

Article

Model-Based Prediction of an Effective Adhesion Parameter Guiding Multi-Type Cell Segregation

Philipp Rossbach ^{1,2}, Hans-Joachim Böhme ^{1,2} , Steffen Lange ^{1,2}  and Anja Voss-Böhme ^{1,2,*} 

¹ DataMedAssist, HTW Dresden, 01062 Dresden, Germany; philipp.rossbach2@htw-dresden.de (P.R.); hans-joachim.boehme@htw-dresden.de (H.-J.B.); steffen.lange@htw-dresden.de (S.L.)

² Faculty of Informatics/Mathematics, HTW Dresden—University of Applied Sciences, 01062 Dresden, Germany

* Correspondence: anja.voss-boehme@htw-dresden.de

Abstract: The process of cell-sorting is essential for development and maintenance of tissues. Mathematical modeling can provide the means to analyze the consequences of different hypotheses about the underlying mechanisms. With the Differential Adhesion Hypothesis, Steinberg proposed that cell-sorting is determined by quantitative differences in cell-type-specific intercellular adhesion strengths. An implementation of the Differential Adhesion Hypothesis is the Differential Migration Model by Voss-Böhme and Deutsch. There, an effective adhesion parameter was derived analytically for systems with two cell types, which predicts the asymptotic sorting pattern. However, the existence and form of such a parameter for more than two cell types is unclear. Here, we generalize analytically the concept of an effective adhesion parameter to three and more cell types and demonstrate its existence numerically for three cell types based on in silico time-series data that is produced by a cellular-automaton implementation of the Differential Migration Model. Additionally, we classify the segregation behavior using statistical learning methods and show that the estimated effective adhesion parameter for three cell types matches our analytical prediction. Finally, we demonstrate that the effective adhesion parameter can resolve a recent dispute about the impact of interfacial adhesion, cortical tension and heterotypic repulsion on cell segregation.

Keywords: differential adhesion hypothesis; differential migration model; cell sorting; cellular automaton; statistical learning methods; high heterotypic interfacial tension hypothesis; pattern formation



Citation: Rossbach, P.; Böhme, H.-J.; Lange, S.; Voss-Böhme, A. Model-Based Prediction of an Effective Adhesion Parameter Guiding Multi-Type Cell Segregation. *Entropy* **2021**, *23*, 1378. <https://doi.org/10.3390/e23111378>

Academic Editor: Haralambos Hatzikirou

Received: 6 September 2021
Accepted: 14 October 2021
Published: 21 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

During the development of multicellular organisms, mechanical forces affect the internal states of cells as well as the interaction between cells, and they are, therefore, an integral part of all morphogenetic processes [1]. These forces are typically driven by molecular motors and transmitted via cytoskeleton elements and adhesion molecules within the cells and between them. Together, molecular motor driven movement and force transmission via adhesion complexes constitute two major self-organizing phenomena that drive tissue morphogenesis.

Since it constitutes a paradigmatic process for tissue morphogenesis, cell-sorting has found much attention both from experimental and theoretical research. It describes the tissue-scale segregation phenomenon that is observed when mixed heterotypic cell populations unmix into spatially confined homotypic cell clusters [2,3]. Over many decades this process has been studied in a number of in-vitro cell-sorting experiments for many different cell types [2–7]. To resolve which types of intercellular interaction guide individual mobile cells to find their homotypic neighbors, several theoretical hypotheses have been proposed and studied. Steinberg put forward the Differential Adhesion Hypothesis (DAH) which exploits the similarity of cell segregation with the unmixing of immiscible fluids such as water and oil. It states that cell-sorting is driven by the minimization of tissue surface tensions which result from quantitative differences in the strengths of cell-type

specific intercellular adhesion [8]. However, Harris (1976) [9] remarked that tissue surface tension can result from several cellular mechanisms besides differential adhesion and suggested the differential contractility of cells as a major driver of cell-sorting. This is also acknowledged by the Differential Interfacial Tension Hypothesis (DITH) of Brodland and Chen (2000) [10]. There, the effects of intercellular adhesion and cellular contractility at cell–cell contacts are subsumed to the concept of differential interfacial tension at homo- and heterotypic cell contacts which determines the degree of mutual attachment between cells and thus drives cell-sorting. Canty et al. (2017) [11] demonstrated experimentally that repulsion at heterotypic cell–cell contacts with little to no contribution from adhesive or tensile differences between cell types can lead to cell-sorting and tissue separation. They propose the High Heterotypic Interfacial Tension Hypothesis (HIT) which asserts that heterotypic repulsion creates a situation where tension is strongly increased at heterotypic contacts compared to homotypic contacts and constitutes the major driver of cell-sorting and tissue separation.

The competing explanatory models DAH, DITH and HIT are unified within the Differential Migration Model (DMM) of Voss–Böhme and Deutsch (2010) [12]. They established that all the above hypotheses have in common that cell–cell contacts affect type-specifically the migration properties of the involved cells and parametrized this effect, independent of its specific nature - adhesion or repulsion, cortical tension or even signaling cascades. For the segregation of two cell types, the existence and form of an effective adhesion parameter (EAP) has been derived mathematically, which predicts the asymptotic sorting behavior [12].

However, while most previous studies focus on the simplest case of two cell types [7,13,14], the case of three and more cell types is more relevant in real biological systems. Studies of cell segregation and pattern formation with three or more cell types are still rare [15]. Although the DMM can be directly generalized to more than two cell types, it has only been studied for the minimal case of two cell types. In particular, the existence and form of the EAP for systems with three or more cell types remains an open question.

We study numerically the impact of the EAP and related parameters on the degree of tissue separation in terms of clear tissue boundaries as well as the separation time scale for mixtures of two cell types. We present new arguments to analytically predict the form of the EAP, thus generalizing the concept of the EAP to mixtures of arbitrary many cell types. We show numerically for three cell types that the analytically predicted EAP guides the asymptotic sorting behavior in the DMM, i.e., the EAP determines the asymptotic value of the normalized sum of heterotypic contacts. The form of the EAP on the asymptotic segregation behavior is independently quantified via two statistical learning methods, Support Vector Machines (SVM) and Logistic Regression model (Logit) [16]. The form estimated by the statistical learning methods matches our analytical prediction both for two and three cell types. Finally, we demonstrate that the EAP can explain different segregation behaviors in two cell-type Cellular Potts models which was previously attributed to the competing hypotheses DAH/DITH and HIT [11].

2. Materials and Methods

2.1. Differential Migration Model

The DMM is a special probabilistic cellular automaton (PCA) based on Voss–Böhme and Deutsch [12], which has been analyzed for two cell types. Here, we present the key features of this model for the case of arbitrary many cell types:

Lattice: The PCA works on a squared lattice $S := \{0, 1, \dots, L\}^2$ with periodic boundaries and we set $L = 25$ throughout our numerical analysis.

Cells: To emulate biological cells, every lattice site $x \in S$ is assigned a cell type $w \in W$ of all possible cell types $W := \{0, 1, \dots\}$ in the system. In this way each cell regardless of its type, has approximately the same size and occupies exactly one lattice site. The number of cells of each type remains constant during the sorting process.

Configurations: A configuration η_t of the lattice S represents the state of the model at time t . It belongs to the set of all possible configurations $X := W^S = \{\eta : S \rightarrow W\}$.

Migration: A configuration η is changed by a cell position switch $\eta \rightarrow \eta_{xy}$ involving two adjacent lattice sites x and y with $x, y \in S$:

$$\eta_{xy}(z) := \begin{cases} \eta(z), & z \neq x \text{ and } z \neq y \\ \eta(y), & z = x \\ \eta(x), & z = y. \end{cases} \tag{1}$$

Position switches between two cells of the same type do not change the configuration and are therefore neglected. Thus, neighboring cells situated at lattice sites x and y only switch their positions if they are heterotypic, i.e., $\eta(x) \neq \eta(y)$.

Differential adhesion: We assume that stronger bonds to neighboring cells hinder cell motility. Accordingly, a cell switch $\eta \rightarrow \eta_{xy}$ occurs with rate $c(x, y, \eta)$. The rate $c(x, y, \eta)$ depends on the parameters $\beta_{\eta(x)\eta(z)}$ and $\beta_{\eta(y)\eta(z)}$, which represent the binding strengths between cells at lattice sites x and y to positions from the von-Neumann-1 neighborhoods $\mathfrak{N}(x)$ and $\mathfrak{N}(y)$, see Figure 1. The von-Neumann-1 neighborhood of a lattice site corresponds to all neighboring lattice sites with Manhattan distance one. The cell switch rate of the two adjacent lattice sites $x, y \in S$ with $x \in \mathfrak{N}(y)$ and $y \in \mathfrak{N}(x)$ is as follows:

$$c(x, y, \eta) := \begin{cases} \exp\left(-\sum_{z \in \mathfrak{N}(x)} \beta_{\eta(x)\eta(z)} - \sum_{z \in \mathfrak{N}(y)} \beta_{\eta(y)\eta(z)}\right), & \eta(x) \neq \eta(y) \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

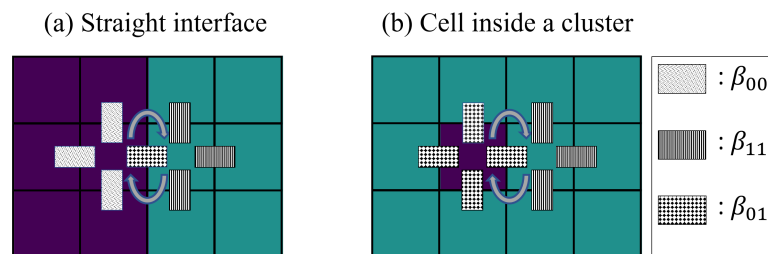


Figure 1. Example of a cell switch. **(a)** A cell switch at a straight interface between two clusters of type $w = 0$ (purple) and $w = 1$ (turquoise) has rate $c = \exp[-3(\beta_{00} + \beta_{11}) - 2\beta_{01}]$ according to Equation (2). **(b)** A single cell of type $w = 0$ inside a cluster of cells of type $w = 1$ switches with rate $c = \exp[-3\beta_{11} - 5\beta_{01}]$. The cell switch between two adjacent cells is highlighted (gray double arrows).

