Distinguishing between mechanisms of cell aggregation using pair-correlation functions

D. J. G. Agnew[a], J. E. F. Green[a], T. M. Brown[a], M. J. Simpson[b,c], B. J. Binder[a,*]

[a] *School of Mathematical Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia.*
[b] *School of Mathematical Sciences, Queensland University of Technology (QUT), Brisbane, Australia.*
[c] *Tissue Repair and Regeneration Program, Institute of Health and Biomedical Innovation, QUT, Brisbane, Australia.*

Abstract

Many cell types form clumps or aggregates when cultured *in vitro* through a variety of mechanisms including rapid cell proliferation, chemotaxis, or direct cell-to-cell contact. In this paper we develop an agent-based model to explore the formation of aggregates in cultures where cells are initially distributed uniformly, at random, on a two-dimensional substrate. Our model includes unbiased random cell motion, together with two mechanisms which can produce cell aggregates: (i) rapid cell proliferation, and (ii) a biased cell motility mechanism where cells can sense other cells within a finite range, and will tend to move towards areas with higher numbers of cells. We then introduce a pair-correlation function which allows us to quantify aspects of the spatial patterns produced by our agent-based model. In particular, these pair-correlation functions are able to detect differences between domains populated uniformly at random (*i.e.* at the exclusion complete spatial randomness (ECSR) state) and those where the proliferation and biased motion rules have been employed - even when such differences are not obvious to the naked eye. The pair-correlation function can also detect the emergence of a characteristic inter-aggregate distance which occurs when the biased motion mechanism is dominant, and is not observed when cell proliferation is the main mechanism of aggregate formation. This suggests that applying the pair-correlation function to experimental images of cell aggregates may provide information about the mechanism associated with observed aggregates. As a proof of concept, we perform such analysis for images of cancer cell aggregates, which are known to be associated with rapid proliferation. The results of our analysis are consistent with the predictions of the proliferation-based simulations, which supports the potential usefulness of pair correlation functions for providing insight into the mechanisms of aggregate formation.

Keywords: cell aggregates; spatial patterns; agent-based model; cell-based experiments; pair-correlation functions; cell proliferation; chemotaxis; cell-cell interaction

*Corresponding author. Email: benjamin.binder@adelaide.edu.au, *Telephone* + 61 429 401 874

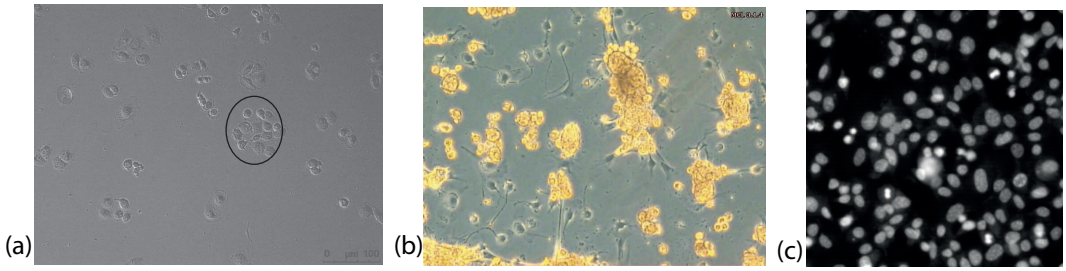

Figure 1: Spatial patterning observed in images from several *in vitro* experiments. (a) Aggregates of MBA MD 231 breast cancer cells. (b) Aggregates of liver cells (reprinted from Thomas et al.[25], with permission from Eur. Cells Mater.). (c) Uniform distribution of 3T3 fibroblast cells.

1. Introduction

The formation of clusters or aggregates is a ubiquitous phenomenon in cell biology; examples include cultures of *myxobacteria*, the slime mould *dictyostelium*, and many other cell types grown *in vitro* for cancer research, studies in developmental biology, or applications in tissue engineering [1, 6, 14, 16, 18, 19, 25, 26]. These aggregates can be produced by rapid cell proliferation [22] (as are, for example, those shown in Fig. 1(a)), or by other mechanisms involving attractive cell-to-cell interactions, such as chemotaxis, or direct physical contact, as is the case for those shown in Fig. 1(b) [14, 25]. Identifying the mechanism of aggregate formation in a particular case provides us with a fundamental understanding of the underlying processes, and can also be of practical importance in, for example, optimising culture conditions to promote cell viability and functionality. In some cases, it may be possible to identify the mechanism of aggregate formation simply by observing the cells in culture - *e.g.* one can observe the physical contact between co-cultured hepatocytes and stellate cells using time-lapse video [25]. Unfortunately, this is not always possible since some observations are made at one point in time only. A complicating factor is that nearly all cell types undergo some degree of unbiased random motion. If a proliferation or attraction mechanism is absent, this type of unbiased motion leads to a uniform cell distribution and no spatial patterning will be observed (as in the case of the mouse cells shown in Fig. 1(c)). However, even when aggregating mechanisms are present, unbiased cell motion can obscure the details meaning that it is not obvious how to assess and measure the aggregation mechanism. Thus a means of analysing spatial patterns that is highly sensitive to the presence of non-random arrangements of cells, and able to provide information regarding the mechanism that drives aggregate formation, would be of considerable practical use.

In this study, we develop agent based models [6, 4, 11, 23] to analyse two common mechanisms of aggregate formation, each of which includes a component of random motion. We consider both a cell proliferation mechanism, and a biased cell motion mechanism, in which cells detect the presence of others within a certain spatial range and attempt to move towards them. The latter mechanism is a generic representation of a variety of attractive interactions between cells, including direct physical contact and chemotaxis. We show that when the agents experience a sufficiently large amount of random motion it is difficult to distinguish the spatial patterns produced by either mechanism from cells distributed uniformly, at random, by visual inspection alone. This motivates us to introduce a modified pair-correlation function [8, 9, 12, 17, 27] to provide a more precise means of characterising these spatial patterns.

