bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

BIO-LGCA: a cellular automaton modelling class for analysing collective cell migration

Andreas Deutsch[1], Josué Manik Nava-Sedeño[1,2], Simon Syga[1], and Haralampos Hatzikirou[1, 3]

[1]Technische Universität Dresden, Centre for Information Services and High Performance Computing, Nöthnitzer Straße 46, 01062 Dresden, Germany.
[2]Universidad Nacional Autónoma de México, Faculty of Sciences, Department of Mathematics, Circuito Exterior, Ciudad Universitaria, 04510 Mexico City, Mexico.
[3]Mathematics Department, Khalifa University, P.O. Box 127788, Abu Dhabi, UAE

1 Abstract

Collective dynamics in multicellular systems such as biological organs and tissues plays a key role in biological development, regeneration, and pathological conditions. Collective dynamics - understood as population behaviour arising from the interplay of the constituting discrete cells - can be studied with mathematical models. Off- and on-lattice agent-based models allow to analyse the link between individual cell and collective behaviour. Notably, in on-lattice agent-based models known as cellular automata, collective behaviour can not only be analysed through computer simulations, but predicted with mathematical methods. However, classical cellular automaton models fail to replicate key aspects of collective migration, which is a central instance of collective behaviour in multicellular systems.

To overcome drawbacks of classical on-lattice models, we introduce a novel on-lattice, agent-based modelling class for collective cell migration, which we call biological lattice-gas cellular automaton (BIO-LGCA). The BIO-LGCA is characterised by synchronous time updates, and the explicit consideration of individual cell velocities. While rules in classical cellular automata are typically chosen ad hoc, we demonstrate that rules for cell-cell and cell-environment interactions in the BIO-LGCA can also be derived from experimental single cell migration data or biophysical laws for individual cell migration. Furthermore, we present elementary BIO-LGCA models of fundamental cell interactions, which may be combined in a modular fashion to model complex multicellular phenomena. Finally, we present a mathematical mean-field analysis of a BIO-LGCA model that allows to predict collective patterns for a particular cell-cell interaction. A Python package which implements various interaction rules and

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

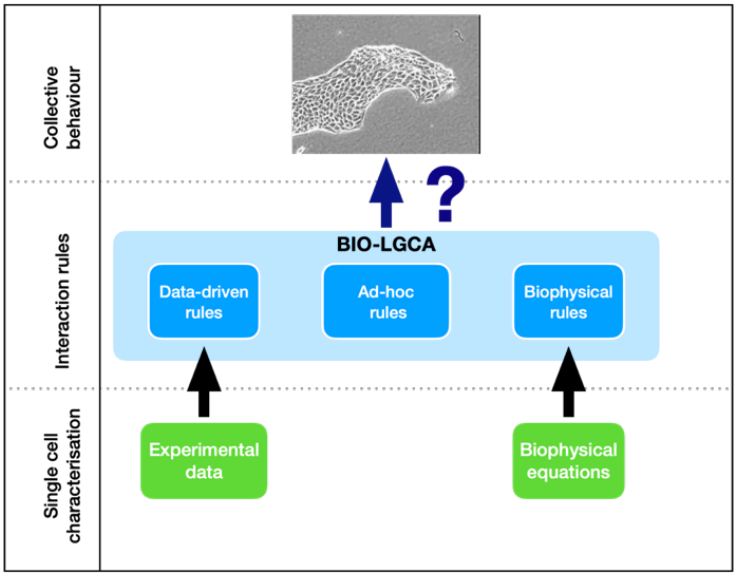

Figure 1: BIO-LGCA modelling: The key question is to extract the interaction rules underlying a particular collective behaviour in a population of cells. Appropriate BIO-LGCA interaction rules can be chosen ad hoc, extracted from experimental single cell migration data, or derived from biophysical equations for single cell migration.

visualisations of BIO-LGCA model simulations we have developed is available at https://github.com/sisyga/BIO-LGCA.

2 Author summary

Pathophysiological tissue dynamics, such as cancer tissue invasion, and structure formation during embryonic development, emerge from individual inter-cellular interactions. In order to study the impact of single cell dynamics and cell-cell interactions on tissue behaviour, one needs to develop space-time-dependent agent-based models (ABMs), which consider the behaviour of individual cells. Typically, in agent-based models there is a payoff between biological realism and computational cost of corresponding model simulations. Continuous time ABMs are typically more realistic but computationally expensive, while rule- and lattice-based ABMs are regarded as phenomenological but computationally efficient and amenable to mathematical analysis. Here, we present the rule- and lattice-based BIO-LGCA modelling class which allows for (i) rigorous derivation of rules from biophysical laws and/or experimental data, (ii) mathematical analysis of the resulting dynamics, and (iii) computational efficiency.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3 Introduction

Systems biology and mathematical modelling is rapidly expanding its scope from the study of single cells to the analysis of collective behaviour in multicellular tissue- and organ-scale systems. In such systems, individual cells may interact with their environment (hapto- and chemotaxis, contact guidance, etc.) or with other cells (cell-cell adhesion, contact inhibiton of locomotion, etc.) and produce collective patterns exceeding the cells' interaction range. To study collective behaviour in such systems theoretically and/or computationally, a mathematical model must be decided upon as a first step. Continuous models describe the average behaviour of the cellular population. Their lack of resolution at the individual level makes them inappropriate to study the role of individuals in collective behaviour. Agent-based models, on the other hand, are particularly suited to the study of collective behaviour in multicellular systems, as they resolve individual cell dynamics, and thus allow for the analysis of large-scale tissue effects of individual cell behaviour.

In the context of multicellular tissue dynamics, various agent-based models have been developed to analyse tissue dynamics as a collective phenomenon emerging from the interplay of individual biological cells. In these models, cells are regarded as separate, individual units, contrary to continuum methods, which neglect the discrete individual cell nature, and where tissue dynamics is derived from conservation and constitutive laws, drawing parallels to physical systems. Since agent-based models represent individual biological cells, distinct cell phenotypes can be taken into account, which may be fundamental for the organisation at the tissue level. For example, it has been shown that cell-to-cell variability plays a key role in tumour progression and resistance to treatment [1]. Moreover, with the advance of high performance computing, agent-based models can be used to analyse *in vitro* systems at a 1:1 basis even for large cell population sizes.

Agent-based models can be classified into on-lattice and off-lattice or "lattice-free" models depending on whether cell movement is restricted to an underlying lattice (see [2] for references). While lattice-free models are typically very detailed in their biophysical description, they are often too complex for mathematical analysis, while lattice models are normally very abstract and phenomenological, thus facilitating their analysis, but making biological data integration at the model definition level challenging.

In lattice models, either (i) a lattice site may be occupied by many biological cells (e.g., [3]), (ii) a site may be occupied by at most one single biological cell (e.g., [4]), or (iii) several neighbouring lattice sites may represent a single biological cell (e.g., [5]). In probabilistic lattice models, the interacting particle system (IPS) is an important example, proliferation, death, and migration of biological cells are modeled as stochastic processes. Model types (i) and (ii) can mimick volume exclusion effects, (iii) can qualitatively capture cell deformation and compression, while each of the three approaches can describe the effects of mechanical forces of one cell on its neighbour, or on a group of neighbouring cells to some extent.

Lattice models are equivalent to cellular automata, which have been introduced by J. v. Neumann and S. Ulam in the 1950s as models for individual (self-)reproduction [6]. A cellular automaton consists of a regular spatial lattice in which each lattice node can assume a discrete, typically finite number of

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

states. The next state of a node solely depends on the states in neighbouring sites and a deterministic or stochastic transition function. Cellular automata provide simple models of self-organising systems in which collective behaviour emerges from an ensemble of many interacting "simple" components - being it molecules, cells or organisms [7, 8, 9, 10].

However, when modeling collective migration phenomena, classic CA and IPS models have major drawbacks which are due to the strict volume exclusion and asynchronous update in such models. Most importantly, they fail to produce collective movement at unit density, since this density implies a "jammed state" due to volume exclusion. These are for example the "traffic jams" in [11]. However, a fluidized state at unit density is important in epithelia. Furthermore, asynchronous update may lead to oscillating density spikes. For example in an IPS model for persistent motion in a crowded environment, cells at the invasion front detach and leave gaps behind that are subsequently filled by following cells [12]. This is an artefact of models with asynchronous update as in reality invasion can happen while cells stay connected all the time. Moreover, classic CA models consider only cell position and not explicitly cell momentum, complicating the modeling of collective cell migration mediated primarily through changes in momentum, rather than density.

The lattice-gas cellular automaton (BIO-LGCA) introduced here, on the other hand, is a cellular automaton in which lattice sites are updated synchronously, and which explicitly considers individual cell velocities. These features make the BIO-LGCA appropriate for modeling collective migration phenomena where cell interactions result in directional changes of velocity, and where high cell densities do not hamper movement. Table 1 presents an overview of on-lattice cell-based models.

