Bsc Thesis Applied Mathematics

Mathematical Modelling of Collective Cell Migration in Wound Healing

Thomas van Vroonhoven (4483804)

Delft Technical University

Supervisors

Dr. Q. Peng (Leiden University) Dr. E. G. Rens

Committee members

Dr. J.L.A. Dubbeldam Dr. N. V. Budko

August, 2022 Delft

Contents

Nomenclature

Cell biology

ECM	extracellular matrix
Fbs	fibroblasts
LLC	large latent complex
Mfs	myofibroblasts
TGF-β	transforming growth factor-β

Cellular Potts Model

CPM	Cellular Potts Model
MCS	Monte Carlo step
T	temperature

Useful quantities

DP	directional persistence
MSD	mean squared displacement
QB	qualitative behaviour score

1 Introduction

This project models collective cell migration and differentiation as part of a wound healing process. The hybrid mathematical model used in this project consists of an agent-based model for the cells combined with a system of coupled partial differential equations for two forms of the growth factor molecule TGF-β. The model is implemented in Morpheus, a software package designed for the simulation and integration of cell-based models with PDEs [1]. We determine some conditions in which the additional release of TGF-β by myofibroblasts is conducive to fibroblast migration and differentiation.

1.1 Wound healing

Wound healing consists of a complex, dynamic and well-orchestrated series of events, in which multiple different cells and cellular products cooperate to restore the injured skin [3]. Since it involves a cascade of cellular activities, it is difficult to study wound healing in its full extent experimentally. For this reason, mathematical modelling has been used as a tool to obtain a more insightful understanding; see for example [2].

A short sketch of the multi-staged wound healing process will serve to identify on which stage the current project focuses. The sketch is necessarily a simplification. Wound healing occurs in four overlapping stages: 1) hemostasis, 2) inflammation, 3) proliferation and 4) remodelling (see Figure 1). During hemostasis (bleeding), the blood in the wound coagulates due to its platelets, forming a blood clot that stops further blood effusion and prevents debris and bacteria from entering the wound. During the inflammatory stages, the platelets degrade and form growth factor molecules that attract leukocytes (white blood cells), which clean the wound and prevent bacterial infection. The subsequent proliferation stage is characterized by the activation of various processes: i) The population of fibroblasts is restored by various growth factors which attract fibroblasts to the wound area. ii) Within the wound, fibroblasts proliferate and produce components such as collagen, which help to construct a new extracellular matrix (ECM). The ECM is of crucial importance, since it regulates the growth, movement and differentiation of cells within it. iii) Proliferating fibroblasts are also the main component of the granulation tissue, the new connective tissue with small blood vessels which eventually replaces the blood clot. iv) Epidermal cells migrate to cover the wound area, restoring the protective barrier of the skin. During the final remodelling state, the former wound area is covered with a mesh of cross-linked collagen bundles [3, 4]. The final stage can take months or even years, depending on the specifics of the wound.

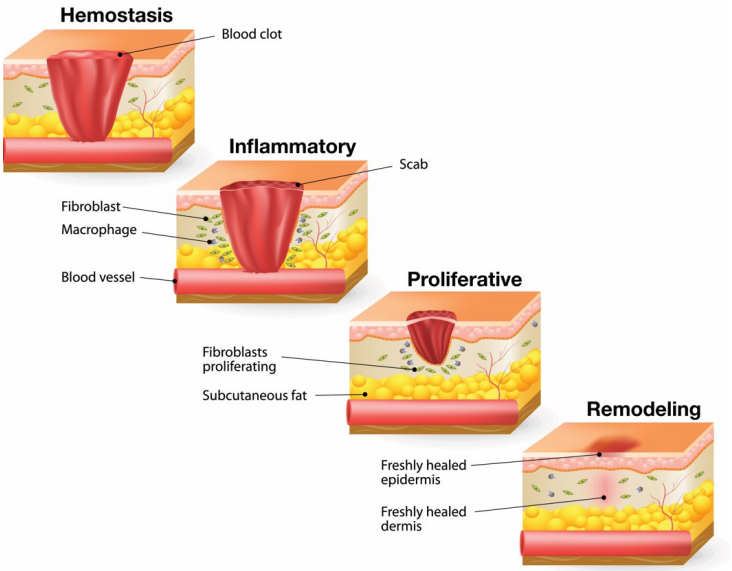


Figure 1: Four stages of wound healing

This project mainly focuses on the role of *fibroblast* cells in the proliferation stage. The fibroblasts

in the wound region can differentiate into *myofibroblasts*, an alternative phenotype which has a cellular structure resembling smooth muscle cells [5]. Both fibroblasts and myofibroblasts produce collagen molecules to reconstruct the ECM and exert pulling forces on the surrounding tissue. However, the pulling force produced by the myofibroblasts is substantially larger [6]. This stress is adaptive for wound repair: it causes the wound to contract. This reduces the surface area of the wound that needs to be covered by newly generated tissue. During a normal healing process, most fibroblasts and myofibroblasts undergo apoptosis in the final remodelling stage, releasing the strain on the dermal tissue [7]. By contrast, excessive proliferation of myofibroblast can cause skin contractures. Both phenotypes have a size of approximately 50 μm, but they have significantly different cell shapes. While fibroblasts have an elongated spindle-like shape, myofibroblasts are typically star shaped (see Figure 2) [4].

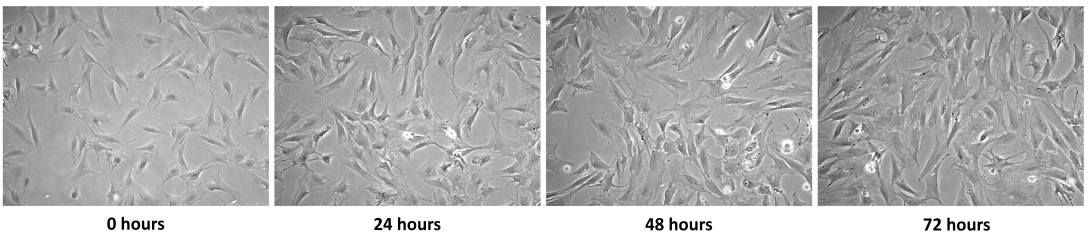


Figure 2: Differentiation of fibroblasts into myofibroblasts (in vitro) [5]

During the proliferation phase, fibroblasts move towards the wound area via *chemotaxis*, the process by which cells move up a concentration gradient. Early research has established that the transforming growth factor-beta (TGF-β) functions as a main chemo-attractant for fibroblasts [8].[1] Recent research has established that the growth factor TGF-β also stimulates the differentiation of fibroblasts into myofibroblasts [9]. However, differentiation only occurs when there is sufficient mechanical tension on the ECM and the ECM is stiff enough (Figure 3).

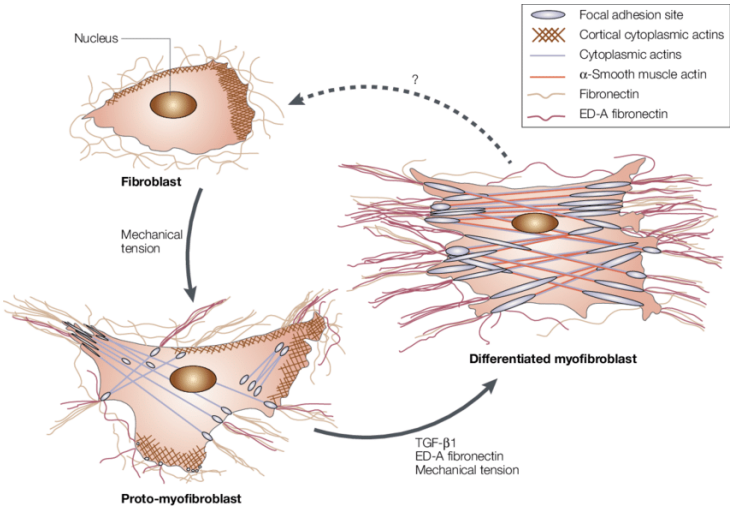


Figure 3: Differentiation of fibroblasts (schematic) [5]

Experimental research has shown that TGF-β induces the differentiation of fibroblasts in 2D and 3D cell cultures. Conversely, mechanical stress alone fails to induce differentiation in absence of TGF-β [6]. However, it is unclear why both TGF-β and mechanical tension are required for the differentiation of fibroblasts into myofibroblasts. At the same time, evidence has been accumulating that there is an extra-cellular structure, the so-called *large latent complex* (LLC), attached to the (myo)fibroblasts and the ECM, which stores quantities of TGF-β together with its binding proteins and other proteins [10].

[1] The platelet-derived growth factor (PDGF), the other main growth factor known to attract fibroblasts, will not be considered in this project.

Combining these findings, Wipff and colleagues proposed a new hypothesis regarding cell differentiation: when the myofibroblasts exert forces on the ECM, the LLC is forced to open and releases the latent TGF-β into the cellular environment. The released TGF-β then promotes further differentiation, which again promotes the release of TGF-β, thus creating a feed-forward loop. The hypothesized release-mechanism is illustrated in Figure 4. The mechanism also explains the rigidity condition for differentiation: if the ECM is not rigid, the pulling forces exerted by the myofibroblasts simply displace the LLC and the ECM, and no additional TGF-β is released [6].

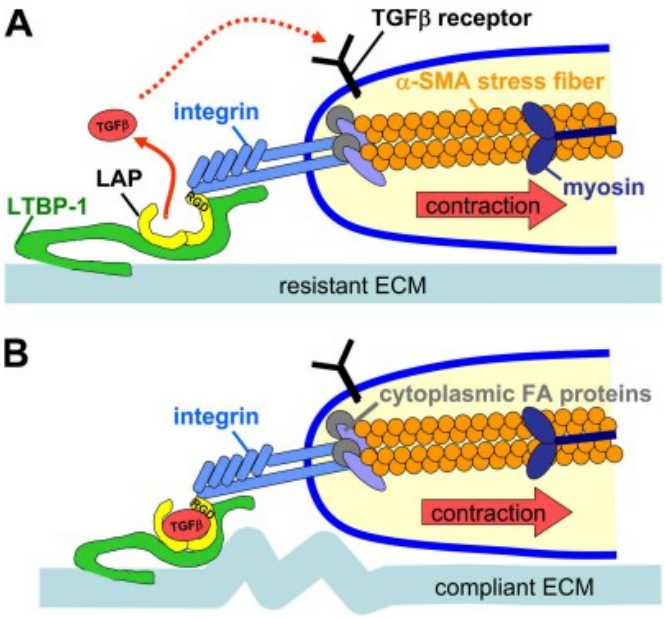


Figure 4: Model of the activation of TGF-β by myofibroblasts reproduced from Wipff *et al.* [6]. (A) When located on a resistant ECM, contraction by a myofibroblast (cell on the right) leads to a liberation of latent TGF-β (in red). (B) By contrast, in a compliant ECM the LLC (yellow-green) is dragged towards the pulling myofibroblast and the TGF-β remains latent.

Wipff *et al.* found empirical confirmation for the release mechanism outlined above [6]. These results include the following: 1) blocking experiments show that contractile activity of myofibroblasts causes the release of TGF-β, 2) Myofibroblasts activate more latent TGF-β from the ECM than fibroblasts (1.5 to 2 fold), 3) Fibroblasts have significantly lower amounts of TGF-β stores in their ECM environment compared to myofibroblasts, 4) Myofibroblasts can de-differentiate into fibroblasts by blocking TGF-β. In this project, we will therefore assume that this release mechanism is the appropriate biological framework for modelling collective migration and differentiation of fibroblasts.

1.2 Research question

While the Wipff-mechanism provides an answer to the question *how* mechanical stress causes the release of TGF-β, it does not explain *why* the release is necessary or effective in the wound healing process. In this sense, the role of the additional release of TGF-β is not yet fully understood.

From a purely theoretical perspective, we can conceive of several possible effects of the additional release of TGF-β. For example, the release of the growth factor during migration can amplify the chemotactic gradient, enhancing both the migration of fibroblasts by chemotaxis and their differentiation. If this is the case, the release of additional growth factor promotes the migration process. Alternatively, the additional release might cause a local maximum in the concentration of the chemo-attractant, which pulls the migrating cells towards it and thereby impedes or stops the cell migration. In this case, the release can be considered as a non-adaptive by-product of the stress exercised on the ECM by myofibroblasts.

This project investigates the role of the additional release of the chemo-attractant using mathematical modelling. By investigating the effects of the release in a reduced model environment, it aims to clarify under which circumstances the additional release of TGF-β promotes the wound healing process.

1.3 Necessity of using mathematical modelling

There are various reasons for using mathematical modelling to understand biological phenomena. For instance, mathematical models can provide insights in processes that are difficult to study experimentally. In addition, when a mathematical model has reached the stage of representing biological reality reasonably well, its output can then be used to suggest new empirical hypotheses [2]. In addition, a mathematical model can act as a 'proof of principle' that an hypothesized biological mechanism produces the expected results.

In this project, the production of new hypotheses is at best a secondary goal. A direct translation of the model to biological reality is not feasible. The main obstacle here is the fact that many biological parameters are still unknown. These include the chemotactic strength, threshold values for differentiation and the average distance to be traversed during migration.[2] Furthermore, an efficient model is preferred, hence, it is unnecessary to include all the details and components in a mathematical model.

Therefore, our first objective is to build a computational model that reproduces the cellular behaviour of migration and differentiation. The second objective is to include the release of latent TGF-β by myofibroblasts in order to study its role in the migration and differentiation process. In other words, the aim is to construct a preliminary model. We expect that the model can give rise to some quantitative and qualitative statements about the circumstances in which the release of additional TGF-β is conducive for the wound healing process.

1.4 Thesis outline

The mathematical model will be presented in Section 2, which includes a description of the Cellular Potts Model (CPM) and the continuum model components. Section 3 then outlines the methods used for analysing the model output. Some preliminary results of the baseline model are presented in section 4. The main numerical results are presented in Section 5. The discussions and conclusions are given in Section 6.

2 Mathematical model

This section outlines the main mathematical model used in this project. First, the model assumptions are discussed. Then the Cellular Potts Model (CPM) and the PDEs are described. Finally, some variants of the mathematical model are introduced.

2.1 Model assumptions

Simplifications and assumptions are needed to construct a model. The assumptions used in this project are listed below:

1. We consider a 2D model for the 3D cellular world. Because cell-cell contacts during wound healing are typically restricted to a small zone close to the skin surface, it is generally considered reasonable to treat cell packing in wound healing as a 2D problem [11].

2. Cellular environment consists only of two types of cell: fibroblasts (Fbs) and myofibroblasts (Mfs). Fibroblasts can differentiate into myofibroblasts, but not vice versa.

3. Only two types of chemicals occur in the model: TGF-β in active and in latent form. The active form functions as chemo-attractant for both cell types. It also promotes cell differentiation of Fbs info Mfs.

[2]Based on survey of the literature mentioned in the bibliography. Some parameters has been determined empirically, such as the diffusion coefficient of TGF-β, which falls within the range $10.6 < D_{TGF} < 100.6\ \mu m^2/h$ [4].

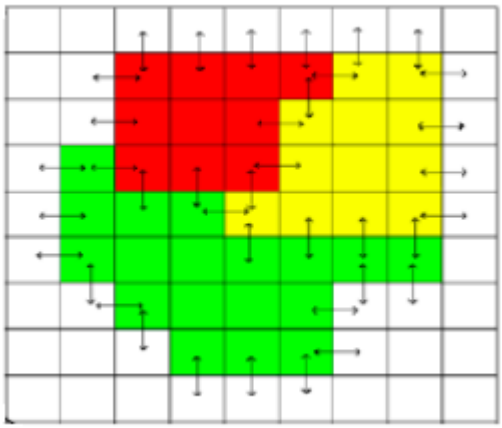


Figure 5: Three cells on the lattice. The non-zero spins are coloured; the ECM ($\sigma = 0$) remains blank.