The pair-correlation function is a summary statistic that provides a quantitative measure of spatial patterning. The function is derived by normalising the counts of the pair-distances between pairs of agents in the domain. The normalisation term ensures that the expected value of the pair-correlation function is unity at all pair-distances for domains that are populated uniformly at random — termed the exclusion complete spatial randomness (ECSR) state [5, 8, 12, 13, 15]. We show that the pair correlation function is sufficiently sensitive to distinguish the ECSR state from the patterns resulting from either of our cluster-forming mechanisms, even when no difference is detectable by eye. The pair-correlation function is also shown to provide important quantitative information on the spatial patterning, such as the multiple length scales of aggregation and segregation [8, 12], and the size of aggregates.

In cases where there is little random motion, we can clearly observe from visual inspection of our simulations using the biased motion rule the emergence of a characteristic separation distance between aggregates, which is not observed when aggregates are formed by rapid cell proliferation. This difference

is easily distinguished when comparing the pair correlation functions for the two types of simulation results. Significantly, we find that this difference in the pair correlation functions is still clearly evident even when the degree of unbiased motion is increased to a level such that it is impossible to distinguish any difference in the simulation results by eye. This demonstrates the potential utility of the pair-correlation function to identify and quantify mechanisms of aggregation in cell-based experiments.

Given the promising results from our simulation data, as a proof of concept, we apply our technique to experimental images of the cancer cells in Fig. 1(a). This cell type is known to form aggregates as a result of rapid cell proliferation [22]. We process the experimental images, rescaling the cell location and cell area data so that it can be approximately represented on a non-dimensional integer lattice in a similar manner to the simulation results. Evaluating the pair-correlation function for this dimensionless data set we obtain a signal that is qualitatively similar to that produced by the proliferation mechanism in the agent-based simulations. This suggests, as expected, that these aggregates are formed predominately by cell proliferation.

2. Agent-based model

We consider the situation, relevant to many types of *in vitro* cell culture, in which cells are seeded onto a two-dimensional surface. It is assumed that initially the cells are distributed uniformly at random within this domain. We develop an agent-based model, discrete in time and space, [4, 23] to simulate two mechanisms of aggregate formation: biased cell motion and cell proliferation, as outlined above. In both cases, cells are assumed to undergo a variable degree of unbiased random motion.

We define a two-dimensional domain, consisting of an integer lattice with unit spacing. We assume that this domain represents part of a larger region containing cells - for example, the region of a culture well within the field of view of microscope; we thus impose periodic boundary conditions in both the x- and y-directions. At any discrete time t the lattice sites (x, y), with $x \in 1, 2, \ldots, X$ and $y \in 1, 2, \ldots, Y$, are either occupied or unoccupied by single agent which is denoted with a matrix

$$M(x,y) = \begin{cases} 0 \text{ if site } (x,y) \text{ is vacant,} \\ 1 \text{ if site } (x,y) \text{ is occupied.} \end{cases} \tag{1}$$

The number of agents within the domain at any given time is

$$n = \sum_{x=1}^{X} \sum_{y=1}^{Y} M(x,y), \tag{2}$$

with corresponding lattice density

$$\rho = \frac{n}{XY}. \tag{3}$$

In all the simulations, the domain is initially populated uniformly at random with agents to some prescribed density. The simulation is updated at discrete time intervals, using the motility and proliferation rules described in detail below, until the desired end time is reached. During each time step, each agent is selected in random order without replacement, and attempts to move with probability $P_m \in [0, 1]$. This is followed by a proliferation event, which occurs with probability $P_p \in [0, 1]$. If at any point during a simulation, an agent attempts to move into an occupied lattice position, the move is aborted; hence the simulation is a volume exclusion process [4, 5, 10, 23].

2.1. Motility rules

When an agent is selected, we first draw a random number $r_0 \in [0, 1]$ to determine if it will attempt to move. If $r_0 < P_m$, an attempt is made; otherwise, the agent remains at its current location. We incorporate both biased motion and unbiased random motion into our model by introducing a bias probability, $0 \leq P_b \leq 1$. For agents which attempt to move, we draw another random number, $r_1 \in [0, 1]$, and compare this to P_b. If $r_1 < P_b$, the agent attempts to move according to the biased motion rule; otherwise the unbiased motility rule is followed. (Hence the probability a move is attempted using the biased motility rule is $P_m P_b$; an unbiased move is attempted with probability $P_m(1-P_b)$.) If the unbiased motility rule is selected, an agent positioned at (x, y) attempts to move with equal probability to one of four neighbouring sites $(x \pm 1, y)$ or $(x, y \pm 1)$. If the biased motility rule is chosen, the agent similarly