Table 1: Comparison of on-lattice cell-based models with respect to time representation, computational efficiency, migration modelling capacity, and availability of analytic methods. CA: cellular automaton, BIO-LGCA: biological lattice-gas cellular automaton, IPS: interacting particle system, CPM: cellular Potts model; for details see also [2].

model	**comput. effic.**	**model. flexibility**	**analytic methods**	**single cell mig.**	**coll. cell mig.**
one site/ one cell (CA/IPS)	++	+/-	++	++	+/-
one site/ many cells (CA/IPS)	++	+/-	++	-	+
one site/ many cells with velocity (BIO-LGCA)	++	+	+	++	++
many sites/one cell (CPM)	-	++	-	++	++

The structure of the paper is as follows: we first formally define the BIO-LGCA model class. Then, we construct *biophysical BIO-LGCA rules* from microscopic Langevin models for selected cases of single and collective cell migration. Subsequently, we demonstrate how to generate *data-driven BIO-LGCA rules* from experimental single cell migration data (Fig. 1). Furthermore, we show that, in specific cases, the biophysical and the data-driven approaches con-

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

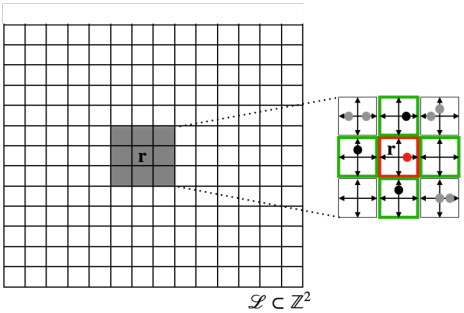

Figure 2: Lattice and neighbourhood in the BIO-LGCA: example of square lattice (left). The node state is represented by the occupation of velocity channels (right); in the example, there are four velocity channels $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4$, corresponding to the square geometry of the lattice, and one "rest channel" $\mathbf{c}_0$. Filled dots denote the presence of a cell in the respective velocity channel; right: von Neumann neighbourhood (green) of the red node.

verge to the same functional form. For this case, we present several simple, but biologically relevant model examples. We end with a critical discussion of the BIO-LGCA modelling framework.

4 Definition

A BIO-LGCA is defined by a discrete spatial lattice $\mathcal{L}$, a discrete state space $\mathcal{E}$, a neighbourhood $\mathcal{N}$ and local rule-based dynamics.

4.1 Lattice

The regular lattice $\mathcal{L} \subset \mathbb{R}^d$ consists of nodes $\mathbf{r} \in \mathcal{L}$. Every node has b nearest-neighbours, where b depends on the lattice geometry. Each lattice node $\mathbf{r} \in \mathcal{L}$ is connected to its nearest neighbours by unit vectors $\mathbf{c}_i, i = 1, \ldots, b$, called velocity channels. In addition, a variable number $a \in \mathbb{N}_0$ of rest channels (zero-velocity channels) $\mathbf{c}_j = \mathbf{0}$, $b < j \leq a + b$, is allowed (Fig. 2). The parameter $K = a + b$ defines the maximum *node capacity*.

4.2 Neighbourhood

The set $\mathcal{N}$, the neighbourhood template, defines the nodes which determine the dynamics of the node $0 \in \mathcal{L}$. Throughout this work, the neighbourhood will be assumed to be a von Neumann neighbourhood (Fig. 2), defined as $\mathcal{N}^b := \mathcal{N}^b(0) = \{\mathbf{c_1}, \mathbf{c_2}, \ldots, \mathbf{c_b}\}$, but other neighbourhood choices are possible. In general, $\mathcal{N}(\mathbf{r}) := \mathcal{N}^b(\mathbf{r}) = \mathcal{N}^b + \mathbf{r}$, specifies the set of lattice nodes which determine the dynamics of the state at node $\mathbf{r} \in \mathcal{L}$. [1]

[1] $\mathcal{N}^b + \mathbf{r} = \{\mathbf{r} + \mathbf{r}', \mathbf{r}' \in \mathcal{N}^b\}$

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4.3 State space

The state space in LGCA is defined through the occupation numbers $s_j \in \{0,1\}$, $j = 1, \ldots, K$. These occupation numbers represent the presence ($s_j = 1$) or absence ($s_j = 0$) of a cell in the channel $\mathbf{c}_j$ within some node. Then, the configuration of a node is given by the state vector

$$\mathbf{s} = (s_1, \ldots, s_K) \in \mathcal{E} = \{0,1\}^K .$$

This reflects an *exclusion principle* which allows not more than one cell at the same node within the same channel simultaneously. As a consequence, each node $\mathbf{r} \in \mathcal{L}$ can host up to K cells, which are distributed in different channels.

It is possible to consider more than one cell phenotype in the BIO-LGCA model. In this case each phenotype is indexed by $\sigma \in \Sigma \subset \mathbb{N}$. Then, the configuration vector is given by

$$\mathbf{s} = (\mathbf{s}_\sigma)_{\sigma \in \Sigma} \in \mathcal{E} = \{0,1\}^{|\Sigma| K},$$

where $|\cdot|$ denotes the cardinality of a set. Each node will be able to support up to $|\Sigma| K$ cells.

Two useful quantities for a given node are the *total number of cells at the node* $n(\mathbf{s})$ and the *momentum/node flux* $\mathbf{J}(\mathbf{s})$, defined as

$$n(\mathbf{s}) := \sum_{\sigma \in \Sigma} n_\sigma(\mathbf{s}) = \sum_{\sigma \in \Sigma} \sum_{j=1}^{K} s_{\sigma,j} \quad \text{and} \quad \mathbf{J}(\mathbf{s}) := \sum_{\sigma \in \Sigma} \sum_{j=1}^{K} s_{\sigma,j}\, \mathbf{c}_j, .$$

where $n_\sigma(\mathbf{s})$ is the σ number of cell phenotypes.

4.4 Dynamics

In general, in cellular automata a new lattice configuration is created according to a local rule that determines the new state of each node in terms of the current states of the node and the nodes in its neighbourhood. In order to determine a new lattice configuration, the local rule is applied independently and simultaneously at every node $\mathbf{r}$ of the lattice. Mathematically, in probabilistic cellular automata, the local rule can be interpreted as a transition probability $P(\mathbf{s} \to \mathbf{s}')$ to replace a current configuration $\mathbf{s}$ with a new node configuration $\mathbf{s}'$.

In a BIO-LGCA, local rules are composed of a particular combination of operators for stochastic reorientation ($\mathcal{O}$), phenotypic switching ($\mathcal{S}$), and stochastic cell birth and death ($\mathcal{R}$), as well as a deterministic propagation operator ($\mathcal{P}$) (see figure 3). The propagation and reorientation operators together define cell movement, while phenotypic switching allows cells to stochastically and reversibly transition between phenotypes. In a BIO-LGCA, the stochastic operators are applied sequentially to every node, such that the transition probability can be expressed as

$$P(\mathbf{s} \to \mathbf{s}') = P_\mathcal{S} P_\mathcal{R} P_\mathcal{O},$$

where $P_i, i \in \{\mathcal{S}, \mathcal{R}, \mathcal{O}\}$ are the transition probabilities of the corresponding operator. In this way, a post-interaction node configuration $\mathbf{s}'$ is defined as the resulting node configuration after subsequent application of the stochastic

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

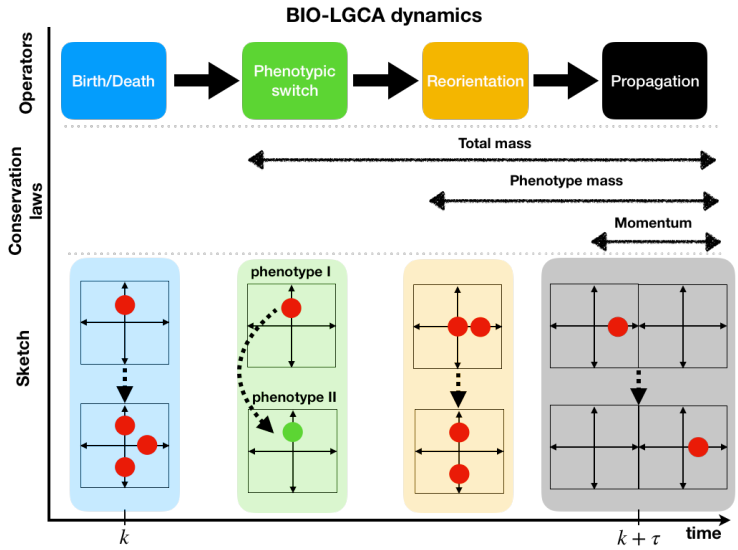

Figure 3: Operator-based dynamics of the BIO-LGCA: Propagation $\mathcal{P}$, reorientation $\mathcal{O}$, phenotypic switch $\mathcal{S}$, and birth/death operators $\mathcal{R}$ (top); conservation laws maintained by the different operators (middle); sketches of the operator dynamics (bottom), see text for explanations.

operators, i.e. $\mathbf{s}' = \mathbf{s}^{\mathcal{S}\circ\mathcal{R}\circ\mathcal{O}}$. Subsequently, the deterministic propagation operator $\mathcal{P}$ is applied: cells in velocity channels at the node, i.e. moving cells, are translocated to neighbouring nodes in the direction of their respective velocity channels. The time step increases once the propagator operator has been applied. Accordingly, the dynamics of the BIO-LGCA can be summarized in the stochastic microdynamical equation

$$s_j\left(\mathbf{r}+\mathbf{c}_j, k+1\right) = s'_j\left(\mathbf{r}, k\right). \tag{1}$$

5 BIO-LGCA rule derivation

In classical cellular automata, transition probabilities are typically chosen *ad hoc*. Here, we show that BIO-LGCA rules can also be rigorously derived from known biophysical equations of motion, and from experimental data reflecting the mean behaviour of individual cells. In the following, we disregard birth/death processes and phenotypic transitions. For the corresponding BIO-LGCA model specified exclusively by transition probabilities for reorientation, we first present a method to derive the BIO-LGCA transition probabilities from a Langevin model describing single-cell migration [13] and then a method where the the transition probabilities are derived from average observations, while the internal dynamics of the cells are assumed to be unknown [14]. In certain cases, independent of the particular method chosen for rule derivation, the functional form of the transition probabilities will be the same (Fig. 4).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5.1 Reorientation dynamics derived from biophysical equations of motion

