UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Multiscale Models for Developing Tissues or Organs in Biological Systems

Permalink

https://escholarship.org/uc/item/6rc1t03c

Author

Toomey, Alysha N

Publication Date

2024

Copyright Information

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Multiscale Models for Developing Tissues or Organs in Biological Systems

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Mathematics

by

Alysha N Toomey

March 2024

Dissertation Committee:

    Dr. Weitao Chen, Co-Chairperson
    Dr. Mark Alber, Co-Chairperson
    Dr. Yiwei Wang

The Dissertation of Alysha N Toomey is approved:

                                    Committee Co-Chairperson

                                    Committee Co-Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Multiscale Models for Developing Tissues or Organs in Biological Systems

by

Alysha N Toomey

Doctor of Philosophy, Graduate Program in Mathematics
University of California, Riverside, March 2024
Dr. Weitao Chen, Co-Chairperson
Dr. Mark Alber, Co-Chairperson

This thesis consists of two parts.

In the first part, we develop a model that will be used to investigate pavement cell morphogenesis. Pavement cells, the leaf epidermal cells in the *Arabidopsis thaliana* plant, have complex jigsaw puzzle piece shapes. The formation of these interlocking shapes relies on mechanical, chemical, and cell to cell signals at different scales. Because of this, pavement cells are an interesting model system used to study the mechanisms involved in cell morphogenesis in plant tissue. Utilizing the local level set method, biochemical dynamics on moving cell boundaries are captured. By incorporating cell-cell adhesion, the model is expanded to a multicellular framework that can be used to investigate the components involved in establishing these intricate cell shapes.

In the second part, we use a combination of experimental and modeling techniques to study new and existing regulations in the Dpp-Rho1-Cdc42 network in the *Drosophila* wing disc. During organogenesis in the wing disc, the regulation of epithelial cell height and curvature is crucial in developing correct tissue shapes. This requires the interplay between mechanical forces and morphogen-mechanogen pathways, at both the cell and tissue levels. Morphogens, such as Decapentaplegic (Dpp), regulate cell growth and division, as well as mechanogen activity. Mechanogens, such as RhoGTPases, are small diffusible molecules

that regulate mechanical components, such as actin and myosin, to coordinate cell shape and tissue geometry. Even though the effect of morphogens in regulating mechanogens is critical for proper tissue formation, insufficient work has been done to understand this in the context of epithelial organogenesis. In this study, a combination of experimental and mathematical modeling approaches are used to study the linkage between Dpp, Rho1, and Cdc42 in the wing disc. By using experiments, new regulations between Dpp and Cdc42 have been identified, as well as the interaction between Cdc42 and Rho1. A mathematical model is developed by using a system of reaction-diffusion equations to model Dpp, Rho1, and Cdc42 dynamics, as well as the newly identified regulations. We utilize Bayesian optimization to explore this model and to investigate the robustness of the proposed networks.

Contents

List of Figures

List of Tables

Chapter 1

Multiscale Model for Morphogenesis of Pavement Cells in Plant Leaves

1.1   Preface

This work was done in collaboration with Dr. Zhenbiao Yang, Dr. Xue Pan, Dr. Betul Senay-Aras, and Dr. Weitao Chen. Dr. Yang and Dr. Pan both contributed the biological knowledge behind the development of this model. Dr. Senay-Aras assisted in the development of the multicellular model.

1.2   Biological Background

Pavement cells (PCs), present in the leaves of *Arabidopsis thaliana*, are a type of epidermal cell that form the outermost layer of the leaf epidermis. PCs are thin, flat and tightly packed together, creating a continuous layer that covers the leaf surface. At early stages of development these cells take on a polygonal shape with their internal components randomly organized. However, throughout morphogenesis the cells develop into complex puzzle-piece

shapes with intricate polarities (i.e., spatial organization of various proteins; reference Fig 1.1). The intricate and dynamic control of cell polarity formation, both within individual cells and in coordination with neighboring cells, is essential for the development of the interlocking PC shape [11]. The shape is regulated by chemical signals, cell wall components, mechanical stress, and cell-to-cell signals, thus making PCs an ideal system to study both intracellular and intercellular mechanisms.

PC morphogenesis can be broken down into three broad stages: polarity establishment, lobe/indentation initiation, and lobe/indentation expansion [11]. Many of the components involved in these stages have been investigated. Although this is the case, the connection between all of the known components and hence the true driving force behind PC morphogenesis remains elusive. There is a divide amongst researchers as to which component, chemical signaling or mechanical forces, drives the cell morphological change.

To start, many groups argue that the chemical signaling network is the driving force behind PC morphogenesis. There are three main components known on this side: auxin, ROP2, and ROP6. Auxin is a highly diffusible, tissue-level molecule that lives in the cell wall. ROP2 and ROP6 are intracellular signals that are localized at the lobes and indentations, respectively, of each cell. Auxin activates the ROP2 and ROP6 pathways via a cell-surface auxin receptor. The ROP2 pathway polarizes PIN1 to the tips of the lobes, where it exports auxin to locally amplify extracellular auxin through the ROP2 positive feedback loop. The auxin can diffuse across the cell wall to activate the ROP6 pathway. ROP polarity is maintained and the formation of lobes and indentations in the cell walls is promoted through the mutual inhibition of the ROP2 and ROP6 pathways [8].

On the other hand, there are many groups that argue mechanical properties are the driving force. For example, Sapala *et al.* suggest that mechanical stress is a key driver of the shape change by focusing on the role of microtubulues in producing the indentations within cells [6]. They utilize a finite element model to argue that microtubulues are utilized to resist stress as the tissue grows, which produces small indentations in the cell. These indentations accumulate more stress which further recruits microtubulues, exaggerating the

indentations. Another group, Higaki *et al.*, also focuses on the mechanical aspects of the system, in addition to some protein dynamics and again uses a finite element model to test their hypothesis [7]. This study suggests that cell growth produces tension in the cell wall, which in turn buckles, contributing to the production of the lobes and indentations. A third study worth noting is the use of a phase field model to suggest that the interdigitated pattern arises as a result of cell wall remodeling [9]. Higaki *et al.* reproduce the maintenance of cell wall thickness and the formation of the jigsaw puzzle shapes. While all of these studies contribute significant insight into mechanisms involved in PC morphogenesis, they neglect one important piece of the puzzle: how the cells initiate that break in symmetry at the beginning of the developmental process. These mechanically-focused models assume symmetry is already broken at the initial state and give no explanation as to what provides the initial cue to cause the break in symmetry.

Much work has been done to understand the mechanisms behind these PC shapes but the puzzle is far from being solved, as most of the pieces largely remain unconnected. A major challenge in making the connections lies in the difficulties in determining the cause and effect between the mechanisms involved and the lack of experimental techniques to investigate such mechanisms. However, hypotheses can be tested by using a combination of experimental and modeling approaches. Known biological components can be incorporated into models, along with proposed mechanisms to be tested. Dr. Yang and Dr. Pan have proposed the following mechanism to incorporate the initial break in symmetry, as well as the role of chemical and mechanical components in the development of the interlocking PC shapes. The proposal can be broken down into three main pieces, that together can make a complete story of how these shapes are produced.

The first piece of the puzzle starts at the beginning of development: explaining how the cells' components break that initial symmetry. This initialization of polarization process is proposed in [3]. The process starts with TMK1 and sterols randomly distributed along the plasma membrane of the PCs. Auxin stimulates these components and causes the sterols and TMK1 to cluster together to form nanoclusters. Once the concentration of those

clusters is high enough, they then activate ROP6. The ROP6 then signals the microtubules to arrange in a manner where they prevent further movement of the particles, creating a positive feedback so the clusters continually become larger. The large clusters then cause the symmetry of the cell wall to break and multiple polar sites for indentations to form. Pan *et al.* found that this auxin-induced, sterol-dependent nanoclustering of TMK1 is critical for the formation of polarized domains at the plasma membrane [3].

With the nanodomains established, the ROP2 and ROP6 pathways stablize polarization and initiate shape change (refer to Fig 1.1) [4]. Auxin activates the ROP2 and ROP6 signaling pathways and these two pathways mutually inhibit each other. ROP6 is localized to the indentations, which accumulates microtubules. Microtubules suppress outgrowth of the cell, furthering indentations. ROP2 localizes to the lobes and recruits F-actin, which enhances outgrowth of the cells. Mutual inhibition of the two ROP pathways maintains ROP polarity and promotes the formation of the lobes and indentations [4]. Neighboring cells coordinate their polarities so that the ROP2 region of one cell matches with the ROP6 area of the neighboring cells, thus matching the lobes of one cell with the indentations of the other. This further exaggerates the shape change of the cells to their final shape.

In order to test this proposed three-stage mechanism for PC development, a modeling framework first needs to be developed that can be applied to this system. Once that framework is established, incorporation of chemical signaling dynamics, mechanical forces, and cell-cell interaction allows for future testing of the proposed mechanism to simulate shape change.

In order to build a modeling framework to accomplish this goal, the following steps are applied:

Step 1: Utilize the local level set method to capture the deforming cell boundaries.

Step 2: Solve reaction-convection-diffusion equations on the deforming interface to simulate protein dynamics
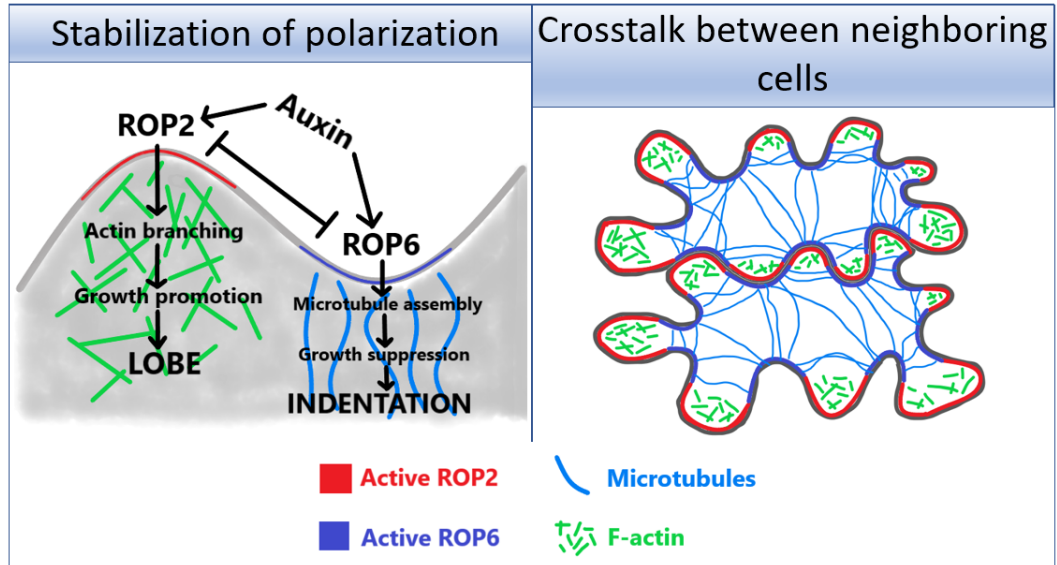
Figure 1.1: Illustration of ROP2 - ROP6 pathway in PCs (left) and interdigitation coordination between neighboring cells (right).

Step 3: Introduce adhesion and cell-cell interaction to simulate multiple cells.

Step 4: Determine boundary conditions (both biological and numerical).

The development of each of these steps is detailed in the next section, with implementation and initial results of the modeling framework shown in Section 1.4.

1.3 Modeling Background

Level Set Method

The level set method [10] is an Eulerian method used to describe and capture the shape and motion of objects throughout time. For simplicity, the following discussion is based in two-dimensional space. Consider a closed, moving curve $\Gamma(t)$ that encloses a region $\Omega(t)$. The corresponding level set function $\phi(x, t)$ is a Lipschitz continuous function such that the zero level set of $\phi(t)$ is always $\Gamma(t)$. In particular, $\phi(x, t) < 0$ in $\Omega(t)$ (inside the enclosed region), $\phi(x, t) = 0$ on $\Gamma(t)$ (the interface), and $\phi(x, t) > 0$ in $\overline{\Omega}(t)^c$ (outside the enclosed
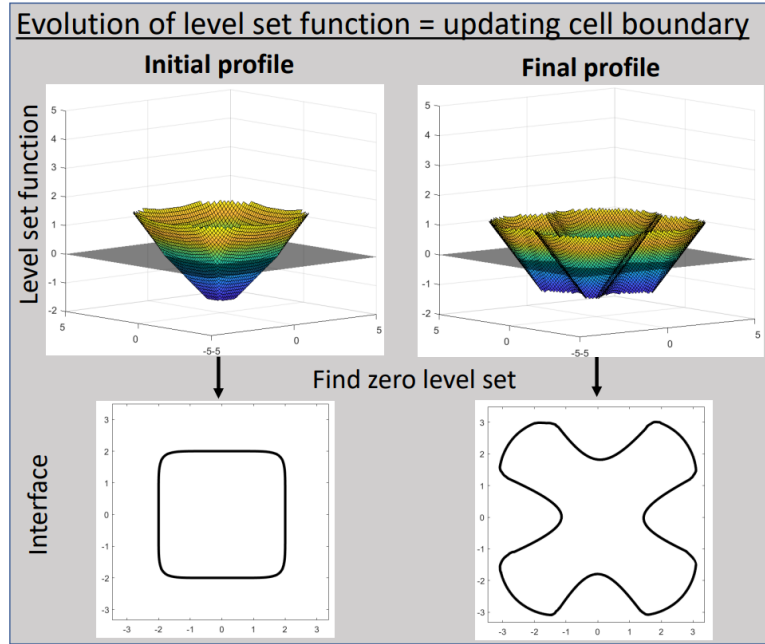
Figure 1.2: Illustration of evolution of level set method.

region), where $x \in \mathbb{R}^2$ and $t \in \mathbb{R}^+$. In this method, moving the interface is equivalent to updating the level set function ϕ (refer to Fig 1.2).

Now, suppose that the interface $\Gamma(t)$ moves under the prescribed velocity field $\vec{u} = \dot{x}(t)$. Since $\phi(x(t), t) = 0$ on $\Gamma(t)$, differentiating the equation with respect to t gives the following Hamilton-Jacobi type equation:

$$\phi_t + \vec{u} \cdot \nabla\phi = 0 \tag{1.1}$$

Simplifying this equation by considering the projection of the velocity onto the direction normal gives the equivalent expression:

$$\phi_t + u_n |\nabla\phi| = 0 \tag{1.2}$$

Thus, the evolution of the level set function is done by solving this equation.

A typical choice for $\phi(x, t)$ is the signed distance function, i.e., the level set value at each point in the space indicates how far that point is from the interface. Points inside the
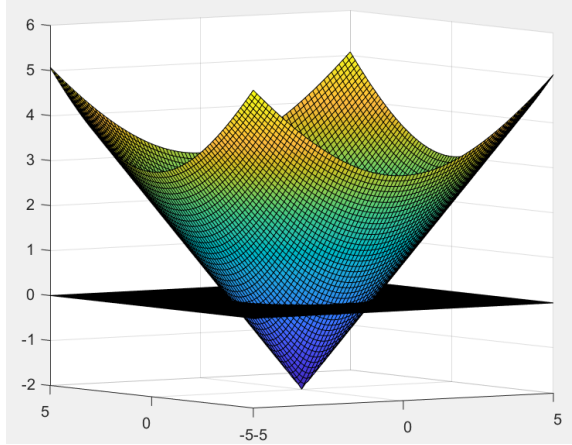
Figure 1.3: A signed distance function for a circle of radius 2 in the region $[-5,5] \times [-5,5]$.

bounded region have negative values, points on the interface have zero value, and points outside the bounded region have positive values. A key characteristic of the signed distance function is that $|\nabla \phi| = 1$. An example of a signed distance function for a circle of radius 2 in the region $[-5,5] \times [-5,5]$ is shown in Fig 1.3. The signed distance function is a common choice due to many advantages, including numerical stability, easy initiation, and topological flexibility.

As the level set function evolves throughout time, it may lose the characteristic of being a signed distance function, in addition to the solutions becoming very flat and/or steep at the interface (i.e., $|\nabla \phi|$ becomes large). A process called reinitialization is performed periodically to reset $\phi(x,t)$ to a signed distance function. Reinitialization is done by solving the following Hamilton-Jacobi equation to steady state:

$$\phi_\tau + S(\phi)(|\nabla \phi| - 1) = 0 \tag{1.3}$$

7

where $S(\phi)$ is the sign function of ϕ defined as:

$$S(\phi) = \begin{cases} -1, & \text{if } \phi < 0 \\ 0, & \text{if } \phi = 0 \\ +1, & \text{if } \phi > 0 \end{cases} \qquad (1.4)$$

The sign function is used to achieve the correct flux direction. In other words, outside of the interface, information from the interface is propagated away and inside the interface, information is propagated towards the interior of the enclosed region. By solving Eqn 1.3 to steady state, $(|\nabla\phi| - 1)$ in the equation ensures that the level set function becomes a signed distance function, i.e., $|\nabla\phi| = 1$.

The level set method is a powerful technique used for capturing the evolution of shapes throughout time. However, like any method, it comes with many advantages and limitations. Some advantages are:

1) The moving interface is captured on a fixed grid so there is no adding or removing grid points.

2) The method turns a geometric problem into a PDE-based problem. The PDEs in the level set method are of the same type so efficient and robust numerical schemes can be utilized to solve the equations in a similar manner, as well as to deal with discontinuities and nonlinearities.

3) The level set method can handle complex shapes and shape changes.

Limitations of the level set method include:

1) The traditional level set method can be computationally expensive. The method adds a dimension to the problem so it can become computationally heavy.

2) In order to achieve accurate results, the grid resolution is usually required to be fine. This adds to the computational expense of the method.

3) Success of the method requires an appropriate choice of factors such as time step, numerical schemes used, and efficient data structures. Because of these factors, implementation of the method can be quite sophisticated.

Since the computational intensity is the main drawback of the level set method, a local level set method has been developed by Peng *et al.* [1]. In this method, the level set evolution and reinitialization are only performed in a small neighborhood around the interface. In this way, simulations are faster and the method is more computationally efficient. In the local level set method, information is recorded in two sets of neighborhoods, or tubes, around the interface: an inner tube of width 2β (i.e., $|\phi| \leq \beta$) and an outer tube of width $2(\gamma - \beta)$ (i.e., $\beta < |\phi| \leq \gamma$). A smooth cut-off function is introduced into the level set evolution equation in order to prevent numerical oscillations. With this, the local level set evolution equation is:

$$\phi_t + c(\phi)u_n|\nabla\phi| = 0 \tag{1.5}$$

where the cutoff function $c(\phi)$ is:

$$c(\phi) = \begin{cases} 1, & \text{if } |\phi| \leq \beta \\ \dfrac{(|\phi| - \gamma)^2(2|\phi| + \gamma - 3\beta)}{(\gamma - \beta)^3}, & \text{if } \beta < |\phi| \leq \gamma \\ 0, & \text{if } |\phi| > \gamma \end{cases}$$

An example of a circular interface of radius 2 with corresponding inner and outer tubes is shown in Fig 1.4. To reiterate, computations are only done in these regions, rather than the entire computational domain.

A summary of the algorithm for the local level set method is as follows:

Step 0 - Initialization: Choose the initial level set function $\phi_0(x, t)$. If ϕ_0 is not a signed distance function, perform reinitialization on ϕ_0.
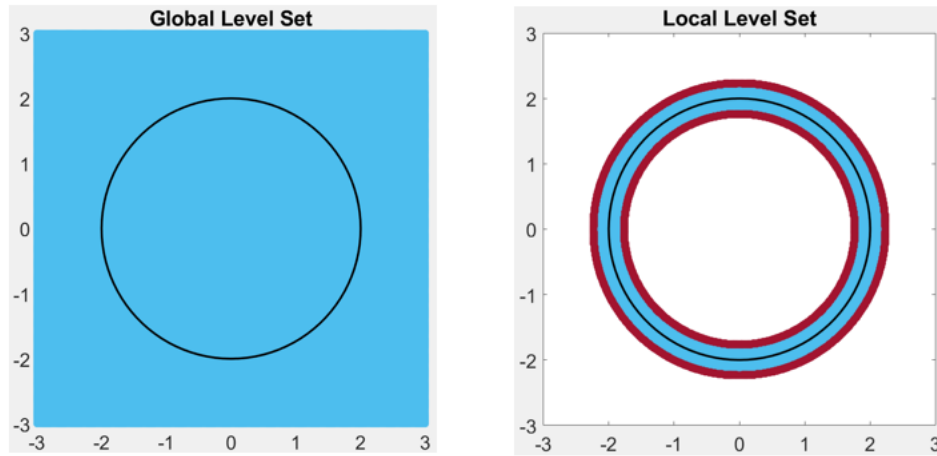
Figure 1.4: Comparison of computational domain in global vs. local level set. Black curve is interface of radius 2. Blue region corresponds to where the level set equation is solved in global (left) vs. local (right) level set method. The red region in the local level set figure (right) is the outer tube, where a slight variation of the level set equation is solved.

Step 1 - Determine velocity and time step: Velocity can be a function of factors such as the space variable x, curvature, or protein concentration on the interface. An appropriate time step is chosen for the numerical method used.

Step 2 - Evolve the local level set equation: Update the local level set function by solving Eqn 1.5 for one time step. Numerical schemes such as the Godunov method with WENO can be used.

Step 3 - Reinitialization: Solve Eqn 1.3 to perform reinitialization near the interface to reset ϕ to a signed distance function. This step is only done periodically and the frequency depends on factors such as the velocity and numerical schemes used.

Step 4 - Repeat steps 1-3 for a maximum number of iterations.

Adhesion Methods for Multicellular Simulations

The local level set method can also be applied to scenarios where multiple shapes are present. The fundamental idea is similar to the case where a single shape is present. However,

10

additional complexity is introduced due to the interaction between shapes. This added difficulty requires careful consideration and the methodology applied depends on the specific application and the nature of the shape interactions.

Pavement cells remain tightly packed as they grow and change shape. With this in mind, adhesion must be incorporated into the local level set method so that gaps or overlaps between neighboring cells never form. The first type of adhesion considered is that discussed by Merriman *et al.* in [5]. The type of adhesion described here is an artificial type of adhesion. That is, there is no biological or mechanical meaning behind the movement of neighboring cells walls. This type of adhesion is useful for the initialization of PC modeling, as in Fig 1.5, as well as other applications.

Adhesion Method 1: To begin, each cell is assigned its own level set function. Those level set functions are evolved simultaneously and independently according to their prescribed velocity. Doing this allows for gaps or overlaps between cell boundaries to potentially form. Merriman *et al.* [5] present the following correction term that can be applied to remove these gaps or overlaps:

$$\phi_i = \phi_i - \max_{i \neq j} \phi_j \tag{1.6}$$

However, the authors assume that the level set function is positive inside the interface and negative outside the interface. Since we use opposite signs, the correction term used here is:

$$\phi_i = \phi_i - \min_{i \neq j} \phi_j \tag{1.7}$$

The general idea behind this function is that it moves the interface (and corresponding level sets) of ϕ_i to the midway point between the original location of the interface of ϕ_i and the closest interface of other ϕ_j. The $\min_{i \neq j} \phi_j$ term locates the "closest" level set function to ϕ_i. Since positive level set values are present outside of an interface, a smaller value at a given point corresponds to being closer to the interface of that ϕ_j. Thus, the farther the interface of ϕ_j is from the interface of ϕ_i, the more positive the value of ϕ_j. A similar idea applies to overlapping interfaces and negative level set values. Once that "closest" ϕ_j is

chosen, performing $\phi_i = \phi_i - \min\limits_{i \neq j} \phi_j$ relocates the zero level set of ϕ_i to the location where $\phi_i = \min\limits_{i \neq j} \phi_j$. This is repeated for all of the level set functions.

This method of adhesion can be applied to correct walls and junctions of cells that are neighboring other cells. However, the method cannot be applied to segments of cells that do not neighbor other cells. There are multiple ways to combat this behavior and the method used depends on the specific application. For the application to PC's, we utilize the following technique:

Step 1: Define a global level set function ϕ_{global} that has zero level set equal to the external walls of the outermost layer of cells, with negative values where the cells are present and positive values outside the collection of cells.

Step 2: Use ϕ_{global} to define an indicator matrix "Bound_mat" so that

$$\text{Bound\_mat} = \begin{cases} 0, & \text{if } \phi_{global} \geq 0 \\ 1, & \text{if } \phi_{global} < 0 \end{cases}$$

Step 3: Perform adhesion using the updated equation:

$$\phi_i = \phi_i - (\min\limits_{i \neq j} \phi_j) \cdot \text{Bound\_mat} \tag{1.8}$$

Performing adhesion in this way allows the external walls of the outermost layer of cells to remain in place when adhesion is applied.

