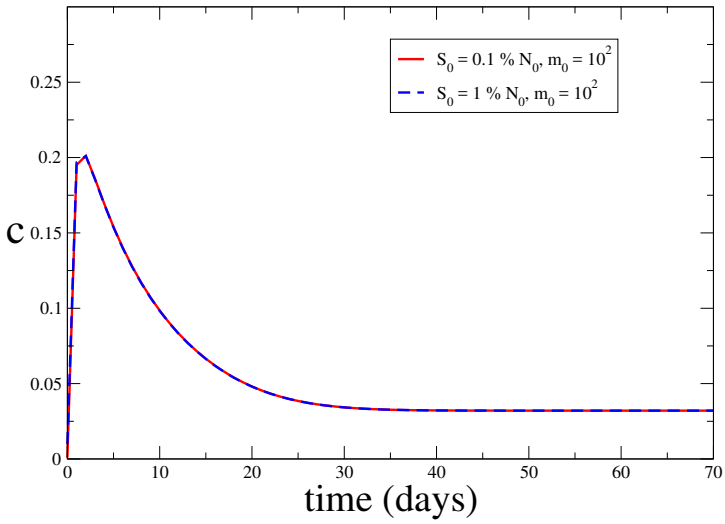

Figure 2.10: Simulation of relative concentration of CSC evolution in the model without delay ($\tau = 0$): overshoot of the CSC relative fraction ($N_0 = 10^6$, $m_0 = 0$). Initial CC cells are equally distributed among the g generations.

Apart from the overshoot being immediate, we can see from Fig.2.10 that the different initial CSC concentrations c_0 lead essentially to the same dynamics of c_t, and thus to the same value for c_{max} , while in the experiments the amount of CSC in the overshoot peak decreases with the fraction $c(0)$ of initial CSC. Thus removing the time delay from our model causes the loss of two macroscopic features of the experimental results (the timing of the overshoot and its dependence from the initial fraction of CSC).

If instead one disregards the ageing of CC, describing them just in terms of their total number N, transforming Eqs. (2.11) into

$$\begin{cases} \dot{m} &= k_{mp}S - k_m m \\ \dot{S} &= (k_{2S} - k_{2N})S + \sigma N\left(1 - \frac{m^h}{K^h+m^h}\right) \\ \dot{N} &= (2k_{2N} + k_{1N1S})S - \left(k_d + \sigma\left(1 - \frac{m^h}{K^h+m^h}\right)\right)N, \end{cases} \tag{2.13}$$

then it is possible to write a nearly closed equation for the concentration $c(t)$ of CSC. Setting $\Delta = \frac{k^h}{k^h+m^h}$, this reads

$$\frac{dc}{dt} = -(\epsilon + \lambda + k_d)c^2 + (k_d + \epsilon - \sigma\Delta)c + \sigma\Delta. \tag{2.14}$$

Since for large time and $\epsilon > 0$, then $m \longrightarrow \infty$ (cf. Appendix C), and so $\Delta = \frac{k^h}{k^h+m^h} \longrightarrow 0$, giving the stationary value

$$c^* = \frac{1}{1 + \frac{2k_{2N}+k_{1N1S}}{k_{2S}-k_{2N}+k_d}} \tag{2.15}$$

For c^* to be of the order of the observed order of magnitude $\sim 10^{-2}$, Eq. (2.15) requires that $3k_{2N} + k_{1N1S} \gg k_{2S} + k_d$. This is unlikely to be the case, because the experimentally observed division rates k_d, k_{2S}, k_{2N}, k_{1S1N} are all of the order of 0.1 days^{-1} (89) , and for sure $k_{2N} < k_{2S}$, because we are in the case $\epsilon > 0$.

Finally, if phenotypic switching occurs in an uncontrolled way, no overshoot is possible. In fact, if one removes the dependence on m in the second of Eqs. (2.13, this is equivalent to setting $\Delta = 1$ in Eq. (2.14. Setting $f(c) = -(\epsilon+\lambda+k_d)c^2 + (k_d+\epsilon-\sigma)c+\sigma$, the stationary points of Eq. (2.14 can be found from $f(c^*) = 0$ and are

$$c^*_{1,2} = \frac{k_d + \epsilon - \sigma \pm \sqrt{(k_d + \epsilon - \sigma)^2 + 4\sigma(\epsilon + \lambda + k_d)}}{2(\epsilon + \lambda + k_d)}. \tag{2.16}$$

These are always real because $\epsilon + \lambda + k_d > 0$, and have opposite sign (let us call, e.g., $c^*_2 < 0$ and $c^*_1 > 0$). The solution c^*_2 is not admissible because it cannot describe a fraction of cells. The stability of c^*_1 is determined by the condition

$$f'(c^*_1) = k_d + \epsilon - \sigma - (\epsilon + \lambda + k_d)c^*_1 < 0, \tag{2.17}$$

which is satisfied for every initial concentration c_0 in the range $[0, 1]$. One can exclude the existence of an overshoot showing that $c(t)$ is monotonous as it approaches c^*_1 for every t and every c_0. Since $-(\epsilon + \lambda + k_d) < 0$, then $f(c) > 0$ if and only if $c^*_2 < c < c^*_1$, that for the case of interest becomes $0 < c(t) < c^*_1$. For $c(t)$ to have an overshoot, starting from a small c_0, $c(t)$ must grow in a finite time to a peak strictly above the attractive equilibrium value c^*_1. But this is not possible since $c'(t) = f(c)$ is strictly negative in the region $c^*_1 < c(t) < 1$. Thus, in the case of undamped (non-inhibited) phenotypic switching at constant rate σ the phenomenon of overshoot of the concentration c is not possible.

2.3.3 Interpretation of the experiments on growth inhibition

In Fig. 2.11 we show that it is possible to inhibit the growth of the total number of tumoral cells after the sort–out of CSC, adding to the system the medium obtained prior to sorting, that arguably contains the inhibiting molecule.

To understand this experiment from a quantitative point of view, we simulated it with the model of Eq. 2.11. For this purpose, we compared the total number of cells at fixed time T when the sorted cells are treated at time $t = 0$ with a solution containing m_0 inhibiting molecules, with the total number of cells in the untreated case (i.e., $m_0 = 0$). We defined the quantity

$$R_{T,m_0} = \frac{[S(T) + N(T)]_{m(0)=m_0}}{[S(T) + N(T)]_{m(0)=0}} \tag{2.18}$$

to quantify the effect of the inhibiting molecule at time T. It should be noted, that the amount of inhibiting molecule varies with time from its initial value, because, besides that provided at time zero, it is produced by CSC and undergoes degradation.

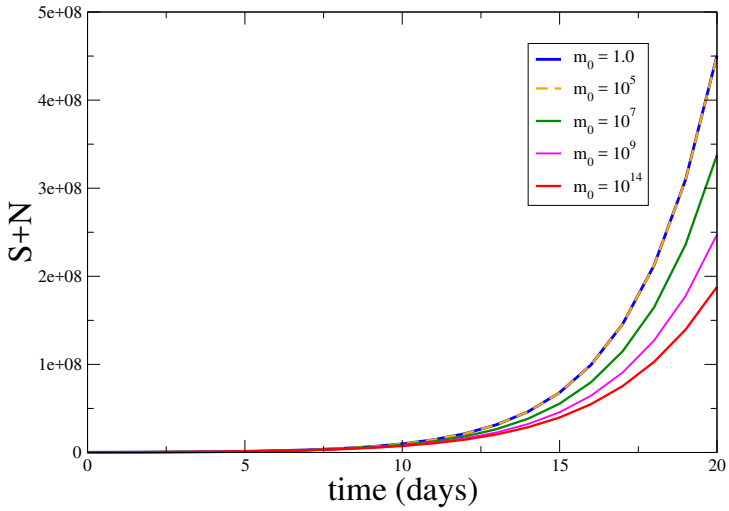

Figure 2.11: Evolution of the total number of cells $S + N$, with different values of m_0, initial number of cells $2 \cdot 10^5$, $c_0 = 2\%$.

In the left panel of Fig. 2.12 it is displayed the value of R_{20days,m_0} as a function of m_0, for different choices of c_0, evaluating the number of cells after 20 days. Independently on the remaining concentration c_0 of CSC, there is a threshold of $\sim 10^5$ molecules, beyond which one can observe an effect of the inhibitor. Assuming that the dish used for the experiment has a volume of ~ 10 mL, this threshold correspond to a concentration of 10^{-17} mol, meaning that even a tiny amount of molecule has a detectable effect on the (exponential) grow of the cells (cf. Fig. S18). The effect of the inhibitor saturates to a c_0–dependent value when its copy number reaches $\sim 10^{12}$, corresponding to ~ 1 nmol.

From the behaviour of R_{T,m_0} as a function of the experimental time T (cf. the right panel of Fig. 2.12, it is possible to notice that the ideal time at which to carry out the experiment to assess the effect of the inhibitor is no less than 20 days after sorting and plating the cells, since only after that amount of time the tumoral mass for highest initial concentrations of the molecule is reduced of about the 50%, and so the reduction is mass is macroscopically quite detectable. From Figure S18 in the Supplementary Material we can see that for $T = 20$ the tumoral mass reaches the maximum capacity of a single Petri dish, i.e. $10^8 - 10^9$ cells. Thus the time $T = 20$ days seems to be the optimal one to perform our proposed experiment.

2.3.4 Delay and phenotypic switching: clues from literature.

The time delay in the switch-related terms of our model (2.11) may so far appear to have been introduced *ad hoc* for purely the phenomenological reasons, in order to reproduce the exact timing of the overshoot observed in the experiments. In (71) it is hinted that cells may require some time to revert their phenotype (this is cited as the main reason for following, in the experiments, the population dynamics of cancer cells for longer times than earlier studies), but the topic is not further discussed in the paper. It is quite obvious that a biological process like phenotypic switching, which presumably involves the reorganizations of cellular circuits that regulates the expression of phenotype, and

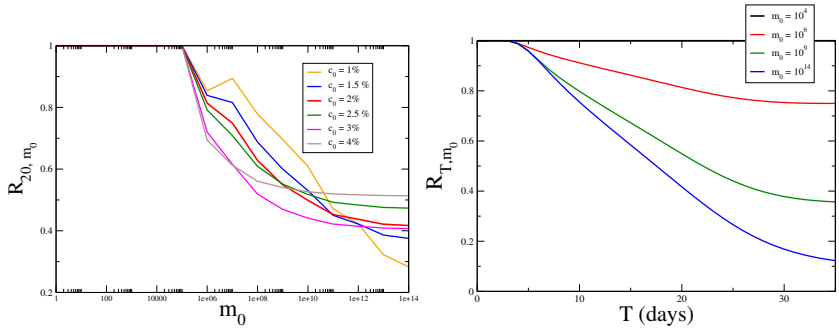

Figure 2.12: (left panel) The quantity R_{T,m_0} for $2 \cdot 10^5$ sorted cancer cells assuming different possibilities for c_0 and making the experiments at $T = 20$ days. (right panel) The same quantity plotted as a function of T for different choices of m_0.

which is subject to control and feedback by a complex molecular regulatory network, should take some time to complete, but no attempts to provide quantitative information about its duration have been made.

In this brief section we plan to show that our model is not in contrast with one of the few mathematical models related to phenotypic switching existing in literature that reproduces experimental data (116). This model is also of particular interest since the model in (116) was build to clarify the connections between phenotypic switching the epithelial-mesenchymal transition (EMT), a topic which has been the subject of much priomising investigations recently ((117; 118; 119; 120)).

EMT is a cellular differentiation process wherein epithelial cells lose expression of genes that facilitate the ordered inter-cellular junctions that are characteristic of epithelial cells and instead adopt mesenchymal features including the capacity for migration. This process plays essential roles in normal development, including facilitating the formation of the neural crest and secondary palate in embryogenesis (121); however, in recent years, the study of EMT has transcended developmental biology as cancer biologists have noted similarities between developmental EMT and the acquisition of invasive properties by cancer cells in various carcinomas (121) . While EMT appears to be important in granting certain cancer cells the ability to metastasize, it is an implication of the CSC hypothesis, as we have already seen (65; 80), that successful metastatic growth requires the dissemination of a CSC from the primary tumour (and its subsequent re-establishment of tumour growth at a secondary site) (116). A link between EMT and CSCs was suggested in (118) and (119), where it was independently found that by inducing both normal and neoplastic (i.e. coming from breast carcinomas) mammary epithelial cells to undergo EMT, the proportion of cells in the population under study displaying stem-like properties, including the ability to form spheres in nonadherent culture and tumorigenicity (for neoplastic cells), increased.

The model proposed in (116) is an attempt to provide an explanation for the results of (118). In fact there are two possible explanations for the observations of enrichment of mammary epithelial/breast cancer CSCs following EMT: that non-CSCs are endowed with a CSC phenotype as a consequence of EMT or that self-renewal of pre-existing CSCs is upregulated by EMT. The first explanation implicates that EMT may be viewed in this context as a phenotypic switching event.

To simulate the effect of the first scenario (which is the one of interest for us), the

authors of (116) developed a hierarchical model for the evolution of the cellular subpopulations (the CSC population and the different generations of CC), adding of course terms that provided the transition from the CC compartment to the CSC compartment. The resulting model is nearly identical to our model (2.11), with the notable exception that switching is not subject to a control mechanism. In order to fit the data of (118) with this model the authors found that, apart from choosing appropriate values from the parameters, it was required to impose an initial delay period (of some days) for the process of switching, after the process of EMT induction. This because experimentally, a delay period of several days passes before any appreciable increase in stem cell activity is observed following EMT induction (118).

In conclusion, one of the few examples of mathematical models available in literature for phenotypic switching does present one of the key feature of our model (2.11) (the time delay of switching, the other being the control mechanism for switching). The delay period proposed for switching in (116) also appears to be of the same order of magnitude of the delay period proposed by us (some days). All the available clues seems to indicate that switching may actually be subject to a delay of several day to be completed since its induction. The connection with the process of epithelial-mesenchymal transition (EMT) is also interesting, since there is a growing number of papers exploring this relationship (including (71)), offering a possible biological mechanism for phenotypic switching.

2.4 The spatial model.

The rate equations (2.11) average the concentration of cells and of molecule over space. To investigate whether specific spatial pattern appear during the replication of cells, we studied an extension of the model in which the spatial degrees of freedom appear explicitly. In recent years advanced techniques of immunohistochemistry [4] have been developed that provide biologically and clinically useful information on protein expression and cellular localization in cancer cells (see for example (123)). A popular technique of immunohistochemistry used to investigate the spatial distribution in tissues is immunofluorescence, which uses the specificity of antibodies to their antigen to target fluorescent dyes to specific biomolecule targets within a cell, and therefore allows visualization of the distribution of the target molecule through the sample. Of course to investigate the distribution of cancer cells in a tissue, the target molecules to be taken are their respective biomarkers. Spatial patterns resulting from the simulation of the model can thus be compared with patterns observed by means of immunohistochemical analysis in the the growth of tumoral masses cultivated in the wet lab. Experiments of such kind are in course of being performed by La Porta's group: these experiments consist in selecting a single positive cell or a mass of negative cells, following the procedures illustrated in 2.2.1, which are then plated to grow. They can be plated either in a Petri dish with a pool of nutrients or cultured in a gel to form three-dimensional spheroids.

Apart for future comparison with experimental results, the inclusion of spatial degrees of freedom is also employed to investigate the role of diffusion of the switching-inhibitor molecule in the growth of the tumoral mass and in the formation of spatial patterns: this could provide some information about the nature of the molecule in question, and for this purpose multiple simulations of the model are performed, for different values of the diffusion coefficient in a realistic range.

2.4.1 Materials and methods.

In the model we assume that cells sit on the vertexes of a regular square lattice, whose elementary length is then $\ell \sim 10$ μm (a typical value for the linear dimension of a cancer cell (124)) and whose linear size is typically $10^3\ell \sim 1$ cm. The 2D lattice is more suitable to model the growth of the tumour mass in a Petri dish, than a 3d-spheroid in gel. The model could be easily adapted to simulate growth in three spatial dimensions, with the introduction of a 3D square lattice, but we concentrated more on the 2D, also or ease of analysing the results of the simulations.

Each site of the domain can host at most one cell, either normal or CSC. It undergoes the same processes (symmetric or asymmetric division, phenotypic switching, cellular ageing and death, production of the switching-inhibitor molecule) as described in Fig. 2.5 for the rate–equation model. Also the numerical values assigned to the rates of the different processes are the same (cf. Table 2.2), with a notable exception that shall be discussed in detail later.

The cellular dynamics is simulated with a Gillespie algorithm (125), starting either with a single CSC located in the centre of the lattice or with a mass of negative cells (CC). Upon cell division, one of the new cells remains in the same site, while the other occupies a neighbouring site; if all of them are occupied, all cells in a random direction are shifted to make room for the new one. Neighbours are defined along the horizontal, vertical

[4] Immunohistochemistry involves the process of selectively imaging antigens (proteins) in cells of a tissue section by exploiting the principle of antibodies binding specifically to antigens in biological tissues.

and diagonal directions, according to the Moore scheme, which minimizes artefacts due to the on–lattice model, such as (126).

A further parameter which was not defined in connection with the rate equations is the diffusion coefficient of the molecule. This is difficult to quantify because of our ignorance about the properties of the molecule. If it were a nm–sized miRNA complex, Stokes' law would suggest a diffusion coefficient $D \sim 100\ \mu\text{m}^2/\text{s}$ (radius of about $2\ nm$, diffusion in water). In the cellular and extracellular environment the diffusion coefficient is, however, lower by about one order of magnitude due to macromolecular crowding resulting in steric constraints on diffusion and influencing diffusion via weak intermolecular interactions ((127)), thus resulting in an effective diffusion coefficient $D_0 = 10\frac{\mu m^2}{s} = 10^{-11}\frac{m^2}{s}$, which shall be our main reference order of magnitude. It should also be considered that miRNAs can be secreted into the extracellular environment through exosomes or in complexes with protein or lipid-based carriers. Accumulating evidence (129; 128) demonstrates that miRNAs as well as proteins can be transferred to neighbouring or distant cells in these secretory forms to modulate cell function. Exosomes are cell-derived extracellular vesicles (EVs) present in many and perhaps all biological fluids, including blood, urine, and cultured media of cell cultures. Exosomes range in diameter, between 40 and 150 nm (129). For miRNA coated in a vesicle, according to Stokes' law, the values of the diffusion coefficient would decrease to about $D \sim 0.1\ \mu\text{m}^2/\text{s}$ (as indicative order of magnitude). In the following list we present all the different values of the diffusion coefficient D employed in the simulations, from lowest to highest, with their physical meaning (if any):

- $D = 6 \cdot 10^{-2}\ \mu\text{m}^2/\text{s} \sim 10^{-2}\ \mu\text{m}^2/\text{s}$;
- $D = 3 \cdot 10^{-1}\ \mu\text{m}^2/\text{s} \sim 10^{-1}\ \mu\text{m}^2/\text{s}$ (miRNA coated in a vesicle of about 150 nm of diameter, diffusing in a crowded environment);
- $D = 1\ \mu\text{m}^2/\text{s}$ (miRNA coated in a vesicle of about 40 nm of diameter, diffusing in a crowded environment);
- $D = 10\ \mu\text{m}^2/\text{s}$ (miRNA of about 2 nm of diameter, diffusing in a crowded environment);
- $D = 10^2\ \mu\text{m}^2/\text{s}$ (miRNA of about 2 nm of diameter, diffusing in water);
- $D = 10^3\ \mu\text{m}^2/\text{s}$.

Simulating the diffusion (and the self-degradation) of the molecule within the Gillespie algorithm is computationally infeasible, because each movement of every single molecule takes place on a time scale which is much smaller than that associated with the cellular processes, and thus essentially all the computational time would be used to simulate the diffusion of the molecule. Consequently, the diffusion of the molecule is simulated in parallel to the Gillespie algorithm by solving a discretized version of the diffusion equations

$$\frac{\partial m(r,t)}{\partial t} \approx \frac{\Delta m(r,t)}{\Delta t} = k_{mp} S(r,t) - k_m m(r,t) + D\,\nabla^2 m(r,t), \tag{2.19}$$

where $m(r,t)$ is the number of molecules at position r of the lattice at time t and $\Delta t = 3 \cdot 10^{-4}$ days is the time step used for numerical differentiation, following a five-point stencil finite-difference explicit method (the maximum value that ensures numerical stability of this numerical scheme in 2D). The Gillespie time-step is inherently

stochastic (and thus variable), and tends to decrease in the course of the simulation, as the number of cells involved increases. Since it rapidly become very tiny, multiple Gillespie steps takes place in the interval $[t, t + \Delta]$. When the sum of successive Gillespie time steps exceeds the value of the diffusion time step, the algorithm solves the equation (2.19) on the lattice advancing with the standard diffusion time step ($\Delta t = 3 \cdot 10^{-4}$ days). Another step of diffusion is necessary to make equal the "Gillespie time" and the "diffusion time", so the algorithm solves once again (2.19) with the appropriate time step to adjust their difference. This process is repeated until the end of the simulation.

Even with this separation of the time scales, the simulations remain computationally quite expensive. Thus, as a a further approximation, we solve the diffusion equation not in the whole square domain, but only in a circle of radius $r_{max} + \delta$, r_{max} being the distance the farthest cell from the center of the square lattice, while δ is a constant (we choose $\delta = 4$, in unit of lattice sites), fixing Dirichlet boundary conditions at the border of the circle and outside of it, in the rest of the lattice: these boundary conditions are reasonable, because since the molecule is produced by the CSCs, outside of the cellular domain the only possibility, except diffusion, is progressive self-degradation, especially at long distances, so this approximation does not significantly affect cellular dynamics. The simulation effectively ends when $r_{max} + \delta = \frac{dim}{2} - 2$. With an algorithm so conceived, each simulation takes approximately 400 hours of CPU time on Intel i7 3.4Ghz processors.

The role of the space in relation to the molecule is strictly related to the only parameter for which, as mentioned, before, we cannot use in the simulation the same numerical value employed in the rate equations, that is the the threshold K of Hill's function $\Delta = \frac{K^h}{K^h + m^h}$. In fact in the spatial simulation the quantity K is a local one, since it has to be compared with the quantity m of molecule present in the specific site of the lattice, while in the rate equations the threshold K of Hill's function is global, i.e. it has to be compared to the total number of molecules m. To obtain comparable results for the two models (the spatial simulations and the rate equations), at least in terms of orders of magnitude, we cannot keep the same values of K in the two models. We therefore chose to divide the threshold K employed in the rate equations by the quantity $dim^2 = 1000^2$, that is the number of different sites in the 2d lattice: $K_{space} = \frac{K_{rate}}{dim^2}$

In the case we want to include in our simulation the time delay τ_{delay} for phenotypic switching we have to add further steps. If, according to the Gillespie algorithm, the switching of a CC into a CSC has to take place at a certain time t, instead of doing the process immediately, we store the coordinates of the site of the CC which has to switch in an array, along with the time in which the switch has to take place ($t + \tau_{delay}$). At each step of the Gillespie algorithm, at the time t_{actual} we search in the array if there are cells which have to switch in a time $t \in [t_{actual} - \tau_{i-1}, t_{actual} + \tau_{i-1}]$, where τ_{i-1} is the Gillespie stochastic time interval generated at the previous step of the algorithm. If any such cell is found, the switching of the cell takes place and then its coordinates are removed from the array. If in the interval of time between the selection of the CC which has to switch and the actual switching the CC dies or goes quiescent, switching does not happen. If in the same interval the cell moves, its coordinates in the array are updated, while if cellular division takes place, one of the two cells resulting from division inherit the future switching. A cell which has to switch is obviously not able to be selected again for switching.

One last thing has to be said about the selection of cells once the biological reaction that has to take place has been selected by the Gillespie algorithm. The coordinates of the CSCs are stored in an array which is constantly updated during the algorithm,

and the same for the CCs (adding to the parameters stored for each cell the number corresponding to its present generation). Once the kind of reaction has been chosen, for example symmetric division of CSCs, a CSC is randomly chosen in the array and then the reaction takes place. Choices are made randomly because, since the tumour grows in a culture medium in wet lab, each cell has the same resources at disposal and therefore there are no cells with a particular reproductive advantage because of their spatial position. If the reaction involved has to do, for CCs, with their age, random choice is made between cells with the same age.

2.4.2 Spatial growth of the tumoral mass.

We first performed a simulation of the model starting from a mass of only N cells (CC). The initial number of cells placed in the center of the lattice (forming a circle) was $N = 10^5$. We can see in Fig. 2.13 that such simulation (for three different values of the diffusion coefficient) display an overshoot in CSC concentration, similar in amplitude and timing to what observed in the rate equations and in the experiments. These results validate our choice of the threshold $K_{space} = \frac{K_{rate}}{dim^2}$

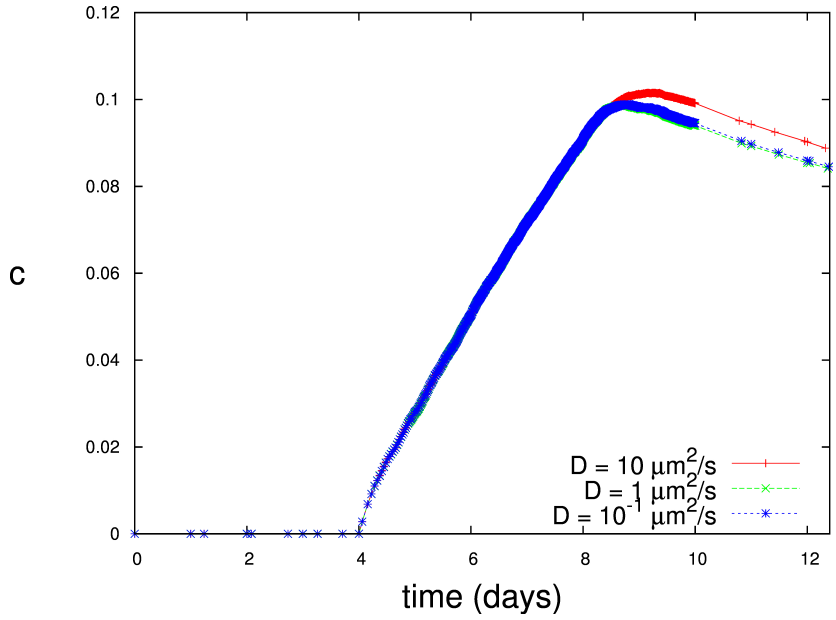

Figure 2.13: The fraction of CSC for a simulation starting from only N cells shows an overshoot similar to that observed in the rate equations (red curve $D = 10\mu m^2/s$, green curve $D = 1\mu m^2/s$, blue curve $D = 0.1\mu m^2/s$)

In Appendix E (Figs. E.1, E.3, E.5) we show the spatial plots at time $t = 11$ d of the simulations for different values of the diffusion coefficient and the corresponding radial distributions of the fraction of CSC (E.2, E.4, E.6), which clearly show that, starting from a relatively large number of CC cells, not only the timing and amplitude of the overshoot does not depend on the diffusion (as it is evident from Fig. 2.13), but that also the spatial organizations of cells in the tumour mass are quite independent from it.

