An In silico Model to Demonstrate the Effects of Maspin on Cancer Cell Dynamics

[*]M A Al-Mamun[a], D M Farid[b], L Ravenhil[c], M A Hossain[d], C Fall [e] and R Bass[c,e]

[a*]Department of Population Medicine & Diagnostic Sciences, College of Veterinary Medicine,

Cornell University, Ithaca, NY 14850, USA

[b]Department of Computer Science & Engineering, United International University, Bangladesh

[c]Department of Applied Sciences, Faculty of Health and Life Sciences,

University of Northumbria at Newcastle, UK

[d]Anglia Ruskin IT Research Institute (ARITI), Anglia Ruskin University, Cambridge, UK

[e]Computational Intelligence Group, Faculty of Engineering and Environment,

University of Northumbria at Newcastle, UK

ma875@cornell.edu;dewanfarid@cse.uiu.ac.bd;alamgir.hossain

@anglia.ac.uk ;{charles.fall; rosemary.bass}@northumbria.ac.uk

*Corresponding author: ma875@cornell.edu

Abstract

Most cancer treatments efficacy depends on tumor metastasis suppression, where tumor suppressor genes play an important role. Maspin (Mammary Serine Protease Inhibitor), an non-inhibitory serpin has been reported as a potential tumor suppressor to influence cell migration, adhesion, proliferation and apoptosis in in vitro and in vivo experiments in last two decades. Lack of computational investigations hinders its ability to go through clinical trials. Previously, we reported first computational model for maspin effects on tumor growth using artificial neural network and cellular automata paradigm with in vitro data support. This paper extends the previous in silico model by encompassing how maspin influences cell migration and the cell-extracellular matrix interaction in subcellular level. A feedforward neural network was used

to define each cell behavior (proliferation, quiescence, apoptosis) which followed a cell-cycle algorithm to show the microenvironment impacts over tumor growth. Furthermore, the model concentrates how the in silico experiments results can further confirm the fact that maspin reduces cell migration using specific in vitro data verification method. The data collected from in vitro and in silico experiments formulates an unsupervised learning problem which can be solved by using different clustering algorithms. A density based clustering technique was developed to measure the similarity between two datasets based on the number of links between instances. Our proposed clustering algorithm first finds the nearest neighbors of each instance, and then redefines the similarity between pairs of instances in terms of how many nearest neighbors share the two instances. The number of links between two instances is defined as the number of common neighbors they have. The results showed significant resemblances with in vitro experimental data. The results also offer a new insight into the dynamics of maspin and establish as a metastasis suppressor gene for further molecular research.

Keywords: SERPINB5, Serine Protease Inhibitor, Cellular Automata, Clustering, Tumor Growth

1. Introduction

Cancer is a complex disease because of its involvement with different biological processes at the cellular and subcellular level (Hanahan and Weinberg, 2011). Genetic mutations lead to abnormal cell proliferation and a mass initially called a benign tumor which has a localized position. When the tumor spreads to distal parts of the body it is called as malignant tumor. This happens by a process whereby control of key biological processes is eroded i.e. invasion, metastasis and angiogenesis. Alternatively, tumor life span can be divided into three stages: avascular phase (diffusion limited stage), vascular phase (angiogenesis stage) and finally metastatic phase (spreading to the distal body parts) (Zou et al., 1994). At the molecular level, these stages involve both intracellular and extracellular mechanisms to influence proliferation, cell-cell adhesion, cell migration, cell-extracellular matrix (ECM) interactions and angiogenesis. Overall, the efficacy of most existing treatments depends on metastasis suppression. Maspin or SERPINB5 is a 42 kDa protein that has been characterized as a type II tumor metastasis

suppressor. Maspin expression is downregulated in breast, prostate, gastric and melanoma cancers but overexpressed in pancreatic, gallbladder, colorectal, and thyroid cancers, suggesting that maspin may play a prognostic roles on different cell types (Bodenstine et al., 2012; Berardi et al., 2013). Maspin decreases tumor growth and metastasis in in vivo (Zou et al., 1994) and invasion in in vitro (Biliran and Sheng, 2001). This is achieved by the ability of maspin to influence aspects of cell behavior including migration, invasion, proliferation, angiogenesis and apoptosis. These effects have been investigated in many in vitro models (Sheng et al., 1996; Biliran and Sheng, 2001; Ngamkitidechakul et al., 2003; Bailey et al., 2005; Yin et al., 2006; Bass et al., 2009; Ravenhill et al., 2010) and in vivo models to investigate the intracellular (Zou et al., 1994; Zhang et al., 2000 and Shi et al., 2001) and extracellular (Cell et al., 2006; Bass et al., 2009; Ravenhill et al., 2010; Endsley et al., 2011) activities of maspin.

The clinical data regarding maspin expression are variable. In this case, computational models can be built to support the in vitro and in vivo hypotheses to establish the effects of maspin on basic cell migration, cell invasion and understand the metastasis suppression mechanisms. Normally, in silico models integrate the complex multiple processes of a biological system and build bridges among multiple spatial and temporal scales with a deeper understanding (Kam et al., 2012; Anderson et al., 2013). From our best knowledge, the diversity of maspin has been investigated for the first time from the computational point of view in our previous model (Al-Mamun et al., 2013b), where the model showed its potential engagement with multiple cellular phenomena using artificial neural network (ANN) and cellular automata (CA) modeling techniques.

1.1. Biological Background

Tumor cell migration is the movement of a cell through or on a surface of ECM. Cells need attachment sites on extracellular matrices in order to re-organize their cytoskeleton and initiate protrusions. Tumor cells require a well regulated peri-cellular proteolysis to migrate. Proteolysis is a process where ECM contents are degraded by cellular enzymes. It is now widely believed that the breakdown of these barriers is catalyzed by proteolytic enzymes released from the invading tumor. Most of these proteases belong to

one of two general classes: matrix metalloproteases (MMPs) and serine proteases. This regulation system consists of the urokinase plasminogen activator (uPA), urokinase receptor (uPAR), ECM and plasminogen activator inhibitors: type-1 (PAI 1) and type-2 (PAI 2). uPA is an extracellular serine protease produced from cells as a single chain pro-enzyme pro-uPA (Andreasen et al., 2000). Generally, the protease moiety activates plasminogen and generates plasmin. Plasmin is a serine protease capable of digesting basement membrane and ECM proteins. Plasmin itself has complex mechanism where it is not only catalyzes the breakdown of many of the known ECM and basement membrane molecules (such as vitronectin, fibrin, laminin and collagens), but also activates metalloproteinases. Naturally, the unrestrained generation of plasmin is potentially hazardous to cells. In a healthy organism the process of plasminogen activation is strictly controlled through the availability of plasminogen activators (PAs), localized activation and interaction with specific inhibitors (PAIs) (Andreasen et al., 2000). But in cancer cells this control mechanism gets affected due to different alterations in the cellular environments.

Maspin is a potential protease inhibitor with its multicellular functionalities. Exogenous maspin decreases proliferation and increases cell adhesion in vitro (Ngamkitidechakul et al., 2001). It inhibits angiogenesis in vivo (Zhang et al., 2000) and causes apoptosis when expressed in endothelial cells (Li et al., 2005). In addition, it has been shown that maspin can inhibit the migration of vascular smooth muscle cells (VSMCs) (Bass et al., 2009). Ravenhill et al., (2010) showed the G-helix is necessary and sufficient for maspin migration effects. Integrins have complex roles in regulating cell motility and migration. Previously, it was shown that maspin binds to cell surface integrins, in particular α5β1. This binding causes inactivating conformational changes in the integrin and leads to the inhibitory effect of maspin on cell migration (Bass et al., 2009). Another anti-invasive mechanism has been suggested that maspin targets the uPA/ uPAR complex at the cell surface. Although maspin does not directly inhibit uPA activity, it has shown to reduce cell surface associated uPA/uPAR by inducing its internalization (Biliran and Sheng, 2001; Yin et al., 2006). This finding has been supported by Amir et al., (2005) which indicated that transfected maspin was able to reduce hypoxia-induced uPA/uPAR expression in

MDAMB-231 cells. It has also been shown that maspin binds the pro-uPA zymogen, inhibiting its activation (Yin et al., 2006).

Recently, the possible interaction of maspin with uPA system was investigated by (Endsley et al., 2011), where it was indicated that two proposed pathways can be utilized by maspin to increase cell-ECM adhesion; that is, the plasminogen activation system and β1 integrin signaling. But still the extracellular presence of maspin is in debate according to a recent study, where it has been stated that maspin has no influence on tumor growth and embryonic development (Teoh et al., 2014). These motivated us to extend our previous cellular level model with cell-cell adhesion and cell-ECM interaction to define how maspin reduces cell migration by showing potential enhancement with ECM components (fibronectin, laminin, collagen etc.).

2. Related work

In silico models have a tremendous ability to handle multiple dynamic interacting variables, such as numerous cell types or various environmental factors. Innovative computational modeling and simulation, in addition to appropriately designed biological experiments can facilitate a powerful tool to refine high-throughput biological data, hypotheses and more accurate predictions (Macklin and Lowengrub, 2007; Kam et al., 2012; Edelman et al., 2010; Deisboeck et al., 2011;Johnson et al., 2013). These models can be formulated from the concept of biological spatial spaces: atomic, molecular, microscopic, and macroscopic (Anderson et al., 2011). In relation with this study, some CA models are essential for understanding tumor dynamics. A CA model of avascular tumor invasion investigated the migratory behavior of the cancer cells using a set of partial differential equations (PDEs) and that it coupled the cell dynamics to continuous fields of oxygen, ECM and matrix degrading enzymes (MDEs) (Anderson, 2005). Another model revealed that MDEs degrade surrounding ECM and tumor cells migrate through ECM gradients during tumor invasion (Conde et al., 2008). Other CA models investigated some other important factors: impact of oxygen concentrations, dynamics of solid tumor growth (Gerlee and Anderson, 2007) and the evolutionary aspects of glycolytic phenotype (Smallbone et al., 2007; Gerlee

and Anderson, 2008; 2009) and emergence of the glycolytic phenotype which suggested that the oxygen concentration and matrix density have combined effects on the tumor morphology. Recently, a hybrid in silico model calculated tumor mass with consideration of oxygen, glucose, ECM, cell-cell adhesion and cell movement as key micro environmental parameters (Kazmi et al., 2012a). The model includes more microenvironment parameters such as protein expression, growth promoters/inhibitors and a bioreductive drug tirapazamine (TPZ) transport model for hypoxic tumor cells on top of the Anderson's model (Anderson, 2005). In another study, the same authors showed that diminished drug transport is one of the key reasons for TPZ failure, which needs further optimization of the drug transport properties in the emerging TPZ generations (Kazmi et al., 2012b). To date, our previous model presented maspin effects on tumor growth dynamics using a hybrid computational framework where maspin was included to show its effects on proliferation, cell migration, invasion, and apoptosis (Al-Mamun et al., 2013b). The model suggested 10-40% cell migration reduction and 20-30% cell proliferation enhancement due to maspin in in silico environment. Later on, several in vitro experiments were presented to show the resemblances with the in silico results.