This type of adhesion is a quick way to eliminate any gaps or overlaps between cells. However, it has two main downfalls. One, this method causes the level sets to degenerate over time. This degeneration causes instability in the numerical simulations over time. Reinitialization can be applied to correct this degeneration enough for the level set algorithm to remain stable but it causes problems when incorporating aspects such as evolution of protein dynamics. The second problem with this type of adhesion is particular to the PC application: there is no physical meaning behind this adhesion. In a way, this method of
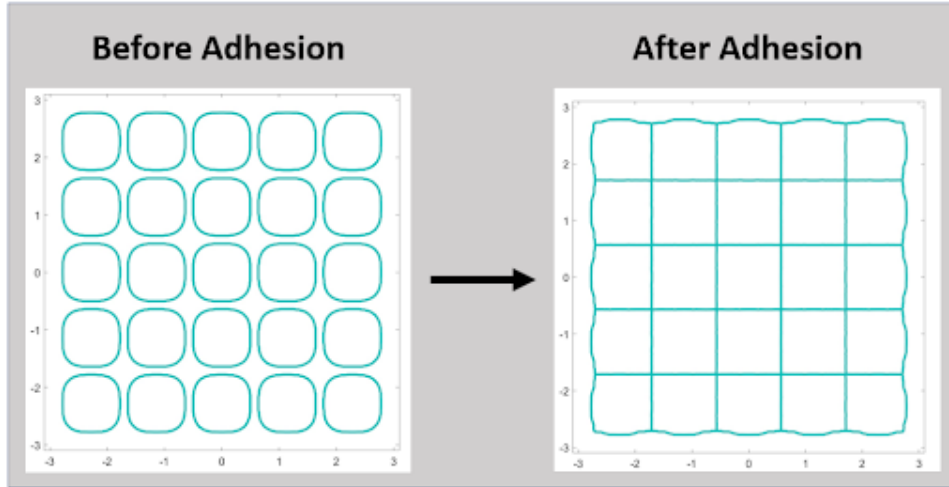
Figure 1.5: Adhesion method 1 applied to a group of 25 cells.

adhesion just picks up and moves the interfaces of the cells so that they are positioned correctly. There is no biological explanation for this process. Due to these two issues, a different type of adhesion should be used during the level set iterations.

Adhesion Method 2: A more physically meaningful and numerically stable method of adhesion is presented by Zhao *et al.* in [14]. This method introduces constraints in the velocity term of the level sets to prevent overlaps and gaps. They recall that the area of an open region Ω is defined by

$$area(\Omega) = \iint H(\phi(x,y))dxdy \tag{1.9}$$

where $H(x)$ is the Heaviside function:

$$H(x) = \begin{cases} 1, & \text{if } x \leq 0 \\ 0, & \text{if } x > 0 \end{cases}$$

With this definition of the Heaviside function, we know $H(\phi_i) = 1$ inside cell i and $H(\phi_i) = 0$ outside cell i. When there is an overlap of cells, $\sum_{i=1}^{n} H(\phi_i(x,y)) > 1$ for (x,y) in the overlapping region. Similarly, when there is a gap between cells, $\sum_{i=1}^{n} H(\phi_i(x,y)) = 0$ for

13

(x, y) in the gap region. Thus to eliminate overlaps or gaps between cells, we need, for each (x, y), to satisfy:

$$\left(\sum_{i=1}^{n} H(\phi_i(x, y)) \right) - 1 \equiv 0 \tag{1.10}$$

Zhao *et al.* then uses this constraint in the minimization of the energy functional to obtain the following evolution equation for the level set function:

$$\frac{\partial \phi_i}{\partial t} = |\nabla \phi_i| \left(\gamma_i \nabla \cdot \left(\frac{\nabla \phi_i}{|\nabla \phi_i|} \right) - V_i - \lambda \left(\sum_{j=1}^{n} H(\phi_j) - 1 \right) \right) \tag{1.11}$$

where γ_i can be thought of as a tuning parameter for the amount of tension and λ is a Lagrange multiplier and can be determined as:

$$\lambda = \frac{\sum_{i=1}^{n} \iint_D \delta(\phi_i)(\gamma_i \Delta \phi_i - e_i)(\sum_{j=1}^{n} H(\phi_j) - 1) dx dy}{\sum_{i=1}^{n} \iint_D \delta(\phi_i)(\sum_{j=1}^{n} H(\phi_j) - 1)^2 dx dy} \tag{1.12}$$

Here $\delta(x)$ is the Dirac delta function, e_i is the area energy density of the regions, and D is a fixed domain.

In the PC application, we use a modified version of Eqn 1.11:

$$\frac{\partial \phi_i}{\partial t} = |\nabla \phi_i| \left(-V_i - \lambda \left(\sum_{j=1}^{n} H(\phi_j) - 1 \right) \right) \tag{1.13}$$

V_i in this equation is the velocity determined by the protein concentration on the boundary of the cells and the last term in the equation is the constraint term to eliminate gaps and overlaps between cells. By utilizing this type of adhesion, cells grow and change shape in a more biologically relevant manner, unlike the first method of adhesion presented.

Incorporating Biochemical Dynamics on a Moving Interface

Now that the boundaries of the cells can be captured using the local level set method, the protein dynamics on the moving interfaces need to be incorporated. ROP2 and ROP6 concentrations will be integrated into the model by solving reaction-convection-diffusion equations on the moving interface. Solving PDEs on a moving interface is a challenging

problem so we begin with a framework for handling convection and diffusion on these moving cell boundaries. For simplicity, the discussion will begin with a single cell.

We develop a convection-diffusion equation modeled after the efforts presented in [2]. Xu and Zhao present a convection-diffusion equation for the concentration of a surfactant on a moving interface that is convected in an incompressible flow (i.e., $\nabla \cdot \vec{u} = 0$). However, we cannot assume ROP2 and ROP6 will be convected in an incompressible flow so the derivation is slightly different. To start, let f be a scalar quantity (e.g., the concentration of a protein) defined on the interface. Convection of f can be written as:

$$f_t + \nabla_s \cdot (f\vec{u}) = 0 \tag{1.14}$$

Here ∇_s is the surface gradient, $\vec{u}$ is the prescribed velocity field, and $\hat{n}$ is the unit outward normal. Decomposing the velocity into normal and tangential components allows the above equation to be written as:

$$f_t + \nabla_s \cdot (f u_s) + (\vec{u} \cdot \hat{n})\kappa f = 0 \tag{1.15}$$

However, since we are assuming that PCs only have velocity in the normal direction, the previous equation can be rewritten as:

$$f_t + (\vec{u} \cdot \hat{n})\kappa f = 0 \tag{1.16}$$

Adding in isotropic diffusion where the diffusion coefficient is normalized to 1, the equation now becomes a convection-diffusion equation:

$$f_t + (\vec{u} \cdot \hat{n})\kappa f - \nabla_s^2 f = 0 \tag{1.17}$$

As shown in [2], if f is defined smoothly in a neighborhood of Γ, the surface Laplacian of f can be decomposed so the convection-diffusion equation becomes:

$$f_t + (\vec{u} \cdot \hat{n})\kappa f - (\nabla^2 f - \frac{\partial^2 f}{\partial \hat{n}^2} - \kappa \frac{\partial f}{\partial \hat{n}}) = 0 \qquad (1.18)$$

This equivalent representation of the surface Laplacian operator works well in the context of the level set method because geometric quantities can be computed easily by writing them in terms of the level set function ϕ. For example, the unit normal $\hat{n}$ and the mean curvature κ are written as:

$$\hat{n} = \frac{\nabla \phi}{|\nabla \phi|} \quad , \quad \kappa = \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} \qquad (1.19)$$

In the level set formulation, the concentration of the protein needs to be extended to a neighborhood of the interface. A natural way to perform this extension of f off the interface is to let f be constant along the normal direction. This extension process is done by solving the following Hamilton-Jacobi equation to steady-state [1]:

$$f_t + S(\phi)\frac{\nabla \phi}{|\nabla \phi|} \cdot \nabla f = 0 \qquad (1.20)$$

where $S(\phi)$ is the sign function expressed in Eqn 1.4.

A summary of the algorithm combining both local level set and protein dynamics is detailed in the next section.

1.4 Numerical Approach and Results

Full Algorithm

The detailed algorithm described below will be expanded upon to investigate the biological hypothesis stated in Section 1.2. This algorithm is designed for multiple cells that are changing shape with a protein convecting and diffusing along the cell boundaries.

Step 0: Initialize

Start by setting up a uniform mesh where $\Delta x = \Delta y$ are sufficiently small on the rectangular domain $[a, b] \times [c, d]$. The rectangular grid must be big enough that the moving interfaces do not leave the domain throughout the duration of the computations. The tube widths used throughout the algorithm need to be determined. These widths are dependent on aspects such as desired computational speed and accuracy and the numerical scheme used. For example, in the computations done in [1] tubes of size $\beta = 6\Delta x$ and $\gamma = 9\Delta x$ are used with a 5th order WENO scheme.

With the computational grid and tubes defined, the variables representing the level sets and protein concentrations for each cell can be initialized. The initial level set functions ϕ_i and initial protein distributions f_i are user-defined. Reinitialization and extension are completed, if necessary.

For the pavement cell application, the cells are initialized as super-ellipses with some small distance between them. Adhesion method 1 is then applied to those cells, as in Fig 1.5. Global reinitialization is run to steady state for each cell. Once these cells are successfully reinitialized, the initial protein distributions f_i are established and extension is completed.

Step 1: Determine velocity and time step

Calculate the velocity and determine a time step that satisfies the CFL condition. These velocities can be constant, or in terms of protein concentration, curvature, etc. If adhesion method 2 is used, velocity will also be in terms of the amount of overlaps or gaps between neighboring cells.

Step 2: Evolve local level set equation

Complete one iteration of the local level set equation (Eqn 1.5) for each cell. To do so, the coordinates of the β and γ tubes need to be determined. A simple find

17

function can be utilized since the ϕ_i are signed distance functions for each cell. This step can be done in parallel for each cell since they evolve independently.

Step 3: Evolve protein concentration on moving interface

Evolution of Eqn 1.18 is performed in the inner β tube of the previous iteration's ϕ_i for each cell. If the time step for the convection-diffusion equation (Eqn 1.18) is the same as the time step for the local level set equation (Eqn 1.5), then complete one iteration of the convection-diffusion equation. However, since the time step necessary for diffusion is often much smaller than the time step necessary for the local level set equation, this is usually not the case. When this happens, perform (local level set time step) / (convection-diffusion time step) number of iterations of Eqn 1.18. This step 3 can also be done in parallel for each cell.

Step 4: Reinitialize level set functions and extend protein off interface

Reinitialization is performed to the updated ϕ_i from step 2 by evolving Eqn 1.3 to steady state. This process is done globally for each cell but is terminated when the values within the tube $|\phi_i| \leq (\gamma + 4 \cdot dx)$ reach steady state.

Values of protein concentration from step 3 are then extended to the reinitialized level set functions for each cell. Extension is done by solving Eqn 1.20 to steady state in a tube slightly smaller than the tube used in reinitialization but larger than the γ tube.

Both reinitialization and extension are done in parallel for each cell.

Step 5: Repeat

Repeat steps 1-4 for some maximum number of iterations.

Similar numerical algorithms can be used to solve all three Hamilton-Jacobi equations. For the local level set equation (Eqn 1.5) and the reinitialization equation (Eqn 1.3), a

TVD Runge-Kutta [12] method of third-order is used for the time discretization along with Godunov's scheme [1] for the flux and fifth-order WENO [13] for the spatial differences. Since numerical accuracy is usually not an issue for the extension step, first order upwind schemes coupled with a forward Euler time discretization is used. For the convection-diffusion equation (Eqn 1.18), quantities derived in [2] are utilized for the spacial discretization and forward Euler is used for time. Reference Appendix A.1 for more details on these schemes.

Simulations

The above algorithm is used to produce the simulations below. Both single cell and multicellular simulations are shown, incorporating first just the local level set method, followed by the addition of adhesion, and finally protein dynamics. In all simulations, $\Delta x = \Delta y = 0.0125$, with tube widths of $\beta = 8\Delta x$ and $\gamma = 10\Delta x$. Figures showing the cell shapes and, when applicable, protein concentration on the cell membranes are provided. The location of the zero level set of each cell is shown in two dimensions and the protein concentrations along the membranes are shown separately in two dimensions as well.

Simulation 1.4.1. We first implement the numerical algorithm for the local level set method for a single cell. The computational domain is chosen to be $[-1.5, 1.5] \times [-1.5, 1.5]$. The initial shape of the cell membrane in 2D space is a super-ellipse centered at $(0, 0)$, with $a = b = 0.57$. The velocity shown in Fig 1.6 is used to run 100 iterations with the time step $\Delta t = 0.03 \frac{\Delta x}{\max(|v|)} = 3$. The results are shown in Fig 1.7.
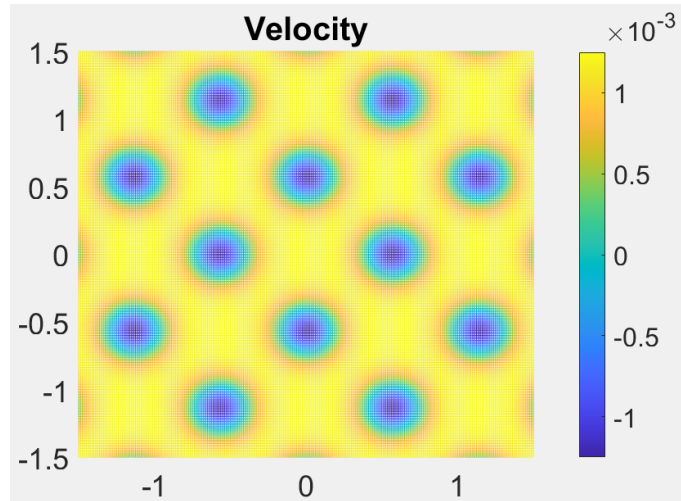
Figure 1.6: Velocity used in Sim 1.4.1. to produce the results shown in Fig 1.7.
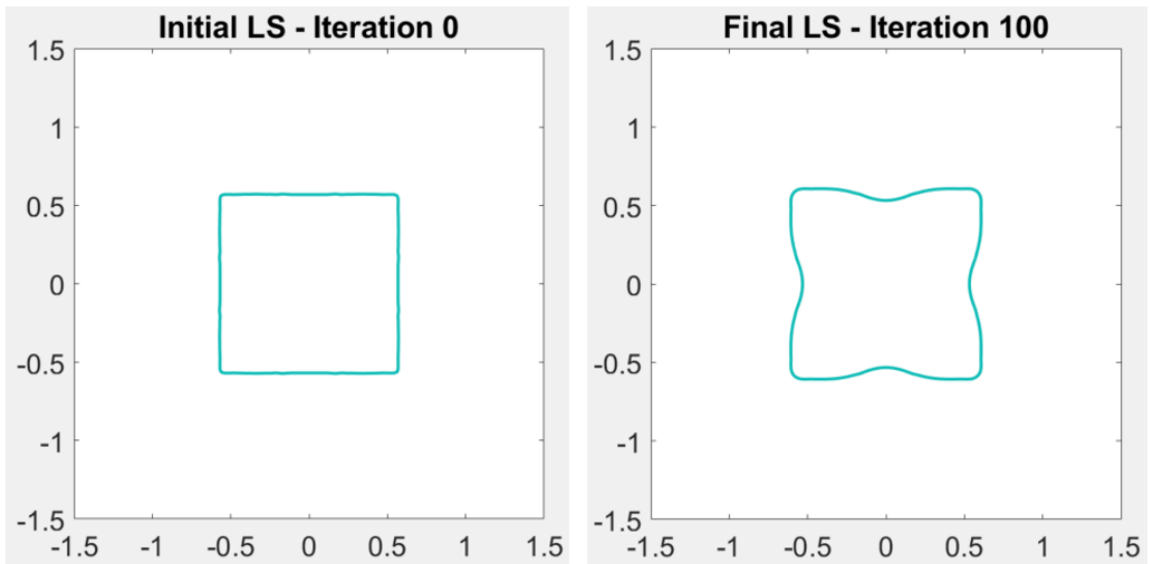
Figure 1.7: Illustration of the local level set method applied to a single cell.

Simulation 1.4.2. The next simulation illustrates the local level set method with adhesion method 2 for 25 cells. The computational domain is now $[-3.1, 3.1] \times [-3.1, 3.1]$. The cells are initiated in the format of a 5×5 grid of super-ellipses, each with $a = b = 0.57$, with the center cell in the grid centered at $(0, 0)$. Adhesion method 1 is applied to adhere the initial cell membranes together, as in Fig 1.5. Starting with the 25 adhered cells (Iteration 0 in Fig 1.9) the velocity field applied to the level sets of each cell i is of the form:

$$V_i = (-1)^i \left(\text{f\_v} - \frac{1}{2} \max(|\text{f\_v}|) \right) \cdot k - \lambda \left(\sum_{j=1}^{25} H(\phi_j) - 1 \right) \tag{1.21}$$

where f_v is force_val given in Fig 1.8, $\lambda = 0.01$, and $k = 0.0005$. The cells are numbered in increasing order from left to right, bottom to top. Notice this velocity field includes adhesion method 2, similar to Eqn 1.13. The simulation is run for 95 iterations with the time step $\Delta t = 0.03 \frac{\Delta x}{\max(|v|)}$ for each cell. The results are shown in Fig 1.9.
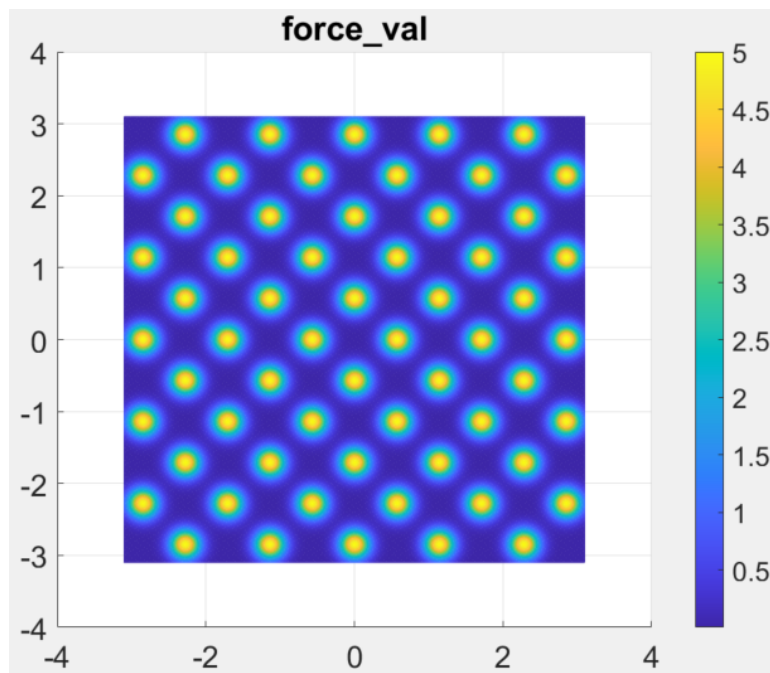
Figure 1.8: Values for the parameter "force_val" in Eqn 1.21 to produce results of Sim 1.4.2 in Fig 1.9.
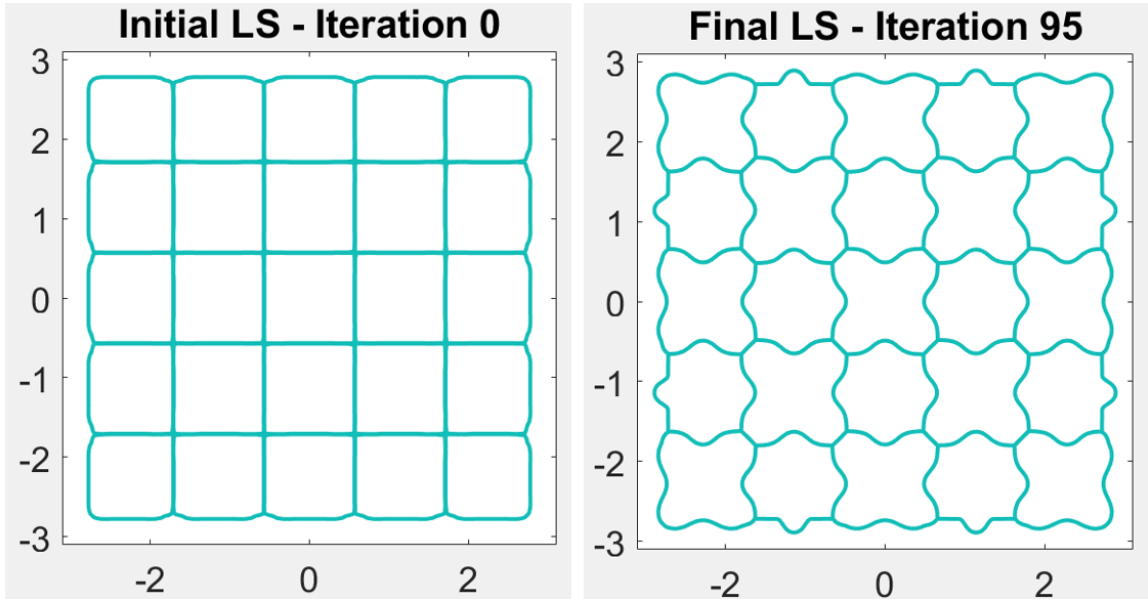
Figure 1.9: Initial and final level sets from Sim 1.4.2.

Simulation 1.4.3. This next simulation tests the role of adhesion. The computational domain is expanded to $[-4.3, 4.3] \times [-4.3, 4.3]$. Similar to Sim 1.4.2., the cells are initiated as a 7×7 grid of super-ellipses, each with $a = b = 0.57$, with the grid centered at $(0,0)$. Adhesion method 1 is applied to achieve the starting profile, Iteration 0 in Fig 1.10. In Sim 1.4.2., the velocity of neighboring cells is linked such that a positive velocity in one cell corresponds to a negative velocity of the same magnitude in the neighboring cell, and vice versa. With such a velocity field, cells grow in a correlated fashion so the role of adhesion is minimally tested. However, to investigate if adhesion will still prevent gaps or overlaps between cell boundaries when cells do not grow in a correlated manner, the following velocity field is applied for each cell i:

$$V_i = \left(\text{f\_v} - \frac{1}{2} \max(|\text{f\_v}|) \right) \cdot k - \lambda \left(\sum_{j=1}^{49} H(\phi_j) - 1 \right) \tag{1.22}$$

where f_v is force_val given in Fig 1.12, $\lambda = 0.01$, and $k = 0.0005$. The simulation is run with this velocity for 200 iterations and a time step of $\Delta t = 0.03 \frac{\Delta x}{\max(|v|)}$. Without adhesion, this

velocity field should cause the cells to grow inwardly at the corners and expand outwardly in the center of the sides, creating gaps and overlaps between neighboring cells. As seen in Fig 1.10, this is not the case because the adhesion term prevents this from happening. The cell boundaries remain in the same location as the initial setup due to adhesion and the fact that the magnitude of the forces are the same in the neighboring cells.
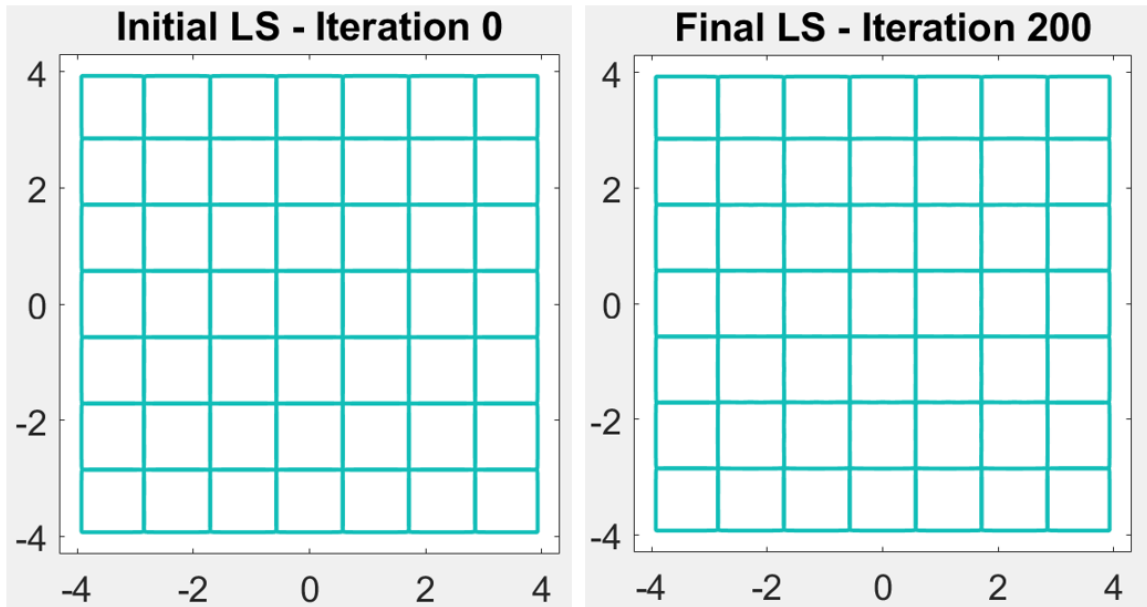
Figure 1.10: Resulting level sets from Sim 1.4.3., where the effect of adhesion is tested.