We then focused our analysis of the growth of a tumoral mass starting from a single CSC, as displayed in Fig. 2.14.

The overall growth of the number of CC and CSC becomes exponential after ~ 5 days with a rate that depends mildly on the diffusion coefficient of the molecule (from 0.4 days^{-1} at low D to 0.5 days^{-1} at large D, see Figs. E.9-E.10 in Appendix E) and is indistinguishable from a model without phenotypic switching (see lower left panel in Fig. 2.14). The fraction $S/(S + N)$ of CSC decreases quite smoothly to a stationary value of few percent, slightly higher at large D, slightly lower at small D and without switching (see lower right panel in Fig. 2.14, and for the complete set of data E.7 – E.70 in Appendix E). The dependence of the values of the overall fraction of CSC on the total

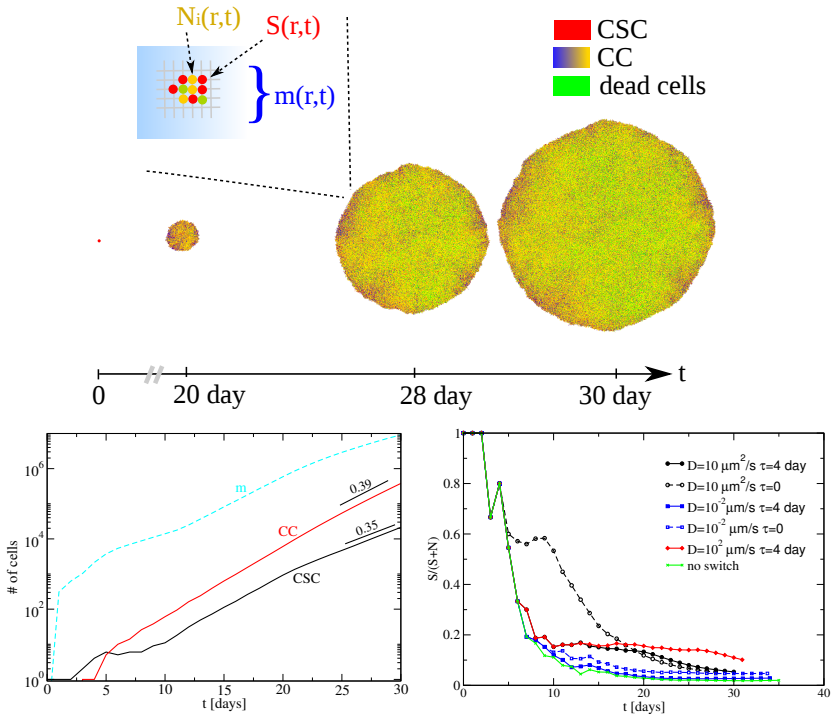

Figure 2.14: (Upper panel) The growth of a cell population with $D = 10\,\mu\text{m}^2/\text{s}$ from a single CSC at time 0. CSC are displayed in red, CC of different age from blue to yellow and dead cells in green. To the left, a sketch of the lattice model used for the calculations. (Lower left panel) The total number of cells as a function of time. (Lower right panel) The total fraction of CSC as a function of time for different choices of the diffusion constant of the molecule, and in absence of phenotypic switching.

number of cells from the diffusion coefficient D can be explained in a relatively simple way: for higher values of D ($D > 10\,\mu\text{m}^2/\text{s}$) the molecule is more likely to travel long distances from the point where it was produced and reach the border of the domain [5], where it dies (Dirichlet conditions at the borders); thus for higher values of D the total amount of molecule in the domain is inferior than in the case with lower values of D (as it is evident from the radial distributions of m in Figs. E.65 – E.67 in Appendix E), because less molecule means more probability of switching for CCs. This behaviour would not change qualitatively if we extended the domain of resolution of the diffusion equation, possibly to the whole lattice. In the case of lower values of D ($D < 10\,\mu\text{m}^2/\text{s}$) the molecule tends to travel shorter distances, and so it is usually concentrated in the place where it has been produced (we can see in Figs. E.65 – E.67 in Appendix E that the radial profiles have peaks located exactly where also the distributions of CSC fraction have peaks, while for $D > 10\,\mu\text{m}^2/\text{s}$ the profiles are parabolic monotonically decreasing), thus it is less subject to be transported to the border and die.

Unlike the cellular growth rate and the fraction of CSC, the spatial organization of

[5] Remember that for any time, if the radius of the tumour (distance of the farthest cell from the centre of the lattice), the domain in which we solve the diffusion equation is a circle of radius $r_{max} + \delta$, with $\delta = 4$), and we fix the amount m of molecule to be zero at the border of the circle (Dirichlet conditions) and outside of it.

CC and CSC depends markedly on the properties of the diffusing molecule (cf. Fig. 2.15, and Appendix E, Figs. E.16 – E.63 for more examples). The most apparent feature of molecule–controlled switching is the accumulation of CSC at the border of the tumoral mass. This effect increases with the diffusion coefficient D of the molecule and is absent if no switching applies. Moreover, the same plot suggests that the peak at the border does not grow monotonously with time, but reaches a maximum (in time) and then converges to a stationary value.

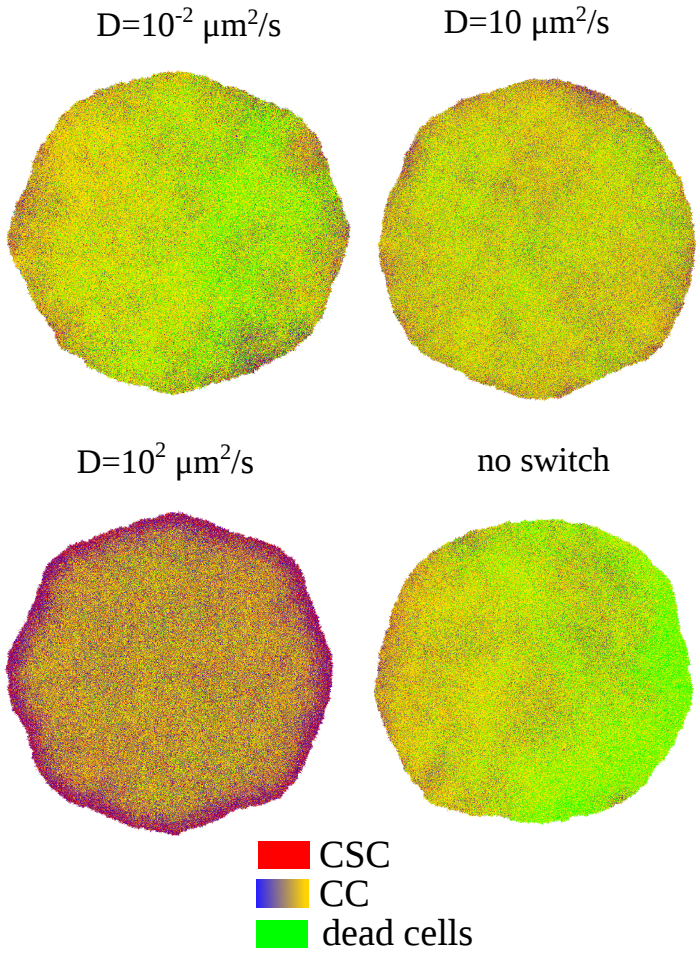

Figure 2.15: The spatial arrangements of CC of different ages and of CSC after 30 days at different values of the diffusion constant of the molecule and without switch.

2.4.3 Enrichment of CSC at the border

The effect of the enrichment of CSC at the border can be better quantified from the radial distribution of the fraction of CSC, a typical example of which is displayed in Fig. 2.16 a (more examples are shown in Figs. E.64 – E.69 in Appendix E). When D is large, the fraction of CSC at the border of the tumoral mass reaches a higher peak than at smaller values of D, while no enrichment at all is observed in absence of switching. This effect is not due to a different characteristic time at which the peak grows, as shown in Fig. 2.16 b, where one can notice that the height of the border peak at small D is inferior to the others at any time. The enrichment in CSC concentration and its location is thus preserved in the dynamics, with the exception of the case with $\sigma = 0$ (where there is no persistent enrichment in CSC concentration at the border, but rather fluctuations in the radial distribution of CSC concentration, see Appendix E).

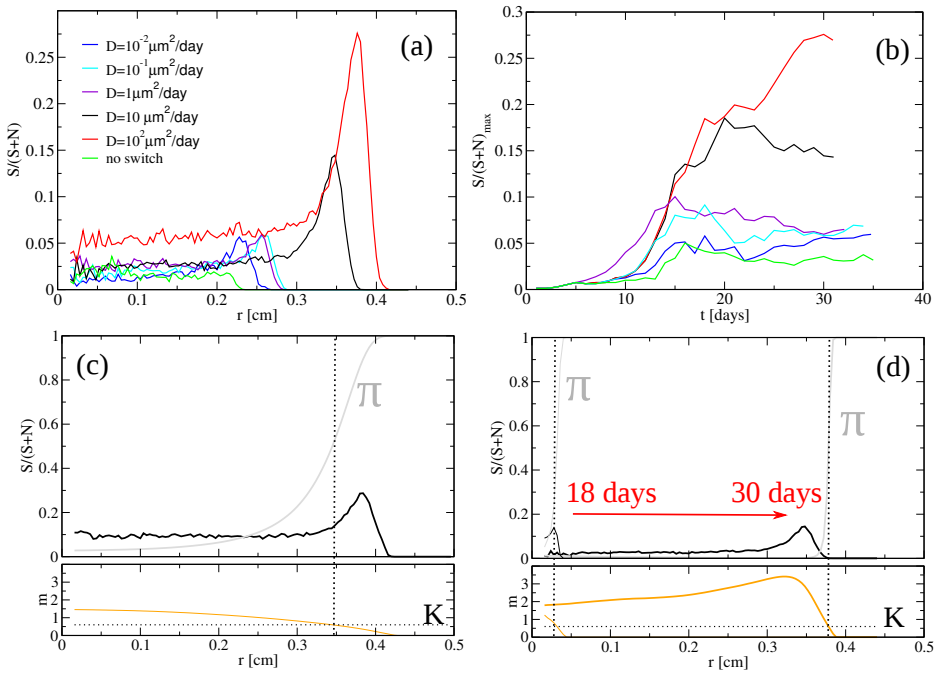

Figure 2.16: (a) The fraction of CSC as a function of the radius of the tumoral mass (upper left panel) at 30 days with $\tau = 4$ d for different values of D and for the model without switching. (b) The height of the maximum reached by the fraction of CSC as a function of time. (c) The black curve displays the fraction of CSC as a function of the distance r from the center of the tumour for $D = 10^2$ μm^2/s; the gray curve is the switching factor π that depends on the local concentration of molecule m, displayed in orange, and on the equilibrium constant K, displayed as a horizontal dotted line. Where $m = K$ then $\pi = 1/2$, and this point is highlighted with a vertical dotted line. (d) The same as in panel (c), but simulated at $D = 10$ μm^2/s; in this case, besides the final distributions at $t = 30$ days, we show with thinner lines the distributions at $t = 18$ days.

In all the previous simulations we used the parameters reported in Tab. 2.2 (which provided a quite good fit with the available data), with the said caveat for Hill's threshold K. In order to minimally test the robustness of the spatial patterns for reasonable alternative values of the parameters (note that the simulation are computationally very

expensive), we simulated the same system with a different value of $\mathbf{k}_m$ (the rate of decay of the signal molecule), under the reasonable assumption that if the switching-inhibitor molecule is secreted into the extracellular environment through vesicles then of course it would be less subject to self-degradation. The results of these simulations for $\mathbf{k}_m = 0.5\,\text{day}^{-1}$ (instead of $\mathbf{k}_m = 4.0\,\text{day}^{-1}$, as in the previous simulations) show that the time evolution of the fraction of CSC, the growth rate of the tumour and, most importantly, the radial distribution of CSC are qualitatively the same as in previous simulations (Figs. E.70 – E.73 in Appendix E).

To get an insight into the appearance of the peak at the border, we plotted in Figs. 2.16 c and d the radial profile of the fraction of CSC at 30 days together with that of diffusing molecule and the switching factor $\pi(r,t)$

$$\pi(r,t) = \frac{m^h(r,t-\tau)}{K^h + m^h(r,t-\tau)} \tag{2.20}$$

that controls the switching probability of CCs for two utmost values of the diffusion coefficient D. At $D = 10^2$ $\mu\text{m}^2/\text{s}$ and $t = 30$ days, the front of the cellular mass is well beyond the radius at which $\pi \gg 0$. The reason is that the distribution of molecule m (orange curve decays smoothly from the center of the system) and assumes the value K (i.e., the switching inhibition constant, at which $\pi = 1/2$) at quite small values of r. Consequently, a large amount of cells near the border of the tumoral mass can undergo phenotypic switching, resulting in the enrichment shown in Fig. 2.16 a. This enrichment increases with time. On the contrary, at lower values of D, the long–term behaviour is different. For $D = 10$ $\mu\text{m}^2/\text{s}$, for example (see Fig. 2.16 d), all cells are in the region where the molecule m is abundant and so inhibits consistently phenotypic switching, causing $\pi \ll 1$. However this is not the case at all times: during an initial period (18 days in Fig. 2.16 d), cells can overcome the front of the molecule, increasing transiently the enrichment at the border: in fact the molecule diffuses in space at a distance r that scales with time as $r \sim \sqrt{D}\cdot t^{\frac{1}{2}}$, while cells grow initially at a super-exponential rate (Figs. E.9 – E.10 in Appendix E). While the two processes of cellular growth and molecule diffusion are coupled, it is evident that, especially for lower values of D, cellular mass can reach a certain distance r before the concentration of molecule at that distance has reached a level necessary to inhibit most of the switching (diffusing from the inner regions of the tumour, where it is supposed to be present, at least initially, in larger numbers).

The height of the border peak of CSCs, after cells overcome the front of the molecule, then decreases to (or reaches from below) a stationary value (Figs. E.74 – E.75 in Appendix E), meaning that the cellular mass grows in size with the border constantly enriched. What is then the reason for the long–time enrichment of CSC at the border of the tumour when $\pi \sim 0$? The answer is that the initial enrichment, caused by the transient lack of molecule at the border, is sustained by the normal cell duplication, without the need of further switching. This can be understood from a simple argument.

When phenotypic switching does not apply, a simple model that captures the fact that CSCs duplicate into other CSCs with rate $\epsilon = k_{2S} - k_{2N}$ and into both CSCs and CCs approximately with rate $K = k_{2S} + k_{2N} + k_{1N1S}$ and CCs duplicate into CCs with rate k_d, pushing the others to make room, can be obtained defining the amount $S(r,t)$ of CSCs in a radial bin of width Δr. Assume that the system starts from a distribution of CSCs $S(r,0) = S_0(r)$ that is peaked at the border. The dynamics of $S(r,t)$ can be roughly described as

$$\frac{\partial S(r,t)}{\partial t} = \left[K\frac{1}{\Delta r}\int_0^{r-\Delta r} dr'\, S(r',t) + k_d\frac{1}{\Delta r}\int_0^{r-\Delta r} dr'\, N(r',t)\right]\cdot[S(r-\Delta r,t) - S(r,t)] + \\ + \epsilon S(r,t). \tag{2.21}$$

stating that the variation of S at point r is proportional to the total amount of CSCs and CCs within r, that duplicate respectively with rates K and k_d pushing the others in radial direction, and to the amount of CSC in the previous bin $r-\Delta r$, that can be pushed to r. Applying a mean–field approximation $\int_0^r dr'\, S(r',t) \approx \overline{S}\cdot r$ and $\int_0^r dr'\, N(r',t) \approx \overline{N}\cdot r$, where $\overline{S}$ and $\overline{N}$ are the average number of CSCs and CCs per bin (which, except in the peak region, tend to be constant and reach a steady state in the long time).

Thus, assuming that Δr is small, the dynamics can be written as a partial differential equation

$$\frac{\partial S(r,t)}{\partial t} = -\epsilon' r\cdot\frac{\partial S}{\partial r} + \Delta\epsilon\Delta r\frac{\partial S}{\partial r} + \epsilon S(r,t), \tag{2.22}$$

where $\epsilon' \equiv K\overline{S} + k_d\overline{N}$ and $\Delta\epsilon \equiv \epsilon' - \epsilon$. Given the initial condition, the solution has the form (see Appendix F)

$$S(r,t) = e^{\epsilon t}\cdot S_0\left(\frac{\epsilon' r - \Delta\epsilon\Delta r}{\epsilon'}\mathrm{e}^{-\epsilon' t} + \frac{\Delta\epsilon\Delta r}{\epsilon'}\right). \tag{2.23}$$

As it is proven in Appendix F, the initial values of the solution are transported along a family of curves (in this case, exponential curves), called *characteristics curves*, according to a mechanism typical of this kind of equations (hyperbolic p.d.e. of the first order, also known as transport equations). This means that a peak that is initially at the border r_b of the tumoral mass is transported exponentially in time in the radial direction, along the curve

$$\exp[-(K\overline{S} + k_d\overline{N})t]r - r_0 = C \tag{2.24}$$

(where C = constant), as observed in the simulation (cf. Fig. 2.17). The radius r_{tumour} of the total mass (i.e. the measure of the support of the solution (2.23)) also increases exponentially, in accordance with Fig. 2.17 (a sketch of the evolution of the profile of radial number of CSCs is shown in Fig. 2.18).

The solution (2.23) provides an explanation of the evolution of the peaks of CSCs density. An initial, consistent enrichment of CSCs at the border of the cellular mass is formed, with the mechanism explained at the beginning of this section, when controlled phenotypic switching is active: cellular growth is faster than the diffusion of the molecule with its inhibitory effect and thus cells at the border lies in a region of space where they can massively switch to CSC state. Instead, if $\sigma = 0$ (no phenotypic switching), tumour subpopulations' growth is driven exclusively by cellular replication, and thus there is not a deterministic mechanism which can allow a persistent and significant enrichment of CSCs isotropically at the border or in some inner region (observed peaks in the radial distribution of CSCs density are of small amplitude and should be considered random fluctuations from the average density), while they can form isolated clusters of higher density in some regions.

Once a consistent and significant enrichment (not a statistical fluctuation in the density) at the border of the tumour mass is formed (in the case when switching is active),

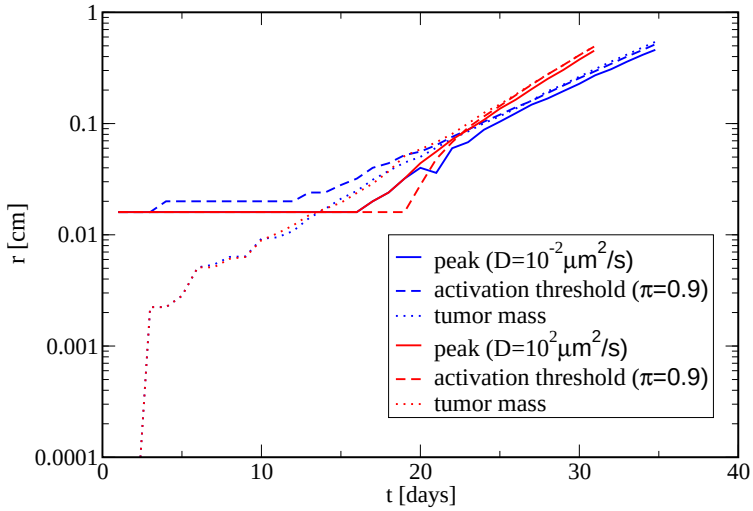

Figure 2.17: The position r of the peak of CSC concentration for $D = 10^{-3}\mu\text{m}^2/\text{s}$ (red lines) and $D = 10^{2}\mu\text{m}^2/\text{s}$ (blue lines). The dashed lines are the position of the activation threshold, defined as the value of r at which the concentration of the molecule gives $\pi = 0.9$, i.e., at which switching inhibition drops to very low values. The dotted lines are the overall radius of the tumoral mass.

the peak of CSCs is transported, because of cellular replication in the inner regions of the tumour mass. The height of the peak tends to increase, because cells at the border divide and are pushed in the outer region where $\pi \sim 1$, and thus they can freely switch *en masse*. As time goes by, diffusion manages to keep up with cellular growth (as shall be explained in more detail later) and from then on the peak lies in the region where $\pi \sim 0$. In this region phenotypic switching is almost totally inhibited and thus the spatial dynamics of CSCs can roughly be described by (2.22). In this case, the peak is sustained purely by normal cellular duplication (and not by invasion of regions where $\pi \sim 0$): in the solution (2.23)

$$S(r,t) = e^{\epsilon t} \cdot S_0\left(\frac{\epsilon' r - \Delta\epsilon\Delta r}{\epsilon'}\mathrm{e}^{-\epsilon' t} + \frac{\Delta\epsilon\Delta r}{\epsilon'}\right) \tag{2.25}$$

the term $S_0\left(\frac{\epsilon' r - \Delta\epsilon\Delta r}{\epsilon'}\mathrm{e}^{-\epsilon' t} + \frac{\Delta\epsilon\Delta r}{\epsilon'}\right)$ represents the effect of transport of the initial profile $S_0(r)$ of CSCs (with a peak at the border) due to growth of the inner cellular mass, while the multiplicative reaction term $e^{\epsilon t}$ accounts for the in–loco replication of CSCs. The transport part of the solution tends to maintain the peculiar shape of the initial radial profile of CSCs (the peak at the border), while causing the support of the solution to grow exponentially (reflecting the observed exponential growth of the tumour mass). Since the multiplicative reaction term $e^{\epsilon t}$ is constant at a given time and monotonous in t, it also preserves the peak of CSCs at the border during the evolution.

Thus the peak continues to advance along an exponential characteristic curve (2.24)

of the hyperbolic equation (2.22), remaining located in the region where $\pi \sim 0$, preserving its shape and also, as can be seen in Fig. 2.16, its height. Our model does not explain the fact that the height of the peak reaches a steady state, but we remark that this is just an approximated qualitative argument; the main approximation in such a space–dependent rate equation is that we neglect that in each bin the maximum value of S is constrained by the the fact that each site can be occupied by only one CSC. We also remark that our model (2.23) describes the evolution of the radial distribution of the number of CSCs, and not of their local fraction $S/(S+N)$, while all our previous analyses are based on this quantity. However we note that a radial distribution of CSC presenting a huge peak at the border would result in a radial density distribution with the same qualitative shape, and vice versa.

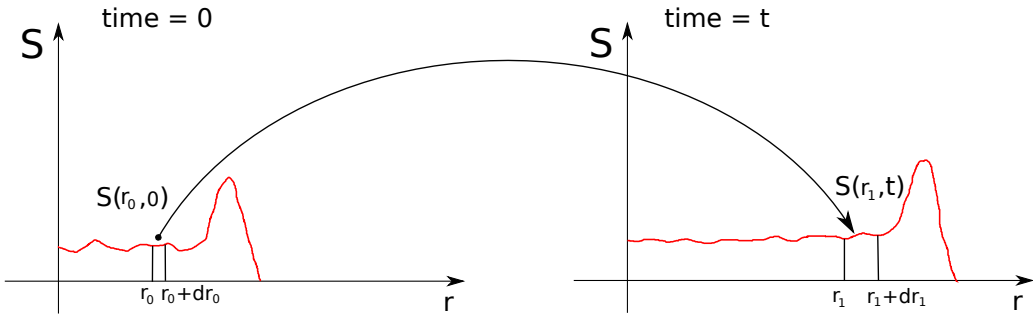

Figure 2.18: Time evolution of the radial distribution of the number of CSCs according to (2.22).

The evolution of the peaks with respect to the diffusion coefficient D could be clarified by Fig. 2.17. We note that, for the lower value of D ($D = 10^{-2}$ μm^2/s), the peak of the CSC fraction is behind the activation threshold for switching (the radius at which $\pi = 0.9$, and behind which switching is rapidly inhibited) [6]; thus switching ceases rapidly to have a significant effect and to enrich too much the fraction of CSC at the border, while transport dynamics described above translates the peak forward keeping it approximately [7] constant. For $D = 10^2$ μm^2/s, the peak goes behind the activation threshold only at about day 25 of the simulation, so for the first 25 days there is persistent enrichment of CSC at the border and increase in the height of the peak (see Figs. E.74 – E.75 in Appendix E). We can explain the relative positions of the peak and the activation threshold recurring to what said in 2.4.2. If D is large the molecule is more likely to travel long distances from the point where it was produced and reach the border of the domain, where it dies: thus its radial distribution will likely be parabolic monotonically decreasing. On the other hand, if D is small the molecule will travel shorter distances and will tend to accumulate where it was produced, forming a peak at the border where also CSC tend to concentrate. Confronting the two radial profiles of molecule, it is evident that in the case with lower D the activation threshold is generally located at larger distances from the centre of the lattice (relative to the tumour radius) than in the case with larger D.

[6] the constant positions in the first days of the dynamics and the fact that they seem to exceed the radius of the tumoral mass reflect the binning of the lattice. The lattice is divided into circular crowns, the most internal of which must have a bigger radius in order to include a sufficient number of points, so that the calculation of the average fraction of CSC in it is not excessively affected by fluctuations

[7] There are of course statistical spatial and temporal fluctuations and a residual switching factor.

2.4.4 Clustering of CSC

From the snapshots displayed in Fig. 2.15 it is apparent that for small and intermediate values of D CSC tend to cluster together, especially at the enriched border. This effect seem to disappear at large D, at which CSC are rather symmetric for rotations. In absence of switching, some clustering seems present as well.

To get a more quantitative insight in this, we performed a clustering analysis of CSC based on their spatial density, using as algorithm of reference the one developed from (130).