Notice that the particular mechanism which affects the cellular motility on intercellular contact is not specified. The parameters β_{ij} can be interpreted as repulsion, if $\beta_{ij} < 0$, enhancing cellular motility, and it can be interpreted as binding strength resulting from the interplay of adhesion and relaxed cortical tension at cell–cell contact, if $\beta_{ij} > 0$, inhibiting cellular motility. The details on the numerical implementation of this PCA are elaborated in the Appendix A. Depending on how many cell types are considered the number of intercellular adhesion parameters varies. For this, the vector $\beta_{(N)}$ of all intercellular adhesion parameters occurring in a system with N cell types is introduced, for instance

$$\begin{aligned} \beta_{(N=2)} &:= (\beta_{00}, \beta_{11}, \beta_{01})^T \text{ for 2-cell-type systems and} \\ \beta_{(N=3)} &:= (\beta_{00}, \beta_{11}, \beta_{22}, \beta_{01}, \beta_{02}, \beta_{12})^T \text{ for 3-cell-type systems.} \end{aligned}$$

For a 2-cell-type model, an effective adhesion parameter (EAP)

$$\beta_{(N=2)}^* := \beta_{00} + \beta_{11} - 2\beta_{01}$$

was analytically predicted in Voss-Böhme and Deutsch [12]. The EAP determines the asymptotic cell segregation behavior of systems with $N = 2$ cell types regardless of the value combinations of $\beta_{(2)}$. The higher the EAP the more segregated the cell types become.

For simulation of the DMM model a Gillespie-related algorithm is used where independent cell switches, see Equation (1), are chosen and the waiting times between these events are calculated based on the cell switch rates, see Equation (2), within lattice configuration η . For details about this algorithm, see Appendix A, Algorithm A1. Each numerical simulation yields a time-series of lattice configurations η_t , since every cell switch, i.e., whenever two neighboring cells switch their position, see Equation (1), results in a new configuration. The variable t refers to the number of cell switches performed.

2.2. Order Indicator

To quantify the level of cell segregation an order indicator is introduced that quantifies the level of cell segregation of a lattice configuration η . The order indicator $\omega(\eta)$ is the normalized sum of homotypic cell–cell contacts within a configuration η . This indicator increases the more cells of the same type are clustered together. In detail, the set of all homotypic von-Neumann-1 neighbors of a cell of type $\eta(x)$ at position $x \in S$ is given by

$$H(x, \eta) := \{z \in S \mid \eta(x) = \eta(z), z \in \mathfrak{N}(x)\}.$$

With $n_=(x, \eta) := |H(x, \eta)|$, the total sum of homotypic connections is defined as

$$d(\eta) := \frac{\sum_{x \in S} n_=(x, \eta)}{2}. \quad (3)$$

The highest amount d_{max} and the lowest amount d_{min} of homotypic lattice site connections are reached in the case of a fully sorted and chessboard-patterned configuration, respectively. The exact values for 2- and 3-cell-type systems are listed in Appendix B. With d_{max} and d_{min} , the order indicator value $\omega(\eta)$ for configuration η is defined as the normalization of $d(\eta)$:

$$\omega(\eta) := \frac{d(\eta) - d_{min}}{d_{max} - d_{min}}$$

Thus, $\omega(\eta)$ takes values in $[0, 1]$, where configurations with almost perfect sorted patterns, such as in Figure 2I, yield $\omega(\eta)$ values above 0.95. In the case of chessboard configurations as in Figure 2III, the order indicator values are below 0.05. For a random lattice configuration, the order indicator value is $\omega(\eta) \approx 0.5$, in the case of a 2-cell-type system, and it is $\omega(\eta) \approx 0.33$ for a 3-cell-type system.

With the order indicator $\omega(\eta)$, a time-series of configurations η_t can be converted into a time-series of order indicator values $\omega(\eta_t)$, in short $\omega(t)$. Furthermore, the asymptotic cell segregation level for a simulation is estimated as the average over the last 10% of the time-series $\omega(t)$ and is denoted $\bar{\omega}$, see Figure 2 for an illustration.

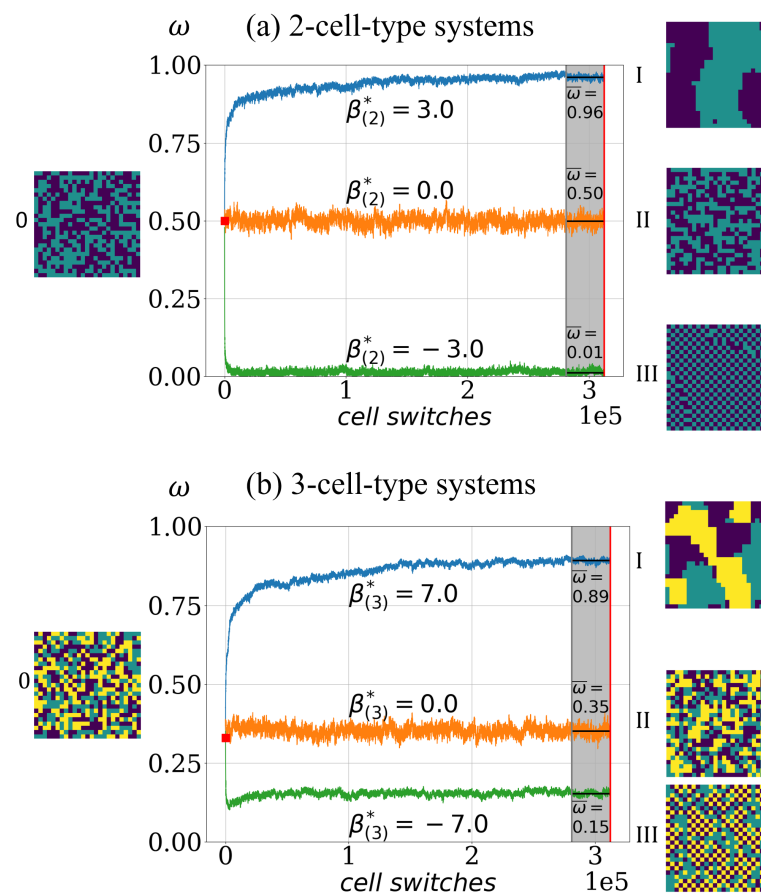


Figure 2. Illustration of the principal cell-sorting behavior in the model. Depending on the effective adhesion parameter the model exhibits asymptotically one of three basic cell-sorting states. (a) Simulated order indicator $\omega(t)$ in the case of 2-cell-type system for three different sets of intercellular adhesion parameters $\beta_{(2)}$: (1.2, -0.2, -1.0) (blue), (1.1, -1.1, 0.0) (orange), (-1.3, 0.3, 1.0) (green) for three different cell-type systems, each corresponding to a distinct asymptotic segregation: a cell population with sorted patterns as in configuration I (blue), a not fully cell-type segregated population as in II (orange) and a cell type mixture population with chessboard pattern as in III (green). Configuration 0 illustrates the random start configuration $\eta_{t=0}$ with an initial order indicator value $\omega(t=0) \approx 0.5$ (marked as red square at $t=0$). The simulation termination condition is $t = 312499$ cell switches (red vertical line). The asymptotic value $\bar{\omega}$ (indicated for each case as black horizontal line at the end) is computed as the average over the last 10% of cell switches (gray area). (b) Analogous for examples of 3-cell-type systems, starting with random configuration, i.e., initial order indicator value $\omega(t=0) \approx 0.33$. The three different sets of intercellular adhesion parameters $\beta_{(3)}$ are: (0.6, 0.6, 2.3, -0.9, -1.1, -1.5) (blue), (-0.8, -0.1, 1.0, -0.1, -0.2, 0.4) (orange) and (-0.6, -0.6, -2.3, 0.6, 2.4, 0.5) (green).

3. Results

3.1. Cell System Parameters for Two Cell Types

We present a new simple argument for the form of the effective adhesion parameter $\beta_{(2)}^*$ guiding the asymptotic behavior in the DMM for two cell types [12] to generalize it to systems with arbitrary many cell types. For this, we introduce two additional system parameters $\beta_{(2)}^s$ and $\beta_{(2)}^\Delta$ besides $\beta_{(2)}^*$, which determine how changes to the intercellular adhesion parameters $\beta_{(2)}$ affect the asymptotic cell segregation behavior as well as the dynamics. Although the parameters $\beta_{(2)}^s$ and $\beta_{(2)}^*$ set the temporal scale of the waiting times and the asymptotic level of segregation, respectively, we argue that the parame-

ter $\beta_{(2)}^\Delta$ determines the number of cell switches required to reach the asymptotic level of segregation.

