UNIVERSITY OF CALIFORNIA SAN DIEGO

Models to study the mechanism of single and collective cell migration

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics

by

Haicen Yue

Committee in charge:

Wouter-Jan Rappel, Chair
Alexander Groisman
Bo Li
Kaixuan Ni
Massimo Vergassola

2018

The dissertation of Haicen Yue is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Wouter-Jan Rappel, for his mentoring and guidance in my path towards a scientist in the field of biophysics. His advice in my research work is insightful and inspiring, without which I would have not accomplished the works in this dissertation. He also gives me valuable advice on my future career as a scientist.

I would also like to thank professor Brian Camley, who was a postdoc in our group, for his active discussion and help in the border cell project. In addition, he has taught me a lot, with himself as a role model, on how to become a young scientist.

I am grateful to Dr. Monica Skoge, Dr. Rachel Lee for the experiments and discussions in our collaborations. I would thank professor Herbert Levine, professor William Loomis, professor Wolfgang Losert and professor Pascale Charest for the useful discussions during the collaborations.

I also want to thank my committee members: professor Alex Groisman, professor Massimo Vergassola, professor Bo Li and professor Kaixuan Ni. It's my great honor to have them as my dissertation committee members.

Finally, I would like to express my gratitude to my parents, for their love and support all these years.

Chapter 2, in full (with minor exceptions), is a reprint of the material in part as it appears in "The cAMP-induced G protein subunits dissociation monitored in live Dictyostelium cells by BRET reveals two activation rates, a positive effect of caffeine and potential role of microtubules. Islam, A.F.M. Tariqul; Yue, Haicen; Scavello, Margarethakay; Haldeman, Pearce; Rappel, Wouter-Jan; Charest, Pascale G., Cellular Signaling, 2018". The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is a reprint of the material in part as it appears in "Cellular memory in eukaryotic chemotaxis. Skoge, Monica; Yue, Haicen; Erickstad, Michael; Bae, Albert; Levine, Herbert; Groisman,Alex; Loomis, William F.; Rappel, Wouter-Jan, Proceedings of the National

Academy of Sciences, 2014". The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full (with minor exceptions), has been accepted by Biophysical Journal in 2018 as "Minimal network topologies for signal processing during collective cell chemotaxis. Yue, Haicen; Camley, Brian. A; Rappel, Wouter-Jan". The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full (with minor exceptions), is a reprint of the material in part as it appears in "Inferring single-cell behaviour from large-scale epithelial sheet migration patterns. Lee, Rachel M.; Yue, Haicen; Rappel, Wouter-Jan; Losert, Wolfgang, Journal of The Royal Society Interface, 2017". The dissertation author was the primary investigator and author of this paper.

VITA

2011	B. S. in Physics, Peking University, China
2018	Ph. D. in Physics, University of California San Diego, USA

PUBLICATIONS

Minimal network topologies for signal processing during collective cell chemotaxis. Accepted by *Biophysical Journal*

The cAMP-induced G protein subunits dissociation monitored in live Dictyostelium cells by BRET reveals two activation rates, a negative effect of caffeine and potential role of microtubules. *Cellular signaling*, 2018

Inferring single-cell behaviour from large-scale epithelial sheet migration patterns. *Journal of The Royal Society Interface*, 2017

Relationship between cancer mutations and parameter sensitivity in Rb pathway. *Journal of theoretical biology*, 2016

Cellular memory in eukaryotic chemotaxis. *Proceedings of the National Academy of Sciences*, 2014

Correlation between oncogenic mutations and parameter sensitivity of the apoptosis pathway model. *PLoS computational biology*, 2014

ABSTRACT OF THE DISSERTATION

Models to study the mechanism of single and collective cell migration

by

Haicen Yue

Doctor of Philosophy in Physics

University of California San Diego, 2018

Wouter-Jan Rappel, Chair

Cell migration plays a vital role in many biological processes and the mechanism for both single cell and collective cell migration remains unclear. It is a complex multi-step procedure, including signal receiving, signal processing, force generation and when considering multi-cell migration, cell-cell interaction. Thus it inspires many scientists in biology as well as interdisciplinary areas to study this topic. This dissertation includes the study of cell migration from different aspects and on different scales, using quantitative models. I write them with a sequence from the smallest scale to the largest. In Chapter 2, I study the kinetics of the activation of G-protein-coupled receptors used in chemotaxis trying to explain the two activation rates observed in the experiments. In Chapter 3, I study single cell chemotaxis, focusing on the

signaling networks to explain the memory effect observed in the experiments. In Chapter 4, I study the collective cell chemotaxis taking cell-cell communication into consideration, trying to find the possible minimal network topologies for the signal processing. Finally, in Chapter 5, I study multi-cell migration on a much larger scale with thousands of cells and use a simplified model which focuses on the different kinds of forces applied on the cells to study the relation between single cell properties and large scale behaviors shown in the multi-cell migration. All these works provide some new knowledge on part of the mechanism controlling cell migration.

Chapter 1

Introduction

In the last decades, with the development in biological techniques, biologists are able to get more quantitative data. And more researchers from other fields such as physics, mathematics, and computer science become interested in the questions underlying the complex biological phenomena. They use mathematical models, theory in physics and computational methods to analyze the data, study the mechanism and generate new ideas or perspectives. Especially, the collaboration between theorists and experimentalists has produced many good results to help people better understand the underlying mechanism of some biological processes. This is also what I have been doing in my graduate study. More specifically, I build theoretical models based on quantitative experimental data to study the mechanism related to cell migration.

Cell migration is essential for many biological processes, such as embryonic development, wound healing, cancer metastasis etc. It has different types. For example, there are single cell migration, such as *Dictyostelium* chemotaxis, collective cell migration in a small free group, such as the border cell cluster in Drosophila and collective cell migration of an epithelial sheet with thousands of cells. And it is also a complex multi-step procedure well orchestrated by some mechanism that are still unclear. So, this is an attractive area for both biologists and

interdisciplinary researchers. Cell migration includes signal receiving, signal processing, and force generation. In addition, when considering the collective cell migration, cell-cell contact should also be considered. Due to this complexity, researchers study this procedure from different perspectives or on different scales. For example, researchers studying the receptor binding may focus on specific proteins and reactions; those interested in signal processing may pay attention to signaling networks which include combinations of many reactions; and those who study the collective cell migration may look at this procedure on the tissue or colony scale and show more concerns on cell-cell contact or cell-matrix contact. And the scales on which we study this procedure also determine how many details we include or to what extent we can simplify the system. I also study cell migration on different scales, from the subcellular scale to the tissue scale.

In Chapter 2, we study the mechanism for the activation of heterotrimeric G proteins through the G protein-coupled receptors (GPCRs), which is a key step for the detection of chemoattractant for *Dictyostelium*. Using bioluminescence resonance energy transfer (BRET), the experimentalists obtain data for the activation of G proteins over time after adding stimulus (cAMP), which is then fitted with single or double exponential function. I find that when the cAMP concentration is below some threshold, the time curve can be fitted with single exponential function, indicating only one explicit activation rate. However, when the cAMP concentration is above this threshold, the time curve need to be fitted with double exponential function, indicating two activation rates. In order to explain the origin of these two rates for high cAMP concentration, I use a simple model of receptor-G protein coupling in which both the free receptors and the receptors pre-coupled to inactive G proteins exist and can bind with the ligand, with different affinities. The simulation results show that the existence of these two types of receptors might be the reason of these two activation rates. I also suggest new experiments based on the prediction of the model to verify this possible reason.

In Chapter 3, we aim to better understand the back-of-the-wave problem that *Dictyostelium*

cells can avoid reversing in the back of the wave even though the spatial gradient reverses. The experimentalists use microfluidic devices to generate traveling waves and changing gradient, such as the change from a positive gradient to a negative one or to zero gradient, and obtain quantitative data related to cell's polarization and movement under different conditions. The cells show some memory effect in these experiments. I build an ODE model based on the local excitation global inhibition (LEGI) mechanism but add ultra-sensitivity and a bistable memory module to try to explain the memory effect. We find that the adaptive dynamics of the LEGI mechanism, or more specifically, the slow decay of the global inhibitor, provides a short-term memory. However, this short-term memory is not enough to explain all the experimental phenomena. We need the bistable memory module to strengthen and stablize this short-term memory. The model predicts that the movement of the cell is controlled by activated Ras, an early response indicating cell's directional sensing, together with the memory module, which is downstream of Ras, and that the behavior of activated Ras and the movement of the cell may not always be consistent. Specifically, our model predicts that in the back of the wave, the cell continues to move forward but the Ras response becomes zero at both the front and the back of the cell. This prediction is then verified by the experiment.

In Chapter 4, we switch our focus to the collective cell chemotaxis, in which cell-cell communication plays an important role. The biological system we study is the border cell cluster in Drosophila egg. We use the experimental data from a previous paper [1] as the standard to find the minimal network topologies that can be used by the cluster to chemotax collectively. As little is known about the signaling network in this system, we start from the simplest case with only four nodes which are necessary. Different from the work in Chapter 3, we do not start from a widely used model, but try to find the minimal network topologies that can explain the experimental results from all possible candidates. We combine an analytical method, which does not rely on specific equations and parameters, and a simulation method, which provides more quantitative comparison and prediction, in the selection procedure and finally get six network

topologies with six links. These six networks perform equally well considering the experimental results in the paper we refer to, but we make predictions with our model about how they can be different in some other conditions and thus suggest experiments for further discrimination. We regard these six topologies as a starting point for further research on the mechanism of collective chemotaxis for the border cell cluster and develop methods that may also be applied in other collective migrating systems.

In Chapter 5, we still study the collective cell migration but look at a system of a larger scale with thousands of cells. We aim to infer the connection between individual cell properties and collective behavior using a model of multicellular migration. The experimentalist analyzes the particle image velocimetry (PIV) results which allow us to extract the entire flow field of cell motion. In the experiments, we observed some changes of the collective behavior in the MCF10A cell sheets in different media. I try to use a modified model based on a previous one ([2, 3]) to explain these changes. In the large-scale model, I neglect the signal processing details within each cell but only treat a single cell as two points. The model includes cell growth and division, motility forces, friction forces, and volume exclusion and adhesion between neighboring cells and adopts the alignment mechanism that the cell's motility force tends to align with its velocity. I use this model to simulate the migration of the cells in an epithelial sheet and by comparing with experimental data, show that a special group of cells on the edge of the monolayer with different properties as other cells (we call it leader cells) are not necessary to explain the observed large-scale change in collective migration. Changes in proliferation are also not likely to be the cause of the observed change. However, the wake rate, indicating the activity of the cells, and the strength of the motility-velocity alignment mechanism may play a role.

In the above four projects, I did all the modeling and theoretical work. In addition, I also did the data analysis work in Chapter 2.

Chapter 2

The cAMP-induced G proteins subunits dissociation reveals two activation rates

2.1 Introduction

Chemotaxis, the ability of cells to sense gradients of chemicals (chemoattractants) and migrate towards their highest concentration, is a cellular behavior central to the embryonic development and immune response and that is deregulated in diseases such as cancer metastasis and inflammatory disorders. However, how cells detect and determine the direction of a chemoattractant gradient is not fully understood. Many chemoattractants are detected by seven transmembrane receptors that signal through heterotrimeric G proteins (G protein-coupled receptors; GPCRs), and studies with the chemotaxis experimental model *Dictyostelium discoideum* have provided important insight into chemoattractant GPCR signaling dynamics [4, 5]. When food is abundant,

Dictyostelium is in the growth (or vegetative) stage and grows as single undifferentiated cells [6]. Upon starvation, *Dictyostelium* enters a multicellular development/differentiation program going through aggregation ($\sim 4-8$ h), mound (~ 12 h), slug (~ 16 h), and culmination ($\sim 18-24$ h) stages that end with the formation of a fruiting body containing spores. Aggregation is mediated by the chemotaxis of cells towards cyclic AMP (cAMP). *Dictyostelium* has four cAMP receptors (cARs) [7, 8]. cAR1 expression is maximal during aggregation and cAR3 levels rise towards the end of aggregation, whereas cAR2 and cAR4 are maximally expressed during the slug and culmination stages, respectively [9–12]. There is little cAR expressed in undifferentiated vegetative cells. cAR1 and cAR3 are most similar, with 56% amino acid identity, and they both mediate the response to cAMP through the heterotrimeric G protein G$\alpha 2\beta\gamma$. Although the two receptors display similar cAMP binding affinities in phosphate buffer, cAR3 is ~ 100 times less efficient than cAR1 in inducing cAMP-stimulated responses and cAR1 is essential for chemotaxis-driven aggregation through G$\alpha 2\beta\gamma$ [7, 12, 13]. Whereas there are twelve G alpha protein subunits in *Dictyostelium*, G$\alpha 2$ (gene *gpab*) is the main G alpha subunit responsible for the chemotactic responses to cAMP and there is only one G beta (gene *gpba*) and gamma subunit (gene *gpga*) [14–19].

The dynamics and mechanisms controlling cAR1-G$\alpha 2\beta\gamma$ coupling are not completely understood. The interaction between G$\alpha 2$ and G$\beta\gamma$ in live cells has previously been studied using a molecular proximity assay based on Förster (or fluorescence) resonance energy transfer (FRET) and has shown that the G protein subunit dissociation reflects their activation [20]. To investigate the dynamics of cAMP-induced G$\alpha 2\beta\gamma$ activation, we have used a similar molecular proximity method, but based on bioluminescence resonance energy transfer (BRET), which we have recently adapted for use in *Dictyostelium* [21]. BRET is similar to FRET but functions with a bioluminescent enzyme (luciferase) as energy donor and, consequently, does not suffer from problems related to fluorescence excitation as in FRET [22]. Several BRET methods have been developed, which differ in the use of different luciferases, luciferase substrates, and fluorescent

acceptor proteins [22, 23]. We have used the BRET^2 method with the *Renilla* luciferase variant 2 (*R*luc2) and GFP variant 2 (GFP^2) [24, 25]. We show that Gα2GFP^2 and *R*luc2Gβ are functional, that the BRET^2 signal detected in cells expressing both constructs reflects the interaction between the two G protein subunits, and that they behave as previously reported in experiments using FRET [20]. Our studies confirm that cAR1 is the main cAR mediating the cAMP-induced activation of Gα2βγ, with negligible role of cAR3, and kinetics analyses coupled to quantitative modeling of the cAMP-induced Gα2βγ subunit dissociation responses suggest that: 1) both uncoupled and Gα2βγ-pre-coupled cAR1 exist in resting cells; 2) the two cAR1 species have differential affinities for cAMP; and 3) the two cAR1 species contribute to the initial response of cells exposed to elevated cAMP concentrations through different kinetics.

2.2 Methods

2.2.1 Kinetics analyses

The net BRET^2 data was normalized using the average basal net BRET^2 measured before stimulation for each condition, which was set to 1. The net BRET^2 data were then fitted using the single exponential function: $\text{BRET}^2 = 1 + a(e^{-rt} - 1)$ or double exponential function: $\text{BRET}^2 = 1 + a_1(e^{-r_1 t} - 1) + a_2(e^{-r_2 t} - 1)$. Residuals of fitting are calculated by subtracting the data from the predicted values of the fit. To quantitatively determine the best fit, the error difference between single and double exponential fit was calculated as $\Delta E = (\text{RMSE}_1 - \text{RMSE}_2)/(\text{RMSE}_2)$, where RMSE_1 is the root mean squared error of the single exponential fit and RMSE_2 is that of the double exponential fit. From our analyses, we determined that a $\Delta E < 0.05$ indicates that the data can be fitted with a single exponential and a $\Delta E > 0.1$ indicates a double exponential function provides the best fit. A ΔE between 0.05 and 0.1 is ambiguous, meaning that the data could be fitted as well using a single or double exponential function. Every time curve obtained for different cAMP concentration was fitted using this standard, generating corresponding rates r

(single rate) or r_1 and r_2 (two rates).

2.2.2 Computational modeling

Six reactions were considered to model the cAMP receptor coupling to and activation of heterotrimeric G protein G$\alpha 2\beta\gamma$. The first reaction is the basal level of receptor (R) pre-coupled to the inactive, GDP-bound heterotrimeric G-protein (GD) in the absence of cAMP ligand (L), with rates k_1 and k_{-1}. The equation for pre-coupled receptor (RGD) without cAMP is defined as:

$$\frac{d[RGD]}{dt} = k_1[R][GD] - k_{-1}[RGD]$$

Without ligand, $[R] = R^{tot} - [RGD]$ and $[GD] = G^{tot} - [RGD]$. Consequently, the stable value of [RGD] before adding L is obtained by setting $\frac{d[RGD]}{dt} = 0$, and this is also the initial value for [RGD] after L is added:

$$[RGD]_{initial} = \frac{R^{tot} + G^{tot} + K_1 - \sqrt{(R^{tot} + G^{tot} + K_1)^2 - 4G^{tot}R^{tot}}}{2}$$

$K_1 = k_{-1}/k_1$ is the equilibrium dissociation constant and its reciprocal $1/K_1$ represents the binding affinity between R and GD. The second reaction is where L binds RGD to form LRGD, with rates k_2 and k_{-2}. The third reaction is where L binds R to form LR, with rates k_3 and k_{-3}. The fourth reaction is where LR then binds GD to form LRGD, with rates k_4 and k_{-4}. The fifth reaction is where the heterotrimeric G protein becomes active, GTP-bound (GT, with G$\alpha 2$ and G$\beta\gamma$ subunits dissociated) and dissociates from the receptor to form LR + GT. In *Dictyostelium* cells exposed to cAMP, because heterotrimeric G protein subunits don't adapt and display a steady state of activation as long as the stimulus is present (this study and [20]), we assume that this reaction is irreversible in the conditions used, with rate k_5. The sixth reaction is considering the cycling of GT back to GD with re-association of the G$\alpha 2$ and G$\beta\gamma$ subunits upon GTP hydrolysis, which

can then bind R or LR again, with rate k_6. To quantitatively explain the observed dissociation kinetics and represent the whole activation/deactivation cycle of the heterotrimeric G protein, the following 4 ordinary differential equations were used:

$$\begin{aligned}
\frac{[RGD]}{dt} =& k_1(R^{tot} - [RGD] - [LR] - [LRGD])(G^{tot} - [RGD] - [LRGD] - [GT]) \\
&+ k_{-2}[LRGD] - k_2[RGD][L] - k_{-1}[RGD] \\
\frac{[LR]}{dt} =& k_3[L](R^{tot} - [RGD] - [LR] - [LRGD]) + k_{-4}[LRGD] + k_5[LRGD] \\
&- k_{-3}[LR] - k_4[LR](G^{tot} - [RGD] - [LRGD] - [GT]) \\
\frac{d[LRGD]}{dt} =& k_2[RGD][L] + k_4[LR](G^{tot} - [RGD] - [LRGD] - [GT]) - k_{-2}[LRGD] \\
&- k_{-4}[LRGD] - k_5[LRGD] \\
\frac{d[GT]}{dt} =& k_5[LRGD] - k_6[GT]
\end{aligned}$$

These equations were normalized for $R^{tot} = 1$, leaving 11 independent parameters. For every value of [L], the equations were solved and the value $\frac{G^{tot} - [GT]}{G^{tot}}$ are compared to the normalized experimental data. The same procedure and standard as those for fitting the experimental data with single and double exponential functions were then used to obtain the exponential rates (r or r_1 and r_2) and amplitude of the response (a or $a_1 + a_2$) for the simulation results. These rates and amplitudes for different [L] are then fitted to experimental data to identify proper values for the 11 parameters. This fitting was carried by employing a simulated annealing method, using the simulannealbnd function in MATLAB(R2015b) with default settings, as we previously reported [26, 27]. First, an error function was defined as:

$$\text{Error} = \sum_i W_1(r^s_{1i} - r^e_{1i})^2 + W_2(r^s_{2i} - r^e_{2i})^2 + W_3(a^s_{1i} + a^s_{2i} - a^e_{1i} - a^e_{2i})^2$$

r_1, r_2, a_1 and a_2 are the exponential rates and response amplitudes obtained from a double exponential fit. When a single exponential fit is good, we set $r_1 = r_2 = r$ and $a_1 = a_2 = a/2$,

where r and a are obtained from a single exponential fit. Subscript "i" represents the ith group with $[L] = [L]_i$. W is the weight of different terms that can be tuned. Superscript "s" means simulation results and "e", experimental results. The simulated annealing prevents trapping the fit in a local minimum by accepting a trial step with some probability dependent on an artificial temperature T, even when this step does not improve the fit. The initial temperature was set to be high to allow for a larger searching area. Then, the temperature was gradually decreased, leading to more selective sampling towards the error decreasing direction. The position in the parameter space where the error function is minimal for each of the 11 parameters in the model was then determined: $k_1 = 0.002$; $K_1 = 0.962$; $k_2 = 0.007$; $K_2 = 1.24$; $k_3 = 0.502$; $K_3 = 78.4$; $k_4 = 0.214$; $K_4 = 1.04$; $k_5 = 0.905$; $k_6 = 0.075$; $G^{tot} = 0.989$. To simulate the heterotrimeric G protein subunit dissociation rates in the absence of pre-coupled receptors (RGD), the initial value for RGD was set to zero in the ODEs indicated above.

2.3 Results and discussion

To investigate the dynamics of cAMP-induced G$\alpha 2\beta\gamma$ activation, we measured G protein subunit dissociation in real-time in response to different cAMP concentrations and compared WT to *carC*$^-$ cells to determine if cAR3 may contribute to the observed kinetics (Figures 2.1 and A.1). Each time curve was fitted using single and double exponential decay and the goodness of fit was determined as described in Section 2.2.1 (Table A.1). Similar kinetics were observed using WT and *carC*$^-$ cells, supporting our previous observation that the detected cAMP-induced G$\alpha 2\beta\gamma$ dissociation is mainly due to cAR1 (Figure 2.1). Interestingly, we found that G protein subunit dissociation induced by cAMP concentrations of $\sim$ 500 nM and below are best fitted with a single exponential whereas that induced by cAMP concentrations above 500 nM are best fitted with double exponentials with two rates (Figures 2.1 and A.1, Table A.1). Similar kinetics were observed using wild-type and *carC*$^-$ cells, supporting our previous observation that the detected

cAMP-induced Gα2βγ dissociation is mainly due to cAR1 and that the appearance of a second rate at higher cAMP concentrations is not due to cAR3 signaling (Figure 2.1) Kinetics analyses of data obtained with Gα2GFP^2/*R*luc2Gβ-expressing *g*α2$^-$ cells also shows similar results (Figure A.2).

Interestingly, the cAMP concentration at which the presence of a second G protein dissociation rate becomes obvious correlates with cAMP concentrations at which most of the low affinity cAR1 binding sites ($K_d \sim 300$nM versus high affinity sites $K_d \sim 30$nM) should be occupied [7]. Evidence suggest that the high affinity cAR1 binding sites in Dictyostelium are receptors already in complex with the G protein, as is the case for many mammalian GPCRs [19, 28–34]. However, whether G protein pre-coupled cAR1 contributes to cAMP signaling is debated [35, 36]. To investigate the cAMP-induced Gα2βγ dissociation kinetics profile, we considered a simple model of receptor-G protein coupling, in which both free cAR1 (R) and cAR1 pre-coupled to inactive Gα2βγ (RGD) potentially exist in an equilibrium in resting cells (Figure 2.2A). In our model, we consider that upon stimulation with cAMP ligand (L), the latter can bind to R or RGD to generate LR and LRGD, respectively. In turn, LR then recruits GD to also lead to LRGD. Formation of LRGD stimulates exchange of GDP for GTP on the G protein, thereby activating it leading to LR + GT. In *Dictyostelium*, evidence suggest that the Gα2βγ subunits dissociate upon activation, so we consider that GT = Gα2 + Gβγ [20, 35, 37, 38]. Upon GTP hydrolysis on Gα2, the subunits then re-associate to form GD that can be activated again as long as cAMP is present. Six reactions total are considered, as shown in Figure 2.2A and described under Section 2.2.1. In our scheme, the appearance of two rates of G protein dissociation for higher cAMP concentrations is due to increased cAMP binding to uncoupled receptors to which it has lower affinity. Our scheme also incorporated that cAMP binding to lower affinity, uncoupled receptors produces a slower G protein activation response due to the additional G protein recruitment step. We generated a quantitative model of this scenario including both uncoupled (R) and pre-coupled receptors (RGD) that can simulate the kinetics data (Model R

+ RGD; Figure 2.2B). Interestingly, using this model, when we simulate the responses in the absence of pre-coupled receptors (Model R), we obtain only one rate throughout the range of cAMP concentrations used, supporting the hypothesis that the second rate comes from binding to pre-coupled receptors (Model R; Figure 2.2B).

In a previous study, Xu et al. considered the role of three possible cAR1 coupling mechanisms during chemotaxis: (1) $R + GD \rightarrow RGD + L \rightarrow LRGD \rightarrow LR + GT$ (with receptor pre-coupling); (2) $L + R \rightarrow LR + GD \rightarrow LRGDtoLR + GT$ (no pre-coupling); and (3) the previous two mechanisms combined (co-existence of uncoupled and pre-coupled receptors) [35]. As their results show that persistent ligand stimulation results in steady-state G protein activation, coupling mechanism (1) was ruled out, as it cannot allow a persistent cycle. However, the model of Xu et al. did not allow quantitatively discriminating between mechanisms (2) and (3) and, thus, did not determine whether pre-coupled receptors contribute to G protein activation during chemotaxis to cAMP. Compared to the model by Xu et al., our model uses fewer intermediate steps, parameters and variables, and does not have any spatial dynamics, which makes it easier to quantitatively fit our experimental data and provide insight into potential coupling mechanisms underlying the two G protein subunits dissociation rates that we observe. Our results suggest that pre-coupled receptors play a role in the rapid activation of the heterotrimeric G protein upon cAMP stimulation. Consequently, our results suggest that mechanism (3) described above, which includes contribution of both uncoupled and pre-coupled receptors, is likely the mechanism for cAMP-induced activation of heterotrimeric G proteins.

To test this prediction, we performed an experiment in which we stimulated the cells with two sequential cAMP concentrations. For the first stimulus, we used 10, 20, and 50 nM and, for the second stimulus, we used a saturating cAMP concentration of 10 μM (Figure 2.2C). We then analyzed the kinetics of the response after the second stimulus and compared to that obtained with cells directly stimulated with 10 μM cAMP. According to our prediction, lower doses of cAMP will bind and activate pre-coupled receptors RGD preferentially, reducing the number

of RGD seen by the second stimulus. Consequently, the double exponential in the G protein dissociation profile from stimulation with 10 μM cAMP should become gradually less obvious as the strength of the first stimulus increases. To determine the best fit for the data, we calculated the error difference (ΔE) between single and double exponential fit. Supporting our hypothesis, we found that the 10 μM cAMP-induced response has decreasing ΔE as the concentration of the first cAMP stimulus increases (Figure 2.2D). Therefore, this result supports the hypothesis that, in resting cells, two species of cAR1 with differential affinities for cAMP are responsible for the two rates observed when cells are stimulated with elevated cAMP concentrations.

cAR1 expressing cells have high and low affinity binding sites for cAMP of $\sim$ 30 and 300 nM, respectively [7]. Previous observations suggest that the high affinity cAMP binding sites are G protein-dependent [33, 34]. The cAMP concentrations inducing G protein subunit dissociation through two distinct rates correlate with cAMP concentrations at which both high and low affinity binding sites should be occupied. Therefore, we propose that the kinetics of cAMP/cAR1-induced G protein subunit dissociation that we observe result from the co-existence, in resting cells, of both uncoupled and pre-coupled cAR1 that have differential affinities for cAMP. At cAMP concentrations of $\sim$ 500 nM and below, mostly high affinity pre-coupled cAR1 sites are occupied and promote G$\alpha 2\beta\gamma$ activation, reflected by only one rate of G protein subunit dissociation. However, at cAMP concentrations above 500 nM, both high affinity pre-coupled cAR1 and low affinity uncoupled cAR1 sites are occupied. In this case, pre-coupled cAR1 leads to rapid G protein activation while uncoupled cAR1 first needs to recruit G$\alpha 2\beta\gamma$ before activating it, reflected by two distinct rates of G protein subunit dissociation.

Although cAMP binding affinities and cAR1 pre-coupling can explain the G protein dissociation rates, other factors may also contribute to the observed kinetics. Previous studies showed the presence of an intracellular pool of G proteins that undergo regulated translocation to the plasma membrane of Dictyostelium cells upon stimulation with higher cAMP concentrations [37, 39]. The cAR1 recruitment and activation of this translocated pool of G proteins at high

cAMP concentrations likely plays a role in the observed second rate of G protein dissociation. Furthermore, although cAR1 has been extensively studied, both the pharmacology and G protein coupling of cAR1 may be more complicated than what we currently know. For instance, many mammalian GPCRs are known to form homo- or hetero-oligomers, which can affect GPCR function in different ways, including modulating ligand binding and G protein coupling [40–42]. Whether cAR1 forms dimers or higher oligomers remains to be determined but it is possible that cAMP stimulation actually alters cAR1's oligomeric state, which could then lead to inducing G protein dissociation through distinct rates. Nonetheless, our findings indicate that initial cAMP stimulation of resting Dictyostelium cells occurs through two distinct rates that correlate with the high and low affinity cAR1 sites. Consequently, we propose that these different signaling dynamics may play a role in initial gradient sensing. On the other hand, during chemotaxis, when cells are constantly exposed to a cAMP gradient and the heterotrimeric G protein is predicted to continuously and rapidly cycle between active and inactive states, it is possible that only uncoupled cAMP-bound cAR1 (LR) exists in this situation, cycling between LR and LRGD forms as previously proposed by Xu et al. [35].

2.4 Conclusion

We used BRET^2 to investigate the dynamics and regulation of heterotrimeric G protein Gα2$\beta\gamma$ activation in response to cAMP stimulation in *Dictyostelium* by monitoring the dissociation of Gα2 and Gβ. Our kinetics analyses and receptor-G protein coupling modeling studies suggest that both uncoupled and Gα2$\beta\gamma$-coupled cAR1 exist in resting cells, and that the different binding affinities of cAMP for these two receptor species likely underlies the observed two rates of G protein subunit dissociation when cells are exposed to elevated cAMP concentrations. Although the study of protein-protein interactions by BRET is currently limited to measuring responses to acute stimulations, the information gained from these studies provide insights into how cells

initially respond to chemoattractant exposure and into mechanisms possibly involved in sensing gradients.

2.5 Acknowledgement

This work was supported by a Research Scholar Grant from the American Cancer Society (grant number RSG-15-024-01-CSM) to PGC. MS was also supported by NIHT32 grant GM008804, and WJR and HY also acknowledge support from NIH grant P01 GM078586. We are grateful to Chris Janetopoulos (University of the Sciences, Philadelphia, PA), Alan Kimmel (NIH, Bethesda, MD), Rick Firtel (University of California San Diego, La Jolla, CA), and Michel Bouvier (Universite de Montréal, Montréal, Canada) for providing material. We also want to thank the Dicty Stock Center for facilitating sharing of materials and the material depositors Peter Devreotes (Johns Hopkins, Baltimore, MD), Douglas Robinson (Johns Hopkins, Baltimore, MD) and Douwe Veltman.

Chapter 2, in full (with minor exceptions), is a reprint of the material in part as it appears in "The cAMP-induced G protein subunits dissociation monitored in live Dictyostelium cells by BRET reveals two activation rates, a positive effect of caffeine and potential role of microtubules. Islam, A.F.M. Tariqul; Yue, Haicen; Scavello, Margarethakay; Haldeman, Pearce; Rappel, Wouter-Jan; Charest, Pascale G., Cellular Signaling, 2018". The dissertation author was the primary investigator and author of this paper.

