

## ABSTRACT

Title of dissertation:      CHARACTERIZING THE COMPLEX  
                                  SPATIAL PATTERNS IN BIOLOGICAL  
                                  SYSTEMS

Joshua M. Parker, Doctor of Philosophy, 2015

Dissertation directed by:  Professor Wolfgang Losert  
                                  Department of Physics

Spatial point patterns are ubiquitous in natural systems, from the patterns of raindrops on a sidewalk to the organization of stars in a galaxy. In cell biology, these patterns can represent the locations of fluorescently-labeled molecules inside or on the surface of cells, or even represent the centers of the cells themselves. These patterns arise due to the signaling activity of the cells which are mediated by a broad range of chemicals, and understanding this activity is vital to investigating these complex systems. Luckily, though each pattern is unique, the statistical properties of the patterns embed information about the underlying pattern formation process.

In this work, I demonstrate techniques to characterize the complex spatial patterns found in unicellular systems. Using topologically-derived measures, I demonstrated a technique to automatically classify sets of point patterns into groups to identify changes in higher order statistical moments due to experimental variation. This technique utilizes functional principal component analysis (FPCA) on the Minkowski functionals of a secondary pattern formed by imposing disks on each

point center. I demonstrate that this better classifies a range of point pattern sets, and then applied this technique to pattern sets representing membrane-bound proteins in human immune cells, showing that this procedure correctly identifies non-interacting proteins.

Further, I demonstrate a simulation-based technique to diminish the statistical impact of large-scale pattern features. In protein patterns, these represent the effects of membrane ruffling during pattern formation. These features dominate correlation measures, obscuring any hint of nanoscale clustering. Using heterogeneous Poisson null models for each cell to re-normalize their pairwise correlation functions, I found that patterns of LAT proteins (“linker for the activation of T-cells”) do indeed cluster, with a characteristic length-scale of approximately 500 nm. By performing clustering analysis at this length scale on both the LAT patterns and their respective null models, I found that clusters are most commonly dimers, but that this clustering is strongly diminished upon T-cell activation. This loss of clustering may be due to the presence of unlabeled molecules that have been recruited to the cell membrane to form complexes with LAT.

I also investigate both molecular and cell-center patterns in *Dictyostellium discoideum* cells, which are a model organism for amoeboid motion and G-protein receptor-mediated chemotaxis. These cells migrate using “autocrine” signal relay in that they both secrete and sense the same chemoattractant, cyclic adenosine monophosphate (cyclic AMP or cAMP). They also secrete phosphodiesterases that degrade the chemoattractant. This leads to streaming patterns of cells towards aggregation centers, which serve as sites of sporulation. To study these cells, I

demonstrate an image analysis technique that statistically infers the local population of fluorescently-labeled mRNA units in fluorescent images of self-aggregating cells. The images were of experiments where two particular mRNAs were labeled along with their respective proteins, the first being adenylyl cyclase A (ACA), a molecule involved in the production of cAMP. ACA itself has already been seen to accumulate at the back of migrating cells. The location of these molecules were compared to that of the locations of cyclic AMP receptor 1 (cAR1), which is the cell's mechanism for gradient sensing. Using my analysis technique, I found that statistically significant proportions of ACA mRNA preferentially locate towards the rear of migrating cells, an asymmetry that was also found to identically correlate with the asymmetry of ACA itself. This asymmetry was not seen in cAR1 mRNA, which tends to distribute uniformly. Further, the asymmetry in ACA was most exaggerated in cells migrating at the rear of streams, with the approach to the local aggregate center diminishing leading to more uniformly distributed molecules. This may suggest that ACA is locally translated at the back of migrating cells, a result requiring further investigation.

I then construct a computational migration model of *D. discoideum* chemotaxis and use it to investigate how the streaming phase is effected by cell-cell adhesion as well as by the global degradation of cAMP. To classify the dynamics of the model with respect to cell density and external chemical gradient, the two relevant phase variables, I develop an order parameter based on the fraction of broken cell-cell contacts over time. This parameter successfully classifies the dynamic steady states of the model (independent motion, streaming, and aggregation), outperforming the

often used “chemotactic index”. I found that the elimination of degradation strongly diminishes any presence of streaming, suggesting that chemical degradation is vital to stream formation. In contrast, the addition of cell-cell adhesion expanded the streaming phase, stabilizing streams that were formed initially through signal relay.

CHARACTERIZING THE COMPLEX SPATIAL  
PATTERNS OF BIOLOGICAL SYSTEMS

by

Joshua M. Parker

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2015

Advisory Committee:  
Professor Wolfgang Losert, Chair/Advisor  
Professor Edward Ott  
Professor Patrick Kanold  
Professor Arpita Upadhyaya  
Professor Doron Levy  
Dr. Lawrence E. Samelson

© Copyright by  
Joshua Parker  
2015

“Fortunate is he, who is able to know the causes of things.”

– Virgil, *Georgics II*, 490

“Truth can be banal, and is often pedestrian, and that is all physics can teach us.

So, more whiskey is, and can be the solution to all problems.”

– Aftaab Dewan, personal communication, April 13th, 2015

## Dedication

There quite literally are dozens of people I'd like to dedicate this work to for supporting, inspiring, and encouraging me to end up here, a Doctor of Philosophy in Physics.

To my undergrad professors at Middle Tennessee State University: To Dr. Bill Robertson, a hearty thank you for taking me on in an undergrad research project despite me having zero experience in science or research, letting me break equipment and code improperly, all the while getting more and more fascinated with science. To Dr. Eric Klumpe and Dr. Victor Montemayor, who both were excellent teachers and mentors, bringing hard the hammer down on this lazy B+ student and shaping me into a student who fought for perfection. And to Dr. Daniel Erenso, who taught me always to do physics with a smile, and for taking a lot of his own time to lecture Quantum Mechanics to me. My hat is off to you, gentlemen!

To my great friends here at Maryland: Brian Calvert, Matt Harrington, Matt Kretschmer, Thomas Rensink, Jeff Magill, and Patrick Stanger – you made many a late night study session produce both passing grades and belly laughs. I wouldn't have survived without you! A special dedication to my great friend and colleague Can Guven, who was the best research partner anyone could ask for. We flowed so fluidly between heated esoteric discussions of scientific analysis and laughing at silly jokes that only we thought were funny. Doing four years of research by yourself? Ain't nobody got time for that!

To my adviser Dr. Wolfgang Losert and the entire Losert research group: thanks for putting up with me! For listening to every hair-brained idea I had with vague interest, for helping me hone my skills, and for keeping me honest.

To my thesis committee, Dr. Ott, Dr. Kanold, Dr. Upadhyaya, Dr. Levy, and Dr. Samelson, thank you for agreeing to wade through my work! I have valued all of your comments, concerns, and suggestions, and you've truly helped me make this thesis much more solid.

And finally, a huge dedication to all of my family for supporting my crazy idea to run off and become a scientist. To my parents Mike and Paula, my sisters Rachael, Anna, Bethany, and Mary, to my best friend and brother-in-law Billy, and to my two other amazing brothers-in-law John and Nathan, a dramatic thank you. Your support was critical to my sanity and I love you all very much. And to my three beautiful and brilliant nieces Isabella, Penelope, and Aubrey – you've provided so much joy to help me get through this whole process. To you three and to my soon-to-be-here nephew Harrison – Uncle Josh will explain all of this stuff to you later!

Portions of this work were funded by the Intramural Research Programs of the National Cancer Institute (The Center for Cancer Research), NIH Grant No. GM085574, NIH grant R01GM085574, and the American Society of Engineering Education's SMART fellowship program. Numerical simulations were done in part using the HPCC at UMD.

# Table of Contents

List of Figures	ix
1 Introduction	1
1.1 The chemical origins of biological pattern formation . . . . .	1
1.2 Microscopic pattern formation in the human immune system . . . . .	5
1.3 From the immune system to the forest floor: signal relay results in the spatial patterning in migrating slime mold cells . . . . .	8
1.4 Gaining insight from spatial patterns . . . . .	11
1.5 Outline of Thesis . . . . .	14
1.5.1 Using morphological analysis techniques to sort point pattern sets . . . . .	15
1.5.2 Cluster analysis of LAT protein pattern sets containing het- erogeneous large scale features . . . . .	17
1.5.3 Quantifying the spatial asymmetry of ACA mRNA in migrat- ing <i>Dictyostelium discoideum</i> cells . . . . .	18
1.5.4 Modeling and analyzing the patterns of externally guided chemo- taxing cells . . . . .	21
1.6 Discussion . . . . .	23
2 Automatic sorting of point pattern sets using Minkowski Functionals	25
2.1 Overview . . . . .	25
2.2 Introduction . . . . .	26
2.2.1 Motivation: Why study point patterns? . . . . .	26
2.2.2 Current methods for sorting patterns . . . . .	28
2.2.3 From points to disks . . . . .	30
2.3 Outlining the Procedure . . . . .	31
2.3.1 Minkowski functional analysis of point patterns . . . . .	31
2.3.2 Sorting the patterns . . . . .	34
2.4 Testing our sorting method . . . . .	35
2.4.1 Data set 1: Two Strauss processes . . . . .	36
2.4.2 Data set 2: Baddeley-Silverman vs. random . . . . .	39
2.4.3 Data set 3: Bi-disperse patterns of inter cellular proteins . . . . .	40
2.5 Conclusions and discussion . . . . .	43

3	Heterogeneous cluster analysis of LAT protein point pattern sets reveal changes in clustering upon T-cell activation	46
3.1	Overview	46
3.2	Introduction	48
3.2.1	Point patterns in cellular systems	48
3.2.2	Does T-cell activation affect apparent clustering in LAT proteins?	48
3.2.3	An obstacle to analysis: the cell membrane	51
3.2.4	Heterogeneous cluster analysis using simulations	54
3.2.5	Data and Methods	54
3.2.6	Results: LAT proteins show clustering at the nanoscale which is diminished upon activation	55
3.3	Discussion	59
4	Adenylyl cyclase mRNA localizes to the back of polarized <i>Dictyostelium</i> cells during chemotaxis	61
4.1	Overview	61
4.2	Introduction	62
4.3	Experimental context for simulation-based analysis: ACA mRNA seems to localize at the back of polarized chemotaxing cells	63
4.4	Quantifying and simulating mRNA distribution in fluorescent images	68
4.5	Results and Discussion	72
4.5.1	ACA mRNA and ACA protein asymmetrically localize to the back of isolated and migrating cells	72
4.6	Conclusions	75
5	A broken-contact order parameter for inferring inter-cellular communication from patterns of externally guided migrating cells	77
5.1	Overview	77
5.2	Introduction	78
5.3	Stochastic cell migration model	81
5.4	Quantifying coordinated motion in migrating cells	84
5.5	Chemical and mechanical perturbations affect stream stability	88
5.6	Conclusions	89
6	Summary and Future Outlook	90
A	Appendix B: Supplemental Material for Chapter 2	94
A.1	Software	94
A.2	Intensity Scaling	95
A.3	Approximating $g(r)$	95
A.4	Pattern Simulation	96

B	Appendix B: Supplemental materials for Chapter 3	97
B.1	Approximating $g(r)$	97
B.2	Theoretical considerations for heterogeneous cluster analysis	97
B.2.1	The heterogeneous pairwise correlation function	97
B.3	Simulating a heterogeneous Poisson null model for an existing point patter	99
C	Appendix C: Supplemental materials for Chapter 4	101
C.1	Experimental procedures	101
C.1.1	Preparation of cells	101
C.1.2	Antibodies and Immunoblotting	101
C.1.3	Chemotaxis and streaming assays	102
C.1.4	Fluorescent in situ hybridization (FISH)	102
C.2	Perturbations	103
C.3	RNA isolation and analysis	103
C.4	Cycloheximide recovery	103
C.5	Ribopuromycylation	104
C.6	Measuring cell polarity	105
C.6.1	ACA translation originates in the cytoplasm and localizes to the back of cells as they acquire polarity	106
D	Appendix: Supplemental materials for Chapter 5	111
D.1	Cell motion experiments	111
D.1.1	Cell culture and development	111
D.1.2	Chemotaxis assay	112
D.2	Parameter selection based on experiment	112
D.2.1	The Concentration Profile	112
D.2.2	Individual cell parameters	114
D.3	The time-scale independence of the fraction of broken links	118
D.4	Condition for signal independent ensemble speed	119
D.5	The parameter regime for two-time scale correlations	121
D.6	Numerical details	123
	Bibliography	126

## List of Figures

1.1	A) From left to right: a zebra (photo by Annie Katz, used with permission), a discus fish (Miandad Rahim, used with permission), aloe polyphylla (photo by Stan Sheb, distributed under the Creative Commons license), and a sunflower (photo by Lucapost, distributed under the Creative Commons license.) B) Diagram of how morphogen gradients can lead to cell differentiation (adapted from [1], used with permission) C) Diagram of Turing’s reaction-diffusion model for pattern formation (adapted from [2], used under Elsevier license 3594940056911) . . . . .	2
1.2	(A-B) Initial chemical signals of wounded tissue releases chemokines which attract primary neutrophils, who in turn secret LTB4 to relay the signal (adapted from [3], used under Elsevier license 3594941307244) C) Diagram of T-cell mediated inflammatory response (adpated from [4], used under the Creative Commons license) D) LAT signaling proteins on the surface of a Jurkat T-cell (adapted from [5], used under Elsevier License 3594950695211) . . . . .	6
1.3	(A) Life cycle of the slime mold cell Dictyostellium Discoideum. Streaming phase occurs between individual motion and sporulation (Adapted from Hideshi, distributed under Creative Commons) (B) The complex spiral structure generated during cell aggregation (Photo by Bruno, distributed under Creative Commons) . . . . .	10
1.4	(A) Differential pattern formation both in the presence (left) and the absence (right) of chemoattract degradation (adapted from [6], used with author’s permission). (B) The different pattern formation “trajectories” of cell populations platted on surfaces with different adhesive properties (adapted from [7], used under Creative Commons license) . . . . .	12
2.1	Using photoactivated localization microscopy, the fluorescencly labeled proteins are localized by fitting a point spread function to the stochastically photoactivated molecules; the final pattern represents thousands of fluoresecent images . . . . .	27

2.2	(a) The Minkowski functionals are calculated by imposing disks on the point pattern. This new secondary structure can be characterized using topological measures, which vary for different radii (b) The three reduced Minkowski functionals for a 2-D Poisson (random) process. These functionals are unitless due to the normalization by the same measure one would expect for a set of non-overlapping disks. . . . .	32
2.3	(a) The regularity of a Strauss process is completely determined by $\gamma$ , the interaction parameter. For $\gamma = 0$ , no overlaps are allowed. For $\gamma = 1$ , all overlaps are allowed. (b) The left most panes display the $g(r)$ and $\chi(r)$ for the 40 simulated Strauss processes. On the right are the results of using FPCA scores to divide the pattern set into two groups. As can be seen, both $g(r)$ and the Minkowski functionals can perfectly separate the set into two groups corresponding to different values of $\gamma$ . . . . .	37
2.4	(a) A Baddeley-Silverman (left) process side-by-side with a Poisson process (right). Despite the visible differences, the pairwise correlation functions are identical (b) The left most panes display the $g(r)$ and $\chi(r)$ for the 58 patterns simulated. On the right are the results of using FPCA scores to divide the pattern set into two groups. As can be seen, FPCA sorting with $g(r)$ creates two perfectly heterogeneous groups, while FPCA sorting with the Minkowski functionals groups the patterns correctly. . . . .	38
2.5	(a) The two proteins are both dispersed in the cell membrane. LAT (blue) and TAC (red) proteins are separately tagged, allowing them to be visualized separately. (b) The left most panes display the $g(r)$ and $\chi(r)$ for the 16 molecular patterns. On the right are the results of using FPCA scores to divide the pattern set into two groups. As can be seen, using Minkowski functionals with FPCA improves the differentiation of the two sets. . . . .	41
2.6	From left to right: Sorting with all three functionals, the <i>area</i> , the <i>perimeter</i> , and the <i>Euler number</i> . Considering the individual functionals (sorting using weighted modularity), the Euler number outperforms the other two, only misidentifying one pattern . . . . .	43
3.1	(a) Experimental setup. Glass surfaces are coated with antigens for cells to settle onto and imaged using super-resolution microscopy. (b) A single sample image of the fluorescence patterns seen in PALM experiments. (c) Three representative patterns of those derived from non-active cells (left and center) and activated T-cells (right) (d) Images of the spreading process of a Jurkat T-cell on a glass slide. Both the cell boundary and the variation in adhesion pattern affect the protein pattern (adapted from [8], used under the Elsevier user license) . . . . .	50
3.2	Standard pairwise correlation function for three sets of protein patterns. Dashed lines represent 95% confidence intervals. . . . .	53

3.3	(a) (left) A protein point pattern derived from a PALM experiment and (right) a heterogeneous Poisson point process with the same intensity distribution. Small changes can be seen in the location of the points on the fringes, but the large scale features are preserved. (b) Heterogeneous pairwise correlation function for three sets of protein patterns. Dashed lines represent 95% confidence intervals. All three pattern sets collapse onto a single function reminiscent of clustering .	56
3.4	Differences between cluster size distributions of data and Heterogeneous Poisson null models for two sets of patterns representing non-active T-cells and one set of active patterns. Dashed lines represent 95% confidence intervals . . . . .	58
4.1	ACA and cAR1 mRNAs are randomly distributed in vegetative cells. (A) Maximum intensity projections of confocal fluorescent images of vegetative ACA-YFP/aca- cells depicting DAPI (nucleus) and ACA mRNA (pink). (B) Maximum intensity projections of confocal fluorescent images of vegetative cAR1-YFP/car1/3-/- cells depicting DAPI (nucleus) and cAR1 mRNA (pink). (C and D) Maximum intensity projections of confocal fluorescent images of 5 hrs differentiated aca- (C) or cAR1/3-/- cells (D) depicting DAPI (nucleus) and ACA or cAR1 mRNA (pink). The data are representative of three independent experiments. . . . .	65
4.2	ACA mRNA is spatially localized to the back of polarized chemotaxing cells. (A) Representative maximum intensity projections of confocal fluorescent images of WT cells depicting F-actin (green), DAPI (nucleus), and ACA mRNA (pink). The dotted red box in the image is enlarged to show a single polarized cell and the asymmetric ACA mRNA distribution. The small white arrows indicate the position of the mRNA spots (B-C) Representative merged phase contrast and maximum intensity fluorescent images depicting DAPI (nucleus), ACA mRNA (pink), in WT and ACA-YFP/aca- cells. The small white arrows indicate the position of the mRNA spots. The white dashed arrows indicate the direction of the stream in these self-aggregation chemotaxis experiments. . . . .	66

4.3	<p>(A-B) Representative merged phase contrast and maximum intensity fluorescent images depicting DAPI (nucleus), and cAR1 mRNA (pink) in WT and cAR1-YFP/car1/3-/- cells. The small white arrows indicate the position of the mRNA spots. The white dashed arrows indicate the direction of the stream in these self-aggregation chemotaxis experiments. (C) Each cell was hand bisected, defining front and back Regions of Interest (ROI) based on both the orientation towards the aggregate center and the relative back enrichment of ACA-YFP in the cell. (G) Simulated and linear estimates of mRNA units across cells is plotted for ACA-YFP/aca- and cAR1-YFP/car1/3- cells. The boxes show the 50% confidence region from the median (red line). The bars cover a region with 99% confidence level from the median. All data points beyond this confidence level are considered as outliers and shown with red dots. The statistical significance is inferred by the t-test, * represents <math>p &lt; 0.05</math> and ** represents <math>p &lt; 0.01</math>. (<math>n_{ACA} = 45</math>, <math>n_{cAR1} = 24</math>).</p>	67
4.4	<p>Simulation and quantification of spatial ACA mRNA localization patterns. (A) For each image, a peak finding routine was run on the mRNA florescent channel (left). Isolated spots were identified by thresholding their size and intensity (right). (B) Peaks were fit to Gaussian point spread functions. The resulting distributions were thresholded from above until fine, unimodal distributions remained for the two fit parameters. The mean of these distributions were termed as units. Both ACA and cAR1 mRNA showed comparable parameters. (C) The sequential images from a single iteration of the image simulation procedure performed on the mRNA fluorescent channel. Areas of yellow represent agreement. (D) The number of units in a particular image was determined by minimizing the squared different between the approximated image and the original. This is equivalent to minimizing the chi-square parameter of the fit. (E) After performing the procedure multiple times, the average image is calculated and used for quantification.</p>	70