In Fig. 2.19 we show the multiplicity of clusters found and the distribution of their density (for the complete set of clustering analysis, see Appendix E, Figs. E.76 – E.95). Without switching, the system display a small number of high–density clusters (mainly, though not necessarily at the border). A similar behaviour is observed with switching at low diffusion constant. Increasing D, we observe many low–density clusters, suggesting that CSCs decrease their tendency to partition in clusters. What is also evident is that, for lower values of D, the distribution of CSC at the border becomes more asymmetric, as it is shown in Fig. 2.20, which reports, for different values of D, the angular distribution of the fraction of CSC, obtained dividing the border of the final plots of tumour mass (a circular crown whose radius is $r_{crown} = 0.2 \cdot r_{max}$) equally into 32 sectors and calculating the local fraction of CSC in each sector. It is evident that the distribution becomes more uneven as the value of D becomes lower. We also note that the main factor contributing to the anisotropy in the distribution of the CSC at the border is the value of D, and not the time delay τ, as the curves for $\tau = 0$ d and $\tau = 4$ d are quite similar.

Summing up, clustering seem to appear as a consequence of CSC duplication, while switching, which is more effective at large D (cf. Sect. 2.4.3), tends to make the distribution of CSC more uniform. This is not unexpected, since cell duplication is a local phenomenon, while switching is random in space, and can only be modulated in space by the molecule; however, the high diffusivity of the molecule makes its distribution to be quite uniform, suppressing most spatial patterns.

This observed effect about the partition of CSCs in clusters could suggest further experiments to test the validity of our hypotheses. Since what we have observed is that the most we suppress phenotypic switching, the most CSCs have a tendency to partition in high-density clusters at the border, biologists could compare the spatial patterns in the distribution of CSCs for a tumoral mass resulting from an initial CSC plated in a neutral medium and for one resulting from an initial CSC in a culture medium repleted with switching-inhibitor molecule (for example, one resulting from a long-time growth of similar cells, collected and concentrated), possibly with continuous enrichment of such medium during the experiment. At the end of the experiment the cellular mass resulting from the non-neutral medium should show a higher degree of CSC clusterization at the border. As usual, the main problem with this approach is whether the difference is so relevant that it can be clearly detected in the lab.

The results obtained in the case for lower values of D are consistent with recent experiments by the Pajonk group at UCLA (132), who observed that, for sufficiently large tumour spheroids in vitro, the CSC population becomes localized in discrete clusters near the tumour boundary, resulting in a heterogeneous distribution of cells. They are also consistent with the more complex models developed by Youssefpour et al. (131), who using a partial differential equation approach to simulate the spatio-temporal dynamics of cell lineage in solid tumours,were able to observe complex pattering with CSCs being predominantly located in individual clusters at the outer rim of the total population, in response to a variety of cellular feedback mechanisms and oxygen tension.

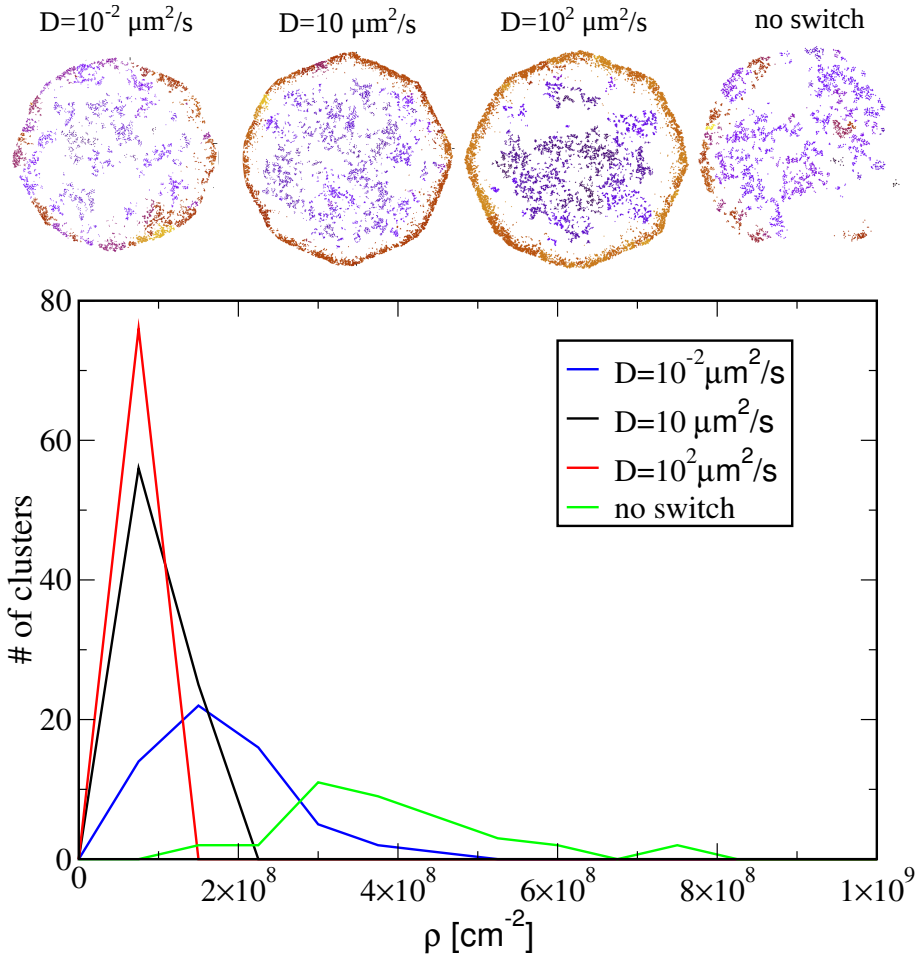

Figure 2.19: (Upper panel) The clusters of CSC, subtracted their density halo and coloured differently for each cluster, at different values of D and without switching. The clustering was done with the algorithm of ref. (130) using as parameters $k - \mathrm{NN} = 5$ and $Z = 1.8$. (Lower panel) The histogram of densities of the different clusters.

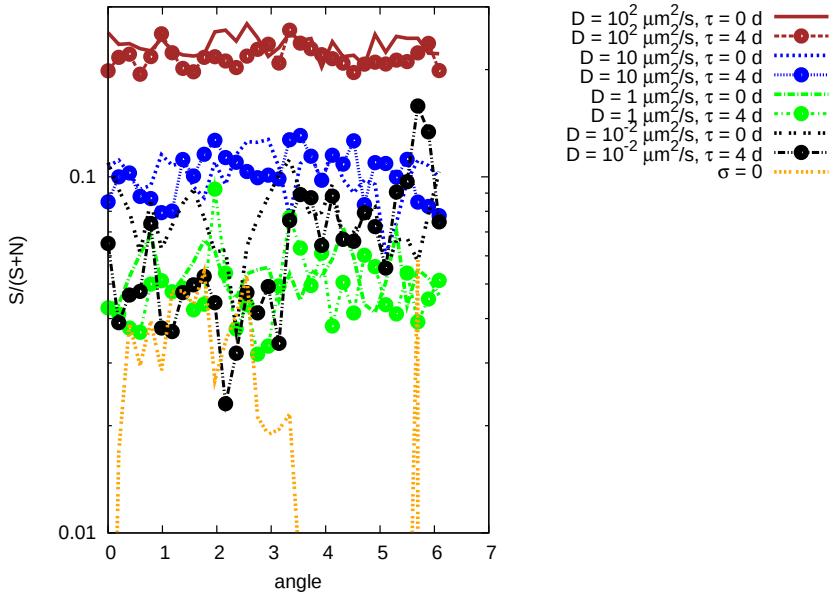

Figure 2.20: Angular distribution of fraction of CSC for 32 sectors of circular crowns ($r_{crown} = 0.2 \cdot r_{max}$) in which the border of final spatial plots of tumour mass has been divided, for some values of the diffusion coefficient D.

2.4.5 Discussions and conclusions

In this work we proposed a model for the evolution of cancer cells from a melanoma tissue plated in vitro as observed in (71). The main feature of the dynamics observed in the experiments reported in (71) is the observation of a large overshoot (up to 25 %) in the fraction of cancer stem cells (CSC), starting from a population where the fraction of CSC has been reduced under the 1 % of the total. Our model for the dynamics of the cell populations, written in terms of rate equations, was grounded in the theory of the aberrant hierarchy of CSC (which assumes that all cancer cells in a tumour are derived from a small subpopulation of cells with stem cell-like properties, with an unlimited capacity to proliferate and to generate normal, differentiated cancer cell (CC)) and included the key hypothesis, still not universally accepted, that normal cancer cells could revert back to CSC-state, undergoing what is called phenotypic switching. The first distinctive and original feature of our model is a dynamic feedback regulation mechanism of the process of phenotypic switching, given by an underlying field of a switching-inhibitor molecule produced by the CSC themselves. This regulatory mechanism is consistent with the experimental results in (71) and the results of a successive series of experiments performed by the same group (which are in contrast with the model proposed by the authors in (71)). The second distinctive characteristic of our model is a time delay of some days in the process of switching and its regulation.

Our model(2.11) with the features described above provides a clear regulatory mechanism for cellular dynamics and can qualitatively reproduce, for at least one realistic set of the parameters, the experimental results observed in (71), in particular the overshoot in the fraction of CSC and the dependence of the fraction of CSC at the overshoot on the fraction of CSC in the initial population 2.7.

Given the new opportunities provided by recent techniques of immunohistochemistry, which offer the possibility of investigating the localization of molecules and cells in a tissue (in our specific case, it would allow to localize CSC measuring the spatial distribution of its biomarkers), we have studied an extension of the model in which the spatial degrees of freedom appear explicitly, where cells are points on a regular 2D lattice and the molecule produced by the CSCs diffuse according to Fourier's law. Apart from observing patterns in the spatial distribution of cancer cells, this extended models allows us to investigate the role of diffusion in the growth of the tumoral mass.

Spatial simulations are performed first starting from a mass of negative cells, which shows that also in the spatial model an overshoot appears (from initial conditions with a low concentration of CSC, see Fig. 2.13). Successive spatial simulations are performed starting from a single CSC. The fraction $S/(S+N)$ of CSC decreases quite smoothly to a stationary value of few percent, slightly higher at large D, slightly lower at small D (Fig. 2.16 a). The main feature resulting from these spatial simulations is that CSC tend to accumulate at the border of the tumoral mass, with a strong dependence on the value of the diffusion coefficient D (see Fig. 2.14): when D is large, the radial distribution of the fraction of CSC reaches a higher peak at the border of the tumoral mass than at smaller values of D, while no enrichment at all is observed in absence of switching (just stochastic fluctuations around the mean fraction). We also note that, especially for smaller values of D, the peak at the border does not grow monotonously with time, but reaches a maximum (in time) and then converges to a stationary value. In order to explain this phenomenon and to predict the future development in the spatial organization of the tumour as it grows further, we proposed a simple mathematical model. As the tumoral mass grows, more and more part of the cellular mass tends to be in the region of space where, due to abundance in the concentration of the switching-inhibitor molecule, phenotypic switching is minimally active: this mechanism is much quicker in the case of low D (Figs. 2.16, 2.17). If we suppose that switching is mostly inhibited, a first-order hyperbolic PDE (2.22) describes the transport of CSC during the growth of the tumour, due to local duplication and pushing from cells duplicating in the inner regions of the tumour. We proved that such dynamics preserves the height of the peak of the initial CSC fraction profile, which is transported exponentially in time in the radial direction, and also that the total fraction of CSC is stationary: this is in perfect accordance with the results of the simulations (Figs. 2.16, 2.17).

Finally, we showed that for lower values of D CSC tends to be localized in high-density discrete clusters near the tumour boundary: since this is consistent with recent experiments (132), and since the lower values for the CSC stationary fraction are obtained for low D, we conclude that under this framework the most likely scenario is for D to assume relatively low values. This means that it if the signalling unknown molecule is effectively miRNA (as supposed by the authors of (71)), it is likely that it is secreted into the extracellular environment through vesicles, whose dimensions can be much larger than that of the simple miRNA molecule.

Further development of the present work could be concerned with the extension of our spatial model to three dimensions, in order to simulate the growth of tumor spheroids (which will require the development of some tools to better analyse the spatial distribution of cells in 3D) and the development of more complex models (starting from the rate equations), reflecting further degrees of heterogeneity in the cellular compartments of CSCs and CCs, and more complex signalling networks for the control of cellular phenotype and proliferation potential. Building such kind of models shall of course rely heavily on results from ongoing and future experimental tests, and hypotheses arising from them. A first and simple example of hypothesis to be tested could be

that of the imperfection of biomarkers for CSCs with relation to the differentiation of CSCs. Emerging evidence supports in fact the conclusion that early generations of CCs share some biomarkers with CSCs (perhaps with decaying statistical correlation as the number of generations in a small range increases) (116; 122; 133). Our present model could be updated to reflect this evidence, to test whether such hypothesis could provide better accordance with experimental data and also some insight into the important question of the reliability of biomarkers for the isolation of CSCs subpopulation.

Appendices

APPENDIX A

Dependence of oscillations on parameters.

In this appendix we collect the plots representing the dependence of Ω (number of oscillations) on the parameters of the system (1.17) for all the time delays τ employed in the simulations ($\tau = 500, 1000, 1500, 2000, 2500$ s) and for temperature $T = 1$. The density of points in each region of the plots reflects the probability that the delay equations display a specific value of Ω for the value of the parameters that define that region.

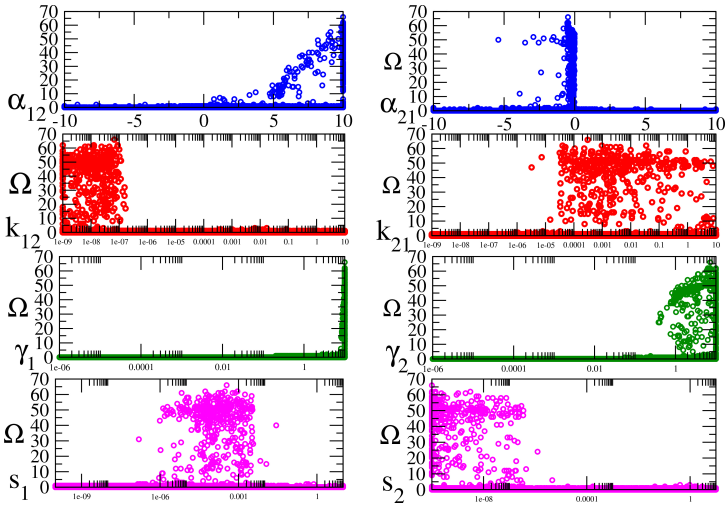

Figure A.1: The values of the number of oscillations Ω associated with the points of parameter space sampled by the simulation carried out at $\tau = 500$ s and $T = 1$ projected, respectively, onto the parameters α_{ij} , k_{ij} , γ_i and s_i . To make the results more readable, α_{ij} is plotted in a linear scale, while the other parameters in logarithmic scale.

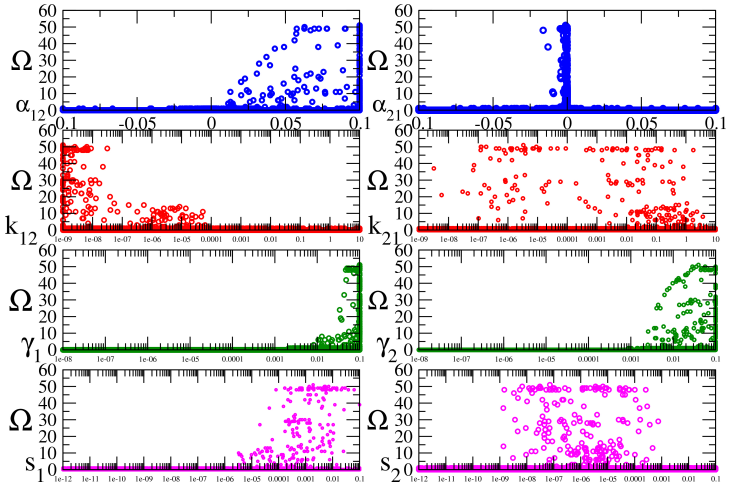

Figure A.2: The values of the number of oscillations Ω associated with the points of parameter space sampled by the simulation carried out at $\tau = 1000$ s and $T = 1$ projected, respectively, onto the parameters α_{ij} , k_{ij} , γ_i and s_i . To make the results more readable, α_{ij} is plotted in a linear scale, while the other parameters in logarithmic scale.

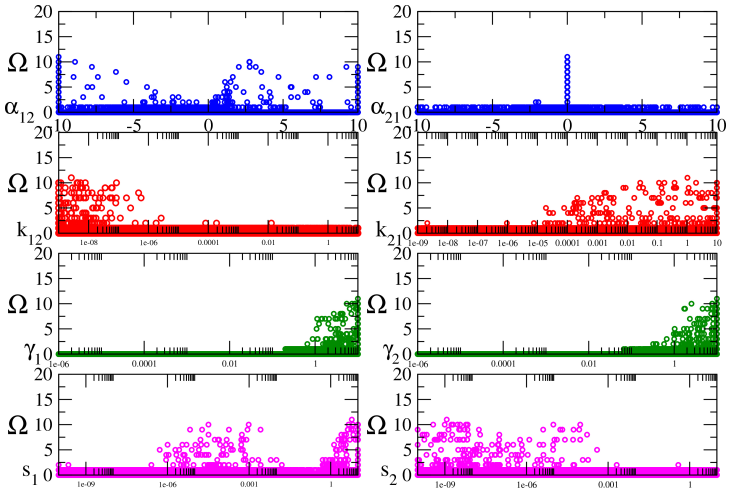

Figure A.3: The values of the number of oscillations Ω associated with the points of parameter space sampled by the simulation carried out at $\tau = 1500$ s and $T = 1$ projected, respectively, onto the parameters α_{ij} , k_{ij} , γ_i and s_i . To make the results more readable, α_{ij} is plotted in a linear scale, while the other parameters in logarithmic scale.

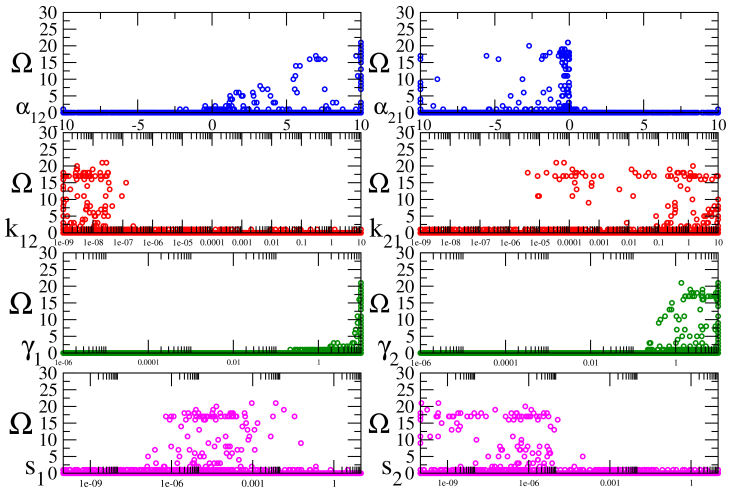

Figure A.4: The values of the number of oscillations Ω associated with the points of parameter space sampled by the simulation carried out at $\tau = 2000$ s and $T = 1$ projected, respectively, onto the parameters α_{ij} , k_{ij} , γ_i and s_i . To make the results more readable, α_{ij} is plotted in a linear scale, while the other parameters in logarithmic scale.

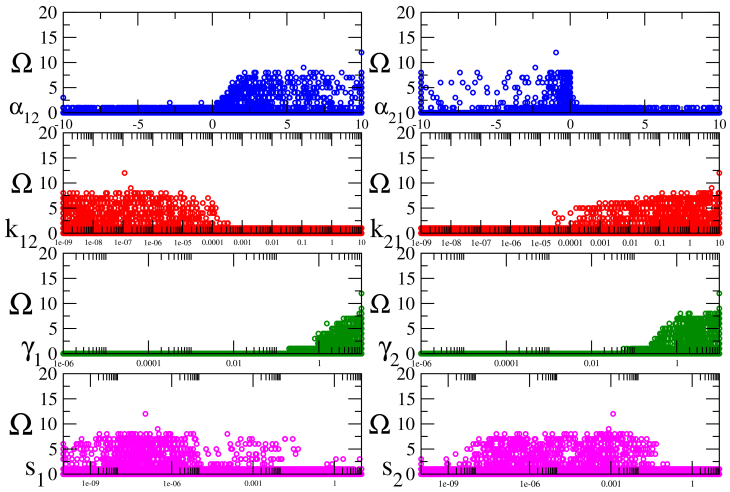

Figure A.5: The values of the number of oscillations Ω associated with the points of parameter space sampled by the simulation carried out at $\tau = 2500$ s and $T = 1$ projected, respectively, onto the parameters α_{ij} , k_{ij} , γ_i and s_i . To make the results more readable, α_{ij} is plotted in a linear scale, while the other parameters in logarithmic scale.

APPENDIX B

Steady state values of concentration c of CSC.

The stationary ratios of S^* and N_j^* (and thus the stationary value of CSC concentration c) of the rate equations (2.11) can be found setting the time derivatives to zero and defining $\epsilon = k_{2S} - k_{2N}$ and $\lambda = 2k_{2N} + k_{1N1S}$, thus obtaining

$$\begin{aligned} N_0^* &= \frac{\lambda}{k_d + \epsilon} S^* \\ N_1^* &= \frac{\lambda}{k_d + \epsilon} \frac{2k_d}{k_d + \epsilon} S^* \\ N_i^* &= \frac{\lambda}{k_d + \epsilon} \left(\frac{2k_d}{k_d + \epsilon} \right)^i S^* \quad \forall i <= g. \end{aligned} \tag{B.1}$$

Thus,

$$N^* = \frac{\lambda}{k_d + \epsilon} S^* \sum_{i=0}^{g} \left(\frac{2k_d}{k_d + \epsilon} \right)^i. \tag{B.2}$$

and one obtains

$$c^* = \frac{S^*}{S^* + N^*} = \frac{1}{1 + \frac{\lambda}{k_d + \epsilon} \sum_{i=0}^{g} \left(\frac{2k_d}{k_d + \epsilon} \right)^i} \tag{B.3}$$

If $k_d = \epsilon$, then $\frac{N_i}{Nj} = 1 \; \forall i, j$, and the CC are equally distributed among generations.

These finding could be useful to determine the initial conditions for our simulations, since, assuming the validity of our model, cancer cells sorted in experiments to be plated and grown are taken from a mature tissue, and therefore their distributions among generations should be consistent with the asymptotic ratios calculated here (of course the more so the more cells are sorted and plated).

In order to test if initial conditions consistent with (B.1) could lead to qualitatively different results in the solution of our model, we solved once again the rate equations (2.11) distributing the initial CC among generations according to the asymptotic ratio. The results of the simulations can be seen in Fig. B.1, from which it is evident that this alternative distribution of the CC among generations does not affect the dynamics of the system qualitatively.

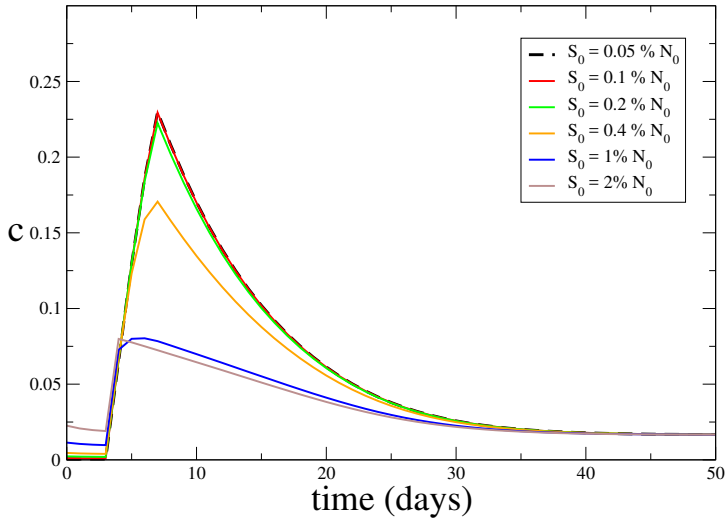

Figure B.1: The dynamics of the concentration $c(t)$ of CSC with different choices of the initial value $c(0)$ ($N_0 = 10^6$, $m_0 = 0$) and initial CC cells distributed among generations according (B.1).

APPENDIX C

Long-time dynamics of the rate equations

C.1 Condition for monotonous and unbounded increase of the total number of cells

We want to show that a necessary and sufficient condition for unlimited tumour growth for our model (2.11) is $\epsilon \geq 0$. In order to do this we must prove that for this values of the parameter ϵ the total number of cancer cells $S + N$ diverges.

Consider the simplified version (without generations of CC cells and without delay, which does not affect the value of the fixed points, as represented in Fig. C.1) of the system (1) ($\epsilon = k_{2S} - k_{2N}$, $p = 2k_{2N} + k_{1N1S}$)

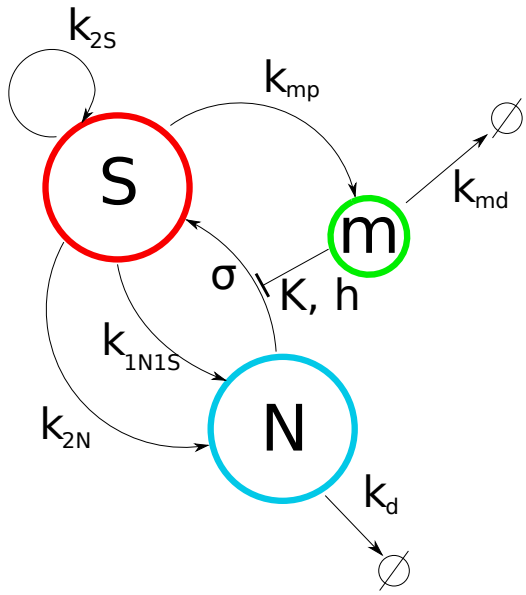

Figure C.1: Reaction network for the simplified mean–field model, involving cancer stem cells (S), normal cancer cells (N) and inhibitor molecule m.

$$\begin{cases} \dot{m} &= k_{mp}S - k_m m \\ \dot{S} &= \epsilon S + \sigma N\left(1 - \frac{m^h}{K^h + m^h}\right) \\ \dot{N} &= p \cdot S - \left(k_d + \sigma\left(1 - \frac{m^h}{K^h + m^h}\right)\right)N \end{cases} \tag{C.1}$$

The only fixed point of the system is $(m, S, N) = (0, 0, 0)$. In the case $\epsilon > 0$ with the simple estimations $\dot{S} > \epsilon S$ and $\dot{N} > p \cdot S - (k_d + \sigma)N$, wh have that the solutions of system (C.1) are bounded from below by the corresponding solutions of the following system (which has only the fixed point $(m, S, N) = (0, 0, 0)$, as (C.1))

$$\begin{cases} \dot{m} &= k_{mp} S - k_m m \\ \dot{S} &= \epsilon S \\ \dot{N} &= p \cdot S - (k_d + \sigma) N \end{cases} \tag{C.2}$$

The Jacobian matrix of the system made up of the last two equations (decoupled from the remaining one) is the following

$$\begin{bmatrix} \epsilon & 0 \\ p & -k_d - \sigma \end{bmatrix}$$

The product of the eigenvalues of the Jacobian matrix $\lambda_1 \cdot \lambda_2 = -\epsilon(k_d + \sigma)$ is negative and thus the eigenvalues are of different sign. Thus the fixed point $(m, S, N) = (0, 0, 0)$ is unstable and all the solutions are divergent, and so are the solutions of (C.1).