In the field of cell migration, different types of cell migration have been described by a set of stochastic differential equations governing the motion of discrete cells in an overdamped situation, e.g in a highly viscous medium. The corresponding off-lattice model is known as self-propelled particle model (SPP), and the stochastic differential equations encoding individual cell motion (as introduced in [15]) are Langevin equations, where a stochastic variable θ_m describes the orientation of the m-th cell in the system which moves with a constant speed $v_0 \in \mathbb{R}^+$ and orientation $\theta_m(t) \in [0, 2\pi)$ varying according to some potential and influenced by noise. Its equations of motion read [16]:

$$\dot{\mathbf{x}}_m = v_0 \mathbf{v}(\theta_m), \tag{2a}$$

$$\dot{\theta}_m = -\gamma \frac{\partial U(\{\mathbf{x}_k\}, \{\theta_k\})}{\partial \theta_m} + \xi_m(t), \tag{2b}$$

where $\mathbf{x}_m \in \mathbb{R}^d$ is the cell's spatial position, $\mathbf{v}(\theta_m) \in \mathbb{R}^d$ is a unit vector pointing in the direction of the cell's displacement, $\gamma \in \mathbb{R}^+$ is a relaxation constant, and $\xi_m(t)$ is a white noise term with zero mean and correlation $\langle \xi_m(t_1) \xi_n(t_2) \rangle = 2D_\theta \delta(t_2 - t_1) \delta_{m,n}$. The heart of the model is the potential $U(\{\mathbf{x}_k\}, \{\theta_k\}) : \mathbb{R}^{Nd} \times [0, 2\pi)^N \mapsto \mathbb{R}$, where N is the number of cells within the central cell's neighbourhood of interaction. This potential encodes all the biophysical mechanisms that dictate the cell's reorientation. The reorientation potential typically only depends on the orientations of neighbouring cells, though a dependence on cell positions is also possible [17].

The probability density function of the stochastic variable θ_m governed by the aforementioned Langevin equations is given by the corresponding Fokker-Planck equation

$$\frac{\partial P(\theta_m, t)}{\partial t} = \gamma \frac{\partial}{\partial \theta_m} \left(\frac{\partial U}{\partial \theta_m} P(\theta_m, t) \right) + D_\theta \frac{\partial^2 P(\theta_m, t)}{\partial \theta_m^2}. \tag{3}$$

If we assume fast relaxation times for the solution of the Fokker-Planck equation, then one can take the p.d.f. of the steady state, $P(\theta_m)$, as the probability of cell m to have an orientation θ_m.
The next step is to relate this probability to a node configuration probability. For this, we identify θ_m with the argument of the unit vector pointing towards the direction of the cell's displacement, i.e. $\theta_m = \arg(\mathbf{v}_m)$, so that the probability can be expressed as $P(\theta_m) = P(\mathbf{v}_m)$.

Given that the velocity is constant and the direction of motion is totally defined by the orientation of the cell, we can identify the (instantaneous) cell displacement $\mathbf{v}(\theta_m)$ by an occupied velocity channel in the BIO-LGCA, $\mathbf{c}_m$, as it fully determines the translocation of the cell during the propagation step. Due to the velocity discretization in the BIO-LGCA, cells in the Langevin model with an orientation $\arg(\mathbf{c}_m) - a \leq \theta \leq \arg(\mathbf{c}_m) + b$; $a, b \in \mathbb{R}$, are described in the LGCA model as occupying velocity channel $\mathbf{c}_m$ (Fig. 5). The probability of occupying velocity channel $\mathbf{c}_m$ in the BIO-LGCA can then be calculated as

$$P(\mathbf{c}_m) = \int_{\arg(\mathbf{c}_m)-a}^{\arg(\mathbf{c}_m)+b} P(\theta) \mathrm{d}\theta.$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

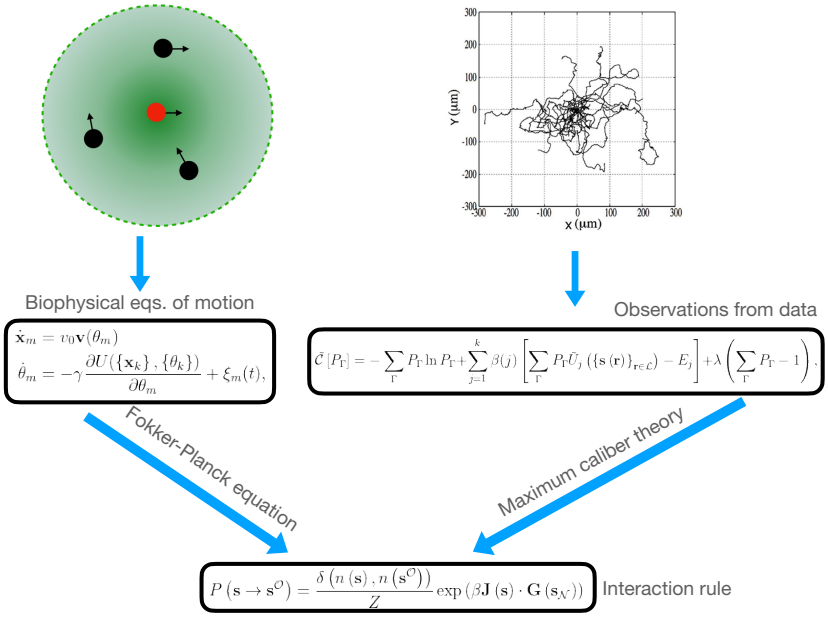

Figure 4: Rule generation in BIO-LGCA models: The transition probability defining the BIO-LGCA interaction rule can be derived from experimental data and biophysical equations of motion (for explanation see text).

We now assume that we can choose the integration interval $[\arg(\mathbf{c}_m)-a, \arg(\mathbf{c}_m)+b]$, such that, by the mean value theorem,

$$P(\mathbf{c}_m) = K \cdot P(\arg(\mathbf{c}_m)),$$

where $K = a + b$ is the size of the integration interval. We shall asume that K is identical for all velocity channels.

If we further assume that the occupation of velocity channels is uncorrelated, then the probability to transition to a post-reorientation configuration $\mathbf{s}^{\mathcal{O}}$ follows a multinomial distribution:

$$P\left(\mathbf{s} \to \mathbf{s}^{\mathcal{O}}\right) = \frac{\delta\left(n\left(\mathbf{s}\right), n\left(\mathbf{s}^{\mathcal{O}}\right)\right)}{Z\left(\mathbf{s}^{\mathcal{O}}\right)} \prod_{m=0}^{b} P(\mathbf{c}_m)^{s_m^{\mathcal{O}}}, \tag{4}$$

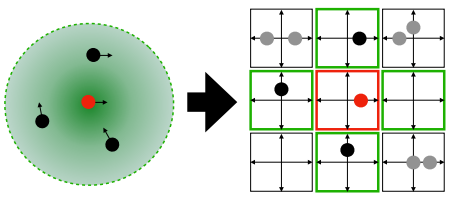

Figure 5: Sketch of velocity discretization in BIO-LGCA models from a biophysical off-lattice Langevin model.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where the Kronecker delta ensures mass conservation, and $Z\left(\mathbf{s}^{\mathcal{O}}\right)$ is the normalisation factor which has the following form:

$$Z\left(\mathbf{s}^{\mathcal{O}}\right) = \sum_{\mathbf{s}^{\mathcal{O}}} \prod_{m=0}^{b} P(\mathbf{c}_m)^{s_m^{\mathcal{O}}}. \tag{5}$$

If the drift term in Eq. 3 is non-zero, we can further simplify before plugging in the interaction potential. With non-zero drift, we have for the steady state:

$$\gamma \frac{\partial}{\partial \theta_m}\left(\frac{\partial U}{\partial \theta_m} P\right) + D_\theta \frac{\partial^2 P}{\partial \theta_m^2} = 0, \tag{6}$$

which, after integration, yields the following:

$$P(\theta_m) = C_0 \exp\{(-\beta U(\theta_m))\}, \tag{7}$$

where $\beta = \gamma / D_\theta$. Inserting this into Eq. 4 and absorbing the integration constant C_0 in the partition function Z, one obtains the following expression:

$$\begin{aligned} P\left(\mathbf{s} \to \mathbf{s}^{\mathcal{O}}\right) =& \frac{\delta\left(n\left(\mathbf{s}\right), n\left(\mathbf{s}^{\mathcal{O}}\right)\right)}{Z} \prod_{m=0}^{b} \left\{C_0 \exp\{[-\beta U(\mathbf{c}_m)]\}\right\}^{s_m^{\mathcal{O}}} \\ =& \frac{\delta\left(n\left(\mathbf{s}\right), n\left(\mathbf{s}^{\mathcal{O}}\right)\right)}{\tilde{Z}} \exp\left\{-\beta \sum_{m=0}^{b} \left[U(\mathbf{c}_m) s_m^{\mathcal{O}}\right]\right\}, \end{aligned} \tag{8}$$

where $\tilde{Z}$ is the normalisation constant in which the integration constant has been absorbed.
In order to account for possible discrepancies between the original Langevin model and the derived BIO-LGCA model (Fig. 6), it is important to emphasise the assumptions made during the derivation:

- The relaxation time of the Fokker-Planck solution is smaller than the BIO-LGCA time step.
- $P(\arg(\mathbf{c}_m))$ is the mean value of $P(\theta_m)$ in the interval $[\arg(\mathbf{c}_m) - a, \arg(\mathbf{c}_m) + b]$.
- The size of the integration interval, K, is identical for all velocity channels.
- The occupation probabilities of all velocity channels are uncorrelated.

