Quantitative approaches to investigating epithelial morphogenesis

Jochen Kursawe

Somerville College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2017

Abstract

Morphogenesis - the generation of biological shape and form - is fascinating, and its study promises to shed light on a wide range of developmental defects and inform strategies for the artificial growth of organs. Recently, the experimental study of morphogenesis has thrived due to a rise in quantitative methods. The resulting avalanche of quantitative data requires us to rethink the scientific method. We need to design quantitative hypotheses through mathematical models, make quantitative experimental predictions, devise methods for quantitative data analysis, and design methods for quantitative inference using models and data. This thesis aims to enable this transition for the integrative analysis of morphogenesis in epithelia, one of the major tissue types in animals. We begin by conducting the first systematic numerical analysis of a widely used cell-based model of epithelia, the vertex model, and estimate to what extent quantitative model predictions may be influenced by parameter values and implementation details. We then apply this model to a key question in developmental biology by constructing a quantitative theory for tissue size control in the embryonic epidermis of the fruit fly *Drosophila*, using the model to predict the outcomes of future experiments. Subsequently, we devise a method for estimating mechanical parameters of vertex models from imaging data and quantifying the uncertainty associated with such estimates. Finally, we propose a novel algorithm for robust cell tracking in live-imaging microscopy videos of epithelial tissues that illustrates how graph theoretic concepts may be used to overcome challenges in quantitative data analysis. Together, the contributions in this thesis will enable the quantitative study of epithelia for a wide range of applications.

Acknowledgements

I would like to thank my supervisors Ruth Baker and Alexander Fletcher for their tireless support throughout the last four years. I am deeply grateful for the countless times when you were there to help me out, and that you went out of your way to patiently teach me in so many aspects of academic work. Working with you was immensely enjoyable and I will miss our regular meetings.

I would like to thank Philip Maini for supporting me during the application process and helping me decide on my DPhil topic. Without your generous support this thesis would not have been possible.

I would also like to thank Radek Erban for supervising me in the first months of my thesis and I further extend my gratitude to the Engineering and Physical Sciences Research Council for funding my studies.

My stay at the maths department was enlightened by my office mates Aaron, Linus, Jessica, Jackie, Abdullah, and Paul. Thank you all for your support and friendship, and for making my time at the office such a cheerful experience independent of the time of day.

My time in Oxford was enriched by lasting friendships with colleagues and housemates. Thank you for the fun times, witty conversation and all the support Doireann, Pete, Chii Fen, Zachary, Teresa, Christoph, Anne, Nele, Kate, and Hannah. I am looking forward to many more shared meals and activities.

I would like thank my friends from home, Benjamin, Sarah, Julian, Lenka, Maria, Annemarie, Tobias, and Kristina for their close friendship despite the long distance.

Am meisten bedanke ich mich bei meinen Eltern und bei meiner Schwester Franzi. Danke für eure unermüdete Unterstützung und Liebe, und dafür, dass ihr immer für mich da seid.

Statement of originality

This thesis is submitted to the University of Oxford in partial fulfillment of the requirements of the degree of Doctor of Philosophy. This thesis contains no material which has been previously submitted for a degree or diploma at this University or any other institution. All experimental data presented in this thesis were obtained by Cody Narciso and Pavel A. Brodskiy working in the laboratory of Prof. Jeremiah J. Zartman at the Department of Chemical and Biomolecular Engineering, University of Notre Dame, Indiana. The contents of Chapters 3, 4 and 6 have been published on *bioRxiv*, in *PLoS Computational Biology*, and in the *Journal of the Royal Society Interface*, respectively. This thesis represents my own original work towards this research degree.

Jochen Kursawe

University of Oxford

March 2017

Contents

1 Introduction **8**

2 Modelling preliminaries **14**

2.1 Cell-based models 14

2.2 Vertex models of epithelia 16

3 Numerical analysis of vertex model behaviour **24**

3.1 Background and motivation 24

3.2 Methods: A vertex model of the growing *Drosophila* wing disc 26

3.3 Results: Dependence of modelling results on implementation parameters 31

3.4 Discussion 47

4 Mechanical modelling of tissue size control **55**

4.1 Background and motivation 55

4.2 Methods: A model of embryonic epidermic P compartments in *Drosophila* 60

4.3 Results: Model analysis and predictions 67

4.4 Discussion 86

5 Parameter inference on vertex models **93**

5.1 Background and motivation 93

5.2 Methods: ABC for vertex models 98

5.3 Results: Estimating uncertainty in vertex model parameter inference . 109

5.4 Discussion 123

6 Cell tracking in epithelia **129**

6.1 Background and motivation 129

6.2 Methods: Design of the cell tracking algorithm 133

6.3 Results: Algorithm performance 144

6.4 Discussion 154

7 Discussion **161**

7.1 Contributions to open-source software projects 164

7.2 Conclusions 166

A Mathematical description of the cell tracking algorithm **167**

A.1 Mathematical formulation 168

A.2 Construction of the conserved MCS 170

A.3 Post-processing 174

A.4 Computational implementation 180

Bibliography **181**

Chapter 1

Introduction

Embryogenesis is the fascinating process by which complex organisms develop from a single cell. While our understanding of embryogenesis is continuously expanding, many of the fundamental mechanisms that guide this process are still poorly understood (Wolpert, 2011). Studying embryogenesis is extremely promising, since unravelling the mysteries of embryonic development will help us to treat developmental defects and artificially engineer organs and tissues. One key question in developmental biology is how embryonic tissues reach, and maintain, the correct size and proportions. Tissue size control is incredibly robust to external or genetic perturbations. For example, tetraploid salamanders of the species *Amblystoma mexicanum* have half the number of cells as their diploid counterparts, yet are the same size (Vernon and Butsch, 1957). The underlying mechanisms that regulate cell growth, proliferation and death to achieve tissue size control remain largely unknown.

A common experimental method to analyse developmental processes, such as tissue size control, is to employ model organisms, for example mice, the frog *Xenopus laevis*, or the worm *Caenorhabditis elegans* (Wolpert, 2011). It is possible to learn about human development by using such organisms since many developmental processes are conserved across the animal kingdom, and genes in one species often have homologues of the same function in other species (Wolpert, 2011). The fruit fly

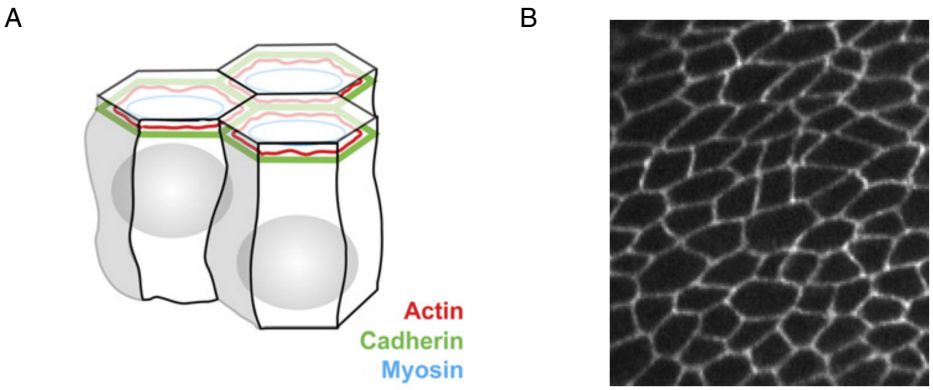

Figure 1.1: Composition of an epithelial monolayer. (A) Schematic of cells in an epithelial monolayer. Cells are tightly bound by molecules at the apical (top) surface, such as E-Cadherin, actin, and myosin. Figure panel adapted from Farhadifar et al. (2007) with permission from Elsevier. (B) Example microscopy image of the apical surface of an epithelial monolayer, the growing *Drosophila* wing. Fluorescent labelling of E-Cadherin renders the cell outlines visible. Figure panel adapted from Staple et al. (2010) with permission from Springer.

Drosophila melanogaster is one of the most prominent and pivotal model organisms of developmental biology. For example, chromosomes were originally identified to be the carriers of genetic information by examining *Drosophila* mutations and the genome of *Drosophila melanogaster* was among the first to be fully sequenced (Adams et al., 2000). The fruit fly has several attributes that make it a well-suited laboratory animal. Fruit flies have quick reproductive cycles, stocks are inexpensive to maintain, and today a large range of techniques are available to study its embryogenesis, including tools for localised genetic perturbations.

Many processes during morphogenesis are governed by the growth and dynamics of epithelia, one of the major tissue types in animals. For example, specification of the anterior-posterior axis in mammals is initiated by directed migration of an epithelial subpopulation (the anterior visceral endoderm) to one side of the embryo (Arnold and Robertson, 2009; Srinivas, 2006), whereas the formation of an epithelial tissue on the outside of the embryo, the epidermis, comprises one of the earliest morphogenetic events in *Drosophila* (Campos-Ortéga and Hartenstein, 1997). Such epithelia form polarized sheets of cells with distinct apical ('top') and basal ('bottom') surfaces,

with tight lateral attachments nearer their apical surface (Figure 1.1).

A recent quantitative transition in data acquisition and analysis paves the way for new insights to developmental processes (Pantazis and Supatto, 2014). While high-resolution imaging and measurement techniques, such as electron microscopy and atomic force microscopy, have existed for several decades (Robinson, 1986; Binnig et al., 1986), their potential to capture material properties and process dynamics has not been fully harnessed by the biological research community. Not only are biological tissues challenging to analyse because of their heterogeneous and dynamic composition, they are also inherently fragile and sensitive to the levels of energy exposure that are traditionally associated with nanoscale measurements (Thach and Thach, 1971). The discovery of green fluorescent protein (GFP) (Shimomura et al., 1962; Shimomura, 2005) and its subsequent application to report gene expression *in vivo* (Chalfie et al., 1994) pioneered a revolution of microscopy techniques that included the development of the confocal, two-photon, and light-sheet microscopes (Pantazis and Supatto, 2014). Using these techniques it is now possible to observe biological processes *in vivo* and *in vitro* with unprecedented detail. This advance in imaging techniques has been accompanied by the development of nanoscale engineering solutions that expand the types of data that can be acquired at small length or time scales. These solutions include, for example, molecules that allow the measurement of forces within living tissues (Meng et al., 2008; Gayrard and Borghi, 2016).

These new experimental techniques generate a host of new quantitative and semi-quantitative data (Pargett and Umulis, 2013). To understand and interpret such data, it is necessary to use mathematical models (Shou et al., 2015; Tomlin and Axelrod, 2007). Models can be used to provide theories for the mechanisms underlying experimental observations and to make predictions for future experiments that may help verify or distinguish these theories. One of the most well-known examples of mathematical modelling in biology is the diffusion-driven instability introduced by

Turing (1952). Turing proposed that heterogeneity in developing embryos may be generated from a uniform initial state by interacting chemicals, and he established his theory using a mathematical model. Sixty years later, a similar mechanism has been shown to regulate digit morphogenesis in mouse limbs (Sheth et al., 2012). Modern applications of computational modelling are increasingly quantitative and include modelling of gene regulatory networks (Davidson and Levin, 2005) and studies that target the mechanical principles of morphogenesis (Guillot and Lecuit, 2013b).

If the data are collected at a cellular scale, cell-based computational models may be applicable (Brodland, 2004; Osborne et al., 2017). Cell-based models have helped understand various embryonic processes that involve the mechanical and biochemical interactions of individual cells. For example, Landsberg et al. (2009) showed that mechanical tension can help maintain boundaries between cells of varying fates, and Monier et al. (2015) showed that the mechanical tension required for furrow formation during *Drosophila* limb development is generated by apoptotic (dying) cells. Both of these examples are applications of vertex models (Fletcher et al., 2014), one of the most widely used class of cell-based model in developmental biology. Vertex models are specifically designed to study epithelia. In Chapter 2, we provide a brief overview of existing cell-based model formalisms, together with a mathematical formulation of the vertex model.

This thesis aims to enable the use of vertex models to study the growth and morphogenesis of epithelia in a quantitative manner. Many cell-based models, such as vertex models, are developed to reproduce qualitative phenomena, for example, cell sorting through differential adhesion between cells (Brodland, 2004). However, it is unclear to what extent cell-based models can be used to analyse biological processes quantitatively. For example, to what extent can computational implementation choices, such as the size of a simulation time step, influence model predictions? We address this question in Chapter 3, where we identify how robust vertex model predic-

tions are to changes in the method of numerical approximation and model parameters that may not have biophysical correlates.

We then apply the vertex model to study tissue size control during embryonic development in Chapter 4. A previous study that analysed tissue size control in *Drosophila* embryos showed that within the skin of the embryo, its epidermis, compartments are of fixed size even in the face of genetic perturbations that alter the number of cells (Parker, 2006). In Chapter 4 we use a vertex model to develop a quantitative theory for these previous observations, and to make predictions that can be tested in future experiments.

The work in Chapter 4 motivates further research questions. The predictions in this chapter are not dependent on the mechanical parameters of the vertex model. However, to model processes that rely on tight mechanical regulation it may be necessary to directly measure the parameters in the vertex model from experimental data. In Chapter 5 we investigate how vertex model parameters can be inferred from imaging data and explore methods to quantify the uncertainty of such parameter estimations.

Chapters 3, 4 and 5 individually contribute to a quantitative transition in different steps of the scientific method (Figure 1.2). We focus on the design of quantitative models in Chapter 3, develop a quantitative theory for tissue size control and make experimental predictions in Chapter 4, and analyse methods for parameter inference in Chapter 5. In Chapter 6 we turn to the most crucial step of scientific investigation, the collection and analysis of experimental data. Mathematical tools can significantly enhance the processing of quantitative data. For epithelia, data analysis often comprises two steps, image segmentation and cell tracking. We show how graph theoretic methods can be used to conduct robust cell tracking in live-imaging microscopy videos of epithelial tissues. Together, Chapters 3, 4, 5 and 6 contribute to the quantitative study of epithelia at all stages of the scientific method.

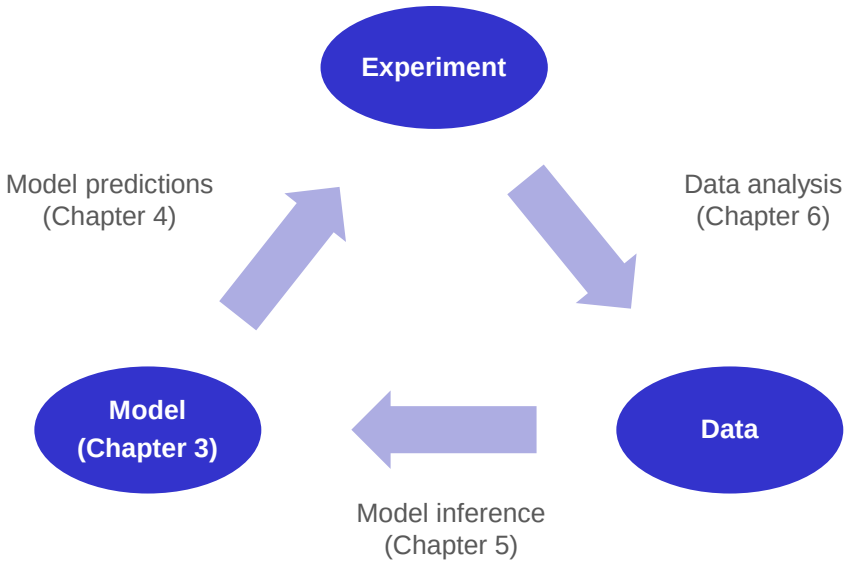

Figure 1.2: This thesis contributes to the quantitative study of epithelia at different stages of the scientific method. The scientific method comprises a cycle of hypothesis building and experimental analysis that iteratively improves our scientific understanding. When the scientific method is applied quantitatively, the hypothesis constitutes a mathematical or computational model. In Chapter 3, we provide a numerical analysis of the well-established vertex model. In Chapter 4, we apply the vertex model to study embryonic tissue size control and make experimental predictions. In Chapter 5, we design a inference method to estimate vertex model parameters. In Chapter 6, we employ graph-theoretical concepts to track cells in live-imaging microscopy videos.

In the next chapter we give a brief overview of existing cell-based models before introducing our implementation of the vertex model, which we shall use in Chapters 3, 4 and 5.

Chapter 2

Modelling preliminaries

2.1 Cell-based models

The characterization of fundamental embryonic mechanisms at a cellular scale is an area of recent progress. The ability to observe how cells interact and self-organise to form organs bears great potential for applications in medicine and synthetic biology. A promising method to form hypotheses about underlying mechanisms is cell-based modelling. Such models can naturally capture stochastic effects and heterogeneity when only few cells are present and can be used to explore tissue behaviour when complex assumptions on the cellular scale prevent straightforward continuum approximations on the tissue scale. The applications of cell-based models range from embryonic development (Farhadifar et al., 2007; Mao et al., 2011; Trichas et al., 2012; Monier et al., 2015; Atwell et al., 2015), to wound healing (Walker et al., 2004) and tumour growth (Anderson et al., 2006).

Multiple cell-based modelling approaches exist. In cellular automata (Ermentrout and Edelstein-Keshet, 1993; Lee et al., 1995), each cell occupies a site on a lattice, and different rules can be implemented for the proliferation, motion, and interaction of the cells. In cellular Potts models each cell is allowed to occupy multiple lattice sites, and energy minimisation is used to propagate the shape of each cell over time.

The cellular Potts model has been used to study biological processes ranging from cell sorting (Glazier and Graner, 1993) and morphogenesis (Izaguirre et al., 2004) to tumour growth (Shirinifard et al., 2009).

Off-lattice approaches to cell-based modelling include centre-based models. Cell shapes are either assumed to be spherical (Drasdo and Höhme, 2005) or polygonal (Meineke et al., 2001). In the latter case, polygonal shapes are derived from the cell centre locations by Voronoi tessellations. It is possible to model more realistic cell shapes by modelling the motion of points along the boundary of the cell. This concept is applied in vertex models (Fletcher et al., 2014) and in the immersed boundary method (Rejniak et al., 2004; Cooper et al., 2016). In the former, epithelia are modelled as a tessellation of polygons or polyhedra, and vertices move due to forces originating from the cells and their interactions. In the latter, cell boundaries are represented as a set of points that move like elastic membranes immersed in a fluid. Note that complex cell shapes can also be modelled on-lattice with the cellular Potts models mentioned earlier by considering an energy equation that takes the shape of the cell boundaries into account (Scianna and Preziosi, 2016). Other cell-based models take the subcellular composition of cells into account, such as the subcellular element model (Newman, 2007) or the finite element vertex model (Brodland et al., 2007).

Each cell-based modelling approach has specific advantages and disadvantages. For example, the immersed boundary method allows a detailed representation of cell shapes, but this benefit comes at an increased computational cost in comparison to other methods. Cellular Potts models are very versatile in the possible effects that can be modelled, but may suffer from lattice artifacts (Marée et al., 2007). When modelling a specific application, it is necessary to weigh the benefits of the existing cell-based models against each other in the context of the application. A comprehensive introduction to the various cell-based modelling frameworks, and a comparison of their predictions regarding a variety of benchmark problems, has recently been conducted

by Osborne et al. (2017).

In case of epithelial monolayers, the most commonly used model framework is the two-dimensional vertex model. Its main advantage is that it directly predicts polygonal cell patterns, the summary statistics of which may be directly compared to experimental data. The vertex model has an advantage over centre-based models by explicitly modelling cell rearrangement and thus allowing larger variety of cell shapes than those that can be described by Voronoi tessellations. In centre-based models cell shapes are solely determined by the locations of the cell centres. The vertex model also has advantages over cellular Potts models since it is lattice-free, and it is possible to explicitly include time as a parameter, which facilitates, for example, the modelling of cell cycle progression. Simulations of cellular Potts models propagate by energy minimisation, and time differences between simulation events need to be inferred implicitly, for example through differences in cell centre positions.

2.2 Vertex models of epithelia

Analysis of cell packing in epithelial tissues often comprises polygonal approximations (Gibson et al., 2006; Farhadifar et al., 2007; Sánchez-Gutiérrez et al., 2016; Escudero et al., 2011) (Figure 2.1A). Here, we give a brief introduction to vertex models, which model the evolution of such polygonal representations (Figure 2.1B).

Vertex models were first introduced to study inorganic structures, such as foams (Nagai et al., 1988) or grain boundaries (Kawasaki et al., 1989; Torres et al., 2015), and have since been applied to study a variety of epithelial tissues (Landsberg et al., 2009; Farhadifar et al., 2007; Nagai and Honda, 2001; Aegerter-Wilmsen et al., 2010; Mao et al., 2013; Sugimura and Ishihara, 2013). Multiple further studies focussed on the *Drosophila* wing imaginal disc (Aigouy et al., 2010; Aegerter-Wilmsen et al., 2010, 2012; Mao et al., 2011), analysing, for example, how uniform tissue growth can be achieved despite non-uniform cell signalling (Hufnagel et al., 2007), or the control

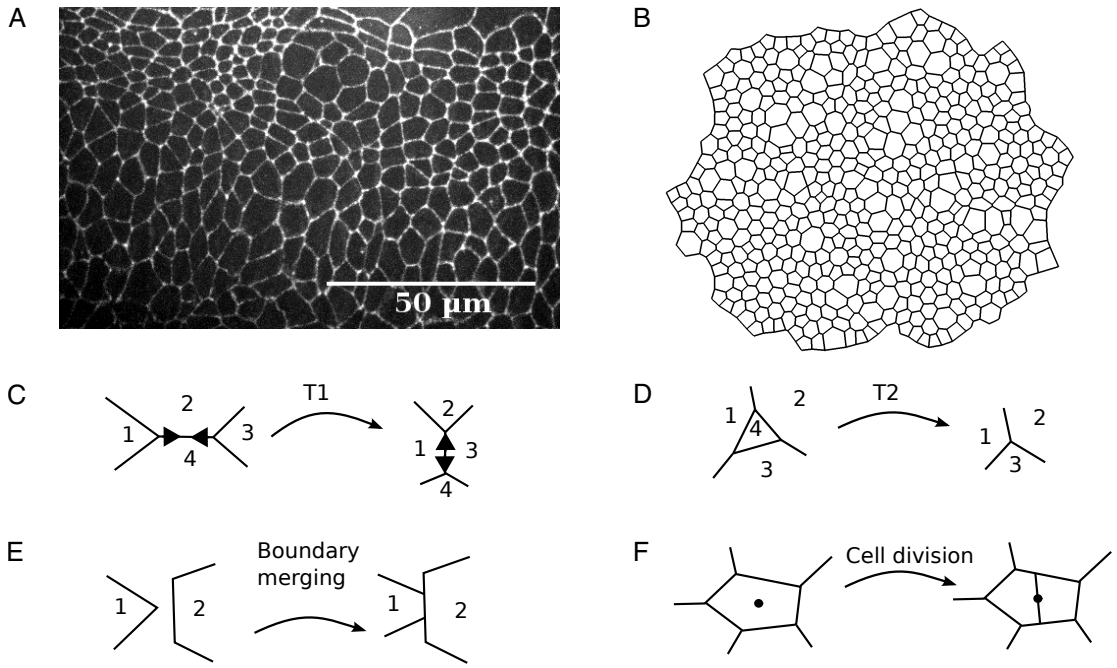

Figure 2.1: Two-dimensional vertex models represent cells in an epithelial tissue as polygons and allow different types of vertex rearrangement. (A) Example microscopy image of an epithelial tissue, the *Drosophila* embryonic epidermis. Cell shapes are distinctly polygonal. Experimental details on the acquisition of this specific image are provided in Chapter 6. (B) Snapshot of a vertex model simulation of a growing tissue. Cell shapes are always polygonal and vertices move due to forces originating from the cells. Details on this specific simulation and the choices of parameters are provided in Chapter 3. Throughout the simulation, vertices may rearrange by T1 transitions (C), T2 transitions (D), boundary merging (E), and cell division (F).

of mitotic cleavage planes, i.e. the direction and formation of newly created cell-cell interfaces (Mao et al., 2011; Patel et al., 2009; Gibson et al., 2011). In other studies, vertex models were used to understand cell sorting and the formation of straight boundaries between tissues of varying fates (Landsberg et al., 2009; Aliee et al., 2012; Canela-Xandri et al., 2011; Schilling et al., 2011; Salbreux et al., 2012), or how tissue size can be controlled by cellular decisions (Hufnagel et al., 2007; Kursawe et al., 2015; Wartlick et al., 2011a,b). Further applications include collective cell migration (Bi et al., 2014; Trichas et al., 2012), cell extrusion and death (Marinari et al., 2012; Monier et al., 2015), wound healing (Nagai and Honda, 2009), and convergent extension (Rauzi et al., 2008; Collinet et al., 2015; Tetley et al., 2016). The represen-

tation of cells as a polygonal packing that is characteristic of vertex models has also been used in attempts to infer force strengths of cell-cell interactions from geometric packings (Chiou et al., 2012; Ishihara and Sugimura, 2012; Ishihara et al., 2013). Reviews of vertex models and their applications are available in (Fletcher et al., 2014; Guillot and Lecuit, 2013b; Honda and Nagai, 2015).

In two-dimensional vertex models, epithelial cell sheets are approximated by tessellations of polygons representing cell apical surfaces, and vertices (corresponding biologically to adherens junctions where three or more cells meet) move in response to forces due to growth, interfacial tension and hydrostatic pressure within each cell (Figure 2.1B). For many applications, a two-dimensional approximation is valid since the forces typically governing apical cell shapes, such as those seen in Figure 2.1, are considered to originate from molecules located near the cell apical surfaces, such as E-Cadherin or myosin (Farhadifar et al., 2007; Guillot and Lecuit, 2013a). Vertex models typically include cell growth and proliferation. In addition, cells exchange neighbours through so-called T1 transitions (Figure 2.1C) whenever the length of a cell-cell interface falls below a threshold, and any triangular cell whose area falls below a threshold is removed by a so-called T2 transition (Figure 2.1D).

A common approach in vertex models is to consider forces on vertices arising as a result of minimizing the total stored energy in the tissue. The functional form for this total stored energy varies between applications, but is typically chosen to reflect the effect of the force-generating molecules which localise at or near the apical surface. This energy function is then used either to derive forces that feed into a deterministic equation of motion for each vertex, which must be integrated over time (Mao et al., 2011; Canela-Xandri et al., 2011; Mao et al., 2013), or else minimized directly assuming the tissue to be in quasistatic mechanical equilibrium at all times (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010). A third approach is to apply a Monte Carlo algorithm to find energy minima (Etournay et al., 2015; Ray et al., 2015). Throughout

this thesis we use the first of these approaches and use force equations to propagate vertex positions over time. This choice is convenient, since it allows us to accurately describe the evolution of the tissue over time without assuming that the tissue is in mechanical equilibrium at all times.

2.2.1 Equations of motion

The terms in the model account for the mechanical effect of the force-generating molecules that accumulate in the apical surface of the cells, such as actin, myosin, and E-cadherin (Guillot and Lecuit, 2013a; Figure 1.1). Vertices correspond to adherens junctions, and throughout this thesis their positions are propagated over time using an overdamped force equation, reflecting the assumption that adherens junctions are not associated with a momentum. The force equation takes the form

$$\mu \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} = -\boldsymbol{\nabla}_i E, \quad i = 1, \ldots, N. \tag{2.1}$$

Here, μ is the friction strength (which we assume to take the same constant value for all vertices), t is time, $\mathbf{x}_i$ is the position vector of vertex i, and E denotes the energy of the whole system. The total number of vertices in the system, N, may change over time due to cell division and cell removal. The symbol $\boldsymbol{\nabla}_i$ denotes the gradient operator with respect to the coordinates of vertex i. The forces act to minimise a phenomenological energy function, based on the contributions thought to dominate epithelial mechanics (Farhadifar et al., 2007):

$$E = \sum_{\alpha} \frac{K}{2}(A_\alpha - A_{0,\alpha})^2 + \sum_{\langle i,j \rangle} \Lambda l_{i,j} + \sum_{\alpha} \frac{\Gamma}{2} P_\alpha^2. \tag{2.2}$$

Here, the first sum runs over every cell in the sheet, A_α denotes the apical surface area of cell α and $A_{0,\alpha}$ is its preferred area, or target area. This energy term penalises deviations from a target area for individual cells, thus imposing cellular bulk elasticity.

The second sum runs over all edges $\langle i, j \rangle$ in the sheet and penalizes long edges (we choose $\Lambda > 0$), thus representing the combined effect of E-cadherin, myosin and actin at the binding interface between two cells. The third sum also runs over all cells, and P_α denotes the perimeter of cell α. This term models the effect of a contractile acto-myosin cable along the perimeter of each cell (Farhadifar et al., 2007). The parameters K, Λ and Γ govern the strength of the individual energy contributions. Although this description of cell mechanics is phenomenological, a variety of studies have demonstrated its ability to match observed junctional movements and cell shapes in epithelial sheets through validation against experimental measurements (Farhadifar et al., 2007; Mao et al., 2013; Landsberg et al., 2009). Note that, while K and Γ are always positive, Λ can in general assume positive as well as negative values. Negative values of Λ can be interpreted as adhesion at cell-cell interfaces dominating over contractile tension along the interface, which could, for example, be generated through contractile myosin molecules.

Before solving the model numerically, we non-dimensionalise it to reduce the number of free parameters (Farhadifar et al., 2007). Rescaling space by a characteristic length scale, L, chosen to be the typical length of an individual cell, and time by the characteristic timescale, $T = \mu/KL^2$, equations (2.1) and (2.2) become

$$\frac{\mathrm{d}\mathbf{x}'_i}{\mathrm{d}t'} = -\boldsymbol{\nabla}'_i E', \tag{2.3}$$

$$E' = \sum_\alpha \frac{1}{2}(A'_\alpha - A'_{0,\alpha})^2 + \sum_{\langle i,j \rangle} \overline{\Lambda} l'_{i,j} + \sum_\alpha \frac{\overline{\Gamma}}{2} P'^2_\alpha, \tag{2.4}$$

where $\mathbf{x}'_i$, A'_α, $A'_{0,\alpha}$, $l'_{i,j}$ and P'_α denote the rescaled i^{th} vertex positions, the rescaled area and target area of cell α, the rescaled length of edge $\langle i, j \rangle$, and the rescaled cell perimeter of cell α, respectively. The symbol $\boldsymbol{\nabla}'_i$ denotes the gradient with respect to the rescaled i^{th} vertex position. In the non-dimensionalised model, cell shapes are governed by the rescaled target area of each cell, $A'_{0,\alpha}$, and the rescaled me-

chanical parameters, $\overline{\Lambda}$ and $\overline{\Gamma}$. For these parameters we use previously proposed values (Farhadifar et al., 2007), unless stated otherwise.

To solve equations (2.3) and (2.4) numerically we use a forward Euler scheme

$$\mathbf{x}'_i(t' + \Delta t') = \mathbf{x}'_i(t') - \boldsymbol{\nabla}'_i E'(t')\Delta t', \tag{2.5}$$

where E' depends on all rescaled vertex positions $\mathbf{x}'_i$ as well as the rescaled time t'. We perform a convergence analysis in Chapter 3 to verify that this discretization is suitable.

2.2.2 Cell neighbour exchange, removal and proliferation

T1 transitions (Figure 2.1C) are executed whenever the length of a given edge decreases below the threshold $l'_{\mathrm{T1}} = 0.01$. The length of the new edge, $l_{\mathrm{new}} = \rho\, l_{\mathrm{T1}}$ ($\rho = 1.5$), is chosen to be slightly longer than this threshold to avoid an immediate reversion of the transition.

A second topological rearrangement in vertex models is a T2 transition, during which a small triangular cell or void is removed from the tissue and replaced by a new vertex (Figure 2.1D). In our implementation any triangular cell is removed if its area drops below the threshold $A'_{\mathrm{T2}} = 0.001$. Note that in equation (2.2) the bulk elasticity or area contribution of a cell α is finite even when the area A_α is zero, allowing individual cells to become arbitrarily small if this is energetically favourable. As cells decrease in area they typically also reduce their number of sides. Hence it is sufficient to remove only small triangular cells instead of cells with four or more sides (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010; Mao et al., 2011).

The energy function (2.2) in conjunction with T2 transitions can be understood as a model for cell removal: cells are extruded from the sheet by a T2 transition if they become mechanically unstable. Note that we do not discriminate between cell removal by cell death or by delamination, since this distinction is immaterial for our purposes.

However, delamination has been shown to provide an alternative way of cell removal from epithelial sheets that is distinct from apoptosis (controlled cell death) (Marinari et al., 2012; Eisenhoffer et al., 2012). Rates of cell removal predicted by previous vertex model applications have coincided with experimental measurements in the *Drosophila* wing imaginal disc (Farhadifar et al., 2007) and notum (Marinari et al., 2012).

We further model the merging of overlapping tissue boundaries (Figure 2.1E) to prevent the occurrence of self-intersecting tissues. Whenever two boundary cells overlap, a new edge of length l_{new} is created that is shared by the overlapping cells. In cases where the cells overlap by multiple vertices, or if the same cells overlap again after a previous merging of edges, the implementation ensures that two adjacent polygons never share more than one edge by removing obsolete vertices. The merging of boundary edges is discussed in further detail by Fletcher et al. (2013).

At each cell division event, a new edge is created that separates the newly created daughter cells (Figure 2.1F). The new edge is drawn along the short axis of the polygon that represents the mother cell (Fletcher et al., 2013). The short axis has been shown to approximate the division direction (cleavage plane) of cells in a variety of tissues (Hofmeister, 1863), including the *Drosophila* wing imaginal disc (Gibson et al., 2011; Li et al., 2012). Tricellular junctions are part of a shape-sensing mechanism that determines the cleavage plane orientation *in vivo* (Bosveld et al., 2016). The short axis of a polygon crosses the centre of mass of the polygon, and it is defined as the axis around which the moment of inertia of the polygon is maximised. Each daughter cell receives half the target area of the mother cell upon division.

The timings of division and cellular growth, as well as the exact choice of boundary and initial conditions, depend on the specific application and hence are not discussed here. Typically, growth is modelled by making the preferred area $A_{0,\alpha}(t)$ time-dependent based on implicit cell-cycle models (Fletcher et al., 2013). Examples of biologically informed growth laws, initial conditions and boundary conditions are

Initialize time $t' = 0$;
Generate initial configuration;
while $t' < t'_{\text{tot}}$ **do**
1. Update cell target areas;
2. Perform cell division on cells that have reached the end of their cell cycle;
3. Perform any T2 transitions;
4. Perform any T1 transitions;
5. Perform boundary merging;
6. Propagate vertex positions using equation (2.5);
7. Increment time by $\Delta t'$;

end

Algorithm 1: Pseudocode of the simulation algorithm.

provided in Chapters 3, 4 and 5.

2.2.3 Computational implementation

We implement the model within Chaste, an open source C++ library that provides a systematic framework for the simulation of vertex models (Fletcher et al., 2013; Mirams et al., 2013). Pseudocode for our implementation is provided in Algorithm 1. Each time step starts by updating the cell target areas. Then, cell division, removal (T2 transitions), rearrangement (T1 transitions), and boundary merging are performed before incrementing the simulation time. The algorithm stops when the end time of the simulation is reached.

Now that we have introduced the vertex model and the algorithm that we use to implement it, we will analyse how the choice of the time step in the numerical approximation, as well as model parameters that do not have a physical correlate, such as thresholds for cell rearrangement, may influence typical model predictions.

Chapter 3

Numerical analysis of vertex model behaviour

Having introduced the vertex model and its implementation in Chapter 2, we now focus on vertex model predictions. Despite the frequent use of vertex models in biology, a rigorous numerical analysis of their implementation is still lacking. In this chapter, we close this gap in the literature by investigating how vertex positions and summary statistics of cell packing may depend on the parameters of numerical approximation, such as the time step, and implementation parameters that do not have explicit biophysical correlates, such as thresholds for cell rearrangement. The contents of this chapter are currently in revision for publication in the *Journal of Computational Physics*.

3.1 Background and motivation

In Chapter 2 we introduced the various types of cell-based models and highlighted the wide range of their applications. Despite the popularity of cell-based models, their numerical solution remains challenging since multi-scale implementations of such models, coupling processes at the subcellular, cellular, and tissue scales, may suffer

from numerical instabilities (Weinan and Engquist, 2003; Miller and Tadmor, 2009) and many such models include parameters of numerical approximation or parameters that have no direct physical correlate. These issues are of growing importance as cell-based models become used in an increasingly quantitative way (Pantazis and Supatto, 2014; Yu and Fernandez-Gonzalez, 2016; Pargett and Umulis, 2013). Here, we analyse the vertex model of cell mechanics and proliferation introduced in Chapter 2, to understand to what extent implementation choices may affect quantitative model predictions.

Previous theoretical analyses of vertex models have elucidated ground state configurations and their dependence on the mechanical parameters of the model (Staple et al., 2010), inferred bulk material properties (Merzouki et al., 2016; Xu et al., 2015a; Davit et al., 2013), and introduced ways to superimpose finite-element schemes for diffusing signals with the model geometry (Smith et al., 2012). Spencer et al. (2017) analysed the stability of fourfold vertices in the vertex model. In other work, vertex models have been compared to lattice-based cellular Potts models and other cell-based modelling frameworks (Osborne et al., 2017; Magno et al., 2015).

In the case of vertex models of grain boundaries (interfaces in polycrystalline materials), Torres et al. (2015) proposed an adaptive time stepping algorithm to accurately resolve vertex rearrangements without the need of ad-hoc rearrangement thresholds and provide a numerical analysis of the algorithm. However, vertex models in that context only consider energy terms that are linear in each grain-grain (or cell-cell) interface length, whereas the energy terms in vertex models of biological cells typically depend non-linearly on cell areas and perimeters.

Importantly, previous studies such as that by Torres et al. (2015) do not analyse to what extent changes in hidden model parameters, such as parameters of numerical approximation, like the size of the time step, or non-physical parameters, such as length thresholds for cell rearrangement, can influence vertex configurations and other

summary statistics. Here, we analyse the force-propagation implementation of vertex models introduced in Chapter 2 as applied to a widely studied system in developmental biology, the larval wing disc of the fruit fly *Drosophila* (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010; Mao et al., 2011). We conduct convergence analyses of vertex positions with respect to all numerical and non-physical model parameters, and further analyse to what extent experimentally measurable summary statistics of tissue morphology, such as distributions of cell neighbour numbers and areas, depend on these parameters.

We find that vertex model predictions are sensitive to the length of cell cycle duration, the time step, and the size of the edge length threshold for cell rearrangement. Counterintuitively, the rate of cell removal is robust to changes in the area threshold for cell removal over multiple orders of magnitude. Further, analysing the active forces within the tissue reveals that vertices are subject to stronger forces during periods when cells grow and divide.

The remainder of this chapter is organised as follows. In Section 3.2, we describe our vertex model implementation of growth in the *Drosophila* larval wing imaginal disc. In Section 3.3 we present our results. Finally, we discuss our results and draw conclusions for the use of cell-based models in quantitative biology in Section 3.4.

3.2 Methods: A vertex model of the growing *Drosophila* wing disc

We consider a vertex model of the wing pouch in the growing *Drosophila* wing imaginal disc, a monolayered epithelial tissue that is one of the most widely used applications of vertex models. The *Drosophila* imaginal wing disc is an epithelial precursor to the wing of the adult fly which grows inside *Drosophila* larvae. This tissue is particularly suited to study tissue growth during morphogenesis, as it can be cultured *ex vivo* (Al-

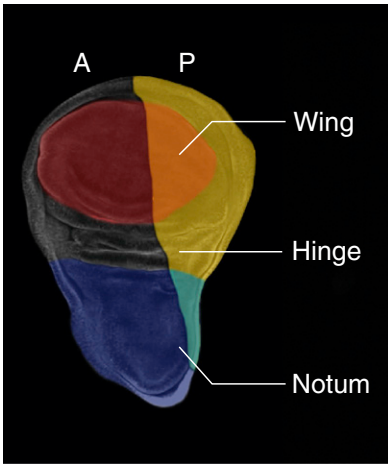

Figure 3.1: Schematic of a *Drosophila* wing imaginal disc. The *Drosophila* wing imaginal disc is located in the body of a *Drosophila* larva and is a precursor organ for the adult wing. It is subdivided into anterior (A) and posterior (P) compartments. Subsections of the precursor organ develop into separate structures in the adult fly, such as the wing itself, its hinge, and the notum, which is located at the back of the adult fly. The wing pouch (red) is widely used to study tissue growth and patterning, such as in this chapter. Image adapted from Aldaz and Escudero (2010) with permission from Elsevier.

daz et al., 2010). The wing pouch is a flat subsection of the wing disc which is devoid of any epithelial folds, making it particularly suitable to study cellular packing during tissue growth. The location of the wing pouch inside the wing imaginal disc is shown in Figure 3.1. The wing pouch initially comprises around 30 cells, and undergoes a period of intense proliferation until there are around 10000 or more cells (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010). Previous applications of vertex models of the *Drosophila* wing imaginal disc have been summarised in Chapter 2 and included, for example, studies on the emergence of characteristic cell packing during epithelial morphogenesis (Farhadifar et al., 2007), the control of cellular growth through mechanical feedback (Hufnagel et al., 2007; Aegerter-Wilmsen et al., 2010), or the formation of straight interfaces between tissues of varying fates (Landsberg et al., 2009). The details of our model implementation are described in Section 2.2. Here, we outline how we apply this model to the growing *Drosophila* wing disc.

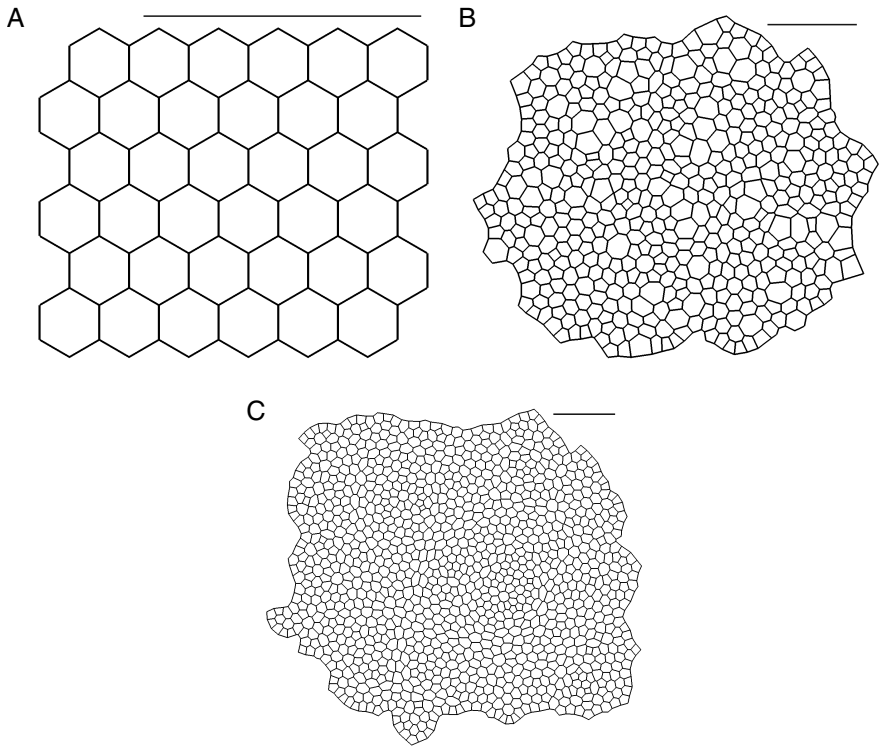

Figure 3.2: Snapshots of an example vertex model simulation used in our analysis. The growing *in silico* tissue undergoes five rounds of cell division. (A) The initial condition is a hexagonal packing of 36 cells. (B) Simulation progress after 6,750 time units at an intermediate stage of tissue growth. The tissue boundary is allowed to move freely and individual cells grow before division. (C) Snapshot of the tissue at the end of the simulation at 27,000 time units. After the fifth (last) round of cell divisions the tissue relaxes into a stable configuration. Simulated tissues in (B-C) are rescaled to fit the view; a scale bar of fixed length is added for comparison. Parameter values are listed in Table 3.1.

3.2.1 Initial and boundary conditions

Initially, the sheet is represented by a regular hexagonal lattice of six by six cells (Figure 3.2A). The boundary of the lattice is allowed to move freely throughout the simulation. Each cell has initial area and target area $A'^s = A'^s_0 = 1$, respectively.

3.2.2 Cell growth and division

Unless stated otherwise the tissue is simulated for $n_d = 5$ rounds of division, i.e. each cell divides exactly n_d times. To facilitate comparison with previous simulations of

the wing disc where vertices were propagated by minimising the energy function (2.2) (Farhadifar et al., 2007; Staple et al., 2010), we model each cell to have two cell cycle phases: quiescent and growing.

The duration of the first, quiescent, phase of the cell cycle is drawn independently from an exponential distribution with mean $2t_l'/3$, where t_l' is the total cell cycle duration. We introduce stochasticity in this phase of the cell cycle to avoid biologically unrealistic synchronous adjacent divisions; this also helps keep the simulations in a quasistatic regime since adjacent divisions are prevented from influencing each other, thus maintaining mechanical equilibrium.

The duration of the second, growing, phase of the cell cycle is fixed at length $t_l'/3$ for each cell. During this time the target area, $A_{0,\alpha}'$, of the cell grows linearly to twice its original value. Upon completion of the growth phase, the cell divides. We choose a fixed duration for the growth phase to ensure that cells grow gradually at a uniform rate throughout the tissue. Two-stage cell cycles with an exponentially distributed and a fixed length contribution have previously been observed in various cell cultures (Smith and Martin, 1973; Shields, 1978) and have been applied to model growth in the *Drosophila* wing imaginal disc (Canela-Xandri et al., 2011).

The assigning of the durations of these cell cycle stages to two and one third of the total cell cycle duration, t_l', allows us to modify the average age of a dividing cell with one single parameter. This decomposition of the cell cycle ensures that cell cycle durations are stochastic, while allowing the growth phase to occupy a significant proportion of the total cell cycle duration. The assumption that the tissue is in a quasi-steady state is commonly used in vertex models (Farhadifar et al., 2007; Marinari et al., 2012; Landsberg et al., 2009; Canela-Xandri et al., 2011) and reflects the fact that the time scales associated with mechanical rearrangements (seconds to minutes) are an order of magnitude shorter than typical cell cycle times (hours) (Farhadifar et al., 2007).

Table 3.1: Description of parameter values used in this chapter.

Parameter	Description	Value	Reference
$\overline{\Lambda}$	Cell-cell adhesion coefficient	0.12	Farhadifar et al. (2007)
$\overline{\Gamma}$	Cortical contractility coefficient	0.04	Farhadifar et al. (2007)
$\Delta t'$	Time step	0.01	Fletcher et al. (2013)
$A'_{\min}$	T2 transition area threshold	0.001	Fletcher et al. (2013)
l'_{T1}	T1 transition length threshold	0.01	Fletcher et al. (2013)
ρ	New edges after a T1 transition have the length $l'_{\mathrm{new}} = \rho l'_{\mathrm{T1}}$	1.5	Fletcher et al. (2013)
A'^s	Initial cell area	1.0	Farhadifar et al. (2007)
A'^s_0	Initial cell target area	1.0	Farhadifar et al. (2007)
N^s	Initial cell number	36	Farhadifar et al. (2007)
t'_l	Mean cell cycle duration	1,750	–
t'_{tot}	Simulation duration	27,000	–
n_d	Total number of divisions per cell	5	–

For parameter values for which no reference is given, please see Section 3.2 for details on how these values were estimated. Spatial and temporal parameters are non-dimensional (see Section 2.2 for details).

Applying this cell cycle model, we let the tissue grow for $n_d = 5$ generations until it contains $36 \times 2^5 \approx 1000$ cells, making it sufficiently large to obtain summary statistics of cell packing (Farhadifar et al., 2007). Note that the precise number of cells at the end of the simulation varies due to variations in the number of T2 transitions by which individual cells get removed from the tissue (see Chapter 2). Each cell of the last generation remains in the quiescent phase of the cell cycle until the simulation stops. We select the total simulation time to be $t'_{\mathrm{tot}} = 27,000$, unless specified otherwise. This duration is chosen such that the tissue can relax into its equilibrium configuration after the final cell division.

3.3 Results: Dependence of modelling results on implementation parameters

We now analyse how model behaviour depends on numerical and non-physical model parameters. Vertex models are typically used to predict summary statistics of cell packing and growth, such as the distribution of cell neighbour numbers and areas (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010). We analyse how these summary statistics depend on simulation parameters. Specifically, we focus on the final number of cells in the tissue, the total tissue area, the numbers of cell rearrangements (T1 transitions) and cell removals (T2 transitions), the distribution of cell neighbour numbers, and the correlation between cell neighbour number and cell area. Note that we exclude cells on the tissue boundary from statistics of cell neighbour numbers in order to avoid boundary artefacts, which can be seen in Figure 3.2C. In Figure 3.2C, cell shapes along the tissue boundary differ from those in the bulk of the tissue, and the cell neighbour number is poorly defined for cells along the tissue boundary, since it does not coincide with the number of cell edges.

3.3.1 Tissue size is sensitive to cell cycle duration

In previous vertex model applications (Farhadifar et al., 2007; Mao et al., 2011; Aegerter-Wilmsen et al., 2010), experimentally measured summary statistics of cell packing were reproduced using an energy minimisation implementation. Such energy minimisation schemes assume quasistatic evolution of the sheet, where the tissue is in mechanical equilibrium at all times. It is unclear to what extent summary statistics are preserved when the tissue evolves in a dynamic regime.

We analyse the dependence of the summary statistics on the cell cycle duration, t_l', in Figure 3.3. The units of rescaled space and time that are used throughout this thesis have been introduced in Chapter 2. The cell number and tissue area at the end

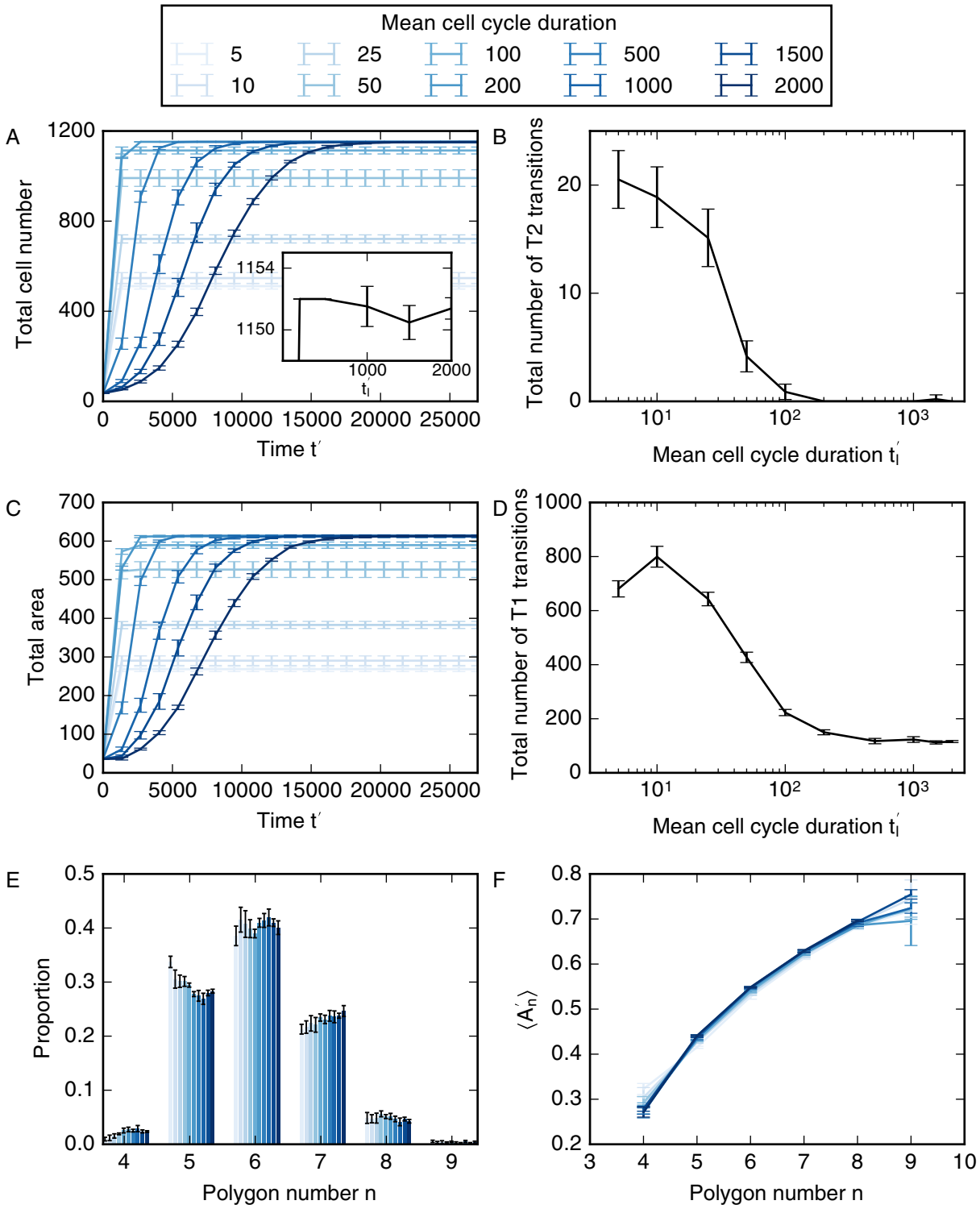

Figure 3.3: Dependence of summary statistics of tissue growth and packing on the mean cell cycle duration. Variation of cell numbers (A), number of T2 transitions (B), tissue area (C), total number of T1 transitions (D), cell neighbour number distribution (E) and mean area per polygon class (F) with mean cell cycle duration. Error bars denote standard deviations across 100 simulations. All simulation parameters are provided in Table 3.1.

of the simulation, and the total number of cell rearrangements, vary by up to a factor of two as the mean cell cycle duration increases from five to 2000 non-dimensional time units (Figure 3.3A-D). The cell number and tissue area increase with the mean cell cycle duration, whereas the amount of rearrangement (T1 transitions) decreases, reflecting a reduction in cell removal events (T2 transitions). The cell number and the tissue area do not increase further for mean non-dimensional cell cycle durations larger than 1000 time units. In this regime, the total number of rearrangements and cell removals also stop decreasing. We thus identify this regime as the quasistatic regime, where the tissue maintains mechanical equilibrium throughout the simulation. Note, however, that neither the total cell number, nor the tissue area, the number of cell rearrangements or the number of cell removals converge numerically as the mean cell cycle duration increases, due to the stochastic nature of the system.