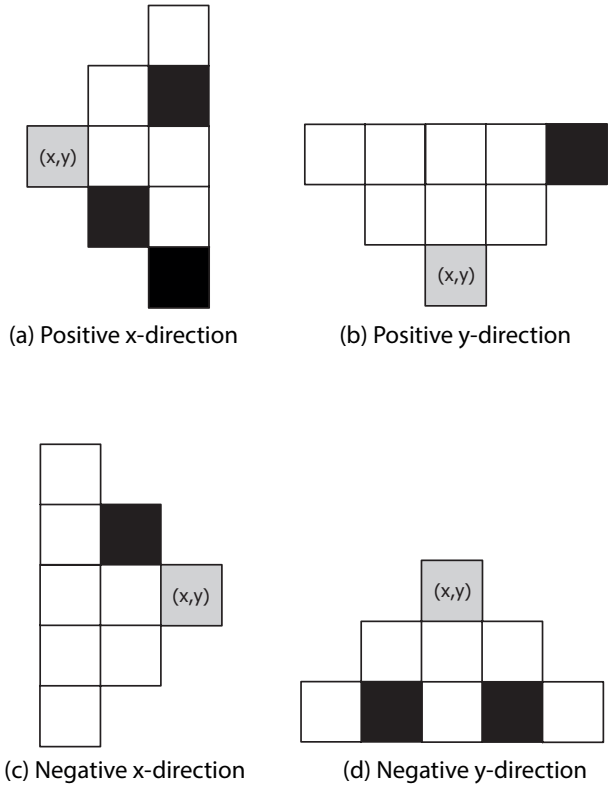

Figure 2: Biased motility rule with $L = 2$. The grey cell indicates the position (x, y) of the cell which is being considered. The black cells indicate the positions on the lattice occupied with agents that are included in the counts. (a) Count in the positive x-direction, $c_{x+} = 3$. (b) Count in the positive y-direction, $c_{y+} = 1$. (c) Count in the negative x-direction, $c_{x-} = 1$. (d) Count in the negative y-direction, $c_{y-} = 2$.

attempts to move to one of the same four sites, but with a probability weighted in favour of the direction that would move it towards the area most densely populated with other agents. The directional bias is quantified by counting the surrounding agents within some integer search radius, L, of the selected agent. The counts are performed in the positive and negative x and y directions, and are defined by

$$c_{y+} = \sum_{l=1}^{L} \sum_{i=x-l}^{x+l} M(i, y+l), \tag{4}$$

$$c_{y-} = \sum_{l=1}^{L} \sum_{i=x-l}^{x+l} M(i, y-l), \tag{5}$$

$$c_{x+} = \sum_{l=1}^{L} \sum_{j=y-l}^{y+l} M(x+l, j), \tag{6}$$

$$c_{x-} = \sum_{l=1}^{L} \sum_{j=y-l}^{y+l} M(x-l, j), \tag{7}$$

as illustrated in Fig. 2 for the case $L = 2$. The total count is $T = c_{x-} + c_{x+} + c_{y-} + c_{y+}$, which leads to defining the directional weights, $c_1 = c_{y+}/T$, $c_2 = c_{y-}/T$, $c_3 = c_{x+}/T$ and $c_4 = c_{x-}/T$. Drawing a random number, r_2, from the interval $[0, 1]$ and comparing it to the concatenation of the weights, $0 \leq c_1 \leq c_1 + c_2 \leq c_1 + c_2 + c_3 \leq c_1 + c_2 + c_3 + c_4 = 1$, determines the direction the agent moves (*e.g.* if $0 < r_2 < c_1$, the agent moves in the positive y-direction).

2.2. Proliferation rule

Following the motility event, we again use the random number drawing method described earlier to determine if the selected agent attempts to proliferate, which occurs with probability P_p [4, 23]. When

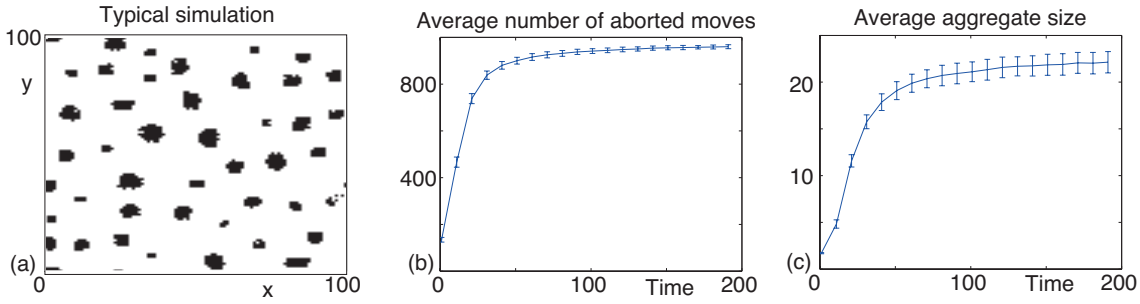

Figure 3: Average number of aborted moves and average aggregate size for the biased motion mechanism, $\rho = 0.10$, $P_p = 0$, $P_m = P_b = 1$ and $L = 5$. (a) Typical simulation at time $t = 190$. (b) and (c) Analysis of data from $N = 100$ simulations. (b) Average and standard deviation of the number of aborted moves. (c) Average and standard deviation of aggregate size.

an agent positioned at (x, y) attempts to proliferate, it deposits an additional daughter agent with equal probability on one of four neighbouring sites $(x \pm 1, y)$ or $(x, y \pm 1)$. If the site selected for the daughter agent is occupied, the proliferation event is aborted.

2.3. Simulation duration

Before we can perform any analysis we must determine when to terminate the simulations. For simulations involving the proliferation mechanism, with $0 < P_p \leq 1$, we choose simply to terminate the simulation when the density of agents in the domain has reached some pre-determined value. However, in the case $P_p = 0$, the density in the simulations remains constant for all time, due to the periodic boundary conditions and the absence of cell proliferation. We then choose to analyse the spatial patterning when the simulation has evolved to a quasi-steady state, and consider two measures to asses this. The first is to count the number of aborted moves in the simulations, which evolves to an approximately constant number with time, indicating that the macroscopic features of the spatial distribution of agents are unchanging (see Fig. 3(b)). While computationally convenient, this approach cannot be applied to experimental images. A second measure, where we examine the time evolution of the average aggregate size, can be used with simulation or experimental data (see Fig. 3(c)). The average aggregate size was computed using Matlab's image processing functions `#bwlabel.m` and `#regionprops.m`.

Typical aggregation patterns associated with the biased motion mechanism and rapid proliferation mechanism are given in Figs. 3, 4 and 5.