Collective migration Let's now consider the case where the reorientation potential is of the form

$$U\left(\{\mathbf{x}_k\}, \{\theta_k\}\right) = -C\left(\{\mathbf{x}_k\}, \{\theta_k\}\right) \cos\left[\theta_m - \varphi\left(\{\mathbf{x}_k\}, \{\theta_k\}\right)\right], \tag{9}$$

that is, the reorientation potential depends only on the cosine of the orientation of the central cell, whose amplitude and shift may depend on the positions and/or orientations of all cells within the neighbourhood of interaction (including the central cell) only through the amplitude (C) and shift (φ) of the cosine

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

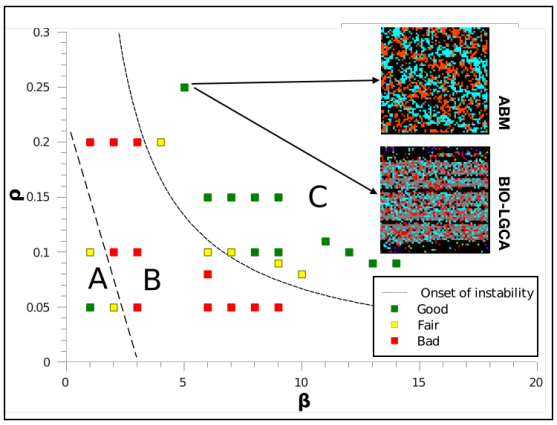

Figure 6: Comparison of Langevin ABM model and BIO-LGCA simulations in the density (ρ) vs interaction sensitivity (β) parameter space (from [13]).

function. Using trigonometric identities, the reorientation potential can then be rewritten as

$$U\left(\{\mathbf{x}_k\},\{\theta_k\}\right) = -\mathbf{v}\left(\theta_m\right)\cdot\mathbf{G}\left(\{\mathbf{x}_k\},\{\theta_k\}\right), \tag{10}$$

where $\mathbf{G}\left(\{\mathbf{x}_k\},\{\theta_k\}\right)$ is called the *local director field*, whose norm and argument are, respectively, $\|\mathbf{G}\left(\{\mathbf{x}_k\},\{\theta_k\}\right)\| = C\left(\{\mathbf{x}_k\},\{\theta_k\}\right)$ and $\arg\left[\mathbf{G}\left(\{\mathbf{x}_k\},\{\theta_k\}\right)\right] = \varphi\left(\{\mathbf{x}_k\},\{\theta_k\}\right)$. Substituting Eq. 10 in Eq. 8, and using the linearity of the internal product, the transition probability of the reorientation operator is

$$P\left(\mathbf{s}\to\mathbf{s}^{\mathcal{O}}\right) = \frac{\delta\left(n\left(\mathbf{s}\right),n\left(\mathbf{s}^{\mathcal{O}}\right)\right)}{Z}\exp\left(\beta\mathbf{J}\left(\mathbf{s}\right)\cdot\mathbf{G}\left(\mathbf{s}_{\mathcal{N}}\right)\right), \tag{11}$$

where $\mathbf{G}\left(\mathbf{s}_{\mathcal{N}}\right)$ is the local director field of the neighbourhood configuration, and $\mathbf{J}\left(\mathbf{s}\right)$ is the node flux, as described previously.

In general, whenever the reorientation potential can be expressed as

$$U\left(\{\mathbf{x}_k\},\{\theta_k\}\right) = -C\left(\{\mathbf{x}_k\},\{\theta_k\}\right)\cos^n\left[\theta_m - \varphi\left(\{\mathbf{x}_k\},\{\theta_k\}\right)\right], \tag{12}$$

with $n\in\mathbb{N}$, the argument of the exponential in the transition probability can be expressed as an internal product of two vectors. In the specific case of $n=2$, using trigonometric functions, one can arrive at the transition probability

$$P\left(\mathbf{s}\to\mathbf{s}^{\mathcal{O}}\right) = \frac{\delta\left(n\left(\mathbf{s}\right),n\left(\mathbf{s}^{\mathcal{O}}\right)\right)}{Z}\exp\left(\beta\mathbf{N}\left(\mathbf{s}\right)\cdot\mathbf{G}\left(\mathbf{s}_{\mathcal{N}}\right)\right), \tag{13}$$

where $\mathbf{N}\left(\mathbf{s}\right)$ is the *local nematic alignment vector*, and is defined as

$$\mathbf{N}\left(\mathbf{s}\right) := \frac{1}{2}\sum_{p=1}^{b}\mathbf{c}_{[2(p-1)\mathrm{mod}b]+1}s_p. \tag{14}$$

Thus, the reorientation probabilities have the same general form whenever the interaction potential is conservative and consists of a pairwise comparison between the angles and/or positions of neighbouring cells.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5.2 Data-driven rules

Due to the complexity of inter- and intracellular processes, it is not uncommon to have little knowledge about biophysical mechanisms mediating single and collective cell migration. Typically, there is partial to no knowledge of the intracellular processes underlying experimental observations, such as signalling pathways. In this case, statistical methods can be used to derive reorientation rules needed to define appropriate BIO-LGCA models.

Suppose that an experimental observation is the only known information about a single or collective cell migration process, with no knowledge about the underlying mechanisms driving such process. These observations are typically averages of a time-dependent quantity (called the observable). The reorientation probabilities of the corresponding BIO-LGCA model should be chosen such that the mean of the observable function matches the experimental observation. For example, a common observable is the comparison between the initial cell velocity v_0, and the velocity at a later time t, v_t, expressed as $v_0 \cdot v_t$. The corresponding experimental observation is called the velocity autocorrelation function, defined as $g(t) = \langle v_0 \cdot v_t \rangle$.

However, this is not enough to completely determine the probability distribution. Take, for example, a zero-mean observation. Most symmetric distributions centered about zero will fulfill the observation. Which functional form of the probabilities to choose, then? We can actually exploit the lack of information on the mechanistic nature of the process to our own advantage. The maximum caliber (and maximum entropy) formalism [18] dictates to maximise entropy (a measure of the lack of information contained in a probabilistic model) while restricting probabilities to reproduce the experimental observation. This translates into optimising the following functional

$$\tilde{\mathcal{C}}\left[P_\Gamma\right] = -\sum_\Gamma P_\Gamma \ln P_\Gamma + \sum_{j=1}^{k} \beta(j) \left[\sum_\Gamma P_\Gamma \tilde{U}_j \left(\{\mathbf{s}\left(\mathbf{r}\right)\}_{\mathbf{r}\in\mathcal{L}}\right) - E_j\right] + \lambda \left(\sum_\Gamma P_\Gamma - 1\right), \tag{15}$$

where P_Γ is the probability of a cell to follow a certain spatial trajectory Γ, $\beta(j)$ and λ are Lagrange multipliers, $\tilde{U}_k\left(\{\mathbf{s}\left(\mathbf{r}\right)\}_{\mathbf{r}\in\mathcal{L}}\right)$ is the value of the observable at the time step k depending on the state of the lattice $\{\mathbf{s}\left(\mathbf{r}\right)\}_{\mathbf{r}\in\mathcal{L}}$, and E_k is the value of the observation at the time step k, which is the average of the observable obtained from experimental data. The first term of the functional is the entropy, which we want to maximise. The second term restricts the resulting probabilities to match the experimental observation. Since we assume the observation, E_k to be a time-dependent function, then a Lagrange multiplier $\beta(k)$ is needed for every time step k. The last term guarantees the normalisation of probabilities, which requires and additional Lagrange multiplier, λ.

The optimisation of this functional yields an optimal value for the path probabilities $P_\Gamma = \frac{1}{Z}\exp\left[\sum_{j=1}^{k}\beta(j)\tilde{U}_j\left(\{\mathbf{s}\left(\mathbf{r}\right)\}_{\mathbf{r}\in\mathcal{L}}\right)\right]$, where the value of $\beta(j)$ is such that $E_k = \sum_\Gamma P_\Gamma \tilde{U}_j\left(\{\mathbf{s}\left(\mathbf{r}\right)\}_{\mathbf{r}\in\mathcal{L}}\right)$.

If the process is Markovian, then one may decompose the path probability into individual channel occupation probabilities for each time step k, as

$$P_{i,k} = \frac{1}{Z}\exp\left[\beta(k)\tilde{U}_k\left(s_i \mid \mathbf{s}_\mathcal{N}\right)\right], \tag{16}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where the observable is dependent on the occupancy of the i-th channel of the node and conditioned to a certain configuration of its interaction neighbourhood. For example, if the observation is the autocorrelation function $g(t) = \langle \mathbf{v}_0 \cdot \mathbf{v}_t \rangle$, where $\mathbf{v}_t$ denotes the normalized velocity of a cell at the time t, determined from experimental data, then the corresponding channel occupation probabilities are found to be

$$P_{i,k} = \frac{1}{z} \exp\left[dg(k)\left(\mathbf{c}_{i_0} \cdot \mathbf{c}_i\right)\right], \tag{17}$$

where z is the normalisation constant for the transition probability, d is the dimension of space, and $\mathbf{c}_{i_0}$ is the initial orientation of the cell.