Simulation 1.4.4. This simulation now illustrates the incorporation of protein dynamics on the moving interfaces. The same computational domain and initial level set values as Sim 1.4.3. are used. The concentration of protein "A" for each cell is initially set to 10 in the inner β tube and 0 elsewhere.

Where this simulation differs from Sim 1.4.2. is in the implementation of boundary conditions. Thinking back to the PC application, PCs rely on information from their neighbors in order to produce the correct ROP patterning. However, we are unable to model an entire system of PCs so both biological and numerical boundary conditions must be imposed on the cells being modeled. To handle this in the simulations the following idea is proposed. All cells are initialized with similar super-ellipse shapes and homogeneous

distribution of the proteins, plus some random noise. The outermost layer of cells, the purple layer of cells in Fig 1.11, will not be updated through level set evolution or convection-diffusion. The next layer, the green cells in Fig 1.11, will be updated but with the incorrect information from the boundary layer of cells. Continuing this trend inwards, there will be a few cells in the center of the 7×7 grid that receive correct information. These cells, the blue and red layers in Fig 1.11, will be the cells studied.
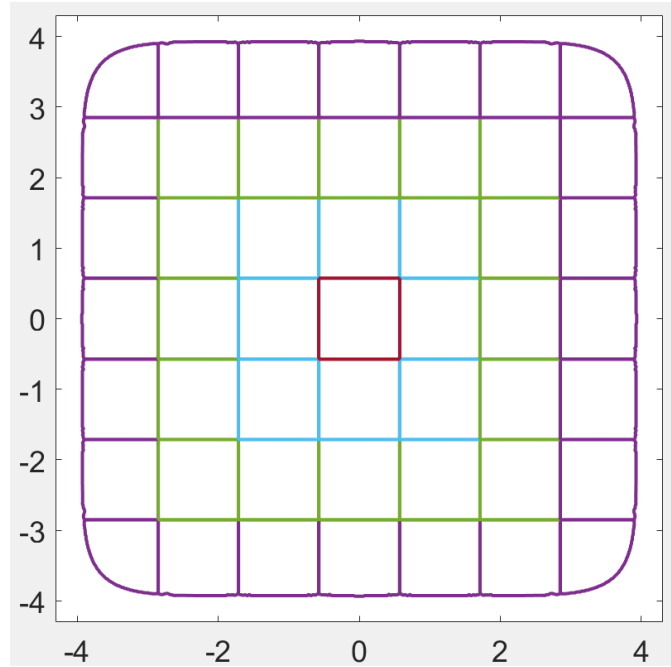
Figure 1.11: Illustration of layers of cells in 49 cell simulation.

To initially test these boundary conditions, the following velocity field is applied for cell i:

$$V_i = \begin{cases} (-1)^i \left(\text{f\_v} - \frac{1}{2} \max(|\text{f\_v}|) \right) \cdot k - \lambda \left(\sum_{j=1}^{49} H(\phi_j) - 1 \right) & \text{if } i \in IL \\ 0 & \text{if } i \in BL \end{cases} \tag{1.23}$$

where f_v is force_val given in Fig 1.12, $\lambda = 0.01$, $k = 0.0005$, BL is the boundary layer of cells, and IL is the interior layers of cells.
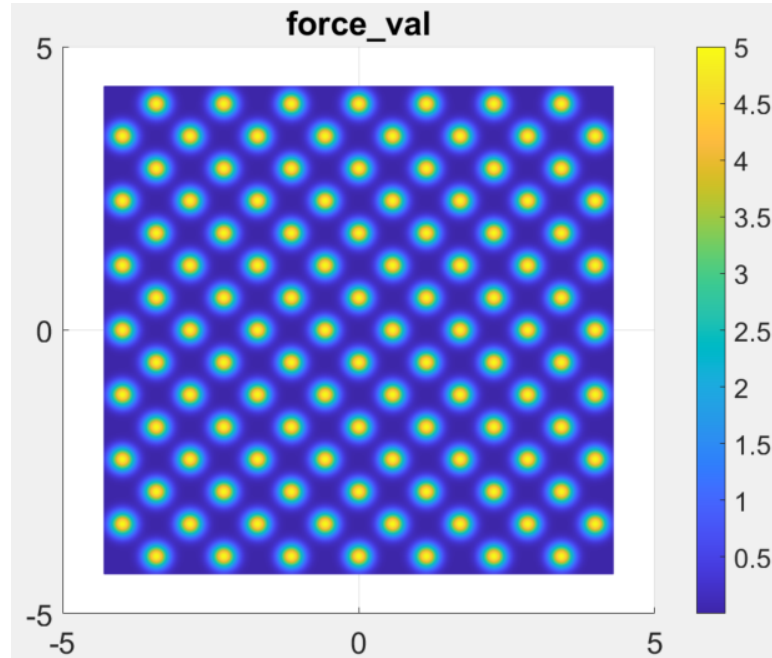
Figure 1.12: Values for the parameter "force_val" in Sim 1.4.3. and Sim 1.4.4.

Imposing these boundary conditions, the simulation is run for 115 iterations with the level set time step $\Delta t = 0.03 \frac{\Delta x}{\max(|v|)}$ for each cell. The time step for the convection-diffusion equation in each cell is $\Delta \tau = 3.75 \cdot 10^{-6}$. For every level set iteration, $\Delta t / \Delta \tau$ number of iterations of the convection-diffusion equation are run.

The resulting concentration profiles of the 49 cells are shown in Fig 1.13. The interface of each cell is not explicitly shown but they are in the same location as the nonzero concentration values for each cell. As expected, the outermost layer of cells remain unchanged in both shape and protein concentration level. For the next layer of cells that share a cell boundary with the outermost layer, the interior walls change shape and thus the concentration levels change as well. The cells that increase in overall size see a decrease in protein concentration, and vice versa.
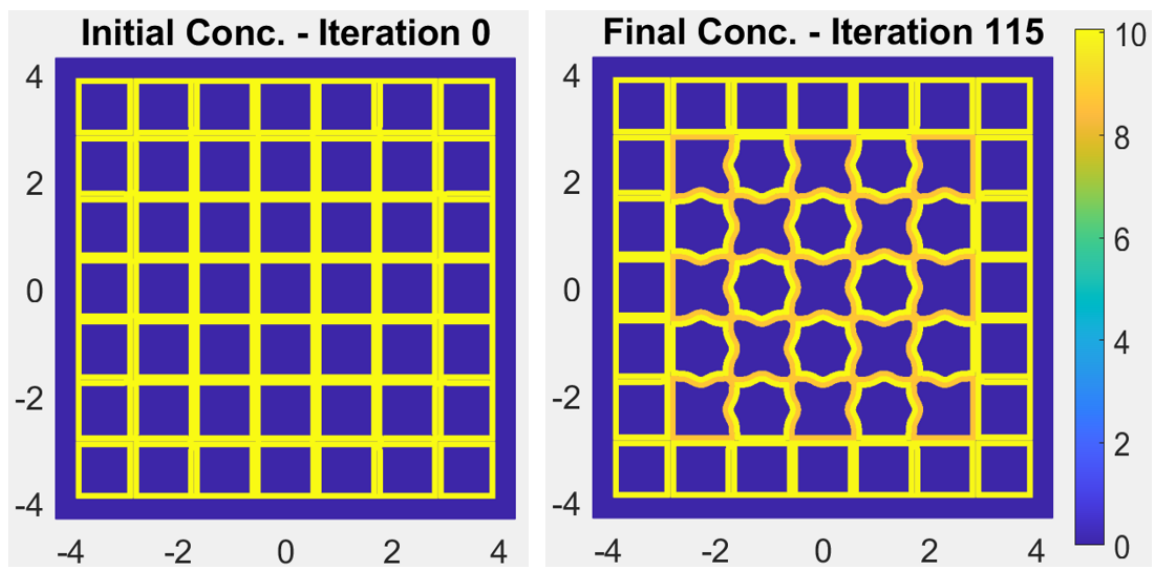
Figure 1.13: Initial and final concentration profiles for Sim 1.4.3. The interface for each cell is in the same location as the nonzero concentration values for that cell.

Simulation 1.4.5. Next is a simulation with the same initial setup and velocity as Sim 1.4.4. but with two changes. First, the initial concentration is now nonhomogeneous along the cell boundaries. The concentration in each cell ranges between 0 and 5, with the highest concentration present in the center of the sides of the cells and lowest concentration in the corners (see Fig 1.14). Second, the number of iterations of the convection-diffusion equation in each level set iteration is reduced from $\Delta t/\Delta \tau$ number of iterations to 100 iterations. This is done to reduce the effect of diffusion.
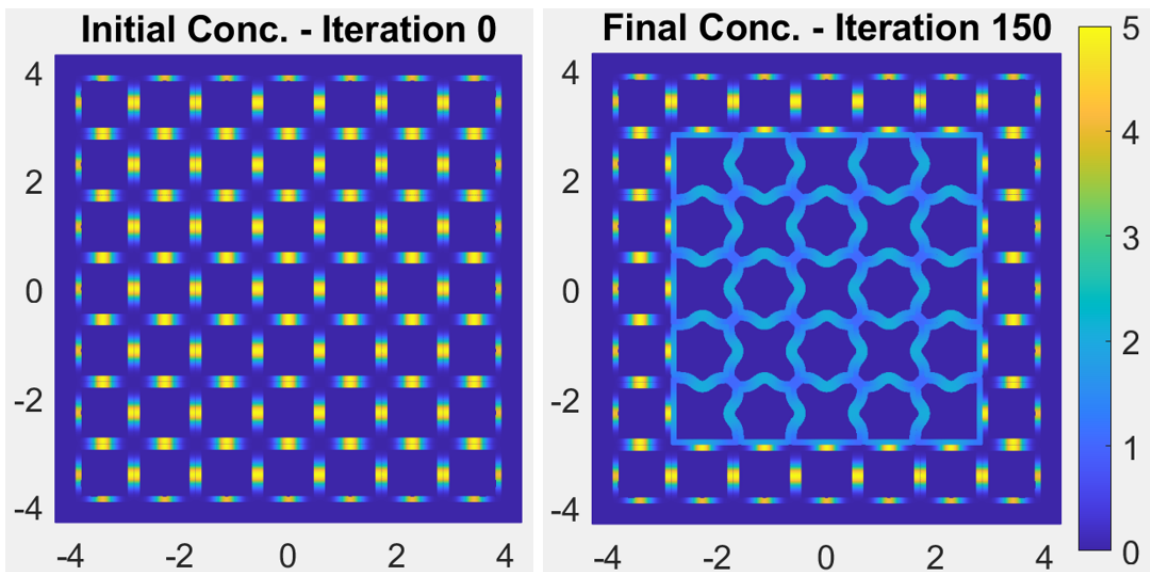
Figure 1.14: Initial and final concentration profiles for Sim 1.4.5.

The outer layer of cells remains unchanged in both level set and concentration values. The inner cells deform in the same manner as in Sim 1.4.4. The protein is convected and diffused to produce a nearly homogeneous final distribution by iteration 150.

Simulation 1.4.6. The final simulation is an example of velocity dependent upon protein concentration levels "u". The initial setup is the same as in Sim 1.4.5. except the initial concentration is slightly different. The concentration is initialized to resemble ROP2 distribution within the cells. The odd numbered cells have concentration levels ranging from about 5 to 10, with the highest located in the center of the cell walls and the lowest in the corners. The even numbered cells have concentrations ranging from about 0 to 5, with the lowest being in the center of the cell walls. In this way, the high concentration region of one cell corresponds with the low concentration region of the neighboring cell.

Since velocity is now dependent upon concentration "u", all parts of the velocity equation must be updated after each iteration. In other words, the velocity for cell i in iteration $m + 1$ is:

$$V_i^{m+1} = \begin{cases} \left(u_i^m - \frac{1}{2}\max_{i \in BL}(u_i)\right) \cdot k - \lambda \left(\sum_{j=1}^{49} H(\phi_j^m) - 1\right) & \text{if } i \in IL \\ 0 & \text{if } i \in BL \end{cases} \tag{1.24}$$

As in the previous simulations, $k = 0.0005$ and $\lambda = 0.01$. The first term in this velocity expression, $\left(u_i^m - \frac{1}{2}\max_{i \in BL}(u_i)\right)$, projects the initial concentration levels $(0, 10)$ to the range $(-5, 5)$. Regions with concentrations in the lower half of the range correspond to negative or inward velocity and concentrations in the upper half of the range correspond to positive or outward velocity. One thing to note is that $\max_{i \in BL}(u_i)$ will remain constant throughout all iterations since the boundary cells are never updated.

Two hundred iterations are run using a level set time step of $\Delta t = 0.03\frac{\Delta x}{\max(|v|)}$ for each cell, convection-diffusion time step of $\Delta \tau = 3.75 \cdot 10^{-6}$, and 100 iterations of convection-diffusion for each level set iteration. Initial and final results are shown in Fig 1.15, along with intermediate results at iterations 50 and 125. By the final time, the protein has homogeneous distribution in the interior cells, with the cells initialized with concentration in the range $(5, 10)$ having overall higher concentration and larger cell size than the cells initialized with

concentration in the range $(0, 5)$. With a few changes, the foundation of this simulation can be used to test the ROP network necessary for PC morphogenesis.
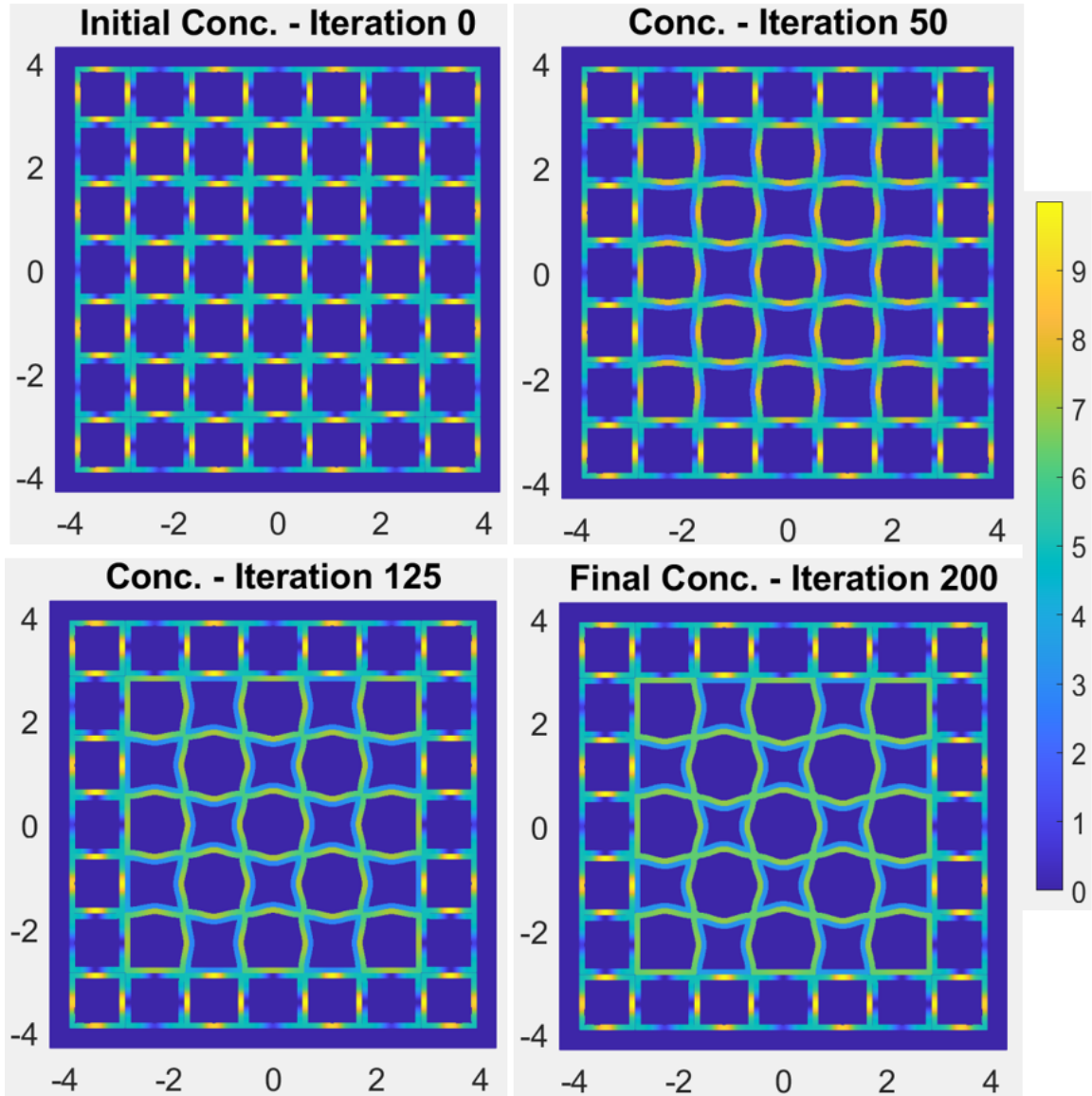
Figure 1.15: Initial, intermediate, and final concentration profiles for Sim 1.4.6.

1.5   Conclusions and Future Directions

In this part, we develop a novel modeling framework that will be used to incorporate mechanical forces, chemical signals, and cell-cell interaction to study pavement cell morphogenesis. The local level set method is utilized to capture the deforming cell boundaries in a less computationally expensive way than the traditional level set method. Two types of adhesion are explored to coordinate cell growth throughout time. Finally, protein dynamics are simulated on the changing cell boundaries by solving a convection-diffusion equation that is coupled with the level set method. Examples of the novel combination of these three aspects are shown.

The model will be used to incorporate the ROP2-ROP6 network to further investigate the role of chemical signaling in the development of the PC shape. A nanoclustering model for the initiation of polarization will be incorporated in the form of initial conditions, with the reaction-convection-diffusion equation on the moving interface used to model the chemical signaling network required between neighboring cells to achieve interdigitation. The update scheme for the velocity term, in terms of protein dynamics and cell-cell adhesion, needs to be determined. To improve computational efficiency of the current model, the numerical scheme used to solve the convection-diffusion equation must be improved. Finally, the model will need to be calibrated using experimental results.

Chapter 2

Modeling Chemical Signaling Network to Regulate Cellular Cytoskeletal Components in the Cross Section of the *Drosophila* Wing Disc

2.1 Preface

This is collaborative work done with members of Dr. Jeremiah Zartman's lab at University of Notre Dame. Dr. Nilay Kumar, Mayesha S. Mim and Dr. Marycruz Flores-Flores generated and analyzed the experimental data with Dr. Zartman's guidance. The modeling was done by me with help from Dr. Weitao Chen and Dr. Mark Alber.

2.2 Biological Background

The *Drosophila* (i.e., fruit fly) wing imaginal disc is a model system to study the mechanisms involved in the development of growing epithelium due to its short cell life, limited number of cells, and large number of common genes shared with mammals. The wing disc is a tissue that becomes the adult wing of the fruit fly. In early stages of development, when the fruit fly is still in the larval stage, the wing disc starts with about 20 cells. The tissue takes on a flat, thin profile where the pouch cells are cuboidal in shape. Over the course of a few days, the cells proliferate and the wing disc grows to come to a final state of around 50,000 cells [18]. During this time, the tissue develops a more curved shape with a flattened midsection and the columnar cells become long and thin. The wing disc develops this shape through the patterning of morphogen and mechanogen signals that prescribe specific characteristics of cells [15].

Due to the wing disc being a model system to study, much work has been done to investigate the components involved in organogenesis. To start, Dpp is a tissue-level morphogen that is expressed along the boundary between the anterior and posterior compartments of the wing disc and diffuses bilaterally across the tissue [20]. Dpp is known to regulate cell growth and division through the downstream regulation of many other morphogens. When combining with the receptor Tkv, the Dpp-Tkv complex activates pMad [21]. pMad then represses Brk, creating a mirror image of the Dpp gradient [22]. Figure 2.2 shows the morphogen profiles in the cross section of the wing disc. This morphogen network regulates Rho-GTPases Rho1 and Cdc42, which are intracellular signals that regulate mechanical components to coordinate cell shape and tissue geometry. Dpp controls Rho1 through activation by pMad and inhibition by Brk. Rho1 and Cdc42 regulate non-muscle myosin II and actin depolymerization, which together control actomyosin contractility, forming the link from chemical to mechanical properties within the tissue [16].

Understanding the regulation of Rho1 and Cdc42 is essential in understanding the development of the shape of the wing disc and therefore the development of a proper adult
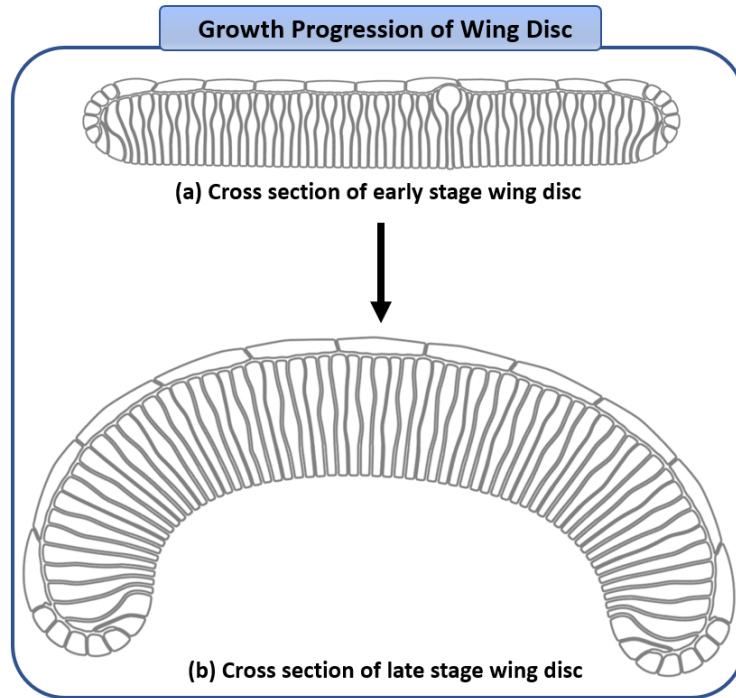
Figure 2.1: Growth progression of *Drosophila* wing imaginal disc. Simulated tissue shapes are produced from existing subcellular element model [19, 25]. (a) Cross section of early stage wing disc. (b) Cross section of late stage wing disc.
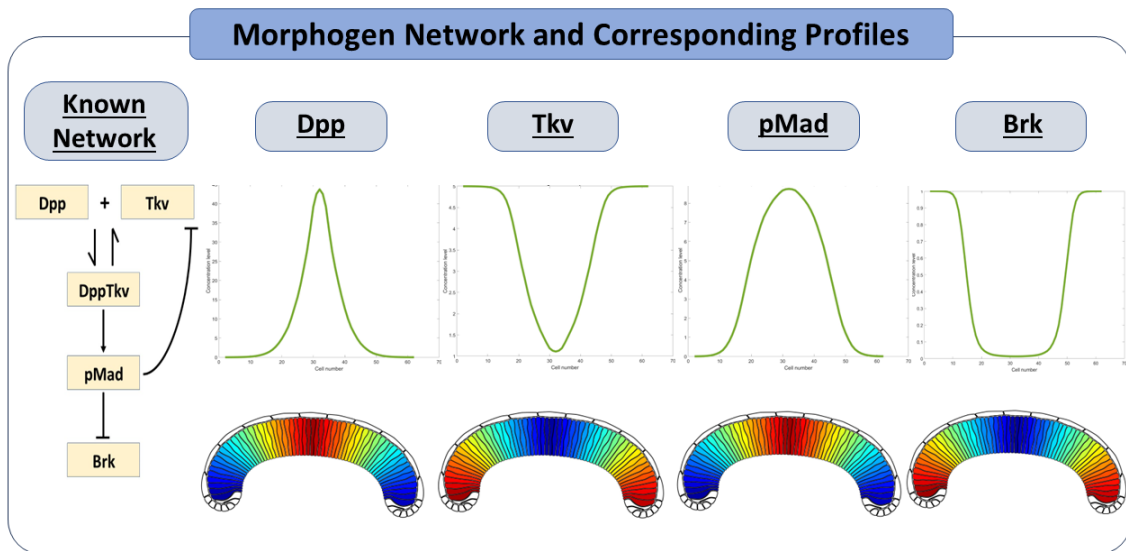
Figure 2.2: Known morphogen network involved in development and patterning of morphogens in the cross section of the wing disc.

wing. Much has been accomplished to investigate the regulation of Rho1. However, there is still much work to be done, in addition to investigating the regulation of Cdc42 in more detail. Thus, the goal of this project is to identify the interactions in the Dpp-Rho1-Cdc42 network through a combination of experimental and modeling techniques. By utilizing new and existing experimental results, mathematical models can be formulated. Calibration and testing is then performed on those models using experimental data. Doing so allows reformulation of ideas based on the results, which promotes convergence on a model that best explains the experimental data. We use this process to investigate the Dpp-Rho1-Cdc42 network in order to connect chemical and cytoskeletal regulations in the wing disc.