Essentially, our previous model did not consider cell-matrix adhesion and cell-cell adhesion during the tumor growth simulation subcellularly. Moreover the data resemblances method was based on statistical t-test, which can only present the variation of mean value of cell migration data. Particularly, this investigation introduces an extension where it considers the impacts of maspin on cell-cell adhesion and cell-ECM phenomenon. Also this model provides a new in silico data verification method to see the resemblances with a previous in vitro model (Ravenhill et al., 2010). The key objective of this paper is not only to provide support the extracellular presence of maspin in tumor cell and but also to provide a data resemblances method with in vitro lab data. More specifically, the assumptions taken into consideration for building model are maspin reduces cell migration while present exogenously (Bass et al., 2009 and Endsley et al., 2011), it increases cell-ECM adhesion while it reduces the cell migration (Ravenhill et al., 2010; Endsley et al., 2011) and it reduces the overall tumor growth (from Zou et al., 1994; Zhang et al., 2000).

The model presented here is investigating the complex interaction of maspin connection with the different microenvironment parameters but especially the cell-ECM components and cell-cell adhesion. For supporting our computational method, we presented a novel in vitro experiment of MCF7 breast cancer cell line adhesion assay to show maspin dynamics over different ECM components like collagen I, laminin and fibronectin. Results of our previous models showed various agreements with biological experiments: HT29 human colon carcinoma cell line growth on spheroids obtained from monolayer cultures (Kazmi et al., 2012a) and in vitro apoptosis assay for transiently transfected DU145 prostate cancer cell line (Al-Mamun et al., 2013b). Moreover, it is clear from the previous papers that maspin reduces cell migration and increase cell-cell adhesion where β1 integrin, ECM, E-cadherin, uPA-uPAR maspin complex have putative roles. These in vitro and in vivo results led us to build this model to show the cell-cell adhesion and cell-ECM interaction due to maspin effect on top of the previous computational models (Kazmi et al., 2012a; Anderson et al., 2013; Domschke et al., 2014). The proposed model was implemented, tested and verified through a set of experiments to demonstrate the merits and capabilities of the scheme. Previous tumor growth models for maspin showed only the dynamics explaining the underlying mechanisms like cell migration, cell invasion, apoptosis (Al-Mamun et al., 2012;2013a), but no model provided any data analysis techniques to establish maspin impacts on tumor growth (Gerlee and Anderson, 2007; Kazmi et al., 2012a;2012b; Al-Mamun et al., 2013b). This model proposed clustering methods to analyze the cell migration data attained from both in vitro and in silico experiments.

3. Tumor growth framework

3.1. Tumor growth model construction

Our extended model is hybrid in nature, where tumor cells are discrete individual entity and chemical fields are continuous. For modeling the basic tumor growth a tumor tissue was developed from four cells at the center of a 2D lattice grid. Each grid element was either occupied by an abnormal cell or was empty. The grid elements had local concentrations of oxygen and maspin. A set of PDEs have been used to demonstrate the interaction between cells and the local chemicals concentrations. Individual cell was

influenced by different tumor microenvironment constraints and each cell can take its own decision during its life time. This complex decision mechanism was applied by using a feedforward neural network (FNN). The output of the FNN defines the individual cell's behaviors: proliferation, quiescence, apoptosis and/ or cell movement. Depending on selected behaviors, then model performs cell cycle algorithm and nutrients consumption with time and space around the neighboring grid points. An abstraction of the model is given in figure 1.

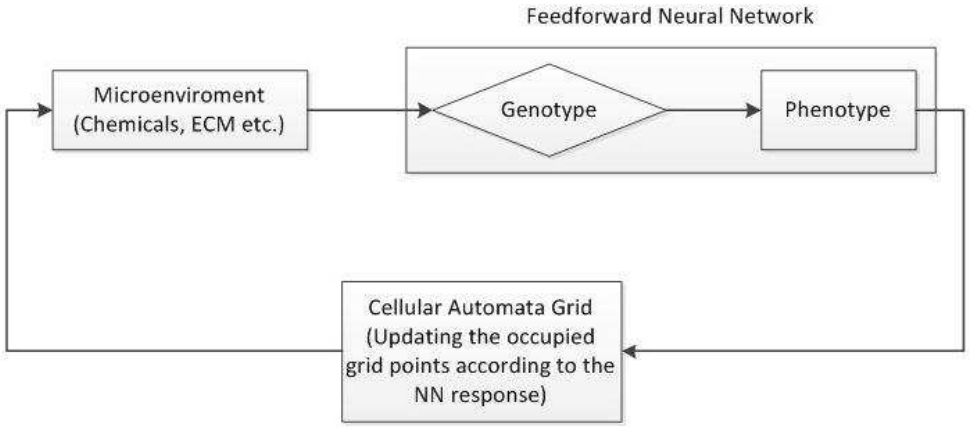


Fig. 1. Tumor growth model construction summary describes how a cell takes the microenvironment as an input and processed (genotype to phenotype) to the ultimate response of the cell. The resulting phenotypic changes have the potential to modify the microenvironment in CA grid which deals with all the parameters

The model considered normal tumor growth with an oxygen supply as oxygen is one of the key nutrients for the tumor growth and progression. For picturing the maspin effect, we have not considered hypoglycemia (lack of glucose) or acidity (lack of hydrogen). The concentrations of glucose and hydrogen ions did not vary in the reported in vitro experiments (Ravenhill et al., 2010). The varying distance of cells from a blood vessel can cause heterogeneity in the tumor microenvironment and the tumor mass. The model introduced external blood vessel by stating boundary conditions to PDEs. The boundary conditions simulated such an environment where the surrounding blood vessel of the tumor

supplied the nutrients and also removed the metabolic waste produced by the tumor cells. The evolution of oxygen and maspin with respect to time was maintained by the following second order PDEs

$$\frac{\partial O_2(x,t)}{\partial t} = D_{O_2}\Delta O_2(x,t) - f_{O_2}(x,t) \quad (1)$$

Where, D_{O_2} is the oxygen diffusion constant. For maspin, we have taken diffusion equation mentioned in (Gerlee and Anderson, 2007).

$$\frac{\partial M(x,t)}{\partial t} = D_M\Delta M(x,t) - f_M(x,t) \quad (2)$$

Where, D_M is the maspin diffusion constant. The term $f_l(x,t)$ was the utilization or production function of oxygen and maspin (l=O_2,M), for each cell at a specific position x and at time t and is described in equation 2.

$$f_l(x,t) = \begin{cases} 0 & \text{no tumor cell at that grid point} \\ cr_l F(x) & \text{grid point occupied by a tumor cell} \end{cases} \quad (3)$$

Where, cr_l is the base consumption/production rates, F(x) is the modulated energy consumed by the cell located at the grid element x and calculated in equation 3. It was used to report the differences for the energy consumptions among different subclones.

$$F(x) = max(k(R - T_r) + 1,0.25) \quad (4)$$

Where, k is the strength of modulation, R is the response of FNN and T_r is the target response. The term max(,0.25) shows that the cell's metabolism was at least a quarter of the base line consumption rate as considered by Gerlee and Anderson (2008). This function also ensured that the cell with the greatest network response would consume more nutrients. As this modelling scheme is taken as a standard procedure for growing avascular tumor in many studies (Anderson, 2005; Gerlee and Anderson, 2007;2008; Kazmi et al., 2012; Al-Mamun et al., 2013b).

3.2. Cell signaling nodes

Each cell behavior is defined by an intracellular signaling network which takes nutrient concentration inputs (oxygen and maspin) as a phenotype from the grid points. After signaling mechanism, it defines cells behavior as a genotypic response. A FNN mimicked this setup and normally it consists of number of nodes and it can take real number values as input. The nodes are constructed into three layers: input layer, hidden layer and output layers. The construction of proposed FNN is shown in figure 2. The functions of each layer

- Input layer nodes take chemical values from a particular grid point and number of empty neighbors of that point
- Hidden layer nodes take the input layer response and perform as an intermediate layer where there can be groups of co-regulated gene complex
- Output layer determines the fate of each cell which can be proliferation, quiescent, dead and/or movement

Different layer nodes are connected together with varying weights, determined by two matrices X and Y, and hidden layer nodes and output layer nodes consist of their internal threshold ϕ, θ (see figure 2). Further description of of X, Y, ϕ, and θ and the functional mechanism can be found in appendix A.1.

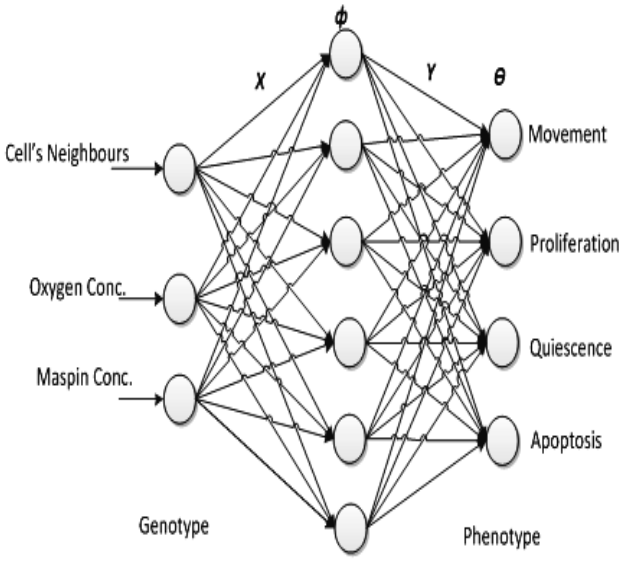


Fig. 2. Cellular signaling response network.

However, in case of real cells, the regulatory mechanisms are much more complex where hundreds of biochemical reactions take place; it has been shown in the literature that FNN with one hidden layer can approximate any continuous function (Bray, 1990; Kazmi et al., 2012b), if we add more intermediate layers nothing will change in the output. The presented network is able to calculate the output for an undetermined sample set that is not specified in the training set. This network was not trained, and instead the weights in the network were constructed by hand and were selected by an iterative process so that the network produced a predefined output. The network parameters have to be adjusted in such a manner that they summarize the behavior of actual cancer cells (Gerlee and Anderson, 2007).