If we assume independence among cells within the same node, we arrive at a reorientation probability of the form

$$P\left(\mathbf{s} \to \mathbf{s}^{\mathcal{O}}\right) = \frac{\delta\left(n\left(\mathbf{s}\right), n\left(\mathbf{s}^{\mathcal{O}}\right)\right)}{Z} \exp\left[\sum_{i=1}^{b} \beta(k) \tilde{U}_k\left(s_i \mid \mathbf{s}_{\mathcal{N}}\right)\right]. \tag{18}$$

Note that, if both observation and observable are time-independent, then the Lagrange multiplier $\beta(k) = \beta$ is also time independent, and the transition probabilities are given by

$$P\left(\mathbf{s} \to \mathbf{s}^{\mathcal{O}}\right) = \frac{\delta\left(n\left(\mathbf{s}\right), n\left(\mathbf{s}^{\mathcal{O}}\right)\right)}{Z} \exp\left[\sum_{i=1}^{b} \beta \tilde{U}\left(s_i \mid \mathbf{s}_{\mathcal{N}}\right)\right]. \tag{19}$$

Furthermore, when $\tilde{U}\left(s_i \mid \mathbf{s}_{\mathcal{N}}\right) = s_i \mathbf{c}_i \cdot \mathbf{G}\left(\mathbf{s}_{\mathcal{N}}\right)$, then Eq. 19 reduces to Eq. 11.

6 BIO-LGCA rules for single and collective cell migration

Now that the BIO-LGCA framework has been defined and data- and equation-based methods for rule derivation have been described, we present key examples of transition probabilities corresponding to reorientation operators, which model important elementary single-cell and collective behaviours. Note that several of these examples' probabilities have the general form of Eq. 11.

6.1 Single cell migration

Random walk Random walks are performed by cells such as bacteria and amoebae in the absence of any environmental cues. Random walk of cells can be modeled by a reorientation operator with the following transition probabilities:

$$P\left(\mathbf{s} \to \mathbf{s}^{\mathcal{O}}\right) = \frac{1}{Z(\mathbf{s})} \cdot \delta\left(n(\mathbf{s}), n(\mathbf{s}^{\mathcal{O}})\right). \tag{20}$$

This rule conserves mass, i.e. cell number.

Chemotaxis Chemotaxis describes the dependence of individual cell movement on a chemical signal gradient field. Accordingly, spatio-temporal pattern formation at the level of cells and chemical signals can be observed. Chemotactic patterns result from the coupling of different spatio-temporal scales at the cell and the molecular level, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

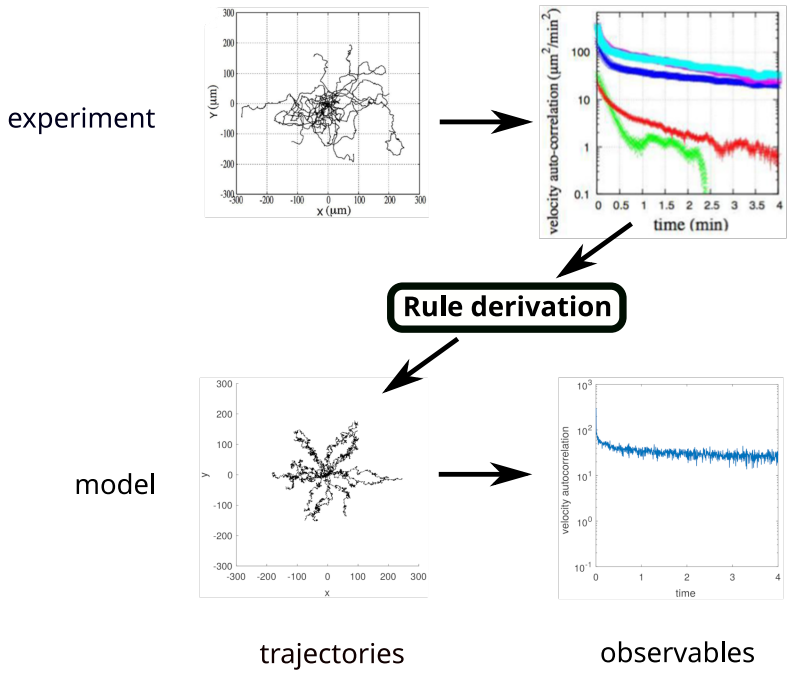

Figure 7: BIO-LGCA modelling with data-driven rules. First, an experiment is performed and data (here: migration trajectories) are gathered. Second, the data is processed and a characteristic observation is selected (here the observation is the velocity autocorrelation function). Third, a BIO-LGCA model is constructed by deriving rules from the experimental observations (see Fig. 4), and data (trajectories) are gathered from simulations. The model can be validated by showing that it reproduces experimental observations (here, the velocity autocorrelation function). Experimental plots of *D. discoideum* trajectories (top) were adapted from [19]. BIO-LGCA simulation plots of single-cell trajectories (bottom) were produced by the authors from a BIO-LGCA model obtained from experimental plots as described within the main text.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To mimick a chemotactic response to the local signal concentration, we define the signal gradient field

$$\mathbf{G}_{\text{sig}}(\mathbf{s}_{\mathcal{N}}) := \sum_{p=1}^{b} \mathbf{c}_p \; c^p_{sig}, \quad \mathbf{s}_{\mathcal{N}} = \left((\mathbf{s}^1, c^1_{sig}), \ldots, (\mathbf{s}^b, c^b_{sig})\right) \in \bar{\mathcal{E}}^{\mathcal{N}}, \tag{21}$$

where $\bar{\mathcal{E}}^{\mathcal{N}} = \mathcal{E} \times \mathbb{R}^+_0$. Chemotaxis can be modeled through a reorientation operator with transition probabilities given by

$$P\left(\mathbf{s} \to \mathbf{s}^{\mathcal{O}} | \mathbf{s}_{\mathcal{N}}; \beta\right) = \frac{1}{Z\left(\mathbf{s}_{\mathcal{N}}, \beta\right)} \exp\left(\beta \, \mathbf{G}_{\text{sig}}(\mathbf{s}_{\mathcal{N}}) \cdot \mathbf{J}(\mathbf{s}^{\mathcal{O}})\right) \delta_{n(\mathbf{s}), n(\mathbf{s}^{\mathcal{O}})}. \tag{22}$$

where β is the *chemotactic sensitivity* of the cells.

With large probability, cells will move in the direction of the external chemical gradient $\mathbf{G}_{\text{sig}}$.

Haptotaxis We consider cell migration in a static environment that conveys directional information expressed by a vector field

$$\mathbf{E} : \mathcal{L} \to \mathbb{R}^2.$$

A biologically relevant example is haptotactic cell motion of cells responding to fixed local concentration differences of adhesion molecules along the extracellular matrix (ECM). In this example, the local spatial concentration differences of integrin ligands in the ECM constitute a gradient field that creates a "drift" $\mathbf{E}$ [20].

The transition probabilities associated to the reorientation operator, given a vector $\mathbf{E} \in \mathbb{R}^2$, is given by

$$P\left(\mathbf{s} \to \mathbf{s}^{\mathcal{O}} | \mathbf{E}\right) = \frac{1}{Z\left(n(\mathbf{s}), \mathbf{E}, \beta\right)} \exp\left(\beta \mathbf{E} \cdot \mathbf{J}\left(\mathbf{s}^{\mathcal{O}}\right)\right) \cdot \delta_{n(\mathbf{s}), n(\mathbf{s}^{\mathcal{O}})},$$

where $\mathbf{E} \in \mathbb{R}^2$.

In this case, cells preferably move in the direction of the external gradient $\mathbf{E}$.

Contact guidance We now focus on cell migration in environments that convey orientational, rather than directional, guidance. Examples of such motion are provided by neutrophil or leukocyte movement through the pores of the ECM, the motion of cells along fibrillar tissues, or the motion of glioma cells along fiber track structures. Such an environment can be represented by a second rank tensor field that encodes the spatial anisotropy along the tissue. In each point, the corresponding tensor informs the cells about the local orientation and strength of the anisotropy and induces a principal (local) axis of movement. Thus, the enviroment can again be represented by a vector field

$$\mathbf{E} : \mathcal{L} \to \mathbb{R}^2.$$

Contact guidance can be modeled through a reorientation operator with transition probabilities defined as

$$P\left(\mathbf{s} \to \mathbf{s}^{\mathcal{O}} | \mathbf{E}\right) = \frac{1}{Z\left(n(\mathbf{s}), \mathbf{E}, \beta\right)} \exp\left(\beta |\mathbf{E} \cdot \mathbf{J}\left(\mathbf{s}^{\mathcal{O}}\right)|\right) \cdot \delta_{n(\mathbf{s}), n(\mathbf{s}^{\mathcal{O}})}.$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

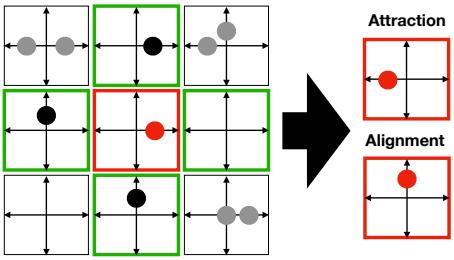

Figure 8: Basic interactions with neighbourhood impact; node configuration (red) before and after application of stochastic interaction rule: cell-cell attraction (top), cell alignment (bottom).