4. We make all assumptions about cellular dynamics implicit in the Cellular Potts Model. For example, it is assumed that cells are highly deformable and that cell movement occurs via local protrusions and retractions (see below).

5. Cell proliferation and apoptosis (programmed death of cells) are not considered in this project.

6. Only Mfs pull out latent TGF-β from the ECM by cellular forces which strain the ECM. This strain is not explicitly modelled.

7. Following Wipff *et al.* [6], the ECM is assumed to be stiff enough so that the contraction of the ECM releases latent TGF-β. ECM stiffness is not modelled explicitly.

8. The intracellular environment (including cell polarization) is not yet considered in this project.

9. An unlimited and continuous supply of TGF-β from the wound is assumed in all simulations unless otherwise specified.

2.2 The Cellular Potts Model

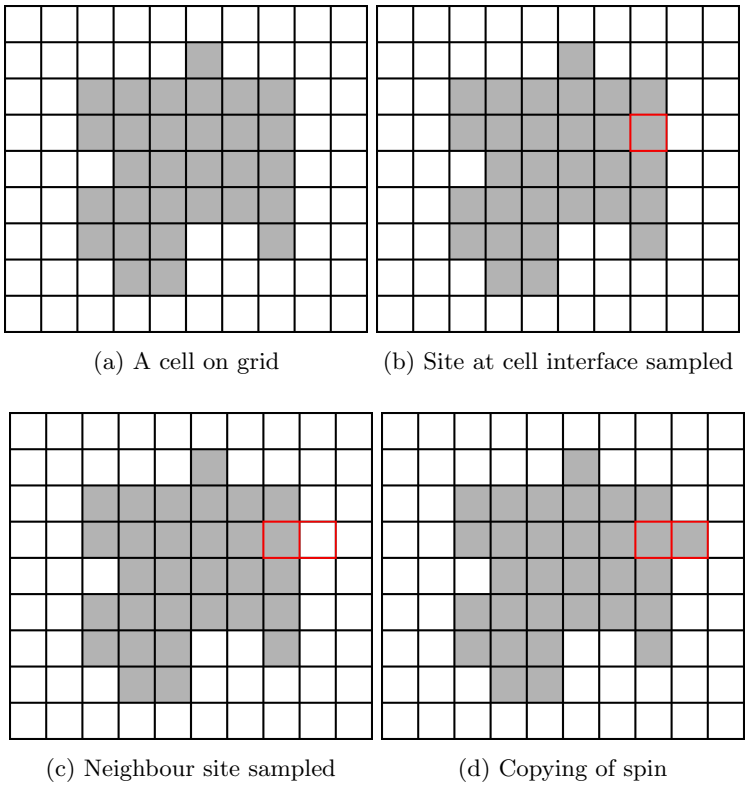
Since we are interested in the migration and differentiation of fibroblasts in a small-scale wound, the Cellular Potts Model (CPM) is selected, which is a typical agent-based cell model. We present a brief description of the CPM; more details of the model can be found in [14].

2.2.1 Representation of cells

Historically, the CPM has been developed as an extension of the large-Q Potts models to simulate the dynamics of biological cells and, more specifically, the biological process of cell sorting [12]. The CPM is a discrete time lattice based model for cellular dynamics. The dynamics are determined by associating energy costs to cellular configurations on the lattice. The total free energy of the system is described by an energy function. During each Monte Carlo time step, the cell configuration changes by cell movements that attempt to minimize the total energy.

In the CPM, a biological cell is represented as a set of several lattice sites. Each biological cell is identified by its characteristic spin $\sigma \in \{1, 2, .., N\}$, where N is the total number of cells in the model. Every biological cell consists of all the lattice sites having identical non-zero spin. For example, Cell 1 contains all lattice sites with spin $\sigma = 1$. These sites typically form a 'patch' on the lattice. Lattice sites that are not occupied by a biological cell are interpreted to be the extracellular matrix (ECM), for which the zero spin has been reserved. Figure 5 presents an example of a lattice configuration with three cells.

Denoting $\boldsymbol{x} = (x, y)$ as an arbitrary lattice site, then the spin of the lattice site is given by $\sigma = \sigma(\mathbf{x})$. Formally, the spin σ defines a mapping from the lattice L to the set of possible spin values: $\sigma : L \rightarrow$



(a) A cell on grid (b) Site at cell interface sampled

(c) Neighbour site sampled (d) Copying of spin

Figure 6: The updating procedure: a successful protrusion.

$\{0, 1, 2, .., N\}$. Note that this mapping is neither injective (lattice sites can have the same spin) nor surjective (cells can apoptose or 'fill up' the extra-cellular matrix). In this project, however, cells do not apoptose and volume restrictions preclude unlimited expansion. The interfaces of two 'patches' are part of a cell boundary. On the lattice, such an interface is represented by two neighbouring lattice sites having different spins. The interfaces together constitute the cell boundary. In other words, the set of all interfaces of lattice sites with spin k (i.e. $\sigma = k \in \{1, 2, ..., N\}$) and lattice sites with a different spin ($\sigma \neq k$) constitute the cell boundary of cell k. This cell boundary or perimeter is interpreted to be the two-dimensional equivalent of the cell surface. Several aspects of the cellular dynamics, such as surface tension and movement by protrusions and retractions, occur at the cell boundary.

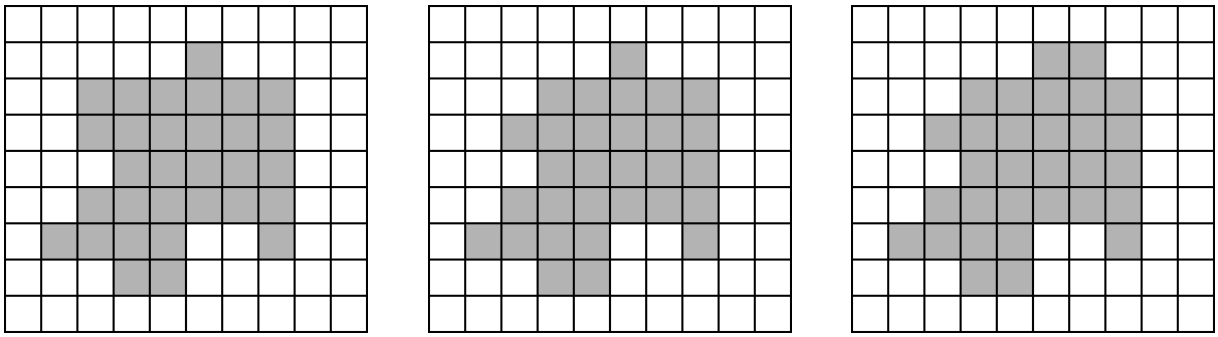
2.2.2 Cell movement

Cells change their position on the lattice when the lattice sites adjacent to their boundary change their spin. During a Monte Carlo time step, lattice sites are sampled randomly. Every sampled lattice site changes its spin into the spin of one of its randomly chosen neighbours with a certain probability. For a second-order neighborhood, this 'update' can only result in changes in the lattice configuration when sites at the cell interface are sampled.

The updating procedure is illustrated in Figure 6, which shows a 'patch' representing a biological cell (gray lattice sites) surrounded by the extracellular matrix (blank sites; see Figure 6a). During the updating process, a site at the cell interface is sampled, as indicated by a red border (Figure 6b). A lattice site in its neighborhood is then selected at random (Figure 6c). An attempt is made to copy the spin of the sampled lattice site into its neighbouring site. This would result in a 'protrusion' of the cell into the ECM. A successful attempt is shown in Figure 6d. Similarly, sites in the ECM can be sampled to copy their spin into an adjacent biological cell. Such a protrusion of the ECM into the cell is also a 'retraction' from the perspective of the cell; see Figure 7 as a schematic.

Given a cell which is surrounded by the ECM, we consider a retraction of that cell. The process is illustrated in the Figures 7a and 7b: a lattice site $\mathbf{x_1} = (3, 8)$ belonging to the ECM copies its spin into

lattice site $\mathbf{x_2} = (3, 7)$ belonging to a neighbouring cell. Therefore, the spin of $\sigma(\mathbf{x_1})$ of the ECM is assigned to the lattice site of the cell: $\sigma(\mathbf{x_2}) = \sigma(\mathbf{x_1}) = 0$. Conversely, a protrusion of the cell into the ECM on lattice site $\mathbf{x_3} = (7, 8)$ is illustrated in Figure 7c.



(a) A cell and the ECM (b) Retraction of cell at site $(3, 7)$ (c) Protrusion of cell at site $(7, 8)$

Figure 7: The updating procedure: retraction and protrusion. The blank pixels represent the ECM.

By the repeated copying of spins at cell interfaces, cells change positions on the lattice. Such displacement effectively constitutes *cell movement* in the CPM. Biologically, the change of spin corresponds to small deformations of cell membrane. These in turn model the protrusions and retractions of filopodia and lamellipodia of migrating cells [13].

The movement of cells is governed by a Monte Carlo algorithm. During one a Monte Carlo step (MCS), the number of randomly sampled updates equals the total number of lattice sites (in this project: $500 \cdot 50 = 25000$ updates). Therefore, every lattice site will, on average, be sampled for an update during one single MCS. We denote a sampled lattice site as $\mathbf{x}_k$, its neighbours as $\mathbf{x}'_k$ and lattice sites of its second-order neighbourhood as $\mathcal{N}(\mathbf{x}_k)$. During every update attempt the following steps are performed:

1. A site $\mathbf{x}_k$ from the lattice is sampled at random with uniform distribution.
2. For every sampled lattice site $\mathbf{x}_k$, a different lattice site $\mathbf{x}_k^{'} \in \mathcal{N}(\mathbf{x}_k)$ is selected at random with uniform distribution.
3. Acceptance or rejection of the spin copy is dependent on an energy function $\mathcal{H}$ (explained below). The algorithm calculates the change in energy $\Delta\mathcal{H}$ which would result if the spin of site $\mathbf{x}_k$ would be copied into site $\mathbf{x}_k^{'}$, i.e. if $\sigma(\mathbf{x}_k^{'})$ becomes $\sigma(\mathbf{x}_k)$. That is: the energy quantity $\mathcal{H}$ is calculated before and after the attempted copy of the spin of site $\mathbf{x}_k$ into site $\mathbf{x}'_k$:

$$\Delta\mathcal{H} = \mathcal{H}_{after} - \mathcal{H}_{before}. \tag{1}$$

4. The copying of the spin is accepted with Boltzmann probability given by:

$$P(\Delta\mathcal{H}) = P(\sigma(\mathbf{x}) \to \sigma(\mathbf{x}^{'})) = \begin{cases} \exp\left(\frac{-\Delta\mathcal{H}}{T}\right), & \text{if } \Delta\mathcal{H} \geq 0, \\ 1, & \text{if } \Delta\mathcal{H} < 0, \end{cases} \tag{2}$$

where $T > 0$ denotes the 'temperature' of the system.

Updates that decrease the energy of the system are always accepted. In short, cellular movement in the CPM is modelled as an attempt to minimize energy.

The temperature T is not the thermal temperature in physical systems (as it is known from precursors to the CPM), but rather models the inherent noisy character of cell movement. Hence, increasing the temperature increases the probability that energetically unfavorable updates will be accepted. When $T \to \infty$, $P(\Delta\mathcal{H}) \to 1$, i.e. every update attempt would be accepted. Note that the spin of a lattice site can be replaced by the spin of its neighbouring cell even when this increases the overall energy. Biologically, the possible acceptance of energetically unfavourable updates corresponds with the random character of exploratory extensions and retractions performed by the cells. The noise term also enables cells to escape local optima, for instance in the context of chemotaxis [13].

2.2.3 The energy function

Changes in the lattice configuration are determined by using an energy function or *Hamiltonian*, which calculates the total energy of a lattice configuration. The Hamiltonian is given by:

$$\mathcal{H} = \sum_{\mathbf{x}} \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} \underbrace{J[\tau(\sigma(\mathbf{x})), \tau(\sigma(\mathbf{x}^{'}))]}_{\text{adhesion}} \underbrace{(1 - \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x}')})}_{\text{on interfaces}} + \underbrace{\lambda_{A,\tau} \sum_{\sigma} (a_\sigma - A_{\tau_\sigma})^2}_{\text{area constraint}} \tag{3}$$
$$+ \underbrace{\lambda_{L,\tau} \sum_{\sigma} (l_\sigma - L_{\tau_\sigma})^2}_{\text{length constraint}} + \underbrace{\lambda_{P,\tau} \sum_{\sigma} (p_\sigma - P_{\tau_\sigma})^2}_{\text{perimeter constraint}}.$$

All terms of the Hamiltonian are explained below. As indicated by Eq. (1), the quantity $\mathcal{H}$ is calculated twice during every update attempt.

The first term on the right-hand side contains a double sum: the first is over all lattice sites $\mathbf{x}$, and the second over all neighbouring sites $\mathbf{x}'$ of $\mathbf{x}$. The term $\delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x}')}$ is the Kronecker delta function. Lattice sites not located at a cell interface cannot change their spin by copying in a second-order neighborhood and therefore cannot change the energy of the system. The term $(1 - \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x}')})$ ensures that only sites at the cell interfaces are considered. A rudimentary version of the Hamiltonian in Eq. (3) can therefore be obtained by the double sum of the term $(1 - \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x}')})$. The cell interfaces now increase the total energy of the system:

$$\mathcal{H}_1 = \sum_{\mathbf{x}} \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} (1 - \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x}')}). \tag{4}$$

Biologically, the energy increase corresponds to the surface tension between cell interfaces.

Adhesion or interaction energies are introduced to specify the bonds between cells. The adhesion energy J of these bonds depends on the cells involved, hence $J = J(\sigma(\mathbf{x}), \sigma(\mathbf{x}'))$. However, adhesion energies typically do not depend on the cell per se, but on the cell *type*, such as fibroblast and myofibroblast. We therefore introduce the function $\tau : S \to T$, which maps the set of spins S to the set of cell types T. The adhesion energy now depends on the cell types of two cells at an interface: $J(\tau(\sigma(\mathbf{x})), \tau(\sigma(\mathbf{x}^{'})))$. Incorporating the adhesion energy in the Hamiltonian yields:

$$\mathcal{H}_2 = \sum_{\mathbf{x}} \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} J[\tau(\sigma(\mathbf{x})), \tau(\sigma(\mathbf{x}^{'}))](1 - \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x}')}). \tag{5}$$

Adhesion energies were introduced into the CPM by Graner and Glazier [12] to reproduce spontaneous cell sorting processes.

Cells typically have more or less constant geometrical properties such as volume and perimeter. The other three terms in Eq. (3) are constraints to maintain these properties. For example, the first term is a volume or *area constraint*, which ensures that cells do not deviate too much from their natural size. The area constraint is given by:

$$\lambda_{A,\tau} \sum_{\sigma} (a_\sigma - A_{\tau_\sigma})^2. \tag{6}$$

Here, a_σ denotes the current area of the cell and A_{τ_σ} the target area which is the same for cells of the same phenotype. The cell area a_σ is simply the number of lattice sites with spin σ. The coefficient $\lambda_{A,\tau}$ is a Lagrange multiplier which determines the strength of the constraint for all cells of the same cell type. Biologically, the multiplier models the cell's incompressibility. The area constraints sums over all cells (indicated by their spin σ). For each cell, it increases the total energy of the system with a quantity proportional to the quadratic deviation of its target area.