3.3. Cell cycle rules

The individual part of the model took the input vector (environment parameters) for each cell and sampled from the grid point. Cell behavior was determined as output from FNN calculation. There were four nodes in output: proliferation, quiescent, apoptosis and movement. Each cell consumed oxygen (considered constant exogenous maspin supply) or maspin (considered constant oxygen supply) according to its behaviour (FNN response) in automaton points. Then the amount of oxygen gets checked and if there is no sufficient oxygen present the cell dies from necrosis. If a cell dies once, it is no longer updated. The life-cycle action decided by the network is carried out:

- If proliferation (P) is chosen, check if the cell has reached proliferation age and if there is space for a daughter cell or not. If both are true the cell divides and the daughter cell get placed in a neighboring grid point, if not the cell does nothing
- If quiescence (Q) is chosen the cell becomes quiescent
- If apoptosis (A) is chosen the cell dies and
- If movement (M) is chosen the cell moves to the empty neighbors. Here we present two kinds of cell movement; each cell can move if FNN output is movement and movement is also depend on the ECM degradation and cell-cell adhesion (describes in section 3.5).

3.4 Maspin as an ECM constraint

In reality, cell-cell adhesion and cell-ECM interactions are responsible for the inside-out cellular signaling involving with different biological factors. The model takes into account tumor growth, ECM remodeling and mechanical interaction with host tissue. We consider ECM as a whole without distinguishing its components (collagen, laminin, fibronectin etc.) to reduce the complexity of the modeling method, though we are aware that they contribute differently to the mechanical and adhesive properties of the matrix and have different production mechanism. Normally, tumor cells make contact with the ECM proteins, release proteases including MMPs which degrade the ECM. In this model, ECM has been taken as a growth constraint for each cell. We formulated a set of equations to define cell-ECM interaction by relating maspin as a controlling factor. In the following equations, $E(x,t)$ represents the ECM density, $M_D(x,t)$ is production of degrading enzyme, $M(x,t)$ is maspin value at particular grid point and $A(x,t)$ is the density of degraded ECM.

$$\frac{\partial E(x,t)}{\partial t} = -\beta M(x,t)N(x,t) + A(x,t) - M(x,t)\beta_m + \beta_f E(x,t) \tag{5}$$

$$\frac{\partial M_D(x,t)}{\partial t} = D_M \Delta M_D(x,t) + \gamma_m N(x,t) - \alpha_m M_D(x,t) \tag{6}$$

$$\frac{\partial A(x,t)}{\partial t} = \gamma M_D(x,t)E(x,t) + \chi_a \Delta A(x,t) - \mu_a A(x,t) \tag{7}$$

$$N(x,t) = \sum_{i=1}^{i=n} I_{B_\varepsilon(x)}(x_i) \tag{8}$$

Where

$$I_{B_\varepsilon(x)} = \begin{cases} 1, & \text{if} x_i \in B_\varepsilon(x) \\ 0, & \text{Otherwise} \end{cases}$$

ECM digestion rate β has got constant value equal to 1. $N(x,t)$ represents the number of neighbors around the cell x at specific time t. $B_\varepsilon(x)$ is a ball of radius ε (approximately equal to the radius of a cancer cell) that is centered at x. N is the total number of tumor cells in the tumor tissue and x_i is the position of the i^{th} cell. A constant was introduced which is the production rate of attractants, χ_a is the diffusion coefficient of the digested ECM and μ_a is the decay rate of the digested ECM. The ECM densities and degradation rates have not been measured in biological experiments. During the tumor growth, ECM density has assumed to be 1 with no degraded ECM, $A(x,0) = 0$. Every proliferating cell had to degrade 7% of the ECM before the placement of a daughter cell in the available neighboring points.

3.5 Cell-cell adhesion and cell movement

For modeling cell-cell adhesion, model took cell-cell adhesive forces function which defines the cell movement called cadherin induced motion biologically. Cells attraction is mimicked as a pulling force or the potential function, $P(x,t)$ during the solid tumor growth. At every time step model calculates the pooling force for each and every cell. The potential function and cell movement direction were calculated using equations (9) and (10) respectively (Conde et al., 2008).

$$P(x_i,t) = I_{B_\varepsilon(x_i)}\left(\frac{1}{d(x_i,x_j)} + e^{\alpha} - he^{-(d(x_i,x_j)-\varepsilon/2)^2}\right) \quad (9)$$

$$D = -\nabla P(x_i,t) \quad (10)$$

Where, $d(x_i,x_j)$ is the distance between two cells, where $B_\varepsilon(x_i)$ is a neighbourhood centred on x_i with a radius ε, e^{α} is the maximum energy and h is the cell's capacity to bond. When the movement node value was greater than 0.5, the cell was allowed to move. After acquiring the required network response the model calculated the attraction forces of the other cells for the selected cell using equation 9. Each cell moved to the specified direction by following equation 10. The model compared each cell's P value to the surrounded cells P values and moved towards the highest value cells. During the movement the cell was

also allowed to proliferate only if it met the proliferation criteria. With the ECM involvement the cell took longer to divide because it could place its daughter cell in the available neighboring grid point only if the surrounding ECM had been degraded sufficiently.

3.6 Model parameters

CA is getting a famous modeling paradigm in different fields of research. Currently, its diversity spreads around the different multidisciplinary areas (Xiao, et al., 2008;2011). CA method has also been used successfully for different aspects of tumor growth modeling (explained in section 2). For this model, a 2D lattice grid has been taken into $N \times N$. All presented PDEs were discritised using standard fivepoint finite central difference formula and used length scale $\Delta d = 0.0025$ (the rescaling of the length gives each square cell an area of $6.25 \times 10^{-6} cm^2$ (Casciari et al., 1988) and time scale $\Delta t = 5 \times 10^{-4}$. Each time step the chemical concentrations are solved using the discretised equations and every one of the tumor cells is updated in a random order. The necrosis has been activated when oxygen concentration goes below certain threshold level C_{ap}. C_{ap} is suggested 15% of the initial oxygen concentration (Gerlee and Anderson, 2007; Al-Mamun et al., 2013b). Cells should not divide if there is no sufficient space for the daughter cell; in this model the value of neighbors $n(\vec{x}, t) > 3$. Previous models have shown different tumor dynamics for different oxygen concentration values (Gerlee and Anderson, 2007; Kazmi et al., 2012; Al-Mamun et al., 2013b), but for the relatively small values we consider this effect is negligible. The parameters used for equation (1-4) are listed in table 1. Table 2 gives a list of parameters used in equations (5-10) to design the cell-cell adhesion and cell-ECM interaction due to maspin. The weights and thresholds of initial neural network setup is given in appendix A.2.

3.7 In vitro Methods: Cell line, Plasmids, Transfection, Time Lapse Video Microscopy

This paper includes an in vitro experiment to determine the adhesion of MCF7 breast cancer cell line which was stably transfected with wild type maspin. We formulated the in vitro hypothesis that maspin

increases cell-cell adhesion to various ECM components like collagen I, laminin and fibronectin. MCF7 cells were obtained from ATCC and grown in MEM supplemented with 10% (v/v) foetal calf serum (FCS), 1% (v/v) nonessential amino acids and 1% (v/v) sodium pyruvate. All cell culture reagents were from Life Technologies (Paisley, UK). Cell lines stably expressing different forms of maspin have been described previously (Ravenhill et al., 2010).

Table 1: A list of parameters and values used during the tumor growth simulation

Parameters	**Definition**	**Value**	**References**
C_0	Oxygen background conc.	1.7×10^{8}mol cm^2	Anderson, (2005)
D_{O2}	Oxygen diffusion constant	$1.8\times10^{5}cm^2s^1$	Grote et al., (1977)
D_M	Maspin diffusion constant	$1\times10^{6}cm^2s^1$	Young et al., (1980)
n_o	Cancer cell density	1.6×10^{5}cellscm^2	Casciari et al., (1988)
cr_{O2}	Base oxygen consumption rate	2.3×10^{16}mol cells^1 s^1	Klieser et al., (1986)
k	Modulation strength	6	Model specific
T_r	Target response	0.675	Model specific

Table 2: A list of parameter used cell-cell adhesion and cell-ECM interaction equations

Parameters	**Definition**	**Value**	**References**
β	Digestion rate of ECM	1	Model specific
β_m	Maspin inhibition constant	400 mol^{1}cm^2	Bass et al., (2002)
B_f	ECM remodeling parameter for ECM	(0,0.015)	Conde et al., (2008)
D_m	Diffusion constant of degrading enzyme	0.08	Toma et al., (2012)
α_m	Decay coefficient	01	Toma et al., (2012)
γ_m	Production constant for MDEs	1	Toma et al., (2012)
γ	Degradation rate of ECM by MDEs (plasmin)	8.15	Andasari et al., (2011)
χ_a	Diffusion constant of digested ECM	0.01	Conde et al., (2008)
μ_a	Decay rate of the digested ECM	0.01	Conde et al., (2008)

4. Data clustering

The model was fed two datasets consisting with in vitro and in silico cell migration data. In vitro data set was composed of control (normal tumor growth without maspin) and maspin (growth with maspin) data set. In silico data set consists of normal tumor growth and tumor growth with maspin data. All the datasets were obtained from independent set of experiments from both in vitro and in silico models. The cell migration data was measured in both in vitro and in silico experiments via velocity. In in vitro model, the cell velocity was measured using cell tracking toolkit integrated in Axiovision 4.7.1 software and Zeiss Examiner 4.0. In in silico model, the distance of each cell was tracked by saving the co-ordinate after every time point. Then in the same way like in vitro, the average velocity was measured for each cell. Normally, in biological experiments, the data from two different sources are analyzed via student t-test or ANOVA test. But most of the time, the data patterns of these independent data samples get ignored, as a results not that much assumptions can be made through this hypothesis tests. In our previous attempt we presented resemblances between in vitro and in silico cell migration data using different statistical measurements (Al-Mamun et al., 2013b). But here we employed several clustering algorithm to reveal the hidden patterns of the datasets in order to see the similarity of our in silico model with in vitro model (Ravenhill et al., 2010). Clustering or data segmentation of unsupervised data is the process of grouping the data instances into clusters, so that instances within a cluster have high similarity in comparison to one another but are very dissimilar to instances in other clusters (Chen and Miao, 2011). Similarity of instances is based on the attribute values describing in the instances (Wang and Chan, 2013).