The cell neighbour number distribution depends on the cell cycle duration in a non-linear fashion (Figure 3.3E). For example, the number of hexagons peaks at cell cycle durations of 10 as well as 1000 time units. For cell cycle durations longer than 1000 time units the numbers of pentagons and heptagons increase as the cell cycle duration increases, while the number of hexagons decreases. We interpret this non-linear dependence as resulting from changes in cell neighbour numbers due to cell division and due to cell neighbour exchanges. As the cell cycle duration exceeds $t_l' = 10$, a decrease in the number of cell removals events leads to an increase in cell division events which, in turn, drives the polygon distribution away from its hexagonal initial condition. As the number of cell divisions ceases to increase the number of cell rearrangements drops as well, and the number of hexagons reaches a second peak. Increasing the time between cell divisions further decreases the number of hexagons. Note that none of the simulated polygon histograms coincide with previously reported histograms in which pentagons outweigh hexagons (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010), despite choosing identical parameters in energy equation (2.2).

We discuss possible reasons for this difference in Section 3.4.

Another common summary statistic of cell packing is the mean area of cells of each polygon number $\langle A'_n \rangle$, where $\langle \cdot \rangle$ denotes an average across all cells in the tissue that are not on the tissue boundary, A' is the rescaled cell area, and n is the polygon number, i.e. the number of neighbours that a cell has. This summary statistic is often used to characterize epithelia (Lewis, 1928; Patel et al., 2009; Kim et al., 2014; Farhadifar et al., 2007). We find that the mean cell area for each polygon number is not sensitive to changes in cell cycle length and increases monotonically with polygon number.

We interpret the data in Figure 3.3 as follows. Differences in tissue size and cell packing arise due to a sensitive interplay between the cell cycle duration and the timescale for mechanical relaxation of the tissue, T. Growing cells push against their neighbours, leading to tissue growth. This outward movement is counteracted by the friction term in the force equation (2.1). As cells grow more quickly, i.e. with smaller cell cycle durations, the force required to push the surrounding cells outward increases. For sufficiently small cell cycle durations, the forces may become strong enough to cause cell extrusion. This finding may not be biologically relevant when studying growth in the *Drosophila* wing imaginal disc, since in this system the time scales for mechanical rearrangement are orders of magnitude smaller than the time scales associated with growth and proliferation (Farhadifar et al., 2007). However, our results suggest that, in other systems, such as the *Drosophila* embryonic epidermis, where cells divide on the time scales of minutes rather than hours, cell extrusion may be induced during periods of fast tissue growth.

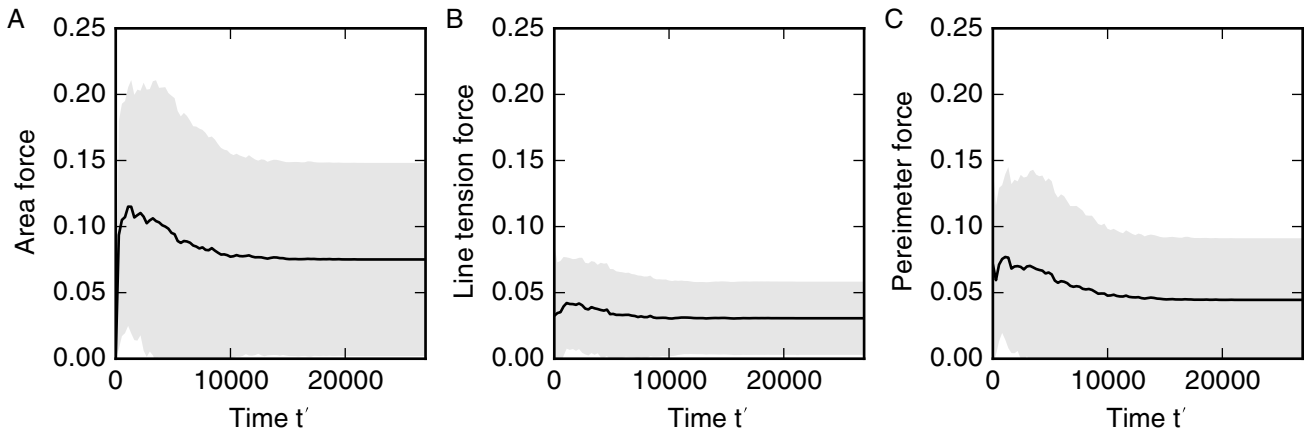

Figure 3.4: Magnitude of area (A), edge (B), and perimeter force contributions (C) over time. The solid lines represent the average of force contribution magnitudes across all vertices for one simulation. The shaded regions represent one standard deviation of the force contribution magnitudes across the tissue. A cell cycle duration of $t_l' = 2000$ is used. All other parameters are listed in Table 3.1.

3.3.2 Cell growth and division increase forces within the tissue

The energy expression (2.2) leads to three different force contributions on each vertex: an area force; an edge force; and a perimeter force. In Figure 3.4 we analyse the magnitude of these contributions for a simulation with mean cell cycle duration $t_l' = 2000$. The solid line represents the average magnitudes for the individual contributions for all forces in the tissue, and the shaded areas mark one standard deviation. The strongest force contribution is the area force (Figure 3.4A), whereas the weakest is the edge force (Figure 3.4B). This relationship is intuitive if one considers the directions of the individual force contributions when both $\overline{\Lambda}$ and $\overline{\Gamma}$ are positive: Most cells in the tissue have areas smaller than their target area of 1.0 (compare with Figure 3.3F), hence for an individual cell, the area force contribution points outwards from the cell. The edge contribution and perimeter contribution (Figure 3.4C) point inwards for individual cells, counteracting the area force. It follows that the area contribution is strongest since, in mechanical equilibrium, it counteracts the sum of the edge and

perimeter contributions. The variation of each force contribution has the same order of magnitude as their mean values, illustrating that the forces on vertices can vary strongly across the tissue. The force magnitudes change throughout the simulation, and they peak at a value that is 50% higher than the final values. For times larger than 15000 time units, the forces do not change with time in Figure 3.4. At this time cells stop dividing and the final cell number is reached, illustrating that the forces are largest when the tissue size is increasing most rapidly. This transient rise in forces emerges because cells in the interior of the simulated tissue push on their neighbours as they grow before division. These observations enable us to predict that cells undergoing active processes, such as growth and division, are subject to significantly higher forces than cells in quiescent (inactive) tissues.

3.3.3 Large time steps suppress cell rearrangement

When using an explicit Euler method to propagate the model forwards in time, such as in equation (2.5), the time step should be chosen sufficiently small to provide a stable and accurate numerical approximation of the model dynamics. To this end, we conduct a convergence analysis. To reduce simulation times, we conduct the convergence analysis on sample simulations in which each cell divides $n_d = 4$ times instead of five, and set total simulation time as $t'_{\text{tot}} = 21000$. We choose a series of decreasing time steps, $\Delta t'_k$, and define the error function

$$\epsilon_k^t = \left\| \sum_j \mathbf{x}_j^k - \sum_j \mathbf{x}_j^{k-1} \right\|, \tag{3.1}$$

where the sums run over all vertex positions, $\mathbf{x}_j^i$, at the end of the simulation with time steps $\Delta t'_k$ and $\Delta t'_{k-1}$, respectively. The error function (3.1) evaluates the differences between the sums of final vertex positions at decreasing values of the time step. To ensure that simulations with consecutive values of the time step follow identical

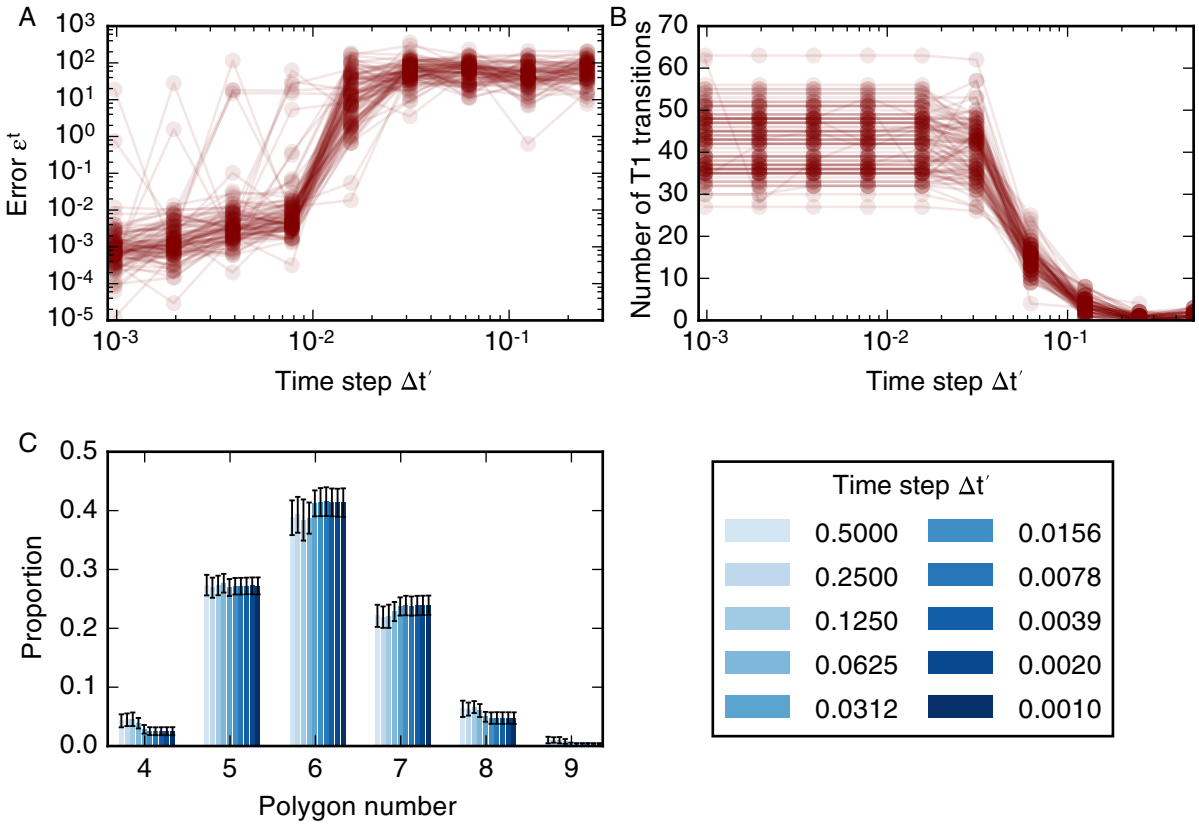

Figure 3.5: Variation in simulation results with the time step. (A) The error function for 100 different realisations of the model plotted as overlapping, opaque curves. The error function (3.1) decreases as the time step is decreased, but does not converge for all simulations. (B) The dependence of the number of the number of T1 transitions on the time step for 100 model realisations. The number of T1 transitions in the simulations is stable for time steps smaller than 0.02 and decreases with time steps greater than 0.002. (C) For time steps $\Delta t' < 0.02$ the cell neighbour number distribution is stable; the means of individual polygon class proportions vary by less than 0.01. In these simulations, cells undergo $n_d = 4$ rounds of division, and the total simulation time is $t'_{\text{tot}} = 21000$. All other parameter values are listed in Table 3.1. Error bars denote standard deviations across 100 simulations.

dynamics we generate fixed series of exponentially distributed random variates from which we calculate the cell cycle durations. We plot results of our analysis of the convergence of the vertex positions with the time step $\Delta t'$ in Figure 3.5. In general, the error function does not converge. However, for most simulations the error function (3.1) assumes values smaller than 10^{-1} for time steps smaller than 10^{-2} (Figure 3.5A). Note that this time step is five orders of magnitude smaller than the average cell cycle

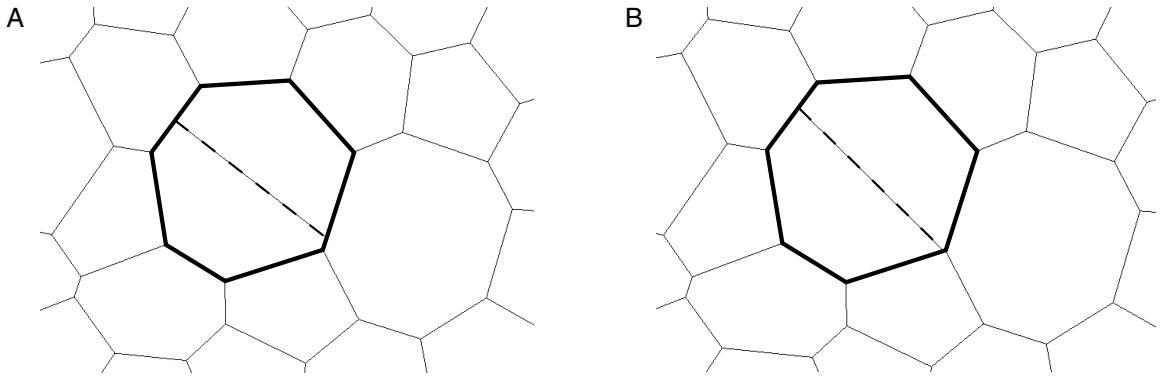

Figure 3.6: Differences in vertex configurations can arise in simulations run with different temporal resolution. A dividing cell in both simulations run with time steps $\Delta t' = 0.004$ (A) and $\Delta t' = 0.002$ (B) is shown in bold. During the cell division, a new cell-cell interface (dashed line) is created along the short axis of the dividing cell by creating new vertices (see Section 2.2 for details). The daughter cells of the dividing cell contain different vertices in the configurations corresponding to the two time steps. This leads to different vertex configurations at the end of the simulations.

duration. When the time step is larger than 10^{-2} the error function (3.1) is larger than unity since a significant number of T1 transitions are suppressed. On rare occasions, for less than five examples out of 100, the error function may be non-negligible even if the time step is smaller than 10^{-2}. These large values of the error function (3.1) reflect changes in the number of T1 transitions as the time step decreases (Figure 3.5B). When the time step is smaller than 10^{-2} summary statistics of cell packing, such as the distribution of cell neighbour numbers (Figure 3.5C) or the total number of cells, do not change as the time step is decreased further. Note, that the distribution of cell neighbour numbers in Figure 3.5C differs from those in Figure 3.3 due to the decreased number of divisions per cell, n_d. Further, we conclude from our analysis in Figure 3.5 that it is necessary to use a time step smaller than 0.01 in order to arrive at physically meaningful solutions of the vertex model, since otherwise the extent of cell rearrangement and summary statistics of cell packing will be affected by the numerical implementation of the model.

An example of how differences in the number of T1 transitions and final vertex positions can emerge when the time step is smaller than 10^{-2} is shown in Figure 3.6.

In this figure, a cell division occurs in two simulations using a time step of 0.004 (Figure 3.6A) and 0.002 (Figure 3.6B). Both simulations use the same, fixed, series of cell cycle times, and vertex positions in both simulations are similar over time up until the illustrated division. Here, and throughout, cells divide along their short axis. In this example, the short axis of the cell intersects the cell boundary close to an existing vertex. Due to differences in the vertex positions of the cell, the new vertex is created on different cell-cell interfaces as the size of the time step varies. As the simulation progresses, these different vertex configurations propagate towards different final tissue configurations, leading to differences in the total number of T1 transitions and the error function. In Figure 3.5, differences in final vertex positions are observed for all considered values of the time step. However, such differences in vertex positions do not propagate through to tissue-level summary statistics such as the distribution of cell neighbour numbers or areas.

Model convergence with time step is not improved if higher-order numerical methods are used

The results in Figures 3.5 and 3.6 were generated by propagating the vertex positions using a forward Euler time-stepping scheme. The choice of a forward Euler scheme over more accurate numerical methods is common in vertex models. For example, in a previous application where a tissue was relaxed starting from a random initial condition, it was shown that, in order to accurately resolve all T1 transitions, sufficiently small time steps had to be chosen that the benefits of higher order numerical methods were negligible (Smith, 2011). However, in Figures 3.5 and 3.6 vertex positions do not converge as the time step is decreased due to differences in T1 transitions and cell divisions for varying values of the time step, suggesting that convergence might be achieved if higher-order numerical methods were used. We test this hypothesis in Figure 3.7, where we record the error function (3.1) when propagating the vertex

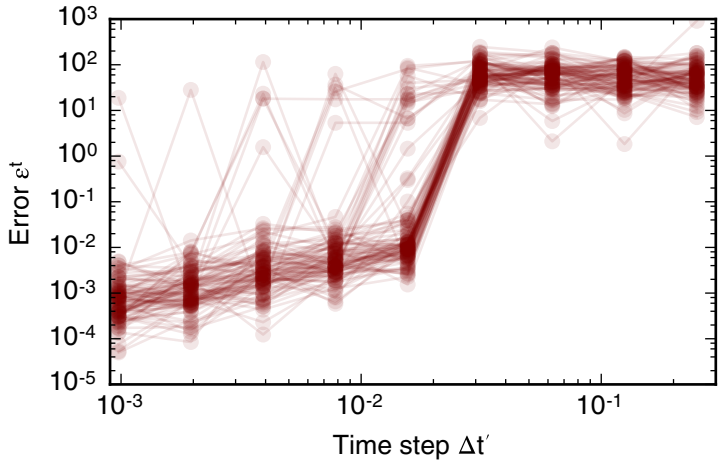

Figure 3.7: Variation in simulation results with the time step if a fourth-order Runge-Kutta scheme is used. The error function (3.1) for 100 different realisations of the model, evaluated using a fourth-order Runge-Kutta scheme, is plotted as overlapping, opaque curves. The error function decreases as the time step is decreased, but does not converge for all simulations. This result is similar for simulations run with a forward Euler scheme in figure 3.5A.

model with a fourth-order Runge-Kutta time-stepping scheme as follows. First, all vertices are accumulated into the vertex vector $\mathbf{x}'$, such that if there are N vertices at time t' then the vector $\mathbf{x}'(t')$ has $2N$ components. We propagate the vertex vector using

$$\mathbf{x}'(t'+\Delta t') = \mathbf{x}'(t') + \frac{\Delta t'}{6}\left(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4\right), \tag{3.2}$$

$$\mathbf{k}_1 = -\boldsymbol{\nabla}' E'(t', \mathbf{x}'(t')), \tag{3.3}$$

$$\mathbf{k}_2 = -\boldsymbol{\nabla}' E'(t' + \frac{\Delta t'}{2}, \mathbf{x}'(t') + \frac{\Delta t'}{2}\mathbf{k}_1), \tag{3.4}$$

$$\mathbf{k}_3 = -\boldsymbol{\nabla}' E'(t' + \frac{\Delta t'}{2}, \mathbf{x}'(t') + \frac{\Delta t'}{2}\mathbf{k}_2), \tag{3.5}$$

$$\mathbf{k}_4 = -\boldsymbol{\nabla}' E'(t' + \Delta t', \mathbf{x}'(t') + \Delta t'\mathbf{k}_3). \tag{3.6}$$

Here, $\boldsymbol{\nabla}'$ denotes the gradient with respect to the vector $\mathbf{x}$.

Similar to the error function obtained using a forward Euler numerical scheme in Figure 3.5A, the error function obtained using a fourth-order Runge-Kutta numerical

scheme in Figure 3.7 assumes values smaller than one for time steps below 0.01, but does not converge as the time step is decreased further. Comparing Figures 3.5A and 3.7 we conclude that a higher-order time-stepping scheme does not improve the accuracy of vertex model propagation, since both the forward Euler and the fourth-order Runge-Kutta scheme require time steps smaller than roughly 0.01 in order for the error function (3.1) to assume values smaller than one on average, while exhibiting a similar degree of variability across all simulations.

3.3.4 Occurrence of cell rearrangements is regulated by rearrangement threshold

We further analyse the dependence of vertex positions and summary statistics on the T1 transition threshold, l'_{T1}. Similar to the time step convergence analysis we define a series of decreasing values of $l'_{\mathrm{T1},k}$ and the error function

$$\epsilon_k^{T1} = \left\| \sum_j \mathbf{x}_j^k - \sum_j \mathbf{x}_j^{k-1} \right\|, \tag{3.7}$$

which measures the difference between the final vertex positions of simulations with decreasing values of the T1 transition threshold, $l'_{\mathrm{T1},k}$ and $l'_{\mathrm{T1},k-1}$, respectively. The variation of the error function with decreasing values of $l'_{\mathrm{T1},k}$ is shown in Figure 3.8A. For all considered values of l'_{T1} the error function does not converge and varies between values of 1 and 10^3. Only for $l'_{\mathrm{T1}} < 10^{-3}$ is the error function (3.7) smaller than unity for some simulations. However, for such small values of l'_{T1}, many simulations fail as the simulation algorithm encounters situations that it cannot resolve, for example due to overlapping or self-intersecting cells (Figure 3.8B).

A large T1 transition threshold of 0.2 length units leads to a large number of T1 transitions, whereas T1 transitions are suppressed for thresholds of 0.003 length units or smaller (Figure 3.8C). This variation in the number of cell rearrangements

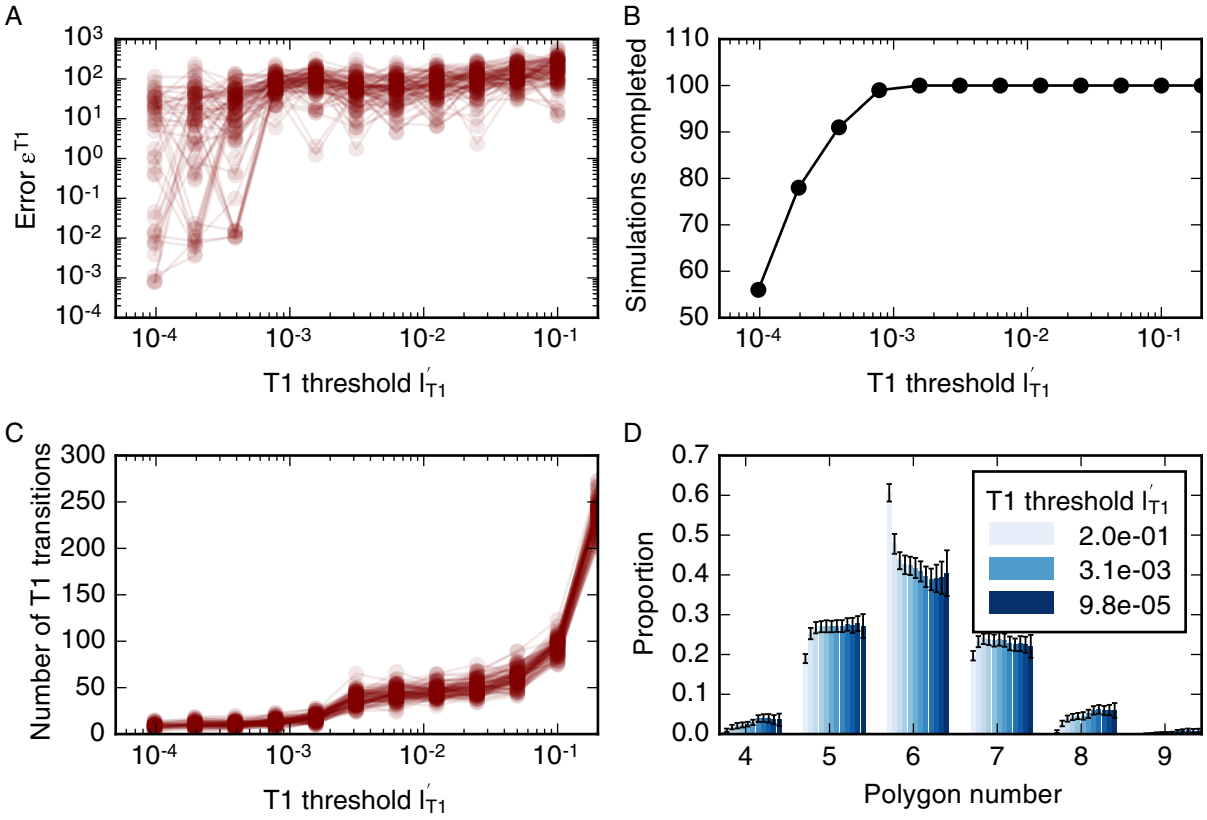

Figure 3.8: Variation of simulation result with size of the T1 transition threshold, l'_{T1}. The error function (3.7) does not converge as l'_{T1} decreases. The number of cell rearrangements is larger than 100 for a large value of the rearrangement threshold, $l'_{T1} > 0.1$, whereas cell rearrangements are suppressed for small values of the rearrangement threshold, $l'_{T1} < 0.001$, with cell rearrangement numbers less than 30. When cell rearrangement is suppressed, some simulations fail to complete (see text). Varying amounts of cell rearrangement lead to different distributions in cell neighbour numbers. Parameter values are listed in Table 3.1. Error bars denote the standard deviations across 100 simulations.

influences summary statistics of cell packing, for example leading to variations in the cell neighbour number distribution. For large rearrangement thresholds, e.g. $l'_{T1} = 0.2$, the number of cell rearrangements is high, leading to a high proportion of hexagons (around 0.6), whereas suppression of cell rearrangements for small cell rearrangement thresholds, for example $l'_{T1} = 0.2$, leads to a wider distribution of cell neighbour numbers with a proportion of hexagons below 0.4. The number of cell rearrangements is stable between T1 transition thresholds of 0.02 and 0.003. In this regime, the

proportion of hexagons varies slightly between 0.425 and 0.409 (Figure 3.8D). Despite the stable number of T1 transitions across this parameter regime between 0.02 and 0.003 the final vertex positions differ for any two values of the T1 transition threshold, as reflected in values of the error function.

As illustrated in Figure 3.8B, if the T1 transition threshold is smaller than 0.001, simulations fail to complete, as the simulation algorithm encounters situations that it cannot resolve, for example due to overlapping or self-intersecting cells. An example of how a simulation can fail due to a small value of the T1 transition threshold is provided in Figure 3.9. A snapshot is taken of the simulation at the last two time steps before simulation failure. Due to a short edge two boundary vertices in the tissue appear merged (arrow in Figure 3.9A). This short edge is magnified for the penultimate (Figure 3.9B) and last time steps (Figure 3.9C) before simulation failure. At this last time step, one of the boundary cells becomes concave. The simulation then fails since our vertex model implementation cannot resolve this configuration. When two boundary cells overlap, the simulation procedure attempts to merge the vertex with its closest cell boundary. This procedure fails because the identified boundary is internal to the tissue rather than a boundary interface.

3.3.5 Simulation results are robust to variation in length of newly formed edges

When cells exchange neighbours by way of T1 transitions, new edges are formed. Each new edge has length $l'_{\text{new}} = \rho l'_{\text{T1}}$. In order to investigate the extent to which changes in the length of newly formed edges can affect simulation results we define a series of increasing values for ρ^k and the error function

$$\epsilon_k^\rho = \left\| \sum_j \mathbf{x}_j^k - \sum_j \mathbf{x}_j^0 \right\|, \tag{3.8}$$

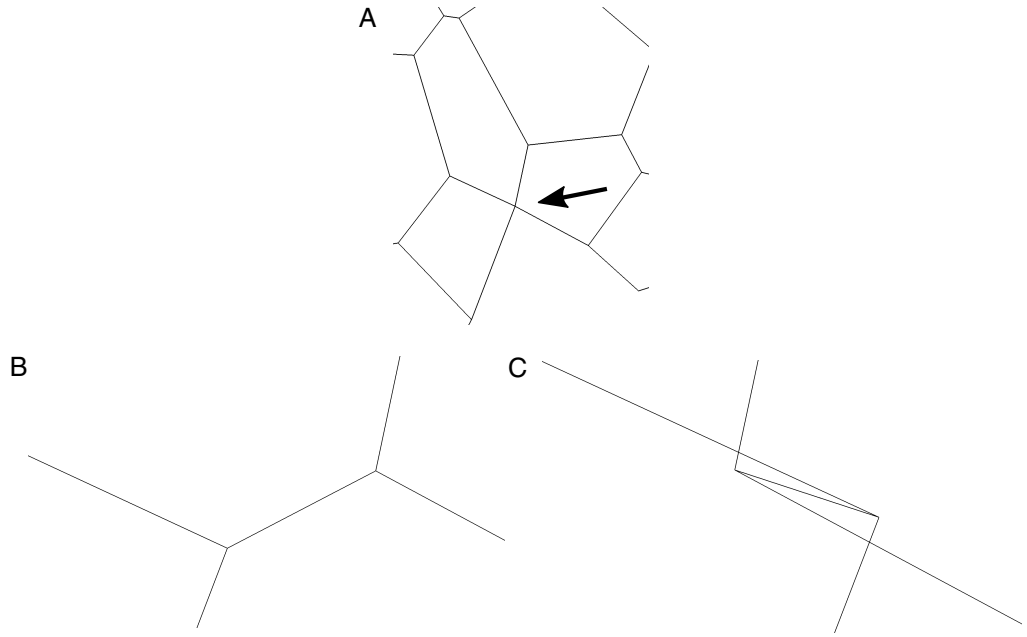

Figure 3.9: Small values of the T1 transition threshold, $l'_{T1} < 10^{-3}$, suppress rearrangement and lead to failure of the simulation algorithm. One of the failing simulations in Figure 3.8 is analysed. The tissue configuration in the last time step before simulation failure contains two vertices that appear to be merged due to a short edge on the tissue boundary. The short edge is indicated by an arrow (A) and magnified for the penultimate (B) and final completed time step (C) of the simulation. Since the short edge in the penultimate time step is prevented from rearranging, the two adjacent boundary cells intersect each other, leading to failure of the simulation.

which measures the difference in vertex positions relative to simulations with $\rho^0 = 1.05$. As shown in Figure 3.10, individual simulations may result in different final tissue configurations than the reference configuration if newly formed edges are twice as long as the rearrangement threshold or longer ($\rho > 2$, Figure 3.10). Such differences in configuration were observed for three out of 100 simulations, illustrating the robustness of simulation results to the length of newly formed edges.

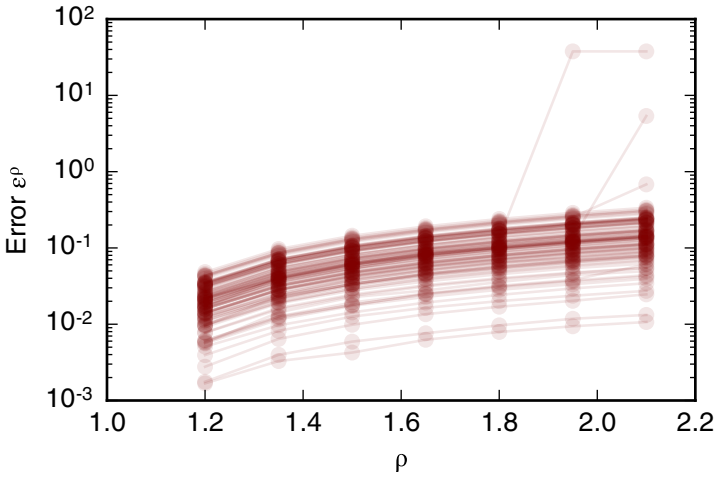

Figure 3.10: Dependence of simulation results on the length of edges created by T1 transitions, $l'_{\mathbf{new}} = \rho l'_{\mathbf{T1}}$. The error function (3.8) is recorded for 100 simulations. All simulation parameters are listed in Table 3.1. The error function is smaller than 1 for $\rho < 2.0$.

3.3.6 Rate of T2 transitions is robust to variation in the T2 transition threshold over five orders of magnitude

Next, we turn to the value of the T2 transition threshold. We define a series of decreasing values of A'^{k}_{T2} and the error function

$$\epsilon_k^{T2} = \left\| \sum_j \mathbf{x}_j^k - \sum_j \mathbf{x}_j^{k-1} \right\|, \tag{3.9}$$

which measures the difference between the final vertex positions of simulations with decreasing values of the T2 transition threshold, A'^{k}_{T2}, and A^{k-1}_{T2}. In order to analyse the value of the error function (3.9) in a simulation with a significant amount of cell rearrangement and removal we run simulations with $n_d = 8$ generations, a cell cycle duration of $t'_l = 700$, and total simulation time $t'_{\mathrm{tot}} = 19600$. All other parameter values are listed in Table 3.1.

The value of the error function, on average, is small in Figure 3.11. However, the error function does not converge for individual simulations and may be large between

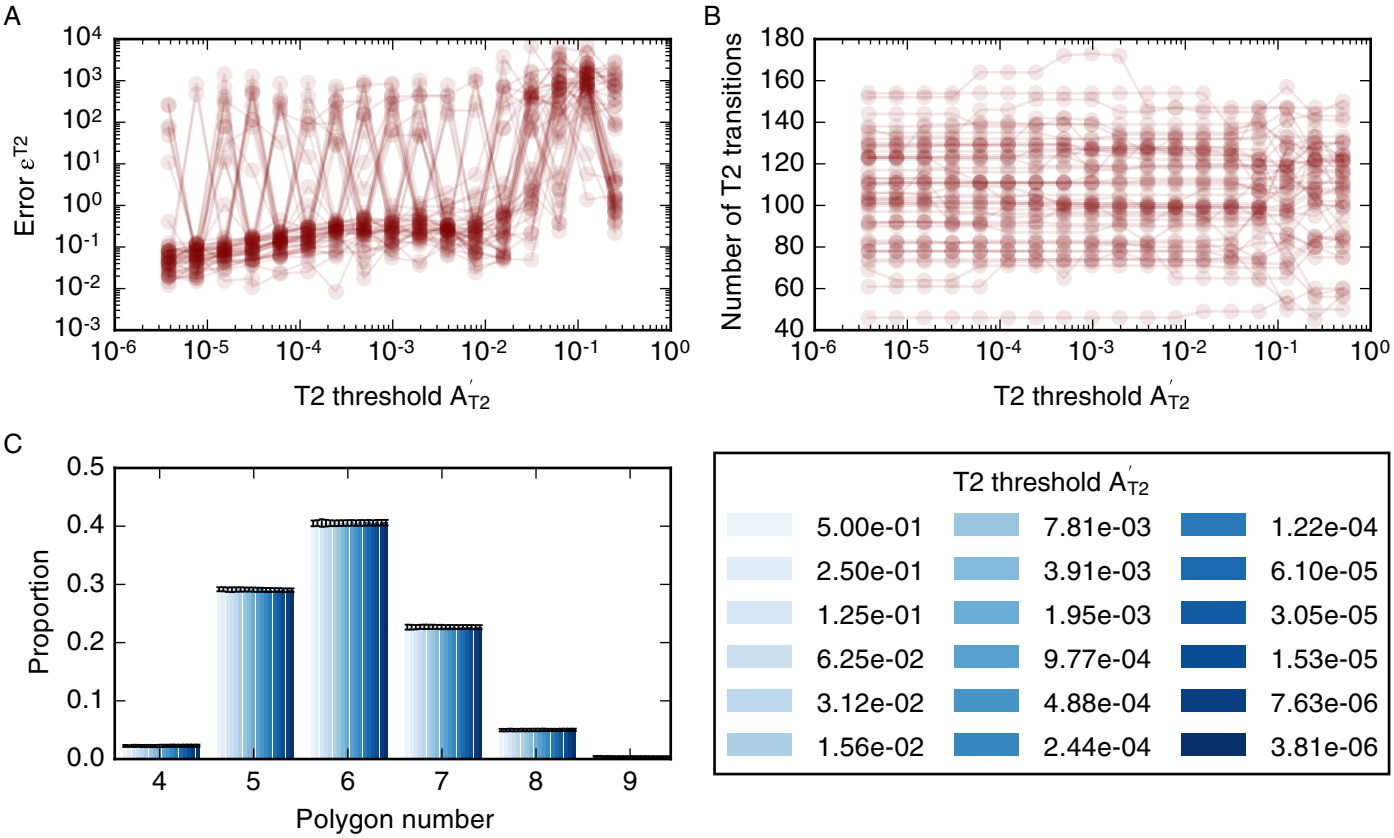

Figure 3.11: Dependence of simulation results on the T2 transition threshold, A'_{T2}. (A) The dependence of the error function (3.9) on the T2 transition threshold for 50 model realisations. The error function assumes values less than unity for $A_{T2} < 10^{-2}$ but does not converge. (B) The total number of T2 transitions for 50 model realisations is stable for all observed values of A_{T2}. (C) Tissue-level summary statistics such as the cell neighbour number distribution are not affected by changes in the threshold. Error bars denote standard deviations across 50 simulations. Simulations are run with $n_d = 8$ rounds of division, a cell cycle duration of $t'_l = 700$, and total simulation time $t'_{\text{tot}} = 19600$. All further simulation parameters are listed in Table 3.1.

consecutive values of the threshold. In particular, the error function does not converge to zero. As the threshold decreases, the overall number of T2 transitions in the simulations is stable at approximately 150 T2 transitions per simulation. However, for individual simulations the total number of T2 transitions may vary by up to 10 as the threshold A'_{T2} is decreased. The overall number of T2 transitions does not change over a large range of T2 transition thresholds that covers multiple orders of magnitude, and all simulations complete without errors even if the T2 transition

threshold is smaller than 10^{-6}, which is three orders of magnitudes smaller than the standard value for this parameter in our simulations. The independence of the number of T2 transitions of the threshold $A_{\mathrm{T2}}^{\prime k}$ is reflected in tissue-level summary statistics, such as the distributions of cell neighbour numbers, which are unaffected by changes in the T2 transition threshold.

Dependence of the simulation results on the update ordering in each time step

Finally, we investigate whether the update ordering within Algorithm 1 may affect simulation results. To this end, we randomise the order in which T1 transitions are conducted during one time step. We find that the update order does not lead to differences in final vertex positions in 100 simulations. This is intuitive, considering that the order in which individual events are conducted is most likely to be relevant in situations where events happen directly adjacent to each other, for example if two adjacent edges undergo T1 transitions at the same time step, if there are two adjacent divisions, or if a dividing cell also participates in cell rearrangement. In these examples, the order in which these events occur during one time step may have an impact on simulation outcomes. Our results imply that no adjacent two edges undergo T1 transitions in 100 sample simulations.

3.4 Discussion

Cell-based models have the potential to help unravel fundamental biophysical mechanisms underlying the growth and dynamics of biological tissues. However, the numerical implementation of such models is rarely analysed, and the dependence of model predictions on implementation details often remains unexplored. Here, we have analysed a widely applied class of cell-based models called a vertex model, and probed to

what extent experimentally relevant summary statistics can depend on implementation details, such as choice of numerical or non-physical model parameters.

For example, we found that the speed at which cells grow and divide relative to the speed of tissue relaxation can significantly alter *in silico* tissue behaviour. The total number cells in the tissue, as well as the tissue area and the number of cell rearrangements, varies by up to a factor of two as the mean cell cycle duration is changed. Summary statistics of cell packing, such as the distribution of cell neighbour numbers, or the correlation between cell neighbour number and area, are less strongly affected by the exact choice of timescale; the main features of these statistics are preserved in all cases. This finding that the total cell number and tissue are dependent on the mean cell cycle duration suggests that cell extrusion may be induced in fast-growing tissues.

The distribution of cell numbers for the case of quasistatic simulations, identified as simulations where increases in the cell cycle duration would not lead to an overall increase in tissue area or cell number, differs from previously reported results (Farhadifar et al., 2007). Specifically, we observe fewer pentagons than hexagons. This discrepancy might arise from a difference in how equation (2.2) is used to evolve the tissue. For example, our implementation of the cell cycle differs from other implementations where the cell cycle duration varies spatially in the tissue (Mao et al., 2011, 2013; Canela-Xandri et al., 2011). Further, in (Farhadifar et al., 2007), a global energy minimisation scheme is used to propagate vertex positions, whereas a more accurate force-based approach is used here. A major difference between the two approaches is the fraction of cells in the tissue that are allowed to grow and divide concurrently. In our implementation, up to one third of the cells undergo cell-growth at any given time, whereas in other implementations all cells grow and divide sequentially. Further analysis is required to understand to what extent synchronous growth and division can affect cell packing in epithelial tissues. Milan et al. (1996) report that up to 1.7%

of cells in the early wing disc are mitotic at any given time. However, mitosis and cell growth may not happen consecutively, hence the optimal choice of the duration of the growth phase in our simulations is unclear. Overall, it is unclear to what extent different choices for the cell cycle model may influence summary statistics of cell packing.

Our analysis of forces throughout simulations, presented in Figure 3.4, reveals that, on average, the area force contribution is stronger than the edge force contribution and the perimeter force contribution on a given vertex. Further, forces on cells increase during phases of proliferation and growth. Our findings may be of relevance in force-inference approaches that estimate forces using segmented microscopy images of epithelial tissues (Ishihara and Sugimura, 2012; Ishihara et al., 2013; Chiou et al., 2012). Force-inference methods often assume that the measured configuration of cells is in equilibrium and it is unclear to what extent errors are introduced if this is not the case. In our simulations, forces are up to 50% higher when simulations are run in a dynamic regime, where cells grow and divide, than in the static regime at the end of the simulation, where cells are relaxed into a static configuration.

The vertex positions, as well as simulation summary statistics, vary as the time step is changed, and differences in vertex positions decrease with the time step. Counterintuitively, large time steps can suppress cell rearrangements in vertex model simulations. This may be explained by considering that, for large time steps, vertex positions move further than the length threshold for cell rearrangements, and instances when the lengths of cell-cell interfaces fall below this threshold may not be resolved. Importantly, in order for differences in simulation results to be negligibly small, a time step has to be chosen that is five orders of magnitude smaller than the average cell cycle duration in our simulation, and six orders of magnitude smaller than the simulation time. For individual simulations, simulation outcomes may change if a smaller time step is chosen, an effect that is preserved even when a higher-order

numerical scheme, such as fourth-order Runge-Kutta, is used. The latter finding confirms that, for vertex model implementations with ad-hoc rules for cell rearrangement and division, such as in this study, the benefits of higher-order numerical schemes diminish, and it is beneficial to reduce the computational cost of the algorithm by using a simpler numerical scheme, such as forward Euler. A forward Euler scheme is more computationally efficient than a fourth-order Runge-Kutta scheme since it requires fewer floating point operations per time step. In our simulations, differences in simulation outcomes with decreasing time steps occurred at all observed choices of the time step for both numerical schemes investigated. More research is required to analyse the extent to which further decreases in the time step can lead to convergence of the simulation results. Here, we stopped investigating the effects of further decreasing the time step due to prohibitive increases in calculation times as the time step is decreased. In previous studies, vertex models have been reported to converge as the time step is decreased (Smith, 2011; Smith et al., 2012). Our analysis differs from these previous studies by considering a tissue undergoing cell division and rearrangement rather than relaxation from an initial condition.

The simulation results are sensitive to the T1 transition threshold chosen in the simulation. The size of the T1 transition threshold can be used to regulate the extent to which the simulated tissue is allowed to rearrange in order to minimise energy. Literature values for this quantity span a range from 0.1 (Mao et al., 2011; Fletcher et al., 2013) to 0.01 (Kursawe et al., 2015). Final vertex positions of individual simulations change with the value for the T1 transition threshold and do not converge as the threshold is decreased.

Our results that both the time step and the cell rearrangement threshold may influence the rate of T1 transitions illustrates that these parameters are interconnected. When the time step is chosen sufficiently large such that vertices move further than the cell rearrangement threshold between time steps, cell rearrangement is suppressed.

This means that if a small cell rearrangement threshold is chosen, a sufficiently small time step needs to be chosen. A careful choice of time steps and cell rearrangement threshold is crucial since an incorrect choice may lead to failure of the simulation algorithm. For vertex models designed to simulate polycrystalline materials an adaptive time stepping scheme has been developed that resolves the exact time at which the end points of a short edge meet, and a T1 transition is performed whenever this happens (Torres et al., 2015). For the vertex model considered here, Spencer et al. (2017) analysed the stability of fourfold vertices and showed that these are only stable if inhomogeneous line tension parameters are applied to the adjacent edges. More work is required to understand how rates of T1 transitions differ if different conditions for rearrangement are implemented, such as the shortening of an edge to a given threshold or the shrinking edge of an edge to a point. Ultimately, the optimal algorithm to simulate cell rearrangement in epithelial tissues can only be chosen through comparison with experimental results.

While simulated vertex model configurations are sensitive to the size of the time step and thresholds for cell rearrangement, they are less sensitive to the length of newly formed edges, and to thresholds for cell removal. We find that the length of newly formed edges may be up to twice as long as the threshold for T1 transitions without affecting final vertex configurations. However, this may change in other parameter regimes, for example if larger values for the cell rearrangement threshold are chosen.

The size of the area threshold for cell removal may be varied over six orders of magnitude without impacting tissue-level summary statistics, even though the exact number of T2 transitions may differ for any two values of the area threshold. In particular, it seems to be possible to choose arbitrarily small values for the T2 transition threshold without causing the algorithm to fail. There are three effects that may contribute to the stability of small elements in our simulations. First, since small cells with areas close to the threshold for cell removal are far away from their preferred

area in our simulations ($A_{0,\alpha} > 1.0$), their area force is larger than that of adjacent neighbours. This makes the cells stiff and prevents them from becoming inverted or otherwise misshapen. Second, the relationship between area and cell neighbour numbers presented in Figure 3.3 shows that small elements are most likely to be triangular. Our simulation algorithm does not permit T1 transitions if the short edge is part of a triangular cell in order to prevent triangular elements from becoming inverted and thus the algorithm from failure. Third, this relationship between cell area and cell neighbour number may also contribute to the stability of the algorithm when the area threshold is large, for example 0.2. In this case, individual cells may be smaller than the area threshold without undergoing T2 transitions if they are not triangular.

The energy equation (2.2) provides a geometrical hyphothesis for the removal of cells from epithelia, in which cells are removed from the tissue if it is energetically favourable. Mechanical effects of cell death are an area of increasing biophysical interest (Ambrosini et al., 2017), and it is the subject of future work to design vertex models that allow alternative hypotheses for cell death to be tested.

When calculating overdamped forces in equation (2.1), the friction on each vertex is independent of its surrounding cells, and the energy required for each vertex displacement simply depends on the displacement distance. The use of force equation (2.1) is hence suitable in this thesis where we focus on summary statistics of cellular packing and vertex rearrangements. However, when applying the vertex model to study cell migration, equation (2.1) may lead to artifacts, since in this case the drag force on each cell would scale with its number of vertices independently of the cell area, which would be unphysical. Such artifacts could be prevented by applying a different friction strength μ_i for each vertex i in equation (2.1), such that μ_i depends on the number of vertices and the areas of adjacent cells. Cell area dependent friction coefficients have previously been used, for example, by van Leeuwen et al. (2009).

Here, we analysed how numerical and non-physical parameters can influence exper-

imentally measurable summary statistics in cell-based models by examining a force-propagation-based implementation of vertex models. Individual results may be relevant to other implementation choices. For example, our finding that the duration of the cell cycle in our model influences simulation outcomes may mean that parameters that control the rate of energy-minimisation may influence results in other vertex model implementations (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010, 2012). In general, further work is required to understand how other choices of implementation schemes may impact computational model predictions. For example, the noise strength in a Monte-Carlo vertex propagation scheme (Etournay et al., 2015; Ray et al., 2015) or the choice of energy-minimisation algorithm may influence vertex model behaviour.

While most of our findings in this chapter are of a numerical nature, some have explicit biological relevance. Our analysis of the dependence of tissue properties and forces on the mean cell cycle duration reveals that the vertex model predicts increased forces in tissues undergoing growth and proliferation, and that fast tissue growth may induce cell extrusion. Our findings further suggest that statistics of cell packing may depend on the nature of the cell cycle or the boundary condition of the tissue. Note that findings that do not make explicit biological predictions, such as the robustness of the vertex model to changes in the area threshold for cell removal, or its sensitivity to changes in the length threshold for cell rearrangement, are nonetheless highly relevant, since these findings highlight that choices of model design and implementation have to be carefully considered when applying vertex models quantitatively.

Throughout this chapter we use non-dimensional parameters that arise when rescaling time and space by the characteristic length and time scales of the model. The use of such rescaled parameters is beneficial in this case since it allows, for example, the comparison of our model parameters to previously used values (Farhadifar et al., 2007; Canela-Xandri et al., 2011; Mao et al., 2011). Further, we identify reference

parameter values for which our simulations are physically reasonable. By providing non-dimensional values for these parameters we facilitate their reuse in other applications where the physical values of the characteristic length or time scales may be different.

Now that we have understood the dependence of the vertex model on its numerical and implementation specific parameters we apply it to a specific biological application, that of tissue size control in the *Drosophila* embryonic epidermis.

Chapter 4

Mechanical modelling of tissue size control

After introducing and analysing the vertex model in Chapters 2 and 3, we now apply it to a key problem in developmental biology, tissue size control. In this chapter, we apply the vertex model to design a theory for the emergence of previously measured experimental data, and use it to suggest future experiments that may be used to test this theory. The contents of this chapter have been published in the journal *PLoS Computational Biology* (Kursawe et al., 2015).

4.1 Background and motivation

The mechanisms underlying tissue size control during embryonic development are extremely robust. There are many cases where the rates of proliferation, growth, or death are perturbed significantly yet patterns are maintained or repaired during later stages of development. For example, even after 80% of the material in a mouse embryo is removed, accelerated growth can give rise to correctly proportioned, albeit non-viable offspring (Snow and Tam, 1979). In fruit fly embryos, overexpressing the maternal effect gene *bicoid* leads to stark overgrowth in the head region, but the excess

tissue is removed during later stages of development through apoptosis (programmed cell death), leading to viable adults (Namba et al., 1997). Tetraploid salamanders of the species *Amblystoma mexicanum* have half the number of cells as their diploid counterparts, yet are the same size (Vernon and Butsch, 1957).