3. Pair-correlation function

We now introduce the average periodic pair correlation function, which we use to characterise the spatial patterning observed in the simulation images obtained from the agent based models described in §2. Essentially, this function displays the counts of the pair distances between the agents, normalised such that the value of the function is unity for a distribution of agents at the ECSR state [5, 8, 12, 13], and averaged over a number of realisations of the agent-based model. The formulation closely follows that of Binder and Simpson [8], with minor modifications to account for periodic boundary conditions.

We begin by considering the set of agent pairs

$$\psi_{ab} = \{(a, b) \mid a = (a_x, a_y), b = (b_x, b_y), a \neq b, M(a_x, a_y) = M(b_x, b_y) = 1, \\ a_x, b_x \in X, a_y, b_y \in Y\}, \tag{8}$$

and the define the periodic pair distances in the x and y directions by

$$i = \min\left\{|a_x - b_x|,\, X - |a_x - b_x|\right\}, \tag{9}$$

$$j = \min\left\{|a_y - b_y|,\, Y - |a_y - b_y|\right\}, \tag{10}$$

respectively. The subsets of agent paired distances are given by

$$S_i = \left\{(a, b) \mid i = 1, \ldots, \left\lceil \frac{X}{2} - 1 \right\rceil, (a, b) \in \psi_{ab}\right\}, \tag{11}$$

and

$$S_j = \left\{ (a,b) \mid j = 1, \ldots, \left\lceil \frac{Y}{2} - 1 \right\rceil, (a,b) \in \psi_{ab} \right\}. \tag{12}$$

The counts of periodic paired distances are given by

$$c_x(i) = |S_i| \quad \text{and} \quad c_y(j) = |S_j|, \tag{13}$$

which are normalised to obtain the pair-correlation functions

$$p_x(i) = \frac{c_x(i)}{Y^2 X \rho\hat{\rho}} \quad \text{and} \quad p_y(j) = \frac{c_y(j)}{X^2 Y \rho\hat{\rho}}, \tag{14}$$

where the subscripts denote the x and y direction. The modified density

$$\hat{\rho} = \frac{n-1}{XY-1} = \frac{\rho - 1/XY}{1 - 1/XY} \tag{15}$$

in Eqn. 14 represents the conditional probability of selecting the second agent in the agent pair given that the probability of selecting the first agent is the usual density ρ, as is the case in a volume exclusion process [5, 8]. Finally, for a square domain with $i = j$ and $X = Y$, we define the average periodic pair-correlation function

$$F(i) = \frac{1}{2N} \sum_{s=1}^{N} p_x^s(i) + p_y^s(i), \tag{16}$$

where p_x^s and p_y^s are pair correlation functions in the x and y direction from the s^{th} realisation in a total of N identically prepared realisations of the same stochastic process.

The ability of the function to accurately identify the ECSR state was verified by evaluating it for spatial domains that were populated uniformly at random with agents (*i.e.* for the initial condition used in the simulations). We found that the signal fluctuated around unity at all pair distances, and these fluctuations decreased as the sample number N was increased (not shown).

We note that the sample size of the periodic counts of Eqn. 13 are on average the same at any pair distance. This is shown by the constant normalisation factor in Eqn. 14, for spatial distributions at the ECSR state. In contrast, the sample size of the non-periodic counts of Binder and Simpson's [8] pair-correlation function decays linearly with increased pair-distance. One advantage of the periodic pair-correlation, Eqn. 16, over the non-periodic pair-correlation function of Binder and Simpson's [8], is that fewer samples are required to characterise the spatial distribution generated by the aggregation mechanisms being studied.

4. Results

We evaluate the average pair correlation function to quantify the spatial patterning in the agent based simulations. We begin with considering the two aggregation mechanisms in isolation from each other; biased motion only, with $P_m = P_b = 1$ and $P_p = 0$, and proliferation only, with $P_m = 0$. The simulations with no unbiased motility [Fig. 4(a) and (c)] produce visibly distinct aggregates. Common to both mechanisms are short length scales of aggregation where the signal of the pair-correlation function is greater than unity [Figs. 4(b) and 4(d)]. At these short length scales the pairs of agents are typically associated with the same distinct aggregate.

Physical intuition leads us to expect that, for the biased motion mechanism, the length scale of aggregates will be approximately L, so that attraction between cells keeps them together within the aggregate. Similarly, we expect that the inter-aggregate distance will tend to be in the range L–$2L$. This is because, if two aggregates were separated by a distance less than L, attraction between the cells on their outer edges would tend to cause them to coalesce. Conversely, between two aggregates spaced more than $2L$ apart, there would be an region outside the range of attraction of the cells of either, in which an intermediate aggregate could form. The pair correlation signal for the biased motion mechanism in Fig. 4(b) contains a maxima at the pair-distance $i \approx 15 = 3L$. This corresponds to pairs of agents typically belonging to distinct, adjacent, neighbouring aggregates, and corresponds to the average separation