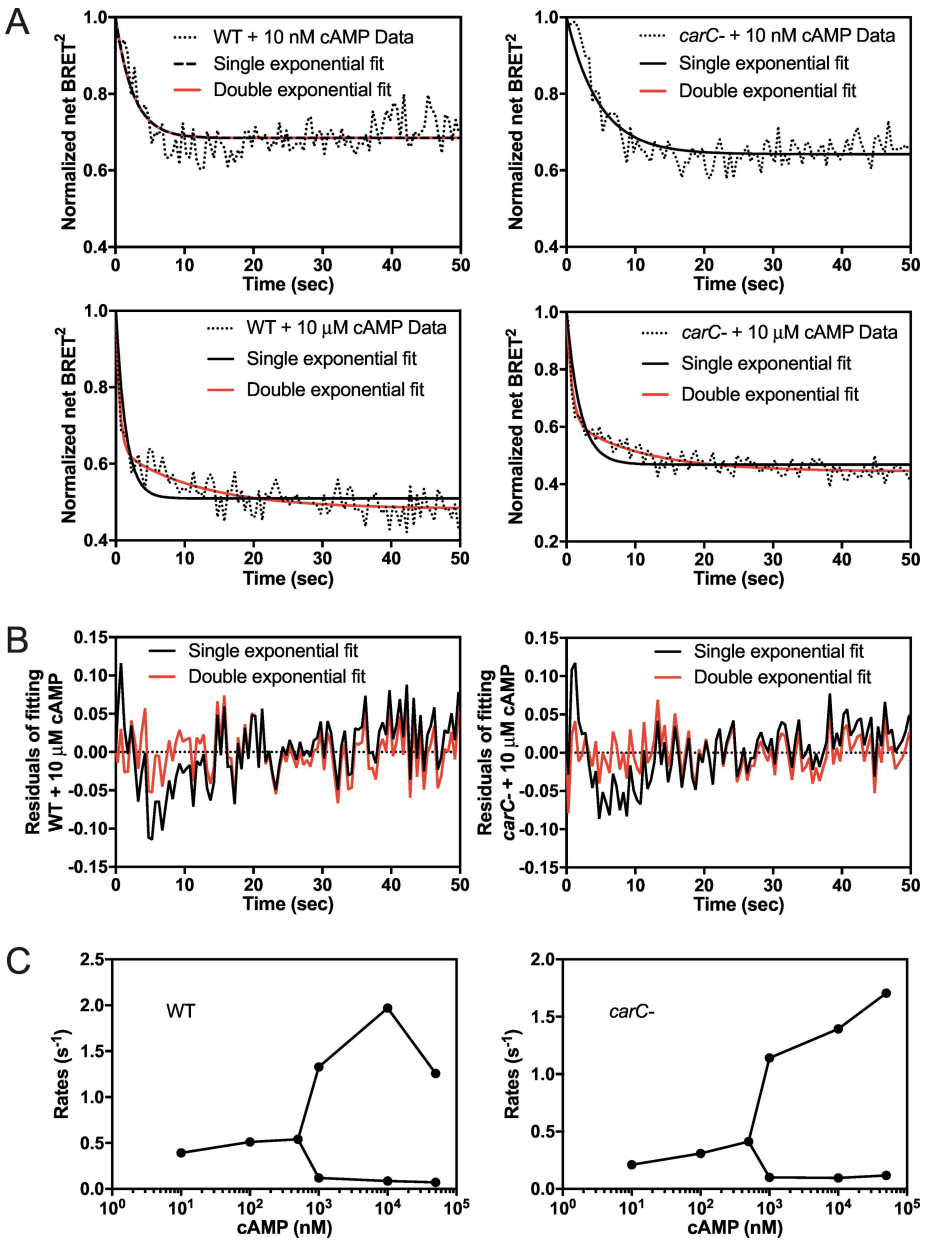
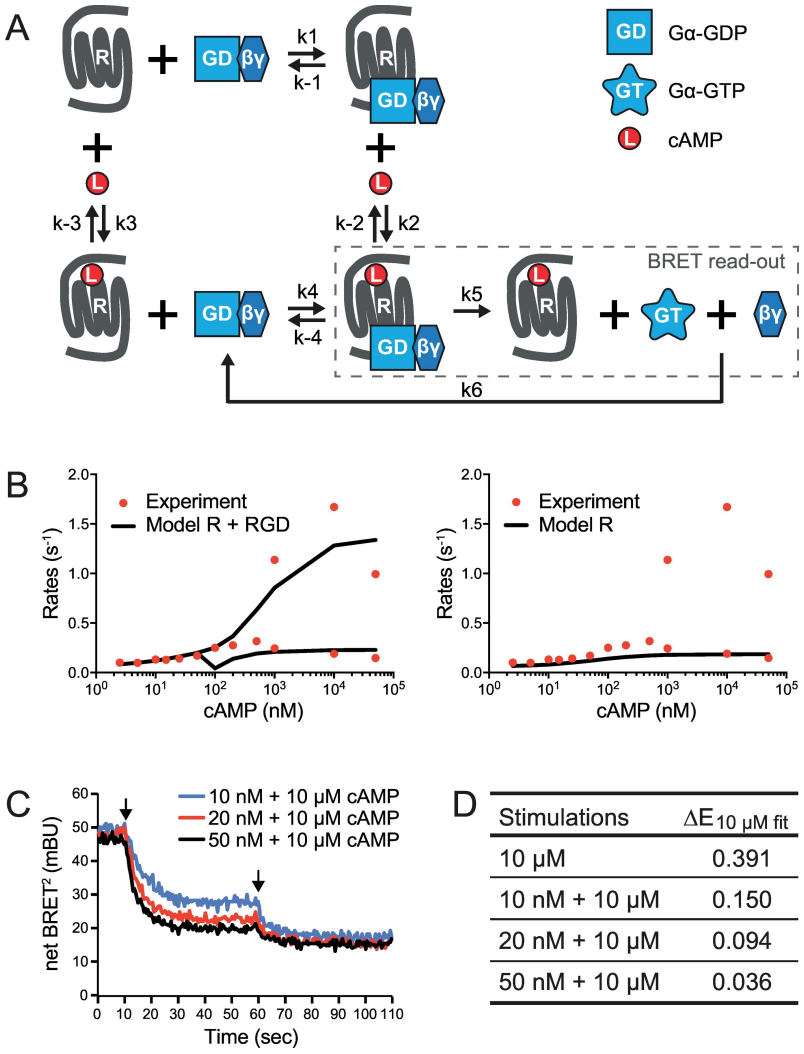


Figure 2.1: Kinetics analyses of cAMP-induced Gα2βγ subunit dissociation. A, net BRET^2 values for measurements taken every 0.5 s after stimulation of wild-type(WT) and *carC*$^-$ cells expressing Gα2GFP^2 and *R*luc2Gβ with 10 nM or 10 µM cAMP, normalized to BRET^2 levels before stimulation. The curves were fitted using either single or both single and double exponential functions. For the 10 nM data, the fits are identical and overlap. B, Residuals for single and double exponential fittings for the 10 µM cAMP-stimulated conditions shown in A were calculated by subtracting the measured values from the fitted values. C, Rates of Gα2βγ subunit dissociation determined from the time curve fittings obtained for 10 nM, 100 nM, 500 nM, 1 µM, 10 µM, and 50 µM cAMP stimulations of WT and *carC*$^-$ cells. Results represent the mean or the analyses of at least three independent experiments.



Stimulations	$\Delta E_{10\ \mu M\ fit}$
10 μM	0.391
10 nM + 10 μM	0.150
20 nM + 10 μM	0.094
50 nM + 10 μM	0.036

Figure 2.2: Quantitative modeling of the observed Gα2βγ subunit dissociation kinetics. A, Conceptual model of cAR1-Gα2βγ coupling mechanism. The constants are defined in the Section 2.2.1. B, The rates of Gα2βγ subunit dissociation induced by different concentrations of cAMP stimulus obtained from analyzing the experimental data (red circles; data from Figure 2.1C, WT cells) and from quantitatively modeling the responses where both uncoupled (R) and pre-coupled (RGD) cAR1 bind the cAMP ligand (L) or where only uncoupled cAR1 is present and mediates the response to cAMP (black curves). C, Representative net $BRET^2$ values obtained from measurements taken continuously every 0.5 s before, during and after injection of a first cAMP stimulus of 10, 20, or 50 nM at 10 s, and of a second cAMP stimulus of 10 μM at 60 s of Gα2GFP^2+*R*luc2Gβ/*g*α2 null cells. Solid arrows indicate when the two sequential stimuli were injected. D, Calculated error differences (ΔE) for single and double exponential fittings of the Gα2βγ subunit dissociation kinetics induced by 10 μM cAMP stimulation alone (from data presented in Figure 2.1A) or after the first stimuli indicated in C. ΔE calculations and interpretations are described in Section 2.2.1 and main text. Results represent the mean or the analyses of at least three independent experiments.

Chapter 3

Cellular memory in eukaryotic chemotaxis

3.1 Introduction

Eukaryotic chemotaxis—the directed motion of cells along spatial gradients of chemicals—plays an essential role in a wide variety of biological processes, including embryogenesis, neuronal patterning, wound healing, and tumor dissemination [43–47], and many of its molecular components are conserved across cell types [48, 49]. Much work has been devoted to understanding chemotaxis in static gradients [50, 51] and has revealed that cells are highly sensitive to spatial cues [52, 53]. Natural chemical gradients, however, are often dynamic [54, 55], and chemotaxis in such environments requires an integration of spatial and temporal cues which is poorly understood. One striking example is the self-organized chemoattractant field arising during the development of the social amoeba *Dictyostelium* following nutrient deprivation. Here, nondissipating waves of chemoattractant travel outward from aggregation centers and provide stable long-range cues to direct the migration of cells toward the wave source. In a symmetric traveling wave, the spatial

gradients in the front and back halves of the wave are equal in strength, but opposite in direction. Hence, if a cell responded simply to spatial information, it would move forward in the front of the wave and backward in the back of the wave. Thus, additional processing is needed for cells to solve the resulting back-of-the-wave problem [56] and to move efficiently toward the wave source.

In principle, cells could distinguish between the front and back of the wave by the temporal gradient—the concentration increases in time in the front of the wave and decreases in time in the back of the wave. Temporal gradient sensing plays a fundamental role in bacterial chemotaxis [57] and entails keeping a short-term memory of the stimulus via an adaptation system. Past studies have suggested that short-term memory also plays a role in eukaryotic chemotaxis. In spatially uniform concentrations of chemoattractant, neutrophils were observed to maintain their polarity when the concentration was increasing, but reverse polarity when the concentration was decreasing [58]. Likewise, *Dictyostelium* and neutrophils were shown to exhibit chemokinetic responses to temporal oscillations of chemoattractant [59, 60] . The mechanisms underlying this behavior are not well understood. However, the directional sensing markers, activated Ras and its downstream targets, have been shown to adapt on a timescale of 10–30 s [26, 61, 62], providing a potential mechanism for shortterm memory.

In addition to short-term memory, migrating cells also exhibit polarity, manifested by an elongated cell shape with a defined front and rear and polarized distributions of signaling molecules. During chemotaxis, cells polarize in the gradient direction and can reverse their polarity when the gradient is changed [63–65]. In uniform chemoattractant, cells undergo a random walk with a persistence time of $\sim$3–10 min [66–69]. This persistence of migration, which is likely tied to maintenance of polarity, is indicative of a long-term memory with a timescale similar to the Dictyostelium wave period ($\sim$6 min). Long-term memory could enable cells to remember the gradient direction experienced in the front of the wave, as the wave passes by. However, the precise roles of short- and long-term memory in solving the back-of-the-

wave problem have not been determined. In particular, it is unclear how memory is coupled to directional sensing to enable chemotaxis toward the source of a traveling chemical wave.

Here, we used microfluidics to gain quantitative insight into the back-of-the-wave problem by measuring Dictyostelium chemotaxis in traveling waves of varying period. We found that, for natural periods, cells exhibited cellular memory and maintained direction toward the wave source in the back of the wave, whereas for longer periods, cells increasingly reversed direction. To connect this cellular behavior to known signaling pathways, we also characterized the response of a gradient sensing marker, activated Ras, to rapid changes in spatial gradients and found evidence for both short- and long-term memory at the level of Ras activation. Our results can be explained by a model that couples an adaptive directional sensing module to a bistable memory module and provides a framework for understanding chemotaxis in spatiotemporal gradients.

3.2 Experimental Results

First, we used microfluidics to study Dictyostelium chemotaxis in traveling waves of cAMP. The microfluidic wave generator (Figure 3.1A) periodically sweeps a hydrodynamically focused stream of cAMP across the chemotaxis channel. Within one period, a bell-shaped profile of cAMP with a peak of $\sim$700 nM, created by the molecular diffusion out of the stream, moves across the chemotaxis channel at a constant speed (Figure 3.1B) and generates spatiotemporal cues (Figure 3.1B) similar to those measured for natural waves of cAMP [70, 71]. During aggregation, the period of natural waves decreases from 10 to 6 min, as stable aggregation centers form [72]. We first tested chemotaxis of 5-h-developed cells in waves with period T = 6 min. We tracked cells as they migrated in response to the periodic traveling wave stimulus for up to 2 h and measured the instantaneous chemotactic index, defined as the velocity in the direction toward the wave source divided by the speed, $CI = V_x/V$, as a function of time from the passage of the peak of the wave. Cells showed excellent chemotaxis during the passage of the front of the wave,

where the chemoattractant gradient pointed toward the source and, surprisingly, maintained their directed migration with only a slight decrease in CI in the back of the wave, where the gradient was reversed (Figure 3.1C).

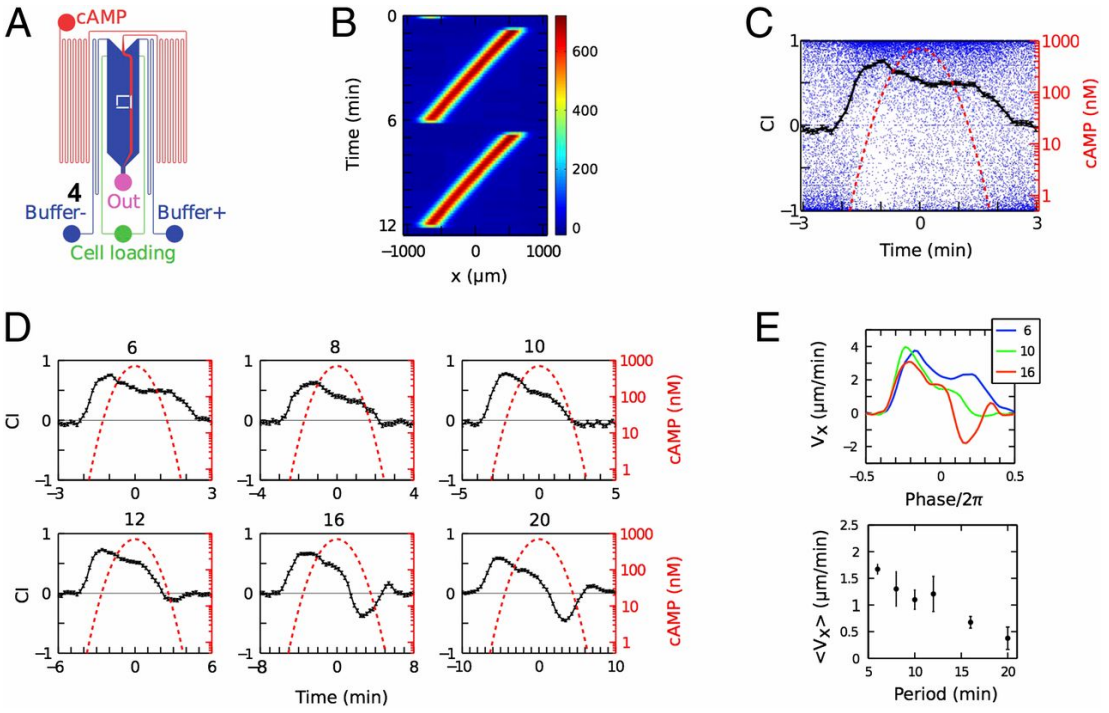


Figure 3.1: Wave behavior. (A) Schematic of microfluidic wave generator, with observation region indicated by white box. (B) Kymograph of cAMP concentration measured with fluorescein dye. (C) Chemotactic index (CI) as a function of time with respect to the peak of the wave (T = 6 min; blue dots: individual cells; black curve: binned average; dotted red line: cAMP concentration). Error bars are SEM. (D) Average CI as a function of time for different wave periods (min). (E) Average velocity toward the wave source as a function of the phase of the wave for different periods (Upper) and average migration velocity as function of wave period (Lower). Error bars represent day-to-day variation.

Next, we increased the wave period by decreasing the wave speed, while keeping the spatial profile constant. In the limit of long period, or small wave speed, the spatial gradient experienced by the cell becomes almost static. Thus, we expected cells to eventually reverse in the back of the wave because of the high sensitivity of cells to static spatial gradients [52, 53]. We found that cells in waves with periods in the range of T = 6 – 10 min maintained directed motion in the back of the waves for ∼2 min,indicative of cellular memory (Figure 3.1D). In T = 12 min

waves, the CI was detectably negative in the back of the wave and in waves with $T \geq 16$ min, cells fully reversed, consistent with our expectations. The average migration velocity $\langle V_x \rangle$, computed as the net average distance traveled per wave period toward the wave source divided by the period T, decreased for increasing wave periods (Figure 3.1E).

To connect the observed behavior of cells in waves to known signaling pathways, we studied the localization of a directional sensing marker, activated Ras, using the Ras-binding domain of Raf tagged to GFP (RBD-GFP) [73]. Ras activation is an early response of the chemotaxis network and activated Ras localizes in patches at the front of the cell in a static gradient [74, 75]. Cells were subjected to rapid, reversible switching between an initial gradient and a final gradient in 2-μm-deep microfluidic gradient chambers, where flattening of the cells facilitated both the imaging and application of well-defined linear gradients [76]. Linear spatial gradients in the chemotaxis chambers were created by diffusion and their strength was a function of the chamber length, L, and the cAMP concentration in the two side channels, C_R and C_L: $(C_R - C_L)/L$.

We explored cellular memory by measuring the spatiotemporal dynamics of Ras activation when a gradient was suddenly replaced with a uniform concentration. In the wave experiments, the maximal response of the CI was observed for concentrations between ∼10 and 100 nM, corresponding to relative spatial gradients of ∼15–30% across a 10-μm cell. To apply comparable conditions, we subjected cells to linear gradients $0 - 100$ nM (corresponding to the side-channel concentrations) and studied cells in the center of the chamber of L = 120 m, experiencing a local mean concentration of 50 nM with a relative spatial gradient of 17% across 10 μm. When cells in this gradient were rapidly exposed to a uniform concentration of equal or higher value, activated-Ras patches were maintained at the old front (Figure 3.2A). When cells experienced a drop in mean concentration, the patches disappeared immediately, but reappeared at the original front following a delay, which increased as the final uniform concentration decreased (Figure 3.2A). Cell movement followed the Ras-activation pattern: cells that experienced an increase in

concentration continued their movement in the original gradient direction until they exited the chamber, whereas those that experienced a decrease in concentration stopped moving when the activated-Ras patches disappeared and resumed movement concurrent with patch reappearance (Figure 3.2A).

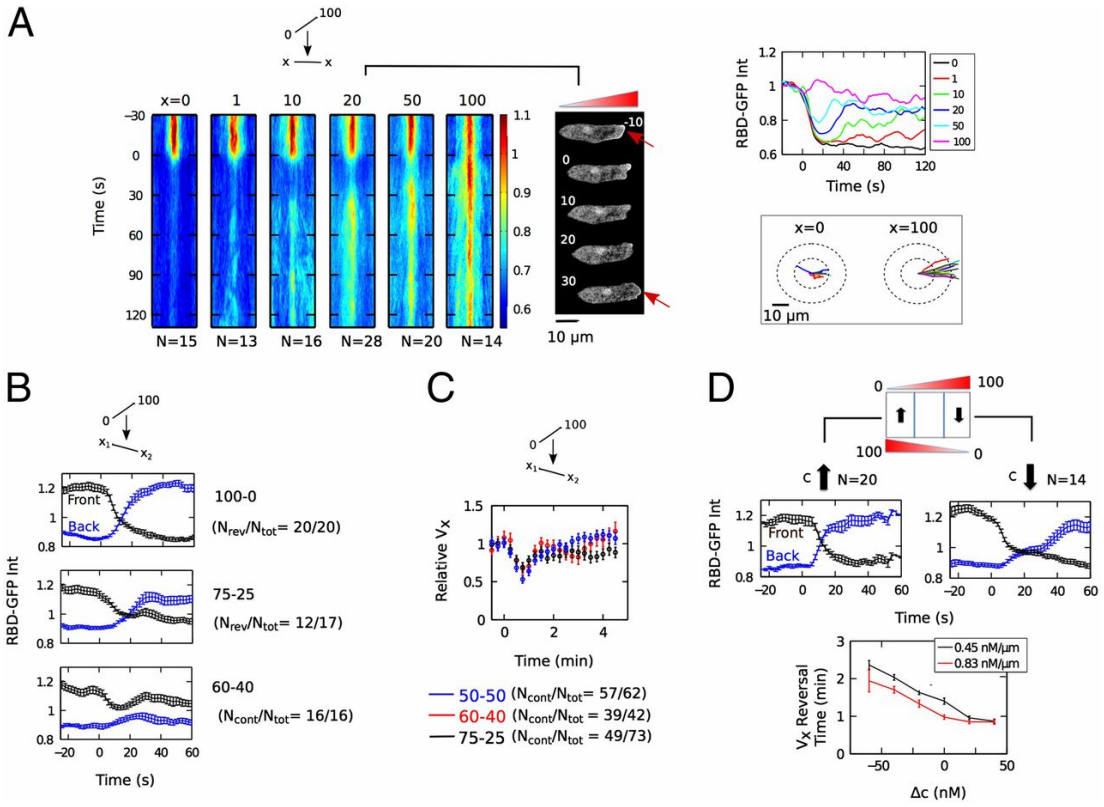


Figure 3.2: Ras activation dynamics during gradient switches. (A) Kymographs of average normalized RBD-GFP membrane-to-cytoplasm intensity ratio, when a 0–100 nM gradient is switched to spatially uniform cAMP at the indicated concentrations x (in nanomolar concentration) and image sequence for x = 20 nM (time in seconds; RBD-GFP patches indicated by arrows). Also shown are the corresponding RBD-GFP patch intensities as a function of time and cell trajectories for x = 0 and 100 nM following the switch. (B) Average RBD-GFP membrane-to-cytoplasm intensity ratios of the front and back halves of cells under gradient reversals. (C) Average normalized cell velocity in the direction of the initial gradient, V_x, following reversal at time 0 of a gradient, 0–100 nM, to a weaker or no gradient (75–25, 60–40, and 50–50 nM). The velocity was averaged only over the cells that failed to reverse to the weaker gradient and instead continued movement in the original gradient direction. The fractions of cells found to continue movement in the original gradient direction, N_{cont}/N_{tot}, is shown in the legend. (D) Average RBD-GFP membrane-to-cytoplasm intensity ratio for increasing and decreasing local concentrations during a 0–100 to 100–0 nM gradient reversal and velocity reversal time as a function of the change in local concentration. All error bars represent SEM.

To quantify the competition between cellular memory and the ability of cells to respond to new gradients, we subjected cells in a 0–100 nM gradient in a L = 70 μm chamber to reversals to increasingly weaker gradients, while keeping the mean concentration experienced by the cell fixed. For equally strong reversed gradients (100–0 nM), all cells reversed their migration, whereas for intermediate gradients (75–25 nM), a fraction of the cells did not reverse (5 of 17) and for weak gradients (60–40 nM), all cells continued moving in the original direction (Figure 3.2B). Furthermore, after reversals to these weak gradients, activated Ras remained predominantly localized in the old gradient direction (Figure 3.2B), indicating that cellular memory overcomes weak gradient cues at the level of Ras activation. To determine the stability of cellular memory, we followed cells in weak reversed gradients and uniform stimuli for longer periods following the switch using longer chambers (L = 220 μm). We quantified the average velocity in the direction of the original gradient, V_x, normalized by its value before gradient reversal. Cell tracking revealed persistence of migration in the original direction, against the new gradient, for at least 5 min (Figure 3.2C), suggesting long-term memory comparable to that observed in the wave experiments.

Next, we measured the response of cells to gradient reversals, 0–100 to 100–0 nM, as a function of the change in the local mean concentration by examining cells in different regions of the chemotaxis chamber. Cells in the left and right regions of the chamber experienced increases and decreases in the mean concentration, respectively (Figure 3.2D). The reversal time, measured by quantifying activated Ras and by computing the time for the cell velocity V_x to drop below a reversal threshold of –2 μm/min, depended strongly on the change in the mean, with a decrease in the mean resulting in a significant delay in reversal (Figure 3.2D). The delay in Ras reversal was comparable to the deadaptation time of the Ras response to uniform stimulation [26, 77], suggesting that adaptive, short-term memory regulates directional sensing, consistent with the local excitation global inhibition (LEGI) mechanism explained below.

Last, we explored the link between cellular memory and polarity by assaying cells, which

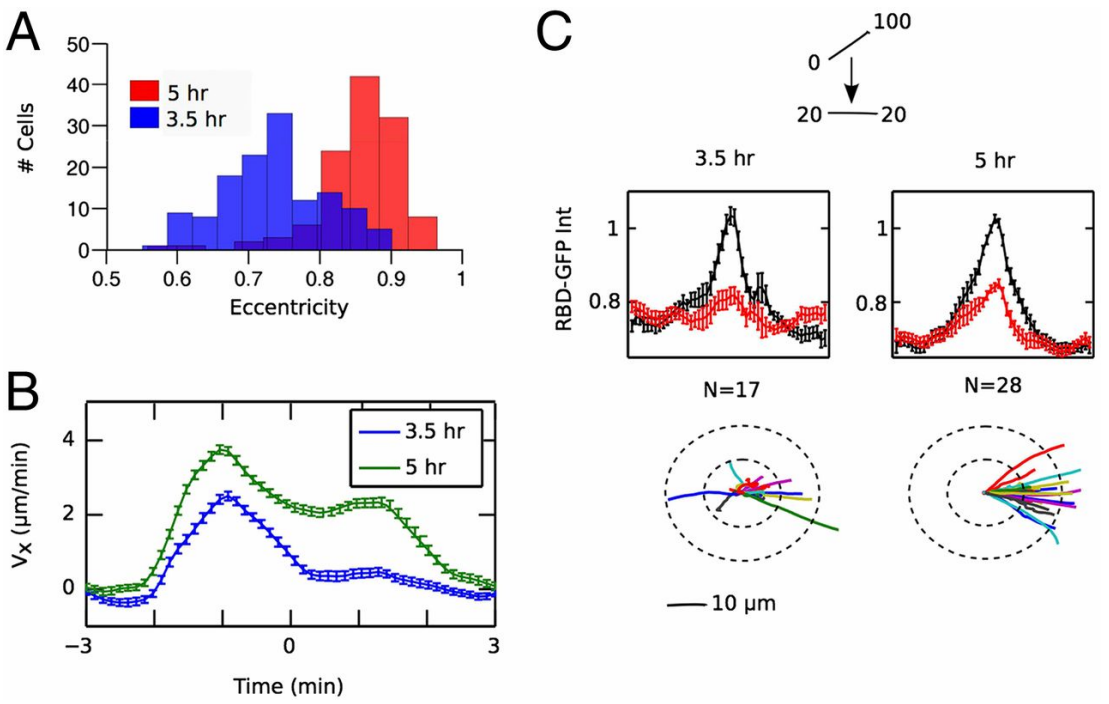


Figure 3.3: Development time and cellular memory. (A) Distributions of eccentricity for 3.5-h-developed cells [blue; mean, 0.73 ± 0.07 (SD)] and 5-h-developed cells [red; mean, 0.86 ± 0.06 (SD)]. Shown are the time-averaged eccentricities of cells during a 6-min wave experiment. (B) V_X as a function of time in 6-min waves for 3.5-h-developed cells and 5-h-developed cells. (C) Spatial distributions of normalized RBD-GFP membrane-to-cytoplasm intensity ratio averaged over the time before (–30 to 0 s; black) and after (30–120 s; red) the switch from 0–100 to 20–20 nM for 3.5- and 5-h-developed cells with the corresponding trajectories of cells following the switch shown below.

had only been developed for 3.5 h. These cells are less polarized than 5-h-developed cells, as manifested by their rounder shapes (Figure 3.3A). Unlike 5-h-developed cells, the 3.5-h-developed cells did not maintain directed movement in the back of 6-min waves (Figure 3.3B). Moreover, 3.5-h-developed cells also did not show return of patches to the old front following the replacement of a 0–100 nM gradient with a lower uniform concentration and, correspondingly, lost their directional movement (Figure 3.3C). Thus, cellular memory, like polarity, is a function of development, consistent with previous observations of developing populations showing that the response of cells to natural waves of cAMP becomes more directed over time [78]. Although the molecular basis of cellular memory remains to be determined, we speculate that the delayed

onset of cellular memory may play an important role in the self-organized aggregation process by enhancing chemotaxis only after coherent aggregation centers have formed.

3.3 A model with an ultrasensitive LEGI module coupled to a bistable memory module

To better understand the coupling between directional sensing and the observed short- and long-term memory, we modified our previous mathematical model of adaptation in which an incoherent feedforward loop regulates RasGTP [26]. This model contains a LEGI module in which ligand-bound receptor (R) activates a membrane localized activator, E, as well as a diffusible inhibitor, I, that both regulate a response element, RE [79]. Quantitative comparisons with the experimental data can be performed by assuming that RE is RasGTP, as in our earlier work [26] or that RE directly controls the level of activated Ras, possibly through an excitable pathway [62, 77]. The ratio of the activator to the inhibitor provides an internal representation of the gradient, whereas the difference in kinetics between the fast activator and slow inhibitor enables adaptive dynamics. Amplification is achieved through ultrasensitive regulation of RE by the local activator E and the global inhibitor I [80]. This LEGI module could not, however, account for the long-term cellular memory revealed by our experiments (Figure 3.8b). Therefore, we coupled the LEGI module to a memory module whose output, M, is regulated by RE, but also positively feeds back to RE (Figure 3.4A). Because the memory observed in our experiments lasted for greater than 5 min (Figure 3.2C), we chose this module to have bistable rather than simple first-order decay dynamics. Bistable dynamics display switch-like behavior and have been incorporated into several models for cell polarity [81–83]. In our module, M is turned on and off through a threshold mechanism dependent on RE and its own history. See Section 3.3.2 for details.

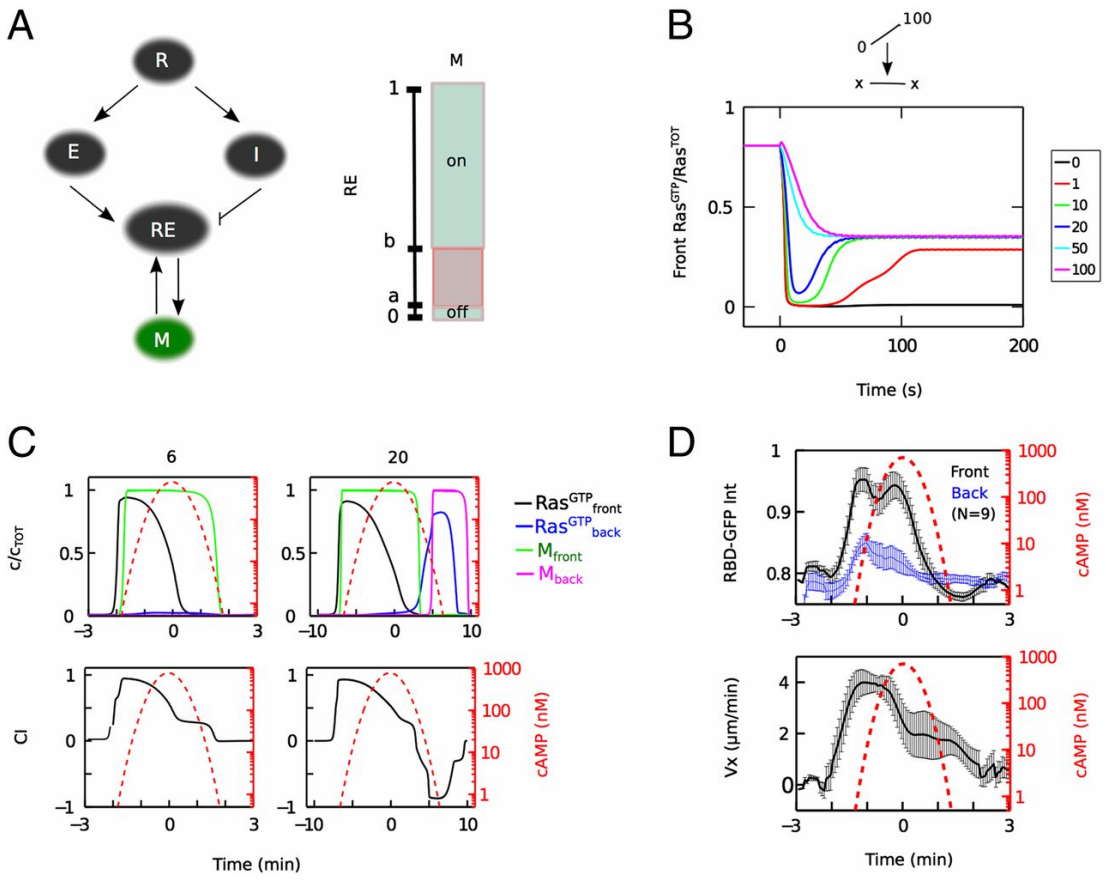


Figure 3.4: Ultrasensitive LEGI coupled to bistable memory model. (A) Schematic of feedforward LEGI regulation of the response element RE by E (local activator) and I (global inhibitor) with receptor occupancy input R, coupled to a bistable memory module M. M is turned on or off as a bistable switch depending on RE relative to the thresholds a and b and its own history. The output of the LEGI module, RE, is assumed to correspond to the experimentally measured RasGTP. (B) Model response of front RasGTP for switching from a gradient (0–100 nM) to a uniform concentration [x (nM)]. Compare with Figure 3.2A. (C) Model predictions for 6- and 20-min waves showing the components at the front and back of the cell (upper row) and chemotactic index, computed as CI = (RE+M)/2 (lower row). Compare with Figure 3.1D. (D) Average RBD-GFP membrane-to-cytoplasm intensity ratio at the front (black) and the back (blue) of the cell (from confocal imaging) and average cell velocity towards the wave source (V_x) measured during a 6-min wave period. Error bars represent SEM.

3.3.1 Modeling Results

The model is described by ordinary differential equations (ODEs) shown in Section 3.3.2 and the parameters were fitted using a subset of the single-cell RBD-GFP data together with existing uniform response data in our earlier work [26]. Details of the data fitting is in Section

3.3.3. With these fitted parameters, our model accounts for the experimental RBD-GFP data (Figure 3.4B and Figure 3.5). In the absence of a chemoattractant, M is in its low state and does not contribute to Ras activation. The kinetics of switching of M are sufficiently slow that uniform step stimuli do not keep Ras activated long enough to turn on M. So, for uniform step stimuli, the model with M module in this paper behaves almost the same as the LEGI model without the M module in our earlier work [26] and shows good agreement with the experimental results (Figure 3.5a). However, in steep static gradients, M gets activated into its high state at the front of the cell, while remaining in its low state at the back of the cell. When the gradient is replaced by a lower uniform concentration, M remains in its high state and serves as a memory of the initial gradient. The model reproduces both the fast disappearance of the RBD-GFP patches, due to the fast activator kinetics, and the concentration-dependent delay before patch reappearance at the front of the cell (Figure 3.4B), due to the slow degradation of the inhibitor. The delay of the reversal time when the mean cAMP decreases, shown in Figure 3.2D, is also repeated by the model in Figure 3.5b. Importantly, the relative contribution of M to Ras activation is small compared with that of the stimulus-controlled activator E. Consequently, memory does not interfere significantly with reversals to strong gradients and is only revealed at the level of Ras activation when initially strong gradients are replaced by weak reversed gradients or uniform stimuli (Figure 3.4B and Figure 3.5c).

We then determined the response of the model, without additional parameter modifications, to traveling wave stimuli (Figure 3.4C). In the 6-min wave, Ras is activated in the anterior of the cell during the passage of the front of the wave but is never activated in the posterior of the cell in the back of the wave due to the slow decay of the inhibitor I (more discussion is in Section 3.4). Thus, the temporal dynamics of the LEGI mechanism prevent a reversal in signaling components. The memory at the anterior, on the other hand, stays on due to its slow switching kinetics (Figure 3.4C). In contrast, for 20-min waves, I has sufficient time to decay as the wave passes over the cell to allow full Ras activation in the posterior of the cell in the back of the wave (Fiugre 3.4C). Our

model can reproduce the experimental wave data (Figure 3.1D) if we assume that cell motility depends linearly on both M and RasGTP (Figure 3.4C and Figure 3.6). Then, for 6-min waves, cell migration in the back of the wave persists in the absence of activated Ras, whereas for 20-min waves cells fully reverse. One prediction of the model is that Ras does not stay activated in the anterior of the cell in the back of the wave. This prediction was experimentally verified using confocal imaging of RBD-GFP localization (Figure 3.4D).