Analogously, a length constraint and perimeter constraint are added to the Hamiltonian. Since most cells naturally have an elliptical shape, the lattice imposes an unnatural geometry upon the cells. The cell length is therefore estimated by calculating an elliptical approximation of the cell (see Appendix A). Cell length then equals twice the length of the semi-major axis of the elliptical approximation. The

perimeter constraint models the rigidity of the cell membrane (the cell cortex). In this two-dimensional case, the perimeter equals the sum of all cell interfaces. Motivation for regarding this sum as a reasonable approximation of the cell perimeter are given by Magno *et al.* [11]. All shape constraints are explained with illustrations in Appendix A.

Adding the terms for the shape constraints completes Eq. (3). However, since the length constraint may cause cells to split up in disconnected patches, an additional constraint is needed to ensure the cohesion of biological cells [13]. This *connectivity constraint* is only applied to cells but not the ECM. Details of the connectivity constraint are explained in Appendix B.

In summary, the CPM models the dynamics of cells in terms of energy minimization. The dynamics are completely described by Eqs. (1)-(3). The CPM does not explicitly model the forces on cells causing the properties of cell geometry, but instead uses additional constraints in the Hamiltonian to model the cell geometry.

2.2.4 Chemotaxis

Chemotaxis is cell movement up the gradient of a concentration field. In this model, both fibroblasts and myofibroblasts are striving for a higher concentration of TGF-β. We write L for the two-dimensional lattice and $u : L \longrightarrow \mathbb{R}_{\geq 0}$ for the concentration field of the active form of TGF-β (suppressing time dependency, see below). Chemotaxis is now modelled by:

$$\Delta\mathcal{H}_{chemotaxis} = \mu(u(\mathbf{x}_i) - u(\mathbf{x}_j)), \tag{7}$$

where $u(\mathbf{x}_i)$ denotes the concentration of active TGF-β at position $\mathbf{x}_i$ at and $\mu > 0$ denotes the chemotactic strength [14]. The difference $\Delta\mathcal{H}$ of Eq. (7) is calculated during an update attempting a spin copy of $\mathbf{x}_i$ into $\mathbf{x}_j$ (i.e. $\sigma(\mathbf{x}_i) \to \sigma(\mathbf{x}_j)$) and added to the left hand side of Eq. (1). When $u(\mathbf{x}_j) > u(\mathbf{x}_i)$, then $\Delta H < 0$, hence, cell movement in the direction of $\mathbf{x}_j$ is favoured.

2.2.5 CPM settings

Throughout this project, the CPM is implemented on a rectangular 500×50 lattice. A second-order neighborhood (8 neighbouring sites) is used for the updating algorithm. Fbs cells ($\tau = 1$) differentiate into Mfs ($\tau = 2$) when two conditions are met: i) the local concentration $u(\mathbf{x})$ of the chemo-attractant exceeds some threshold value (default $ths = 0.1$) and ii) $x > 0.8$ where X is a uniform random variable $X \sim U[0, 1]$.

All adhesion energies between cells J_{τ_1,τ_2} are set to 1 and all cell-ECM energies are set to 2. If the cell-ECM energies are smaller than the cell-cell energies, larger cell aggregates will dissociate into cells consisting of isolated spins, which hardly mimics natural circumstances [12]. The temperature is set to $T = 1$, therefore, acceptance of energetically unfavourable updates is relatively uncommon. The area constraint and connectivity constraints will be used in all models for both fibroblasts and myofibroblasts (with target $A_{\tau=1} = A_{\tau=2} = 40$) and multiplier $\lambda_A = 1$. Note that no geometrical constraints are chosen for the ECM due to its ambient nature.

2.3 Continuous model components

This section outlines the equations governing the chemo-attractant TGF-β. We distinguish i) an unbound or 'active' form $u(\mathbf{x}, t)$, which can diffuse freely through the ECM, and ii) the 'bound' form $b(\mathbf{x}, t)$, which is assumed to be roughly attached to the ECM. Both Fbs and Mfs only chemotact towards higher concentrations of the *active* form $u(\mathbf{x}, t)$.

Writing simply u and b for $u(\mathbf{x}, t)$ and $b(\mathbf{x}, t)$, the coupled model equations are given by:

$$\frac{\partial u}{\partial t} = -\epsilon u + \gamma(x)b - \beta u + \alpha b + K_u\nabla^2 u, \tag{8}$$

$$\frac{\partial b}{\partial t} = p_{sc}(x) - \gamma(x)b + \beta u - \alpha b + K_b\nabla^2 b. \tag{9}$$

Eq. (8) models the concentration of the active form u of TGF-β. The term $\gamma(x)b$ represents the release of the bound form b of TGF-β into the cellular environment by Mfs, which in this model amounts to a

conversion of the bound form b into the active form u. The terms α and β in Eqs. (8) and (9) represent the base rate conversion of TGF-β from its bound form b to the unbound form u (parameter α) and vice versa (parameter β), respectively. Here ϵ is a degradation term.

Eq. (9) models the concentration of the bound form b. It depends on the secretion of TGF-β denoted by $p_{sc}(x)$, which is assumed to occur only in the 'wound' at the end of the domain, i.e. $x > 498$. The release $\gamma(x)b$ occurs on two conditions: 1) the release is caused only by Mfs ($\tau = 2$) and 2) does not occur near the wound itself (i.e. $x \geq 490$). Formally:

$$p_{sc}(x) = p_0 \cdot \mathbf{1}_{\mathbf{x}>498}, \qquad \gamma(x) = \tilde{\gamma} \cdot \mathbf{1}_{\mathbf{x}<490, \tau_\sigma(\mathbf{x})=2} \tag{10}$$

where p_0 and $\tilde{\gamma}$ are positive secretion and release constants. Although both u and b diffuse through the cellular environment, the relatively stable character of the bound form b is modelled by choosing its diffusion coefficient small in comparison, viz. $K_b = 100$ and $K_u = 400$.

A few other model assumptions are implicit in the Eq. (8)-(10). The TGF-β is assumed to be bound to the LCC or ECM immediately after its release from the wound. Hence the secretion term $p_{sc}(x)$ only features in Eq. (9). The base rate conversion of the bound to the active form ensures that the active form diffuses through the domain. Secretion $p_{sc}(x)$ has no time-dependency or depletion. Eq. (10) reflects the assumption that there is no conversion of TGF-β from the bound form into the active form at the end of the domain (i.e. 'wound' itself). If admitted, this conversion would render the solution unstable and erratic. This can be explained by the cellular dynamics. Cells that reach the wound through chemotaxis differentiate into Mfs.[3] In the wound area, these cells encounter an inexhaustible supply of bound TGF-β. The Mfs then convert this into the active form, causing a surge in active TGF-β. This unnaturally high concentration then diffuses but also attracts the cells, causing erratic patterns of cell movement. To rule out this effect, it is assumed that there is no conversion near the end of the domain (i.e. $x \geq 490$).

The parameter values used throughout this project (unless indicated otherwise) are listed in the table below. The initial conditions for u and b are chosen to be: $u_0 = 10^{-5}$ and $b_0 = 10^{-4}$. Note that all the parameters are dimensionless so far, since the model is still in the preliminary phase.

Parameter List		
Parameter	*Description*	*Value*
ϵ	degradation (only for u)	$2.5 \cdot 10^{-6}$
α	base rate conversion $b \rightarrow u$	$5 \cdot 10^{-3}$
β	base rate conversion $u \rightarrow b$	$5 \cdot 10^{-3}$
K_u, K_b	diffusion coefficients of u and b	400, 100
p_0	release of bound form b by 'wound'	0.1

Homogeneous Neumann boundary conditions are imposed on the vertical domain boundaries ($x = 0, x = 500$) and periodic boundary conditions are used on the horizontal domain boundaries ($y = 0, y = 50$). Eqs. (8) and (9) are numerically solved in the Morpheus package using the Heun solver (improved Euler) with $\Delta t = 1.0$.

2.4 Model variants

This section outlines the three models with increasing complexity used in this project.

2.4.1 Model 1: baseline

The baseline the model is represented by Eqs. (8)-(10) with parameters as listed in the table above. When running multiple simulations, the varying parameters are the release term $\gamma \in \{0.1, 0.2, ..., 1\}$ and the chemotactic strength $\mu \in \{50, 60, .., 100\}$. Both Fbs and Mfs have identical connectivity and volume constraint ($A_{\tau_\sigma=1} = A_{\tau_\sigma=2} = 1, \lambda_A = 1$) and no other constraints on geometry. The model is initialized with 10 Fbs and 1 Mf with an off-set of $x = 10$. The simulation end time is $t_{end} = 2000$ and the differentiation threshold is set to $ths = 0.1$.

[3]The exact percentage of Mfs in the wound region in biological wounds is as yet unknown.

2.4.2 Model 2: variable release

In the baseline model Fbs differentiate into Mfs instantaneously. In reality differentiation is a gradual process, since it involves a gradual development of the cytoskeleton. In Model 2 the release of TGF-β by Mfs is therefore made to depend on time:

$$\gamma_{var}(t) = \begin{cases} 0 & t < \tau_d \\ \gamma \cdot (1 - e^{-k(t-\tau_d)}), & t \geq \tau_d, \end{cases} \tag{11}$$

where γ is a constant and τ_d indicates the time when differentiation starts, i.e. the time when the two conditions for differentiation have been met (cf. 2.2.5). We choose $k = 1/150$ such that at $t = \tau_d + 300$, we have $\gamma_{var} = \gamma(1 - \frac{1}{e^2}) \approx 0.86\gamma$. The release term $\gamma_{var}(t)$ now changes gradually, mimicking the gradual differentiation of Mfs. The Fbs still do not change shape upon differentiation. The model is initialized with 10 Fbs (and no Mfs) with an off-set of $x = 30$. The simulation end time is $t_{end} = 4000$ and the differentiation threshold is set to $ths = 0.05$.

2.4.3 Model 3: constraints on cell shape

Model 3 also includes the geometrical differences between Fbs and Mfs. Both cell types already share the volume constraint and connectivity constraint. In addition, a length constraint is now imposed on the Fbs (target $L_{\tau_\sigma=1} = 20, \lambda_L = 1$) and a perimeter constraint on the Mfs (target $P_{\tau_\sigma=2} = 40, \lambda_P = 2$) to mimic the elongated shape and star-shape of Fbs and Mfs respectively. To model a gradual development, the target perimeter for Mfs is also made to depend on time. The perimeter constraint is given by:

$$\lambda_{P,\tau=2} \sum_{\sigma} (p_\sigma - P_{var}(t))^2, \text{ and } P_{var}(t) = \begin{cases} 0, & t < \tau_d, \ \ \lambda_P = 0 \\ P_{\tau_\sigma=2} \cdot (1 - e^{-k_2(t-\tau_d)}), & t \geq \tau_d, \ \ \lambda_P = 2. \end{cases} \tag{12}$$

where the sum is over all Mfs ($\tau = 2$) and $k_2 = 1/75$. The release term γ is variable as in Eq. (11). The model is initialized with 10 Fbs (and no Mfs) and an off-set of $x = 20$. The simulation end time is $t_{end} = 7000$ and the differentiation threshold is set to $ths = 0.2$. All CPM interaction energies are set to 1.

3 Analytical tools

This section outlines the measures and indicators used to quantify the model output.

3.1 Mean squared displacement

The *mean squared displacement* (MSD) is used to quantify cell displacement. It indicates the mean displacement of a population of cells as a function of time:

$$MSD(t) := \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{x_t^{(i)}} - \mathbf{x_0^{(i)}}||^2, \tag{13}$$

where $\mathbf{x_t}$ denotes a cell's position at time t, $\mathbf{x_0}$ represents its initial position and N denotes the total number of cells.

The MSD provides a useful measure for cell motility, since several relations are known to hold between the MSD and the nature of the movement [15]. In an one-dimensional unbiased random walk, it can be verified that MSD increases linearly in time: $MSD(t) = 2Dt$, where D denotes the 'diffusion' of the cell over a one-dimensional lattice. However, our current CPM-based model contains both random cellular movement, which can be understood as diffusion of cells, and the movement in a preferred direction (bias) due to chemotaxis, which can be modelled as drift. Cell movement in the hybrid model can therefore be described by the *drift-diffusion* equation. Writing $p(x,t)$ for the probability that a cell is at position x at time t, the equation is given by:

$$\frac{\partial p}{\partial t} = -u\frac{\partial p}{\partial x} + D\frac{\partial^2 p}{\partial x^2}, \tag{14}$$

where the first term on the right represents drift due to chemotaxis and the second term represents cell diffusion. Here $u > 0$ is the drift (advection) velocity and $D > 0$ the cell diffusion constant. The solution of Eq. (14) is given by:

$$p(x,t) = \frac{1}{\sqrt{4\pi Dt}} \exp\Big(\frac{-(x-ut)^2}{4Dt}\Big), \tag{15}$$

which is a normal distribution with mean ut and variance $2Dt$ [15]. Hence, the first en second moments are given by:

$$\mathbf{E}[X_t] = ut, \tag{16}$$

$$\mathbf{E}[X_t^2] = (ut)^2 + 2Dt \tag{17}$$

The mean displacement $\mathbf{E}[X_t]$ equals ut as expected: random movements average out and the average cell displacement therefore depends solely on drift, which is linear in time. The mean squared displacement $\mathbf{E}[X_t^2]$ depends quadratically on time.

This implies that cell motility in the current model can be estimated by calculating a quadratic fit of the measured MSD values. In this project, we use a polynomial regression to estimate the cell motility u_{mot}. Based on Eq. (17) the constant term (intercept) of the regression is set to zero.

3.2 Directional persistence

The *directional persistence* (DP) quantifies how erratic the average cell movement is. It is the quotient of the effective displacement over the total distance traversed, averaged over the cell population:

$$DP(t_{end}) := \Big\langle \frac{||\mathbf{x}_{t_{end}} - \mathbf{x}_0||}{\sum_{t=0}^{t_{end}-1} ||\mathbf{x}_{t+1} - \mathbf{x}_t||} \Big\rangle, \tag{18}$$

where the brackets indicate the cell population average, the subscript t the time steps and t_{end} the simulation end time (see below). Note that $0 \leq DP \leq 1$, where DP increases towards 1 when cells move in a straight line. In contrast, the DP is low when cells traverse a long distance for a small effective displacement. The DP decreases to zero when the cells traverse an infinitely long trajectory.

3.3 Clustering and success criterion

In some simulations cells split up into two separate clusters. In this case, either both clusters migrate but at different times or only one cluster migrates (see e.g. Figure 10 below). When migration occurs in two clusters, measures such as the MSD are calculated by taking the weighted average over both clusters. A simulation in this project is considered to be *successful* when all initialized cells migrate and reach the right domain boundary, either in one cluster or in two consecutive clusters.

4 Preliminary results of the baseline model

The role of the baseline is to provide a first exploration of cell behaviour in the hybrid CPM-PDE model. We run the baseline model with varying parameters $\gamma \in \{0.1, 0.2, ..., 1\}$ (i.e. a constant release of TGF-β) and chemotactic strength $\mu \in \{50, 60, .., 100\}$. Some typical results of the baseline model are illustrated in Figures 8-10 for parameter values $\gamma = 0.2$, $\mu = 50$ at several times. Fbs are depicted as gray, Mfs as green cells. The concentration of the active form of the chemo-attractant u has been coloured.

Due to the randomness present in the initialization and the updating algorithm, three typical patterns of cell behaviour are observed:

1. *Migration and differentiation.* Fbs migrate over the domain and subsequently differentiate into Mfs (Figure 8). The cells either migrate simultaneously or in different clusters.

2. *Stationary differentiation.* Fbs do not migrate and differentiate at the left side of the domain (Figure 9).

(a) After initialization, $t = 25$

(b) Differentiation starts, $t = 1000$

(c) Full differentiation, $t = 1025$

(d) Migration and differentiation completed, $t = 1600$

Figure 8: Baseline model: migration & differentiation ($\gamma = 0.2, \mu = 50$).