The robustness of tissue size control relies on tight coordination of cellular processes including growth, proliferation, apoptosis and movement at a tissue level. However, the fundamental mechanisms underlying such coordination remain largely unknown. In particular, the mechanical implementation of tissue size control is not well understood. The regulation of cellular mechanical properties is known to play a key role during morphogenetic events, such as tissue folding, elongation and cell sorting (Guillot and Lecuit, 2013a; Heisenberg and Bellaïche, 2013). For example, upregulation of myosin II generates tension that helps to straighten compartment boundaries in the *Drosophila* wing imaginal disc (Landsberg et al., 2009), while controlled cell death provides the tension required for invagination during *Drosophila* leg development (Monier et al., 2015). It has been illustrated theoretically how mechanical feedback might facilitate uniform growth in epithelia in the face of morphogen gradients (Shraiman, 2005) and how it might contribute to size control through growth-inhibition (Hufnagel et al., 2007). It is further known that cellular crowding can trigger cellular signalling through mechanical feedback (Pan et al., 2016). Could mechanical forces also play a significant role in robustly maintaining tissue size?

4.1.1 Previous work

To explore questions of tissue maintenance, in this chapter we develop a computational model of a patterned epithelium, with application to the segments of the *Drosophila* embryonic epidermis (Figure 4.1C). These tissues define the body plan along the head-tail axis. They are first defined during stage six of embryonic development and are visible as stripes in the epidermis of the larva (Ashburner, 2011). The segments are

subdivided into anterior (A) and posterior (P) compartments, which are marked by distinct gene expression patterns. In particular, cells in the P compartment express the gene *engrailed* (Kornberg et al., 1985) (Figure 4.1D). While the initial specification and establishment of segments is relatively well studied (Hughes and Krause, 2001), the maintenance of segment identities has received much less attention. However, it is known that compartment dimensions can be robustly restored in the presence of genetic manipulations made during earlier developmental patterning events (Namba et al., 1997; Li et al., 1999; Pazdera et al., 1998; Parker, 2006). Both the conserved epidermal growth factor receptor (EGFR) and Wnt/Wingless (WG) pathways have been implicated in regulating apoptosis to achieve pattern repair for perturbations made in each of the compartments, and they are known to antagonize each other (Namba et al., 1997; Parker, 2006).

A major strength of *Drosophila* as a model organism is the availability of genetic tools that enable the ectopic expression of gene products or RNA interference (RNAi) constructs to manipulate cell growth, proliferation and signalling in a spatiotemporally controlled manner (Perrimon et al., 2010; Duffy, 2002; Elliott and Brand, 2008). For example, the bipartite GAL4-UAS system can be used to drive expression of ectopic genes in embryos through a cross of one line containing a tissue-specific enhancer driving expression of the heterologous yeast transcription factor GAL4 with a second line that activates expression of a transgene upon binding of GAL4 to the UAS promoter region. Using this approach, Parker (2006) investigated P compartment size using the GAL4 driver line as the control genotype *engrailed-GAL4*, *UAS-GFP*, in the following referred to as *wt* (wild type). Due to this genetic modification the P compartments were fluorescently labelled in *wt*. The compartments in *wt* organisms were compared to these with various perturbations (Figure 4.1B). In particular, these included crosses between the driver line and *UAS-CyclinE* (which we shall term *en>CycE*) and *UAS-dacapo* lines (further specified as *en>dap*), in which the amount

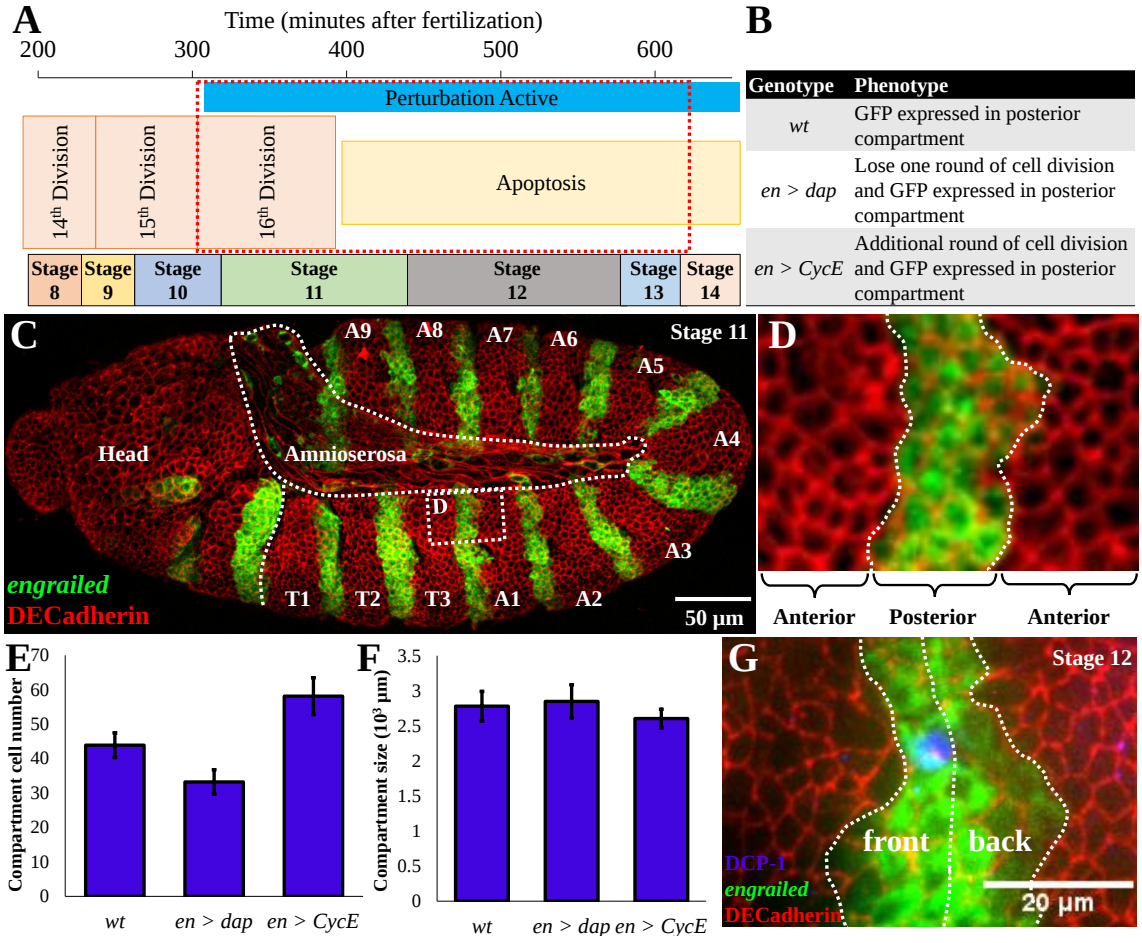

Figure 4.1: The *Drosophila* embryo as a model system for tissue size control. (A) Specification of embryonic stages over time; the red boxed region represents the time period of simulations (Foe, 1989). (B) Summary of genetic perturbations simulated in this chapter. The *wt* genotype is *engrailed>GAL4*, *UAS>GFP*, exhibiting fluorescent marking of the posterior compartments. The perturbations are crosses between the *wt* and *UAS>CyclinE* and *UAS>dacapo* lines, leading to increased or decreased amounts of proliferation, respectively. (C) Stage 11 embryo expressing GFP in the posterior compartment, stained for DE-cadherin to show cell boundaries. (D) High magnification image of simulation domain. (E, F) Data extracted from (Parker, 2006) demonstrating that compartment dimensions are robust to manipulations that change the number of cells. (G) Cell death, indicated by cleaved *Drosophila* caspase-1 antibody staining (Song, 1997), is statistically more likely to occur in the front half of the posterior compartment in *en>CycE* embryos (Parker, 2006). Figure reproduced from Kursawe et al. (2015).

of final proliferation events is perturbed towards the end of the normal range of proliferation in the epidermis (Figure 4.1A).

Parker (2006) observed an increase in final cell number of more than 30% (Figure 4.1E, right bar) in the P compartments of *en>CycE* embryos, which exhibited an additional round of cell division. However, the size of the P compartment was much less affected by this perturbation (Figure 4.1F, right bar), as measured in first instar larvae (Parker, 2006). Conversely, in *en>dap* embryos that exhibited a loss of one round of cell division, Parker (2006) observed a reduction in cell number of 25% while, again, the compartment size was relatively unchanged (Figures 4.1E and 4.1F, middle bars).

Parker's findings also suggest that epidermal growth factor receptor (EGFR) signalling, through the activating ligand Spitz, patterns apoptosis inside the P compartment. Spitz is released by a column of cells inside the anterior (A) compartment that is directly adjacent to the P compartment. Identifying cell death events through TUNEL staining (Gavrieli, 1992; Parker, 2006), led to the observation that apoptosis is much more frequent in the 'front' (more anterior) half of the P compartment, away from the Spitz source (Figure 4.1G), than the 'back' half. These numbers differed by a factor of nearly 40 in *wt* (Parker, 2006). Counter-intuitively, inhibiting apoptosis by expressing the caspase inhibitor p35 inside the P compartment of *en>CycE* embryonic segments resulted in compartment shrinkage by nearly 10%.

The above findings shed light on tissue size control in the *Drosophila* embryonic epidermal tissues, suggesting a reliance of the control mechanism on the regulation of apoptosis rather than proliferation. However, the cell-level interactions governing size control remain poorly understood. In particular, potential roles of cellular mechanics in augmenting or repairing growth defects in patterned tissues remain unexplored. To address this, we develop a vertex model of an embryonic segment to test hypotheses about the emergence of size control. Comparing the model to previously published

data across *wt* and genetic perturbations, we investigate the extent to which passive mechanical forces might suffice to explain the observed size control and asymmetries in cell death frequencies across the P compartment. Our results suggest that size control can emerge to a significant degree from the passive mechanical responses of cells. However, the observed spatial asymmetry in cell death frequencies requires patterning of mechanical properties by inter-cellular communication. These results also provide a basis for differentiating experimentally how extracellular signalling pathways like EGFR and WG might impact cellular decision making processes through predictions of observable cellular morphologies, and tissue behaviour after cell bond ablation.

4.2 Methods: A model of embryonic epidermic P compartments in *Drosophila*

We use a vertex model to simulate cell movement, intercalation, shape changes and apoptosis during the sixteenth round of divisions in a segment of the *Drosophila* embryonic epidermis. In this section we describe how the vertex model introduced in Chapters 2 and 3 is adapted to model tissue homeostasis in the *Drosophila* embryonic epidermis.

4.2.1 Geometry and boundary conditions

To simulate the subsections of the P compartment we consider a spatial domain comprising two adjacent cell populations, the cells in the P compartment and parts of the adjacent tissue in the anterior compartment on either side of it. Sample simulation images are shown in Figure 4.2A and Figure 4.2C. For simplicity, we assume that cells initially have regular hexagonal shapes.

Motivated by the repeated pattern of A and P compartments along the anterior-posterior axis of the embryo, as well as by the fact that single P compartments stretch

farther dorso-ventrally than the simulated region, doubly periodic boundary conditions are applied (Figure 4.2A). These boundary conditions keep the simulated region of interest at a fixed size. Compartment size changes are analysed as changes in the relative proportions of the anterior and posterior compartment within the simulated region. We explore the robustness of our simulation results to the choice of initial and boundary conditions later in this chapter.

To enable comparison of cell death rates in the front and back halves of the P compartment (see Figure 4.1G), a cell is defined to be in the front or back half of a compartment if its centroid is located to the anterior ('left') or posterior ('right') side of the centre of the tissue, respectively. The tissue centre is defined to be the horizontal midpoint of the sheet at time $t = 0$ and is held fixed at all times.

When computing measures of cell shape in our analysis of simulation results, we define the area and perimeter of a cell to be those of the associated polygon in the vertex model, while 'cell elongation' is defined as the square root of the ratio of the largest to the smallest eigenvalues of the moment of inertia of that polygon. This latter measure provides a robust way to measure elongations of arbitrary shapes (Fletcher et al., 2013) and is comparable to the ratio of the lengths of the long and short axis of the best fit ellipse to a cell.

4.2.2 Cell proliferation

All simulations start with $N_P^s = 24$ cells in the P compartment and $N_A^s = 40$ cells in the A compartment, to approximately match observed cell numbers (Parker, 2006) and to ensure that there are similar amounts of anterior tissue on either side of the P compartment.

In the case of a *wt* embryonic segment each cell divides once, with cell cycle times drawn independently from the uniform distribution on 0 to $t'_{wt} = 600$ time units. This corresponds to the duration of the sixteenth division cycle in the epidermis, which

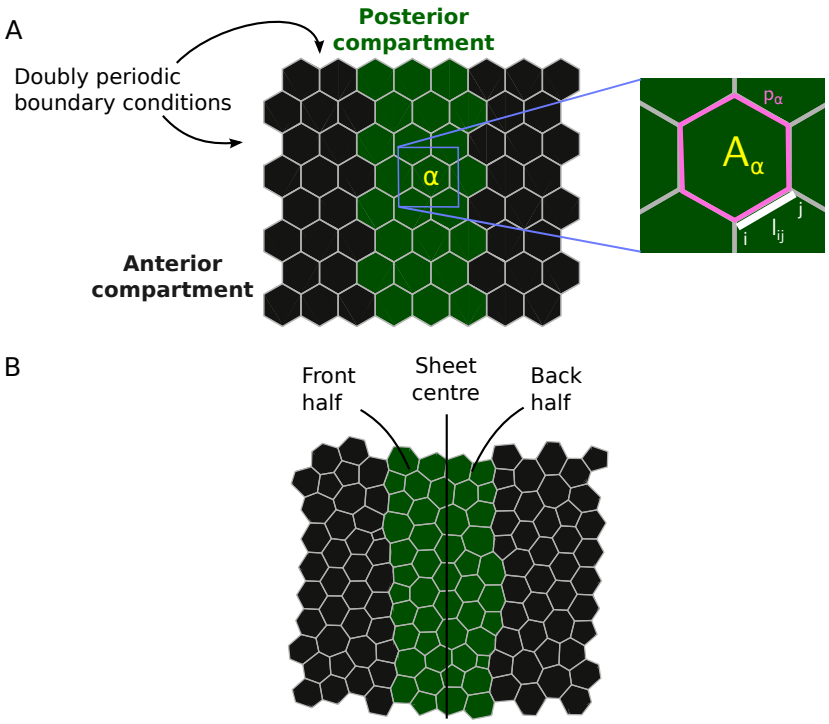

Figure 4.2: Vertex model of posterior compartment dynamics during the last division cycle in the *Drosophila* embryonic epidermis. (A) Snapshot of the initial tissue configuration for each simulation, with mechanical parameters in equation (2.4) annotated. (B) Snapshot of a *wt* simulation at the final time point, once all cell divisions have occurred, with annotation for the front and back halves of the P compartment. Parameter values are listed in Table 4.1.

occurs during late stage 10 and early stage 11 and takes roughly 50 minutes (Foe, 1989). After the round of divisions is complete, the system is allowed to relax for 200 more time units, corresponding to a total simulation time of $t'_{wt} = 800$ time units.

For an *en>CycE* embryonic segment, the first round of divisions is implemented as for *wt*, but each cell in the P compartment then has a probability $P_{CycE} = 0.54$ of dividing a second time once the first round of divisions is complete, with cell cycle times drawn independently from the uniform distribution from $t'_{CycE} = 600$ to $t'_{CycE} = 1200$ time units. This probability is inferred from published data on the *en>CycE+p53* perturbation (Parker, 2006); in this case apoptosis is blocked, allowing us to infer the average number of cell division events. The second period of 600 time

units corresponds to the duration of the ectopic divisions in the *en>CycE* embryos, which occur during late stage 11 and early stage 12 (Parker, 2006). After the second round of divisions is complete, the system is allowed to relax for 200 more time units, corresponding to a total simulation time of $t'_{CycE} = 1400$ time units.

For an *en>dap* embryonic segment, each cell in the P compartment has a fixed probability $P_{dap} = 0.6$ of not participating in the single round of divisions. This probability is inferred from published data on the *en>dap* perturbation (Parker, 2006). As with *wt*, divisions occur uniformly at random during the first $t'_{wt} = 600$ time units, after which the system is allowed to relax for 200 more time units, corresponding to a total simulation time of $t'_{dap} = 800$ time units.

In each case each daughter cell receives half the target area of the mother cell upon division unless stated otherwise. These simulation times are chosen such that the system is at quasi-steady state at each time point, similar to the simulations in Chapter 3.

4.2.3 Parameter choices

We choose the characteristic length scale $L = 11$ µm such that L^2 is the mean cell area in the P compartment at the start of the simulation period, i.e. 121 µm^2; the P compartment occupies a total area of 2.76×10^3 µm^2 (Parker, 2006) and is initialized with 24 cells. The precise value of the characteristic time scale, T, depends on tissue properties (μ and K) and could be inferred from the duration of vertex recoil after laser ablation experiments, for example. In the non-dimensionalised model, cell shapes are governed by the rescaled target area of each cell, $A'_{\alpha,0}$ and the rescaled mechanical parameters, $\overline{\Lambda}$ and $\overline{\Gamma}$. For these parameters we use previously proposed values (Farhadifar et al., 2007), unless stated otherwise. A complete list of parameters used in this chapter is given in Table 4.1.

In our implementation T1 swaps are executed whenever the length of a given edge

Table 4.1: Description of parameter values used in this chapter.

Parameter	Description	Value	Reference
$\overline{\Lambda}$	Adhesion parameter between cells in same compartment	0.12	Farhadifar et al. (2007)
$\overline{\Lambda}_r$	Reference adhesion parameter for asymmetry simulations	0.12	Farhadifar et al. (2007)
$\overline{\Lambda}_b$	Adhesion parameter between cells in different compartments	2×0.12	Landsberg et al. (2009)
$\overline{\Gamma}$	Cell perimeter contractility	0.04	Farhadifar et al. (2007)
$\overline{\Gamma}_r$	Reference perimeter contractility for asymmetry simulations	0.04	Farhadifar et al. (2007)
$\Delta t'$	Time step	0.01	Fletcher et al. (2013)
$A'_{\min}$	T2 transition threshold	0.001	Fletcher et al. (2013)
$l'_{\min}$	T1 swap threshold	0.01	Fletcher et al. (2013)
l'_{new}	Distance between new edge nodes after swaps	$1.05\ l'_{\min}$	Fletcher et al. (2013)
t_{wt}	End time of the simulation for *wt* and *en>dap*	800	–
t_{CycE}	End time of the simulation for *en>CycE*	1400	–
t'_{wt}	Time when first round of divisions finishes	600	–
$t^{s\prime}_{CycE}$	Time when second round of divisions starts in *en>CycE*	600	–
t'_{CycE}	Time when second round of divisions finishes in *en>CycE*	1200	–
A'^s	Initial cell area	1.0	Farhadifar et al. (2007)
A'^s_0	Initial cell target area	1.0	Farhadifar et al. (2007)
N^s_P	Initial cell number inside P compartment	24	Parker (2006)
N^s_A	Initial cell number inside A compartment	64	Parker (2006)
P_{dap}	Probability for P cells to not divide in *en>dap*	0.6	Parker (2006)
P_{CycE}	Probability for P cells to divide twice in *en>CycE*	0.54	Parker (2006)
R_A	Ratio between target areas of mother cells and daughter cells	0.5	–
λ_A	Area asymmetry	0.0	–
λ_l	Line tension asymmetry	0.0	–
λ_p	Perimeter asymmetry	0.0	–

For parameter values for which no reference is given, please see Section 4.2 for details on how these values were estimated.

decreases below a threshold $l_{\min} = 0.11$ µm, which is 100 times smaller than the approximate length of a cell at the beginning of the simulation. Any triangular cell is removed if its area drops below a threshold $A_{\min} = 0.121$ µm^2, which is 100 times smaller than the initial area of each cell.

In contrast to many previous vertex model applications (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010; Mao et al., 2011), we allow the mechanical parameters Λ, Γ, and A_0 to vary between cells as a consequence of underlying tissue patterning. In particular, we consider A_0 to be a function of cell generation and introduce the parameter

$$R_A = A_{0,\text{daughter}}/A_{0,\text{mother}}, \tag{4.1}$$

as the ratio of target areas of daughter cells to mother cells. To ensure that the target areas of all cells add up to the total size of the spatial domain, which is assumed to be fixed, we choose $R_A = 0.5$ unless stated otherwise. Throughout this chapter, variation of the parameter R_A is used to account for cellular growth of daughter cells as well as changes in total target area upon division. In each simulation, the initial area of each cell, A^s, equals its initial target area, A_0^s, with $A^s = A_0^s = 121$ µm^2 (see discussion below for the choice of length scales in the model).

In contrast to several previous applications (Farhadifar et al., 2007; Mao et al., 2013) of the vertex model the spatial domain in this chapter is constrained due to the fact that there is no net organism growth during this period of embryogenesis.

Unless stated otherwise, the line tension along the compartment boundaries is set to $\Lambda_b = 2\Lambda$, twice the value of the line tension in the remainder of the tissue. High tension along compartment boundaries is known to promote cell sorting and boundary straightness (Landsberg et al., 2009; Canela-Xandri et al., 2011), and the presence of myosin cables that can generate this increased tension between A and P compartments in the *Drosophila* embryonic epidermis has been reported (Monier et al., 2010).

4.2.4 Incorporating mechanical asymmetry

To investigate the consequences of asymmetries in cell mechanical properties on P compartment size control and the patterning of apoptosis, we consider three distinct cases.

In the first case, we allow for asymmetry in cell target areas in the P compartment. This is implemented by modifying the target area of each cell in the P compartment to take the form

$$A_0' = (R_A)^g(1 \pm \lambda_A), \tag{4.2}$$

where $R_A = 0.5$ as listed in Table 4.1, $g \in \{0, 1, 2\}$ denotes the generation of the cell, and the $-$ and $+$ signs apply to cells located in the front and back halves of the compartment, respectively. We refer to the parameter λ_A as the *area asymmetry*.

In the second case, we allow for asymmetry in line tensions in the P compartment. This is implemented by modifying the line tension of each cell-cell interface (edge) inside the P compartment to take the form

$$\Lambda = \Lambda_r(1 \pm \lambda_l), \tag{4.3}$$

where Λ_r is the value of the line tension when no asymmetry is imposed. The $+$ sign applies to all edges between P compartment edges whose midpoint is the front half of the compartment, while the $-$ sign applies to all edges whose midpoint is in the back half of the compartment. We refer to the parameter λ_l as the *line tension asymmetry*.

In the third case, we allow for asymmetry in perimeter contractility in the P compartment. This is implemented by modifying the perimeter contractility of each cell in the P compartment to take the form

$$\Gamma = \Gamma_r(1 \pm \lambda_p), \tag{4.4}$$

where Γ_r is the value of the perimeter contractility when no asymmetry is imposed, and the $+$ and $-$ signs apply to cells in the front and the back halves of the P compartment, respectively. We refer to the parameter λ_p as the *perimeter asymmetry*.

The asymmetry parameters are all fixed at zero in Figures 4.3-4.7, and are varied in Figures 4.8, 4.9 and 4.10.

4.2.5 Experimental procedures

All experimental data collection and processing was conducted by Cody Narciso and Pavel A. Brodskiy in the laboratory of Jeremiah J. Zartman at the University of Notre Dame, Indiana. A *Drosophila* line expressing GAL4 under the engrailed (en) promotor and CD8::GFP under the UAS promotor was used. Immunohistochemistry on embryos was performed as described in Zartman et al. (2009) with rabbit anti-dpERK (1:100, Cell Signaling), rat anti-DCAD2 (1:100, DSHB) and DAPI (5µg/ml, Invitrogen DU1306), with goat anti-rat IgG 561 (1:500, Invitrogen), and goat anti-rabbit IgG 647 (1:500, Invitrogen). Confocal z -stacks were collected using an Andor spinning disc confocal microscope with a piezo stage at 1.0µm intervals. For each embryo, six by six grids with thirty-three percent overlap were collected for each of the four channels. Image collection was performed with MetaMorph version 7.0.11. A costum vignetting correction algorithm was applied as a part of the stitching process.

4.3 Results: Model analysis and predictions

We first analyse the extent to which passive mechanical forces can lead to stable tissue size control as observed by Parker (2006). We then investigate the effect of spatial regulation of cellular mechanical properties on P compartment sizes, cell numbers, and cell death locations.

4.3.1 Compartment size control can emerge from passive mechanical forces

As an initial study, we analyse simulations where compartment size is governed solely by the passive mechanical properties of individual cells, and no further regulatory mechanism for size control is assumed. In particular, all cells in the tissue are specified to have the same mechanical properties, with the exception of interfaces shared by cells at the compartment boundary. As we shall show, such passive mechanical interactions are sufficient to explain the robustness of compartment size to hyperplastic manipulations.

Figure 4.3A shows snapshots of individual simulations of *wt*, *en>dap* and *en>CycE* embryonic segments. We observe cells that are larger but fewer in number in *en>dap* than in *wt*, while the *en>CycE* compartment contains more smaller cells. Generating statistical distributions by running 100 simulations in each case, we obtain the summary statistics visualized in Figure 4.3B-C. To allow for comparison with observed values we superimpose on each panel in Figure 4.3B-C either the upper and lower bounds in observed P compartment areas (Parker, 2006) across the three perturbations (shaded gray) or the upper and lower limits in cell numbers for each perturbation separately (blue, green, red for *wt*, *en>CycE* and *en>dap*, respectively). We do not plot the distinct shaded regions in the case of P compartment areas since the regions for individual perturbations overlap. Figure 4.3B shows that, for *wt* and *en>dap*, the average P compartment sizes and cell numbers at the end of the final round of divisions predicted by the model closely match observed values. The difference in cell number between simulation and experiment for *en>CycE* is statistically significant (17%), indicating that the model underestimates the number of cell deaths in this perturbation.

A summary of the frequency of T1 swaps occurring in the model simulations is provided in Table 4.2. There are very few cell intercalation events in our simula-

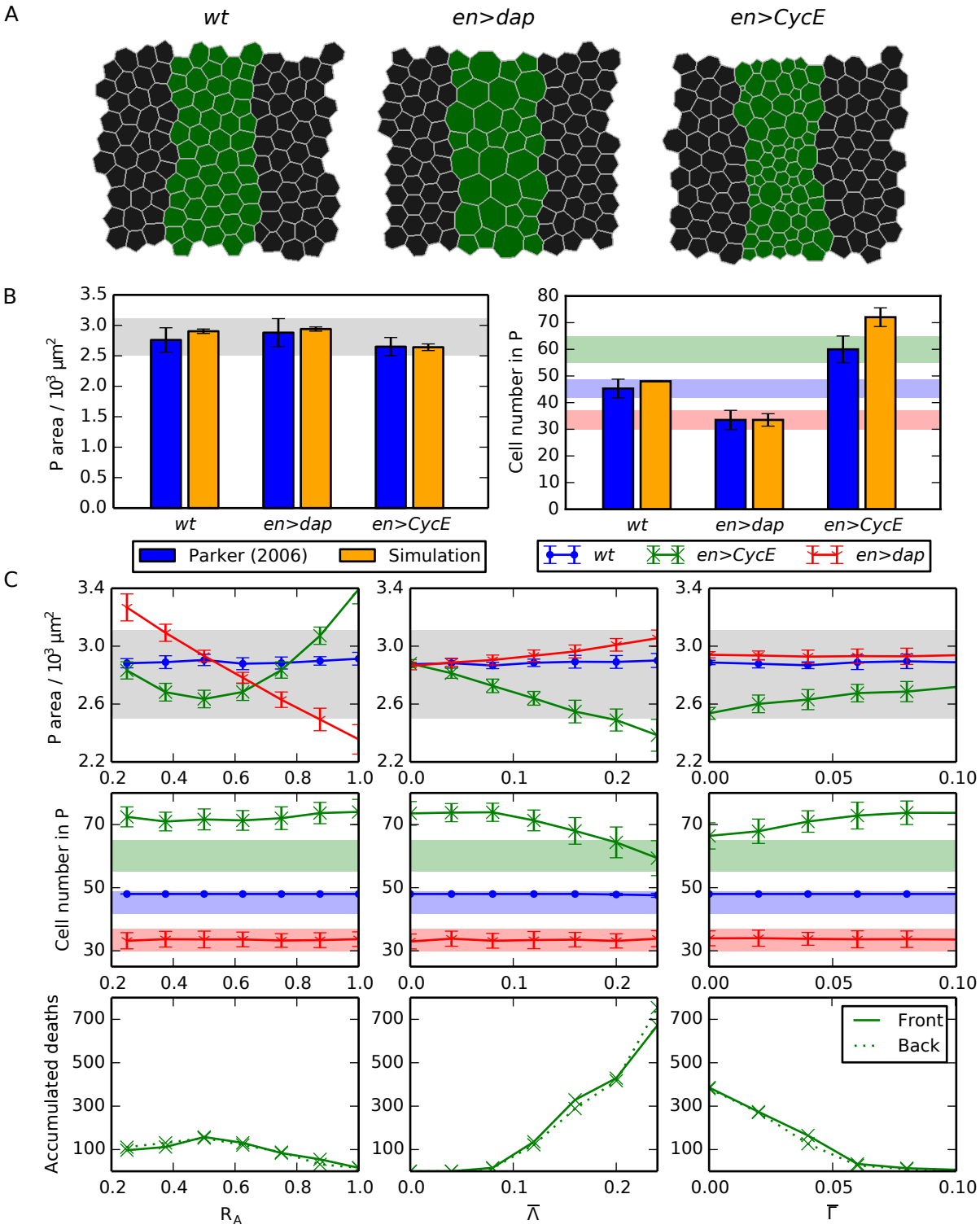

Figure 4.3: Compartment size control can emerge from passive mechanical forces. (A) Snapshots of final tissue configurations for *wt*, *en>dap* and *en>CycE* simulations, i.e. each following the final round of division. Parameter values are listed in Table 4.1. (B) Comparison of simulated P compartment areas and cell numbers with observed values (Parker, 2006). Mean values from 100 simulations are shown and error bars are standard deviations. (C) Variation in P compartment area (upper row) and cell number (middle row), and in the number of accumulated cell deaths in the *en>CycE* perturbation over 100 simulations in the front and back halves of the P compartment (lower row), as each mechanical parameter is varied individually, holding all other parameter values as listed in Table 4.1. Shaded areas in (B) and (C) mark the ranges of experimentally observed values, for reference (see Section 4.3 for details).

Table 4.2: Occurrence of T1 swaps in the simulations.

wt	*en>dap*	*en>CycE*
0	0.02 ± 0.14	3 ± 1.7

For each case of the passive mechanical model considered in Figure 4.3, we present the mean number of T1 swaps across 100 simulations. The indicated errors are standard deviations.

tions, with no T1 swaps observed for *wt*, in line with experimental observations of germ-band retraction (Schöck and Perrimon, 2002). The germ-band is a part of the *Drosophila* epidermis that retracts concurrently with the last round of cell divisions.

These simulation results were achieved using literature values of the parameters $\overline{\Lambda}$ and $\overline{\Gamma}$ (Farhadifar et al., 2007), and by assigning daughter cells to have half the target area of their mother cells ($R_A = 0.5$). Although the model is a drastic simplification of epithelial compartment size homeostasis, the *in silico* results provide a close match to experimental values without any parameter tuning. The model thus provides a simple explanation for the emergence of P compartment size control (Parker, 2006): size control can be achieved through passive mechanical forces without any further regulation of cellular properties through signalling gradients.

To explore how robust the observed size control is to the model parameters, we performed a single parameter sensitivity analysis while fixing the remaining parameters at their values listed in Table 4.1 (see Figure 4.3C). For most parameter values considered, the simulation results fall within the bounds of experimentally observed values, except for values of the target area ratio R_A smaller than 0.4 and larger than 0.9, and for values in Λ larger than 0.2.

Focusing on the results of *en>CycE* simulations, the model exhibits some counter-intuitive behaviour. In particular, uniformly low perimeter contractility, $\overline{\Lambda}$, or high line tension, $\overline{\Gamma}$, leads to mechanically induced P compartment shrinkage. In addition, an increase or decrease of R_A away from 0.5 will increase compartment sizes for the *en>CycE* perturbation. We may interpret these results as follows.

Mechanically induced P compartment shrinkage can be understood as a result of the balance between the energy terms in equation (2.2). The perimeter contractility and line tension terms act to minimise edge lengths and perimeters of cells. These force contributions can be counteracted by the area term, which acts to keep the cell close to its target area, or by stretching forces exerted by neighbouring cells. Upon division, a new edge is created, which adds an inward contractile force that any expansive forces must counteract. Therefore, daughter cells occupy a smaller area than their mother cell once they reach mechanical equilibrium. The observation that an increased rate of cell division leads to tissue shrinkage is counter-intuitive, yet not unrealistic; data from (Parker, 2006) for *en>CycE* and *en>CycE+p53* embryonic compartments show a similar trend, in which the more cells are present, the smaller the compartment area. Inhibition of cell death in the *en>CycE+p53* embryonic compartments lead to more cells, but smaller compartments. Further, this counter-intuitive experimental result, which cannot be explained by a simpler hypothesis where EGFR signalling leads to size control through direct patterning of apoptosis and growth, may be explained by a simple mechanical argument.

A similar mechanism explains the dependence of the size of the *en>CycE* compartment on the target area ratio, R_A. Mitosis-induced shrinkage is a result of the perimeter contractility and line tension terms in the mechanical model. If we choose a value for R_A that is not equal to 0.5, then the target areas of all cells will no longer add up to the total area of the tissue, and more cells have areas that are far away from their actual target areas. This increases the absolute value of the area elasticity term in the energy equation, and hence reduces the relative strength of the perimeter contractility and line tension terms. As the relative strength of these two terms decreases, the extent of mitosis-induced shrinkage is also reduced. In the case $R_A<0.5$, the additional line tension and perimeter forces due to the new edge during division are not strong enough to stretch the cells surrounding the division further away from

their target area, and if R_A>0.5 the forces originating from the new edge are not strong enough to further oppose the strength of the target area terms of the new cells. Hence, mitosis-induced shrinkage occurs only if $R_A \approx 0.5$. In our simulations, P compartment size is relatively robust to the value of R_A, despite the fact that the areas of many cells differ widely from their target values. The bulk elasticity energy term in equation (2.2) varies quadratically with deviations between cell area and cell target area. Thus, one might expect significant changes in P compartment areas or cell numbers when target areas are perturbed upon proliferation. Our simulation results suggest that P compartment areas or cell numbers are not affected by such changes in total tissue energy.

A further counter-intuitive result shown in Figure 4.3C is that increasing the line tension parameter $\overline{\Lambda}$ and increasing the perimeter contractility parameter $\overline{\Gamma}$ have opposing effects on P compartment size in the *en>CycE* perturbation. Increasing line tension results in a stronger contractile force on the cell, resulting in more T2 transitions and hence a smaller P compartment (Figure 4.3C, central panel). In contrast, although increasing perimeter contractility also results in a stronger contractile force for each cell, in this case the mechanical interactions between adjacent cells (a contracting cell acts to stretch its neighbours) result in fewer T2 transitions and hence a larger P compartment.

All the observed changes in P compartment sizes and cell numbers remain within experimentally measured values (Figure 4.3, shaded regions), the exception being the P compartment cell numbers for the *en>CycE* perturbation. The discrepancy between observed values and *in silico* results for the P compartment cell numbers in *en>CycE* is insensitive to parameter variation. The robustness of the simulation results in Figure 4.3B to parameter values provides further confirmation that size control is a natural outcome of passive mechanical cellular interactions in our model. Size control is preserved in the face of small amounts of cell growth or shrinkage (variations in

R_A) or perturbations of cellular mechanical properties (variations in $\overline{\Lambda}$ and $\overline{\Gamma}$).

However, this model fails to capture the observed asymmetry in cell death locations, as measured by the ratio of accumulated cell death occurrence between the front and the back half of the P compartment across multiple embryos. The third row of Figure 4.3C shows that the total number of cell deaths across 100 simulations is the same between the front half and the back half of the P compartment. Here we only plot the cell death occurrences of the *en>CycE* simulations, since no cell deaths were observed in any *wt* or *en>dap* simulations. This is in close agreement with experimental results (Parker, 2006), where only 0.7 (*wt*) or 0.2 (*en>dap*) cell deaths were identified by TUNEL staining per embryo.

We draw two main conclusions from the simulations presented in Figure 4.3: (i) mechanical interactions between identical cells can explain robust size control of all considered genetic perturbations (*wt*, *en>CycE*, *en>dap*), even if the model parameters are varied significantly; (ii) passive mechanical interactions of cells with uniform mechanical properties cannot explain the observed asymmetry in cell death occurrence, nor completely recapitulate the changes in cell numbers for the *en>CycE* perturbation.

4.3.2 Robustness of results to initial and boundary conditions

We analyse the influence of the choice of boundary condition on the compartment sizes and compartment cell numbers in the simulations in Figure 4.4. We compare the results of simulations using the doubly periodic boundary conditions described in Section 4.2 to those where fixed boundary conditions are imposed on the tissue as follows. Vertices on the boundary of the tissue experience the same forces as non-boundary vertices. At each time step, after applying the forces to all vertices and updating their positions, we map the boundary vertices perpendicularly back onto

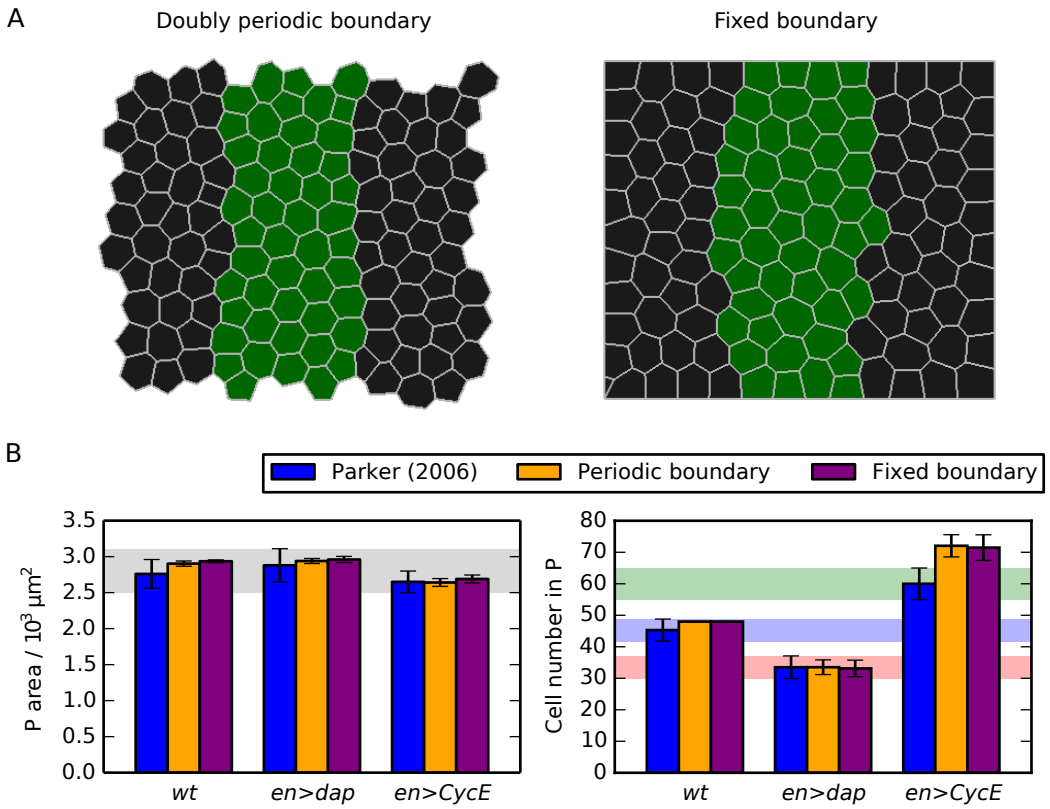

Figure 4.4: Choice of boundary condition does not affect P compartment sizes and cell numbers. (A) Snapshots of a *wt* simulation at the final time point, once all cell divisions have occurred, where either doubly periodic (left) or fixed (right) boundary conditions are imposed. Parameter values are listed in Table 4.1. (B) Comparison of P compartment areas and cell numbers for *wt*, *en>dap* and *en>CycE* simulations where doubly periodic or fixed boundary conditions are imposed. Mean values from 100 simulations are shown and error bars are standard deviations. Shaded areas mark the ranges of experimentally observed values and are added for reference and comparison with Figure 4.3.

the boundary. If a cell on the boundary undergoes a T2 transition, then the newly created vertex is also mapped onto the boundary. The dimensions of this fixed boundary are chosen such that the total area of the tissue equals the sum of the target areas of all cells of the tissue. The tissue occupies a total of 64 rescaled area units, or $7.744 \times 10^3 \mu m^2$. The length of each fixed boundary is 8L, i.e. 88µm. As Figure 4.4 shows, we find that imposing this alternative boundary condition gives similar predicted P compartment areas and cell numbers to those obtained using doubly periodic boundary conditions. This is true across *wt*, *en>dap* and *en>CycE* simulations.

The precise choice of boundary condition imposed in the model simulations does not significantly affect predicted compartment sizes and cell numbers.

We next analyse the influence of the choice of initial tissue geometry on the final P compartment sizes and cell numbers in our model. In particular, we compare the results of simulations initiated with randomly generated cell shapes to those of simulations with regular hexagonal cell shapes. The random initial conditions are created as follows. First, 64 seeds are randomly placed in a spatial domain occupying 64 rescaled area units ($7.744 \times 10^3 \mu m^2$). Second, the random seeds are mirrored along each tissue boundary to ensure periodicity, and the Voronoi tessellation of all seeds is computed. Third, we relax the Voronoi tessellation using five steps of Lloyd's relaxation algorithm (Lloyd, 1982). During a Lloyd's relaxation the centroid of each cell is calculated and a new Voronoi tessellation is generated using these centroids as seeds. This relaxation homogenises the cellular packing. Sánchez-Gutiérrez et al. (2016) have shown that the cellular packing achieved after five Lloyd's relaxation steps is similar to those found in developing epithelia, such as the *Drosophila* wing disc.

After generating the initial cell packing through Voronoi tessellations in this way, cells are assigned to the P compartment if their centroid falls within a central vertical band of width 2.9 rescaled length units (32µm). Figure 4.5 shows that these irregular initial conditions result in similar mean P compartment sizes and cell numbers to those obtained using our regular hexagonal initial conditions, while increasing the standard deviation in each quantity. Thus, irregularity in the initial geometric configuration does not significantly affect simulation outcomes.

For both the boundary conditions considered in this chapter, the initial area of each cell A^s takes its target value A_0^s . To test whether our results depend strongly on this assumption, we ran control simulations in which the initial area of each cell was half its target value at the beginning of the simulations. We found that resulting P compartment sizes and cell numbers were not strongly affected by this modification,

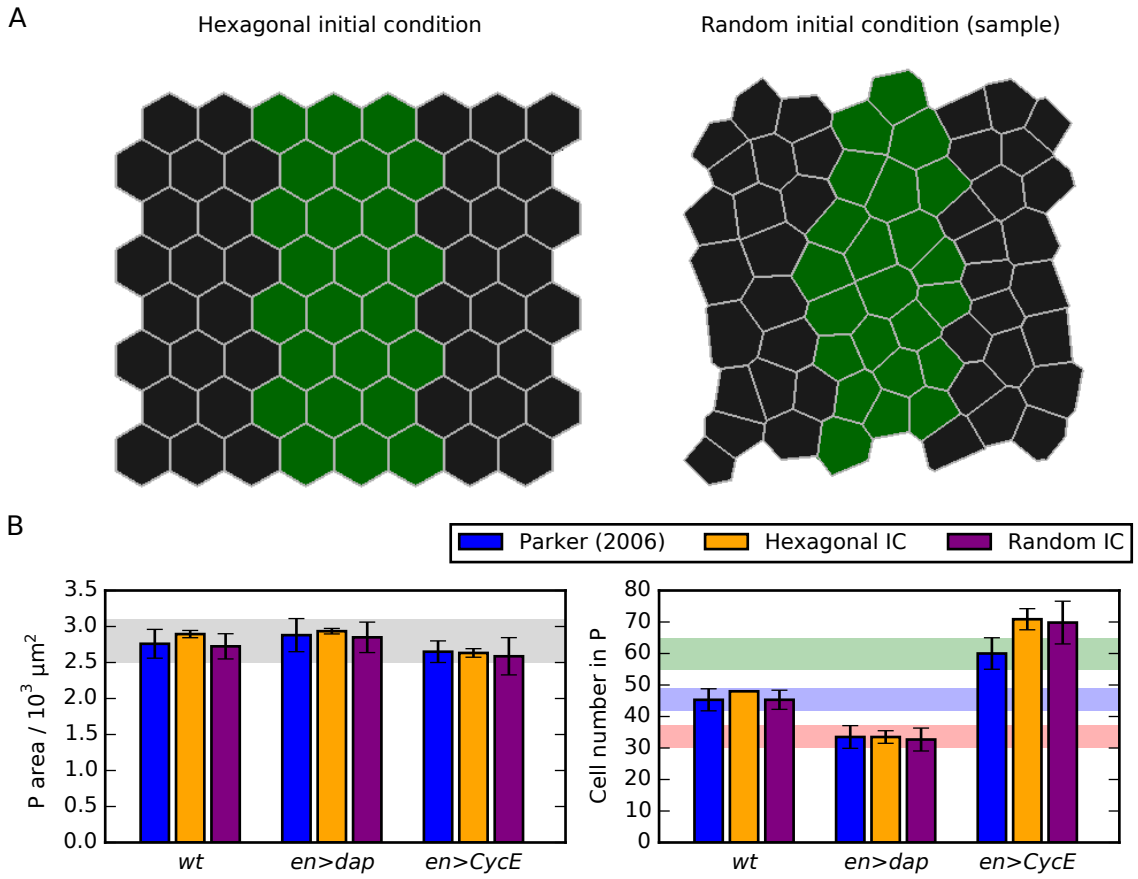

Figure 4.5: Initial cell shapes do not significantly affect P compartment sizes and cell numbers. (A) Snapshots of a hexagonal (left) initial condition, and a sample random (right) initial condition, as described in Section 4.3.2. The cells assigned to the posterior compartment occupy a similar area in both images. (B) Comparison of P compartment areas and cell numbers for *wt*, *en>dap* and *en>CycE* simulations where either a hexagonal or random initial condition (IC) was used. Mean values from 100 simulations are shown and error bars are standard deviations. Shaded areas mark the ranges of experimentally observed values and are added for reference and comparison with Figure 4.3. Parameter values are listed in Table 4.1.

and compartment size control was preserved. The results of this analysis are shown in Figure 4.6, and they differ from Figure 4.3B in two ways. First, the tissue areas at the end of simulations vary slightly from the experimental observations made by Parker; in particular, the P compartment for the *en>cycE* perturbation is larger than for the *en>dap* perturbation. This is to be expected, since in this set of simulations most cell areas are far from their target values, hence the area energy term dominates and mitosis-induced shrinkage does not occur. Second, *en>CycE* P compartment cell

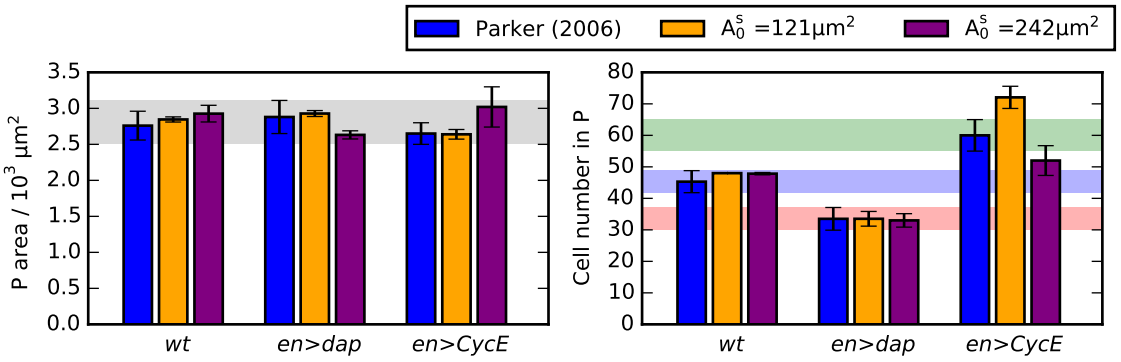

Figure 4.6: Influence of initial cell target area on P compartment size and cell numbers. (B) Comparison of P compartment areas and cell numbers for *wt*, *en>dap* and *en>CycE* simulations where either initial target areas $A_0^s = 121\mu m^2$ or $A_0^s = 242\mu m^2$ were used. Mean values from 100 simulations are shown and error bars are standard deviations. Shaded areas mark the ranges of experimentally observed values and are added for reference and comparison with Figure 4.3. Parameter values are listed in Table 4.1.

numbers are smaller than observed by Parker and in the results of Figure 4.3B (52 vs 60 and 72 cells, respectively), illustrating that more cells die in the simulations where cell target areas have been doubled. In the simulations summarized in Figure 4.6, the overall tissue target area is significantly larger than the tissue area, and therefore more cells are removed by apoptosis (T2 swaps) to decrease the corresponding energy term. Although we consider the initial condition for the simulations with doubled initial target areas A_0^s unrealistic, with cell areas being far from their target values, the model continues to exhibit P compartment size control.

Finally, Figure 4.7 shows that while the increase in line tension along compartment boundaries does affect the straightness of the boundary between A and P compartments in the model simulations, it does not significantly affect compartment sizes or cell numbers.

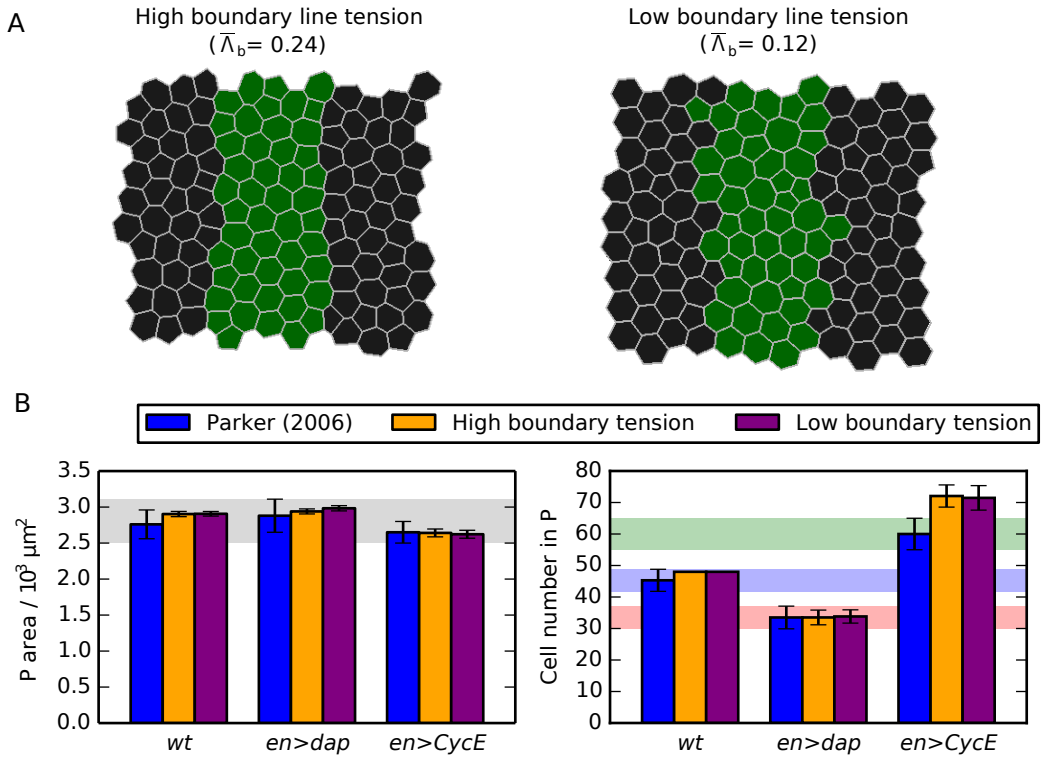

Figure 4.7: Compartment boundary line tension does not affect P compartment sizes and cell numbers. (A) Snapshots of a *wt* simulation at the final time point, once all cell divisions have occurred, where either a high (left) or low (right) line tension, $\overline{\Lambda}$, is imposed at the boundary between A and P compartments. Parameter values are listed in Table 4.1. Compartment boundary line tension promotes cell sorting and straightness of the boundary, but does not affect compartment sizes. (B) Comparison of P compartment areas and cell numbers for *wt*, *en>dap* and *en>CycE* simulations where a high (left) or low (right) compartment boundary line tension is imposed. Values for $\overline{\Lambda}$ at the compartment boundary are those given in (A). Mean values from 100 simulations are shown and error bars are standard deviations. Shaded areas mark the ranges of experimentally observed values and are added for reference and comparison with Figure 4.3.