6.2 Collective cell migration

Several kinds of organisms, as well as biological cells, e.g. fibroblasts, can align their velocities globally through local interactions. Here, we introduce a reorientation operator where the local director field is a function of the states of several channels and nodes, reflecting the influence of neighbouring cells during collective cell migration.

$$P\left(\mathbf{s}, \rightarrow \mathbf{s}^{\mathcal{O}} | \mathbf{s}_{\mathcal{N}}\right) = \frac{1}{Z(\mathbf{s}_{\mathcal{N}})} \exp\left(\beta\, \mathbf{D}(\mathbf{s}_{\mathcal{N}}) \cdot \mathbf{J}(\mathbf{s}^{\mathcal{O}})\right) \delta\left(n(\mathbf{s}), n(\mathbf{s}^{\mathcal{O}})\right), \tag{23}$$

where $\mathbf{D}(\mathbf{s}_{\mathcal{N}}) = \sum_{p=1}^{b} \mathbf{J}(s^p)$ is the *local cell momentum* This particular reorientation probability triggers cell alignment [21].

6.3 Attractive interaction

Biological cells can interact via cell-cell adhesion, through filpodia cadherin interaction, for example. Agent attraction/adhesion can be modeled with a reorientation operator with the following probability distribution.

$$P\left(\mathbf{s}, \rightarrow \mathbf{s}^{\mathcal{O}} | \mathbf{s}_{\mathcal{N}}, \beta\right) = \frac{1}{Z(\mathbf{s}_{\mathcal{N}})} \exp\left(\beta\, \mathbf{G}(\mathbf{s}_{\mathcal{N}}) \cdot \mathbf{J}(\mathbf{s}^{\mathcal{O}})\right) \delta\left(n(\mathbf{s}), n(\mathbf{s}^{\mathcal{O}})\right), \tag{24}$$

where $\mathbf{G}(\mathbf{s}_{\mathcal{N}}) = \sum_{p=1}^{K} n(\mathbf{s}^p)\mathbf{c}_p$ is the *density gradient field.*

This reorientation probability favors cell agglomeration. A similar rule has been introduced in [22].

7 Mean-field analysis

We here demonstrate the mean-field analysis of the BIO-LGCA for the example of the attractive interaction eq. (24). This analysis allows to predict collective behaviour. In particular, we calculate the critical sensitivity β_c, such that aggregation occurs for $\beta > \beta_c$, while a homogeneous initial conditions is stable for $\beta < \beta_c$. Under "mean-field" we here understand that we neglect correlations between the occupation numbers of different channels and that we approximate

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the mean value of any function f of a random variable X by the function evaluated at the mean value of the random variable, i.e. $\langle f(X)\rangle \approx f(\langle X\rangle)$. As we are interested in the onset of aggregation from a homogeneous initial state with low density $\bar{\rho} := \frac{1}{|\mathcal{L}|}\sum_{\mathbf{r}\in\mathcal{L}} n(\mathbf{r}) \ll 1$ and weak interaction $\beta \ll 1$ we can linearize the transition probabilities. We further assume that there is at most one cell at each node and therefore only consider single-cell transitions ($n = 1$). For the partition function Z we then obtain

$$Z = \sum_{i=1}^{K} \exp \beta \mathbf{G}(\mathbf{s}_{\mathcal{N}}) \cdot \mathbf{c}_i \approx \sum_{i=1}^{K} 1 + \beta \mathbf{G}(\mathbf{s}_{\mathcal{N}}) \cdot \mathbf{c}_i = K, \tag{25}$$

due to the symmetry of the lattice. For the single-cell transition probability we obtain

$$P(s_i^{\mathcal{O}} = 1|\mathbf{s}_{\mathcal{N}}) \approx \frac{1 + \beta \mathbf{G}(\mathbf{s}_{\mathcal{N}}) \cdot \mathbf{c}_i}{K}. \tag{26}$$

Since the transition probability only depends on the number of cells on the neighbouring nodes, but not on their distribution on the channels, we analyze the mean local density

$$\rho(\mathbf{r}, k) := \langle n(\mathbf{r}, k)\rangle = \left\langle \sum_{i=1}^{K} s_i(\mathbf{r}, k) \right\rangle. \tag{27}$$

According to the propagation rule the cell number $n(r, k+1)$ is given by

$$n(\mathbf{r}, k+1) = \sum_{i=1}^{K} s_i^{\mathcal{O}}(\mathbf{r} - \mathbf{c}_i, k). \tag{28}$$

We calculate the expected value under the mean-field assumption in terms of numbers of cells $n(\mathbf{r})$ at node $\mathbf{r} \in \mathcal{L}$, and the number of cells $\mathbf{n}_{\mathcal{N}(\mathbf{r})}$ in the neighbourhood of $\mathbf{r} \in \mathcal{L}$ as

$$\rho(\mathbf{r}, k+1) = \sum_{i=1}^{K} \sum_{\mathbf{n}_{\mathcal{N}}} P\big(s_i^{\mathcal{O}} = 1|\mathbf{n}_{\mathcal{N}(\mathbf{r}-\mathbf{c}_i)}\big) P\big(\mathbf{n}_{\mathcal{N}(\mathbf{r}-\mathbf{c}_i)}\big). \tag{29}$$

As in the low-density regime $P(n(\mathbf{r}) > 1) \approx 0 \quad \forall \mathbf{r} \in \mathcal{L}$, we can use the single-cell transition probability eq. (26) and the factorizing probability distribution under our mean-field assumption to obtain

$$\begin{aligned}
\rho(\mathbf{r}, k+1) &= \sum_{i=1}^{K} \sum_{\mathbf{n}_{\mathcal{N}}} \frac{1 + \beta \mathbf{c}_i \cdot \sum_{j=1}^{K} \mathbf{c}_j n(\mathbf{r} - \mathbf{c}_i + \mathbf{c}_j, k)}{K} P\big(\mathbf{n}_{\mathcal{N}(\mathbf{r}-\mathbf{c}_i)}\big) \\
&= \frac{1}{K} \sum_{i=1}^{K} \left[1 + \beta \mathbf{c}_i \cdot \sum_{j=1}^{b} \mathbf{c}_j \rho(\mathbf{r} - \mathbf{c}_i + \mathbf{c}_j, k) \right] P(n(\mathbf{r} - \mathbf{c}_i) = 1) \\
&\approx \frac{1}{K} \sum_{i=1}^{K} \left[1 + \beta \mathbf{c}_i \cdot \sum_{j=1}^{b} \mathbf{c}_j \rho(\mathbf{r} - \mathbf{c}_i + \mathbf{c}_j, k) \right] \rho(\mathbf{r} - \mathbf{c}_i, k)
\end{aligned} \tag{30}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To proceed, we assume a one-dimensional lattice, where $b = 2, c_{1,2} = \pm 1$ with a rest channels to obtain the finite-difference equation (FDE)

$$\rho(r,k+1) = \frac{a}{K}\rho(r,k) + \frac{1}{K}\big[\rho(r-1,k) + \rho(r+1,k)\big] + \frac{\beta}{K}\Big\{\rho(r-1,k)\big[\rho(r,k) - \rho(r-2,k)\big] + \rho(r+1,k)\big[\rho(r,k) - \rho(r+2,k)\big]\Big\}. \quad (31)$$

This FDE can be analyzed by means of linear stability analysis. To do this, we first rewrite eq. (31) in terms of the density difference $\Delta\rho(r,k) := \rho(r,k+1) - \rho(r,k)$, which we linearize around the steady state for a small perturbation of the form $\rho(r,k) = \bar{\rho} + \delta\rho(r,k)$

$$\Delta\rho(r,k) \approx \Delta\rho(r,k)\Big|_{\bar{\rho}} + \sum_{r'} \frac{\partial\Delta\rho(r,k)}{\partial\rho(r',k)}\Big|_{\bar{\rho}} \delta\rho(r',k) \quad (32)$$

$$= K^{-1}\{\delta\rho(r,k)[-2 + 2\beta\bar{\rho}] + [\delta\rho(r+1,k) + \delta\rho(r-1,k)] - [\delta\rho(r+2,k) + \delta\rho(r-2,k)]\beta\bar{\rho}\}. \quad (33)$$

We now apply the discrete Fourier transform

$$\delta\tilde{\rho}(q,k) = \sum_{r=0}^{L-1} \delta\rho(r,k) \exp\frac{-2\pi i q r}{L}, \quad (34)$$

and obtain the mode-dependent FDE

$$\Delta\tilde{\rho}(q,k+1) = \frac{2\delta\tilde{\rho}(q,k)}{K}\left\{\beta\bar{\rho} - 1 + \cos\frac{2\pi q}{L} - \beta\bar{\rho}\cos\frac{4\pi q}{L}\right\} \quad (35)$$

$$= \frac{2\delta\tilde{\rho}(q,k)}{K}\left\{2\beta\bar{\rho} - 1 + \cos\frac{2\pi q}{L} - 2\beta\bar{\rho}\cos^2\frac{2\pi q}{L}\right\}, \quad (36)$$

using $2\cos x = e^{ix} + e^{-ix}$ and $\cos 2x = 2\cos^2 x - 1$. Note that the system becomes unstable when the r.h.s. of the equation is larger than 0, meaning the perturbation grows, while it is stable with a decreasing perturbation if it is smaller than 0. To find the dominant Fourier mode q that maximizes the r.h.s. we assume an infinite lattice $L \to \infty$ so that we can define the quasi-continuous wave number $\kappa := \frac{2\pi q}{L}$ and use the derivative with respect to κ to calculate the maxima of the bracket on the r.h.s. of eq. (36),

$$-\sin\kappa + 4\beta\bar{\rho}\cos\kappa\sin\kappa = 0, \quad (37)$$

$$-1 + 4\beta\bar{\rho}\cos\kappa = 0, \quad (38)$$

$$\cos\kappa = \frac{1}{4\beta\bar{\rho}}. \quad (39)$$

Note that we divided by $\sin\kappa$ here, neglecting the trivial solutions $\kappa = 0, \pi$. Clearly the solution $\cos\kappa_c = \frac{1}{4\beta\bar{\rho}}$ is only valid for $\beta\bar{\rho} \geq \frac{1}{4}$ and it is the dominant wave number in this case. This in turn allows us to define the critical parameter combination $(\beta\bar{\rho})_c := \frac{1}{4}$. We can also calculate the dominant wave length in dependence of $\beta\bar{\rho}$ as $\lambda_c = \frac{2\pi}{\kappa_c}$, which diverges at $\beta\bar{\rho} \to (\beta\bar{\rho})_c \Rightarrow \lambda_c \to \infty$ and approaches $\lambda_c \to 4$ for $\beta\bar{\rho} \to \infty$.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