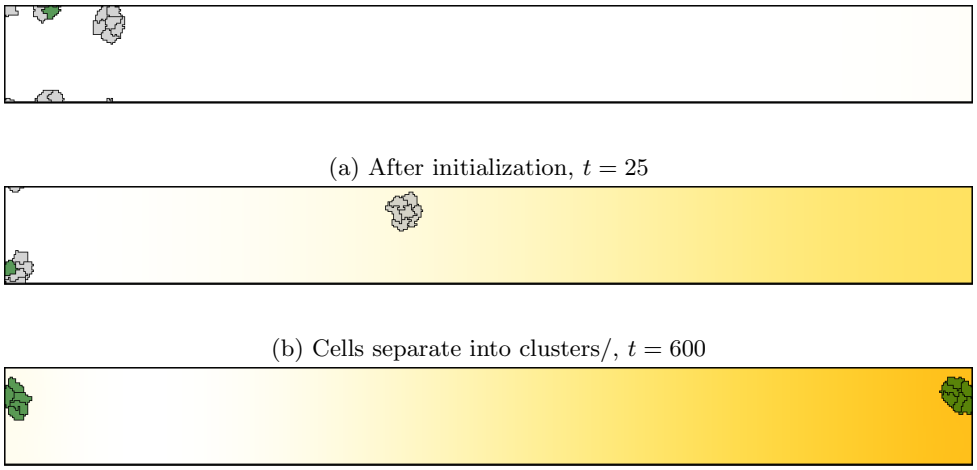
3. *Fission.* Fbs split up into a migrating and non-migrating cluster. Both clusters then differentiate at different sides of the domain (Figure 10).

(a) After initialization, $t = 25$

(b) No migration, $t = 600$

(c) Full differentiation, $t = 1600$

Figure 9: Baseline model: differentiation, no migration ($\gamma = 0.2, \mu = 50$).



(a) After initialization, $t = 25$

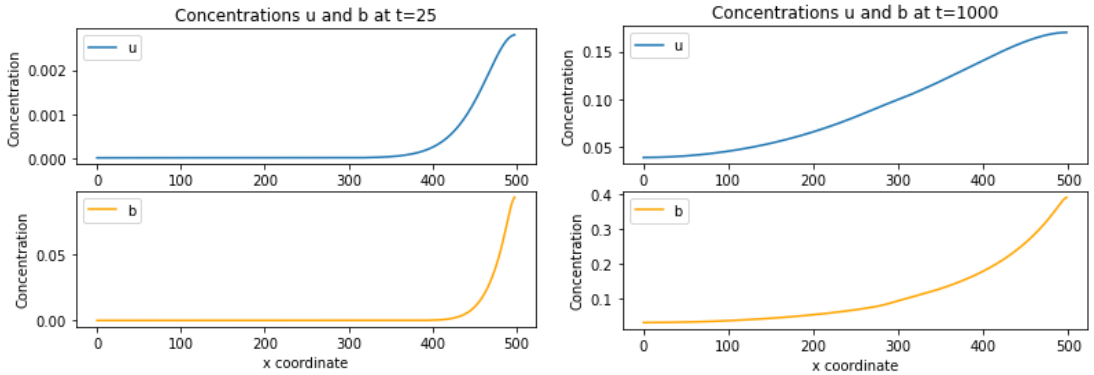
(b) Cells separate into clusters/, $t = 600$

(c) Full differentiation, $t = 1600$

Figure 10: Baseline model: fission and differentiation ($\gamma = 0.2, \mu = 50$).

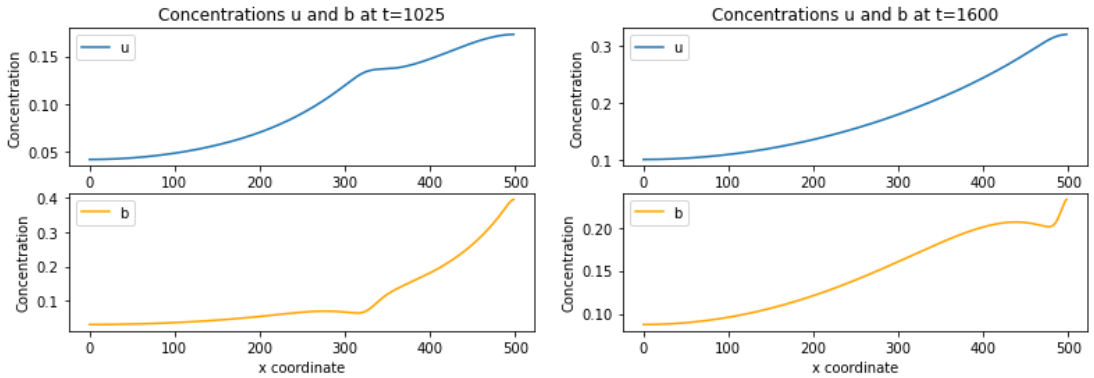
In all scenario's the cells eventually differentiate due to the diffusion of TGF-β. However, in both the second and third scenario the wound healing process is 'unsuccessful' in the sense that a part of the Mfs has not arrived at the 'wound' at the right domain end.

The concentrations of the active form u and the latent form b of TGF-β for the first scenario (Figure 8) are depicted in Figure 11. The concentration values have been averaged over the y-coordinates.



(a) After initialization, $t = 25$

(b) Differentiation starts, $t = 1000$



(c) Differentiation complete, $t = 1025$

(d) Migration complete, $t = 1600$

Figure 11: Average concentrations for a baseline model run for various times ($\gamma = 0.2, \mu = 50$).

At $t = 25$ (Figure 11a), a higher concentration of the latent form b of TGF-β is present at the right side of the domain compared to the active form u (note the difference in scales). This is due to the fact that the TGF-β excreted by the wound is immediately bound to the LLC/ECM and the base rate of conversion (α) to the active form is low. The active form has diffused more due to its higher diffusion coefficient. This difference is also manifest in the plot for $t = 1000$ (Figure 11b), which shows the concentration of u approaching a linear gradient. At $t = 1025$ (Figure 11c), all Fbs have differentiated into Mfs which convert the bound form b of TGF-β into the active form u, causing a bulge and a cavity in the graphs of the concentrations of u and b respectively near $x = 330$. At $t = 1600$, all Mfs have migrated to the right domain boundary. Conversion from bound to active form causes a rise in u and a decrease of b, excepting the wound itself ($x \geq 490$), where no conversion occurs.

Since different migration patterns are observed, the baseline model is at times successful in reproducing the natural cell behaviour of migration and differentiation.

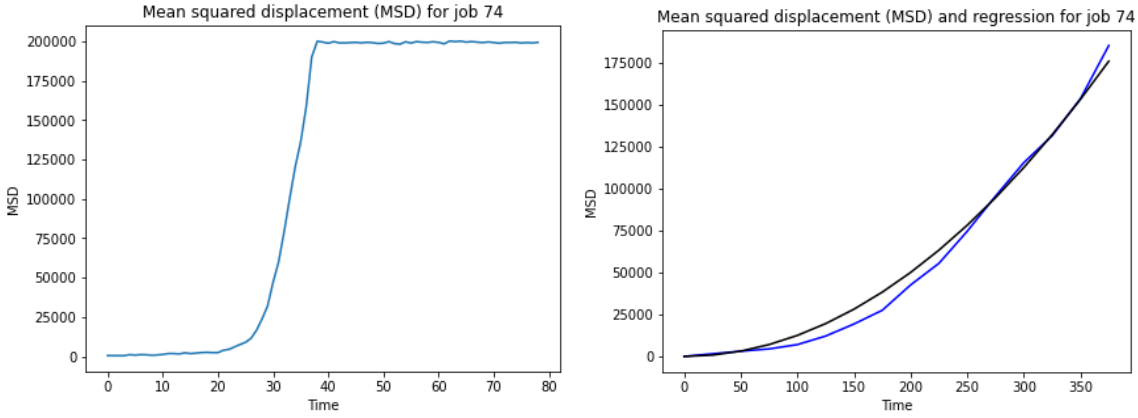
5 Numerical results

5.1 Cut-off values

To assess the effects of the additional release of TGF-β on cell migration, the first quantity to consider is cell motility. In order to obtain a reasonable estimate of cell motility, cut-off times for the regression have to be determined. On the one hand, the active form of TGF-β, which is mainly released from the wound at the right side, needs sufficient time to diffuse through the domain. Therefore, the cluster initially barely moves towards the wound. On the other hand, since the MSD is a non-decreasing function of time, the MSD stays more or less stable after the cells have reached the wound boundary. We therefore consider only a part of the simulation for the regression. We choose as beginning time T_b the first time when the MSD exceeds 5000 units for four consecutive times steps (as multiples of 25); as the end time T_e we choose the first time when the average $x-$coordinate of the cells exceeds 480, i.e.:

$$T_b = \arg\min_{t>0}\{MSD(\tau) > 5000, \forall\tau \in \{t, t+25, t+50, t+75\}\},$$
$$T_e = \arg\min_{t>T_b>0}\left\{\frac{1}{M}\sum_{\sigma} x_\sigma(t) > 480\right\}.$$

where $x_\sigma(t)$ represents the x-coordinate of cell σ at time t and the sum is over all cells in a considered cluster ($\sigma \in \{1, 2, .., M\}$). Effectively, T_b and T_e now represent the times when cells start and stop migrating substantially.



(a) MSD over time

(b) MSD (blue) and regression curve (black)

Figure 12: MSD for the baseline model

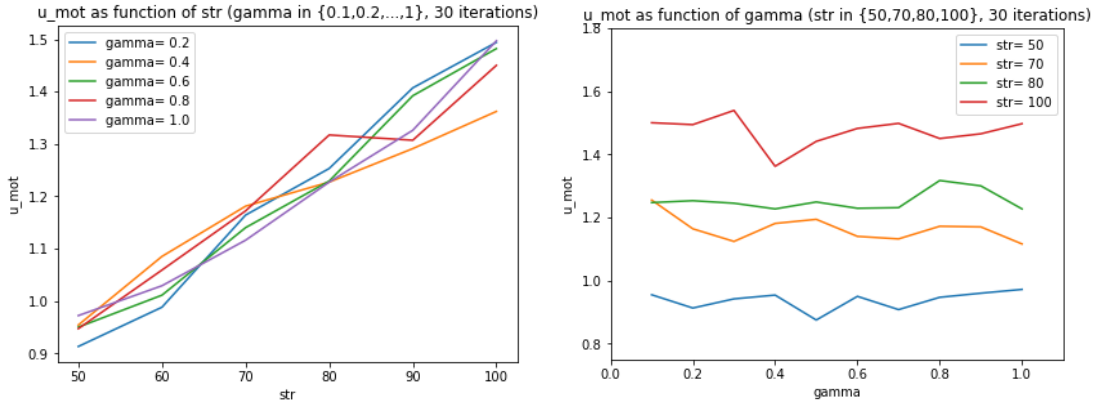
Figure 12a shows the typical MSD profile of a successful simulation. At first, the cells hardly move from their initial position. However, when the chemo-attractant diffuses and reaches the cells, migration

starts and the MSD increases until it plateaus after migration. Figure 12b shows both the MSD and the quadratic regression, where the time-axis has been shifted to start at the beginning time of the regression. For successful runs, we observe that the quadratic regression neatly matches the MSD profile (generally $R^2 > 0.91$)[4] on the domain determined by the cut-off values, as was expected based on Eqs. (14)-(17).

For example, for the run plotted in Figure 12, cut-off values are calculated to be $T_b = 575$ and $T_e = 950$, which yields the regression coefficients $a = 1.25231804$ and $b = 1.51665855 \cdot 10^{-16}$ for the fitted curve $y_{fit} = at^2 + bt$. In accordance with Eq. (17), the cell motility is now calculated to be $u_{mot} = \sqrt{a} \approx 1.119$. In this case, when the cells have started migrating substantially, they traverse approximately 1.12 grid units per time unit.

5.2 Results of Model 1: baseline

This section presents the results of the baseline model, i.e. Eqs. (8)-(10), with parameters values $\gamma \in \{0.1, 0.2, ..., 1\}$ and $\mu \in \{50, 60, .., 100\}$. We conducted 30 runs for each pair of (γ, μ), subsequently there are $N = 1800$ runs in total in this section. For the sake of studying cell motility, only successful runs will be considered. In total, in 821 runs out of 1800 all cells eventually migrate. For these successful runs only, Figure 13 shows the average estimated cell motility as a function of the release term γ and of the chemotactic strength μ.



(a) Cell motility and chemotactic strength μ

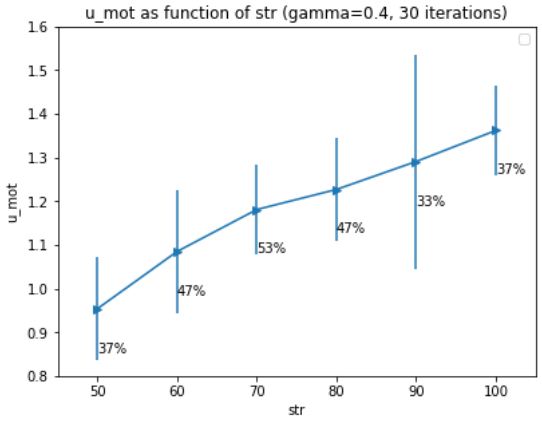
(b) Cell motility and release term γ

Figure 13: Average cell motility (u_mot) of the 821 successful runs out of 1800 of the baseline model.

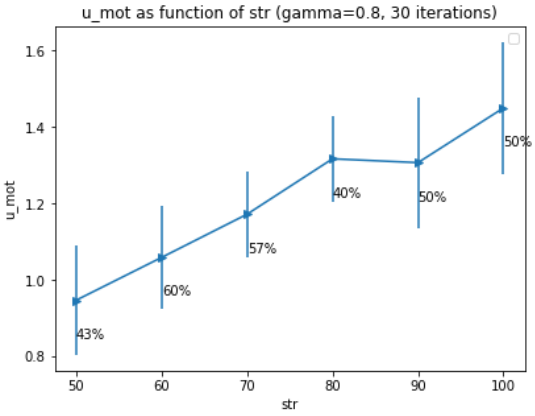
Figure 13a indicates that the average cell motility depends linearly on chemotactic strength. This follows from Eq. (7): higher chemotactic strength incurs a higher energy penalty for every movement which is not directed up the gradient. Figure 13b suggests that there is no obvious relation between the release term γ and average cell motility in the baseline model. This can be explained by the fact that the chemotactic strength is set too high, which causes local increases of TGF-β to have relatively little effect compared to the global TGF-β gradient.

Figure 14 shows the average cell motility for two fixed release parameters. The percentages indicate the number of successful runs out of 30 iterations for a specific (γ, μ)-pair. Figures 14a and 14b do not indicate a relation between the number of successful runs and chemotactic strength.

[4]For the 821 successful runs out of the 1800 runs of the baseline model, 4 runs with clustering have $R^2 < 0.91$, but at least $R^2 \geq 0.86$.



(a) Cell motility for $\gamma = 0.4$.



(b) Cell motility for $\gamma = 0.8$.

Figure 14: Average cell motility of the 821 successful runs of the baseline model as function of chemotactic strength with $\gamma \in \{0.4, 0.8\}$. Vertical error bars indicate standard deviations. The percentages indicate the number of valid runs out of 30 iterations.

We also study the migration process by investigating the amount of time required for a certain percentage the Fbs to differentiate. Figure 15 shows the average time required for at least 75% of the Fbs to differentiate into Mfs. As expected, the trend in Figure 15 is decreasing. Since average cell motility increases with chemotactic strength, cells need less time to migrate towards higher concentrations and reach the differentiation threshold.