4.5	The ACA mRNA is asymmetrically distributed at the back of the streaming cells (A) Cartoon depicting the distribution of cells within a stream. Each cell was characterized as either being at the back of a stream, in the middle of stream or near an aggregate based on its position from the aggregate center. (B) The simulated estimate of mRNA units across the cell is plotted for ACA-YFP/ <i>aca-</i> (grey box) and cAR1-YFP/ <i>car1/3</i> (white box) cells. The boxes show the 50% confidence region from the median (red line). The bars cover a region with 99% confidence level from the median. All data points beyond this confidence level are considered as outliers and shown with red dots. The statistical significance is inferred by the t-test, * represents $p < 0.05$ and ** represents $p < 0.01$ , $n = 15 - 52$ . (C-D) The correlation between mRNA and its corresponding protein at the back of cells is presented for ACA-YFP/ <i>aca-</i> and cAR1-YFP/ <i>car1/3-/-</i> cells (see Experimental Procedures for details), $n = 12 - 45$ . . . . .	74
5.1	(a) Initial state of cells (uniformly distributed). (b-c) Cells moving in streams. (c) Final configuration of wild-type cells is a large aggregate. (d) Mutant cells that cannot stream form smaller aggregates in their final configuration. . . . .	80
5.2	(a-c) Snapshots of the simulations demonstrating the three steady state dynamic modes: individual motion, streaming, and aggregates ( $\rho = 750 \text{ mm}^{-2}$ ) (d) The mean progression index (left) and the fraction of broken links (right) for a variety of $\rho$ and $C_0$ values. (e) The fraction of broken links for cells with intercellular adhesion (f) The fraction of broken links for cells that do not degrade cAMP . . . . .	87
C.1	. Loss of ACA-YFP but not cAR1-YFP after CHX treatment Western analysis showing protein levels of ACA-YFP from ACA-YFP/ <i>aca</i> cells or cAR1-YFP from cAR1-YFP/ <i>car1/3-/-</i> cells in the presence of 1.6 mM CHX and during the recovery time points. DMSO-treated cells were used as control for this experiment. Representative data of two independent experiments are shown. . . . .	107

C.2 ACA translation occurs in the cytoplasm followed by localization to the rear of the polarized cells (A) Representative phase contrast (upper panel) and confocal fluorescent (lower panel) images of ACA-YFP/*aca-* treated with 1.6 mM CHX for 2 hrs. Fluorescent recovery is monitored after CHX removal. (B) Representative maximum intensity projections of confocal fluorescent images of ACA-YFP/*aca-* cells depicting DAPI (nucleus) and ACA mRNA (red). Red arrows indicate the nascent ACA transcription sites after CHX removal. The white star indicates the location of the aggregation center. (C) Representative phase contrast (upper panel) and confocal fluorescent (lower panel) images of cAR1-YFP/*car1/3-/-* treated with 1.6 mM CHX for 2 hrs. (D) Representative maximum intensity projections of confocal fluorescent images depicting DAPI (nucleus) and cAR1 mRNA (red). For (A) and (C), red arrows indicate the appearance of the nascent cellular ACA translational sites while white stars indicate the location of the aggregation center. For (B) and (D), white stars indicate the location of the aggregation center. . . . . 108

C.3 (A) The simulated estimate of ACA mRNA units 60 and 120 min after CHX removal across the cell is plotted for ACA-YFP/*aca-* cells. The boxes show the 50% confidence region from the median (red line). The bars cover a region with 99% confidence level from the median. All data points beyond this confidence level are considered as outliers and shown with red dots. The statistical significance is inferred by the t-test, \* represents  $p < 0.05$ . The data excludes the 0 min time point as these cells are not polarized,  $n = 6 - 15$ . (B) Western analysis of C-terminally puromycylated nascent chains released from ribosomes with and without puromycin (PMY) treatment in ACA-YFP/*aca-* and WT cells (upper panel). The complex was immunoprecipitated using the anti-PMY mAb 12D10 and immunoblotted with the anti-rpL8 polyclonal antibody to detect associated ribosomes (middle panel). The associated RNA was dissociated from the complex and the presence of specific mRNAs corresponding to *acaA* and *carA* were detected by RT-PCR using specific primers (bottom panel). This figure is a representative of at least four independent experiments. . . . . 109

D.1 The y-component of the signal vector seen by the cell in the absence of cell secretion (i.e.  $s_{cell} = 0$ ). As the external concentration is decreased, the signal goes to a constant. . . . . 113

D.2 The signal-to-noise as seen by a cell as a function of its position in the chamber. The noise strength  $\eta_\phi$  is chosen so that the signal-to-noise reaches unity as the cell approaches the end of the chamber (i.e. for  $y = L_y$ ) . . . . . 115

D.3 The fraction of broken links is determined by the time scales of the model and not the actual timescale of the temporal network window . 118

## Chapter 1: Introduction

### 1.1 The chemical origins of biological pattern formation

There are few things more ubiquitous in the biological world than patterns. In the animal kingdom, patterns range from the simple spots on animal hides and the clever camouflages of moths to the dizzying patterns seen on rain forest dwellers and even those under the sea. In plant species, the complex and seemingly fractal growth patterns seen in leaf arrangements and flower petals show wild variation (see Fig.1.1.A).

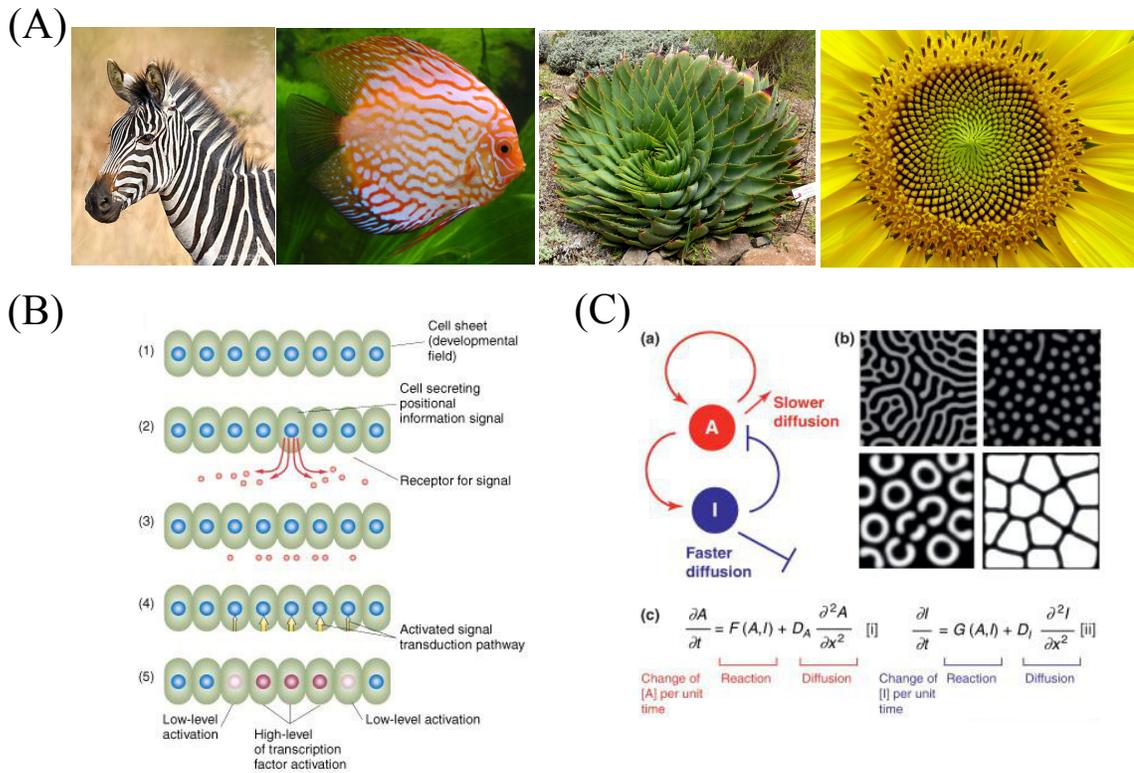


Figure 1.1: A) From left to right: a zebra (photo by Annie Katz, used with permission), a discus fish (Miandad Rahim, used with permission), aloe polyphylla (photo by Stan Sheb, distributed under the Creative Commons license), and a sunflower (photo by Lucapost, distributed under the Creative Commons license.) B) Diagram of how morphogen gradients can lead to cell differentiation (adapted from [1], used with permission) C) Diagram of Turing's reaction-diffusion model for pattern formation (adapted from [2], used under Elsevier license 3594940056911)

The structure and functional purpose of these patterns have fascinated scientists for centuries. For instance, much complexity can be seen in the different spiral patterns found in plant leaves. Depending on the angle of the leaf relative to its stem, the type of offset each leaf will take relative to its neighbor, and the number of leaves on each stem, spirals of varying leaves of complexity and repetition can be generated, distinguishing whole classes of plants, making them visibly recognizable to animal and man alike [9]. This class of behavior, called “phyllotaxis” (greek for “*leaf arrangement*”), has been intensely studied by mathematicians, who have connected angular deviations to the Fibonacci numbers [10] or to fractal growth patterns [11]. This patterning has been suggested to maximize the project surface leaf surface area, yielding the best overall plant exposure to sunlight [12]. There is even more complexity in the variation of the patterns of animal hides. From striped animals such as tigers and zebras to spotted animals such as dalmatians, leopards, and giraffes, the alternate orientation of light and dark regions of hair or scales have been suggested to arise from complex predator-prey cycles where the best patterned creatures are the ones who survive [13].

Though patterns vary so wildly, the primary determinant of pattern structure is often chemical in nature, governed by signaling molecules called “morphogens”. During development, morphogens are released, forcing gradients to arise across an organism (see Fig.1.1.B). Depending on the local concentration, different signaling pathways are triggered in a cell which alter gene expression, resulting in spatially heterogeneous cell differentiation [14]. This serves to pattern the organism with

diverse cell groups, with the specific biochemistry and physical domain of the morphogens designing the global pattern.

Famed Mathematician Alan Turing showed [15] that if a two morphogen system contained a self-induced short range source of positive feedback on one species, which itself was long-range inhibited by the other species, then simple reaction-diffusion kinetics of the two molecules could lead to spatial patterning of cell differentiation (see Fig.1.1.C). He suggested this as a potential mechanism for the spiral structures of plants as well as the varying patterns on animal hides. Many such chemical reactions have been identified, with chemically induced spirals and spots being dubbed “Turing patterns ” [16].

However, the issue has turned out to be much more complex. Indeed, the physical mechanism for the structure of leaf spiral has been shown to be the presence or absence of the growth hormone auxin. A leaf begin to grow where auxin is in low concentration, which then draws auxin to it and away from other areas, leaving places for the next leaf to grow [17]. However, despite being chemically-based, auxin gradients are not simply reaction-diffusion systems. The chemical migration is regulated by the organism itself through transport proteins that pump auxin from cell to cell [17, 18], resulting in the variation of plant spiraling. Further, in *Drosophila* development, the cell-cell pumping of morphogens has been shown to be modified by a process called “lateral inhibition”, where a cell triggers a neighbor’s surface receptors, resulting in the inhibition of a gene in the absence of a morphogenic cue [19]. This means that both external chemical signals and cell-cell mediated signaling end up affecting the overall global pattern formation.

## 1.2 Microscopic pattern formation in the human immune system

Another system where external gradients and cell-cell signaling produce complex biological processes is the human immune system. When a wound forms, the compromised cells release cytokines which diffuse away from them. Leukocytes, more commonly known as white blood cells, sense the presence of the cytokines through receptor activity and begin migrating up the gradient towards the wound site, a process known as “chemotaxis”, to engulf and/or kill any invading pathogens as well as to form a barrier at the wound site. This is a feature of the innate immune response [20], which acts swiftly and automatically to target all foreign invaders (see Fig.1.2.A-B). Two types of leukocytes, macrophages and dendritic cells, consume or “phagocytose” the pathogens, disassemble them, and taxi pieces to the cell membrane to signal the presence of an invader to lymphocytes (T-cells and B-cells), which make up the adaptive part of the immune system [21] (see Fig.1.2.C). When the pieces of the pathogen (called “antigens”) are presented to pathogen-specific lymphocytes, both T-cells and B-cells further respond by releasing more cytokines, releasing antibodies, and hunting down the pathogen in the blood stream to kill it. Thus, the entire human inflammatory response requires the chemical coordination of a handful of cell types who signal to each other with a variety of chemical methods.

Not surprisingly, these complex chemical processes also lead to pattern formation. Upon antigen presentation to a T-cell, the “T-cell receptor” (TCR) activates a signaling cascade that results in the phosphorylation of the “linker for the activation of T-cell” protein (LAT), which in turn draws other molecules to the surface of the

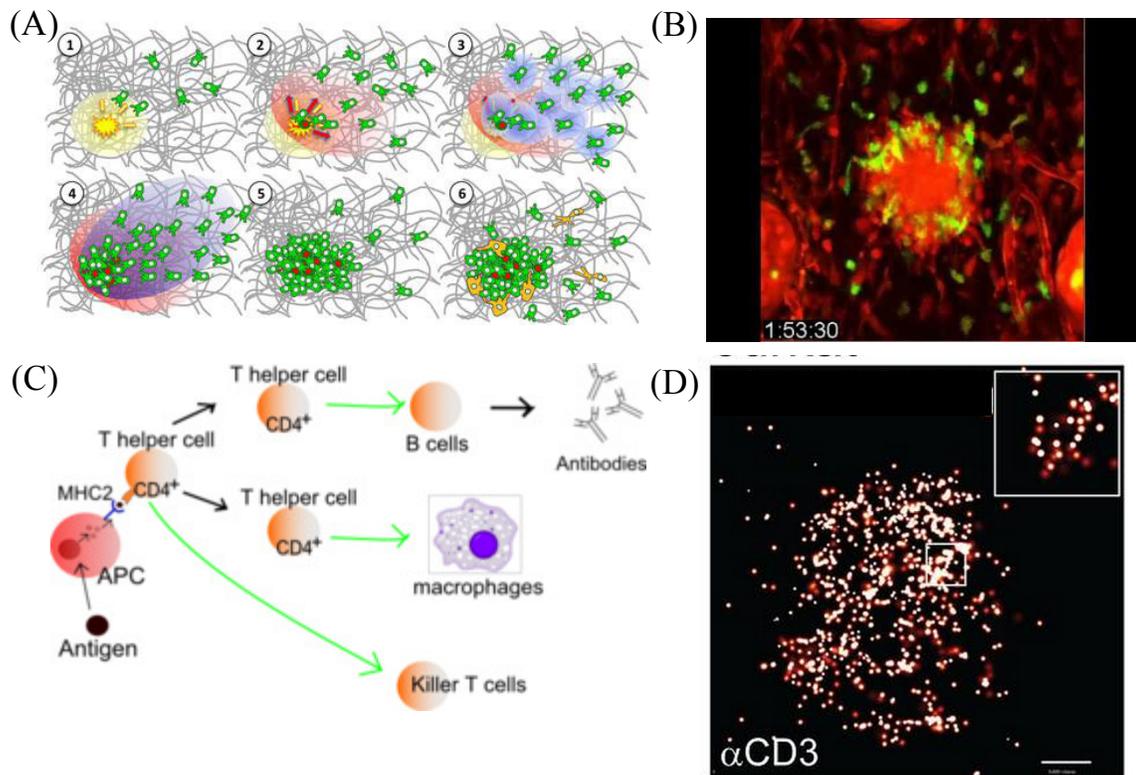


Figure 1.2: (A-B) Initial chemical signals of wounded tissue releases chemokines which attract primary neutrophils, who in turn secrete LTB4 to relay the signal (adapted from [3], used under Elsevier license 3594941307244) C) Diagram of T-cell mediated inflammatory response (adapted from [4], used under the Creative Commons license) D) LAT signaling proteins on the surface of a Jurkat T-cell (adapted from [5], used under Elsevier License 3594950695211)

membrane to form multimolecular complexes [22]. Depending on the type of T-cell, further downstream effects result in changes in migratory patterns and secretion of cytokines. Like morphogenic triggers, the activation of membrane-bound signaling receptors creates whole cell changes. Recent developments in microscopy allow the imaging of these membrane proteins to within tens of nanometers [23], yielding spatial point patterns, the statistics of which are indicative of protein-protein and protein-membrane interactions (see Fig.1.2.D). Studying these signaling pathways is paramount for understanding both the basic functions of the immune system, but also aberrant TCR-mediated responses of T-cells lead to various health concerns, ranging from hair loss [24] to serious illness or death [25].

Further, the ability for leukocytes to migrate up chemical gradients can itself result in spatial patterning. Neutrophils, the most common leukocyte in the blood stream, are the body's first responders, swarming the site of infection and creating a barrier to further penetration while other leukocytes migrate slower and consume pathogens to present to T-cells [3,26,27]. This creates spatial heterogeneous patterns surrounding the infection site, with macrophages clustering outside of the already-present neutrophils (see Fig.1.2.B). Further, neutrophils have been shown to secrete the cytokine leukotriene b4 (LTB4) which enhances the original cytokinetic signal, a process known as "paracrine" signal relay. This results in a swarming pattern, drawing cells to the site of inflammation that were too far away to respond to the first cytokinetic burst [26]. Just as morphogens regulate the spatial patterns of cell populations by cell-cell signaling, neutrophils can utilize cell-cell signaling to modify external chemical cues and change their spatial arrangements. Therefore, changes

in these dynamic patterns can reveal the effect of different drug treatments and cell mutations on leukocyte response efficiency.

Further, neutrophil mis-regulation is involved in autoimmune disorders such as arthritis and lupus, with targeted deletions in the signaling pathways and cell surface markers leading to reductions in the prevalence of the disease in mice models [28]. The chemotaxis of neutrophils is mediated by G-protein coupled receptors (GPCRs), a large class of proteins involved in receptor-based signal transduction across the eukaryotic kingdom. Current attempts to provide therapeutic intervention consists of investigations into which drugs can target expression factors along the GPCR signaling pathways [29].

### 1.3 From the immune system to the forest floor: signal relay results in the spatial patterning in migrating slime mold cells

Another eukaryotic cell that exhibits GPCR mediated chemotaxis is *Dicystostellium Discoideum*, an amoeboid slime mold cell. A much simpler organism, *D. Discoideum* both responds to and secretes the same chemical, cyclic adenosine monophosphate (cAMP), which leads to “autocrine” signal relay. To prevent self-signaling, the cells secrete cAMP at the back and respond most strongly to ligand binding and unbinding at their leading edge [30]. Cells also secrete a phosphodiesterase to degrade the external signal [31]. Despite these differences, the downstream kinases that transmit chemical signals into mechanical responses are highly conserved, resulting in *D. Discoideum* being a model cell for studying factors that

affect chemotaxis [32].