2.3 Modeling Background

Multiscale mathematical models are developed using a system of reaction-diffusion equations to model the dynamics of the Dpp-Rho1-Cdc42 network. The modeling framework used is similar between all biological networks tested so the following discussion is a general case that can be applied to all chemical signaling networks in Section 2.4. Let molecule A be a tissue-level morphogen (such as Dpp) and let molecule B and molecule C be two intracellular signals (such as Rho1 and Cdc42).

The signaling model uses a discrete node structure from an existing multiscale subcellular element model for the cross section view of the *Drosophila* wing disc [19, 25]. Reference Fig 2.1 for two different discretized tissue shapes, with Fig 2.1(b) being the shape used in the simulations presented in Section 2.4. The apical nodes for all columnar cells are used to model the tissue-level morphogen dynamics (molecule A). Additionally, since all columnar cells are elongated in the cross section view, the width of individual cells is ignored and the distributions of molecule B and C are only considered along the apical-basal axis, i.e., the cell midline, of each cell.

Equations representing the dynamics of molecules A, B, and C are all of similar type:

$$\frac{\partial[A]}{\partial t} = D_A \Delta_s^{tis}[A] + R(A)$$

$$\frac{\partial[B]}{\partial t} = D_B \Delta_s^{cell}[B] + R(B) \tag{2.1}$$

$$\frac{\partial[C]}{\partial t} = D_C \Delta_s^{cell}[C] + R(C)$$

where

- $[X]$ represents the concentration of molecule X

- D_X is the diffusion coefficient for molecule X

- Δ_s^{tis} and Δ_s^{cell} are the diffusion operators at the tissue level (i.e., along the one-dimensional curve comprised of the apical nodes of the columnar cells) and the cell level (i.e., along the midline of each cell), respectively

- $R(X)$ accounts for all reactions involving molecule X

Since diffusion is done in one dimension, the numerical approximation for the diffusion operator is simple. However, the curves on which diffusion is performed (i.e., either the line of apical nodes or midlines of each cell) are not straight so the following approximation is used:
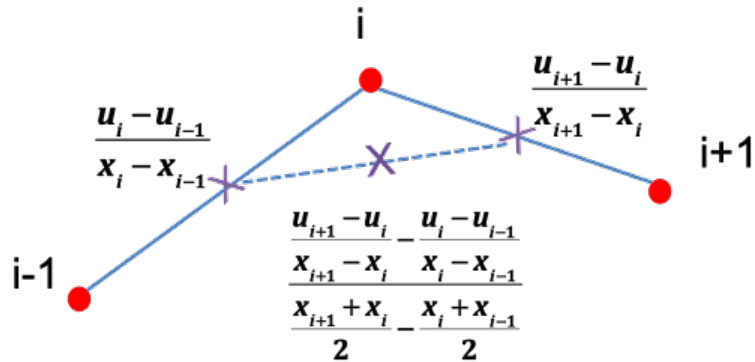
Figure 2.3: Numerical approximation of one-dimensional diffusion operator.

The reaction term for molecule A includes activation and degradation. At the morphogen level, these reaction terms are usually space-dependent, meaning that activation rates of molecule A depend on the location in the tissue. This will be further explained in the next section. The reaction-diffusion equation for molecule A is run to steady state before the equations for molecules B and C are evolved.

As for the intracellular components, molecules B and C, every cell has its own system of equations since molecules B and C do not diffuse between cells. Molecules B and C are produced at both the apical node and basal node of each cell and then diffused into the interior or the cell (Fig 2.4). Production is modeled using Hill functions that are in terms of the regulating components (i.e., tissue-level molecule A, the other intracellular molecule, etc.). When the concentration of molecule A is present in the production term for molecule B and C, the steady state values of molecule A are averaged across the apical nodes for a cell and used in the Hill function for that cell at both apical and basal nodes.
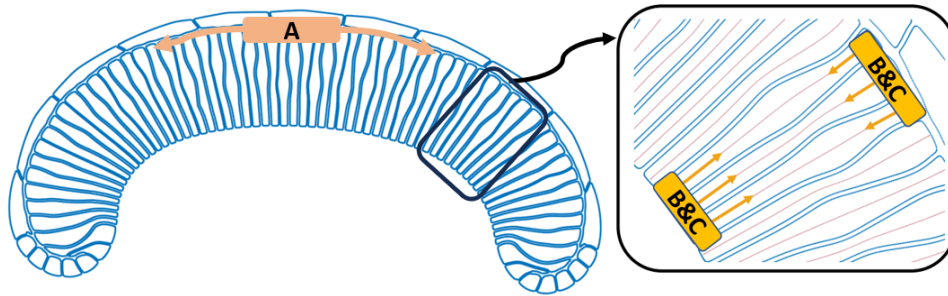
Figure 2.4: Node structure and illustration of one-dimensional curves where the reaction-diffusion equations are solved.

This modeling framework is the basis for which different signaling networks are tested, regulations are explored, and sensitivity analysis is performed.

2.4 Model Selection, Parameter Estimation, and Sensitivity Analysis

At a qualitative level, Dpp is known to regulate Rho1 in the wing disc and is likely an important regulator of cell mechanics as a key mechanogen [16]. However, how Dpp regulates other Rho-GTPases such as Cdc42 in context to tissue growth or morphogenesis remains unclear. Therefore, members of Dr. Zartman's lab investigated and quantified the impact of perturbing Dpp/pMad signaling on the localized distribution of Rho1 and Cdc42 in the wing disc. To do so, a constitutively active form of thickvein (Tkv) receptor (Tkv$^{\text{CA}}$) was expressed in the posterior compartment of the wing disc. Dpp binds to Tkv receptors and pMad, resulting in an increase of pMad in the nucleus. As expected a higher fluorescence of pMad was observed in the posterior compartment (Fig 2.5 A-ii', B-ii'). An antibody staining was then carried out for both Rho1 and Cdc42. Rho1 showed a statistically significant increase on the basal surface of the posterior compartment (Fig 2.5 A-iii'). As for Cdc42, there was an increase in both apical and basal levels in the posterior compartment (Fig 2.5 B-iii').

In summary, the results demonstrate that overexpression of Tkv in the posterior compartment of the *Drosophila* wing disc leads to a non-compartment specific increase in Rho1 across the basal surface and a compartment-specific increase in Cdc42 at both apical and basal surfaces of the pouch. A further study on if and how the Rho-GTPases regulate the expression of each other is required to fully understand the complex interaction between Dpp and Rho-GTPases such as Rho1 and Cdc42.

For establishing the closed-loop network between Dpp, Rho1 and Cdc42, an understanding of existing interactions between Rho-GTPases is crucial. In previous studies, Rho-GTPases often have been shown to suppress each other, but the nature of interaction has been reported differently across systems. An insufficient amount of work has been done to understand this inhibition in the context of *Drosophila* wing disc morphogenesis, and mutual inhibition between Rho1 and Cdc42 in the wing disc has not been previously reported in literature.

Dpp Regulates Localization of Rho1 and Cdc42
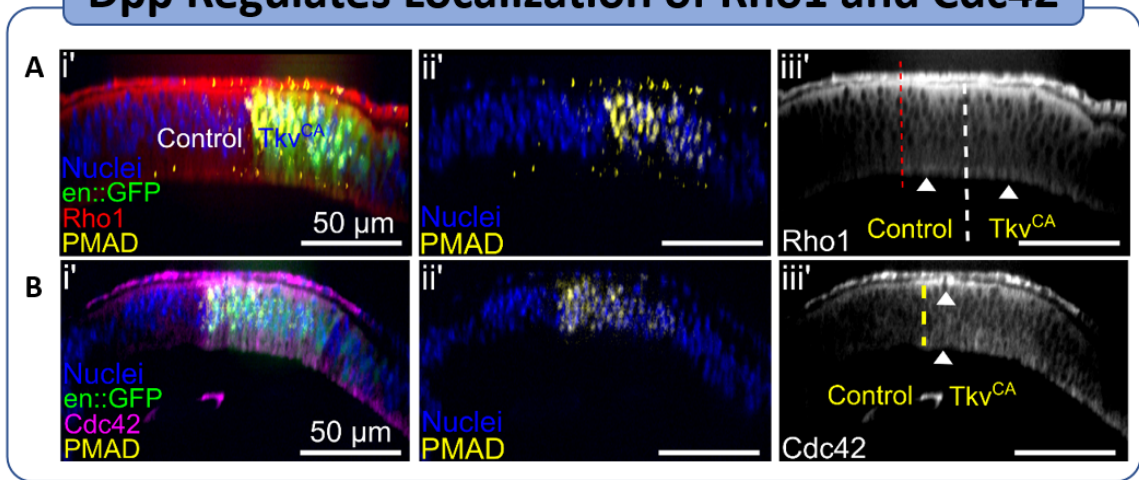
Figure 2.5: Dpp signaling regulates subcellular localization of both Rho1 and Cdc42. Images courtesy of Dr. Nilay Kumar.

To investigate the interaction between Rho1 and Cdc42, the experimentalists blocked the activity of the Rho-GTPases one at a time using the Gal4 UAS system and measured the expression of the other using immunohistochemistry assays.

As a first study, to test if Rho1 regulates Cdc42, Rho1$^{\text{RNAi}}$ was expressed in the posterior compartment of the wing disc, followed by a Cdc42 antibody staining (Fig 2.6 B). A knockdown of Rho1 in the posterior compartment led to an increase in basal localization of Cdc42 in a compartment-specific manner (Fig 2.6 B-ii'). A strong increase in basal levels of Cdc42 was seen, while there was no significant change in the apical patterning of Cdc42.

To study the interaction in the opposite direction, a dominant negative form of Cdc42 (Cdc42DN) was introduced in the posterior compartment and Rho1 antibody staining was carried out (Fig 2.6 A). A strong phenotypic change in terms of wing disc morphology was reported, along with a reduction in the height of the pouch cells (Fig 2.6 A-ii'). Along with the shortening of cells, Rho1 expression was seen to be more uniform across the apical and basal surfaces compared to the wildtype Rho1 profiles. Additionally, there was a significant
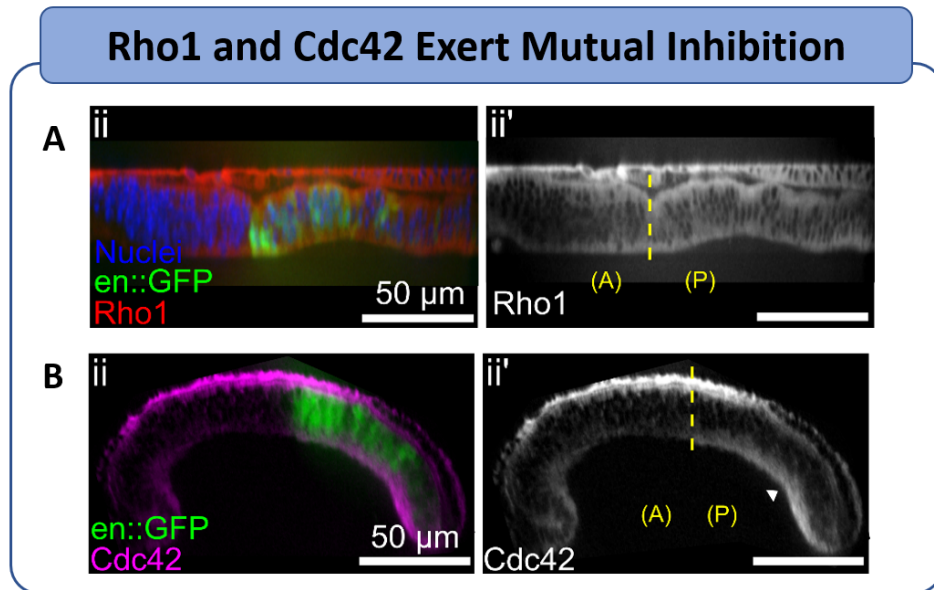
Figure 2.6: Rho1 and Cdc42 mutually inhibit each other. Images courtesy of Dr. Nilay Kumar.

increase in basal levels of Rho1 when Cdc42 was inhibited. Thus, it can be concluded that Rho1 and Cdc42 mutually inhibit each other in the *Drosophila* wing disc.

Since Rho1 and Cdc42 are directly linked to cytoskeletal components that control the shape of the wing disc, it is beneficial to be aware of the wildtype patterning of the Rho-GTPases. To do so, late stage wing discs were imaged and immunofluorescence was carried out to visualize the spatial localization of Rho1 and Cdc42 (Fig 2.7 A, B). The results indicate that both Rho1 and Cdc42 primarily localize along the apical and basal surfaces of the wing disc. However, the manner in which they localize differs between each Rho-GTPase. Rho1 is predominantly localized to the center of the apical surface. Cdc42, on the other hand, is predominantly localized to the basal surface and is highly concentrated near the center and regions of high curvature of the tissue. The intensity profiles of Rho1 and Cdc42 were extracted, smoothed, and normalized to produce the plots in Fig 2.7 B and C. Data for each species is normalized by dividing values by the maximum value of that species. One high concentration peak can be seen for apical Rho1, while there are two less significant
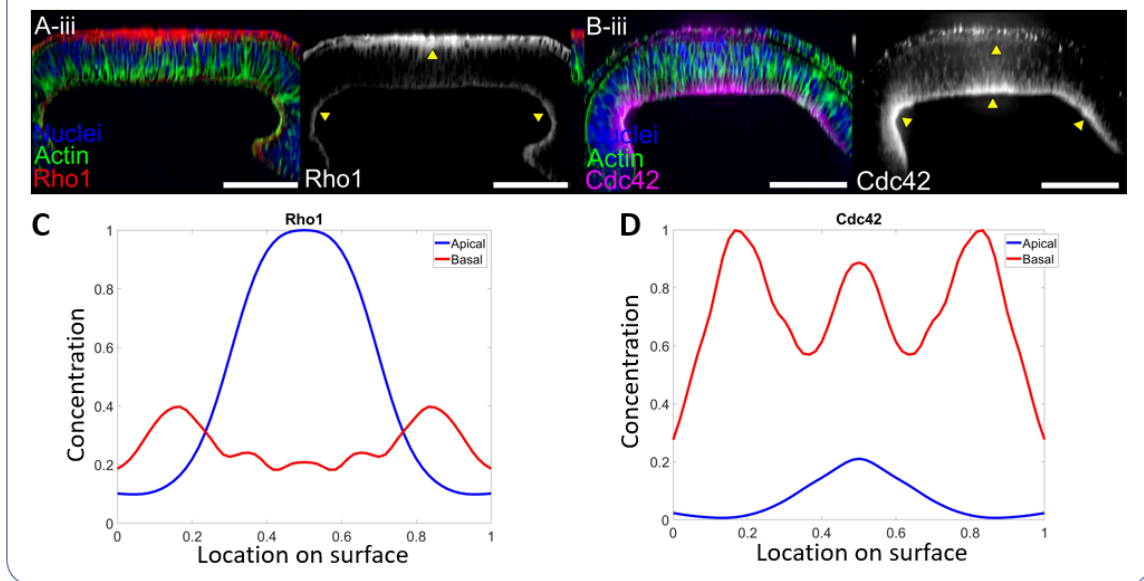
Figure 2.7: Wildtype concentration profiles for Rho1 and Cdc42. Rho1 is apically concentrated while Cdc42 is basally localized. An optical section showing Rho1 (A-iii) and Cdc42 (B-iii) are above the plots exhibiting the quantification of the concentration along the apical (blue) and basal (red) surfaces. Images courtesy of Dr. Nilay Kumar.

concentration peaks for basal Rho1. Similarly, there are three high concentration peaks for basal Cdc42 and one less significant concentration peak for apical Cdc42.

With these new and existing results in mind, a model is developed to test if these regulations alone are sufficient to reproduce the wildtype profiles of Rho1 and Cdc42. The network in Fig 2.8 shows the interactions between Dpp, Rho1, and Cdc42 that are tested. Dpp activates both Rho1 and Cdc42, while Rho1 and Cdc42 inhibit each other.
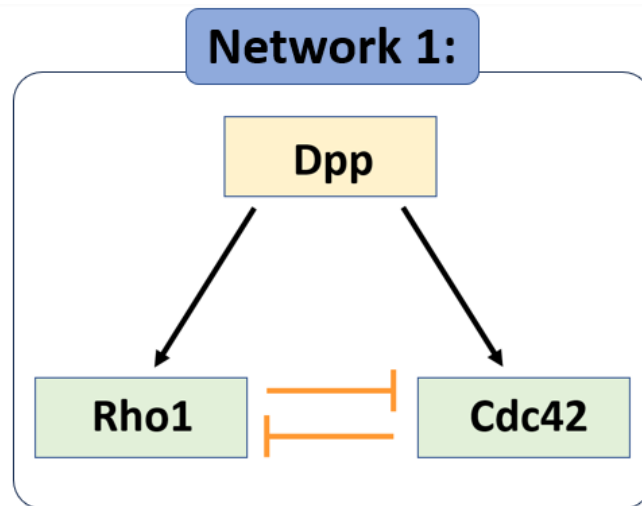
Figure 2.8: First network investigated. Tan boxes represent tissue-level morphogens and green boxes indicate intracellular mechanogens. Black arrows signify regulations that are fixed and orange arrows represent regulations perturbed in the parameter estimation/sensitivity analysis process. Lines with pointed arrows indicate activation and lines with perpendicular ends indicate inhibition.

Using the modeling framework described in Section 2.3, a system of reaction-diffusion equations (Eqn 2.2) describing the regulations in network 1 are developed. Assumptions of the developed model are as follows: 1) Dpp concentrations are identical along the apical and basal surfaces, 2) the regulation of Rho1 and Cdc42 by Dpp and the other Rho-GTPase is the same for both the apical and basal surfaces, 3) the source term for basal Rho1 is half of the source term for apical Rho1, and the source term for apical Cdc42 is half of the source term for basal Cdc42, and 4) the initial concentrations for Rho1 and Cdc42 is zero for all nodes in the tissue. The tissue shape shown in Fig 2.1(b) is used in all simulations.

Model 1:

$$\frac{\partial [Dpp]}{\partial t} = D_d \Delta_s^{tis}[Dpp] + s_d(x,y) - d_d[Dpp]$$

$$\frac{\partial [Rho1]}{\partial t} = D_r \Delta_s^{cell}[Rho1] - d_r[Rho1]$$

$$+ s_{ra} \cdot I_{apical} \cdot \left(\frac{c_{dr,max}}{1 + ([Dpp]/k_{dr})^{-n_{dr}}}\right)\left(\frac{c_{cr,max}}{1 + ([Cdc42]/k_{cr})^{n_{cr}}}\right)$$

$$+ s_{rb} \cdot I_{basal} \cdot \left(\frac{c_{dr,max}}{1 + ([Dpp]/k_{dr})^{-n_{dr}}}\right)\left(\frac{c_{cr,max}}{1 + ([Cdc42]/k_{cr})^{n_{cr}}}\right) \tag{2.2}$$

$$\frac{\partial [Cdc42]}{\partial t} = D_c \Delta_s^{cell}[Cdc42] - d_c[Cdc42]$$

$$+ s_{ca} \cdot I_{apical} \cdot \left(\frac{c_{dc,max}}{1 + ([Dpp]/k_{dc})^{-n_{dc}}}\right)\left(\frac{c_{rc,max}}{1 + ([Rho1]/k_{rc})^{n_{rc}}}\right)$$

$$+ s_{cb} \cdot I_{basal} \cdot \left(\frac{c_{dc,max}}{1 + ([Dpp]/k_{dc})^{-n_{dc}}}\right)\left(\frac{c_{rc,max}}{1 + ([Rho1]/k_{rc})^{n_{rc}}}\right)$$

In Eqn 2.2, there is a reaction-diffusion equation representing each of the three components in network 1. Dpp is produced in the center of the tissue (indicated by the source term $s_d(x,y)$), diffuses bilaterally across the apical nodes of the tissue ($D_d \Delta_s^{tis}[Dpp]$), and degrades at the rate d_d. Rho1 and Cdc42, on the other hand, are produced at the first and last node of each cell and diffused into the interior of the cell. The source region and diffusion of Dpp, Rho1, and Cdc42 are illustrated in Fig 2.4 by A, B, and C, respectively. The source terms for Rho1 and Cdc42 are decomposed into apical and basal activation by using indicator functions I_{apical} and I_{basal}. The regulations by Dpp and the other Rho-GTPase are incorporated using Hill functions, where the sign of the exponent dictates if the regulation is activation or inhibition.

With the biological network now described by differential equations, calibration of Eqn 2.2 is needed to determine the correct parameter values to reproduce the wildtype Rho1 and Cdc42 profiles. Some parameter values are attained from existing literature (see values in Appendix A.2). However, the 4 parameters in the model representing the newfound biological regulations (represented in blue color in Eqn 2.2) need to be calibrated. To do so, a uniform sampling of the parameter space is produced and the model is run for each of those parameter sets. For each result, the number of high concentration peaks produced

are counted and shown in the heat maps in Fig 2.9. In this figure, the plots are separated into 4 different sections: apical Rho1, apical Cdc42, basal Rho1, and basal Cdc42. For each section of plots, the parameter c_{cr_max} is perturbed in the columns of plots and k_{cr} is perturbed in the rows of the plots (indicated by the large purple axes). In each subplot, k_{rc} and c_{rc_max} are perturbed in the minor axes. The color depicts the number of peaks produced for that specific parameter set. In this way the results of the four-dimensional parameter perturbation is shown.
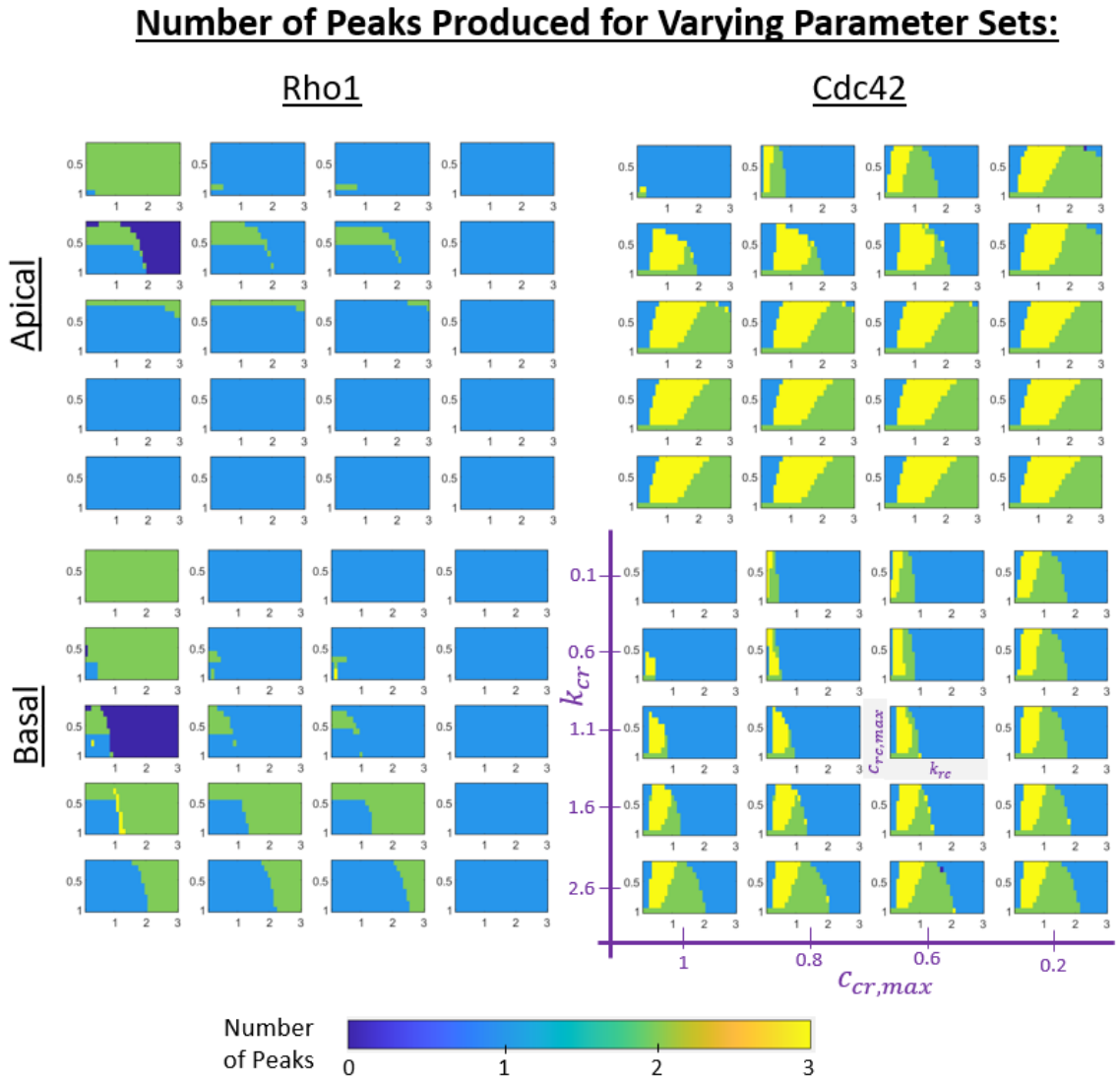
Figure 2.9: Sensitivity analysis for network 1.