3.3.2 Equations

Our model, shown schematically in Figure 3.4A, contains two coupled modules. The first module is similar to our earlier work [26] and contains an incoherent feedforward local excitation global inhibition (LEGI) mechanism. A recent study has found that activated Ras, RasGTP, can exhibit excitable dynamics [77]. Because the LEGI module is not excitable, this suggests that the output of the LEGI module is a component that is upstream from RasGTP, rather than RasGTP itself. Thus, our module is written in general terms where its output, the response element RE, is activated by a local activator E and is deactivated by a global inhibitor I. Comparison with the experiments can be made by assuming that RE directly regulates activated Ras. In other words, RasGTP data from the experiments were directly fitted to RE levels from the model. To account for the observed amplification of RasGTP localization, we use an ultrasensitive version of the LEGI model, which naturally amplifies the external ligand gradient. The output of the LEGI module RE feeds into the memory module, whose output in turn feeds back into the LEGI module and provides the long-term cellular memory. The complete model is written in terms of differential equations:

$$\frac{d[R_1]}{dt} = k_{R1}(cAMP + r_1)\left(R_1^{tot} - [R_1]\right) - k_{-R1}[R_1] \tag{3.1}$$

$$\frac{d[R_2]}{dt} = k_{R2}(cAMP + r_2)\left(R_2^{tot} - [R_2]\right) - k_{-R2}[R_2] \tag{3.2}$$

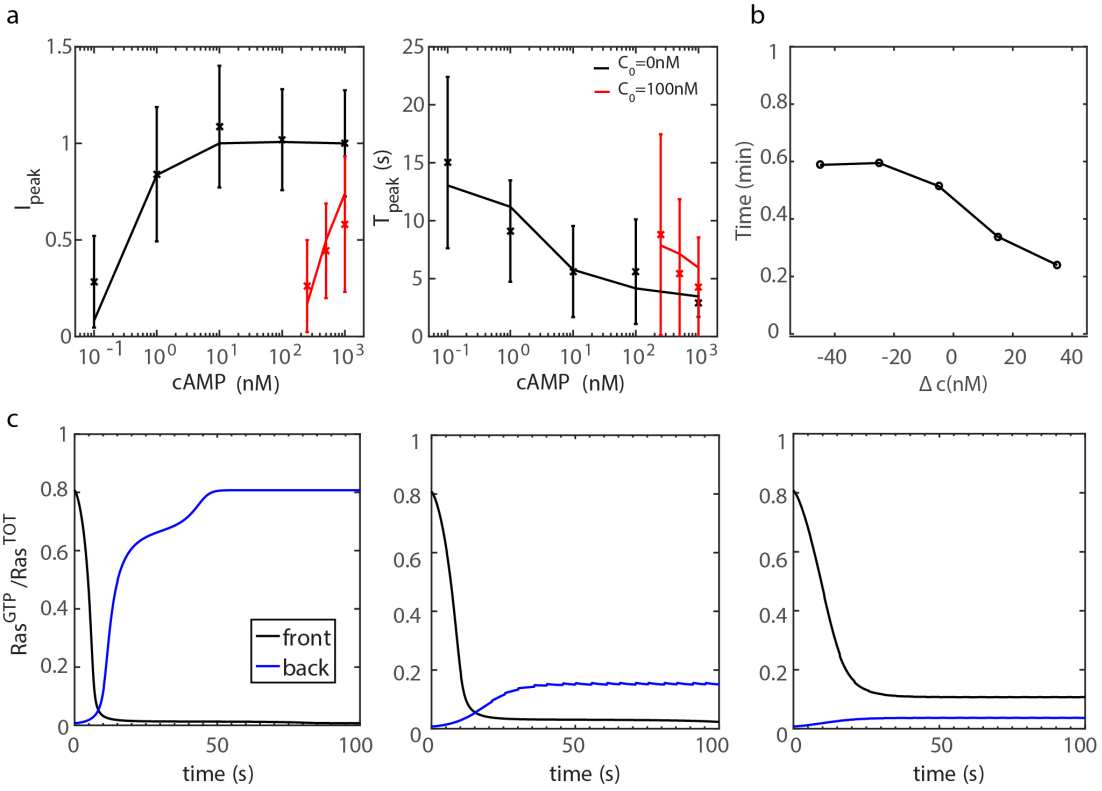


Figure 3.5: Modeling results for experiments of uniform step stimuli and gradient-reversal. (a)Comparison between experimental results of [26] (symbols) and our simulation results (lines) for rapid changes in uniform cAMP concentration. Cells were pretreated with a cAMP concentration of 0 nM (black) and of 100 nM (red). Both the numerically computed intensity peak, I_{peak}, as a function of the stimulus strength and the time corresponding to the maximum peak amplitude show good agreement with the experimental results. Essential to this agreement is the small value of the activation rate of M, k_{Mem}, which ensures that M remains in its low stable state after a uniform step-change in stimulus. (b) Reversal time in the model shown as a function of the change in local mean concentration for a 0–100 to 100–0 nM gradient reversal in a chamber of L = 200 μm. The reversal time in our numerical simulations was determined as the time point for which RasGTP at the back has reached 50% of its final, steady-state value following gradient reversal at time 0. As in the experiments (Figure 3.2D), this reversal time increases for larger decreases in the mean. (c) Activated Ras at the front (black) and back (blue) in our model, computed after a 0–100 to 100–0 nM reversal (Left), a 0–100 to 75–25 nM reversal (Center), and a 0–100 to 60–40 nM reversal (Right). As required, and consistent with our experiments, the activated-Ras distribution reverses in the 100–0 and 75–25 nM gradients but fails to reverse in the 60–40 nM gradient.

$$[R] = [R_1] + [R_2] \tag{3.3}$$

$$\frac{d[E]}{dt} = k_E[R] - k_{-E}[E] \tag{3.4}$$

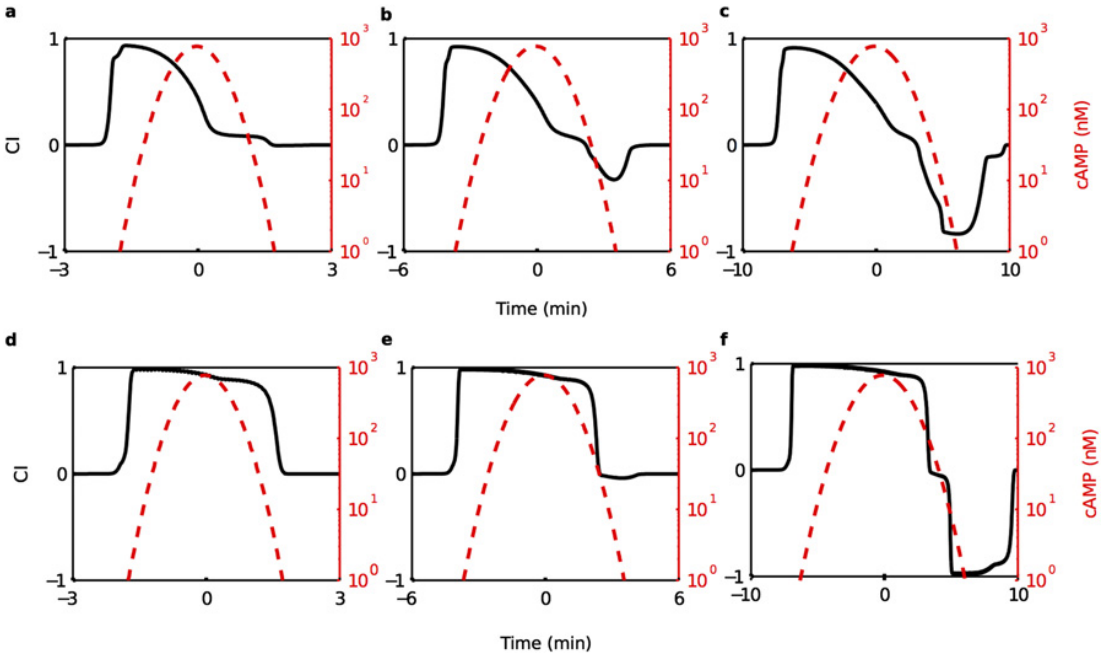


Figure 3.6: Motility response of our model shown for different wave periods. CI was computed by assuming that motility is a linear combination of the RE level representing RasGTP, and the value of M: CI=$\lambda(\alpha RE+\beta M)$, where α and β are parameters determining the relative weight of activated Ras and M to cell movement and where λ is an overall scale factor. The assumption that motility is not solely determined by activated Ras is motivated in part by the observation that cells in waves continue their motion even though RasGTP is no longer localized at the old front (Figure 3.4C). Furthermore, because M is downstream from Ras activation it is plausible that M affects the actin machinery and, hence, motility. Shown in a–c is a parameter combination in which RE contributed most to the motility of the cell ($\alpha = 0.9$ and $\beta = 0.1$), resulting in CI curves that are inconsistent with experimental data: CI drops to very low values after the peak of the wave has passed and does not exhibit the plateau that is prominently present in Figure 3.1D. In d–f, the CI is shown assuming that M contributes most prominently to motility ($\alpha = 0.1$ and $\beta = 0.9$). Compared with the experimental curves (Figure 3.1D), the CI remains large for a much longer period after the peak of the wave. Through numerical investigations, we have determined that the ratio presented in Figure 3.4C ($\alpha = 0.5$ and $\beta = 0.5$) leads to CI curves that qualitatively fit the experimental data best.

$$\frac{d[I]}{dt} = k_I\left(k_{basal} + [R]_{avg}\right) - k_{-I}[I] \tag{3.5}$$

$$\frac{d[M]}{dt} = -k_{Mem}[M]\left([M] - M^{tot}\right)\left([M] - M^{tot}\frac{b-[RE]}{b-a}\right) \tag{3.6}$$

$$\begin{aligned}\frac{d[RE]}{dt} = {} & k_{RE}[E]\frac{RE^{tot}-[RE]}{K_{m1}+RE^{tot}-[RE]} - k_{-RE}I\frac{[RE]}{K_{m2}+[RE]} \\ & + k_{RE2}[M][R]\frac{RE^{tot}-[RE]}{K_{m3}+RE^{tot}-[RE]}\end{aligned} \tag{3.7}$$

Equations 3.1 and 3.2 describe two types of receptors, R_1 and R_2, with different binding affinities for cAMP (one with a large K_d value and one with a small K_d value) and different constitutive activations r_1 and r_2. These two types of receptors are assumed to have the same downstream effect so that only R, the sum of R_1 and R_2, appears in the downstream equations. In our earlier work, we have shown that numerical fits with two types of receptors are more accurate than those with only a single receptor population. We have verified, however, that fits with only one receptor population with a K_d value that falls in between the K_d values for R_1 and R_2 result in good agreement with the observed data.

Equations 3.4 and 3.5 describe the E and I dynamics, respectively. E is the local activator and I is the global inhibitor and they are both activated by receptors occupied by cAMP. For simplicity, we assume that the diffusion of I in the cytosol is sufficiently fast so that its concentration is uniform and has the same value at the front and back of the cell. In addition, we neglect the detailed morphology of the cell and model it as a 10-μm line with the two endpoints representing the front and back, respectively. Thus, the concentration of I is controlled by the average receptor occupancy of front and back, $[R]_{avg}$, along with a small basal activity k_{basal}.

Equations 3.6 and 3.7 describe the memory module, represented by M, along with its coupling to the output of the LEGI module, RE. The equation for the memory module is bistable and has two stable steady states for M that are 0 and M^{tot}. The selected state depends on the value of RE with respect to the two thresholds, a and b. Specifically, when RE is smaller than the lower threshold, a, M is at its low state while when RE is larger than the upper threshold b, M is at its high state. When the value of RE is in between the two thresholds, M is bistable and its state is determined by the history. In addition, the switch between the two stable states is not instantaneous so that a change of the M state requires that RE is below a or above b for a sufficiently long time period. This time period is determined by the parameter k_{Mem} and needs to be small enough to prevent fast switching between the two stable states. The equation for RE, the output of the model, contains three terms. The first two terms describe the Michaelis-

Menten kinetics of the E-regulated activation and I-regulated deactivation, respectively. We set the Michaelis constants, K_{m1} and K_{m2}, to be small to achieve near zero-order ultrasensitivity, making RE sensitive to shallow cAMP gradients. The third term describes the feedback from the memory module to the activation of RE.

3.3.3 Data Fitting

Our model contains 24 parameters, of which we fixed 11 to the values used in our earlier work [26]: $k_{-R1} = 0.16\ s^{-1}$, $k_{R1} = 0.00267\ nM^{-1}s^{-1}$, $k_{-R2} = 1.1\ s^{-1}$, $k_{R2} = 0.00244\ nM^{-1}s^{-1}$, $r_1 = 0.012$ nM, $r_2 = 0.115$ nM, $R_1^{tot} = 0.1$ nM, $R_2^{tot} = 0.9$ nM, $RE^{tot} = 1$ nM and $k_E/k_{-E} = k_I/k_{-I}$. In addition, we fix $M^{tot} = 1$ nM. The remaining 12 parameters were fitted to experimental data using simulated annealing. This method avoids trapping in local minima of the parameter space. It samples a large region of parameter space and does not necessarily reject a trial step when this step cannot improve the fit. Instead, there is an artificial temperature to control the acceptance for the trial steps [84]. We start from a high temperature so that the program is more likely to accept the trial steps, which leads to a larger searching area. As the temperature is decreased during the search, sampling becomes more and more selective. To prevent the memory from becoming trapped in an unstable fixed point, we occasionally perturbed M through the addition of a small constant (0.01 nM).

To fit the RasGTP experimental data, we chose 29 discrete experimental points that best characterized the experimental data set and 19 points determining the stable state of M in different conditions. The latter 19 points are inferred from the experimental results: if in a certain experiment the cells show memory, we assume that M is in its high stable state. The resulting numerical values of the n = 48 fit points, x_i^{sim}, were then compared with the experimental values,

x_i^{exp}, using the following error function:

$$E = \frac{1}{N}\sum_{i=1}^{48}\left(\frac{x_i^{sim} - x_i^{exp}}{\sigma_i^{exp}}\right)^2$$

where σ_i^{exp} are parameters that characterize the uncertainty of the data. For the experiments, they were chosen to be the SDs of the data, whereas for the memory data we used the value 0.1. The fitting algorithm determined the parameter set that minimizes the error function.

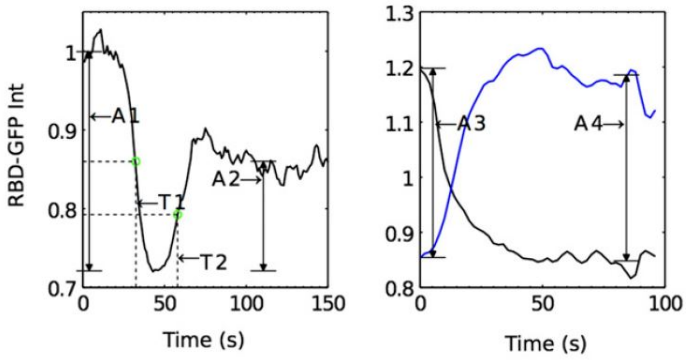


Figure 3.7: Representative example of experimental time traces, along with the amplitudes and time points used in our fitting procedure.

We required our model to fit the uniform dose-response data of our earlier work [26]. Specifically, we chose the peak amplitudes (normalized by the peak amplitude of the nonpretreated cells stimulated by 1000 nM cAMP) and peak times for experiments with cAMP uniform concentration switching from 0 to 0.1, 1, 10, 100, and 1000 nM, from 100 to 250, 500, and 1000 nM. In addition to these 16 data points, we also required that M is not activated to the high state for these uniform steps. The remainder of the fit points were derived from the data in the current study. First, we required that cells in the middle of a 0–100 nM gradient in a L = 70-μm-long gradient chamber, corresponding to a gradient of 1.43 nM/μm, have a front with a high M state and a back with a low M state. Then, for the gradient to uniform experiments, we chose two time points representing how fast the localized RasGTP patches disappeared and reappeared, and one

amplitude point, as illustrated in Figure 3.7. Data points from the experiments switching from 0–100 nM to uniform 20, 10, and 1 nM, quantifying the RBD-GFP intensity of the front half of the cell, normalized by their preswitch value, were included. Specifically, for these experiments we determined the amplitude of the intensity decrease following the switch to uniform concentration (A1 in Figure 3.7). We also determined the difference in fluorescence intensity of this minimum and the level after long-term exposure to a uniform concentration (A2 in Figure 3.7). The amplitude point used in our fitting was the ratio of these two quantities (A1/A2). The two time points used in our fitting correspond to the time it takes to reach the intensity minimum (T1 in Figure 3.7 and corresponding to a quantification of the timescale of patch disappearance) and the time required to reach a steady state (T2 in Figure 3.7 and corresponding to a quantification of the timescale of patch reappearance). Additionally, we required that the front concentration of M be at its high stable state even after the external stimulus concentration has been switched to a uniform level in the simulations from 0–100 nM to 50, 20, 10, and 1 nM, respectively. In contrast, for the case of a complete removal of the chemoattractant we required that M at the front should drop from the high state to the low state after the switch. For the gradient reversal experiments 0–100 to 100–0 and 75–25 nM, we chose the ratio of the amplitudes of RBD-GFP intensity before and after reversal (corresponding to A3/A4 in Figure 3.7). In addition, we required that M reverses in the 0–100 to 100–0 nM experiment. In contrast, in the 0–100 to 60–40 nM experiment, we required that M does not reverse. The resulting fitted parameters are listed in Table 3.1.

3.4 Discussioin of the Model

In this part, we will show the roles that the two modules in the model play respectively in the cellular memory. We first show that without the memory module, the ultrasensitive LEGI module can suppress the response at the back of the cell in the back of the wave and prevent

Table 3.1: Model parameters

Parameter	Value	Parameter	Value
k_E	0.159 s^{-1}	k_I	0.139 s^{-1}
k_{-E}	0.159 s^{-1}	k_{-I}	0.139 s^{-1}
k_{RE}	2,000 s^1	k_{RE}	2,048 s^{-1}
a	0.058 nM	b	0.5 nM
k_{RE2}	116.7 s^{-1}	k_{Mem}	0.360 $nM^{-2} \cdot s^{-1}$
K_{m1}	0.018 nM	K_{m2}	0.001 nM
K_{m3}	0.127 nM	k_{basal}	0.000015 nM
k_{R1}	0.00267 $nM^{-1} \cdot s^{-1}$	k_{R2}	0.00244 $nM^{-1} \cdot s^{-1}$
k_{-R1}	0.16 s^{-1}	k_{-R2}	1.1 s^{-1}
r_1	0.012 nM	r_2	0.115 nM
R_1^{tot}	0.1 nM	R_2^{tot}	0.9 nM
RE^{tot}	1 nM		

cells reversing direction. In Figure 3.8, we examine the kinetics of the ultrasensitive LEGI model during a traveling wave with a small (Left), an intermediate (Center), and a large period (Right). As the wave passes by, the activator level increases and then decreases, roughly following the profile of the local cAMP concentration (black line). The activator in the back experiences the same cAMP profile but delayed by L/ν, where L is the length of the cell and ν is the speed of the wave (green line). Assuming fast cytosolic diffusion, we can take the inhibitor to be uniform. Because experiments have demonstrated that the kinetics of the inhibitor is slower than the kinetics of the activator, the resulting profile is similar to the activator profiles but exhibits a delay with respect to the activator profiles. For large wave speeds, corresponding to small wave periods, the time profiles of the front and back activator, E_f and E_b, respectively, are nearly indistinguishable whereas the inhibitor I is delayed due to its intrinsic kinetics (Left). As a result, the front and back of the cell respond almost identically. For small wave speeds, the delay between E_f and E_b becomes large (Right) and the cell's response is similar to the one due to two sequential and independent gradients: a positive gradient followed by a negative one. For intermediate values of the wave speeds and corresponding periods the inhibitor's delay can be such that its time course coincides with the time course of the activator at the back of the cell (Center). Consequently, the

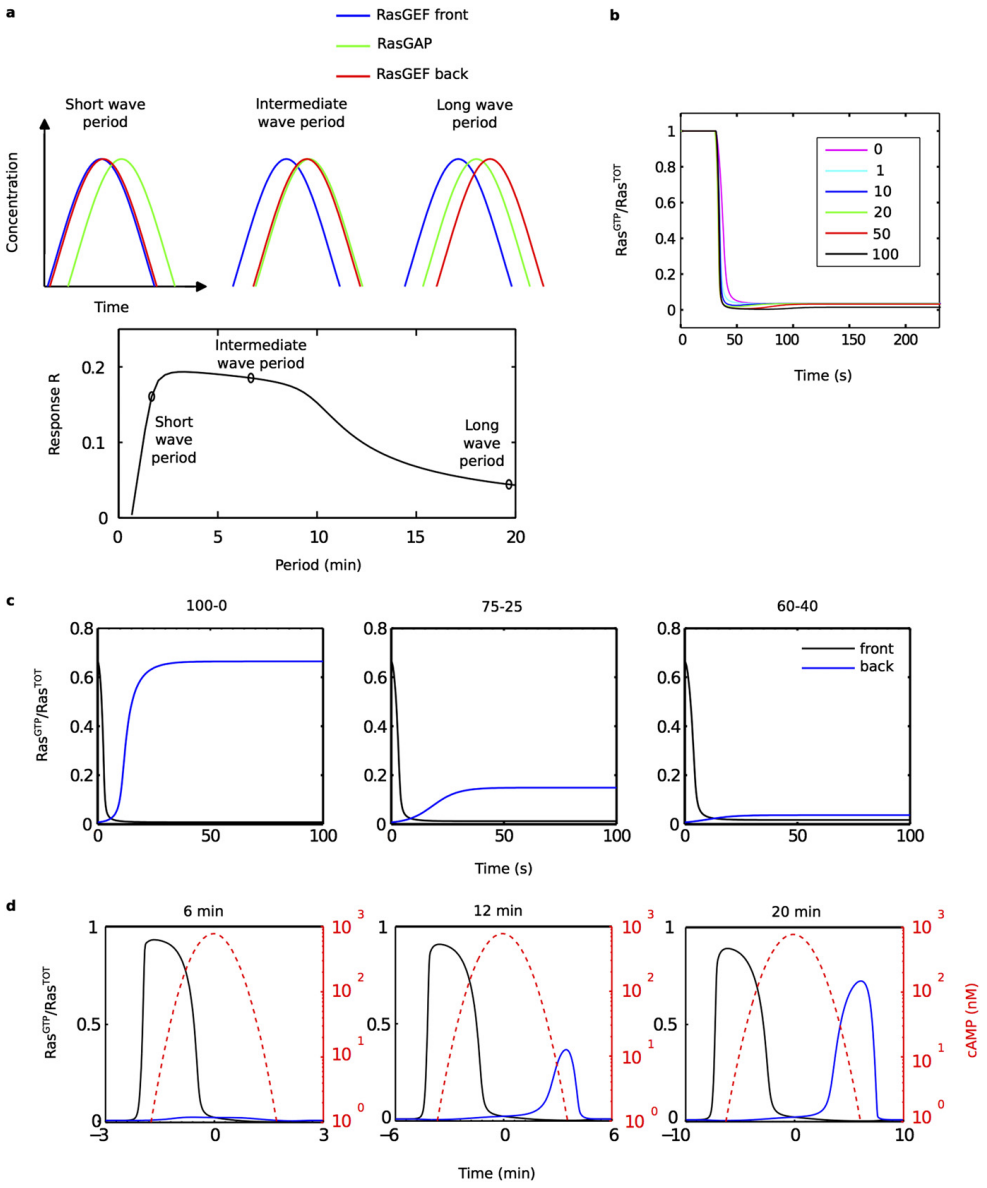


Figure 3.8: Analysis of the ultrasensitive LEGI model. (a) Schematic response of the LEGI model during a traveling wave with a short (Left), an intermediate (Center), and a long period (Right). The lower panel shows the response of the ultrasensitive LEGI model (without the memory module) using the parameter values of Table 3.1 as a function of the wave period. (b) Response of front RasGTP in our model without the memory module for switching from a gradient (0–100 nM) to a uniform concentration (x in nM). (c) Activated Ras at the front (black) and back (blue) in our model without memory, computed after a 0–100 to 100–0 nM reversal (Left), a 0–100 to 75–25 nM reversal (Center), and a 0–100 to 60–40 nM reversal (Right). (d) Response of our model without the memory module to a wave with a period of 6 min (Left), 12 min (Center), and 20 min (Right) showing the mean cAMP concentration (dashed line), and RasGTP at the front (black) and back (blue) of the cell.

response element RE at the front of the cell will be activated in the rising part of the wave while RE at the back remains low during the falling phase by the inhibitor's activity. Thus, for a range of wave speeds and corresponding periods, the dynamics of the ultrasensitive LEGI model is able to suppress the response at the back of the cell. A quantification of this is provided in the lower panel of Figure 3.8a where we have computed the response, R, for our 1D model cell according to:

$$R = \frac{1}{T}\int (RE_f - RE_b)dt$$

In the computation, we have used the ultrasensitive LEGI model without memory and the parameter values of Table 3.1.

However, the ultrasensitive LEGI model without the memory module cannot explain other experimental results. In Figure 3.8b, we show the response of front RE, representing RasGTP, to switching from a gradient (0–100 nM) to a uniform concentration (x). Contrary to the experimental results (Figure 3.2A) and the modeling results with the memory module (Figure 3.4B), the RasGTP response rapidly decreases following the removal of the gradient and does not show any recovery. In Figure 3.8c, we show that without the memory module, the activated-Ras distribution reverses for all cases. Finally, the response of our model without the memory module to a traveling wave is also at odds with experimental results (Figure 3.8d), with no persistent forward movement in the back of the wave, for 6-min wave.

Our model shows that the origin of the memory in the back-of-the-wave problem is the adapative dynamics of the LEGI mechanism, specifically the slow decay of the global inhibitor, and this short-term memory is then strengthed and stablized by a bistable memory module. In other words, what the cell remembers is not necessarily the gradient that comes first, but the gradient with increasing mean concentration. So, we can make a prediction that if we use an inverted wave, in which the cell first experiences the decreasing side of the wave and then the

increasing side, it will keep moving in the gradient direction in the increasing side of the wave. More specifically, as shown in Figure 3.9, in the decreasing side of the first wave, neither the RasGTP at the front of the cell nor at the back is activated and then in the increasing side, the RasGTP at the front of the cell is activated and this activation partly maintains even after the wave pulse passes. This prediction is verified by the experiment in [85].

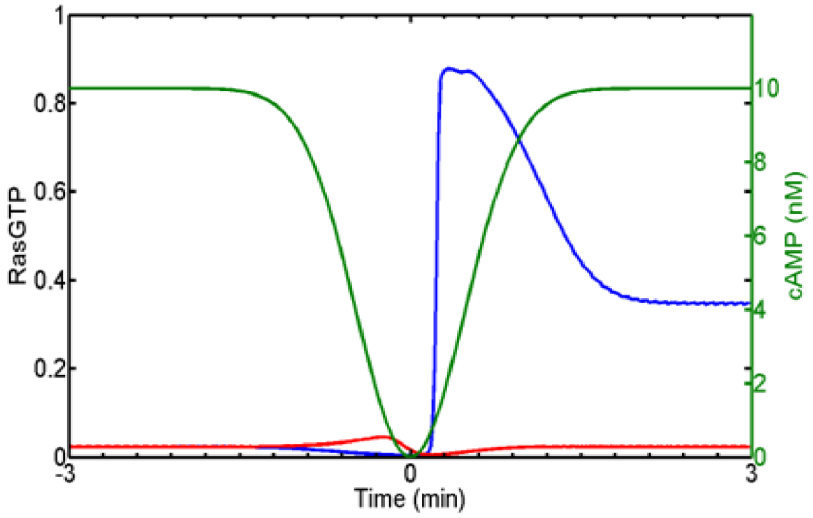


Figure 3.9: Prediction with inverted wave. The green line is the cAMP wave and the RasGTP at the front of the cell is in blue while that at the back is in red.

3.5 Summary

In summary, we have used microfluidics to study chemotaxis in traveling waves of chemoattractant to gain quantitative insight into the classic back-of-the-wave problem—how cells avoid reversing direction in the back of the wave, where the spatial gradient reverses. By varying the wave period, we have shown that for natural wave periods, cells not only avoid reversing direction in the back of the wave, but in fact continue movement toward the source for $\sim$2 min, indicative of cellular memory (Figure 3.1). However, for longer wave periods, cells increasingly reverse direction in the back of the wave, consistent with the high sensitivity of cells to static

spatial gradients. The latter finding shows that cells do not simply lock in the direction of the spatial gradient in the front of the wave and that cells are still capable of directional sensing even when the concentration is decreasing.

Our results suggest that the solution to the back-of-the-wave problem consists of two parts. First, the ability of cells to not reverse in the back of the wave can be explained by the adaptive dynamics of the LEGI mechanism, specifically the slow decay of the global inhibitor, which provides a short-term memory. For short wave periods, and correspondingly large wave speeds, the fast temporal decrease in concentration causes the inhibitor to stay elevated during the back of the wave and suppress directional sensing, whereas for long wave periods this inhibitor has sufficient time to decay (Section 3.4 and Figure 3.8). Second, the continued directed movement in the back of the wave can be explained by a bistable memory mechanism. We have provided evidence that such a long-term cellular memory is present at the level of Ras activation, even when actin polymerization is inhibited (Figure 3.2) and only in sufficiently developed, polarized cells (Figure 3.3).

Together, the proposed coupling between the LEGI directional sensing module and the bistable memory module enables cells to maintain direction when the local concentration decreases rapidly, while keeping them sensitive to slow gradient reversals. Future work will be required to further validate this mechanism, including biochemical identification of the memory module and its coupling to intracellular oscillators [86, 87]. Nonetheless, the essential elements and general framework of adaptive short-term and long-term cellular memory may be relevant to the directed migration of other types of chemotactic cells in dynamic gradients.

3.6 Acknowledgements

We thank D. Fuller for technical assistance and A. Kortholt and P. J. van Haastert for providing the RBD-GFP plasmid. We also thank dictyBase for providing the PH domain plasmid.

This work was supported by National Institutes of Health Grant P01 GM078586.

Chapter 3, in part, is a reprint of the material in part as it appears in “Cellular memory in eukaryotic chemotaxis. Skoge, Monica; Yue, Haicen; Erickstad, Michael; Bae, Albert; Levine, Herbert; Groisman,Alex; Loomis, William F.; Rappel, Wouter-Jan, Proceedings of the National Academy of Sciences, 2014”. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Minimal network topologies for signal processing during collective cell chemotaxis

4.1 Introduction

Collective cell migration plays an important role in many biological processes, including development, wound healing, and cancer metastasis and has been the focus of much recent experimental and theoretical work [88–94]. Interestingly, collective migration is not merely the result of many independent cells moving around but can exhibit unique behavior. For example, measurements of several different cell types have shown that cells cooperate to sense a gradient such that cell clusters follow gradients while single cells are unable to detect this gradient [95, 96]. Furthermore, expanding cell monolayers often reveal the spontaneous formation of finger-like instabilities with specialized leader cells at their tips [97, 98]. These results point towards the

important role that cell-cell interactions can play in collective migration. Here, we are particularly interested in collective chemotaxis: how clusters of cells work together to follow a chemical gradient, as measured in malignant lymphocytes [95], neural crest migration in developing embryos [96], and border cell migration in Drosophila [1, 99, 100].

Collective chemotaxis problems have been modeled by many groups, with widely varying assumptions for cell-cell interactions. These include a focus on intercellular forces such as cell-cell repulsion and adhesion [101], mechanical communication via contact inhibition of locomotion (CIL) [102–105], or biochemical communication through a diffusible inhibitor [106, 107]. In particular, several earlier theoretical studies show that, depending on the biochemical details of how the cluster processes the signal, cluster speeds and directionalities can be profoundly different [106–109]. Current experiments have not yet provided a way to clearly determine these signal processing mechanisms, and most papers have assumed only a minimal multicellular local-excitation-global-inhibition (LEGI) signal processing mechanism[106, 107, 109].

In this paper, we develop a method to use existing experimental data to constrain possible multicellular signaling networks. Our approach is a qualitative perturbation analysis, in which we use mathematical analysis to determine how experimental interventions such as photo-activation of one element of the signaling network will affect measured outcomes. These behaviors will often only depend on the topology of the signaling network—which elements are connected to which—and not the biochemical details. We use this technique to determine potential minimal network topologies for border cell cluster migration in the Drosophila egg chamber. We can show that existing experimental data rules out all networks with five or fewer interactions (including the aforementioned LEGI model) and we can deduce that only six possible six-link topologies are able to explain the experimental data. In this process, we introduce two methods to circumvent the complexity caused by a stochastic, spatially extended problem like border cell chemotaxis. We first use analytical methods for the selection of possible topologies to avoid exhaustively simulating all topologies. Then, for selected topologies,we carry out quantitative simulations

in which we solve non-linear stochastic differential equations using a master equation approach instead of direct numerical integration. We verify through explicit numerical sampling that these networks are fully consistent with the data and carry out a parameter fit for the six topologies. Finally, we propose additional experiments to further narrow down the possible options.

4.2 Border cell migration: available experimental data

We apply our method to a characteristic example of collective chemotaxis: the guided movement of border cell clusters during the embryonic stage of Drosophila [110–112]. A border cell cluster consists of two polar cells and 4–8 border cells (Figure 4.1a) and migrates during development between the nurse cells towards the oocyte as an interconnected group. During this migration only the border cells show protrusions while the polar cells do not protrude and remain at the center of the cluster. The directional migration of the cluster is essential for Drosophila's oogenesis and is guided by chemoattractant gradients [113, 114]. Using genetic manipulations of the signal receptors and photo-activatable Rac, a Rho GTPase known to be pivotal in the control and formation of protrusions during cell migration [43], experiments show that the change of signaling state in one cell of the cluster can guide the movement of the whole cluster [115]. Light-mediated activation and inhibition of Rac in a single cell can not only change the direction of the cluster, but can also inhibit or promote the protrusion of other cells in the cluster [1].

We use experimental data obtained from [1], which is summarized in Table 4.1. This data quantifies the number of border cells protruding and the directionality index of the cluster, which can be positive, corresponding to more protruding cells at the front of the cluster, or negative, corresponding to more protruding cells which are located away from the highest chemoattractant concentration [1]. (A detailed definition of the directionality index will be given later.) In wild-type clusters, the average number of cells containing a protrusion is about two while the directionality index is large and positive. The authors of [1] also obtained additional data using

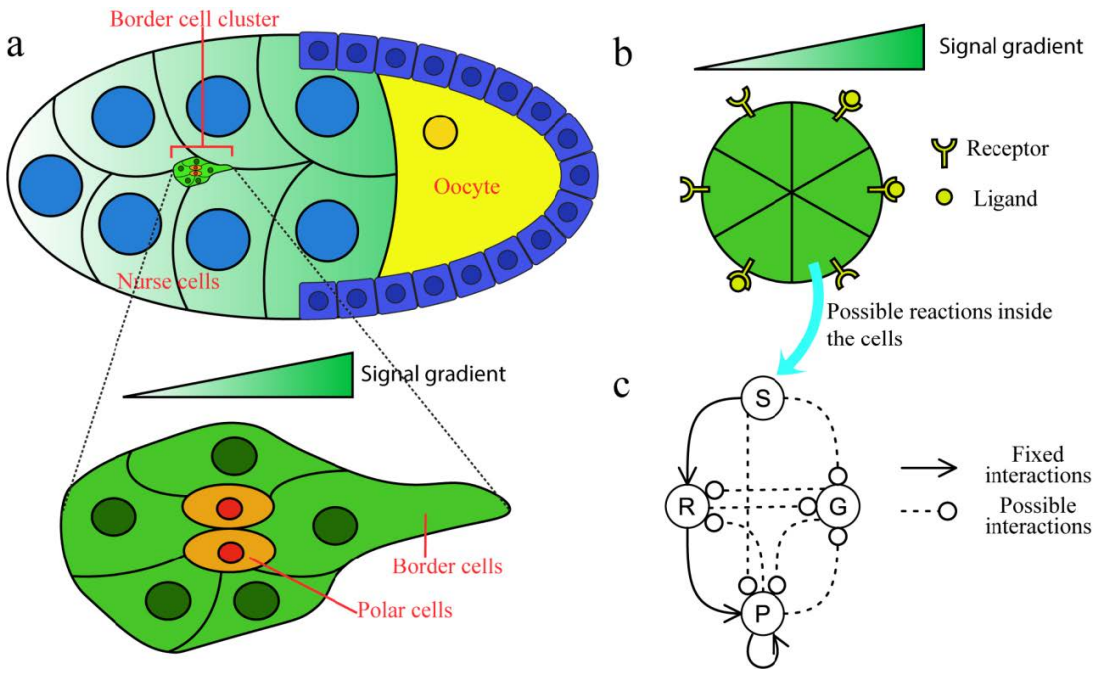


Figure 4.1: Border cell cluster in Drosophila's egg chamber and model simplification. (a) Schematic drawing of a Drosophila's egg chamber and a border cell cluster composed of border cells and polar cells. The cluster moves towards the oocyte guided by chemical signals. (b) Our simplified model of a border cell cluster, consisting of six non-deformable cells arranged in a rigid array. (c) Basic structure of the signal processing network consisting of four nodes, three fixed, positive interactions (solid arrows), and seven possible interactions (dashed arrows with round heads meaning the sign of the interactions is undetermined).

transgenic flies in which Rac can be selectively and locally activated or inhibited using light. Photo-inhibition of Rac of a border cell at the front of the cluster, for example, results in a larger number of average protrusions in the non-inhibited cells and a directionality close to zero. In addition, the number of protrusions was quantified in cells expressing dominant negative forms of the chemoattractant receptors, denoted here by Receptor^{DN}. In these cells, ligands can still bind to receptors but fail to deliver a downstream signal. Clusters expressing Receptor^{DN} were found to have more protrusions on average than wildtype clusters but with roughly zero directionality index, as would be expected in the absence of any directional signal.