Most of the existing clustering algorithms work well on small datasets containing fewer than several hundred data instances with few attributes; however, a large data set may contain millions of data instances with a large number of attributes (Lee and Olafsson, 2011). Generally, clustering methods structures the set of instances into groups based on their similarity, and then assign labels to the relatively small number of groups. It has been widely applied in numerous real world applications, including biology, medicine, anthropology, marketing, pattern recognition, and image processing (Wang and Chan,

2013). For analyzing these unsupervised datasets, initially we calculated similarity and distance measurement and then implemented nearest neighbor clustering and k-Means clustering. As these methods are already well established, we described the methods in brief with algorithms in appendix section B. After using the common clustering approached we proposed a density based clustering approach to solve the holes of the existing clustering algorithms for defining our experimental datasets. The main objective was to see that how much similar was our in silico model to the in vitro model in case of cell migration data.

4.2 Proposed density based clustering

After exploring the existing k-Means and NN clustering techniques, we formulated a density based clustering technique to identify similarity based on the number of links between instances. A pair of instances are said to be neighbors if their similarity exceeds some threshold. Our proposed clustering algorithm first finds the nearest neighbors of each instance, and then redefines the similarity between pairs of instances in terms of how many nearest neighbors the two instances share. This need not be defined based on a precise metric, but rather a more intuitive approach using domain experts knowledge. The number of links between two instances is defined as the number of common neighbors they have. The objective of our clustering algorithm is to group instances that have shared links. We used Jaccard coefficient instead of using euclidean distance to measure similarity that shown in equation 11.

$$sim(x_i, x_j) = \left|x_i \cap x_j\right| / |x_i \cup x_j| \tag{11}$$

A goodness measure was used to merge clusters which pair of instances is merged at each step that shown in equation 12.

$$g(K_i, K_j) = \frac{link(K_i, K_j)}{(x_i + x_j)^{1+2f(\Theta)} - x_i^{1+2f(\Theta)} - x_j^{1+2f(\Theta)}} \tag{12}$$

In equation 12, $link(K_i, K_j)$ is the number of links between the two clusters. Also, x_i and x_j are the number of instances in each cluster. $x_i^{1+2f(\Theta)}$ is an estimate for the number of links between pairs of

points in K_i when the threshold used for the similarity measure is Θ. The function $f(\Theta)$ depends on the data that each $x_i \in K_i$ has approximately $x_i^{f(\Theta)}$ neighbors in the cluster.

5. Computational Experimental Setup

The in silico model was coded using custom way in MATLAB environment. The clustering experiments were conducted using a MacBook Pro with Retina display with 2.7 GHz quadcore Intel Core i7 Processor and 16 GB of RAM. We implement the clustering algorithm in Java. We use NetBeans IDE 7.3.1 for Java coding. NetBeans IDE is the first IDE providing support for JDK 7 and Java EE 6 (http://netbeans.org/index.html). The code for the basic versions of NN and k-Means clustering are adopted from Weka version 3.0, which is open source data mining software (Hall et al., 2009). It is a collection of machine learning algorithms for data mining tasks. Weka 3.0 contains tools for data preprocessing, classification, regression, clustering, association rules, and visualization. The clustering algorithms in Weka 3.0 can be either applied directly to a dataset or called from our own coding.

5.1 Adhesion Assays

Adhesion assays were performed as described previously (Ravenhill et al., 2010). With the exception that here individual matrix components at 5μg/ml were used to coat the wells of tissue culture treated 96 well plates overnight at 4∘ C, prior to assay.

6. Results and discussion

6.1. Tumor growth dynamics due to maspin

Tumor growth started from four cells in our simulation and grew spherically in layered structure consisting of dead region in the center and quiescent and proliferating cells surrounding the necrotic region respectively. At every time step each cell completed its own lifecycle actions depending on the NN response. Figure 3 shows compact spherical morphology for normal tumor growth and tumor growth with maspin respectively. This round morphology is matched with a basic study of multicellular tumor

spheroids of EMT6/Ro model proposed by (Freyer and Sutherland, 1986). The growth dynamics showed three compartments at 20^{th} day of growth: necrotic region is surrounded by quiescent and proliferating cells sequentially. For both case, the tumor size was almost same because during first few days the tissue was small in size and cells got sufficient amount of nutrients to survive. On 60^{th} and 100^{th} day, the apoptosis region appeared throughout the center area because decreased diffusion in the central region. Moreover, diameter of the maspin growth was smaller than the diameter of the normal growth on 60^{th} and 100^{th} days and also the compactness and roundness on 60^{th} and 100^{th} days shows the effects of maspin in figure 3. It has to be noted that previous papers showed the experimental evidences that tumor shows round morphology in the presence of ECM in different circumstances (Härmä *et al*., 2010; Chan *et al*., 2013).

From the early age of maspin investigations, it has been suggested that maspin reduces cell migration while it enhances the cell adhesion to ECM components especially fibronectin (Seftor *et al.,* 1998; Cella *et al.,* 2006). The layered structures in figure 3b suggest the reduced cell migration and increased cell adhesion in an abstract format when maspin was present in the model. This indicates that maspin possibly enhanced the cell-ECM adhesion to show this compact layered structure. It corroborates the notion presented in previous literature (Al-Mamun *et al.,* 2013b) about reduced proliferation, migration and invasion mechanism of maspin.

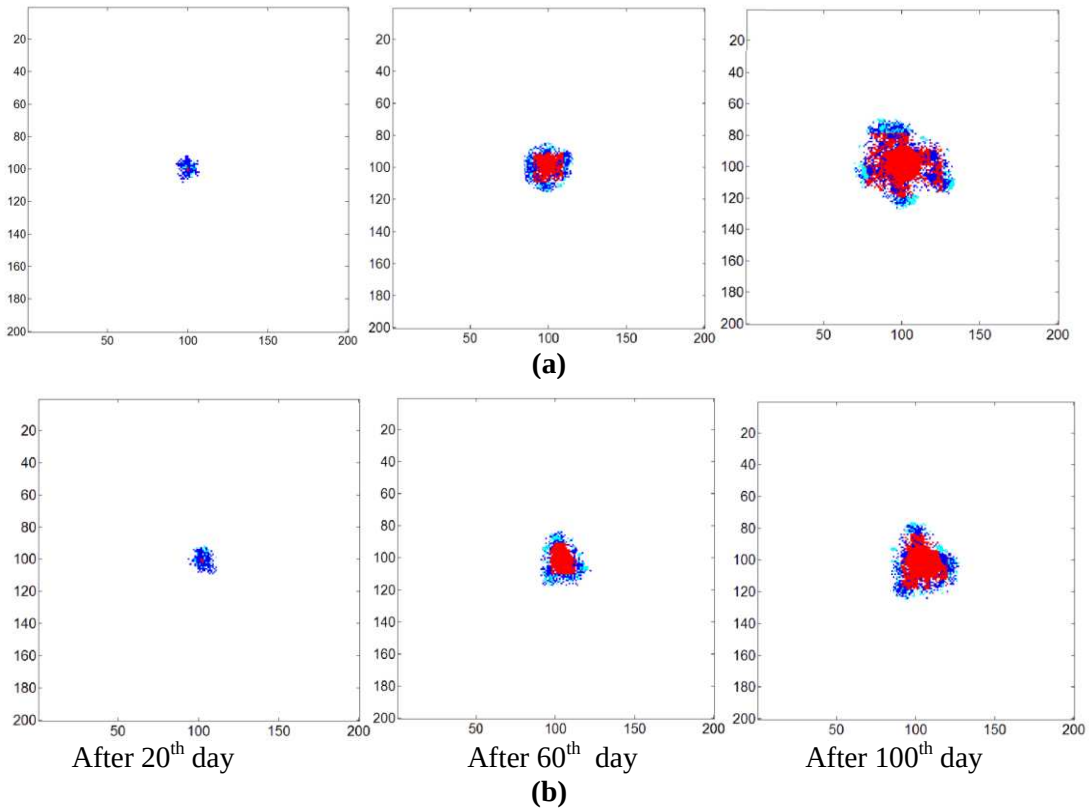


Fig. 3.a. Normal tumor growth morphology and b. Solid tumor growth morphology due to maspin, where red color represents dead regions, blue color represents quiescent cells, cyan color represents proliferating cells, and white color represents empty grid points during the growth evolution

6.2. Maspin as an ECM Constraint

For simulating the overall maspin impact, we calibrated the model to see whether our ECM equations (5-8) present the real scenario or not. The microenvironment parameter of ECM density (E) evidently influenced tumor growth rate (or proliferation of tumor cell). The abnormal cells grow at a slower rate because they are not allowed to accommodate daughter cells unless their surrounding ECM is sufficiently degraded (shown in figure 4). We simulated the model where first, tumor grew without ECM and maspin and then compared the growth with the scenario while tumor grew with both ECM and maspin. The inclusion of the ECM and maspin constraint in the model causes a very low growth rate. This time the tumor takes more time to reach the same size as observed when no ECM was included in the model. It is also important to mention that this kind of behavior has been seen from previous studies (Conde et al.,

2008; Kazmi et al., 2012b) which confirm the ECM inclusion in our extended model signifies that ECM itself has an impact on the tumor growth. The tumor growth rate without integrating ECM also quite similar to the rate mentioned in Al-Mamun et al., (2013b).

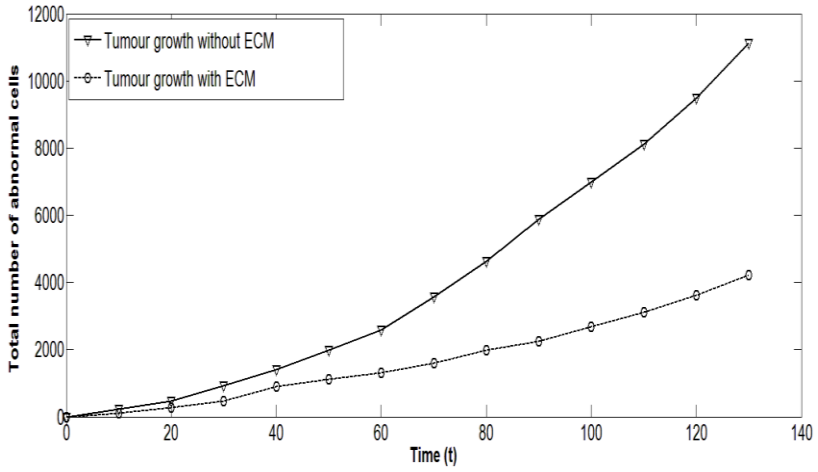


Fig. 4. The tumor growth rate comparison in terms of abnormal cells before ECM inclusion and after ECM inclusion

Normally, a tumor cell needs to invade the surrounding ECM in order to go to the metastasis stage. The model measured the ECM density for different maspin concentration during the tumor growth (figure 5). Initially, ECM density E(x,t) is taken zero and afterwards cells created the ECM layer surrounding them which restricted cells to move the neighboring cells. After every iteration ECM density has calculated from equation 5 which is related to MDEs/uPA production and amount of degrading parameters. ECM density is highest when maspin concentration has given 1.5uM.