4.3.3 Spatial patterning of cell death emerges from differential growth or differential mechanical regulation

We next use the model to analyse how asymmetries in cellular mechanical properties across the P compartment may lead to the observed spatial patterning of apoptosis. We consider three cases (Figure 4.8A): (i) 'area regulation', which refers to patterning of the cell target areas, $A_{0,\alpha}$, through the parameter λ_A; (ii) 'line tension regulation',

which refers to patterning of the line tension, $\overline{\Lambda}$, through the parameter λ_l; and (iii) 'perimeter regulation', which refers to the patterning of the perimeter contractility, $\overline{\Gamma}$, through the parameter λ_p. These parameters are defined in Section 4.2. The 'area regulation' scenario can be interpreted as a patterned growth scenario, whereas the 'line tension regulation' and 'perimeter regulation' scenarios correspond to the patterning of cellular mechanical properties. The biochemical process leading to such patterning could, for example, be Spitz-mediated EGFR-activation; this pathway has previously been identified to affect cell properties in the P compartment by Parker (2006).

Figures 4.8B-D shows the impact of small amounts of asymmetry on P compartment dynamics. In each of the cases (i)-(iii), we set the relevant asymmetry parameter to 0.2, while keeping the other two asymmetry parameters fixed at zero. Figure 4.8B shows snapshots of simulation outcomes for each case. A visual inspection suggests that these three cases give rise to P compartments with similar cell sizes and shapes as in Figure 4.3A.

Figure 4.8C shows that P compartment sizes and cell numbers are not affected by these small amounts of asymmetry in the tissue. In each case, the *in silico* compartment area and cell number is as close to the observed values (Parker, 2006) as the passive mechanical model. Although cellular properties are now patterned, compartment size control still emerges within the model. Figure 4.8D shows the total number of cell deaths in the front and the back halves of the P compartment across 100 simulations in each asymmetry case. We find that each case can explain the observed spatial asymmetry in cell death locations.

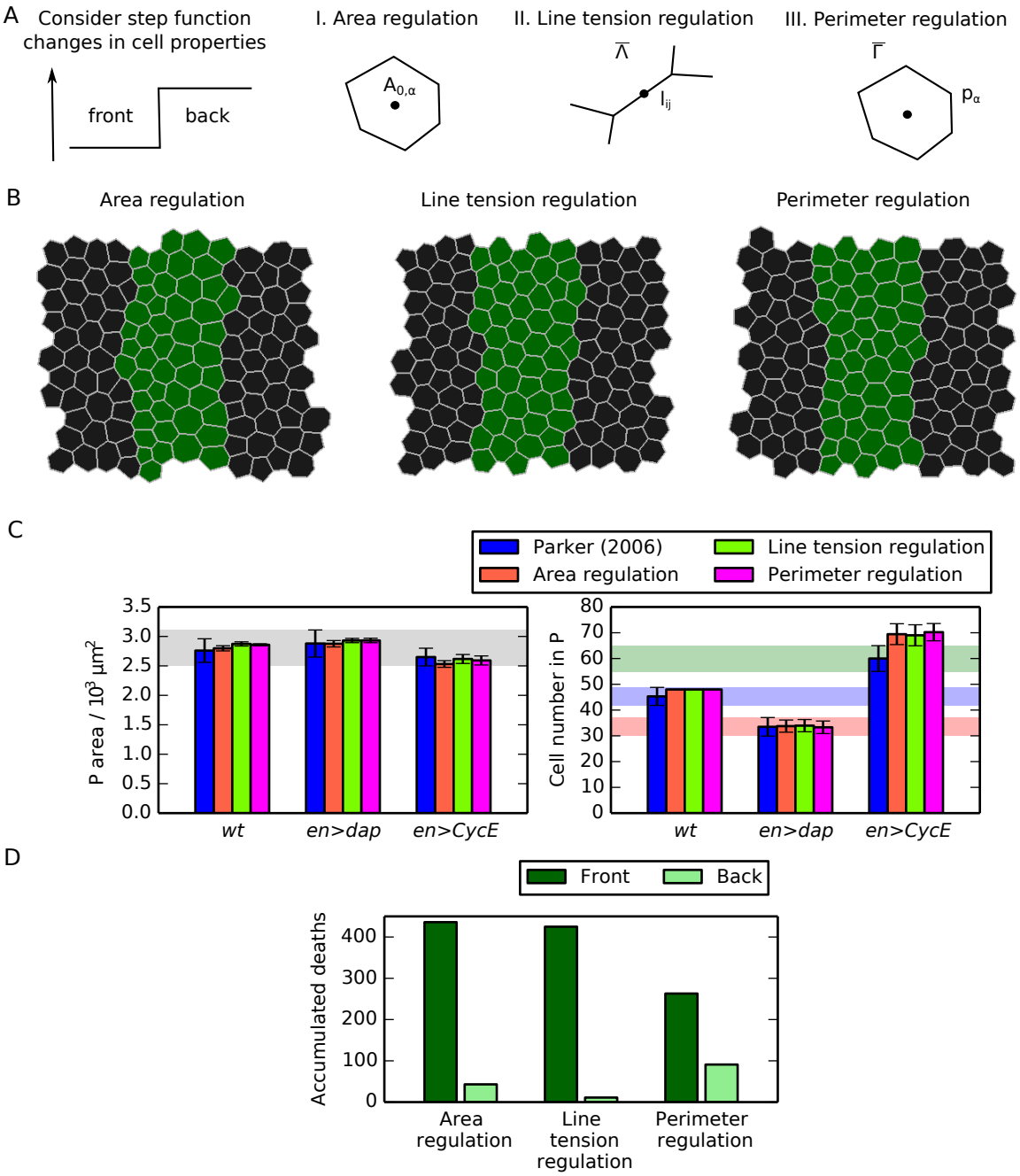

Figure 4.8: Spatial regulation of mechanical cell properties can induce asymmetries in cell death occurrences inside posterior compartments. (A) Schematic of the distinct forms of mechanical asymmetries considered in this work. (B) Snapshot of final configuration of simulations for each considered perturbation. (C) Comparison of P compartment areas and cell numbers for each of the considered perturbations with experimental values. Mean values from 100 simulations are shown and error bars are standard deviations. Parameter values are given in Table 4.1 and in Section 4.3. Shaded areas mark the ranges of experimentally observed values and are added for reference and comparison with Figure 4.3. (D) Comparison of accumulated numbers of cell deaths over 100 simulations in the front and back halves of the P compartment for each of the considered perturbations.

4.3.4 Robustness of compartment size and compartment cell number to cellular asymmetry

To assess to what extent P compartment sizes and cell numbers are robust to spatial asymmetry in cell mechanical properties, we next vary each of the three asymmetry parameters in turn while keeping the others fixed at zero. Figure 4.9 shows that increases in asymmetry lead to decreases in P compartment sizes and cell numbers (top and middle row) and the degree of asymmetry in cell death across the front and back halves of the compartment (bottom row).

In the model, P compartment sizes and cell numbers are most sensitive to asymmetry in cell target areas; for example, a value of $\lambda_A > 0.9$ can result in loss of the entire P compartment. In contrast, P compartment sizes and cell numbers remain within experimentally measured regimes for values of λ_p or λ_l from 0 up to 0.4.

4.3.5 Differential growth and mechanical regulation generate distinct distributions of cell shapes in *wt*

To identify experimentally observable signatures to differentiate between modes of regulating compartment homeostasis, we examined the distributions of four measures of cellular morphology for the scenarios described in Figure 4.8. We extract the distributions of cell areas, cell perimeters, lengths of edges between cells, and cell elongations within the P compartment at the end of each simulation. We observe distributions of these four measurements in the posterior compartment as a whole, and in the front and the back halves of the compartment separately. The results of this investigation are summarized in Figure 4.10.

The top two rows of Figure 4.10 show that the distributions of cell areas and cell perimeters (row 1 and 2) for the area regulation scenario are distinct from the corresponding distributions for the line tension and perimeter regulation scenarios.

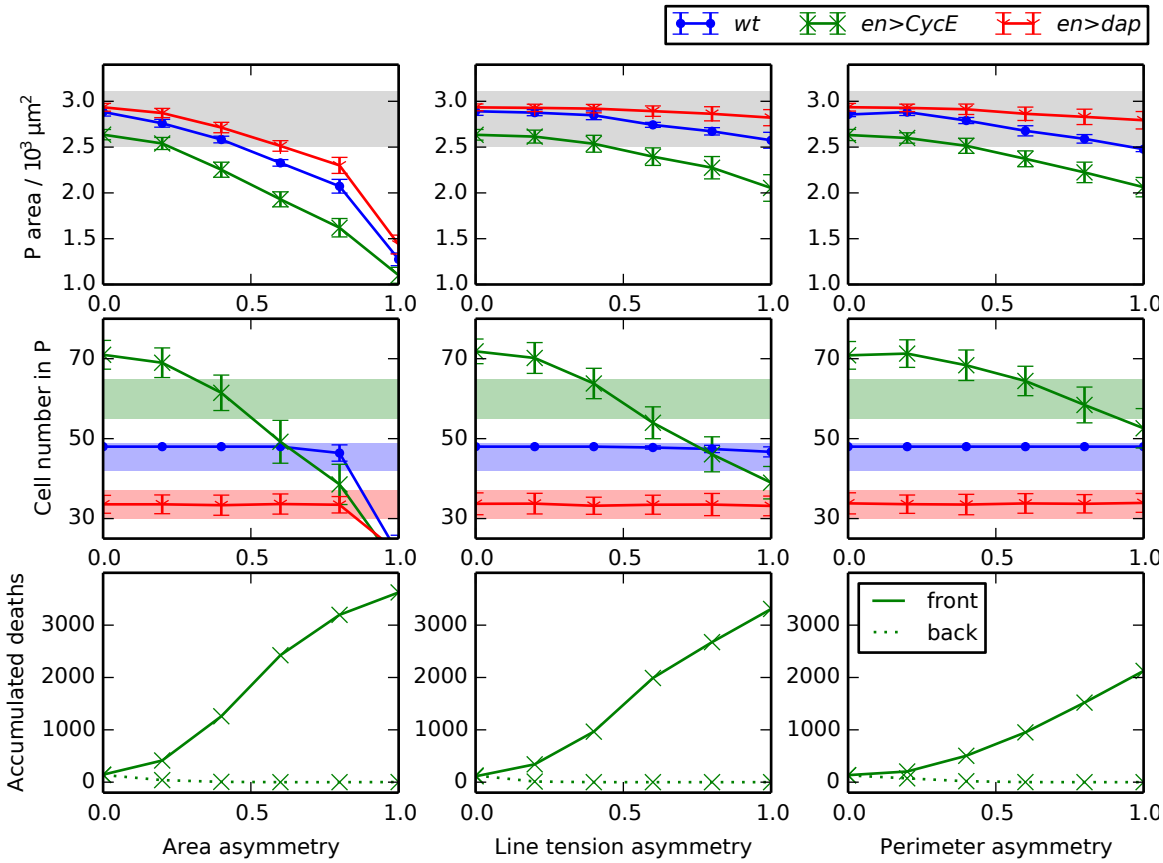

Figure 4.9: Sensitivity of P compartment size and cell number to asymmetry. Variation of P compartment area (upper row) and cell number (middle row), and of the number of accumulated cell deaths over 100 simulations in the front and back halves of the P compartment (bottom row), as the asymmetry parameters λ_A, λ_l, and λ_p are varied individually while holding all other parameters at their values listed in Table 4.1. Shaded areas are added for comparison with Figure 4.3.

In particular, the distribution of all areas is bimodal for the area regulation scenario, whereas it is not bimodal for the line tension and perimeter regulation scenarios. A similar distinction can be made for the perimeter distributions, which is bimodal for the 'area regulation' scenario and not bimodal for the 'line tension regulation' and 'perimeter regulation' scenarios. These bimodal distributions are marked by nearly non-overlapping distributions of cell areas and cell perimeters in the front and the back halves of the compartment for the area regulation scenario, whereas these distributions are overlapping in the line tension and perimeter regulation scenarios. Upon decomposing cell area distributions into contributions from the front and back

halves of the P compartment, we see that the mean cell area is different between these two halves in the area and perimeter regulation scenarios, and the same holds for the cell perimeter distributions. Cell elongations and edge lengths have similar shapes and mean values for all three asymmetry scenarios (rows 3 and 4 of Figure 4.10).

The results in Figure 4.10 suggest that it is possible to distinguish between the 'area regulation' scenario (differential growth across the compartment) from the 'line tension regulation' scenario (regulation of apical mechanical properties) by measuring the distributions of cell areas or perimeters in the front and the back halves of the posterior compartment separately. The distribution of cell areas or perimeters across the P compartment may further allow one to distinguish the 'area regulation' scenario from the 'perimeter regulation' scenario, since this distribution is bimodal in the former scenario, but not clearly bi- or unimodal in the latter. However, multiple sources of noise in an experimental setup may make this distinction between the 'area regulation' and 'perimeter regulation' scenarios less clear. Measuring edge lengths or cell elongations will not reveal differences between the scenarios.

4.3.6 Characteristics of cell area distributions for the *en>dap* and *en>CycE* perturbations are preserved across asymmetry scenarios

While cell area distributions in *wt* simulations may allow the different asymmetry scenarios considered to be distinguished from one another, these distributions in the *en>dap* and *en>CycE* cases provide model predictions that are preserved across all scenarios. In each case, we find that the cell area distribution is multimodal. In particular, the *en>dap* cell area distribution is trimodal in the 'area regulation' scenario, whereas it is bimodal in the other cases considered.

This multi-modality in areas arises from overlapping cell generations. Since we assume that cell target areas decrease upon division ($R_A < 1$), each successive gener-

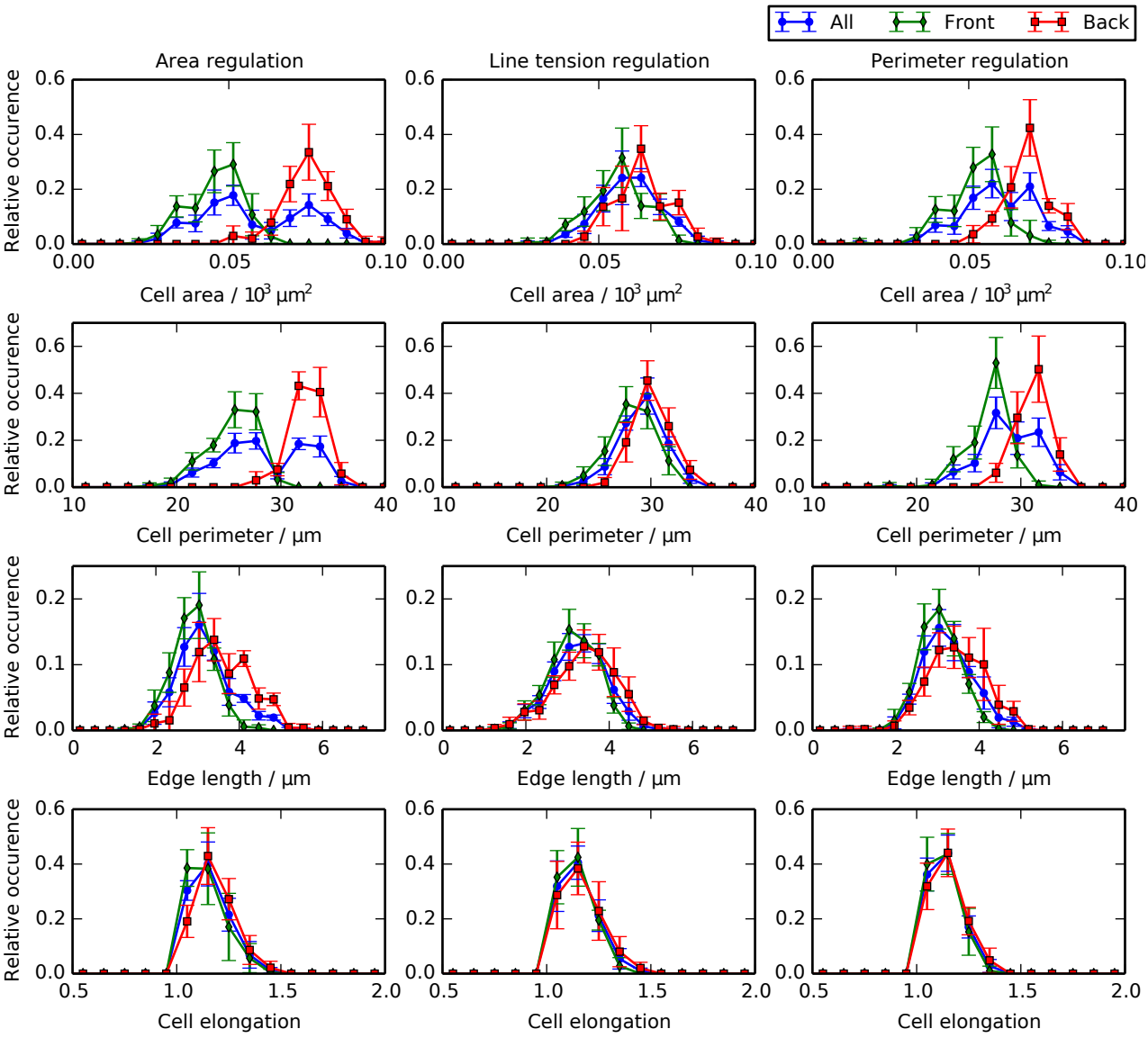

Figure 4.10: Differential growth and mechanical regulation generate distinct distributions of cell shapes. Distributions of cell areas (row 1), cell perimeters (row 2), cell edge lengths (row 3), and cell elongations (row 4) for the *wt* simulations of each scenario of cellular asymmetry. We distinguish distributions for all cells in the posterior compartment ('All'), for cells in the front half only ('Front'), and cells in the back half only ('Back').

ation of cells has a smaller target area. In simulations of the *en>CycE* perturbation, some cells divide twice while others only divide once, resulting in a bimodal cell area distribution. Similarly, for the *en>dap* perturbation, some cells divide once while others don't divide at all; however, we also observe area differences between cells in the front and the back half of the P compartment (Figure 4.10). These effects combine to yield a trimodal cell area distribution.

In summary, the area distributions of the genetic perturbations *en>dap* and *en>CycE* may be used as a measure to validate the model assumptions, and provide a further tool to distinguish the 'area regulation' scenario from the 'perimeter regulation' and 'line tension regulation' scenarios.

4.3.7 Simulated laser ablation experiments allow discrimination between asymmetry scenarios

As a further analysis of the model, we performed a laser ablation analysis on the final configuration of our *wt*, *en>dap* and *en>CycE* simulations. In 100 simulations for each perturbation, we 'cut' a randomly selected cell-cell interface (edge) in the P compartment. This was implemented by setting the line tension parameter $\overline{\Lambda}$ for this edge, as well as the perimeter tension parameter $\overline{\Gamma}$ for the cells adjacent to the edge, to zero. We then ran each simulation for 200 further time units and recorded the average initial vertex recoil velocity and total vertex recoil distance. Results for each of the three asymmetry scenarios are shown in Figure 4.12. We find that under the 'perimeter regulation' scenario, the average initial vertex recoil velocity and total vertex recoil distance are both smaller in each perturbation than in *wt*. In contrast, under the two other asymmetry scenarios there is no significant difference in these statistics across *wt*, *en>dap* and *en>CycE* simulations. These results offer a further experimentally testable prediction that, in conjunction with the cell area distribution results summarised in Figure 4.10, allows for discrimination between the

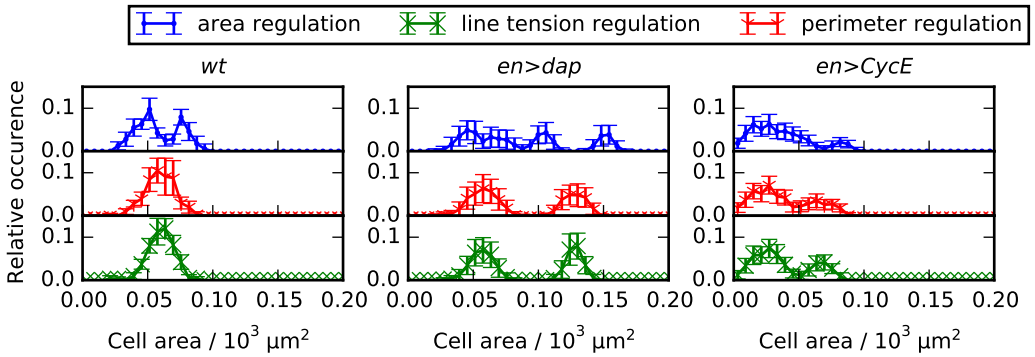

Figure 4.11: Cell area distributions in the *en>dap* and *en>CycE* perturbations are multimodal. Distributions of cell areas for each perturbation of cell division events (*wt*, *en>dap* and *en>CycE*) and each scenario of cellular asymmetry. Cell areas are recorded at the end of each simulation and error bars denote standard deviations across 100 simulations. Parameter values are given in Table 4.1.

three asymmetry scenarios considered.

4.4 Discussion

In this chapter we have applied a vertex model of a *Drosophila* embryonic segment to test hypotheses about the emergence tissue of size control. A comparison of the *in silico* segment with extant literature values indicated that passive mechanical forces suffice to explain the observed size control. However, the observed spatial heterogeneity in cell death frequencies requires some form of patterning of mechanical properties across the tissue. Several conceptually distinct modifications of the model can explain size control while also recapitulating the spatially varying rates of cell death: first, individual cells could regulate their sizes through differential growth; and second, cells could regulate their apical mechanical properties through differential expression of tension regulating protein activities. It is possible to distinguish these two scenarios within the model by the spatial distribution of P compartment cell areas and perimeters, as well as by the speed of vertex recoil after laser ablations. These results hint at two possible mechanistic functions of trophic (growth-promoting) signalling pathways,

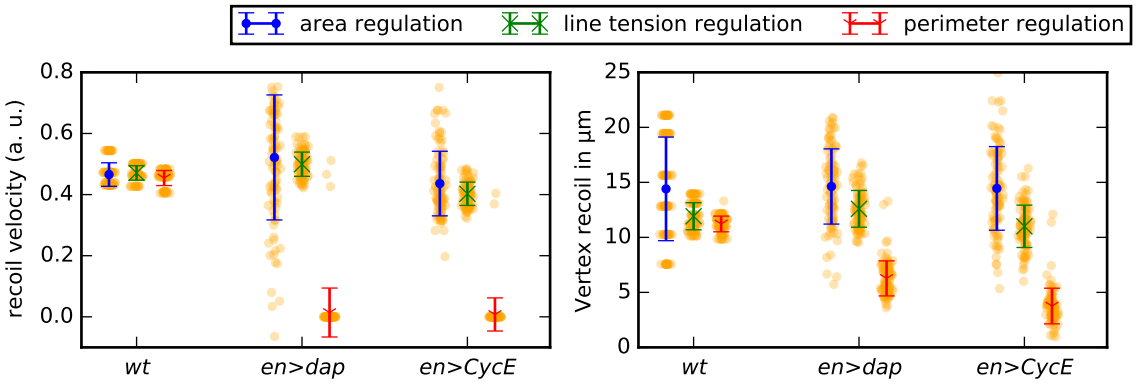

Figure 4.12: Simulated laser ablation experiments allow discrimination between asymmetry scenarios. Average initial vertex recoil velocities and total recoil distances across simulations of *wt* and perturbations. Error bars denote standard deviations across 100 simulations. Parameter values are given in Table 4.1 and in Section 4.3.

such as EGFR or Wg (Parker, 2006; O'Keefe et al., 1997; Szüts et al., 1997): they could either cause growth of individual cells, or else modulate cell shape through the regulation of contractile cytoskeletal activity, either of which would explain the experimentally observed shrinkage or growth when the pathways are perturbed (Parker, 2006).

4.4.1 Connecting robustness of proportional size control to cell mechanics

Understanding the mechanism of tissue size control is particularly challenging due to the interconnected and complex nature of cell signalling and the high degree of feedback between cell- and tissue-level processes. Computational models therefore offer an important tool for investigating and testing hypothesised mechanisms and to abstract the principles underlying developmental robustness (Iber et al., 2015; Schwarz and Dunlop, 2012; Xiong and Megason, 2015).

The development of multicellular organisms requires control of total cell numbers and relative proportions of cell types with tissues. Size control can be divided into

two steps: initial specification and maintenance (Li et al., 1999). Much work has focused on the spatial regulation of cellular fates during early embryonic development. Traditionally, tissue size specification has been associated with signalling gradients (Wolpert, 1969; Crick, 1970; Rogers and Schier, 2011). However, the mechanisms that ensure the maintenance of tissue size and of boundaries between tissues is less well understood. In particular, the physics of size homeostasis for patterned epithelia are not well understood, yet they are a recurring theme in development (Vakulenko et al., 2009; von Dassow and Davidson, 2011) and it is increasingly recognised that mechanical feedback plays a role in controlled tissue behaviour (Buchmann et al., 2014; Shraiman, 2005).

A gradient growth model has previously been proposed for the regulation of P compartment size in the *Drosophila* embryonic epidermis (Parker, 2006). This conceptual model requires the correct maintenance of a morphogen gradient in the face of multiple genetic perturbations. The present study demonstrates that an alternative, passive mechanical model can partially explain robustness of P compartment sizes and cell numbers in the *Drosophila* embryonic epidermis, eliminating the need for a tightly controlled intermediary morphogen gradient. More detailed cell-level analysis and modelling is required in the future to fully understand how morphogen signals are established, maintained, and interpreted (Hufnagel et al., 2007; Wartlick et al., 2011b), especially in the face of genetic or environmental perturbations.

Advancing our knowledge of how embryos achieve robustness to defects or damage to the initial patterning of tissue domains is important for understanding the underlying causes of birth defects, as well as diseases with an underlying basis of misregulated growth, such as cancers.

4.4.2 Providing predictions to guide future experimental inquiries into pattern repair

Although several studies have investigated the robustness of sizes of patterned epidermal segments of *Drosophila*, quantification has been somewhat sporadic and diffuse. This quantification will, in general, require a thorough systems-level characterization of later stages of *Drosophila* morphogenesis for multiple experimental perturbations. The present study provides a basis for guiding future experiments that seek to identify possible modes of size control in late stages of epithelial development in *Drosophila*.

How could model predictions be validated against such experiments? Several previous vertex models of developing epithelia have been validated against key summary statistics. Such studies have focused primarily on the *Drosophila* wing imaginal disc, which undergoes up to nine rounds of divisions to arrive at a distinct distribution of cell polygon numbers (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010). In these studies, it is safe to assume that the initial distribution of cell polygon numbers will not affect simulation outcomes, due to the high levels of proliferation. Here, we considered one or two rounds of divisions; over such a short developmental timespan we expect the initial sheet topology to influence final polygon distributions. Hence, for a quantitative comparison of this summary statistic between model and data, experimentally informed cell shapes of late stage 10 segments may be required. Such summary statistics remain lacking for the *Drosophila* embryonic epidermis during its development, and collecting them poses an experimental challenge due to the small system size (20-60 cells). Large sample sizes will be required to obtain accurate distributions of cell polygon numbers. Figures 4.11 and 4.12 in this study suggest that distributions of cell areas, and characterization of vertex recoils following standard laser ablation experiments, for genetic perturbations of the P compartment may be used to validate the underlying computational model. Thus, future iterations of the model may be further constrained through inference of mechanical parameters from

laser ablation (Landsberg et al., 2009) or less invasive experimental protocols (Ishihara and Sugimura, 2012).

4.4.3 Current limitations

Embryogenesis is an extremely complex process. To make headway into understanding the factors that influence robustness of tissue size maintenance, there needs to be conscious decoupling and abstraction of these factors through studies of simpler systems. This is also part of the rationale for studies in genetic model organisms from the worm and fly to mouse (Xiong and Megason, 2015).

Due to the lack of kinematic data on cell shapes and compartment sizes during the latter stages of embryogenesis, we have not included an analysis of such kinematic data for this initial study and have focused on more local mechanisms. In particular, we assumed that the overall tissue dimensions are constant during the considered time frame, since the epidermis forms at the outside of the embryo during stage five of *Drosophila* development and as a whole does not change dimensions for the remainder of development. However, larger scale tissue morphogenetic movements, which are undoubtedly important for some aspects of morphogenesis (Butler et al., 2009), may affect the exact size of a given subsection of the tissue. For example, dorsal closure occurs during the considered time frame, which leads to an extension of the tissue that we study (Gorfinkiel et al., 2009). The assumption that this extension should not affect the relative proportions of A and P compartment size requires future experimental validation. In addition, our finding that elevated tension along compartment boundaries does not affect compartment sizes may be contrasted with theoretical and experimental studies showing how differential line tension, either at compartment boundaries or across tissues, may drive convergent extension (Rauzi et al., 2008; Vroomans et al., 2015). A key conceptual difference between the present work and these studies is the assumption of a fixed, or free, boundary to the tissue.

In vertex models with a free boundary, contractile forces along cell perimeters may lead to deviations of cell areas from their respective target areas. The analysis of simulations with changed initial target areas presented in Figure 4.6 reveals that such deviations between cell target areas and their absolute areas may lead to increases in predicted apoptotic rates. Further investigation is required to understand the boundary conditions that best represent the effect of adjacent tissues in different epithelia, and the effects that forces along tissue boundaries can have on different summary statistics. It may be possible to gain insights to this question by quantifying tissue-level kinematics of germ-band retraction for the *wt* and developmental perturbations.

Our model employs the quasi-steady state assumption that the tissue is at mechanical equilibrium at each time point. We justified this assumption on the basis that individual cell cycle times of the 16th division cycle in *Drosophila* development are around an hour (Ashburner, 2011). However, if cell divisions occur highly synchronously, then this assumption might not hold. In *en>CycE* embryonic segments, the numbers of cell division events in the model were inferred from data where apoptosis was blocked by expressing the protein p35 in the P compartments (Parker, 2006). It has previously been reported that epithelial sheets can extrude cells that are not undergoing apoptosis (Marinari et al., 2012); if this occurs to a great extent in the *Drosophila* embryonic epidermis, then our inferred numbers of mitotic events would require adjustment. In this case, an *in vivo* cell tracking study would be necessary to measure the levels of cell division and extrusion events. Such data would also help to shed light, for example, on the possible impact of mitotic cell rounding on local cell shapes and possible short-range correlations between mitosis and apoptosis events. Since apoptosis in the vertex model is a passive process, we cannot extend our model analysis to p35 mutants in which apoptosis is blocked. How to adapt vertex models in such a way as to prevent the occurrence of T2 transitions while retaining tissue

integrity remains an open question.

Due to a current lack of data in the literature, our model does not include a description of upstream patterning of cell types. Instead, we infer the necessity of patterning of cell mechanics through simulations. This study is timely as it provides some guidance into important parameters and considerations that should be taken into account in future quantitative analyses of the late stages of epidermal development, including germband retraction and head involution. From the results presented here, further questions arise. If a passive mechanical model is sufficient to explain compartmental size control, then what is the functional role of Spitz-mediated EGFR regulation? It is known that EGFR signalling is required for dorsal closure during *Drosophila* development (Shen et al., 2013). Hence, it is possible that the influence of EGFR signalling on larval compartment sizes reflects the role of EGFR signalling in convergent extension during dorsal closure. If the asymmetry in our model reflects patterning of mechanical properties through trophic signalling, then a more detailed experimental analysis of the spatio-temporal dynamics of cellular signalling will allow more detailed modelling of how these properties may be patterned.

The results in this chapter are not sensitive to the mechanical parameters $\overline{\Lambda}$ and $\overline{\Gamma}$ of the model, enabling us to make experimental predictions without knowing the exact value of these parameters. However, in other contexts it may be necessary to measure these parameters from experimental data. Parameter inference for cell-based models, such as the vertex model, is not trivial and requires a careful choice of summary statistics. Specifically, the uncertainty quantification of such parameter estimation is challenging. For this reason, inference of the mechanical vertex model parameters $\overline{\Lambda}$ and $\overline{\Gamma}$ from experimental data will be the focus of the next chapter.

Chapter 5

Parameter inference on vertex models

Having applied the vertex model to tissue size control as a fundamental question in developmental biology in Chapter 4, we next investigate methods for inferring the mechanical parameters of the vertex model from microscopy imaging data.

5.1 Background and motivation

Cell-based models typically include mechanical parameters. For example, spring constants may define the interaction strength between cells in a cell-centre model (Meineke et al., 2001) or the elasticity of cell boundaries in immersed-boundary models (Cooper et al., 2016). The vertex model considered in Chapters 2, 3 and 4 includes two mechanical parameters, $\overline{\Gamma}$ and $\overline{\Lambda}$, in equation (2.4). Such parameters may strongly influence the behaviour of a model. For example, the interaction strength between cells in a cell-centre model has an impact on the overall cell density (Murray et al., 2009), and in the vertex model, $\overline{\Gamma}$ and $\overline{\Lambda}$ affect cell shapes and their packing properties (Farhadifar et al., 2007). However, the optimal choice for mechanical parameters to model a specific system accurately is often not intuitive, since these mechanical

parameters reflect properties of single cells or pairwise interactions. To measure the mechanical properties of a single cell it may be necessary to remove it from the surrounding tissue, which may, in turn, influence its properties.

In the case of vertex models, efforts have been made to estimate mechanical parameters from tissue-scale measurements. Farhadifar et al. (2007) analyse *Drosophila* wing imaginal discs and estimate the parameters $\overline{\Lambda}$ and $\overline{\Gamma}$ by manual fitting using a combination of summary statistics of cellular packing, as well as laser ablation experiments. In particular, Farhadifar et al. (2007) employed the distribution of cell neighbour numbers and the average area of cells for each polygon class, as well as the recoil of the vertices adjacent to an ablated edge, and the changes in cell area and perimeter for cells sharing the ablated edge. Analysing the same tissue, i.e. *Drosophila* wing imaginal discs, Canela-Xandri et al. (2011) and Schilling et al. (2011) use the same summary statistics of cellular packing, but do not employ laser ablation experiments for the fitting of $\overline{\Lambda}$ and $\overline{\Gamma}$, whereas Mao et al. (2011) use the cell neighbour number distribution only. Nestor-Bergmann et al. (2016) study animal cap tissue from *Xenopus laevis* and identify the average circularity (or elongation) for cells of each polygon class as a preferable summary statistic over other choices. All of these authors arrive at different parameter estimates, which is not surprising given that the analysed tissues, summary statistics, and model implementations differ.

Different approaches have been taken to fit summary statistics obtained from vertex models to experimental data. Canela-Xandri et al. (2011) and Schilling et al. (2011) use an ad-hoc approach to obtain best-fit parameter values, whereas Mao et al. (2011) and Nestor-Bergmann et al. (2016) use a least-squares approach to estimate the mechanical parameters of the vertex model. Merzouki et al. (2016) infer parameters of the vertex model through tissue deformation and comparing stress-strain curves obtained from simulations with those obtained in experiments on monolayers of Madine-Darby Canine Kidney cells, a common model tissue for the analysis

of epithelia (Harris et al., 2012). In these experiments, a free monolayer was suspended between rods and the stress curve was recorded as the distance of the rods was increased. A similar approach is taken by Xu et al. (2015a,b), who highlight that stress-strain curves obtained in this way are affected by the amount and orientation of cell divisions. Similarly, Wyatt et al. (2015) show that cell divisions in tissues under stress orient such that the stress is reduced.

Studies that seek to measure cellular mechanical properties in epithelia also include force-inference methods (Chiou et al., 2012; Ishihara and Sugimura, 2012; Ishihara et al., 2013). Force-inference methods typically assume that the tissue is in a steady state and fit a heterogeneous generalisation of the vertex model to microscopy data of epithelia in order to find the values for apical tension and pressure forces on each cell that can best explain its shape in relation to the shapes of its neighbours. Such force-inference approaches are particularly suitable to estimate mechanical heterogeneity in a tissue without the need to physically manipulate the sample.

None of the studies mentioned above quantify the uncertainty for estimations of mechanical parameters in the vertex model introduced in Chapter 2, or rigorously investigate the most informative choice of summary statistic or experimental setup for parameter estimation *in vivo*. While allowing an accurate estimation of parameters, tissue manipulations such as those discussed by Merzouki et al. (2016), Xu et al. (2015a) and Harris et al. (2012) are not directly applicable to developing embryos *in vivo*, since they require the removal of the tissue under investigation from its substrate. Force-inference methods have proven successful at estimating spatial mechanical heterogeneity and give access to the uncertainty associated with such measurements. However, it is unclear how force-inference results could be translated into optimal choices for the parameters in a vertex model, since they rely on a much larger set of parameters.

The aim of this chapter is to investigate inference methods for vertex models

that are applicable *in vivo* and allow uncertainty estimation. Specifically, we aim to identify which summary statistics and *in vivo* measurements are most informative in estimating vertex model parameters. To do so, we return to the simulations of growth in the *Drosophila* wing disc that we introduced in Chapter 3.

When estimating parameters in stochastic biological systems, approximate Bayesian computation (ABC) (Beaumont et al., 2002; Beaumont, 2010) is often used. In Bayesian statistics it is common to consider the joint probability distribution $p(\Theta, D)$ of parameters Θ and data D in order to calculate the posterior distribution $p(\Theta|D)$, the probability distribution of the parameters given the data. The calculation of the posterior is achieved by applying Bayes' rule

$$p(\Theta|D) = \frac{p(D|\Theta)p(\Theta)}{p(D)}. \tag{5.1}$$

Here, $p(D|\Theta)$ is the probability of the data D given the parameter Θ, and is usually referred to as the likelihood, and $p(D)$, the probability of observing the data, is the marginal likelihood. The likelihood of the parameter $p(\Theta)$ is referred to as the prior. In equation (5.1) the data D and the parameter Θ are one-dimensional. However, this does not need to be the case in general and it is common that both the data and parameter are vector-valued.

For complex computational models, such as the vertex model, the likelihood is not analytically tractable. In such cases, ABC provides a method to approximate the posterior through random sampling of parameters from the prior and evaluating the model for each sample. Samples for which a chosen summary statistic is sufficiently close to the observed data are considered samples of the posterior, otherwise they are simply discarded or 'rejected'. It is possible to refine the estimate of the posterior by conducting a regression adjustment step (Beaumont et al., 2002), which applies local-linear regression to move all accepted samples closer to the real posterior distribution as well as assigning accepted parameters that are closest to the observed data higher

weights in the estimated posterior.

Various adaptations to such rejection-based ABC exist. Well-established examples include: Markov Chain Monte Carlo approaches (MCMC-ABC) (Marjoram et al., 2003), which aim to reduce the number of rejected samples by generating each new sample in proximity to previously accepted samples; sequential Monte Carlo techniques (SMC-ABC) which iteratively redefine the prior to eventually converge to the posterior (Sisson et al., 2007); and approximate approximate Bayesian computation (AABC) (Buzbas and Rosenberg, 2015), which uses statistical surrogate models to enable the fast generation of a large number of samples. Here, we apply standard rejection-based ABC with regression adjustment as introduced by Beaumont et al. (2002), who showed on selected examples that the regression adjustment leads to better estimates of the posterior than rejection sampling alone. In contrast to more sophisticated methods, rejection-based ABC allows us to use a fixed set of sample simulations to evaluate the suitability of various summary statistics of the data at different points in parameter space.

We show how ABC methods can be applied to computational models of the growing *Drosophila* wing imaginal disc. We find that parameter estimates from summary statistics of cellular packing or from tissue responses to laser ablations are associated with high parameter uncertainty. We further identify the mean area of cells of each polygon class as the most suitable summary statistic for vertex model parameter inference and analyse the quality of the arising parameter estimates at different points in parameter space. The remainder of this chapter is structured as follows. In Section 5.2 we describe the simulations of vertex model tissue growth for which we infer parameters and describe the inference method in detail. In Section 5.3 we compare a range of summary statistics and their utility in inferring parameters. In Section 5.4 we discuss the implications of our findings.

5.2 Methods: ABC for vertex models

Here, we use the vertex model of growing *Drosophila* wing imaginal discs that we introduced in Chapter 3 to generate static cellular packings that correspond to minima of the energy (2.4).

5.2.1 Adaptations to the vertex model

Throughout this chapter, we use a modified version of the model described in Chapter 3, which leads to shorter computation times. This is required because we need to simulate the model using a large number of samples from the prior distribution. In the following, we describe the model modifications in detail, discuss the impact of these modifications to summary statistics of cell packing generated by the model, as well as the calculation times. Our modifications to the model include adjustments to the boundary and initial condition, as well as the model parameters.

Instead of initialising the simulation with 36 cells as in Chapter 3, we start our simulations with four cells (Figure 5.1A), and let the tissue grow to around 500 cells over the turn of $n_d = 7$ rounds of divisions (Figure 5.1B). We also introduce a new parameter, $\overline{\Lambda}_B$, which is used to calculate the line tension, i.e. the second term in equation (2.4), of cell interfaces with the tissue boundary. If the line tension parameter $\overline{\Lambda}$ is negative, we set the line tension of edges along the boundary to zero, $\overline{\Lambda}_B = 0$. Otherwise, we set $\overline{\Lambda}_B = \overline{\Lambda}$ as in Chapter 3. This adjustment of the boundary condition helps prevent boundary artefacts that lead to unphysical tissue shapes (Figure 5.1C). In Figure 5.1C, the simulation parameters are $\overline{\Lambda} = -0.85$ and $\overline{\Gamma} = 0.1$, and the line tension for cell edges along the tissue boundary is the same as throughout the tissue, i.e. $\overline{\Lambda}_B = \overline{\Lambda} = -0.85$. In simulations with this parameter choice, cells disconnect from each other or self-intersect, leading to failure of the simulation algorithm. In contrast, for simulations using $\overline{\Lambda} = -0.85$, $\overline{\Gamma} = 0.1$, $\overline{\Lambda}_B = 0$, the tissue grows normally and leads to a physically meaningful tissue configuration (Figure 5.1D). If $\overline{\Lambda}$ is negative,

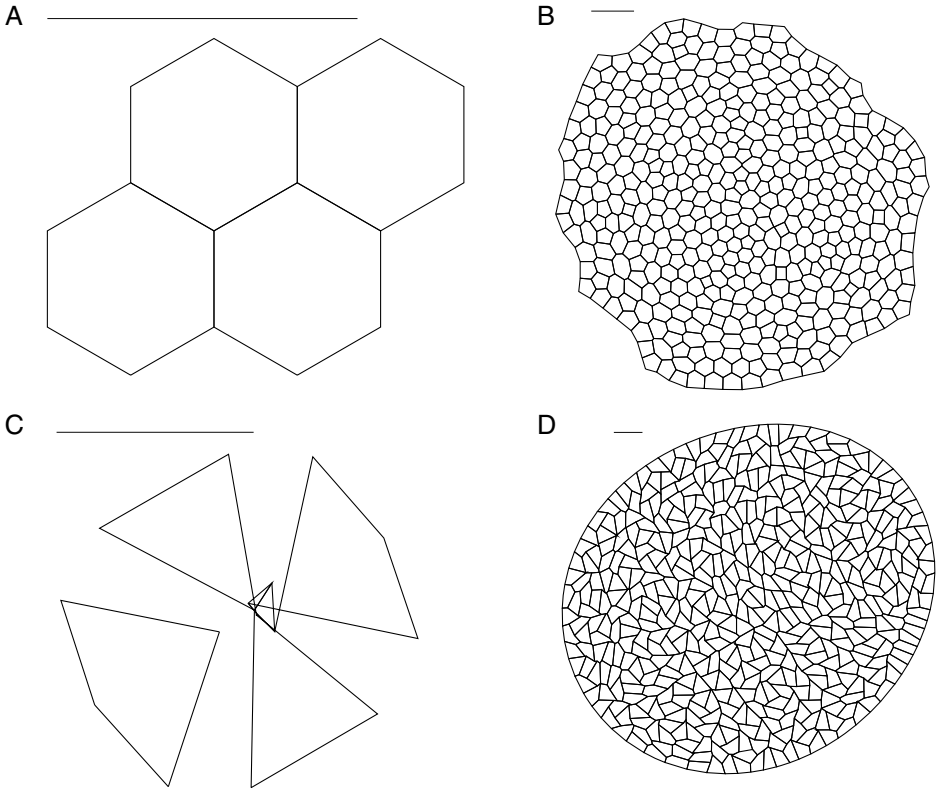

Figure 5.1: Simulation setup and boundary conditions (A) The initial condition comprises four hexagonal cells. (B) The tissue grows for $n_d = 7$ rounds of divisions until it comprises approximately 500 cells. The parameter values $\overline{\Lambda} = \overline{\Lambda}_B = 0.12$ and $\overline{\Gamma} = 0.04$ were used, all other parameter values are listed in Table 5.1. (C) If the line tension parameter $\overline{\Lambda}$ is negative in the bulk of the tissue as well as on the tissue boundary, the tissue may assume unphysical configurations. The parameter values $\overline{\Lambda} = \overline{\Lambda}_B = -0.85$ and $\overline{\Gamma} = 0.1$ were used, all other parameter values are listed in Table 5.1. (D) We prevent such boundary artifacts by setting the line tension along the tissue boundary to zero when $\overline{\Lambda}$ is negative, leading to well-defined tissue geometries. The parameter values $\overline{\Lambda} = -0.85$, $\overline{\Lambda}_B = 0.0$ and $\overline{\Gamma} = 0.1$ were used, all other parameter values are listed in Table 5.1. Panels (A-D) are rescaled to fit the view; a scale bar is added for comparison.

the energy contribution of individual edges decreases as the edge length increases. Since the motion of boundary vertices is unopposed by neighbouring cells, cells at the boundary may thus grow to arbitrarily large sizes or self-intersect if the associated gain in energy from the edges is sufficiently large. Such boundary effects are prevented if $\overline{\Lambda}_B = 0$. An adjustment of the boundary tension is not necessary if $\overline{\Lambda}$ is positive, since in this case it is energetically favourable for edges to shorten.

Further to adjusting the boundary condition, we choose a different set of simulation parameters to those proposed in Chapter 3, as listed in Table 5.1. The differences in parameter values between simulations in this chapter and those in Chapter 3 are: (i) the initial condition consists of four hexagonal cells instead of 36; (ii) the mean cell cycle duration is set to 20 instead of 1750; (iii) each cell undergoes seven rounds of cell division instead of five, and (iv) the simulation duration is 700 instead of 27000. Note that we non-dimensionalised the model in equations (2.4) and (2.3). Hence, the vertex model parameters used throughout this thesis are unit-less.

We chose these parameters so that the final tissue is sufficiently large to obtain tissue-level summary statistics of cell packing while minimising the amount of time that is required to run a single simulation. Ensuring that individual simulations of the model take a small amount of time to run is a prerequisite for conducting rejection-based ABC, since ABC requires simulation of the model using a large number of samples from the prior distribution. For example, throughout this chapter we use 100,000 samples. We will discuss simulation times further later in this section.

The described change in model parameters accelerates the propagation of the tissue from its initial to its final configuration. This change is justified, since, in this chapter, we do not study study temporal changes within the tissue during growth. We instead focus on the packing geometry and tissue properties at the end of the simulation, when the tissue has equilibrated into a stable, static configuration. In Chapter 3 a key descriptor of cellular packing, the average area of cells of each cell neighbour (or

polygon) number, was unaffected by changes in the cell cycle duration over multiple orders of magnitude. This suggests that characteristics of cellular packing in the final tissue configuration are not affected by an acceleration of tissue growth. In previous studies, summary statistics of cellular packing after tissue growth originally reported by Farhadifar et al. (2007) were similar in simulations that used alternative approaches to generate cellular packings, for example where the vertex model was relaxed from initial random Voronoi tessellations (Nestor-Bergmann et al., 2016). This emphasises that static tissue configurations in the vertex model are governed by the energy equation (2.4) and are less strongly affected by the dynamic processes leading up to the final configuration.

To estimate to what extent our adaptations to the model may influence tissue properties, we compare outcomes between our simulation procedure and the one described by Farhadifar et al. (2007) for three points in parameter space in Figure 5.2. In this figure, cell packings resemble those simulated previously by (Farhadifar et al., 2007). Further, the average area per polygon class as a sample summary statistic of cell packing is similar to those previously reported for all three considered parameter combinations. In Figure 5.2F, the mean area per polygon class deviates between our simulations and those reported by Farhadifar et al. (2007) for seven- and eight-sided cells. This difference may, however, not originate in the differences in parameter choices between the simulations in this chapter and those described in Chapter 3. Instead, it might originate from differences in boundary conditions between our simulations and those by Farhadifar et al. (2007), or from differences in the choice of cell-cycle model. We previously encountered differences in cell-neighbour number distributions between our simulations and those by Farhadifar et al. (2007) in Chapter 3. We conclude that the final vertex configuration in our model represent typical cell packing geometries of the vertex model for different points in $(\overline{\Gamma}, \overline{\Lambda})$ parameter space. Next, we discuss how we conduct ABC on the vertex model. We begin by motivating

Table 5.1: Description of parameter values used in this chapter.

Parameter	Description	Value	Reference
$\Delta t'$	Time step	0.01	Fletcher et al. (2013)
$A'_{\min}$	T2 transition area threshold	0.001	Fletcher et al. (2013)
l'_{T1}	T1 transition length threshold	0.01	Fletcher et al. (2013)
ρ	New edges after a T1 transition have the length $l'_{\mathrm{new}} = \rho l'_{\mathrm{T1}}$	1.5	Fletcher et al. (2013)
A'^s	Initial cell area	1.0	Farhadifar et al. (2007)
A'^s_0	Initial cell target area	1.0	Farhadifar et al. (2007)
N^s	Initial cell number	4	–
t'_l	Mean cell cycle duration	20	–
t'_{tot}	Simulation duration	700	–
n_d	Total number of divisions per cell	7	–
N_T	Total number of samples investigated	100,000	–
M	Number samples used for posterior estimates	1000	–

The parameter values provided in this table are used throughout this chapter, unless stated otherwise. For parameter values for which no reference is given, please see Section 5.2 for details on how these values were chosen. Spatial and temporal parameters are non-dimensional (see Section 2.2 for details).

the shape of our prior and our implementation of rejection-based ABC with regression adjustment.

5.2.2 Definition of the prior

In this chapter, we aim to infer the mechanical parameters of the vertex model from snapshots of simulated tissues without any prior knowledge of what the possible model parameters might be. Thus, in order to define the prior distribution of parameters it is necessary to understand which combinations of parameters in the vertex model may lead to realistic modelling outcomes, and where in parameter space the model is unphysical.