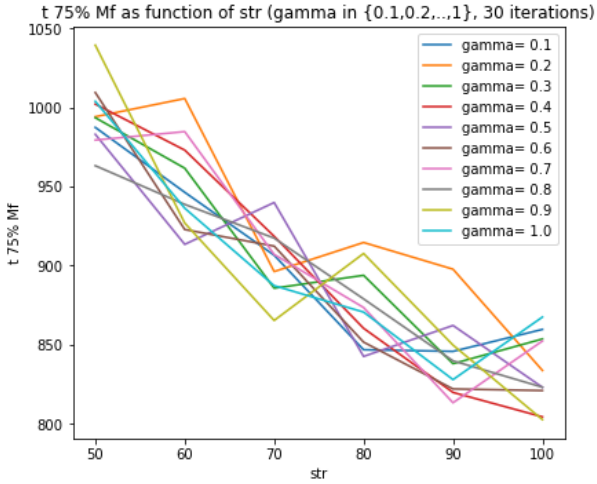
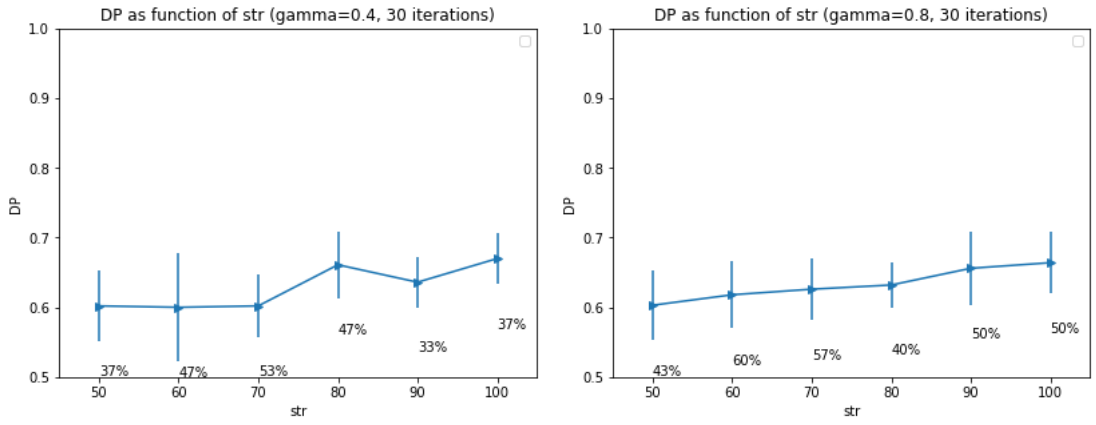


Figure 15: Average time when $\geq 75\%$ of Fbs have differentiated, in the 821 successful runs of the baseline for fixed $\gamma \in \{0.1, 0.2, .., 1\}$.

Figure 16 shows that the directional persistence (DP) is slightly increasing with increasing chemotactic strength. Movements in any other direction than up the gradient become increasingly unlikely as the chemotactic strength increases.

QB	Migratory pattern	DP	Number
2	All cells migrate collectively.	$0.48 \leq DP \leq 0.79$	711
1	All cells migrate, sometimes in multiple clusters.	$0.44 \leq DP \leq 0.71$	110
0	There is a non-migrating cell cluster.	$0.08 \leq DP \leq 0.63$	388
-1	No cells migrate.	$DP < 0.08$	591

Table 1: Qualitative behaviour and range of average DP scores for the baseline (1800 runs)



(a) DP for $\gamma = 0.4$

(b) DP for $\gamma = 0.8$

Figure 16: Average DP of the 821 successful runs of the baseline model for fixed $\gamma \in \{0.4, 0.8\}$. The percentages indicate the number of successful runs out of 30 iterations.

For the sake of quantifying and categorizing migratory behaviour, we can introduce a *Qualitative Behaviour* (QB) score for different patterns of cell migration. We might be interested in the question whether these migration patterns can be associated to the DP scores. Table 1 tabulates four migratory patterns (e.g. 'all cells migrate collectively') and the associated range for the DP scores for the simulations of the baseline. The DP ranges overlap: runs displaying different migratory patterns can have the same DP score. This means that there is no straightforward mapping from the DP to the migratory pattern (QB).

Figure 17 shows a scatter plot of DP and QB scores in the baseline model. The overlap of DP scores for different QB scores is clearly visible. Thus, the DP does not fully determine the migratory pattern (QB). However, some threshold values can be determined for the baseline model. Cells collectively migrate when the average $DP > 0.71$, all cells migrate when $DP > 0.63$ and at least one cluster migrates when $DP > 0.06$.

There are several reasons for the overlap in DP scores. For example, the initial population might split up in a small non-migrating cluster and a large migrating cluster (so $QB = 0$). In this case, the weighted average DP of both clusters will still be relatively high even though the migration is considered unsuccessful. Alternatively, the initial cell population might migrate entirely to the right boundary ($QB = 2$) only after first migrating to the left and clinging to the left boundary, which produces a relatively low DP score.

We conclude that the baseline model succeeds to some extent in reproducing collective cell migration. However, a disadvantage is the large percentage of runs in which cells do not migrate at all, but instead cling to the left boundary. Also, the release factor γ does not yet have a significant effect in this model.

5.3 Results of Model 2: variable release

In this model, we compare the results for a fixed and variable release term. In the variable case, the release term γ_{var} increases to a constant value $\gamma \in \{0, 0.2, ..., 1\}$ according to Eq. (11). The off-set is increased to $x = 30$ to reduce the tendency of cells to cling to the left boundary. We compare the results

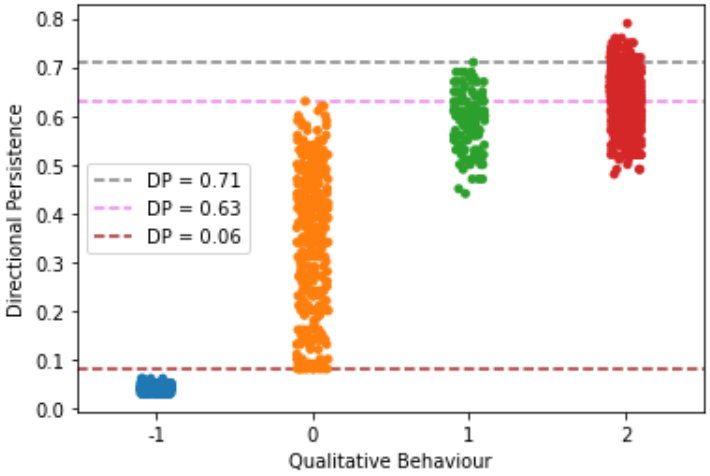
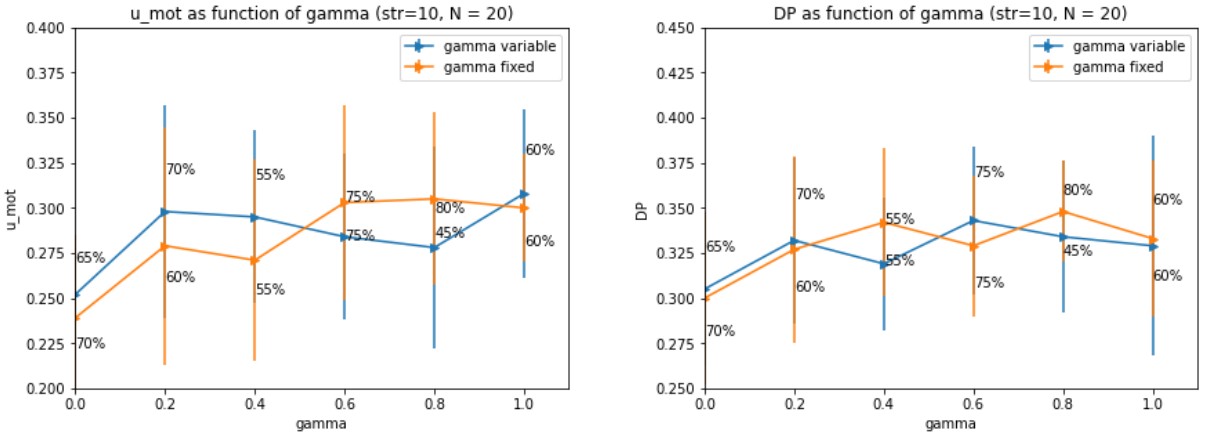


Figure 17: Scatter plot of average DP and QB of all baseline simulations (30 iterations of parameter pairings, 1800 runs in total). Dashes lines indicate threshold values (see text).

for a fixed and variable release term with $\gamma \in \{0, 0.2, ..., 1\}$ and fixed chemotactic strength $\mu = 10$. The results are presented in Figure 18 (20 iterations for each value of γ, both in the variable and the fixed case). Only the successful simulations are used to calculate the average results.



(a) Cell motility

(b) DP

Figure 18: Average cell motility and DP for variable (blue) and fixed (orange) release term γ (20 iterations). Only the successful runs are used for calculating the averages. The percentages denote the success rates.

We observe that both cell motility (Figure 18a) and directional persistence (Figure 18b) are about 10% higher when $\gamma > 0$ as opposed to when $\gamma = 0$ - both for a variable and fixed release term. Cells migrate and faster and more efficiently when there is an additional release of TGF-β. Cell motility is lower for runs with a fixed release term when $\gamma < 0.5$, but higher when $0.5 < \gamma < 1$. There is no clear difference in either the DP or the percentage of valid runs between the cases of fixed and variable release term (Figure 18b).

The difference in cell motility might be partly explained as follows. For higher values of γ, the fixed release term causes a significant addition of TGF-β starting right after the time of differentiation. The increase of TGF-β then increases cell motility. By contrast, when the release term is variable (i.e. it take time for the release term to increase to the value γ), less additional TGF-β is released, so cell motility is not increased as much. Therefore cell motility is higher for a fixed release term for high values of γ.

However, given that only the successful runs are used to calculate the average results in Figure 18, more iterations are needed to confirm these results.

We conclude that using a positive release term increases cell motility and DP, even when the release term is a function of time - which is more in line with biological reality.

5.4 Results of Model 3: cell shape constraints

5.4.1 Adding a length constraint

We first investigate the effect of adding a length constraint to the Fbs (while maintaining the volume constraint with $A = 40$). Setting the release term $\gamma = 0$, fixing the chemotactic strength μ and excluding cell differentiation in the model, we compare the cell motility of a cluster of 11 cells both with and without length constraint (target length $L_{\tau=1} = 20, \lambda_{L,\tau=1} = 1$). All cells are initialized in a 50×50 block with off-set $x = 20$. The results are listed in Table 2.

	Cell motility	DP	% Success
Constraints of Fb ($\mu = 50$)	0.88	0.62	63 %
No constraint on Fb ($\mu = 50$)	0.98	0.58	47 %
Constraint on Fb ($\mu = 80$)	1.18	0.69	70 %
No constraint on Fb ($\mu = 80$)	1.46	0.63	63 %

Table 2: Average cell motility with and without length constraint on Fbs (30 iterations). Chemotactic strength is fixed at $\mu = 50$ (top) and $\mu = 80$ (bottom).

The results consistently show that cells with a length constraint have a lower motility and a higher percentage of valid runs (DP is also somewhat higher). We consider two possible explanations:

1. *Lower motility of elongated cells*
   Both elongated and non-elongated cells are on average more oriented in parallel to the gradient than perpendicular to it. In other words: the major axis of the elliptical approximation of both cell types points in the direction of the gradient more often than not.

   This can be verified in Morpheus by outputting the cell orientation ϕ, which is the angle between the x-axis and the major axis of the elliptical approximation of the cell. We introduce an *orientation score* (OS) such that $OS = 1$ if $\phi \mod \pi < \frac{\pi}{4}$ or $\phi \mod \pi > \frac{3\pi}{4}$ and $OS = 0$ otherwise. The OS therefore measures the extent to which cells are oriented along the x-axis, which runs parallel to the gradient in this scenario. Calculating the average OS, we find that $OS = 0.7$ for elongated cells and $OS = 0.67$ for non-elongated cells ($\mu = 50$ over 30 iterations). Thus, both cell types are on average more oriented horizontally than vertically.

   Now since cell motility is determined by chemotaxis (Eq. (7)), the energy change $\Delta\mathcal{H}$ depends on local concentration differences of the chemo-attractant. Since its concentration decays in a not (yet) linear fashion over the domain (see e.g. Figure 11), pixels in the rear of the cell sense a smaller concentration difference than pixels in the front. For pixels in the rear $\Delta\mathcal{H}$ is low, therefore, movement by chemotaxis is only favoured slightly compared to pixels in the front. The pixels in the rear therefore slow down the total cell movement. This effect is more significant when the distance between front and rear increases, i.e. for elongated cells. By contrast, cells with no length constraint have a more circular shape, therefore this impeding effect is less strong, which explains their higher motility.

2. *Higher success rate of elongated cells*
   The higher success rate for cells with a length constraint is explained by the fact that a higher percentage of the runs migrates and does not cling to the left domain boundary (13% of cells with length constraint cling to the boundary as opposed to 37% of cells without the constraint, for $\mu = 50$). Cells cling to the left boundary because this reduces the energy penalty for interfaces. This behaviour can therefore be regarded as an artefact of the CPM. Here we offer a tentative explanation by arguing that the cell shape determines the cellular behaviour. When a pixel of a

more circular cell touches the left boundary, the probability that another pixel in the same cell will also touch the boundary will be higher than when the cell is elongated. Non-elongated cells therefore reduce energy more readily by clinging to the boundary than elongated cells.

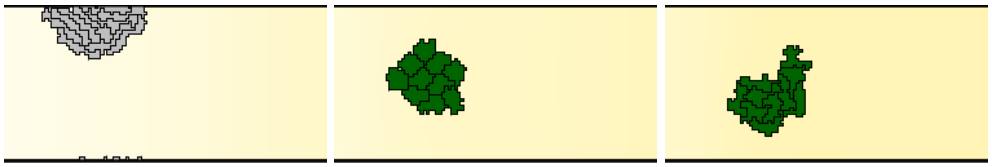
5.4.2 Model 3: two shape constraints

We now model cell behaviour with length and perimeter constraints on the Fbs and Mfs (the model settings are described in section 2.4.3). The release factor is a function of time as in Eq. (11). All CPM interaction energies are set to 1. Below we consider three different cases.

5.4.2.1 Case I: equal conversion rates Recall that in the model Eqs. (8)-(9) the terms α and β denote the standard conversion rates of the two forms of the chemo-attractant. The term α indicates the conversion rate of the latent form to the active form ($b \to u$) and β vice versa ($u \to b$). We first consider the case of equal conversion rates $\alpha = \beta = 5 \times 10^{-3}$. This implies that differences in concentration, e.g. caused by differences in initial conditions, will even out over time across the domain.

We run a preliminary simulation with parameter pairs (γ, μ) with release term $\gamma \in \{0, 0.2, .., 1\}$ and chemotactic strength $\mu \in \{10, 20, .., 100\}$. It turns out that for $\mu > 10$, the cells either do not migrate or the cells differentiate very late (or only after cells have already migrated to the right boundary). For these values, the role of the release term during migration cannot be investigated. In the following we therefore fix $\mu = 10$.

For $\mu = 10$ and $\gamma \in \{0, 0.2, .., 0.8, 1\}$, most cells migrate and differentiate mid way. During differentiation into Mfs, the cells becoming increasingly 'star-shaped', as shown in Figure 19. Most cells slow down slightly during differentiation and eventually migrate towards the right boundary. Figure 20 shows an example.