*D. discoideum* cells *also* exhibit complex spatial patterning during sporulation. When in a nutrient rich environment, *Dicty* cells migrate individually, but upon starvation polarize and begin to secrete cAMP from their rear [33], creating local chemoattract gradients (see Fig.1.3.A). The cells themselves begin to migrate up the local gradients, secreting all the while, meaning that the cell ensemble is both generating and responding to a temporally varying chemical signal. Eventually instabilities in the cell density create local aggregation centers that begin to dominate the chemical signal, drawing other cells to it. Due to the autocrine signaling, however, the responding cells form head-to-tail streams, which result in large, transient spiral structures that end in an ensemble sized aggregate (see Fig.1.3.B). Then, different cell cues take over to draw the ensemble up off of the surface and form a spore, which is then picked up by the wind and taken to a new, hopefully nutrient-rich environment.

Again, these spatial patterns depend on experimental conditions. Cells which do not externally or internally degrade cAMP do not chemotax efficiently and therefore show aberrant stream formation [34, 35]. Cells which migrate on surfaces with different adherent properties also show differences in streaming capabilities, with cells on strongly adhesive surfaces not being able to properly form large aggregates [7]. And by inhibiting the production of adenylate cyclase A (ACA), the cyclase actively involved in the production of secreted cAMP, streaming disappears entirely, completely destroying the ability of the cell ensemble to find the appropriate aggregation center and sporulate [36]. Pattern formation in *D. Discoideum* is

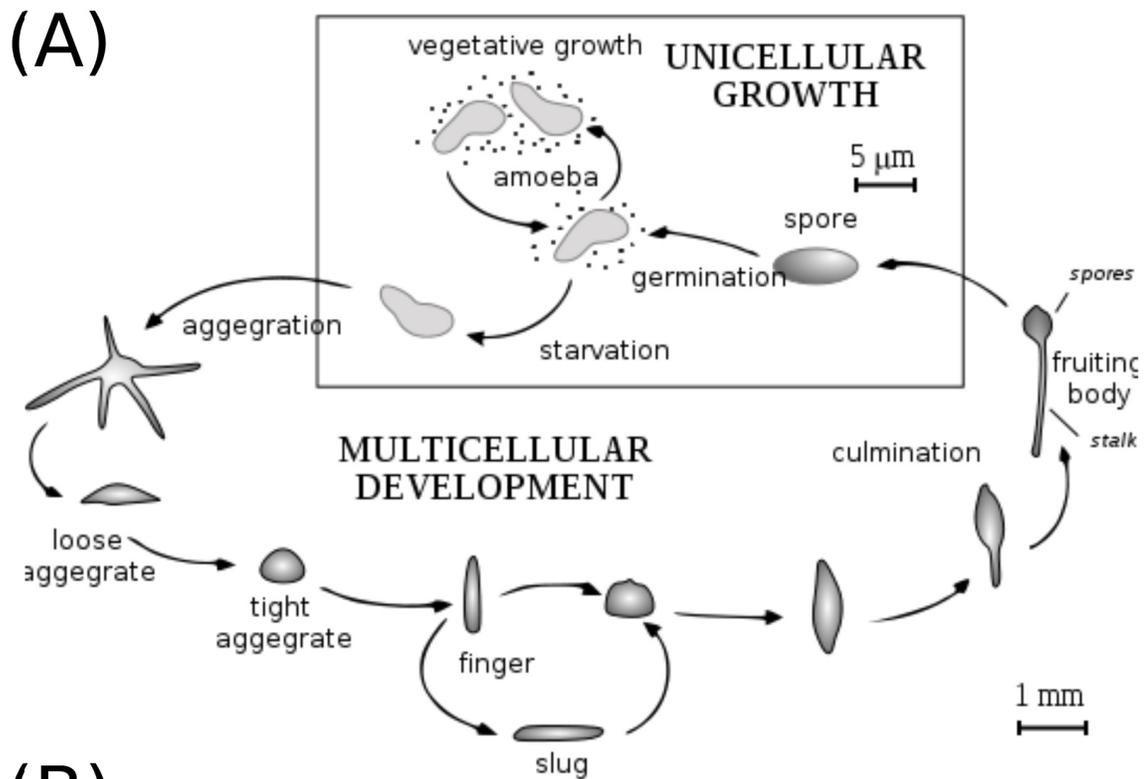


Figure 1.3: (A) Life cycle of the slime mold cell *Dictyostellium Discoideum*. Streaming phase occurs between individual motion and sporulation (Adapted from Hideshi, distributed under Creative Commons) (B) The complex spiral structure generated during cell aggregation (Photo by Bruno, distributed under Creative Commons)

therefore very sensitive to experimental variation, both in external environmental cues as well as cell mutations.

## 1.4 Gaining insight from spatial patterns

As has been discussed, spatial patterns arise in a variety of biological systems, and of particular interest is the patterning of membrane-bound signaling molecules and the spatio-temporal patterns that they introduce in migrating cells. These patterns, both static and dynamic, are stochastic – with probability 1, no two patterns will ever be identical. However, the underlying statistical process that generates the pattern is demonstrative of the underlying interactions involved, and therefore each pattern can be used to characterize the particular system.

Whether we’re discussing the absolute positions of proteins or the centers of migrating cells, one can classify these patterns as “point” patterns, in that the center position information is what one intends to characterize. Standard analysis begins with the defined centers,  $\{x_i\}$ , within a viewing window,  $\Omega$ . A researcher applies different statistical measures and then compares them with the same measures of the null hypothesis that the pattern’s points are uniformly distributed, i.e. the points are completely spatially random (CSR). Confidence intervals are defined, statistical tests are verified, parameters are fit and the pattern is considered characterized.

These measures can vary by their statistical “order”, as in the number of points needed to calculate the measure. The 0th order measure is just the number of points,  $n$ . This contains no spatial information other than the average density,  $\lambda = \frac{n}{|\Omega|}$ . A

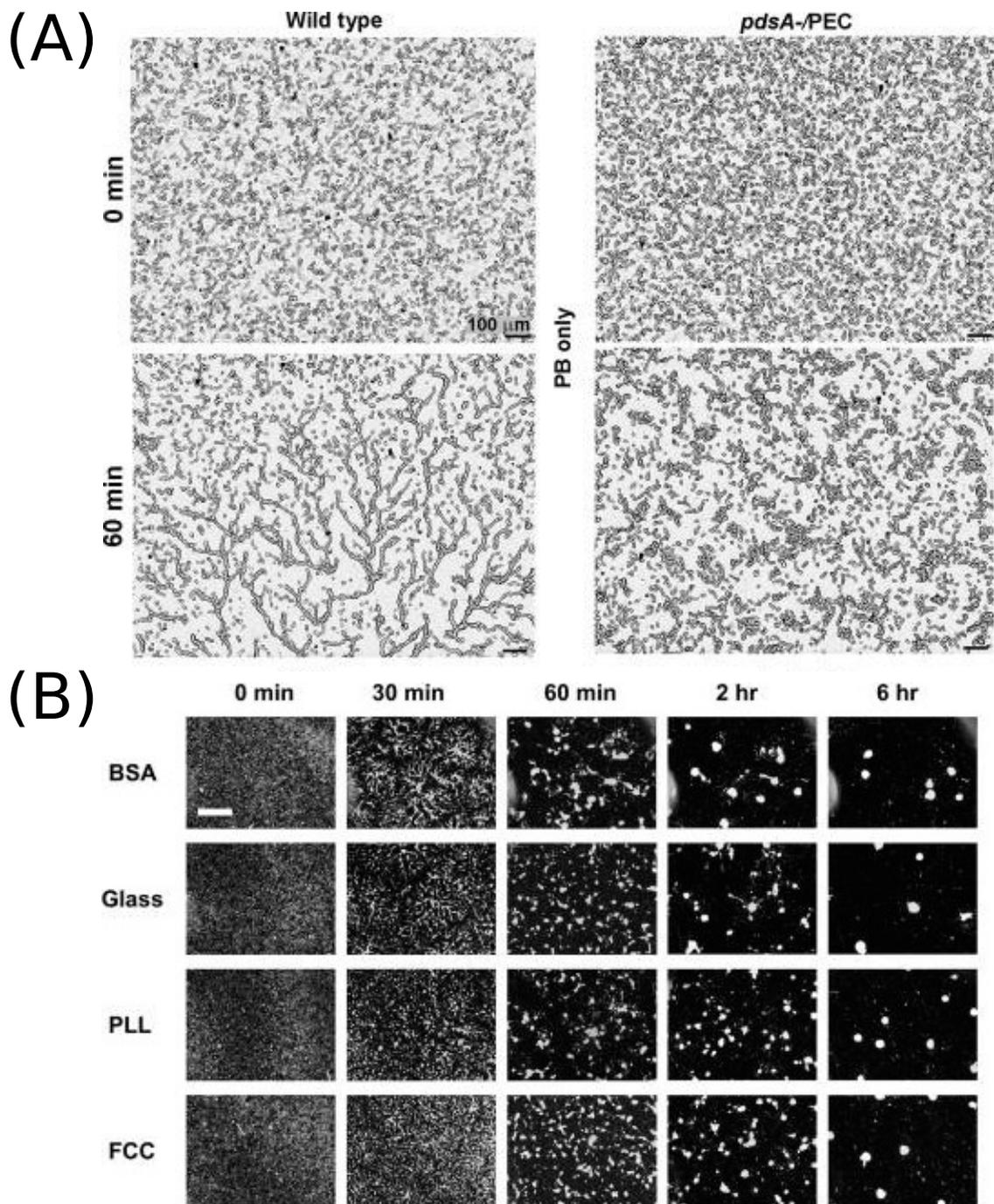


Figure 1.4: (A) Differential pattern formation both in the presence (left) and the absence (right) of chemoattract degradation (adapted from [6], used with author’s permission). (B) The different pattern formation “trajectories” of cell populations plated on surfaces with different adhesive properties (adapted from [7], used under Creative Commons license)

1st order measure of the pattern is a quantity that only involves individual points in the viewing window, such as the local number intensity  $\lambda(\vec{r})$ . This quantity can be approximated using the entire set of points, but does not contain statistical information about correlations. A 2nd order process involves measures on inter-particle distances, such as the pairwise correlation function or Ripley's K function [37]. These by far are the most common types of pattern analysis techniques since they both include enough information to discuss interactions between points but are simple to understand because they only consider two points at a time. There exist 3rd order measures which considering three points at a time [38,39], 4th order which consider four [40], and so on. The majority of research in point pattern statistics involves building phenomenological models founded in statistical theory and then applying the theory to data sets to fit parameters.

If the pattern formation process is a linear superposition of pairwise interactions, such as those formed under electrical or gravitational forces, second order statistics are often enough. However, patterns in biological systems are not so simple. For protein patterns, if a protein complex is formed due to a catalyzed chemical reaction (due to the presence of an enzyme), then the fluorescent images of the protein pattern could show higher order moments due to the non-linear chemical reaction process and the presence of unlabeled molecules. Further, the protein-membrane effects will force the pattern to already be non-random [8]. This means that the appropriate null model for the patterns is not CSR. In migrating cell populations, cell-cell signaling via chemical secretion is mediated by the modified external chemical gradient, which is both spatially varying and time dependent. Migrating cells

are also generally navigating complex environments which would arbitrarily clump or separate otherwise non-interacting points [3]. Since these features themselves vary from cell to cell and experiment to experiment, appropriately handling these issues become doubly important issue when dealing with *sets* of point patterns.

Additionally, no single order correlation function can *characterize* a point pattern since it only describes that point pattern at that statistical level. This means that processes appear random at one order can be seen to have structure or clustering at other orders [37, 41]. What is being more and more addressed in applied point pattern analysis is the use of “secondary” structures where the point pattern is transformed into another type of mathematical object, such as a graph or topological set, and analyzed using non-standard statistical technique [37, 42]. These measures focus not on a single order, but include many orders at once to better quantify a point pattern. This often leads to measures with more statistical power [43], a vital quantity when analyzing sets of biological data.

## 1.5 Outline of Thesis

This thesis contains four contributions representing my work to develop and apply statistical techniques that quantify and characterize the complex spatial and spatio-temporal point patterns that arise in biological systems, particularly those that are related to the human immune system. These techniques aim to help researchers answer the “are” and the “how” questions of spatial pattern analysis: “*Are the patterns different?*” and “*How are the patterns different?*”. What follows is a

brief overview of each project and the results therein. Whereas much of standard point pattern analysis relies heavily on analytic but phenomenological models, I leaned heavily on direct simulation to both remove the assumptions of the statistical approach as well as to better approximate the biological processes considered.

### 1.5.1 Using morphological analysis techniques to sort point pattern sets

The first contribution [44] was aimed at answering the “Are they different?” question by improving techniques to perform unsupervised classification of point pattern sets. Prior techniques relied on using standard 2nd order functionals of point patterns – ones that consider only pairwise interaction – along with functional principle component analysis (FPCA) to divide a set of patterns into different groups [45]. However, these functionals have been shown in several instances to not fully characterize spatial point patterns [37]. Most definitively, Baddelley and Silverman explicitly designed a point process that is indistinguishable from a Poisson process when only considering 2nd order measures [41]. This suggests that to fully characterize a spatial point pattern for sorting purposes, different measures should be used.

To address the issue of the failure of 2nd order measures, Mecke and Stoyan suggested transforming a point pattern into a set of overlapping discs of radius  $r$  and calculating the “Minkowski numbers” of the pattern [43]. In two dimensions, these numbers are the projected *area* and *perimeter* of the shapes resulting from

the overlapping discs, as well as the *Euler number*, which is the difference between the number of connected objects and the number of holes that are present in the topological set. These numbers contain at once all the statistical moments of the pattern, and have been known to completely characterize convex sets, a result of Hadwiger's Theorem [46]. Mecke and Stoyan showed that by considering a range of  $r$  values, the resultant Minkowski functionals were able to distinguish between Poisson processes and the process generated by Baddeley and Silverman.

I, therefore, implemented a sorting technique that combined FPCA with Minkowski Functional analysis. I first showed that both standard 2nd order functionals and Minkowski functionals perfectly distinguish between sets of Strauss processes [47], which simulate randomly distributed discs with varying levels of overlap permitted; this result was consistent with the original results of [45]. However, when I attempted to distinguish between a set of Poisson patterns and a set of patterns representing Baddeley and Silverman's process, I showed that 2nd order functionals completely failed to distinguish the two pattern sets while the Minkowski functionals sorted the sets perfectly. I then applied the technique to bi-disperse point patterns derived from super-resolution images of two trans-membrane proteins, LAT and TAC, on the surface of T-cells. Previous results show strong differences in the molecular distribution of these molecules, with LAT forming clusters and TAC distributing more randomly along the membrane [48]. Again, I showed that FPCA with Minkowski functionals outperformed FPCA with 2nd order functionals, which again completely mis-sorted the pattern sets. Then, investigating FPCA sorting with each functional individually, I showed that FPCA with only the *Euler number*

best sorted the protein pattern sets. I discussed this result in the context of future analysis of multi-color pattern sets.

### 1.5.2 Cluster analysis of LAT protein pattern sets containing heterogeneous large scale features

My second contribution [49] was aimed at the “How are they different?” question, namely by developing a technique to perform cluster analysis on pattern sets that have heterogeneous large scale structures. Specifically, I was interested in looking at the small-scale clustering hierarchy of point patterns derived from super-resolution images of LAT protein patterns on the surface of T-cells (utilizing photoactivation localization microscopy, or PALM [23]). These experiments included ones where the T-cells had been chemically activated and where they had not. The main obstacle to performing this analysis, however, is large scale heterogeneities in the individual point patterns. These can be due to larger scale clusters, voids in the pattern due to lack of membrane adhesion, or by protein-membrane interactions, where random voids in the point pattern form due to high membrane fluctuations [8]. These large scale features dominate correlation measures, making it impossible to isolate small features.

A way of dealing with this issues is to simulate heterogeneous Poisson processes (HPPs) for each cell to use as individualized statistical baselines. Prior work with protein pattern sets has shown that these null models are better adapted to establish randomness than CSR when looking at the pairwise correlation function of a single

pattern [5]. To facilitate cluster analysis of the set of point patterns, I simulated multiple HPPs for each pattern and used their average pairwise correlation function to calculate the heterogeneous pairwise correlation function through simple division  $g_{i,inh}(r) = g_i(r) / \langle g_{i,hpp}(r) \rangle$ . I demonstrated that this strongly diminished cell-to-cell variation, revealing a recognizable correlation function for a clustered pattern with a characteristic length scale of approximately 500 nm. I then performed cluster analysis at that length scale on both the cell data and the simulated null models, constructed the cluster size distributions  $\rho(k)$ , and looked for statistical deviations from randomness. Doing so, I showed that the LAT proteins in inactive T-cells tend to cluster in smaller structures, preferring to be found in dimers. However, upon T-cell activation, these structures all but vanish, likely representing the recruitment of other signaling molecules to the membrane [22]. I then discuss extensions of this technique to multi-colored pattern sets, i.e. patterns with multiple tagged molecules, as well as to spatially and temporally heterogeneous pattern sets.

### 1.5.3 Quantifying the spatial asymmetry of ACA mRNA in migrating *Dictyostelium discoideum* cells

My third contribution [50] represents answering both questions for sets of spatial patterns, the difference being that the actual physical location of each point pattern was unknown. Specifically, I was interested in the perceived asymmetry of ACA mRNA in *Dictyostelium discoideum* cells. The data was derived from Fluorescent *in situ* Hybridization experiments (FISH) of “streaming” *D. Discoideum*

cells, which resulted in diffraction-limited fluorescent images of both ACA mRNA and cAR1 mRNA molecules. It is already known that ACA localizes to the back of migrating cells [33]. Across dozens of images from multiple experiments, there seemed to also be a pattern where ACA mRNA fluorescence was strongly localized towards the rear of migrating cells, whereas the cAR1 mRNA was distributed more uniformly. Further, this seemed by eye to be a trend most highlighted in cells that were just joining streaming structures, and was less accentuated for cells that were near an aggregate center.

I sought to quantify this trend by inferring the underlying mRNA locations from the diffraction-limited images. Because of the amorphous shapes that *D. discoideum* take, I determined that simple moment analysis of the fluorescent intensity was inappropriate. For all the images, I performed a peak-finding algorithm and fit Gaussian point-spread functions to the peaks found, measuring both the peak intensity,  $I_0$ , and the standard deviation,  $s$ . I then thresholded these values until I achieved finely peaked distributions whose average was indicative of a fluorescent “unit” of the pattern, likely representing a small number of mRNA molecules. Then, for each cell, I took a region-of-interest approach (ROI) to quantify localization, dividing the cell front and back with arbitrarily shaped polygons and quantifying the number of mRNA units present in an ROI using two different techniques. The first technique employed standard error propagation using the means and standard deviations of the distributions of  $I_0$  and  $s$  to calculate both the average number of mRNA present and an estimate of the error. The second technique used a simulation approach, rebuilding each image dozens of times with randomly placed mRNA

units to match the observed fluorescence, each time yielding a potential realization of the underlying molecular pattern. Then, the mRNA population in a specified ROI was measured by the average number of mRNA units placed in that region, with the error being the standard error of the mean. While both techniques yielded similar results, I found that the simulation method had a stronger statistical power in differentiating populations.