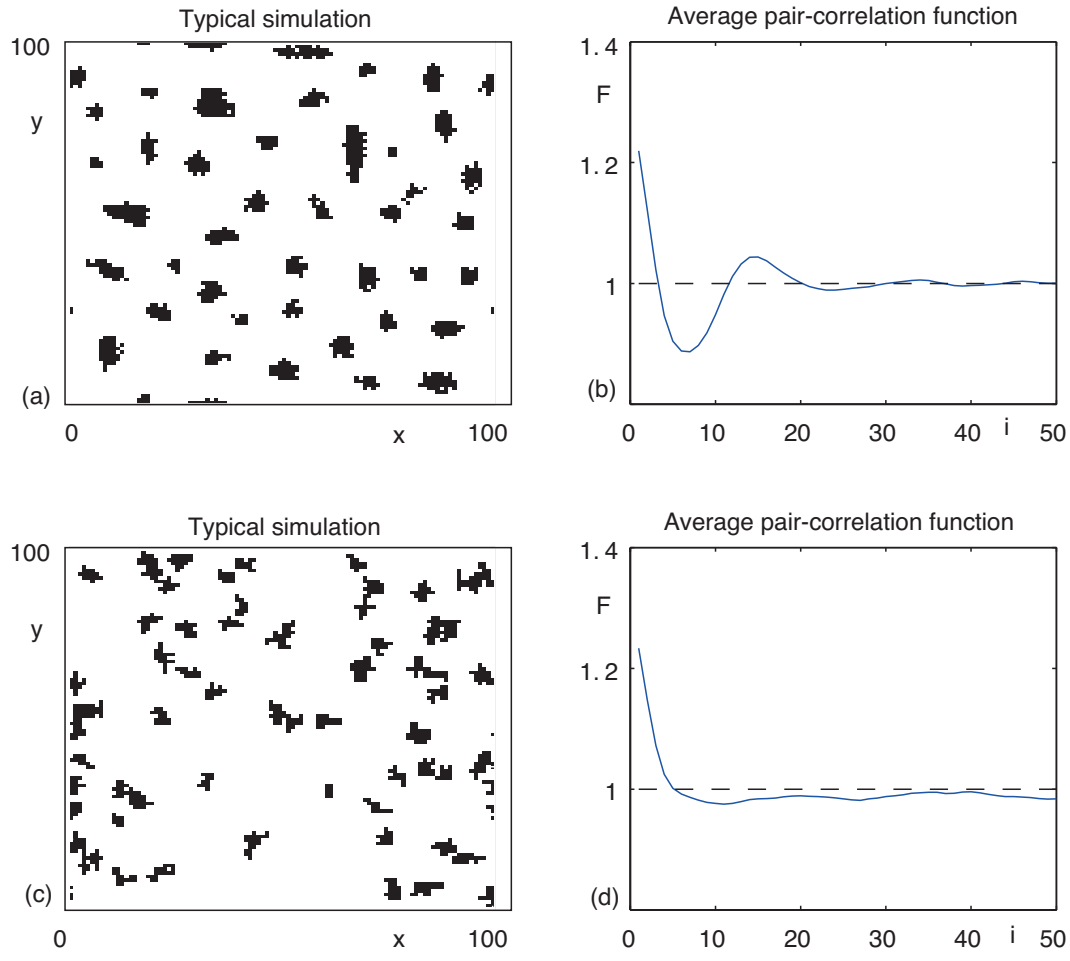

Figure 4: Aggregation patterns from the agent-based models. (a) and (b) Biased motion mechanism. (c) and (d) Rapid proliferation mechanism. (a) Typical simulation, $L = 5$, $P_p = 0$, $P_m = P_b = 1$, $\rho = 0.1$, $t = 190$. (b) Average pair-correlation function for (a), $N = 50$. (c) Typical simulation, $P_m = 0$, $P_p = 1$. The initial density in each simulation is $\rho(0) = 0.007$. The simulations are terminated when $\rho = 0.1$. (d) Average pair-correlation function for (c), $N = 50$.