Indeed, the addition of a constant θ to all intercellular adhesion parameters simultaneously like

$$\beta'_{(2)} = \beta_{(2)} + \theta \mathbf{a}_s \text{ with } \mathbf{a}_s = (1, 1, 1)^T \tag{4}$$

rescales the cell switch rates $c_\beta(x, y, \eta)$ in the DMM by a factor $\exp(-8\theta)$, since

$$\begin{aligned} c_{\beta'_{(2)}}(x, y, \eta) &= \exp\left(-\sum_{z \in \mathfrak{N}(x)} (\beta_{\eta(x)\eta(z)} + \theta) - \sum_{z \in \mathfrak{N}(y)} (\beta_{\eta(y)\eta(z)} + \theta)\right) \\ &= \exp(-8\theta) \cdot c_{\beta_{(2)}}(x, y, \eta), \end{aligned} \tag{5}$$

independent of the actual configuration $\eta \in X$ and the considered lattice sites $x, y \in S$. This factor does not alter the relation between the cell switch rates, but only rescales all waiting times between cell switches by $\exp(-8\theta)$. Thus, the spatio-temporal order of the cell sorting dynamics and the asymptotic behavior are not affected. In order to parametrize this rescaling, a scaling parameter $\beta_{(2)}^s$

$$\beta_{(2)}^s = \beta_{00} + \beta_{11} + \beta_{01} = \langle \beta_{(2)}, \mathbf{a}_s \rangle \tag{6}$$

is introduced, where the functional, which maps $\beta_{(2)} \rightarrow \beta_{(2)}^s$, is given as a scalar product with the vector $\mathbf{a}_s = (1, 1, 1)^T$.

Secondly, we notice that the choice which cell type is denoted by type $w = 0$ and which one by $w = 1$ is arbitrary and the dynamics in the model is invariant against relabeling of the cell types. This invariance is reflected in the cell switch rates $c(x, y, \eta)$ as well as in the order indicator ω , which quantifies the level of segregation of a configuration η as sum of homotypic lattice site connections. Accordingly, the functional which maps the intercellular adhesion parameters $\beta_{(2)}$ to the effective adhesion parameter $\beta_{(2)}^*$ must reflect this invariance as well. This implies that an effective adhesion parameter $\beta_{(2)}^*$, which controls the asymptotic sorting behavior, assuming it exists and is determined by a linear functional on the intercellular adhesion parameters, has the form

$$\beta_{(2)}^* = \mathbf{a} \left(\beta_{00} + \beta_{11} \right) + \mathbf{b} \beta_{01} = \langle \beta_{(2)}, \mathbf{a}_* \rangle \tag{7}$$

with two real-valued constants \mathbf{a}, \mathbf{b} and a vector $\mathbf{a}_* = (\mathbf{a}, \mathbf{a}, \mathbf{b})^T$ introduced analogously to Equation (6) to express the functional. The constants \mathbf{a}, \mathbf{b} are set by the condition that the asymptotic level of segregation and thus the differential adhesion parameter $\beta_{(2)}^*$ must not be affected by the temporal scaling parameter $\beta_{(2)}^s$, which implies

$$\langle \mathbf{a}_*, \mathbf{a}_s \rangle = 0 \quad \Rightarrow \quad 2\mathbf{a} = -\mathbf{b}. \tag{8}$$

Without loss of generality, we can neglect an additional offset and factor in Equation (7) and choose $\mathbf{a} = 1$ for which follows

$$\begin{aligned} \beta_{(2)}^* &= \beta_{00} + \beta_{11} - 2\beta_{01} = \langle \beta_{(2)}, \mathbf{a}_* \rangle \\ \mathbf{a}_* &= (1, 1, -2)^T. \end{aligned} \tag{9}$$

We present an additional heuristic argument to make the assumption of a linear dependency in Equation (7) plausible and to estimate that the critical value of the effective adhesion parameter, at which the system remains randomly mixed, should be zero: Consider two limit scenarios of a cell switch, first two cells at a straight interface between clusters and secondly a single cell of one type moving inside a cluster of opposite type. As illustrated in Figure 1a, two cells of different type at the interface switch their position and

thus cause a less segregated configuration with a rate $c_{\text{mix}} = \exp[-3(\beta_{00} + \beta_{11}) - 2\beta_{01}]$. In contrast, a cell of type $w = 0$ inside a cluster of type $w = 1$ switches position with a given neighbor with rate $c_{\text{unmix},0} = \exp[-3\beta_{11} - 5\beta_{01}]$, see Figure 1b. Analogously, for a single cell of type $w = 1$ in a cluster of type $w = 1$ the rate is $c_{\text{unmix},1} = \exp[-3\beta_{00} - 5\beta_{01}]$. Both types of switches are necessary to move a single cell of opposite type out of a cluster. Since the former switch with rate c_{mix} reduces the level of segregation while the latter two switches with rates $c_{\text{unmix},0}$, $c_{\text{unmix},1}$ are required to increase segregation, the ratio between these types of switches $c_{\text{unmix},0}/c_{\text{mix}} \cdot c_{\text{unmix},1}/c_{\text{mix}} = \exp[3(\beta_{00} + \beta_{11} - 2\beta_{01})]$ should determine the level of the asymptotic segregation. In particular, the point at which the system remains mixed is expected where mixing and unmixing rates are equal, i.e., where the ratio of the rates becomes 1. Since the exponential function is monotonous, it is sufficient to focus on the exponent, which is a multiple of the effective adhesion parameter $\beta_{(2)}^* = \beta_{00} + \beta_{11} - 2\beta_{01}$. Thus, the critical value 1 of the ratio of rates translates into a critical parameter $\beta_{(2)}^* = 0$ for which a two-cell-type system is expected to remain mixed.

After establishing the system parameters $\beta_{(2)}^s$ and $\beta_{(2)}^*$, which are defined via scalar products with the corresponding vectors $\mathbf{a}_s, \mathbf{a}_*$, we point out that there is a third system parameter $\beta_{(2)}^\Delta$ defined by the vector perpendicular to \mathbf{a}_s and \mathbf{a}_*

$$\beta_{(2)}^\Delta = |\beta_{00} - \beta_{11}| = |\langle \boldsymbol{\beta}_{(2)}, \mathbf{a}_\Delta \rangle| \tag{10}$$

$$\mathbf{a}_\Delta = (1, -1, 0)^T \sim \mathbf{a}_s \times \mathbf{a}_* .$$

The parameter $\beta_{(2)}^\Delta$ quantifies the difference between the homotypic adhesion strengths of the two cell types. It becomes zero if both types have the same strength of homotypic adhesion. We use the absolute value in the definition of the parameter $\beta_{(2)}^\Delta$ as we are not interested in which cell type has the stronger homotypic adhesion. We find numerically that the parameter $\beta_{(2)}^\Delta$ affects how many cell switches are required to reach the asymptotic level of segregation, see Figure 3. For a fixed effective adhesion parameter $\beta_{(2)}^* = 3$, which means for the same asymptotic level of cell segregation, Figure 3 shows the number of cell switches required to reach certain thresholds of cell segregation, here $\omega \geq 0.7, \geq 0.8$ and ≥ 0.9 . This number increases with increasing $\beta_{(2)}^\Delta$ independently of the chosen thresholds.

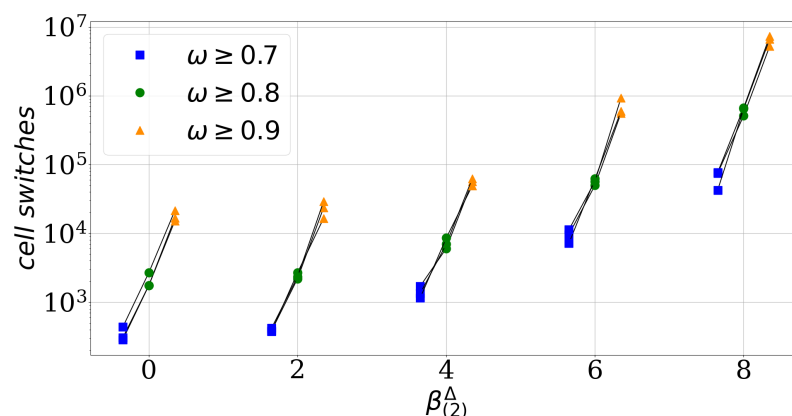


Figure 3. The convergence parameter slows down the sorting process. For each $\beta_{(2)}^\Delta$ value a different intercellular adhesion parameter $\boldsymbol{\beta}_{(2)}$ with $\beta_{(2)}^* = 3.0$ is used and every simulation is repeated three times, see black lines. When a simulation reached an order indicator value $\omega \geq 0.7, \geq 0.8$ and ≥ 0.9 for the first time the corresponding cell switch number is marked with a blue square, green dot and orange triangle. Each simulation has a random start configuration $\eta_{t=0}$ with an initial order indicator value $\omega(\eta_{t=0}) \approx 0.5$ and the intercellular adhesion parameters for each convergence speed parameter $\beta_{(2)}^\Delta$ are: (0.5, 0.5, -1) for 0, (-0.5, 1.5, -1) for 2, (-1.5, 2.5, -1) for 4, (3.5, -2.5, -1) for 6 and (-3.5, 4.5, -1) for 8.