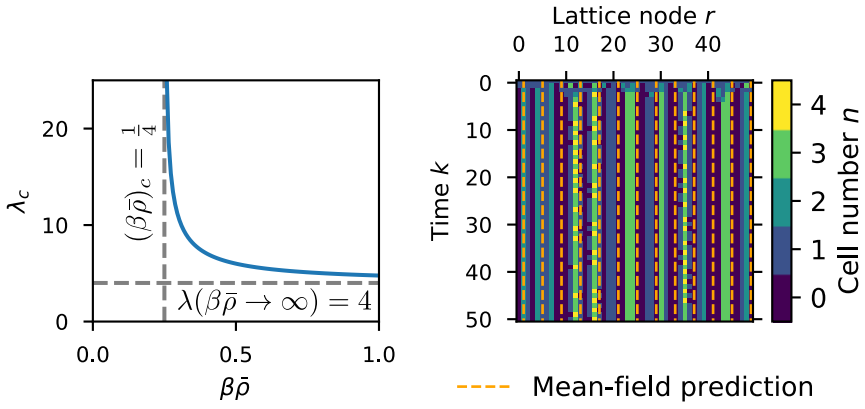

Figure 9: Pattern formation in the LGCA aggregation rule. Left: critical wave length obtained from the mean-field analysis. The critical wave length diverges for $(\beta\bar{\rho}) \to (\beta\bar{\rho})_c = 1/4$, and $\lim_{(\beta\bar{\rho})_c \to \infty} \lambda_c = 4$. Right: emergence of a periodic pattern from a homogeneous initial state. The horizontal distance of the orange dashed lines is equal to the critical wave length predicted by mean-field analysis. Parameters: $\beta = 100, \bar{\rho} = 1, a = 2$.

8 Discussion

In contrast to "continuumsystems" and their canonical description with partial differential equations, there is no standard model for describing interactions of discrete objects, particularly interacting discrete biological cells. In this paper, the BIO-LGCA is proposed as a lattice-based model class for a spatially extended system of interacting cells. The BIO-LGCA idea can be expanded to multispecies models with different cell phenotypes where cell phenotypes may differ in their migration or interaction behaviours reflected in the specific interaction rule (cp. also [23]). It is also possible to extend the BIO-LGCA idea to heterogeneous populations and environments, e.g. cells which differ in their adhesivities and/or which interact with a heterogeneous non-cellular environment [24, 25]. BIO-LGCA models are appropriate for low and moderate cell densities. For higher densities e.g. in epithelial issues, cell shape may matter and other models, such as the Cellular Potts model, may be better choices (see [2, 26] for reviews of on- and off-lattice models). It is also important to be aware of lattice artefacts inherent to the spatial discretization considered in every cellular automaton model, e.g. the checkerboard artefact (cp. [27]). In two spatial dimensions, the hexagonal lattice possesses less artefacts than the square lattice. A major advantage of BIO-LGCA models compared to other on- and off-lattice cell-based models for interacting cell systems, such as interacting particle systems, e.g. [28, 29], asynchronous cellular automata, e.g. [30, 31, 32], further cell-based models [33] or systems of stochastic differential equations [34], is their computational efficiency, and their synchronicity and explicit velocity consideration, which enables the modeling of moderately packed cell collectives while minimizing model artifacts.

Most BIO-LGCA models do not conserve momentum, especially if they

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

model active cells in highly viscous environments and spend energy to bypass momentum conservation. In particular, cell motion may be influenced by the interaction of cells with components of their microenvironment through haptotaxis or differential adhesion, contact guidance, contact inhibition, and processes that involve cellular responses to signals that are propagated over larger distances (e.g. chemotaxis). BIO-LGCA models have already been used in the study of several biological processes, namely angiogenesis [35], bacterial rippling [36], Turing pattern formation [37], active media [38], epidemiology [39] and various aspects of tumor dynamics [24, 40, 41, 42, 43, 44, 45, 46, 47], among others.

BIO-LGCA models allow multiscale analysis of behaviours emerging at multiple temporal and spatial scales. The microscopic scale is much smaller than the typical cell size and is not explicitly considered in BIO-LGCA models. The macroscopic scale is much larger than the cell size and refers to the behaviour of the cell population. Thus, a BIO-LGCA operates at a mesoscopic scale between the microscopic and the macroscopic scale: the mesoscopic scale coarse-grains microscopic properties but distinguishes individual cells. BIO-LGCA as "mesoscopic" models can be regarded either as coarse-grained microscopic models, or discretized macroscopic models. The BIO-LGCA framework facilitates theoretical analysis of emergent, tissue-scale (macroscopic) behaviours [27]. In many cases, the macroscopic behaviour of the mesoscopic BIO-LGCA can be analyzed using reasonable approximations, such as through a spatial mean-field description based on a partial differential equation [27, 48, 49, 50] (see also sec. 7). In particular, BIO-LGCA have been used for analysing collective behaviours at the macroscopic biological tissue level that result from local cellular interactions. Typical examples of observables at a macroscopic scale are cell density patterns and quantities related to the dynamics of moving cell fronts and cluster size distributions [51, 52, 53]. Cell density patterns can often be assessed experimentally and provide, therefore, a means to relate BIO-LGCA model predictions to experimental observations.

Meanwhile, microscopic model descriptions in the form of stochastic differential equations have been derived from a mesoscopic individual-based BIO-LGCA formulation as well [54]. Note that in standard LGCA and BIO-LGCA individual agents can not be distinguished and therefore followed.

The BIO-LGCA modeling strategy is "modular": starting from "basic models", which include those explored in this paper such as alignment, contact guidance/repulsion, hapto- and chemotaxis. Coupling them is required to design models for complex biological problems. The focus of future activities is the analysis of further model combinations for selected biological problems, which are not necessarily restricted to cells but could also comprise interactions at subcellular and tissue scales. The resulting multi-scale models will contain a multitude of coupled spatial and temporal scales and will impose significant challenges for their analytic treatment.

9 Simulator

https://imc.zih.tu-dresden.de//biolgca/
https://github.com/sisyga/BIO-LGCA

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10 Acknowledgments

HH gratefully acknowledges the funding support of the Helmholtz Association of German Research Centers—Initiative and Networking Fund for the project on Reduced Complexity Models (ZT-I-0010). HH is supported by MulticellML (01ZX1707C) of the Federal Ministry of Education and Research (BMBF) and the Volkswagenstiftung within the "Life?" programm (96732). A.D. acknowledges support by the EU-ERACOSYS project no. 031L0139B. The authors thank the Centre for Information Services and High Performance Computing at TU Dresden for providing an excellent infrastructure.

References

[1] F. Bertaux, S. Stoma, D. Drasdo, and G. Batt, "Modeling dynamics of cell-to-cell variability in trail-induced apoptosis explains fractional killing and predicts reversible resistance," *PLoS Comput Biol*, vol. 10, no. 10, p. e1003893, 2014.

[2] P. V. Liedekerke, M. M. Palm, N. Jagiella, and D. Drasdo, "Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results," *Computat. Part. Mechan.*, vol. 2, pp. 401–444, 2015.

[3] M. Radszuweit, M. Block, J. Hengstler, E. Schöll, and D. Drasdo, "Comparing the growth kinetics of cell populations in two and three dimensions," *Physical Review E*, vol. 79, no. 5, p. 051907, 2009.

[4] A. Voss-Boehme and A. Deutsch, "The cellular basis of cell sorting kinetics," *Journal of theoretical biology*, vol. 263, no. 4, pp. 419–436, 2010.

[5] F. Graner and J. A. Glazier, "Simulation of biological cell sorting using a two-dimensional extended potts model," *Physical review letters*, vol. 69, no. 13, p. 2013, 1992.

[6] A. W. Burks, *Essays on Cellular Automata*. University of Illinois Press, Urbana IL, 1970.

[7] J. L. Casti, *Alternate realities*. New York: John Wiley, 1989.

[8] B. Chopard and M. Droz, *Cellular automata modeling of physical systems*. Cambridge University Press, New York, 1998.

[9] S. Wolfram, *A new kind of science*. Wolfram Media, Inc, 2002.

[10] A. Deutsch and S. Dormann, *Cellular automaton modeling of biological pattern formation*. Boston: Birkhauser, 2017.

[11] F. Peruani, T. Klauss, A. Deutsch, and A. Voß-Böhme, "Traffic jams, gliders, and bands in the quest for collective motion of self-propelled particles," *Physical Review Letters*, vol. 106, March 2011.