The parameter space of the vertex model with energy (2.2) has previously been investigated by Farhadifar et al. (2007) and Staple et al. (2010) and can be subdivided into three regions labelled A, B, and C in Figure 5.3. In region A, stable configurations,

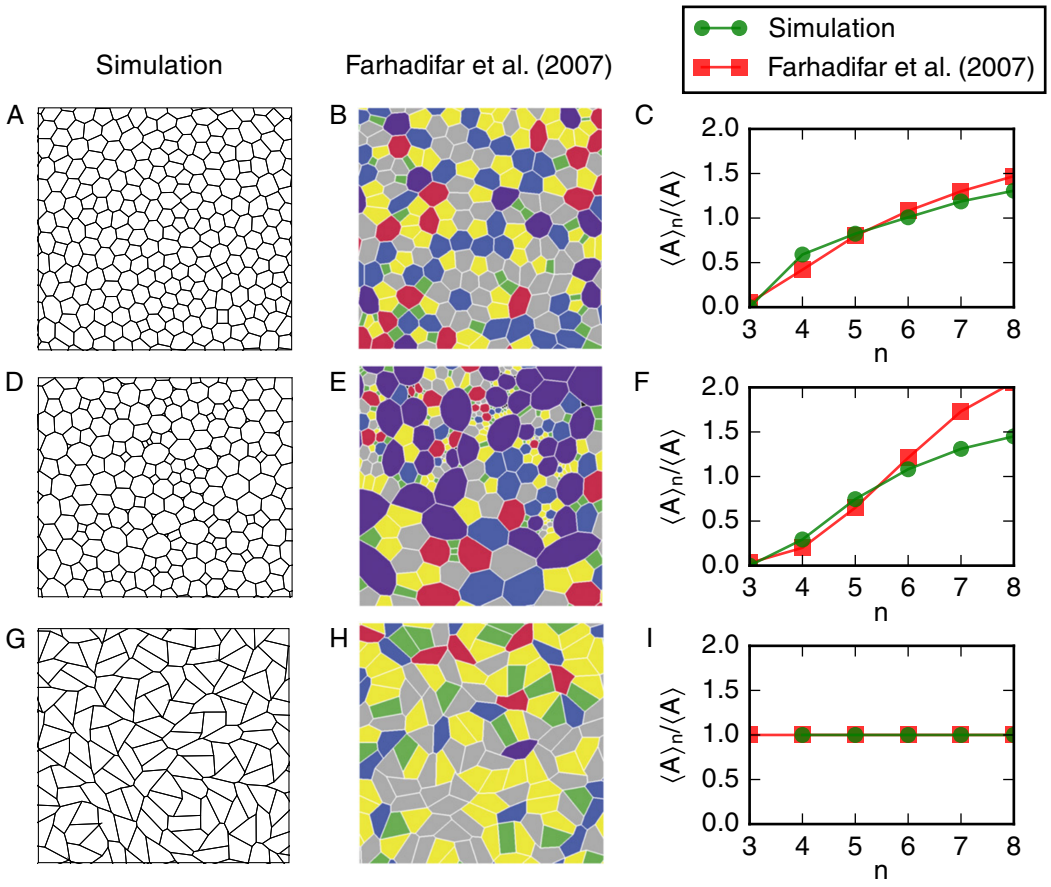

Figure 5.2: Cellular packings depend on the mechanical vertex model parameters $\overline{\Lambda}$ and $\overline{\Gamma}$. (A,D,G) Sample cell shapes at the end of the simulation for different parameter values. (B,E,H) Cell shapes from simulations reported by Farhadifar et al. (2007), reprinted with permission from Elsevier. (C,F,I) Cell shapes are quantified by the mean area of cells of each polygon class, and compared between our simulation outcomes and those by Farhadifar et al. (2007). Parameter combinations are $\overline{\Lambda} = 0.12$, $\overline{\Gamma} = 0.04$ (A-C); $\overline{\Lambda} = 0.0$, $\overline{\Gamma} = 0.1$ (D-F) and $\overline{\Lambda} = -0.85$, $\overline{\Gamma} = 0.1$ (D-F).

i.e. stable minima of the energy equation (2.4) do not exist, since in this region the minimal energy is degenerate and infinitely many vertex configurations are possible at the energy minimum. This region is characterized by zero shear-stress of the tissue. We were unable to generate simulations of growing *Drosophila* wing discs in this regime, in agreement with previous results (Staple et al., 2010; Farhadifar et al., 2007; Nestor-Bergmann et al., 2016). Simulations in region A tend to stop prematurely due to self-intersecting or overlapping cells. In region B, well-defined energy minima exist, and cell packings vary as the parameters are changed within this region (see, for

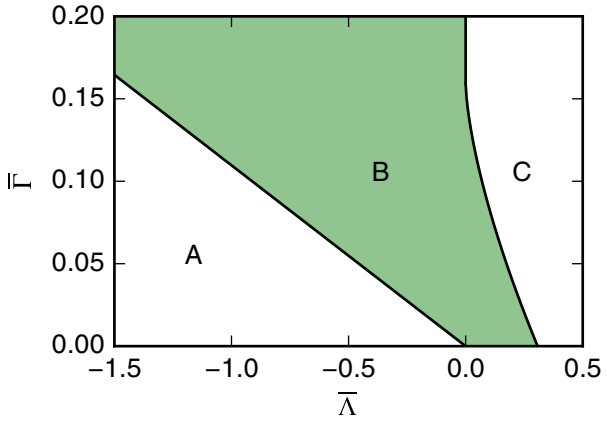

Figure 5.3: Definition of the prior. The parameter space of the vertex model with a free boundary and energy equation (2.4) can be subdivided into three regions denoted A, B, and C. Region A marks configurations of zero shear stress, where infinitely many vertex configurations are possible at the energy minimum of equation (2.4). In region B, vertex configurations can be stable, corresponding to a well-defined energy minimum. In region C, the energy is minimal when all cells have zero area and zero perimeter. Throughout this chapter, the prior is chosen to be a uniform distribution covering region B.

example, Figure 5.2). In region C, all cell areas are zero at the energy minimum, i.e. tissues initialised in this region shrink until all cells are removed from the tissue by T2 transitions. Thus, we define the prior as a uniform distribution across region B of Figure 5.3, i.e. a uniform distribution across the region in parameter space where the model leads to physical results, restricted to the ranges $-1.5 \leq \overline{\Lambda}$ and $0 \leq \overline{\Gamma} \leq 0.2$. The restriction of the parameter values to these ranges is common in the literature, and a wide variety of biophysically realistic model behaviours has been observed within this range (Farhadifar et al., 2007; Staple et al., 2010; Nestor-Bergmann et al., 2016).

The borders of region B can be estimated analytically by calculating the energy minima of equation (2.4) for tissues containing single, regular n-gons (Staple et al., 2010). Different polygon classes give rise to different boundaries for the transition between stable configurations in region B and shear-free configurations in region A. Among all polygon classes, this boundary occupies the 'left-most' position for tri-

angular cells, which allow the lowest values of $\overline{\Lambda}$. For cells of triangular shape, the boundary between region A and region B is described by the line (Staple et al., 2010)

$$\overline{\Lambda} = -4\overline{\Gamma}\sqrt{3\tan\pi/3}. \tag{5.2}$$

The boundary between region B and C assumes positions at increasing values of $\overline{\Lambda}$ as the polygon number increases. For large values of n, this position converges to (Staple et al., 2010)

$$\begin{aligned} \overline{\Lambda} &= \sqrt{\frac{8}{\pi}}\left(\frac{1-2\pi\overline{\Gamma}}{3}\right)^{3/2} \quad && (2\pi\overline{\Gamma} < 1), \\ \overline{\Lambda} &= 0 && (2\pi\overline{\Gamma} \geq 1). \end{aligned} \tag{5.3}$$

Using equations (5.2) and (5.3) we thus define the prior to exclude regions in parameter space where all simulations are unphysical. However, unphysical solutions may also occur in region B in Figure 5.3 for parameters that are close to the boundaries (5.2) and (5.3). Note that these boundaries are derived using free boundary conditions in the vertex model, such as in this study.

5.2.3 Sample generation

Within the region B of Figure 5.3 we generate $N_T = 100,000$ samples of our wing disc simulations using a supercomputer. Using the parameters listed in Table 5.1 a single simulation requires approximately five minutes to run on a single machine with a Intel(R) Core(TM) i5-5200U CPU with 2.20GHz and 8GB memory. Note that running all simulations requires approximately 83,000 hours of calculation time. Simulations in Chapter 3 take five to six times longer to run than those reported here. Of the 100,000 generated sample simulations, 10,762 simulations did not run to completion due to unphysical events, for example the generation self-intersecting or overlapping cells.

5.2.4 Implementation of ABC regression

In Section 5.3 we investigate various summary statistics and their utility in inferring vertex model parameters. Specifically, we evaluate the model at a reference parameter vector $\boldsymbol{\Theta}_{\text{obs}} = (\overline{\Lambda}_{\text{obs}}, \overline{\Gamma}_{\text{obs}})$. The samples used for parameter inference using ABC have parameter vectors $\boldsymbol{\Theta}_k = (\overline{\Lambda}_k, \overline{\Gamma}_k),\ k = 1, \ldots, N_T$. A given vector of summary statistics computed with $\boldsymbol{\Theta}_{\text{obs}}$ takes the vector form $\boldsymbol{s}_{\text{obs}} = (s_1, \ldots, s_S)^T$ with a total of S components, and the summary statistic at sample $\boldsymbol{\Theta}_k$ assume the values $\boldsymbol{s}_k = (s_{k1}, \ldots, s_{kS})^T$. Throughout this chapter, we use summary statistics where individual vector entries $s_1, \ldots, s_S$ have been rescaled by their respective standard deviations across all samples. This is a common procedure in ABC, since it reduces the impact of vector components with high variance onto the parameter estimate (Beaumont et al., 2002; Beaumont, 2010).

Inference is then conducted by first accepting the $M = 2,000$ closest samples based on an Euclidean distance measure to the reference statistic, $d_i = ||\boldsymbol{s}_i - \boldsymbol{s}_{\text{obs}}||_2$. Note that, when conducting rejection-based ABC, it is common to accept a small proportion of the total number of samples rather than all samples within a chosen, small distance. If the accepted proportion is sufficiently small, all accepted samples will be close to the observed summary statistic, thus avoiding the necessity of selecting an appropriate acceptance distance. The acceptance threshold, δ, can be identified as the maximal distance of all accepted samples

$$\delta = \max_{i=1,\ldots,M} \{d_i\}. \tag{5.4}$$

We then proceed by adjusting the accepted parameters $\boldsymbol{\Theta}_i = (\overline{\Lambda}_i, \overline{\Gamma}_i)$ using local-linear regression, as introduced by Beaumont et al. (2002). Specifically, we minimise

the penalty function

$$f(\boldsymbol{\alpha}, B) = \sum_{i=1}^{M} ||\boldsymbol{\Theta}_i - \boldsymbol{\alpha} - (\boldsymbol{s}_i - \boldsymbol{s}_{\text{obs}})^T B||^2 K_\delta(||\boldsymbol{s}_i - \boldsymbol{s}_{\text{obs}}||), \tag{5.5}$$

providing the regression parameters $\boldsymbol{\alpha}$ and B for a local-linear regression of the parameters dependent on the summary statistics,

$$\boldsymbol{\Theta}_i = \boldsymbol{\alpha} + (\boldsymbol{s}_i - \boldsymbol{s}_{\text{obs}})^T B + \boldsymbol{\epsilon}_i. \tag{5.6}$$

Here, K_δ refers to the Epanechnikov kernel function

$$K_\delta(||\boldsymbol{s}_i - \boldsymbol{s}_{\text{obs}}||) = \max\left\{\frac{3}{4\delta}\left(1 - \left(\frac{||\boldsymbol{s}_i - \boldsymbol{s}_{\text{obs}}||}{\delta}\right)^2\right), 0\right\}, \tag{5.7}$$

and the $\boldsymbol{\epsilon}_i$ are uncorrelated random variables with zero mean and a common variance. Note that no further assumptions are made as to the distribution of the $\boldsymbol{\epsilon}_i$. In practice, the penalty function (5.5) can be minimised by calculating the matrix C

$$C = \begin{pmatrix} \boldsymbol{\alpha} \\ B \end{pmatrix} = (X^T W X)^{-1} X^T W \hat{\Theta}. \tag{5.8}$$

This matrix C has two columns and $S + 1$ rows. The matrix W has diagonal entries $K_\delta(||\boldsymbol{s}_i - \boldsymbol{s}_{\text{obs}}||)$ in the i^{th} row and zero entries everywhere else, whereas the matrix X has the form

$$X = \begin{pmatrix} 1 & s_{11} - s_1 & \cdots & s_{1S} - s_S \\ \vdots & \vdots & \ddots & \vdots \\ 1 & s_{M1} - s_1 & \cdots & s_{MS} - s_S \end{pmatrix}. \tag{5.9}$$

The matrix $\hat{\Theta}$ is of dimension $2 \times M$ and contains all accepted samples, $\hat{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_M)^T$.

After calculating the matrix B using equation (5.8) we can perform the regression

adjustment on all samples,

$$\boldsymbol{\Theta}_i^* = \boldsymbol{\Theta}_{\boldsymbol{i}} - (\boldsymbol{s}_i - \boldsymbol{s})^T B; \tag{5.10}$$

thus shifting the samples towards the posterior at $\boldsymbol{s}_{\text{obs}}$ instead of $\boldsymbol{s}_i$. Note that the vector $\boldsymbol{\alpha}$ estimates the mean of the posterior distribution

$$\boldsymbol{\alpha} = \frac{\sum_{i=1}^{M} \boldsymbol{\Theta}_i^* K_\delta(||\boldsymbol{s}_i - \boldsymbol{s}||)}{\sum_{i=1}^{M} K_\delta(||\boldsymbol{s}_i - \boldsymbol{s}||)}. \tag{5.11}$$

To estimate posterior distributions, we use kernel density estimation with the Epanechnikov kernel. Specifically, each parameter vector $\boldsymbol{\Theta}_i^*$ is assigned the weight $w_i = K_\delta(||\boldsymbol{s}_i - \boldsymbol{s}_{\text{obs}}||)$, and the posterior density at parameter $\tilde{\boldsymbol{\Theta}}$ can be written as

$$p(\tilde{\boldsymbol{\Theta}}) = \frac{\sum_{i=1}^{M} K_\rho(\overline{\Lambda}_i^* - \tilde{\overline{\Lambda}}) K_\sigma(\overline{\Gamma}_i^* - \tilde{\overline{\Gamma}}) w_i}{\sum_{i=1}^{M} w_i}. \tag{5.12}$$

The kernel thresholds ρ and σ are used calculate the density in $\overline{\Lambda}$ and $\overline{\Gamma}$ parameter space, respectively. Note, that these thresholds are different from the previously used kernel threshold, δ, which is a measure of the distance between observed and simulated summary statistics. When using kernel-density estimation, the estimated density at a given parameter point may depend on the kernel thresholds ρ and σ. Throughout this chapter, we use least-squares cross-validation to select ρ and σ automatically (Li and Racine, 2007). Least-squares cross-validation, as described by Li and Racine (2007), is an established method to select kernel thresholds and aims to minimise the mean squared error of the density estimate.

5.3 Results: Estimating uncertainty in vertex model parameter inference

In the following, we investigate the efficacy of different summary statistics when inferring the parameters of a reference simulation that is run with a known set of parameters. We start with a reference simulation using the parameters $\overline{\Lambda}_{\mathrm{obs}} = 0.12$, $\overline{\Gamma}_{\mathrm{obs}} = 0.04$, a commonly used set of parameter values in the vertex model (Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010).

First, we analyse the posterior distributions generated using different summary statistics of cell packing. We start by focussing on first-order summary statistics that characterize the shapes of individual cells, before turning to second-order summary statistics of cell packing that reflect the relationships between shapes of adjacent cells. Note that vertex positions in our model are measured in units of the characteristic length scale of the tissue, which we set equal to the square root of the target area of cells in the tissue, $L = \sqrt{A_0}$. Since this target area, A_0, may differ from the actual area of cells it is not experimentally accessible. Throughout this chapter we thus consider dimensionless summary statistics. Specifically, we divide all area-based summary statistics by the average area of cells in the tissue, and all length-based summary statistics by its square root. A full list of all summary statistics used in this chapter and their mathematical description is provided in Table 5.2.

5.3.1 Parameter estimates from summary statistics of cell shapes have high uncertainty

First-order summary statistics of cell packing are often used for parameter estimation in vertex models. The most common examples of these are the distribution of polygon (cell neighbour) numbers and the average area of cells for each polygon class. Here, we employ both of these summary statistics for parameter inference, in addition to

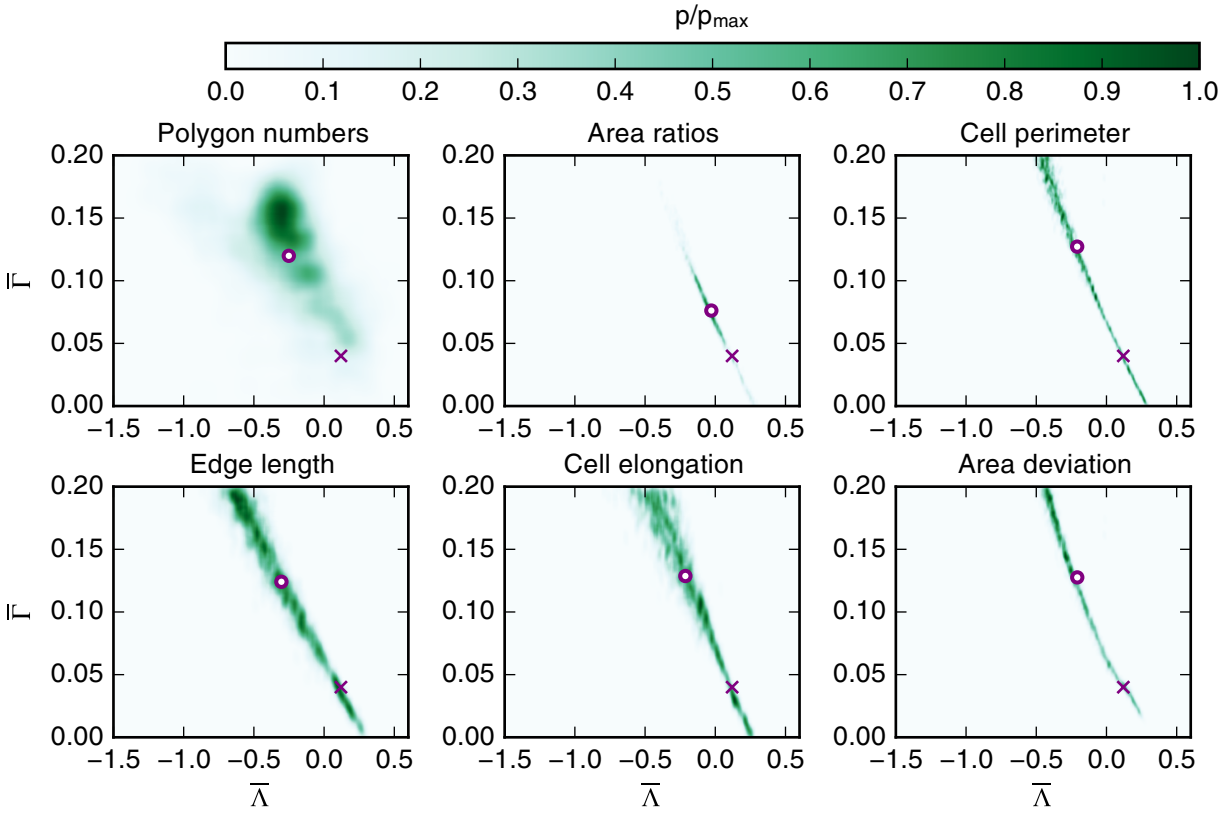

Figure 5.4: Posterior distributions generated using different summary statistics of cell shape spread across parameter space. Posterior distributions of first-order summary statistics of cell shape are shown in green. Model and inference parameters are listed in Table 5.1 and mathematical descriptions of the summary statistics are provided in the text and listed in Table 5.2. Crosses mark the reference parameter vector of the observed data for which the posterior is estimated. Circles mark the means of the posterior distributions.

the average cell perimeter, the average edge length (i.e. the average length of cell-cell interfaces), the average cell elongation, and the standard deviation of cell areas within the tissue. Note that we introduced the cell elongation summary statistic in Chapter 4, and provide its mathematical description in Section 4.2. Similar to our analysis in Chapter 3, all summary statistics of cell shapes are averaged over all cells in the tissue that do not share an edge with the tissue boundary, since cell shapes along the tissue boundary differ from those in the bulk of the tissue.

The approximate posterior distributions generated by ABC using these first-order

summary statistics are shown in Figure 5.4. In this figure, the value of the reference parameter vector is indicated by a purple cross, whereas the mean of the posterior distribution is indicated by a circle. Out of the posterior distribution in Figure 5.4, none estimates the parameters of the reference simulation correctly, based on the means of the posterior distributions. The closest, i.e. the most accurate, parameter estimate is achieved by the summary statistic 'Area ratios', with a parameter estimate of $(\overline{\Lambda}, \overline{\Gamma}) = (-0.28, 0.076)$. For this summary statistic, the posterior distribution is also the most narrow and concentrated, i.e. it is the posterior distribution with the lowest degree of uncertainty. The marginal standard deviations are 0.14 and 0.035 in $\overline{\Lambda}$ and $\overline{\Gamma}$, respectively. All of the posterior distributions spread across parameter space instead of concentrating in an area close to the true parameter vector. Thus, the parameter estimates are all associated with a high degree of uncertainty, and therefore without knowing the reference parameter set it would be difficult for us to evaluate the quality of parameter estimates, including the cases of the commonly used polygon distribution and the cell area ratios.

5.3.2 Posterior distributions are robust to the total number of samples and the number of accepted samples

When conducting ABC, it is important to check that the method is applied correctly. It is necessary to check that the total number of samples, N_T, is sufficiently large, and that the number of accepted regression samples, M, is sufficiently small in order to generate accurate representations of the posterior. In Figure 5.5 we compare the posterior distributions obtained with $N_T = 100,000$ and $M = 2,000$ (Figure 5.5A), to those obtained with $N_T = 50,000$ and $M = 1,000$ (Figure 5.5B), and those obtained with $N_T = 100,000$ and $M = 1,000$ (Figure 5.5C). In Figure 5.5B, both N_T and M are reduced while fixing the acceptance ratio M/N_T at 0.02. As Figures 5.5A-C show, the posterior distributions obtained using the summary statistics 'Polygon numbers'

Table 5.2: Description of summary statistics used in this chapter.

Abbreviation	Summary statistic	Description
Polygon numbers	$(p_4, \ldots, p_8)$	Cell neighbour number distribution
Area ratios	$(\langle A\rangle_4/\langle A\rangle, \ldots, \langle A\rangle_8/\langle A\rangle)$	Mean area of cells of each polygon class
Cell perimeter	$\langle P\rangle/\sqrt{\langle A\rangle}$	Average cell perimeter
Edge length	$\langle l_{i,j}\rangle/\sqrt{\langle A\rangle}$	Average edge length
Cell elongation	$\langle\lambda\rangle$	A measure for the elongation of each cell, cf. Section 4.2.
Area deviation	$\sigma(A)/\langle A\rangle$	Standard deviation of the cell areas
Area correlation	$\mathrm{corr}(A_i, A_j)$	Correlation between areas of adjacent cells i and j
Polygon number correlation	$\mathrm{corr}(N_i, N_j)$	Correlation between neighbour numbers of adjacent cells i and j
Neighbour areas	$(\langle A_{\mathrm{neigh}}\rangle_4/\langle A\rangle, \ldots, \langle A_{\mathrm{neigh}}\rangle_8/\langle A\rangle)$	Mean neighbour area of cells for each polygon class
Neighbour numbers	$(\langle N_{\mathrm{neigh}}\rangle_4, \ldots, \langle N_{\mathrm{neigh}}\rangle_8)$	Mean neighbour number of cells for each polygon class
Laser recoil	$\langle d_{\mathrm{recoil}}\rangle/\sqrt{\langle A\rangle}$	Recoil distance of vertices adjacent to an ablated edge
Area asymmetry	$\langle\Delta A\rangle/(d_{\mathrm{recoil}}\sqrt{\langle A\rangle})$	Relative change in area of the adjacent cells over the vertex recoil
Perimeter asymmetry	$(\langle\Delta P\rangle - 2d_{\mathrm{recoil}})/d_{\mathrm{recoil}}$	Relative change in perimeter of the adjacent cells over the vertex recoil

Further motivation for each summary statistic is provided in the text. p_n denotes the proportion of cells of polygon class n. The notations A_{neigh} and N_{neigh} denote the average area of a cell's neighbours and the average polygon number of a cell's neighbours, respectively. The notation $\langle\cdot\rangle_n$ denotes the average over all cells of polygon class n, whereas the notation $\langle\cdot\rangle$ denotes the average over all cells in the tissue. Both averages exclude cells at the tissue boundary. For the summary statistics 'Polygon numbers', 'Area ratios', 'Neighbour areas' and 'Neighbour numbers' only cells of polygon classes four to eight are considered, since cells with different neighbour numbers are rare in the tissue, compare, for example, Figure 3.3.

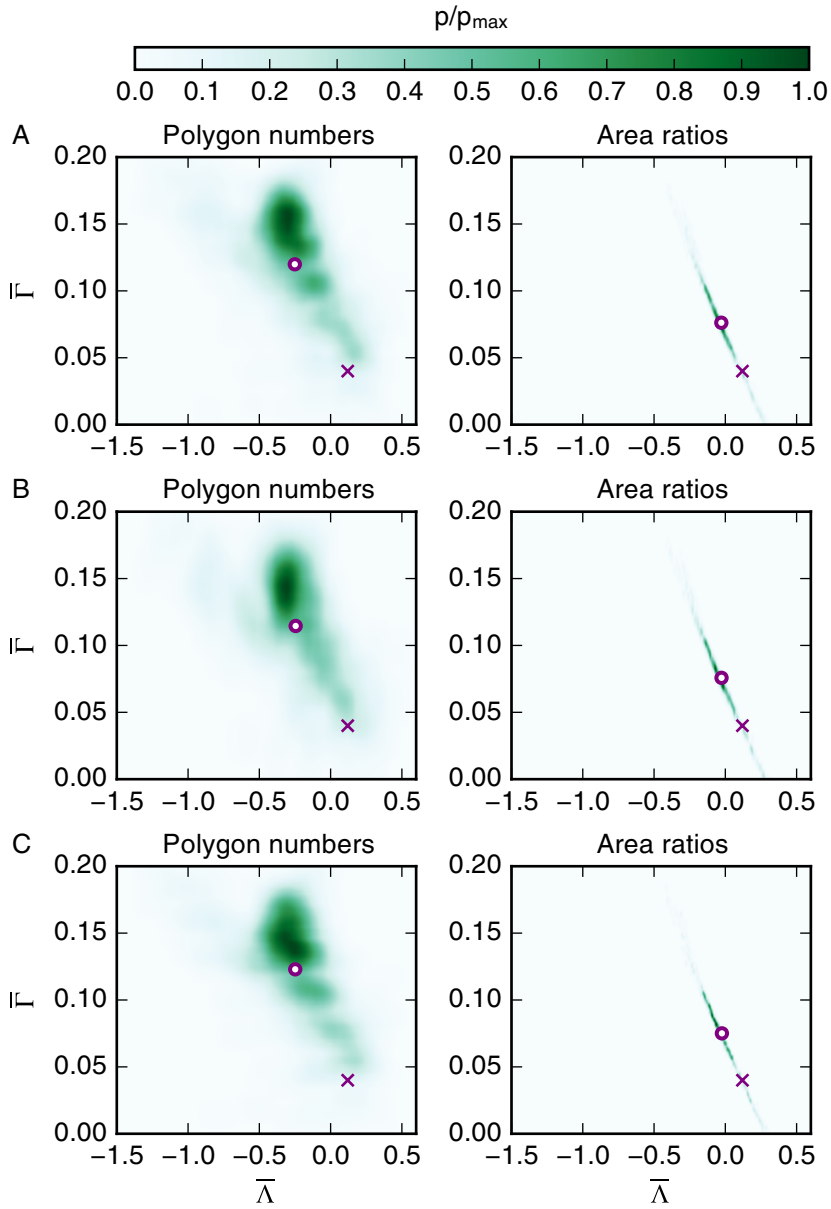

Figure 5.5: Posterior distributions are robust to inference parameters N_T and M. (A) Posterior distributions of the summary statistics 'Polygon numbers' and 'Area ratios' obtained using the inference parameters $N_T = 100,000$ and $M = 2,000$. (B) Posterior distributions obtained by using half the total number of samples, $N_T = 50,000$. We also halved the acceptance threshold to $M = 1,000$ in order to ensure that the acceptance ratio M/N_T is unaffected. (C) The posterior distributions obtained when halving the acceptance ratio, i.e. $N_T = 100,000$ and $M = 1000$. Model and inference parameters are listed in Table 5.1 and mathematical descriptions of the summary statistics are provided in the text and listed in Table 5.2. Crosses mark the reference parameter vector of the observed data for which the posterior is calculated. Circles mark the means of the posterior distributions.

and 'Area ratios' are not strongly affected by such changes of the inference parameters N_T and M and the characteristics of these distributions, including their mean value, are preserved. For example, for the 'Polygon numbers' summary statistics, the mean values of $(\overline{\Lambda}, \overline{\Gamma})$ are $(-0.251, 0.12)$, $(-0.245, 0.115)$ and $(-0.248, 0.123)$ for cases A, B and C in Figure 5.5, respectively. We conclude that, for our choice of N_T and M, the ABC method has converged and that the obtained posterior distributions are sensible.

5.3.3 Parameter estimates from second-order summary statistics share characteristics of parameter estimates from first-order summary statistics

In Figure 5.6 we investigate the efficacy of second-order summary statistics in estimating vertex model parameters. We focus on four quantities: the 'Area correlation', i.e. the correlation coefficient between areas of adjacent cells; the 'Polygon number correlation', i.e. the correlation coefficient between neighbour numbers of adjacent cells; the 'Neighbour areas', i.e. the mean area of neighbours for cells of each polygon class; and the 'Neighbour numbers', i.e. the mean polygon number of neighbours for cells of each polygon class. Throughout this chapter, we calculate correlations between two random variables A and B by

$$\mathrm{corr}(A, B) = \frac{\langle (A - \langle A \rangle)(B - \langle B \rangle) \rangle}{\sigma_A \sigma_B}, \tag{5.13}$$

where σ_X denotes the standard deviation of a random variable X. In the case of 'Area correlation' and 'Polygon number correlation', the random variables are the areas and polygon numbers of adjacent cells, respectively. In these cases, the average in equation (5.13) is taken over all pairs of adjacent cells.

In general, the second-order statistics in Figure 5.6 suffer the same drawbacks as

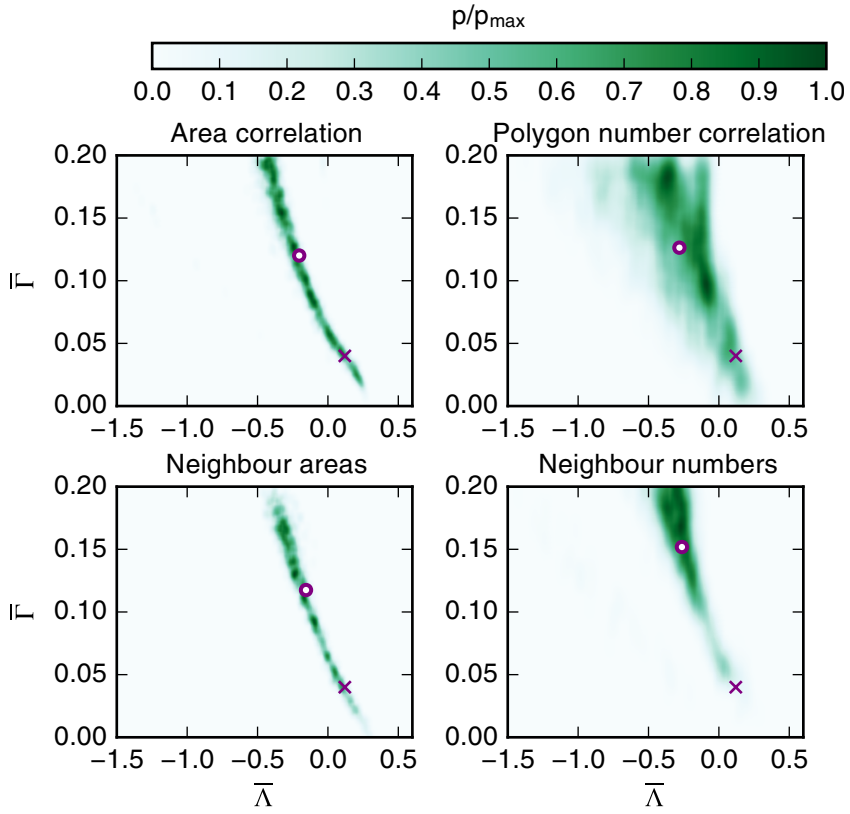

Figure 5.6: Posterior distributions obtained using second-order summary statistics of cell shape have similar shapes to those using first-order statistics. Posterior distributions of second-order summary statistics are shown in green. Model and inference parameters are listed in Table 5.1 and mathematical descriptions of the summary statistics are provided in the text and listed in Table 5.2. Crosses mark the reference parameter of the observed data for which the posterior is calculated. Circles mark the means of the posterior distributions.

the first-order summary statistics. In particular, they all have drawn-out posterior distributions covering wide regions in parameter space and the mean value of the posterior distributions lie in a different region of parameter space than the parameters of the reference simulation.

The shapes of the posterior distributions in Figures 5.4 and 5.6 have striking similarities. For many of the summary statistics, such as the 'Area correlation' or 'Cell elongation', the posterior distributions spread along a narrow, elongated region in parameter space. This indicates that cell shapes and neighbour relationships gen-

erated by the vertex model within such regions are similar and indistinguishable by the employed summary statistics. The reason for this might be that the parameters $\overline{\Lambda}$ and $\overline{\Gamma}$ have distinct but related roles in the vertex model. The parameter $\overline{\Lambda}$ regulates the strength of an energy contribution linear in the edge length, and $\overline{\Gamma}$ regulates the strength of an energy contribution quadratic in the perimeter. In this way, both parameters affect the overall contraction force along cell-cell interfaces, which might explain why stable configurations in the vertex model appear similar in regions of decreasing $\overline{\Lambda}$ and increasing $\overline{\Gamma}$. This relationship can be illustrated by considering the stability conditions for tissues containing single cells derived by Staple et al. (2010), who considered tissues containing single polygons of target area 1, area a, and perimeter p. Such polygons are energetically stable if

$$2a(a-1) + \overline{\Gamma}p^2 + \overline{\Lambda}p/2 = 0. \tag{5.14}$$

If a polygon with area a and perimeter p is stable for fixed values of $\overline{\Lambda}$ and $\overline{\Gamma}$, then there are infinitely many other parameter combinations for which equation (5.14) is fulfilled for the same values of a and p. These parameter combinations all lie along a line in $(\overline{\Lambda}, \overline{\Gamma})$ parameter space,

$$\overline{\Gamma} = -2a(a-1)/p^2 - \overline{\Lambda}/(2p). \tag{5.15}$$

The shapes of the posterior distributions in Figures 5.4 and 5.6 indicate that it might be possible to generalise this rule to energy minima of tissues containing multiple cells.

5.3.4 Parameter estimates for summary statistics from laser ablation experiments have high uncertainty

In Figures 5.4 and 5.6 we investigated the utility of summary statistics of cell packing to infer vertex model parameters. Using such image-based summary statistics we had limited success in inferring the mechanical parameters reliably, revealing a possible interdependence of the mechanical vertex model parameters. It is intuitive that such static image-based methods have limitations, since they use cell shapes as a means to infer mechanical properties, and one may assume that invasive methods that record the dynamics of tissue response to a manipulation may produce more reliable results for the inference of mechanical parameters.

To test whether localised mechanical manipulation can be helpful in inferring parameters we conduct laser ablation experiments, similar to those introduced in Chapter 4, where we used laser ablation experiments to distinguish several scenarios of mechanical regulation. For each sample simulation we randomly select three cell-cell interfaces in the final tissue configuration and conduct *in silico* laser ablations on these interfaces. To ensure that the laser ablations do not interfere with each other or with the tissue boundary, we pick cell interfaces that are separated by 60° angles along a circle sharing the tissue centre and with a radius of a quarter of the tissue width. To simulate laser ablation, we remove the line tension contribution in equation (2.4) from the ablated edge and the perimeter contribution from both of its adjacent cells, and further set the threshold for vertex rearrangement l'_{T1} to zero. The latter prevents vertex rearrangements in the cells adjacent to the ablation, which we consider biologically unrealistic since we expect recoil after laser ablation to happen on faster timescales than vertex rearrangement. It also helps us keep track of the length of the ablated edge, since it will not be involved in any vertex rearrangements. After each laser ablation, we simulate the tissue for another 150 time units, which is sufficiently long to allow the tissue to settle in a static equilibrium.

In Figure 5.7 we analyse the utility of typical summary statistics of such laser ablation experiments to measure vertex model parameters. Specifically, we investigate the recoil distance, the area asymmetry, and the perimeter asymmetry. Each of these summary statistics is averaged across all three laser ablations per parameter vector. The latter two asymmetry measures were used by Farhadifar et al. (2007) to constrain their parameter choices in manual parameter fitting. The area asymmetry measures how much the combined areas of the cells adjacent to the ablation change in comparison to changes in the length of the ablated edge. The perimeter asymmetry similarly considers changes in the combined cell perimeter of the two cells. Both of these measures aim to investigate whether the cell shape changes upon laser ablation are limited to an extension of the ablated edge, or whether the cells respond as a whole by changing their total areas or perimeters. Similar to our previous observations of summary statistics of cell packing, the posterior distributions shown in Figure 5.7 have mean values in different regions of parameter space than the reference parameter, and a high degree of uncertainty is associated with this parameter estimate. For example, the posterior estimated using the 'Area asymmetry' summary statistic has a mean value of $(\overline{\Lambda}, \overline{\Gamma}) = (-0.33, 0.11)$ and marginal standard deviations of 0.33 and 0.06 in $\overline{\Lambda}$ and $\overline{\Gamma}$, respectively.

Note that Farhadifar et al. (2007) restricted their parameter analysis to positive $\overline{\Lambda}$ values, since the vertex distances always increased after the ablation, rather than contracted, indicating that the edges are under tension. Here, of the 88479 parameter points for which we successfully conducted laser ablation experiments, only 982 lead to a contraction of vertex distances after cutting, illustrating that increasing vertex distances after ablation are common even when $\overline{\Lambda}$ is negative. Hence, we cannot confirm the assumption that positive vertex recoils dictate positive values of $\overline{\Lambda}$.

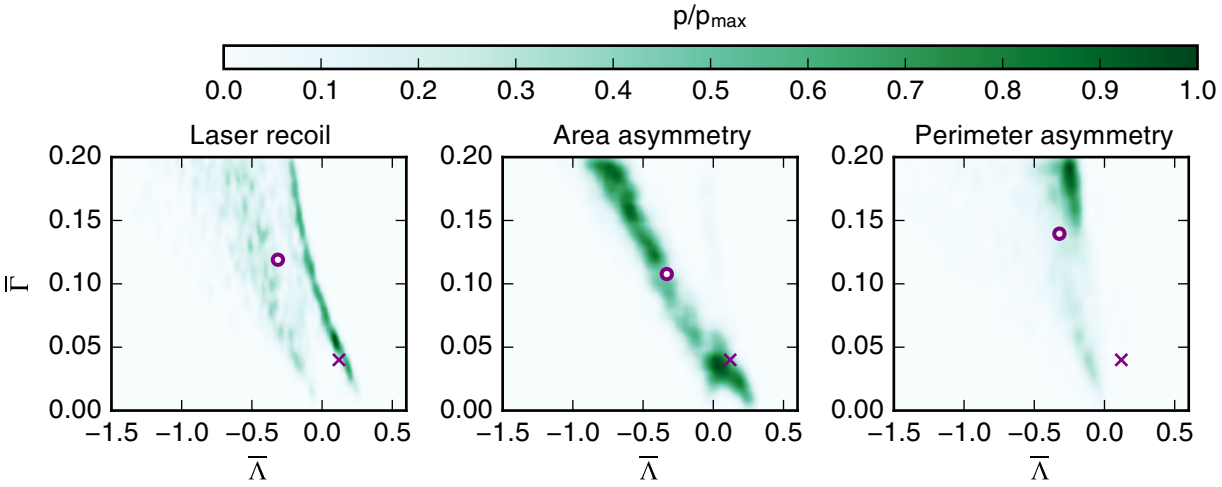

Figure 5.7: Using summary statistics associated with laser ablation experiments does not lead to improved parameter estimates. Posterior distributions estimated using summary statistics from laser ablation experiments are shown in green. For each parameter point, three laser ablations were conducted. Model and inference parameters are listed in Table 5.1 and mathematical descriptions of the summary statistics are provided in the text and listed in Table 5.2. Crosses mark the reference parameter of the data for which the posterior is estimated. Circles mark the means of the posterior distributions.

5.3.5 Selected combinations of summary statistics do not improve parameter estimates significantly

After attempting to infer the mechanical parameters of the vertex model using summary statistics of cell packing and from laser ablation experiments, we now seek to improve the parameter estimation by combining summary statistics. Summary statistics are combined by creating vector-valued summary statistics. If at least one of the summary statistics is already vector-valued, the vector is simply extended with entries for the other statistic. As described in Section 5.2, the individual entries of the combined, vector-valued summary statistics are rescaled by their respective standard deviations.

The posterior distributions from three such combinations of summary statistics are plotted in Figure 5.8. First, we combine the laser ablation summary statistics 'Laser recoil' and 'Area asymmetry' to test whether this combination leads to a posterior that

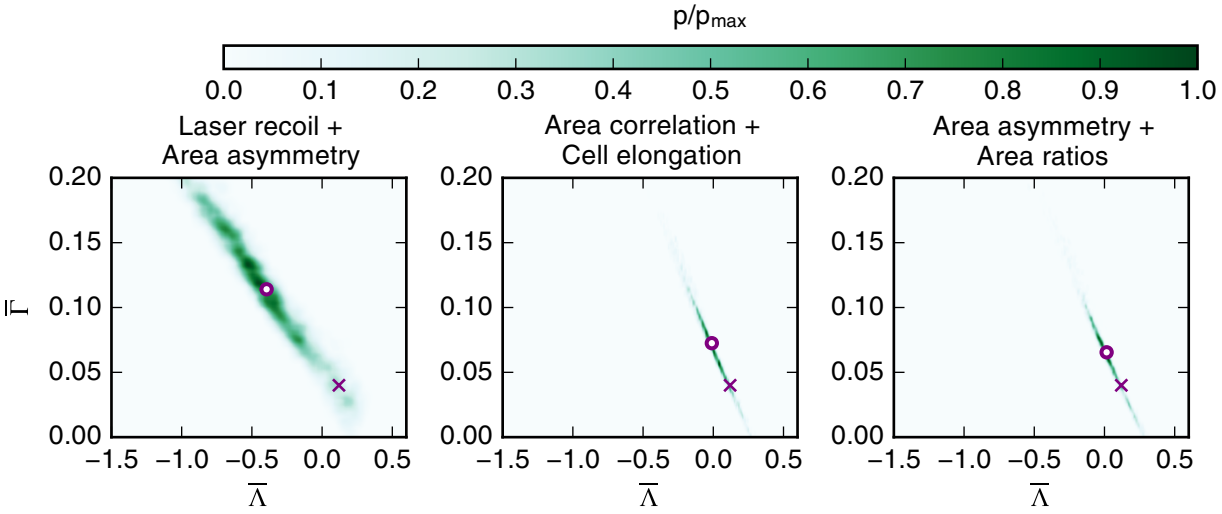

Figure 5.8: Combinations of summary statistics do not lead to improved parameter estimates. Posterior distributions obtained when combining summary statistics are shown in green. The method for combining summary statistics is explained in the text. Model and inference parameters are listed in Table 5.1 and mathematical descriptions of the summary statistics are provided in the text and listed in Table 5.2. Crosses mark the reference parameter of the data for which the posterior is estimated. Circles mark the means of the posterior distributions.

is concentrated around the region where both the individual posterior distributions have high values, which would be close to the reference parameter value. However, the posterior from the combined summary statistic is instead widespread, which leads to a parameter estimate in a different region in parameter space than the reference parameter, similar to the posteriors from either individual summary statistic.

Next, we combine the 'Area ratios' and 'Cell elongation' summary statistics to test whether the contributions from the 'Cell elongation' summary statistic may help in 'shifting' the parameter estimate of the 'Area ratios' statistic closer to the reference parameter value. However, the posterior distribution obtained from the combined summary statistic resembles the posterior from the 'Area ratios' statistic, with a marginal improvement of the $(\overline{\Lambda}, \overline{\Gamma})$ estimate from (-0.028, 0.076) to (0.01, 0.072).

Finally, we combine the 'Area ratios' summary statistic with the 'Area asymmetry'. This is motivated by (Farhadifar et al., 2007), who used summary statistics of cell packing in combination with laser ablation experiments to identify ver-

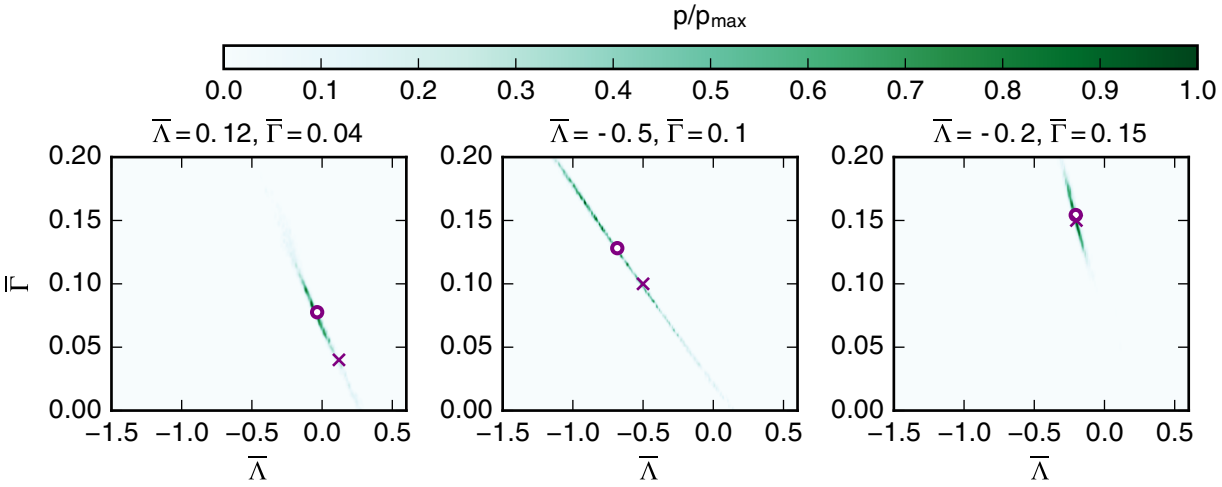

Figure 5.9: Characteristics of posterior distributions are preserved for different reference parameters. Posterior distributions of the 'Area ratios' summary statistic are shown in green. The method for combining summary statistics are explained in the text. Model and inference parameters are listed in Table 5.1 and mathematical descriptions of the summary statistics are provided in the text and listed in Table 5.2. Crosses mark the reference parameter of the data for which the posterior is estimated. Circles mark the means of the posterior distributions.

tex model parameters. This combination of summary statistic again leads to a marginally better parameter estimate than the 'Area ratios' summary statistic, with $(\overline{\Lambda}, \overline{\Gamma}) = (0.015, 0.065)$.

5.3.6 Characteristics of posterior distributions are preserved for different reference parameters

Among all the summary statistics we considered, one the most accurate parameter estimates is achieved when using the 'Area ratios' summary statistic. Combinations of this summary statistic with other summary statistics of cell shapes or from laser ablation experiments lead to marginal improvements on the parameter estimate. Note, however, that even for these summary statistics the parameter estimate does not coincide with the reference parameter. A benefit of using the 'Area ratios' summary statistic is that it does not require laser ablation experiments to be conducted, i.e. it

can be experimentally measured using non-invasive imaging methods.

We next test to what extent this 'Area ratios' summary statistic can be used to infer parameters of sample simulations with reference parameters elsewhere in parameter space, namely $(\overline{\Lambda}, \overline{\Gamma}) = (-0.5, 0.1)$ and $(\overline{\Lambda}, \overline{\Gamma}) = (-0.2, 0.15)$ (Figure 5.9). In the first case, the parameter estimate is similarly close to the reference value as for the previously analysed parameter point, but the posterior distribution is more widespread. In the second case, the reference and the target parameter practically coincide and the parameter uncertainty is similar to the parameter uncertainty at $(\overline{\Lambda}, \overline{\Gamma}) = (0.12, 0.04)$.

5.3.7 Parameter estimates depend on the observed data at the reference parameter

In Figures 5.4, 5.6, 5.7 and 5.9 the parameter estimates are associated with high parameter uncertainty for each of the observed summary statistics, and the estimated parameters generally do not coincide with the reference parameter set. One possible reason for the deviation of the parameter estimates from the reference parameter could be that the observed data at the reference parameter is an outlier that is not representative of the model behaviour at this parameter. In order to test how strongly the parameter estimate depends on the observed data we plot the parameter estimates obtained from ten different samples at the reference parameter $\overline{\Lambda}_{\text{obs}} = 0.12$, $\overline{\Gamma}_{\text{obs}} = 0.04$ in Figure 5.10, using the 'Area ratios' summary statistic. The estimated parameter varies for the different samples between $(\overline{\Lambda}, \overline{\Gamma}) = (-0.23, 0.12)$ and $(\overline{\Lambda}, \overline{\Gamma}) = (0.01, 0.06)$. Similar to the posterior distribution observed for the 'Area ratios' summary statistic, the estimated parameters appear to align along a line in parameter space. We conclude that the parameter estimate depends on the observed data at the reference parameter, which is not surprising considering the uncertainty associated with the posterior distributions obtained in this chapter. We further note

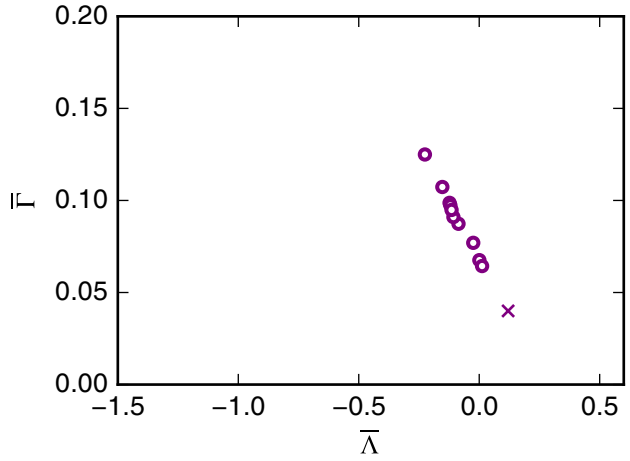

Figure 5.10: Parameter estimates depend on the observed data at the reference parameter. Circles mark the parameter estimates obtained using the 'Area ratios' summary statistic evaluated for ten different realisations of the model at the reference parameter set $\overline{\Lambda}_{\text{obs}} = 0.12$, $\overline{\Gamma}_{\text{obs}} = 0.04$. The reference parameter is marked by a cross. Model and inference parameters are listed in Table 5.1 and mathematical descriptions of the summary statistics are provided in the text and listed in Table 5.2.

that the reference sample used throughout this chapter is not an outlier in Figure 5.10, with a parameter estimate of $(\overline{\Lambda}, \overline{\Gamma}) = (-0.03, 0.08)$, obtained using the 'Area ratios' summary statistic . In this sense, the sample is representative of the model behaviour at the reference parameter. We highlight that none of the estimated parameters in Figure 5.10 coincides with the reference parameter.

5.4 Discussion

In this chapter, we have investigated the utility of a range of summary statistics in identifying the mechanical parameters of the vertex model. For the first time, we provide uncertainty estimates arising from the use of such summary statistics by using ABC. We find that neither first-order summary statistics of cell shapes, nor second-order summary statistics of cell shapes, nor summary statistics from laser ablation experiments, nor combinations of summary statistics are fully sufficient to

reliably infer vertex model parameters. Among all analysed summary statistics, we identified the summary statistic 'Area ratios' as the most suitable summary statistic for measuring vertex model parameters when using non-invasive methods, i.e. the average area of cells for each polygon class. Marginal improvements on parameter estimates from this summary statistic were achieved for a sample at the reference parameter when combining the 'Area ratios' statistic with other measures of cell shape, or summary statistics from laser ablation experiments. We identified that posterior estimates are robust to changes in the total number of samples and the acceptance ratio, and that characteristics of the posterior distribution are preserved if different reference parameters are used. Further, we confirmed that the parameter estimate may depend on the observed data at the reference parameter.

Summary statistics of cell shape have previously been used to constrain vertex model parameter space by Farhadifar et al. (2007), who found that large regions of parameter space could give rise to experimentally observed values for such summary statistics. Despite this finding, it has been a common procedure to only use, for example, the distribution of cell neighbour numbers to select parameter values. Other summary statistics of cell shape have since been proposed, for example the cell elongation (or circularity) (Nestor-Bergmann et al., 2016). Here, we confirm previous findings by Farhadifar et al. (2007) that high uncertainty is associated with classical summary statistics of cell packing, as well as with alternative descriptors of cell shape, such as cell elongation or perimeter length.