For the case $\epsilon = 0$, we can see that S diverges, since $\dot{S} = \sigma N \left(1 - \frac{m^h}{K^h + m^h}\right) \geq 0$, and so does the total number $S + N$ of cancer cells, since N is always positive (of course this argument is also valid for $\epsilon > 0$.

In the case with $\epsilon < 0$ the solutions of the system (C.1) are bounded from above by the corresponding solutions of the system

$$\begin{cases} \dot{m} &= k_{mp} S - k_m m \\ \dot{S} &= \epsilon S + \sigma N \\ \dot{N} &= p \cdot S - k_d \cdot N \end{cases} \tag{C.3}$$

The Jacobian matrix of the system made up of the last two equations (decoupled from the remaining one) is the following

$$\begin{bmatrix} \epsilon & \sigma \\ p & -k_d \end{bmatrix}$$

It is easy to see that the eigenvalues of the Jacobian matrix are complex conjugate with negative real parts. the corresponding solutions of system (C.3) are linear combinations of sines and cosines modulated by decreasing exponential functions. Therefore the solutions are definitely bounded from above (and also from below, since they need to be positive), and so are the solutions of system (C.1). In the case $\epsilon < 0$, therefore, the growth of cancer cells population is bounded.

A more refined analysis in the case $\epsilon < 0$ is possible. In this case, (C.1) has two fixed points, one being the usual $(m, S, N) = (0, 0, 0)$, and the other

$$(m, S, N) = \left(\frac{1}{r_{staz}} \frac{\sigma \cdot k}{-\epsilon} - k, \frac{k_{md}}{k_{mp}} \cdot \left(\frac{1}{r_{staz}} \frac{\sigma \cdot k}{-\epsilon} - k \right), \frac{k_{md}}{k_{mp} \cdot r_{staz}} \cdot \left(\frac{1}{r_{staz}} \frac{\sigma \cdot k}{-\epsilon} - k \right) \right),$$

where

$$r_{staz} = \left(\frac{S}{N} \right)_{staz} = \frac{k_d}{k_{2S} + k_{1N1S} + k_{2N}}. \tag{C.4}$$

The fixed point $(m, S, N) = (0, 0, 0)$ is globally attractive (extinction of the tumour) when

$$\sigma < \sigma_{crit} = \frac{-\epsilon k_d}{p + \epsilon} = \sigma_{crit.},$$

while the positive fixed point (C.4) is globally attractive when $\sigma > \sigma_{crit}$.

We can therefore conclude that a necessary and sufficient condition the total number of cancer cells $S + N$ to be an increasing and unbounded function of time is $\epsilon \geq 0$.

C.2 The case $\epsilon < 0$ and benign tumours

In the previous section, considering the simplified equations (C.1) we have come to the conclusion that a sufficient and necessary condition for unlimited tumour growth is $\epsilon \geq 0$. In the case $\epsilon < 0$, provided that the parameter σ is above a certain critical value, the system reaches a positive steady state (m^*, S^*, N^*). Once we rename

$$\left(\frac{S}{N}\right)_{staz} = r_{staz} = \frac{k_d}{k_{2S} + k_{1N1S} + k_{2N}}; \tag{C.1}$$

the values of (m^*, S^*, N^*) are given by

$$m^* = \left(\frac{k^h}{r_{staz}}\frac{\sigma}{-\epsilon} - k^h\right)^{\frac{1}{h}}.$$

$$S^* = \frac{k_{md}}{k_{mp}} \cdot \left(\frac{k^h}{r_{staz}}\frac{\sigma}{-\epsilon} - k^h\right)^{\frac{1}{h}}.$$

$$N^* = \frac{S_{staz}}{r_{staz}} = \frac{k_{md}}{r_{staz} \cdot k_{mp}} \cdot \left(\frac{k^h}{r_{staz}}\frac{\sigma}{-\epsilon} - k^h\right)^{\frac{1}{h}}.$$

Since in this case the mass of the tumour does not exceeds a finite value, we could think of this case as a very basic model for the benign version of the tumour under consideration. A benign tumour is a mass of cancer cells that lacks the ability to invade neighbouring tissue or metastasize. In the case of unbounded growth of the tumour (that is $\epsilon > 0$), which we associate with the malignant version of the tumour, it is inevitable cancer cells would spread into and invade nearby tissues, causing metastases. A benign tumour could instead stay dormant or, as the cells divide and accumulate mutations, undergoing phenotypical changes that cause the tumour to become malignant. In our case the tumour progression would be triggered by a mutation that causes the potential for symmetrical division into two CSC of a CSC (the rate k_{2S}) to exceed its potential for symmetrical division into two CC (the rate k_{2N}), since that would cause the transition from $\epsilon < 0$ ($k_{2S} < k_{2N}$) to $\epsilon > 0$ ($k_{2S} > k_{2N}$).

APPENDIX D

Supporting figures of Chap. 2

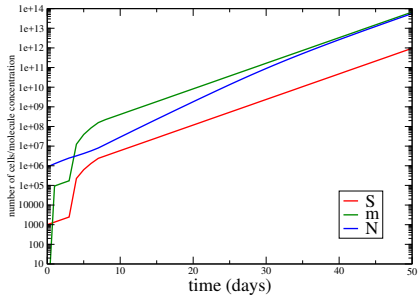

Figure D.1: Example of dynamics of S, N and m obtained solving numerically the rate equations.

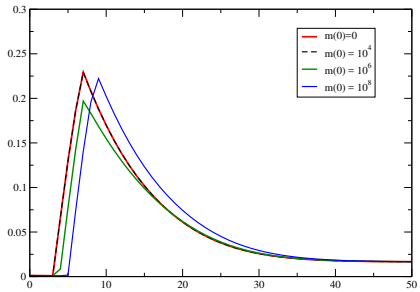

Figure D.2: Dynamics of $c(t)$ for $S_0 = 10^{-4}N_0$, varying the initial value $m(0)$.

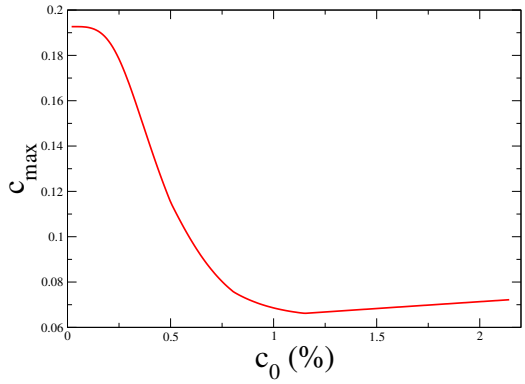

Figure D.3: The maximum c_{max} of CSC concentration depending on the initial concentration c_0 ($N_0 = 10^6$, $m_0 = 0$ for simulations carried out with the parameters listed in Table 1).

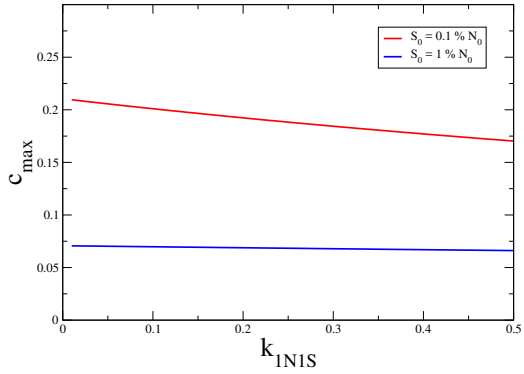

Figure D.4: The maximum value c_{max} as a function of k_{1N1S} ($k_{1N1S} = 0.2$ for simulations carried out with the parameters listed in Table 1).

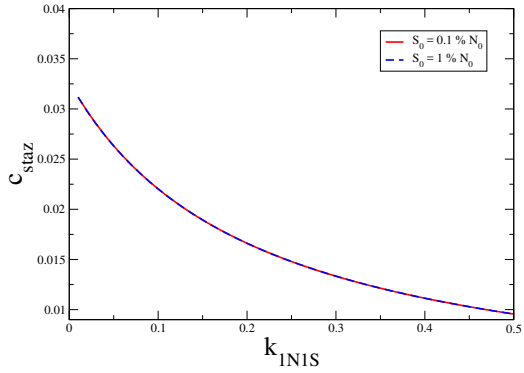

Figure D.5: The stationary value c^* as a function of k_{1N1S} ($k_{1N1S} = 0.2$ for simulations carried out with the parameters listed in Table 1).

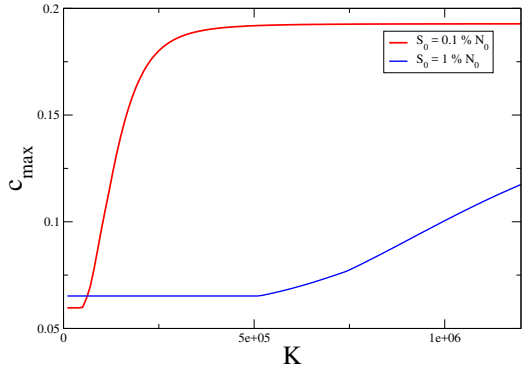

Figure D.6: c_{max} variation with respect to K ($K = 6 \cdot 10^5$ for simulations carried out with the parameters listed in Table 1).

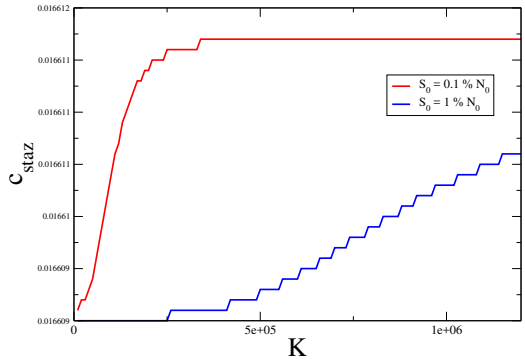

Figure D.7: c_{staz} variation with respect to K ($K = 6 \cdot 10^5$ for simulations carried out with the parameters listed in Table 1).

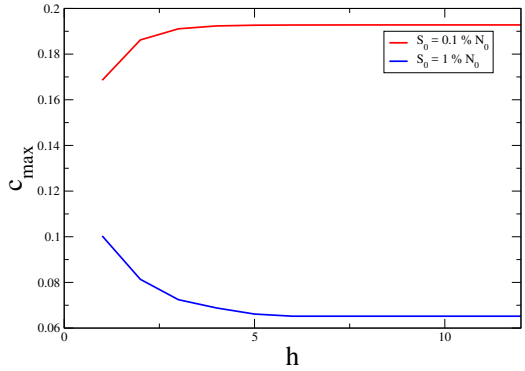

Figure D.8: c_{max} variation with respect to h ($h = 4$ for simulations carried out with the parameters listed in Table 1).

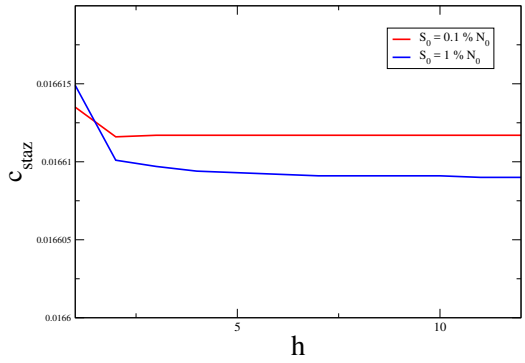

Figure D.9: c_{staz} variation with respect to h ($h = 4$ for simulations carried out with the parameters listed in Table 1).

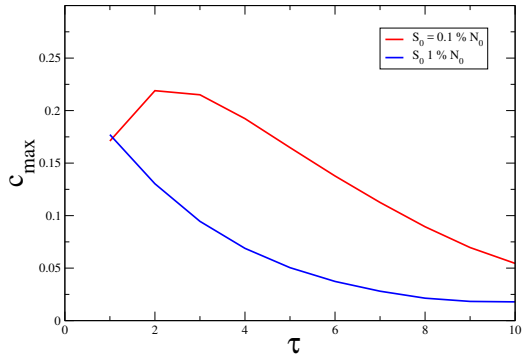

Figure D.10: c_{max} variation with respect to τ ($\tau = 4$ in Fig.**??**).

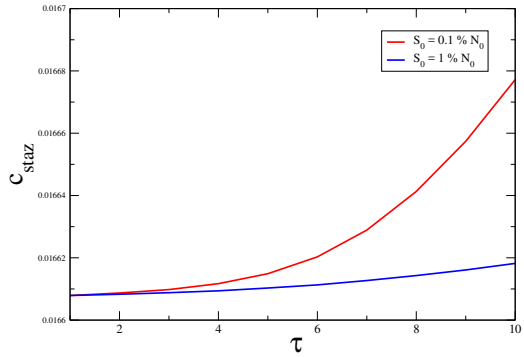

Figure D.11: c_{staz} variation with respect to τ ($\tau = 4$ for simulations carried out with the parameters listed in Table 1).

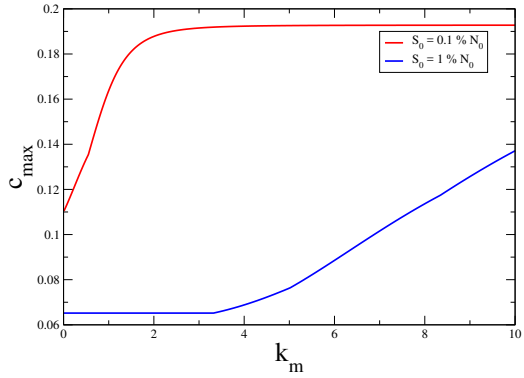

Figure D.12: c_{max} variation with respect to k_m ($k_m = 4.0$ for simulations carried out with the parameters listed in Table 1).

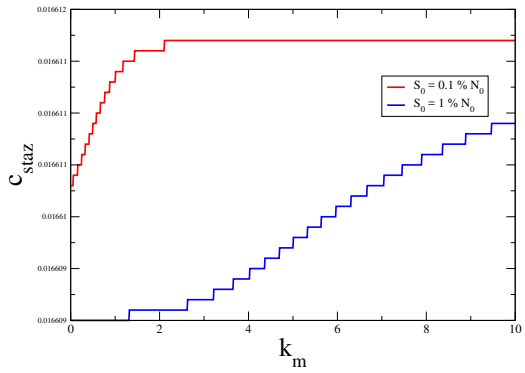

Figure D.13: c_{staz} variation with respect to k_m ($k_m = 4.0$ for simulations carried out with the parameters listed in Table 1).

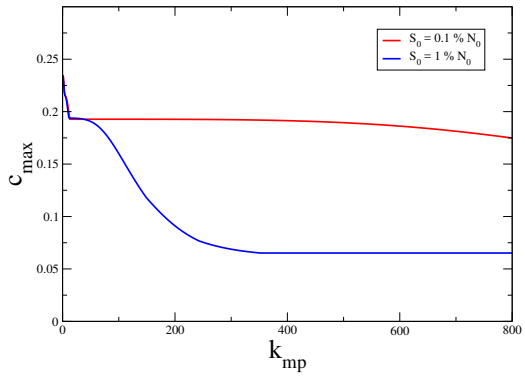

Figure D.14: c_{max} variation with respect to k_{mp} ($k_{mp} = 300.0$ for simulations carried out with the parameters listed in Table 1).

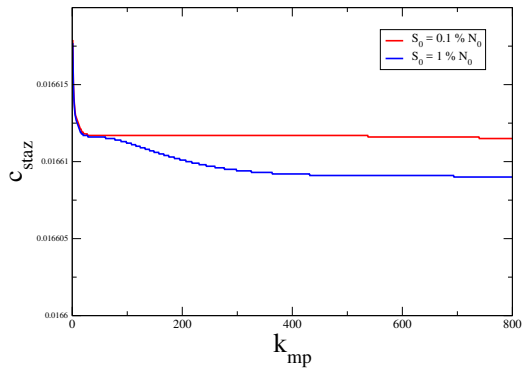

Figure D.15: c_{staz} variation with respect to k_{mp} ($k_{mp} = 300.0$ for simulations carried out with the parameters listed in Table 1).

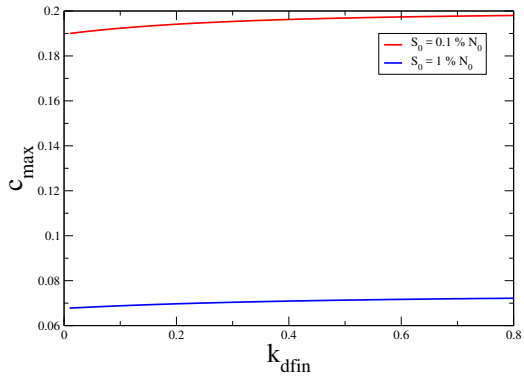

Figure D.16: c_{max} variation with respect to k_{dfin} ($k_{dfin} = 0.1$ for simulations carried out with the parameters listed in Table 1).

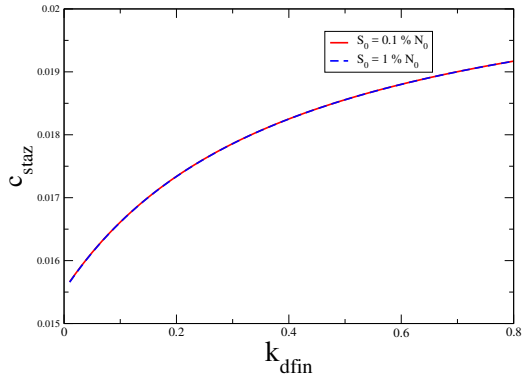

Figure D.17: c_{staz} variation with respect to k_{dfin} ($k_{dfin} = 0.1$ for simulations carried out with the parameters listed in Table 1).

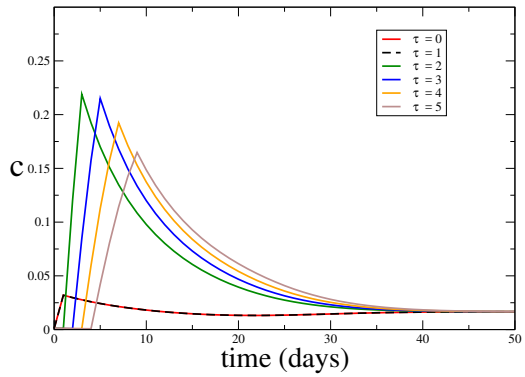

Figure D.18: The dependence of the time associated with the overshoot peak on the time delay τ ($c_0 = 0.1\%$).

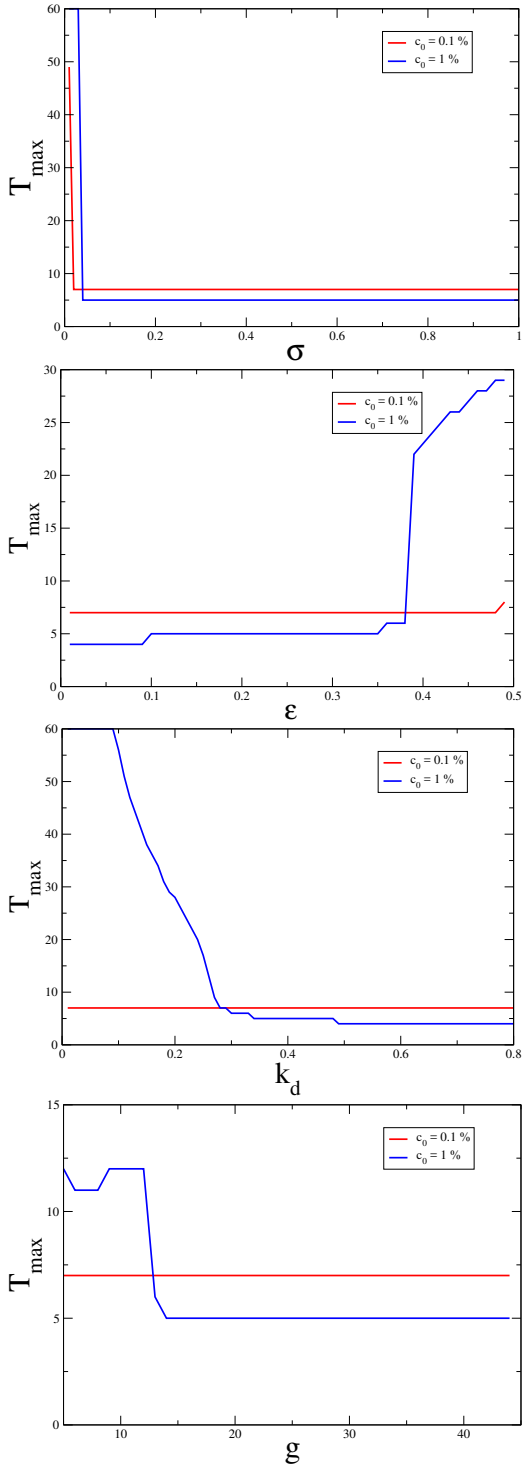

Figure D.19: The dependence of the time associated with the overshoot peak on some of the parameters of the system.

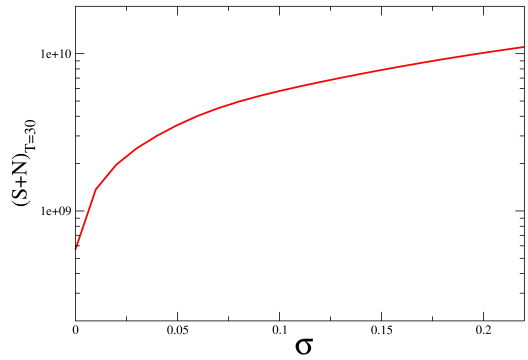

Figure D.20: Dependence of the total mass of the tumour at a fixed time on σ, according to our model (standard initial conditions).

APPENDIX E

Spatial simulations.

E.1 Spatial overshoot.

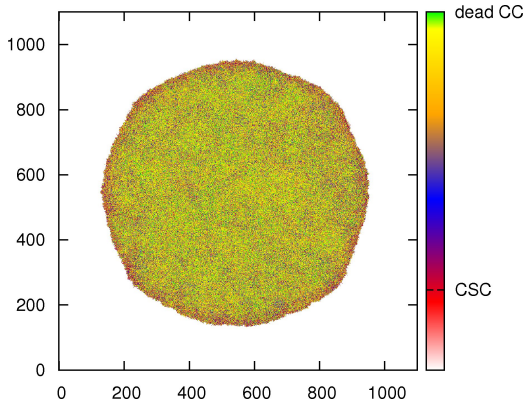

Figure E.1: The spatial arrangements of CC of different ages and of CSC after 11 days, $D = 10^1\ \mu\mathrm{m}^2/\mathrm{s}$.

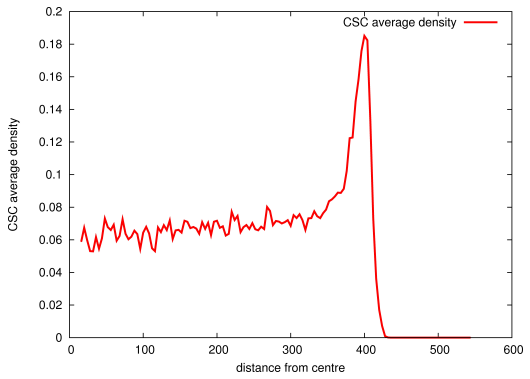

Figure E.2: Radial distribution of the fraction of CSC after 11 days, $D = 10^1\ \mu\mathrm{m}^2/\mathrm{s}$.

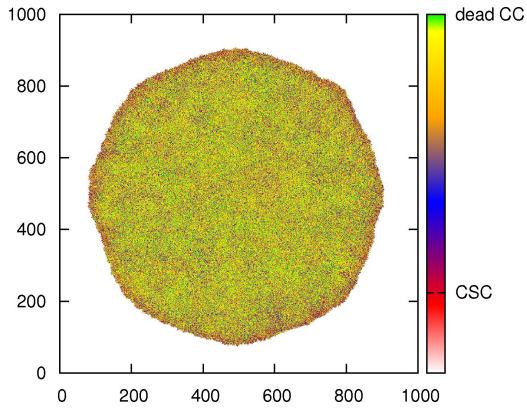

Figure E.3: The spatial arrangements of CC of different ages and of CSC after 11 days, $D = 1\ \mu\mathrm{m}^2/\mathrm{s}$.

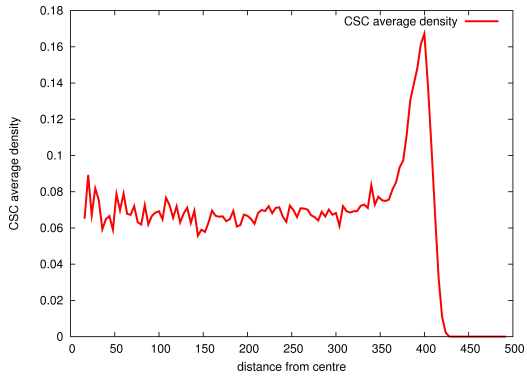

Figure E.4: Radial distribution of the fraction of CSC after 11 days, $D = 1\ \mu\mathrm{m}^2/\mathrm{s}$.

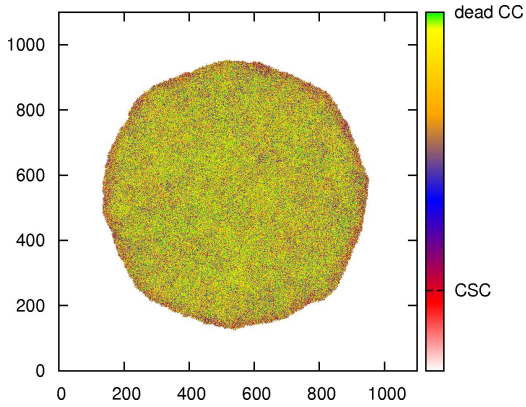

Figure E.5: The spatial arrangements of CC of different ages and of CSC after 11 days, $D = 10^{-1}\ \mu\text{m}^2/\text{s}$.