Using these two techniques, I quantified the front/back localization of both ACA mRNA and cAR1 mRNA in migrating cells. I found no real differences in size or fluorescent intensity between the two molecular units, and indeed found that cAR1 mRNA had no overall trend of localization in the cell, preferring to distribute randomly. In sharp contrast, ACA mRNA showed a strong tendency to localize towards the back of migrating cells, with the strongest trend being found in cells towards the back of streams. When the number of mRNA units was compared also to the local concentration of their respective molecule (as measured by the integrated local intensity), a strong correlation was found between the rear proportion of ACA mRNA and the proportion of the cell concentration of ACA localized at the rear, with both being strongly localized at the rear of cells that are just joining streams. This phenomena was further investigated by looking at time lapse images of cells recovering from treatment with cycloheximide which inhibits protein synthesis. Initially, ACA mRNA were not visible due to the inhibition. Upon recovery, the return of ACA mRNA to normal levels maintained the assymetry, suggesting that the translation machinery of ACA localizes at the back of polarized cells, allowing local replenishment of ACA during cell migration.

## 1.5.4 Modeling and analyzing the patterns of externally guided chemotaxing cells

The fourth and final contribution [51] represents a turning of the two questions on their head, seeking first to design a temporally varying spatial point pattern with visibly distinguishable behavior that could not be quantified with available techniques. I then demonstrate a new analysis technique, and use both the model and technique to make theoretical predictions. Specifically, I was interested in the chemotactic motion of *D. Discoidium* cells in the presence of external chemoattractant gradient. When external chemical signals are very strong, the autocrine signal relay of *D. Discoidium* is washed out and cells migrate towards the source of the signal individually. When the external signal is weak, the self-generated chemical signals of the ensemble dominate and the cells form non-migrating aggregates. However, for moderate concentrations of external chemical signals, the autocrine signaling leads to cells coordinating their migration, resulting in streams of cells that move together towards the signal. As is to be expected, the definitions of what is “strong”, “moderate”, and “weak” depend on the density of the cells, since more cells secrete more chemoattractant. Though this motion is often measured, it is hard to quantify. Standard measures applied to cell migration experiments are various “chemotactic indices” (abbreviated *CI*). Always a first order quantity, a common definition [36] for *CI* is the average over the ensemble of the projection of the cell’s front-to-back alignment onto the direction of the external chemical source. Perfect migration yields  $CI \approx 1$ , random migration yields  $CI \approx 0$ , with most real

experiments showing values in between. This is similar to Viscek’s mean progression index [52], which also measures the mean orientation of the ensemble.

I wanted to investigate these three stages of dynamic steady state and how different model perturbations might effect the critical concentrations, ( $C_0$ ) and densities ( $\rho_0$ ) necessary for their observation. I began by developing a computational model for individual *D. Discoidium* migration, viewing the cell’s motion as the result of three modules: directional sensing, protrusion generation, and center-of-mass motion. These cells individually effect the global chemical signal through local secretion and global degradation. They also interact with each other mechanically through soft-core repulsive forces. I simulated these cells in a box with a fixed concentration at the top and bottom, applied periodic boundary conditions to the cell’s motion, and let the system reach steady state. I found many combinations of  $C_0$  and  $\rho_0$  that yielded one of the three dynamic steady states. By the construction of the simulation, the difference between streaming and individual motion is lateral coordination, with both types of cells still migrating upward. This means the for both conditions,  $CI \approx 1$ , demonstrating that it is not an adequate method for exploring perturbations to the different dynamic steady states.

I instead demonstrated a technique for quantification based on measures on the ensemble’s spatio-temporal contact network. Our rationale was that randomly migrating cells will break more contacts then those that are stream, which will break more than those that have collapsed into an aggregate. To show this, at each time,  $t$  I calculated a contact network, where each cell was a node and edges were present if  $d_{ij} < r_{cell}$ . I then looked at a time  $t + \Delta t$  to see what fraction of

these contacts had been broken,  $F_s$  ( $\Delta t$  was always chosen to be longer than the individual cell's persistence time). I then ran the simulation for the same density and concentration with cell secretion removed and calculated the fraction of broken links in that context,  $F_0$ . I showed then that the ratio,  $F_s/F_0$  correctly identifies the three dynamic regions, allowing us to create a psuedo phase diagram as a function of the two parameters  $\rho_0$  and  $C_0$ . I then used the model and the fraction of broken links to investigate the effects of two perturbations: the addition of adhesion and the removal of degradation. I found that because adhesion opens up another mechanism for motion coordination, it both strengthened coordination in the absence of chemical secretion and in the presence. In contrast, removing the ability for the cells to degrade the chemical signal all but destroyed the streaming phase. This occurs due to old signals not being removed in place of new ones while the global concentration continues to increase. I discuss the extension of the model to paracrine signal relay, which is relevant for understanding immune cell migration.

## 1.6 Discussion

Whether sets of static point patterns, fluorescent images, or migrating points, my work has aimed to push forward the field of pattern set analysis for biological applications. The complex patterns analyzed represent structures whose size span three orders of magnitude: nanometer scale patterning of proteins, micron scale assymetries in the molecular distribution of mRNA molecules, and millimeter scale patterning of chemotaxing cells. However, the techniques developed were not re-

stricted to any one size. Contact and proximity networks brought insight to both protein clustering and cell streaming. For heterogenous cluster analysis of proteins, I leveraged the local number intensity to recreate randomized patterns for statistical comparison, a technique I similarly employed to infer the approximate numbers of ACA mRNA units in migrating cells.

These techniques, however, were adapted in each setting for the system at hand, and in particular involved mono-disperse patterns. Further extensions of my work involve interactions between different populations. For measuring clustering in signaling proteins, techniques are already being implemented that allow for the fluorescent labeling of many different proteins at once, meaning that both pattern set distinction and cluster analysis could be employed to investigate multi-molecular complexes. For FISH experiments with labeled mRNA, co-localization of different mRNA units could identify complementary processes in the cell machinery. And cell migration experiments with cell mixtures could highlight sorting in collective migration, whether through differential adhesion or the suggested bi-stability of receptor sensitivity [53]. Further, extending my migration model to paracrine chemotaxis would allow for the study of the group migration of neutrophils, who already show cell sorting with regard to defects in the sensing of LTB4 gradients.

## Chapter 2: Automatic sorting of point pattern sets using Minkowski Functionals

### 2.1 Overview

*This chapter is adapted from the paper in Physical Review E by Parker, Sherman, van de Raa, van der Meer, Samelson, and Losert [44] in which I developed and implemented a technique to use topologically derived measures to divide point pattern sets into like groups.*

Point pattern sets arise in many different areas of physical, biological, and applied research, representing many random realizations of underlying pattern formation mechanisms. These pattern sets can be heterogeneous with respect to underlying spatial processes, which may not be visually distinguishable. This heterogeneity can be elucidated by looking at statistical measures of the patterns sets and using these measures to divide the pattern set into distinct groups representing like spatial processes. We introduce here a numerical procedure for sorting point pattern sets into spatially homogeneous groups using functional principal component analysis (FPCA) applied to the approximated Minkowski functionals of each pattern. We demonstrate that this procedure correctly sorts pattern sets into similar groups both

when the patterns are drawn from similar processes and when the second-order characteristics of the pattern are identical. We highlight this routine for distinguishing the molecular patterning of fluorescently labeled cell membrane proteins, a subject of much interest in studies investigating complex spatial signaling patterns involved in the human immune response.

## 2.2 Introduction

### 2.2.1 Motivation: Why study point patterns?

Spatial points patterns naturally arise in many areas of research in both the physical and life sciences, including ecology [54, 55], crime statistics [56], epidemiology [57], economics [58], seismology [59], material science [60], and astronomy [61]. Whether the points represent molecules, trees, cell phone users, or entire galaxies, the spatial distributions of point patterns belie the underlying stochastic processes that govern the pattern's formation.

A new area of point pattern analysis involves studying the molecular patterning of proteins on the surfaces of cells. Due to photo-activated localization microscopy (PALM) [23], a new super-resolution microscopy technique, cell biologists are now able to measure the spatial distribution of fluorescently-tagged membrane proteins and determine the response of the molecules to different stimuli (see 2.1). By fixing the cells on a slide and exposing them to laser light, researchers can activate molecules one by one in multiple cells, locating the center to within 20 nm. This new technique has resulted in a wealth of new point pattern data representing different

molecules and surface treatments, and quantitative analysis of these patterns can contribute much to understanding protein-protein and protein-membrane interactions [48, 62]. From a theoretical standpoint, each pattern is a pure realization of

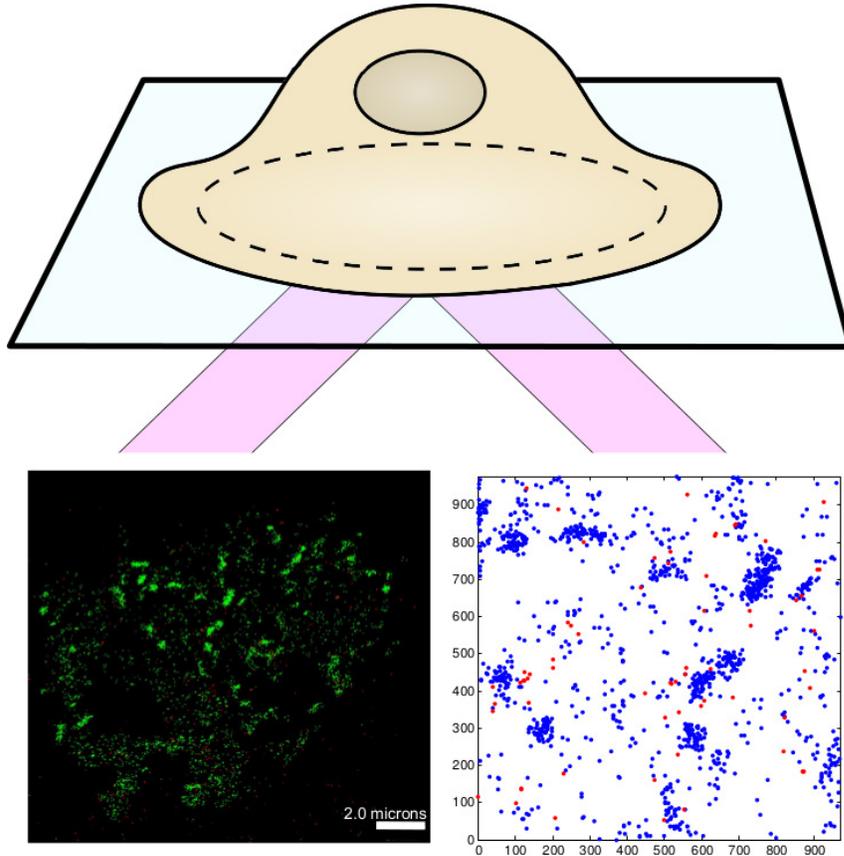


Figure 2.1: Using photoactivated localization microscopy, the fluorescently labeled proteins are localized by fitting a point spread function to the stochastically photoactivated molecules; the final pattern represents thousands of fluorescent images

an underlying spatial process and can be used to characterize that process. From a practical perspective, however, it takes many experimental realizations with finite systems to discern the underlying structure. Furthermore, if the point interactions are complex or the patterns are formed in complicated environments (such as the membrane of a cell), the amount of data needed to confidently quantify a process becomes large and cumbersome to analyze. This gives rise to the need to be able to

confidently divide large sets of patterns, sorting the patterns into smaller, homogeneous groups that can be analyzed further. In addition to simplifying analysis, this type of sorting can also provide researchers with quick information about the homogeneity of a process and the experimental parameters that affect this homogeneity.

### 2.2.2 Current methods for sorting patterns

The standard method of sorting pattern sets is as follows: For each pattern, one calculates a list of numerical summary characteristics (e.g. index of dispersion, Clark-Evans index). These can be regarded as the "coordinates" of a pattern, to which distance-based clustering algorithms can be applied [37]. This approach presents the researcher with the task of deciding which characteristics to use, how to compare them (normalizing, z-scores, etc) as well as how many: too few may result in missing information, too many could result in redundancy. This adds nuance to the sorting, limiting the statistical conclusions that can be drawn, and making trustworthy automation of the sorting procedure for large pattern sets difficult to accomplish.

A more robust sorting technique has been developed, where patterns are sorted using functional principal component analysis (FPCA) on smoothed second-order functionals of the patterns. This routine treats the point set as a set of functions,  $\{a_i(t)\}$ , such as the pairwise correlation function,  $g(r)$  [45]. The coordinates of each pattern are then calculated by finding the eigenfunctions and corresponding

eigenvalues of the equation:

$$\int v(s, t)w_i(s)ds = \lambda_i w_i(t). \quad (2.1)$$

Here  $v(s, t)$  is the variance-covariance function of the set of functionals  $a(r)$ , defined as

$$v(s, t) = (N - 1)^{-1} \sum_{i=0}^N (a_i - \bar{a}_i(t))(a_i(s) - \bar{a}_i(s)). \quad (2.2)$$

The “score” of the  $i$ -th pattern on the  $j$ -th principal component is then  $\int a_i(t)w_j(t)dt$  (see Refs. [63,64]). Like standard PCA, the eigenvalues form a positive decreasing set whose truncated sum represents the total variance encapsulated in the included principal components. For automation, one can simply set a threshold for the amount of variance to be included, which in turn prescribes the number of coordinates to be used. This feature removes the arbitrariness of sorting patterns via the standard method, making FPCA an easily automatable way of quantifying the difference between patterns. Illian et al [45] showed that this routine was robust to point location uncertainty approaching 20% of the window size. <sup>1</sup>

However, spatial processes can create patterns with more structure than second-order functionals can measure. The Neyman-Scott process (NS), introduced to study galaxy clustering, involves randomly distributed parent points generating clusters of varying size. The complexity of the parent/daughter interaction gives rise to families of NS processes with the same pairwise correlation function [37, 42, 65], despite

---

<sup>1</sup>See Supplemental Material at <http://link.aps.org/supplemental/DOI> for more information regarding how to incorporate spatial uncertainty into point pattern analysis.

underlying spatial differences in the patterns.

Baddeley and Silverman [41] also introduced a cell process which is built by partitioning a domain into cells of equal size which are then filled with a varying number of uniformly distributed points. Though the process is rather regular, they showed analytically that their process was indistinguishable from a Poisson process when considering second-order functionals of the pattern [41], meaning that higher-order functionals must be used to resolve this ambiguity.

### 2.2.3 From points to disks

In this paper, we apply the proximity measure of FPCA to the approximated Minkowski functionals of point patterns [43]. These functionals are calculated by centering a disk on each point and analyzing the topology of this secondary pattern of overlapping disks as a function of the radius. Since the overlap can be very complex, involving all possible combinations of individual points, these functionals depend on all orders of interaction simultaneously. This makes them a more complete “fingerprint” for pattern comparison [43, 66]. These functionals have enjoyed marked success in astrophysics [67], soft matter [68], and fluid turbulence [69].

For completeness, we first explain the Minkowski functionals and how they are applied to point pattern analysis. We then demonstrate the sorting procedure by clustering sets of patterns of both synthetically generated data and biological data representing the spatial distributions of membrane proteins. Using both agglomerative and divisive clustering algorithms, we show that this procedure outperforms

FPCA clustering with second-order functionals, and in general we demonstrate it to be a viable method for automatically sorting point pattern sets.

## 2.3 Outlining the Procedure

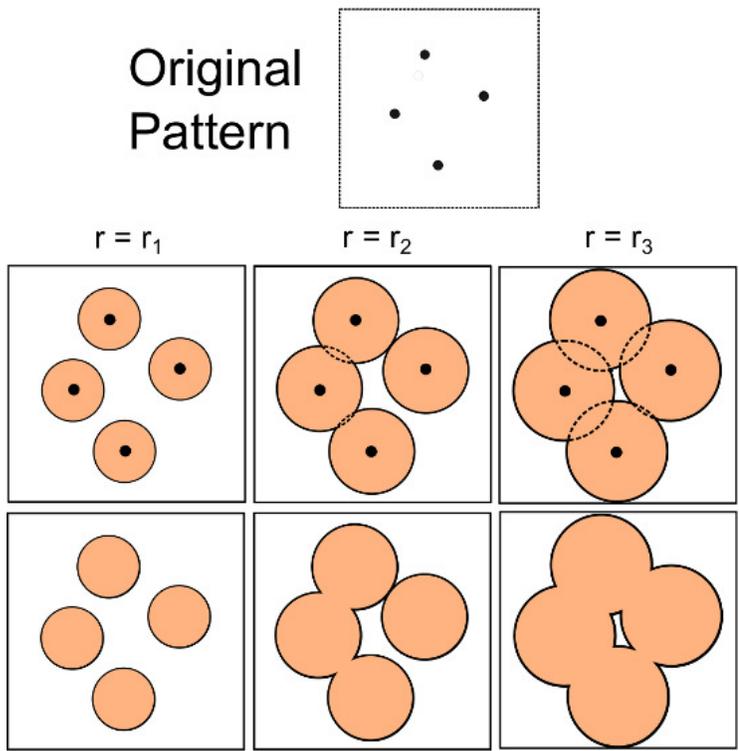
### 2.3.1 Minkowski functional analysis of point patterns

The first step in two-dimensional (2D) Minkowski functional analysis [43, 66] is to turn a point pattern into a “secondary pattern” by centering a disk of radius  $r$  at the center of every point (see Fig. 2.2).<sup>2</sup>

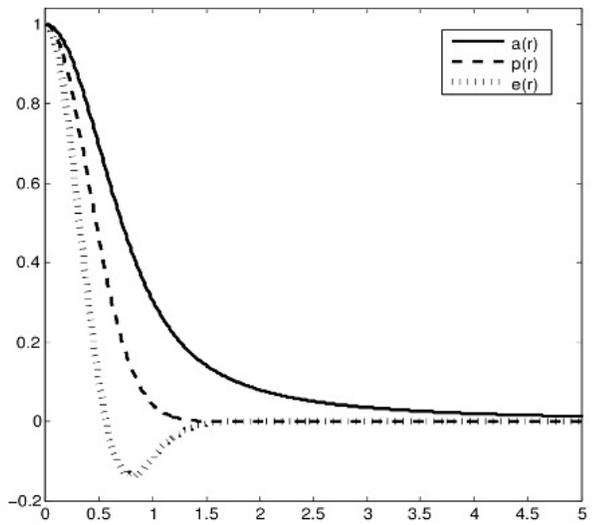
If the radius is large enough, some of these disks will overlap. By combining the overlapping disks, a pattern of differently shaped objects is formed. The total area,  $A$ , of this collection of objects is then just the total area of the disks excluding any overlapping area. This is the first Minkowski measure. The second Minkowski measure, the total perimeter,  $P$ , of the pattern is the perimeter of all of the shapes, which is again reduced from the perimeter of the individual disks because of overlaps. The Euler number,  $\chi$ , is the final Minkowski measure, defined as the total number of distinct shapes or components in the window (created by the overlapping disks) minus the number of holes.

---

<sup>2</sup>In this paper, we deal only with 2D patterns, but our procedure is easily generalizable to patterns of any dimension.



(a)



(b)

Figure 2.2: (a) The Minkowski functionals are calculated by imposing disks on the point pattern. This new secondary structure can be characterized using topological measures, which vary for different radii (b) The three reduced Minkowski functionals for a 2-D Poisson (random) process. These functionals are unitless due to the normalization by the same measure one would expect for a set of non-overlapping disks.

By calculating each of these measures first for small radii, where the disks do not overlap, and growing the radius after each calculation until the entire pattern window is covered, the three Minkowski functionals,  $A(r), P(r), \chi(r)$  are approximated. Because at each radius, the Minkowski measures depend on the locations of all of the points simultaneously, these functionals include information about every type of spatial structure present in the pattern, completely characterizing it (a consequence of Hadwiger's theorem from integral geometry, see Ref. [70]). This feature makes the Minkowski functionals a more complete measure of the underlying point interactions, including information from all possible groupings of points.