This numerical observation is supported by a heuristic argument: Consider again the limit scenarios of a cell switch, i.e., two cells at a straight interface between clusters which switches with a particular neighbor with rate $c_{\text{mix}} = \exp[-3(\beta_{00} + \beta_{11}) - 2\beta_{01}]$ and a single cell of one type within a cluster of opposite type which switches with rate $c_{\text{unmix},0} = \exp[-3\beta_{11} - 5\beta_{01}]$, or $c_{\text{unmix},1} = \exp[-3\beta_{00} - 5\beta_{01}]$ respectively. In the symmetric case of equal homotypic adhesion parameters $\beta_{00} = \beta_{11}$, which implies $\beta_{(2)}^\Delta = 0$ and $c_{\text{unmix},0} = c_{\text{unmix},1}$, cell switches of the two single cells in either cluster are equally frequent and these switches occur in a certain ratio with cell switches at the interface. If the symmetry is broken by increasing $\beta_{(2)}^\Delta$ at constant $\beta_{(2)}^*$ and $\beta_{(2)}^s$, for instance $\beta_{00} > \beta_{11}$, implying that $c_{\text{unmix},0} > c_{\text{unmix},1}$, the single cell of type $w = 0$ is favored to leave the cluster of opposite type in comparison to the symmetric case and in particular compared to the other single cell of type $w = 1$. Thus, segregation is now limited by the slower process of the single cell of type $w = 1$ moving out of clusters of type $w = 0$. Moreover, the rate c_{mix} of a cell switch at the interface is constant if only the parameter $\beta_{(2)}^\Delta$ is altered. Thus, in between two switches of a single cell of type $w = 1$ (rate $c_{\text{unmix},1}$) out of a type $w = 0$ cluster there are more cell switches at the interface (rate c_{mix}) than in the symmetric case. Please note that these back-and-forth switches at the interface do not progress the segregation but increase the number of total cell switches. This scenario illustrates how an increase of the parameter $\beta_{(2)}^\Delta$ increases the number of cell switches required to reach the asymptotic level of segregation. Thus, we denote $\beta_{(2)}^\Delta$ as convergence speed parameter.

Please note that the convergence speed parameter $\beta_{(2)}^\Delta$ in Equation (10) is proportional to the Euclidean distance between $\beta_{(2)}$ and the related symmetric adhesion parameters $\beta_{(2)}^{\text{sym}}$ with the same values of $\beta_{(2)}^*$ and $\beta_{(2)}^s$. Indeed, for a given vector of adhesion parameters $\beta_{(2)} = (\beta_{00}, \beta_{11}, \beta_{01})^T$ with system parameters $\beta_{(2)}^s, \beta_{(2)}^*$, and $\beta_{(2)}^\Delta$, the corresponding symmetric adhesion parameters $\beta_{(2)}^{\text{sym}} = (\frac{1}{2}[\beta_{00} + \beta_{11}], \frac{1}{2}[\beta_{00} + \beta_{11}], \beta_{01})$ have the same system parameters $\beta_{(2)}^s$ and $\beta_{(2)}^*$ but a convergence speed parameter of zero. The Euclid distance of the adhesion parameters $\beta_{(2)}$ to this symmetric case $\beta_{(2)}^{\text{sym}}$ is

$$\|\beta_{(2)} - \beta_{(2)}^{\text{sym}}\| = \sqrt{\frac{1}{2}}|\beta_{00} - \beta_{11}| \sim \beta_{(2)}^\Delta. \tag{11}$$

Since the convergence speed parameter is zero, the symmetric case $\beta_{(2)}^{\text{sym}}$ leads to the fastest convergence to the asymptotic level of segregation for the given system parameters $\beta_{(2)}^s, \beta_{(2)}^*$. In contrast to the definition of the convergence speed parameter $\beta_{(2)}^\Delta$ in Equation (10), the definition of $\beta_{(2)}^\Delta$ via Equation (11) can be generalized to N cell types, see below.

3.2. The Effective Adhesion Parameter for Arbitrary Number of Cell Types

For a mix of more than two cell types neither the form nor the existence of an effective adhesion parameter are known. Assuming that for an arbitrary number N of cell types an effective adhesion parameter $\beta_{(N)}^*$ exists, we postulate its form based on the generalization of our arguments for the case of two cell types. We also postulate the form of the convergence speed parameter $\beta_{(N)}^\Delta$ as introduced for two cell types.

The scaling parameter is directly generalized from Equation (6) to N cell types

$$\beta_{(N)}^s = \sum_{i=0}^{N-1} \beta_{ii} + \sum_{i<j}^{N-1} \beta_{ij} = \langle \beta_{(N)}, \mathbf{a}_s \rangle \tag{12}$$

$$\mathbf{a}_s = (1, \dots, 1)^T.$$

and only scales all waiting times between cell switches as in the case of two cell types.

The invariance against relabeling of the cell types applies as well to N cell types which implies, analogously to Equation (7),

$$\beta_{(N)}^* = \mathbf{a} \sum_{i=0}^{N-1} \beta_{ii} + \mathbf{b} \sum_{i<j}^{N-1} \beta_{ij} = \langle \boldsymbol{\beta}_{(N)}, \mathbf{a}_* \rangle \quad (13)$$

where the constants \mathbf{a} , \mathbf{b} are set again by the condition

$$\langle \mathbf{a}_*, \mathbf{a}_s \rangle = 0 \quad \Rightarrow \quad N\mathbf{a} = -\frac{N(N-1)}{2} \mathbf{b}. \quad (14)$$

Choosing again $\mathbf{a} = 1$, the effective adhesion parameter has the form

$$\beta_{(N)}^* = \sum_{i=0}^{N-1} \beta_{ii} - \frac{2}{(N-1)} \sum_{i<j}^{N-1} \beta_{ij} \quad (15)$$

with corresponding vector \mathbf{a}_* . For $N = 2$ this equals the result for two cell types from Equation (9) and for $N = 3$ this takes the form

$$\beta_{(3)}^* = \beta_{00} + \beta_{11} + \beta_{22} - (\beta_{01} + \beta_{02} + \beta_{12}). \quad (16)$$

In contrast to the case of two cell types, the space perpendicular to the vectors \mathbf{a}_s and \mathbf{a}_* for $N > 2$ is not one- but $[N + N(N-1)/2 - 2]$ -dimensional, i.e., two-dimensional for $N = 3$. Thus, the definition of a convergence speed parameter $\beta_{(N)}^\Delta$, analogous to Equation (10), is ambiguous. One option is to consider all pairs of cell types and their corresponding convergence speed parameters, e.g., $|\beta_{00} - \beta_{11}|$, $|\beta_{00} - \beta_{22}|$, and $|\beta_{11} - \beta_{22}|$ for $N = 3$. Since the segregation of all cell types requires the segregation of each pair of cell types, these pair-wise parameters should predict the convergence speed on subsets of intercellular adhesion parameters analogously to Figure 3. For instance, when two vectors of intercellular adhesion parameters are equal except they differ in one of these pair-wise convergence speed parameters, the one with the smaller parameter leads to a faster convergence. However, these pair-wise parameters do not allow comparison if several of them differ between two vectors of adhesion parameters and the influence of the heterotypic parameters is not even considered. Thus, the convergence speed parameter $\beta_{(N)}^\Delta$ is instead defined by the generalization of Equation (11)

$$\begin{aligned} \beta_{(N)}^\Delta &= \left\| \boldsymbol{\beta}_{(N)} - \boldsymbol{\beta}_{(N)}^{\text{sym}} \right\| \\ \boldsymbol{\beta}_{(N)}^{\text{sym}} &= (\bar{\beta}_{\text{hom}}, \dots, \bar{\beta}_{\text{hom}}, \bar{\beta}_{\text{het}}, \dots, \bar{\beta}_{\text{het}}) \\ \bar{\beta}_{\text{hom}} &= \frac{1}{N} \sum_{i=0}^{N-1} \beta_{ii}, \quad \bar{\beta}_{\text{het}} = \frac{2}{N(N-1)} \sum_{i<j}^{N-1} \beta_{ij}. \end{aligned} \quad (17)$$

We postulate, in analogy to the case of two cell types, that the parameter $\beta_{(N)}^\Delta$ determines the number of cell switches required to reach the asymptotic level of segregation for a given effective adhesion parameter $\beta_{(N)}^*$.

3.3. Numerical Evidence for the Effective Adhesion Parameter

We test our analytical prediction for the effective adhesion parameter for more than two cell types, by comparison with numerical simulations of the DMM model for two and three cell types. We find in both cases that the proposed effective adhesion parameter determines the asymptotic level of cell segregation. However, there are deviations due to the limited simulation time, from which the asymptotic level of the order indicator $\bar{\omega}$ is extrapolated. This is revealed by additionally linking the simulation results to the convergence speed parameter β^Δ , introduced above. For two cell types, we observe that for

small values of $\beta_{(2)}^\Delta$, which refers to simulations which converge quickly to their asymptotic value, the dependency between $\beta_{(2)}^*$ and $\bar{\omega}$ agrees almost perfectly with the analytically proven prediction. Deviations from the analytical prediction are the most pronounced the larger the convergence speed parameter is. For three cell types, we observe an analogous relation between deviations from the postulated dependency on the effective adhesion parameter $\beta_{(3)}^*$ and a convergence speed parameter $\beta_{(3)}^\Delta$ as introduced above.