Table 4.1: Summary of experimental data extracted from figures in [1]

Experiment	# of protruding cells	Directionality index	Comments
Wild type (WT)	2.0 ± 0.2	0.435 ± 0.035	
Light-inhibition of WT	4.5 ± 0.6	-0.025 ± 0.095	
Receptor^{DN}	4.15 ± 0.35	0.015 ± 0.085	Cells expressing the dominant negative form of the chemoattractant receptors.
Light-activation of WT	2.0 ± 0.3	-0.405 ± 0.055	Photo-activation of Rac in the cell at the back of the wild type cluster
Light-activation of Receptor^{DN}	2.1 ± 0.4	0.245 ± 0.045	Photo-activation of Rac in one cell of the Receptor^{DN} cluster

4.3 Model

To construct our model, as the experimental measurements we study are the protrusions in the cluster measured over a relatively short time, we will not consider cell deformation, cluster motion, or cluster rearrangement. Furthermore, since polar cells do not protrude and remain at the center of the cluster, we will only consider border cells in our simulations. We fix the number of these cells to six but note that the analytical methods do not rely on the number of cells in a cluster and we have verified that the numerical simulation results remain similar for clusters with more or fewer cells. The resulting model is schematically shown in Figure 4.1b and consists of a rigid cluster with cells making an angle of $\vartheta_i = \frac{\pi}{3}i$, i=1,..,6 with the gradient direction. Each border cell i contains an identical pathway and our goal is to determine a minimal set of components or nodes required to capture experimental findings. Since the cell is responding to an external chemoattractant in the form of protrusions and since the activation and inhibition of Rac affect the cell's response we assume that the pathway has as local components the input chemoattractant signal S_i, the protrusion level P_i, and the Rac activation level R_i. In addition,

we introduce a fourth, global component, G, representing cell-cell communication. We do not specify the nature of this communication, but note that it can be in the form of a small molecule as suggested by experiments on branching morphogenesis [106] or in the form of mechanical interactions [116, 117]. Since in either case communication between cells is fast compared to protrusion dynamics, which occur on timescales of 20–30 minutes [1, 118], we will assume that all cells in the cluster share the same value of G. Note that this assumption is valid as long as the degradation of G is not too fast [107, 108]. The output of the model, and the quantities that can be compared to experimental data, include the total number of cells with protrusions, represented by NP, and the direction of the protrusions, quantified by a directionality index specified below.

We will consider three possible interactions or links between each node of the network, including self-interactions: 1) positive, for which the output of one node increases the activity of the other node or of itself, 2) negative, corresponding to a repression of the activity, and 3) no interaction. As a result, we have to consider 316 possible topologies for the model network. One potential technique to determine if these networks are consistent with data would be to follow the approach of [119], who exhaustively searched three-node networks, sampling a broad range of parameters for each, to find those with perfect adaptation to a changing signal. However, this is not computationally feasible given the more than 40,000,000 possibilities at hand. Such an approach is further complicated by the fact that our model describes a stochastic, spatially extended system. Thus, although we describe the pathway as a four-node network, in which one node is (fixed) chemotactic input, the equations for the entire cluster contain 13 variables and 13 equations (2 local equations for each cell and 1 global equation). Instead of the sampling approach of [119], we will pursue a qualitative perturbation analysis method.

4.4 Qualitative selection approach

We will now reduce the number of possible topologies based on experimental data and without specifying specific functional forms of the interactions. In this qualitative selection procedure, we do not take into account the precise number of protrusions following a specific experimental perturbation. Instead, we only take into account if the perturbation increases or decreases the number of protrusions, along with the location of the protrusions. The flow of this procedure is schematically shown in Figure 4.2a. As we will see below, we only have to consider topologies with fewer than seven links. Therefore, we start with all possible network configurations with five-link (amounting to 139,776) and six-link (totaling 512,512) topologies.

Before addressing the detailed results in Wang et al. we can already rule out a large amount of these topologies. First, we note the general result that Rac is known to be activated by an external signal and activates downstream effectors of cell protrusion [43]. Consequently, positive interactions from S to R and from R to P are required. Second, experiments have shown that positive feedback loops that are downstream of activated Rac help maintain the protrusion [43]. These loops will be captured by a positive self-activation link for P. In addition, as there is no experimental evidence for the amplification of R or G, and as we are trying to find the minimal topologies, we assume that there is no self-activation or self-inhibition for these nodes. Furthermore, since S represents the external signal, we assume that there is no feedback to S. This reduction results in the networks are shown in Figure 4.1c, where the three fixed, positive interactions are shown as black solid arrows and where the seven remaining possible interactions are shown using black dotted arrows with round heads meaning the sign of the interactions is undetermined. Finally, we note that the experiments of Wang et al. show that local photo-activation and inhibition affect the protrusion number of the entire cluster. This automatically implies that there must be cell-cell communication. In our model, this communication is achieved through the global factor G and we thus need at least two more links in addition to the three

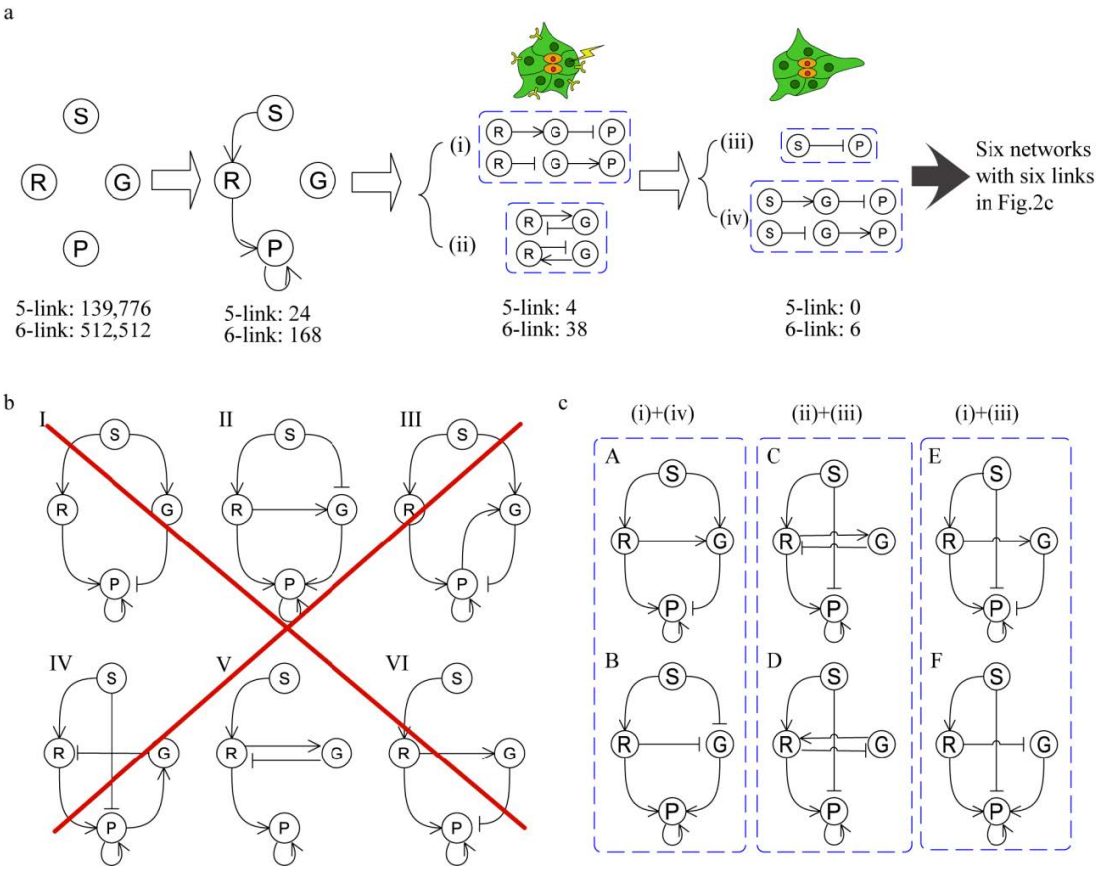


Figure 4.2: Network selection procedure. (a) Schematic flow of the qualitative selection procedure. In the first step, the number of candidates is reduced by fixing three links based on other experimental results and by the requirement of at least one input to G and one output from G. In the second and third steps, photo-inhibition of Rac and Receptor^{DN} experiments are used respectively to further reduce the number of candidates. (b) Examples of networks that are excluded using our qualitative selection procedure. Networks I–IV are excluded based on light-inhibition of Rac while the exclusion of V and VI is based on Receptor dominant-negative mutant experiments. Positive interactions are shown as arrows and negative interactions as bars. For further details, see text. (c) Final results of the qualitative selection procedure, consisting of 6 networks, each with 6 links.

shown in Figure 4.1c: one input to G and one output from G. Taking into account the restrictions mentioned above we find that we have 24 possible five-link topologies and 168 possible six-link topologies, schematically indicated by the second step in Figure 4.2a. Of course, there will be even more topologies with a larger number of links. However, as we will see below, several six-link topologies are consistent with the experimental data and, since we are searching for minimally complex networks, we will not consider topologies with more than six links.

4.4.1 Photo-inhibition experiment requires one of four possible interaction motifs

Photo-inhibition of Rac in the front cell of a wild-type cluster decreases the probability of that cell protruding, but increases the protrusion probability of the other cells in the cluster [1]. This information creates two requirements for the network. First, because inhibiting Rac in one cell affects a separate cell's protrusions, there should be an interaction from one cell's R to the other cell's P through the global factor G. This gives us four possible motifs within the network ($R \rightarrow G \rightarrow P$, $R \rightarrow G \rightarrow R \rightarrow P$, $R \rightarrow P \rightarrow G \rightarrow P$, and $R \rightarrow P \rightarrow G \rightarrow R \rightarrow P$), and topologies lacking these motifs can be excluded. An example of such an excluded network is shown in Figure 4.2b, I. Second, since inhibition corresponds to a reduction of R, the interaction between R and P through G should be negative. Thus, the product of the signs of the links in these four possible motifs should be negative. (We will not consider the possibility of having more than one motif, as this would be redundant, and we are trying to find the minimum possible networks.) For instance, in the $R \rightarrow G \rightarrow P$ motif, if both links are positive, the net effect from one cell's R to another cell's P will be positive which contradicts experiments. Network II in Figure 4.2b is an example of a network that can be excluded using this logic. The other three motifs contain more than two links, including a direct link between R and P. Since this $R \rightarrow P$ interaction is required by other experimental results to be positive [43, 120], there are also only two possibilities of sign assignments. For example, for the pathway $R \rightarrow G \rightarrow R \rightarrow P$, the sign assignment can only be either $R \xrightarrow{+} G \xrightarrow{-} R \xrightarrow{+} P$ or $R \xrightarrow{-} G \xrightarrow{+} R \xrightarrow{+} P$.

After further analysis, we can rule out both the (a) $R \rightarrow P \rightarrow G \rightarrow P$ and the (b) $R \rightarrow P \rightarrow G \rightarrow R \rightarrow P$ motifs entirely. Two examples including these two excluded motifs are shown in Figure 4.2b as III and IV. In these motifs, communication from one cell to another occurs through G, thus using protrusion as an intermediate step. The photoinhibition experiments show that Rac inhibition in one cell leads to a significant increase of protrusions in other cells (see Table 4.1).

Within these motifs, this implies that a protrusion decrease in one cell (P_i) promotes protrusions in other cells (P_j). As the signal S does not change in the light-inhibition experiments, the following logic holds regardless of how S is connected to the other components of the network. So, in the following analysis, when we rule out a motif, we mean that all the networks containing this motif (without redundant motifs) are ruled out. And as $R \rightarrow P$ is fixed to be a positive effect by experiments, we only need to consider the signs of the following two motifs: (a) $P \rightarrow G \rightarrow P$ and (b) $P \rightarrow G \rightarrow R$.

As is mentioned in the first paragraph of this section, the phenomenon that inhibiting R_i leads to a decrease in P_j has ruled out the possibility when the effect of motif (a) is positive. On the other hand, if the effect of motif (a) is negative (as in Figure 4.2b, III), we take $P \xrightarrow{+} G \xrightarrow{-} P$ as an example since the logic is similar if we switch the signs. After light-inhibition, experiments show that more cells display protrusions. This results in an increase in the global factor G since the sign of the link between P to G is positive. However, since G to P is negative, this will lead to lower P_j's which contradicts the experimental results. So, for motif (a), neither positive effect nor negative effect match the experiments and thus motif (a) is excluded.We can follow a similar line of logic to exclude motif (b). For this motif, positive effect is also ruled out in the first paragraph of this section by the phenomenon that inhibiting R_i promotes P_j. Consider the case where the effect of motif (b) is negative; we take $P \xrightarrow{+} G \xrightarrow{-} R$ (as in Figure 4.2b, IV) as an example because switching the signs does not influence the logic. As NP increases after light inhibition, G increases and thus R_j decreases. R_j's decrease, however, will lead to P_j's decrease which contradicts the experiments. So, motif (b) is also excluded.Therefore, only two of the four possible motifs remain. These motifs, when taking into account the different sign assignments, result in four possibilities listed in (i) and (ii) in Figure 4.2a. If combined with the fixed links mentioned earlier, they generate four five-link and 38 six-link networks.

4.4.2 Receptor^{DN} experiment requires one of three possible interaction motifs

Wang et al. found that clusters expressing Receptor^{DN} have more protrusions on average than wildtype clusters [1]. We model Receptor^{DN} by taking the signal to be zero—meaning all cells will see identical environments. The comparison to wildtype protrusion numbers will be simpler if we similarly assume wildtype clusters see a constant but nonzero signal. This will be reasonable in predicting the number of protrusions if the gradient is relatively small. This assumption is supported by measurements of the chemoattractant PVF1 along the border cell cluster's migration path which do not show a significant gradient [99]. Furthermore, as we will show below, numerical simulations which take the gradient into account, give consistent results.

In the absence of a gradient, the problem is reduced into a homogeneous one, and the average number of protrusions is simply given by $\langle \mathrm{NP} \rangle$ = (number of cells) $\times$ Prob where Prob is the probability of a protrusion in any cell. The result of the Receptor^{DN} experiment (protrusion number increases when [S] decreases) can be translated into an inequality: $\frac{\mathrm{dProb}}{\mathrm{d[S]}} < 0$. Using the total derivative's chain rule, we can express $\frac{\mathrm{dProb}}{\mathrm{d[S]}}$ in terms of partial derivatives and total derivatives which are experimentally known (see Section B.1 for detailed derivation):

$$\left(1 - \frac{\partial \mathrm{Prob}}{\partial [\mathrm{G}]} \frac{\partial [\mathrm{G}]}{\partial \mathrm{Prob}}\right) \frac{\mathrm{dProb}}{\mathrm{d[S]}} = \frac{\partial \mathrm{Prob}}{\partial [\mathrm{S}]} + \frac{\partial \mathrm{Prob}}{\partial [\mathrm{G}]} \frac{\partial [\mathrm{G}]}{\partial [\mathrm{S}]} + \left(1 - \frac{\partial \mathrm{Prob}}{\partial [\mathrm{G}]} \frac{\partial [\mathrm{G}]}{\partial \mathrm{Prob}}\right) \frac{\mathrm{dProb}}{\mathrm{d[R]}} \frac{\mathrm{d[R]}}{\mathrm{d[S]}} \tag{4.1}$$

Experiments indicate that both total derivatives appearing in the last term are positive: FRET measurements show that Rac activity increases after adding EGF [1], implying $\frac{\mathrm{d[R]}}{\mathrm{d[S]}} > 0$, while experiments using photoactivatable Rac show that Rac activation induces cell protrusion [121], implying $\frac{\mathrm{dProb}}{\mathrm{d[R]}} > 0$. In order to get $\frac{\mathrm{dProb}}{\mathrm{d[S]}} < 0$ in equation (4.1), there are two options:

$$\text{a)} \quad 1 - \frac{\partial \mathrm{Prob}}{\partial [\mathrm{G}]} \frac{\partial [\mathrm{G}]}{\partial \mathrm{Prob}} < 0 \quad \text{and} \quad \frac{\partial \mathrm{Prob}}{\partial [\mathrm{S}]} + \frac{\partial \mathrm{Prob}}{\partial [\mathrm{G}]} \frac{\partial [\mathrm{G}]}{\partial [\mathrm{S}]} > 0 \quad \text{or}$$

$$\text{b)} \quad 1-\frac{\partial \text{Prob}}{\partial [\text{G}]}\frac{\partial [\text{G}]}{\partial \text{Prob}} > 0 \quad \text{and} \quad \frac{\partial \text{Prob}}{\partial [\text{S}]}+\frac{\partial \text{Prob}}{\partial [\text{G}]}\frac{\partial [\text{G}]}{\partial [\text{S}]} < 0.$$

Option a) requires at least three links, namely i) G$\rightarrow$P, P$\rightarrow$G and S$\rightarrow$P or ii) G$\rightarrow$P, P$\rightarrow$G and S$\rightarrow$G as $\frac{\partial \text{Prob}}{\partial [\text{G}]}\frac{\partial [\text{G}]}{\partial \text{Prob}}$ should be non-zero. Combining the links in these two possibilities with the five links already determined above, and taking into account possible redundant links, leads to networks that have either 7 or 8 links. Thus, these networks will not be further considered. Option b) requires $\frac{\partial \text{Prob}}{\partial [\text{S}]} < 0$ or $\frac{\partial \text{Prob}}{\partial [\text{G}]}\frac{\partial [\text{G}]}{\partial [\text{S}]} < 0$ which can be accomplished with either one or two additional links shown in Figure 4.2a (iii) and (iv). Two examples of excluded networks based on the Receptor$^{\text{DN}}$ experiments are shown in Figure 4.2b as V and VI.

4.4.3 Final construction of possible topologies

We can now assemble the simplest possible networks by choosing one possible motif mandated by the photo-inhibition experiment and one possible motif mandated by the Receptor$^{\text{DN}}$ experiment. Specifically, combining motifs from group (i) and motifs from group (iv) of Figure 4.2a results in the two networks A and B shown in the first group of Figure 4.2c. Similarly, combining motifs from (ii) and (iii) leads to networks C and D in Figure 4.2c and combining motifs from group (i) and (iii) produces networks E and F. All of these candidate networks have six links: our qualitative selection procedure has ruled out any five-link networks. Choosing motifs from group (ii) and group (iv) will result in networks with seven links, which we will ignore, as we have found that six-link networks are sufficient to meet our requirements. Finally, its worth emphasizing again that the above analysis did not specify functional forms of the interactions in the networks; the results will hold regardless of the quantitative details of the interactions.

4.5 Parameter sampling results are consistent with qualitative selection analysis

In the above analysis of the Receptor^{DN} experiment, we made one key approximation: we assumed that the gradient is shallow enough that protrusion numbers are independent of the gradient. To verify that this approximation does not influence our results and to study the selected networks in a quantitative way, we cast our candidate networks into ordinary and stochastic differential equations. Specifically, we have one deterministic equation for R in each cell and a single global deterministic equation for G. These equations are listed in Section 4.10. The equation for the protrusion in cell i, $[P_i]$, is cast as a stochastic differential equation (SDE) of the form:

$$\frac{d[P_i]}{dt} = F([G], [P_i], [R_i], [S_i]) + \eta(t)$$

where F represents a non-linear function and where η is a Gaussian noise term with zero mean and standard deviation σ: $\langle \eta(t_1)\eta(t_2) \rangle = 2\sigma^2\delta(t_1 - t_2)$. We have chosen F to be such that P is bistable and can be either in a low state, $[P]_a$, corresponding to a non-protruding state, or in a high state, $[P]_c$, representing the protruding state (Details in Section 4.10). Using a bistable equation for P results in a clear distinction between cells that are protruding (in the high state) and those that are not (low state). To solve the resulting set of 13 coupled equations, we assume that all seven deterministic equations reach equilibrium quickly compared to the time scale of protruding (20–30 min) [1, 118]. The resulting stable states for R and G can then be substituted into the six SDEs for $[P_i]$.

To solve the SDEs one could in principle use a SDE solver to determine the steady state of the cluster and thus the average protrusion number. However, this would require capturing a large number of transitions between the high and low states. These transitions could become

prohibitively slow, especially if the barrier in the bistable protrusion state is large compared to the strength of fluctuations, requiring an excessive amount of computational resources. Therefore, rather than directly solving the SDEs, we treat the problem as a transition between the low state $[P]_a$ and high state $[P]_c$ and compute transition rates between these two states. This is possible since the equation for P can be written as:

$$\frac{d[P]}{dt} = -U'([P]) + \eta(t)$$

which is precisely the equation of an over-damped particle in a double-well potential U([P]). Thus, we can calculate the transition rates between states $[P]_a$ and $[P]_c$ using Kramers approximation as:

$$k_{on} = \frac{\sqrt{-U''([P]_a)U''([P]_b)}}{2\pi} e^{-\frac{U([P]_b)-U([P]_a)}{\sigma^2}}$$
$$k_{off} = \frac{\sqrt{-U''([P]_c)U''([P]_b)}}{2\pi} e^{-\frac{U([P]_b)-U([P]_c)}{\sigma^2}}$$

Here, $[P]_b$ is the maximum of the potential U([P]) between $[P]_a$ and $[P]_c$ (see also Figure B.1). Note that the Kramers approximation needs the prerequisites that $\Delta U \gg \sigma^2$, where $\Delta U \equiv U([P]_b) - U([P]_{a \text{ or } c})$. When this prerequisite does not hold, we use a more accurate method based on the mean first passage time in a one-dimensional potential to calculate the transition rates (details are in Section B.2). Because the transition rates depend on the ratio of ΔU and σ^2, rather than the specific value of σ^2, and the value of ΔU is controlled by other parameters, we can fix $\sigma = 0.1$ in our simulation without losing generality.

Given the transition rates between two states, we can compute the average number of protrusions under many circumstances. Each cell can be either protruding or not; there are then $2^6 = 64$ possible states for a cluster with six cells; each state α (with α=1,2,,64) then has a total number of protrusions, NP_α. Associated with each state is a probability W which can be computed

using a master equation:

$$\frac{dW_\alpha}{dt} = \sum_{\beta=1}^{64} q_{\alpha\beta} W_\beta$$

Here, q is a matrix with $q_{\alpha\beta}(\alpha \neq \beta)$ describing the transition rate from state β to state α while $q_{\alpha\alpha}$ representing the transition rate out of state α, which can be defined as: $q_{\alpha\alpha} = -\sum_{\beta=1,\beta\neq\alpha}^{64} q_{\beta\alpha}$. The transition rates are given by the escape rates k_{on} and k_{off} shown above. We will only consider transitions where one cell changes its state. In the long-time limit, the probabilities tend to the stationary solution, that is $\frac{dW_\alpha}{dt} = 0$, resulting in a set of 64 coupled linear equations which can be easily solved (37). Once these probabilities are found, it is straightforward to find the average number of protrusion, computed as $\langle NP \rangle = \sum_\alpha NP_\alpha W_\alpha$, the number of protrusions in the direction of the gradient, or other relevant experimental measurements.

Table 4.2: Range of Model Parameters used in Sampling Simulations in Log10 Space

Name	Range	Comments	Name	Range	Comments
S_{grad}	(-3,0)		S_{mean}	(-2,1)	
$basal_{AR}$	(-2,1)		k_1	(-2,2)	Parmeters corresponding to the links except the three fixed ones. k_3 is only used in 6-link networks.
P^{tot}	(-2,2)		k_2	(-2,2)	
k_{RP}	(-2,2)		k_3	(-2,2)	
PI	(-2,2)	photo-inhibition strength	PA		photo-activation strength (only used in fitting, not sampling tests).

To reduce the computational expense, we study all 24 networks with 5 links resulting from our initial qualitative selection (shown in the second step of Figure 4.2a), but limit ourselves to a subset of 12 6-link networks, out of 168 possible networks, including the ones shown in Figure 4.2c. In our simulations, we use Latin hypercube sampling to generate 500,000 sets of parameters; because parameters correspond to positive rates, we sample uniformly over the log of the parameters. The parameters and their sampling ranges are listed in Table 4.2. Then, we apply

the same perturbations corresponding to the Receptor^{DN} and photo-inhibition experiments and compare the pre- and post-perturbation protrusion numbers.

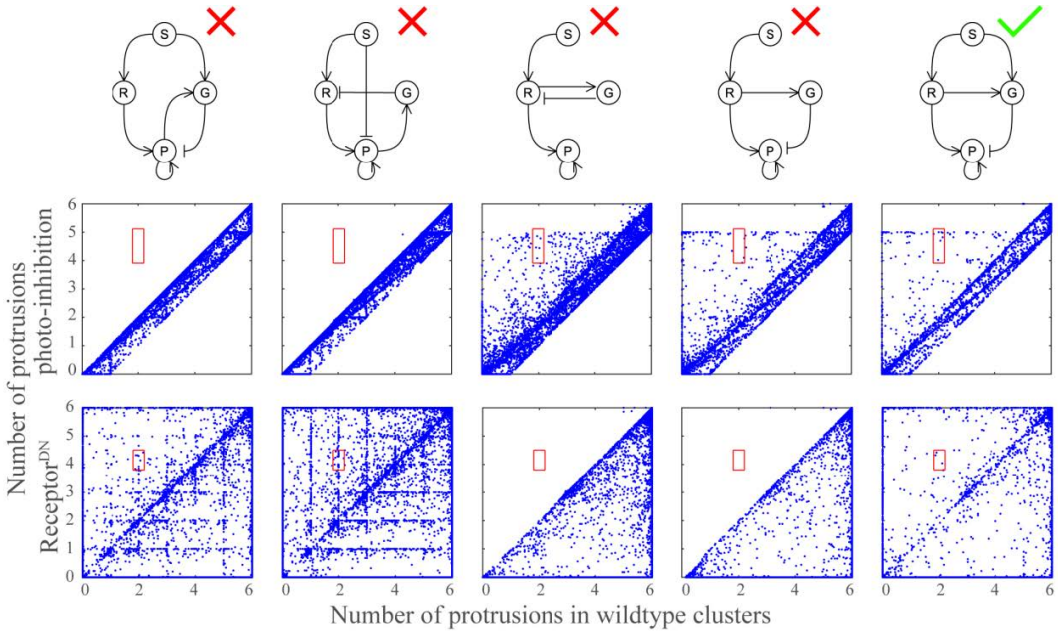


Figure 4.3: Parameter sampling results of five networks. The x-axes show the number of protrusions for wildtype clusters and the y-axes show the number of protrusions for photo-inhibition case (middle row) and Receptor^{DN} case (lower row). The red boxes show the region quantitatively consistent with the experimental data while a dot represents the output of one set of parameters.

In Figure 4.3 we show the sampling results for the five networks illustrated in the upper row. The first four (two five-link and two six-link topologies) were ruled out by our selection procedure while the right-most network is not excluded by the qualitative selection. The results of the numerical sampling are shown in the middle and lower row where each dot corresponds to one set of parameters. The x-axis represents the number of protrusions in wild-type clusters while the y-axis represents the protrusion number for the lead cell photo-inhibition (middle row) and the Receptor^{DN} experiments (lower row). The red box shows the experimentally observed protrusion number with the center as the mean and the length and width as the observed standard deviations. The existence of dots within the red box means that the network is consistent with the experimental data. We see that the first four networks fail to meet the requirements for one

of the two experiments while the last network shows parameter combinations that can satisfy both.We emphasize that in Figure 4.3, and in the networks we have sampled from explicitly, we find results that are consistent with our qualitative models. Models which fail the Receptor^{DN} test qualitatively fail it in sampling while models that can be ruled out by the qualitative selection procedure based on the photo-inhibition experiments are also inconsistent with these experiments when analyzing the sampling results. Its worth noting that the points within the red box in the middle row may not be the same sets of parameters with those in the lower row: having points in both red boxes is necessary, but not sufficient to be consistent with experiment. Below, we will fit the parameters for the six selected networks to make sure that we can find at least one set of parameters for each network that is consistent with the experimental results listed in Table 4.1. More sampling results can be found in Appendix B, Figures B.2, B.3 and B.4.

4.6 Further discrimination between the six minimal networks with six interactions

After finding the six minimal networks with six interactions, the remaining question is how to discriminate between these six networks or these three groups. In addition to the protrusion statistics we have already used, Wang et al. also defined and measured a directionality index, defined as $D_{exp} = \frac{\sum_i \vec{P}_i \cdot \vec{d}}{\sum_i \|\vec{P}_i\|}$, where $\vec{d}$ is unit vector in the signal gradient direction and $\vec{P}_i$ is the protrusion vector for cell i, determined from experimental images. To compare this to our model results, we take the direction of the protrusion in the model to be normal to the edge of the round cluster and of length 1 for the protruding state and 0 for the non-protruding state. Thus, we can define a directionality index for the model as $D_{mod} = \frac{\sum_i^6 P_i \cos\vartheta_i}{\sum_i^6 P_i}$ which can range from +1 (only the front cell shows a protrusion) to –1 (only the back cell protrudes). Among the six possible networks in Figure 4.2c, networks C, D, E, and F have a negative $S \rightarrow P$ interaction while networks A and B do not. This negative $S \rightarrow P$ interaction is determined by the Receptor^{DN}

experiments requiring $\langle \mathrm{NP}_{\mathrm{Rec}^{\mathrm{DN}}} \rangle - \langle \mathrm{NP}_{\mathrm{wild}} \rangle > 0$. However, if the chemoattractant suppresses protrusions, it would tend to make the cluster move against the signal gradient, contradicting the experiments. This is not a critical problem as there is also a positive effect $\mathrm{S} \to \mathrm{R} \to \mathrm{P}$ to balance the negative effect. But for networks C, D, E, and F, positive chemotaxis of the cluster is not a robust feature but one requiring fine-tuning of the parameters. In contrast, for networks A and B, the negative effect from S to P is through the global factor G which has the same value for all the cells—leading to robust chemotaxis. This difference is shown in Figure 4.4a by plotting the results of 500,000 samples with the y-axis as the directionality for the wildtype cluster and the x-axis as the difference between the average protruding number for Receptor$^{\mathrm{DN}}$ clusters and that for wildtype clusters ($\langle \mathrm{NP}_{\mathrm{Rec}^{\mathrm{DN}}} \rangle - \langle \mathrm{NP}_{\mathrm{wild}} \rangle$). Within Figure 4.4a, parameters that pass the Receptor$^{\mathrm{DN}}$ standard are in the right half of the plot, while parameters with positive directionality are in the top half of the plot; ideally, models should have points in the top right quadrant. Figure 4.4a shows more dots in the upper half for networks A, B, and F, which means that it is more likely to obtain good parameters for these three networks compared to the others. More importantly, there are no dots below the x-axis in networks A and B, meaning that regardless of parameters, these two networks always make the wildtype cluster follow the gradient. In other words, networks A and B robustly have positive directionality. The absence of robust directionality in networks C–F suggests that, if these are correct, further experimental interventions could convert chemotaxis to chemo-repulsion.