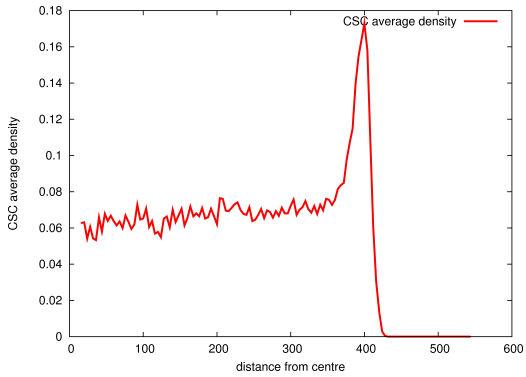

Figure E.6: Radial distribution of the fraction of CSC after 11 days, $D = 10^{-1}\ \mu\text{m}^2/\text{s}$.

E.2 CSC fraction evolution.

Here we consider the evolution of the fraction of CSC in the tumour mass. Defined as $c = \frac{S}{S+N}$, it is clearly $c = 1$ for $t = 0$, since our initial condition is invariably a single CSC in the centre of an otherwise empty 2D lattice. In the following pictures we report the evolution of $c(t)$, first in the case where delay in switching is active, and then in the case where there is no delay in switching. In both cases we have included for comparison the curve of $c(t)$ in the case where $\sigma = 0$ (because when switching is inactive the delay loses significance).

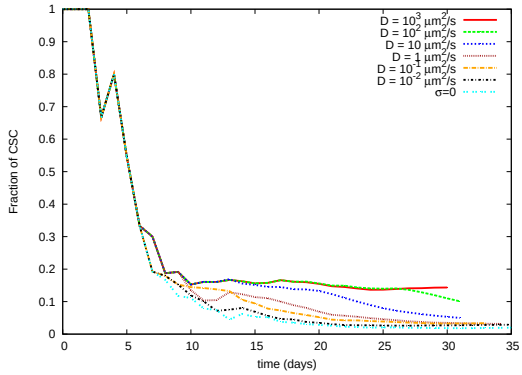

Figure E.7: Evolution of fraction of CSCs on total number of cells with $\tau = 4$ d.

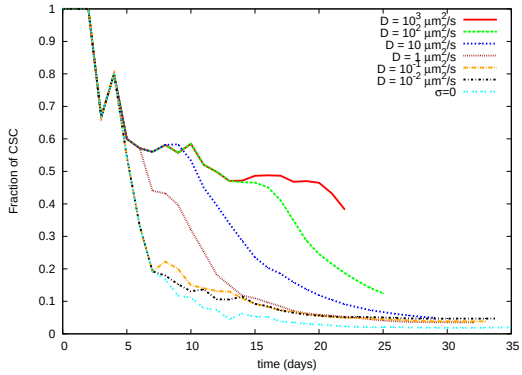

Figure E.8: Evolution of fraction of CSCs on total number of cells with $\tau = 0$ d.

The behaviour observed in both cases is in accordance with the observation (resulting from experiments reported, for example, in (88)) that when a cancer stem cell population is isolated, purified and maintained in culture, the percentage of stem cells rapidly decreases and finally stabilizes at a certain level, characteristic for this given cell culture. This is evident especially in the cases with lower diffusion coefficients, where $c(t)$ reaches an undoubted steady state. From the pictures we clearly see that when delay in switching is present, the fraction of CSCs stabilizes faster than in the case with no delay

and generally to a lower equilibrium value. Also we note that the fraction of CSCs in the cases with low values of the diffusion coefficients is virtually indistinguishable, in its evolution, from $c(t)$ corresponding to the case with $\sigma = 0$ (no switching), while of course the actual numbers of cells would be quite different in the two cases.

E.3 Total growth rate of the tumour and CSC growth rate.

Here we perform the analysis of the total rate of growth of the tumour mass. If the total number of cells grows according the exponential law

$$S(t) + N(t) = c_1 \exp(K \cdot t) \tag{E.1}$$

then from the data of the simulations we can estimate the total instantaneous growth rate K as $K(t) = \frac{\log(S(t)+N(t))}{t}$.

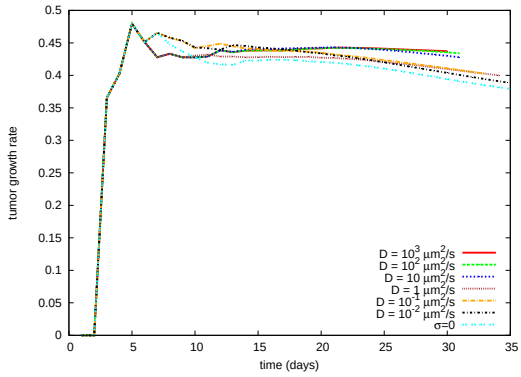

Figure E.9: Instantaneous growth rate of the tumour mass with time delay in switching.

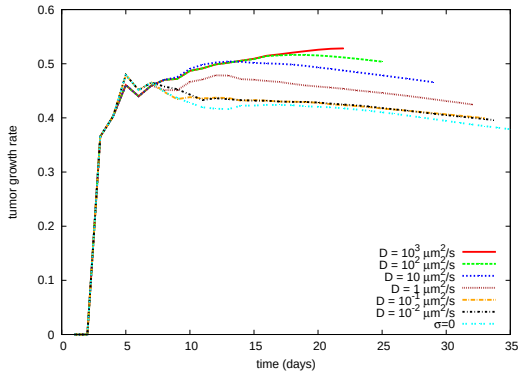

Figure E.10: Instantaneous growth rate of the tumour mass with no time delay in switching.

What is evident here is that after an initial period of super-exponential growth, the growth rate become approximately constant, to mildly decline for larger times. We see

wider gap of growth rate between high and low values of the diffusion coefficients in the case with no time delay in switching. In the case with delay, the growth rate is generally smaller, and again we observe no great difference between the curves of $K(t)$ for the lowest values of the diffusion coefficient and the curve for the case $\sigma = 0$.

E.4 The case $\sigma = 0$.

In this section we collect two examples of spatial plots (with the cellular spatial organization of the tumoral mass) and their respective radial densities of the fraction of CSC obtained during the simulation in absence of switching ($\sigma = 0$). It is evident that no permanent and isotropic enrichment of CSC at the border of the tumoral mass is present, and that the maxima in the radial density of CSCs are assumed in the inner regions of the tumour. These maxima, with respect to the case of active controlled switching, appear to be more statistical fluctuations from the average density, especially at large times (when they are smoothed). CSC can however form (see Section E.9) individual clusters of higher density, whose distribution is of course anistotropic.

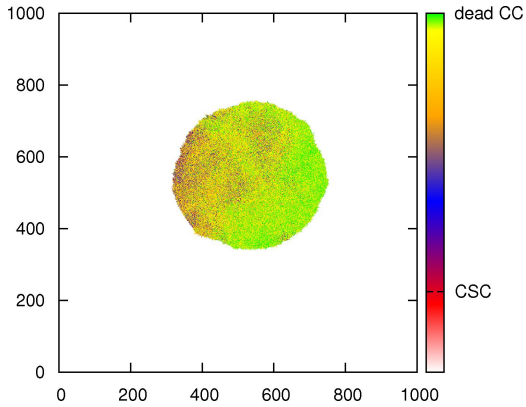

Figure E.11: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\sigma = 0$.

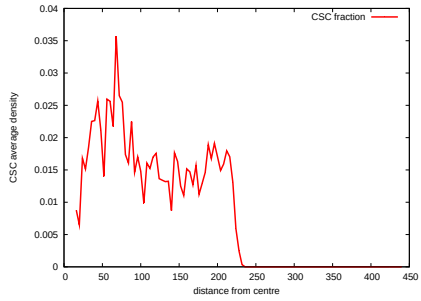

Figure E.12: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\sigma = 0$.

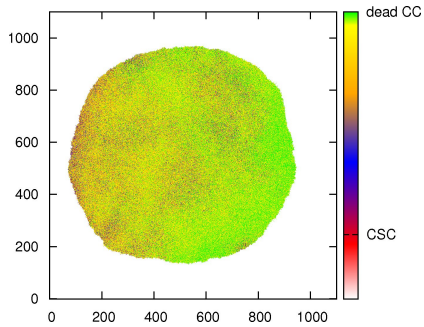

Figure E.13: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 35$ d, with $\sigma = 0$.

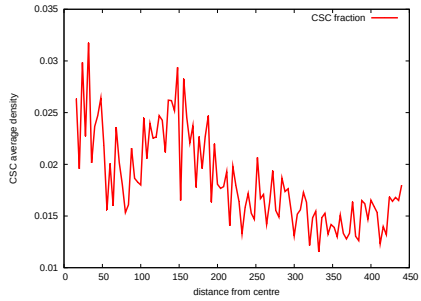

Figure E.14: Snapshot of radial distribution of CSC at time $t = 35$ d, with $\sigma = 0$.

This is also evident from Fig. E.15, where it can be seen that r_{max} is not always close to r_{tumour}.

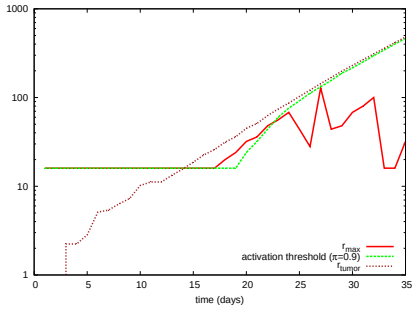

Figure E.15: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5 Spatial plots and radial densities.

In this section we collect, for each simulation of interest, a sample of the output as follows:

- a spatial plot with the cellular spatial organization at a certain time t (in general close to the end of the simulation);
- the radial distribution of the fraction of CSC on total cells at time t;
- a plot comparing (at the time T) the radial distribution of the switching inhibitor molecule, the radial distribution of the fraction of CSC on total cells, the threshold K and the effective switching factor $\pi(r,t) = m^h(r,t)/[K^h + m^h(r,t)]$;
- a plot showing the evolution of the position r_{max} of the peak in CSC concentration, the evolution of the activation threshold, defined as the value of r at which the concentration of the molecule gives $\pi = 0.9$, i.e., at which switching inhibition drops to very low values, and the evolution of the overall radius r_{tumour} of the tumoral mass during the entire course of the simulation.

E.5.1 $D = 10^{-2}$ μm^2/s, $\tau = 0$ d.

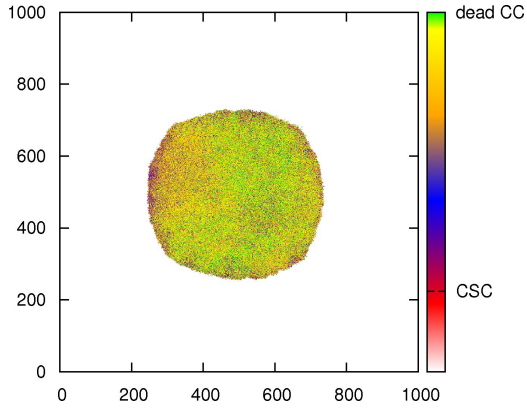

Figure E.16: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 0$ d and $D = 10^{-2}$ μm^2/s.

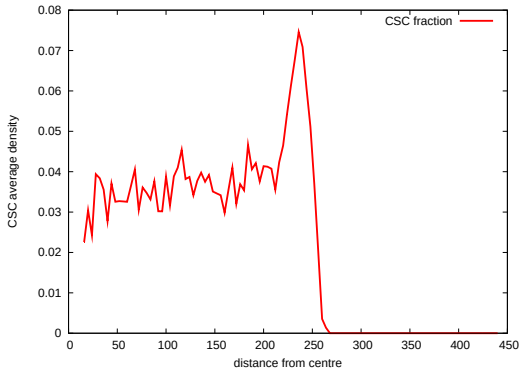

Figure E.17: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 0$ d and $D = 10^{-2}$ μm^2/s.

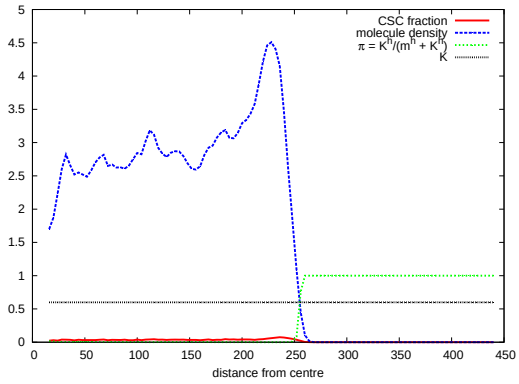

Figure E.18: Snapshot of radial distributions at time $t = 30$ d, with $\tau = 0$ d and $D = 10^{-2}$ μm^2/s.

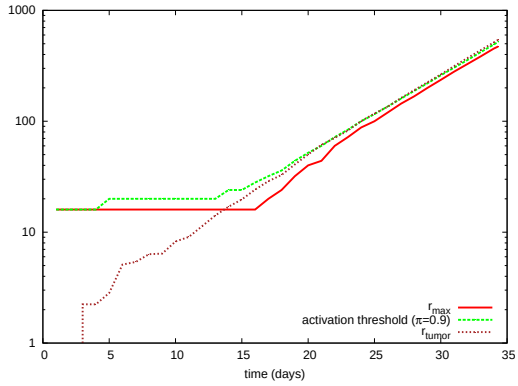

Figure E.19: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.2 $D = 10^{-2}$ μm^2/s, $\tau = 4$ d.

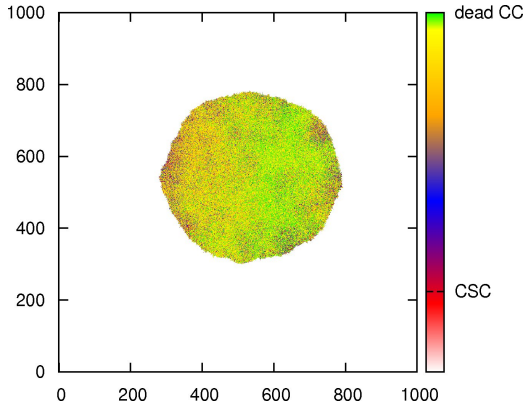

Figure E.20: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 4$ d and $D = 10^{-2}$ μm^2/s.

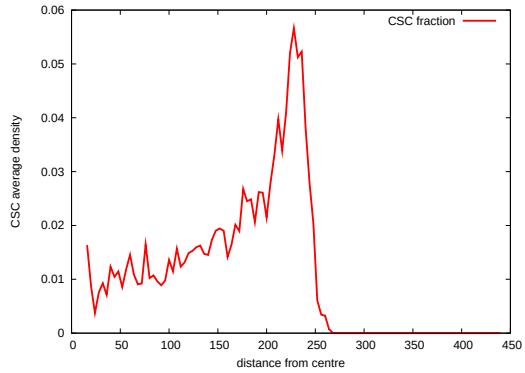

Figure E.21: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 4$ d and $D = 10^{-2}$ μm^2/s.

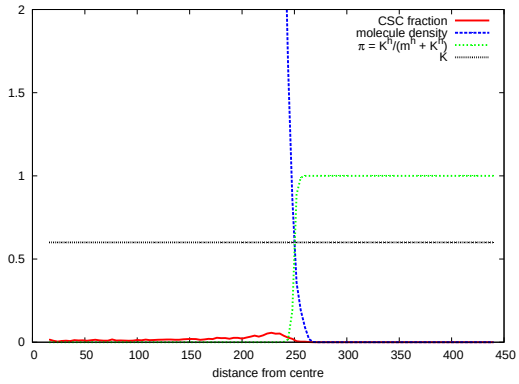

Figure E.22: Snapshot of radial distributions at time $t = 30$ d, with $\tau = 4$ d and $D = 10^{-2}$ μm^2/s.

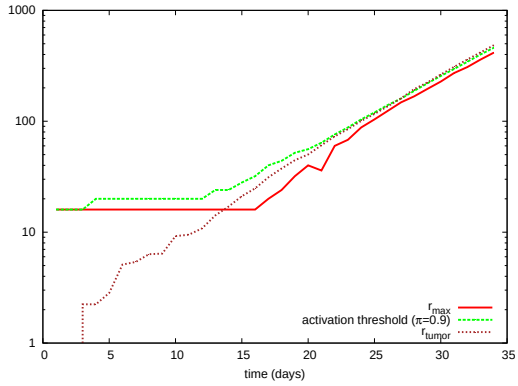

Figure E.23: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.3 $D = 10^{-1}$ μm^2/s, $\tau = 0$ d.

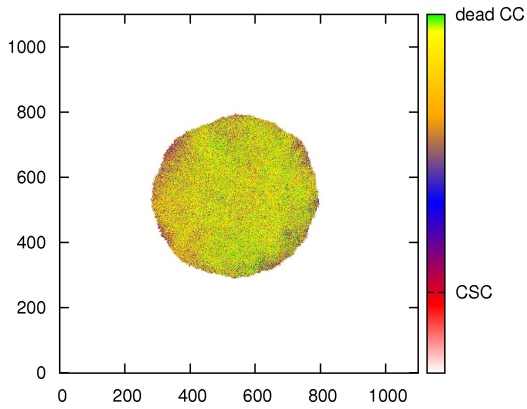

Figure E.24: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 0$ d and $D = 10^{-1}$ μm^2/s.

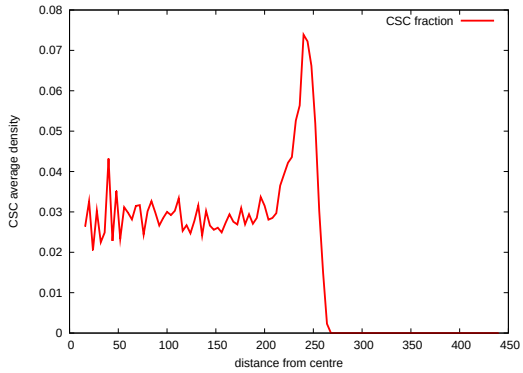

Figure E.25: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 0$ d and $D = 10^{-1}$ μm^2/s.

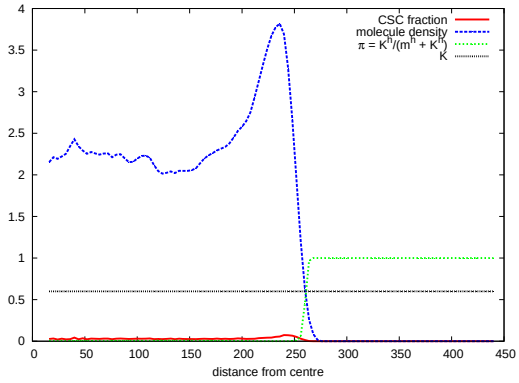

Figure E.26: Snapshot of radial distrbutions at time $t = 30$ d, with $\tau = 0$ d and $D = 10^{-1}$ μm^2/s.

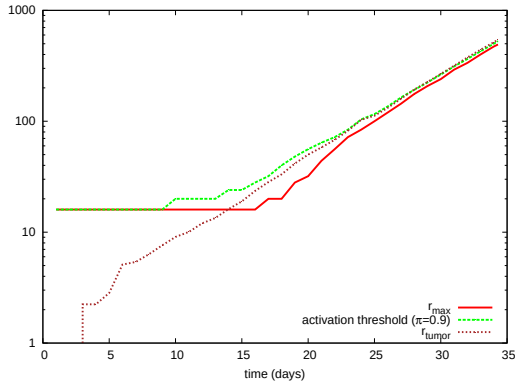

Figure E.27: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.4 $D = 10^{-1}$ μm^2/s, $\tau = 4$ d.

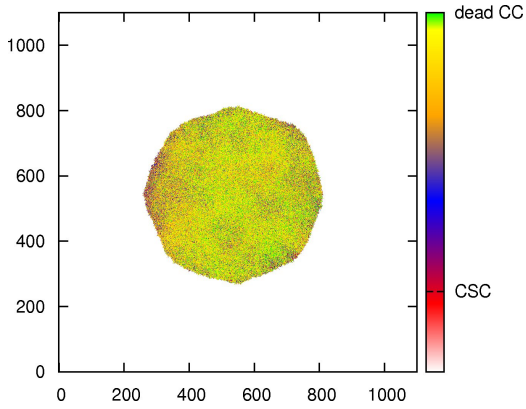

Figure E.28: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 4$ d and $D = 10^{-1}$ μm^2/s.

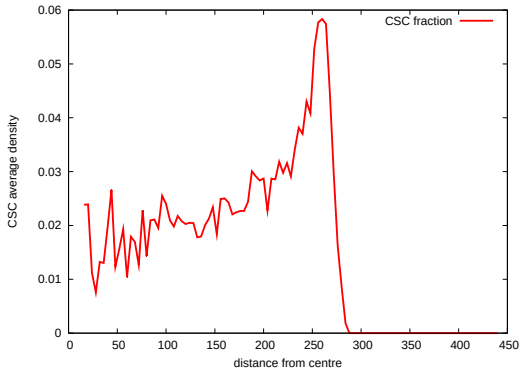

Figure E.29: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 4$ d and $D = 10^{-1}$ μm^2/s.

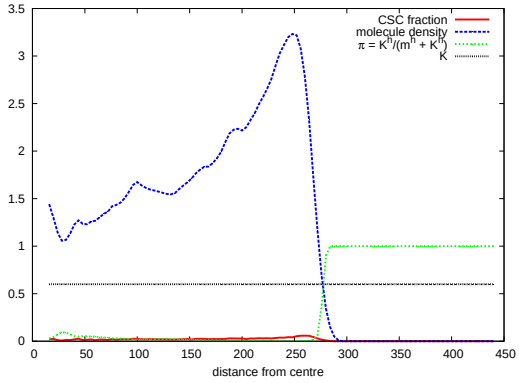

Figure E.30: Snapshot of radial distrbutions at time $t = 30$ d, with $\tau = 4$ d and $D = 10^{-1}$ μm^2/s.

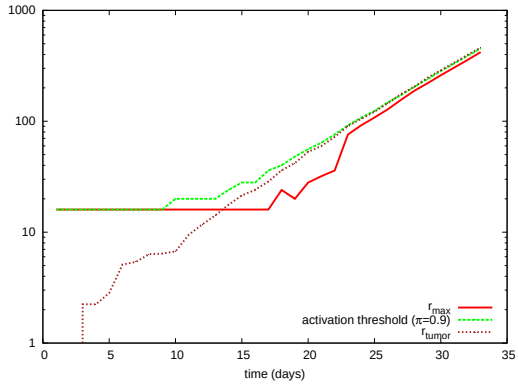

Figure E.31: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.5 $D = 1\ \mu\text{m}^2/\text{s}$, $\tau = 0$ d.

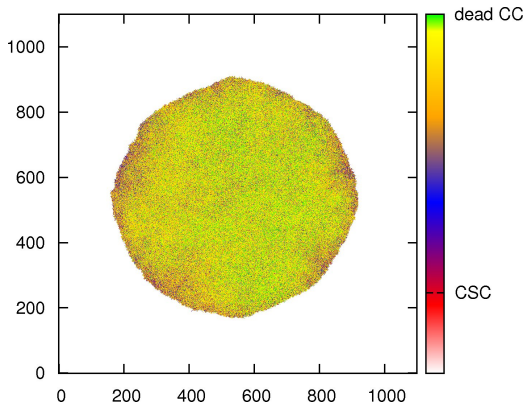

Figure E.32: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 0$ d and $D = 1\ \mu\text{m}^2/\text{s}$.

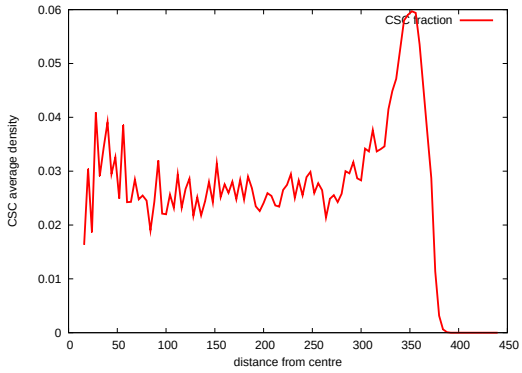

Figure E.33: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 0$ d and $D = 1\ \mu\text{m}^2/\text{s}$.

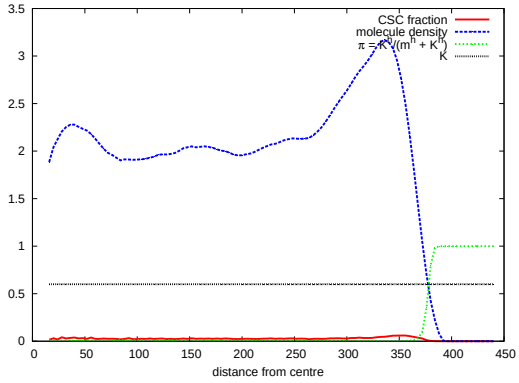

Figure E.34: Snapshot of radial distrbutions at time $t = 30$ d, with $\tau = 0$ d and $D = 1\ \mu\text{m}^2/\text{s}$.

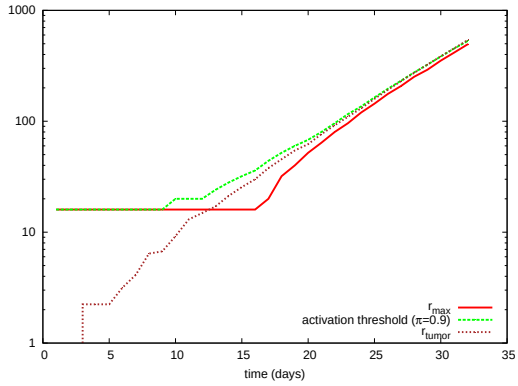

Figure E.35: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.6 $D = 1\ \mu\text{m}^2/\text{s}$, $\tau = 4$ d.

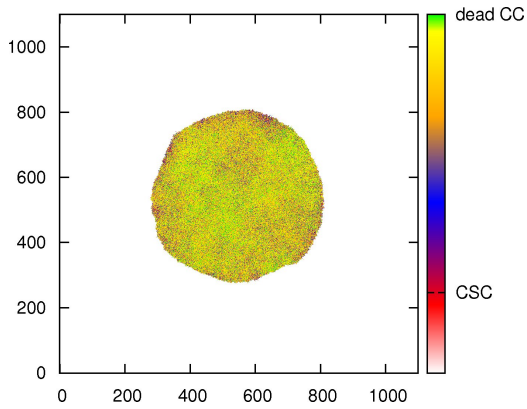

Figure E.36: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 4$ d and $D = 1\ \mu\text{m}^2/\text{s}$.