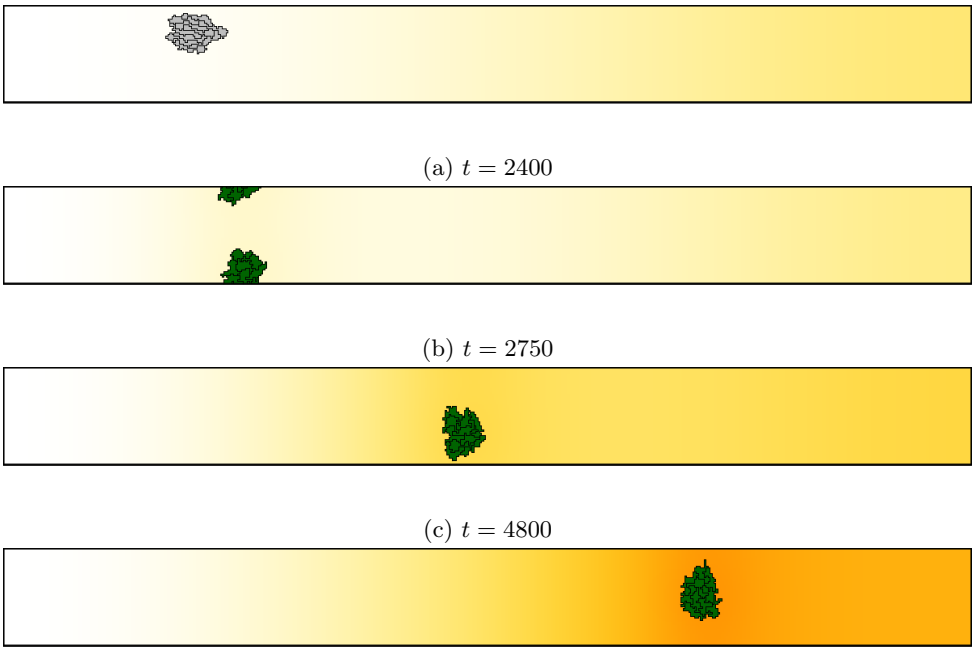


(a) $t = 2400$ (b) $t = 2500$ (c) $t = 2600$

Figure 19: Cell cluster before (a) and during (b)-(c) differentiation of Fbs (gray) into Mfs (green). Differentiation starts at $t = 2450$.

The results for this scenario are listed in Table 3 (10 iterations). We observe that the release of TGF-β indeed influences cell motility: the average cell motility almost doubles for $\gamma = 0.2$ compared to the case $\gamma = 0$. However, increasing the release term beyond $\gamma > 0.2$ does not increase cell motility consistently. In other words, a higher cell motility is not necessarily associated with an lower end time of migration, since the migration can be less efficient (as measured by the DP, see also Figure 21), as is evident from comparing the case $\gamma = 0.2$ with $\gamma = 0.8$. The value $\gamma = 0.4$ seems to represent an optimum for cell migration when all measures are taken into account. By contrast, a high release term such as $\gamma = 1$ yields a high average cell motility, while only a small percentage of the runs are successful. This suggests a trade-off, governed by the release term γ, between cell motility on the one hand and efficiency and migratory success on the other.

In theory, cells might stop migrating because they cannot escape the local maximum of chemo-attractant released by Mfs mid way. We turn to an explanation why cell clusters still migrate successfully (even for higher values of γ). Figure 22 shows the concentrations of both active and latent form of the chemo-attractant form for various times. We first note that irrespective of the choice of $\gamma \in \{0, 0.2, .., 1\}$, the difference between the concentrations of active and latent form are relatively small: both concentrations remain within the same order of magnitude. This is the case even though the 'wound' only provides TGF-β in the bound form, creating a big difference in concentration at the right end of the domain in the early stage of the simulation (see e.g. the curve for $t = 50$). As time progresses, however, these differences even out across the domain due to the equal choice of conversion rates α and β.



(a) $t = 2400$

(b) $t = 2750$

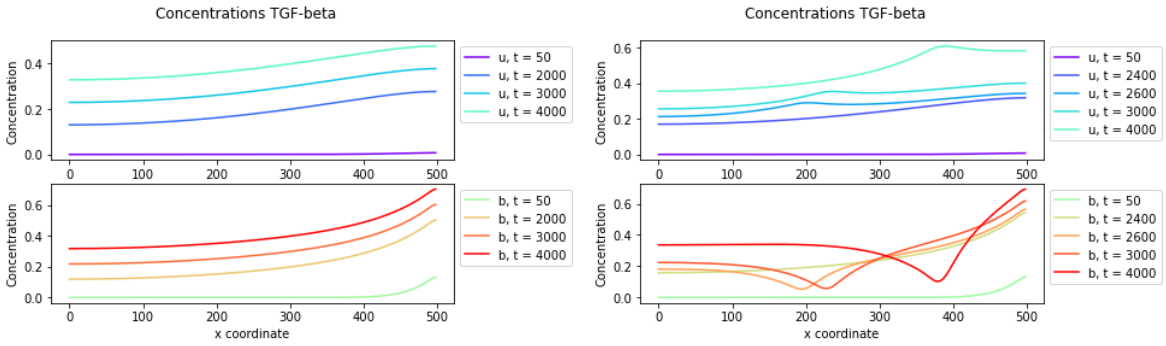
(c) $t = 4800$

(d) $t = 6100$

Figure 20: Model 3: successful migration and differentiation ($\mu = 10, \gamma = 0.8$). Differentiation here starts at $t = 2575$.

Results Model 3 - equal conversion rates					
γ	cell motility	DP	t 75% Mfs	t_{end} migration	% Valid
0	0.07	0.23	2283	5350	30
0.2	0.12	0.26	2338	4950	40
0.4	0.12	0.31	2134	3938	80
0.6	0.12	0.28	2279	4654	70
0.8	0.11	0.28	2275	4769	40
1	0.13	0.28	2275	4305	50

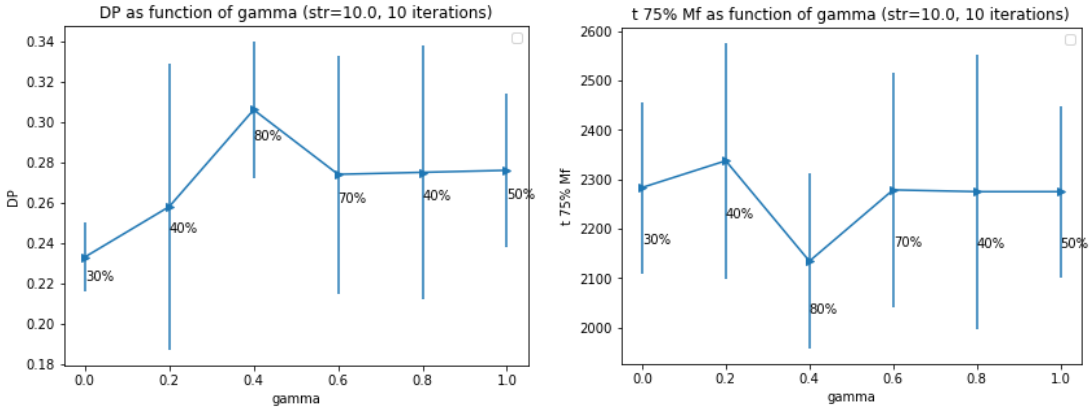
Table 3: Average measures for $\mu = 10, \gamma \in \{0, 0.2, ..., 1\}$. Based on 10 iterations. The averages are calculated based on the successful runs only.



(a) $\gamma = 0$

(b) $\gamma = 1$

Figure 22: Concentrations TGF-β in active form u (above) and latent form b (below) for a successful run with $\gamma = 0$ (a) and $\gamma = 1$ (b) at various times.



(a) Average DP

(b) Average time of $\geq 75\%$ Mfs

Figure 21: Average DP (a) and time of $\geq 75\%$ differentiation (b) for variable release term γ (10 iterations). The averages are calculated based on the successful runs only. Release value $\gamma = 0.4$ represents an optimum.

For $\gamma = 1$, Figure 22b shows an increase in concentration of active form u (and concomitant drop in the latent form b) after $t = 2425$ when differentiation has started. This increase in concentration, caused by the release of the active chemo-attractant by developing Mfs, constitutes a local (but not a global) maximum for $t \leq 3675$ and a global maximum for $t > 3675$. For other values of $\gamma \in \{0.2, 0.4, ..., 1\}$ the pattern is similar: during the migration phase, the local maximum of the active form of the chemo-attractant (released by the Mfs) becomes the global maximum.

The local maximum of the active form attracts the cell cluster. Despite this attraction to the local maximum, the cells still complete a full migration. This is explained by the fact that in the neighborhood of the local maximum, the concentration curve of the active form is slanted, with higher concentrations at the right hand side of the local maximum. Figure 23 shows the concentrations and the minimal and maximal cell center positions for $t = 4000$. Since the Mfs convert the latent form of TGF-β into the active form, the minimum concentration of the latent form is located within the range of the minimal and maximal cell center position. However, the x-position of the local maximum of the active form is located on the right-hand side of this range. The maximum attracts the cells, resulting in chemotaxis towards the right, which explains the complete migration of the cells. Indeed, the concentration of active form is caused by both the release by Mfs and the base rate conversion of the diffused latent form, which has a higher concentration on the right hand side due to its proximity to the wound.

To summarize: given equal conversion rates, the difference in the concentrations of latent and active form become small over time. Consequently, when $0 < \gamma < 1$, the increase of the active form caused by Mfs is also small. Nonetheless, a positive release term increases cell motility substantially. There exists an optimal value for the release term (here: $\gamma = 0.4$), for which the the efficiency of migration is maximal and the starting time of differentiation and the end time of migration are minimal.

5.4.2.2 Case II: unequal conversion rates In this case, we assume that the quantity of latent growth factor stored in the ECM is substantially higher than the active form in the cellular environment. This assumption is reflected in the choice of conversion parameters: $\alpha = 5 \times 10^{-3}$ and $\beta = 3 \times 10^{-2}$. Due to the difference in conversion parameters, the concentration of the active form of TGF-β is relatively low. To compensate, the threshold for differentiation in this model is set to $ths = 0.1$.

We find that for $0.1 \leq \gamma < 0.2$ the cells migrate, while for $\gamma \geq 0.2$ the cells hardly migrate: the cells then differentiate and migration stops almost directly. The concentrations for the cases $\gamma = 0$ and $\gamma = 0.2$ are presented in Figure 24. We first note that, due to the unequal conversion terms, the concentration of the latent form b exceeds the concentration of the active form u by almost a factor 5 before the Fbs start differentiating (Figure 24). Due to this large difference in concentrations, even a small positive release term results in a large increase in the active form of the chemo-attractant after differentiation. Therefore, a global maximum of the active form is created shortly after differentiation. As is visible in Figure 24b,

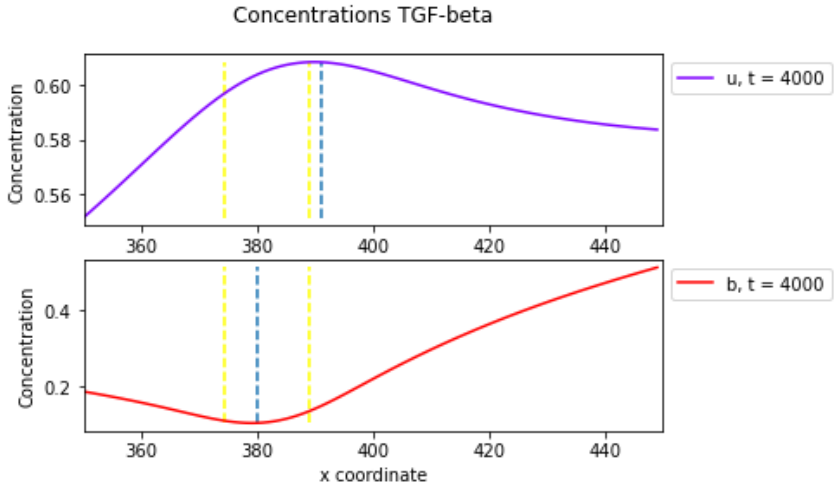
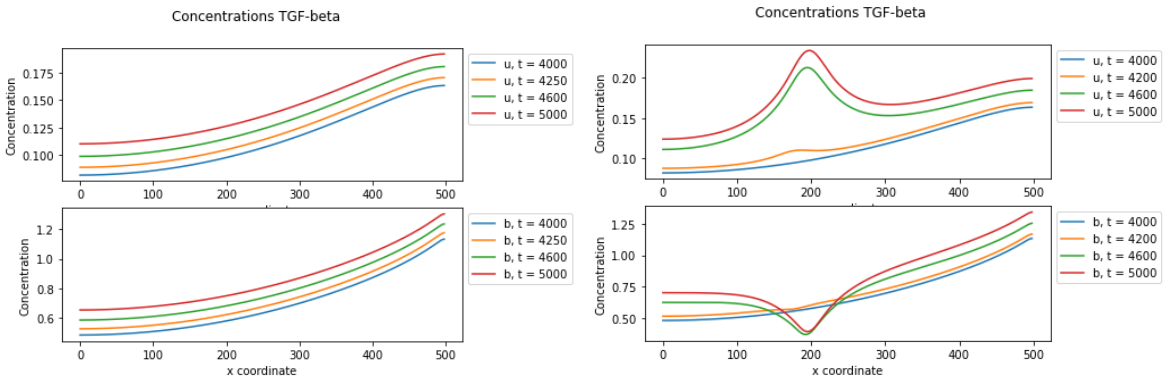


Figure 23: Slanted concentration curve around a global maximum (for u) and the global minimum (for b), indicated by the blue dashed lines for $t = 4000$ ($\gamma = 1$). The yellow dashed lines indicate the minimal and maximal x-position of the cell centers at $t = 4000$.

this global maximum is steep and symmetrical, making it impossible for cells to escape it. The location of the maximum therefore hardly changes in time. Cells are attracted to the global maximum of TGF-β that they have released themselves, and therefore migration stops.



(a) $\gamma = 0$ (b) $\gamma = 0.2$

Figure 24: Concentrations TGF-β in active form u (above) and latent form b (below) for with $\gamma = 0$ (a) and $\gamma = 0.2$ (b) at various times. Differentiation starts at $t = 4200$ in both cases.

5.4.2.3 Case III: limited source of TGF-β In the previous models, the wound at the right boundary secretes the latent form of TGF-β and is never depleted. This is a simplification of biological reality, since TGF-β is produced by cells in the wound area such as macrophages (in the inflammatory and early proliferation phase of wound healing) and the supply is therefore limited. This suggests an alternative function for the latent TGF-β and its release by Mfs. The storage of TGF-β might function as a natural 'back-up' for when the secretion of TGF-β from the wound is impeded. In this case, the additional release of TGF-β by Mfs compensates for this reduction and ensures that the processes of cell migration and differentiation continue successfully.

If this is the case, and if the supply of TGF-β from the wound is limited, simulations which would be otherwise unsuccessful would become successful when adding the release term. We test this hypothesis by limiting the period in which TGF-β is supplied from the source to $t_d \cdot t_{end}$, where t_{end} is the end time of the simulation and $t_d \in \{0.5, 0.6, ..., 1\}$. The threshold for differentiation is set to $ths = 0.2$, with

$t_{end} = 6000$ and equal conversion rates $\alpha = \beta = 5 \times 10^{-3}$.