In addition, we are able to suggest experiments to help further discriminate between the six selected networks. One significant difference between networks C and D of Figure 4.2c and the other networks is how the protrusion of one cell is affected by the activation of R in another cell. For networks C and D, this control is through R, which means changing R in one cells influences not only the protrusions of other cells but also the level of R. Therefore, measurement of Rac activation simultaneously with photo-activation would show whether Rs activity can be controlled by light-treatment on other cells. Results of these experiments can then be used to

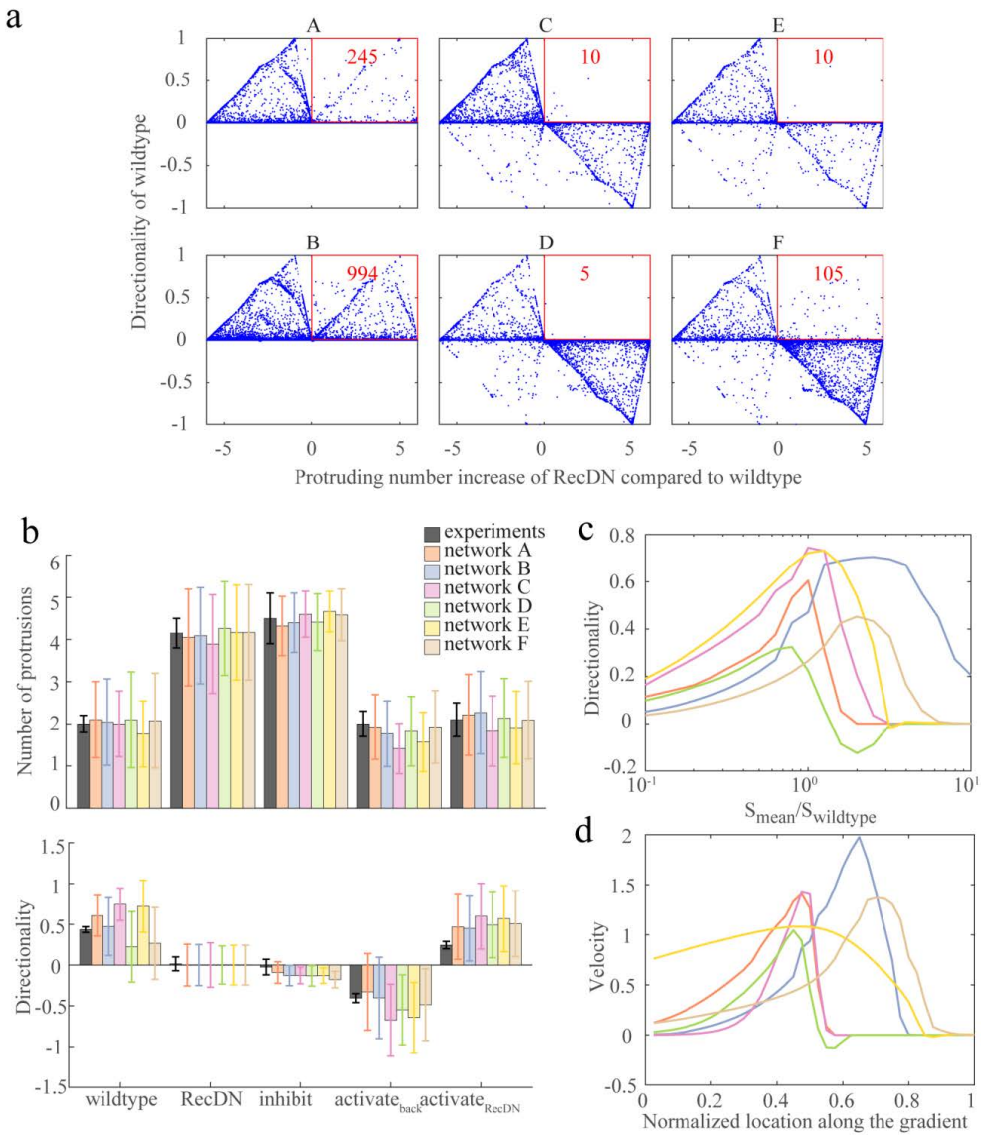


Figure 4.4: Simulation results. (a) Parameter sampling results for the six selected networks showing average increase of protrusion number under Receptor^{DN} condition compared to wildtype cluster in x-axes and the wildtype clusters directionality in y-axes. The red box shows the region that is qualitatively consistent with the experimental results and the number represents the number of parameter sets (out of 500,000) that produce results that are consistent with the experiments. (b) Comparison of experimental data, reported in Table 4.1, and simulation results for the six selected networks with fitted parameters. A–F corresponds to the network A to F in Figure 4.2c. Error bars show the standard deviation. (c) Trends of the clusters directionality for the six selected networks with fitted parameters when the mean signal is changed. (d) Trends of the clusters speed for the six selected networks with fitted parameters along the migration path.

further discriminate between topologies.

4.7 Parameter fitting

To make sure that parameters exist to meet the requirement of all the experiments simultaneously and as a prerequisite to making predictions, we fit parameters for each of the six networks using a simulated annealing algorithm. We fit not only the Receptor^{DN} and photo-inhibition experiments, but also the other protrusion number data found in Table 4.1. We also mandate that our fits display signs of directionality that are consistent with the experimental results listed in Table 4.1. Details of the procedure can be found in Section B.3. The comparison of protrusion number and directionality between the simulation results with fitted parameters and the experimental data, for the six networks, under all five conditions listed in Table 4.1, are shown in Figure 4.4b. Clearly, our best-fit parameters are able to produce results that are within similar levels of accuracy for each of the six networks. The values of the fitted parameters for the six networks are listed in Table B.1.

One interesting feature of this figure is that for most conditions, the standard deviations (SD) of the simulations are much larger than those of the experiment. For all the six minimal networks, there is no feedback from P to G so that the protrusion of one cell is independent of the protruding states of other cells. With this independence prerequisite, as the signal for every cell is the same in the Receptor^{DN} case, the protruding probabilities for all the cells in a cluster of n cells are given by Prob = $\frac{4.15}{n}$ as the experimental measurement shows an average of 4.15 protrusions under this condition. Now we can calculate the SD for protrusion number easily, as the probability of having n protrusions is simply $\binom{n}{i}\text{Prob}^i(1-\text{Prob})^{n-i}$ where $\binom{n}{i}$ is the binomial coefficient. Therefore, SD = $\sqrt{\sum_{i=0}^{n}(i-4.15)^2\binom{n}{i}\text{Prob}^i(1-\text{Prob})^{n-i}}$,giving SD ≈ 1.13 for n = 6. This is consistent with our simulation result and is much larger than the experimental SD which is about 0.35. This means that within a model with 6 cells where the probability of a cell protruding

is independent of whether other cells have protruded, the SD of the protrusion number cannot be reduced by merely changing the fit parameters. In order to get a SD close to the experimental result, we have to change either the cell number or the network topology to remove independence. When n = 5, SD $\approx$ 0.84 and when n = 4, in order to get an average protrusion number close to 4, Prob = 1 and thus SD = 0. Usually there are 4–8 border cells in the cluster [111], so one possible reason for the small SD in the experiment is that a significant fraction of cluster samples has only 4 border cells. The other possibility is to change the independence prerequisite by adding a link from P to G to form a negative feedback between P and G on the six selected networks. Networks with a P$\rightarrow$G link show reduced variance, because the feedback can constrain the average number of protrusions to be very close to a whole number. This is clear in the sampling results of Receptor^{DN} for the first two networks in Figure 4.3 showing pattern with grids at the whole numbers. If the variance is to be tightly constrained in this way, however, we need at least 7 interactions for the minimal networks.

4.8 Model predictions

Although we do not know which of six selected networks is the real one, we can still make some reasonable predictions based on their common simulation results with the fitted parameters. By changing the mean value of the signal while keeping the relative gradient constant, we can get the trends of the directionality. This is shown in Figure 4.4c which reveals that all networks display a maximal directionality near the original mean value of signal. This result is qualitatively robust, regardless of network topologies or parameters, as long as they meet the requirements of our selecting process. Figure 4.4c also shows that the directionality of network D (green curve) can become negative for some range of signal strengths. This is consistent with our previous conclusion that the networks, except networks A and B, do not guarantee chemotaxis. We note that a direct quantitative comparison between experimental and model results is difficult. First,

experiments have not established the concentration and gradient strength of chemoattractant in vivo. Secondly, we did not consider the length of the protrusion in our modelour protrusions were either on or off. For example, in our model, if all the cells in the cluster have protrusions, the directionality is exactly zero but in reality it is possible that the protrusions of the cells at the front half of the cluster are larger, resulting in a positive directionality.

Despite the lack of quantitative comparison, we can still compare our model prediction of Figure 4.4c with experimental data. Increasing the mean of signal in a wildtype cluster while keeping the relative gradient constant is similar to PVF1 overexpression. This overexpression was found to decrease the directionality of the cluster [122]. Furthermore, experiments have measured the speed profile of the cluster along its path, during which the mean signal is assumed to increase [105]. In order to compare with this experimental data, we define a velocity $V = \sum_i \vec{P}_i \cdot \vec{d}$, where $\vec{d}$ is unit vector in the signal gradient direction and $\vec{P}_i$ is the protrusion vector for cell i. Note that in this definition, we do not consider the length of the protrusion nor do we include any mechanical aspect of migration. We assume that the signal profile across the egg chamber is exponential, as suggested in [105] and the signal profile for a cluster at location Lx is $S(x) = S_{mean} e^{-S_{grad}(C-Lx)}(1 + S_{grad}\cos\vartheta_i)$. Here L is the total length of the migration path and x is the normalized location. We set the radius of the cluster to be 1 and by rough measurement of the length of the clusters travel path from images in [110] we will take L = 20. For C, the location where the data in [1] is measured, we will take C = 8, corresponding to measurements around 1/3–1/2 region of the clusters travel path (X. Wang, private communication). S_{mean} and S_{grad} are the parameters fitted in the previous part for the six selected networks and correspond to the mean and relative gradient of the signal at the measured position in the experiments. We plot V versus x curves for all the six networks in Figure 4.4d. We can see that the main trend that velocity first accelerates and then decelerates is consistent with the experimental curves in [105]. This behavior emerges from our model, which has been fit to the perturbation experiments, but uses no information from [105] other than the proposed exponential signal profile.

4.9 Discussion

In this study, we used mathematical analysis, combined with sampling simulations and parameter fitting, to determine the simplest network topology that is consistent with a set of in-vivo experiments on border cell migration. The mathematical analysis relied solely on qualitative experimental observations and is thus independent of specific functional forms of the interactions between the pathway nodes. Key to this analysis is the fact that specific experimental perturbations, either in the form of photo-inhibition, photo-activation, or mutations, resulted in clear phenotypic changes. Using these changes, we were able to reduce the amount of possible simple networks from a very large number to only six.

The complexity of our problem makes application of standard approaches infeasible. A number of existing techniques exist to infer network topologies from experimental data on single cells [106–108]. Often it is possible to gain insight into the topology on the basis of mutant analysis, drug responses, or clever experimentation [26, 61]. Furthermore, for topologies that contain a limited number of components, it is feasible to perform an exhaustive numerical search through parameter space to determine all possible topologies that are consistent with a particular feature observed in experimental data [119]. We have not directly applied these approaches to our problem, because inferring network topologies for collective migration is more challenging for multiple reasons. First, cells communicate with each other to sense signal gradient, resulting in a spatially extended system. Second, the output of the system is often a stochastic variable and is only known as an average quantity. In addition to these system-specific challenges, the number of network components might be too large to carry out exhaustive numerical searches. Therefore, we have developed a qualitative perturbation analysis method that can greatly reduce the number of candidate topologies number. We then carried out numerical simulations only on the remaining networks for further study.

The final result, six possible six-link pathways shown in Figure 4.2c, were verified using

both sampling simulations and by a fitting procedure that produced parameter sets for which the model results are fully consistent with all experimental data. We should point out that obtaining a satisfactory fit was not necessarily expected. Our fitting procedure takes into account directionality data (Table 4.1) which was not used in our qualitative selective procedure. Since we found parameter sets for all topologies that were consistent with the experimental data, we are not able to further distinguish between the six networks. However, additional sampling simulations show that some topologies may be more robust than others (Figure 4.4a). Furthermore, we propose additional experiments such as combining light-treatment with Rac activity measurement that should be able to further discriminate between possible topologies.

Despite these limitations, our results provide valuable information about the multicellular signal processing network in border cells collective migration, showing the defect of the multicellular LEGI model (network I in Figure 4.2b) which was studied in [106–109] and narrowing the topology to only six possibilities. These models may be interesting candidate models to further study collective migration in the border cell system as well as in other systems. For example, in [100, 123], Rab11 and Cdc42s function in cell-cell communication are studied using mutations and light-treatment of Rac. With their results, we can check what roles they may play in our topologies and further discrimination of the six topologies may be possible, or we may need a larger network including some new components. The six topologies that we have found are also potential candidates to study in contexts where minimal multicellular LEGI models have been previously applied [106–109]. These topologies may potentially provide more gradient sensing accuracy than the LEGI scheme [106, 107], or may have different chemotactic dependence on cluster sizes [105, 108, 109]. In addition, the photo-activatable Rac has been used in other cell types, such as mesendoderm cells [116] and neural crest cells [117], which also show collective cell migration. Application of our approach, combined with Rac activation experiments in these systems, could show whether the topologies we have found for the border cell cluster migration are robust across different cell types, or whether different topologies are optimal for differen-

tial biological circumstances. We should point out that if photo-activatable Rac, an excellent tool for the reversible manipulation of proteins, is not practicable, traditional genetic methods may also provide data that can be analyzed by our qualitative approach. For example, in [123], DNA-programmed epithelial cell assembly is used to study multicellular protrusions when Ras is activated in all or some of the cells. The comparison of homogeneous and heterogeneous Ras activity can provide similar kind of information as the Receptor^{DN} (homogeneous inhibition) and locally light-treatment of Rac (heterogeneous activation or inhibition) experiments we use in our work. Thus, our work provides a starting point for future research on signaling networks of collective chemotaxis and the methods we have developed here may be applied in other collective migrating systems as well.

4.10 Methods

The ordinary differential equation (ODE) for Rac is:

$$\frac{d[R_i]}{dt} = \frac{basal_{AR} + k_{SR}[S_i] + k_{GR}[G]}{1 + PI_i}(R^{tot} - [R_i]) - (basal_{DR} + k_{\_GR}[G])[R_i]$$

Here $[R_i]$ represents the concentration of Rac in cell i and $[S_i]$ is the signal sensed by cell i. No matter what the signal profile across the whole egg chamber looks like, as the size of the cluster is small compared to the length of the egg chamber, it is reasonable to approximate the $[S_i]$ within the cluster in a linear form: $[S_i] = S_{mean}(1 + S_{grad}\cos\vartheta_i)$ with ϑ_i as the angle of cell i in the cluster. In this equation, the first term is the activation of Rac and the second term is the deactivation of Rac. $basal_{AR}$ and $basal_{DR}$ are the basal activation and deactivation rate respectively. The activation or deactivation terms with $[S_i]$ or $[G]$ are the rates regulated by the signal or the global factor. PI_i is the photo-inhibition effect implemented on Rac in cell i. The photo-inhibition works with a photoactivatable form of the dominant-negative mutants of Rac (PA-RacT17N). RacT17N strongly binds to upstream GEFs and blocks wildtype Rac from being activated. We thus divide

the activation term in this equation by $(1+\mathrm{PI_i})$ to describe the photo-inhibition effect.

The ODE for the global factor G is:

$$\begin{aligned}\frac{d[G]}{dt} =&(\mathrm{basal_{AG}} + \mathrm{k_{RG}}([R]_{avg} + \mathrm{PA_{avg}}) + \mathrm{k_{SG}}[S]_{avg} + \mathrm{k_{PG}} \times \mathrm{NP}) \\ &-(\mathrm{basal_{DG}} + \mathrm{k_{\_RG}}([R]_{avg} + \mathrm{PA_{avg}}) + \mathrm{k_{\_SG}}[S]_{avg} + \mathrm{k_{\_PG}} \times \mathrm{NP})[G]\end{aligned}$$

Here $[R]_{avg}$ and $[S]_{avg}$ are the average concentration of Rac and signal over all the cells. In this equation, the first term is the production of the global factor and the second term is the degradation of this factor. $\mathrm{basal_{AG}}$ and $\mathrm{basal_{DG}}$ are the basal rates of G. $\mathrm{PA_i}$ is the photo-activation effect on Rac in cell i. In the photoactivatable form of Rac, the binding site of downstream effector on Rac is blocked without light of certain wavelength and light-induced conformational change would release this block and make Rac available to the effectors [121]. This means light treatment provides more activated Rac available to induce downstream event. This light activation effect does not go into the equation of [R]. Instead, we add the additional activated Rac (PA) provided by the light treatment to wherever there is [R] in the equations of [G] and [P] as a form of $[R]+\mathrm{PA}$. $\mathrm{PA_{avg}}$ is the average effect of photo-activation and is defined as $\mathrm{PA_{avg}} = \frac{1}{n}\sum_i \mathrm{PA_i}$ where n is the number of cells in a cluster. In addition, we haven't introduced a $\mathrm{G^{tot}}$ in [G]'s equations as we assume that [G] is far from saturation.

The stochastic differential equation (SDE) for protrusion is:

$$\begin{aligned}\frac{d[P_i]}{dt} =&(\mathrm{k_{RP}}([R_i] + \mathrm{PA_i}) + \mathrm{k_{GP}}[G] + \mathrm{k_{SP}}[S_i])\frac{[P_i]^2}{[P_i]^2 + 1^2}(\mathrm{P^{tot}} - [P_i]) \\ &-(\mathrm{basal_{DP}} + \mathrm{k_{\_GP}}[G] + \mathrm{k_{\_RP}}([R_i] + \mathrm{PA_i}) + \mathrm{k_{\_SP}}[S_i])[P_i] + \eta(t)\end{aligned}$$

Here, the first term is the activation of P and the second-order Hill function represents the self-activation of P. The second term is P's deactivation and the basal deactivation rate $\mathrm{basal_{DP}}$ is normalized to 1 in the simulation. The third term is the Gaussian noise term.

If we assume that the equations for [R] and [G] reach equilibrium quickly compared to the time scale of protruding (20–30 min) [1, 118], we can get the steady states for [R] and [G] as follows and then plug them into [P]'s equations:

$$[R_i]^s = \frac{basal_{AR} + [S_i] + k_{GR}[G]^s}{basal_{AR} + [S_i] + k_{GR}[G]^s + (basal_{DR} + k_{-GR}[G]^s)(1 + P_i)} R^{tot}$$

$$[G]^s = \frac{k_{RG}([R]^s_{avg} + PA_i) + k_{SG}[S]_{avg} + k_{PG} \times NP + basal_{AG}}{k_{-RG}([R]^s_{avg} + PA_i) + k_{-SG}[S]_{avg} + k_{-PG} \times NP + basal_{DG}}$$

With this fast equilibrium assumption, we are only interested in the steady states of [R] and [G] so that we can normalize $basal_{DR}$, R^{tot}, $basal_{AG}$ and $basal_{DG}$ to 1.

It's worth noting that these equations are the general forms including all the possible interactions and given certain network, only the parameters related to the existing interactions are non-zero. For networks with five or six interactions, three of the interactions are fixed and their related parameters are listed in Table 4.2 and the rest parameters related to the undetermined interactions are represented by k_1, k_2 and k_3 in Table 4.2. For one specific network, such as network A in Figure 4.2c, $k_1 \equiv k_{SG}$, $k_2 \equiv k_{RG}$ and $k_3 \equiv k_{-GP}$.

4.11 Acknowledgements

We thank Denise Montell and Xiaobo Wang for useful conversations. This work was supported by National Institutes of Health grant number P01 GM078586.

Chapter 4, in full (with minor exceptions), has been accepted by Biophysical Journal in 2018 as "Minimal network topologies for signal processing during collective cell chemotaxis. Yue, Haicen; Camley, Brian. A; Rappel, Wouter-Jan". The dissertation author was the primary investigator and author of this paper.

Chapter 5

Inferring single-cell behavior from large-scale epithelial sheet migration patterns

5.1 Introduction

Collective cell migration is an important biological phenotype used in many biomedical assays. For example, in a wound healing assay, the speed at which two monolayers of cells migrate towards each other is often measured to determine a cell migration response to drug treatments. The use of collective migration as a biomedical phenotype stems from its important role in many biological processes; collective migration is essential for development [90, 124] and wound healing [125, 126], but its misregulation plays a role in diseases such as metastatic cancer [91, 127]. It is non-trivial, however, to infer single-cell migration behaviors from metrics such as the monolayer boundary displacement. Despite many studies on the behavior of individual cells,

which have looked at features of migration such as the influence of the surrounding microenvironment [128–130] or the flow of actin within a migrating cell [131, 132], the connection between individual cell properties and collective behavior remains unclear.

To infer single-cell phenotypes from collective migration behavior requires additional information beyond that provided by metrics such as time to wound closure. Particle image velocimetry (PIV) allows us to extract not only boundary motion from time-lapse imaging data but also the entire flow field of cell motion, which includes features at the scale of single cells or smaller. Since their introduction into cell migration research several years ago, PIV flow fields have rapidly emerged as a powerful tool and have been used to analyse a wide variety of collective cell behaviors including correlated motion [98, 133], vortices [134], patterns of stresses within the cell sheet [135] and changes to collective migration during malignancy [136]. Notable prior work on large-scale flow fields generated by coupled polar entities was carried out in the context of soft condensed matter, starting with liquid crystals and most recently active matter [137, 138]. Our focus is on the heterogeneity and time variability that is a hallmark of living systems, including most notably sources of heterogeneity such as cellular activity and leader cells that are thought to control collective behavior in living systems.

Here, we link detailed metrics of collective behavior derived from PIV data with simulations of collective cell motion that explicitly model the behavior of individual cells. By linking experiments and simulations, we can infer likely single-cell behavior from collective motion phenotypes. Collective cell migration has been studied using a wide variety of modelling techniques [3, 98, 139–144]. These modelling techniques have been used to explore a variety of factors involved in collective migration that are difficult to access experimentally, including the effect of matrix geometry on migration strategies [145] and the maturation of cell contacts within a monolayer [146]. Several studies have noted the important link between cell motility and polarity [146–148], a feature of migration that we explore further.

Our work links previously published experimental observations on collective migration of

MCF10A breast epithelial cells [149] and a previously published model of collective migration [2, 3, 150, 151] to elucidate which properties of individual cells are most consistent with the observed multi-cell collective migration behavior. The experimental data show changes in collective behavior on large length scales that span the cell monolayer. In the cases of collective behavior, cell motion contributes to the radial expansion of the monolayer across millimeter scales, while in other, less collective cases, only those cells near the outer edge contribute to monolayer expansion. We show that these changes in large-scale migration patterns can be recreated in our model without requiring large-scale gradients or leader cells (a subpopulation of cells at the edge of the monolayer that has different properties from cells within the bulk of the monolayer). Our modelling results suggest that the experimentally observed changes in collective behavior are consistent with simply decreasing the activity of individual cells as long as the cells have a strong coupling between their velocity and preferred motility direction.

5.2 Material and methods

5.2.1 MCF10A dataset

We analyse a previously published set of time-lapse images of MCF10A (breast epithelial cells) migrating in collective sheets [149, 152]. These cells, which were plated in a circular monolayer, migrate on a collagen IV coated glass surface and phase contrast images were taken every 3 min for a total of 1000 min (16.6 h). The dataset includes cells migrating in normal cell culture media (referred to as $1:1$) and cells migrating in a dilution of this culture media (referred to as $1:5$). This change decreases the horse serum, insulin, EGF, hydrocortisone and cholera toxin concentrations to 20% of their full media values (e.g. horse serum at a concentration of 1% instead of 5%).

5.2.2 Migration analysis

Phase contrast images were analysed with PIV using the MatPIV toolbox (J. Kristian Sveen, GNU general public licence) for Matlab (MathWorks, Inc.). Multiple iterations of interrogation window sizes were used: two iterations of 64×64 pixel windows were followed by two iterations using 32×32 pixel windows. A 50% overlap was used for each interrogation step. Several steps were taken to increase the quality of our PIV flow fields. The size of our interrogation windows was chosen to provide of the order of 10 features per window, suggested as a good rule of thumb for a clear PIV signal [153]. We also run a multi-pass algorithm that first uses a larger window size to reduce noise. The resulting flow field uses information about the relative height of the chosen cross-correlation peak and all possible correlation peaks as a signal-to-noise ratio to further filter the PIV flow field; outliers were detected using a signal-to-noise threshold of 1.3.

Custom Matlab segmentation code was used to find the leading front of the cell monolayer. The phase contrast images were Sobel filtered followed by median filtering and morphological opening to clean the binary image before finding the perimeter of the objects in the image. The edge coordinates were then found using a Matlab implementation of Dijkstras algorithm (dijsktra path finder by Sebastien PARIS, available on the Matlab File Exchange at mathworks.com). In combination with the microscope stage positions, this edge was used to fit the cell monolayer to a circle (See Figure C.1 in Appendix C). The effective radius of the monolayer and the centre position were used to define regions of the cell monolayer for later migration analysis. Speed and radial velocity values were averaged over theta to create radial profiles of motion.

Velocity correlations were calculated as

$$C(\Delta r) = \frac{\sum_{r,t}(\vec{v}(r) - \bar{v}) \cdot (\vec{v}(r - \Delta r) - \bar{v})}{\sigma_v^2}.$$

In this case, $\vec{v}(r)$ is the velocity at a location r within the cell sheet. The correlation values are averaged over all r within the region 0.5 , r/R , 0.75 and over all times. In this equation and throughout, r refers to the location within the cell monolayer with respect to the monolayer centre and R refers to the size of the monolayer. These correlation values were fitted to a double exponential of the form

$$C(\Delta r) = Ae^{-r/L_{c1}} + Be^{-r/L_{c2}} + (1 - A - B).$$

5.2.3 Simulations

Our model is based on earlier work [2, 3] and includes cell growth and division, motility forces, friction forces, and volume exclusion and adhesion between neighboring cells, as shown in Figure 5.2a. In the model, two point particles represent a single cell with a size given by the inter-particle distance r. The equation of motion for each particle is given by

$$\frac{d\vec{p}}{dt} = \vec{m} + \vec{F}_{exp} + \vec{F}_{int} + \vec{F}_B + \sum_{r \leq R_{CC}} (\vec{F}_{rep/ad} + \vec{F}_{df})$$

Here, $\vec{m}$ is the motility force (discussed further below) and its value, along with all other parameter values, is given in Appendix C, Table C.1. Cell growth is simulated through a repulsion force between the two particles within the same cell: $\vec{F}_{exp} = -B/(r+1)^2\hat{r}$, when $r \leq R_{exp}$, where B is an expansion factor, $\hat{r}$ is the unit vector parallel to the line connecting the two particles and R_{exp} is the range of cell expansion. Cell division is incorporated as follows: after the cell size reaches a threshold R_{div}, it divides at a constant rate k_{div}, after which two new particles are placed a distance r_{div} away from the particles constituting the old cell. $\vec{F}_{int}$ represents the intracellular friction force between particles constituting the same cell and is determined by $\vec{F}_{int} = -\xi_{int}(\vec{v}_{ij} \cdot \vec{r}_{ij})\hat{r_{ij}}$ when $r_{ij} \leq R_{exp}$. Cell-substrate friction is assumed to be proportional to the cell velocity, resulting in $\vec{F}_B = -\xi_B v$. The last two forces in the equation of motion act only within a distance of $r \leq R_{CC}$

and represent volume exclusion and adhesion of cells ($\vec{F}_{rep/ad}$) and friction forces that oppose the relative motion of cells with a friction coefficient ξ_{df} ($\vec{F}_{df}$). The specific form of these two forces can be written as $\vec{F}_{rep/ad} = -(f_0(1/r_{ij} - 1) - f_1)\hat{r_{ij}}$ with a short-range repulsive term, which prevents cell-cell overlap and is parameterized by f_0, and a long-range adhesive force with a strength determined by f_1 $\vec{F}_{df} = -\xi_{df}(1 - r_{ij}/R_{CC})^2(\vec{v}_{ij} \cdot \vec{r}_{ij})\hat{r_{ij}}$. More details can be found in Basan et al. [2, 3].

The orientation of the motility force $\vec{m}$ is such that it tends to align with the cell's velocity v. Specifically, this motility-velocity alignment coupling is implemented by assuming two states of the cell: a motile state and a non-motile state. Cells in the nonmotile state exert no motility force while cells in the motile state generate a motility force with a fixed magnitude in a random direction. The transition from the non-motile state to the motile state is determined by a constant rate k_{wake}. The transition rate from the motile state back to the non-motile state depends on the motility-velocity alignment, quantified by $\vec{m} \cdot \vec{v}$. In Basan et al. [3], this transition rate could take on only two values: low for aligned cells ($\vec{m} \cdot \vec{v} > 0$) or high for non-aligned cells ($\vec{m} \cdot \vec{v} < 0$). This bias results in a higher ratio of aligned cells versus non-aligned cells in the motile state and provides an alignment mechanism. In this study, we use a continuous form of the transition rate to the non-motile state that depends on the degree of alignment

$$k'_{sleep} = \frac{(k_0 - k_1)(1 + \tanh(-(\vec{m} \cdot \vec{v})/k_{mv}))}{2} + k_1$$

Thus, this rate ranges from k_1 for aligned cells to k_0 for nonaligned cells while the transition between these two rates is determined by k_{mv}; when $k_{mv} \ll |\vec{m}||\vec{v}|$ the transition rate approaches the aligned and not-aligned binary case of Basan et al. [3].

Our transition rate was further modified to account for the effect of leader cells, i.e. cells near the edge of the colony that preferentially move outward. To this end, we multiplied the above transition rate with a spatially dependent prefactor so that the final form for particle i is

$k_{sleep} = k'_{sleep} e^{k_{Rm} \vec{m}_i \cdot \vec{R}_i}$. Here $\vec{R}_i$ is a spatial average of the particle's neighbors: $\vec{R}_i = \sum_{neighbors} \vec{r}_{ij}$, where $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$ is the relative position of cells i and j. The parameter k_{Rm} determines the strength of the leader cell effect. Neighbor cells are defined as cells within the range of non-zero cell interactions, i.e. $r \leq R_{CC}$. Cells far from the edge are surrounded by other cells such that $\vec{R} \approx \vec{0}$ and the prefactor is close to one. Cells near the edge, however, have only neighbors inside the colony, resulting in a value of $\vec{R}$ that points inward. Consequently, k_{sleep} is smaller for cells near the edge that have their motility vector pointing outwards, leading to 'leader' cells that move outward. In the standard set of parameters See Appendix C, Table C.1), $k_{RM} = 0$ and there is no leader cell effect.

The simulation starts with 200 spheres (and thus 100 cells) randomly located within a square box of 10×10, measured in units of l_0. The initial velocity for all the spheres is taken to be zero. The first 10 frames of the simulated data ($100t_0$) were discarded before migration analysis to remove artefacts from low cell number monolayers. Particle positions and velocities were interpolated to a grid to compare with experimental PIV results. The edge of the simulated monolayer was found based on the particle positions and fitted to a circle. As with the experimental results, this edge was used to define radial regions of the monolayer and to create profiles of speed and radial velocity. Migration analysis was then carried out in the same manner as previously described for the experimental data. We should note that the length scale of the model, l_0, can be related to experimental values by comparing it to the experimental division size. In our simulations, we choose the threshold distance for cell division to be $0.4l_0$. This should be of the same order of magnitude as the average cell size, which is approximately 40 μm, resulting in $l_0 \approx 100\mu m$. The simulation time can then be determined by comparing the mean speed of cells in the simulation ($\approx 0.025 l_0/t_0 = 2.5\mu m/t_0$) to the mean speed in the experiment ($\approx 0.25\mu m/min$) which leads to $t_0 \approx 10min$. We can then verify that the relative growth rate in the simulation is consistent with the relative growth rate observed in the experiments, as shown in Appendix C, Figure C.2. This comparison suggests that the experiments and simulations are on the same scale

for growth, but note that we do not attempt to fit the multiple model parameters to the data.

5.3 Results

To study large spatial scale migration patterns in epithelial MCF10A cells, we analyse a previously published set of time-lapse images of a circular cell monolayer [149]. The cells were imaged near the edge (Figure 5.1 a) and in the center of the monolayer; a schematic of the imaging fields of view is shown in Figure 5.1b. The multiple fields of view allow us to investigate migration behavior as a function of location within the monolayer; Figure 5.1c shows an example kymograph of speed within a monolayer. There are heterogeneities in speed over both the 16 h time course and approximately 4 mm spatial scale of the cell sheet, yet there is a trend towards higher speed at the edge of the monolayer. This trend can be seen in the time-averaged speed curves in Figure 5.1d. In addition to measuring speed, which indicates how active the cells are, we also determine the radial velocity as a measure of how well the cells' motion contributes to the overall expansion of the cell sheet. As shown in Figure 5.1e, when the overall motion is slow, we see a decreased radial velocity, as would be expected.

In addition to this quantitative change, however,we also see a qualitative change in which some experiments exhibit a concave curve (black squares in Figure 5.1e) while others—those with lower cell speeds—show a convex curve (red circles in Figure 5.1e). Since radial velocity is the component of cell motion that contributes to the collective expansion of the monolayer, this suggests changes in the collective behavior of the cell sheet. In those cases with lower overall cell speed, the cells near the center of the monolayer move in all directions; since radial velocity can be positive or negative, these values average out to near zero in the center region. In the cases with higher overall speed, however, cells throughout large regions of the cell monolayer contribute to the overall expansion. Thus, we use the shape of the radial velocity curve as one indicator of how cooperatively the cells move—when the curve is concave, cooperative motion

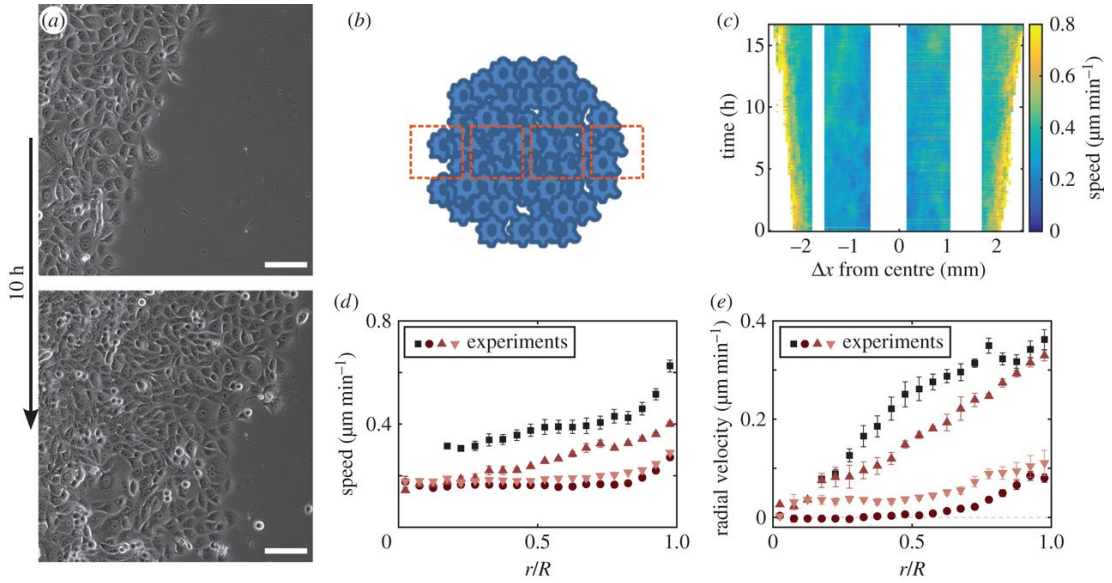


Figure 5.1: MCF10A cell sheets migrate collectively over 10 h (a; scale bar 100 μm). The cells are plated in a circular monolayer and multiple fields of view scanning the diameter are imaged (b; see also electronic supplementary material, figure S1). Over time, these regions show different speed profiles with larger speeds seen near the edge of the cell sheet (c). The experiments show variability in overall speed, but cells near the edge remain faster (d). In low speed cases, the radial velocity decreases and also qualitatively changes pattern (e). Error bars in (d,e) represent the standard error of the mean of four technical replicates performed on the same day in full media (1 : 1, black squares) or diluted media (1 : 5, red circles and triangles).

spreads through the cell monolayer, but when the curve is convex the cells move less collectively.

To explore how individual cell properties affect the qualitative large-scale changes seen in this dataset, we compare the experimental results to simulations using a previously published model [2, 3, 150]. In this model, each cell is represented by two particles that move according to the forces acting upon them (shown schematically in Figure 5.2a). These forces, further detailed in the Section 5.2.3, include friction forces, motility forces, cell-cell interaction forces and a force representing cell growth. Cell division is initiated once the distance between the particles constituting the same cell exceeds a threshold. Cells can be either motile or non-motile, and the transition rate between these two states is chosen such that motile cells tend to align their motility force (the force that pushes a cell in a preferred direction) with their velocity. Alignment between the cells motility force and velocity has previously been shown to lead to ‘go with the

flow' dynamics [3]; alignment between velocity and a preferred direction is also a feature of active matter models of cell migration [154].