We further analysed, for the first time, whether second-order statistics of cell shapes, such as the correlation between areas of adjacent cells, can lead to improved parameter estimates. However, these second-order statistics suffer a similar drawback to first-order statistics, namely a spread of the posterior distribution along an approximately linear line in parameter space, and similar posterior distributions were observed for summary statistics from laser ablation experiments.

We further analysed possible combinations of summary statistics and showed that even combinations cannot improve the parameter estimates considerably and established that the main characteristics of the posterior distributions are preserved when reference parameters in different regions in parameter space are inferred. We further identified that the estimated parameter may depend on the observed data.

All of the reported summary statistics are associated with a significant amount of parameter uncertainty. There are two possible reasons for the emergence of this parameter uncertainty. First, the mechanical vertex model parameters might not be identifiable, i.e. the same model behaviour might be achieved by multiple parameter combinations. Second, the analysed tissues and numbers of laser ablations might be too small to lead to accurate values of the summary statistics. In future efforts, it may be possible to reduce the parameter uncertainty by conducting multiple 'experiments' per parameter point, i.e. measuring averages from more than three laser ablations or from multiple tissues. Instead of simulating multiple tissues, it may also be possible to gain estimates of cell packing from single, larger tissues. If the model parameters can be shown to be unidentifiable, it may be possible to design altered versions of the vertex model that depend on a smaller number of parameters.

That summary statistics obtained from laser ablation experiments do not lead to a concise parameter estimate is surprising, since such summary statistics have previously been used to significantly constrain possible vertex model parameters (Farhadifar et al., 2007). Farhadifar et al. (2007) excluded negative $\overline{\Lambda}$ from their analysis with the argument that negative values of the line tension would require vertex distances to contract instead of extend after laser cutting. Here, we confirmed that vertices tend to recoil even if $\overline{\Lambda}$ is negative, possibly due to the contractile nature of cell perimeters of adjacent cells.

When simulating laser ablation experiments, we set the line tension of the ablated edge, as well as the perimeter contractility of the adjacent cells, to zero. This

procedure has originally been proposed by Farhadifar et al. (2007). Canela-Xandri et al. (2011) set the line tension of the ablated edge to zero as well, but remove the perimeter contractility contribution only from the ablated edge and not the remaining edges of the adjacent cells. More research is required to determine which method to implement laser ablations *in silico* is the most biologically realistic.

Laser ablations are a common procedure to measure the mechanical properties of a tissue *in vivo*. Other ways of tissue manipulation exist, for example through stretching or compressing the tissue (Merzouki et al., 2016). Future work may include applying ABC to this type of experiment and measure the uncertainty of the associated parameter estimation. Experimentally, a possible avenue for future work might be to develop corresponding methods of tissue manipulation that are applicable *in vivo*, for example through locally perturbed tissue growth.

Here, we analysed the packing geometry and response to perturbations of tissues that emerge after periods of uniform growth, such as in the *Drosophila* wing disc. However, packing geometries emerge from from a wide range of epithelia under different conditions. Future work may investigate whether other ways of generating tissue packing lead to similar packing geometries, and indeed if the same summary statistics, or even the same simulation procedure presented here, may be used to measure vertex model parameters in different tissues.

Throughout this chapter, we use rejection-based ABC and identified the mean area per polygon class as the most powerful summary statistic to constrain vertex model parameters. Using this summary statistic it may be possible to apply adaptations of ABC in order to implement more efficient ABC methods that have lower sample rejection rates, for example SMC-ABC (Sisson et al., 2007), MCMC-ABC (Marjoram et al., 2003), or AABC (Buzbas and Rosenberg, 2015). Specifically, the simulation procedure used in this chapter is designed to minimise the time that is required to generate the tissue. Using this simulation procedure, the tissue does not arrive at its final

configuration through slow, quasistatic growth. Instead, tissue growth occurs within a dynamic regime of the model. We confirmed that summary statistics of cell packing are similar to what has previously been reported. Using alternative inference methods, such as SMC-ABC or AABC, it may be possible to use quasi-static timescales in the sample simulations. When using a biologically realistic implementation of the growth phase of the tissue it may be possible to improve the parameter estimates obtained here by using dynamic data of tissue growth throughout the simulation, instead of focussing only at the final configuration.

Throughout this chapter, we rescaled components of vector-valued summary statistics by their respective standard deviations. This is a common procedure when conducting parameter inference, since it minimises the impact of summary statistics of high variability on the parameter estimate (Beaumont et al., 2002; Beaumont, 2010). However, the optimal choice for weights when combining summary statistics for parameter estimation is a matter of active research (Harrison and Baker, 2017; Prangle, 2017; Jung and Marjoram, 2011). Here, we considered a large number of summary statistics and combinations, and concluded that in each case the parameter estimates are associated with high uncertainty in the posterior distributions. It is unclear whether better parameter estimates could be achieved in our work if different weights were used in vector-valued summary statistics.

Future methods of investigating vertex model parameters may include measurements of the mechanical relaxation time, which has previously been proposed by Canela-Xandri et al. (2011) and Mao et al. (2011). This may, in turn, allow the precise implementation of cell cycle times in the vertex model.

Here, we have investigated parameter inference for vertex models propagated using the energy equation (2.4). However, different energy equations have been proposed, for example by Nagai and Honda (2001), Weliky and Oster (1990), Hufnagel et al. (2007) and Smith et al. (2012). Common alterations to equation (2.4) include terms that take

the height of the cells into account, or omit individual terms, such as the perimeter contribution. In order to conduct parameter inference on vertex models using these altered energy equations, it would be necessary to conduct a preliminary analysis investigating the different possible model behaviours, and the parameter regions where the models are physical. Future work may include using ABC for model selection on experimental data to investigate which of these energy equations provides the most plausible description of biological phenomena. Such work could, for example, include alterations to equation (2.4) that prevent mechanically-induced cell removal.

Having investigated methods to infer vertex model parameters from static snapshots of epithelial tissues, we next turn to cell tracking as a method to analyse dynamic changes in epithelia in Chapter 6.

Chapter 6

Cell tracking in epithelia

In Chapters 3, 4 and 5 we have focussed on the design of quantitative cell-based models for epithelia, on how to apply these models to make experimental predictions, and on how to compare cell-based models to experiments. To fully unveil the power of cell-based models in understanding biological systems it is essential to combine them with experimental data. However, the extraction of quantitative data from, for example, live-imaging microscopy videos is challenging. In this chapter, we illustrate how mathematical concepts can contribute to quantitative data analysis in biology. We develop a cell-tracking algorithm that applies graph theoretical concepts to cell tracking in live-imaging datasets of planar, approximately two-dimensional epithelia. The contents of this chapter have been published in the *Journal of the Royal Society Interface* (Kursawe et al., 2016b).

6.1 Background and motivation

Live-imaging microscopy is a powerful, and increasingly quantitative, tool for gaining insight into fundamental processes during embryonic development (Stephens and Allan, 2003; Pantazis and Supatto, 2014; Truong and Supatto, 2011). Quantitative information on cell growth, proliferation, death, shape changes and movement ex-

tracted from live-imaging experiments reveals how such processes are regulated to give correct tissue-level behaviour. This approach has been particularly successful in characterizing the growth and patterning of embryonic epithelial tissues in a number of model organisms (Mao et al., 2011; Gibson et al., 2006; Rauzi et al., 2008; Collinet et al., 2015; Ritsma et al., 2014; Parker, 2006).

A common experimental technique for visualising cell shapes in an epithelial sheet is to fluorescently tag a molecule marking cell boundaries, such as E-cadherin (Figure 6.1A). The analysis of time-lapse microscopy data obtained from such tissues is extremely challenging (Pantazis and Supatto, 2014; Truong and Supatto, 2011), especially in cases of imaging data of rapidly evolving tissues, and when limitations of, for example, microscope speed, imaging resolution or phototoxicity prohibit the creation of datasets with high temporal and spatial resolution.

The analysis of time-lapse microscopy data comprises two major steps: segmentation and tracking (registration). Segmentation must be performed for each frame of a video and involves the identification of objects and landmarks, such as cell shapes (Figure 6.1B). Automated segmentation is hindered by various factors such as noise in fluorescent signals, uneven illumination of the sample, or overlapping cells in a two-dimensional projection. Often, manual correction is necessary to address over-segmentation, where too many cells are detected, or under-segmentation, where too few cells are detected (Mashburn et al., 2012; Cilla et al., 2015; Schiegg et al., 2013). Tracking involves the association of segmented cells across video frames (Figure 6.1C) and requires resolving cell movement, cell division, cell death, and cells entering and leaving the field of view (Schiegg et al., 2013).

Numerous algorithms are available for the segmentation and tracking of cellular-resolution microscopy data (Mashburn et al., 2012; Cilla et al., 2015; Heller et al., 2016). Common methods for cell tracking utilize optimization techniques to minimise differences in cellular properties between two frames (Padfield et al., 2011; Cilla et al.,

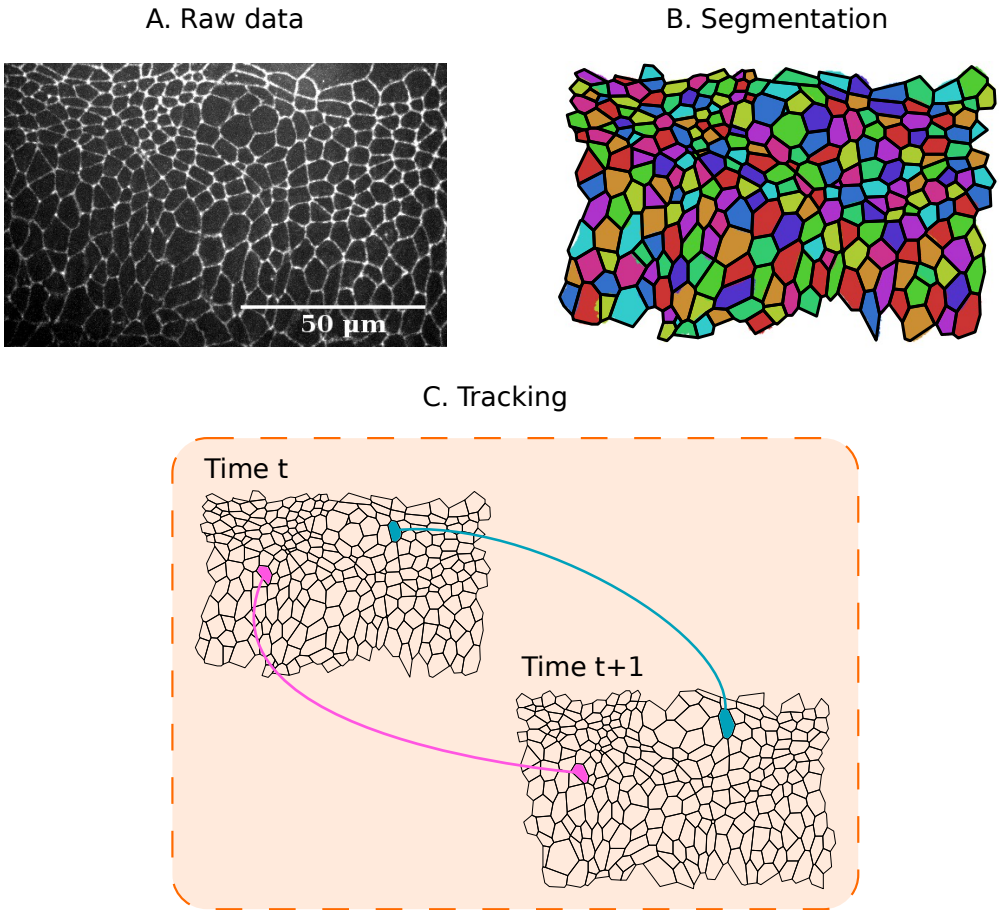

Figure 6.1: Pipeline for analysing epithelial tissues. (A) Example raw data. Frame of a live-imaging microscopy video of the lateral epidermis of a stage-11 *Drosophila* embryo, expressing DE-Cadherin::GFP. (B) Segmentation of this image, showing cell shapes (coloured regions) and polygonal approximation based on three-cell junctions (black lines). (C) Cell tracking involves registering individual cells across consecutive segmented images. Details of the experimental system and segmentation process are given in Section 6.2.

2015; Youssef et al., 2011; Wait et al., 2014; Winter et al., 2011). The min-cost max-flow algorithm (Padfield et al., 2011) uses linear integer programming to minimise differences in cell areas, perimeters, orientations, and locations between frames, whereas multiple-parameter tracking (Youssef et al., 2011) employs global optimization to minimize differences in cell shapes as well as locations. In contrast, multitemporal association tracking (Wait et al., 2014; Winter et al., 2011) minimises differences in cell locations and sizes by using a probabilistic approach that finds the most likely extension to existing cell trajectories. Chain-graph models (Sommer et al., 2011) minimise

differences in cell velocity while overcoming mis-segmentation by verifying that each segmented object continues or begins a cell trajectory in successive frames. Optical flow ('warping') between successive frames can be used to guide cell tracking as well as segmentation (Liu et al., 2014). It is also possible to combine segmentation and tracking of two-dimensional microscopy videos by interpreting time as a third spatial dimension and employing three-dimensional segmentation techniques (Bellaïche et al., 2011), or by fitting a deformable graph model to the microscopy data (Zou and Tomasi, 2016), where the deformable graph represents the dynamically changing network of cell-cell interfaces in epithelia. The nearest-neighbour method associates two cells in consecutive frames with each other if their respective centroids have minimal distance within the field of view (Mashburn et al., 2012), or if their overlap in pixels within the field of view is maximal (Aly et al., 2014; Wang et al., 2010). Particle image velocimetry, a technique originally developed to analyse fluid flow (Raffel et al., 2007), has also been employed to track cells in epithelial tissues (Puliafito et al., 2012).

Software implementations and computational tools for cell tracking include FARSIGHT (Al-Kofahi et al., 2010) (segmentation only), SeedWaterSegmenter (Mashburn et al., 2012) (nearest-neighbour tracking), ilastik (Sommer et al., 2011) (chain-graph models), Tufts Tissue Tracker (Cilla et al., 2015) (min-cost max-flow algorithm), Tracking with Gaussian Mixture Models (Amat et al., 2014) (nearest-neighbour tracking), Packing Analyzer (Aigouy et al., 2010) (particle image velocimetry) and EpiTools (Heller et al., 2016) (nearest-neighbour tracking). These algorithms and software tools primarily rely on there being small differences in cell positions and shapes across consecutive images. Their performance is therefore hindered when analysing data from *in vivo* studies where phototoxicity provides a barrier to high temporal resolution imaging (Hoebe et al., 2007; Wood and Jacinto, 2005; Mavrakis et al., 2008). To address this limitation, we propose a novel algorithm for cell tracking that uses only the connectivity of cell apical surfaces (Figure 6.1). By representing the cell sheet

as a physical network in which each pair of adjacent cells shares an edge, we show that cells can be tracked between successive frames by finding the *maximum common subgraph* (MCS) of the two networks: the largest network of connected cells that is contained in these two consecutive frames. It is then possible to track any remaining cells based on their adjacency to cells tracked using the MCS. Our algorithm does not require the tuning of parameters to a specific application, and scales in subquadratic time with the number of cells in the sheet, making it amenable to the analysis of large tissues.

We demonstrate here that our algorithm resolves tissue movements, cell neighbour exchanges, cell division, and cell removal (for example, by delamination, extrusion, or death) in a large number of *in silico* datasets, and successfully tracks cells across sample segmented frames from *in vivo* microscopy data of a stage-11 *Drosophila* embryo. We further show how our algorithm may be used to gain insight into tissue homeostasis by measuring, for example, the rate of cell rearrangement in the tissue. In particular, we find a large amount of cell rearrangement within the observed dataset despite the absence of gross morphogenetic movement. The remainder of the chapter is structured as follows. In Section 6.2 we describe the algorithm for cell tracking. In Section 6.3 we analyse the performance of the algorithm on *in silico* and *in vivo* datasets. Finally, in Section 6.4 we discuss future extensions and potential applications.

6.2 Methods: Design of the cell tracking algorithm

Here we provide a conceptual overview of the core principles underlying our cell tracking algorithm. We focus on providing an accessible, non-technical description rather than including all details required to implement the algorithm from scratch. A comprehensive mathematical description of the algorithm is provided in Appendix A.

The input to the algorithm is a set of segmented images obtained from a live-

imaging microscopy dataset of the apical surface of an epithelial cell sheet. For each image, the segmentation is assumed to have correctly identified which cells are adjacent and the locations of junctions where three or more cells meet. Various publicly available segmentation tools can be used for this segmentation step, for example SeedWaterSegmenter (Mashburn et al., 2012) or ilastik (Sommer et al., 2011). The segmentation is used to generate a polygonal approximation to the cell tessellation (Figure 6.1B-C). Such approximations are an adequate assumption for many epithelia (Farhadifar et al., 2007; Cilla et al., 2015; Escudero et al., 2011; Sánchez-Gutiérrez et al., 2016; Sáez et al., 2013).

Our algorithm tracks cells by interpreting the polygonal representations arising from the segmentation as networks ('graphs') of cells. Examples of such networks are shown in Figure 6.2A. In this representation, each cell corresponds to a vertex of the network, and two vertices are connected by an edge if the corresponding cells are adjacent. Our algorithm tracks cells across consecutive images by aligning the networks of cells that correspond to these images. This network alignment is achieved in three steps. First, we generate an initial tracking for subsets of the cells in each pair of consecutive images by finding the MCS between the two corresponding networks (Figure 6.2B). Second, this MCS is reduced to avoid tracking errors (Figure 6.2C). Third, remaining untracked cells are tracked based on their adjacency to cells within the MCS (Figure 6.2D). In the final output of the algorithm, each tracked cell of a frame is paired with exactly one cell in the subsequent frame.

The key step in this network alignment approach is the identification of a MCS (Ullmann, 1976; Krissinel and Henrick, 2004). A MCS comprises the largest subnetwork that is contained in two larger networks; thus finding an MCS can be understood as recognising patterns of connections that are preserved between two networks. In the present work, the structure of the MCS roughly corresponds to cells that do not rearrange between consecutive images, except for a few cells at its boundaries.

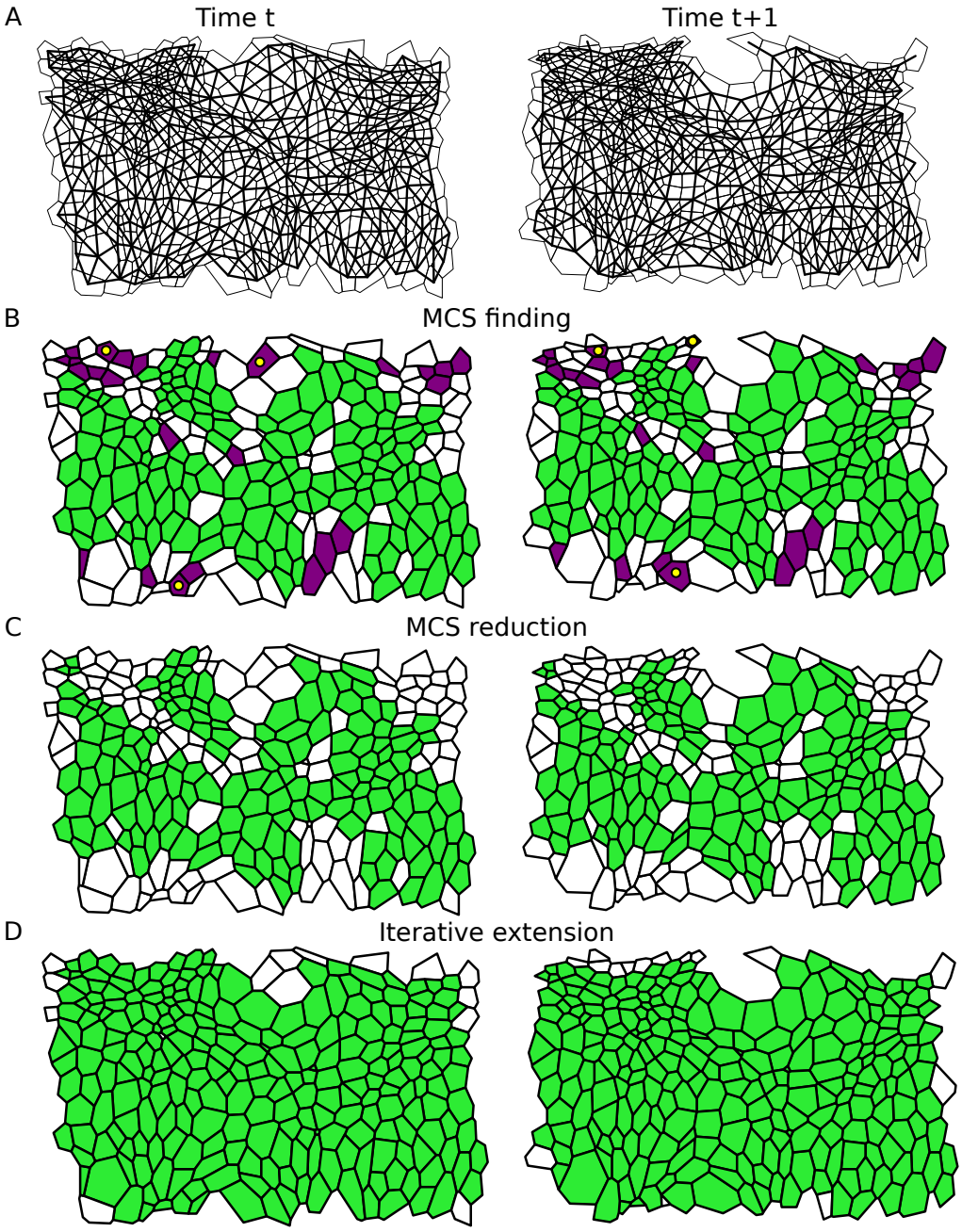

Figure 6.2: Illustration of our cell tracking algorithm. (A) Grey: Two consecutive segmented time-lapse images (left and right columns) of the lateral epidermis of a stage-11 *Drosophila* embryo, taken five minutes apart. See Section 6.2 for details. There are several cell neighbour exchanges between these images. Black: Overlay of the network of cells that the algorithm uses for cell tracking. Cells in the tessellation correspond to network vertices that are connected by an edge if the cells are adjacent. (B) We first identify a cell mapping between the two graphs based on the MCS. This includes correctly tracked (green/light) cells and cells that have only few tracked neighbours (purple/dark). Here, the MCS incorrectly tracks three cells (yellow/light dots). (C) Weakly connected cells and small isolated clusters of cells are removed from the MCS to prevent mismatches. (D) An extended tracking mapping is constructed, which includes the maximum possible number of cells. See Section 6.2 for details. The remaining white cells have entered or left frame of view between images and therefore are not tracked.

In Figure 6.2B, we visualize the MCS generated by our algorithm as a collection of green (light) and purple (dark) cells. Most of the highlighted cells in Figure 6.2B are tracked correctly by the MCS. Three cells in each frame are marked by a yellow (bright) dot. Within the two cell networks, these cells are members of the MCS. However, these cells are not tracked correctly by the MCS. This mismatch arises since the MCS is based on the connectivity of cells within the network alone. The fewer connections a cell has to other cells in the MCS, the less information about the cell's position and shape is encoded by these network connections, and so the greater the possibility of mismatches. To avoid such tracking errors, we remove any cells that have only a few connections within the MCS, as well as small isolated clusters of cells. All cells that are removed from the tracking in the second step of our algorithm are shown in purple (dark) in Figure 6.2B. In Figure 6.2C we highlight the cells that are tracked after applying this second step of our algorithm.

Cells that are untracked after reducing the MCS are then tracked based on their connections to previously tracked cells. This last step of our algorithm comprises starting from the MCS and iteratively 'growing' the set of tracked cells by adding cell-to-cell matches to the tracking that maximise the number of preserved connections to other tracked cells. In this step, the algorithm also resolves cell neighbour exchanges, cell removals, and cell divisions by identifying changes in the network structure that are characteristic of these events. For example, a cell neighbour exchange corresponds to the deletion of a network connection while a new connection is added.

6.2.1 The MCS is identified through repeated seeding and iterative extension

In computer science, MCS finding has been known to be an NP-hard problem (Ullmann, 1976; Krissinel and Henrick, 2004): the time to find an exact MCS of two networks increases exponentially with the size of the networks, which poses a compu-

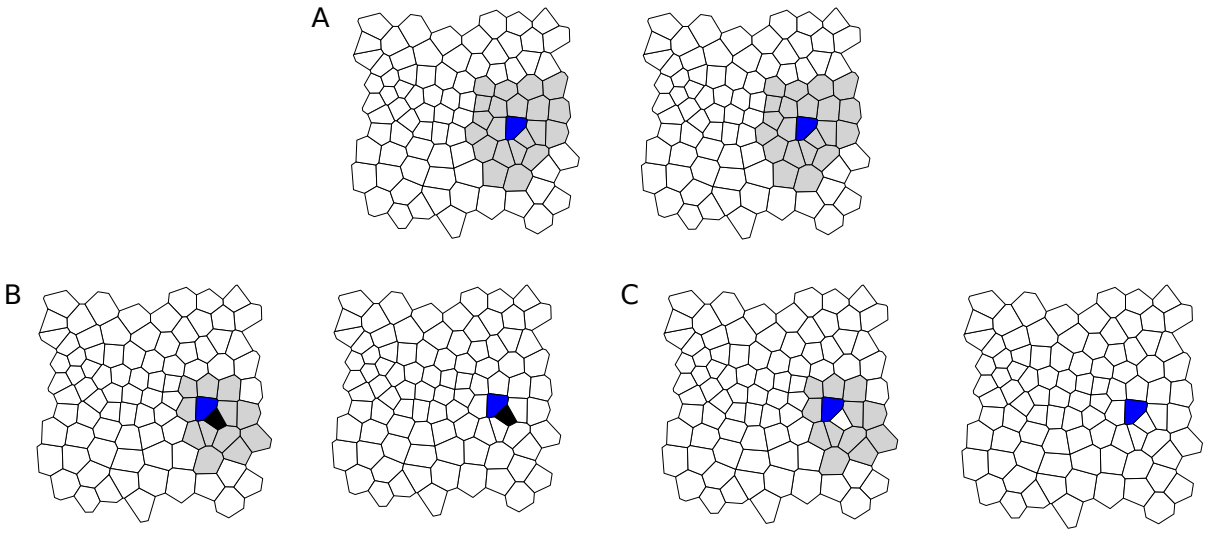

Figure 6.3: Construction of the MCS. (A) The algorithm picks a first match of cells for the MCS (blue) if their neighbourhoods form identical networks. The considered neighbourhood (grey) includes all neighbours and second nearest neighbours. (B-C) Additional cells are added to the MCS iteratively by inspecting the MCS between the grey area on the left, and the white area on the right. In (B), where the black cell is paired correctly, the local MCS is larger than in (C), where the selected cell is not considered for mapping. Hence, the pairing of black cells is added to the MCS.

tational barrier to the use of MCS-finding algorithms in applications. We overcome this computational barrier in the present work by constructing the MCS iteratively from the MCSs of smaller subgraphs, exploiting the planar structure of our cell networks to reduce the complexity of the problem.

To start the construction of the MCS, the algorithm identifies a match between two cells in the consecutive images for which the structure of the network of their surrounding cells is identical. Here, the network of surrounding cells is restricted to the network formed by a cell's neighbours and its second nearest neighbours (see Figure 6.3A). If no such initial match can be found, the algorithm instead searches for an initial match where only the first-order neighbourhood is preserved, under the condition that this neighbourhood does not touch the boundary of the tissue. This latter condition avoids tracking errors that can occur on the tissue boundary where cells have few neighbours.

Once the initial match (a 'seed') is found, the algorithm iteratively adds further cells to the MCS. At each step of this iteration, a cell in the first network is picked that is adjacent to the existing MCS and has a minimal number of potential matches. This number of potential matches is determined based on how many cells in the second network have the same number of neighbours as the considered cell while preserving connections to already tracked cells. Among the choice of potential matches the algorithm identifies an optimal match based on the local network structure of these cells' neighbours. A cell in the second network is identified as an optimal match if the network structure of its neighbourhood is most similar to the cell in the first match. This choice is made based on local MCSs between the neighbourhood-networks of the cell in the first image and each potential match. Note that the optimal choice may exclude the cell from the tracking entirely. In this case, most neighbours are included in the local MCS when the considered cell is not tracked, indicating for example a cell removal event. In this case the cell is not mapped and the algorithm proceeds by inspecting another cell in the first match. Cells in the first frame for which no match in the second frame has been found may be re-inspected at later stages of the algorithm as the size of the identified MCS increases. Once no more adjacent cells can be added to the MCS through this iterative extension, the iteration continues the search among untracked cells in the first network that are not adjacent to the existing MCS. As soon as at least one cell has been added to the MCS in this way, the algorithm again restricts its search to adjacent cells. The algorithm halts once no further cell-to-cell matches can be found. During the construction of the MCS the algorithm ignores any potential cell-to-cell matches where the corresponding cell centroids are more than a cutoff distance $d_{\max}$ apart within the field of view. Throughout this chapter, we choose $d_{\max}$ to be ten average cell lengths.

Once the MCS is complete, any cells that have less than three isolated connections to other cells in the MCS are removed from the tracking. Any clusters of ten or fewer

cells are also removed from the tracking result. Both of these steps help to minimise tracking errors (Figure 6.2B-C).

6.2.2 Cells are added to the tracking result by inspecting connections to previously tracked neighbours

Through the identification of the MCS, the algorithm tracks most of the cells that do not rearrange between consecutive frames. Next, the algorithm tracks any remaining cells, and identifies cell rearrangements, cell removal, and cell division events. Similar to the construction of the MCS, the tracking of remaining cells is iterative. At each iteration, the algorithm identifies a cell-to-cell match that maximises the number of connections to already tracked cells, thus 'growing' the set of tracked cells from the intermediate tracking result of the MCS. When adding cells to the tracking, the algorithm ensures that a cell cannot gain more tracked neighbours between consecutive frames than the number of tracked neighbours preserved between these frames. The algorithm also requires a cell to have at least two tracked neighbours in order to be added to the tracking in this way.

Once all possible cells have been tracked, the algorithm resolves division events. Division locations can be identified as regions in the second frame that contain more cells than the corresponding region in the first frame. Since the algorithm will have found exactly one match in the second network for each tracked cell in the first network and *vice versa*, there are thus untracked cells in the second frame wherever a cell divides between two consecutive frames. The algorithm attempts to resolve division events by identifying changes in cell-to-cell connectivity that are characteristic of dividing cells (Figure 6.4). For example, two cells adjacent to each division must gain a neighbour (grey cells in Figure 6.4A), and in many cases the mother and daughter cells are easily identified as the cells that are shared neighbours of these cells adjacent to the division event. However, one of the daughter cells may be four- or three-sided

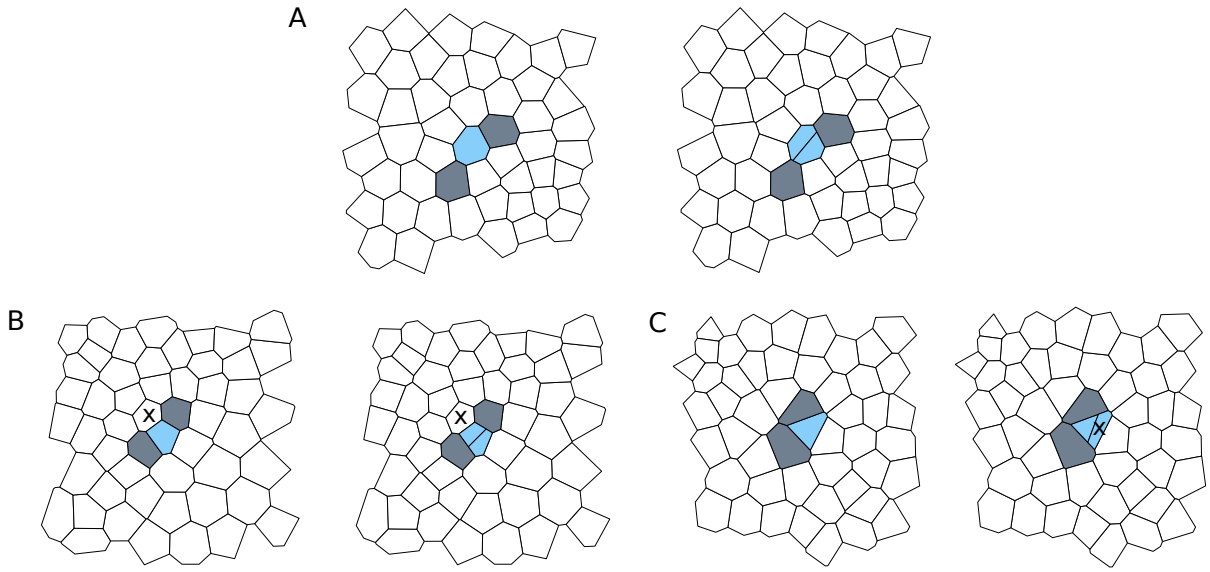

Figure 6.4: Resolving division events. Dividing cells are coloured blue. (A) Division events are resolved by identifying cells that gain an edge between the time frames (grey). The dividing cell and the daughter cells are shared neighbours of such cells. (B) When one of the daughter cells is four-sided, two mother cells are possible, the blue marked mother cell, and the cell marked by an 'x'. (C) If one of the daughter cells is three-sided, the mother cell can be mistaken for having gained an edge if it is identified with the daughter cell labelled 'x'. Our algorithm correctly resolves each of the types of division events shown in (A)-(C).

(Figure 6.4B-C). In these cases, the algorithm is not able to determine the mother- and daughter cells based on their network properties alone. Instead, the algorithm takes the geometric shape of the cells into account. The mother and daughter cells are chosen by identifying which pair of potential daughter cells has the closest position to their potential mother cell.

Cell deaths are identified as cells in the first frame that do not have a tracked match in the second frame and that are not on the boundary of the region of tracked cells.

6.2.3 Code availability

The code used in this chapter is publicly available under the 3-clause BSD license as the MCSTracker project (`https://github.com/kursawe/MCSTracker`). The project

is implemented in pure Python, employs unit testing (Osborne et al., 2014) and is fully documented. Graphs in our code are represented using the Python package NetworkX (Hagberg et al., 2008).

6.2.4 Generation of *in silico* datasets

To test the algorithm, we generate *in silico* datasets that include examples of cell divisions, removals and neighbour exchanges, as well as tissue movement. These datasets are generated using Voronoi tessellations modified using Lloyd's relaxation, which has previously been described in Chapter 4. Datasets generated with this method resemble cell packings in a variety of epithelial tissues (Sánchez-Gutiérrez et al., 2016; Honda, 1978).

To generate polygonal patterns of size $m \times n$, where m and n are natural numbers, $(m + g) \times (n + g)$ Voronoi seeds are distributed uniformly at random in a two-dimensional domain Ω of width $m + g$ and height $n + g$ (Figure 6.5A). Here, g denotes the size of a boundary region that is introduced to reduce the impact of the Voronoi boundary on the patterns. The tissue dimension of $m \times n$ arbitrary length units is chosen for convenience such that cells occupy one area unit on average. The domain Ω is surrounded by two additional rows of evenly spaced seeds on each side. The inner row is a distance of 0.5 length units to Ω, and the seed-spacing is 1.0. The outer row has a distance of 1.5 to Ω, and the seeds are shifted parallel to the first row by a distance of 0.5. The Voronoi tessellation of all these seeds is then constructed.

In each Lloyd's relaxation step, the polygons (or infinitely large areas) corresponding to the regularly spaced seeds outside Ω are removed from the tessellation. Next, the centroid of each remaining polygon is calculated and registered as a new seed. Further seeds are added that again correspond to two rows of evenly spaced seeds outside Ω. A new Voronoi tessellation is then constructed (Figure 6.5B). This procedure is repeated for n_L relaxation steps, after which all generated polygons are

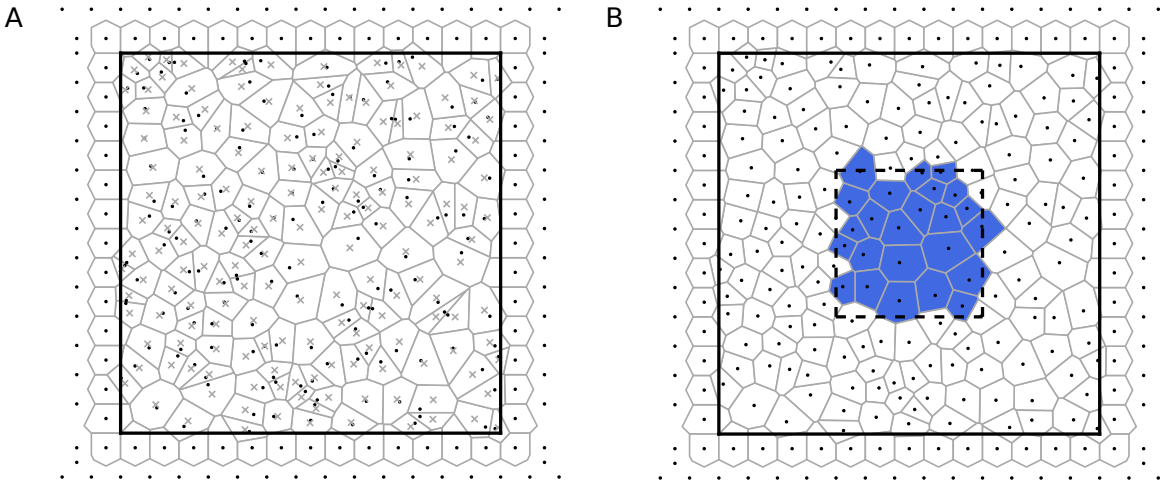

Figure 6.5: Generation of *in silico* data. (A) Random seeds (black dots) are placed inside a domain Ω (black line). Additional seeds are placed outside Ω. The Voronoi tessellation of all seeds is shown in grey, excluding Voronoi regions corresponding to the outermost row of seeds, since these are large or unbounded. The centroids of the Voronoi regions (grey crosses) differ from the seeds. (B) The centroids of the Voronoi regions in (A) are used as seeds for a new Voronoi tessellation, for which evenly spaced seeds are again added outside the domain Ω. Voronoi regions whose centroids lie within a central rectangle (dashed black line) are collected to form the *in silico* tissue (blue). In this figure, one such Lloyd's relaxation step is shown. Throughout this chapter, we generate *in silico* tissues using four Lloyd's relaxation steps.

discarded except those whose centroids lie within a rectangular domain of size $n \times m$ area units whose centroid coincides with that of Ω (Figure 6.5B).

The polygonal tessellations have approximately $m \times n$ polygons of average area 1.0. During the generation of the tessellations, evenly spaced seeds outside Ω are added to prevent the occurrence of infinitely large polygons inside Ω. The boundary of size g is added in between the generated tessellation and the evenly spaced seeds to reduce the effect of the evenly spaced boundary seeds on the tessellation. Throughout this chapter, we use $g = 8$ and $n_L = 4$, resulting in cell packings similar to those observed, for example, in the *Drosophila* wing imaginal disc (Sánchez-Gutiérrez et al., 2016). We provide further details of how tissue rearrangements are implemented in the Section 6.3.

6.2.5 Experimental methods

Experiments were conducted by Cody Narciso and Pavel A. Brodskiy in the laboratory of Jeremiah J. Zartman at the University of Notre Dame, Indiana. Live-imaging of cell proliferation was performed in stage-11 *Drosophila* embryos expressing a tagged version of DE-Cadherin (DE-Cadherin::GFP) using a spinning disc confocal microscope, as described by Narciso et al. (2015). For the embryo setup, a modified version of the standard live-imaging protocol was used (Parton et al., 2010).

6.2.6 Data segmentation

Microscopy images were segmented using pixel classification in ilastik (Sommer et al., 2011). The classifier was trained to recognise cell outlines and the segmentation of each frame was manually corrected. A watershed algorithm was used to identify the precise shape of the cell outlines. Each segmented frame was converted to a 16-bit grayscale image where pixels belonging to different cells had different integer values. Polygonal tessellations for the tracking algorithm were generated from the segmented image in two steps. First, all junctions between three or more cells were identified as points where pixels of three or more different cells met; second, vertices were assigned to cells. Then, edges shorter than two pixels (0.5 µm) were removed and replaced by a single vertex at the midpoint of the edge. Finally, polygons at the boundary of the tissue were removed from the simulation. This removal was necessary since cell shapes at the tissue boundary are poorly approximated by polygons due to missing vertices. Note that our algorithm can interpret segmentations saved using either ilastik (Sommer et al., 2011) or SeedWaterSegmenter (Mashburn et al., 2012).

6.3 Results: Algorithm performance

6.3.1 *In silico* testing of the algorithm

To assess the performance of the algorithm, we begin by applying it to *in silico* datasets that include cell neighbour exchanges, tissue movement, cell removal and cell division, respectively. In each case, we compare the outcome of the tracking algorithm to the ground truth.

We begin by assessing the ability of the algorithm to resolve permutations in otherwise identical tissues (Figure 6.6A). In this test, a random tessellation of size nine by nine cells is created as described in Section 6.2, and integer identifiers c_i are assigned to each cell. Next, an identical copy of the tissue is created in which the integer identifiers are randomly shuffled. A ground truth mapping from the first to the second integer identifiers is generated. Next, the algorithm is applied. Upon conducting 100 such tests, we find that all identified cell-to-cell mappings are matched correctly, as compared to the ground truth. In rare examples, isolated cells at the boundary of the tissue are not tracked. In these examples, either a single cell has only one adjacent cell in the tissue, or two cells of identical polygon number are adjacent and share exactly one neighbour. Neither the MCS detection algorithm, nor the post-processing algorithm are able to resolve such mappings, which involve fewer than four cells in each dataset (fewer than five percent of the tissue).

We design four further tests of tissue rearrangements (Figure 6.6B-E). The first test comprises tissue movements between images (Figure 6.6B). In this test, a tissue of size fifteen by eight cells is generated as described in Section 6.2. Two smaller tissues of width seven units are cut out of this tissue, which each cover the full height of the tissue, and which are horizontally translated relative to each other by a distance of two cell lengths. The position of each three-cell junction in both tissues is shifted such that the x-coordinate of the left-most junction in each tissue is zero.

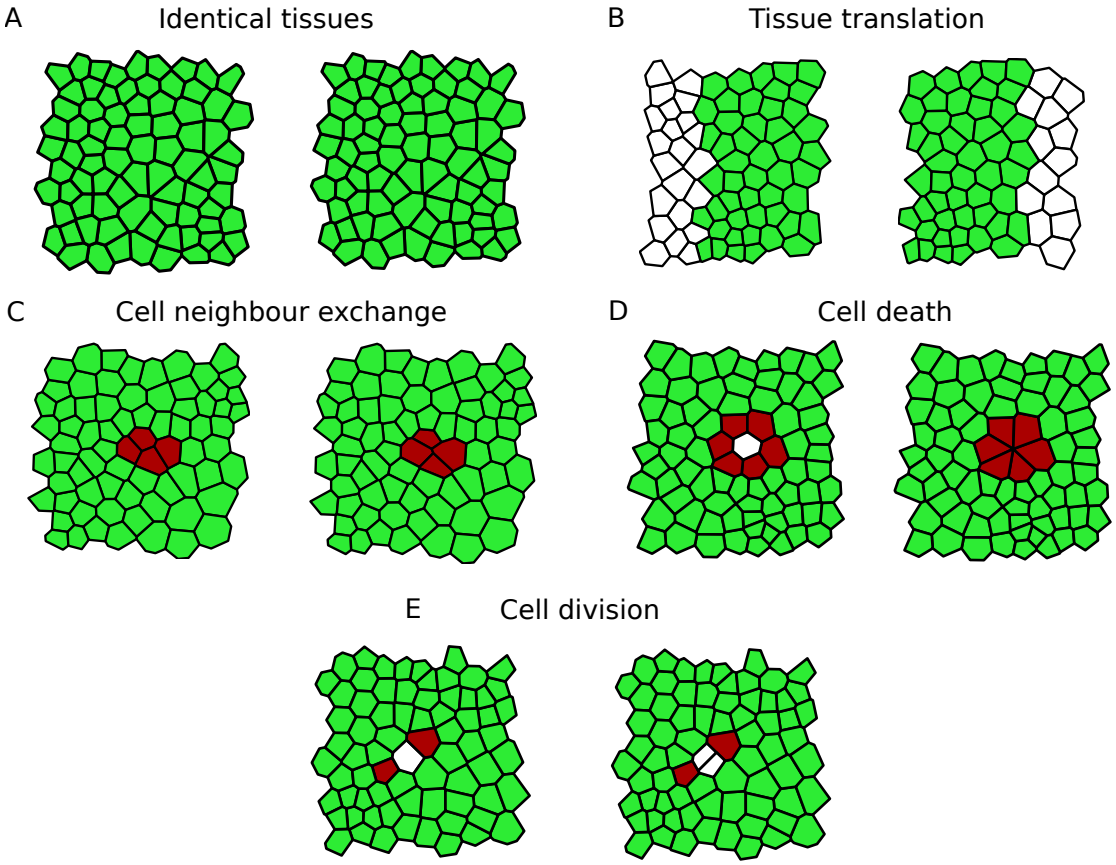

Figure 6.6: Examples of *in silico* test cases. In each image, cells identified by the MCS algorithm are highlighted in green (light), whereas cells that have been filled in by the post-processing steps are highlighted in red (dark). The algorithm tracks cells between identical tissues (A), in tissues undergoing translation (B), cell neighbour exchange (T1 transition) (C), cell removal (D) and cell division (E).

The second test (Figure 6.6C) generates cell neighbour exchanges, also called T1 transitions, as defined in Chapter 2 (Nagai et al., 1988; Etournay et al., 2015). In our implementation of T1 transitions, an edge shared by two cells is replaced by a new perpendicular edge such that the local cell connectivity changes (Figure 6.2B). The length of the new edge ($l_{\mathrm{T1}} = 0.2$ length units) is chosen as a non-zero value smaller than the average edge length in the tissue (see Figure 6.6C). Note, that the precise length of the new edge is not relevant since the algorithm performance is determined by neighbour relationships only, independent of interface lengths. We create two identical copies of a tissue of size nine by nine cells. In the second copy, a T1 transition is performed on an edge in the centre of the tissue.

The third test involves cell removal (Figure 6.6D). In this test, we first generate two identical copies of a tissue of size nine by nine cells. In the second copy, we replace the central cell by a vertex shared by its neighbouring cells, a rearrangement similar to so-called T2 transitions introduced in Chapter 2 (Nagai et al., 1988). The final test involves cell divisions (Figure 6.6E). Here, we once again create two identical copies of size nine by nine cells. In the second copy, a cell in the centre of the tissue is bisected by introducing a straight line in a random direction through the centroid of that cell.

For all tests generated in this way, integer cell identifiers in the second tissue are randomly shuffled, and a ground truth is generated. We run 100 realisations of each test case, and compare the tracking outcome to the ground truth. In all cases we find that cells are tracked correctly, with at most three unmatched cells at the boundary of the sheet.

In Figure 6.6, all cells identified after the cleaning step, in which weakly connected cells are removed from the MCS, are coloured green, whereas cells that are identified by the post-processing algorithm are coloured red. Note that the exact number of cells that are identified by the post-processing algorithm varies between individual realisations of the tests. In many cases, the cells identified by the post-processing algorithm include cells that are adjacent to those undergoing division, removal or neighbour exchange.

We next analyse the extent to which the success of our tracking algorithm depends on the number of Lloyd's relaxation steps, n_L, used to generate the *in silico* datasets. To investigate this we iteratively increase n_L, thus generating tissues with increasingly homogeneous graph structures, and repeat all tests. We find that the algorithm successfully passes all tests for all values of n_L from 4 up to 14.

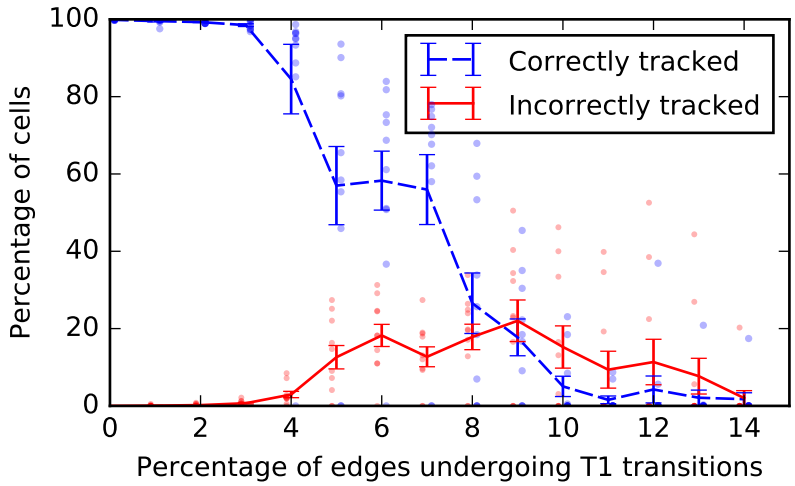

Figure 6.7: Success rate of the algorithm for *in silico* tissues with increasing frequency of cell rearrangement. Virtual tissues spanning 20 cell lengths in each dimension are generated, and T1 transitions are applied to an increasing proportion of the inner edges of the tissue. For each ratio of T1 transitions, 10 repetitions of the test are run, and the ratio of correctly and incorrectly tracked cells in the tissue is recorded. The dashed blue and solid red lines correspond to mean values of correctly and incorrectly tracked cells, respectively. Error bars denote the standard deviation of the mean, and results of individual runs of the test are represented by dots. When 3% of the edges in the tissue undergo T1 transitions, roughly 25% of the cells exchange neighbours.

6.3.2 Algorithm performance for large numbers of cell neighbour exchanges

To assess the performance of the algorithm when applied to tissues exhibiting large numbers of cell neighbour exchanges, we next apply the algorithm to *in silico* datasets with increasing numbers of cell neighbour exchanges between frames (Figure 6.7). The number of correctly tracked cells decreases as the number of cell neighbour exchanges increases. However, the number of incorrectly tracked cells remains below 20% throughout the analysed range of neighbour exchanges, and decreases to zero as the number of edge swaps exceeds 10%.

The number of untracked cells increases rapidly as the percentage of cell-cell interfaces that are swapped between successive images increases from five to ten percent.

Note that the percentage of cells involved in these neighbour exchanges is larger than the percentage of cell-cell interfaces that are swapped, since an individual T1 transition changes the cell neighbour relations of four cells, and each cell shares multiple inner edges. For example, rearranging five percent of the inner edges of the tissue affects roughly 40% of the cells in the tissue, while rearranging ten percent of the tissue edges affects up to 70% of the cells. The number of (correctly or incorrectly) tracked cells drops to zero if the tissue rearranges so much that the neighbourhood of each cell changes; in this case a first match cannot be found to initialise the MCS construction algorithm.

6.3.3 Application of the algorithm to *in vivo* data

Figure 6.8 shows the first three of 21 segmented image frames of the lateral epidermis of a stage-11 *Drosophila* embryo to which the algorithm was applied. During stage 11, gross morphogenetic movements do not occur but the tissue is very active with a large number of cell proliferation events occurring within a short duration, making this a much more challenging tissue on which to perform cell segmentation and tracking than the wing imaginal disc, where many previous efforts have been made (Heller et al., 2016; Farhadifar et al., 2007; Aegerter-Wilmsen et al., 2010; Mao et al., 2011). Cell delamination is also more common than in the wing imaginal disc during normal development. This stage of development thus offers a true test of the proposed algorithm.