3.3.1. The Asymptotic Level of Cell Segregation Depends on the Effective Adhesion Parameter

At first the dependence of the level of asymptotic cell segregation $\bar{\omega}$ on the effective adhesion parameter β^* is tested. For the 2-cell-type, the effective adhesion parameter clearly determines the asymptotic sorting state, see dark blue points in Figure 4a, i.e., $\bar{\omega}$ increases with increasing $\beta_{(2)}^*$, as predicted by Voss-Böhme and Deutsch [12], and all configurations stay randomly mixed at $\beta_{(2)}^* = 0$, i.e., $\bar{\omega}(\beta^* = 0) = 0.5$ (see black dashed lines), as predicted by the heuristic argument above. The monotonic dependence of the asymptotic order parameter $\bar{\omega}$ on the proposed effective parameter $\beta_{(2)}^*$ is supported by a value of 0.81 of the Spearman rank correlation coefficient (Kendall rank correlation coefficient of 0.58). Additionally, Figure 4a shows the asserted influence of the convergence speed parameter $\beta_{(2)}^\Delta$ on level of cell segregation dynamics. More precisely, the higher the $\beta_{(2)}^\Delta$ value is the slower is the sorting process and the closer the respective $\bar{\omega}$ value remains to $\omega(t = 0) \approx 0.5$ until the end of the simulation. For these intercellular adhesion parameters $\beta_{(2)}^*$, the asymptotic value of the order indicator has not yet been reached, which is why they are scattered above and below the increasing line, indicated by the dark blue points in Figure 4a. Note, that also the effective adhesion parameter β^* shows impact on the amount of cell switches needed for a system to reach its asymptotic cell segregation $\bar{\omega}$. This is further investigated in Appendix D.

Analogous results for the case of three cell types with the postulated effective adhesion parameter are shown in Figure 4b. Qualitatively the same functional relation between the hypothesized effective adhesion parameter $\beta_{(3)}^*$ and $\bar{\omega}$ is visible as in the case of two cell types, see dark blue points in Figure 4b. This includes the prediction that all configurations stay randomly mixed at $\beta_{(3)}^* = 0$, i.e., $\bar{\omega}(\beta^* = 0) = 0.33$ (see black dashed lines). The monotonic dependence of the asymptotic order parameter $\bar{\omega}$ on the proposed effective parameter $\beta_{(3)}^*$ is supported by a value of 0.79 of the Spearman rank correlation coefficient (Kendall rank correlation coefficient of 0.53), which is of similar quality as in the case of two cell types. Based on the analogy of the numerical results for two and three cell types, we conclude that $\beta_{(3)}^*$ predicts the level of sorting reached asymptotically such as $\beta_{(2)}^*$ does in the 2-cell-type case. Correspondingly, the asymptotic cell segregation exhibits the postulated influence of the convergence speed parameter $\beta_{(3)}^\Delta$. Note that the restriction in Figure 4b to simulations with an $\beta_{(3)}^\Delta < 3.0$ ensures a sufficient sampling of data points with smaller $\beta_{(3)}^\Delta$ parameter values. Smaller β^Δ values result in more simulations reaching their asymptotic cell segregation.

For the 2-cell-type system, we can additionally visualize the effective adhesion parameter $\beta_{(2)}^*$ as a plane in 3D, see Figure 5, with $E : (\beta_{00} + \beta_{11}) - 2\beta_{01} = 0$ for $\beta_{(2)}^* = 0$ and its normal vector $\mathbf{a}_* = (1, 1, -2)$. This plane separates simulated data points based on their level of asymptotic cell segregation $\bar{\omega}$. This means, the direction \mathbf{a}_* indeed describes the direction of order progression, see color transition in Figure 5.

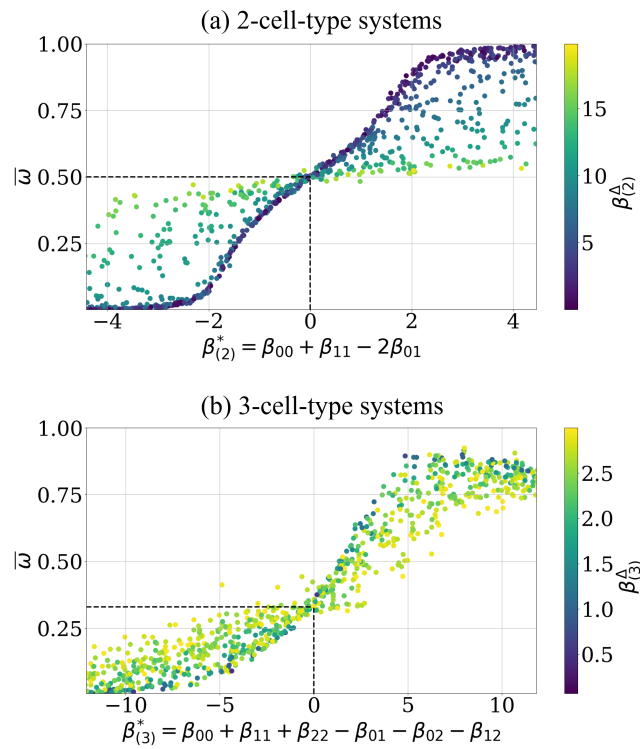


Figure 4. The level of the asymptotic cell segregation is predicted by the effective adhesion parameter. **(a)** Each point corresponds to a simulation whose intercellular adhesion parameter is drawn uniform, i.e., $\beta_{(2)} \sim U[-10, 10]^3$, with a random start configuration, i.e., initial order indicator $\omega(t = 0) \approx 0.5$. The curve $\omega(\beta_{(2)}^*)$ intersects with the predicted point $\omega(\beta_{(2)}^* = 0) = 0.5$ (highlighted by black dashed lines). A data point represents the asymptotic estimate $\bar{\omega}$ of this simulation after $\approx 300,000$ cell switches, along with the corresponding effective adhesion parameter $\beta_{(2)}^*$ and convergence parameter $\beta_{(2)}^\Delta$ (color bar). The corresponding shift parameter $\beta_{(2)}^s$ is not shown. **(b)** Analogous representation for the data points of 3-cell-type systems starting with random configurations, i.e., initial order indicator values $\omega(t = 0) \approx 0.33$. The curve $\omega(\beta_{(3)}^*)$ intersects with the predicted point $\omega(\beta_{(3)}^* = 0) = 0.33$ (highlighted by black dashed lines). The intercellular adhesion parameters are also drawn uniform, i.e., $\beta_{(3)} \sim U[-10, 10]^6$ and have an additional condition such as $\beta_{(3)}^\Delta < 3$, see Equation (17), to reduce scattering for visibility, see text for details.

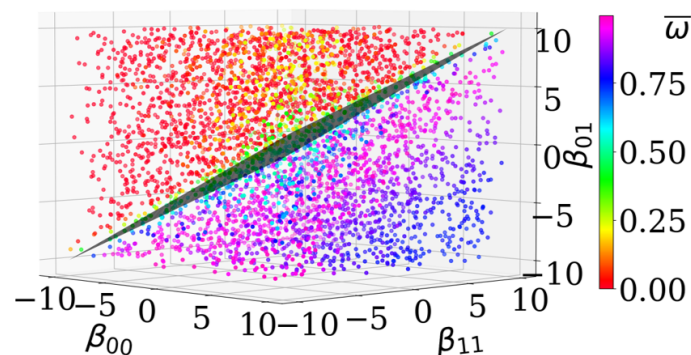


Figure 5. A plane with the analytically derived normal vector separates data points respective to their estimated asymptotic level of cell segregation. The points refer to the same data as in Figure 4a. The coloration represents the level of asymptotic cell segregation $\bar{\omega}$. The analytically predicted plane (gray rhomb in the middle) with the normal vector \mathbf{a}_* is $E : \beta_{00} + \beta_{11} - 2\beta_{01} = \langle \beta_{(2)}, \mathbf{a}_* \rangle = 0$.

We present analogous analysis for 4- and 5-cell-type systems, see Appendix C, Figure A1 and find a strong support of the analytically predicted effective adhesion parameter as well.

3.3.2. Estimating the Effective Adhesion Parameter Using Statistical Learning Methods

We confirmed qualitatively the compliance between our prediction and the numerical data in the previous section. Now we confirm additionally our prediction by comparing the analytical effective adhesion parameters for two and three cell types to the corresponding estimates obtained via two statistical learning methods, Support Vector Machines (SVM) and Logistic Regression (Logit-Model) [16]. More precisely, for the general case of N cell types, the learning methods provide a hyperplane equation of the effective adhesion parameter $\beta_{(N)}^*$. Therefore, we quantify statistically the deviation from our analytical prediction. This is done for $\beta_{(2)}^*$ and $\beta_{(3)}^*$ based on the numerical data $\bar{\omega}(\beta)$ referred to in the previous section and displayed in Figure 4a,b, respectively. Both methods are suited for this task because they fit the linear dependency of the parameter β^* on the intercellular adhesion parameters β , are well documented and directly available from libraries such as scikit-learn for Python [17], and, in contrast to some other approaches like deep learning, both quantify the factors that lead to classification.

The learning methods provide a hyperplane equation as a linear predictor to separate classified data points. Thus, the numerical data points $\bar{\omega}(\beta)$ are first classified based on their asymptotic order indicator $\bar{\omega}$ and a classification threshold. As thresholds, we chose the random configurations, i.e. $\bar{\omega} \geq 0.5$ for two cell types and $\bar{\omega} \geq 0.33$ for three cell types. Due to the symmetry of the effective adhesion parameter β^* , see Equations (7) and (13), the number of coefficients, which must be estimated, can be reduced to the constant a for the homotypic and the constant b for the heterotypic adhesion parameters. The resulting hyperplane equations for both 2-cell-type and 3-cell-type cases are presented in the headlines of Table 1a,b.