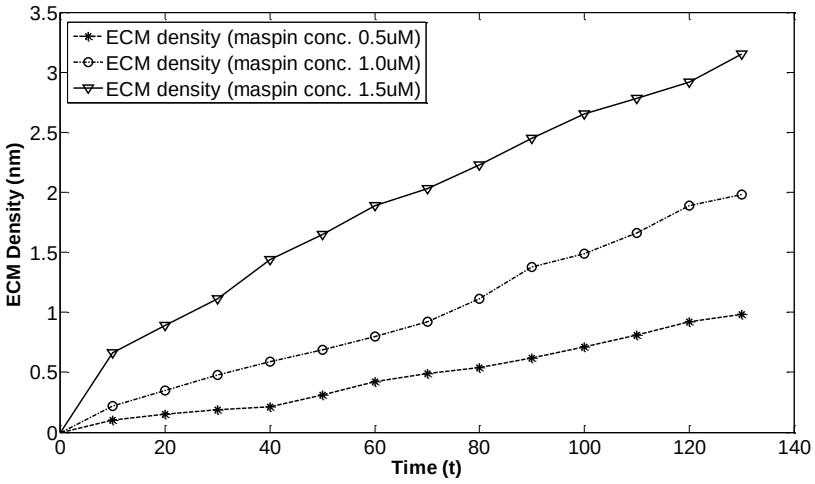


Fig. 5. ECM density measurements during the tumor growth for different maspin concentration 0.5uM,1.0uM and 1.5uM.

The model also measured the degraded ECM density during the tumor growth (figure 6). For showing the maspin impacts on ECM density, every proliferating cell degrades 7-10% of the ECM before the placement of a daughter cell in the available neighboring points. The model calculates amount of degrading ECM density from equation 7 which defines a complex biological where MDEs production, amount of created ECM density and maspin involved themselves. Degraded ECM density measurements are shown in figure 6 that when maspin has given in low concentration in the model, the ECM degradation increased whereas the high maspin concentration showed minimal ECM degradation.

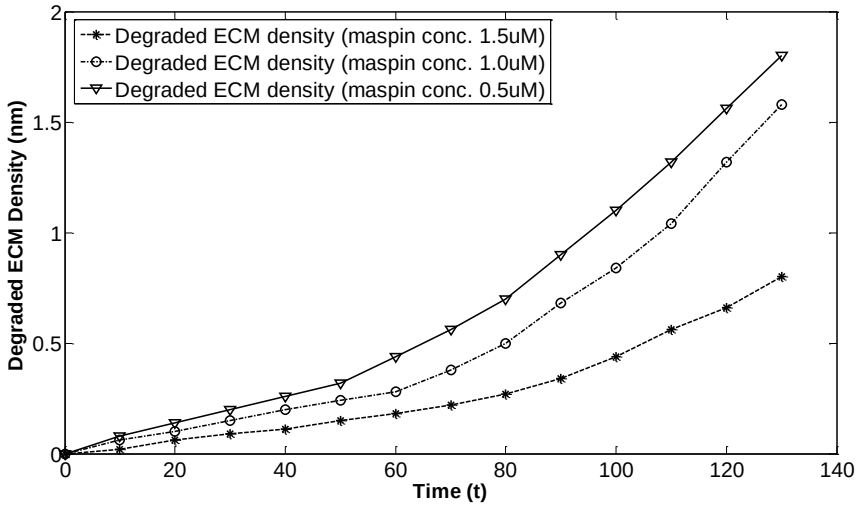


Fig. 6. Degraded ECM density measurements during the tumor growth for different maspin concentration 0.5uM,1.0uM and 1.5uM.

Collectively, figure 5 and 6 suggest that maspin certainly has a role in regulating ECM concentration and also present in the supra-molecular structure of the adhesion plaque. This mechanism involves direct engagement of extracellular matrix ligands to mediate cell adhesion, which can be modulated by "inside-out" signals to alter the affinity for these ligands, and also to transduce "outside-in" signals to the cytoplasm (Bass et al., 2009). Density of ECM at extracellular surface and amount of degraded density refines the hypothesis made earlier in (Seftor et al., 1998; Cella et al., 2006) that maspin enhances cell adhesion. Though in both in vitro and in vivo studies, the ECM turnover has not been measured, the results presented here predict the possible impact of maspin over ECM turnover.

6.3. Maspin effects on cell matrix components

To support the maspin effects on cell-ECM enhancement, a novel *in vitro* experiment has been adopted to see the percentage of adhesion to the individual matrix component in the presence of maspin. Here we examined the effect on cell adhesion when cells were transfected with wild type pcDNA3.2maspin from the cell adhesion assay of MCF7 cell lines. Using an *in vitro* assay to measure cell adhesion, it was found that wild type maspin enhanced MCF7 cell adhesion to various ECM components (Figure 7). Specifically,

MCF7 cells stably expressing wild type maspin significantly increased cell adhesion by 113±5% on a laminin matrix and by 45±6% on either collagen I or fibronectin matrices, in comparison to cells expressing vector only. Figure 7 shows MCF7 cells stably transfected with pcDNA3.2 (open bars) or pcDNA3.2maspin (hatched bars), plated onto 5µg/ml matrix components. Adhered cells were stained with methylene Blue and absorbance was measured at 630 nm. 100% adhesion is defined by pcDNA3.2 cells; an average OD630 of 0.7, 0.25, 0.7 for collagen I, laminin or fibronectin respectively. Three independent experiments were performed in triplicate. Statistical significance was compared to cells transfected with pcDNA3.2 and measured by Students t-test ($p^*<0.05$). These results confirm the undertaken hypothesis in our computational model that maspin has direct influence in enhancing the cell-ECM adhesion. The real dataset for this experiment has given supplementary file 1 .

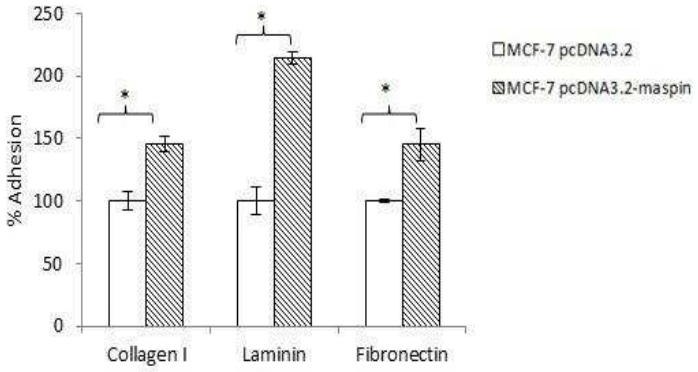


Fig. 7. Adhesion of MCF7 stable cells on matrix components.

6.4. Experimental analysis

The main motivation of implementing the clustering method was to establish the fact that maspin reduces cell migration. Another purpose was to see whether the produced cell migration data from our in silico model was really resemblance with the supplied in vitro data (Ravenhill et al., 2010) or not. For this purpose, we proposed several clustering algorithms in section 4. Table 3 presents the basic mean and standard deviation for both in vitro and in silico datasets. By using student t-test it has been shown in our

previous attempt that in vitro result showed that maspin reduces the cell migration up to 70%, whereas in silico model reduced cell migration up to 73%. As the results produced from the raw data set for both models, it is clear that the scaling was not same for both models. Scaling is one of well discussed issue in previous models (Gerlee and Anderson, 2007; Kam et al., 2012; Anderson et al., 2013).

Table 3: Basic statistical parameters of in vitro and in silico data

Dataset	**Minimum**	**Maximum**	**Mean**	**StdDev**
In vitro control	0.2	8.176	2.256	1.298
In vitro maspin	0.104	5.899	1.576	1.071
In silico control	0.234	3.087	1.996	1.019
In silico maspin	0.126	2.774	1.457	0.707

The experimental analysis in vitro and in silico data using clustering algorithms are shown in tables 4, 5, 6, and 7. k-Means clustering centroid result has been measured and it also supports our statistical hypothesis (shown in table 3). Three clusters also give the centroid variations between two datasets (shown in table 4). It is clearly seen that centroid of first cluster in case of in vitro maspin and in silico maspin reduce cell migration significantly in compared with second cluster. We have seen, the shape of the clusters were circular in shape with overlapped data points for both in vitro and in silico datasets (shown in fig 8). This is one of the major problems of k-Means clustering, as it needs apriori specification of the number of clusters and does not work well with non-globular clusters. Another problem is that the intra and inter cluster distance vary in each iteration. One example is shown in figure 7 where the distribution of data points in each cluster is shown in case of in silico control and maspin datasets.

Table 4: Cluster centroids of in vitro and in silico data using k-Means clustering

Dataset	**Mean**	**Cluster 1**	**Cluster2**	**Cluster 3**
In vitro control	2.2563	5.0825	2.6434	1.271
In vitro maspin	1.5764	0.7858	3.5007	1.7996
In silico control	1.9957	3.242	0.8545	2.089
In silico maspin	1.4567	0.6798	2.3082	1.5279

Table 5 tabulates the percentage of instances in each cluster. It shows that, in vitro clusters have unequal distribution of instances because of their frequent nature values and it is 11%, 42% and 47% respectively, whereas in silico clusters have almost equal distribution of instances, that is 30%, 35% and 35% respectively. In case of maspin data for both models also show the same assumptions which again specify that control data for both models are scattered in nature and spread more than maspin data.