By examining Fig 2.9, a few trends can be seen in the choice of parameter sets. First, as k_{cr} increases, the proportion of parameter sets that produce a higher number of Cdc42 peaks grows. Also, lower values of k_{cr} and higher values of c_{cr_max} produce 2 peaks in basal Rho1 (the desired number of peaks). And lastly, the number of peaks produced is more sensitive to k_{cr} and k_{rc}.

Using the parameter exploration in Fig 2.9, we can find the regions of parameter sets that produce the correct number of peaks for all 4 wildtype profiles. That is, one peak for both apical Rho1 and apical Cdc42, two peaks for basal Rho1, and three peaks for basal Cdc42. Extracting parameter sets that meet that criteria, we can examine the qualitative behavior of the results, i.e., do the model results resemble the experimental results. A representative example of profiles that satisfy the correct number of peaks criteria for model 1 is shown in Fig 2.10. The Dpp profile shown in Fig 2.10 is used in all simulations for model 1. For Rho1 and Cdc42, the blue curves represent the apical profiles and the red curves represent the basal profiles. Comparing these curves to the wildtype experimental profiles in Fig 2.7 shows the qualitative difference between the model 1 outputs and the experimental results. Model 1 can capture the correct number of peaks for all 4 profiles. However, model 1 is unable to produce the correct location of basal Rho1 peaks, produce clearly defined basal Rho1 peaks, or capture the correct width of basal Cdc42 peaks. Model 1 is also incapable of reproducing the mutant cases detailed above.

With this in mind, the biological regulations incorporated into the model need to be revisited to determine if a critical component necessary for the production of correct peak behavior was ignored. As stated previously, it is known that Dpp regulates both Rho1 and Cdc42 through activation by pMad, and Dpp also regulates Rho1 through Brk [16]. To test the regulation of Cdc42 by Brk, the Gal4-UAS system was used to inhibit Brk with Brk[RNAi] in the posterior compartment of the wing disc using engrailed-Gal4 driver for Cdc42 antibody analysis. This led to compartment-specific increase in Cdc42 in both the apical and basal surfaces (Fig 2.11). The inhibition of both Rho1 and Cdc42 by Brk is an important component that can be introduced into the model.
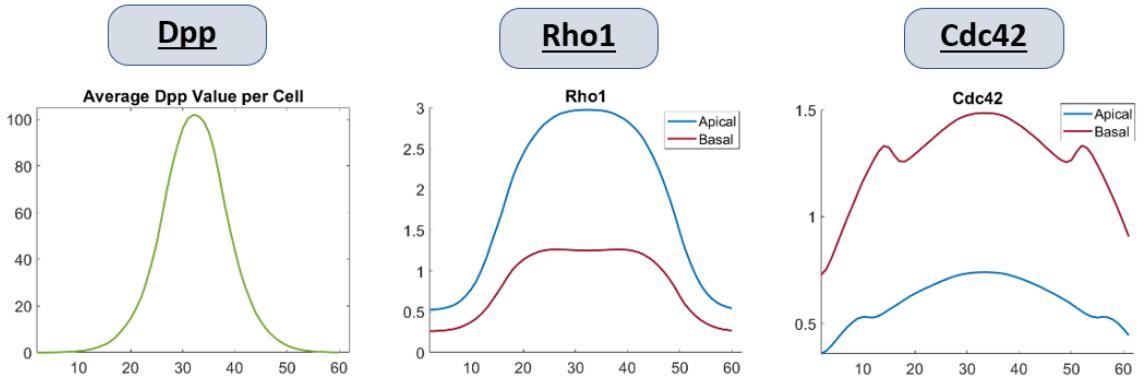
Figure 2.10: Resulting profiles of apical and basal Rho1 and Cdc42 from model 1 using a parameter set that produces the correct number of peaks for all 4 wildtype profiles.
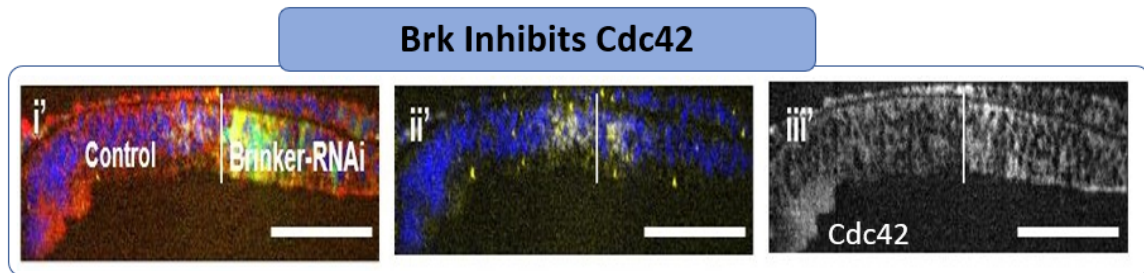
Figure 2.11: Brk inhibits Cdc42. Inhibition of Brk in the posterior compartment led to compartment-specific increase in Cdc42 in both apical and basal surfaces. Images courtesy of Mayesha S. Mim.

Taking a closer look at the mutant case presented in Fig 2.6(B), it can be seen that Rho1 only significantly inhibits Cdc42 on the basal surface. When Rho1 was inhibited, basal Cdc42 increased while there was no significant increase in apical Cdc42 values. This suggests that not only are the apical and basal interactions of Rho1 and Cdc42 different, but that all regulations on the apical and basal surfaces need not be the same. This fact is an important component to incorporate into the models.

With these results in mind, the next phase of networks and corresponding models are developed (Fig 2.12). Dpp, Rho1 and Cdc42 are modeled in a similar fashion to network 1. Tkv, DppTkv complex, pMad, and Brk are produced on the apical surface of the tissue but not diffused. Equations for the tissue level components are as follows:

Morphogen equations for networks 2-4:

$$\frac{\partial[Dpp]}{\partial t} = D_d \Delta_s^{tis}[Dpp] + s_d(x,y) - d_d[Dpp] - k_{on}[Dpp][Tkv] + k_{off}[DppTkv]$$

$$\frac{\partial[Tkv]}{\partial t} = \left(\frac{v_t}{1+([pMad]/k_{pt})^{n_{pt}}}\right) - d_t[Tkv] - k_{on}[Dpp][Tkv] + k_{off}[DppTkv]$$

$$\frac{\partial[DppTkv]}{\partial t} = k_{on}[Dpp][Tkv] - k_{off}[DppTkv] - d_{dt}[DppTkv] \tag{2.3}$$

$$\frac{\partial[pMad]}{\partial t} = \frac{v_p}{1+([DppTkv]/k_{dtp})^{-n_{dtp}}} - d_p[pMad]$$

$$\frac{\partial[Brk]}{\partial t} = \frac{v_b}{1+([pMad]/k_{pb})^{n_{pb}}} - d_b[Brk]$$

These equations are run to steady state for a fixed parameter set that produces the profiles in Fig 2.2. Parameter values can be found in Table A.2 in Appendix A.2. Assumptions for networks 2-4 are: 1) all tissue level components have identical profiles on apical and basal surfaces, 2) regulation of Rho1 and Cdc42 is not assumed to be the same on the apical and basal surfaces, and 3) the morphogen value used for pMad and Brk in the Rho1 and Cdc42 equations for each cell is the average value of the pMad and Brk concentrations on the apical nodes for that cell.

Starting with network 2, it is assumed that Dpp and pMad regulate Rho1 and Cdc42 through Brk. Brk directly regulates both apical and basal Rho1 and Cdc42. For the apical components, inhibition is no longer mutual. That is, only Cdc42 inhibits Rho1 on the apical surface. Mutual inhibition is still present in the basal surface. These interactions are modeled in the following Rho1 and Cdc42 equations.
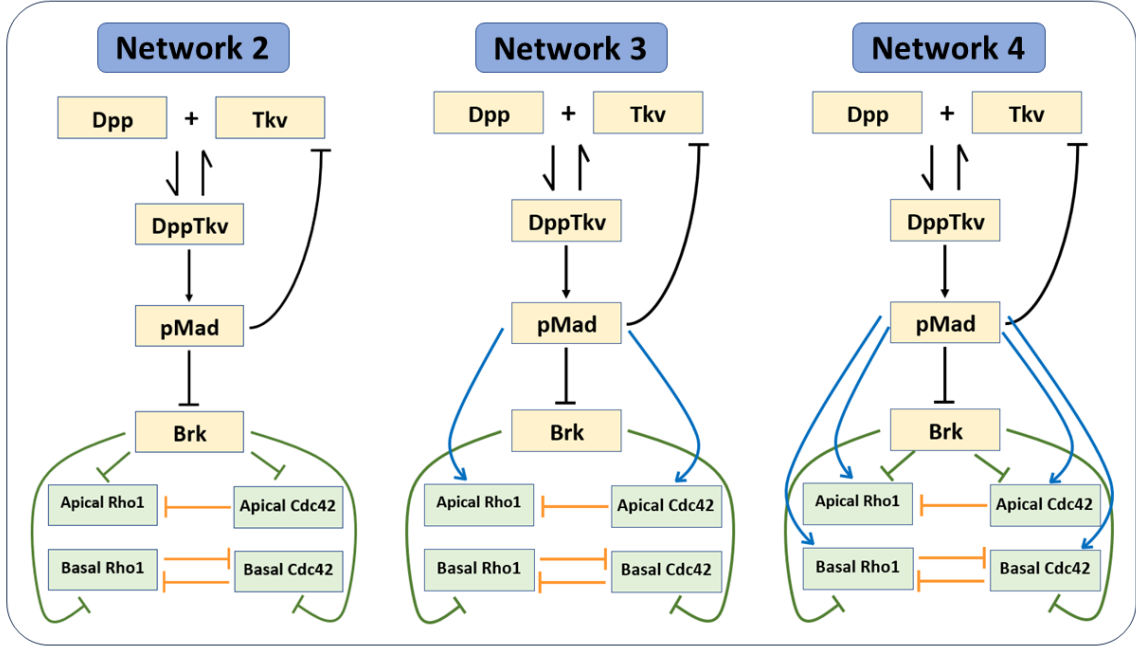
Figure 2.12: Network diagrams for the second phase of model testing, networks 2 through 4.

Model 2:

$$
\begin{aligned}
\frac{\partial [Rho1]}{\partial t} &= D_r \Delta_s^{cell}[Rho1] - d_r[Rho1] \\
&+ s_{r\_ap} \cdot I_{apical} \cdot \left(\frac{1}{1 + ([Brk]/k_{br\_ap})^{n_{br}}} \right) \left(\frac{1}{1 + ([Cdc42]/k_{cr\_ap})^{n_{cr}}} \right) \\
&+ s_{r\_ba} \cdot I_{basal} \cdot \left(\frac{1}{1 + ([Brk]/k_{br\_ba})^{n_{br}}} \right) \left(\frac{1}{1 + ([Cdc42]/k_{cr\_ba})^{n_{cr}}} \right) \\
\frac{\partial [Cdc42]}{\partial t} &= D_c \Delta_s^{cell}[Cdc42] - d_c[Cdc42] \\
&+ s_{c\_ap} \cdot I_{apical} \cdot \left(\frac{1}{1 + ([Brk]/k_{bc\_ap})^{n_{bc}}} \right) \\
&+ s_{c\_ba} \cdot I_{basal} \cdot \left(\frac{1}{1 + ([Brk]/k_{bc\_ba})^{n_{bc}}} \right) \left(\frac{1}{1 + ([Rho1]/k_{rc\_ba})^{n_{rc}}} \right)
\end{aligned}
\tag{2.4}
$$

For network 3, the direct regulation of apical Rho1 and apical Cdc42 by Brk is removed and replaced with activation by pMad. This is due to the fact that wildtype apical Rho1 and Cdc42 closely resemble the pMad profile. So Brk is removed from the apical terms in the Rho1 and Cdc42 equations and replaced by pMad, as seen in the following equations.

Model 3:

$$\frac{\partial [Rho1]}{\partial t} = D_r \Delta_s^{cell}[Rho1] - d_r[Rho1]$$

$$+ s_{r\_ap} \cdot I_{apical} \cdot \Big(\frac{1}{1 + ([pMad]/k_{pr\_ap})^{-n_{pr}}}\Big)\Big(\frac{1}{1 + ([Cdc42]/k_{cr\_ap})^{n_{cr}}}\Big)$$

$$+ s_{r\_ba} \cdot I_{basal} \cdot \Big(\frac{1}{1 + ([Brk]/k_{br\_ba})^{n_{br}}}\Big)\Big(\frac{1}{1 + ([Cdc42]/k_{cr\_ba})^{n_{cr}}}\Big)$$

$$\frac{\partial [Cdc42]}{\partial t} = D_c \Delta_s^{cell}[Cdc42] - d_c[Cdc42]$$

$$+ s_{c\_ap} \cdot I_{apical} \cdot \Big(\frac{1}{1 + ([pMad]/k_{pc\_ap})^{-n_{pc}}}\Big)$$

$$+ s_{c\_ba} \cdot I_{basal} \cdot \Big(\frac{1}{1 + ([Brk]/k_{bc\_ba})^{n_{bc}}}\Big)\Big(\frac{1}{1 + ([Rho1]/k_{rc\_ba})^{n_{rc}}}\Big) \tag{2.5}$$

Finally for network 4, both activation by pMad and inhibition by Brk is applied to both apical and basal Rho1 and Cdc42.

Model 4:

$$\frac{\partial [Rho1]}{\partial t} = D_r \Delta_s^{cell}[Rho1] - d_r[Rho1]$$

$$+ s_{r\_ap} \cdot I_{apical} \cdot \Big(\frac{1}{1 + ([pMad]/k_{pr\_ap})^{-n_{pr}}}\Big)$$

$$\cdot \Big(\frac{1}{1 + ([Brk]/k_{br\_ap})^{n_{br}}}\Big)\Big(\frac{1}{1 + ([Cdc42]/k_{cr\_ap})^{n_{cr}}}\Big)$$

$$+ s_{r\_ba} \cdot I_{basal} \cdot \Big(\frac{1}{1 + ([pMad]/k_{pr\_ba})^{-n_{pr}}}\Big)$$

$$\cdot \Big(\frac{1}{1 + ([Brk]/k_{br\_ba})^{n_{br}}}\Big)\Big(\frac{1}{1 + ([Cdc42]/k_{cr\_ba})^{n_{cr}}}\Big)$$

$$\frac{\partial [Cdc42]}{\partial t} = D_c \Delta_s^{cell}[Cdc42] - d_c[Cdc42]$$

$$+ s_{c\_ap} \cdot I_{apical} \cdot \Big(\frac{1}{1 + ([pMad]/k_{pc\_ap})^{-n_{pc}}}\Big)\Big(\frac{1}{1 + ([Brk]/k_{bc\_ap})^{n_{bc}}}\Big)$$

$$+ s_{c\_ba} \cdot I_{basal} \cdot \Big(\frac{1}{1 + ([pMad]/k_{pc\_ba})^{-n_{pc}}}\Big)$$

$$\cdot \Big(\frac{1}{1 + ([Brk]/k_{bc\_ba})^{n_{bc}}}\Big)\Big(\frac{1}{1 + ([Rho1]/k_{rc\_ba})^{n_{rc}}}\Big) \tag{2.6}$$

We want to determine which, if any, of networks 2-4 can reproduce the wildtype behavior of Rho1 and Cdc42. From the successful models, the parameter space can then be refined

using the mutant cases. Parameter estimation must be done on all 3 models. As seen in model 1, calibrating the model to simply achieve the correct number of peaks is not sufficient. Factors such as peak location, height and width also need to be captured. Because of this and the increased number of parameters involved in models 2-4, a more sophisticated approach is needed for parameter estimation. The number of parameters increased from 4 perturbed parameters in model 1 to 11 or 14 parameters in models 2-4 (once again emphasized in blue in the model equations).

In order to produce results that are qualitatively similar to the experimental profiles, the peak and valley location, height and width need to be captured, as well as the relative concentration levels. In other words, the difference between the simulated output curve and the experimental profile curve needs to be sufficiently small. There are many different error metrics that can be used but we consider 3 possibilities: Frechet distance, mean-squared error, and sum of squared error. To determine which error metric to use in the parameter estimation process, we pick two simulated profiles to compare to the experimental profile (reference Fig 2.13). The first curve exhibits one wide peak and the second curve has 3 peaks, 2 narrow peaks on the sides of one wide center peak. We compare the three error metrics to see which one identifies the profile with 3 peaks as the profile with the smallest error. As seen in Fig 2.13, the sum of squared errors (SSE) clearly indicates that the profile with 3 peaks (profile 2) is the curve that is most similar to the desired experimental profile. Thus, the parameter estimation technique used must be capable of selecting parameter sets that minimize the SSE between the model output and the experimental Rho1 and Cdc42 wildtype profiles.

The first parameter estimation approach we investigate is Bayesian optimization (BO) [27]. This method uses a probabilistic surrogate model that runs faster than the physics-based model to test a large number of parameter sets quickly. The goal of BO is to find the input that produces the optimal output of an expensive, "black-box" objective function, which in this case is the chemical signaling model (models 2-4). Details of the BO algorithm used on models 2-4 are as follows:

Profile #	Frechet distance	Mean-squared error	Sum of squared errors
1	0.11519	0.0359	2.1899
2	0.11786	0.0097219	0.59304
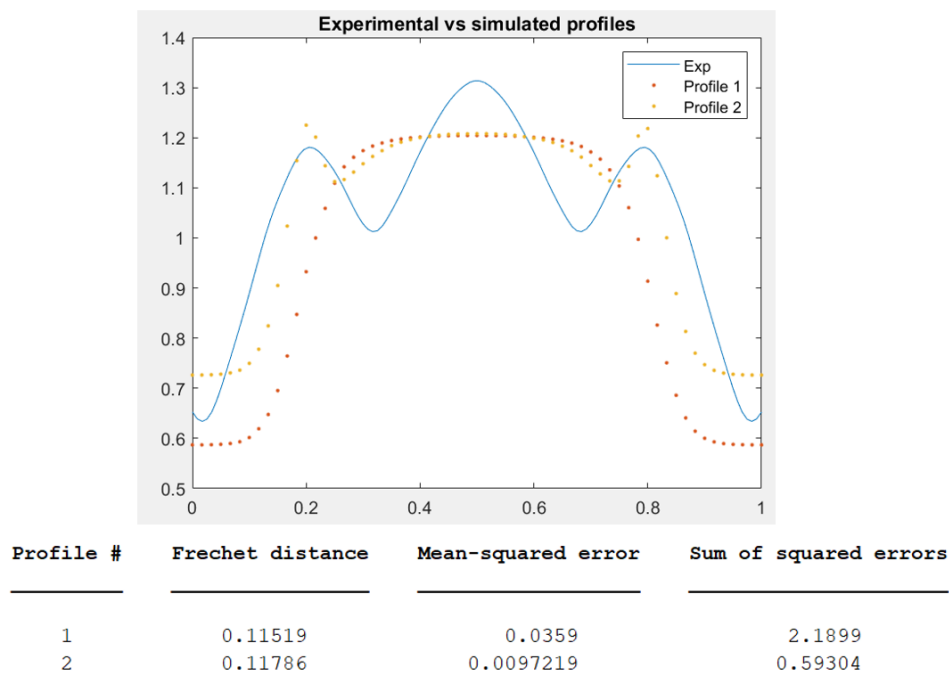
Figure 2.13: Brief error metric comparison to be used in parameter estimation techniques for models 2-4.

Step 1: Produce initial set of training data by running the chemical signaling model on a large number of parameter sets. A good rule of thumb for determining this number of starting parameter sets is for every 10 parameters perturbed, about 200 training data sets are needed. Once the model is run on each of the initial parameter sets that sufficiently covers the parameter space, the training data is generated by calculating the sum of squared error between the normed experimental data and the normed chemical signaling model outputs. Both experimental data and model outputs are normed in the same fashion: apical and basal Rho1 values are normed with respect to the maximum apical Rho1 value, and apical and basal Cdc42 values are normed with respect to the maximum basal Cdc42 value.

Step 2: Use the results from the physics-based model to train the Gaussian process regression (GPR) model. The MATLAB function "fitrgp" is utilized for this procedure.

Step 3: Sample a large number of parameter sets (on the scale of one to two million) and run the trained GPR model on these parameter sets. The "predict" function in MATLAB is used for this step. Use sampling techniques such as random, Sobol, or Latin hybercube sampling. For models 2-4, random sampling is used.

Step 4: Apply an acquisition function to select the next parameter sets to evaluate with the chemical signaling model. The Expected Improvement (EI) acquisition function is used. EI measures how much better the objective function value at a given point is compared to the current best observed value. It balances the exploitation of known promising regions of the search space with the exploration of uncertain regions. The EI is defined as [28]:

$$EI(x) = \mathbb{E} max(f(x) - f(x^+), 0) \tag{2.7}$$

where $f(x^+)$ is the value of the best sample so far. The EI is evaluated as:

$$EI(x) = \begin{cases} (\mu(x) - f(x^+) - \xi)\Phi(Z) + \sigma(x)\phi(Z) & \text{if } \sigma(x) > 0 \\ 0 & \text{if } \sigma(x) = 0 \end{cases} \tag{2.8}$$

where

$$Z = \begin{cases} \frac{\mu(x) - f(x^+) - \xi}{\sigma(x)} & \text{if } \sigma(x) > 0 \\ 0 & \text{if } \sigma(x) = 0 \end{cases} \tag{2.9}$$

and $\mu(x)$ and $\sigma(x)$ are the mean and standard deviation of the GPR posterior predictive at x, Φ and ϕ are the cumulative distribution function and probability

density function of the standard normal distribution, and ξ is a parameter that determines the amount of exploration during optimization.