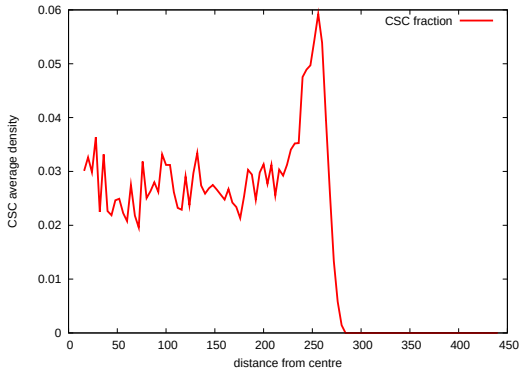

Figure E.37: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 4$ d and $D = 1\ \mu\text{m}^2/\text{s}$.

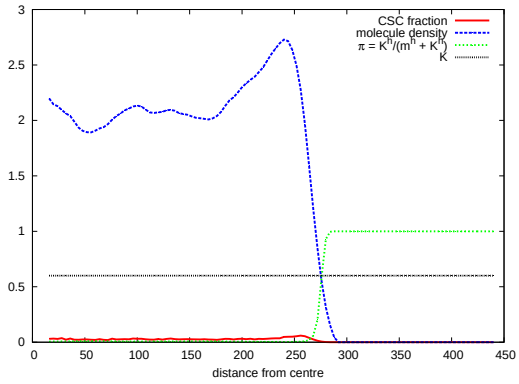

Figure E.38: Snapshot of radial distrbutions at time $t = 30$ d, with $\tau = 4$ d and $D = 1\ \mu\text{m}^2/\text{s}$.

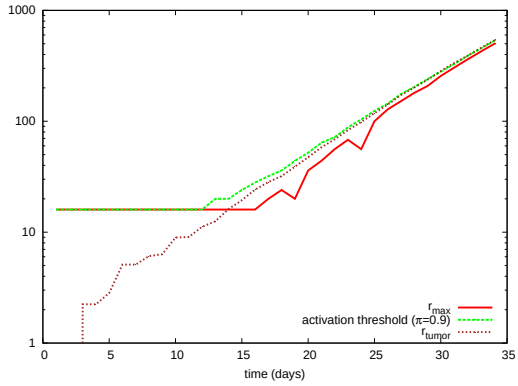

Figure E.39: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.7 $D = 10\ \mu\text{m}^2/\text{s}$, $\tau = 0$ d.

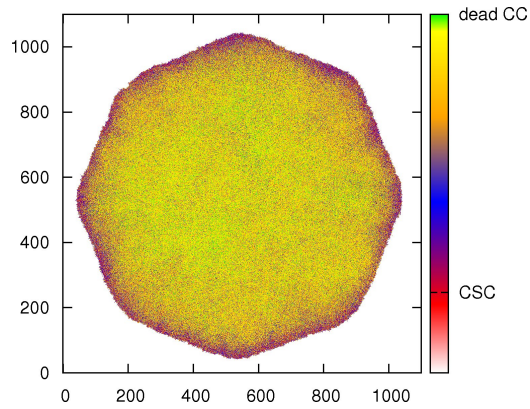

Figure E.40: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 29$ d, with $\tau = 0$ d and $D = 10\ \mu\text{m}^2/\text{s}$.

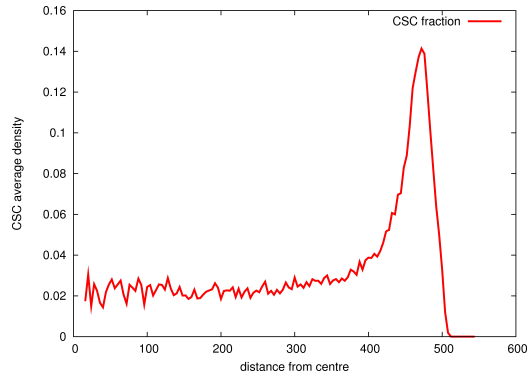

Figure E.41: Snapshot of radial distribution of CSC at time $t = 29$ d, with $\tau = 0$ d and $D = 10\ \mu\text{m}^2/\text{s}$.

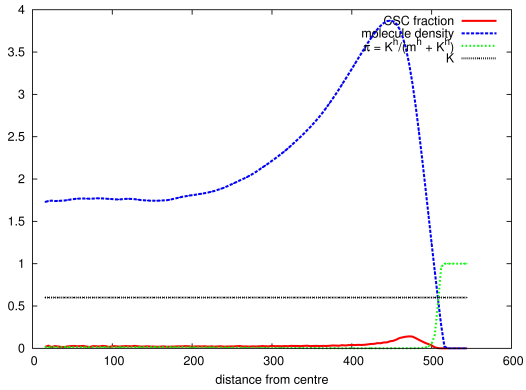

Figure E.42: Snapshot of radial distrbutions at time $t = 29$ d, with $\tau = 0$ d and $D = 10\ \mu\text{m}^2/\text{s}$.

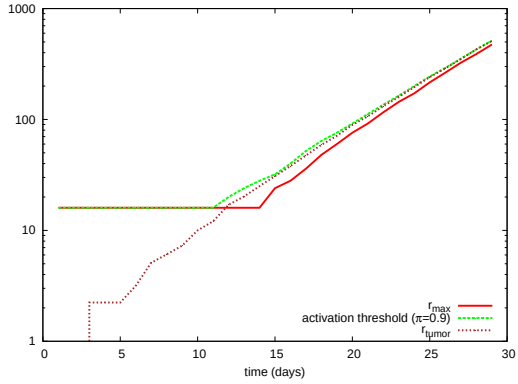

Figure E.43: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.8 $D = 10\ \mu\text{m}^2/\text{s}$, $\tau = 4$ d.

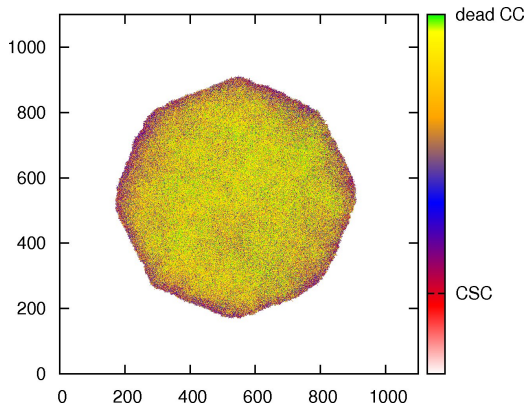

Figure E.44: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 4$ d and $D = 10\ \mu\text{m}^2/\text{s}$.

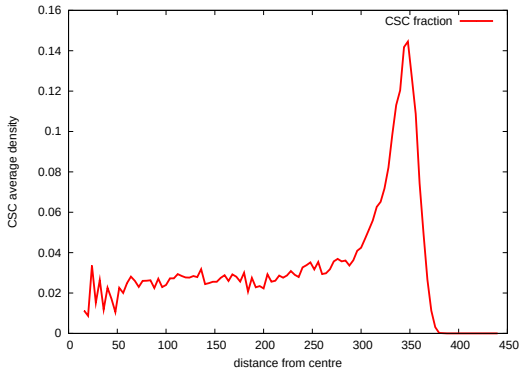

Figure E.45: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 4$ d and $D = 10\ \mu\text{m}^2/\text{s}$.

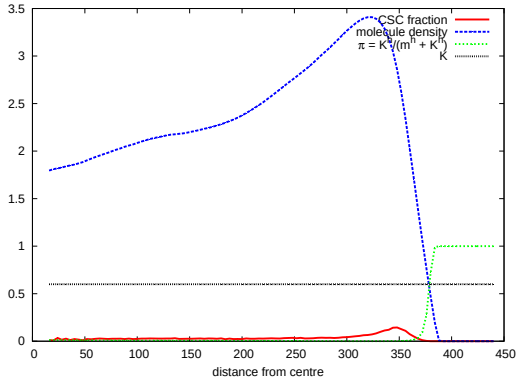

Figure E.46: Snapshot of radial distrbutions at time $t = 30$ d, with $\tau = 4$ d and $D = 10\ \mu\text{m}^2/\text{s}$.

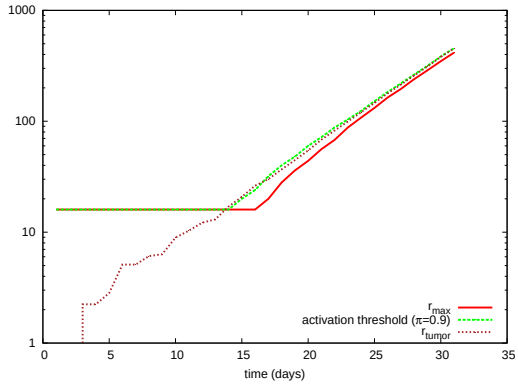

Figure E.47: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.9 $D = 10^2$ μm^2/s, $\tau = 0$ d.

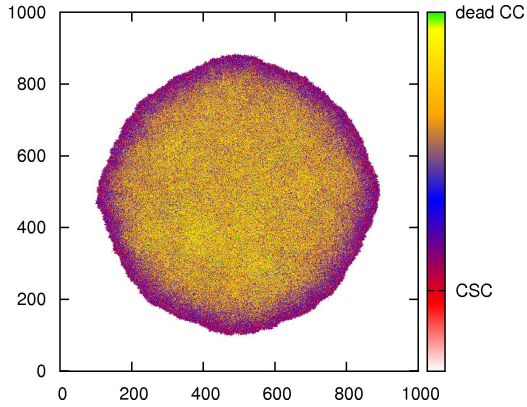

Figure E.48: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 26$ d, with $\tau = 0$ d and $D = 10^2$ μm^2/s.

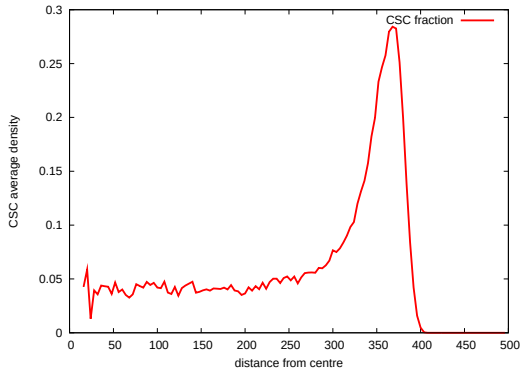

Figure E.49: Snapshot of radial distribution of CSC at time $t = 26$ d, with $\tau = 0$ d and $D = 10^2$ μm^2/s.

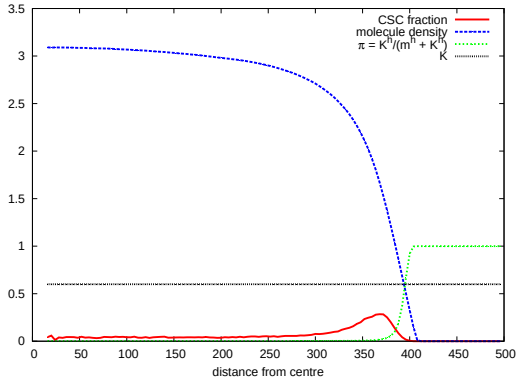

Figure E.50: Snapshot of radial distrbutions at time $t = 26$ d, with $\tau = 0$ d and $D = 10^2$ μm^2/s.

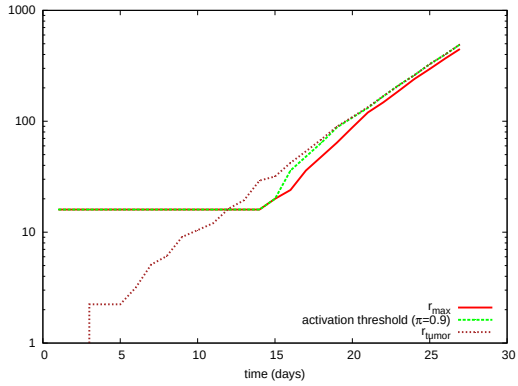

Figure E.51: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.10 $D = 10^2$ μm^2/s, $\tau = 4$ d.

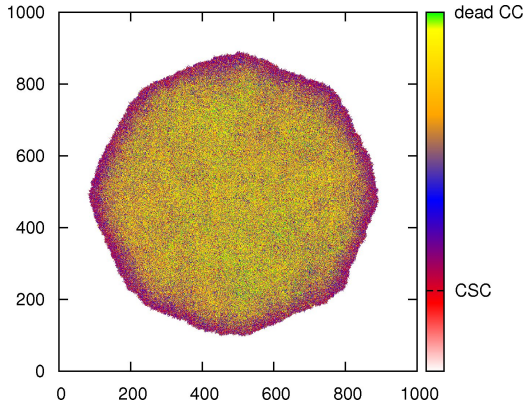

Figure E.52: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 4$ d and $D = 10^2$ μm^2/s.

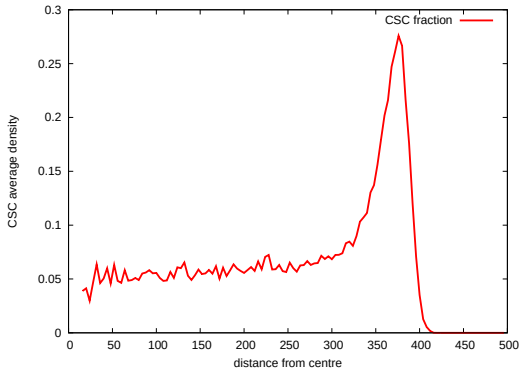

Figure E.53: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 4$ d and $D = 10^2$ μm^2/s.

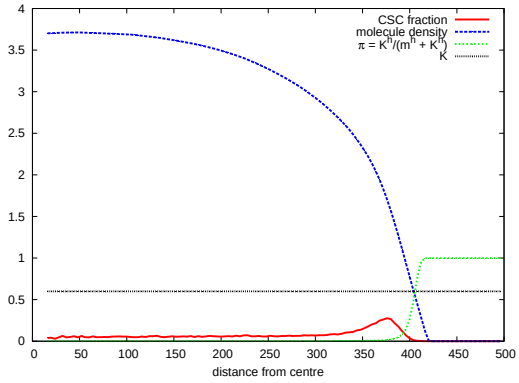

Figure E.54: Snapshot of radial distrbutions at time $t = 30$ d, with $\tau = 4$ d and $D = 10^2$ μm^2/s.

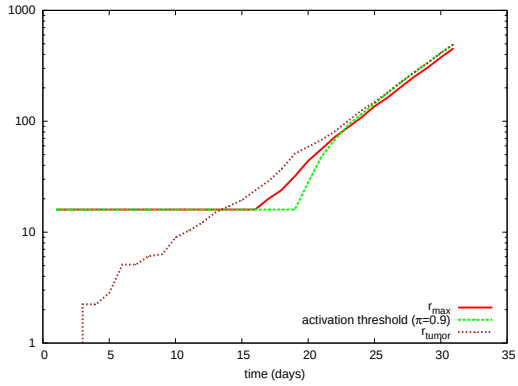

Figure E.55: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.11 $D = 10^3$ μm^2/s, $\tau = 0$ d.

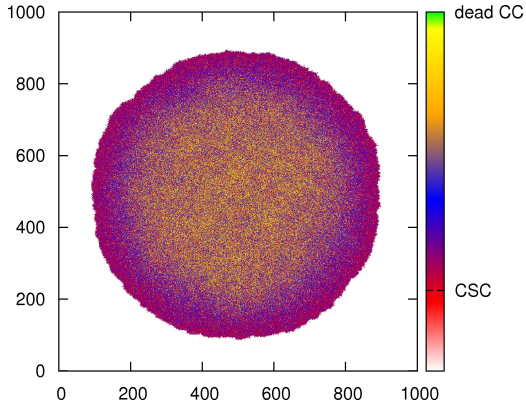

Figure E.56: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 26$ d, with $\tau = 0$ d and $D = 10^3$ μm^2/s.

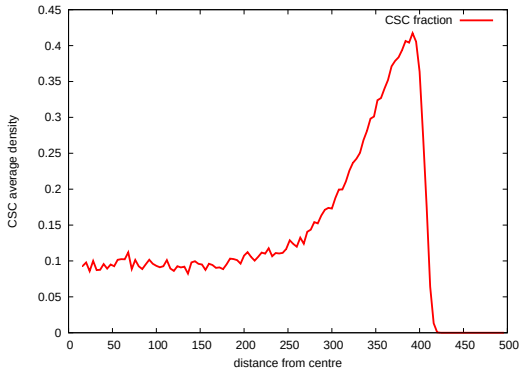

Figure E.57: Snapshot of radial distribution of CSC at time $t = 26$ d, with $\tau = 0$ d and $D = 10^3$ μm^2/s.

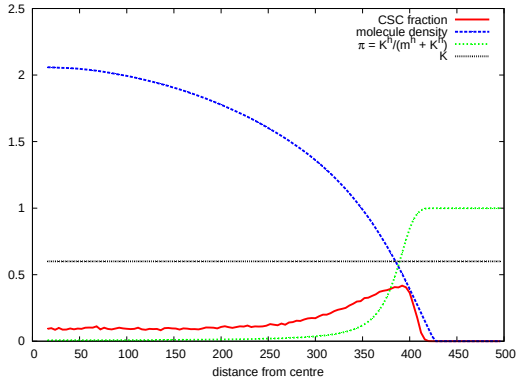

Figure E.58: Snapshot of radial distrbutions at time $t = 26$ d, with $\tau = 0$ d and $D = 10^3$ μm^2/s.

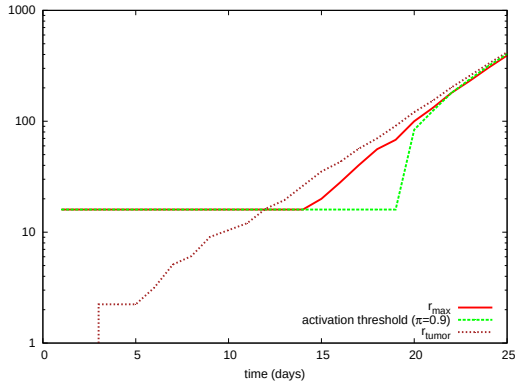

Figure E.59: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.5.12 $D = 10^3$ μm^2/s, $\tau = 4$ d.

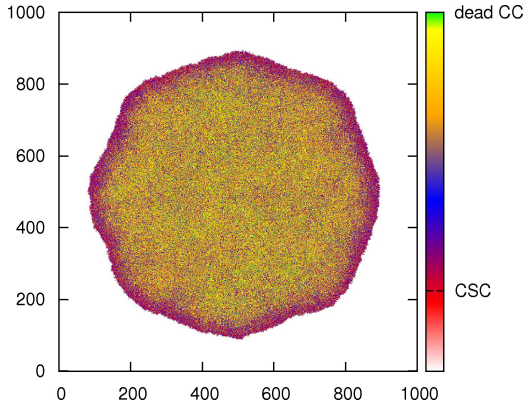

Figure E.60: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 30$ d, with $\tau = 4$ d and $D = 10^3$ μm^2/s.

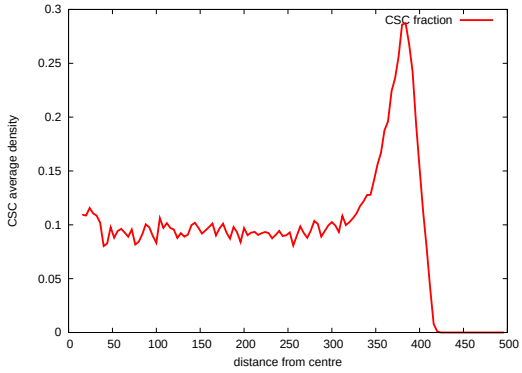

Figure E.61: Snapshot of radial distribution of CSC at time $t = 30$ d, with $\tau = 4$ d and $D = 10^3$ μm^2/s.

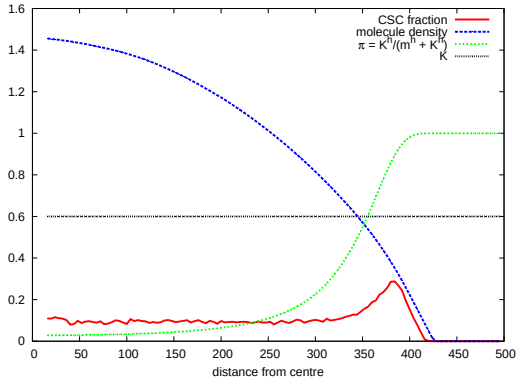

Figure E.62: Snapshot of radial distrbutions at time $t = 30$ d, with $\tau = 4$ d and $D = 10^3$ μm^2/s.

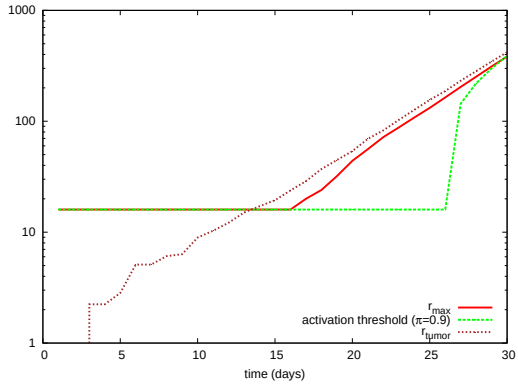

Figure E.63: Evolution in the simulation of the position r_{max} of the peak in CSC concentration, the activation threshold (position for which $\pi = 0.9$), and the overall radius r_{tumour} of the tumoral mass.

E.6 Radial distributions of CSC fractions and molecule density compared.

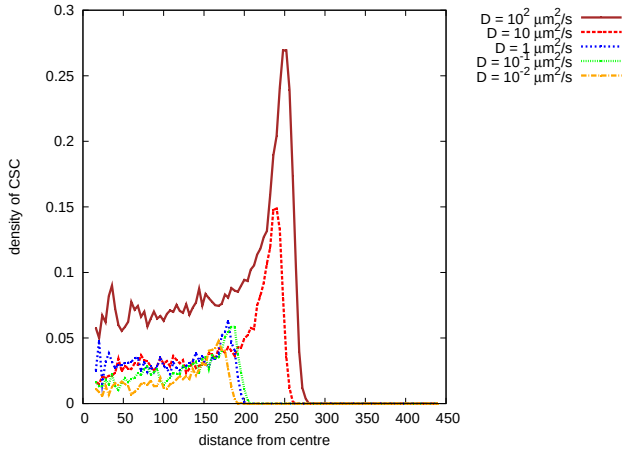

Figure E.64: Radial distribution of fraction of CSC at time $t = 28$ d, with $\tau = 4$ d.

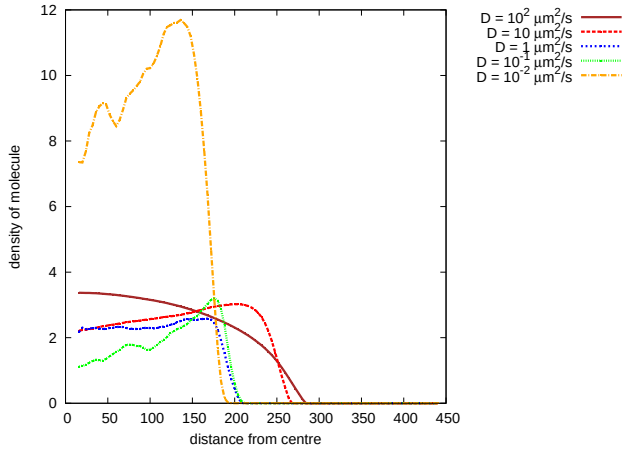

Figure E.65: Radial distribution of molecule at time $t = 28$ d, with $\tau = 4$ d.

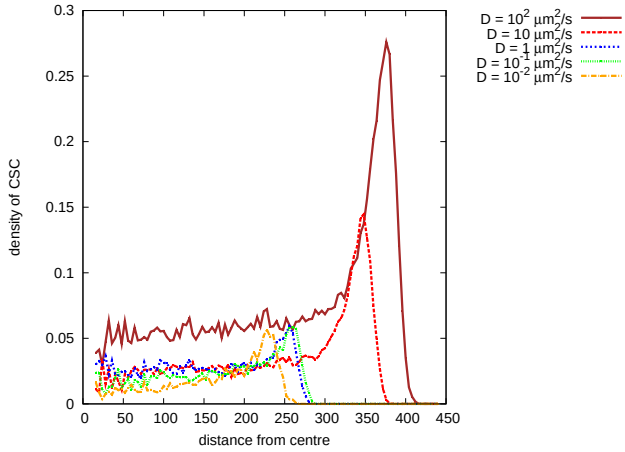

Figure E.66: Radial distribution of fraction of CSC at time $t = 30$ d, with $\tau = 4$ d.

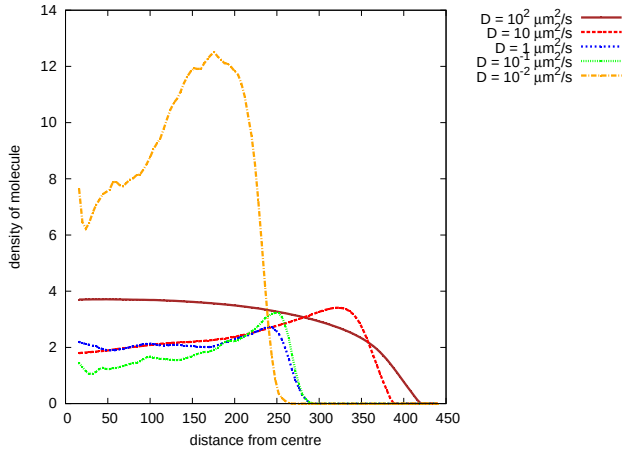

Figure E.67: Radial distribution of molecule at time $t = 30$ d, with $\tau = 4$ d.

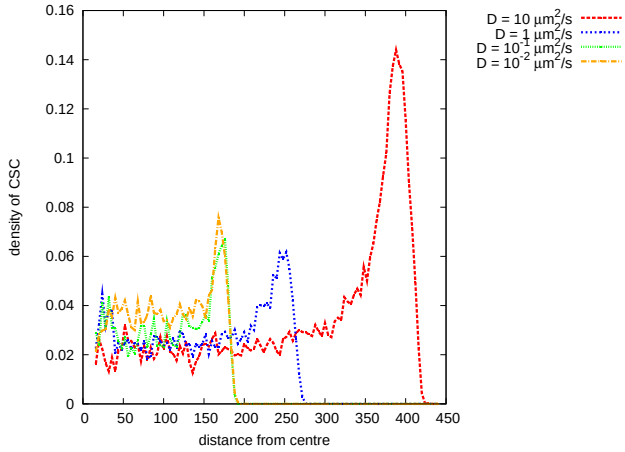

Figure E.68: Radial distribution of fraction of CSC at time $t = 28$ d, with $\tau = 0$ d.