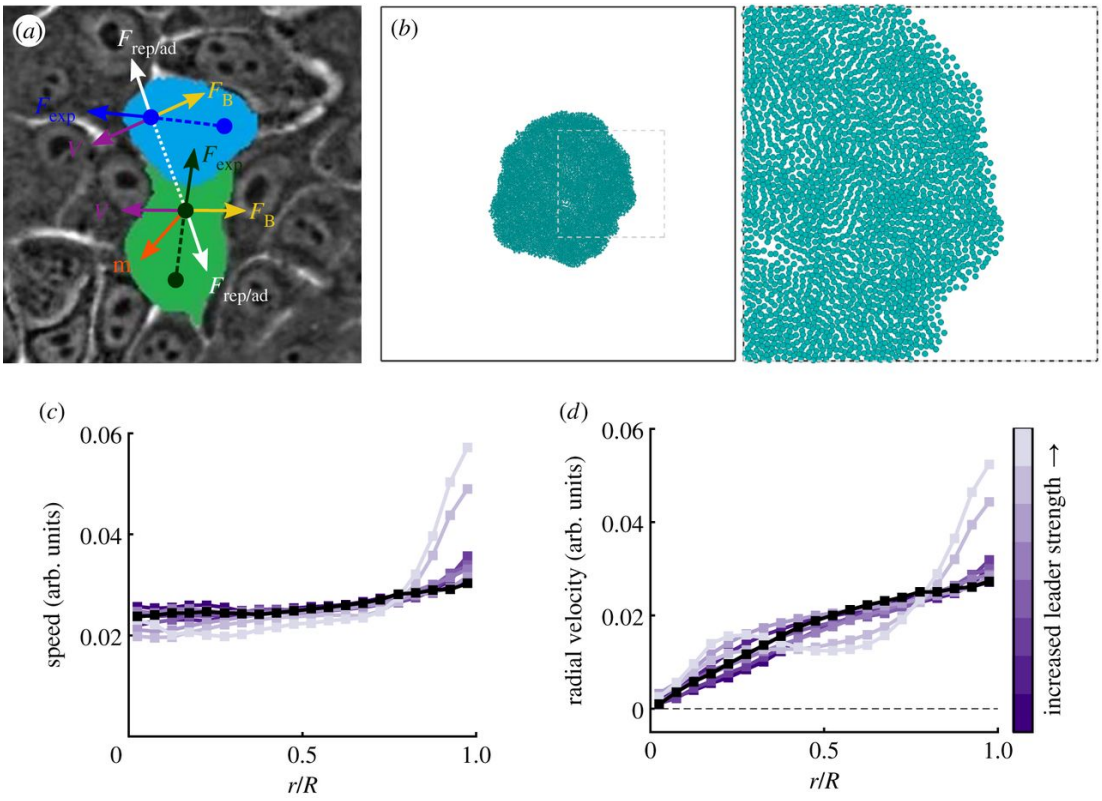


Figure 5.2: Multiple forces act on the simulated cells (a). Each cell is composed of two particles, which feel a cell expansion force (F_{exp}). The particles also have a velocity (V) and experience friction between the cell and substrate (F_B). Cells interact through a force that is repulsive at short distances and attractive at intermediate distances ($F_{rep/ad}$). Cells may be in a sleep state (a, blue cell) or awake state (a, green cell); only awake cells experience a motility force (m) that provides a preferred direction of motion. Particles also experience friction forces that are not shown. A still image from the movie of simulated cell motion is shown in (b). Simulated cells show a similar speed profile to the experimental results; the addition of leader cells increases the speed at the edge of the monolayer (c). Simulated leader cells increase radial velocity, but do not cause the qualitative change in radial velocity seen experimentally (d). From dark purple to light purple, k_{Rm}—which controls the leader cell strength—is set to 0.01, 0.05, 0.1, 0.25, 0.5, 1, 5 and 10, while the black curve represents the standard simulation set with no leader cells, $k_{Rm} = 0$ (c,d).

Simulated monolayers in this model using the standard parameters (See Appendix C, Table C.1) and in the absence of leader cells show a radial expansion similar to the MCF10A experimental system, and have increased speed near the edge of the monolayer (see Figure 5.2b,

black curve). Simulations with the standard parameter set also show an increased radial velocity near the leading edge of the cell sheet and a concave radial velocity profile in the centre of the cell monolayer (Figure 5.2c, black curve), similar to the experimental results with faster cells (Figure 5.2c, black squares).

In a first computational test, we explore whether introducing an active subpopulation of cells near the edge, leader cells, may be sufficient to cause the experimentally observed change in velocity profile shown in Figure 5.1. Leader cells near the leading edge of migrating cell groups are seen in many collective migration systems [93] and have often been observed at the tip of finger-like protrusions of epithelial sheets [97, 155]. Many studies have also found that multicellular rows of leader cells can emerge in epithelial wound healing [93, 156]. The exposure to extracellular matrix on one side of the cells and cell-cell adhesion on the other, as found at the leading edge of cell sheets, can cause cells to change their morphology or up-regulate distinct signalling pathways [93, 97, 155, 156].

In our simulation, leader cells are included by making the transition rate between the motile and non-motile state of motion dependent on the spatial location of the cell (see Section 5.2.3). As seen in Figure 5.2c, increasing the strength of this leader cell effect can lead to increased speeds near the monolayer edge. However, this increase is not accompanied by a qualitative change from concave to convex radial velocity profiles (Figure 5.2d), suggesting that another mechanism is responsible for the experimentally observed change in the radial velocity profiles. At high leader cell strengths, the profile does qualitatively change in that it shows a flat profile, in part, of the sheet, but this is due to an instability of the boundary shape. This instability results in the formation of fingering structures at the leading edge, similar to the behavior seen in a previous study which found leading tip cells in finger-like protrusions of epithelial sheets [97].

We next hypothesized that the proliferation rate may affect the phenotype. Changing proliferation rates in our simulations, however, does not change the simulated speed profile (See Appendix C, Figure C.3a) or the radial velocity profile (Figure 5.3a).We should note that for low

proliferation rates (less than 10% of the standard parameter value for division), the monolayer begins to break apart (See Appendix C, Figure C.5). In this case, the cell sheet can no longer be fitted to a circle for radial analysis. The disassociation of the monolayer at low proliferation rates agrees with a hypothesis that cell proliferation is used to fill in the gaps left by a migrating monolayer rather than as a mechanism for pushing migration forward [97, 157].

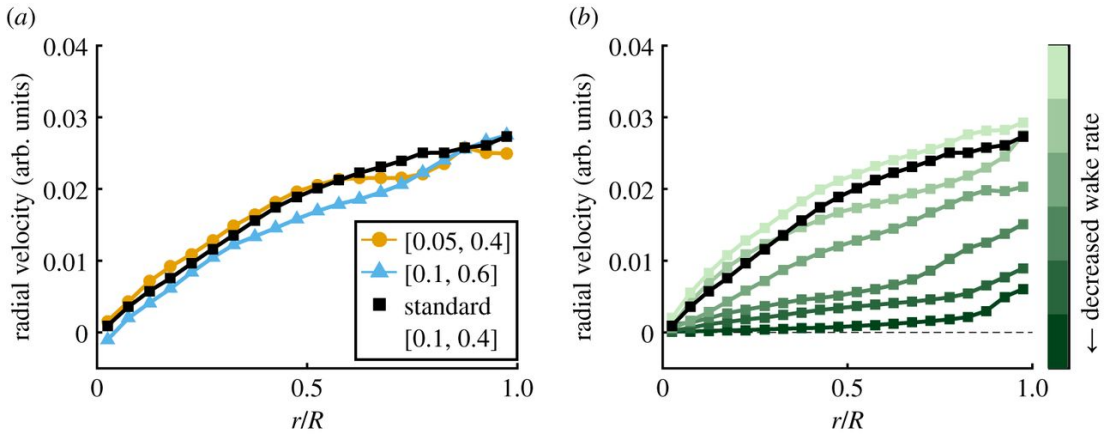


Figure 5.3: Decreasing proliferation rate does not qualitatively change the radial velocity (a); legend pairs indicate [k_{div}, R_{div}] parameter values. Decreasing the rate at which non-motile cells 'wake' causes a transition from a concave to a convex curve (b). Wake rate parameter values (k_{wake}) in (b) are 0.03, 0.05, 0.1, 0.2, 0.3 and 0.5 from dark green to light green with the standard wake rate of 0.4 shown in black. Speed profiles for these simulations are shown in Appendix C, Figure C.3.

In a third set of computational trials, we determined how changes in the 'wake rate' parameter affect velocity profiles. The wake rate parameter controls how often cells switch from a state without a preferred direction of motion to a 'motile' state in which the cells experience an additional force that indicates a preferred direction of motion and thus controls how 'active' the cells are within the monolayer. The motility force plays the role of a biomechanical polarity within the cell, but does not distinguish which of the many mechanical and chemical stimuli that lead to polarity changes [158] are experienced by the cell. It may be expected that the cases of reduced speed in the experimental data correspond to decreased activity in the cells.

Our simulations reveal that decreasing the wake rate leads to an overall lower speed (See Appendix C, Figure C.3c) and, as shown in Figure 5.3b, a qualitatively different radial velocity

profile. Specifically, decreasing the wake rate leads to a transition from behavior where cells across the radius of the monolayer move outward, showing radially expanding motion, to a state where only those cells near the edge move cooperatively outward. As a result, the radial velocity profile changes from a concave one to a convex one consistent with the experimentally observed changes in radial velocity and collective behavior.

We next determined how the coupling between the cell's velocity and motility force, parametrized by k_{mv}, affects the velocity profiles. This alignment coupling leads to higher sleep rates for those cells that have velocity and motility forces pointing in opposing directions. Interestingly, if the coupling between the motility force and velocity is removed, the simulations always show convex radial velocity profiles (i.e. decreased collective migration), even for large values of the wake rate (Figure 5.4a). This suggests the strength of this coupling may also play a role in migration. Indeed, using otherwise standard parameters but decreasing the strength of the coupling between velocity and motility can cause a transition from a concave to a convex radial velocity curve (Figure 5.4b). Thus, causing each cell to have strong alignment between its motility direction and its current velocity leads to more active behavior (concave velocity profiles), while motility that is not strongly coupled leads to convex velocity profiles similar to those seen in the less active experiments.

To probe further the changes in collective behavior that may accompany the observed changes in radial velocity, we calculate velocity correlations in the region between the centre and the edge of the monolayer ($0.5 < r/R < 0.75$). This range was chosen to correspond to the region where we see experimental changes in the radial velocity profile. The spatial autocorrelation of velocity does not distinguish between types of behavior such as divergence or rotation, but rather provides a metric for similarity of motion across the cell sheet; this similarity would be expected when cells migrate cooperatively. As shown in Figure 5.5a, experimental conditions with decreased speed and convex radial velocity correspond to the cases of less correlated motion within the cell sheet. To determine whether this result is consistent with our computational

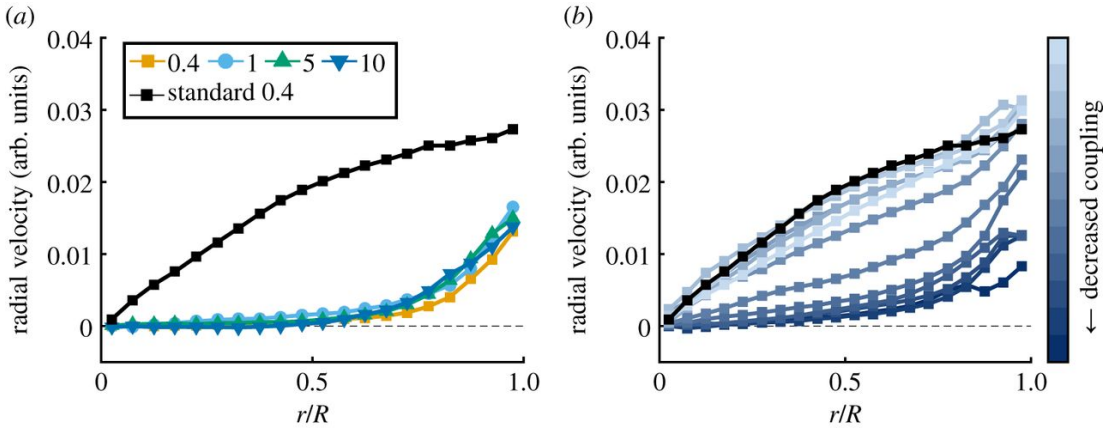


Figure 5.4: If velocity-motility coupling is removed (a), changing the wake rate cannot capture the experimentally seen results (k_{wake} value shown in legend; standard parameter set with coupling and $k_{wake} = 0.4$ shown in black). Radial velocity undergoes a qualitative change as the coupling is increased which is similar to the change seen in the MCF10A experiments (b); from dark blue to light blue the coupling parameter k_{mv} is set to 0.1, 0.05, 0.04, 0.035, 0.03, 0.025, 0.02, 0.015, 0.01, 0.005, 0.001 and 0.0001. The standard k_{mv} value of 0.00001 is shown in black. Smaller values of k_{mv} indicate a stronger coupling strength. Speed profiles for these simulations are shown in Appendix C, Figure C.4

model we compute the spatial correlation in our simulated trials. We focus on the perturbations—reducing the wake rate and decreasing the motility-velocity coupling—that are able to reproduce the qualitative change in radial velocity profiles. We find that both decreasing the wake rate (Figure 5.5b) and decreasing the coupling (Figure 5.5c) result in a reduction of the spatial correlation consistent with the experimental results.

We compare the observed changes in radial velocity and correlated motion within the cell sheet by calculating characteristic length scales from the velocity autocorrelation. Both the experimental PIV flow fields and the simulated cell motion naturally include two length scales: the experimental data include both subcellular and multicellular flows and the simulated data include the motion of both particles that make up a cell. To address this, we fit the correlation curves to a double exponential and indeed find that the smaller of the two experimental length scales (approx. 15 μm) is of the order of the cell size (Figure 5.5d). We also find that the larger length scale, which indicates multicellular cooperation, shows an increasing trend with increasing radial velocity for both the experimental data and model results (Figure 5.5d-f), further suggesting

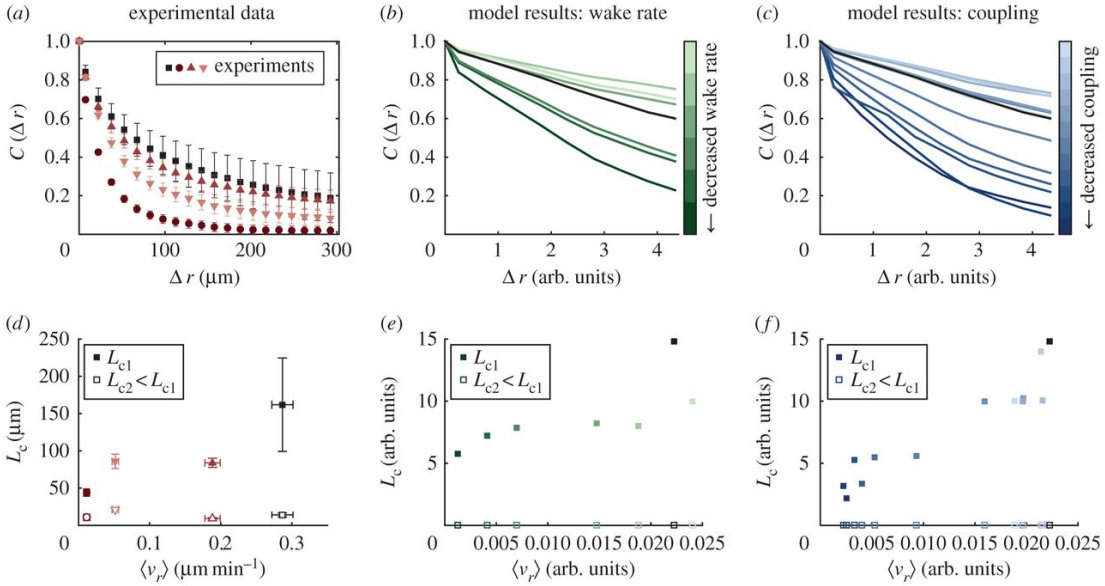


Figure 5.5: In the case of decreased experimental speed (a,d), decreased wake rate (b,e) and decreased coupling (c,f), the cells are less correlated in their motion. (a–c) Time-averaged velocity correlations (see Section 5.2) in the transition region, defined as $0.5 < r/R < 0.75$. (df) The characteristic length scales (L_c) determined from a double exponential fit to the curves in (a–c). Error bars on the experimental data represent the standard error of the mean of four technical replicates. Wake rate parameter values (k_{wake}) in (b,e) are 0.03, 0.05, 0.1, 0.2, 0.3 and 0.5 from dark green to light green. From dark blue to light blue in (c,f), the coupling parameter kmv is set to 0.1, 0.05, 0.04, 0.035, 0.03, 0.025, 0.02, 0.015, 0.01, 0.005, 0.001 and 0.0001. Smaller values indicate a stronger coupling strength. The standard parameter set ($k_{wake} = 0.4$ and $k_{mv} = 0.00001$) is shown in black in (b,c,e,f).

that the observed change in radial velocity corresponds to a change in multicellular cooperation.

5.4 Discussion

Here we provide insight into the single-cell behavior underlying collective migration by comparing experimental data to a model of collective migration. The emergent behavior of collective cell sheets that results from changing the properties of individual cells is difficult to predict and to probe experimentally. To investigate the connection between individual and collective behavior, we use PIV-based migration data to measure motion that spans a wide variety of scales; the length scales studied vary from that of a single cell to a monolayer composed of

thousands of cells. By comparing simulation and experiment across these scales,we provide insight into collective behavior changes observed in epithelial MCF10A cells. Cells moving at higher speeds show a larger region of radial expansive motion than slowly moving cells (Figure 5.1), suggesting changes in their collective behavior.

Leader cells have been found in a variety of experimental systems and in previous studies have been found to recruit following cells to move directionally outward through a Viscek-like coupling [159]. We show here that they are not necessary, however, to recapture large-scale changes in collective migration in our model of epithelial sheet migration (Figure 5.2).We also show that changes in proliferation are unlikely to be the cause of the observed large-scale changes (Figure 5.3a). Experimental changes in activity can be compared with the wake rate in the simulations; this parameter is related to the simulated cells ability to switch between a non-motile and motile state. Here we find that changes in the wake rate can lead to the experimentally observed changes in collective behavior (Figure 5.3b).

Interestingly, changing the wake rate has no effect without implementing alignment coupling between the cells' motility and velocity (Figure 5.4a). Previous work on this collective migration model has shown that this coupling leads to a 'go with the flow' dynamic [3]; the coupling is implemented such that cells with misaligned motility and velocity vectors are more likely to transition to a non-motile state. The motility force gives the cells a preferred direction of motion, and in that sense, it can be viewed as a biomechanical cell polarity. The ability to align this biomechanical polarity with the direction of motion of a cell indicates that a cell is able to sense resistance to motion and adapt its biomechanical machinery to push in a direction in which the cell is actually able to move. Thus, the strength of this coupling indicates how sensitive a cell is to its mechanical environment, and it is not surprising that changing this motility to velocity coupling changes the collective behavior of the cells (Figure 5.4b). This result agrees with previous work which suggests that propagating waves of cell stretching, which leave cells with more elongated and polarized shapes, are followed by waves of increased directionality in

sheets of epithelial cells [160].

Ultimately, the motility coupling strength provides a simple sensor of the surrounding tissue and leads to cells that are best able to follow a path of least resistance, since increased motion along the path of least resistance will lead to feedback with alignment of the motility force. Changing this coupling strength can cause large-scale changes in cell migration without explicitly changing the interactions between neighboring cells or invoking leader cells. Increasing activity within the cell sheet has a similar effect as it gives the cells more opportunities to be motile at the same time and thus follow other cells on paths of least resistance. When combined with a strong sense of polarity, this activity can then lead to increasing cooperativity over time as the cells have more opportunities to align their motion.

We thus find in our simulations that the emergent behavior of a simple model system can explain large-scale changes in collective behavior without the need to specify large-scale features explicitly. We identify cell activity and the coupling between biomechanical polarity and motion as two interesting features of individual cell behavior that can lead to large-scale collective behaviors. Our study demonstrated that PIV measurements may be used in conjunction with simulations to infer detailed biomechanical single-cell phenotypes from the types of collective migration assays that are commonly used in drug discovery and cancer research.

5.5 Acknowledgements

We thank the members of the Losert lab and Carole Parents lab (LCMB, CCR, NCI, National Institutes of Health) for their discussion and suggestions. This work was carried out with financial support to R.M.L. and W.L. from a NSF-Physics of Living Systems grant (PHY1205965). R.M.L. was additionally supported by the JCM Foundation through an ARCS/MWC Scholar Award. W.-J.R. and H.Y. were supported by NIH grant no. P01 GM078586 and NSF grant no. DMS 1309542.

Chapter 5, in full (with minor exceptions), is a reprint of the material in part as it appears in "Inferring single-cell behaviour from large-scale epithelial sheet migration patterns. Lee, Rachel M.; Yue, Haicen; Rappel, Wouter-Jan; Losert, Wolfgang, Journal of The Royal Society Interface, 2017". The dissertation author was the primary investigator and author of this paper.

Appendix A

Appendix to Chapter 2

Table A.1: Calculated thresholds used to determine the function that best fit the G protein dissociation kinetics obtained for different cAMP concentrations.Error differences (ΔE) were calculated as indicated in the Materials and Methods section for wild-type cells (WT) and cells lacking cAR3 (*carC*$^{-}$). $\Delta E < 0.05$ indicates a single exponential function fits the data, $\Delta E > 0.1$ indicates a double exponential function best fits the data. The dashed line indicates where the fits switch between single and double exponential.

cAMP(nM)	ΔE(WT)	ΔE(*carC*$^{-}$)
10	-0.010	-0.010
100	-0.010	0.037
1000	0.173	0.275
10000	0.391	0.483
50000	0.225	0.346

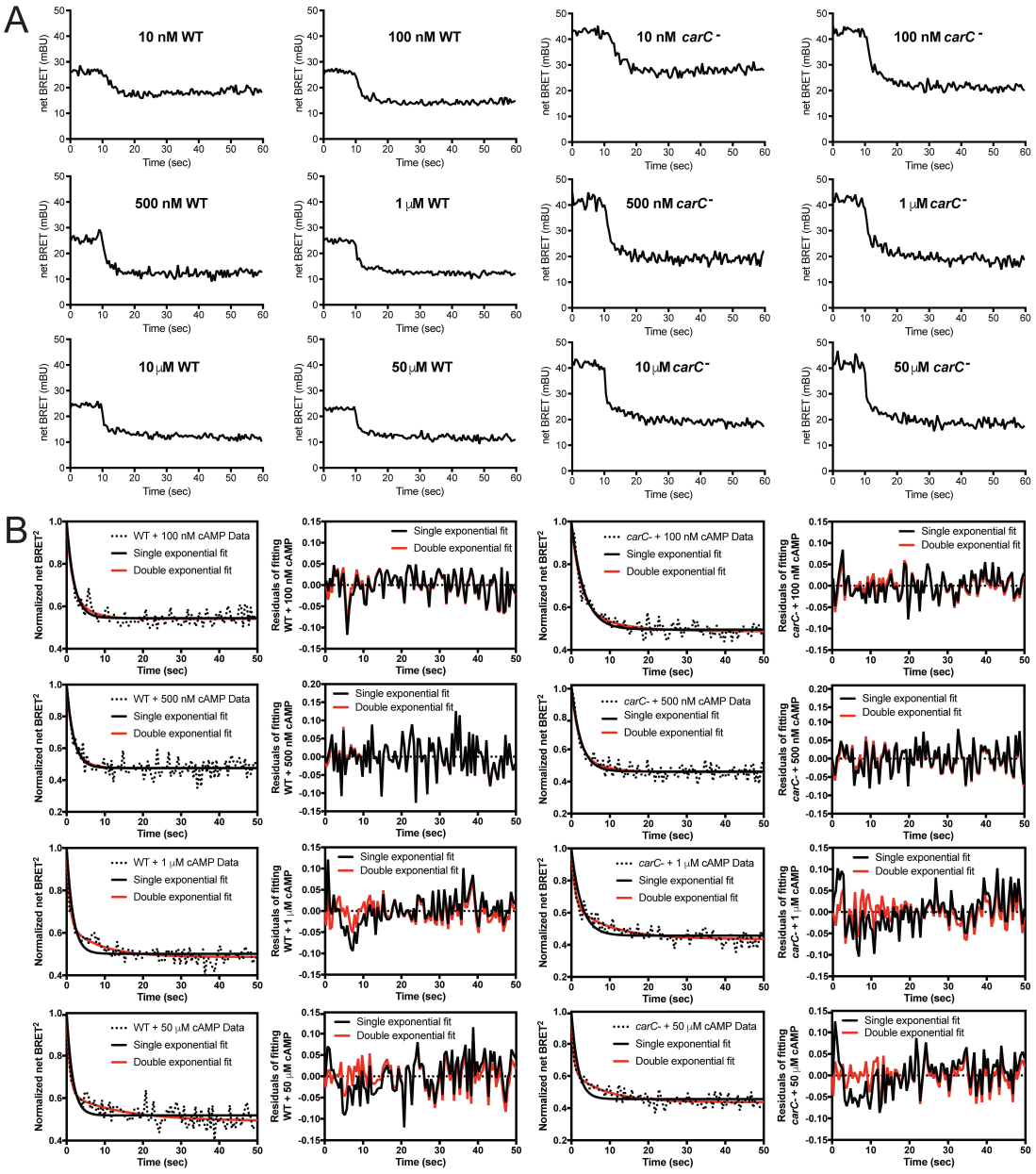


Figure A.1: net BRET^2 data and additional kinetics analyses of cAMP-induced Gα2$\beta\gamma$ subunit dissociation in WT and *carC* null cells used to generate the data presented in main Figure 2.1. A, net BRET^2 data for measurements taken every 0.5 s before and after stimulation with the indicated cAMP concentrations. B, net BRET^2 data normalized to levels before stimulation and the curves were fitted using single (black) and double (red) exponential functions. The residuals analysis for each curve fitting is included to the right of corresponding time curves. Results represent the analyses of at least three independent experiments. mBU, milli BRET units.

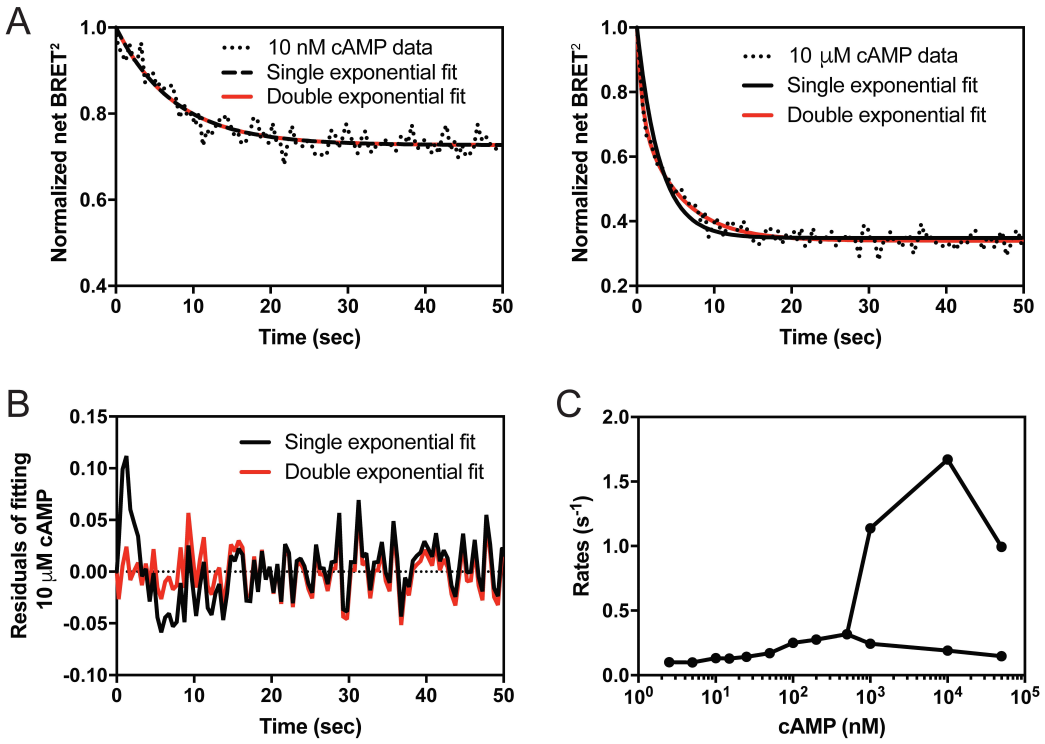


Figure A.2: net BRET^2 data and kinetics analyses of cAMP-induced Gα2$\beta\gamma$ subunit dissociation in Gα2GFP^2/Rluc2Gβ- expressing gα2 null cells. A, net BRET^2 measurements taken every 0.5 s after stimulation with 10 nM or 10 μM cAMP, normalized to BRET^2 levels before stimulation. The curves were fitted using both single and double exponential functions. For the 10 nM data, the fits are identical and on top of each other. B, Residuals for single and double exponential fittings for the 10 μM cAMP-stimulated conditions shown in A were calculated by subtracting the fitted values from the measured values. C, Gα2$\beta\gamma$ subunit dissociation rates determined from the time curve fittings of data obtained for 2.5 nM, 5 nM, 10 nM, 12.5 nM, 25 nM, 50 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, 10 μM, and 50 μM cAMP stimulations. Results represent the analyses of at least three independent experiments.

Appendix B

Appendix to Chapter 4

B.1 Derivation of Equation 4.1

According to the chain rule of differentiating a function with multiple variables,

$$\frac{\mathrm{dProb}}{\mathrm{d[S]}} = \frac{\partial \mathrm{Prob}}{\partial \mathrm{[S]}} + \frac{\partial \mathrm{Prob}}{\partial \mathrm{[G]}}\frac{\mathrm{d[G]}}{\mathrm{d[S]}} + \frac{\partial \mathrm{Prob}}{\partial \mathrm{[R]}}\frac{\mathrm{d[R]}}{\mathrm{d[S]}} \tag{B.1}$$

$$\frac{\mathrm{d[G]}}{\mathrm{d[S]}} = \frac{\partial \mathrm{[G]}}{\partial \mathrm{[S]}} + \frac{\partial \mathrm{[G]}}{\partial \mathrm{[R]}}\frac{\mathrm{d[R]}}{\mathrm{d[S]}} + \frac{\partial \mathrm{[G]}}{\partial \mathrm{Prob}}\frac{\mathrm{dProb}}{\mathrm{d[S]}} \tag{B.2}$$

$$\frac{\mathrm{dProb}}{\mathrm{d[R]}} = \frac{\partial \mathrm{Prob}}{\partial \mathrm{[R]}} + \frac{\partial \mathrm{Prob}}{\partial \mathrm{[G]}}\frac{\mathrm{d[G]}}{\mathrm{d[R]}} \tag{B.3}$$

$$\frac{\mathrm{d[G]}}{\mathrm{d[R]}} = \frac{\partial \mathrm{[G]}}{\partial \mathrm{[R]}} + \frac{\partial \mathrm{[G]}}{\partial \mathrm{Prob}}\frac{\mathrm{dProb}}{\mathrm{d[R]}} \tag{B.4}$$

Substitute $\frac{\mathrm{d[G]}}{\mathrm{d[S]}}$ in equation B.1 with equation B.2 and we get:

$$\frac{\mathrm{dProb}}{\mathrm{d[S]}} = \frac{\partial \mathrm{Prob}}{\partial \mathrm{[S]}} + \frac{\partial \mathrm{Prob}}{\partial \mathrm{[G]}}\left(\frac{\partial \mathrm{[G]}}{\partial \mathrm{[S]}} + \frac{\partial \mathrm{[G]}}{\partial \mathrm{[R]}}\frac{\mathrm{d[R]}}{\mathrm{d[S]}} + \frac{\partial \mathrm{[G]}}{\partial \mathrm{Prob}}\frac{\mathrm{dProb}}{\mathrm{d[S]}}\right) + \frac{\partial \mathrm{Prob}}{\partial \mathrm{[R]}}\frac{\mathrm{d[R]}}{\mathrm{d[S]}}$$

After rearrangement, we can get:

$$\left(1-\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial \mathrm{Prob}}\right)\frac{\mathrm{dProb}}{\mathrm{d[S]}}=\frac{\partial \mathrm{Prob}}{\partial[\mathrm{S}]}+\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial[\mathrm{S}]}+\left(\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial[\mathrm{R}]}+\frac{\partial \mathrm{Prob}}{\partial[\mathrm{R}]}\right)\frac{\mathrm{d[R]}}{\mathrm{d[S]}} \tag{B.5}$$

Substituting $\frac{\mathrm{d[G]}}{\mathrm{d[R]}}$ in equation B.3 with equation B.4 results in:

$$\left(1-\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial \mathrm{Prob}}\right)\frac{\mathrm{dProb}}{\mathrm{d[R]}}=\frac{\partial \mathrm{Prob}}{\partial[\mathrm{R}]}+\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial[\mathrm{R}]} \tag{B.6}$$

Then we substitute $\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial[\mathrm{R}]}+\frac{\partial \mathrm{Prob}}{\partial[\mathrm{R}]}$ in the third term of the right-hand-side of equation B.5 with equation B.6 and get equation 4.1 in the main text.

B.2 Calculation of Mean First Passage Time (MFPT)

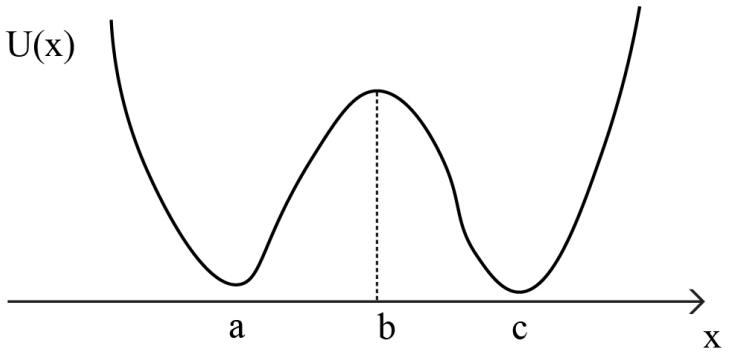


Figure B.1: Sketch of the double-well potential U(x).