Step 5: The parameter sets with the 5 highest EI values are selected, run through the chemical signaling model, and inserted into the training data. Steps 2-4 are then repeated.
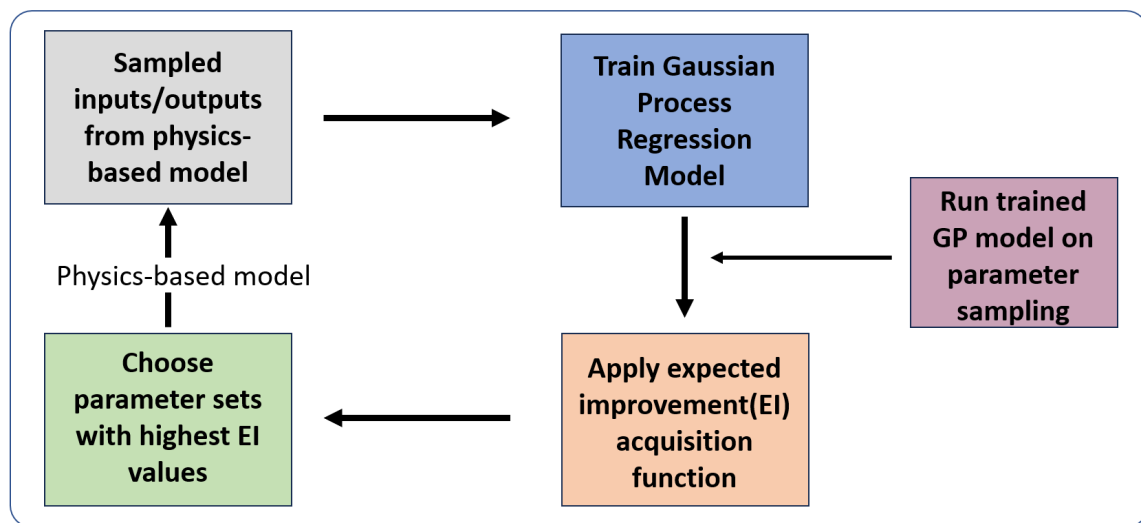
Figure 2.14: Summary of Bayesian optimization algorithm.

With the general algorithm for BO established, we then apply the optimization framework to each of models 2-4. Starting with model 2, about 2050 training data sets are generated using a uniform sampling of the parameter space for the 11 perturbed parameters. A value of $\xi = 0.05$ is used for the exploration parameter. One hundred iterations of BO are run, each testing 1,000,000 random parameter sets with the trained GPR model (refer to Appendix A.2 for parameter ranges). The 5 resulting profiles with the lowest SSE are shown in Fig 2.15.

Examining the profiles produced by the 5 parameter sets with the lowest error shows that network/model 2 is incapable of producing the desired results. The apical Rho1 and Cdc42 profiles exhibit one peak for each parameter set but they are more square in shape than the experimental profiles. Basal Rho1 is able to produce 2 peaks in roughly the correct location
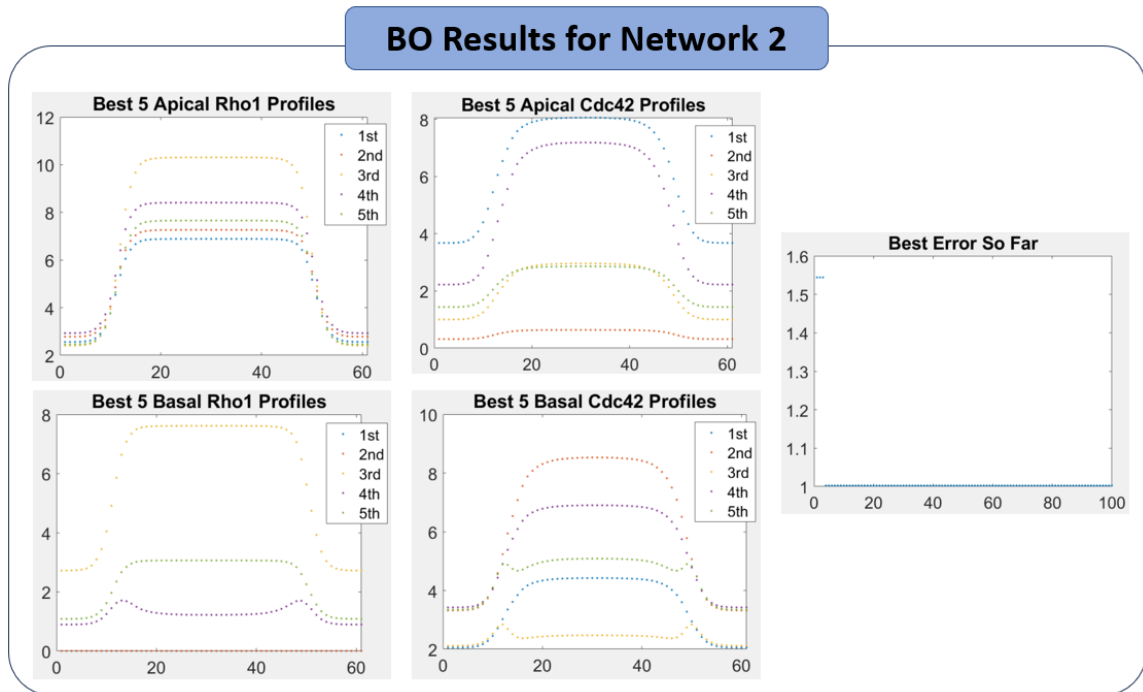
Figure 2.15: Results of BO for network 2. The 5 parameter sets that produce the lowest error are plotted in different colors for each parameter set. For the 4 profile plots, the x-axis is the cell number and the y-axis is the concentration value before norming.

for one parameter set but basal Cdc42 is unable to produce 3 peaks for that same parameter set. The best error so far is also plotted for each iteration. This is the lowest error that has been produced up through a specific iteration number. Referring to the plot in Fig 2.15, the best error so far starts at about 1.55 and by iteration 4 it decreases to about 1 and holds that value for the remaining iterations. This seems to suggest that BO is confident that it selected the parameter set that produces the lowest error possible, given the parameter space provided and BO parameter selection. Thus, it is reasonable to conclude that network 2 is incapable of producing the wildtype Rho1 and Cdc42 profiles.

Moving onto model/network 3, Brk is removed from apical regulations and replaced by pMad. The same set up is used to run BO for model 3 as model 2 (i.e., same exploration parameter, number of training data sets, etc.). Results are shown in Fig 2.16.
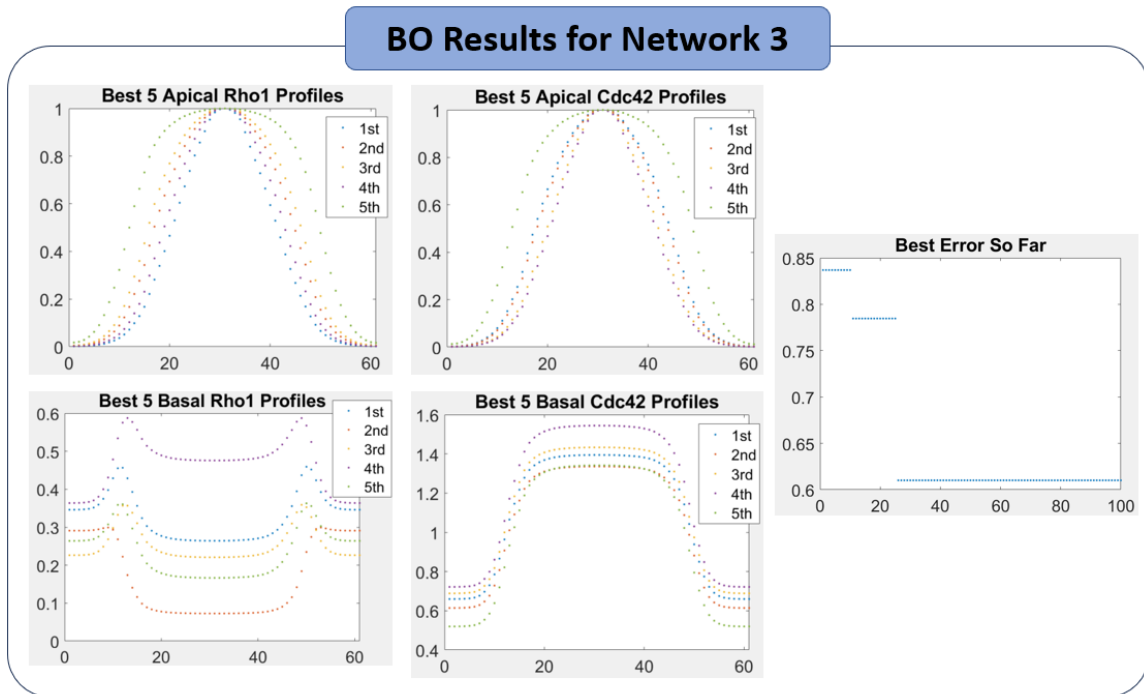
Figure 2.16: Results of BO for network 3. For the 4 profile plots, the x-axis is the cell number and the y-axis is the normed concentration value.

From an initial glance, the simulated profiles from model 3 are much closer to the experimental results than model 2. The apical Rho1 and Cdc42 profiles take on a more exponential-type profile with the regulation from pMad. The basal Rho1 profiles all exhibit 2 peaks at the correct location with a valley in the middle. Basal Cdc42 is still unable to capture the desired 3 peak behavior but the other 3 profiles improved significantly from model 2 results. The best error so far has a more step-like behavior, which is desired. However, by iteration 26 the error becomes fixed once again. Thus, network 3 is better able to capture the desired profiles than model 2 but there still seems to be key regulations missing from the network.

For model 4, regulation by both Brk and pMad are incorporated on the apical and basal surfaces. To start, the training data generation is improved to include about 2000 random parameter sets for the 14 perturbed parameters. The same value for the exploration

parameter, number of BO iterations, and number of parameter sets tested in each iteration remain the same as the previous two cases.
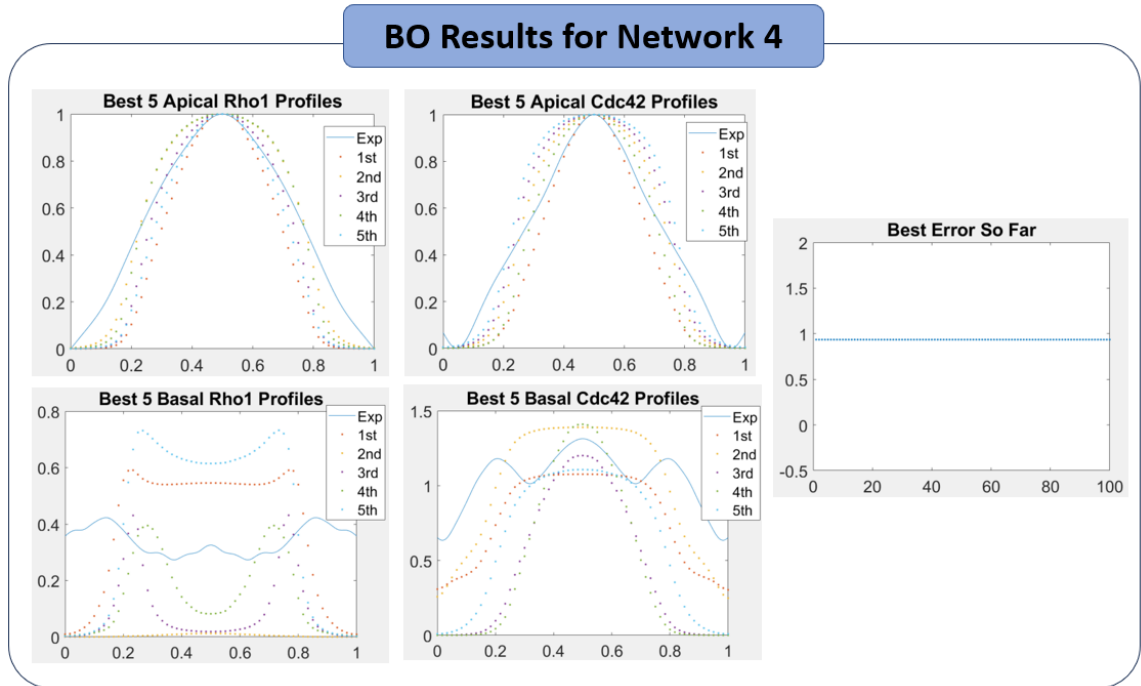
Figure 2.17: Results of BO for network 4. Solid curves show experimental profiles. For the 4 profile plots, the x-axis is the normalized tissue location and the y-axis is the normed concentration value.

Similar to model 3, the correct apical profiles are achieved using model 4 (Fig 2.17). However, both the basal Rho1 and basal Cdc42 profiles do not improve. Examining the best error so far shows that the error remains constant for all 100 iterations of BO. This is most likely due to the improved training data set provided. By using a random sampling of the parameter space, the BO model is able to investigate a wider variety of parameter sets from the beginning. If random sampling was used to produce the training data in model 2 and 3, it's reasonable to assume the behavior would be similar. Comparing the results of models 2-4, network 3 (pMad activation on apical surface and Brk inhibition on basal surface) seems to give the most promising results.

With network 3 selected as the "best" choice of models 2-4, the next step is to decide if BO is the parameter estimation technique best fitted for our needs. BO is designed to handle computationally expensive functions. However, the chemical signaling model, models 2-4, are very inexpensive functions to run. Between this and the overall error not decreasing in the simulated results, a second parameter estimation technique is investigated on network 3: nonlinear regression (NLR). The focus of NLR is on fitting a parametric model to the observed data by minimizing the sum of squared residuals:

$$\sum_{i=1}^{N}(y_i - f_i(\mathbf{b}))^2 \tag{2.10}$$

where

- y_i are the experimental values (e.g., basal Cdc42) for cell i
- $f_i(\mathbf{b})$ are the chemical signaling model outputs for cell i
- $\mathbf{b}$ are the regression coefficients/model parameters
- N is the number of points of comparison (i.e., 4 profiles * 61 cells = 244)

The general idea of NLR is that the signaling model is run multiple times in each optimization iteration and factors such as how small changes in the parameters affect the model predictions and the precision of the parameter estimates are considered. In other words, NLR runs the chemical signaling model many times and strategically chooses the next parameter set, rather than utilizing a surrogate model like BO. The NLR framework works best when used on a rather inexpensive model, such as models 2-4. For a rough time frame, with a model evaluation taking 60 seconds with 10 parameters, the NLR model takes about $(60 \text{ sec/eval}) \cdot (10 \cdot 2 + 1 \text{ function evals/NLR iter}) = 21 \text{ mins/NLR iter}$.

A major downfall of NLR is that it is a local optimization algorithm. It may be susceptible to local minima and may not guarantee global optimization. The method is highly dependent upon the choice of initial parameters. However, using a multi-start NLR method can better cover the parameter space [26]. This is done by sampling the full parameter space and then running the NLR in parallel on each of the samples. Once each case is done running, the

resulting profiles produced for each initial condition can be compared and the outputs that most closely resemble the desired results are chosen. As clarification, initial conditions in this context refers to the starting parameter set used in the model when NLR is run. For the implementation of NLR, the "nlinfit" function in MATLAB is utilized.

To apply multi-start NLR to model 3, 30 random parameter sets are first chosen from the biologically relevant parameter range to use as the initial conditions. Then NLR is run in parallel for each of those 30 initial conditions. Figure 2.18 shows the optimal output of NLR for each initial condition. As expected, there is a large variation between results.
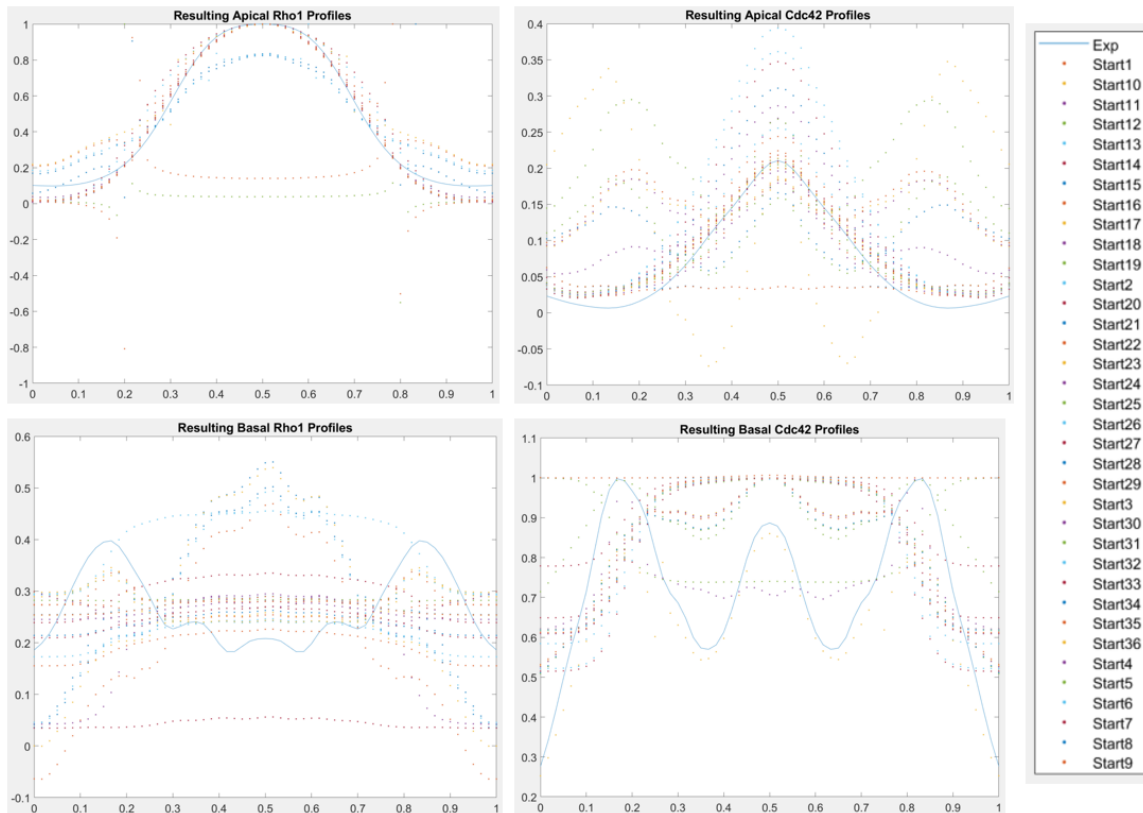
Figure 2.18: Resulting profiles from NLR with 30 different random initial conditions for network 3. Solid curves are experimental results.

Outputs from the multi-start NLR are then filtered by extracting the final parameter sets that produce the lowest SSE. For model 3, this is initial parameter set, or start, 3. The

resulting profiles are shown in Fig 2.19, with the initial and final parameter sets for this run of NLR shown in Fig 2.20.
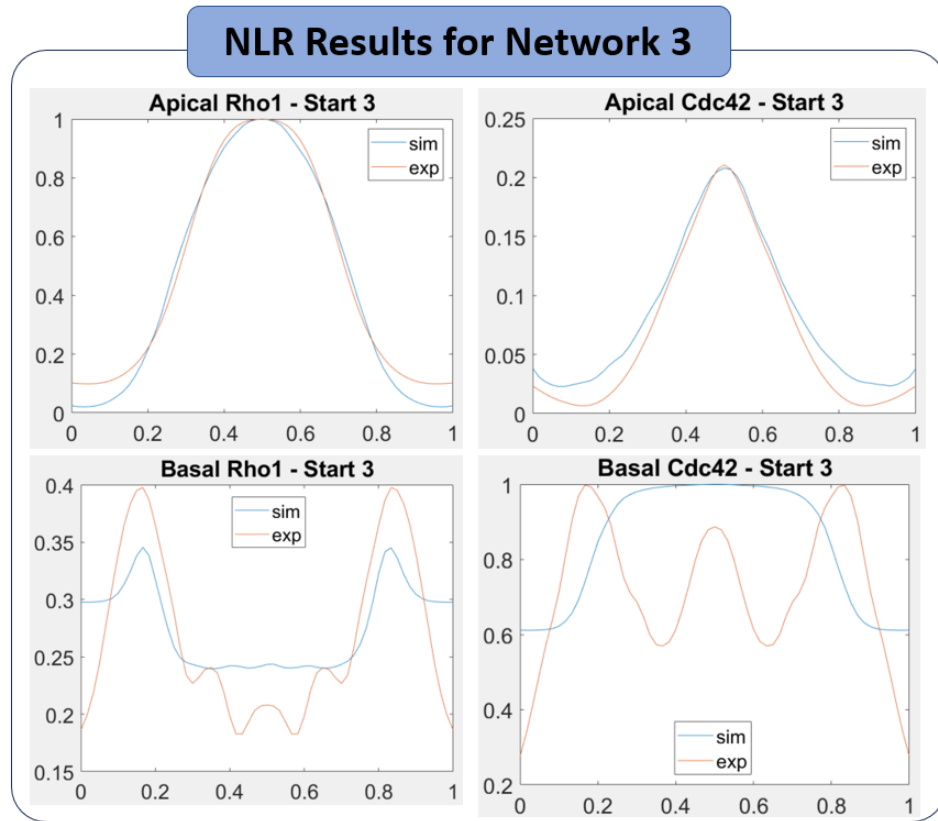
Figure 2.19: Rho1 and Cdc42 profiles from initial condition 3 in Fig 2.18. The red curve is the desired experimental profile and the blue curve is the NLR output.

Figure 2.20: Initial parameter set for model 3 that is passed into NLR (top) and the final parameter set that produces the optimal results (bottom).

The resulting Rho1 and Cdc42 profiles of NLR closely resemble those of BO for model 3. One main difference in the results between the two parameter estimation techniques lies in the output parameter values. Examining the final parameter values for NLR in Fig 2.20 shows that some of the parameters take on non-biologically relevant values. For example, the parameter k_{cr_ap}, which is a concentration threshold, takes on a negative value. However, concentrations cannot take on negative values so this cannot be used to draw any biological conclusions. BO, on the other hand, forces the parameters to remain in a biologically relevant range. This main factor leads to the choice of BO for parameter estimation.

Using both BO and NLR, it is apparent that the regulations modeled in networks 2-4 are not sufficient to produce wildtype profiles of Rho1 and Cdc42. Thus, it is reasonable to assume that there are more critical biological components involved in the regulation of Rho1 and Cdc42.

Two more known regulators of Rho1 and Cdc42 are the transcription factors Omb and Sal. Although Omb and Sal have been shown to regulate Rho1 and Cdc42, the manner in which they regulate (up-regulate vs. down-regulate) is still unclear. Therefore, studying these components may provide insight into the chemical signaling network at play. Omb and Sal are downstream of Brk [20] and have concentration profiles as those depicted in Fig 2.21. The equations to produce these profiles are shown in Eqn 2.11.
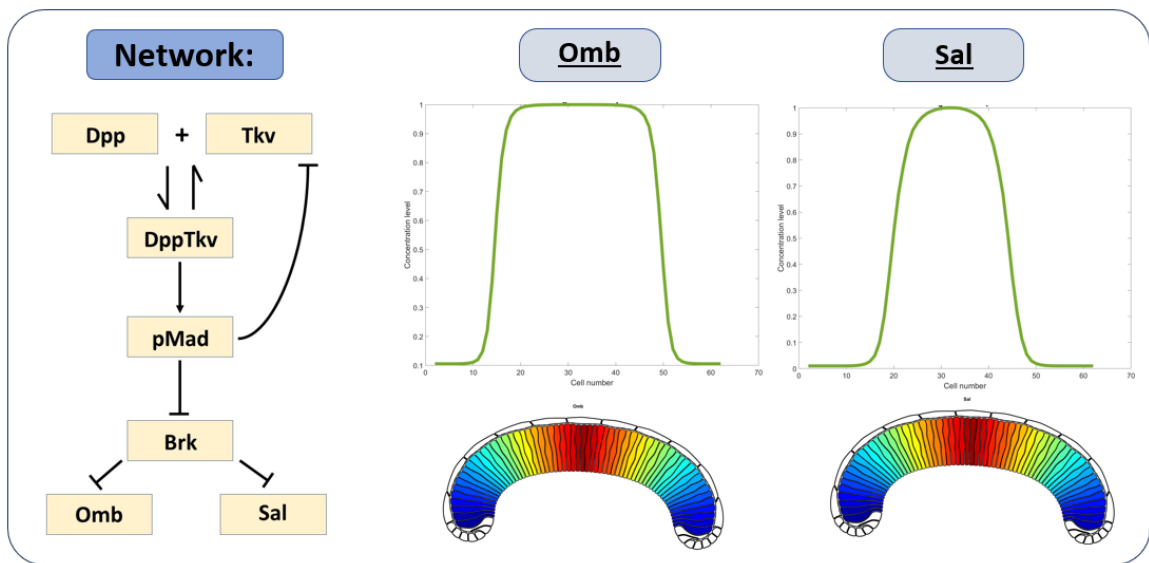
Figure 2.21: Morphogen network with Omb and Sal included, along with profiles of Omb and Sal produced by Eqn 2.11.