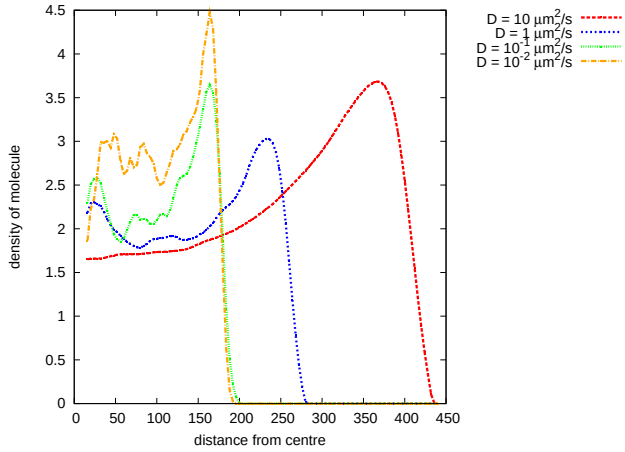

Figure E.69: Radial distribution of molecule at time $t = 28$ d, with $\tau = 0$ d.

E.7 Simulations with $\mathbf{k}_m = 0.5\,\mathbf{d}^{-1}$.

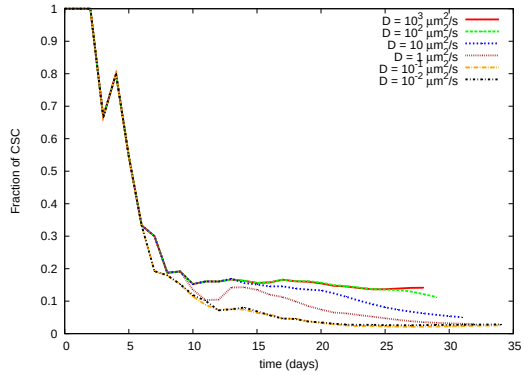

Figure E.70: Evolution of fraction of CSCs on total number of cells with $\tau = 4$ d.

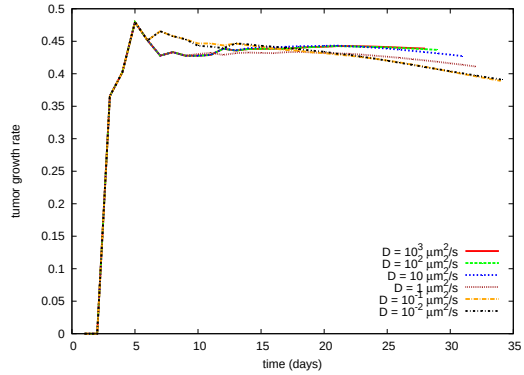

Figure E.71: Instantaneous growth rate of the tumour mass with $\tau = 4$ d.

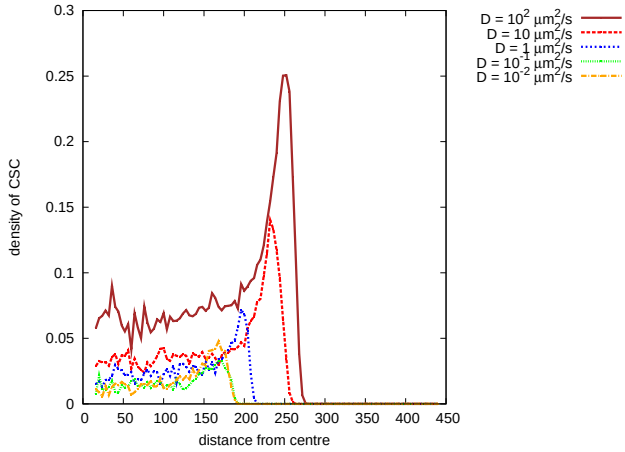

Figure E.72: Radial distribution of fraction of CSC at time $t = 28$ d, with $\tau = 4$ d.

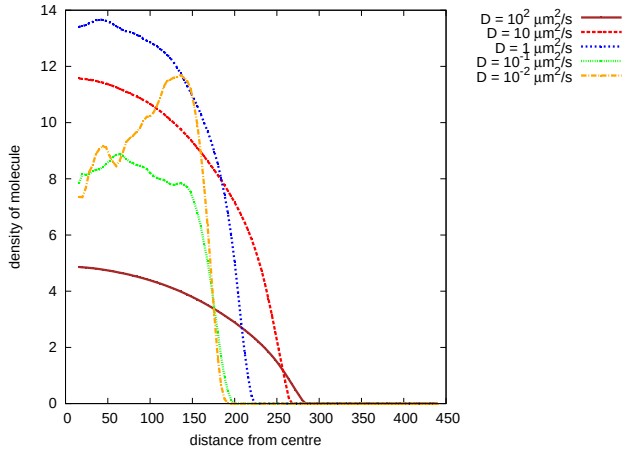

Figure E.73: Radial distribution of molecule at time $t = 28$ d, with $\tau = 4$ d.

E.8 Evolution of maxima of CSC fraction.

In this section we show the evolution of the maxima of the radial fraction of the CSC, both in the case with E.74 and without delay E.75. In is worth remembering that, while for $\sigma \neq 0$ the maximum, after the first 5–10 days of growth, is always located in the proximity of the border of the tumoral mass, this is not the case for $\sigma = 0$.

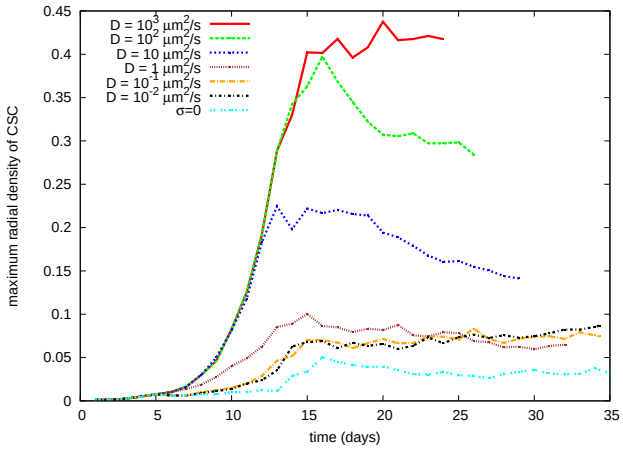

Figure E.74: Evolution of maximum radial density of CSCs with no time delay in switching.

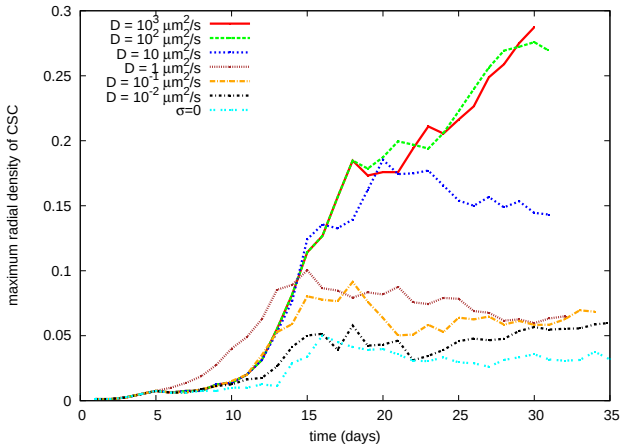

Figure E.75: Evolution of maximum radial density of CSCs with time delay in switching.

What is evident from the picture is that in both cases the maximum of the local density (almost always located on the border of the tumour mass) grows in time until it reaches a maximum value and then decreases. Observing the behaviour of the curves corresponding the lowest diffusion coefficients, it seems reasonable to infer that the maximum of CSC fraction converges to a stationary value. It would be interesting to determine whether the curves for the highest diffusion coefficients, once extended the time of

simulation, would display the same behaviour.

E.9 Clustering analysis.

E.9.1 $\sigma = 0$, $t = 34$ d.

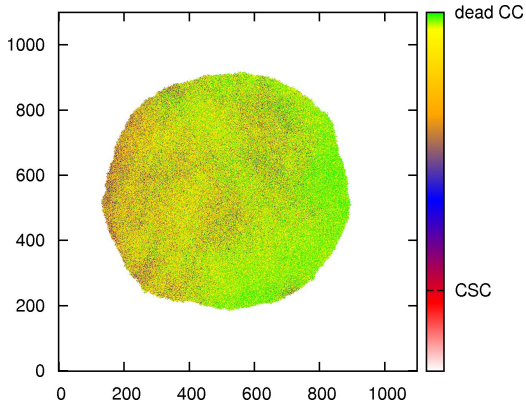

Figure E.76: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 34$ d, with $\sigma = 0$.

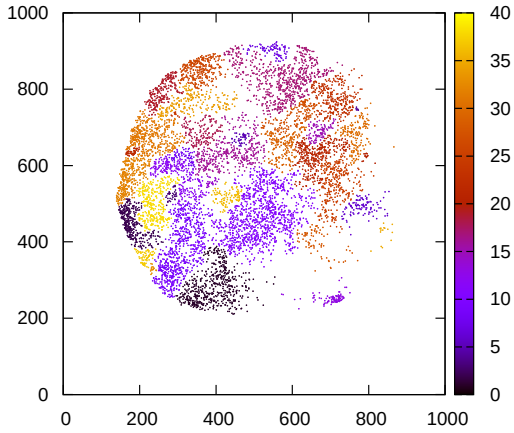

Figure E.77: The clusters of CSC, colored differently for each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

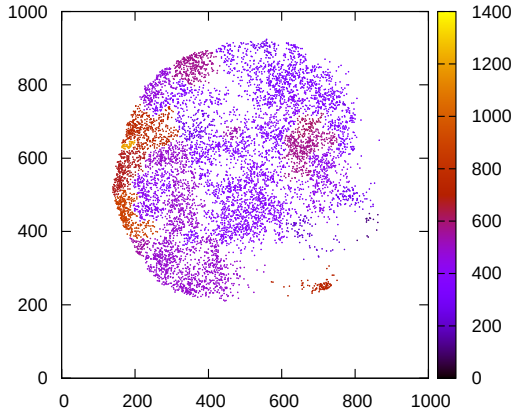

Figure E.78: The clusters of CSC, colored differently according to the spatial density of each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

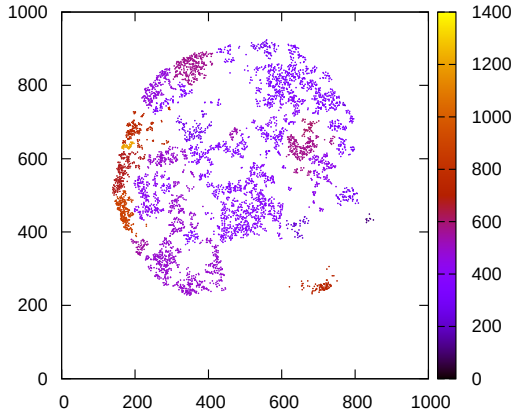

Figure E.79: The clusters of CSC, subtracted their density halo and colored differently according to the spatial density of each cluster (density calculated without the halo points), obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

E.9.2 $D = 10^{-2}$ μm^2/s, $\tau = 4$ d, $t = 34$ d.

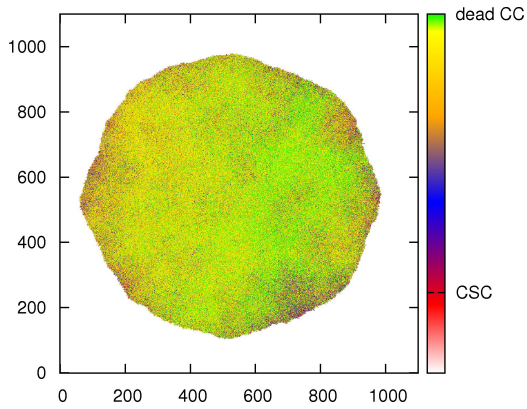

Figure E.80: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 34$ d, with $D = 10^{-2}$ μm^2/s.

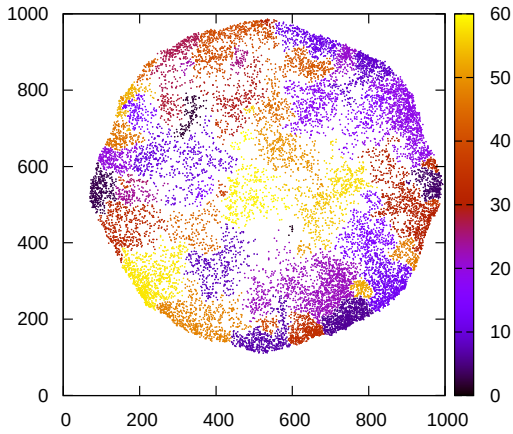

Figure E.81: The clusters of CSC, colored differently for each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

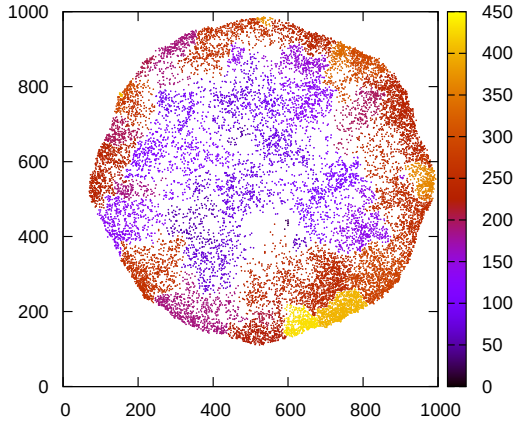

Figure E.82: The clusters of CSC, colored differently according to the spatial density of each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

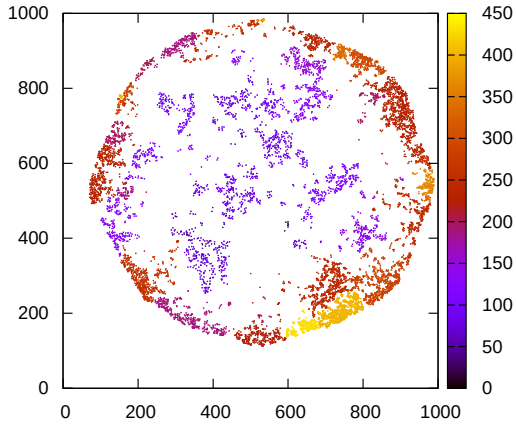

Figure E.83: The clusters of CSC, subtracted their density halo and colored differently according to the spatial density of each cluster (density calculated without the halo points), obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

E.9.3 $D = 10^{-1}$ μm^2/s, $\tau = 4$ d, $t = 34$ d.

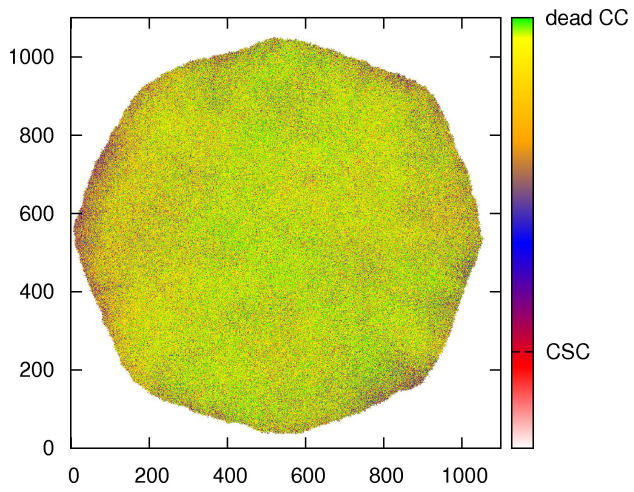

Figure E.84: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 34$ d, with $D = 10^{-1}$ μm^2/s.

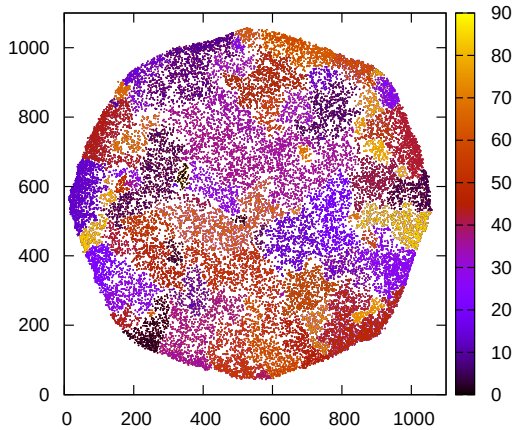

Figure E.85: The clusters of CSC, colored differently for each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

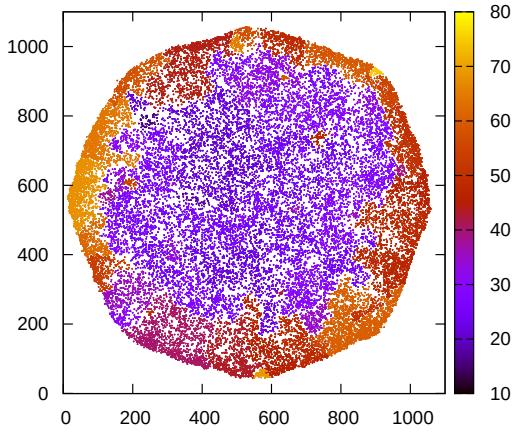

Figure E.86: The clusters of CSC, colored differently according to the spatial density of each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

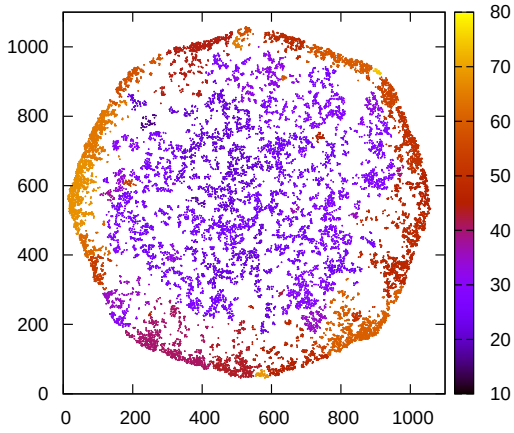

Figure E.87: The clusters of CSC, subtracted their density halo and colored differently according to the spatial density of each cluster (density calculated without the halo points), obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

E.9.4 $D = 10\ \mu\mathbf{m}^2\mathbf{/s}$, $\tau = 4$ **d**, $t = 32$ **d**.

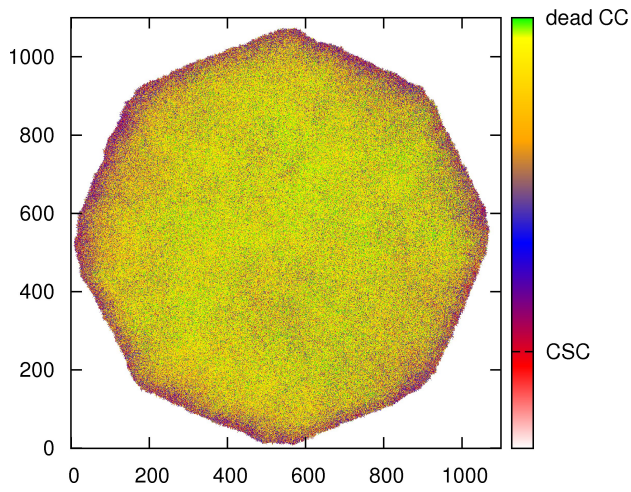

Figure E.88: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 34$ d, with $D = 10\ \mu\text{m}^2/\text{s}$.

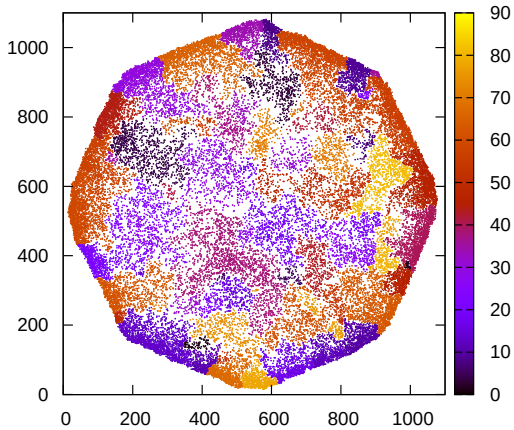

Figure E.89: The clusters of CSC, colored differently for each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

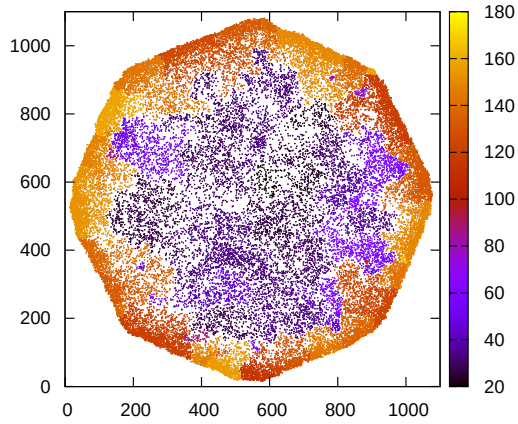

Figure E.90: The clusters of CSC, colored differently according to the spatial density of each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

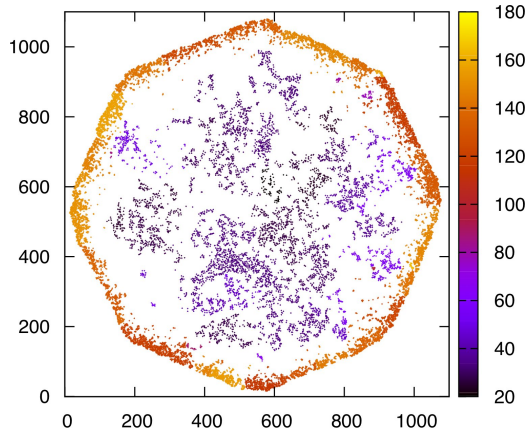

Figure E.91: The clusters of CSC, subtracted their density halo and colored differently according to the spatial density of each cluster (density calculated without the halo points), obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

E.9.5 $D = 10^2$ μm^2/s, $\tau = 4$ d, $t = 3$ d.

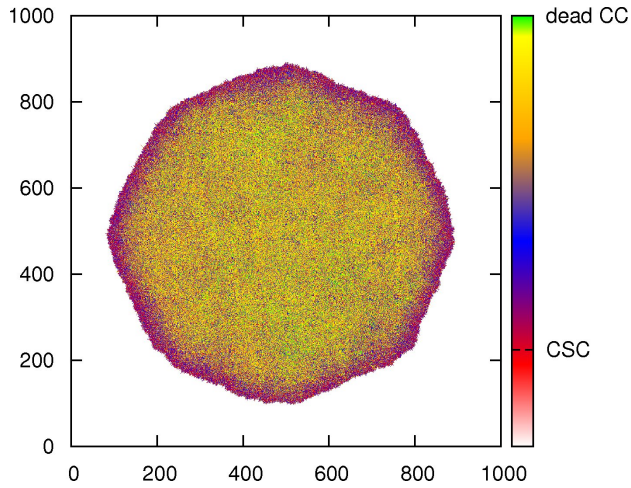

Figure E.92: Spatial plot of the tumoral mass with cellular spatial organization, at time $t = 3$ d, with $D = 10^2$ μm^2/s.

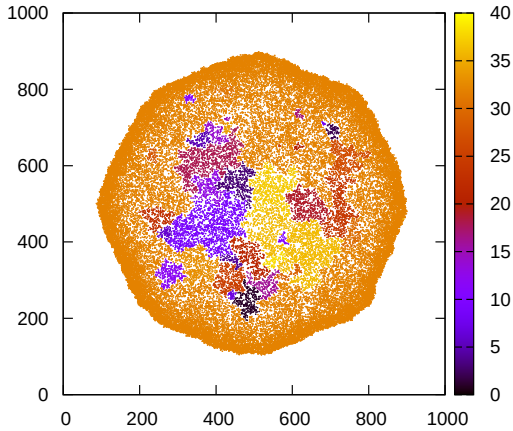

Figure E.93: The clusters of CSC, colored differently for each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

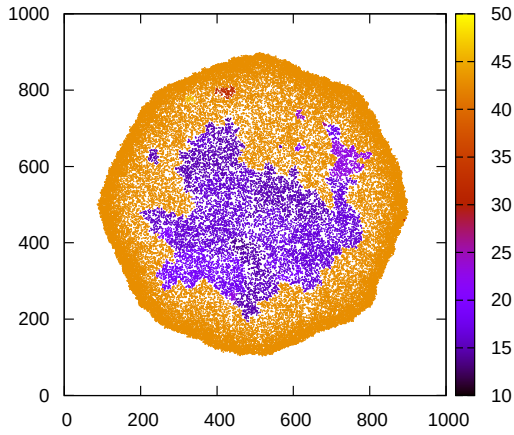

Figure E.94: The clusters of CSC, colored differently according to the spatial density of each cluster, obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

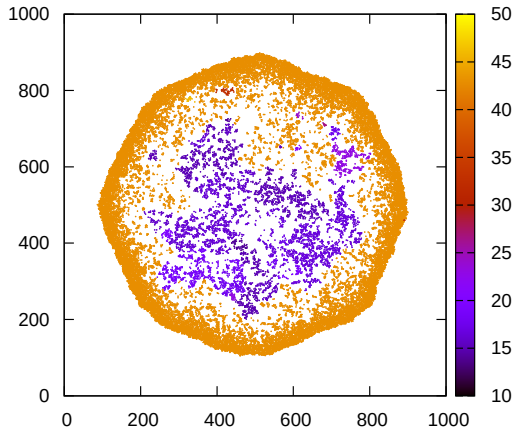

Figure E.95: The clusters of CSC, subtracted their density halo and colored differently according to the spatial density of each cluster (density calculated without the halo points), obtained with the algorithm of ref. (130) using as parameters k-NN $= 5$ and $Z = 1.8$.

APPENDIX F

Persistence of enrichment at the border: mathematical details.