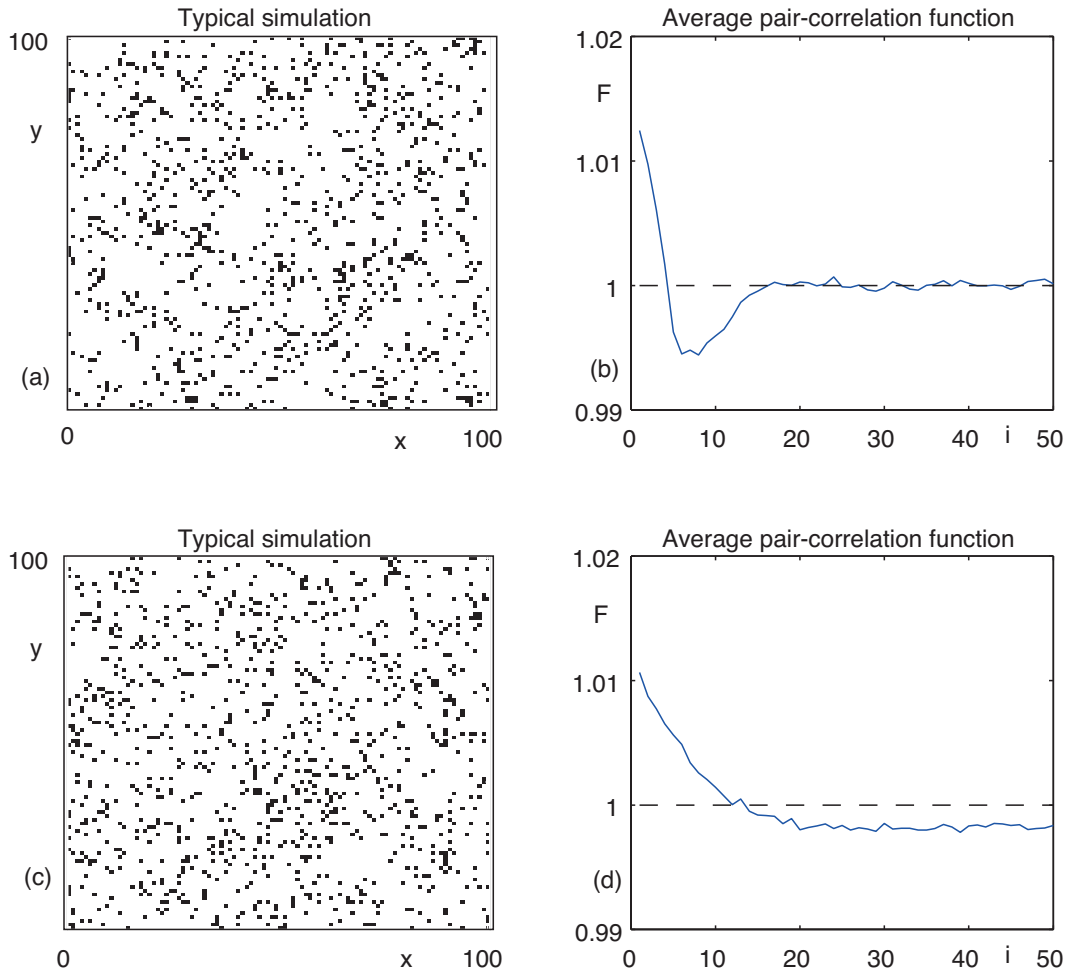

Figure 5: Aggregation patterns from the agent-based models. (a) and (b) Biased motion mechanism (c) and (d) Rapid proliferation mechanism with unbiased motion. (a) Typical simulation, $L = 5$, $P_p = 0$, $P_m = 1$, $P_b = 0.15$, $\rho = 0.1$, $t = 190$. (b) Average pair-correlation function for (a), $N = 1000$. (c) Typical simulation, $P_m = 1$, $P_b = 0$, $P_p = 0.01$. The initial density in each simulation is $\rho(0) = 0.05$. The simulations are terminated when $\rho = 0.1$. (d) Average pair-correlation function for (c), $N = 1000$.

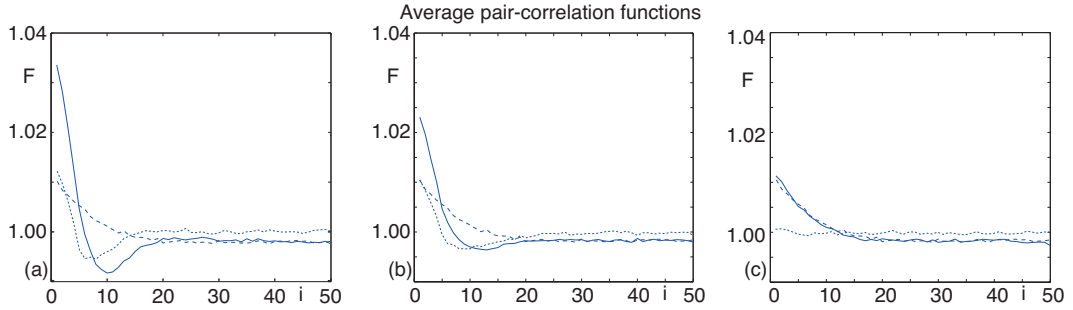

Figure 6: Average pair-correlation functions (solid curves) for the combined biased motion and rapid proliferation mechanisms, with $P_m = 1$, $P_p = 0.01$, $L = 5$, $\rho(0) = 0.05$, $N = 1000$. The simulations were visibly indistinguishable from those shown in Figs. 5(a) and (c), being terminated when $\rho = 0.1$. (a) $P_b = 0.15$. (b) $P_b = 0.1$ (c) $P_b = 0.01$. The dotted curves are for biased motion only, with $P_p = 0$, $\rho = 0.1$ and $t = 190$. The broken curves are for proliferation with unbiased motion only, with $P_b = 0$.

distance between these neighbouring aggregates. This is consistent with the search radius of $L = 5$ and the minimum of the signal at the pair distance $i \approx 7.5 = 3L/2$ where maximum segregation occurs. In contrast, the pair correlation signal for the rapid proliferation mechanism simulations confirms that there is no spatial structure at intermediate length scales, as the pair correlation function decays monotonically to a constant value [Fig. 4(d)]. For both the biased motion and rapid proliferation mechanisms we observe no spatial structure at large length scales [Fig. 4(b) and 4(d)].

Comparing the pair correlation signals Fig. 4(b) and (d) provides a quantitative framework for distinguishing between the two spatial patterns associated with these two aggregation mechanisms. The sensitivity of the pair-correlation function to detect the aggregation mechanism is further demonstrated in Fig. 5, where the spatial distribution of agents are visually indistinguishable owing to the increase in unbiased motility. We note that a larger number of realisations is required to determine a reliable average signal when the amount of random motion is increased in the simulations. A practical implication of this is that it maybe infeasible to use the pair-correlation function to analyse images of cell-based experiments which are dominated by unbiased or random cell motion.

When both aggregation mechanisms operate together the characteristic shape of the pair-correlation function [solid curves, Fig. 6] can provide an indication of which mechanism is dominating the aggregation in the simulations. This is further illustrated on comparison with signals [dotted and broken curves, Fig. 6] for the two isolated aggregation mechanisms, with the three signals being distinguishable in Figs. 6(a) and (b). However, some care needs to be taken when interpreting signals, as the pair-correlation function may not always be able to detect the individual mechanisms of aggregation when multiple mechanisms are acting and no one particular mechanism dominates [solid and broken curves, Fig. 6(c)].

Given that we have been able to make a distinction between two very different mechanisms which can lead to aggregates which appear to be very similar, we repeat the analysis for $N = 17$ experimental images of highly proliferative breast cancer cells [Fig. 1(a)]. The experiments were performed by placing the cells uniformly, at random, on a two-dimensional substrate and then imaging the cell population with time [22]. The location [Fig. 7(a)] and area [Fig. 7(b)] of the cells in the images was estimated using Matlab's imaging processing toolbox [8, 22, 24]. With this data we calculated the average area per cell μ and the average length scale $\sqrt{\mu}$. The average length scale was used to non-dimensionalise the experimental images so that the population of cells could be represented using an integer lattice [Fig. 7(c)]. As expected, the average pair-correlation function [Fig. 7(d)] for these experimental images [Fig. 7(c)] reflects the same characteristic shape that we observed for the rapid cell proliferation mechanism [Fig. 4(d) and 5(d)].