When comparing patterns, one actually uses the reduced Minkowski functionals, namely the Minkowski functionals for the pattern divided by what is expected for a set of non-overlapping disks. These are given by

$$a(r) = \frac{A(r)}{\pi N r^2} \tag{2.3}$$

$$p(r) = \frac{P(r)}{2\pi N r} \tag{2.4}$$

$$e(r) = \frac{\chi(r)}{N} \tag{2.5}$$

The functionals for a Poisson process are shown in Fig. 2.2.b. The analysis in this paper relies exclusively on these reduced functionals, so we do not differentiate between the two.

### 2.3.2 Sorting the patterns

Our aim is to automatically sort patterns by performing FPCA on their approximated Minkowski functionals, clustering the patterns with their individual scores on the principal components. We do the same with the pairwise correlation function so that we can directly compare our method with that of Ref. [45]. For each pattern set, we use enough principal components to account for 95% of the variation. For the Minkowski functionals, we calculate the principal component scores individually for the area, perimeter, and Euler number and then concatenate the scores into a larger vector and use a Euclidean norm (which corresponds to the Mahalanobis distance [71]). Then, we use these scores as coordinates, applying two different clustering algorithms:

- Ward's method [72]. An agglomerative technique which seeks to minimize the total intercluster variance of the distances between objects. We chose this method because it is well known to the pattern analysis community and allows us to directly compare our method with that of Illian *et al* [45].
- Fast weighted modularity [73,74]. To implement this routine, we first calculate the pair-wise Euclidean distance between all patterns,  $D_{ij}$ , and transform our pattern set into a weighted graph with edge weights

$$W_{ij} = \max(D_{ij}) - D_{ij}. \quad (2.6)$$

Then, this algorithm aims to maximize “modularity” of this weighted network, by dividing the set into groups where the total weight of edges between members of the same group is higher than the total weight of the edges between members of different groups. We chose this method because it is a global clustering routine with large popularity in the cluster analysis literature, and because the software implementation is able to work with very large data sets (millions of objects).

By utilizing both cluster analysis algorithms, we can verify whatever results we obtain, and more completely demonstrate the efficacy of our sorting method using the Minkowski functionals.

## 2.4 Testing our sorting method

We now apply this procedure to three different data sets. These sets of patterns highlight three possible situations in which one would sort patterns: (i) comparing different systems, (ii) varying a parameter in an experiment, and (iii) comparing different components of a bi-disperse system<sup>3</sup>. Each set is comprised of two groups which have an a priori cluster structure. We then apply the sorting technique, which allows us to calculate the percentage that is miss-classified ( $P_{MC}$ ) by looking at the fraction of patterns that are assigned to a group that is dominated by a different pattern type.

---

<sup>3</sup>See Supplemental Material at <http://link.aps.org/supplemental/DOI> for specifics regarding software and computational methods used

### 2.4.1 Data set 1: Two Strauss processes

The Strauss process [47] is a germ-grain pattern simulation model specified by two parameters, a radius  $r \in \mathbb{R}^+$  and an interaction parameter  $\gamma \in [0, 1]$ . The interaction parameter determines if grains of radius  $r$  will be allowed to overlap during the formation of the pattern (see Fig. 2.3.a).

If  $\gamma$  is small, there is a strong repulsion between grains, where  $\gamma = 0$  yield a hard-core process. If  $\gamma$  is close to 1, the repulsion is weak, where  $\gamma = 1$  yields a completely random process.

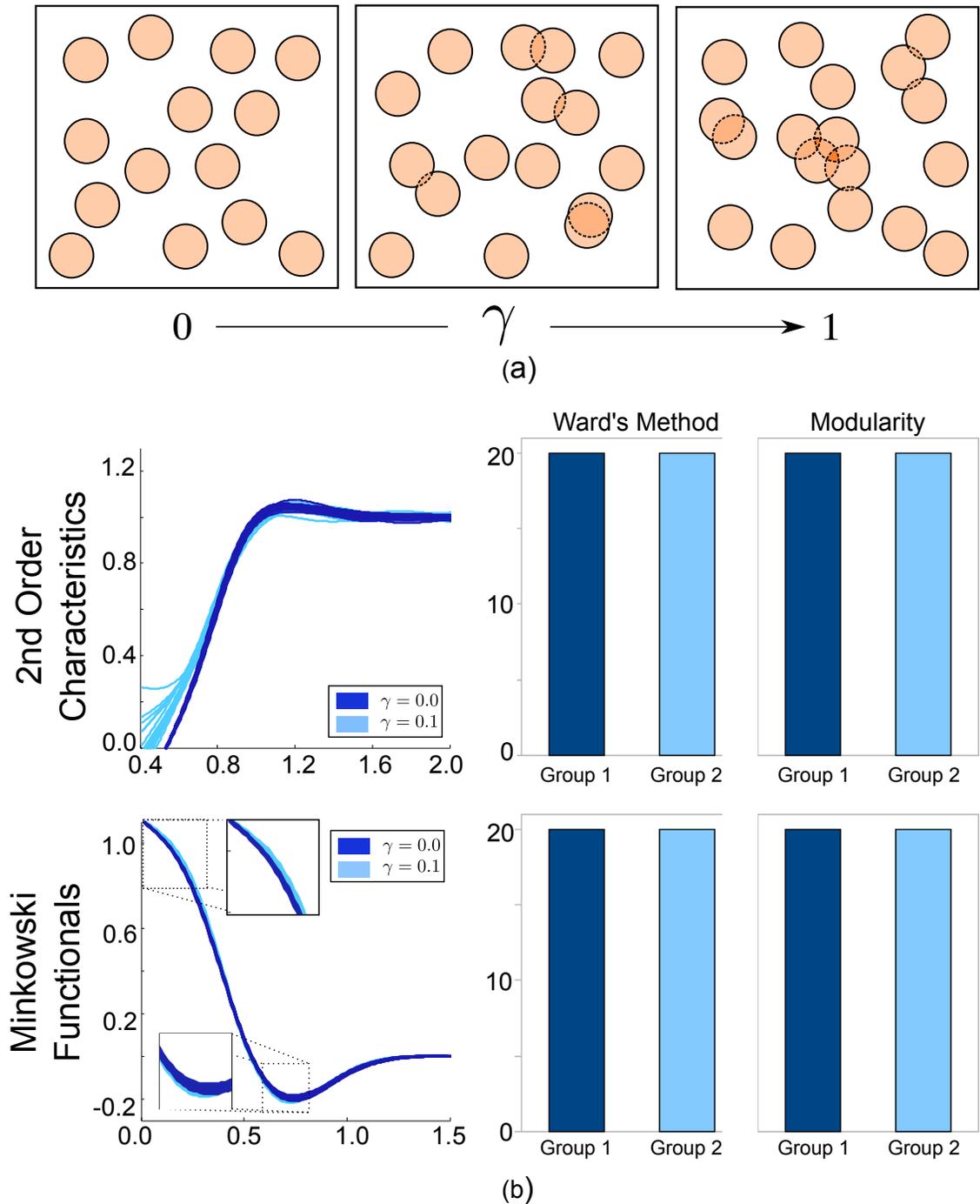


Figure 2.3: (a) The regularity of a Strauss process is completely determined by  $\gamma$ , the interaction parameter. For  $\gamma = 0$ , no overlaps are allowed. For  $\gamma = 1$ , all overlaps are allowed. (b) The left most panes display the  $g(r)$  and  $\chi(r)$  for the 40 simulated Strauss processes. On the right are the results of using FPCA scores to divide the pattern set into two groups. As can be seen, both  $g(r)$  and the Minkowski functionals can perfectly separate the set into two groups corresponding to different values of  $\gamma$ .

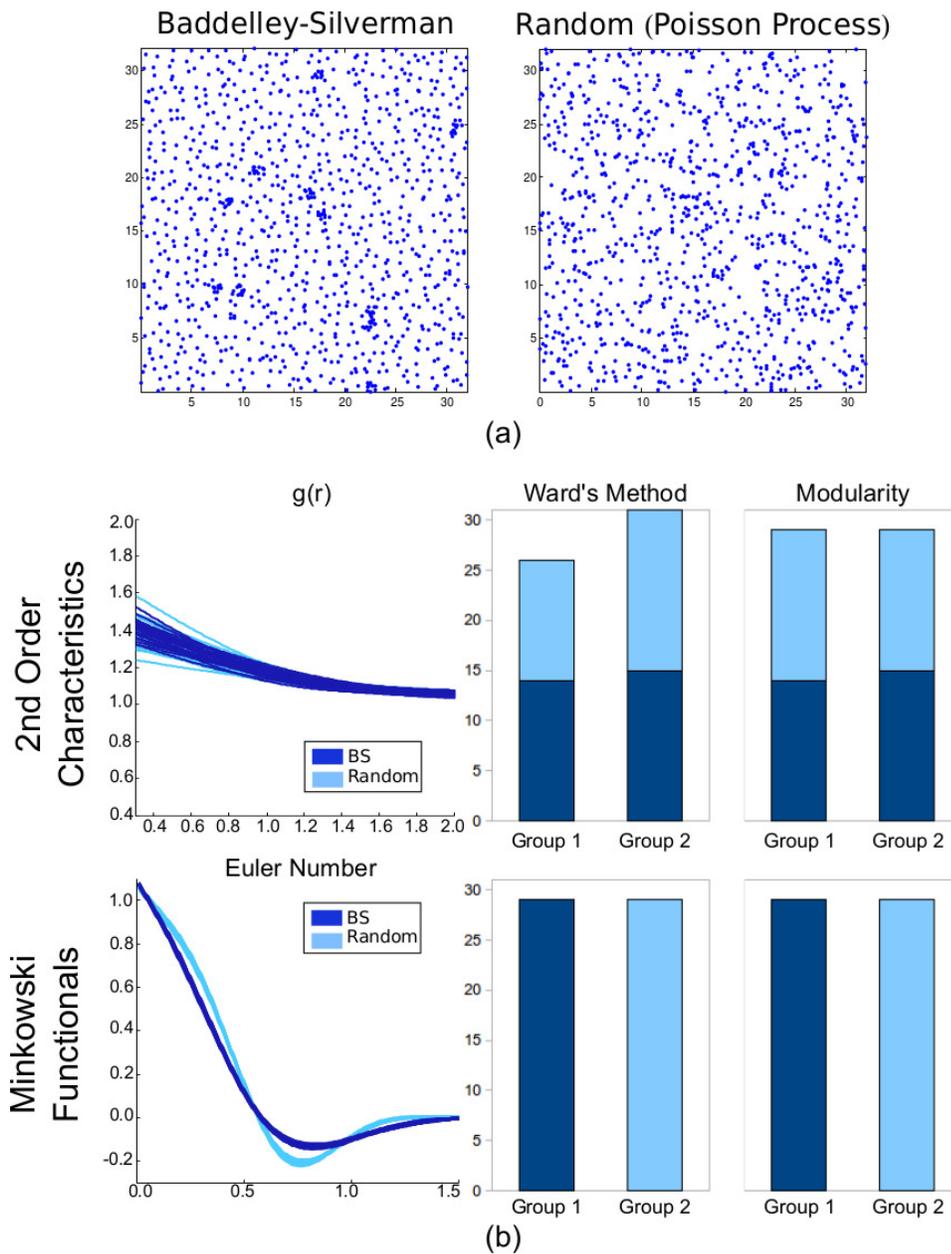


Figure 2.4: (a) A Baddeley-Silverman (left) process side-by-side with a Poisson process (right). Despite the visible differences, the pairwise correlation functions are identical (b) The left most panes display the  $g(r)$  and  $\chi(r)$  for the 58 patterns simulated. On the right are the results of using FPCA scores to divide the pattern set into two groups. As can be seen, FPCA sorting with  $g(r)$  creates two perfectly heterogeneous groups, while FPCA sorting with the Minkowski functionals groups the patterns correctly.

In Ref. [45], it was reported that even for comparison of pattern sets with similarly strong repulsion ( $\gamma = 0.0$  and  $\gamma = 0.1$ ) the pairwise correlation function was able to effectively distinguish different Strauss processes. We here repeat this test with 20 patterns each, fixing the number of points at  $N = 1000$ , and letting  $r = 0.025$ . We also fix the number intensity to be  $\lambda = (2r)^{-2}$ , which forces interaction between the points.

As can be seen in Fig. 2.3.(b), both  $g(r)$  and the Minkowski functionals are able to distinguish the two Strauss processes, separating the pattern set into two homogenous groups. This is to be expected for  $g(r)$ , as second-order interactions dominate the process, and is consistent with the findings of Ref. [45].

#### 2.4.2 Data set 2: Baddeley-Silverman vs. random

The Baddeley-Silverman process [41] is built by partitioning a domain into a grid and moving from box to box, distributing  $N$  points in each box uniformly.  $N$  itself is a random number, taking on the values 0,1, and 10 with probabilities 1/10, 8/9, and 1/90, respectively. This causes the process to be rather regular, but with some strong clustering occurring every now and then (see Fig. 6).

Since  $E[N] = \text{Var}[N] = 1$ , it can be shown that the Baddeley-Silverman process shares all of the same second-order characteristics as a Poisson process. In ref. [43], the Minkowski functionals were shown to be able to distinguish these two processes. Therefore, we expect to see proper sorting when using  $a(r)$ ,  $p(r)$ , and  $e(r)$ , and failure using  $g(r)$ .

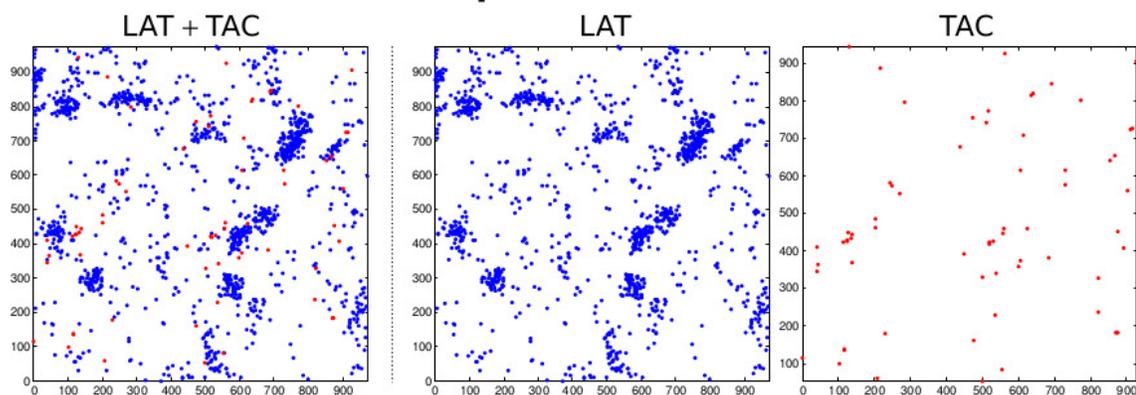
We simulated 29 Baddeley-Silverman processes and 29 Poisson processes, fixing the point number  $N = 1024$ . Using both the pairwise correlation function and the Minkowski functionals, we sorted the pattern sets into two groups. As can be seen in Fig. 2.4.b, the pairwise correlation function fails to sort the patterns correctly, creating two heterogeneous groups ( $P_{MC} > 40$ ). However, the Minkowski functionals successfully divide the pattern set into two homogeneous groups.

### 2.4.3 Data set 3: Bi-disperse patterns of inter cellular proteins

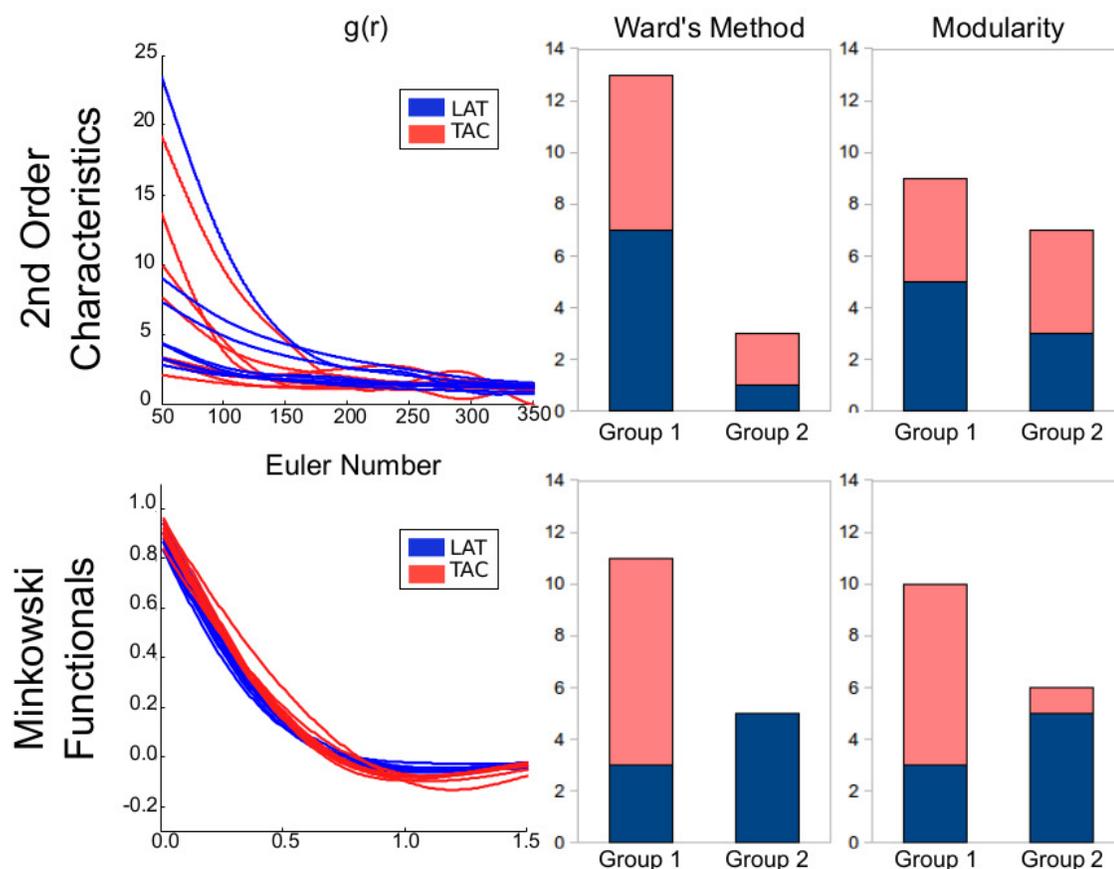
For an application to an experimental data set, we look at super-resolution images of two proteins residing at the membrane of immune cells [see Fig. 2.5.(a)].

One protein under study is LAT, short for “linker for activation of T cells”, a naturally occurring protein crucially involved in the reactions that regulate T cell antigen-dependent activation, a critical event in the adaptive immune response. LAT proteins have been seen to form clusters on the membrane with potentially complicated hierarchies [48]. However, the membrane of the cell can have a first-order effect on the molecular patterning of membrane proteins. It has been found that the location of other membrane protein clusters often correlates with how close the membrane is to the surface, and anti-correlates with regions of high membrane fluctuations [8].

## Bi-disperse Molecules



(a)



(b)

Figure 2.5: (a) The two proteins are both dispersed in the cell membrane. LAT (blue) and TAC (red) proteins are separately tagged, allowing them to be visualized separately. (b) The left most panes display the  $g(r)$  and  $\chi(r)$  for the 16 molecular patterns. On the right are the results of using FPCA scores to divide the pattern set into two groups. As can be seen, using Minkowski functionals with FPCA improves the differentiation of the two sets.

Another protein, TAC (the alpha chain of the interleukin-2 receptor), can also be localized and differentiated from LAT by tagging with a different fluorescent molecule and using two different lasers with different wavelengths. TAC is a membrane protein that does not form clusters, instead distributing uniformly in regions where protein-membrane interactions have not excluded proteins. This means that TAC can serve as a membrane marker when studying the clustering of other proteins. Since LAT and TAC are part of separate signaling pathways, they also do not interact biochemically [75]. Therefore, upon sorting, we should get two homogeneous groups representing the two different molecules.