Table 5: In vitro and in silico data points in each cluster using k-Means clustering

Dataset	**Full data**	**Cluster 1**	**Cluster2**	**Cluster 3**
In vitro control	500	54	209	237
In vitro maspin	500	246	81	173
In silico control	500	148	176	176
In silico maspin	500	180	150	170

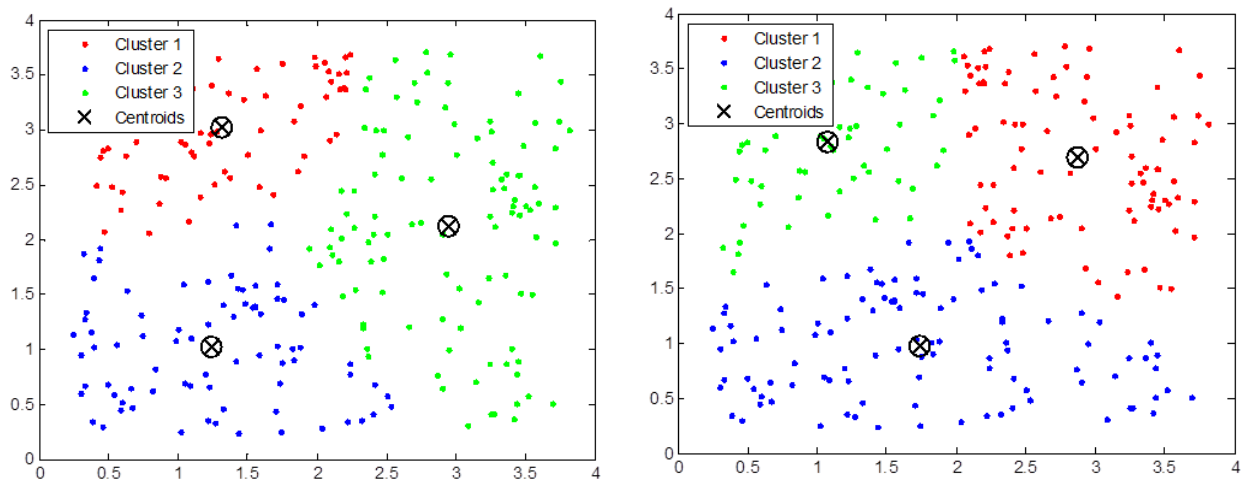
Table 6 and table 7 represent the centroid and data points distribution obtained from density based clustering algorithm respectively. Our algorithm produced two clusters for both datasets and show significant resemblances between both models. Proposed clustering algorithm finds the nearest neighbors for each instance and redefines the similarity between pairs of instances. The number of links showed the successive common neighbors they have. We can get the similarity measurement from equation 11 and we calculated that in silico control data points are almost 65% similar to in vitro data points and in silico maspin data points are almost 62% similar with in vitro data points. These specify that our computational model is mimicking the maspin effects with 60-65% similarity. The density based clustering algorithm normally does not need any apriori specification of number of clusters and has the ability to identify then noise. The main advantage of using density based clustering is that it can find arbitrarily sized and arbitrarily shaped clusters.

Table 6: Cluster centroids of in vitro and in silico data using density based clustering

Dataset	Mean	Cluster 1	Cluster2
In vitro control	2.2563	4.3766	1.759
In vitro maspin	1.5764	1.066	3.029
In silico control	1.9957	2.8334	1.0736
In silico maspin	1.4567	0.6149	2.0636

Table 7 :In vitro and in silico data points in each cluster using density based clustering

Dataset	Full data	Cluster 1	Cluster2
In vitro control	500	95	405
In vitro maspin	500	370	130
In silico control	500	262	238
In silico maspin	500	249	251



(a) (b)

Fig. 8. Distribution of data points in clusters using k-Means clustering approach for a) *in vitro* control and b) *in silico* control datasets.

Previous computational models for maspin growth dynamics did not show any in vitro data verification (Al-Mamun et al., 2013b), but this paper provides a clustering approach to see the similarity among the unsupervised data collected from both models for the first instance. Our proposed density based clustering

algorithm which provides the similarity measurement of both in vitro and in silico datasets. The proposed algorithm shows up to 60-65% similarity of the presented in silico model with in vitro data. It offers better agreement of both models that the other clustering algorithms like k-Means and nearest neighbor. It is worth to mention that clustering algorithm presented here has shown good agreement with our datasets obtained from Ravenhill et al., (2010); it can also work for the other databases of raw cell migration datasets.

6.5. Discussion

Tumor growth results (in section 6.1) suggest that maspin reduces cell movement and also increases cell-cell adhesion. Figure 3 indicate compact circular morphology of avascular tumor when maspin is present in the growth model, where red color represents dead regions, blue color represents quiescent cells, cyan color represents proliferating cells, and white color represents empty grid points. It matches with the previous findings of (Kazmi *et al.,* 2012a) where multicellular spheroids experiments of HT29 showed the solid tumor morphology. Moreover, the growth pattern depended on the ECM degradation for the presence of maspin. It has been suggested that maspin can use two proposed pathway for increasing cell-ECM adhesion: regulating the plasminogen activation system (PAS) and β_1 integrin signaling pathway (Endsley *et al.,* 2011). PAS is strictly maintained by uPA which is an extracellular serine protease. uPA binds to uPAR and activates plasminogen which basically localized ECM degradation. ECM degradation is a crucial component in cancer invasion and metastasis. It has been presented that maspin has an active role to reduce invasion and migration by incorporation of cell surface maspin-uPA-uPAR complexes (Endsley *et al.,* 2011). Maspin has been reported to bind to β_1 integrin cell adhesion receptors and maspin has a certain effects to β_1 integrin activation status. Al-Mamun *et al.,* (2013b) developed a computational model which showed maspin effects computationally and simulation results were verified by *in vitro* raw data supplied by (Ravenhill *et al.,* 2010).

Results presented in section 6.2 also suggest that maspin increases the cell-ECM layer which restrict cell movement and increase cell-cell adhesion. Also the *in vitro* results presented in section 6.3 showed an

evidence of increasing cell-cell adhesion for the individual components of ECM. It is also seen elswhere that collagen type I and II showed binding percentage to maspin 70% and 25% respectively (Ngamkitidechakul *et al.,* 2001).Moreover, our current model also added the data verification method using clustering approach. The results showed in section 6.4 provide a density based clustering method which showed significant similarity between *in vitro* and *in silico* data pattern. These corroborate that maspin has a potential role in reducing cell migration and this might be happened via enhancement of cell-ECM adhesion. It is worth to mention that data clustering has been used in this study to see the similarity and dissimilarity between the datasets. The extended model is differs from our previous attempt is several ways, firstly, this model includes ECM in presence of maspin to see whether it regulates the tumor growth or not. The results suggest that tumor growth decreased in terms of round morphology in presence of ECM with maspin at extracellular environment. It is required to mention that our previous model gave the tumor suppressive features of maspin like proliferation, apoptosis, cell migration and invasion whether we did not consider microenvironment constraints ECM in the model. Secondly, this model concluded an important behavior of maspin that it might have impact on ECM turnover, though currently no study showed any potential engagement of ECM turnover with maspin. This can help us to build a new *in vitro* method to investigate this assumption. Thirdly, the computation result of the extended part of the model was supported by a novel *in vitro* experiment. Fourthly, the data verification method gives our model more confidence that our model resemblances the *in vitro* biologics of maspin.

Overall, the paper contributes to the subtle maspin behavior in different ways; firstly, the results presented in figure 6.1 to 6.3 suggest that extracellular maspin has potential engagement with cell-ECM adhesion enhancement which again matches our previous lab results (Bass *et al.,* 2009 and Ravenhill *et al.,* 2010). It also indicates to develop new *in vitro* and *in vivo* studies for ECM turnover alteration due to maspin engagement. Secondly, *in vitro* results presented in figure 6 provides a direct resemblances with *in silico* results that maspin enhances cell-ECM components engagement by which it reduces cell migration (proposed in Endsley *et al.,* 2011). It also provide clear indication that maspin resides exogenously in cellular environment to display its potential influences tumor metastasis suppression. Thirdly, the results

presented in section 6.4 clearly showed a new data verification method using data clustering method. The results are plausible to show up to 60-65% similarities between both *in vitro* (Ravenhill *et al.,* 2010) and *in silico* models. This leads this modeling framework one step closer to the reality where the real data from the wet lab experiments can be compared with *in silico* data. It is worth to mention that there are some limitations which need to be considered in case of data verification methods. The selection of clustering algorithms totally depends of the nature of the produced data, the density based clustering method is not a solely option for other problems.

7. Conclusion

This paper presents an *in silico* model composed of an *in silico* tumor growth model and a data verification method using clustering. First, the extended *in silico* model upgraded our previous model for maspin dynamics (Al-Mamun *et al.,* 2013b) and added cell-ECM interactions, cell-cell adhesion and cell movement constraints in presence of maspin. Our results suggest that maspin has influence on microenvironment constraint including cell-ECM, cell-cell adhesion and cell movements which show good agreement with the previous *in vitro* model hypotheses (Ngamkitidechakul *et al.,* 2003; Bass *et al.,* 2009; Ravenhill *et al.,* 2010: Endsley *et al.,* 2011). The model predicts that maspin can play an important role when ECM is present in the model. It can slow down the growth dynamics. More significantly, we presented a novel in *vitro* adhesion assay on matrix components (fibronectin, laminin and collagen I) of MCF7 cell line transfected with wild type pcDNA3.2maspin. Collectively, the results suggested the extracellular presence of maspin at the cell surface along with ECM components and MDEs.

Second, the later part of the model showed similarity between *in vitro* and *in silico* cell migration datasets using clustering method. The model presented a density based clustering algorithm to find the similarity between two models. Results showed that our *in silico* model's data matched about 60-65% with the *in vitro* data (Ravenhill *et al.,* 2010). From our best knowledge, this is the first attempt of adopting clustering approach to show the resemblance of an *in silico* model with real *in vitro* experimental data. The implication of this data clustering attempt increases the fidelity and superiority of the computational

model over the biological experiments.

It is important to mention that although this hybrid model does not cover the whole range of maspin effects, but it informs the maspin mechanism on basic biological constraints like cell migration, cell-ECM interaction, and cell-cell adhesion. This model provides a justified subcellular investigation of maspin by supporting novel *in vitro* experiments. Also, introduction of clustering approach to data analysis can bring our effort to one step closer to the real experiments, as it is still far from being a precise model of the subcellular dynamics. The impact of this data verification would be to confirm more real data to see the similarity using this model as a tool in case of *in vivo* data. It is worth to mention that treatment with maspin has been investigated from the very beginning of its discovery. Re-expression of Maspin or treatment with rMaspin decreases tumor growth and metastasis using *in vivo* experiments (Zou *et al.,* 1994; Bodenstine *et al.,* 2012; Berardi *et al.,* 2013). Recently, a paper by Bodenstine *et al.,* (2014) considered recombinant Maspin (rMaspin), which alters invasive properties when directly applied to cancer cells. This study indicates that cellular processing of rMaspin plays a key role by affecting its biologic activity and highlights the need for new approaches aimed at increasing the availability of rMaspin when used to treat cancer. These give us a new clue that our presented model can be usable from different aspects of experiments and treatments while maspin will be taken as a constraint.