5. Discussion

In this paper, we have implemented new agent-based models to simulate the formation of cell aggregates by two different mechanisms. The aim of our study was to determine if these different mechanisms produced different spatial patterns of aggregates, and thus if it was possible to infer the mechanism of aggregation by analyzing the spatial distribution. To investigate this possibility, we introduced a pair correlation function, and showed it is a potentially powerful tool for characterising and quantifying such spatial patterns. In particular, the pair correlation function is sufficiently sensitive enough to identify cell distributions that are not at the ECSR state, even when this is not obvious through visual inspection. Furthermore, we found that the pair correlation function could reliably distinguish between the aggregation patterns produced by different mechanisms even when the spatial pattern appeared to be very similar. This suggests that the pair-correlation function could be a useful tool which can be used to identify and distinguish between different aggregation mechanisms, providing insight into biological processes that may otherwise be unclear. We have taken the first steps towards demonstrating the practicality of such an approach by analysing experimental images, such as those shown in Figure 7; our results were consistent with aggregates forming as a result of cell proliferation, which is the accepted mechanism in this case.

Ideally, to confirm the utility of the pair correlation function for determining the precise details of the aggregation mechanisms, it should be used to analyse images where the underlying aggregation mechanism is unknown *a priori*, in order to remove any form of potential experimenter bias. The theoretical prediction should then be separately verified by experimental methods, *e.g.* inhibition of cell proliferation, or use of Boyden chamber assays where chemotaxis is implicated. For each type of mechanism, such verification experiments ought to be repeated under various conditions, such as varying the cell line or varying the initial density of cells in the experiment. Clearly, an extensive verification process is a considerable undertaking and has not been attempted here.

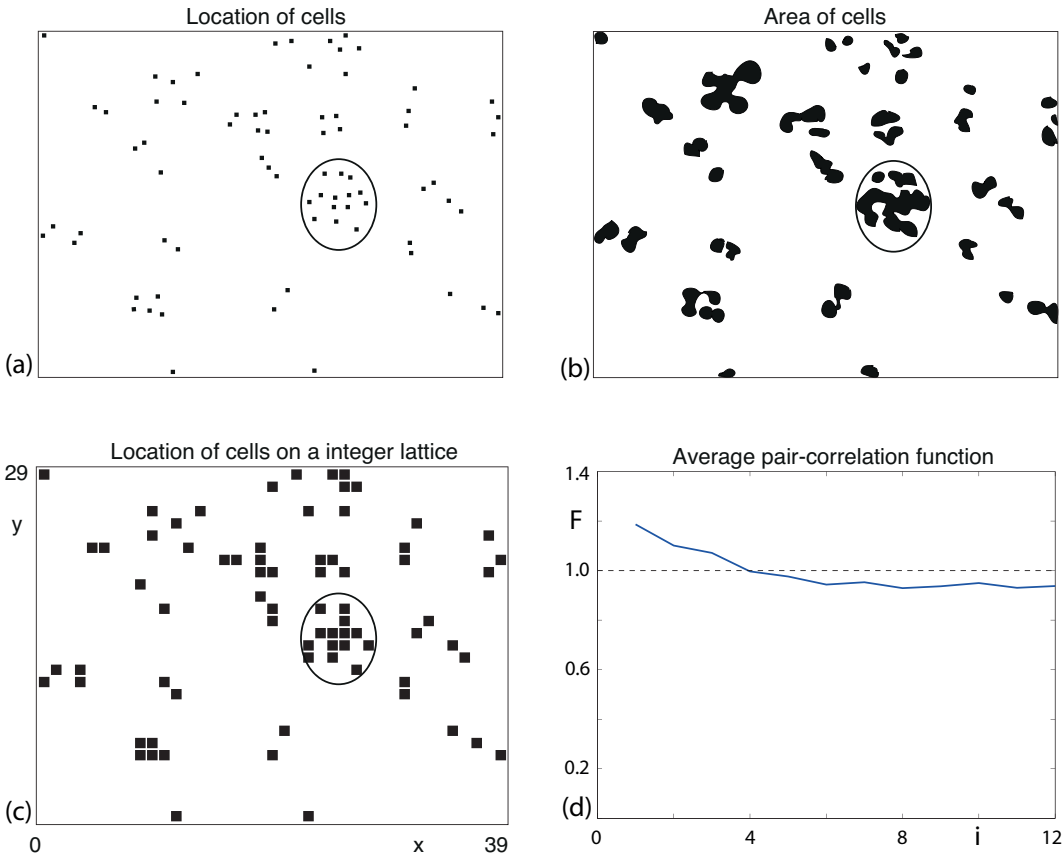

Figure 7: Spatial analysis of breast cancer cells at time $t = 17$hrs. (a)–(c) Image processing for the experimental image in Fig. 1(a). (a) Location of cells. (b) Area of image occupied with cells. (c) Locations of unit square agents on a lattice, representing the cells in Fig. 1(a). (d) Average pair-correlation function, $N = 17$.

There are several options for extending this work with an obvious extension being the consideration of other mechanisms, such as incorporating cell adhesion, cell death and cell processes within a developing tissue [3, 7, 21]. Similarly, this work has only considered cell populations composed of a single cell type; however, the formation of aggregates consisting of two or more cell types is also of interest [2, 14, 19, 25]. A notable feature of these multispecies systems is that, depending upon the strengths of cell–to–cell attraction, different distributions of the various subpopulations are possible within the aggregates, *e.g.* intermixing or segregation. In the future, we aim to extend our pair correlation function analysis to deal with discrete simulations and experimental images of such multispecies processes with the aim of investigating the degree to which the underlying mechanism can be inferred from the observed spatial pattern. This could be achieved by implementing inferencing algorithms, for example approximate Baysian computation [20], that use the pair-correlation function as a summary statistic, to estimate the rates involved in the biological system.