Applying FPCA on the approximated pairwise correlation functions of these data sets again yields strongly heterogeneous groups ( $P_{MC} \approx 50\%$ ). This is visible in the pair-correlation functions [Fig. 2.5.(b)], where the individual patterns exhibit large variability. In contrast, because the Minkowski functionals consider more than just second-order interactions, the *Euler number* is able to visibly distinguish the molecular patterns, and the pattern sorting is improved ( $P_{MC} \approx 25\%$ ).

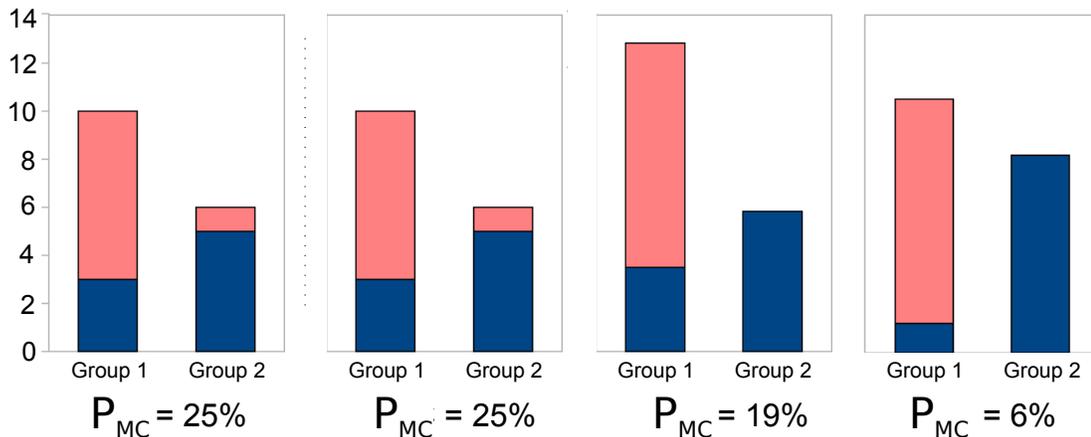


Figure 2.6: From left to right: Sorting with all three functionals, the *area*, the *perimeter*, and the *Euler number*. Considering the individual functionals (sorting using weighted modularity), the Euler number outperforms the other two, only misidentifying one pattern

Further success is achieved if we look at how FPCA sorting with each functional performs on its own [Fig. 2.6]. When only using the *area*, the sorting is identical to the sorting based on all three Minkowski functionals. However, sorting the LAT-TAC protein pattern set improves when just using the *perimeter*, and we achieve almost perfect classification when using the *Euler number*, only misidentifying one pattern. This is not surprising, since the *area* and *perimeter* are constrained to be smooth, positive, and monotonically decreasing, and thus cannot vary as much while the *Euler number* can vary more wildly.

## 2.5 Conclusions and discussion

In this work, we have introduced the procedure to automatically sort point pattern sets using the approximated Minkowski functionals and FPCA. Using Strauss processes with strong repulsion, we have shown that this method can accurately

sort point pattern sets drawn from very similar processes. Further, this method also distinguishes Baddeley-Silverman processes from Poisson processes, a task which the pairwise correlation function perfectly fails to accomplish. We then found that when looking at experimental point patterns representing proteins, FPCA sorting using the Minkowski functionals outperformed FPCA sorting with the pairwise correlation function. This sufficiently demonstrates that the Minkowski functionals can successfully quantify the differences between pattern sets showing complex behavior.

We also found that FPCA sorting using only the *Euler number* strongly outperforms the other two. While mathematically the three functionals do completely classify a pattern, the *area* and *perimeter* may only be slightly different for different spatial processes. This means that error introduced when approximating the functionals numerically may blur these differences, resulting in improper sorting. Since the *Euler number* is allowed to vary more dramatically as the disks combine and holes form, it can visually distinguish very similar pattern sets, and therefore leads to better sorting.

Though we have presented this technique as a way to sort patterns into distinct sets for further analysis, the sorting itself can serve as an analysis tool. We are currently working to apply this tool to examine how the presence of different chemical cues effect the clustering of LAT proteins, as well as how T-cell activation perturbs the patterns. Because of the Minkowski functional's ability to robustly characterize a pattern, we can treat the membership of a pattern in a particular group as a sign of similarity between it and its co-members. We can therefore look at group statistics to determine what experimental variables change the molecular

patterns, and to what degree, allowing for systematic large-scale investigations of the membrane proteins and their response to different stimuli.

## Chapter 3: Heterogeneous cluster analysis of LAT protein point pattern sets reveal changes in clustering upon T-cell activation

### 3.1 Overview

*This chapter is adapted from the paper in preparation by Parker, Sherman, Barr, Samelson, and Losert [49] in which I developed and implemented a technique to perform clustering analysis on sets of LAT point patterns with heterogeneous intensity distributions.*

Recent advances in super-resolution microscopy techniques have allowed biologists to image the spatial locations of trans-membrane signaling proteins on the surface of T-cells to within tens of nanometers [23]. By studying these spatial point patterns over a variety of experimental perturbations, including both non-active and active cells, biologists interested in better understanding the human immune system can study the interactions of signaling molecules, allowing them to investigate the signaling hierarchy of the T-cell [48]. A large obstacle in looking at the clustering hierarchy of these proteins is the presence of the cell membrane, which during cell spreading creates a heterogeneous and constrained environment for the proteins to

settle in [8, 48]. This makes the point pattern of each cell arbitrarily different from other cells studied under the same experiment conditions. Indeed, the variation due to the changes in the membrane-imposed boundary conditions dominate any attempt to investigate features at the smaller scale. In this work, we demonstrate a technique that simulates individualized heterogeneous Poisson null models for each cell. Using these patterns, we calculate an average pairwise correlation function (PCF) to create a membrane-specific statistical baseline and then re-normalize the pairwise correlation functions of the LAT protein point patterns. We show that this reduces the cell-cell variability in the pattern set, resulting in a recognizable PCF of a cluster process with a length scale of roughly 500 nm. Using this length scale, we perform cluster analysis on both the original patterns and the simulated null models by generating proximity graphs and looking at the cluster size distribution for both the data and the simulation patterns. We find that in non-active T-cells, LAT proteins are more likely than random to form smaller structures at this length scale, in particular clustering in dimers. However upon activation, become similar to their random controls. A possible explanation for this is the presence of non-fluorescently tagged proteins being recruited to the membrane upon activation, thus causing the LAT proteins to not be found in nanoclusters.

## 3.2 Introduction

### 3.2.1 Point patterns in cellular systems

Spatial point patterns are a ubiquitous data type in scientific research. No matter the length scale or the object under investigation, understanding how to analyze located objects crops up in research on the fall patterns of rain drops [76], the distribution of gold nano-particles in materials [60], and even the clustering of galactic structures [61]. Recent advances in microscopic techniques have allowed for robust single molecule localization of proteins that reside on the membrane of cells [23]. This technique has been used to study trans-membrane proteins involved in the signaling cascade of T-cell activation. In these experiments, T-cells are allowed to settle on a glass slide and are imaged from below (see Fig. 3.1.A-B).. When the glass slide is coated with an antibody, fluorescently-tagged proteins bind to the surface and are stimulated sequentially with laser light. Through post-processing, these fluorescent events can be associated with single molecules, resulting in a spatial point pattern. These patterns can then be analyzed to infer interactions, both with each other and other physical structures.

### 3.2.2 Does T-cell activation affect apparent clustering in LAT proteins?

Of special interest to researchers is the LAT protein, a key play in human T-cell activation. In the body, helper T-cells are activated by coming into contact

with “antigen-presenting cells”, whose surface coating of antigens bind to the T-cells membrane-bound TCR proteins (“T-cell antigen receptors”). If the antigen presented is what that particular T-cell is adapted to recognize, the cell’s LAT proteins become phosphorylated and cluster which induces T-cell activation. This can be experimentally modeled by allowing T-cells to settle on glass slides coated in different antigens, which allows for fluorescent imaging of the LAT patterns both when the cell is active and when it is not. Fig. 3.1.c shows three representative LAT protein patterns: the left and center patterns are from non-activated T-cells and the right pattern is from an activated T-cells. Recent experiments have suggested that these proteins cluster on small scales [48], and a key question is whether the observed clusters are affected by T-cell activation and in what ways. By performing analysis on sets of these patterns, representing multiple cells, we could statistically confirm if clustering exists and to what extent it is affected by T-cell activation.

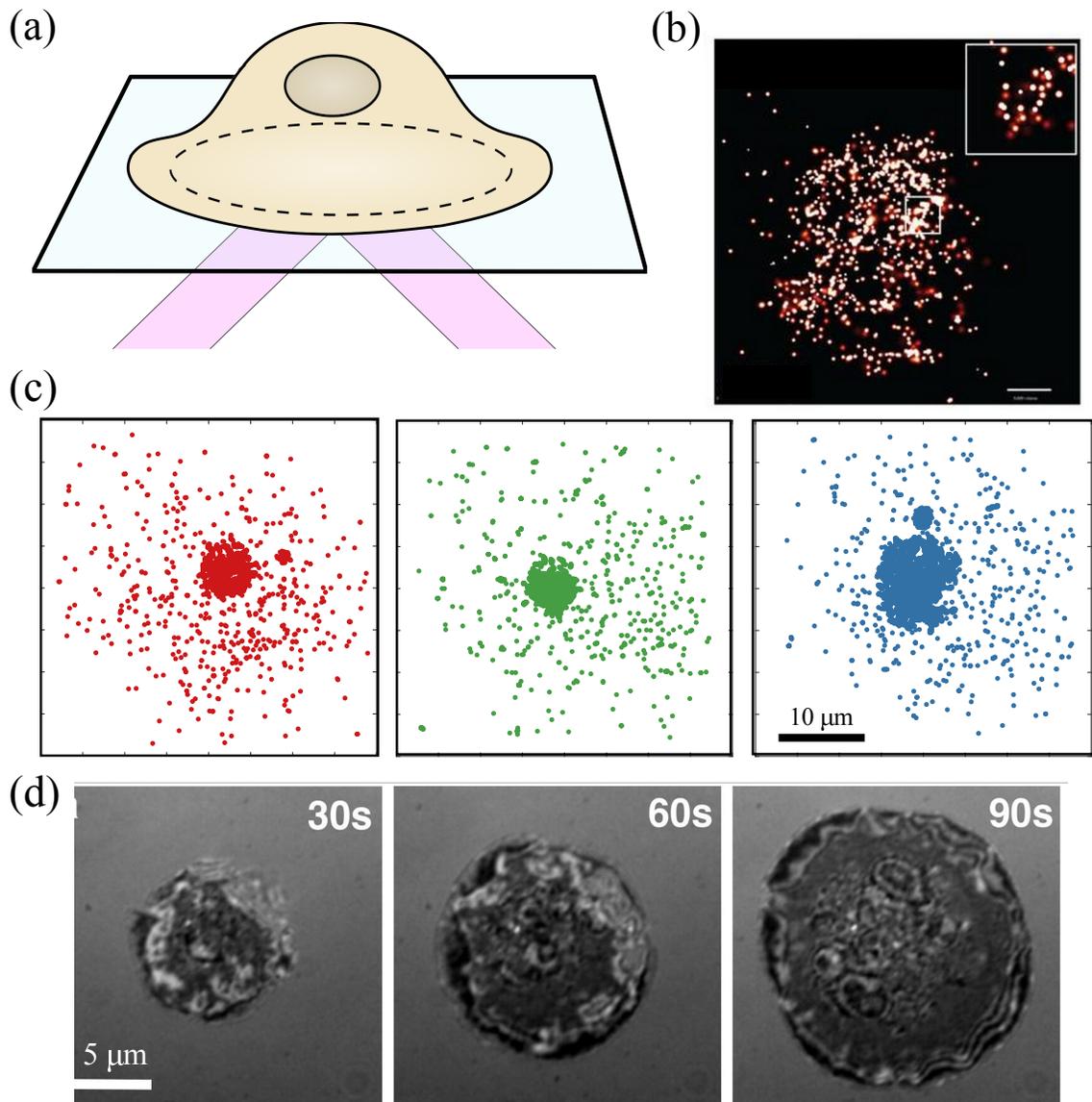


Figure 3.1: (a) Experimental setup. Glass surfaces are coated with antigens for cells to settle onto and imaged using super-resolution microscopy. (b) A single sample image of the fluorescence patterns seen in PALM experiments. (c) Three representative patterns of those derived from non-active cells (left and center) and activated T-cells (right) (d) Images of the spreading process of a Jurkat T-cell on a glass slide. Both the cell boundary and the variation in adhesion pattern affect the protein pattern (adapted from [8], used under the Elsevier user license)

### 3.2.3 An obstacle to analysis: the cell membrane

A naive approach to investigating clustering interactions between points would be to start with the pairwise correlation function (PCF) of the pattern,  $g(r)$ , which measures the probability of the distance between two points being within the range  $d = r \pm dr$ . For a pattern whose point locations are uncorrelated, the probability of a point being anywhere within a viewing window is constant, so  $g(r) = 1$ . Therefore,  $r$  values where  $g(r) > 1$  or  $g(r) < 1$  demonstrate deviations from the assumption that the point's locations are completely uncorrelated, with  $g(r) = 1$  here functioning as the the statistical “baseline” of the measure. In particular, one can investigate clustering by looking at the first minima, where the approximated  $g(r)$  and it's confidence intervals all decrease below the baseline of  $g(r) = 1$ . This distance then corresponds to  $1.5 \times r_c$ , where  $r_c$  is the characteristic length scale for clustering. Then one can connect all points who are nearer to each other than  $r_c$ , creating a proximity graph, and look at the distribution size distribution,  $\rho(k)$ , i.e. the distribution of the number of points,  $k$ , in each of the graph's connected component. For small  $r_c$ , the distributions should be approximately exponential [77]. To determine if clustering is present, a randomized null pattern can be simulated and analyzed similarly, and plots of  $\rho(k) - \rho_r(k)$  are made. For small  $r_c$ , the cluster size distribution should be approximately exponential [77], so the difference plot can highlight any deviations from random clustering present in the pattern. Doing this then for each pattern in a set, one can observe the general behavior of an underlying point process to the variation in experimental conditions.

However, for patterns of membrane bound proteins, there are many different factors controlling the point pattern formation. The process of cell spreading itself is variable, with both the location of the boundary and adhesion contact area forming over time (see Fig.3.1.d). The membrane not only imposes an arbitrary and amorphous boundary, creating a “virtual cluster”, but areas of poor adhesion can cause certain regions of the membrane to not be imaged, resulting in large scale heterogeneities [8, 48]. Looking at  $g(r)$  for three sets of LAT experiments (Fig. 3.2), we see that the pattern sets are strikingly heterogeneous, resulting in large confidence regions (region between the dashed lines). We also see that, despite these large confidence regions, none of the correlation functions approach random, i.e.  $g(r) \approx 1$ . These features are due to both the dominance of large scale pattern features and the feature variability from pattern to pattern. This means that any attempt to extract small scale details of the protein-protein interactions using standard techniques would be unsuccessful.

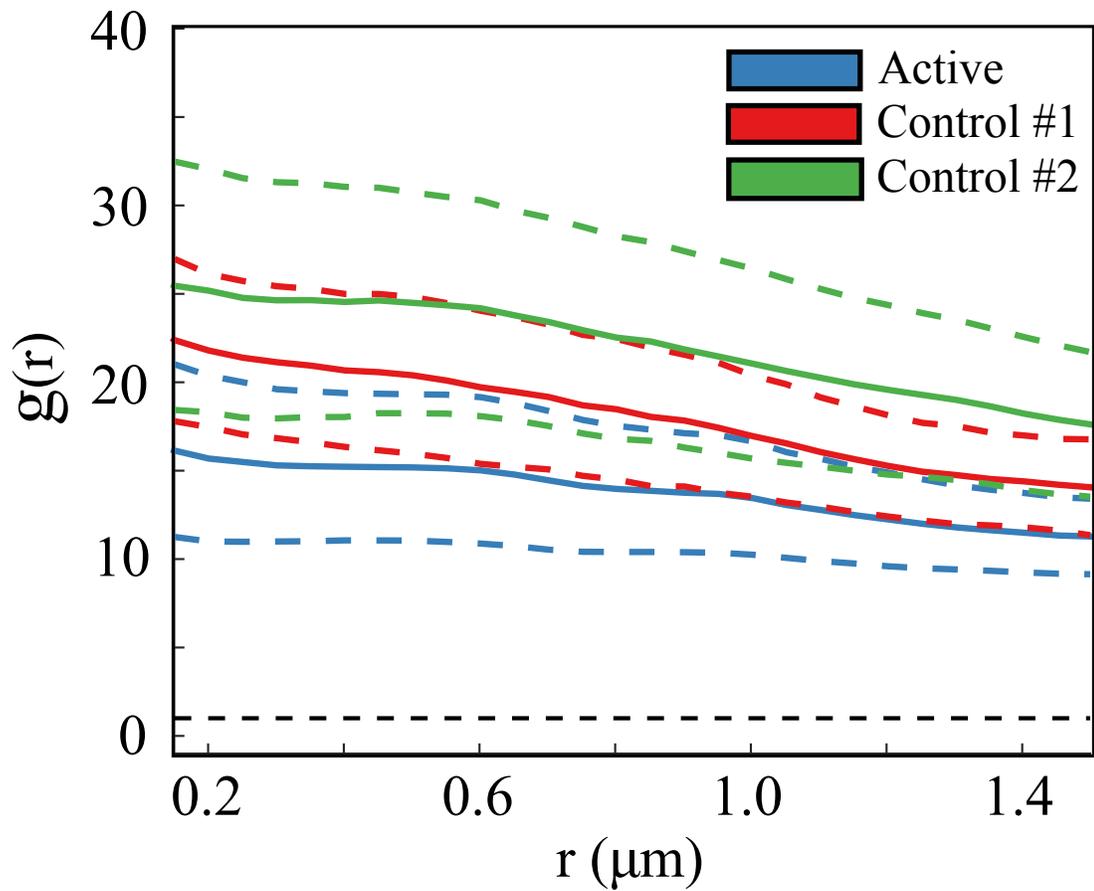


Figure 3.2: Standard pairwise correlation function for three sets of protein patterns. Dashed lines represent 95% confidence intervals.

### 3.2.4 Heterogeneous cluster analysis using simulations

In this work, we address this issue with a simulation based approach. For each pattern, we develop an individualized statistical baseline by simulating multiple heterogeneous Poisson point patterns that recreate the original pattern’s large scale intensity variations but blur small scale structure. Using the PCFs of these patterns, we re-normalize the original pattern’s  $g(r)$ , achieving the heterogeneous pairwise correlation function,  $g_{inh}(r)$  [78, 79]. This new quantity can be interpreted as the probability of two points having a distance  $d = r \pm dr$  given that the process was formed while interacting with a heterogeneous environment, which here is the membrane. Then, if a clustering length scale is observed, we can perform cluster analysis on both the cell data *and* the individualized null models to determine the clustering structure of the LAT proteins.