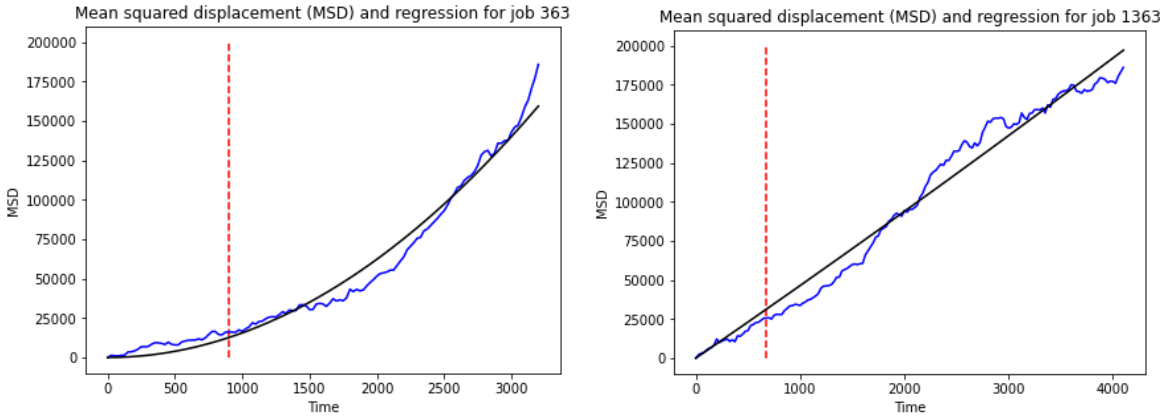
Results Model 3 - limited source						
γ/t_d	0.5	0.6	0.7	0.8	0.9	1
0	0	10	30	25	30	15
0.2	10	30	35	35	30	55
0.4	25	30	30	20	55	55
0.6	15	35	50	50	70	50
0.8	35	20	40	55	30	60
1	30	40	30	55	50	50

Table 4: Percentages of successful runs with limited source term (20 iterations).

The results are tabulated in Table 4 (20 iterations). Note that if the source is only active for 50% of the simulation time, none of the runs are successful (cells never reach the right boundary). Therefore, if the additional release term functions as a back-up for a hampering secretion of TGF-β from the wound, we expect an increase of the success rate when $\gamma > 0$ compared to the case $\gamma = 0$. The results in Table 4 indicate that for all values of t_d, the optimal value of γ increases the success rate of migration by at least a factor $\frac{2}{3}$ compared to the case $\gamma = 0$ (see e.g. the case $t_d = 0.7$). Moreover, this factor increases for higher values of γ. Although perfect success rates are not attained, the results indicate that the additional release can contribute substantially to compensate for a hampering secretion of TGF-β from the wound.

The results of Model 3 with unlimited secretion of TGF-β (5.4.2.1) suggested an optimum for the release term γ. In the current case of limited secretion, we find that for the values $t_d \in \{0.5, 0.7, 0.9, 1\}$ the success rate is highest when $\gamma = 0.6$ or $\gamma = 0.8$. Here, again, the optimum value of γ is not the maximal value. However, when $t_d = 0.6$ and $t_d = 08$ (the wound only secretes TGF-β for 60% and 80% of the simulation time), the highest success rate is achieved when the release term is also the largest ($\gamma = 1$).

Some caution is needed in analysing the cell motility in the model with limited secretion. On some occasions, the cells make short 'retrograde' motions instead of moving smoothly towards the right boundary (here in 15 of 720 simulations). In these cases, the concentration of TGF-β on the left side of the cell cluster is temporarily higher. This is caused by the fact that the wound on the right boundary has stopped secreting TGF-β and the Mfs are releasing latent TGF-β while migrating towards the wound. This causes the regression curve to become linear instead of quadratic. Since the quadratic term of the regression is used to approximate cell motility, the approximation of cell motility then decreases to zero. This is illustrated by Figure 25, which shows the MSD plots for successful runs with and without retrograde motion. In both cases, the MSD drops under the regression curve after cells have started differentiating. When the additional TGF-β caused cells to move backwards, the regression curve becomes linear (Figure 25b). When such retrograde motions occur, cell migration can no longer be understood as simply a advection-diffusion process. The quadratic regression then no longer provides an adequate estimate of cell motility.



(a) MSD pattern ($\gamma = 0.6, t_d = 0.9$)

(b) MSD pattern ($\gamma = 0.6, t_d = 0.8$)

Figure 25: MSD plots for successful migrations without (left) and with (right) short retrograde motion. The red dashed line indicates the time when $\geq 75\%$ of the cells have started differentiating into Mfs.

6 Conclusions and Discussions

The first aim of this project is to reproduce the process of migration and differentiation of fibroblasts (Fbs) during wound healing. The baseline model (Model 1) reproduces the natural behaviour of migration and differentiation of Fbs to some extent. The baseline displays some expected properties such as an increasing cell motility with increasing chemotactic strength. The baseline also shows that there is no simple mapping between directional persistence (DP) and the migration pattern of the simulation (QB). This is due to the stringent criterion used for determining the success of a simulation (one non-migrating cell makes the entire migration unsuccessful). Furthermore, the tendency of cells to cling to the left boundary (hence not migrating) is an undesirable artifact of the CPM. This tendency can be prevented by increasing the off-set and shrinking the subdomain of initialization. Alternatively, a more realistic model might be obtained by expanding the domain to a rectangular grid, imposing periodic boundary conditions on both axes and locating the wound area in the middle (see e.g. [2]).

In this project, Model 2 demonstrated for the first time that the additional release of TGF-β caused by myofibroblasts (Mfs) can have a significant effect on the dynamics of migration. Fixing chemotactic strength at $\mu = 10$, Model 2 investigated the migration patterns for both a fixed and a variable release term, where the variable release term models the gradual development of Fbs into Mfs. In both cases, a positive additional release term increased cell motility and DP with about 10%. One might expect that a variable release term is associated with a lower increase in cell motility compared to a fixed release term, since the positive effect of the additional release of TGF-β on cell motility needs time to develop. However, no clear differences between the effects of the fixed and variable release term on cell motility and DP were observed. It might be related to the fact that out of 20 iterations only the successful runs were used to calculate these measures, making the sample size too small for these differences to become manifest.

The geometrical constraints in the CPM succeed reasonably well in mimicking the shapes of Fbs and Mfs (Model 3). Imposing a length constraint on the Fbs has the effect of decreasing cell motility and increasing the number of successful runs. The lower cell motility of elongated Fbs was explained by the 'lagging' rear end of the cell. This can therefore be considered as an effect of cell geometry. However, this explanation needs to be verified quantitatively. Moreover, by varying target length L, further simulations can reveal in detail how cell motility depends on cell elongation. An interesting question is whether this relation between motility and elongation corresponds to the biological reality of cell movement. If this is not the case, a compensation of this effect would be a relevant addition to the CPM.

When length and perimeter constraints are imposed on the Fbs and Mfs respectively (Model 3) and the conversion base rates α and β are equal, there is an optimal release value for which migration is most efficient and most successful. The results seem to indicate a trade-off between cell motility on the one hand and efficiency and success rate on the other. The results need to be replicated with a larger sample

size to corroborate this finding (cf. the error bars in Figure 22).

However, when the conversion base rates α and β differ by an order of magnitude (and when $\gamma \in [0.2, 0.4, ..1]$), migration stops. The Mfs then create a global maximum of the chemo-attractant from which cells cannot escape. This happens also for small values of the release term γ, which suggests that the difference in the concentrations of active and latent form is the relevant cause for this phenomenon. This result gives rise to the hypothesis that in the biological context, both base rates of conversion will have the same order of magnitude - a hypothesis that might be tested experimentally. The result also underlines the importance of empirically establishing the (ratio of) concentrations in the wound area for future modelling endeavours.

Both in the models with and without perimeter and length constraints, migrating cells have a higher cell motility and DP when the release term $\gamma > 0$ (Models 2 and 3). This can be explained by the fact that a local maximum of the chemo-attractant is usually created in front of the migrating cell cluster. In our model, this is caused by the fact that the concentration of the bound TGF-β, which models the TGF-β stored in the large latent complex (LLC), is higher in the vicinity of the wound area. In other words: the quantity of latent TGF-β is not constant in the domain, but instead forms a gradient (just as the active form does). This is a model assumption that needs empirical confirmation if the results are to be interpreted as representing biological reality.

Further simulations with Model 3 show that the additional release of TGF-β can increase the success rate of migration by at least a factor $\frac{2}{3}$ when the secretion of TGF-β is limited. This corroborates the idea that the additional release could function as a biological 'back-up' for when the secretion of TGF-β by macrophages is somehow impeded. Comparing the results of Model 3, more specifically of case 1 and case 3, we find that there usually exists an optimal value of the release term γ which is *not* the maximal value. However, this is not the case for $t_d = 0.6$ and $t_d = 0.8$, and further research needs to elucidate why these cases constitute exceptions. For the scenario of limited secretion, cell motility has not been analysed in this project. In this scenario cells sometimes move in short retrograde motions, which makes the estimate of cell motility by the quadratic regression of the MSD curve inadequate. Other measures of cell motility will have to be used to analyse cell motility in this scenario.

From a biological perspective, a shortcoming of the current model is that it does not distinguish the cell medium and the ECM. More specifically, it does not distinguish the medium and the LLC which supposedly stores TGF-β. In the current model, the latent TGF-β is modelled as a separate chemical diffusing through the domain. However, following the hypothesis of Wipff *et al.*, the latent TGF-β is stored in the LLC *attached* to the (myo)fibroblast. This could be incorporated in the CPM by defining the LLC as a new cell type, both for Fbs and for Mfs, and choosing the adhesive energies such that the 'mixing'-condition of the CPM is satisfied, i.e. $J_{LLC-cell} < J_{LLC-LLC} = J_{cell-cell}$ [12]. The LLCs then cling to their corresponding (myo)fibroblasts. The next step is to model explicitly how the LLC stores the ambient TGF-β, in other words, how the active form is reduced to the latent form in the LLC.

Furthermore, the current model does not take into account the ECM stiffness, although research suggests that ECM stiffness regulates the release of TGF-β. Since forces are not modelled explicitly in the CPM, the incorporation of stiffness is not straightforward. One might implement the ECM, LLC, Fbs and Mfs as different cell types in the CPM linked via appropriate adhesive energies. Elastic or compliant ECMs impede the release of TGF-β. The easiest way to model this is to define an elasticity parameter E for the ECM and consequently make the release term γ_{var} of a cell inversely proportional to the elasticity parameter of the corresponding ECM.

The realism of the model can be improved by adding more model components. These include, for example, the production of TGF-β by macrophages for a limited period, the production of TGF-β by both Fbs and Mfs, the de-differentiation of Mfs into Fbs in the absence of TGF-β and the apoptosis of Mfs in the wound area. Above all, the model can be improved by using empirically confirmed parameters (such as chemotactic strength).

Some improvements can also be made to the CPM. For instance, the approximation of the cell perimeter can be improved by using the circumference of the elliptical approximation of the cell, since this approximation is already calculated in the CPM to determine the cell length. Although no elementary formula exists, the circumference of the ellipse P can be approximated by e.g. Ramanujan's formula: $P \approx \pi[3(a+b]) - \sqrt{(3a+b)(a+3b)}]$, with a and b the semi-major and semi-minor axis of the elliptical approximation.

References

[1] J. Starruss, W. de Back, L. Brusch and A. Deutsch, *Morpheus: a user-friendly modeling environment for multiscale multicellular systems biology*, Bioinformatics, 30:9: 1331-1332, 2014.

[2] Q. Peng, *Mathematical Aspects of Cell-Based and Agent-Based Modelling for Skin Contraction after Deep Tissue Injury*, PhD-thesis at Delft University of Technology, 2021.

[3] S. Enoch, D.J. Leaper, *Basic science of wound healing*, Surgery (Oxford), 26:31, 2008.

[4] Koppenol, D., *Biomedical implications from mathematical models for the simulation of dermal wound healing*, PhD-thesis at Delft University of Technology, 2017.

[5] J. Tomasek, G. Gabbiani, B. Hinz, C. Chaponnier C and R. Brown, *Myofibroblasts and mechano-regulation of connective tissue remodelling*, Nature Reviews Molecular Cell Biology, 3:349–363, 2002.

[6] P-J. Wipff, D.B. Rifkin, J-J. Meister and B. Hinz, *Myofibroblast contraction activates latent TGF-β1 from the extracellular matrix*, Journal of Cell Biology, 179:6: 1311-1323, 2007.

[7] A. Desmoulière, M. Redard, I. Darby and G. Gabbiani, *Apoptosis mediates the decrease in cellularity during the transition between granulation tissue and scar*, American Journal of Pathology, 146:56–66, 1995.

[8] A. Postlethwaite, J. Keski-Oja, H.L. Moses, and H.A. Kang, *Stimulation of the chemotactic migration of human fibroblasts by transforming growth factor β*, Journal of Experimental Medicine, 165: 251-256, 1987.

[9] A. Desmoulière, A. Geinoz, F. Gabbiani and G. Gabbiani, *Transforming growth factor-β1 induces α-smooth muscle actin expression in granulation tissue myofibroblasts and in quiescent and growing cultured fibroblasts*, Journal of Cell Biology, 122:103–111, 1993.

[10] J.P. Annes, J.S. Munger and D.B. Rifkin, *Making sense of latent TGF-beta activation*, Journal of Cell Science, 116:217–224, 2003.

[11] R. Magno, V.A. Grieneisen, A.F. Marée, *The biophysical nature of cells: potential cell behaviours revealed by analytical and computational studies of cell surface mechanics*, BMC Biophysics, 8:8, 2005.

[12] F. Graner and J. Glazier, *Simulation of Biological Cell Sorting Using a Two-Dimensional Extended Potts Model*, Physical Review Letters, 69:13, 1992.

[13] R.M.H. Merks, S.V. Brodsky, M. S. Goligorksy, S. A. Newman and J. A. Glazier, *Cell elongation is key to in silico replication of in vitro vasculogenesis and subsequent remodeling*, Developmental biology 289:1: 44-54, 2006.

[14] N.J. Savill and P. Hogeweg, *Modelling morphogenesis: from single cells to crawling slugs*, Journal of Theoretical Biology, 184:229–235, 1997.

[15] E. A. Codling, M. Plank and S. Benhamou, *Random walk models in biology*, Journal of the Royal Society Interface, 5: 813-834, 2008.

[16] N. B. Ouchi, J. A. Glazier, J-P Rieu, A. Upadhyayad and Y. Sawadae, *Improving the realism of the cellular Potts model in simulations of biological cells*, Physica A, 329:451–458, 2003.

[17] M. Zajac, G.L. Jones and J.A. Glazier, *Simulating convergent extension by way of anisotropic differential adhesion*, Journal of Theoretical Biology 222; 247-259, 2003.

7 Appendices

7.1 Appendix A: shape constraints

This Appendix first explains and illustrates the length and surface constraints added to the main Hamiltonian in Eq. (3). Then, the elliptical approximation of cell length is explained in some detail.

The *length constraint* is given by:

$$\lambda_{L,\tau_\sigma} \sum_{\sigma} (l_\sigma - L_{\tau_\sigma})^2. \tag{19}$$

Here l_σ denotes the estimated length of cell σ (see details below), L_{τ_σ} the target length and λ_{L,τ_σ} the strength of the constraint. Deviations from the target length incur an positive energy penalty.

The *surface constraint* is in 2D effectively a constraint on the cell perimeter. The perimeter of cell k is estimated as:

$$p_{\sigma=k} = \sum_{\mathbf{x}:\sigma(\mathbf{x})=k} \sum_{\mathbf{x}' \in \mathcal{N}_1(\mathbf{x})} (1 - \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x}')}), \tag{20}$$

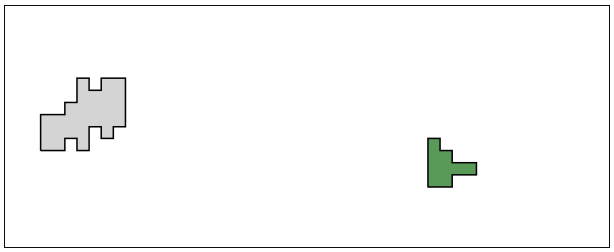
where now $\mathcal{N}_1(\mathbf{x})$ denotes the first-order neighborhood of $\mathbf{x}$. The perimeter constraint reads:

$$\lambda_{P,\tau_\sigma} \sum_{\sigma} (p_\sigma - P_{\tau_\sigma})^2. \tag{21}$$

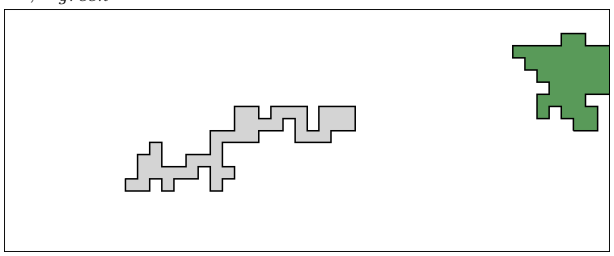
where P_{τ_σ} is the target perimeter and λ_{P,τ_σ} the constraint strength [16].