Table 1. The effective adhesion parameter is confirmed by statistical learning methods. (a) For the estimation of the effective adhesion parameter $\beta_{(2)}^*$, 2000 simulations are conducted under the same conditions as described in Figure 4a. After simulation, each data point is classified: $\bar{\omega} > 0.5$ as class 1, else class -1 . The classified data points are used by the SVM and Logit-Model to estimate the homotypic a and the heterotypic b coefficient as well as the intercept i according to the table head. The model accuracy is tested via 5-fold cross-validation. (b) Analogous for the 3-cell-type systems the simulation conditions are described in Figure 4b. For data point classification applies: $\bar{\omega} > 0.33$ as class 1, else class -1 . For details on the estimation process as well as the model accuracy test, see Appendix E.

(a) 2-cell-type systems	$0 = a(\beta_{00} + \beta_{11}) + b\beta_{01} + i$			Model
	a/a	b/a	i/a	Accuracy
SVM	1.0000	-1.9764	-0.0374	0.9965
Logit	1.0000	-1.9763	-0.0455	0.9965
Theoretical prediction	1	-2	0	
(b) 3-cell-type systems	$0 = a(\beta_{00} + \beta_{11} + \beta_{22}) + b(\beta_{01} + \beta_{02} + \beta_{12}) + i$			Model
	a/a	b/a	i/a	Accuracy
SVM	1.0000	-0.9994	0.2835	0.9895
Logit	1.0000	-0.9935	0.2624	0.9900
Theoretical prediction	1	-1	0	

For the 2-cell-type case in Table 1a, both relative coefficients, for homotypic $a/a \approx 1.00$ and heterotypic $b/a \approx -1.99$ intercellular adhesion parameters, are close to their predicted values 1 and -2 , respectively. The relative intercept i/a , i.e. the value of $\beta_{(2)}^*$ which corresponds to the chosen threshold $\bar{\omega} = 0.5$, is for both learning methods near zero,

as predicted by our heuristic argument. These predictions of the SVM and the Logit-Model are of high quality as indicated by model accuracy values close to one. This quantitative result of the statistical learning methods in Table 1a support the analytical prediction for $\beta_{(2)}^* = \beta_{00} + \beta_{11} - 2\beta_{01}$.

Analogous for three cell types, the results of Table 1b support the analytical results by estimating $\beta_{(3)}^* = \beta_{00} + \beta_{11} + \beta_{22} - \beta_{01} - \beta_{02} - \beta_{12}$ with convincing model accuracy of about 0.99 for both learning methods. The relative intercept i/a , i.e., the value of $\beta_{(3)}^*$ which corresponds to the chosen threshold $\bar{\omega} = 0.33$, is for both learning methods close to zero, as predicted by our heuristic argument. We attribute the deviation of the relative intercept i/a from zero as well as the lower model accuracy compared to two cell types to insufficient convergence of the order indicators as predicted by the convergence speed $\beta_{(3)}^\Delta$.

We present analogous analysis for 4- and 5-cell-type systems, see Appendix C, Table A2 and again find a strong support of the analytically predicted parameters defining the effective adhesion parameter.

3.4. The Impact of Interfacial Tension, Adhesion or Repulsion on Cell Segregation

For two cell types, the EAP resolves previous discussions about the impact of interfacial tension on cell segregation compared to interfacial adhesion or repulsion [11]: We predict that a higher level of segregation is reached when the EAP is large, which can be achieved by both heterotypic repulsion or differential adhesion. In fact, transforming the relative contact tensions used in ref. [11] to the corresponding EAP by identifying high tensions T_{ij} in the experiments and the Cellular Potts model (CPM) with low adhesion parameters $\beta_{ij} = -T_{ij}$ in the DMM, we can predict the experimental and numerical observations there, see Figure 6. In particular, we find that low lengths of heterotypic interfaces (LHI) are observed for high EAP values, analogous to the DMM. Figure 6 shows that the EAP is a better predictor for the asymptotic segregation behavior than the classification according to the DAH/DITH and HIT hypotheses.

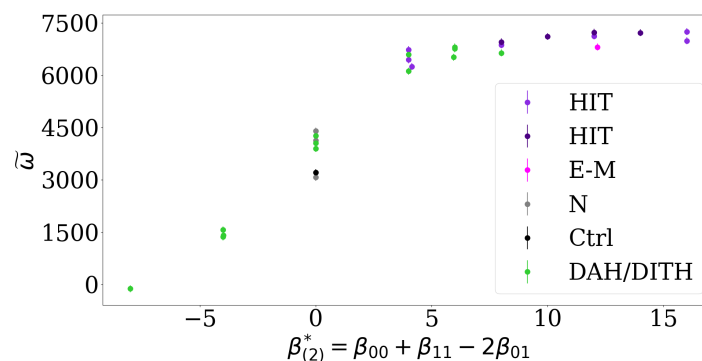


Figure 6. The effective adhesion parameter resolves previous discussions about the impact of interfacial tension compared to adhesion or repulsion. The indicator $\tilde{\omega}$ is analogous to the order indicator ω and is calculated by subtracting the length of heterotypic interfaces (LHI) reported in Canty et al. (2017) [11] from an upper bound of 8000 of all lattice site connections, the EAP is computed from the relative contact tensions T_{ij} used there: $\beta_{(2)}^* = -T_{00} - T_{11} + 2T_{01} = \beta_{00} + \beta_{11} - 2\beta_{01}$. In contrast, Canty et al. [11] propose that the asymptotic level of segregation is determined by a discrete number of scenarios into which all relative contact energies can be sorted. These scenarios are: Differential Adhesion Hypothesis / Differential Interfacial Tension Hypothesis (DAH/DITH), control (Ctrl), two cases of High Heterotypic Interfacial Tension Hypothesis (HIT), ectoderm-mesoderm energies (E-M), negative control (N). The prediction in [11] was that HIT leads to the highest level of segregation, while E-M, N, Ctrl and DAH/DITH lead to lower segregation.

Thus, we propose that rather than differentiating between intercellular adhesion and contact tension, the combined effect of intercellular contact on cellular motility, which is quantified by the EAP, should be focused on. Furthermore, the correct prediction of the

asymptotic level of segregation in the CPM simulations of Ref. [11] using the effective adhesion parameter β^* derived for the DMM suggests that our results may be directly applicable to CPM models of cell segregation. This is plausible due to the analogous structure of the exponent of the cell switch rates in the DMM and the energy functional in the CPM.

4. Discussion

We study analytically and numerically the differential migration model of Voss-Böhme and Deutsch (2010) [12], which is a cell-based model incorporating differential hypothesis [8], differential interfacial hypothesis [10], and High Heterotypic Interfacial Tension Hypothesis [11] within a unified framework. We generalize the existence and form of an effective adhesion parameter (EAP) guiding the asymptotic level of segregation, which is already known for systems with two cell types [12], to systems with an arbitrary number of cell types. We additionally predict the critical value of this effective adhesion parameter at which the system remains randomly mixed. For the case of two and three cell types, we confirm these theoretical predictions numerically and quantify the form of the effective adhesion parameter independently using statistical learning methods. The analogous results for 4- and 5-cell types suggest that our findings are valid in systems with even higher number of cell types. For two cell types, we show that the EAP resolves previous discussions about the impact of interfacial tension on cell segregation compared to interfacial adhesion or repulsion [11]. Thus, we propose that rather than differentiating between intercellular adhesion and contact tension, the combined effect of intercellular contact on cellular motility, which is quantified by the EAP, should be focused on.

Most previous studies focus on the simplest case of two cell types [7,13,14]. Although three and more cell types are more relevant in real biological systems, studies of cell segregation and pattern formation with three or more cell types are still rare [15]. By analytically and numerically demonstrating the existence of cell segregation for three cell types and characterizing its dynamics, we extend results from two cell types to the more general case.

The differential migration model has recently been applied successfully to reproduce experimental cell segregation data for two cell types [18]. In particular, the match between model and experimental observation demonstrated that contact tension and adhesion are sufficient to explain segregation data without additional mechanisms of collective motion. Analogously, our analytical and numerical results can be applied to cell segregation experiments with three or more cell types.

Author Contributions: Conceptualization, H.-J.B., S.L. and A.V.-B.; Data curation, P.R. and S.L.; Formal analysis, P.R., S.L. and A.V.-B.; Funding acquisition, H.-J.B. and A.V.-B.; Investigation, P.R., S.L. and A.V.-B.; Methodology, S.L. and A.V.-B.; Project administration, H.-J.B., S.L. and A.V.-B.; Resources, H.-J.B. and A.V.-B.; Software, P.R. and S.L.; Supervision, H.-J.B., S.L. and A.V.-B.; Validation, P.R., S.L. and A.V.-B.; Visualization, P.R., S.L. and A.V.-B.; Writing—original draft, P.R., S.L. and A.V.-B.; Writing—review & editing, P.R., H.-J.B., S.L. and A.V.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is co-financed by the EU, by the European Social Fund (ESF) and by tax funds on the basis of the budget passed by the Saxon state parliament: SAB-Nr. 100382145. This article is additionally funded by the Open Access Publication Fund of Hochschule für Technik und Wirtschaft Dresden University of Applied Sciences and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—432908064.

Data Availability Statement: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are contained within the manuscript and its appendix.