The diffusion process in an external potential U(x) can be described by the quasilinear Fokker-Planck equation:

$$\frac{\partial \mathrm{P(x,t)}}{\partial \mathrm{t}}=\frac{\partial}{\partial \mathrm{x}}\mathrm{U'(x)P}+\sigma^2\frac{\partial^2 \mathrm{P}}{\partial \mathrm{x}^2}$$

If the potential U(x) is double well as shown in Figure B.1, the MFPT from point “a” to “c” is

[161]:

$$\tau_{ca} = \frac{1}{\sigma^2}\int_a^c e^{\frac{U(x')}{\sigma^2}} dx' \int_{-\infty}^{x'} e^{-\frac{U(x'')}{\sigma^2}} dx''$$

If the shape of U(x) satisfies that $e^{U(x')/\sigma^2}$ is large when x' is near b and otherwise exponentially smaller, it can be replaced with $\exp\left[\frac{U(b)}{\sigma^2} - \frac{|U''(b)|}{2\sigma^2}(x'-b)^2\right]$ in the first integral. Similarly, in the second integral, as $x' \approx b$, the main contribution comes from the neighborhood $x'' \approx a$ and may be approximated with:

$$\int_{-\infty}^{\infty} \exp\left[-\frac{U(a)}{\sigma^2} - \frac{U''(a)}{2\sigma^2}(x''-a)^2\right] dx'' = \sqrt{\frac{2\pi\sigma^2}{U''(a)}} \exp\left[-\frac{U(a)}{\sigma^2}\right]$$

So, the MFPT $\tau_{ca} \approx \frac{2\pi}{\sqrt{U''(a)|U''(b)|}} \exp\left[\frac{U(b)-U(a)}{\sigma^2}\right]$. This is the Kramers approximation.

If the shape of U(x) does not satisfy the above prerequisite, we numerically calculate the double integrals for MFPT. The time cost for numerical integrations is large if we repeatedly implement this for every set of parameters and every state transition (total number is 64×64). We can make significant savings by applying the Kramers approximation where appropriate, and where the prerequisite for the approximation is not satisfied, by pre-computing a table of these rates numerically. We find that the MFPT only directly depends on $[R]^s$, P^{tot} and y (a function of $[R]^s$, [S] and $[G]^s$). The superscript s here means the steady state. So, we make a table of the numerical computation results of the double integrals on a three-dimensional $25 \times 200 \times 200$ grid of $[R]^s$, P^{tot} and y and use linear interpolation to get the MFPTs when needed.

B.3 Fitting procedure

Our model contains 10 parameters to be fitted with two of them being PA and PI which only are relevant in light-treatment condition. We use simulated annealing as our method of fitting. During the fitting, this method accepts a trial step with some probability dependent on an

artificial temperature T, even when this step does not improve the fit. This can help avoid being trapped in a local minimum [84]. The initial temperature is usually high to have a larger searching area and then, the temperature is gradually decreased, leading to more selective sampling towards the error decreasing direction. We use the simulannealbnd function in MATLAB (R2015b, The MathWorks, Natick, MA) with default settings.

Even with the simulated annealing method, it is still not guaranteed that the samplings are not trapped in the local minima. So, we run three rounds of fitting for each network. In the first round, the error function we use is:

$$\text{error}_1 = \frac{1}{1000}\left(\frac{\left(\text{NP}^{\text{sim}}_{\text{wild}} - \text{NP}^{\text{exp}}_{\text{wild}}\right)^2}{0.2^2} + \frac{\left(\text{NP}^{\text{sim}}_{\text{Rec}^{\text{DN}}} - \text{NP}^{\text{sim}}_{\text{wild}} - \left(\text{NP}^{\text{exp}}_{\text{Rec}^{\text{DN}}} - \text{NP}^{\text{exp}}_{\text{wild}}\right)\right)^2}{0.1^2} + \text{H}\left(0.1 - \text{Dir}^{\text{sim}}_{\text{wild}}\right)\frac{\left(\text{Dir}^{\text{sim}}_{\text{wild}} - \text{Dir}^{\text{exp}}_{\text{wild}}\right)^2}{0.01^2}\right)$$

Here, H is the Heaviside step function ensuring that only a directionality that is not significantly positive (smaller than 0.1) increases the error function. Specifically, when $\text{Dir}^{\text{sim}}_{\text{wild}} > 0.1$, $\text{Dir}^{\text{sim}}_{\text{wild}}$s value is not important at all as the whole term is zero and when $\text{Dir}^{\text{sim}}_{\text{wild}} < 0.1$, the term $(\text{Dir}^{\text{sim}}_{\text{wild}} - \text{Dir}^{\text{exp}}_{\text{wild}})^2$ gives a bias towards larger $\text{Dir}^{\text{sim}}_{\text{wild}}$ during the fitting. As there are no data for light-treatment experiment in the first round, we only fit for the eight parameters except for PA and PI. We randomly choose 100 starting points in the parameter space and run 100 fittings with $\text{T}_{\text{initial}} = 100$. Then we pick the first five fitting results with the smallest error functions and then use them as the starting points for the next round of fitting. The error function used in the second round includes the data for light-treatment conditions:

$$\text{error}_2 = \text{error}_1 + \frac{1}{1000}\left(\frac{\left(\text{NP}^{\text{sim}}_{\text{PA}_{\text{back}}} - \text{NP}^{\text{exp}}_{\text{PA}_{\text{back}}}\right)^2}{0.3^2} + \frac{\left(\text{NP}^{\text{sim}}_{\text{PA}_{\text{RecDN}}} - \text{NP}^{\text{exp}}_{\text{PA}_{\text{RecDN}}}\right)^2}{0.4^2}\right.$$

$$+\frac{\left(\mathrm{NP}_{\mathrm{PI}}^{\mathrm{sim}}-\mathrm{NP}_{\mathrm{wild}}^{\mathrm{sim}}-\left(\mathrm{NP}_{\mathrm{PI}}^{\mathrm{exp}}-\mathrm{NP}_{\mathrm{wild}}^{\mathrm{exp}}\right)\right)^2}{0.1^2}$$
$$+\mathrm{H}\left(\mathrm{Dir}_{\mathrm{PA}_{\mathrm{back}}}^{\mathrm{sim}}+0.1\right)\frac{\left(\mathrm{Dir}_{\mathrm{PA}_{\mathrm{back}}}^{\mathrm{sim}}-\mathrm{Dir}_{\mathrm{PA}_{\mathrm{back}}}^{\mathrm{sim}}\right)^2}{0.01^2}\Bigg)$$

In the second round, we keep the eight parameters in the first round constant and only fit for PA and PI and $\mathrm{T}_{\mathrm{initial}} = 1$. Then we put all the ten parameters together for the third round of fitting with the five results from the second round as the new starting points. Then, $\mathrm{error}_3 = 100 \times \mathrm{error}_2$ and $\mathrm{T}_{\mathrm{initial}} = \mathrm{error}_3(\mathrm{startingpoint})$ for the third round and among the five results, we choose the one with the smallest error function as the final fitting result. The fitting results are listed in Table B.1.

B.4 Other sampling results

In Figures B.2, B.3 and B.4, we present additional sampling results. Specifically, Figure B.2 and Figure B.3 show the results for networks with five interactions while Figure B.4 shows the sampling outcomes for networks with six interactions. Note that all networks in Figures B.2 and B.3 can be excluded even though the network in the black box in Figure B.3 appears to work well for both experiments. But in fact, this network should be excluded. In this network, R is only positively dependent on S and R is the only entrance of the external signal. This means that $\frac{\mathrm{dProb}}{\mathrm{d[S]}} = \frac{\mathrm{dProb}}{\mathrm{d[R]}}\frac{\mathrm{d[R]}}{\mathrm{d[S]}}$ and $\frac{\mathrm{d[R]}}{\mathrm{d[S]}} > 0$. So the requirement of the Receptor$^{\mathrm{DN}}$ experiments that $\frac{\mathrm{dProb}}{\mathrm{d[S]}} < 0$ is equivalent to $\frac{\mathrm{dProb}}{\mathrm{d[R]}} < 0$, which is contradictory to the prerequisite $\frac{\mathrm{dProb}}{\mathrm{d[R]}} > 0$ based on other experiments. For the six possible networks that we have selected, there is not this problem. Because in these six networks, R is not the only entrance of S for the network so that P's dependence on R is not necessarily equivalent to P's dependence on S. So, for these networks it is possible that $\frac{\mathrm{dProb}}{\mathrm{d[S]}} < 0$ while $\frac{\mathrm{dProb}}{\mathrm{d[R]}} > 0$.

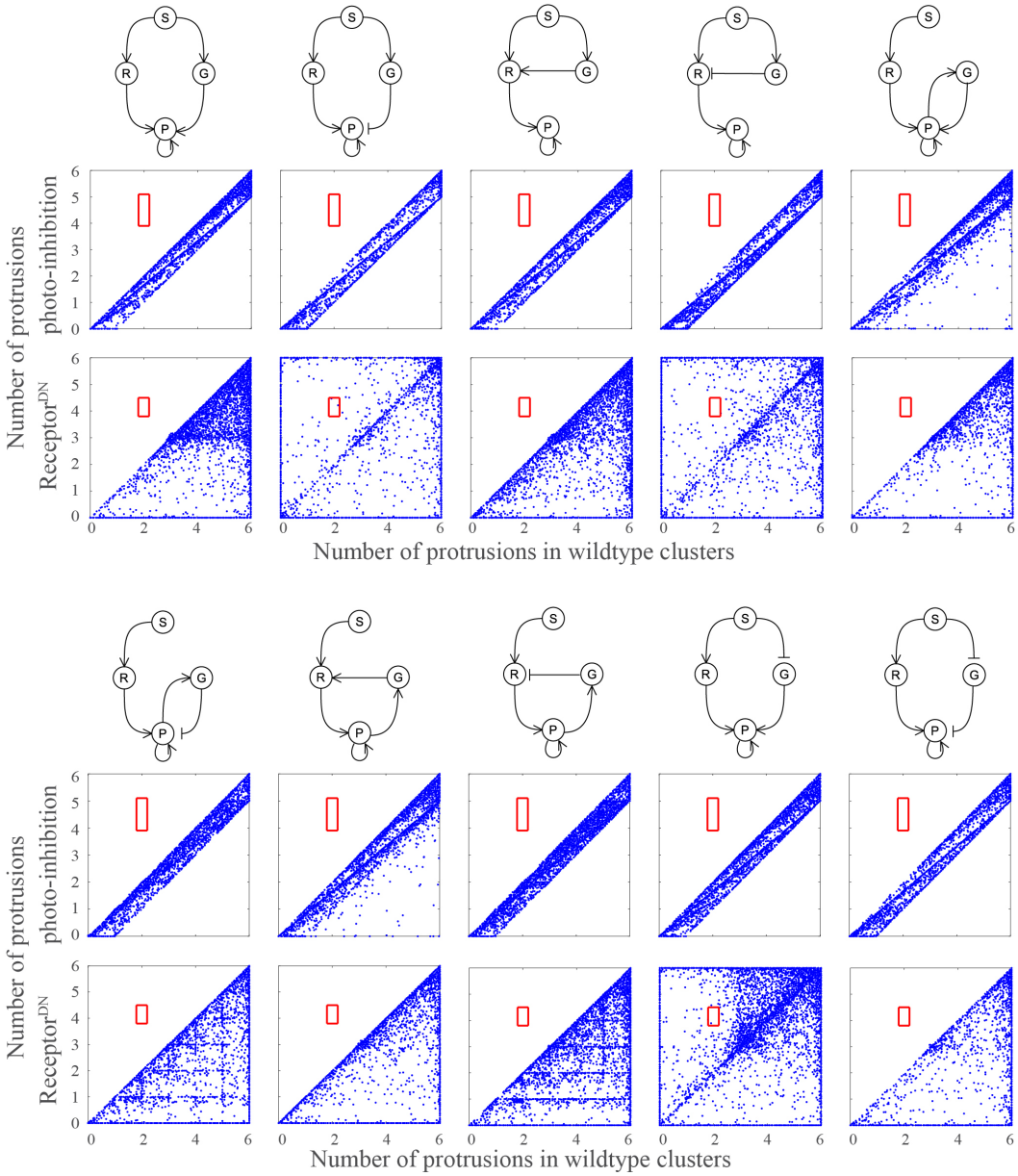


Figure B.2: The first ten sampling results for networks with five interactions.The other twelve are in Figure B.3.

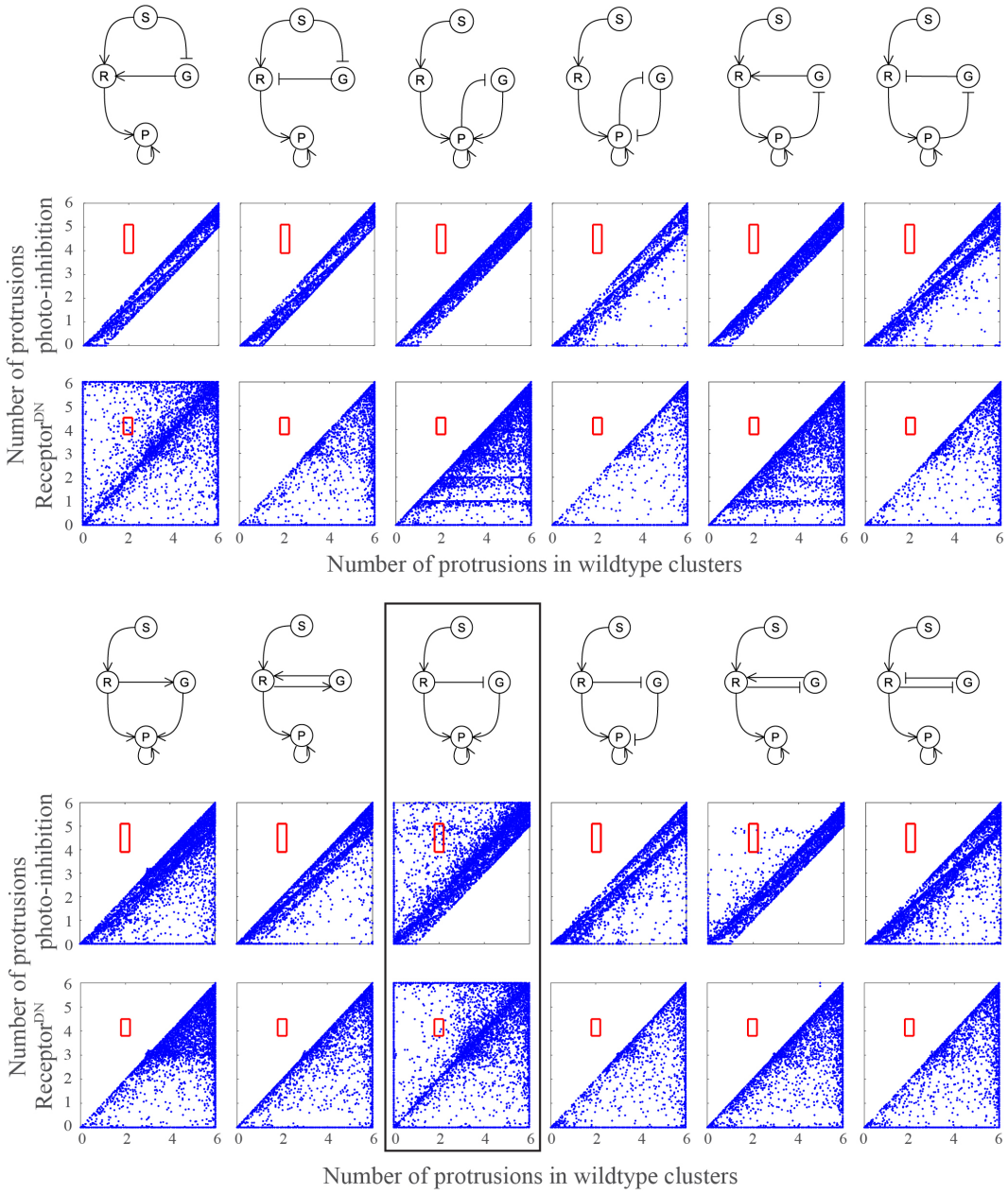


Figure B.3: The other twelve sampling results for networks with five interactions. The first ten are in Figure B.2. It's worth noting that the network in the black box appears to fit both experiments. However, it can be ruled out when considering the Receptor^{DN} experiments and the dProb/d[R] > 0 prerequisite together. The details are in Section B.4

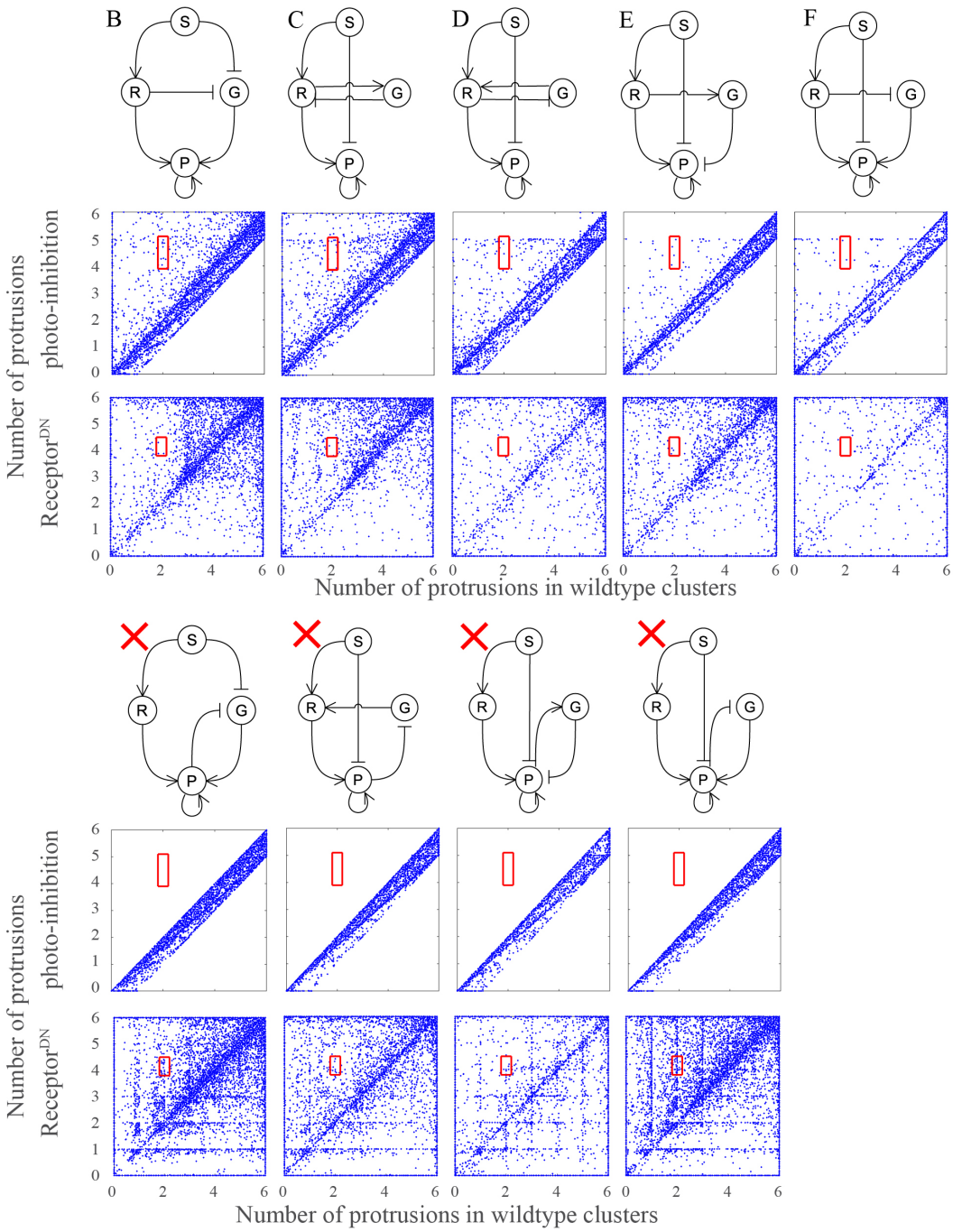


Figure B.4: The sampling results for networks with six interactions. The upper row corresponds to the networks B-F of Figure 4.2c that are not excluded by the qualitative selection. The lower row shows examples of networks that are not consistent with the experimental data.

Table B.1: Parameter fitting results for six-link networks

	Network A	Network B	Network C	Network D	Network E	Network F
S_{grad}	0.4109	0.4814	0.8746	0.5147	0.1590	0.2290
S_{mean}	3.982	0.01205	0.1686	0.06754	0.02587	0.01093
$basal_{AR}$	2.064	0.04961	0.5324	0.01124	0.7171	0.02782
k_{RP}	52.40	9.356	3.825	6.417	0.2470	9.535
P^{tot}	0.8892	2.411	7.475	5.097	96.32	5.223
k_1	k_{SG}	k_{-SG}	k_{RG}	k_{-RG}	k_{RG}	k_{-RG}
	0.1228	0.05322	47.81	2.897	62.25	77.79
k_2	k_{RG}	k_{-RG}	k_{-GR}	k_{GR}	k_{-GP}	k_{GP}
	1.490	92.65	0.8540	0.09783	0.1259	0.7983
k_3	k_{-GP}	k_{GP}	k_{-SP}	k_{-SP}	k_{-SP}	k_{-SP}
	2.125	5.101	1.159	8.702	1.229	10.22
PI	1034	4.939	7561	4580	0.05685	27.14
PA	0.2777	0.02406	0.05794	0.1339	0.004729	0.01876
σ	0.1(not fitted but fixed)					

Appendix C

Appendix to Chapter 5

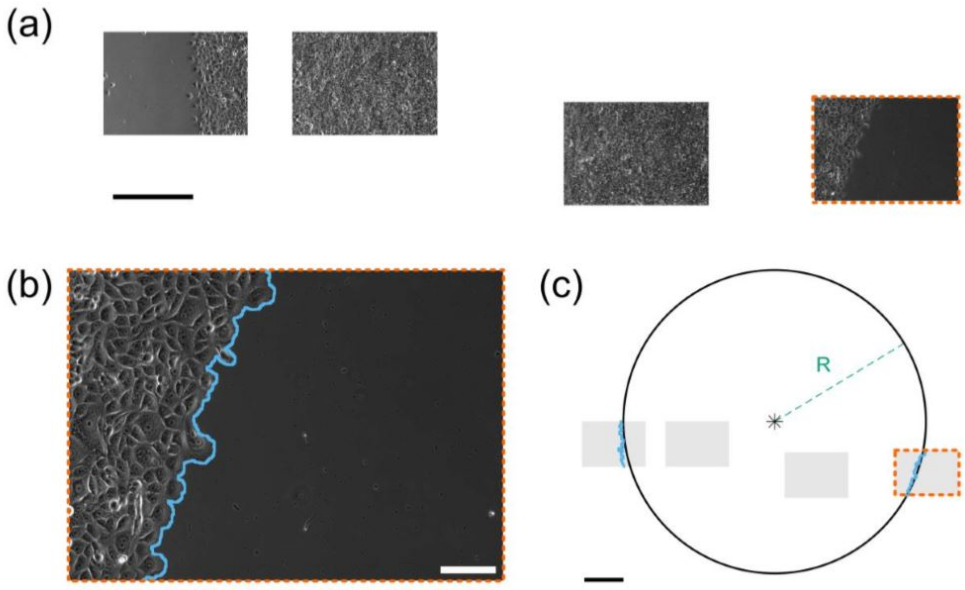


Figure C.1: Finding the monolayer center. Imaged fields of view from a single monolayer are shown to scale (a, scale bar 500 μm). The image outlined with a dashed orange line is magnified and overlaid with the detected edge of the monolayer (b, scale bar 100 μm). In combination with the imaging positions, the detected edge is used to fit the monolayer to a circle of radius R, indicated by the dashed green line (c, scale bar 500 μm). The dashed orange line in (a-c) shows the same region.

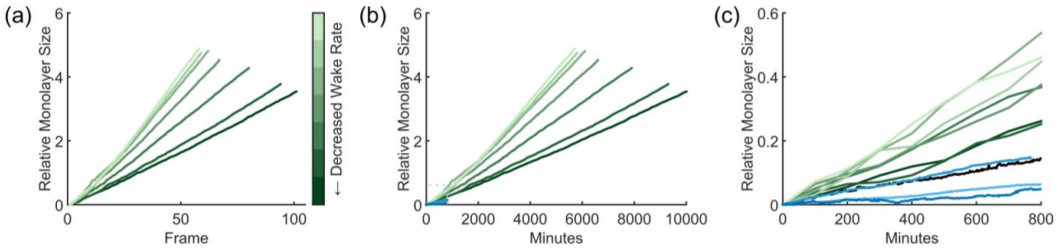


Figure C.2: Comparison of growth rate in the simulation and in the experiments. In all subfigures, monolayer size is plotted with respect to the initial frame ($R_{relative}(t) = (R(t) - R(0))/R(0)$). The simulated monolayers (a) grow linearly as time increases. Each frame of the experiment is two minutes long; the frames of the simulation are recorded every $10t_0$, which can be converted to a time scale of approximately 100 min (see Methods Section). The monolayer growth in both experiment and simulations is shown in (b); the small region indicated by a dashed gray line is enlarged in (c). Wake rate parameter values in all subparts are 0.03, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5 from dark green to light green.

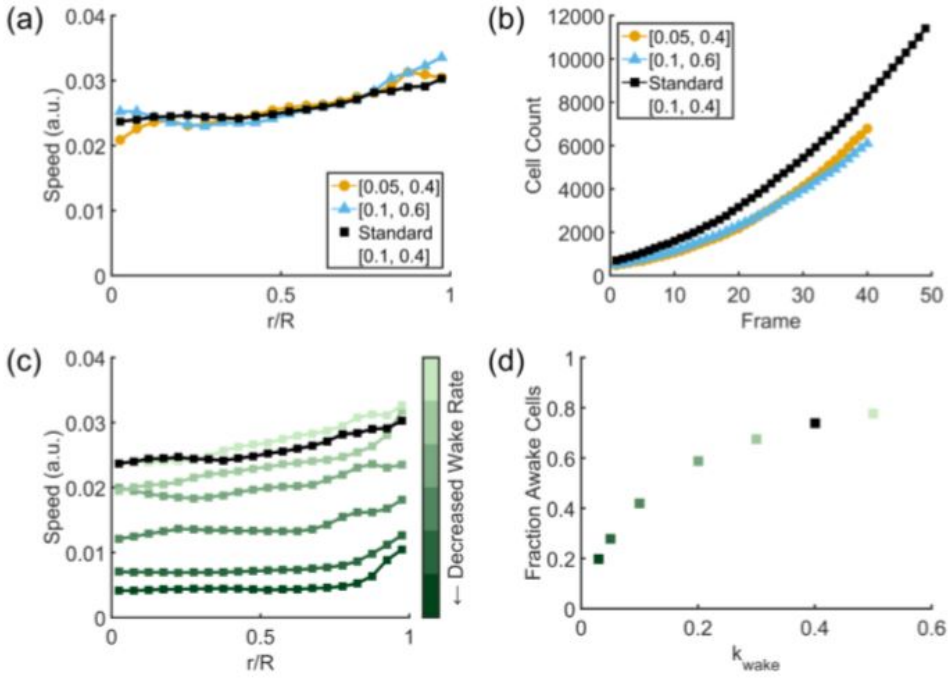


Figure C.3: In (a), legend pairs indicate [division rate, division radius] parameter values. Changing the proliferation rate has no clear effect on speed. Cell growth for these curves is shown in (b). Increasing the rate at which non-motile cells 'wake' increases the overall mean speed while maintaining a similarly shaped curve (c). Wake rate parameter values (k_{wake}) in (c) are 0.03, 0.05, 0.1, 0.2, 0.3, and 0.5 from dark green to light green with the standard parameter value of 0.4 shown in black. The fraction of the cell sheet that is awake at any given time is shown in (d) for each of these parameter values.

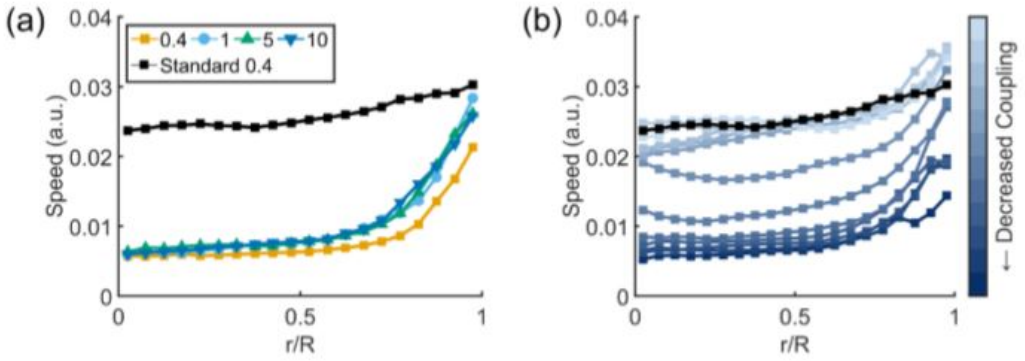


Figure C.4: If velocity motility coupling is removed (a), changing the wake rate does not change the mean speed (wake rate parameter, k_{wake}, value is shown in legend); the standard parameter set with coupling implemented and $k_{wake} = 0.4$ is shown in black. Speed increases as the coupling is increased (b); from dark blue to light blue the coupling parameter, k_{mv}, is set to 0.1, 0.05, 0.04, 0.035, 0.03, 0.025, 0.02, 0.015, 0.01, 0.005, 0.001, and 0.0001. The standard value of 0.00001 is shown in black. Smaller values of k_{mv} indicate a stronger coupling strength.

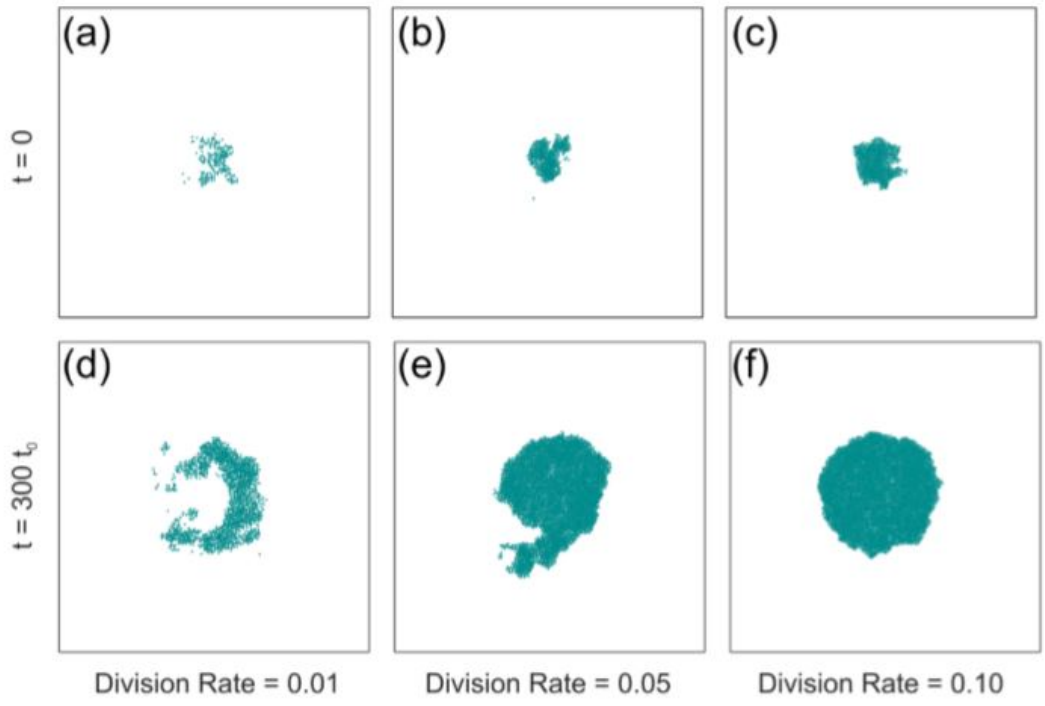


Figure C.5: Snapshots of simulated cell monolayers at t = 0 (a–c) and t = $300t_0$ (d–f). The columns from left to right show increasing division rates with (c,f) corresponding to the standard division rate of 0.1. At the lowest rate of 0.01, the monolayer no longer remains circular or confluent.

Table C.1: Standard Value of Parameters

Parameter	Value	Unit	Description
m	1.2	$p_0 l_0^2$	Magnitude of motility forces
k_1	0.1	t_0^{-1}	Minimal transition rate to non-motile state
k_0	0.3	t_0^{-1}	Maximal transition rate to non-motile state
k_{wake}	0.4	t_0^{-1}	Transition rate to motile state
f_0	2.4	$p_0 l_0^2$	Repulsive cell-cell potential coefficient
f_1	1.0	$p_0 l_0^2$	Attractive cell-cell potential coefficient
R_{cc}	1.0	l_0	Range of pair potentials and dissipative forces
B	5.0	$p_0 l_0^4$	Cellular expansion coefficient
R_{exp}	0.8	l_0	Range of intracellular expansion forces
R_{div}	0.4	l_0	Threshold distance for cell division
r_{div}	0.01	l_0	Distance of new particles after division
k_{div}	0.1	t_0^{-1}	Division rate for cells surpassing size threshold
ξ_{int}	50.0	$p_0 l_0 t_0$	Intracellular dissipation
ξ_{df}	50.0	$p_0 l_0 t_0$	Intercellular dissipation
ξ_B	10.0	$p_0 l_0 t_0$	Background friction coefficient
M	1.0	$p_0 l_0 t_0^2$	Mass of particles
τ	5.0	t_0	Relaxation time for velocity averaging
k_{mv}	10^{-5}	$p_0 l_0^3 t_0^{-1}$	Sharpness of k_{sleep} dependence on alignment
k_{Rm}	0	$p_0^{-1} l_0^{-3}$	Strength of leading cells at the edge

Bibliography

[1] Xiaobo Wang, Li He, Yi I Wu, Klaus M Hahn, and Denise J Montell. Light-mediated activation reveals a key role for rac in collective guidance of cell movement in vivo. *Nature cell biology*, 12(6):591, 2010.

[2] Markus Basan, Jacques Prost, Jean-François Joanny, and Jens Elgeti. Dissipative particle dynamics simulations for biological tissues: rheology and competition. *Physical biology*, 8(2):026014, 2011.

[3] Markus Basan, Jens Elgeti, Edouard Hannezo, Wouter-Jan Rappel, and Herbert Levine. Alignment of cellular motility forces with tissue flow as a mechanism for efficient wound healing. *Proceedings of the National Academy of Sciences*, 110(7):2452–2459, 2013.

[4] Marjon E Kamp, Youtao Liu, and Arjan Kortholt. Function and regulation of heterotrimeric g proteins during chemotaxis. *International journal of molecular sciences*, 17(1):90, 2016.

[5] John ME Nichols, Douwe Veltman, and Robert R Kay. Chemotaxis of a model organism: progress with dictyostelium. *Current opinion in cell biology*, 36:7–12, 2015.

[6] Rex L Chisholm and Richard A Firtel. Insights into morphogenesis from a simple developmental system, 2004.

[7] Ronald L Johnson, PJ Van Haastert, Alan R Kimmel, Charles L Saxe, Bernd Jastorff, and Peter N Devreotes. The cyclic nucleotide specificity of three camp receptors in dictyostelium. *Journal of Biological Chemistry*, 267(7):4600–4607, 1992.

[8] John M Louis, Gail T Ginsburg, and Alan R Kimmel. The camp receptor car4 regulates axial patterning and cellular differentiation during late development of dictyostelium. *Genes & Development*, 8(17):2086–2096, 1994.