Appendix A

A.1. Feedforwad Neural Network Construction and Mechansim

The FNN consisted of a number of nodes which can take real numbe values. Thease nodes are organised intro three layers: Input layer (I), Hidden layer (H) and Output layer (O). The input parameters are transferred to the hidden layer and this layer does its processing using standard transfer function $T(x)$, weight matrix X and threshold vector ϕ. X_{ij} determines the conncetion strengh between node j in the the input layer (I) and i in the hidden layer (H). The hidden layer then passes the calculated values to the output layer that performs its calculation using $T(x)$ as a transfer function, Y matrix and threshold vector θ. Y_{ij} defines the connection strength between node j in the hidden layer (H) and node i in the output layer (O). Input nodes are summed together in equation 6 as Ij and output nodes are summed together as θ_i.

$$T(x) = \frac{1}{1 + e^{-2x}} \qquad \text{A.1}$$

$T(x)$ is the standard transfer fuction considered, as it gaurantees that the reuslting node values of hidden lyaer are the range of [0, 1].

$$H_j = I_j = T\left(\sum_k X_{jk} I_k - \theta_j\right) \qquad \text{A.2}$$

I_j is the sum of the input nodes weighted with connection matrix X and the threshold θ_j. In same way, the node i of the output layer is given by

$$O_j = T\left(\sum_j Y_{ij} I_j - \phi_j\right)$$

$$= T\left(\sum_j Y_{ij}\, T\left(\sum_k X_{jk} I_k - \theta_j\right) - \phi_j\right) \qquad \text{A.3}$$

A.2. Network Parameterisation

X is the weight matrix between input (I) and hidden layer (H)

$$X = \begin{Bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \\ 0.25 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & -1 & -2 \end{Bmatrix}$$

Y is the weight matrix between hidden layer and Output layer (O)

$$Y = \begin{Bmatrix} 0.25 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 1 & -0.25 & 0 \\ 0 & 0 & 0 & 0.25 & -0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{Bmatrix}$$

ϕ and θ are two threshold vectors of hidden and output layer respectively

$$\varphi = (0.25 \quad 0 \quad 0 \quad 0 \quad 0)$$

$$\theta = (0 \quad 0 \quad 0.55 \quad 0 \quad 0.7 \quad -0.75)$$

Therefore, the behavior of the network is determined by these parameters and during cell division, the model allowed these parameters to be copied to the daughter cells with the mutation. The number of parameters to be mutated is chosen from a Poisson distribution with parameter p. For all those parameters x the mutations are modelled by letting$x \to x + s$, where s is random number from the normal distribution$N(N, \sigma)$, where σ determines the strength of the mutation.

Appendix B

B.1. Similarity and distance measures

Given a data set, $D = \{x_1, x_2, \cdots, x_n\}$, of instances, a similarity measure, $sim(x_i, x_l)$, defined between any two instances, $x_i, x_l \in D$, and an integer value k, the clustering problem is to define a mapping $f: D \rightarrow 1, \cdots, k$ where each x_i is assigned to one cluster K_j, $1 \leq j \leq k$. Given a cluster, $K_j, \forall x_{jl}, x_{jm} \in K_j$, and $x_i \notin K_j, sim(x_{jl}, x_{jm}) > sim(x_{jl}, x_i)$. A good clustering is that instances in the same cluster are "close" or related to each other, whereas instances of different clusters are "far apart" or very different from one another, which together satisfy some requirements: (1) each cluster must contain at least one instance, and (2) each instance must belong to exactly one cluster.

A distance measure,$dis(x_i, x_l)$, as opposed to similarity measure, is often used in clustering. The clustering method then has the desirable property that given a cluster,$K_j, \forall x_{jl}, x_{jm} \in K_j$, and $x_i \notin K_j, dis(x_{jl}, x_{jm}) > dis(x_{jl}, x_i)$. Equations B.1 to B.3 are used to calculate centroid, radius, and diameter of a given cluster, K_m of N data points $\{x_{m1}, x_{m2}, \cdots, x_{mN}\}$.

$$centroid = C_m = \frac{\sum_{i=1}^{N} (x_{mi})}{N} \quad \text{B.1}$$

$$radius = R_m = \sqrt{\frac{\sum_{i=1}^{N} (x_{mi} - C_m)^2}{N}} \quad \text{B.2}$$

$$diameter = D_m = \sqrt{\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (x_{mi} - x_{mj})^2}{(N)(n-1)}} \quad \text{B.3}$$

B.2. Nearest neighbor (NN) clustering

The instances are iteratively merged into the existing clusters that are closest. In NN clustering a threshold, t, is used to determine if instances will be added to existing clusters or if a new cluster is created. The complexity of the NN clustering algorithm depends on the number of instances in the

dataset. For each loop, each instance must be compared to each instance already in a cluster. Thus, the time complexity of NN clustering algorithm is $O(n^2)$. Since, we do need to calculate the distance between instances often, we assume that the space requirement is also $O(n^2)$. Algorithm 1 outlines the NN based clustering method.

Algorithm 1 Nearest Neighbor Clustering

Input: $D = \{ x_1, x_2 \dots \dots, x_n \}$ // A set of instances

A// Adjacency matrix shows distance between instances.

Output: A set of K clusters.

Method:

1: $K_1 = \{x_1\}$
2: $K = \{K_1\}$;
3: $k = 1$;
4: ***for*** $i = 2$ *to* n ***do***
5: $find\ x_m\ i\ some\ clusters\ K_m\ in\ K\ so\ that\ dis(x_i, x_m) is\ the\ smallest$
6: ***if*** $dis(x_i, x_m) \geq t, threshold_{value}$ ***then***
7: $K_m = K_m \cup x_i$
8: ***else***
9: $k = k + 1$;
10: $K_k = \{x_i\}$;
11: **end if**
12: ***end for***

B.3. k-Means clustering

Given a data set, $D = \{x_1, x_2, \cdots, x_n\}$, of n number of instances in euclidean space. Partitioning method distributes the $x_i \in D$ into k clusters, $\{C_1, \cdots, C_k\}$, that is, $C_i \subset D$ and $C_i \cap D = \phi$ for $1 \leq i, j \leq k$. An objective function is used to assess the partitioning quality so that instances within a cluster are similar to one another but dissimilar to instances in other clusters. A high degree of similarity among instances in clusters is obtained, while a high degree of dissimilarity among instances is different clusters is achieved simultaneously. The cluster mean of $K_i = \{x_{i1}, x_{i2}, \cdots, x_{im}\}$ is defined in equation B.4.

$$Mean, m_i = \frac{1}{m}\sum_{j=1}^{m} x_{ij} \quad \text{B.4}$$

The k-Means algorithm requires that some definition of cluster mean exists. Note that the initial values for the means are arbitrarily assigned. These could be assigned randomly. Algorithm 2 outlines the k-Means clustering method.

Algorithm 2 k-Means Clustering

Input: $D = \{ x_1, x_2 \ldots \ldots, x_n \}$ // A set of instances

k // the number of clusters

Output: A set of K clusters.

Method:

1: $arbitrariliy\ choose\ x_{im} \in D\ instances\ as\ the\ initial\ k\ centers$
2: **repeat**
3: (**re**)$assign\ each\ x_i \rightarrow k\ to\ which\ the\ x_i is\ the\ most\ similar, based\ on\ the\ mean\ value\ of\ the\ x_{im} \in k$;
4: **update**, $the\ k\ means, that\ is, calculate\ the\ mean\ value of\ the\ instances\ foreach\ clusters$;
5: **until**, $no\ change$

References

Al-Mamun M.A., Kazmi, N., Hossain, M. A., Vickers, P., and Jiang, Y., 2012. An intelligent decision support system for personalized cancer treatment. In: CIS2012: 11th IEEE Conference on Cybernetic Intelligent Systems, 23-24 August 2012, Limerick, Ireland. doi: 10.1109/CIS.2013.6782159.

Al-Mamun, M.A., Hossain, M.A., Alam, M.S., Bass, R., 2013a. A Cellular automaton model of the effects of maspin on cell migration. Advances in Intelligent Systems and Computing Volume, 222, pp 53-60. doi: 10.1007/978-3-319-00578-2_8.

Al-Mamun, M.A., Brown, L., Hossain, M.A., Fall, C., Wagstaff, L., and Bass, R., 2013b. A hybrid computational model for the effects of maspin on cancer cell dynamics. Journal of Theoretical Biology 337, 150–160. doi: 10.1016/j.jtbi.2013.08.016.

Amir, S., Margaryan, N.V., Odero-Marah, V., Khalkhali-Ellis, Z., and Hendrix, M.J., 2005. Maspin regulates hypoxia-mediated stimulation of uPA/uPAR complex in invasive breast cancer cells. Cancer Biol Ther 4(4), 400-406. doi: 10.4161/cbt.4.4.1617.

Andreasen, P.A., Egelund, R., Petersen, H.H ., 2000. The plasminogen activation system in tumor growth, invasion, and metastasis. Cell Mol Life Sci 57(1):25–40. doi:10.1007/s000180050497.

Andasari, V., Gerisch, A., Lolas, G., South, A., and Chaplain, M.A.J., 2011. Mathematical modeling of cancer cell invasion of tissue: biological insight from mathematical analysis and computational simulation. Journal of Mathematical Biology 63(1), 141–171. doi: 10.1007/s00285-010-0369-1.

Anderson, A.R.A., 2005. A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. Mathematical Medicine and Biology 22 (2), 163–186. doi: 10.1093/imammb/dqi005.

Anderson, A.R.A., Tomlin, C.J., Couch, J., and Gallahan, D., 2013. Mathematics of the integrative cancer biology program. Interface Focus 3(4).

Bailey, C. M., Ellis, Z. K., Kondo, S., Margaryan, N. V., Seftor, R. E. B., Wheaton, W. W., Amir, S., Pins, M. R., Schutte, B. C. and Hendrix, M. J. C., 2005. Mammary serine protease inhibitor (maspin) binds directly to interferon regulatory factor 6: identification of a novel SERPIN partnership. Journal of Biological Chemistry 280(40), 34210–34217. doi: 10.1074/jbc.M503523200.

Bass, R., Fernandez, A.M., Ellis, V., 2002. Maspin inhibits cell migration in the absence of protease inhibitory activity. Journal of Biological Chemistry 277, 46845–46848. doi: 10.1074/jbc.C200532200.

Bass, R., Wagstaff, L., Ravenhill, L., Ellis, V., 2009. Binding of extracellular maspin to beta1 integrins inhibits vascular smooth muscle cell migration. Journal of Biological Chemistry 284, 27712–27720. doi: 10.1074/jbc.M109.038919.

Berardi, R., Morgese, F., Onofri, A., et al., 2013. Role of maspin in cancer. Clinical and Translational Medicine 2(1), 8. doi: 10.1186/2001-1326-2-8.