Acknowledgments: We thank Fabian Schwarzenberger for helpful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. PCA Implementation

The Differential Migration Model (DMM) [12] is implemented as a rule-based probabilistic cellular automaton (PCA). Instead of the classical discrete-time implementation via Monte-Carlo steps, a Gillespie-like algorithm is used which simulates the PCA in continuous time. For this, a support data structure is employed which tracks all heterotypic cell switches. These switches correspond to all possible events of the system for the corresponding lattice configuration η at a given point in time. The code effort is higher compared to more intuitive time-discrete algorithms due to the necessary update mechanics of the data structure but it allows the performance of one heterotypic cell switch per iteration step, since there are iteration steps without actual events in the analogous time-discrete Monte-Carlo simulation. At each iteration step, one entry from the support data structure is randomly chosen according to a weight which is calculated from the corresponding cell switch rate $c(x, y, \eta)$, see Equation (2). Afterwards, the associated lattice sites x, y interchange their occupations. Then all entries of the support data structure that are affected by the configuration change are updated.

Algorithm A1 The PCA Algorithm.

```

lattice = initializeLattice(size, cellTypeDistribution);
sds = initializeSupportDataStructure(lattice, cellContactBindingStrength);
while checkSimulationRunCondition do
    cellSwitch = getOneRandomCellSwitch(sds);
    x, y = getLatticeSitesAssociatedToCellSwitch(cellSwitch);
    interchangeLatticeOccupationsAtSites(x, y);
    affectedLatticeSites = getAffectedLatticeSites(x, y, lattice, sds);
    for z in affectedLatticeSites do
        neighbors = getNeighborsOfLatticeSite(z, lattice);
        for n in neighbors do
            c = getCellSwitchRate(z, n, lattice);
            heterotypic = checkLatticeSitesAreHeterotypic(z, n);
            data = [n, c, heterotypic];
            updateSdsEntryForLatticeSiteWithData(sds, z, data);
        end
    end
end

```

Appendix A.1. initializeLattice

Initially, the square lattice S has the start configuration $\eta_0 \in X$. The lattice S is described by $S = \{0, 1, \dots, L\}^2$ where $L = 25$. The parameter called "cellTypeDistribution" determines the total amount of cells for each cell type $w \in W$ within the lattice.

Appendix A.2. initializeSupportDataStructure

The support data structure holds the information about the neighbors of all lattice sites together with the corresponding cell switch rates and whether a cell-cell contact is heterotypic or not. The parameter named "cellContactBindingStrength" provides the intercellular adhesion parameters $\beta_{(N)}$ to calculate the cell switch rate.

Appendix A.3. checkSimulationRunCondition

To finalize a simulation a run condition is met. For this, either an iteration step count or a normalized order indicator value is used. In most cases, the simulation is stopped at 312,500 cell switches except for the data in Figure 3 where thresholds of the order indicator are used.

Appendix A.4. getOneRandomCellSwitch

One heterotypic entry is drawn from the support data structure. This heterotypic cell switch identified by k has the corresponding cell switch rate $c^{(k)}(x, y, \eta)$. This rate belongs to the set $\zeta(\eta)$ of all heterotypic cell switch rates, given the lattice configuration $\eta \in X$. The probability $p^{(k)}$ that a cell switch k is chosen correlates directly with $c^{(k)}(x, y, \eta)$. That is shown in Equation (A1):

$$\zeta(\eta) := \{c(x, y, \eta) \mid x, y \in S, \eta(x) \neq \eta(y)\},$$

$$p^{(k)} := \frac{c^{(k)}(x, y, \eta)}{\sum_{j=1}^n c^{(j)}(x, y, \eta)} \quad \text{with } n = |\zeta(\eta)| \quad \text{and} \quad \sum_{j=1}^n p^{(j)} = 1 \quad (\text{A1})$$

Appendix A.5. getLatticeSitesAssociatedToCellSwitch

The lattice sites $x, y \in S$ from the chosen cell switch are obtained.

Appendix A.6. interchangeLatticeOccupationsAtSites

The new occupation $\eta(x) \in W$ of lattice site $x \in S$ becomes the former occupation $\eta(y) \in W$ of lattice site $y \in S$ and vice versa, according to Equation (1).

Appendix A.7. getAffectedLatticeSites

The set $Z_{\eta \rightarrow \eta_{xy}}$ of all lattice sites affected by the cell switch and the resulting configuration change $\eta \rightarrow \eta_{xy}$, see Equation (1), is composed of the direct neighbors $u(x, y)$ of the lattice sites $x, y \in S$ associated with the cell switch $\eta \rightarrow \eta_{xy}$ and their neighbors. Thus, it is

$$Z_{\eta \rightarrow \eta_{xy}} := \bigcup_{e \in u(x, y)} \mathfrak{N}(e) \quad \text{with} \quad u(x, y) := \mathfrak{N}(x) \cup \mathfrak{N}(y).$$

Appendix A.8. getNeighborsOfLatticeSite

For a lattice site $z \in S$ the neighborhood $\mathfrak{N}(z)$ is obtained, see description to Equation (2).

Appendix A.9. getCellSwitchRate

The cell switch rate $c(x, y, \eta)$ between two lattice sites x and y is calculated according to Equation (2).

Appendix A.10. checkLatticeSitesAreHeterotypic

Two lattice sites $x \in S$ and $y \in S$ are heterotypic if their corresponding cell types $\eta(x), \eta(y)$ are not equal.

Appendix A.11. updateSdsEntryForLatticeSiteWithData

The updated data in terms of cell switch rate and heterotypic cell contact of the neighborhood member $n \in \mathfrak{N}(z)$ is assigned to the neighborhood host z within the support data structure.

Appendix B. Order Indicator

Table A1. Highest and lowest number of homotypic connections for ideal 2- and 3-cell-type systems. These values are used to normalize the total sum of homotypic connections, see Equation (3).

	2-Cell-Type Systems	3-Cell-Type Systems
d_{max}	1200	1175
d_{min}	50	0

If we consider a lattice $S = \{0, 1, \dots, L\}^2$ then the ideal case of a fully sorted configuration consists of stripes for each of the N cell types, each with a height of $\frac{L}{N}$ and a length of L . In this scenario the highest amount d_{max} of homotypic lattice site connections is $d_{max} = 2 \cdot L^2 - n \cdot N$, since every lattice site has two unique homotypic connections except for those at cell-type interfaces. These only have one homotypic connection. For the same ideal scenario, a chessboard configuration where no homotypic connections exist can be achieved for the right lattice side length L . Therefore, $d_{min} = 0$.

In practice, a discrete lattice with 25^2 lattice sites cannot be separated into even stripes. Due to this imperfection, small deviations from these values occur in data. In the case of 2-cell-type systems, a lattice side of $L = 25$ causes $d_{min} = 2 \cdot L = 50$ because every cortical lattice site has one homotypic connection due to periodicity.

In other publications such as Canty et al. (2017) [11], the level of cell segregation and clustering is quantified by the length of the heterotypic interface, here we use the amount of homotypic lattice site connections, see Equation (3).

Appendix C. Numerical Evidence for the Effective Adhesion Parameter with Extended Cell-Type Number

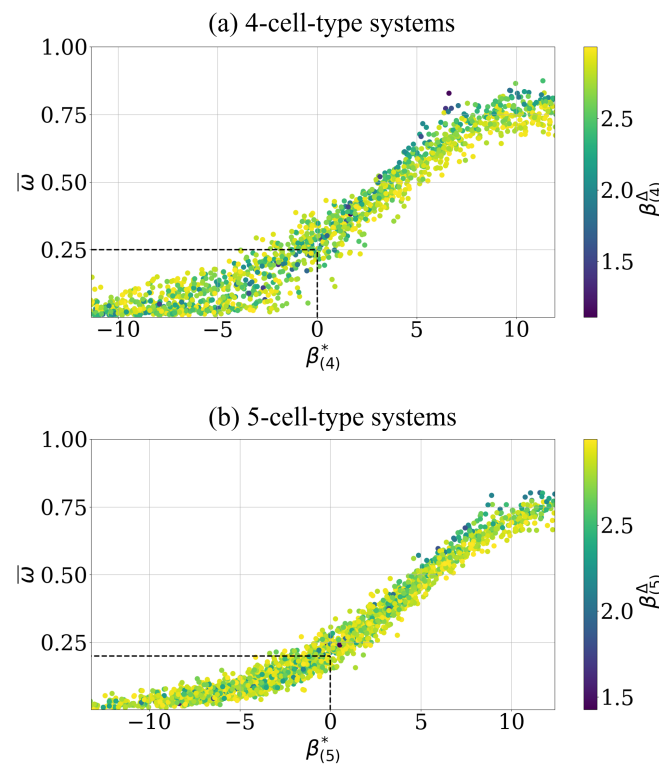


Figure A1. The level of the asymptotic cell segregation is predicted by the effective adhesion parameter for 4- and 5-cell-type systems. (a) Each point corresponds to a simulation whose intercellular adhesion parameter is drawn uniform, i.e., $\beta_{(4)} \sim U[-4, 4]^{10}$ with an additional condition such as $\beta_{(4)}^\Delta < 3$, see Equation (17), to reduce scattering for visibility, see text of Figure 4 for details. The start configuration is random, i.e., initial order indicator $\omega(t = 0) \approx 0.25$. The curve $\omega(\beta_{(4)}^*)$ intersects with the predicted point $\omega(\beta_{(4)}^* = 0) = 0.25$ (highlighted by black dashed lines). A data point represents the asymptotic estimate $\bar{\omega}$ of this simulation after ≈ 300000 cell switches, along with the corresponding effective adhesion parameter $\beta_{(4)}^*$ and convergence parameter $\beta_{(4)}^\Delta$ (color bar). (b) Analogous representation for the data points of 5-cell-type systems, starting with random configurations, i.e., initial order indicator values $\omega(t = 0) \approx 0.20$. The curve $\omega(\beta_{(5)}^*)$ intersects with the predicted point $\omega(\beta_{(5)}^* = 0) = 0.20$ (highlighted by black dashed lines). The intercellular adhesion parameters are also drawn uniform, i.e., $\beta_{(5)} \sim U[-2.5, 2.5]^{15}$ with the additional condition $\beta_{(5)}^\Delta < 3$.