[9] Ronald L Johnson, CL 3rd Saxe, Rachel Gollop, Alan R Kimmel, and Peter N Devreotes. Identification and targeted gene disruption of car3, a camp receptor subtype expressed during multicellular stages of dictyostelium development. *Genes & development*, 7(2):273–282, 1993.

[10] Peter S Klein, Tzeli J Sun, Charles L Saxe, Alan R Kimmel, Ronald L Johnson, and Peter N Devreotes. A chemoattractant receptor controls development in dictyostelium discoideum.

Science, 241(4872):1467–1472, 1988.

[11] Charles L Saxe, Ronald L Johnson, Peter N Devreotes, and Alan R Kimmel. Expression of a camp receptor gene of dictyostelium and evidence for a multigene family. *Genes & Development*, 5(1):1–8, 1991.

[12] Robert H Insall, RD Soede, Pauline Schaap, and Peter N Devreotes. Two camp receptors activate common signaling pathways in dictyostelium. *Molecular biology of the cell*, 5(6):703–711, 1994.

[13] RD Soede, Robert H Insall, Peter N Devreotes, and Pauline Schaap. Extracellular camp can restore development in dictyostelium cells lacking one, but not two subtypes of early camp receptors (cars). evidence for involvement of car1 in aggregative gene expression. *Development*, 120(7):1997–2002, 1994.

[14] M Pupillo, P Klein, R Vaughan, G Pitt, P Lilly, T Sun, P Devreotes, A Kumagai, and R Firtel. camp receptor and g-protein interactions control development in dictyostelium. In *Cold Spring Harbor symposia on quantitative biology*, volume 53, pages 657–665. Cold Spring Harbor Laboratory Press, 1988.

[15] Ning Zhang, Yu Long, and Peter N Devreotes. Gγ in dictyostelium: its role in localization of g$\beta\gamma$ to the membrane is required for chemotaxis in shallow gradients. *Molecular biology of the cell*, 12(10):3204–3213, 2001.

[16] A Kumagai, M Pupillo, R Gundersen, R Miake-Lye, PN Devreotes, and RA Firtel. Regulation and function of gα protein subunits in dictyostelium. *Cell*, 57(2):265–275, 1989.

[17] Akiko Kumagai, JA Hadwiger, M Pupillo, and Richard A Firtel. Molecular genetic analysis of two g alpha protein subunits in dictyostelium. *Journal of Biological Chemistry*, 266(2):1220–1228, 1991.

[18] Pamela Lilly, LIJUN Wu, Dennis L Welker, and Peter N Devreotes. A g-protein beta-subunit is essential for dictyostelium development. *Genes & Development*, 7(6):986–995, 1993.

[19] Lijun Wu, Romi Valkema, PJ Van Haastert, and Peter N Devreotes. The g protein beta subunit is essential for multiple responses to chemoattractants in dictyostelium. *The Journal of cell biology*, 129(6):1667–1675, 1995.

[20] Chris Janetopoulos, Tian Jin, and Peter Devreotes. Receptor-mediated activation of heterotrimeric g-proteins in living cells. *Science*, 291(5512):2408–2411, 2001.

[21] AFM Tariqul Islam, Branden M Stepanski, and Pascale G Charest. Studying chemoattractant signal transduction dynamics in dictyostelium by bret. *Chemotaxis: Methods and Protocols*, pages 63–77, 2016.

[22] Fadi F Hamdan, Yann Percherancier, Billy Breton, and Michel Bouvier. Monitoring protein-protein interactions in living cells by bioluminescence resonance energy transfer (bret). *Current protocols in neuroscience*, pages 5–23, 2006.

[23] Thomas Machleidt, Carolyn C Woodroofe, Marie K Schwinn, Jacqui Mendez, Matthew B Robers, Kris Zimmerman, Paul Otto, Danette L Daniels, Thomas A Kirkland, and Keith V Wood. Nanobret a novel bret platform for the analysis of protein–protein interactions. *ACS chemical biology*, 10(8):1797–1804, 2015.

[24] Andreas Markus Loening, Timothy David Fenn, Anna M Wu, and Sanjiv Sam Gambhir. Consensus guided mutagenesis of renilla luciferase yields enhanced stability and light output. *Protein Engineering, Design and Selection*, 19(9):391–400, 2006.

[25] Lucie Bertrand, Stéphane Parent, Mireille Caron, Mireille Legault, Erik Joly, Stéphane Angers, Michel Bouvier, Mike Brown, Benoit Houle, and Luc Ménard. The bret2/arrestin assay in stable recombinant cells: a platform to screen for compounds that interact with g protein-coupled receptors (gpcrs). *Journal of Receptors and Signal Transduction*, 22(1-4):533–541, 2002.

[26] Kosuke Takeda, Danying Shao, Micha Adler, Pascale G Charest, William F Loomis, Herbert Levine, Alex Groisman, Wouter-Jan Rappel, and Richard A Firtel. Incoherent feedforward control governs adaptation of activated ras in a eukaryotic chemotaxis pathway. *Science Signaling*, 5(205):ra2–ra2, 2012.

[27] Monica Skoge, Haicen Yue, Michael Erickstad, Albert Bae, Herbert Levine, Alex Groisman, William F Loomis, and Wouter-Jan Rappel. Cellular memory in eukaryotic chemotaxis. *Proceedings of the National Academy of Sciences*, 111(40):14448–14453, 2014.

[28] Céline Galés, R Victor Rebois, Mireille Hogue, Phan Trieu, Andreas Breit, Terence E Hébert, and Michel Bouvier. Real-time monitoring of receptor and g-protein interactions in living cells. *Nature methods*, 2(3):177, 2005.

[29] Céline Galés, Joost JJ Van Durm, Stéphane Schaak, Stéphanie Pontier, Yann Percherancier, Martin Audet, Hervé Paris, and Michel Bouvier. Probing the activation-promoted structural rearrangements in preassembled receptor–g protein complexes. *Nature Structural and Molecular Biology*, 13(9):778, 2006.

[30] Matthew R Whorton, Michael P Bokoch, Søren GF Rasmussen, Bo Huang, Richard N Zare, Brian Kobilka, and Roger K Sunahara. A monomeric g protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its g protein. *Proceedings of the National Academy of Sciences*, 104(18):7682–7687, 2007.

[31] Muriel Nobles, Amy Benians, and Andrew Tinker. Heterotrimeric g proteins precouple with g protein-coupled receptors in living cells. *Proceedings of the National Academy of Sciences*, 102(51):18706–18711, 2005.

[32] Finly Philip, Parijat Sengupta, and Suzanne Scarlata. Signaling through a g protein-coupled receptor and its corresponding g protein follows a stoichiometrically limited model. *Journal of Biological Chemistry*, 282(26):19203–19216, 2007.

[33] Pim MW Janssens, Peter LJ van der Geer, Jos C Arents, and Roel van Driel. Guanine nucleotides modulate the function of chemotactic cyclic amp receptors indictyostelium discoideum. *Molecular and cellular biochemistry*, 67(2):119–124, 1985.

[34] Paul W Schenk, Saskia Van Es, Fanja Kesbeke, and B Ewa Snaar-Jagalska. Involvement of cyclic amp cell surface receptors and g-proteins in signal transduction during slug migration of dictyostelium discoideum. *Developmental biology*, 145(1):110–118, 1991.

[35] Xuehua Xu, Tobias Meckel, Joseph A Brzostowski, Jianshe Yan, Martin Meier-Schellersheim, and Tian Jin. Coupling mechanism of a gpcr and a heterotrimeric g protein during chemoattractant gradient sensing in dictyostelium. *Science Signaling*, 3(141):ra71–ra71, 2010.

[36] Freek van Hemert, Milena D Lazova, B Ewa Snaar-Jagaska, and Thomas Schmidt. Mobility of g proteins is heterogeneous and polarized during chemotaxis. *Journal of Cell Science*, 123(17):2922–2930, 2010.

[37] Carrie A Elzie, Jennifer Colby, Morgan A Sammons, and Chris Janetopoulos. Dynamic localization of g proteins in dictyostelium discoideum. *Journal of cell science*, 122(15):2597–2603, 2009.

[38] Xuehua Xu, Martin Meier-Schellersheim, Xuanmao Jiao, Lauren E Nelson, and Tian Jin. Quantitative imaging of single live cells reveals spatiotemporal dynamics of multistep signaling events of chemoattractant gradient sensing in dictyostelium. *Molecular biology of the cell*, 16(2):676–688, 2005.

[39] Yoichiro Kamimura, Yukihiro Miyanaga, and Masahiro Ueda. Heterotrimeric g-protein shuttling via gip1 extends the dynamic range of eukaryotic chemotaxis. *Proceedings of the National Academy of Sciences*, 113(16):4356–4361, 2016.

[40] Paul S-H Park, David T Lodowski, and Krzysztof Palczewski. Activation of g protein–coupled receptors: beyond two-state models and tertiary conformational changes. *Annu. Rev. Pharmacol. Toxicol.*, 48:107–141, 2008.

[41] Sergi Ferré, Vicent Casadó, Lakshmi A Devi, Marta Filizola, Ralf Jockers, Martin J Lohse, Graeme Milligan, Jean-Philippe Pin, and Xavier Guitart. G protein–coupled receptor oligomerization revisited: functional and pharmacological perspectives. *Pharmacological reviews*, 66(2):413–434, 2014.

[42] Stefan Gahbauer and Rainer A Böckmann. Membrane-mediated oligomerization of g protein coupled receptors and its implications for gpcr function. *Frontiers in physiology*, 7:494, 2016.

[43] Anne J Ridley, Martin A Schwartz, Keith Burridge, Richard A Firtel, Mark H Ginsberg, Gary Borisy, J Thomas Parsons, and Alan Rick Horwitz. Cell migration: integrating signals from front to back, 2003.

[44] Dirk Dormann and Cornelis J Weijer. Chemotactic cell movement during development, 2003.

[45] Marc Tessier-Lavigne and Corey S Goodman. The molecular biology of axon guidance, 1996.

[46] Marco Baggiolini. Chemokines and leukocyte traffic. *Nature*, 392(6676):565, 1998.

[47] Evanthia T Roussos, John S Condeelis, and Antonia Patsialou. Chemotaxis in cancer, 2011.

[48] Peter N Devreotes and Sally H Zigmond. Chemotaxis in eukaryotic cells: a focus on leukocytes and dictyostelium. *Annual review of cell biology*, 4(1):649–686, 1988.

[49] Tian Jin, Xuehua Xu, and Dale Hereld. Chemotaxis, chemokine receptors and human disease, 2008.

[50] Peter JM Van Haastert and Peter N Devreotes. Chemotaxis: signalling the way forward, 2004.

[51] Kristen F Swaney, Chuan-Hsiang Huang, and Peter N Devreotes. Eukaryotic chemotaxis: a network of signaling pathways controls motility, directional sensing, and polarity. *Annual review of biophysics*, 39:265–289, 2010.

[52] Danny Fuller, Wen Chen, Micha Adler, Alex Groisman, Herbert Levine, Wouter-Jan Rappel, and William F Loomis. External and internal constraints on eukaryotic chemotaxis. *Proceedings of the National Academy of Sciences*, 107(21):9656–9659, 2010.

[53] Loling Song, Sharvari M Nadkarni, Hendrik U Bödeker, Carsten Beta, Albert Bae, Carl Franck, Wouter-Jan Rappel, William F Loomis, and Eberhard Bodenschatz. Dictyostelium discoideum chemotaxis: threshold for directed motion. *European journal of cell biology*, 85(9-10):981–989, 2006.

[54] Tim Lämmermann, Philippe V Afonso, Bastian R Angermann, Ji Ming Wang, Wolfgang Kastenmüller, Carole A Parent, and Ronald N Germain. Neutrophil swarms require ltb4 and integrins at sites of cell death in vivo. *Nature*, 498(7454):371, 2013.

[55] Philipp Niethammer, Clemens Grabher, A Thomas Look, and Timothy J Mitchison. A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish. *Nature*, 459(7249):996, 2009.

[56] Raymond E Goldstein. Traveling-wave chemotaxis. *Physical review letters*, 77(4):775, 1996.

[57] Victor Sourjik and Ned S Wingreen. Responding to chemical gradients: bacterial chemotaxis, 2012.

[58] Eric Albrecht and Howard R Petty. Cellular memory: neutrophil orientation reverses during temporally decreasing chemoattractant concentrations. *Proceedings of the National Academy of Sciences*, 95(9):5039–5044, 1998.

[59] Deborah Wessels, John Murray, and David R Soll. Behavior of dictyostelium amoebae is regulated primarily by the temporal dynamic of the natural camp wave. *Cytoskeleton*, 23(2):145–156, 1992.

[60] Jeremy Geiger, Deborah Wessels, and David R Soll. Human polymorphonuclear leukocytes respond to waves of chemoattractant, like dictyostelium. *Cytoskeleton*, 56(1):27–44, 2003.

[61] C Joanne Wang, Adriel Bergmann, Benjamin Lin, Kyuri Kim, and Andre Levchenko. Diverse sensitivity thresholds in dynamic signaling responses by social amoebae. *Science Signaling*, 5(213):ra17–ra17, 2012.

[62] Masatoshi Nishikawa, Marcel Hörning, Masahiro Ueda, and Tatsuo Shibata. *Biophysical journal*, 106(3):723–734, 2014.

[63] Börn Meier, Alejandro Zielinski, Christoph Weber, Delphine Arcizet, Simon Youssef, Thomas Franosch, Joachim O Rädler, and Doris Heinrich. Chemotactic cell trapping in controlled alternating gradient fields. *Proceedings of the National Academy of Sciences*, 108(28):11417–11422, 2011.

[64] Daniel Irimia, Su-Yang Liu, William G Tharp, Azadeh Samadani, Mehmet Toner, and Mark C Poznansky. Microfluidic system for measuring neutrophil migratory responses to fast switches of chemical gradients. *Lab on a Chip*, 6(2):191–198, 2006.

[65] Jérémie Dalous, Emmanuel Burghardt, Annette Müller-Taubenberger, Franz Bruckert, Günther Gerisch, and Till Bretschneider. Reversal of cell polarity and actin-myosin cytoskeleton reorganization under mechanical and chemical stimulation. *Biophysical journal*, 94(3):1063–1074, 2008.

[66] Liang Li, Simon F Nørrelykke, and Edward C Cox. Persistent cell motion in the absence of external signals: a search strategy for eukaryotic cells. *PLoS one*, 3(5):e2093, 2008.

[67] Robert M Cooper, Ned S Wingreen, and Edward C Cox. An excitable cortex and memory model successfully predicts new pseudopod dynamics. *PloS one*, 7(3):e33528, 2012.

[68] Leonard Bosgraaf and Peter JM Van Haastert. The ordered extension of pseudopodia by amoeboid cells in the absence of external cues. *PloS one*, 4(4):e5253, 2009.

[69] Leonard Bosgraaf and Peter JM Van Haastert. Navigation of chemotactic cells by parallel signaling to pseudopod persistence and orientation. *PloS one*, 4(8):e6842, 2009.

[70] Peter N Devreotes, Michael J Potel, and Stephen A MacKay. Quantitative analysis of cyclic amp waves mediating aggregation in dictyostelium discoideum. *Developmental biology*, 96(2):405–415, 1983.

[71] Marten Postma and Peter JM van Haastert. Mathematics of experimentally generated chemoattractant gradients. In *Chemotaxis*, pages 473–488. Springer, 2009.

[72] Thomas Gregor, Koichi Fujimoto, Noritaka Masaki, and Satoshi Sawai. The onset of collective behavior in social amoebae. *Science*, 328(5981):1021–1025, 2010.

[73] Arjan Kortholt, Rama Kataria, Ineke Keizer-Gunnink, Wouter N Van Egmond, Ankita Khanna, and Peter JM Van Haastert. Dictyostelium chemotaxis: essential ras activation and accessory signalling pathways for amplification. *EMBO reports*, 12(12):1273–1279, 2011.

[74] Atsuo T Sasaki, Cheryl Chun, Kosuke Takeda, and Richard A Firtel. Localized ras signaling at the leading edge regulates pi3k, cell polarity, and directional cell movement. *The Journal of Cell Biology*, 167(3):505–518, 2004.

[75] Arjan Kortholt, Ineke Keizer-Gunnink, Rama Kataria, and Peter JM Van Haastert. Ras activation and symmetry breaking during dictyostelium chemotaxis. *Journal of Cell Science*, 126(19):4502–4513, 2013.

[76] Monica Skoge, Micha Adler, Alex Groisman, Herbert Levine, William F Loomis, and Wouter-Jan Rappel. Gradient sensing in defined chemotactic fields. *Integrative Biology*, 2(11-12):659–668, 2010.

[77] Chuan-Hsiang Huang, Ming Tang, Changji Shi, Pablo A Iglesias, and Peter N Devreotes. An excitable signal integrator couples to an idling cytoskeletal oscillator to drive cell migration. *Nature cell biology*, 15(11):1307, 2013.

[78] Douwe M Veltman and Peter JM van Haastert. The role of cgmp and the rear of the cell in dictyostelium chemotaxis and cell streaming. *Journal of cell science*, 121(1):120–127, 2008.

[79] Carole A Parent and Peter N Devreotes. A cell's sense of direction, 1999.

[80] H Levine, WF Loomis, and WJ Rappel. Eukaryotic chemotaxis and its limitations due to stochastic sensing. *Fields Institute Communications, ed Sivaloganathan S (American Mathematical Society, Providence, RI)*, pages 1–20, 2010.

[81] Changji Shi, Chuan-Hsiang Huang, Peter N Devreotes, and Pablo A Iglesias. Interaction of motility, directional sensing, and polarity modules recreates the behaviors of chemotaxing cells. *PLoS computational biology*, 9(7):e1003122, 2013.

[82] Alexandra Jilkine and Leah Edelstein-Keshet. A comparison of mathematical models for

polarization of single eukaryotic cells in response to guided cues, 2011.

[83] Pablo A Iglesias and Andre Levchenko. Modeling the cell's guidance system. *Science Signaling*, 2002(148):re12–re12, 2002.

[84] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, September 2007.

[85] Akihiko Nakajima, Shuji Ishihara, Daisuke Imoto, and Satoshi Sawai. Rectified directional sensing in long-range cell migration. *Nature communications*, 5:5367, 2014.

[86] Mineko Maeda, Sijie Lu, Gad Shaulsky, Yuji Miyazaki, Hidekazu Kuwayama, Yoshimasa Tanaka, Adam Kuspa, and William F Loomis. Periodic signaling controlled by an oscillatory circuit that includes protein kinases erk2 and pka. *Science*, 304(5672):875–878, 2004.

[87] Satoshi Sawai, Peter A Thomason, and Edward C Cox. An autoregulatory circuit for long-range self-organization in dictyostelium cell populations. *Nature*, 433(7023):323, 2005.

[88] Brian A Camley and Wouter-Jan Rappel. Physical models of collective cell motility: from cell to tissue. *Journal of physics D: Applied physics*, 50(11):113002, 2017.

[89] Vincent Hakim and Pascal Silberzan. Collective cell migration: a physics perspective. *Reports on Progress in Physics*, 80(7):076601, 2017.

[90] Cornelis J Weijer. Collective cell migration in development. *Journal of cell science*, 122(18):3215–3223, 2009.

[91] Peter Friedl and Darren Gilmour. Collective cell migration in morphogenesis, regeneration and cancer. *Nature reviews Molecular cell biology*, 10(7):445, 2009.

[92] Pernille Rørth. Collective cell migration. *Annual Review of Cell and Developmental*, 25:407–429, 2009.

[93] Roberto Mayor and Sandrine Etienne-Manneville. The front and rear of collective cell migration. *Nature reviews Molecular cell biology*, 17(2):97, 2016.

[94] Brian A Camley and Wouter-Jan Rappel. Cell-to-cell variation sets a tissue-rheology–dependent bound on collective gradient sensing. *Proceedings of the National Academy of Sciences*, page 201712309, 2017.

[95] Gema Malet-Engra, Weimiao Yu, Amanda Oldani, Javier Rey-Barroso, Nir S Gov, Giorgio Scita, and Loïc Dupré. Collective cell motility promotes chemotactic prowess and resistance to chemorepulsion. *Current Biology*, 25(2):242–250, 2015.

[96] Eric Theveneau, Lorena Marchant, Sei Kuriyama, Mazhar Gull, Barbara Moepps, Maddy Parsons, and Roberto Mayor. Collective chemotaxis requires contact-dependent cell polarity. *Developmental cell*, 19(1):39–53, 2010.

[97] Mathieu Poujade, Erwan Grasland-Mongrain, A Hertzog, J Jouanneau, Philippe Chavrier, Benoît Ladoux, Axel Buguin, and Pascal Silberzan. Collective migration of an epithelial monolayer in response to a model wound. *Proceedings of the National Academy of Sciences*, 104(41):15988–15993, 2007.

[98] Néstor Sepúlveda, Laurence Petitjean, Olivier Cochet, Erwan Grasland-Mongrain, Pascal Silberzan, and Vincent Hakim. Collective cell motion in an epithelial sheet can be quantitatively described by a stochastic interacting particle model. *PLoS computational biology*, 9(3):e1002944, 2013.

[99] Danfeng Cai, Shann-Ching Chen, Mohit Prasad, Li He, Xiaobo Wang, Valerie Choesmel-Cadamuro, Jessica K Sawyer, Gaudenz Danuser, and Denise J Montell. Mechanical feedback through e-cadherin promotes direction sensing during collective cell migration. *Cell*, 157(5):1146–1159, 2014.

[100] Damien Ramel, Xiaobo Wang, Carl Laflamme, Denise J Montell, and Gregory Emery. Rab11 regulates cell–cell communication during collective cell movements. *Nature cell biology*, 15(3):317, 2013.

[101] David P Stonko, Lathiena Manning, Michelle Starz-Gaiano, and Bradford E Peercy. A mathematical model of collective cell migration in a three-dimensional, heterogeneous environment. *PloS one*, 10(4):e0122799, 2015.

[102] M Abercrombie and EJ Ambrose. The surface properties of cancer cells: a review. *Cancer research*, 22(5 Part 1):525–548, 1962.

[103] Carlos Carmona-Fontaine, Helen K Matthews, Sei Kuriyama, Mauricio Moreno, Graham A Dunn, Maddy Parsons, Claudio D Stern, and Roberto Mayor. Contact inhibition of locomotion in vivo controls neural crest directional migration. *Nature*, 456(7224):957, 2008.

[104] Brian A Camley, Juliane Zimmermann, Herbert Levine, and Wouter-Jan Rappel. Emergent collective chemotaxis without single-cell gradient sensing. *Physical review letters*, 116(9):098101, 2016.

[105] Danfeng Cai, Wei Dai, Mohit Prasad, Junjie Luo, Nir S Gov, and Denise J Montell. Modeling and analysis of collective cell migration in an in vivo three-dimensional environment. *Proceedings of the National Academy of Sciences*, 113(15):E2134–E2141, 2016.

[106] David Ellison, Andrew Mugler, Matthew D Brennan, Sung Hoon Lee, Robert J Huebner, Eliah R Shamir, Laura A Woo, Joseph Kim, Patrick Amar, Ilya Nemenman, et al. Cell–cell communication enhances the capacity of cell ensembles to sense shallow gradients during

morphogenesis. *Proceedings of the National Academy of Sciences*, 113(6):E679–E688, 2016.

[107] Andrew Mugler, Andre Levchenko, and Ilya Nemenman. Limits to the precision of gradient sensing with spatial communication and temporal integration. *Proceedings of the National Academy of Sciences*, 113(6):E689–E695, 2016.

[108] Brian A Camley, Juliane Zimmermann, Herbert Levine, and Wouter-Jan Rappel. Collective signal processing in cluster chemotaxis: Roles of adaptation, amplification, and co-attraction in collective guidance. *PLoS computational biology*, 12(7):e1005008, 2016.

[109] Julien Varennes, Bumsoo Han, and Andrew Mugler. Collective chemotaxis through noisy multicellular gradient sensing. *Biophysical journal*, 111(3):640–649, 2016.

[110] Denise J Montell, Wan Hee Yoon, and Michelle Starz-Gaiano. Group choreography: mechanisms orchestrating the collective movement of border cells. *Nature reviews Molecular cell biology*, 13(10):631, 2012.

[111] Denise J Montell. Border-cell migration: the race is on. *Nature reviews Molecular cell biology*, 4(1):13, 2003.

[112] Pernille Rørth. Initiating and guiding migration: lessons from border cells. *Trends in cell biology*, 12(7):325–331, 2002.

[113] Peter Duchek, Kálmán Somogyi, Gáspár Jékely, Simone Beccari, and Pernille Rørth. Guidance of cell migration by the drosophila pdgf/vegf receptor. *Cell*, 107(1):17–26, 2001.

[114] Peter Duchek and Pernille Rørth. Guidance of cell migration by egf receptor signaling during drosophila oogenesis. *Science*, 291(5501):131–133, 2001.

[115] Mikiko Inaki, Smitha Vishnu, Adam Cliffe, and Pernille Rørth. Effective guidance of collective migration based on differences in cell states. *Proceedings of the National Academy of Sciences*, 109(6):2027–2032, 2012.

[116] Pooja R Sonavane, Chong Wang, Bette Dzamba, Gregory F Weber, Ammasi Periasamy, and Douglas W DeSimone. Mechanical and signaling roles for keratin intermediate filaments in the assembly and morphogenesis of xenopus mesendoderm tissue at gastrulation. *Development*, 144(23):4363–4376, 2017.

[117] Elena Scarpa, András Szabó, Anne Bibonne, Eric Theveneau, Maddy Parsons, and Roberto Mayor. Cadherin switch during emt in neural crest cells leads to contact inhibition of locomotion via repolarization of forces. *Developmental cell*, 34(4):421–434, 2015.

[118] Mohit Prasad and Denise J Montell. Cellular and molecular mechanisms of border cell migration analyzed using time-lapse live-cell imaging. *Developmental cell*, 12(6):997–1005, 2007.

[119] Wenzhe Ma, Ala Trusina, Hana El-Samad, Wendell A Lim, and Chao Tang. Defining network topologies that can achieve biochemical adaptation. *Cell*, 138(4):760–773, 2009.

[120] Anne Marie Murphy and Denise J Montell. Cell type-specific roles for cdc42, rac, and rhol in drosophila oogenesis. *The Journal of cell biology*, 133(3):617–630, 1996.

[121] Yi I Wu, Daniel Frey, Oana I Lungu, Angelika Jaehrig, Ilme Schlichting, Brian Kuhlman, and Klaus M Hahn. A genetically encoded photoactivatable rac controls the motility of living cells. *Nature*, 461(7260):104, 2009.

[122] Ambra Bianco, Minna Poukkula, Adam Cliffe, Juliette Mathieu, Carlos M Luque, Tudor A Fulga, and Pernille Rørth. Two distinct modes of guidance signalling during collective migration of border cells. *Nature*, 448(7151):362, 2007.

[123] Jennifer S Liu, Justin T Farlow, Amanda K Paulson, Mark A Labarge, and Zev J Gartner. Programmed cell-to-cell variability in ras activity triggers emergent behaviors during mammary epithelial morphogenesis. *Cell reports*, 2(5):1461–1470, 2012.

[124] Eric Theveneau and Roberto Mayor. Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. *Developmental biology*, 366(1):34–54, 2012.

[125] Paul Martin. Wound healing–aiming for perfect skin regeneration. *Science*, 276(5309):75–81, 1997.

[126] Ester Anon, Xavier Serra-Picamal, Pascal Hersen, Nils C Gauthier, Michael P Sheetz, Xavier Trepat, and Benoît Ladoux. Cell crawling mediates collective cell migration to close undamaged epithelial gaps. *Proceedings of the National Academy of Sciences*, 109(27):10891–10896, 2012.

[127] Kevin J Cheung, Edward Gabrielson, Zena Werb, and Andrew J Ewald. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell*, 155(7):1639–1651, 2013.

[128] Andrew D Doyle, Francis W Wang, Kazue Matsumoto, and Kenneth M Yamada. One-dimensional topography underlies three-dimensional fibrillar cell migration. *The Journal of cell biology*, 184(4):481–490, 2009.

[129] Maurizio Ventre, Carlo Fortunato Natale, Carmela Rianna, and Paolo Antonio Netti. Topographic cell instructive patterns to control cell adhesion, polarization and migration. *Journal of the royal society Interface*, 11(100):20140687, 2014.

[130] Guillaume Charras and Erik Sahai. Physical influences of the extracellular environment on cell migration. *Nature reviews Molecular cell biology*, 15(12):813, 2014.

[131] Jun Allard and Alex Mogilner. Traveling waves in actin dynamics and cell motility. *Current*

opinion in cell biology, 25(1):107–115, 2013.

[132] Paolo Maiuri, Jean-François Rupprecht, Stefan Wieser, Verena Ruprecht, Olivier Bénichou, Nicolas Carpi, Mathieu Coppey, Simon De Beco, Nir Gov, Carl-Philipp Heisenberg, et al. Actin flows mediate a universal coupling between cell speed and cell persistence. *Cell*, 161(2):374–386, 2015.

[133] Laurence Petitjean, Myriam Reffay, Erwan Grasland-Mongrain, Mathieu Poujade, Benoıt Ladoux, Axel Buguin, and Pascal Silberzan. Velocity fields in a collectively migrating epithelium. *Biophysical journal*, 98(9):1790–1800, 2010.

[134] Thomas E Angelini, Edouard Hannezo, Xavier Trepat, Jeffrey J Fredberg, and David A Weitz. Cell migration driven by cooperative substrate deformation patterns. *Physical review letters*, 104(16):168104, 2010.

[135] Xavier Serra-Picamal, Vito Conte, Romaric Vincent, Ester Anon, Dhananjay T Tambe, Elsa Bazellieres, James P Butler, Jeffrey J Fredberg, and Xavier Trepat. Mechanical waves during tissue expansion. *Nature Physics*, 8(8):628, 2012.

[136] Michael C Weiger, Vidya Vedham, Christina H Stuelten, Karen Shou, Mark Herrera, Misako Sato, Wolfgang Losert, and Carole A Parent. Real-time motion analysis reveals cell directionality as an indicator of breast cancer progression. *PLoS One*, 8(3):e58859, 2013.

[137] M Cristina Marchetti, Jean-François Joanny, Sriram Ramaswamy, Tanniemola B Liverpool, Jacques Prost, Madan Rao, and R Aditi Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85(3):1143, 2013.

[138] Michael F Hagan and Aparna Baskaran. Emergent self-organization in active materials. *Current opinion in cell biology*, 38:74–80, 2016.

[139] Tamás Vicsek and Anna Zafeiris. Collective motion. *Physics Reports*, 517(3-4):71–140, 2012.

[140] Jakob Löber, Falko Ziebert, and Igor S Aranson. Collisions of deformable cells lead to collective migration. *Scientific reports*, 5:9172, 2015.

[141] François Graner and James A Glazier. Simulation of biological cell sorting using a two-dimensional extended potts model. *Physical review letters*, 69(13):2013, 1992.

[142] Brian A Camley and Wouter-Jan Rappel. Physical models of collective cell motility: from cell to tissue. *Journal of physics D: Applied physics*, 50(11):113002, 2017.

[143] Sri Ram Krishna Vedula, Man Chun Leong, Tan Lei Lai, Pascal Hersen, Alexandre J Kabla, Chwee Teck Lim, and Benoît Ladoux. Emerging modes of collective cell migration induced by geometrical constraints. *Proceedings of the National Academy of Sciences*,

109(32):12974–12979, 2012.

[144] Nir S Gov. Collective cell migration. *Cell Matrix Mech*, pages 219–238, 2014.

[145] Melda Tozluoğlu, Alexander L Tournier, Robert P Jenkins, Steven Hooper, Paul A Bates, and Erik Sahai. Matrix geometry determines optimal cancer cell migration strategy and modulates response to interventions. *Nature cell biology*, 15(7):751, 2013.

[146] Simon Garcia, Edouard Hannezo, Jens Elgeti, Jean-François Joanny, Pascal Silberzan, and Nir S Gov. Physics of active jamming during collective cellular motion in a monolayer. *Proceedings of the National Academy of Sciences*, 112(50):15314–15319, 2015.

[147] Alexandre J Kabla. Collective cell migration: leadership, invasion and segregation. *Journal of The Royal Society Interface*, page rsif20120448, 2012.

[148] András Szabó, R Ünnep, Eld Méhes, WO Twal, WS Argraves, Y Cao, and András Czirók. Collective cell motion in endothelial monolayers. *Physical biology*, 7(4):046007, 2010.

[149] Rachel M Lee, Douglas H Kelley, Kerstin N Nordstrom, Nicholas T Ouellette, and Wolfgang Losert. Quantifying stretching and rearrangement in epithelial sheet migration. *New journal of physics*, 15(2):025036, 2013.

[150] Juliane Zimmermann, Ryan L Hayes, Markus Basan, José N Onuchic, Wouter-Jan Rappel, and Herbert Levine. Intercellular stress reconstitution from traction force data. *Biophysical journal*, 107(3):548–554, 2014.

[151] Juliane Zimmermann, Brian A Camley, Wouter-Jan Rappel, and Herbert Levine. Contact inhibition of locomotion determines cell–cell and cell–substrate forces in tissues. *Proceedings of the National Academy of Sciences*, 113(10):2660–2665, 2016.

[152] Rachel M Lee, Haicen Yue, Wouter-Jan Rappel, and Wolfgang Losert. Data from: Inferring single cell behavior from large-scale epithelial sheet migration patterns. *Digital Repository at the University of Maryland (DRUM).(doi:10.13016/M2855R)*, 2017.

[153] Ronald J Adrian and Jerry Westerweel. *Particle image velocimetry*. Number 30. Cambridge University Press, 2011.

[154] Silke Henkes, Yaouen Fily, and M Cristina Marchetti. Active jamming: Self-propelled soft particles at high density. *Physical Review E*, 84(4):040301, 2011.

[155] Nir S Gov. Collective cell migration patterns: follow the leader. *Proceedings of the National Academy of Sciences*, 104(41):15970–15971, 2007.

[156] Antoine A Khalil and Peter Friedl. Determinants of leader cells in collective cell migration. *Integrative biology*, 2(11-12):568–574, 2010.

[157] Rizwan Farooqui and Gabriel Fenteany. Multiple rows of cells behind an epithelial wound

edge extend cryptic lamellipodia to collectively drive cell-sheet movement. *Journal of Cell Science*, 118(1):51–63, 2005.

[158] Nathan W Goehring and Stephan W Grill. Cell polarity: mechanochemical patterning. *Trends in cell biology*, 23(2):72–80, 2013.

[159] Victoria Tarle, Andrea Ravasio, Vincent Hakim, and Nir S Gov. Modeling the finger instability in an expanding cell monolayer. *Integrative Biology*, 7(10):1218–1227, 2015.

[160] Assaf Zaritsky, Doron Kaplan, Inbal Hecht, Sari Natan, Lior Wolf, Nir S Gov, Eshel Ben-Jacob, and Ilan Tsarfaty. Propagating waves of directionality and coordination orchestrate collective cell migration. *PLoS computational biology*, 10(7):e1003747, 2014.

[161] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, August 2011.