Morphogen equations for networks 5-6:

$$\frac{\partial[Dpp]}{\partial t} = D_d\Delta_s^{tis}[Dpp] + s_d(x,y) - d_d[Dpp] - k_{on}[Dpp][Tkv] + k_{off}[DppTkv]$$

$$\frac{\partial[Tkv]}{\partial t} = \left(\frac{v_t}{1+([pMad]/k_{pt})^{n_{pt}}}\right) - d_t[Tkv] - k_{on}[Dpp][Tkv] + k_{off}[DppTkv]$$

$$\frac{\partial[DppTkv]}{\partial t} = k_{on}[Dpp][Tkv] - k_{off}[DppTkv] - d_{dt}[DppTkv]$$

$$\frac{\partial[pMad]}{\partial t} = \frac{v_p}{1+([DppTkv]/k_{dtp})^{-n_{dtp}}} - d_p[pMad]$$

$$\frac{\partial[Brk]}{\partial t} = \frac{v_b}{1+([pMad]/k_{pb})^{n_{pb}}} - d_b[Brk]$$

$$\frac{\partial[Omb]}{\partial t} = \frac{v_o}{1+([Brk]/k_{bo})^{n_{bo}}} - d_o[Omb]$$

$$\frac{\partial[Sal]}{\partial t} = \frac{v_s}{1+([Brk]/k_{bs})^{n_{bs}}} - d_s[Sal]$$

(2.11)
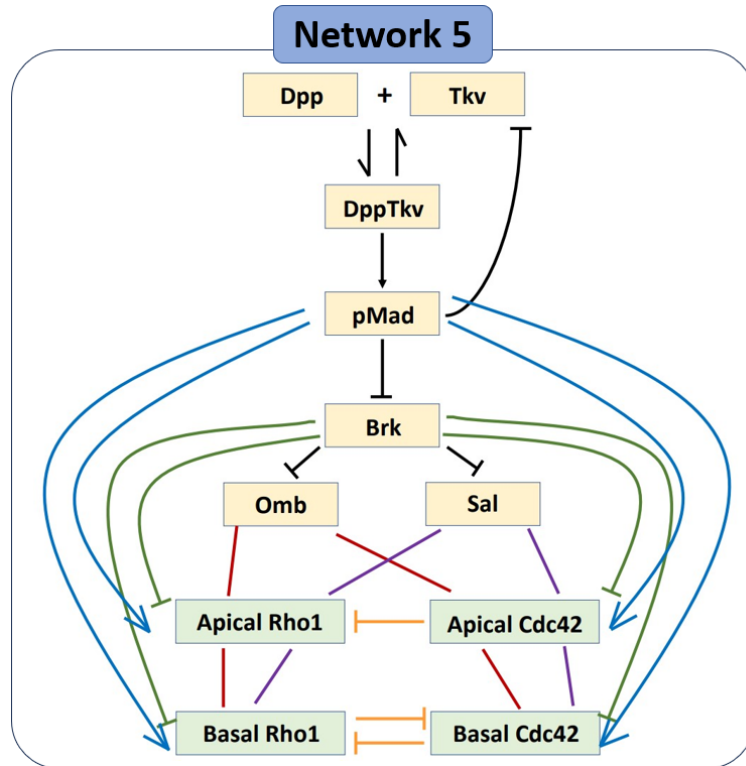
Figure 2.22: Network 5 with introduction of transcription factors Omb and Sal.

Network 5 incorporates the regulation by Omb and Sal. Network 5 includes activation by pMad and inhibition by Brk on both apical and basal Rho1 and Cdc42, as well as the interaction between the Rho-GTPases. The regulation by Omb and Sal is incorporated into the network but the manner in which they interact with Rho1 and Cdc42 is not specified.

For the next step in the process, we want to determine if the introduction of these new biological components can give rise to the wildtype Rho1 and Cdc42 profiles before running the reaction-diffusion model, as well as investigate the regulation of Rho1 and Cdc42 by Omb and Sal. Since building and running the BO model can be time consuming, we want to develop a method in which we can more quickly test if a network can give rise to the desired experimental profiles. To do this, we use a linear regression model to determine if a model can fit the wildtype Rho1 and Cdc42 profiles by weighting the concentration of the upstream components in the network. This is done by solving a constrained linear least-squares problem of the form:

$$\min_w \frac{1}{2}||C \cdot w - d||_2^2 \qquad (2.12)$$

such that

$$lb \leq w \leq ub$$

where

- C are the concentrations of the regulators
- w are the weights
- d are the experimental values
- lb, ub are the limits on the weights

For example, to determine the necessary weights to achieve the basal Cdc42 profiles in model 5, the following system is solved:

$$w_1[pMad] + w_2[Brk] + w_3[Omb] + w_4[Sal] + w_5[Basal\_Rho] = [exp\_Basal\_Cdc42] \quad (2.13)$$

such that

$$
\begin{cases}
0 \leq w_1 \leq \infty \\
-\infty \leq w_2, w_5 \leq 0 \\
-\infty \leq w_3, w_4 \leq \infty
\end{cases}
\tag{2.14}
$$

Here positive weights correspond to activation/up-regulation and negative weights to inhibition/down-regulation. Using linear regression allows us to quickly test numerous networks, in addition to making predictions on unknown regulations and the significance of the contribution of each component in the model. Only the networks that can achieve the desired profiles will be thoroughly tested using the BO model on the corresponding reaction-diffusion equations (Fig 2.23). This linear regression method reduces the number of models that are put through BO, while also refining the parameter space for when BO is run. For example, the sign and magnitude of the predicted weights can give insight into the values of the parameters in the reaction-diffusion model. The sign on a weight dictates if the Hill exponent is positive or negative, indicating activation or inhibition. The magnitude of the weight determines how sensitive the output of the model is to that input. Higher weights in the linear regression model translate to a higher numerator value in the Hill functions in the reaction-diffusion model and possibly higher exponent values as well. Using the linear regression model in this way refines the parameter space so BO runs more effectively and efficiently.

Applying this simplified linear regression method to network 5 produces promising results. Normed profiles from Fig 2.2 are used for pMad and Brk, normed profiles from Fig 2.21 are used for Omb and Sal, and experimental profiles from Fig 2.7 are used for Rho1 and Cdc42. The weights for pMad are constrained to be positive (inferring activation), weights for Brk and the interaction between Rho1 and Cdc42 are negative (inferring inhibition), and the weights for Omb and Sal are unbounded.

As seen in Fig 2.24(A), model 5 can achieve the correct number of peaks for all 4 profiles. The basal peaks for both Rho1 and Cdc42 are in the incorrect location but the magnitude
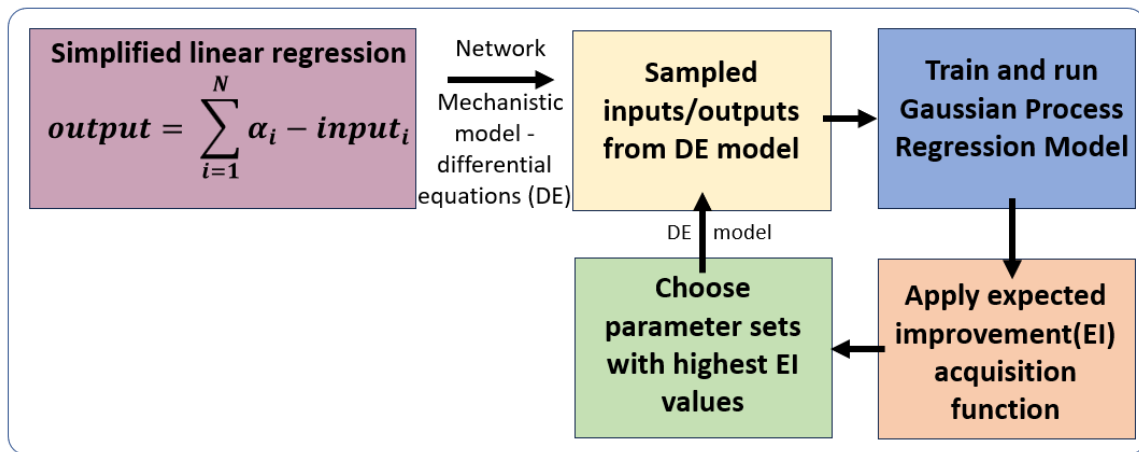
Figure 2.23: Summary of linear regression with Bayesian optimization algorithm.

of the peaks is much closer to the desired result than in models 2-4. Examining the values and signs of the weights produced (Fig 2.24(B)) illuminate the significant and insignificant regulations in network 5. The first and surprising finding is that the regulations from Brk and mutual inhibition are insignificant. The model also predicts that Omb activates both Rho1 and Cdc42, given the other regulators present. Similarly, Sal is seen to be an activator on the apical surface but an inhibitor on the basal surface. Between the inconsistent regulations by Sal, the insignificance of Brk and mutual inhibition, and the incorrect basal peak location, the model assumptions and other biological components need to be revisited since these results suggest that regulations are still missing or that the type of regulations present in network 5 are not correct.

Taking a step back, we need to think about what could possibly answer these important questions: 1) How are the two outer basal Cdc42 peaks in the same location as the basal Rho1 peaks, despite mutual inhibition? 2) Why are the basal peaks concentrated in the areas of high curvature? 3) What is the difference between the apical vs basal surfaces?

One possible answer to these three questions may be the regulation from integrin. Integrins are proteins involved in the attachment of cells to the extracellular matrix (ECM) and they can act as signaling receptors. When integrins bind to specific ligands in the ECM,

Figure 2.24: Results of the linear regression model for network 5. (A) Experimental profiles (red) compared to the output of the linear regression model (blue). (B) Weights of regulations predicted by the linear regression model.

they can activate intracellular signaling pathways such as Rho1 and Cdc42. At late stages, integrin is predominantly localized to the lateral regions of the basal surface (see Fig 2.26) [25]. Feedback between integrin and Rho1 and Cdc42 may help achieve the correct location of the basal peaks.

To test if the incorporation of integrin aids in the production of correct Rho1 and Cdc42 behavior, activation of basal Rho1 and Cdc42 by integrin is introduced into model 6 (Fig 2.27). In this initial test, regulation of integrin by basal Rho1 and Cdc42 is ignored.

With network 6 established, the next step is to apply the linear regression model. Similar to network 5, normed profiles from Fig 2.2 are used for pMad and Brk, normed profiles from
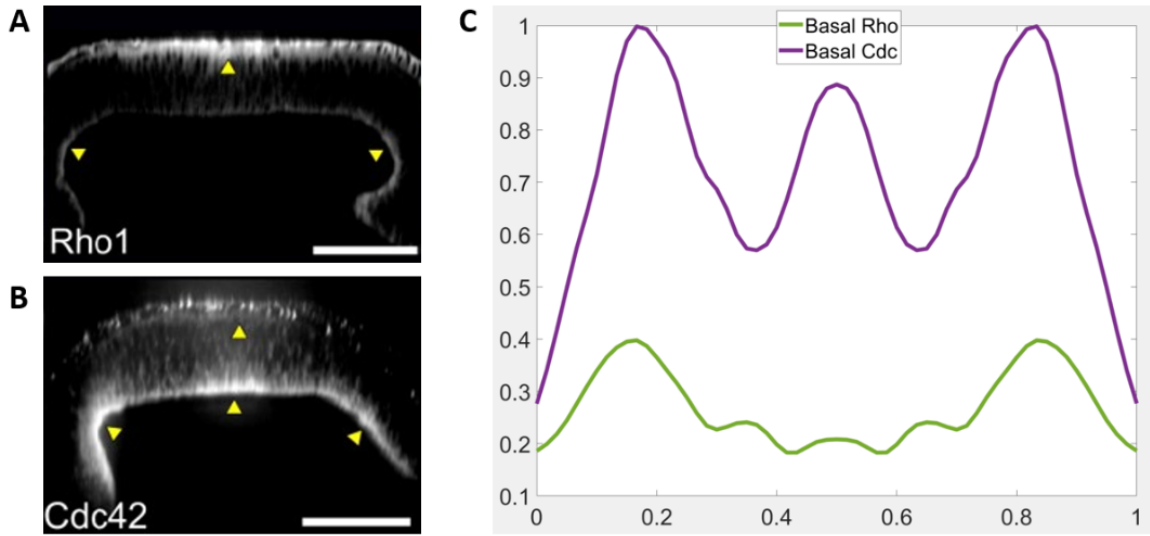
Figure 2.25: (A) Experimental image of cross section of wing disc showing concentration profile of Rho1. (B) Similarly for concentration profile of Cdc42. (C) Comparison of basal Rho1 and basal Cdc42 profiles.
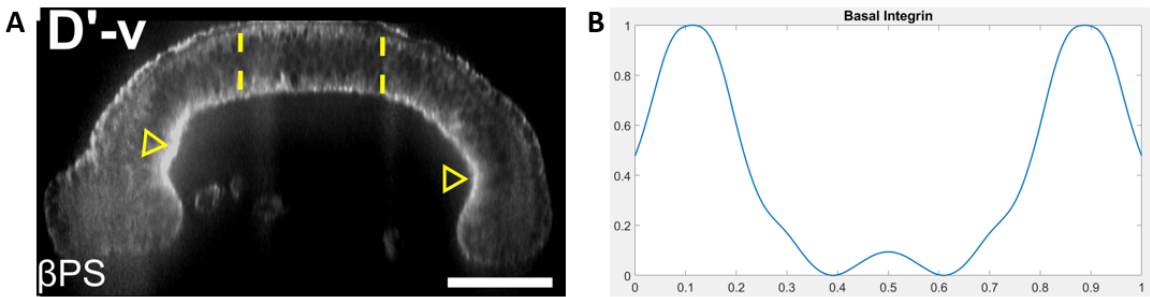
Figure 2.26: Integrin concentration in cross section of wing disc. (A) Experimental image of concentration profile of integrin. Image is from Kumar *et al.* [25]. (B) Smoothed and normed experimental basal integrin profile.

Fig 2.21 are used for Omb and Sal, experimental profiles from Fig 2.7 are used for Rho1 and Cdc42, and the experimental profile in Fig 2.26(B) is used for basal integrin. The weights for pMad and integrin are constrained to be positive (inferring activation), weights for Brk and the interaction between Rho1 and Cdc42 are negative (inferring inhibition), and the weights for Omb and Sal remain unbounded.
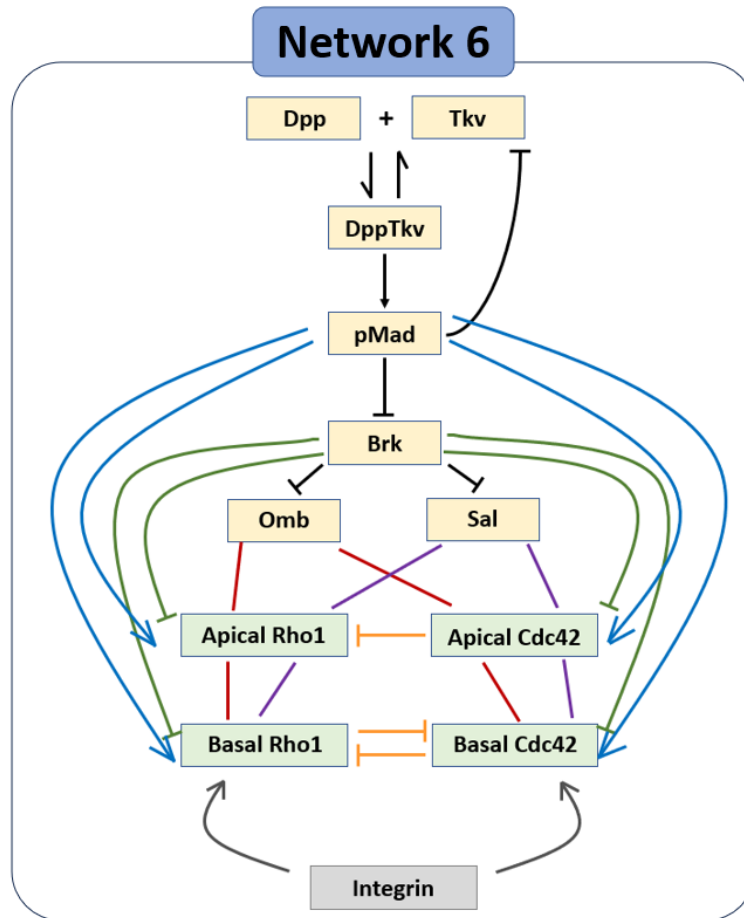
Figure 2.27: Diagram for network 6.

Results of the linear regression model for network 6 are shown in Fig 2.28, along with the weights for each regulator. All profiles closely match the experimental results. For basal Cdc42, there are two peaks where each of the outer single peaks should be but we expect that is due to the data used for integrin in the linear regression. As seen in Fig 2.26, integrin is concentrated in the regions of high curvature, as are basal Rho1 and Cdc42, but the plotted values of integrin do not align with the plotted values of basal Rho1 and Cdc42. This is due to the data being collected from different sets of flies grown at different times but in similar conditions. The natural variance can be eliminated by increasing the sample size used to produce the integrin profiles utilized in the simulations.
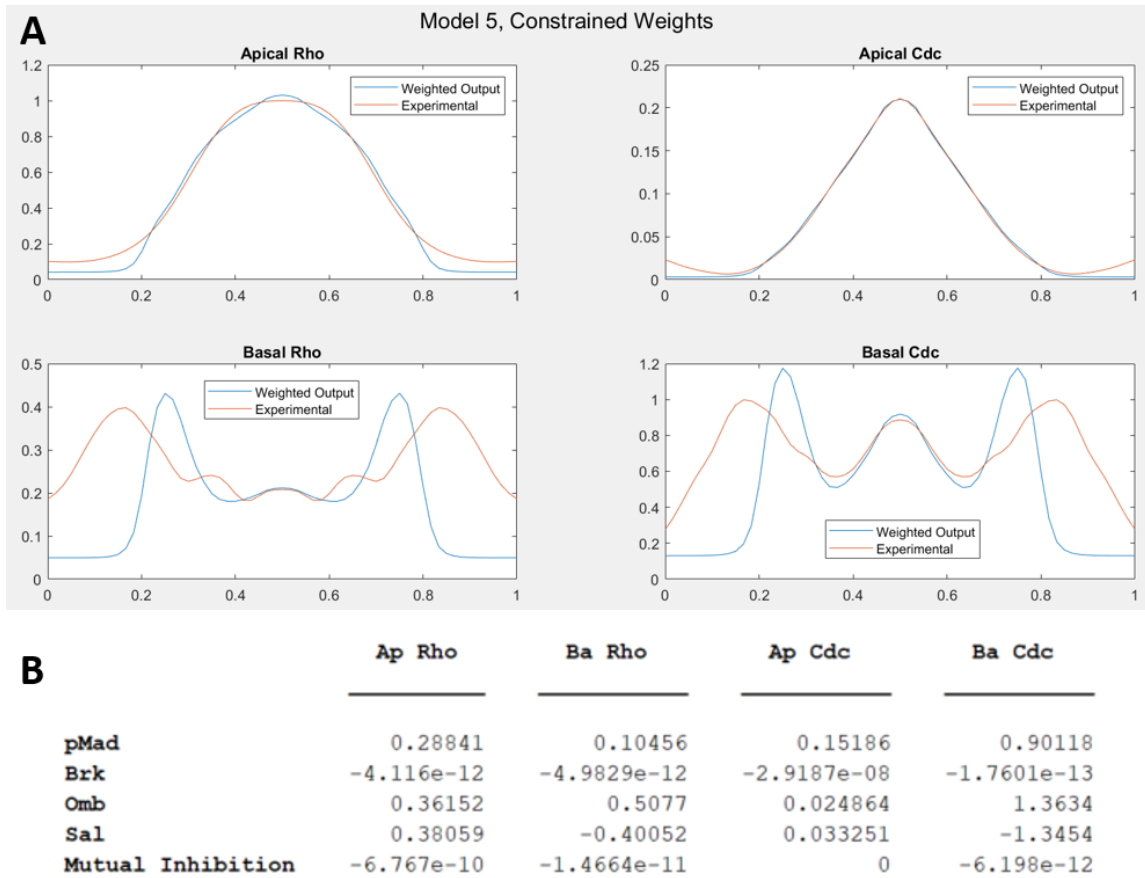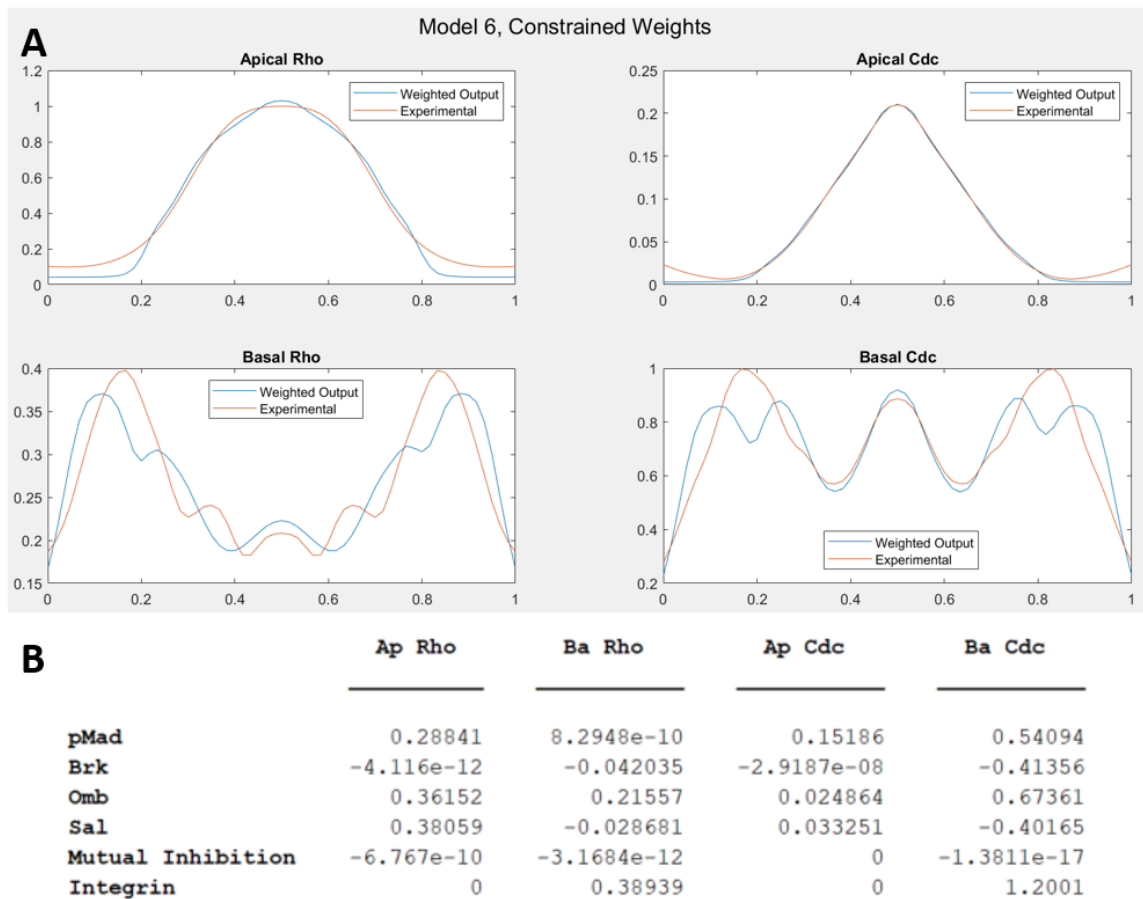
Figure 2.28: Results of the linear regression model for network 6. (A) Experimental profiles (red) compared to output of the linear regression model (blue). (B) Weights of regulations predicted by the linear regression model.

Examining the weights produced by the linear regression model can once again be used to form a hypothesis as to which components play a more significant role in the regulation of Rho1 and Cdc42. For the apical Rho1 profile, pMad, Omb and Sal are weighted positively with about the same magnitude, while Brk and apical Cdc42 weights are negative and insignificant in comparison. For basal Rho1, the significant regulations come from Brk and integrin. pMad is suggested to be the main component for regulation of apical Cdc42, with regulation from Omb and Sal suggested to be positive. Finally, for basal Cdc42, integrin is weighted significantly higher than the other regulators, with pMad, Brk, Omb, and Sal weighted with roughly the same magnitude, and inhibition by basal Rho1 weighted

insignificantly. Across all 4 profiles, the linear regression predicts that Omb is a positive regulator of Rho1 and Cdc42. As in the results for network 5, Sal is predicted to up-regulate Rho1 and Cdc42 on the apical surface and down-regulate on the basal surface.

This preliminary investigation into the role of integrin in the production of the wildtype Rho1 and Cdc42 profiles produces promising results. Incorporation of integrin aids in the achievement of correct basal Rho1 and Cdc42 profiles, despite the mutual inhibition between the Rho-GTPases. To fully investigate the resulting network produced by the linear regression model, a reaction-diffusion model must be written and tested using BO. The predicted role of Omb and Sal can also be tested experimentally.