The images were taken five minutes apart over a time span of 100 minutes, and these first three images comprise 271, 263 and 263 cells, respectively. Our algorithm tracks 247 cells between the first and second images, 245 cells between the second and third images and 234 cells across all three images. The centroids of cells of previous images are superimposed on the tracking results in Figure 6.8, illustrating that the algorithm successfully tracks cells in situations where it is difficult to match cells

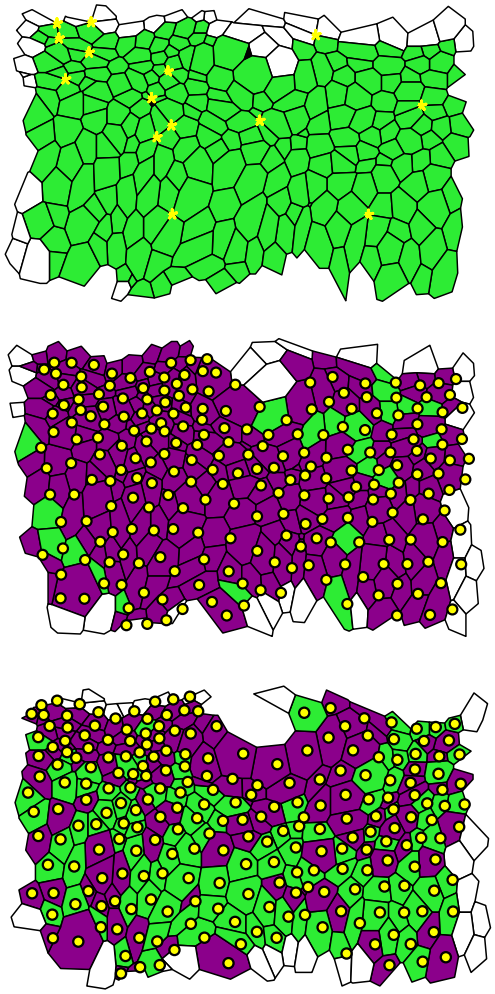

Figure 6.8: Three segmented data frames of an *in vivo* time-lapse microscopy video of the lateral epidermis of a stage-11 *Drosophila* embryo. Cells that are tracked across all frames are coloured green or purple, and cells that leave or enter the tissue at the boundary are white. Dying cells are black. The centroids of tracked cells of the respective previous frames are included as yellow dots, and cells that contain only their centroid from the previous frame are coloured green, whereas cells that do not contain their centroid from the previous frame, and cells that contain multiple centroids, are coloured purple. Together, the centroid information and the colouring illustrate that it is challenging to track cells between the data frames using solely centroid positions. Yellow asterisks in the first frame denote higher-order junctions where more than three cells meet.

between images based on their centroid positions alone. Cells that include only their corresponding centroid from the previous image are coloured in green, while cells that do not include their corresponding centroid from the previous image, and cells that include multiple centroids from the previous image, are coloured in purple. In the first frame we highlight 'higher-order' junctions (shared by four or more cells) by yellow asterisks. Such junctions occur frequently throughout the dataset, illustrating that such higher order junctions do not pose a challenge to our algorithm.

On average, cell centroids move 0.75 cell lengths between the first and second images, with a maximal displacement of 1.17 cell lengths. Between the first and second images 36 cells undergo a net gain in edges, whereas 20 cells have a net loss of edges. In total, four cell deaths and no cell divisions are observed across the three data images. Manual inspection of all individual cell tracks reveals that none of the cells are tracked incorrectly.

The data in Figure 6.8 are the first three out of 21 frames. In Figure 6.9 we show the results of the analysis of the full dataset, including all 21 frames. During the period of measurement the total number of cells increases from 280 to 330 cells, whereas the total number of tracked cells increases from 270 to roughly 310 cells (Figure 6.9A). As the number of cells in the tissue rises, the total number of cell rearrangements increases (Figure 6.9B), whereas the average cell area decreases (Figure 6.9C). Here, the number of cell rearrangements is measured by counting how many cells change their cell-neighbour number between consecutive frames. For all frames, the number of tracked cells is lower than the number of cells in the tissue (Figure 6.9D). Visual inspection of the tracked data reveals that the difference between the total number of cells and the number of tracked cells is largely due to cells entering or leaving the field of view. The percentage of cells that our algorithm tracks is lowest (84%, Figure 6.9D) when the rates of cell division (Figure 6.9A) and cell rearrangement (Figure 6.9B) are highest, which occurs at 70 minutes. Here, the number of tracked cells

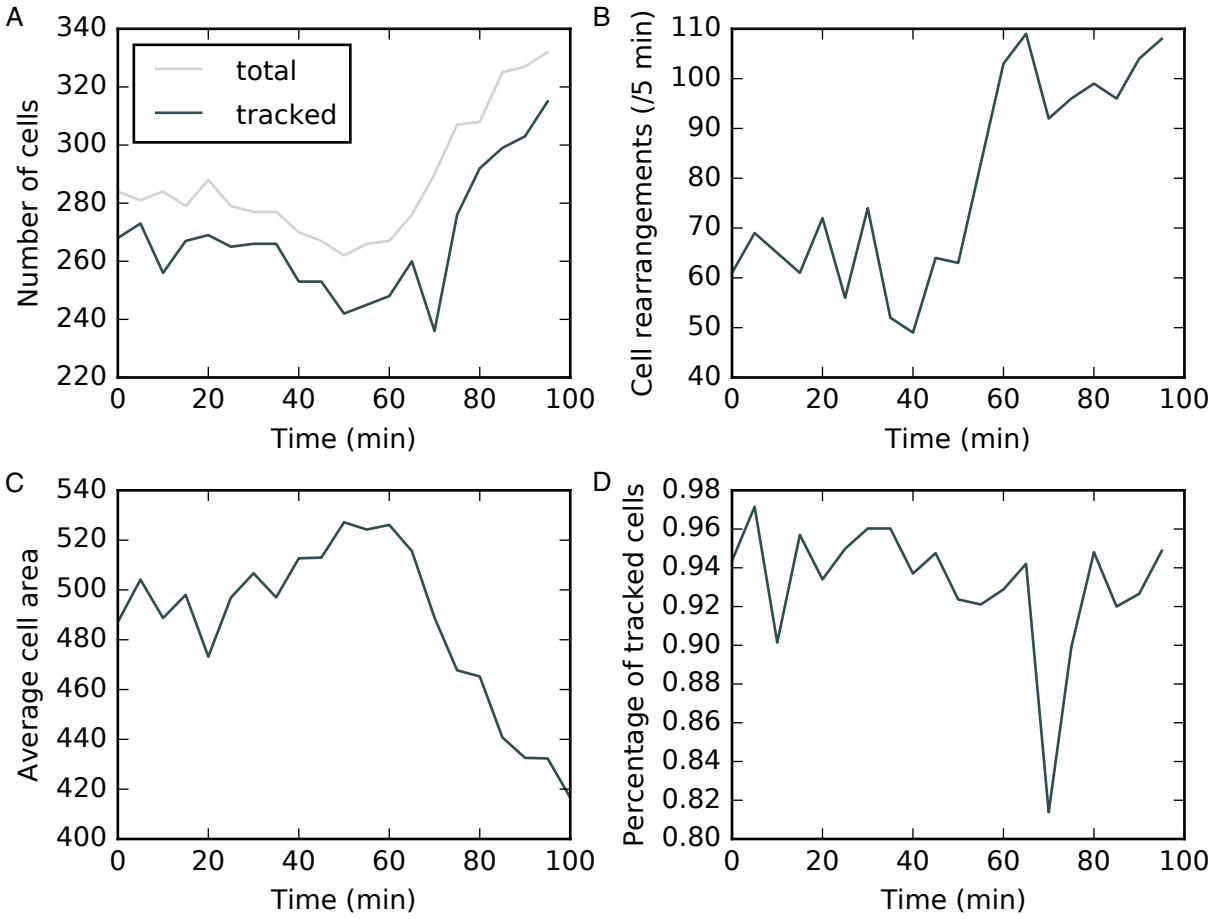

Figure 6.9: Tracking statistics of the *in vivo* dataset. (A) The total number of cells is constant initially and increases from 60 min onwards. The total number of tracked cells correlates with the total number of cells in the tissue. (B) The total number of rearrangements between successive time frames is measured by our algorithm. We record the total number of rearrangements as the total number of cells that either gain or loose neighbours between successive frames. (C) The average cell area in each frame decreases as the total number of cells increases. (D) The percentage of tracked cells decreases at around 70 min, when the amounts of cell rearrangement and division are highest.

decreases since the algorithm is not yet able to resolve division events immediately adjacent to rearrangements as well as multiple adjacent divisions.

Since cell rearrangements are one of the most difficult aspects of cell tracking, and our *in vivo* data exhibit a high frequency of such events, it is natural to ask what percentage of cells are correctly tracked. To estimate this percentage, we compare the results in Figure 6.9, where up to 30% of cells are involved in neighbour exchange between frames in a population of up to 340 cells, with the results shown in Figure 6.7, where in the case of 400 cells and 4% of edges undergoing T1 transitions between frames (corresponding to 30% of cells involved in neighbour exchanges) we find the percentage of correctly tracked cells to be 85%. This provides a lower bound for the success rate of the algorithm on the *in vivo* frames. When up to 3% of edges undergo T1 transitions (corresponding to 25% of cells in the tissue involved in neighbour exchanges), the success rate of the algorithm is approximately 98%.

The tracking of epithelial *in vivo* data enables quantitative assessment of dynamic changes in cellular morphology. The tracking results in Figure 6.9 reveal that the analysed section of the epidermis undergoes 60 cell rearrangements per five minutes initially and around 100 cell rearrangements per five minutes at the end of the observed time interval. The average ratio of the maximal area and the minimal area observed for individual cells during the period of measurement is 4.2, indicating that on average cells increase their apical area by a factor of four during mitotic rounding. A total of 18 cell deaths are tracked in the dataset. A striking feature is the level of T1 transitions occurring during this stage of development, even in the absence of the kinds of gross morphogenetic movements found in earlier or later stages.

6.3.4 Calculation times

To analyse the scaling of the calculation times with tissue size we repeat the permutation test with tissues of square dimension of varying size on a desktop computer

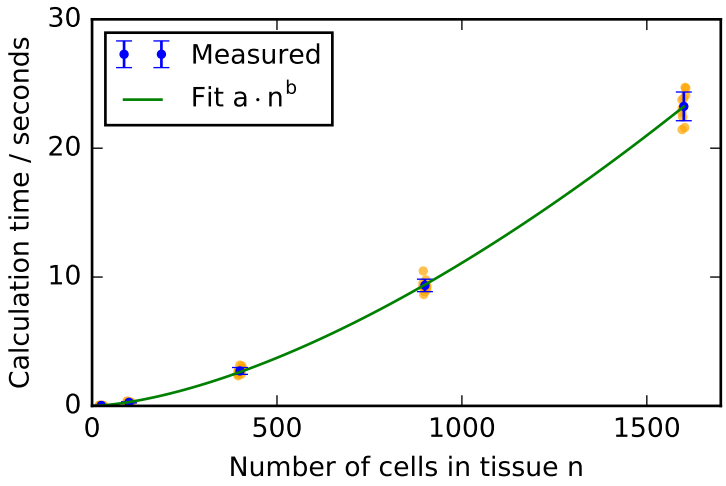

Figure 6.10: Scaling of the calculation times with tissue size. Virtual tissues of varying sizes were generated and the calculation times of the algorithm under the permutation test in Figure 6.6A recorded. Orange dots represent calculation times for individual realisations of the test and error bars denote the standard deviation. The exponent b of the polynomial fit is 1.6. The calculation times were measured on a desktop computer with an Intel i5-6500T CPU (2.5GHz) and 8GB RAM.

with an Intel i5-6500T CPU (2.5GHz) and 8GB RAM. We find that the calculation times scale subquadratically with cell number (Figure 6.10).

The calculation times for the experimental images analysed in Figure 6.8 vary more widely than for the *in silico* datasets. For the tracking between the first and second frames in Figure 6.8, the algorithm required 43 seconds to run, whereas between the second and the third frames the algorithm required nine seconds. This is due to differences in the time required to find the first correct mapping; in the first example 154 cells were searched before the first correct mapping was found, whereas in the second example only 12 cells were searched. This means that the number of cells considered when finding the initial mappings depends on the graph structure of the analysed frames and impacts on the calculation time of the algorithm. In total, analysing all 21 frames of the *in vivo* data presented in Figures 6.8 and 6.9 requires 19 minutes of calculation time.

6.4 Discussion

Cell tracking in epithelial sheets has the potential to generate a vast amount of quantitative data to inform our understanding of the contributions of different cellular processes to tissue morphogenesis. However, cell tracking is notoriously difficult, especially for the complex morphogenetic processes that occur as embryogenesis proceeds. In this chapter, we have developed an algorithm based on MCS detection for the tracking of cells in segmented images of epithelial sheets. Our algorithm successfully tracks cells in *in vivo* images of the *Drosophila* embryonic epidermis, a challenging dataset compared to other tissues, as well as in randomly generated *in silico* datasets, without the need for the adjustment of tissue specific parameters such as weights for individual terms in a global minimisation scheme (Padfield et al., 2011). The use of *in silico* data to test our algorithm allows us to analyse its performance for a large range of experimentally observed cell rearrangements and tessellations.

The tracking of cells in *in vivo* datasets such as presented in Figures 6.8 and 6.9 provides quantitative insight into tissue homeostasis. Using our algorithm we measure example quantities that would not be accessible without a robust cell tracking method. The amount of cell rearrangement, the extent of mitotic rounding, and the occurrence of cell death in the observed frames each can be used to learn about tissue homeostasis in developing epithelia. Within the analysed dataset, we find a significant number of T1 transitions despite the absence of gross morphogenetic movements. This may be driven by the large number of proliferation events that occur. Further, using *in vivo* imaging together with our tracking algorithm allows the observation of cell death or cell delamination without the need for fluorescent markers of apoptosis. Future applications of the algorithm to such processes may, for example, provide novel insight to tissue size control in the *Drosophila* embryonic epidermis (Parker, 2006; Kursawe et al., 2015), for example through the study of growth and division of individual cells, and the algorithm can also be adapted to study the dynamics of epithelial wound

closure. In this and other systems cell tracking may enable the distinction between cell death due to delamination as opposed to apoptosis (Marinari et al., 2012).

Our algorithm is able to track cells that undergo significant movement and neighbour exchanges between frames. For example, we can correctly track cells in tissues where more than 40% of the cells rearrange between successive movie frames (Figure 6.7). In addition, even comparably large gaps in the initial MCS can be filled in during the post-processing step (Figures 6.2 and 6.8). For example, in the first tracking step in Figure 6.1, only 182 of the 246 tracked cells were identified by the MCS algorithm, and it was possible to track the 64 remaining cells during the post-processing step. For comparison, Heller et al. (2016) report 15 cell rearrangements per 1000 cells per hour at an imaging interval of six minutes for their time-lapse microscopy data of *Drosophila* wing imaginal discs. In addition, the experimental data shown in Figures 6.2, 6.8, and 6.9 include junctions shared by four or more cells (yellow asterisks in Figure 6.9) while our *in silico* data include multiple instances of such junctions (Figure 6.6D). Therefore, higher-order junctions, such as multicellular rosettes (Blankenship et al., 2006; Trichas et al., 2012), do not pose a challenge to our algorithm. Access to quantification of cell rearrangement and area changes has recently provided insight to wing morphogenesis in *Drosophila* (Etournay et al., 2015).

Our algorithm is able to correctly track cells in all considered test cases. However, on rare occasions a few cells at the tissue boundary cannot be tracked. It may be possible to adapt the algorithm to track these cells, if this is considered necessary for the application at hand. In the current version of the algorithm, two connections to already tracked cells that are preserved between two time frames are a condition to add a cell-to-cell mapping in the post-processing algorithm. Further analysis of cases where this condition is not fulfilled may reveal ways to relax it.

When generating *in silico* data to test the algorithm, we used Voronoi tessellations in combination with Lloyd's relaxation to generate data that resembles tissues in the

Drosophila wing imaginal disc (Sánchez-Gutiérrez et al., 2016). We expect the algorithm to perform less well on tissues whose network structure is nearly homogeneous. For example, in an epithelial sheet where cells are arranged in a hexagonal fashion, such as the early *Drosophila* embryonic epidermis (Warn and Magrath, 1983) or the late pupal *Drosophila* wing (Classen et al., 2005), the local adjacency network of each cell is identical, and hence a network-based tracking algorithm may not be able to distinguish cells. When generating *in silico* tissues, we use four Lloyd's relaxation steps after Voronoi tessellation. With each Lloyd's relaxation step, the homogeneity of the tissue increases. We were able to successfully repeat all *in silico* tests on virtual tissues that were generated using up to 14 Lloyd's relaxation steps. Hence, we expect the algorithm to be suitable for tissues that can be well described with 14 or fewer Lloyd's relaxation steps, such as the chick neural tube embryonic epithelium, or the *Drosophila* eye disc (Sánchez-Gutiérrez et al., 2016).

The algorithm relies on being able to generate polygonal tessellations from segmented video microscopy data. In particular, all *in silico* tests we conducted considering tissues where each cell has at least three neighbours. Conceptually, it would be possible to apply the algorithm to tissues in which individual cells may have only two neighbours (Ishimoto and Morishita, 2014), although such examples have not been included in the present analysis.

In microscopy videos including division events we expect the algorithm to perform well in tissues in which no adjacent divisions occur between successive movie frames, and in which cells adjacent to the dividing cell do not undergo rearrangements before the next frame is captured. Our algorithm is designed to identify mother and daughter cells of a division event by establishing the bordering cells that gain an edge during the division event. In the case of two adjacent divisions, and if cells adjacent to a division event gain edges due to cell rearrangements, the dividing cell cannot be correctly identified. An example of a typical tracking error for two adjacent divi-

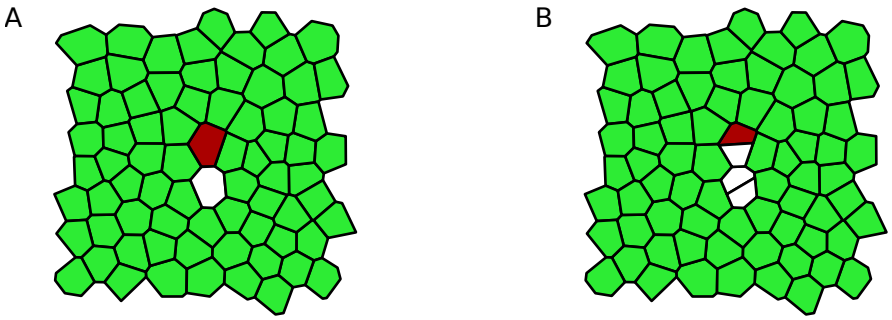

Figure 6.11: Tracking errors can occur if adjacent cells divide. Here, all green (light) cells are tracked correctly. One of the mother cells (red/dark) of the division events has been incorrectly associated with one of the daughter cells of the division.

sions is shown in Figure 6.11. In cases where the division resolution step fails, our Python implementation returns all tracked cells of the post-processing step, and gives a warning that the division has not been resolved. In these cases, manual correction methods could be used for incorrectly tracked cells in the vicinity of division events.

The parameters of the algorithm are chosen to maximise its robustness and avoid the necessity to adjust the parameters to individual applications. For example, the cutoff length, $d_{\max}$, that determines the distance below which two cells in consecutive movie frames are considered mappable to each other, was chosen to be 10 times the average cell length in the tissue, which is significantly larger than the movement that is to be expected between consecutive frames of a live-imaging microscopy video. However, parameter adjustments may be possible for individual applications in order to decrease the algorithm calculation times. For example, the size of the extended neighbourhood considered in the initial step or the iterative extension could be reduced to include only nearest neighbours instead of nearest neighbours and second nearest neighbours in case the tissue is sufficiently heterogeneous. Similarly, for maximal efficiency, one might decrease the $d_{\max}$ for possible cell pairings if the cell positions are not expected to vary significantly between time frames.

Adjustments may be possible to extend the applicability of the algorithm to a wider range of tissues. For example, instead of automatic detection of the initial seeds for

the MCS detection algorithm, a small set of seeds could be manually supplied to guide the tracking. This should improve the performance of the algorithm on homogeneous tissues. In such cases, irregular boundaries may also help to aid the initial seeding. During the construction of the MCS, non-adjacent cells are considered for addition to the MCS whenever the extension of the MCS by adjacent cells is not possible. An alternative option to extend an intermediate MCS may be to repeat the initial seeding algorithm.

In the present work, we have sought to keep geometrical input to the algorithm to a minimum. Cases where geometric data are taken into account comprise division events where one of the daughter cells is four- or three-sided, since in these cases we are not able to make a decision as to which cell is the second daughter cell-based on network adjacency alone. If future applications reveal cases where the algorithm performs poorly due to a large number of cell neighbour exchanges or high degree of tissue homogeneity, then it may be possible to construct algorithms that combine information on the network topology with data on cell shapes, cell positions and cell movements to improve performance. For example, information on network topology could be integrated into previous algorithms that minimise differences between geometric properties of cells, such as cell size and location (Puliafito et al., 2012), with information about network connectivity.

In cell tracking applications, the scaling of the algorithm with tissue size is crucial. Potential applications range from systems of 30 cells, such as in Chapter 4 (*Drosophila* embryonic epidermal P compartments (Parker, 2006)), to 10,000 cells, such as in Chapter 3 (*Drosophila* imaginal wing disc (Farhadifar et al., 2007); the wing pouch has about 3,000 cells (Narciso et al., 2015)). Calculation times in the presented algorithm scale subquadratically with cell number, making it suitable for applications of varying sizes. For example, extrapolating the data in Figure 6.10, a tissue of 10,000 cells could be tracked across two frames within 10 minutes. The scaling of

the algorithm is polynomial despite the fact that it is based on MCS detection, which is known to scale exponentially in the general case. MCS detection has a wide range of research applications, including protein interaction networks (Ciriello et al., 2012; Aladağ and Erten, 2013) and finding the binding sites of chemical structures (Raymond and Willett, 2002). Our approach of reducing the MCS search to a localised search may have applications in other areas where the networks are inherently planar.

Our algorithm is designed to track cells in segmented microscopy videos of epithelial sheets in two dimensions. However, it may be possible to apply the algorithm to datasets of epithelial sheets that are embedded in a three-dimensional environment, such as the *Drosophila* imaginal wing disc (Mao et al., 2011), or the *Drosophila* embryonic epidermis (Parker, 2006; Rauzi et al., 2008), including tissues that can be mapped onto a cylinder or ellipsoid, such as the mouse visceral endoderm (Trichas et al., 2012).

A large number of cell tracking algorithms have been developed for various applications (Mashburn et al., 2012; Cilla et al., 2015; Padfield et al., 2011; Youssef et al., 2011; Schiegg et al., 2013; Heller et al., 2016; Liu et al., 2014; Raffel et al., 2007; Al-Kofahi et al., 2010; Amat et al., 2014; Wait et al., 2014; Winter et al., 2011; Sommer et al., 2011; Bellaïche et al., 2011; Aly et al., 2014; Wang et al., 2010; Puliafito et al., 2012; Aigouy et al., 2010). Further efforts are required to compare these algorithms with our own, and to identify the algorithm best suited for an individual dataset. In the cell tracking challenge, Maška et al. (2014) provide microscopy videos from a variety of *in vitro* cell cultures, including, for example, mouse embryonic stem cells and human squamous lung carcinoma cells, together with ground-truth segmentation and tracking data as benchmarks for cell tracking and segmentation algorithms. However, many of the published algorithms listed above have not yet been applied to the challenge, and benchmark datasets for epithelial sheets are not currently available. Chenouard et al. (2014) use *in silico* datasets as benchmarking datasets for particle

tracking algorithms. The fully segmented dataset published within the MCSTracker project can provide a benchmark for future epithelial cell tracking applications.

In the next chapter, we will summarise this thesis and discuss potential avenues of future research.

Chapter 7

Discussion

Throughout this thesis we have made contributions that enable the quantitative study of epithelia. In Chapter 3, we analysed the extent to which typical vertex model predictions may depend on implementation parameters. In Chapter 4, we used the vertex model to explore mechanical mechanisms for size control in the *Drosophila* embryonic epidermis and to predict the outcome of future experiments. In Chapter 5, we investigated whether vertex model parameter values can be experimentally inferred using static imaging data and how the uncertainty of such estimates can be quantified. In Chapter 6, we used concepts from network theory to track cells in live-imaging microscopy videos.

Our results in Chapter 3 illustrate that care needs to be taken when making predictions using cell-based computational models because implementation details such as the size of the time step, or non-physical parameters, such as length thresholds for cell rearrangement, may influence model predictions significantly. With the rise of quantitative analysis and model-data comparison in biophysical applications, choices of model implementation become increasingly relevant. To enable the use of cell-based models in quantitative settings, it is important to be aware of any influences that implementation choices may have on model predictions when analysing a specific biophysical phenomenon. Understanding model behaviour in detail is crucial to

prevent modelling artefacts from influencing experimental predictions and clouding our biophysical understanding and, as such, our findings emphasise the need to fully document algorithms for simulating cell-based models. Close attention to implementation details is required in order to unravel the full predictive power of cell-based models.

In future applications, cell-based models will be developed further. As the resolution of quantitative data that can be collected increases, computational models will be adapted accordingly. Future cell-based models will include more detailed representations of cell shapes, for example through immersed boundary approaches (Cooper et al., 2016) or other models that take the composition of cells into account, such as subcellular element models (Newman, 2007). In the case of vertex models, future efforts will include advancing the modelling concept to three dimensions. This may include two-dimensional surfaces embedded in three-dimensional space (Monier et al., 2015), or monolayers of three-dimensional polyhedral cells (Okuda et al., 2015b) that may be used to model, for example, tissue buckling, or three-dimensional cells in a three-dimensional cell-aggregate (Okuda et al., 2015a). Another important avenue for future research on vertex models is the proposal of energy equations for which individual cells do not undergo T2 transitions (cell removal), even when exposed to high pressure from the surrounding cells. Such a model would allow us to test hypotheses regarding the control of cell removal and death in epithelial tissues.

Chapter 4 serves as an example of using computational models as an abstraction of the maintenance of tissue sizes with implications for a broad range of studies. Significant advances in stem cell engineering have resulted from understanding how to unlock the potential for multicellular aggregates to self-organize. Recent examples include the morphogenesis of optic eye cups in organ culture conditions (Eiraku et al., 2011) and the engineering of beating mini-hearts (Ma et al., 2015). We posit that great success in developing tissue repair strategies will come through the reverse engineering

of pattern repair mechanisms in situations where pattern repair is perturbed. Such reverse engineering will require guiding experimental efforts through modelling studies that identify the information needed to distinguish between mechanisms.

Future work on tissue size control will include studies on other model systems. For example, the mouse blastocyst has stable proportions of differentiated cells even if the system size is changed (Saiz et al., 2016). For the example of the *Drosophila* embryonic epidermis, future efforts may include the modelling of perturbations of the EGFR pathway, which have previously been shown to affect tissue sizes (Parker, 2006).

Our work in Chapter 5 indicated that the mechanical vertex model parameters might not be identifiable using tissue-level summary statistics or local perturbations, thus illustrating the importance of quantifying uncertainty when measuring parameters of stochastic models in biology. While our findings highlight that it is difficult to identify a specific vertex model parameter set corresponding to a given tissue configuration, it also follows that the exact choice of mechanical vertex model parameters will not strongly influence the model behaviour, and many modelling results will be robust to changes in these parameters. Future efforts for parameter inference will include methods that go beyond measuring the non-dimensionalised parameters of the vertex model, and instead aim to measuring all dimensional parameters of the model, including, for example, the cellular stiffness, the dimensional target area of cells, the timescales of mechanical rearrangement, and parameters concerning cell cycle progression. Initial efforts in this direction have already been conducted by Merzouki et al. (2016) and Xu et al. (2015a). It will be important to extend such efforts to methods that quantify the individual parameter uncertainty, and that are applicable in living tissues.

The proposed cell tracking algorithm in Chapter 6 provides a two-dimensional tracking solution specialised for epithelial sheets that attempts to maximise the information that can be gained from the packing that is typical to such tissues. It

may, however, be possible to extend this algorithm to applications of cell tracking where cells are not physically connected by constructing adjacency networks from Voronoi tessellations that use the cell locations as seeds. We hope that, as segmentation tools are developed further, the combination of our algorithm with these tools will lead to further insights into cellular behaviour in epithelial tissues. Specifically, next-generation segmentation techniques will go beyond traditional image analysis through, for example, intensity thresholds, filters, and binary image operations, and embrace modern techniques for data mining, such as deep learning algorithms (LeCun et al., 2015).

7.1 Contributions to open-source software projects

In order to use computational models, such as in Chapters 2 to 5, or algorithms for data analysis, such as in Chapter 6, it is essential to write computer software. All software used in this has been shared either through the Chaste software project (Mirams et al., 2013), or as supplementary material to journal publications (Kursawe et al., 2015, 2016a), or as a github project (Kursawe et al., 2016b). Making software available in this way enables the reproducibility of computational results and reduces the effort that may be required to write new software that uses similar methods (Osborne et al., 2014).

While developing the software used in this thesis we have made multiple contributions to the Chaste project, which we now list briefly. For Chapters 3 to 5 we designed a new method to run multiple Chaste simulations with identical or distinct parameters in parallel either on single computers or on a research cluster. This method has since become a standard approach among Chaste users and involves the writing of executable programs for specific modelling applications. We have used this method to write a program that allows the simulation of growing *Drosophila* wing imaginal discs in Chapters 3 and 5 and enables the user to set all relevant parameters, such as

those listed in Table 3.1, through command-line options. We have written a similar program to simulate tissue size control in Chapter 4. We further accelerated vertex model simulations in Chaste by refactoring the method for boundary merging, which led to a reduction in the calculation time of a typical simulation of a growing wing disc by 35%. Additionally, we added Chaste features that enable the use of user-defined rules for cell division orientation, as well as user-defined rules for updating cell target areas in vertex models based on, for example, the cell cycle progression or the simulation progress. In both cases, we have written multiple such rules and added them to the Chaste repository. Finally, we have contributed to various maintenance and refactoring projects of Chaste that include, for example, updates to the user-interface for specifying the output of a Chaste simulation.

The work for Chapter 6 comprises a separate, open-source python software project which is available as a separate online repository (`https://github.com/kursawe/MCSTracker`). Features of this project include a set of python classes which allow the user to quickly convert segmented microscopy images into network representations. These network representations give access to cell and tissue properties, such as cell areas and cell neighbour relationships. The project also contains a user-friendly python interface that allows the user to run the cell tracking algorithm developed in Chapter 6 on a given data set and to quickly collect summary statistics on tracked data sets, for example the amount of cell rearrangement. Parts of this code have been used in the Tiff2Chaste project, which aims to enable the use of microscopy data sets to initialise Chaste simulations (`https://github.com/Chaste/Tiff2Chaste`).

As the quantitative techniques become increasingly established within the biological and medical community, and computational models continue to be refined, the two will be increasingly connected. Future versions of modelling software will include close integration with experimental data, for example through projects such as Tiff2Chaste mentioned above. At the same time, the user-interface of modelling software will be

improved. For example, even though Chaste is developed in the programming language C++ there exists now a python interface (`https://github.com/jmsgrogan/PyChaste`) that facilitates the quick prototyping of new simulations. Eventually, cell-based modelling will become a standard tool among biomedical researchers.

7.2 Conclusions

While the motivation for the work in this thesis were fundamental questions in developmental biology, the methods developed in Chapters 3 to 6 are applicable to the study of other phenomena, as well, such as wound healing (Walker et al., 2004) and tumour growth (Anderson et al., 2006). In addition, the study of epithelia is not limited to developmental systems but also comprises, for example, the study of tissue mechanics (Nier et al., 2016) and tissue renewal (van Leeuwen et al., 2009).

New findings and techniques, such as those presented in Chapters 3 to 6, will contribute to unravelling the most fundamental principles of life, and help achieve breakthroughs in the cure of diseases and in synthetic biology. The combination of the newest experimental techniques and mathematical or computational modelling will be the basis for such findings.

Appendix A

Mathematical description of the cell tracking algorithm

A conceptual overview of our cell tracking algorithm is given in Chapter 6. Here, we provide a detailed description of each step of the algorithm. The input to the algorithm is a set of segmented images obtained from a live-imaging microscopy data set of the apical surface of an epithelial cell sheet. For each image, the segmentation is assumed to have correctly identified which cells are adjacent and the locations of junctions where three or more cells meet. This information is used to generate a polygonal approximation to the cell tessellation (Figure A.1A, Figure 6.1B).

Our algorithm tracks cells between each pair of consecutive images in three steps (Figure 6.1). First, we use a MCS approach (Ullmann, 1976; Krissinel and Henrick, 2004) to generate an initial bijection between the two images that includes every cell whose connections to its neighbours do not change between images, e.g. due to cell rearrangements (Figure 6.2B). Second, we remove from the bijection any cells that have less than three isolated connections to other cells in the MCS (Figure 6.2B-C), as well as isolated clusters of fewer than ten cells, since these cells are likely to have been matched incorrectly. Third, we extend the MCS to track any remaining cells that were not included in the bijection and we identify cell division and 'removal' (delamination,

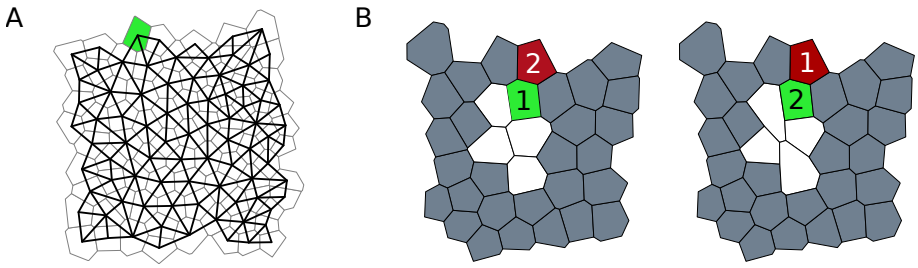

Figure A.1: Construction of the MCS. (A) Overlay of a polygonal tessellation (grey) and the corresponding cell network (black). Each cell corresponds to one vertex in the network, and two vertices share an edge if the corresponding cells are adjacent. The network of cells is used by the algorithm to determine the MCS between tessellations corresponding to consecutive time frames in a microscopy video. Note that the network degree of a cell and its polygon number differ at the boundary of the tissue. For example, the highlighted cell has polygon number five and network degree three. (B) The dark grey cells are members of the conserved MCS between the two *in silico* tissues. In this example, two distinct MCSs are possible. Both MCSs include all highlighted grey, green, and red cells. The two MCSs differ in the way the numbered cells are mapped. The first MCS includes the cell pairings as indicated by the green (light) and red (dark) cells. The second MCS includes the pairings as indicated by the numbers 1 and 2. White cells are not members of the two MCSs.

extrusion or death) events (Figure 6.2D) through characteristic changes to the local cell network under these events.

In the first of the three steps shown in Figure 6.2, the MCS is constructed by iterative extension from an initial seed. The full MCS is then constructed by iteratively adding cells after inspecting MCSs of the cells' extended neighbourhoods.

A.1 Mathematical formulation

We begin by introducing the graph theoretic terminology and notation (Wilson, 2010) used to describe our algorithm. We consider each pair of successive segmented images as vertex-labelled graphs[1] $G = (V, E)$ and $G' = (V', E')$, respectively. Here and throughout, we use a prime symbol $'$ to refer to the latter of the consecutive images. Each vertex in G or G' corresponds to one cell in the respective segmentation, and two

[1] A *graph* is an ordered pair $G = (V, E)$, where $V \subseteq \mathbb{N}$ and $E \subseteq \{A \subseteq V : |A| = 2\}$. The elements of V and E are called the *vertices* and *edges* of G, respectively. Given a graph $G = (V, E)$, a *vertex labelling* is a function of V to a set of labels. With this function, G is called a *vertex-labelled* graph.

vertices share an edge in the graph if the corresponding cells are adjacent. Throughout, we assume the graphs G and G' to be simple, planar and connected; we emphasise that these graphs represent the dual of the polygonal cell packing (Figure A.1A). These assumptions are reasonable in the case of simple epithelial cell sheets.

The vertex labelling of G is defined by three functions, $p_G : V \to \mathbb{N}$, $x_G : V \to \mathbb{R}$, $y_G : V \to \mathbb{R}$, and $r_G : V \to T_G$. For a vertex $v \in V$, we refer to $p_G(v)$, $x_G(v)$, $y_G(v)$, $r_G(v)$ as the *polygon number*, *x coordinate*, *y coordinate*, and *neighbour order* of v, respectively. For a given vertex, the polygon number is the number of neighbours of the corresponding cell, and the x and y coordinates are defined by the centroid of that cell. T_G is defined as the space of ordered sets of arbitrary length on V, and the neighbour order $r_G(v)$ is the set of vertices that are adjacent to v, ordered by the clockwise appearance of their corresponding cells in the segmented image with an arbitrary starting point. An overlay of a polygonal tessellation with the corresponding graph structure is shown in Figure A.1A.

Let ϕ be an isomorphism[2] from $A \subseteq V$ to $B \subseteq V'$ such that for all $v \in A$, we have $p_G(v) = p_{G'}(\phi(v))$ and for all $x, y \in A$, we have $\{x, y\} \in E \Leftrightarrow \{\phi(x), \phi(y)\} \in E'$. We call ϕ a *cell mapping* from G to G' and define the *size* of ϕ to be $|\phi| = |A|$. We say the cell mapping ϕ *preserves rotational order* if for each vertex $v \in A$ the images of $r_G(v) \cap A$ have identical cyclic order to the set $r_{G'}(\phi(v)) \cap B$.

Let S denote the set of cell mappings from subgraphs of G to subgraphs of G'. Suppose that $\phi_{MCS} \in S$ has maximum size, i.e. $|\phi_{MCS}| \geq |\phi| \quad \forall \phi \in S$, and let $V_{MCS} \subseteq V$ denote the domain of ϕ_{MCS}. We call the subgraph induced[3] by V_{MCS} a *MCS* of G and G' (this may not be unique). A non-trivial, i.e. non-empty, MCS exists if there are two vertices $v \in V$ and $v' \in V'$ that have the same polygon number, which is always true in our test cases. Our definition of a MCS differs slightly from

[2]Graphs $G = (V, E)$ and $G' = (V', E')$ are *isomorphic* if there exists a bijection $\phi : V \to V'$ such that, for each $x, y \in V$, we have $\{x, y\} \in E \Leftrightarrow \{\phi(x), \phi(y)\} \in E'$. We say that ϕ is an *isomorphism*.

[3]A graph $G' = (V', E')$ is a *subgraph* of $G = (V, E)$ if $V' \subseteq V$ and $E' \subseteq E$. The subgraph G' of G is *induced* by the vertices $A \subseteq V$ if it contains all edges whose endpoints are both in A.

previous definitions since it requires equivalence of the polygon number in addition to equivalence of edges (Ullmann, 1976; Raymond and Willett, 2002). Note that the polygon number and degree[4] of a vertex may not coincide for cells at the tissue boundary (Figure A.1A). A MCS is said to *preserve rotational order* if ϕ_{MCS} preserves rotational order.

Suppose that G and G' have k MCSs that preserve rotational order, with associated cell mappings $\phi_1, \ldots, \phi_k$. Let V_c denote the set of vertices in V that are mapped to the same vertex in V' by every cell mapping $\phi_1, \ldots, \phi_k$, and let ϕ_c denote the restriction of ϕ_1 (or, equivalently, any of the cell mappings) to V_c. We call V_c the *conserved MCS* of G and G'. In contrast to MCSs, conserved MCSs are unique. Examples of MCSs and conserved MCSs are illustrated in Figure A.1B.

A.2 Construction of the conserved MCS

In general, finding a MCS between two graphs is an NP-hard problem (Ullmann, 1976). Here we adapt an efficient MCS detection algorithm (Krissinel and Henrick, 2004) by exploiting graph planarity to reduce computational complexity. Instead of exploring all possible combinations of vertex-to-vertex matches (Krissinel and Henrick, 2004) we construct the conserved MCS iteratively by finding the MCSs of small subgraphs of G and G'. To describe this construction we make use of the following definitions.

For a graph $G = (V, E)$, we define the *extended neighbourhood*[5] of a vertex $v \in V$ to be the set $\Gamma_G^{(2)}(v) = \{w \in V : d_G(v, w) \leq 2\}$, where d_G denotes graph distance[6]. The extended neighbourhood contains v, all neighbours of v, and all second nearest neighbours of v. An example of an extended neighbourhood is illustrated in Figure 6.3A as the set of highlighted blue and grey cells.

[4]The *degree* of a vertex v of a graph $G = (V, E)$ is the number of incident edges, $\deg_G(v) = |\{w \in V : \{v, w\} \in E\}|$.

[5]The *neighbourhood* $\Gamma_G(v)$ contains all $u \in V$ for which $(u, v) \in E$.

[6]The *distance* $d_G(v, w)$ between two vertices v, w of a graph G is the number of edges in a shortest path connecting them. If no such path exists, then the distance is set equal to ∞.

Let $\rho : A \to B$ be a cell mapping, $v \in V \setminus A$ and $v' \in V' \setminus B$ be vertices in successive graphs, and S^{ρ}_{LM} be the set of cell mappings whose domains lie in $\Gamma_G(v)$, whose images lie in V', which map v to v', and which map v_a to $\rho(v_a)$ for all $v_a \in A \cap \Gamma_G(v)$. Suppose that $\phi^{\rho}_{LM} \in S^{\rho}_{LM}$ has maximum size, i.e. $|\phi^{\rho}_{LM}| \geq |\phi| \ \ \forall \phi \in S^{\rho}_{LM}$, and let V_{LM} denote the domain of ϕ_{LM}. We call the subgraph induced by V^{ρ}_{LM} a *local MCS* (LM) of v and v' under ρ.

Further, let S^{ρ}_{ELM} be the set of cell mappings whose domains lie in $\Gamma^{(2)}_G(v)$, whose images lie in V', which map v to v', and which map v_a to $\rho(v_a)$ for all $v_a \in A \cap \Gamma^{(2)}_G(v)$. Suppose that $\phi^{\rho}_{ELM} \in S^{\rho}_{ELM}$ has maximum size, i.e. $|\phi^{\rho}_{ELM}| \geq |\phi| \ \ \forall \phi \in S^{\rho}_{ELM}$, and let V_{ELM} denote the domain of ϕ_{ELM}. We call the subgraph induced by V^{ρ}_{ELM} a *extended local MCS* (ELM) of v and v' under ρ.

Further, let S^{ρ}_{RLM} denote the set of cell mappings whose domains lie in the extended neighbourhood of v excluding v, whose images lie in V', and which map v_a to $\rho(v_a)$ for all $v_a \in A \cap \Gamma^{(2)}_G(v)$. Suppose that $\phi^{\rho}_{RLM} \in S^{\rho}_{RLM}$ has maximum size and let V^{ρ}_{RLM} denote the domain of ϕ^{ρ}_{RLM}. We call the subgraph induced by V^{ρ}_{RLM} a *reduced local MCS* (RLM) of v under ρ.

Finally, we say that $v' \in V' \setminus B$ is *mappable* to $v \in V \setminus A$ under ρ if $p_G(v) = p_{G'}(v')$, $d_G(w, v) = 1 \Leftrightarrow d_{G'}(\rho(w), v') = 1$ for all $w \in A$, and if $(x_G(v) - x_{G'}(v'))^2 + (y_G(v) - y_{G'}(v'))^2 < d^2_{\max}$, where throughout this thesis we choose the threshold $d_{\max}$ to be ten times the average cell diameter in the tissue (defined as the square root of the average area of the polygonal approximations of the cells in the segmented microscopy image). The threshold $d_{\max}$ is used in our MCS finding algorithm to restrict any possible vertex pairings to those that are in physical proximity. This restriction reduces the size of the search space.

A.2.1 Initial step

To construct the conserved MCS, we first define a cell mapping ϕ_1 between single vertices of the consecutive graphs (Figure 6.3A). Formally, we search through vertices in V and V' to find $v_1 \in V$, $v_1' \in V'$ such that the order[7] of any extended local MCS of v_1 and v_1' under the cell mapping[8] $\phi_0 : \emptyset \to \emptyset$ is equal to $|\Gamma_G^2(v_1)|$ and, for any vertex $v_2' \in V' \setminus \{v_1'\}$ that is mappable to v_1 under ϕ_0, the order of any extended local MCS of v_1 and v_2' is strictly less than $|\Gamma_G^2(v_1)|$. If no such v_1 can be found, we relax our condition and instead search for any $v_1 \in V$, $v_1' \in V'$ such that the order of any local MCS of v_1 and v_1' under the cell mapping $\phi_0 : \emptyset \to \emptyset$ is equal to $|\Gamma_G(v_1)|$ and, for any vertex $v_2' \in V' \setminus \{v_1'\}$ that is mappable to v_1 under ϕ_0, the order of any local MCS of v_1 and v_2' is strictly less than $|\Gamma_G(v_1)|$. We then define a first cell mapping $\phi_1 : V_1 \to V_1'$ with $V_1 = \{v_1\}$, $V_1' = \{v_1'\}$ and a first *set of inspected vertices* $V_1^{\text{ins}} = \emptyset$. Since we wish to use the MCS to aid our cell tracking, the equivalence of the (extended) neighbourhoods of v_1 and v_1' gives us confidence that the corresponding cells are correctly tracked under ϕ_1. If we cannot find an initial cell mapping, then the algorithm halts; this means that the cell connectivity changes so quickly that the neighbourhood of every cell differs between consecutive images.

A.2.2 Iterative extension

Our next step is to iteratively construct a cell mapping $\phi_{\text{cell}} : V_{\text{cell}} \to V_{\text{cell}}'$ for the conserved MCS between G and G', as follows.

For $n = 1, 2, \ldots$, given a cell mapping $\phi_n : V_n \to V_n'$ and a set of already inspected vertices $V_n^{\text{ins}} \subseteq V$, we determine the set of vertices $S_n \subseteq \Gamma_G(V_n) \setminus V_n^{\text{ins}}$ with at least one mappable vertex in $V' \setminus V_n'$ under ϕ_n. If there are no such vertices ($S_n = \emptyset$), then we simply define $\phi_{n+1} = \phi_n$, $V_{n+1} = V_n$, $V_{n+1}' = V_n'$, and set $V_{n+1}^{\text{ins}} = \emptyset$. Otherwise, if there are such vertices ($S_n \neq \emptyset$), then we find a vertex $v_{n+1} \in S_n$ with a smallest

[7]The *order* of G is the number of its vertices, $|V|$.

[8]Here and throughout, $\emptyset$ denotes the empty set.

set of mappable vertices $M'_{n+1} \subseteq V' \setminus V'_n$ under ϕ_n. We then find all RLMs of v_{n+1} under ϕ_n and, for each vertex $v'_m \in M'_{n+1}$, we find all ELMs of v_{n+1} and v'_m under ϕ_n. Next, we find if there is a vertex $v'_{n+1} \in M'_{n+1}$ for which all ELMs of v_{n+1} and v'_{n+1} are larger than all ELMs of v_{n+1} and $v'_m \in M'_{n+1} \setminus \{v'_{n+1}\}$, and larger than all RLMs of v_{n+1}. Finally, we distinguish between the cases (i) where v'_{n+1} exists and the ELM preserves rotational order or (ii) where v'_{n+1} either does not exist or it exists and the corresponding ELM does not preserve the rotational order. If such a vertex v'_{n+1} exists, then we define a new cell mapping $\phi_{n+1} : V_n \cup \{v_{n+1}\} \rightarrow V'_n \cup \{v'_{n+1}\}$ such that $\phi_{n+1}(v_{n+1}) = v'_{n+1}$ and $\phi_{n+1}(v) = \phi_n(v)\ \forall v \in V_n$, and define a new set of inspected vertices $V^{\text{ins}}_{n+1} = V^{\text{ins}}_n$. If there is no such vertex $v'_{n+1} \in \Gamma_G(V_n) \setminus V^{\text{ins}}_n$, then we construct an extended set of inspected vertices $V^{\text{ins}}_{n+1} = V^{\text{ins}}_n \cup \{v_{n+1}\}$, and set $\phi_{n+1} = \phi_n$, $V_{n+1} = V_n$, and $V'_{n+1} = V'_n$. We then increment n and return to the start of the iteration. Note that at each iteration the algorithm proceeds even if there are no non-trivial ELMs or RLMs for a given vertex v_{n+1}.

Whenever we encounter the set of adjacent cells with at least one mappable vertex $S_n = \emptyset$ for two consecutive values of n, the iteration continues with altering the definition of this set to the set of unmapped cells which have at least one mappable vertex, $S_n \subseteq V \setminus (V^{\text{ins}}_n \cap V_n)$, thus removing the restriction of the search to connected subgraphs. As soon as at least one cell mapping has been added to the MCS under this weakened condition, we set $V^{ins}_n = \emptyset$ and again continue the iteration among adjacent cells only, i.e. using $S_n \subseteq \Gamma_G(V_n) \setminus V^{\text{ins}}_n$ and again restrict the addition of new cells to cells adjacent to V_n.

The iteration halts whenever $S_n = \emptyset$ for three consecutive values of n. We then define $\phi_{\text{cell}} = \phi_n$, $V_{\text{cell}} = V_n$ and $V'_{\text{cell}} = V'_n$. Figure 6.3B-C illustrates the cells considered when searching for the RLMs and ELMs of a given vertex.

A.3 Post-processing

The cell mapping ϕ_{cell} is intended to correctly track as many cells as possible between consecutive images. Nevertheless, it is possible that some members of V_{cell} may be tracked incorrectly, while the cell mapping may have excluded some vertices in V that could have been tracked correctly. To eliminate tracking errors and track cells that are not included in the conserved MCS, we construct a *tracking mapping*, ψ_{track}, from $\tilde{V}_{\text{track}} \subseteq V$ to $\tilde{V}'_{\text{track}} \subseteq V'$. We call a mapping $\psi : \tilde{V} \subseteq V \to \tilde{V}' \subseteq V'$ a *tracking mapping* if it is an isomorphism from $\tilde{V}$ to $\tilde{V}'$. In contrast to a cell mapping, a tracking mapping need not preserve polygon numbers or edges between vertices of the subgraphs induced by $\tilde{V}$ and $\tilde{V}'$.

We begin by defining a first tracking mapping $\psi_1 = \phi_{\text{cell}}$ from $\tilde{V}_1 = V_{\text{cell}}$ to $\tilde{V}'_1 = V'_{\text{cell}}$. In the following, we describe how we iteratively refine the tracking mapping by first removing vertices from the domain that we suspect to correspond to incorrectly tracked cells (Figure 6.1B-C), and then we add vertices to the domain to track cells that are not members of the MCS (Figure 6.1D).

A.3.1 Removing weakly connected cells

Let ψ be a tracking mapping from $\tilde{V} \subseteq V$ to $\tilde{V}' \subseteq V'$. We define $v \in \tilde{V}$ to be *weakly connected* with respect to ψ if the set $\Gamma_G(v) \cap \tilde{V}$ contains either exactly one vertex; or exactly two vertices that are not adjacent. We remove any weakly connected vertices from the tracking mapping since the corresponding cells may have been tracked incorrectly by the MCS (Figure 6.1). Once the weakly connected cells have been removed, we remove any cells from the tracking mapping that belong to connected components[9] which lie within the support of the tracking mapping and that contain less than ten vertices.

[9] A *connected component* $C \in V$ in $G = (V, E)$ is a set of vertices $C \in V$ such that all pairs of vertices $v_1, v_2 \in C$ are connected by paths in V and that has no connections to vertices outside C.

In practice, we first find the set of vertices $\tilde{S}_w \subseteq \tilde{V}_1$ that are weakly connected with respect to ψ_1. Next, we let $\tilde{V}_2 = \tilde{V}_1 \setminus \tilde{S}_w$, $\tilde{V}_2' = \tilde{V}_1' \setminus \{\psi_1(w) : w \in \tilde{S}_w\}$, and define a new tracking mapping $\psi_2 : \tilde{V}_2 \to \tilde{V}_2'$ to be the restriction of ψ_1 to $\tilde{V}_2$. Note that this step accounts for the possibility that $\tilde{S}_w = \emptyset$; in this case, we simply have $\psi_2 = \psi_1$.