Bodenstine, T.M., Seftor, R.E., Khalkhali-Ellis, Z., Seftor, E.A., Pemberton, P.A., Hendrix, M.J., 2012. Maspin: molecular mechanisms and therapeutic implications. Cancer Metastasis Review 31 (3–4), 529–551. doi: 10.1007/s10555-012-9361-0.

Bodenstine TM, Seftor RE, Seftor EA, et al., 2014. Internalization by multiple endocytic pathways and lysosomal processing impact maspin-based therapeutics. Mol Cancer Res. 12(10):1480-91. doi: 10.1158/1541-7786.MCR-14-0067.

Biliran, H., Sheng, S., 2001. Pleiotrophic inhibition of pericellular urokinase-type plasminogen activator system by endogenous tumor suppressive maspin. Cancer Research 61, 8676–8682.

Bray, 1990. Intracellular signaling as a parallel distributed process. J. Theor. Biol., 143, pp. 215-231.

Casciari, J.J., Sotirchos, S.V., Sutherland, R.M., 1988. Glucose diffusivity in multicellular tumour spheroids. Cancer Research 48, 3905–3909.

Cella, N, Contreras, A, Latha, K, Rosen, JM, Zhang, M., 2006. Maspin is physically associated with β1 integrin regulating cell adhesion in mammary epithelial cells. FASEB Journal 20, 1510–1512. doi: 10.1096/fj.05-5500fje.

Chan, H.F., Zhang, Y., Ho, Y.P., Chiu, Y.L., Jung, Y. and Leong, K.W., (2013). Rapid formation of multicellular spheroids in double-emulsion droplets with controllable microenvironment- Scientific reports, 3:3462. doi:10.1038/srep03462.

Chen, M., and Miao, D., 2011. Interval set clustering. Expert Systems with Applications 38, 2923–2932. doi:10.1016/j.eswa.2010.06.052.

Conde, I.R., Chaplain, M.A.J., Anderson, A.R.A., 2008. Mathematical modeling of cancer cell invasion of tissue. Mathematical and Computer Modelling 47, 533–545. doi:10.1016/j.mcm.2007.02.034.

Deisboeck, T.S. Wang, Z., Macklin, P., and Cristini, V., 2011. Multiscale cancer modeling. Annual Review of Biomedical Engineering 13(1), 127–155. doi: 10.1146/annurev-bioeng-071910-124729.

Domschke P., Trucu D., Gerisch A. and A J Chaplain M., 2014. Mathematical modelling of cancer invasion: Implications of cell adhesion variability for tumour infiltrative growth patterns. Journal of Theoretical Biology 361:41-60. doi: 10.1016/j.jtbi.2014.07.010.

Edelman, L.B., Eddy, J.A., Price, N.D, 2010. In silico models of cancer. Wiley Interdisciplinary Reviews Systems Biology and Medicine 2 (4), 438–459. doi: 10.1002/wsbm.75.

Endsley, M.P., Yanqiu, Hu., Yong, Deng, et al., 2011. Maspin, the molecular bridge between the plasminogen activator system and β1 Integrin That Facilitates Cell Adhesion. Journal of Biological Chemistry 286 (28), 24599–24607. doi: 10.1074/jbc.M111.235788.

Freyer, J.P., Sutherland, R.M., 1986. Regulation of growth saturation and development of necrosis in emt6/ro multicellular spheroids by the glucose and oxygen supply. Cancer Research 46, 3513–3520.

Gerlee, P., Anderson, A.R.A., 2007. An evolutionary hybrid cellular automaton model of solid tumour growth. Journal of the Theoretical Biology 246 (4), 583–603. doi: 10.1016/j.jtbi.2007.01.027.

Gerlee, P., Anderson, A.R.A., 2008. A hybrid cellular automaton model of clonal evolution in cancer: the emergence of the glycolytic phenotype. Journal of the Theoretical Biology 250 (4), 705–722. doi:10.1016/j.jtbi.2007.10.038

Gerlee, P. and Anderson, A.R.A., 2009. Evolution of cell motility in an individual-based model of tumour growth, Journal of the Theoretical Biology 259(1):67-83. doi: 10.1016/j.jtbi.2009.03.005.

Grote, J., Susskind, S., Vaupel, P., 1997. Oxygen diffusivity in tumour tissue (DS- carcinosarcoma) under temperature conditions within the range of 20–40 1C. Pflugers Architecture 372, 37–42.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H., 2009. The weka data mining software: An update. SIGKDD Explorations 11(1), 2009, http://www.cs.waikato.ac.nz/ml/weka/. Accessed 20.11.2014. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/.

Hanahan, D., and Weinberg, R.A., 2011. The hallmarks of cancer:The next generation. Cell, 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013.

Härmä,V., Virtanen, J., Mäkelä, R., Happonen, A., Mpindi, J-P., Knuuttila, M, et al., 2010. A Comprehensive Panel of Three-Dimensional Models for Studies of Prostate Cancer Growth, Invasion and Drug Responses. PLoS ONE 5(5): e10431. doi:10.1371/journal.pone.0010431.

Johnson, D., McKeever, S., Stamatakos, G., Dionysiou, D., Graf, N., Sakkalis, V., Marias, K., Wang, Z., and Deisboeck, T.S., 2013. Dealing with diversity in computational cancer modeling. Cancer Informatics 12(5), 115–124. doi: 10.4137/CIN.S11583.

Kam, Y., Rejniak, K.A., and Anderson, A.R.A., 2012. Cellular modeling of cancer invasion: Integration of in silico and in vitro approaches. Journal of Cellular Physiology 227(2), 431–438. doi: 10.1002/jcp.22766.

Kazmi, N., Hossain, M.A., Phillips, R.M., Al-Mamun, M.A., Bass, R., 2012a. Avascular tumour growth dynamics and the constraints of protein binding for drug transportation. Journal of Theoritcal Biology 313, 142–152. doi: 10.1016/j.jtbi.2012.07.026.

Kazmi, N., Hossain, M.A., Phillips, R.M., 2012b. A hybrid cellular automaton model of solid tumor growth and bioreductive drug transport. IEEE/ACM Trans Comput Biol Bioinform 9 (6), 1595–1606. doi: 10.1109/TCBB.2012.118.

Klieser, W.M., Freyer, J.P., and Sutherland, R.M., 1986. Influence of glucose and oxygen supply conditions on the oxygenation of multicellular spheroids. British Journal of Cancer 53(3), 345–353.

Lee, J.S., and Olafsson, S., 2011. Data clustering by minimising disconnectivity. Information Sciences 181, 732–746. doi:10.1016/j.ins.2010.10.028.

Li, Z., Shi, H.Y., Zhang, H.Y., 2005. Targeted expression of maspin in tumour vasculatures induces endothelial cell apoptosis. Oncogene 24, 2008–2019. doi:10.1038/sj.onc.

Macklin, P., Lowengrub, J., 2007. Nonlinear simulation of the effect of microenvironment on tumor growth. Journal of Theoretical Biology 245 (4), 677–704. doi:10.1016/j.jtbi.2006.12.004.

Ngamkitidechakul, C., Burke, J.M., OBrien, W.J., and Twining, S.S., 2001. Maspin: Synthesis by human cornea and regulation of in vitro stromal cell adhesion to extracellular matrix. Investigative Ophthalmology and Visual Science, 42(13), 3135–3141.

Ngamkitidechakul, C., Warejcka, D. J., Burke, J. M., O'Brien, W. J. and Twining, S.S., 2003. Sufficiency of the reactive site loop of maspin for induction of cell-matrix adhesion and inhibition of cell invasion. Conversion of ovalbumin to a maspin-like molecule'. J Biol Chem 278(34), 31796–31806. doi: 10.1074/jbc.M302408200.

Ravenhill, L., Wagstaff, L., Edwards, D.R., Ellis, V., Bass, R., 2010. G-helix of Maspin mediates effects on cell migration and adhesion. Journal of Biological Chemistry 285 (47), 36285–36292. doi: 10.1074/jbc.M110.177253.

Seftor, R.E., Seftor, E.A., Sheng, S., Pemberton, P.A., Sager, R. and Hendrix, M.J., 1998. Maspin Suppresses the Invasive Phenotype of Human Breast Carcinoma.Cancer Research 58 (24), 5681–5685.

Sheng S, Carey J, Seftor E.A, Dias L, Hendrix M.J and Sager R., 1996. Maspin acts at the cell membrane to inhibit invasion and motility of mammary and prostatic cancer cells. Proc Natl Acad Sci U S A.; 93(21):11669-74.

Shi, H. Y., Zhang, W., Liang, R., Abraham, S., Kittrell, F. S., Medina, D. and Zhang, M., 2001. Blocking tumour growth, invasion, and metastasis by maspin in a syngeneic breast cancer model. Cancer Research 61(18), 6945–6951.

Teoh, S.S., Vieusseux, J., Prakash, M., et al., 2014. Maspin is not required for embryonic development or tumour suppression. Nature communications 5, 3164.

Toma, A., Mang, A., Schuetz, T A., Becker, S., and Buzug, T.M., 2012. A novel method for simulating the extracellular matrix in models of tumour growth. Computational and Mathematical Methods in Medicine, 109019, 11. http://dx.doi.org/10.1155/2012/109019.

Wang, Y., and Chan, S., 2013. Soft large margin clustering. Information Sciences 232, 116–129. doi:10.1016/j.ins.2012.12.040.

Xiao, X., Wang, P., et al., 2008. Predicting protein structural classes with pseudo amino acid composition: an approach using geometric moments of cellular automaton image. Journal of Theoretical Biology 254: 691-696. doi:10.1016/j.jtbi.2008.06.016.

Xiao, X., Wang, P., et al., 2011. Cellular automata and its applications in protein bioinformatics. Current Protein & Peptide Science 12, 508–519. doi: 10.2174/138920311796957720.

Yin, S., Lockett, J., Meng, Y., et al., 2006. Maspin retards cell detachment via a novel interaction with the urokinase-type plasminogen activator/urokinase-type plasminogen activator receptor system. Cancer Res 66(8), 4173–4181.

Young, M.E., Carroad, P.A. and Bell, R.L. 1980. Estimation of diffusion coefficients of proteins. Biotechnology and Bioengineering 22(5), 947–955.

Zhang, M., Volpert, O., Shi, Y.H., Bouck, N., 2000. Maspin is an angiogenesis inhibitor. Nature Medicine 6, 196–199. doi:10.1038/72303.

Zou, Z., Anisowicz, A., Hendrix, M.J., et al., 1994. Maspin, a serpin with tumour suppressing activity in human mammary epithelial cells. Science 263, 526–529. doi: 10.1126/science.8290962.