[12] E. Gavagnin and C. A. Yates, "Modeling persistence of motion in a crowded environment: The diffusive limit of excluding velocity-jump processes," *Physical Review E*, vol. 97, no. 3, p. 032416, 2018.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[13] J. Nava-Sedeño, H. Hatzikirou, F. Peruani, and A. Deutsch, "Extracting cellular automaton rules from physical Langevin equation models for single and collective cell migration," *Journal of mathematical biology*, vol. 75, no. 5, pp. 1075–1100, 2017.

[14] J. M. Nava-Sedeño, H. Hatzikirou, R. Klages, and A. Deutsch, "Cellular automaton models for time-correlated random walks: derivation and analysis," *Scientific reports*, vol. 7, no. 1, pp. 1–13, 2017.

[15] K. Kawasaki, "Simple derivations of generalized linear and nonlinear Langevin equations," *Journal of Physics A: Mathematical, Nuclear and General*, vol. 6, pp. 1289–1295, sep 1973.

[16] F. Peruani, M. Bär, and A. Deutsch, "A mean-field theory for self-propelled particles interacting by velocity alignment mechanisms," *Eur. Phys. J.*, vol. 157, no. 111, 2008.

[17] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, "Active brownian particles: From individual to collective stochastic dynamics," *Europ. Phys. J. Spec. Top.*, vol. 202, no. 1, pp. 1–162, 2012.

[18] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, "Principles of maximum entropy and maximum caliber in statistical physics," *Reviews of Modern Physics*, vol. 85, no. 3, p. 1115, 2013.

[19] H. Takagi, M. J. Sato, T. Yanagida, and M. Ueda, "Functional analysis of spontaneous cell movement under different physiological conditions," *PloS one*, vol. 3, no. 7, p. e2648, 2008.

[20] R. B. Dickinson and R. T. Tranquillo, "A stochastic model for cell random motility and haptotaxis based on adhesion receptor fuctuations," *J. Math. Biol.*, vol. 31, pp. 563–600, 1993.

[21] H. Bussemaker, A. Deutsch, and E. Geigant, "Mean-field analysis of a dynamical phase transition in a cellular automaton model for collective motion," *Phys. Rev. Lett.*, vol. 78, pp. 5018–5021, 1997.

[22] H. J. Bussemaker, "Analysis of a pattern forming lattice-gas automaton: mean-field theory and beyond," *Phys. Rev. E*, vol. 53, no. 2, pp. 1644–1661, 1996.

[23] R. Kapral, A. T. Lawniczak, and P. Masiar, "Reactive dynamics in a multispecies lattice-gas automaton," *J. Chem. Phys.*, vol. 96, p. 2762, 1992.

[24] D. Reher, B. Klink, A. Voss-Boehme, and A. Deutsch, "Cell adhesion heterogeneity reinforces tumour cell dissemination: novel insights from a mathematical model," *Biol. Direct*, 2017.

[25] K. Talkenberger, A. Cavalcanti-Adam, A. Voss-Boehme, and A. Deutsch, "Amoeboid-mesenchymal migration plasticity promotes invasion only in complex heterogeneous microenvironments," *Scientif. Rep.*, 2017.

[26] T. Boekhorst, L. Preziosi, and P. Friedl, "Plasticity of cell migration in vivo and in silico," *Annu. Rev. Cell Dev. Biol.*, vol. 32, pp. 491–526, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[27] A. Deutsch and S. Dormann, *Cellular automaton modeling of biological pattern formation.* Birkhauser. 334 p, 2005, 2017 (2nd ed.).

[28] T. M. Liggett, *Interacting Particle Systems.* New York: Springer, 1985.

[29] A. Voß-Böhme and A. Deutsch, "The cellular basis of cell sorting kinetics," *Journal of Theoretical Biology*, vol. 263, no. 4, pp. 419–436, 2010.

[30] M. Badoual, C. Deroulers, M. Aubert, and B. Grammaticos, "Modelling intercellular communication and its effects on tumour invasion," *Physical biology*, vol. 7, no. 4, p. 046013, 2010.

[31] B. J. Binder, K. A. Landman, D. F. Newgreen, J. E. Simkin, Y. Takahashi, and D. Zhang, "Spatial analysis of multi-species exclusion processes: application to neural crest cell migration in the embryonic gut," *Bulletin of mathematical biology*, vol. 74, no. 2, pp. 474–490, 2012.

[32] J. Bloomfield, J. Sherratt, K. Painter, and G. Landini, "Cellular automata and integrodifferential equation models for cell renewal in mosaic tissues," *Journal of the Royal Society Interface*, vol. 7, no. 52, pp. 1525–1535, 2010.

[33] J. Galle, M. Hoffmann, and G. Aust, "From single cells to tissue architecture—a bottom-up approach to modelling the spatio-temporal organisation of complex multi-cellular systems," *Journal of mathematical biology*, vol. 58, no. 1-2, p. 261, 2009.

[34] K. A. Rejniak and A. R. Anderson, "Hybrid models of tumor growth," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 3, no. 1, pp. 115–125, 2011.

[35] C. Mente, I. Prade, L. Brusch, G. Breier, and A. Deutsch, "Parameter estimation with a novel gradient-based optimization method for biological lattice-gas cellular automaton models," *J. Math. Biol.*, vol. 63, no. 1, pp. 173–200, 2010.

[36] M. Alber, M. Kiskowski, and Y. Jiang, "Lattice gas cellular automata model for rippling in myxobacteria," *Physica D*, vol. 191, p. 343, 2004.

[37] S. Dormann, A. Deutsch, and A. T. Lawniczak, "Fourier analysis of Turing-like pattern formation in cellular automaton models," *Fut. Comput. Gener. Syst.*, vol. 17, pp. 901–909, 2001.

[38] S. Syga, J. M. Nava-Sedeño, L. Brusch, and A. Deutsch, "A lattice-gas cellular automaton model for discrete excitable media," in *Spirals and Vortices*, pp. 253–264, Springer, 2019.

[39] H. Fuks and A. T. Lawniczak, "Individual-based lattice model for spatial spread of epidemics," *Discrete Dynamics in Nature and Society*, vol. 6, 2001.

[40] J. Moreira and A. Deutsch, "Cellular automaton models of tumour development – a critical review," *Adv. Compl. Syst. (ACS)*, vol. 5, no. 2, pp. 1–21, 2002.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[41] H. Hatzikirou, K. Böttger, and A. Deutsch, "Model-based comparison of cell density-dependent cell migration strategies," *Mathematical Modelling Natural Phenomena*, vol. 10, no. 1, pp. 94–107, 2015.

[42] K. Böttger, H. Hatzikirou, A. Voß-Böhme, E. A. Cavalcanti-Adam, M. A. Herrero, and A. Deutsch, "An emerging Allee effect is critical for tumor initiation and persistence," *PLoS Comput. Biol.*, vol. 11, no. 9, p. e1004366, 2015.

[43] S. Dormann and A. Deutsch, "Modeling of self-organized avascular tumor growth with a hybrid cellular automaton," *In Silico Biol.*, vol. 2, p. 0035, 2002.

[44] H. Hatzikirou, L. Brusch, C. Schaller, M. Simon, and A. Deutsch, "Prediction of traveling front behavior in a lattice-gas cellular automaton model for tumor invasion," *Comput. Mathem. Applic.*, vol. 59, pp. 2326–2339, April 2010.

[45] M. Tektonidis, H. Hatzikirou, A. Chauviere, M. Simon, K. Schaller, and A. Deutsch, "Identification of intrinsic in vitro cellular mechanisms for glioma invasion," *J. Theor. Biol.*, vol. 287, pp. 131–147, October 2011.

[46] C. Mente and A. Deutsch, "A cellular automaton model to investigate emergent behavior of heterogeneous cell populations," in *Proceedings of the International Conference on Scientific Computing (CSC)*, p. 1, The Steering Committee of The World Congress in Computer Science, Computer ..., 2012.

[47] O. Ilina, P. Gritsenko, S. Syga, A. Deutsch, and P. Friedl, "Cell–cell adhesion and 3d matrix confinement determine jamming transitions in breast cancer invasion," *Nat Cell Biol*, vol. 22, p. 1103–1115, 2020.

[48] U. Frisch, B. Hasslacher, and Y. Pomeau, "Lattice-gas automata for the navier-stokes equation," *Phys. Rev. Lett.*, vol. 56, no. 14, pp. 1505–1508, 1986.

[49] D. Wolf-Gladrow, *Lattice-gas cellular automata and lattice Boltzmann models: an introduction*. New York: Springer, 2000.

[50] S. Wolfram, *Cellular Automata and Complexity - collected papers*. Addison-Wesley, 1994.

[51] K. Böttger, H. Hatzikirou, A. Chauviere, and A. Deutsch, "Investigation of the migration/proliferation dichotomy and its impact on avascular glioma invasion," *Mathem. Model. Natur. Phenom.*, vol. 7, no. 1, pp. 105–135, 2012.

[52] H. Hatzikirou, L. Brusch, and A. Deutsch, "From cellular automaton rules to an effective macroscopic mean-field description," *Acta Phys. Polon. B Proc. Suppl.*, vol. 3, pp. 399–416, 2010.

[53] C. Mente, I. Prade, L. Brusch, G. Breier, and A. Deutsch, "A lattice-gas cellular automaton model for in vitro sprouting angiogenesis," *Acta Phys. Pol. B*, vol. 5, no. 1, pp. 99–115, 2012.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.29.360669; this version posted October 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[54] C. Mente, A. Voß-Böhme, and A. Deutsch, "Analysis of individual cell trajectories in lattice-gas cellular automaton models for migrating cell populations," *Bull. Mathem. Biol.*, vol. 77, no. 4, pp. 1–38, 2015.