Acknowledgments

We acknowledge support from an Australian Research Council (ARC) Discovery Early Career Researcher Award (DE130100031) to JEFG, and an ARC Discovery Grant (DP120100551) to MJS. BJB was supported by an Australian Government National Health and Medical Research Council Project Grant (APP1069757), and TMB by a Vacation Research Scholarship from the Australian Mathematical Sciences Institute.

References

[1] M. S. Alber, Y. Jiang, and M. A. Kiskowski. Lattice gas cellular automaton model for rippling and aggregation in myxobacteria. *Physica D*, 191(3-4):343358, 2004.
[2] N. J. Armstrong, K. J. Painter, and J. A. Sherratt. A continuum approach to modelling cell-cell adhesion. *J. Theor. Biol.*, 243 (1):98–113, 2006.
[3] R. E. Baker and M. J. Simpson. Correcting mean-field approximations for birth-death-movement processes. *Phys. Rev. E.*, 82:041905, 2010.
[4] B. J. Binder and K. A. Landman. Exclusion processes on a growing domain. *J. Theor. Biol.*, 259:541–551, 2009.
[5] B. J. Binder and K. A. Landman. Quantifying evenly distributed states in exclusion and nonexclusion processes. *Phys. Rev. E*, 83:041914, 2011.
[6] 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. *Bull. Math. Biol.*, 74:474–490, 2011.
[7] B. J. Binder, K. A. Landman, M. J. Simpson, M. Mariani, and D. F. Newgreen. Modeling proliferative tissue growth: A general approach and an avian case study. *Phys. Rev. E.*, 78:031912, 2008.
[8] B. J. Binder and M. J. Simpson. Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions. *Phys. Rev. E*, 88:022705, 2013.
[9] F. Cardarelli and E. Gratton. *In vivo* imaging of single-molecule translocation through nuclear pore complexes by pair correlation functions. *PLoS One*, 5(5):10475, 2010.
[10] D. Chowdhury, A. Schadschneider, and K. Nishinari. Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms. *Phys. Life Rev*, 2:318–352, 2005.
[11] E. A. Codling, M. J. Plank, and S. Benhamou. Random walk models in biology. *J. R. Soc. Interface*, 5:813–834, 2008.
[12] U. Dieckmann, R. Law, and J. A. J. Metz. *The geometry of ecological interactions: simplifying spatial complexity.* Cambridge University Press, 2000.
[13] P. J. Diggle. *Statistical analysis of spatial point patterns.* Arnold, 1983.
[14] J. E. F. Green, S. L. Waters, J. P. Whiteley, L. Edelstein-Keshet, K. M. Shakesheff, and H. M. Byrne. Non-local models for the formation of hepatocyte - stellate cell aggregates. *J. Theor. Biol.*, 267(1):106–120, 2010.
[15] E. J. Hackett-Jones, K. J. Davies, B. J. Binder, and K. A. Landman. Generalized index for spatial data sets as a measure of complete spatial randomness. *Phys. Rev. E.*, 85:061908, 2012.
[16] T. Hofer, J. A. Sherratt, and P. K. Maini. Cellular pattern formation during *Dictyostelium* aggregation. *Physica D*, 85:425–444, 1995.
[17] R. Law, D. J. Murrell, and U. Dieckmann. Population growth in space and time: spatial logistic equations. *Ecology*, 84:252–262, 2003.
[18] Savill N.J. and Sherratt J.A. Control of epidermal stem cell clusters by notch-mediated lateral induction. *Dev. Biol.*, 258:141–153, 2003.
[19] K. J. Painter and J. A. Sherratt. Modelling movement of interacting cell populations. *J. Theor. Biol.*, 225:327–339, 2003.
[20] D. B. Rubin. Bayesianly justifiable and relevant frequency calculations for the applies statistician. *Ann. Stat.*, 12:1151–1156, 1984.
[21] M. J. Simpson, C. Towne, D. L. S. McElwain, and Zee Upton. Migration of breast cancer cells: Understanding the roles of volume exclusion and cell-to-cell adhesion. *Phys. Rev. E.*, 82:041901, 2010.
[22] M.J. Simpson, B.J. Binder, P. Haridas, B.K. Wood, K.K. Treloar, D.L.S. McElwain, and R. E. Baker. Experimental and modelling investigation of monolayer development with clustering. *Bull. Math. Biol.*, 75:871–889, 2013.

[23] M.J. Simpson, A. Merrifield, K.A. Landman, and B.D. Hughes. Simulating invasion with cellular automata: Connecting cell-scale and population-scale properties. *Phys. Rev. E*, 76:021918, 2007.

[24] M.J. Simpson, K.K. Treloar, B.J. Binder, P. Haridas, K.J. Manton, D.I. Leavesley, D.L.S. McElwain, and R. E. Baker. Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *J. R. Soc. Interface*, 10:20130007, 2013.

[25] R. J. Thomas, A. Bennett, B. Thomson, and K. M. Shakesheff. Hepatic stellate cells on poly(DL-lactic acid) surfaces control the formation of 3D hepatocyte co-culture aggregates *in vitro*. *Euro. Cell. Mater.*, 11:16–26, 2006.

[26] B. Vasiev and C. J. Weijer. Modelling of *Dictyostelium discoideum* slug migration. *J. Theor. Biol.*, 223:347–359, 2003.

[27] W.R. Young, A.J. Roberts, and G Stuhne. Reproductive pair correlations and the clustering of organisms. *Nature*, 412(6844):328–331, 2001.