Dividing the lattice radially with bin Δr, one can assume that the number $S(r,t)$ of CSC in a bin follows

$$\frac{\partial S(r,t)}{\partial t} = \left[K\frac{1}{\Delta r}\int_0^{r-\Delta r} dr'\, S(r',t) + k_d\frac{1}{\Delta r}\int_0^{r-\Delta r} dr'\, N(r',t)\right]\cdot[S(r-\Delta r,t) - S(r,t)] + \\ + \epsilon S(r,t). \tag{F.1}$$

where the first describes the transport of CSC to site r: the duplication of CSC and CC at any site $r' < r - \Delta r$ shift all cells in a radial direction and, if combined with the presence of CSC at site $r - \Delta r$, shift this to site r. The second term is the increase of CSC at site r due to duplication at site $r - \Delta r$. Summing and subtracting $\epsilon S(r,t)$ and multiplying and dividing by Δr, one obtains

$$\frac{\partial S(r,t)}{\partial t} = -\left[K\int_0^{r-\Delta r} dr'\, S(r',t) + k_d\int_0^{r-\Delta r} dr'\, N(r',t)\right]\cdot\frac{\partial S}{\partial r} - \epsilon\Delta r\frac{\partial S}{\partial r} + \\ + \epsilon S(r,t). \tag{F.2}$$

Let's make a mean–field approximation, setting

$$\int_0^{r-\Delta r} dr'\, S(r',t) \approx (r-\Delta r)\cdot\overline{S} \\ \int_0^{r-\Delta r} dr'\, N(r',t) \approx (r-\Delta r)\cdot\overline{N}. \tag{F.3}$$

where the constants $\overline{S}$ and $\overline{N}$ are respectively the average population of CSC. The kinetic equation simplifies to

$$\frac{\partial S(r,t)}{\partial t} = -(K\overline{S} + k_d\overline{N})r\cdot\frac{\partial S}{\partial r} + (K\overline{S} + \epsilon\overline{N} - \epsilon)\Delta r\frac{\partial S}{\partial r} + \epsilon S(r,t), \tag{F.4}$$

which can be written, setting $\epsilon' \equiv K\overline{S} + k_d\overline{N}$, as

$$\frac{\partial S(r,t)}{\partial t} = -\epsilon' r\cdot\frac{\partial S}{\partial r} + (\epsilon' - \epsilon)\Delta r\frac{\partial S}{\partial r} + \epsilon S(r,t). \tag{F.5}$$

The equation written in the following form (setting $\Delta\epsilon \equiv \epsilon' - \epsilon$)

$$\frac{\partial S(r,t)}{\partial t} + \epsilon' r \cdot \frac{\partial S}{\partial r} - \Delta\epsilon\Delta r \frac{\partial S}{\partial r} = \epsilon S(r,t). \tag{F.6}$$

is a first-order transport partial differential equation, with a reaction term ($\epsilon S(r,t)$) and with a transport current going in the positive direction of the r axis (since the transport velocity $\epsilon' r - \Delta\epsilon\Delta r$ is positive, in the equation written in the form (F.6), as $\Delta r << r$).

The solution of this p.d.e. has the form

$$S(r,t) = \left[\epsilon' r - \Delta\epsilon\Delta r\right]^{\epsilon/\epsilon'} \cdot f\left(\frac{\epsilon' r - \Delta\epsilon\Delta r}{e^{\epsilon' t}}\right), \tag{F.7}$$

where f is an arbitrary function.

Assuming an initial condition $S(r,t) = S_0(r)$, using the above solution at $t = 0$ one has

$$S_0(r) = \left[\epsilon' r - \Delta\epsilon\Delta r\right]^{\epsilon/\epsilon'} \cdot f\left(\epsilon' r - \Delta\epsilon\Delta r\right), \tag{F.8}$$

from which

$$f(r) = \frac{1}{r^{\epsilon/\epsilon'}} S_0\left(\frac{r + \Delta\epsilon\Delta r}{\epsilon'}\right) \tag{F.9}$$

that, substituted into Eq. (F.7) gives

$$S(r,t) = e^{\epsilon t} \cdot S_0\left(\frac{\epsilon' r - \Delta\epsilon\Delta r}{\epsilon'} e^{-\epsilon' t} + \frac{\Delta\epsilon\Delta r}{\epsilon'}\right). \tag{F.10}$$

The characteristics of the solution, that is the lines at equal S, proceed in time as

$$\frac{\epsilon' r - \Delta\epsilon\Delta r}{\epsilon'} e^{-\epsilon' t} + \frac{\Delta\epsilon\Delta r}{\epsilon'} = G, \tag{F.11}$$

where G is a constant. Consequently, the maximum enrichment point moves as

$$r(t) = \left(G - \frac{\Delta\epsilon\Delta r}{\epsilon'}\right) e^{\epsilon' t} + \frac{\Delta\epsilon\Delta r}{\epsilon'} \tag{F.12}$$

Bibliography

[1] F. Marchetti, 2015. Modelling genetic regulatory networks with delay (Master thesis).

[2] D. Latchman, 2005. Gene regulation: a eukaryotic perspective. Garland Science.

[3] J. van Helden, 2000. Representing and analysing molecular and cellular function using the computer. Biological Chemistry, vol. 381.

[4] F. Kolpakov, 1999. Genenet: A gene network database and its automated visualization. Bioinformatics, vol. 14.

[5] M. Samsonova, 1998. Genet: A database of genetic networks. Proc. First Int. Conf. Bioinformatics Genome Regulation Struct..

[6] M. Kanehisa, S. Goto, 2000. Kegg: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research, vol. 28.

[7] H. Salgado, J. van Helden, 2000. Regulondb (version 2.0): A database on transcriptional regulation in escherichia coli. Nucleic Acids Research, vol. 27.

[8] Elowitz MB, Levine AJ, Siggia ED, Swain PS; 2002. Stochastic gene expression in a single cell. Science. 297 (5584): 1183-6.

[9] Blake WJ, Kaern M, Cantor CR, Collins JJ; 2003. Noise in eukaryotic gene expression. Nature. 422 (6932): 633-7.

[10] Gershenson, Carlos, 2004. Introduction to Random Boolean Networks. In Bedau, M., P. Husbands, T. Hutton, S. Kumar, and H. Suzuki (eds.) Workshop and Tutorial Proceedings, Ninth International Conference on the Simulation and Synthesis of Living Systems (ALife IX), 160-173.

[11] H. M. Adams, L. Shapiro, 1995. Circuit simulation of genetic networks. Science, vol. 269, no. 5224.

[12] Ashkenazi R, Gentry SN, Jackson TL, 2008. Pathways to tumorigenesis-modeling mutation acquisition in stem cells and their progeny. Neoplasia 10: 1170-1182.

[13] Alon, Uri, 2007. An Introduction to Systems Biology: Design Principles of Biological Circuits. ed. Boca Raton, FL: Chapman & Hall.

[14] Chu, Dominique; Zabet, Nicolae Radu; Mitavskiy, Boris; 2009. Models of transcription factor binding: Sensitivity of activation functions to model assumptions. Journal of Theoretical Biology. 257 (3): 419-429.

[15] A. V. Hill, 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. Journal of Physiology, vol. 40.

[16] G. Tiana, M. Jensen, 2007. Oscillations and temporal signalling in cells. Physical Biology, vol. 18, no. 4, 2007.

[17] Y. Haupt, R. Maya, A. Kazaz, M. Oren, 1997. Mdm2 promotes the rapid degradation of p53. Nature, 387, 296.

[18] G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, M. B. Elowitz, U. Alon, 2004. Dynamics of the p53-Mdm2 feedback loop in individual cells. Nature Genetics 36, 147 .

[19] Hirata H, Yoshiura S, Ohtsuka T, Bessho Y, Harada T, Yoshikawa K, Kageyama R, 2002. Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. Science 298, 840-3.

[20] A. Hoffman, A. Levchenko, M. L. Scott, D. Baltimore, 2002. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science 298, 1241 .

[21] M. L. Dequeant, E. Glynn, K. Gaudenz, M. Wahl, J. Chen, 2006. A complex oscillating network of signaling genes underlies the mouse segmentation clock. Science 314, 1595 .

[22] Vogelstein, B., Lane, D. and Levine, A.J., 2000. Surfing the p53 network. Nature, 408, 307-310.

[23] Chadwick D J, Goode J A (ed) 2003. Mechanisms and Biological Significance of Pulsatile Hormone Secretion Series: Novartis Foundation Symp. (New York: Wiley).

[24] Pourquié O, 2003. The segmentation clock: converting embryonic time into spatial pattern. Science 301, 328-30.

[25] M. Moreno-Risueno, 2010. Oscillating gene expression determines competence for periodic arabidopsis root branching. Science, vol. 329.

[26] R. Thomas, 1998. Laws for the dynamics of regulatory networks. International Journal Of Developmental Biology, vol. 42.

[27] T. Cover, J. Thomas, 2006. Elements of Information Theory, 2nd Edition. Wiley.

[28] G. Tkacik, W. Bialek, 2008. Information flow and optimization in transcriptional regulation. Proceedings of the National Accademy of Science, vol. 105.

[29] C. Shannon, 1949. "Communication in the presence of noise," Proceedings of the IRE, vol. 37.

[30] G. Tiana, M. H. Jensen, 2013.The dynamics of genetic control in the cell: the good and bad of being late. Phylosophical Transaction A, vol. 19.

[31] Mallet-Paret J., Sell G.R., 1996. The Poincaré-Bendixson Theorem for Monotone Cyclic Feedback Systems with Delay. Journal of Differential Equations, 125, 441-489.

[32] R. Guantes, J. Poyatos, 2006. Dynamical principles of two-component genetic oscillators. PLos Computational Biology, vol. 2.

[33] J. Lewis, 2003. Autoinhibition with transcriptional delay: A simple mechanism for the zebrafish somitogenesis oscillator. PLos Computational Biology, vol. 13.

[34] Krishna S, Andersson A M C, Semsey S, Sneppen K, 2006. Structure and function of negative feedback loops at the interface of genetic and metabolic networks. Nucl. Acids Res. 34, 2455-62.

[35] Schnarr M et al, 1991. DNA binding properties of the LexA repressor. Biochimie, 73, 423-31.

[36] Strogatz S, 1994. Nonlinear Dynamics and Chaos (Reading, MA: Addison-Wesley).

[37] Snoussi E H, 1998. Necessary conditions for multistationarity and stable periodicity. J. Biol. Sys. 6, 3-9.

[38] Gouzé J L, 1998. Positive and negative circuits in dynamical systems J. Biol. Syst. 6, 11-5.

[39] Mallet-Paret J, Smith H L, 1990. The Poincaré-Bendixson theorem for monotone cyclic feedback systems. J. Dyn. Diff. Eqns 2, 367-421.

[40] Hetherington, JPJ; Warner, A; Seymour, RM; 2006. Simplification and its consequences in biological modelling: conclusions from a study of calcium oscillations in hepatocytes. J R Soc Interface , 3 (7) 319-331.

[41] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, 1953. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. 21, 1087.

[42] B. Schwanhusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, 2011. Global quantification of mammalian gene expression control. Nature, 473, 337.

[43] M. Galassi et al., 2009. GNU Scientific Library Reference Manual, 3rd Ed. (Network Theory Ltd, London).

[44] S. Pigolotti, 2009. Symbolic dynamics of biological feedback networks. Physical Review Letters, vol. 102.

[45] S. Pigolotti and M. Jensen, 2007. Oscillation patterns in negative feedback loops. Proceedings of the National Accademy of Science, vol. 104, no. 16.

[46] R. Swendsen J. Wang, 1986. Replica Monte Carlo Simulation of Spin-Glasses. Phys. Rev. Lett. 57, 2607.

[47] D. Frenkel, 2002. Understanding Molecular Simulations, 2nd Edition (Academic Press, San Diego)

[48] A. Ferrenberg, R. Swendsen, 1989. Optimized Monte Carlo data analysis. Phys. Rev. Lett. 63, 1195.

[49] G. Tiana, L. Sutto, 2011. Equilibrium properties of realistic random heteropolymers and their relevance for globular and naturally unfolded proteins. Phys. Rev. E 84 , 061910.

[50] C. Simon, G. Brandenberger, 2002. Ultradian oscillations of insulin secretion in humans. Diabetes 51, S258.

[51] F. Vasile, E. Pechkova, C. Nicolini, 2008. Solution structure of the beta-subunit of the translation initiation factor aIF2 from archaebacteria Sulfolobus solfataricus. Proteins 70, 1112.

[52] T. Ikegami, Y. Maruyama, H. Doi, A. Hattori, H. Ando, 2015. Ultradian oscillation in expression of four melatonin receptor subtype genes in the pineal gland of the grass puffer, a semilunar-synchronized spawner, under constant darkness. Front. Neurosci. 9, 9.

[53] H. Han et al., 2015. TRRUST: a reference database of human transcriptional regulatory interactions. Sci. Rep. 5, 11432.

[54] M. Ronen, R. Rosemberg, B. I. Shraiman, U. Alon, 2002. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. Proc. Natl. Acad. Sci. USA 99, 10555.

[55] Preetha Anand, Ajaikumar B. Kunnumakara, Chitra Sundaram, Kuzhuvelil B. Harikumar, Sheeja T. Tharakan, Oiki S. Lai, Bokyung Sung, Bharat B. Aggarwal, 2008. Cancer is a Preventable Disease that Requires Major Lifestyle Changes. Pharm Res. 25(9): 2097-2116.

[56] Global status report on noncommunicable diseases 2014. World Health Organization.

[57] Fearon ER, Vogelstein B.(1990). A genetic model for colorectal tumorigenesis. Cell; 61:759-767.

[58] Nowell PC.(1976). The clonal evolution of tumor cell populations. Science; 194: 23-28.

[59] Sell S., 2004. Stem cell origin of cancer and differentiation therapy. Crit Rev Oncol Hematol; 51:1-28.

[60] Durante F., 1874. Nesso fisio-pathologico tra la struttura dei nei materni e la genesi di alcuni tumori maligni. Arch Memor Observ Chir Pract; 11:217-26.

[61] Cohnheim J., 1867. Ueber entzundung und eiterung. Path Anat Physiol Klin Med; 40:1-79.

[62] Lapidot, T. et al. 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 367, 645-648 .

[63] Bonnet, D., Dick, J. E. Human, 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 3, 730-737.

[64] S. Bao, Q. Wu, R.E. McLendon, Y. Hao, Q. Shi, A.B. Hjelmelandet al, 2006. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature, 444, 756-760.

[65] L Vermeulen, MR Sprick, K Kemper, G Stassi and JP Medema, 2008. Cancer stem cells - old concepts, new insights. Cell Death and Differentiation 15, 947-958.

[66] Al Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF, 2003. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci USA; 100:3983-3988.

[67] Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T et al., 2004. Identification of human brain tumour initiating cells. Nature; 432: 396-401.

[68] Dingli, D., Michor, F., Antal, T., Pacheco, J. M., 2007. The emergence of tumor metastases. CancerBiol. Ther. 6, 383-390.

[69] O' Brien CA, Pollett A, Gallinger S, Dick JE. 2007. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature; 445: 106-110.

[70] Wicha MS, Liu S, Dontu G., 2006. Cancer stem cells: an old idea-a paradigm shift. Cancer Res; 66: 1883-1890.

[71] Alessandro L. Sellerio, Emilio Ciusani, Noa Bossel Ben-Moshe, Stefania Coco, Andrea Piccinini, Christopher R. Myers, James P. Sethna, Costanza Giampietro, Stefano Zapperi, Caterina A. M. La Porta, 2015. Overshoot during phenotypic switching of cancer cell populations. *Sci Rep*, **5** 15464.

[72] Driessens, G., Beck, B., Caauwe, A., Simons, B. D., Blanpain, C., 2012. Defining the mode of tumour growth by clonal analysis. Nature 488, 527-30.

[73] Chen, J. et al., 2012. A restricted cell population propagates glioblastoma growth after chemotherapy. Nature 488, 522-6.

[74] Hill RP., 2006. Identifying cancer stem cells in solid tumors: case not proven. Cancer Res ; 66: 1891-1895.

[75] Matthew D. Johnston, Philip K. Maini, S. Jonathan Chapman , Carina M. Edwards, Walter F. Bodmer, 2010. On the proportion of cancer stem cells in a tumour .Journal of Theoretical Biology 266, 708-711.

[76] Ricci-Vitiani, L., et al., 2007. Identification and expansion of human colon-cancer-initiating cells. Nature 445, 111-115.

[77] Quintana, E., Shackleton, M., Sabel, M. S., Fullen, D. R., Johnson, T. M., Morrison, S. J., 2008. Efficient tumour formation by single human melanoma cells. Nature, 456, 593-598

[78] Kelly, P.N., Dakic, A., Adams, J.M., Nutt, S.L., Strasser, A., 2007. Tumor growth need not be driven by rare cancer stem cells. Science 317, 337.

[79] Yeung, T.M., Gandhi, S.C., Wilding, J.L., Muschel, R., Bodmer, W.F., 2010. Cancer stem cells from colorectal cancer-derived cell lines. Proc. Natl. Acad. Sci. USA 107, 3722-3727.

[80] Andrea Sottoriva, Joost J.C. Verhoeff, Tijana Borovski, et al., 2010. Cancer Stem Cell Tumor Model Reveals Invasive Morphology and Increased Phenotypical Heterogeneity. Cancer Res;70:46-56.

[81] J.E. Visvader, G.J. Lindeman, 2008. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nature Reviews Cancer 8, 755-768.

[82] Rodolfo Molina-Pena, Mario Moises Alvarez, 2012. A Simple Mathematical Model Based on the Cancer Stem Cell Hypothesis Suggests Kinetic Commonalities in Solid Tumor Growth. PLoS One; 7(2): e26233.

[83] Morrison SJ, Kimble J, 2006. Asymmetric and symmetric stem-cell divisions in development and cancer. Nature 441: 1068-1074.

[84] Soltysova A, Altanerova V, Altaner C, 2005. Cancer stem cells. Neoplasma 52:435-440.

[85] Pacheco JM, Traulsen A, Dingli D, 2009. The allometry of chronic myeloid leukemia. J Theor Biol 259: 635-640.

[86] Morrison SJ, Shah NM, Anderson DJ,1997. Regulatory mechanisms in stem cell biology. Cell 88: 287-298.

[87] C. Ginestier, M.S. Wicha, 2007. Mammary stem cell number as a determinate of breast cancer risk. Breast Cancer Res., 9, 109.

[88] Beretta, E.; Capasso V.; Morozova V., 2012. Mathematical Modelling of Cancer Stem Cells Population Behavior. Mathematical Modelling of Natural Phenomena 7, 1, 279-305.

[89] La Porta, C. A. M., Zapperi, S. and Sethna, J. P., 2012. Senescent cells in growing tumors: population dynamics and cancer stem cells. PLoS Comput Biol 8, e1002316.

[90] Boman BM, Wicha MS, Fields JZ, Runquist OA, 2007. Symmetric division of cancer stem cells-a key mechanism in tumor growth that should be targeted in future therapeutic approaches. Clin Pharmacol Ther 81: 893-898.

[91] Heiko Enderling, Lynn Hlatky, Philip Hahnfeldt, 2013. Cancer Stem Cells: A Minor Cancer Subpopulation that Redefines Global Cancer Features. Front Oncol.; 3:76.

[92] Caussinus, E., and Hirth, F., 2007. Asymmetric stem cell division in development and cancer. Prog. Mol. Subcell. Biol. 45, 205-225.

[93] Lathia, J. D., Hitomi, M., Gallagher, J., Gadani, S. P., Adkins, J., Vasanji, A., et al., 2011. Distribution of CD133 reveals glioma stem cells self-renew through symmetric and asymmetric cell divisions. Cell Death Dis. 2, e200.

[94] C. Turner, A.R.Stinchcombe, M. Kohandel, S. Singh, S. Sivaloganathan, 2009. Characterization of brain cancer stem cells: a mathematical approach. Cell Prolif. 42, 529-540.

[95] Gao, X., McDonald, J. T., Hlatky, L., Enderling, H., 2013. Acute and fractionated irradiation differentially modulate glioma stem cell division kinetics. Cancer Res. 73, 1481-1490.

[96] Sottoriva, A., Spiteri, I., Shibata, D., Curtis, C., Tavaré, S., 2013. Single-molecule genomic data delineate patient-specific tumor profiles and cancer stem cell organization. Cancer Res. 73, 41–49.

[97] d'Onofrio A, Tomlinson IP., 2006. A nonlinear mathematical model of cell turnover, differentiation and tumorigenesis in the intestinal crypt. J Theor Biol. 2007 Feb 7; 244(3):367-74.

[98] Leder K, Holland EC, Michor F., 2010. The therapeutic implications of plasticity of the cancer stem cell phenotype. PLoS One; 5(12).

[99] Weekes, S.L., Barker, B., Bober, S. et al., 2014. A Multicompartment Mathematical Model of Cancer Stem Cell-Driven Tumor Growth Dynamics. Bull Math Biol, 76: 1762.

[100] Monika Joanna Piotrowska, Heiko Enderling, Uwe an der Heiden, Michael C. Mackey, 2008. Mathematical Modeling of Stem Cells Related to Cancer. In: Cancer and Stem Cells, Editor: Thomas Dittmar and Kurt S. Zänker. Nova Science Publishers, Inc.

[101] J S Lowengrub et al, 2010. Nonlinear modelling of cancer: bridging the gap between cells and tumours. Nonlinearity 23 R1.

[102] Enderling H, Anderson AR, Chaplain MA, Beheshti A, Hlatky L, Hahnfeldt P., 2009. Paradoxical dependencies of tumor dormancy and progression on basic cell kinetics. Cancer Res. 2009 69(22):8814-21.

[103] H. Enderling, D. Park, L. Hlatky, P. Hahnfeldt, 2009. The Importance of Spatial Distribution of Stemness and Proliferation State in Determining Tumor Radioresponse. Math. Model. Nat. Phenom. Vol. 4, No. 3, pp. 117-133.

[104] Gupta, P. B. et al., 2011. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 146, 633-44.

[105] Roesch, A. et al. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. Cell 141, 583-94.

[106] Quintana, E. et al, 2010. Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. Cancer Cell 18, 510-23.

[107] Tonekaboni S.A. M., Dhawan A. Kohandel M.(2016). Mathematical modelling of plasticity and phenotype switching in cancer cell populations. Mathematical Biosciences 283, 30-37.

[108] N. D. Marjanovic, R. A. Weinberg, C. L. Chaffer, 2013. Cell Plasticity and Heterogeneity in Cancer. Clinical Chemistry 59:1 168-179.

[109] Chaffer, C. L. et al, 2013. Poised chromatin at the zeb1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. Cell 154, 61-74.

[110] Monzani, E. et al.,2007. Melanoma contains CD133 and ABCG2 positive cells with enhanced tumourigenic potential. Eur J Cancer 43, 935-946.

[111] Schatton, T. et al., 2008. Identification of cells initiating human melanomas. Nature 451, 345-349.

[112] Held, M. A. et al., 2010. Characterization of melanoma cells capable of propagating tumors from a single cell. Cancer Res 70, 388-97.

[113] S. Zapperi, C. A. M. La Porta, 2012. Do cancer cells undergo phenotypic switching? The case for imperfect cancer stem cell markers. Scientific Reports 2:441.

[114] A Eramo, F Lotti, G Sette et al., 2008. Identification and expansion of the tumorigenic lung cancer stem cell population. Cell Death and Differentiation volume 15, 504-514.

[115] Vlachos, I. S. et al., 2012. Diana mirpath v.2.0: investigating the combinatorial effect of micrornas in pathways. Nucleic Acids Research 40, W498-W504.

[116] C. Turner, M. Kohandel, 2010. Investigating the link between epithelial-mesenchymal transition and the cancer stem cell phenotype: A mathematical approach. Journal of Theoretical Biology 265,329335.

[117] Hollier, B.G., Evans, K., Mani, S.A., 2009. The Epithelial-to-Mesenchymal Transi-

tion and Cancer Stem Cells: A Coalition Against Cancer Therapies. J Mammary Gland Biol Neoplasia (2009) 14:29.

[118] Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al., 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 133:704-15

[119] A.P. Morel, M. Lièvre, C. Thomas, G. Hinkal, S. Ansieau, A. Puisieux, 2008. Generation of breast cancer stem cells through epithelial–mesenchymal transition. PLoS One, 3, p. e2888.

[120] V.L. Battula, K.W. Evans, B.G. Hollier, Y. Shi, F.C. Marini, A. Ayyanan, et al., 2010. Epithelial-mesenchymal transition-derived cells exhibit multilineage differentiation potential similar to mesenchymal stem cells. Stem Cells, 28, pp. 1435-1445.

[121] J. Yang, R.A. Weinberg, 2008. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Dev. Cell, 14, pp. 818-829.

[122] C. Turner, M.Kohandel, 2012. Quantitative approaches to cancer stem cells and epithelial-mesenchymal transition. Seminars in Cancer Biology, 22, 5-6, 374-378.

[123] Sanjay de Mel, Shou- Nian Choo, Shuangyi Fan, Hoang Mai Phuong, Michal Hoppe, Wee Joo Chng, Anand Jeyasekharan, Siok Bian Ng, 2017. Quantitative Analysis of Multiplexed Immunofluorescence in T-Cell Lymphoma. Blood 130:2800.

[124] Qiuhui Li, Kiera Rycaj, Xin Chen, Dean G. Tang, 2015. Cancer stem cells and cell size: A causal link? Seminars in Cancer Biology, Volume 35, 191-199.

[125] Gillespie, D. T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics, 22(4), 403-434.

[126] Tzedakis, G., Tzamali, E., Marias, K., Sakkalis, V. (2015). The Importance of Neighborhood Scheme Selection in Agent-based Tumor Growth Modeling. Cancer Informatics, 14 (Suppl 4), 67-81.

[127] Misteli T., 2008. Physiological importance of RNA and protein mobility in the cell nucleus. Histochem Cell Biol. 129(1):5-11.

[128] Pulliam L, Gupta A., 2015. Modulation of cellular function through immune-activated exosomes. DNA and cell biology; 34:459-463. [

[129] Qiwei Yang, Michael P. Diamond, Ayman Al-Hendy, 2016. The emerging role of extracellu-lar vesicle-derived miRNAs: implication in cancer progression and stem cell related diseases. J Clin Epigenet; 2(1):13.

[130] Rodriguez, A., & Laio, A., 2014. Machine learning. Clustering by fast search and find of density peaks. Science, 344(6191), 1492-1496. http://doi.org/10.1126/science.124207

[131] H. Youssefpour, X. Li, A.D. Lander, J.S. Lowengrub, 2012. Multispecies model of cell lineages and feedback control in solid tumors. J Theor Biol 304: 39-59.

[132] Vlashi E, Kim K, Lagadec C, Donna LD, Mcdonald JT, Eghbali M, Sayre JW, Stefani E, Mcbride W, Pajonk F, 2009. In Vivo Imaging, Tracking, and Targeting of Cancer Stem Cells. JNCI Journal of the National Cancer Institute; 101(5):350-359.

[133] Dhawan A, Kohandel M, Hill R, Sivaloganathan S, 2014. Tumour Control Probability in Cancer Stem Cells Hypothesis. PLoS ONE 9(5).