### 3.2.5 Data and Methods

For this work, we used three sets of LAT protein point patterns. These patterns were derived from PALM experiments with Jurkat cells, which are an immortalized line of the human helper T-cells [80]. Cells with fluorescently-tagged LAT proteins (PAmCherry) were allowed to settle on glass slides coated with antibodies to mimic an antigen presenting cell. For two of the pattern sets, a non-activating antibody was used ( $\alpha$ CD45 or  $\alpha$ CD28). For the remaining pattern set,  $\alpha$ CD3 $_{\epsilon}$  was used so that the T-Cell would activate. The non-active sets had 10 patterns each and the active set had 15. For each pattern, 100 heterogeneous Poisson patterns were

simulated and their average  $g(r)$  was calculated (for full details, see Supplemental Materials). Then  $g_{inh}(r)$  was calculated through simple division,  $g_{i,inh}(r) = g_i(r) / g_{i,ing}(r)$ . These were then averaged for visualization, with the confidence intervals being determined through error propagation. Bandwidth selection for the intensity approximation will, of course, effect the results of the analysis, with too small a bandwidth reproducing the original pattern and too large a bandwidth not correctly reproducing the intensity profiles. Since we were interested in clustering at the length scale of a few hundred nanometers, we used a bandwidth of 2.5 microns to ensure appropriate separation of length scales but not to be too comparable to the size of the cell. Fig. 3.3.A shows an original pattern and one of its heterogeneous Poisson null models. The absolute location of points are visibly different, but the overall number intensity has been preserved.

### 3.2.6 Results: LAT proteins show clustering at the nanoscale which is diminished upon activation

Fig. 3.3 shows the average  $g_{inh}(r)$  along with 95% confidence intervals. Compared to Fig. 3.2, we can see that the breadth of confidence regions (between the dashed lines) have been strikingly reduce, showing that our method has dramatically reduced the effects of larger scale heterogeneities. The average function that remains is recognizable as that of a cluster process, with all three sets showing a first minima around 750 nm, which gives a characteristic clustering scale of  $r_c \approx 500$  nm.

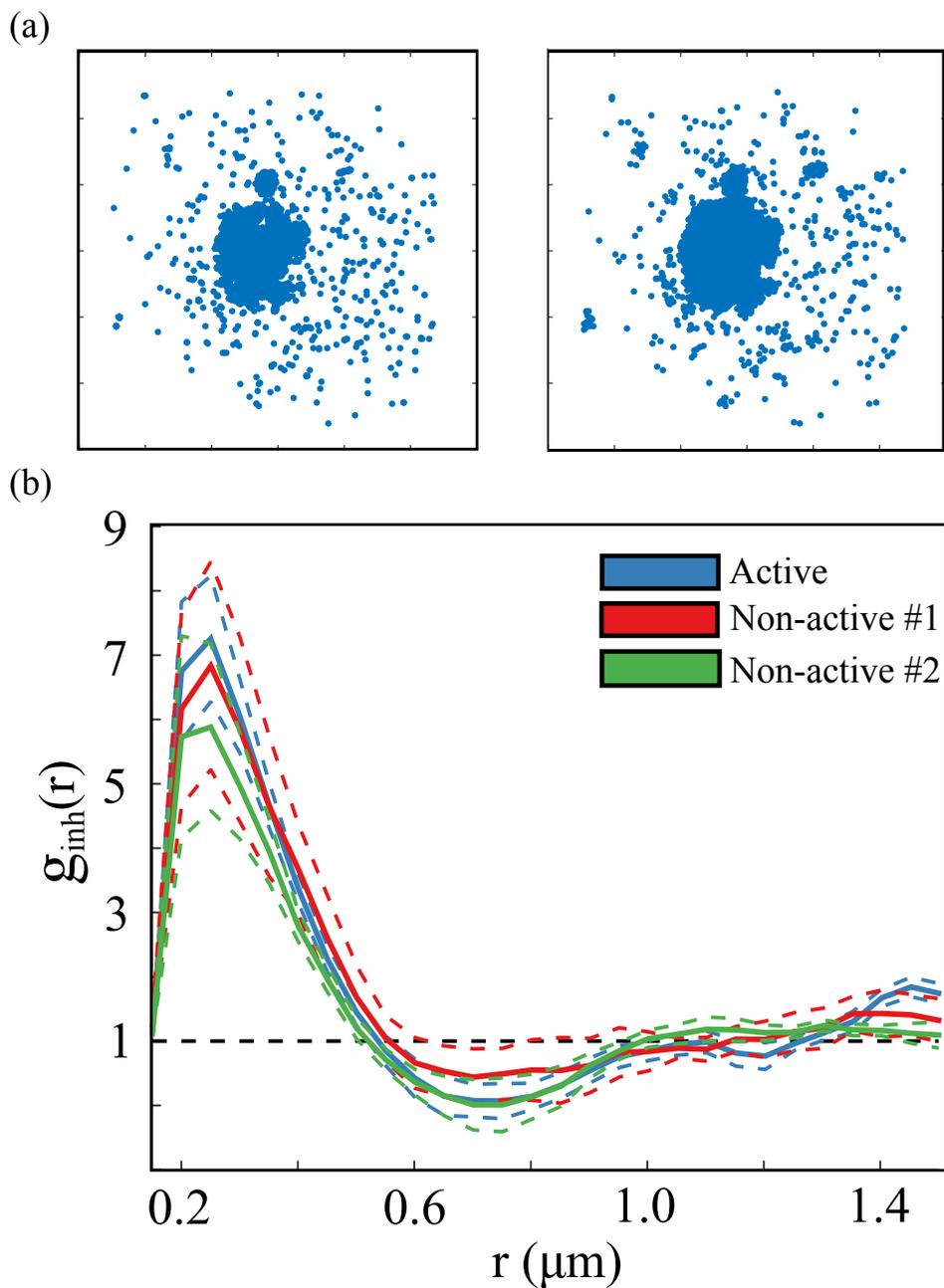


Figure 3.3: (a) (left) A protein point pattern derived from a PALM experiment and (right) a heterogeneous Poisson point process with the same intensity distribution. Small changes can be seen in the location of the points on the fringes, but the large scale features are preserved. (b) Heterogeneous pairwise correlation function for three sets of protein patterns. Dashed lines represent 95% confidence intervals. All three pattern sets collapse onto a single function reminiscent of clustering

Knowing that protein patterns can exhibit higher-order patterning [44], we performed cluster analysis with this length scale by connecting points to each other if their inter-point distance  $d_{ij} < r_c$  and then calculating the cluster size distribution,  $\rho(k)$ . We did the same for each of the cell's randomized patterns, and calculated  $\rho(k) - \rho(k)_{rand}$ . Points that were farther than  $r_c$  away from all other points were ignored, meaning that clusters must have two or more members ( $\rho(k = 1) = 0$ ). Looking at the quantity  $\rho(k) - \rho(k)_{rand}$  for the three data sets (Fig.3.4), we see a statistically significant presence of smaller clusters in non-activated T-cells, particularly dimers ( $k = 2$ ) and a diminished amount of larger sized clusters. This confirms prior inferences that LAT proteins cluster in small groups, primarily dimers [48]. Upon T-cell activation, the same measure shows that these dimers sharply diminish at this length scale in favor of slightly larger clusters, with the overall deviation of the patterns from random diminishing as well. This difference between the three cases is statistically significant, with a two-tailed t-test on  $\rho(2) - \rho_{rand}(2)$  for the active sets being statistically significant from both non-active sets with 99% confidence ( $p < .01$ ).

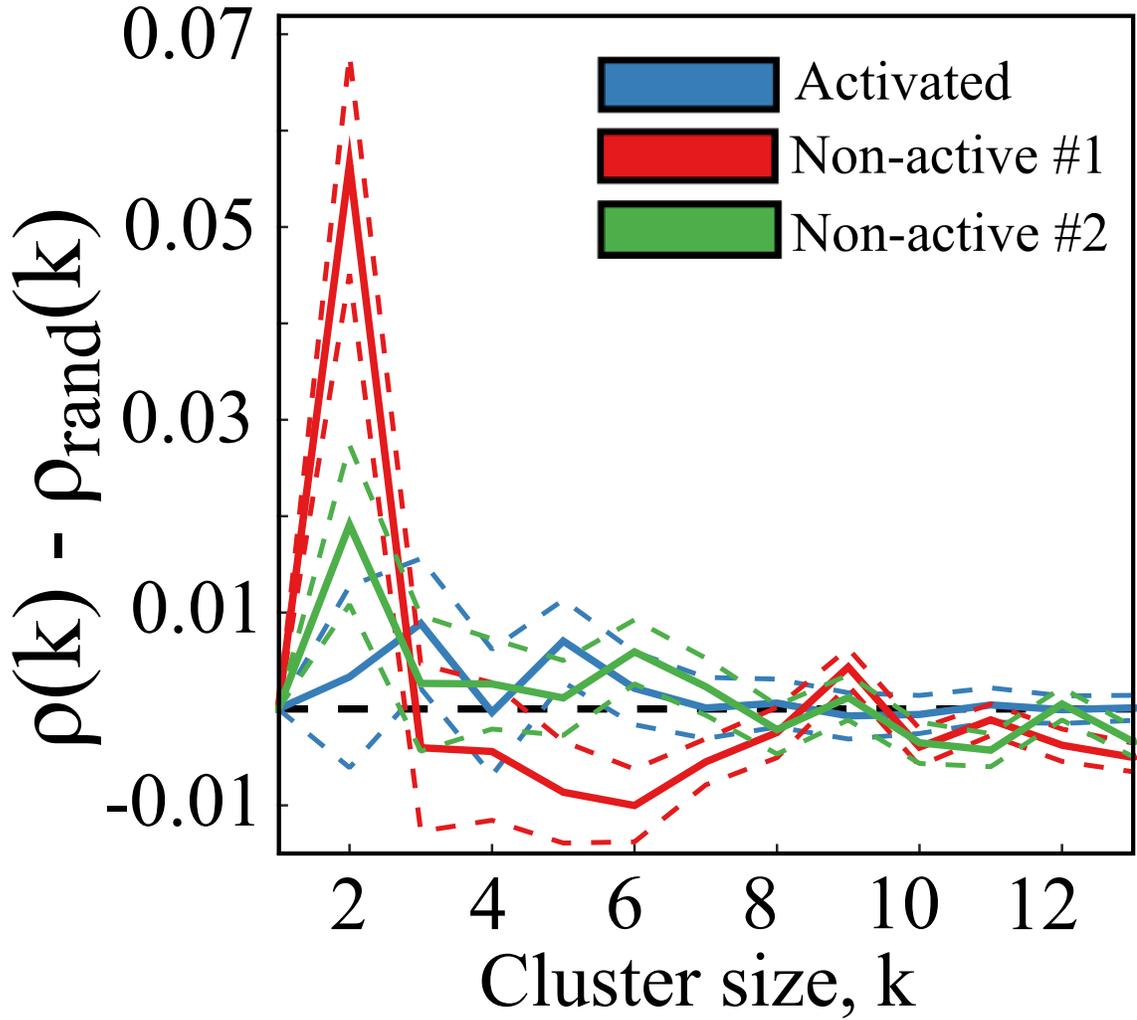


Figure 3.4: Differences between cluster size distributions of data and Heterogeneous Poisson null models for two sets of patterns representing non-active T-cells and one set of active patterns. Dashed lines represent 95% confidence intervals

### 3.3 Discussion

In this work, we've demonstrated a way of performing cluster analysis on heterogenous sets of point patterns by providing individualized baselines for each pattern, both when calculating the pairwise correlation function and when performing cluster analysis. This allowed us to perform cluster analysis on sets of protein point patterns derived from PALM experiments with human immune cells, revealing that LAT proteins do indeed cluster in dimers when the T-cell is non-active. Further, we've demonstrated that this clustering is diminished by T-cell activation. This may represent the presence of other molecules involved in the signaling pathway being drawn to the membrane, crowding out smaller clusters and forcing the LAT proteins to more trace out the membrane. This may also represent a shift towards more micro-scale clustering due to T-cell activation. Further analysis will utilize a larger bandwidth for the estimation of the pattern's local number intensity when building the HPP null models. Additionally, another interest is whether the clusters themselves are spatially segregated based on cluster size. This can be investigated by looking at spatial distribution of cluster centers of different types. Additionally, intuition needs to be gained about the pattern differences between the two control sets, namely where the distinction in clustering hierarchy seen in 3.4 cannot be attributed to activation.

Further, this work can be extended to multi-color patterns, i.e. patterns with more than one fluorescently tagged molecule. We predict that not only will the technique presented here work to elucidate the common structure of the multi-

molecular complexes, but that comparing the heterogeneous baseline of the different colors would elucidate any differences in the protein-membrane interactions of each molecule. Recent work has also extended techniques to analyze spatially heterogeneous patterns to include temporally varying patterns [81, 82], opening up a completely new avenue as well for our clustering to be applied. With the most recent advances in high-speed fluorescent tracking of molecules [83], we anticipate applying both techniques to study temporal varying molecular patterns as well.

## Chapter 4: Adenylyl cyclase mRNA localizes to the back of polarized *Dictyostelium* cells during chemotaxis

### 4.1 Overview

*This chapter is adapted from the paper submitted to Current Biology by Das, Parker, Kriebel, Guven, Larson, Losert, and Parent [50] in which I developed and implemented a technique to infer the spatial localization of mRNA molecules in fluorescent images. I then applied this technique to images from experiments performed by Das and Kriebel which represent the locations of ACA mRNA and cAR1 mRNA in migrating Dictyostelium discoideum cells.*

During directed cell migration front-back polarity is achieved by chemical gradients that promote the asymmetric localization of otherwise evenly distributed proteins. In *Dictyostelium discoideum*, vesicular transport of the adenylyl cyclase A (ACA) to the back of polarized cells is essential to relay exogenous cAMP signals during chemotaxis and for the collective migration of cells in head-to-tail arrangements called streams. Using fluorescence in situ hybridization (FISH), we observed that the ACA mRNA is asymmetrically distributed at the back of polarized cells. This localization requires an intact actin cytoskeleton as well as protein synthesis.

Using both standard estimators and Monte Carlo simulation methods, we found that ACA mRNA enrichment depends on the position of the cell within a stream, with the back localization of ACA mRNA being strongest for cells at the end of a stream. Our findings suggest that the asymmetric distribution of ACA mRNA allows the local translation and accumulation of ACA protein at the back of the cells. These data represent a novel functional role for localized translation in the relay of chemotactic signal during chemotaxis.

## 4.2 Introduction

The transmission of chemotactic signals to neighboring cells is a spatially regulated process. When exposed to an external point source of cAMP, *Dictyostelium* cells lacking ACA migrate directionally, but do not relay chemotactic signals to neighboring cells or align in a head-to-tail fashion to form streams a process that increases recruitment range during aggregation [33, 36]. This streaming behavior not only depends on the presence of ACA, but most remarkably, on its enrichment at the back of polarized cells. Indeed, ACA is distributed in two distinct pools in polarized cells: one is restricted to the plasma membrane, the other is localized on highly dynamic intracellular vesicles that coalesce at the back of polarized cells [36]. We have shown that the spatial enrichment of ACA vesicles at the back of cells and their subsequent secretion are essential for streaming during chemotaxis [30]. We proposed that the asymmetric distribution of vesicular ACA provides a compartment from which cAMP is locally released from the back of cells to attract

neighboring cells. One mechanism to achieve the polarized cellular distribution of proteins involves translation of localized mRNAs [84–86]. Consistent with this, we showed that the asymmetric distribution of ACA requires de novo protein synthesis [30]. We hypothesize that localized ACA synthesis is required to maintain the active ACA pool at the back of polarized cells for streaming during chemotaxis.

### 4.3 Experimental context for simulation-based analysis: ACA mRNA seems to localize at the back of polarized chemotaxing cells

We examined the cellular distribution of ACA transcripts using fluorescent in-situ hybridization (FISH) [87, 88]. We used 48 different fluorescently labeled oligonucleotide probes that span the entire *acaA* gene, thereby creating a sufficient signal-to-noise ratio to allow for mRNA detection [89]. We acquired diffraction limited confocal image slices and reduced them to a maximum intensity projection to facilitate data analysis. As a control for these studies, we followed the distribution of the seven transmembrane cAMP receptor 1 (cAR1) transcripts. cAR1 is the main GPCR that mediates chemotactic responses in *Dictyostelium* cells, it is uniformly distributed on the plasma membrane and does not localize to intracellular vesicles [89]. We found that the FISH signals appeared as specific punctae, i.e. randomly distributed spots likely representing multiple individual transcripts, within the cytoplasm of individual, vegetative, non-polarized *aca-* cells expressing ACA-YFP (ACA-YFP/*aca-*) as well as *car1/3-/-* cells expressing cAR1-YFP (cAR1-YFP/*car1/3-/-*) (Fig. 4.1.A-B). The hybridization of our ACA and cAR1

FISH probes to the *acaA* and *carA* genes was specific as no hybridization signal was observed in *aca-* and *car1/3-/-* cells, respectively (Fig. 4.1.C-D). We next assessed the distribution of ACA transcripts in chemotaxis competent and polarized WT and ACA-YFP/*aca-* cells. Whereas F-actin localized to the leading edge of WT cells, we observed that ACA mRNA was enriched at the back of cells (Fig. 4.2.A). This asymmetric localization of ACA mRNA was observed in both WT and ACA-YFP/*aca-* cells after they were starved and pulsed for 5 hrs and allowed to spontaneously chemotax in a chamber slide (Fig. 4.2.B-C). In contrast both endogenous *cAR1* mRNA (Fig. 4.3.A) and *cAR1*-YFP mRNA (Fig. 4.3.B) appeared uniformly distributed in the cytoplasm of polarized, chemotactically competent cells. For both *cAR1* and ACA, higher-intensity FISH spots were also observed in the nucleus, co-localizing with the DAPI signal, likely representing nascent transcripts associated with the *acaA* and *carA* genes (Fig. 4.1.A-D, Fig. 4.3.A-B).

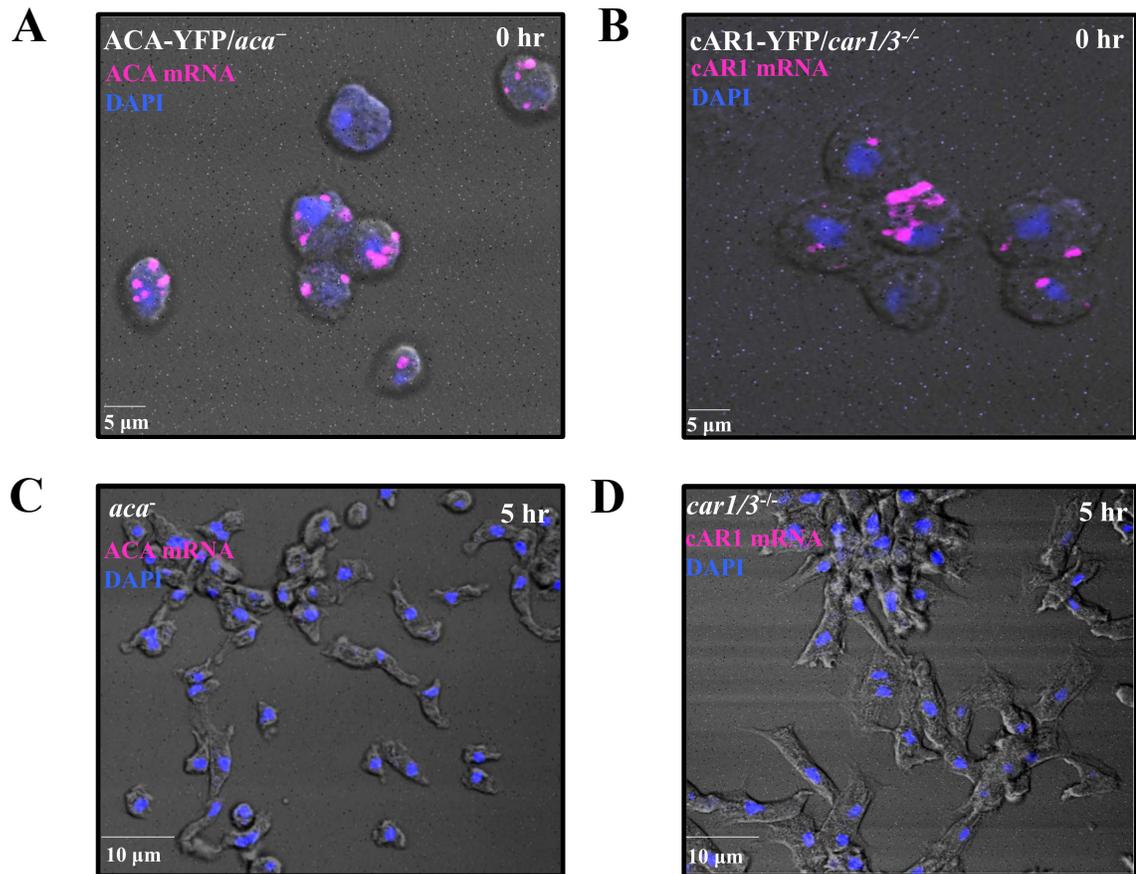


Figure 4.1: ACA and cAR1 mRNAs are randomly distributed in vegetative cells. (A) Maximum intensity projections of confocal fluorescent images of vegetative ACA-YFP/*aca*<sup>-</sup> cells depicting DAPI (nucleus) and ACA mRNA (pink). (B) Maximum intensity projections of confocal fluorescent images of vegetative cAR1-YFP/*car1/3*<sup>-/-</sup> cells depicting DAPI (nucleus) and cAR1 mRNA (pink). (C and D) Maximum intensity projections of confocal fluorescent images of 5 hrs differentiated *aca*<sup>-</sup> (C) or *car1/3*<sup>-/-</sup> cells (D) depicting DAPI (nucleus) and ACA or cAR1 mRNA (pink). The data are representative of three independent experiments.