Table A2. The effective adhesion parameter is confirmed by statistical learning methods for 4 and 5 cell types. (a) For the estimation of the effective adhesion parameter $\beta_{(4)}^*$, 2000 simulations are conducted under the same conditions as described in Figure A1a. After simulation, each data point is classified: $\bar{\omega} > 0.25$ as class 1, else class -1 . The classified data points are used by the SVM and Logit-Model to estimate the homotypic \mathbf{a} and the heterotypic \mathbf{b} coefficient as well as the intercept \mathbf{i} according to the table head. The model accuracy is tested via 5-fold cross-validation. (b) Analogous for the 5-cell-type systems the simulation conditions are described in Figure A1b. For data point classification applies: $\bar{\omega} > 0.20$ as class 1, else class -1 . For details on the estimation process as well as the model accuracy test, see Appendix E.

(a) 4-cell-type systems	$\mathbf{0} = \mathbf{a} \sum_{i=0}^3 \beta_{ii} + \mathbf{b} \sum_{i<j}^3 \beta_{ij} + \mathbf{i}$			Model
	\mathbf{a}/\mathbf{a}	\mathbf{b}/\mathbf{a}	\mathbf{i}/\mathbf{a}	Accuracy
SVM	1.0000	-0.6553	0.8022	0.9615
Logit	1.0000	-0.6625	0.7330	0.9610
Theoretical prediction	1	-2/3	0	
(b) 5-cell-type systems	$\mathbf{0} = \mathbf{a} \sum_{i=0}^4 \beta_{ii} + \mathbf{b} \sum_{i<j}^4 \beta_{ij} + \mathbf{i}$			Model
	\mathbf{a}/\mathbf{a}	\mathbf{b}/\mathbf{a}	\mathbf{i}/\mathbf{a}	Accuracy
SVM	1.0000	-0.5030	0.5889	0.9555
Logit	1.0000	-0.4913	0.5992	0.9565
Theoretical prediction	1	-0.5	0	

Appendix D. The Impact of the Effective Adhesion Parameter on the Convergence Speed

Figure A2 shows that the larger the EAP β^* is the more cell switches are needed before a simulation reaches its asymptotic segregation state which is a fully sorted pattern in the case of $\beta_{(2)}^* = 12$ and $\beta_{(2)}^* = 24$ according to the DMM, see Section 2.1. This effect is visible from about $\beta^* > 4$ and the cause is still unknown.

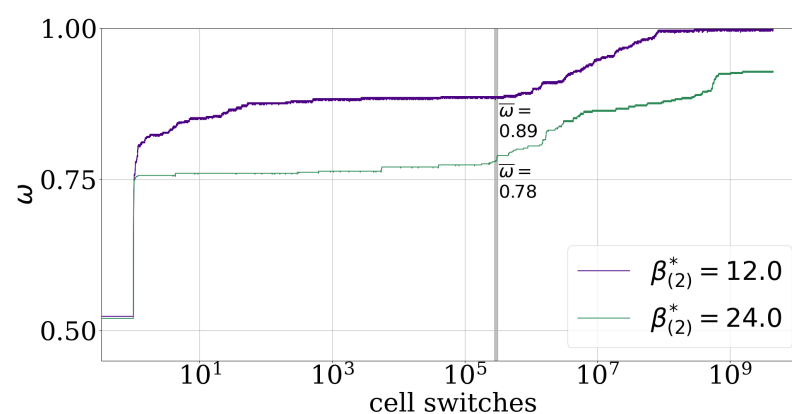


Figure A2. The effective adhesion parameter also influences the convergence speed exemplified with 2-cell-type systems. In the case of $\beta_{(2)}^* = 12$ (violet) the intercellular adhesion parameters are (2.0, 2.0, -4.0) and for $\beta_{(2)}^* = 24$ (green) they are (4.0, 4.0, -8.0). Both simulations have a random start configuration, i.e., initial order indicator $\omega(t = 0) \approx 0.5$ and run for about 7.2 million cell switches. The asymptotic estimate $\bar{\omega}$ for the last 10% of ≈ 300000 cell switches (gray area) is written on the corresponding time-series.

Appendix E. Statistical Learning Methods

All methods, SVM and Logistic regression as well as the 5-fold cross-validation, are executed in default mode and as recommended by the library scikit-learn. This includes among others the hyperparameter $C = 1.0$. The smaller this inverse regularization parameter is the more regularization is applied and the lower is the risk of model over fitting. This also can increase the numerical stability. The default penalization is in both cases l2-penalty [17]. Since all intercellular adhesion parameters are from the same distribution, see caption of Figure 4, preprocessing steps such as data scaling are omitted. Other execution details are:

- SVM: used algorithm implementation “SVC”
- SVM: kernel “linear”
- Logit: used training algorithm/solver “lbfgs”

References

1. Heisenberg, C.P.; Bellaïche, Y. Forces in Tissue Morphogenesis and Patterning. *Cell J.* **2013**, *153*, 948–962. [[CrossRef](#)] [[PubMed](#)]
2. Holtfreter, J. Gewebsaffinität, ein Mittel der embryonalen Formbildung. *Arch. Exp. Zellforsch.* **1939**, *23*, 169–209.
3. Townes, P.L.; Holtfreter, J. Directed movements and selective adhesion of embryonic amphibian cells. *J. Exp. Zool.* **1955**, *128*, 53–120. [[CrossRef](#)]
4. Armstrong, P.B. Light and electron microscope studies of cell sorting in combinations of chick embryo neural retina and retinal pigment epithelium. *Wilhelm Roux Arch. Entwickl. Mech. Org.* **1971**, *168*, 125–141. [[CrossRef](#)] [[PubMed](#)]
5. Armstrong, P.B. Cell sorting out: The self-assembly of tissues in vitro. *Crit. Rev. Biochem. Mol. Biol.* **1989**, *24*, 119–149. [[CrossRef](#)] [[PubMed](#)]
6. Beysens, D.A.; Forgacs, G.; Glazier, J.A. Cell sorting is analogous to phase ordering in fluids. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9467–9471. [[CrossRef](#)] [[PubMed](#)]
7. Méhes, E.; Mones, E.; Németh, V.; Vicsek, T. Collective Motion of Cells Mediates Segregation and Pattern Formation in Co-Cultures. *PLoS ONE* **2012**, *7*, e31711. [[CrossRef](#)] [[PubMed](#)]
8. Steinberg, M.S. On the Mechanism of Tissue Reconstruction by Dissociated Cells, Iii. Free Energy Relations and the Reorganization of Fused, Heteronomic Tissue Fragments. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1769–1776. [[CrossRef](#)] [[PubMed](#)]
9. Harris, A.K. Is Cell sorting caused by differences in the work of intercellular adhesion? A critique of the Steinberg hypothesis. *J. Theor. Biol.* **1976**, *61*, 267–285. [[CrossRef](#)]
10. Brodland, G.W.; Chen, H.H. The mechanics of heterotypic cell aggregates: Insights from computer simulations. *J. Biomech. Eng.* **2000**, *122*, 402–407. [[CrossRef](#)] [[PubMed](#)]
11. Cauty, L.; Zarour, E.; Kashkooli, L.; François, P.; Fagotto, F. Sorting at embryonic boundaries requires high heterotypic interfacial tension. *Nat. Commun.* **2017**, *8*, 157. [[CrossRef](#)] [[PubMed](#)]
12. Voss-Böhme, A.; Deutsch, A. The cellular basis of cell sorting kinetics. *J. Theor. Biol.* **2010**, *263*, 419–436. [[CrossRef](#)] [[PubMed](#)]
13. Ventrella, R.; Kaplan, N.; Getsios, S. Asymmetry at cell-cell interfaces direct cell sorting, boundary formation, and tissue morphogenesis. *Exp. Cell Res.* **2017**, *358*, 58–64. [[CrossRef](#)] [[PubMed](#)]
14. Ko, J.M.; Lobo, D. Continuous Dynamic Modeling of Regulated Cell Adhesion: Sorting, Intercalation, and Involution. *Biophys. J.* **2019**, *117*, 2166–2179. [[CrossRef](#)] [[PubMed](#)]
15. Sivakumar, N.; Warner, H.V.; Peirce, S.M.; Lazzara, M.J. Agent-Based Model of Multicellular Spheroid Pattern Formation Driven by Synthetic Cell Adhesion Signaling Circuits. *bioRxiv* **2021**. [[CrossRef](#)]
16. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009.
17. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
18. Franke, F.; Aland, S.; Böhme, H.J.; Voss-Böhme, A.; Lange, S. Is Cell segregation just like oil and water? *arXiv* **2021**, arXiv:2109.00364.