The effect of these geometrical constraints are illustrated in Figures 26a-c. Two cells of different cell types (coloured gray and green) are displayed on a 50×20 grid. The cells are both subject to the connectivity constraint, but differ in their geometrical constraints (all multipliers λ have been set to 1):

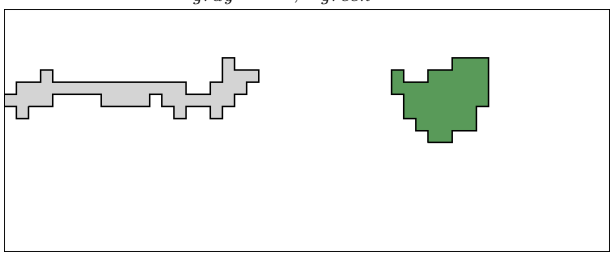
1. Figure 26a shows the gray and green cell with an area constraint having target area $A_{gray} = 30$ and $A_{green} = 10$ (no length or perimeter constraints).
2. In Figure 26b, the cells are subject to an identical area constraint ($A_{green} = A_{gray} = 40$), with different length constraints, with target length $L_{gray} = 30$ and $L_{green} = 10$ (and no perimeter constraints).
3. In Figure 26c, the cells are subject to an identical area constraint ($A_{green} = A_{gray} = 40$), with different perimeter constraints, with target perimeter $P_{gray} = 50$ and $P_{green} = 20$ (and no length constraints).



(a) Two cells only have area constraint of $A_{gray} = 30, A_{green} = 10$.



(b) Area constraint with $A_{green} = A_{gray} = 40$, length constraint with $L_{gray} = 30, L_{green} = 10$.



(c) Area constraint with $A_{green} = A_{gray} = 40$, perimeter constraint with $P_{gray} = 50, P_{green} = 20$.

Figure 26: Different constraints on two cells (colored in gray and green) to illustrate the constraints on cell geometry.

Cell perimeter $p(a)$ increases as the cell area a increases. The order of the increase depends on the length constraint. When a cell length is maximized (the cell is maximally stretched), the cell perimeter depends linearly on the cell area, so $p(a) = \mathcal{O}(a)$. However, when cell length is minimized (the cell is squeezed into a rectangle), $p(a) = \mathcal{O}(\sqrt{a})$. The latter case is more adequate, since biologically realistic scenarios use a length constraint.

7.1.1 Ellipsoid approximation of cell length

Biological cells naturally have an ellipsoid shape. However, the lattice grid imposes an artificial shape on the cells. Therefore a transformation is required to switch from the axis of lattice coordinates to the symmetry axes of ellipsoid cells.

Following Zajac [17], the long and short axes of the ellipses are calculated using the inertia tensor. The elongation of the cell is estimated by the moments of inertia around the center of mass. The centre of mass is calculated by the averages of the cells positions in both coordinate axes of the lattice and is therefore given by $\bar{\mathbf{x}} = (\bar{x}, \bar{y})$. The moments of inertia are given by:

$$
\begin{aligned}
I_{xx} &= \sum_{i=1}^{N}(y_i - \bar{y})^2, \\
I_{xy} &= -\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}), \\
I_{yx} &= I_{xy}, \\
I_{yy} &= \sum_{i=1}^{N}(x_i - \bar{x})^2.
\end{aligned} \tag{22}
$$

where the sum ranges over all the N lattice sites $\mathbf{x} = (x_i, y_i)$ occupied by the cell. The inertia tensor now reads:

$$
\begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} \tag{23}
$$

The moments of inertia give the orientation and length of the semi-major and semi-minor axis of the elliptical approximation.

Since the inertia tensor (in 2D) can be regarded as a symmetric matrix, it has a pair of real eigenvalues $\{\lambda_a, \lambda_b\}$ and there exists an orthogonal basis of eigenvectors. Diagonalization of the symmetric matrix corresponds to transforming the axes of the (x, y)-coordinates of the lattice to a new orthonormal coordinate system (c_1, c_2), where the new axes are aligned with the eigenvectors. By the Principle Axis Theorem, ellipses in the new coordinate system are now described by:

$$
\lambda_a c_1^2 + \lambda_b c_2^2 = 1. \tag{24}
$$

When $\lambda_b > \lambda_a$, the distance from the origin of the ellipse attains its maximum when $c_2 = 0$. Solving equation for c_1 in (24) then gives: $c_1 = \frac{1}{\sqrt{\lambda_a}}$. Therefore this maximal length corresponding to the *smaller* eigenvalue equals the length of the semi*major* axis of the ellipse. Analogously, setting $c_1 = 0$ gives the length of the semi*minor* axis, which equals $\frac{1}{\sqrt{\lambda_b}}$.

Calculating the eigenvectors and eigenvalues of the inertial tensor (23) yields:

$$
\lambda_b = \frac{1}{2}(I_{xx} + I_{yy}) + \frac{1}{2}\sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}, \tag{25}
$$

$$
\mathbf{b} = \hat{\mathbf{x}} I_{xy} + \hat{\mathbf{y}}(\lambda_b - I_{xx}) \tag{26}
$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the unit vectors in the x- and y-direction of the lattice. Since λ_b is the larger eigenvalue, the corresponding vector $\mathbf{b}$ points along the semi-minor axis of the ellipse. The other solution of the characteristic equation λ_a is obtained by switching the sign of the radical in equation (25). Substitution in equation (26) yields the corresponding eigenvector $\mathbf{a}$ which points along the semi-major axis of the ellipse.

Example. We illustrate the elliptical approximation with an example. Consider the cell in Figure 27 comprising four blue lattice sites ($N = 4$). Coordinates of the centre of mass $C = (\overline{x}, \overline{y})$ are given by:

$$
\begin{aligned}
\overline{x} &= \frac{2 \cdot 2 + 2 \cdot 3}{4} = \frac{10}{4} \\
\overline{y} &= \frac{2 + 2 \cdot 3 + 4}{4} = 3
\end{aligned}
$$

Calculating the moments of inertia:

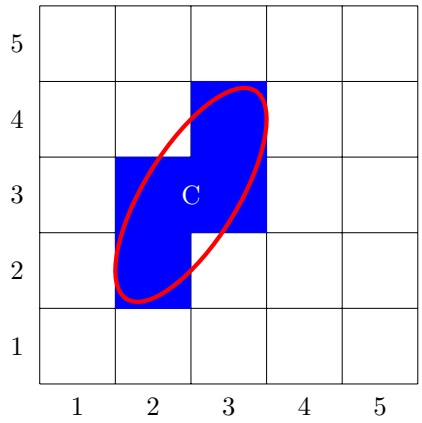


Figure 27: A cell (blue) and its elliptical approximation (red)

$$
\begin{aligned}
I_{xx} &= \sum_{i=1}^{N}(y_i-\overline{y})^2 = (2-3)^2+2(3-3)^2+(4-3)^2 = 2\\
I_{yx} &= I_{xy} = -\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})\\
&= -\Big[(2-\frac{10}{4})(2-3)+(2-\frac{10}{4})(3-3)+(3-\frac{10}{4})(3-3)+(3-\frac{10}{4})(4-3)\Big] = -1\\
I_{yy} &= \sum_{i=1}^{N}(x-\overline{x})^2 = 2(2-\frac{10}{4})^2+2(3-\frac{10}{4})^2 = 1.
\end{aligned}
$$

Therefore the inertia tensor is given by:

$$
\begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}
$$

Eigenvalues and eigenvectors:

$$
\begin{aligned}
\lambda_b &= \frac{1}{2}(I_{xx}+I_{yy}) + \frac{1}{2}\sqrt{(I_{xx}-I_{yy})^2+4I_{xy}^2},\\
&= \frac{1}{2}(2+1)+\frac{1}{2}\sqrt{(2-1)^2+4(-1)^2} = \frac{3+\sqrt{5}}{2}.\\
\lambda_a &= \frac{3-\sqrt{5}}{2}\\
\mathbf{b} &= \hat{\mathbf{x}}I_{xy} + \hat{\mathbf{y}}(\lambda_b - I_{xx}),\\
&= \begin{pmatrix}1\\0\end{pmatrix}\cdot -1 + \begin{pmatrix}0\\1\end{pmatrix}\left(\frac{3+\sqrt{5}}{2}-2\right) = \begin{pmatrix}-1\\ \frac{\sqrt{5}-1}{2}\end{pmatrix}\\
\mathbf{a} &= \begin{pmatrix}-1\\ \frac{-(1+\sqrt{5})}{2}\end{pmatrix}.
\end{aligned}
$$

The length of the semi-major and semi-minor axis of the ellipse are given by: $a = \frac{1}{\sqrt{\lambda_a}} = \sqrt{\frac{2}{3-\sqrt{5}}}$ and $b = \sqrt{\frac{2}{3+\sqrt{5}}}$. The angle of the semi-major axis with the y-axis of the grid equals $\tan^{-1}(\frac{1+\sqrt{5}}{2}) \approx 1.017$ rad. The elliptical approximation of the cell is the red ellipse drawn in Figure 27.

The cell perimeter p is estimated in Morpheus' CPM by counting the cell interfaces on the grid, so here $p_1 = 10$. Using instead Ramanujan's approximation of the circumference gives $p_2 \approx 11.61$.

7.2 Appendix B: connectivity constraint

The connectivity constraint implemented in the Morpheus package is the one proposed by Merks[13]. The algorithm can be described as follows. Let $\mathcal{N}_2$ be the 2-neighborhood of lattice site $\mathbf{x}$ which contains all the neighbours $\mathbf{x_i}'$ of $\mathbf{x}$, with $i = 1, 2, .., 8$. For a neighbour $\mathbf{x_i}'$ of $\mathbf{x}$, denote the next clockwise neighbour around $\mathbf{x}$ as $\mathbf{x}'_{i+1}$.

Now, when evaluating a copy of a spin σ into lattice site $\mathbf{x}$, count the number of times that the neighbours $\mathbf{x_i}'$ of $\mathbf{x}$ have a spin *identical* to the spin of $\mathbf{x}$, i.e. so $\sigma(\mathbf{x}'_i) = \sigma(\mathbf{x}')$, while the next or previous location clockwise around $\mathbf{x}$ has a spin *different* from $\mathbf{x}$, i.e. $\sigma(\mathbf{x}_{i+1}) \neq \sigma(\mathbf{x})$ or $\sigma(\mathbf{x}_{i-1}) \neq \sigma(\mathbf{x})$. Such an instance can be called a 'collision' of cells.

Formally, the number of collisions C is calculated by:

$$C(\mathbf{x}) = \sum_{\mathbf{x'_i}} \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x_i}')}(2 - \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x'_{i+1}})} - \delta_{\sigma(\mathbf{x}),\sigma(\mathbf{x'_{i-1}})}). \tag{27}$$

Here the 2 reflects the fact that for each neighbour $\mathbf{x}'$, at most 2 locations can have a spin differing from $\sigma(\mathbf{x})$, since only the next and previous locations clockwise around $\mathbf{x}$ are considered. Based on the value of C, rejection or acceptance of an updated is determined as follows:

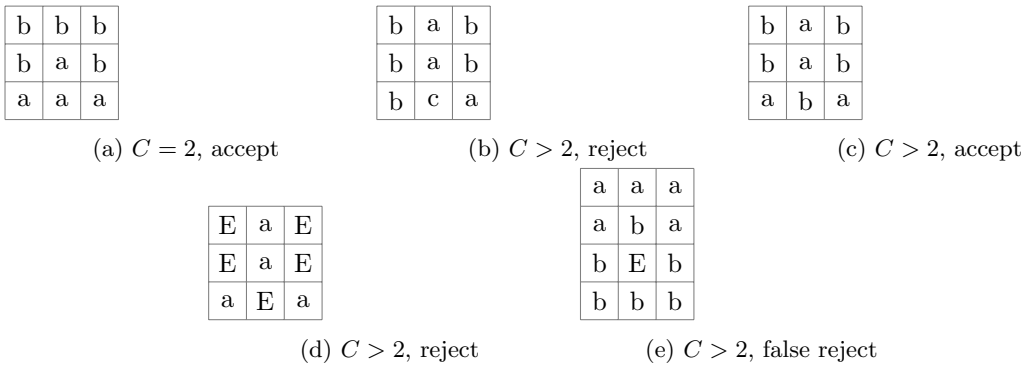
1. When $C \leq 2$, the update is accepted.
2. When $C > 2$ and the neighborhood of $\mathbf{x}$ is populated by precisely 2 non-ECM cells (and no ECM), the update is accepted.
3. When $C > 2$ otherwise, the update is rejected.

Some applications of the constraint are illustrated in Figure 28 (reproduction based on Merks [13]). The sub-figures display possible lattice configurations of the central lattice site $\mathbf{x}$ and its neighbors $\mathbf{x}'$ within the 2−neighborhood. We write A and B for the cells consisting of sites with spin a and b respectively. The connectivity constraint does not apply to the extracellular matrix, whose spin is denoted by E.

1. In Figure 28a, $C = 2$. Hence an update attempt for the central lattice site $\mathbf{x}$ changing its spin $\sigma(\mathbf{x})$ from a to b is accepted. Indeed A and B remain connected.
2. In Figure 28b, $C > 2$ and the neighborhood contains more than 2 non-ECM cell. Hence an update attempt changing $\sigma(\mathbf{x})$ from a to either b or c is rejected, since this would break the connectivity of A.
3. In Figure 28c, $C > 2$ for precisely 2 non-ECM cells in the neighborhood. Hence the update attempt is accepted.[5]
4. In Figure 28d, $C > 2$ and the neighborhood contains only 1 non-ECM cell. Hence an update attempt changing $\sigma(\mathbf{x})$ from a to E is rejected, since this would break the connectivity of A.
5. In Figure 28e, $C > 2$ and the neighborhood does not consist of precisely 2 non-ECM cells. Hence an update attempt changing $\sigma(\mathbf{x})$ from b to either a or E is rejected, since this would break the connectivity of B *locally* (i.e. in the 2-neighborhood of $\mathbf{x}$). However, this is a false rejection, since B remains connected via a global path.

The rejection of an update is enforced by incurring a high energy penalty. In this way, the connectivity constraints precludes a breach of connectivity of connected lattice sites.

[5] But note that this breaks A. This has been brought to the attention of dr. Merks (personal communication). This issue is still unaddressed. It remains to be investigated how this influences the efficiency of Merks' connectivity constraint.



(a) $C = 2$, accept (b) $C > 2$, reject (c) $C > 2$, accept

(d) $C > 2$, reject (e) $C > 2$, false reject

Figure 28: Examples of connectivity constraint

However, the constraint also precludes the copying of spins when the lattice site would remain connected via a non-local path (Figure 28e). Merks points out that, in contrast to this local connectivity constraint, a *global* connectivity constraint would be computationally expensive ($\mathcal{O}(N^2)$), while the results would not be significantly different.[6]

[6] O.c. and personal communication.