Further, we identify the set S_i of all vertices that belong to any connected component $\tilde{C}$ within the graph induced by $\tilde{V}_2$ and for which $|\tilde{C}| < 10$. Next, we let $\tilde{V}_3 = \tilde{V}_2 \setminus \tilde{S}_i$, $\tilde{V}_3' = \tilde{V}_2' \setminus \{\psi_2(w) : w \in \tilde{S}_i\}$, and define a new tracking mapping $\psi_3 : \tilde{V}_3 \to \tilde{V}_3'$ to be the restriction of ψ_2 to $\tilde{V}_3$.

The removal of weakly connected cells and small isolated connected components prevents tracking errors in our algorithm. The network structure occasionally allows small connected components to match cells that do not physically correspond to each other. The same is true for weakly connected cells which tend to lie on the boundary of the conserved MCS.

A.3.2 Adding cells that were not tracked by the MCS

We next add cells to the tracking mapping. This is necessary since any cells that have undergone neighbour exchanges between the consecutive images may have changed their polygon numbers, or their adjacency to each other. This means that their corresponding vertices cannot be members of the conserved MCS, and so regions of cell neighbour exchanges will leave gaps of untracked cells in the MCS (Figure 6.1B-C).

In the following, we iteratively extend the domain of the tracking mapping to include vertices that have neighbours within the domain of the tracking mapping. Possible images of a given vertex can be identified by the aid of the images of the neighbours of the vertex. In this way, we track as many remaining cells as possible based on their neighbour relationships to cells that have been tracked by the conserved MCS. The more mapped neighbours that are preserved between a newly added vertex and its image, the higher our confidence that the corresponding cells are correctly

tracked.

Formally, we start with a tracking mapping $\psi_n : \tilde{V}_n \to \tilde{V}'_n$ (initially with $n = 3$). For each vertex in $\tilde{V} \setminus \tilde{V}_n$ we define the *conserved connectivity* $c_n(v)$ as the maximal number of unique neighbours of v in $\tilde{V}_n$ whose images under ψ_n have exactly one shared neighbour v'. Since $c_n(v)$ is the maximal number of *unique* shared neighbours, we set $c_n(v) = 0$ if there are multiple sets of neighbours of $S_i(v)$ that have a exactly one shared neighbour $S_i(v')$ and whose size is maximal.

In practice, we determine $c_n(v)$ as follows. For a given vertex v in $\tilde{V} \setminus \tilde{V}_n$, let $T_n(v) = \{\psi_n(w) : w \in \Gamma_G(v) \cap \tilde{V}_n\}$ denote the set of images of all adjacent vertices of v in the domain of the current tracking mapping. If $|T_n(v)| < 2$ we let $c_n(v) = 0$. Otherwise, we construct the set of vertices in $V' \setminus \tilde{V}'_n$ that elements of $T_n(v)$ share as neighbours,

$$W_n^{(0)}(v) = \bigcup_{v' \in T_n(v)} \Gamma_{G'}(v') \setminus \tilde{V}'_n. \tag{A.1}$$

If $W_n^{(0)}(v)$ is empty and $|T_n(v)| > 2$, we consider reduced sets of images of the form $T_n(v) \setminus \{w'\}$, where one element w' is removed from $T_n(v)$, and we define the set of all shared neighbours of each reduced image set that are not in the image of ψ_n:

$$W_n^{(1)}(v) = \bigcup_{w' \in T_n(v)} \left(\bigcap_{v' \in T_n(v) \setminus \{w'\}} \Gamma_{G'}(v') \setminus \tilde{V}'_n \right). \tag{A.2}$$

By construction, the set $W_n^{(1)}(v)$ contains those vertices in $V' \setminus \tilde{V}'_n$ that are shared neighbours of images of neighbours of v, each excluding one such neighbour. If $W_n^{(1)}(v)$ is empty, we analogously define $W_n^{(2)}(v)$ as the set of vertices in $V' \setminus \tilde{V}'_n$ that are shared neighbours of images of neighbours of v, each excluding two such neighbours. If $W_n^{(2)}(v)$ is empty, we continue analogously to define $W_n^{(k)}(v)$ as long as $0 < k < |T_n(v)| - 2$. If (i) $W_n^{(k)}(v), 0 < k < |T_n(v)| - 2$ contains exactly one vertex v', then we consider v' to be a possible match of v and set $c_n(v) = |T_n(v)| - k$ as the number of neighbours of v' that are images of neighbours of v under ψ_n. If no unique v' can be

found, we set $c_n(v) = 0$.

If the maximal conserved connectivity $c_n(v)$ among all vertices $v \in \tilde{V} \setminus \tilde{V}_n$ is greater than two, the algorithm attempts to find a vertex v and its unique possible match v' such that $c_n(v)$ is maximal and such that the number of neighbours of v' in $\tilde{V}'_n$ that are not images of neighbours of v is less than $c_n(v) - 2$. If no such vertex can be found, the algorithm attempts to find a vertex and its possible image such that this number of gained gained mapped neighbours is less than $c_n(v) - 1$. We then let $\tilde{V}_{n+1} = \tilde{V}_n \cup \{v\}$, $\tilde{V}'_n = \tilde{V}'_n \cup \{v'\}$ and define a new tracking mapping $\psi_{n+1} : \tilde{V}_{n+1} \to \tilde{V}'_{n+1}$ to be the extension of ψ_n for which $\psi_{n+1}(v) = v'$. The algorithm then increments n and attempts to find a next match v and v' with maximal conserved connectivity.

As soon as the maximal conserved connectivity $c_n(v)$ among all vertices $v \in \tilde{V} \setminus \tilde{V}_n$ is less than two or if no new pair v and v' can be identified, the algorithm halts.

A.3.3 Resolving division events

If a cell divides between consecutive frames, then the tracking mapping ψ_n we have constructed thus far may incorrectly map the mother cell with one of its daughter cells (Figure 6.4). To address this issue, we construct a tracking mapping ψ_{track} in which incorrectly tracked mother cells are removed. To resolve division events, we first identify *boundary vertices* to be those vertices $v \in V$ whose polygon number and degree differ. This corresponds to cells that are at the physical boundary of the sheet, where polygon number and network degree do not coincide (Figure A.1A). We then identify all connected sets of vertices $M' \subseteq V' \setminus \tilde{V}'_n$ that satisfy $\Gamma'_G(M') \subseteq \tilde{V}'_n$ and that contain no boundary vertices of V'. Each such set M' corresponds to one division event, and in the following we treat each M' individually.

For each M', we define $S_{M,1} = \psi_n^{-1}(\Gamma'_G(M'))$ to be the set of inverse images of the mapped neighbours of M' under ψ_n. Next, we identify the set $S_{\text{border}} \subseteq S_{M,1}$ of potential bordering cells of the division, i.e. cells that are adjacent to the division, by

finding those vertices $v \in S_{M,1}$ that gain an edge under the tracking mapping ψ_n:

$$S_{\text{border}} = \{v \in S_{M,1} : p_{G'}(\psi_n(v)) = p_G(v) + 1\}. \tag{A.3}$$

We also identify the set S_{mother} of potential mother cells by finding any shared neighbours of potential bordering cells:

$$S_{\text{mother}} = \bigcap_{v \in S_{\text{border}}} \Gamma_G(v). \tag{A.4}$$

Based on the sets S_{border} and S_{mother} we decide which cells are the mother and daughter cells of the division event, distinguishing between the following cases:

(i) If S_{mother} contains exactly one vertex, then this is identified as the mother cell of the division, and M' must contain exactly two vertices, which are identified as the daughter cells. In this case, neither the mother nor daughter cells are three- or four-sided.

(ii) If $S_{\text{mother}} = \emptyset$, then one of the daughter cells must be three-sided (Figure 6.4C). In this case, a geometry-inferred selection of mother and daughter cells is required. To this end, we define a set of potential daughter cells,

$$S'_{\text{daughter}} = \psi_n(S_{\text{border}}) \bigcup \left(\bigcap_{v' \in \psi_n(S_{\text{border}})} \Gamma_{G'}(v') \right), \tag{A.5}$$

that contains the images of the potential bordering cells and all shared neighbours of these images in V'. Next, we find a definite daughter cell as an element $v' \in S_{\text{daughter}}$ that is three-sided ($p_{G'}(v) = 3$). The geometry-inferred selection of the second daughter cell proceeds as follows. For each $w' \in S'_{\text{daughter}} \setminus \{v'\}$, we construct the *geometrically merged cell* of v' and w' by removing the edge between the polygons that corresponds to v' and w' in the segmentation of the microscopy video frame from which the graph G' was generated, as well as the

cell junctions where three or more cells meet at the end of this edge. We then calculate the distance of the centroid of the geometrically merged cell to the centroid of the cell associated with vertex $\psi_n^{-1}(w')$. The vertex w' for which this distance is minimal is identified as the second daughter cell, and the mother cell is identified as its inverse image under ψ_n.

(iii) If S_{mother} contains more than one vertex, then we define a set of potential daughter cells as any shared neighbours of images of the potential bordering cells

$$S'_{\text{daughter}} = \bigcap_{v' \in \psi_n(S_{\text{border}})} \Gamma_{G'}(v'). \tag{A.6}$$

If S'_{daughter} contains exactly four vertices, then the mother cell and both daughter cells are four-sided, and the mother cell can be identified as the single vertex in the set $S_{M,2}$, which we define as the set of cells which are shared neighbours of all cells in $S_{M,1}$ (the inverse images of neighbours of the division), and which are not in the domain of ψ_n, i.e.

$$S_{M,2} = \bigcap_{v \in S_{M,1}} \Gamma_G(v) \setminus \tilde{V}_n. \tag{A.7}$$

The daughter cells correspond to the only two vertices in M'.

If S'_{daughter} contains exactly three vertices, then one of the daughter cells is four-sided, and we identify this cell as the definite daughter cell of the division v', i.e. we identify $v' \in S'_{\text{daughter}} : p_{G'}(v') = 4$. In this case, geometry-inferred selection of the second daughter cell is required, and we achieve this in a similar way to that described for three-sided daughter cells above. For each cell $w' \in S'_{\text{daughter}} \setminus \{v'\}$, we construct the merged cell of v' and w', and calculate the distance of its centroid to the centroid of $\psi_n^{-1}(w')$. The cell $w' \in S'_{\text{daughter}} \setminus \{v'\}$ for which this distance is smallest is identified as the second daughter cell. Since in this case S_{mother} contains more than one cell, S'_{daughter} must contain at least

three cells[10].

Once each set M' has been inspected and the associated division event has been resolved by identifying the mother and daughter cells, we construct a tracking mapping in which any incorrectly tracked mother cells are removed. To this end, we define the set of all mother cells for which geometry-inferred selection has been used as S_{geo}, and we construct a final tracking mapping $\psi_{\text{track}} : \tilde{V}_n \setminus S_{\text{geo}} \to \tilde{V}'_n \setminus \psi_n(S_{\text{geo}})$ such that $\psi_{\text{track}}(v) = \psi_n(v)\ \forall v \in \tilde{V}_n \setminus \psi_n(S_{\text{geo}})$.

In general, the division resolution step may incorrectly track cells in cases where there is a cell neighbour exchange next to the division, or if there are two adjacent divisions between frames. For example, if each of the bordering cells, i.e. the cells adjacent to the division, were to undergo a neighbour exchange in which they lose an edge between images, then our algorithm would fail to correctly resolve the division event.

A.3.4 Resolving remaining events

At this stage, the tracking algorithm for the two consecutive time frames is completed, and it is straightforward to identify cell neighbour exchanges by finding any cells that have changed their polygon number from one frame to the next. Cell removal events correspond to any vertices $v \in V$ that are not in the domain of ψ_{track}, for which $\Gamma_G(v) \subseteq V_{\text{track}}$, and that do not correspond to mother cells of a division event.

A.4 Computational implementation

We use Krissinel's MCS finding algorithm (Krissinel and Henrick, 2004) to find all RLMs and ELMs in the above steps. This algorithm will always halt eventually. In

[10] If S'_{daughter} contains more than four cells, then our algorithm fails; however, this was never encountered in our test cases.

particular, since the domains on which the RLMs and ELMs are calculated only contain extended neighbourhoods of individual cells, the MCS finding does not pose computational barriers. We adapt the procedure for MCS finding proposed in (Krissinel and Henrick, 2004) in two ways: (i) whenever a next vertex is considered for mapping, we pick a vertex that is adjacent to already mapped cells, hence the adapted algorithm only finds connected subgraphs; (ii) since the RLMs and ELMs are small, we do not implement subgraph-size dependent conditions to interrupt the search early.

When finding the initial mapping, for any two possible matches ELMs are first calculated by considering nearest neighbours only rather than extended neighbourhoods. Once the neighbourhoods[11] of two matching vertices are found to be isomorphic, the extended neighbourhood is considered. This step reduces the time that is needed to find the initial match. During post-processing, the conserved connectivity is not calculated for each vertex. Instead, vertices with maximal numbers of neighbours in the support of ψ_n are observed first and, once the maximal conserved connectivity and the set of corresponding vertices has been identified, no further conserved connectivities are calculated. This reduces the computational cost of the post-processing algorithm.

In the computational implementation of the tracking algorithm we use a further vertex-label $c_G : V \to \mathbb{N}$, which we call the *cell identifier*. In practice, integer identifiers for a given vertex v arise naturally in the segmentation step. Cell identifiers allow us to easily identify vertices and relate them to a cell in a given image independent of how they are stored in the graph structure.

[11]The set of adjacent vertices, $\Gamma_G(v) = \{w \in V : \{v, w\} \in E\}$ is called the *neighbourhood* of v, so the degree of v is $|\Gamma_G(v)|$. We define the neighbourhood of a subset $V' \subseteq V$ to be $\Gamma_G(V') = \{w \in V \setminus V' : \exists\, v \in V' \text{ with } d(w, v) = 1\}$.

Bibliography

M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides et al. The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461):2185–2195, 2000.

T. Aegerter-Wilmsen, M. B. Heimlicher, A. C. Smith, P. B. de Reuille, R. S. Smith, C. M. Aegerter et al. Integrating force-sensing and signaling pathways in a model for the regulation of wing imaginal disc size. *Development*, 139(17):3221–3231, 2012.

T. Aegerter-Wilmsen, A. C. Smith, A. J. Christen, C. M. Aegerter, E. Hafen, and K. Basler. Exploring the effects of mechanical feedback on epithelial topology. *Development*, 137(3):499–506, 2010.

B. B. Aigouy, R. Farhadifar, D. B. Staple, A. Sagner, J. C. Röper, F. Jülicher et al. Cell flow reorients the axis of planar polarity in the wing epithelium of *Drosophila*. *Cell*, 142(5):773–786, 2010.

Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering*, 57(4):841–852, 2010.

A. E. Aladağ and C. Erten. SPINAL: scalable protein interaction network alignment. *Bioinformatics*, 29(7):917–924, 2013.

S. Aldaz and L. M. Escudero. Imaginal discs. *Current Biology*, 20(10):R429–R431, 2010.

S. Aldaz, L. M. Escudero, and M. Freeman. Live imaging of *Drosophila* imaginal disc development. *Proceedings of the National Academy of Sciences of the United States of America*, 107(32):14217–14222, 2010.

M. Aliee, J.-C. Röper, K. P. Landsberg, C. Pentzold, T. J. Widmann, F. Jülicher et al. Physical mechanisms shaping the *Drosophila* dorsoventral compartment boundary. *Current Biology*, 22(11):967–976, 2012.

A. A. Aly, S. Bin Deris, and N. Zaki. Intelligent algorithms for cell tracking and image segmentation. *International Journal of Computer Science and Information Technology*, 6(5):21–37, 2014.

F. Amat, W. Lemon, D. P. Mossing, K. McDole, Y. Wan, K. Branson et al. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. *Nature Methods*, 11(9):951–958, 2014.

A. Ambrosini, M. Gracia, A. Proag, M. Rayer, B. Monier, and M. Suzanne. Apoptotic forces in tissue morphogenesis. *Mechanisms of Development*, 144:33–42, 2017.

A. R. Anderson, A. M. Weaver, P. T. Cummings, and V. Quaranta. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell*, 127(5):905–915, 2006.

S. J. Arnold and E. J. Robertson. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nature Reviews Molecular Cell Biology*, 10(2):91–103, 2009.

M. Ashburner. *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Press, 2nd edition, 2011.

K. Atwell, Z. Qin, D. Gavaghan, H. Kugler, E. J. Hubbard, and J. M. Osborne. Mechano-logical model of *C. elegans* germ line suggests feedback on the cell cycle. *Development*, 142(22):3902–3911, 2015.

M. A. Beaumont. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41(1):379–406, 2010.

M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.

Y. Bellaïche, F. Bosveld, F. Graner, K. Mikula, M. Remešíková, and M. Smíšek. New robust algorithm for tracking cells in videos of *Drosophila* morphogenesis based on finding an ideal path in segmented spatio-temporal cellular structures. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6609–6612. 2011.

D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning. Energy barriers and cell migration in densely packed tissues. *Soft Matter*, 10(12):1885–1890, 2014.

G. Binnig, C. F. Quate, and C. Gerber. Atomic force microscope. *Physical Review Letters*, 56(9):930–933, 1986.

J. T. Blankenship, S. T. Backovic, J. S. Sanny, O. Weitz, and J. A. Zallen. Multicellular rosette formation links planar cell polarity to tissue morphogenesis. *Developmental Cell*, 11(4):459–470, 2006.

F. Bosveld, O. Markova, B. Guirao, C. Martin, Z. Wang, A. Pierre et al. Epithelial tricellular junctions act as interphase cell shape sensors to orient mitosis. *Nature*, 530(7591):495–498, 2016.

G. W. Brodland. Computational modeling of cell sorting, tissue engulfment, and related phenomena: A review. *Applied Mechanics Reviews*, 57(1):47–76, 2004.

G. W. Brodland, D. Viens, and J. H. Veldhuis. A new cell-based FE model for the mechanics of embryonic epithelia. *Computer Methods in Biomechanics and Biomedical Engineering*, 10(2):121–128, 2007.

A. Buchmann, M. Alber, and J. J. Zartman. Sizing it up: The mechanical feedback hypothesis of organ growth regulation. *Seminars in Cell and Developmental Biology*, 35:73–81, 2014.

L. C. Butler, G. B. Blanchard, A. J. Kabla, N. J. Lawrence, D. P. Welchman, L. Mahadevan et al. Cell shape changes indicate a role for extrinsic tensile forces in *Drosophila* germ-band extension. *Nature Cell Biology*, 11(7):859–864, 2009.

E. O. Buzbas and N. A. Rosenberg. AABC: Approximate approximate Bayesian computation for inference in population-genetic models. *Theoretical Population Biology*, 99:31–42, 2015.

J. A. Campos-Ortéga and V. Hartenstein. *The Embryonic Development of Drosophila melanogaster*. Springer, 1997.

O. Canela-Xandri, F. Sagués, J. Casademunt, and J. Buceta. Dynamics and mechanical stability of the developing dorsoventral organizer of the wing imaginal disc. *PLoS Computational Biology*, 7(9):e1002153, 2011.

M. Chalfie, Y. Tu, G. Euskirchen, W. Ward, and D. Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805, 1994.

N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong et al. Objective comparison of particle tracking methods. *Nature Methods*, 11(3):281–289, 2014.

K. K. Chiou, L. Hufnagel, and B. I. Shraiman. Mechanical stress inference for two dimensional cell arrays. *PLoS Computational Biology*, 8(5):e1002512, 2012.

R. Cilla, V. Mechery, B. Hernandez de Madrid, S. Del Signore, I. Dotu, and V. Hatini. Segmentation and tracking of adherens junctions in 3D for the analysis of epithelial tissue morphogenesis. *PLoS Computational Biology*, 11(4):e1004124, 2015.

G. Ciriello, M. Mina, P. H. Guzzi, M. Cannataro, and C. Guerra. AlignNemo: A local network alignment method to integrate homology and topology. *PLoS ONE*, 7(6):e38107, 2012.

A.-K. Classen, K. I. Anderson, E. Marois, and S. Eaton. Hexagonal packing of *Drosophila* wing epithelial cells by the planar cell polarity pathway. *Developmental Cell*, 9(6):805–817, 2005.

C. Collinet, M. Rauzi, P.-F. Lenne, and T. Lecuit. Local and tissue-scale forces drive oriented junction growth during tissue extension. *Nature Cell Biology*, 17(10):1247–1258, 2015.

F. R. Cooper, R. E. Baker, and A. G. Fletcher. Numerical analysis of the immersed boundary method for cell-based simulation. *bioRxiv*, 2016. DOI:10.1101/071423.

F. Crick. Diffusion in embryogenesis. *Nature*, 225(5231):420–422, 1970.

M. von Dassow and L. A. Davidson. Physics and the canalization of morphogenesis: a grand challenge in organismal biology. *Physical Biology*, 8(4):045002, 2011.

E. Davidson and M. Levin. Gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14):4935, 2005.

Y. Davit, J. M. Osborne, H. M. Byrne, D. Gavaghan, and J. Pitt-Francis. Validity of the Cauchy-Born rule applied to discrete cellular-scale models of biological tissues. *Physical Review E*, 87(4):042724, 2013.

D. Drasdo and S. Höhme. A single-cell-based model of tumor growth *in vitro*: monolayers and spheroids. *Physical Biology*, 2(3):133–147, 2005.

J. B. Duffy. GAL4 system in *Drosophila*: A fly geneticist's swiss army knife. *Genesis*, 34(1-2):1–15, 2002.

M. Eiraku, N. Takata, H. Ishibashi, M. Kawada, E. Sakakura, S. Okuda et al. Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature*, 472(7341):51–56, 2011.

G. T. Eisenhoffer, P. D. Loftus, M. Yoshigi, H. Otsuna, C.-B. Chien, P. A. Morcos et al. Crowding induces live cell extrusion to maintain homeostatic cell numbers in epithelia. *Nature*, 484(7395):546–549, 2012.

D. A. Elliott and A. H. Brand. The GAL4 system. In *Methods in Molecular Biology*, volume 420, pages 79–95. Humana Press, Totowa, NJ, 2008.

G. Ermentrout and L. Edelstein-Keshet. Cellular automata approaches to biological modeling. *Journal of Theoretical Biology*, 160(1):97–133, 1993.

L. M. Escudero, L. da F. Costa, A. Kicheva, J. Briscoe, M. Freeman, and M. M. Babu. Epithelial organisation revealed by a network of cellular contacts. *Nature Communications*, 2:526, 2011.

R. Etournay, M. Popović, M. Merkel, A. Nandi, C. Blasse, B. Aigouy et al. Interplay of cell dynamics and epithelial tension during morphogenesis of the *Drosophila* pupal wing. *eLife*, 4:e07090, 2015.

R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, 17(24):2095–2104, 2007.

A. G. Fletcher, J. M. Osborne, P. K. Maini, and D. J. Gavaghan. Implementing vertex dynamics models of cell populations in biology within a consistent computational framework. *Progress in Biophysics and Molecular Biology*, 113(2):299–326, 2013.

A. G. Fletcher, M. Osterfield, R. E. Baker, and S. Y. Shvartsman. Vertex models of epithelial morphogenesis. *Biophysical Journal*, 106(11):2291–2304, 2014.

V. E. Foe. Mitotic domains reveal early commitment of cells in *Drosophila* embryos. *Development*, 107(1):1–22, 1989.

Y. Gavrieli. Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation. *The Journal of Cell Biology*, 119(3):493–501, 1992.

C. Gayrard and N. Borghi. FRET-based molecular tension microscopy. *Methods*, 94:33–42, 2016.

M. C. Gibson, A. B. Patel, R. Nagpal, and N. Perrimon. The emergence of geometric order in proliferating metazoan epithelia. *Nature*, 442(7106):1038–1041, 2006.

W. T. Gibson, J. H. Veldhuis, B. Rubinstein, H. N. Cartwright, N. Perrimon, G. W. Brodland et al. Control of the mitotic cleavage plane by local epithelial topology. *Cell*, 144(3):427–438, 2011.

J. A. Glazier and F. Graner. Simulation of the differential adhesion driven rearrangement of biological cells. *Physical Review E*, 47(3):2128–2154, 1993.

N. Gorfinkiel, G. B. Blanchard, R. J. Adams, and A. Martinez Arias. Mechanical control of global cell behaviour during dorsal closure in *Drosophila*. *Development*, 136(11):1889–1898, 2009.

C. Guillot and T. Lecuit. Adhesion disengagement uncouples intrinsic and extrinsic forces to drive cytokinesis in epithelial tissues. *Developmental Cell*, 24(3):227–241, 2013a.

C. Guillot and T. Lecuit. Mechanics of epithelial tissue homeostasis and morphogenesis. *Science*, 340(6137):1185–1189, 2013b.

A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15. Pasadena, CA USA, 2008.

A. R. Harris, L. Peter, J. Bellis, B. Baum, A. J. Kabla, and G. T. Charras. Characterizing the mechanics of cultured cell monolayers. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41):16449–16454, 2012.

J. U. Harrison and R. E. Baker. An automatic adaptive method to combine summary statistics in approximate Bayesian computation. 2017. ArXiv ID:1703.02341.

C.-P. Heisenberg and Y. Bellaïche. Forces in tissue morphogenesis and patterning. *Cell*, 153(5):948–962, 2013.

D. Heller, A. Hoppe, S. Restrepo, L. Gatti, A. L. Tournier, N. Tapon et al. EpiTools: An open-source image analysis toolkit for quantifying epithelial growth dynamics. *Developmental Cell*, 36(1):103–116, 2016.

R. A. Hoebe, C. H. Van Oven, T. W. J. Gadella, P. B. Dhonukshe, C. J. F. Van Noorden, and E. M. M. Manders. Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. *Nature Biotechnology*, 25(2):249–253, 2007.

W. Hofmeister. Zusätze und Berichtigungen zu den 1851 veröffentlichen Untersuchungengen der Entwicklung höherer Kryptogamen. *Jahrbücher für wissenschaftliche Botanik*, 3:259–293, 1863.

H. Honda. Description of cellular patterns by Dirichlet domains: The two-dimensional case. *Journal of Theoretical Biology*, 72(3):523–543, 1978.

H. Honda and T. Nagai. Cell models lead to understanding of multi-cellular morphogenesis consisting of successive self-construction of cells. *Journal of Biochemistry*, 157(3):129–136, 2015.

L. Hufnagel, A. A. Teleman, H. Rouault, S. M. Cohen, and B. I. Shraiman. On the mechanism of wing size determination in fly development. *Proceedings of the National Academy of Sciences of the United States of America*, 104(10):3835–3840, 2007.

S. C. Hughes and H. M. Krause. Establishment and maintenance of parasegmental compartments. *Development*, 128(7):1109–18, 2001.

D. Iber, S. Tanaka, P. Fried, P. Germann, and D. Menshykau. Simulating tissue morphogenesis and signaling. In *Tissue Morphogenesis: Methods and Protocols*, pages 323–338. 2015.

S. Ishihara and K. Sugimura. Bayesian inference of force dynamics during morphogenesis. *Journal of Theoretical Biology*, 313(0):201–211, 2012.

S. Ishihara, K. Sugimura, S. J. Cox, I. Bonnet, Y. Bellaïche, and F. Graner. Comparative study of non-invasive force and stress inference methods in tissue. *The European Physical Journal E*, 36(4):45, 2013.

Y. Ishimoto and Y. Morishita. Bubbly vertex dynamics: A dynamical and geometrical model for epithelial tissues with curved cell shapes. *Physical Review E*, 90(5):052711, 2014.

J. A. Izaguirre, R. Chaturvedi, C. Huang, T. Cickovski, J. Coffland, G. Thomas et al. CompuCell, a multi-model framework for simulation of morphogenesis. *Bioinformatics*, 20(7):1129–1137, 2004.

H. Jung and P. Marjoram. Choice of summary statistic weights in approximate Bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 10(1):1–23, 2011.

K. Kawasaki, T. Nagai, and K. Nakashima. Vertex models for two-dimensional grain growth. *Philosophical Magazine Part B*, 60(3):399–421, 1989.

S. Kim, M. Cai, and S. Hilgenfeldt. Lewis' law revisited: the role of anisotropy in size-topology correlations. *New Journal of Physics*, 16(1):015024, 2014.

T. Kornberg, I. Sidén, P. O'Farrell, and M. Simon. The *engrailed* locus of *Drosophila*: *in situ* localization of transcripts reveals compartment-specific expression. *Cell*, 40(1):45–53, 1985.

E. B. Krissinel and K. Henrick. Common subgraph isomorphism detection by backtracking search. *Software: Practice and Experience*, 34(6):591–607, 2004.

J. Kursawe, R. E. Baker, and A. G. Fletcher. Impact of implementation choices on quantitative predictions of cell-based computational models. *bioRxiv*, 2016a. DOI:10.1101/092924.

J. Kursawe, R. Bardenet, J. J. Zartman, R. E. Baker, and A. G. Fletcher. Robust cell tracking in epithelial tissues through identification of maximum common subgraphs. *Journal of The Royal Society Interface*, 13(124):20160725, 2016b.

J. Kursawe, P. A. Brodskiy, J. J. Zartman, R. E. Baker, and A. G. Fletcher. Capabilities and limitations of tissue size control through passive mechanical forces. *PLoS Computational Biology*, 11(12):e1004679, 2015.

K. P. Landsberg, R. Farhadifar, J. Ranft, D. Umetsu, T. J. Widmann, T. Bittig et al. Increased cell bond tension governs cell sorting at the *Drosophila* anteroposterior compartment boundary. *Current Biology*, 19(22):1950–1955, 2009.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

Y. Lee, S. Kouvroukoglou, L. McIntire, and K. Zygourakis. A cellular automaton model for the proliferation of migrating contact-inhibited cells. *Biophysical Journal*, 69(4):1284–1298, 1995.

I. M. M. van Leeuwen, G. R. Mirams, A. Walter, A. G. Fletcher, P. J. Murray, J. Osborne et al. An integrative computational model for intestinal tissue renewal. *Cell Proliferation*, 42(5):617–636, 2009.

F. T. Lewis. The correlation between cell division and the shapes and sizes of prismatic cells in the epidermis of *Cucumis*. *The Anatomical Record*, 38(3):341–376, 1928.

Q. Li and J. S. Racine. Density estimation. In *Nonparametric Econometrics*, pages 1–55. Princeton University Press, 2007.

Q. J. Li, T. M. Pazdera, and J. S. Minden. *Drosophila* embryonic pattern repair: how embryos respond to cyclin E-induced ectopic division. *Development*, 126(10):2299–2307, 1999.

Y. Li, H. Naveed, S. Kachalo, L. X. Xu, and J. Liang. Mechanisms of regulating cell topology in proliferating epithelia: Impact of division plane, mechanical forces, and cell memory. *PLoS ONE*, 7(8):e43108, 2012.

K. Liu, S. S. Lienkamp, A. Shindo, J. B. Wallingford, G. Walz, and O. Ronneberger. Optical flow guided cell segmentation and tracking in developing tissue. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 298–301. 2014.

S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

Z. Ma, J. Wang, P. Loskill, N. Huebsch, S. Koo, F. L. Svedlund et al. Self-organizing human cardiac microchambers mediated by geometric confinement. *Nature Communications*, 6:7413, 2015.

R. Magno, V. A. Grieneisen, and A. F. Marée. The biophysical nature of cells: potential cell behaviours revealed by analytical and computational studies of cell surface mechanics. *BMC Biophysics*, 8(1):8, 2015.

Y. Mao, A. L. Tournier, P. A. Bates, J. E. Gale, N. Tapon, and B. J. Thompson. Planar polarization of the atypical myosin Dachs orients cell divisions in *Drosophila*. *Genes and Development*, 25(2):131–136, 2011.

Y. Mao, A. L. Tournier, A. Hoppe, L. Kester, B. J. Thompson, and N. Tapon. Differential proliferation rates generate patterns of mechanical tension that orient tissue growth. *The EMBO Journal*, 32(21):2790–2803, 2013.

A. F. M. Marée, V. A. Grieneisen, and P. Hogeweg. *The Cellular Potts Model and Biophysical Properties of Cells, Tissues and Morphogenesis*, pages 107–136. Birkhäuser Basel, Basel, 2007.

E. Marinari, A. Mehonic, S. Curran, J. Gale, T. Duke, and B. Baum. Live-cell delamination counterbalances epithelial growth to limit tissue overcrowding. *Nature*, 484(7395):542–545, 2012.

P. Marjoram, J. Molitor, V. Plagnol, and S. Tavare. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):15324–8, 2003.

D. N. Mashburn, H. E. Lynch, X. Ma, and M. S. Hutson. Enabling user-guided segmentation and tracking of surface-labeled cells in time-lapse image sets of living tissues. *Cytometry A*, 81A(5):409–418, 2012.

M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra et al. A benchmark for comparison of cell tracking algorithms. *Bioinformatics*, 30(11):1609–1617, 2014.

M. Mavrakis, R. Rikhy, M. Lilly, and J. Lippincott-Schwartz. Fluorescence imaging techniques for studying *Drosophila* embryo development. In *Current Protocols in Cell Biology*, pages 1–43. John Wiley and Sons, Inc., Hoboken, NJ, USA, 2008.

F. A. Meineke, C. S. Potten, and M. Loeffler. Cell migration and organization in the intestinal crypt using a lattice-free model. *Cell Proliferation*, 34(4):253–266, 2001.

F. Meng, T. M. Suchyna, and F. Sachs. A fluorescence energy transfer-based mechanical stress sensor for specific proteins *in situ*. *FEBS Journal*, 275(12):3072–3087, 2008.

A. Merzouki, O. Malaspinas, and B. Chopard. The mechanical properties of a cell-based numerical model of epithelium. *Soft Matter*, 12(21):4745–4754, 2016.

M. Milan, S. Campuzano, and A. Garcia-Bellido. Cell cycling and patterned cell proliferation in the wing primordium of *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, 93(2):640–645, 1996.

R. E. Miller and E. B. Tadmor. A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. *Modelling and Simulation in Materials Science and Engineering*, 17(5):053001, 2009.

G. R. Mirams, C. J. Arthurs, M. O. Bernabeu, R. Bordas, J. Cooper, A. Corrias et al. Chaste: An open source C++ library for computational physiology and biology. *PLoS Computational Biology*, 9(3):e1002970, 2013.

B. Monier, M. Gettings, G. Gay, T. Mangeat, S. Schott, A. Guarner et al. Apico-basal forces exerted by apoptotic cells drive epithelium folding. *Nature*, 518(7538):245–248, 2015.

B. Monier, A. Pélissier-Monier, A. H. Brand, and B. Sanson. An actomyosin-based barrier inhibits cell mixing at compartmental boundaries in *Drosophila* embryos. *Nature Cell Biology*, 12(1):60–65, 2010.

P. J. Murray, C. M. Edwards, M. J. Tindall, and P. K. Maini. From a discrete to a continuum model of cell dynamics in one dimension. *Physical Review E*, 80(3):24–29, 2009.

T. Nagai and H. Honda. A dynamic cell model for the formation of epithelial tissues. *Philosophical Magazine Part B*, 81(7):699–719, 2001.

T. Nagai and H. Honda. Computer simulation of wound closure in epithelial tissues: Cell-basal-lamina adhesion. *Physical Review E*, 80(6):061903, 2009.

T. Nagai, K. Kawasaki, and K. Nakamura. Vertex dynamics of two-dimensional cellular patterns. *Journal of the Physical Society of Japan*, 57(7):2221–2224, 1988.

R. Namba, T. M. Pazdera, R. L. Cerrone, and J. S. Minden. *Drosophila* embryonic pattern repair: how embryos respond to *bicoid* dosage alteration. *Development*, 124(7):1393–1403, 1997.

C. Narciso, Q. Wu, P. Brodskiy, G. Garston, R. E. Baker, A. G. Fletcher et al. Patterning of wound-induced intercellular Ca^{2+} flashes in a developing epithelium. *Physical Biology*, 12(5):056005, 2015.

A. Nestor-Bergmann, G. Goddard, S. Woolner, and O. Jensen. A vertex-based model relating cell shape and mechanical stress in an epithelium. 2016. ArXiv ID:1611.04744.

T. J. Newman. *Modeling Multicellular Structures Using the Subcellular Element Model*, pages 221–239. Birkhäuser Basel, 2007.

V. Nier, S. Jain, C. T. Lim, S. Ishihara, B. Ladoux, and P. Marcq. Inference of internal stress in a cell monolayer. *Biophysical Journal*, 110(7):1625–1635, 2016.

L. O'Keefe, S. T. Dougan, L. Gabay, E. Raz, B. Z. Shilo, and S. DiNardo. Spitz and wingless, emanating from distinct borders, cooperate to establish cell fate across the *engrailed* domain in the *Drosophila* epidermis. *Development*, 124(23):4837–45, 1997.

S. Okuda, Y. Inoue, and T. Adachi. Three-dimensional vertex model for simulating multicellular morphogenesis. *Biophysics and Physicobiology*, 12(0):13–20, 2015a.

S. Okuda, Y. Inoue, T. Watanabe, and T. Adachi. Coupling intercellular molecular signalling with multicellular deformation for simulating three-dimensional tissue morphogenesis. *Interface Focus*, 5(2):20140095, 2015b.

J. M. Osborne, M. O. Bernabeu, M. Bruna, B. Calderhead, J. Cooper, N. Dalchau et al. Ten simple rules for effective computational research. *PLoS Computational Biology*, 10(3):e1003506, 2014.

J. M. Osborne, A. G. Fletcher, J. M. Pitt-Francis, P. K. Maini, and D. J. Gavaghan. Comparing individual-based approaches to modelling the self-organization of multicellular tissues. *PLoS Computational Biology*, 13(2):e1005387, 2017.

D. Padfield, J. Rittscher, and B. Roysam. Coupled minimum-cost flow cell tracking for high-throughput quantitative analysis. *Medical Image Analysis*, 15(4):650–668, 2011.

Y. Pan, I. Heemskerk, C. Ibar, B. I. Shraiman, and K. D. Irvine. Differential growth triggers mechanical feedback that elevates Hippo signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 113(45):E6974–E6983, 2016.

P. Pantazis and W. Supatto. Advances in whole-embryo imaging: a quantitative transition is underway. *Nature Reviews Molecular Cell Biology*, 15(5):327–339, 2014.

M. Pargett and D. M. Umulis. Quantitative model analysis with diverse biological data: Applications in developmental pattern formation. *Methods*, 62(1):56–67, 2013.

J. Parker. Control of compartment size by an EGF ligand from neighboring cells. *Current Biology*, 16(20):2058–2065, 2006.

R. M. Parton, A. M. Vallés, I. M. Dobbie, and I. Davis. Collection and mounting of *Drosophila* embryos for imaging. *Cold Spring Harbor Protocols*, 2010(4), 2010.

A. B. Patel, W. T. Gibson, M. C. Gibson, and R. Nagpal. Modeling and inferring cleavage patterns in proliferating epithelia. *PLoS Computational Biology*, 5(6):e1000412, 2009.

T. M. Pazdera, P. Janardhan, and J. S. Minden. Patterned epidermal cell death in wild-type and segment polarity mutant *Drosophila* embryos. *Development*, 125(17):3427–3436, 1998.

N. Perrimon, J.-Q. Ni, and L. Perkins. *In vivo* RNAi: Today and tomorrow. *Cold Spring Harbor Perspectives in Biology*, 2(8):a003640, 2010.

D. Prangle. Adapting the ABC distance function. *Bayesian Analysis*, 12(1):289–309, 2017.

A. Puliafito, L. Hufnagel, P. Neveu, S. Streichan, A. Sigal, D. K. Fygenson et al. Collective and single cell behavior in epithelial contact inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, 109(3):739–744, 2012.

M. Raffel, C. E. Willert, S. Wereley, and J. Kompenhans. *Particle Image Velocimetry: A Practical Guide*. Springer-Verlag Berlin Heidelberg, 2007.

M. Rauzi, P. Verant, T. Lecuit, and P.-F. Lenne. Nature and anisotropy of cortical forces orienting *Drosophila* tissue morphogenesis. *Nature Cell Biology*, 10(12):1401–1410, 2008.

R. P. Ray, A. Matamoro-Vidal, P. S. Ribeiro, N. Tapon, D. Houle, I. Salazar-Ciudad et al. Patterned anchorage to the apical extracellular matrix defines tissue shape in the developing appendages of *Drosophila*. *Developmental Cell*, 34(3):310–322, 2015.

J. W. Raymond and P. Willett. Maximum common subgraph isomorphism algorithms for matching of chemical structures. *Journal of Computer-Aided Molecular Design*, 16(7):521–533, 2002.

K. Rejniak, H. J. Kliman, and L. J. Fauci. A computational model of the mechanics of growth of the villous trophoblast bilayer. *Bulletin of Mathematical Biology*, 66(2):199–232, 2004.

L. Ritsma, S. I. J. Ellenbroek, A. Zomer, H. J. Snippert, F. J. de Sauvage, B. D. Simons et al. Intestinal crypt homeostasis revealed at single-stem-cell level by *in vivo* live imaging. *Nature*, 507(7492):362–365, 2014.

A. L. Robinson. Electron microscope inventors share nobel physics prize. *Science*, 234(4778):821–822, 1986.

K. W. Rogers and A. F. Schier. Morphogen gradients: From generation to interpretation. *Annual Review of Cell and Developmental Biology*, 27(1):377–407, 2011.

A. Sáez, B. Acha, A. Montero-Sánchez, E. Rivas, L. M. Escudero, and C. Serrano. Neuromuscular disease classification system. *Journal of Biomedical Optics*, 18(6):066017, 2013.

N. Saiz, K. M. Williams, V. E. Seshan, and A.-K. Hadjantonakis. Asynchronous fate decisions by single cells collectively ensure consistent lineage composition in the mouse blastocyst. *Nature Communications*, 7:13463, 2016.

G. Salbreux, L. K. Barthel, P. A. Raymond, and D. K. Lubensky. Coupling mechanical deformations and planar cell polarity to create regular patterns in the zebrafish retina. *PLoS Computational Biology*, 8(8):e1002618, 2012.

D. Sánchez-Gutiérrez, M. Tozluoglu, J. D. Barry, A. Pascual, Y. Mao, and L. M. Escudero. Fundamental physical cellular constraints drive self-organization of tissues. *The EMBO Journal*, 35(1):77–88, 2016.

M. Schiegg, P. Hanslovsky, B. X. Kausler, L. Hufnagel, and F. A. Hamprecht. Conservation tracking. *2013 IEEE International Conference on Computer Vision*, pages 2928–2935, 2013.

S. Schilling, M. Willecke, T. Aegerter-Wilmsen, O. A. Cirpka, K. Basler, and C. von Mering. Cell-sorting at the A/P boundary in the *Drosophila* wing primordium: A computational model to consolidate observed non-local effects of Hh signaling. *PLoS Computational Biology*, 7(4):e1002025, 2011.

F. Schöck and N. Perrimon. Molecular mechanisms of epithelial morphogenesis. *Annual Review of Cell and Developmental Biology*, 18(1):463–493, 2002.

U. S. Schwarz and C. M. Dunlop. Developmental biology: A growing role for computer simulations. *Current Biology*, 22(11):R441–R443, 2012.

M. Scianna and L. Preziosi. A node-based version of the cellular Potts model. *Computers in Biology and Medicine*, 76:94–112, 2016.

W. Shen, X. Chen, O. Cormier, D. C.-P. Cheng, B. Reed, and N. Harden. Modulation of morphogenesis by EGFR during dorsal closure in *Drosophila*. *PLoS ONE*, 8(4):e60180, 2013.

R. Sheth, L. Marcon, M. F. Bastida, M. Junco, L. Quintana, R. Dahn et al. Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism. *Science*, 338(6113):1476–1480, 2012.

R. Shields. Further evidence for a random transition in the cell cycle. *Nature*, 273:755–758, 1978.

O. Shimomura. The discovery of aequorin and green fluorescent protein. *Journal of Microscopy*, 217(1):3–15, 2005.

O. Shimomura, F. H. Johnson, and Y. Saiga. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *Journal of Cellular and Comparative Physiology*, 59(3):223–239, 1962.

A. Shirinifard, J. S. Gens, B. L. Zaitlen, N. J. Popławski, M. Swat, and J. A. Glazier. 3D multi-cell simulation of tumor growth and angiogenesis. *PLoS ONE*, 4(10):e7190, 2009.

W. Shou, C. T. Bergstrom, A. K. Chakraborty, and F. K. Skinner. Theory, models and biology. *eLife*, 4:e07158, 2015.

B. I. Shraiman. Mechanical feedback as a possible regulator of tissue growth. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9):3318–3323, 2005.

S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.

A. M. Smith. *Vertex Model Approaches to Epithelial Tissues in Developmental Systems*. D.Phil. thesis, University of Oxford, 2011.

A. M. Smith, R. E. Baker, D. Kay, and P. K. Maini. Incorporating chemical signalling factors into cell-based models of growing epithelial tissues. *Journal of Mathematical Biology*, 65(3):441–463, 2012.

J. A. Smith and L. Martin. Do cells cycle? *Proceedings of the National Academy of Sciences of the United States of America*, 70(4):1263–1267, 1973.

M. H. L. Snow and P. P. L. Tam. Is compensatory growth a complicating factor in mouse teratology? *Nature*, 279(5713):555–557, 1979.

C. Sommer, C. Straehle, U. Kothe, and F. A. Hamprecht. Ilastik: Interactive learning and segmentation toolkit. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 230–233. 2011.

Z. Song. DCP-1, a *Drosophila* cell death protease essential for development. *Science*, 275(5299):536–540, 1997.

M. A. Spencer, Z. Jabeen, and D. K. Lubensky. Vertex stability and topological transitions in vertex models of foams and epithelia. *The European Physical Journal E*, 40(1):2, 2017.

S. Srinivas. The anterior visceral endoderm - turning heads. *Genesis*, 44(11):565–572, 2006.

D. B. Staple, R. Farhadifar, J. C. Röper, B. Aigouy, S. Eaton, and F. Jülicher. Mechanics and remodelling of cell packings in epithelia. *The European Physical Journal E*, 33(2):117–127, 2010.

D. J. Stephens and V. J. Allan. Light microscopy techniques for live cell imaging. *Science*, 300(5616):82–86, 2003.

K. Sugimura and S. Ishihara. The mechanical anisotropy in a tissue promotes ordering in hexagonal cell packing. *Development*, 140(19):4091–4101, 2013.

D. Szüts, M. Freeman, and M. Bienz. Antagonism between EGFR and wingless signalling in the larval cuticle of *Drosophila*. *Development*, 124(16):3209–19, 1997.

R. J. Tetley, G. B. Blanchard, A. G. Fletcher, R. J. Adams, and B. Sanson. Unipolar distributions of junctional myosin II identify cell stripe boundaries that drive cell intercalation throughout *Drosophila* axis extension. *eLife*, 5:1–35, 2016.

R. E. Thach and S. S. Thach. Damage to biological samples caused by the electron beam during electron microscopy. *Biophysical Journal*, 11(2):204–210, 1971.

C. J. Tomlin and J. D. Axelrod. Biology by numbers: mathematical modelling in developmental biology. *Nature Reviews Genetics*, 8(5):331–340, 2007.

C. E. Torres, M. Emelianenko, D. Golovaty, D. Kinderlehrer, and S. Ta'asan. Numerical analysis of the vertex models for simulating grain boundary networks. *SIAM Journal on Applied Mathematics*, 75(2):762–786, 2015.

G. Trichas, A. M. Smith, N. White, V. Wilkins, T. Watanabe, A. Moore et al. Multicellular rosettes in the mouse visceral endoderm facilitate the ordered migration of anterior visceral endoderm cells. *PLoS Biology*, 10(2):e1001256, 2012.

T. V. Truong and W. Supatto. Toward high-content/high-throughput imaging and analysis of embryonic morphogenesis. *Genesis*, 49(7):555–569, 2011.

A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 237(641):37–72, 1952.

J. R. Ullmann. An algorithm for subgraph isomorphism. *Journal of the ACM*, 23(1):31–42, 1976.

S. Vakulenko, Manu, J. Reinitz, and O. Radulescu. Size regulation in the segmentation of *Drosophila*: Interacting interfaces between localized domains of gene expression ensure robust spatial patterning. *Physical Review Letters*, 103(16):168102, 2009.

J. A. Vernon and J. Butsch. Effect of tetraploidy on learning and retention in the salamander. *Science*, 125(3256):1033–1034, 1957.

R. M. A. Vroomans, P. Hogeweg, and K. H. W. J. ten Tusscher. Segment-specific adhesion as a driver of convergent extension. *PLoS Computational Biology*, 11(2):e1004092, 2015.

E. Wait, M. Winter, C. Bjornsson, E. Kokovay, Y. Wang, S. Goderie et al. Visualization and correction of automated segmentation, tracking and lineaging from 5-D stem cell image sequences. *BMC Bioinformatics*, 15(1):328, 2014.

D. C. Walker, G. Hill, S. M. Wood, R. H. Smallwood, and J. Southgate. Agent-based computational modeling of wounded epithelial cell monolayers. *IEEE Transactions on Nanobioscience*, 3(3):153–63, 2004.

Q. Wang, J. Niemi, C.-M. Tan, L. You, and M. West. Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. *Cytometry A*, 77A(1):101–110, 2010.

R. Warn and R. Magrath. F-actin distribution during the cellularization of the *Drosophila* embryo visualized with FL-phalloidin. *Experimental Cell Research*, 143(1):103–114, 1983.

O. Wartlick, P. Mumcu, F. Jülicher, and M. Gonzalez-Gaitan. Understanding morphogenetic growth control - lessons from flies. *Nature Reviews Molecular Cell Biology*, 12(9):594–604, 2011a.

O. Wartlick, P. Mumcu, A. Kicheva, T. Bittig, C. Seum, F. Jülicher et al. Dynamics of Dpp signaling and proliferation control. *Science*, 331(6021):1154–1159, 2011b.

E. Weinan and B. Engquist. Multiscale modeling and computation. *Notices of the AMS*, 50(9):1062–1070, 2003.

M. Weliky and G. Oster. The mechanical basis of cell rearrangement. I. Epithelial morphogenesis during *Fundulus* epiboly. *Development*, 109(2):373–386, 1990.

R. J. Wilson. *An Introduction to Graph Theory*. Prentice Hall, 2010.

M. Winter, E. Wait, B. Roysam, S. K. Goderie, R. A. N. Ali, E. Kokovay et al. Vertebrate neural stem cell segmentation, tracking and lineaging with validation and editing. *Nature Protocols*, 6(12):1942–1952, 2011.

L. Wolpert. Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology*, 25(1):1–47, 1969.

L. Wolpert. *Developmental Biology: A Very Short Introduction*. Oxford University Press, 2011.

W. Wood and A. Jacinto. *Imaging Cell Movement During Dorsal Closure in Drosophila Embryos*, pages 203–210. Humana Press, New Jersey, 2005.

T. P. J. Wyatt, A. R. Harris, M. Lam, Q. Cheng, J. Bellis, A. Dimitracopoulos et al. Emergence of homeostatic epithelial packing and stress dissipation through divisions oriented along the long cell axis. *Proceedings of the National Academy of Sciences*, 112(18):5726–5731, 2015.

F. Xiong and S. G. Megason. Abstracting the principles of development using imaging and modeling. *Integrative Biology*, 7(6):633–642, 2015.

G.-K. Xu, Y. Liu, and B. Li. How do changes at the cell level affect the mechanical properties of epithelial monolayers? *Soft Matter*, 11(45):8782–8788, 2015a.

G. K. Xu, Y. Liu, and Z. Zheng. Oriented cell division affects the global stress and cell packing geometry of a monolayer under stretch. *Journal of Biomechanics*, 49(3):401–407, 2015b.

S. Youssef, S. Gude, and J. O. Rädler. Automated tracking in live-cell time-lapse movies. *Integrative Biology*, 3(11):1095, 2011.

J. C. Yu and R. Fernandez-Gonzalez. Quantitative modelling of epithelial morphogenesis: integrating cell mechanics and molecular dynamics. *Seminars in Cell and Developmental Biology*, 2016.

J. J. Zartman, J. S. Kanodia, L. S. Cheung, and S. Y. Shvartsman. Feedback control of the EGFR signaling gradient: superposition of domain-splitting events in *Drosophila* oogenesis. *Development*, 136(17):2903–2911, 2009.

R. S. Zou and C. Tomasi. Deformable graph model for tracking epithelial cell sheets in fluorescence microscopy. *IEEE Transactions on Medical Imaging*, 35(7):1625–1635, 2016.