2.5 Conclusions and Future Directions

We use a combination of experimental and modeling techniques to further investigate the Dpp-Rho1-Cdc42 network in the *Drosophila* wing imaginal disc. Wildtype Rho1 and Cdc42 exhibit localized peaks of high concentration on the apical and basal surfaces, respectively, and mutual inhibition between Rho1 and Cdc42 is recognized for the first time in any system. Dpp is also shown to regulate Rho1 and Cdc42 through Brk. Modeling confirms that regulation of Rho1 and Cdc42 is not uniform across the apical and basal surfaces and suggests that the role of integrin is crucial in achieving the correct wildtype behavior.

To further confirm the proposed network 6, a few avenues must be explored. First, the regulation of Rho1 and Cdc42 by Omb and Sal needs to be established through experiments. The integrin profile used in the models needs to be improved to align better with the experimental Rho-GTPase values. The essential role of integrin can be further explored in both experiments and through the use of Bayesian optimization on the corresponding reaction-diffusion model. Other characteristics such as the role of initial conditions and intracellular diffusion in the model can be studied. Actin and myosin can then be incorporated into the network to create a direct link to the mechanical components in the wing disc. The established chemical signaling model can then be coupled with the existing SCE model [25] to incorporate shape change and interactions with the mechanical components.

Bibliography

[1] Peng D and Merriman B and Osher S and Zhao H and Kang M, *A PDE-based Fast Local Level Set Method*, J. Comput. Phys. **155** (1999), no. 2, 410–438.

[2] Xu JJ and Zhao HK, *An Eularian Formulation for Solving Partial Differential Equations along a Moving Interface*, J. Sci. Comp. **19** (2003), no. 1, 573–594.

[3] Pan X and Fang L and Liu J et al., *Auxin-induced signaling protein nanoclustering contributes to cell polarity formation*, Nat. Commun. **11** (2020), no. 3914.

[4] Fu Y and Gu Y and Zheng Z and Wasteneys G and Yang Z, *Arabidopsis interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis*, Cell. **120** (2005), no. 5, 687–700.

[5] Merriman B and Bence JK and Osher S, *Motion of Multiple Junctions: A Level Set Approach*, J. Comp. Phys. **112** (1994), no. 2, 334–363.

[6] Sapala A and Runions A et al., *Why plants make puzzle cells, and how their shape emerges*, Comp. and Sys. Bio. **7** (2018), no. e32794.

[7] Higaki T et al., *Exogenous Cellulase Switches Cell Interdigitation to Cell Elongation in a RIC1-dependent Manner in Arabidopsis thaliana Cotyledon Pavement Cells*, Plant and Cell Physiology **58** (2017), no. 1, 106–119.

[8] Chen J and Wang F and Zheng S and Xu T and Yang Z, *Pavement cells: a model system for non-transcriptional auxin signaling and crosstalks*, J. Exp. Bot. **66** (2015), no. 16, 4957–70.

[9] Higaki T and Kutsuna N and Akita K and Takigawlmamura H and Yoshimura K and Miura T, *A Theoretical Model of Jigsaw-Puzzle Pattern Formation by Plant Leaf Epidermal Cells*, PLoS Comput. Biol. **12** (2016), no. 4.

[10] Osher S and Sethian J, *Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations*, J. Comput. Phys. **79** (1988), no. 1, 12–49.

[11] Lin W and Yang Z, *Unlocking the mechanisms behind the formation of interlocking pavement cells*, Curr. Opin. in Plant Biol. **57** (2020), 142–154.

[12] Gottlieb S and Shu CW, *Total Variation Diminishing Runge-Kutta Schemes*, Math. of Comp **67** (1998), no. 221, 73–85.

[13] Jiang GS and Shu CW, *Efficient Implementation of Weighted ENO Schemes*, J of Comp Phys **126** (1996), 202–228.

[14] Zhao H-K and Chan T and Merriman B and Osher S, *A Variational Level Set Approach to Multiphase Motion*, J. Comp. Phys. **127** (1996), no. 167, 179–195.

[15] Martin FA and Morata G, *Compartments and the control of growth in the Drosophila wing imaginal disc*, Development **133** (2006), no. 22, 4421–26.

[16] Widmann TJ and Dahmann C, *Dpp signaling promotes the cuboidal-to-columnar shape transition of Drosophila wing disc epithelia by regulating Rho1*, J. Cell Sci. **122** (2009), no. 9, 1362–73.

[17] Gibson M and Perrimon N, *Extrusion and Death of Dpp/BMP-Compromised Epithelial Cells in the Developing Drosophila Wing*, Science **307** (2005), no. 5716, 1785–89.

[18] Martin FA and Herrera SC and Morata G, *Cell competition, growth and size control in the Drosophila wing imaginal disc*, Development **136** (2009), no. 2, 3747–56.

[19] Nematbakhsh A and Levis M and Kumar N and Chen W and Zartman JJ and Alber M, *Epithelial organ shape is generated by patterned actomyosin contractility and maintained by the extracellular matrix*, PLoS Comput. Biol. **16** (2020), no. 8.

[20] Restrepo S and Zartman JJ and Basler K, *Corrdination of Patterning and Growth by the Morphogen DPP*, Current Biology **24** (2014), no. 6, 245–255.

[21] Rogers KW and Schier AF, *Morphogen gradients: from generation to interpretation*, Annu. Rev. Cell. Dev. Biol **27** (2011), 377–407.

[22] Affolter M and Basler K, *The Decapentaplegic morphogen gradient: from pattern formation to growth regulation*, Nat. Rev. Genet **8** (2007), 663–674.

[23] Warner SJ and Longmore GD, *Cdc42 antagonizes Rho1 activity at adherens junctions to limit epithelial cell apical tension*, J. Cell Biol. **187** (2009), 119–133.

[24] Duan L et al., *Distinct Roles for Rho Versus Rac/Cdc42 GTPases Downstream of Vav2 in Regulating Mammary Epithelial Acinar Architecture*, J. Biol. Chem. **285** (2010), 1555–1568.

[25] Kumar N and Rangel Ambriz J and Tsai K and Sahir Mim M and Flores-Flores M and Chen W and Zartman JJ and Alber M, *Balancing competing effects of tissue growth and cytoskeletal regulation during Drosophila wing disc development*, Nat Commun **15** (2024), no. 2477.

[26] Reali F and Priami C and Marchetti L, *Optimization Algorithms for Computational Systems Biology*, Front. in Appl. Math. and Stats. **3** (2017).

[27] Kumar N and Dowling A and Zartman J, *Reverse engineering morphogenesis through Bayesian optimization of physics-based models*, bioRxiv (2023).

[28] Jones D and Schonlau M and Welch W, *Efficient Global Optimization of Expensive Back-Box Functions*, J. Global Optim. **13** (1998), 455–492.

[29] Ramezani A and Britton S and Zandi R and Alber M and Nematbakhsh A and Chen W, *A multiscale chemical-mechanical model predicts impact of morphogen spreading on tissue growth*, npj Syst Biol Appl **9** (2023), no. 16.

[30] Zhu Y and Qiu Y and Chen W and Nie Q and Lander A, *Scaling a Dpp morphogen gradient through feedback control of receptors and co-receptors*, Dev. Cell **53** (2020), no. 6, 724–739.

A

Appendix

A.1 Description of Numerical Schemes for Level Set Method

For the local level set equation (Eqn 1.5) and the reinitialization equation (Eqn 1.3), a **TVD Runge-Kutta** [12] method of third-order is used for the time discretization:

$$\begin{cases} \phi_{temp1} = \phi^m + \Delta t L(\phi^m) \\[2mm] \phi_{temp2} = \frac{3}{4}\phi^m + \frac{1}{4}\phi_{temp1} + \frac{\Delta t}{4}L(\phi_{temp1}) \\[2mm] \phi^{m+1} = \frac{1}{3}\phi^m + \frac{2}{3}\phi_{temp2} + \frac{2\Delta t}{3}L(\phi_{temp2}) \end{cases} \tag{A.1}$$

where $L(\phi)$ is the flux of the equation $\frac{d\phi}{dt} = L(\phi)$

For the local level set equation,

$$L(\phi) = -c(\phi)u_n|\nabla\phi| \tag{A.2}$$

Similarly, the flux for the reinitialization equation is

$$L(\phi) = -S(\phi)(|\nabla\phi| - 1) \tag{A.3}$$

For the spacial discretization of the flux operators, **Godunov's scheme** [1] with **fifth-order WENO** [13] is used. For the level set equation this is:

$$L(\phi) = - \left[\max(c(\phi)u_n, 0) \cdot \sqrt{\max[(a^+)^2, (b^-)^2] + \max[(c^+)^2, (d^-)^2]}\right.$$
$$\left. + \min(c(\phi)u_n, 0) \cdot \sqrt{\max[(a^-)^2, (b^+)^2] + \max[(c^-)^2, (d^+)^2]} \right] \tag{A.4}$$

For the reinitialization equation this is:

$$L(\phi) = - \left[\max(S(\phi), 0) \cdot \left(\sqrt{\max[(a^+)^2, (b^-)^2] + \max[(c^+)^2, (d^-)^2]} - 1 \right) \right.$$
$$\left. + \min(S(\phi), 0) \cdot \left(\sqrt{\max[(a^-)^2, (b^+)^2] + \max[(c^-)^2, (d^+)^2]} - 1 \right) \right] \tag{A.5}$$

where $(x)^+ = \max(x, 0)$, $(x)^- = \min(x, 0)$, and $a = D_x^- \phi_{i,j}$, $b = D_x^+ \phi_{i,j}$, $c = D_y^- \phi_{i,j}$, and $d = D_y^+ \phi_{i,j}$ are the one-sided spatial differences computed using WENO5. The sign function $S(\phi)$ is approximated by:

$$S(\phi) = \frac{\phi}{\sqrt{\phi^2 + |\nabla \phi|^2 \Delta x^2}} \tag{A.6}$$

For the extension equation (Eqn 1.20), a first order upwind scheme with a forward Euler time discretization is used:

$$f_{ij}^{n+1} = f_{ij}^n - \Delta t \left[\max(s_{ij} n_{ij}^x, 0) \frac{f_{ij} - f_{i-1j}}{\Delta x} + \min(s_{ij} n_{ij}^x, 0) \frac{f_{i+1j} - f_{ij}}{\Delta x} \right.$$
$$\left. + \max(s_{ij} n_{ij}^y, 0) \frac{f_{ij} - f_{ij-1}}{\Delta y} + \min(s_{ij} n_{ij}^y, 0) \frac{f_{ij+1} - f_{ij}}{\Delta y} \right] \tag{A.7}$$

where

$$s_{ij} = \frac{\phi_{ij}}{\sqrt{\phi_{ij}^2 + \Delta x^2}} \tag{A.8}$$

and

$$\hat{n} = (n^x, n^y) = \left(\frac{\phi_x}{\sqrt{\phi_x^2 + \phi_y^2}}, \frac{\phi_y}{\sqrt{\phi_x^2 + \phi_y^2}} \right) \tag{A.9}$$

Numerical Approximations for Convection-Diffusion Equation

For the convection-diffusion equation:

$$f_t + (\vec{u} \cdot \hat{n})\kappa f - (\nabla^2 f - \frac{\partial^2 f}{\partial \hat{n}^2} - \kappa \frac{\partial f}{\partial \hat{n}}) = 0 \tag{A.10}$$

forward Euler is used for time discretization while central difference schemes are used to compute $\nabla^2 f$, $\frac{\partial^2 f}{\partial \hat{n}^2} = \hat{n} \cdot D^2 f \cdot \hat{n}$, κ, and $\frac{\partial f}{\partial \hat{n}} = \hat{n} \cdot \nabla f$, as in [2].

Approximation for $\nabla^2 f$:

$$\begin{aligned}
\nabla^2 f &= f_{xx} + f_{yy} \\
&= \frac{f_{i+1,j} - 2f_{i,j} + f_{i-1,j}}{\Delta x^2} + \frac{f_{i,j+1} - 2f_{i,j} + f_{i,j-1}}{\Delta y^2}
\end{aligned} \tag{A.11}$$

Approximation for $\frac{\partial^2 f}{\partial \hat{n}^2} = \hat{n} \cdot D^2 f \cdot \hat{n}$

$$\begin{aligned}
\frac{\partial^2 f}{\partial \hat{n}^2} &= \hat{n} \cdot D^2 f \cdot \hat{n} \\
&= (n^x)^2 f_{xx} + (n^y)^2 f_{yy} + 2n^x n^y f_{xy} \\
&= (n^x)^2 \frac{f_{i+1,j} - 2f_{i,j} + f_{i-1,j}}{\Delta x^2} + (n^y)^2 \frac{f_{i,j+1} - 2f_{i,j} + f_{i,j-1}}{\Delta y^2} \\
&\quad + 2n^x n^y \frac{f_{i+1,j+1} + f_{i-1,j-1} - f_{i-1,j+1} - f_{i+1,j-1}}{4\Delta x \Delta y}
\end{aligned} \tag{A.12}$$

Approximation for κ

$$\begin{aligned}
\kappa &= \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} \\
&= \frac{1}{\sqrt{\phi_x^2 + \phi_y^2}} \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \right) \\
&= \frac{\left(\frac{\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j}}{\Delta x^2} + \frac{\phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1}}{\Delta y^2} \right)}{\sqrt{(\frac{\phi_{i+1,j} - \phi_{i-1,j}}{2\Delta x})^2 + (\frac{\phi_{i,j+1} - \phi_{i,j-1}}{2\Delta y})^2}}
\end{aligned} \tag{A.13}$$

Upwind scheme for $\frac{\partial f}{\partial \hat{n}} = \hat{n} \cdot \nabla f$

$$\begin{aligned}
\frac{\partial f}{\partial \hat{n}} &= \hat{n} \cdot \nabla f \\
&= (n_x)^+ D_x^- f_{i,j} + (n_x)^- D_x^+ f_{i,j} + (n_y)^+ D_y^- f_{i,j} + (n_y)^- D_y^+ f_{i,j}
\end{aligned} \tag{A.14}$$

Numerical Approximations Used in Adhesion Method 2

For the adhesion method 2 from [14], the following numerical approximations are used for the Heaviside and Dirac delta functions:

Heaviside function:

$$H_\alpha(x) = \begin{cases} 1, & \text{if } x > \alpha \\ 0, & \text{if } x < -\alpha \\ \frac{1}{2}\left[1 + \frac{x}{\alpha} + \frac{1}{\pi}\sin\left(\frac{\pi x}{\alpha}\right)\right], & \text{if } |x| \leq \alpha \end{cases}$$

Dirac delta function:

$$\delta_\alpha(x) = \begin{cases} 0, & \text{if } |x| > \alpha \\ \frac{1}{2\alpha}\left[1 + \cos\left(\frac{\pi x}{\alpha}\right)\right], & \text{if } |x| \leq \alpha \end{cases}$$

A.2 Tables of Parameter Values for Chemical Signaling Models for the *Drosophila* Wing Disc

Model 1 parameters: To achieve the results shown in Fig 2.10, the following parameter values are used in Eqn 2.2.

Parameter	Value	Description	Source
S_d	1.5	Dpp production rate	Calibrated
D_d	20	Dpp diffusion rate	[30]
d_d	0.001	Dpp degradation rate	Calibrated
D_r	1	Rho1 diffusion rate	Calibrated
d_r	1	Rho1 degradation rate	Calibrated
s_{ra}	1	Rho1 apical activation rate	Calibrated
$c_{dr,max}$	5	max val of Rho1 produced by Dpp	Calibrated
k_{dr}	5	EC50	Calibrated
$-n_{dr}$	-1	Hill coefficient	Calibrated
$c_{cr,max}$	0.8	max val of Rho1 produced by Cdc42	Calibrated
k_{cr}	1.4	EC50	Calibrated
n_{cr}	8	Hill coefficient	Calibrated
s_{rb}	0.5	Rho1 basal activation rate	Calibrated
s_{ca}	0.5	Cdc42 apical activation rate	Calibrated
$c_{dc,max}$	2.5	max val of Cdc42 produced by Dpp	Calibrated
k_{dc}	1	EC50	Calibrated
$-n_{dc}$	-0.2	Hill coefficient	Calibrated
$c_{rc,max}$	0.3	max val of Cdc42 produced by Rho1	Calibrated
k_{rc}	0.5	EC50	Calibrated
n_{rc}	8	Hill coefficient	Calibrated
s_{cb}	1	Cdc42 basal activation rate	Calibrated

Table A.1: Parameter values for Model 1 to produce results shown in Fig 2.10.

Models 2-4 morphogen parameters: The following parameter values are used in Eqn 2.3 to achieve the profiles shown in Fig 2.2.

Parameter	Value	Description	Source
S_d	1.5	Dpp production rate	Calibrated
D_d	30	Dpp diffusion rate	Calibrated
d_d	0.0001	Dpp degradation rate	Calibrated
k_{on}	0.0025	Dpp, Tkv binding rate	Calibrated
k_{off}	0.000025	Dpp, Tkv unbinding rate	[30]
v_t	0.5	Tkv activation rate	Calibrated
k_{pt}	8	EC50	Calibrated
n_{pt}	2	Hill coefficient	Calibrated
d_t	0.1	Tkv degradation rate	[30]
d_{dt}	0.01	DppTkv degradation rate	Calibrated
v_p	1	pMad activation rate	[30]
k_{dtp}	1	EC50	[30]
$-n_{dtp}$	-1	Hill coefficient	Calibrated
d_p	0.1	pMad degradation rate	[30]
v_b	1	Brk activation rate	Calibrated
k_{pb}	1	EC50	Calibrated
n_{pb}	2	Hill coefficient	Calibrated
d_b	0.1	Brk degradation rate	Calibrated

Table A.2: Parameter values for morphogen equations in Models 2-4.

Model 2 Rho1 and Cdc42 parameters: The following parameter values are used in Eqns 2.4. Parameter ranges are those explored in BO.

Parameter	Value	Description	Source
D_r	1	Rho1 diffusion rate	Calibrated
d_d	1	Rho1 degradation rate	Calibrated
s_{r_ap}	$[1, 20]$	Rho1 apical activation rate	Calibrated
k_{br_ap}	$[0, 10]$	EC50	Calibrated
n_{br}	2	Hill coefficient	Calibrated
k_{cr_ap}	$[0, 20]$	EC50	Calibrated
n_{cr}	8	Hill coefficient	Calibrated
s_{r_ba}	$[1, 20]$	Rho1 basal activation rate	Calibrated
k_{br_ba}	$[0, 10]$	EC50	Calibrated
k_{cr_ba}	$[0, 20]$	EC50	Calibrated
D_c	1	Cdc42 diffusion rate	Calibrated
d_c	1	Cdc42 degradation rate	Calibrated
s_{c_ap}	$[1, 20]$	Cdc42 apical activation rate	Calibrated
k_{bc_ap}	$[0, 10]$	EC50	Calibrated
n_{bc}	1	Hill coefficient	Calibrated
s_{c_ba}	$[1, 20]$	Cdc42 basal activation rate	Calibrated
k_{bc_ba}	$[0, 10]$	EC50	Calibrated
k_{rc_ba}	$[0, 20]$	EC50	Calibrated
n_{rc}	8	Hill coefficient	Calibrated

Table A.3: Parameter values used in Rho1 and Cdc42 equations in Model 2. Ranges of parameters are those explored in BO.

Model 3 Rho1 and Cdc42 parameters: The following parameter values are used in Eqns 2.5. Parameter ranges are those explored in BO.

Parameter	Value	Description	Source
D_r	1	Rho1 diffusion rate	Calibrated
d_d	1	Rho1 degradation rate	Calibrated
s_{r_ap}	$[1, 20]$	Rho1 apical activation rate	Calibrated
k_{pr_ap}	$[0, 10]$	EC50	Calibrated
$-n_{pr}$	-1	Hill coefficient	Calibrated
k_{cr_ap}	$[0, 20]$	EC50	Calibrated
n_{cr}	4	Hill coefficient	Calibrated
s_{r_ba}	$[1, 20]$	Rho1 basal activation rate	Calibrated
k_{br_ba}	$[0, 10]$	EC50	Calibrated
n_{br}	2	Hill coefficient	Calibrated
k_{cr_ba}	$[0, 20]$	EC50	Calibrated
D_c	1	Cdc42 diffusion rate	Calibrated
d_c	1	Cdc42 degradation rate	Calibrated
s_{c_ap}	$[1, 20]$	Cdc42 apical activation rate	Calibrated
k_{pc_ap}	$[0, 10]$	EC50	Calibrated
$-n_{pc}$	-1	Hill coefficient	Calibrated
s_{c_ba}	$[1, 20]$	Cdc42 basal activation rate	Calibrated
k_{bc_ba}	$[0, 10]$	EC50	Calibrated
n_{bc}	1	Hill coefficient	Calibrated
k_{rc_ba}	$[0, 20]$	EC50	Calibrated
n_{rc}	4	Hill coefficient	Calibrated

Table A.4: Parameter values used in Rho1 and Cdc42 equations in Model 3. Ranges of parameters are those explored in BO.

Model 4 Rho1 and Cdc42 parameters: The following parameter values are used in Eqns 2.6. Parameter ranges are those explored in BO.

Parameter	Value	Description	Source
D_r	1	Rho1 diffusion rate	Calibrated
d_d	1	Rho1 degradation rate	Calibrated
s_{r_ap}	$[0, 20]$	Rho1 apical activation rate	Calibrated
k_{pr_ap}	$[0, 10]$	EC50	Calibrated
$-n_{pr}$	-1	Hill coefficient	Calibrated
k_{br_ap}	$[0, 10]$	EC50	Calibrated
n_{br}	2	Hill coefficient	Calibrated
k_{cr_ap}	$[0, 20]$	EC50	Calibrated
n_{cr}	4	Hill coefficient	Calibrated
s_{r_ba}	$[0, 20]$	Rho1 basal activation rate	Calibrated
k_{pr_ba}	$[0, 10]$	EC50	Calibrated
k_{br_ba}	$[0, 10]$	EC50	Calibrated
k_{cr_ba}	$[0, 20]$	EC50	Calibrated
D_c	1	Cdc42 diffusion rate	Calibrated
d_c	1	Cdc42 degradation rate	Calibrated
s_{c_ap}	$[0, 20]$	Cdc42 apical activation rate	Calibrated
k_{pc_ap}	$[0, 10]$	EC50	Calibrated
$-n_{pc}$	-1	Hill coefficient	Calibrated
k_{bc_ap}	$[0, 10]$	EC50	Calibrated
n_{bc}	1	Hill coefficient	Calibrated
s_{c_ba}	$[1, 20]$	Cdc42 basal activation rate	Calibrated
k_{pc_ba}	$[0, 10]$	EC50	Calibrated
k_{bc_ba}	$[0, 10]$	EC50	Calibrated
k_{rc_ba}	$[0, 20]$	EC50	Calibrated
n_{rc}	4	Hill coefficient	Calibrated

Table A.5: Parameter values used in Rho1 and Cdc42 equations in Model 4. Ranges of parameters are those explored in BO.

Models 5-6 morphogen parameters: The following parameter values are used in Eqn 2.11.

Parameter	Value	Description	Source
S_d	1.5	Dpp production rate	Calibrated
D_d	30	Dpp diffusion rate	Calibrated
d_d	0.0001	Dpp degradation rate	Calibrated
k_{on}	0.0025	Dpp, Tkv binding rate	Calibrated
k_{off}	0.000025	Dpp, Tkv unbinding rate	[30]
v_t	0.5	Tkv activation rate	Calibrated
k_{pt}	8	EC50	Calibrated
n_{pt}	2	Hill coefficient	Calibrated
d_t	0.1	Tkv degradation rate	[30]
d_{dt}	0.01	DppTkv degradation rate	Calibrated
v_p	1	pMad activation rate	[30]
k_{dtp}	1	EC50	[30]
$-n_{dtp}$	-1	Hill coefficient	Calibrated
d_p	0.1	pMad degradation rate	[30]
v_b	1	Brk activation rate	Calibrated
k_{pb}	1	EC50	Calibrated
n_{pb}	2	Hill coefficient	Calibrated
d_b	0.1	Brk degradation rate	Calibrated
v_o	1	Omb activation rate	Calibrated
k_{bo}	7	EC50	Calibrated
n_{bo}	6	Hill coefficient	Calibrated
d_o	0.1	Omb degradation rate	Calibrated
v_s	1	Sal activation rate	Calibrated
k_{bs}	1	EC50	Calibrated
n_{bs}	2	Hill coefficient	Calibrated
d_s	0.1	Sal degradation rate	Calibrated

Table A.6: Parameter values to produce morphogen profiles used in Networks 5 and 6.