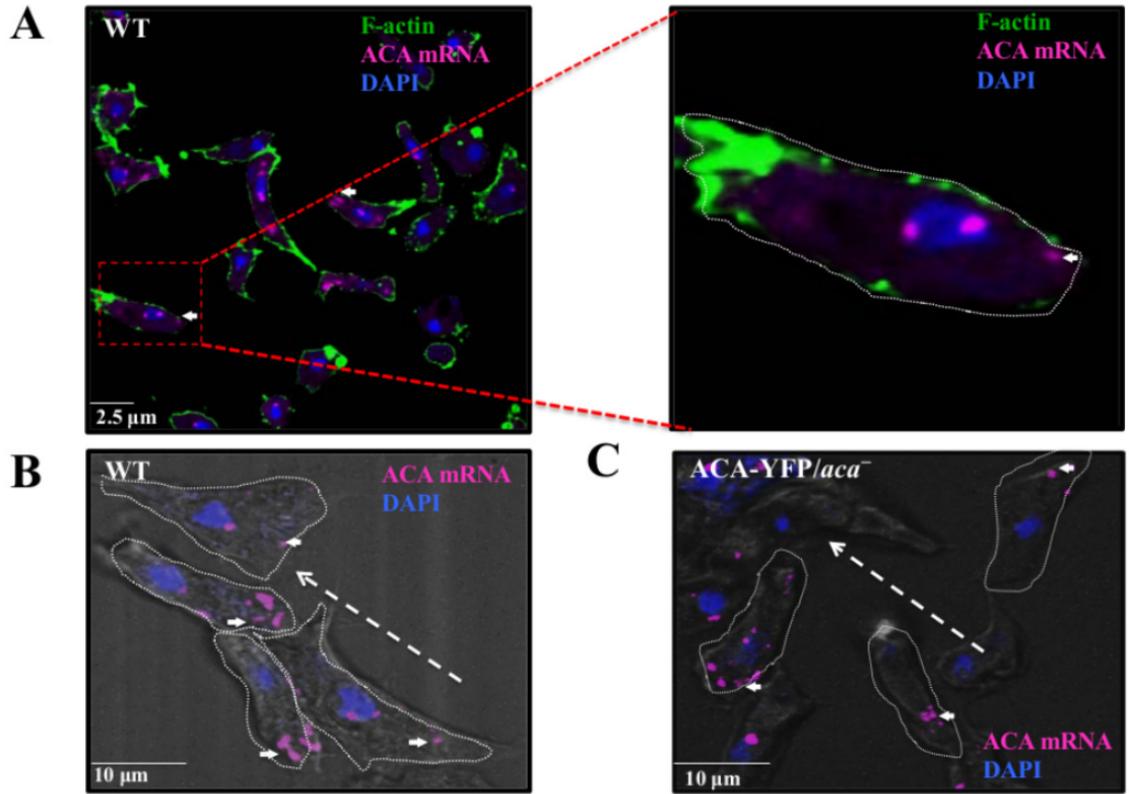


Figure 4.2: ACA mRNA is spatially localized to the back of polarized chemotaxing cells. (A) Representative maximum intensity projections of confocal fluorescent images of WT cells depicting F-actin (green), DAPI (nucleus), and ACA mRNA (pink). The dotted red box in the image is enlarged to show a single polarized cell and the asymmetric ACA mRNA distribution. The small white arrows indicate the position of the mRNA spots (B-C) Representative merged phase contrast and maximum intensity fluorescent images depicting DAPI (nucleus), ACA mRNA (pink), in WT and ACA-YFP/*aca*<sup>-</sup> cells. The small white arrows indicate the position of the mRNA spots. The white dashed arrows indicate the direction of the stream in these self-aggregation chemotaxis experiments.

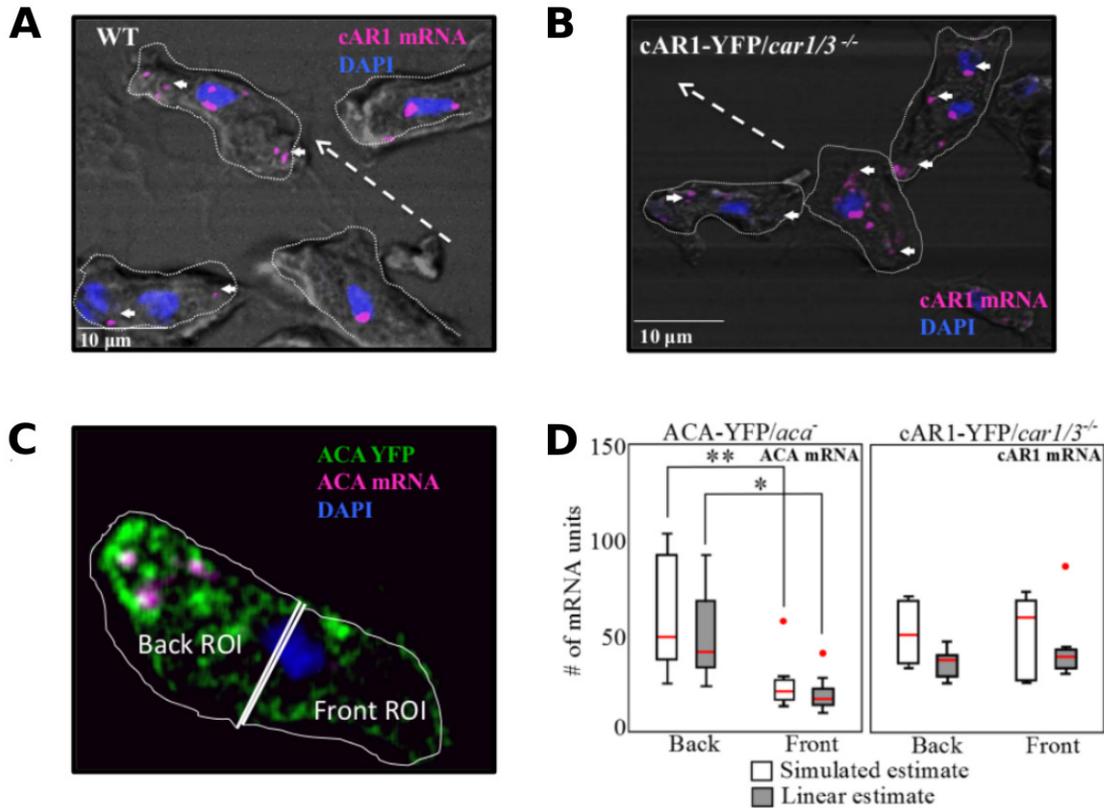


Figure 4.3: (A-B) Representative merged phase contrast and maximum intensity fluorescent images depicting DAPI (nucleus), and cAR1 mRNA (pink) in WT and cAR1-YFP/car1/3<sup>-/-</sup> cells. The small white arrows indicate the position of the mRNA spots. The white dashed arrows indicate the direction of the stream in these self-aggregation chemotaxis experiments. (C) Each cell was hand bisected, defining front and back Regions of Interest (ROI) based on both the orientation towards the aggregate center and the relative back enrichment of ACA-YFP in the cell. (G) Simulated and linear estimates of mRNA units across cells is plotted for ACA-YFP/aca<sup>-</sup> and cAR1-YFP/car1/3<sup>-/-</sup> cells. The boxes show the 50% confidence region from the median (red line). The bars cover a region with 99% confidence level from the median. All data points beyond this confidence level are considered as outliers and shown with red dots. The statistical significance is inferred by the t-test, \* represents  $p < 0.05$  and \*\* represents  $p < 0.01$ . ( $n_{ACA} = 45$ ,  $n_{cAR1} = 24$ ).

## 4.4 Quantifying and simulating mRNA distribution in fluorescent images

We sought to quantify the cellular distribution of ACA and cAR1 mRNA transcripts in polarized chemotaxing cells. Since phase contrast images of fixed cells did not show a sharp cell boundary, we were not able to utilize active contour algorithms to define the cell boundary [90]. We instead took a Region of Interest (ROI) approach, where the cell was hand segmented into visually equal front and back regions, with the identification of the cell's polarity being based on both observed ACA protein enrichment at the back and the position of the local aggregation center (Fig. 4.3.C). The mRNA content of each ROI was then calculated using two methods: a linear estimate and full image simulation. In the first method, simple integrated intensity measurements were used to quantify the spatial distribution of mRNAs. We estimated the ACA and cAR1 mRNA intensity in a region in terms of Gaussian point spread function (i.e. units) with a peak value,  $I_0$ , and a spread value,  $s$ , representing at most a few mRNA [91, 92]. To find candidate variables for the mRNA units, we located the peak intensities in the images to find local bright regions and fit 2-dimensional Gaussian point spread functions to each one. By thresholding the distributions of their spatial size and absolute intensity, we achieved finely peaked, uni-modal distributions of  $I_0$  and  $s$  representing 274 isolated units (Fig. 4.4.A-B). We then estimated the number of points in an ROI by dividing the integrated fluorescent intensity in that region with the total intensity of a

single unit,  $I_T = 2\pi s^2 I_0$ . The second method involved deconvolving the image with Gaussian subunits via Monte Carlo simulations. To simulate the pattern, we used the mean values for  $I_0$  and  $s$  in the following accept/reject Monte Carlo pattern formation procedure:

- 1) Calculate the maximum fluorescent intensity in the image,  $I_{max}$ .
- 2) Pick a random location in the image and measure the total fluorescent intensity at that location,  $I_{local}$ .
- 3) Draw a uniformly distributed random number,  $r \in [0, 1]$ . If  $I_{local}/I_{imax} < r$ , then consider that point to be a location of a subunit and subtract a subunit's intensity profile from that location.
- 4) Repeat until no suitable locations remain

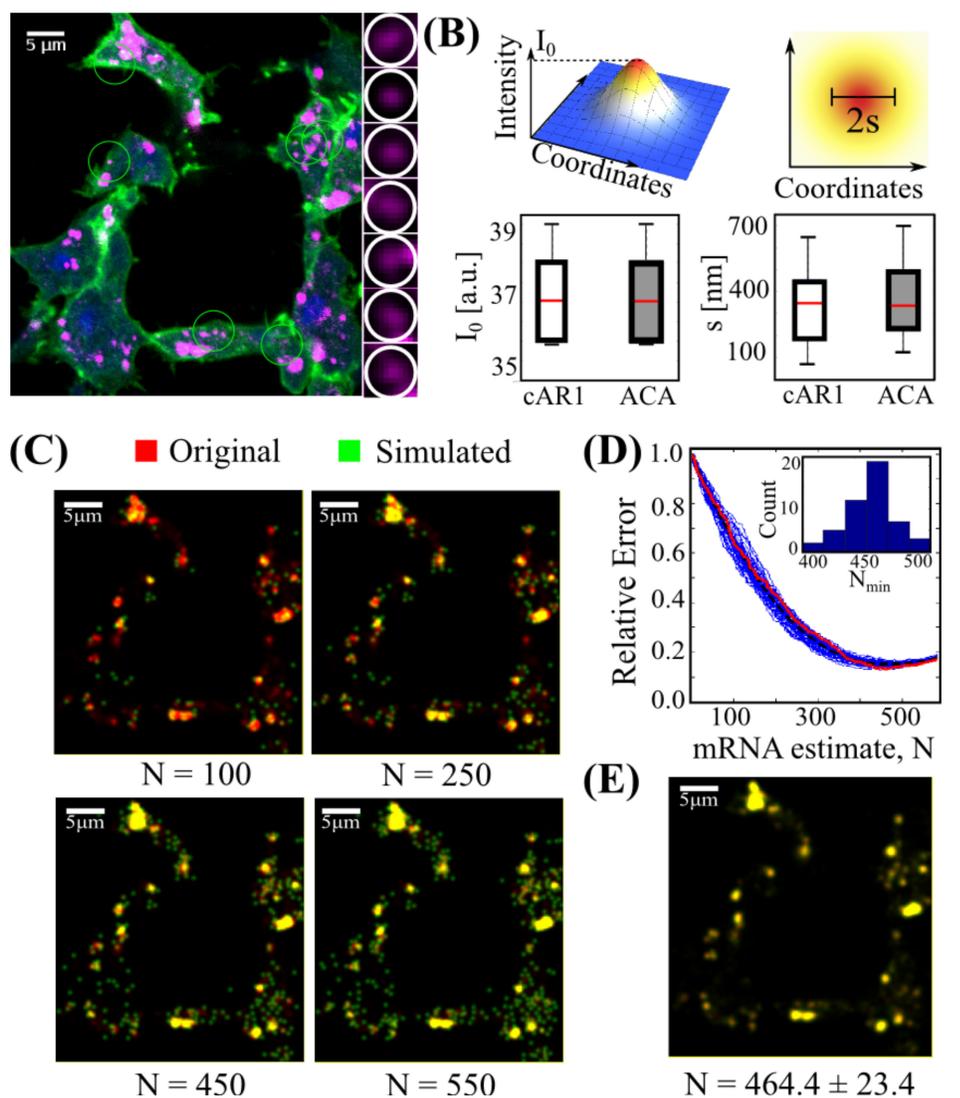


Figure 4.4: Simulation and quantification of spatial ACA mRNA localization patterns. (A) For each image, a peak finding routine was run on the mRNA fluorescent channel (left). Isolated spots were identified by thresholding their size and intensity (right). (B) Peaks were fit to Gaussian point spread functions. The resulting distributions were thresholded from above until fine, unimodal distributions remained for the two fit parameters. The mean of these distributions were termed as units. Both ACA and cAR1 mRNA showed comparable parameters. (C) The sequential images from a single iteration of the image simulation procedure performed on the mRNA fluorescent channel. Areas of yellow represent agreement. (D) The number of units in a particular image was determined by minimizing the squared different between the approximated image and the original. This is equivalent to minimizing the chi-square parameter of the fit. (E) After performing the procedure multiple times, the average image is calculated and used for quantification.

This technique is a simple modification to pattern formation simulation techniques utilized in the analysis of spatial point patterns [37, 55]. Performed once, this procedure provides an estimated underlying pattern of mRNA subunits that created the image. A repeat of this procedure will provide subtly different locations for mRNAs, as well as a different total number present (Fig. 4.4.C), representing the variety of configurations that would still correspond to a good match with the original image. By repeating this process multiple times, the number of subunits in an ROI is therefore estimated by the mean number of points placed in the ROI during the various simulations, and the uncertainty is given by the standard error (Fig. 4.4.D-E). This procedure therefore distinguishes itself from strict peak localization in that it does not require the mRNA to be well-separated to infer the local number density. Thus, these two complimentary procedures provide us with reliable estimates of the local number of mRNA subunits throughout the fluorescent image data set.

Both of these estimates modeled the image as being comprised of a discrete number of fluorescent spots, which we refer to as mRNA units. These are by no means single molecules and likely represent multiple individual transcripts. Using peak finding algorithms on all of the images of both ACA and cAR1 mRNA and thresholding their size and intensity, we identified a characteristic unit (Fig. 4.4.A-B). We found that both cAR1 and ACA mRNA units were comparable in size and intensity (Fig. 4.4.B). The linear estimate of the mRNA content of an ROI, then, is the integrated intensity inside the region divided by the intensity of a single unit. The simulated estimate, by contrast, rebuilds the image one unit at a time until the

sum squared difference between the simulated image (green) and the original confocal image (red) is minimized (Fig. 4.4.C-D). This rebuilding is performed multiple times to obtain an average image, representing a spatial map of all of the mRNA units (Fig. 4.4.E). Using this method, the mRNA content of an ROI is then the average number of mRNA units contained inside the ROI. Since both methods treat the ROI as a container, this method of characterization only depends on the accurate bisecting of the cell into front and back, being insensitive to the accuracy of defining the cell boundary. We found that the simulated methods broader distribution in estimates yields a stronger statistical power. This demonstrates the facility of including a simulated estimate, as the linear estimates do not reflect the uncertainty inherent in estimating the number of mRNA units over a cell population.

## 4.5 Results and Discussion

### 4.5.1 ACA mRNA and ACA protein asymmetrically localize to the back of isolated and migrating cells

Using both methods, we measured no preferential cellular distribution of cAR1 mRNA units in cAR1-YFP/car1/3-/- single cells (Fig. 4.3.D). In sharp contrast, a large proportion of ACA mRNA localized to the back of both WT and ACA-YFP/aca- single cells (Fig. 4.3.D). For migrating cells, Dictyostelium cells can be classified in different groups based on their location in the line of a stream with respect to the aggregation center (Fig. 4.5.A). Cells arrange themselves either at the front (near the aggregate), in the middle, or at the back of a stream of multiple

cells. Using simulated images (Fig. 4.4), we quantified the number of ACA and cAR1 mRNA units localized within the cell front and back in relationship to the position of the cell in a stream (Fig. 4.5.B). We only considered cells that were well polarized allowing hand segmentation, non-polarized cells touching the aggregate center were not taken into the analysis (see eccentricity measurements, see Fig 4.5.A, right panel). We did not measure a preferential distribution of cAR1 mRNA within individual cells in all stream positions analyzed. In contrast, the enrichment of ACA mRNA at the back of cells became stronger in cells positioned in the middle and back of streams (Fig. 4.5.B). Indeed, we found that as cells migrate closer to an aggregation center, the ACA mRNA acquires a random cellular distribution (Fig. 4.5.B). It's well known that the mean local concentration of cAMP controls the ability of the cell to adjust to changes in the local gradient [93,94], a phenomena known as "adaptation". We envision that as cells get closer and form an aggregation center, they are exposed to higher mean cAMP signals, resulting in the inefficient gradient sensing contributing to a loss in the asymmetric distribution of ACA mRNA. However, cells at the far end of the aggregation center experience a lower mean concentration, which allows better cAMP gradient measurement, allowing the maintenance of a polarized state and the ensuing enrichment of ACA mRNA at their back.

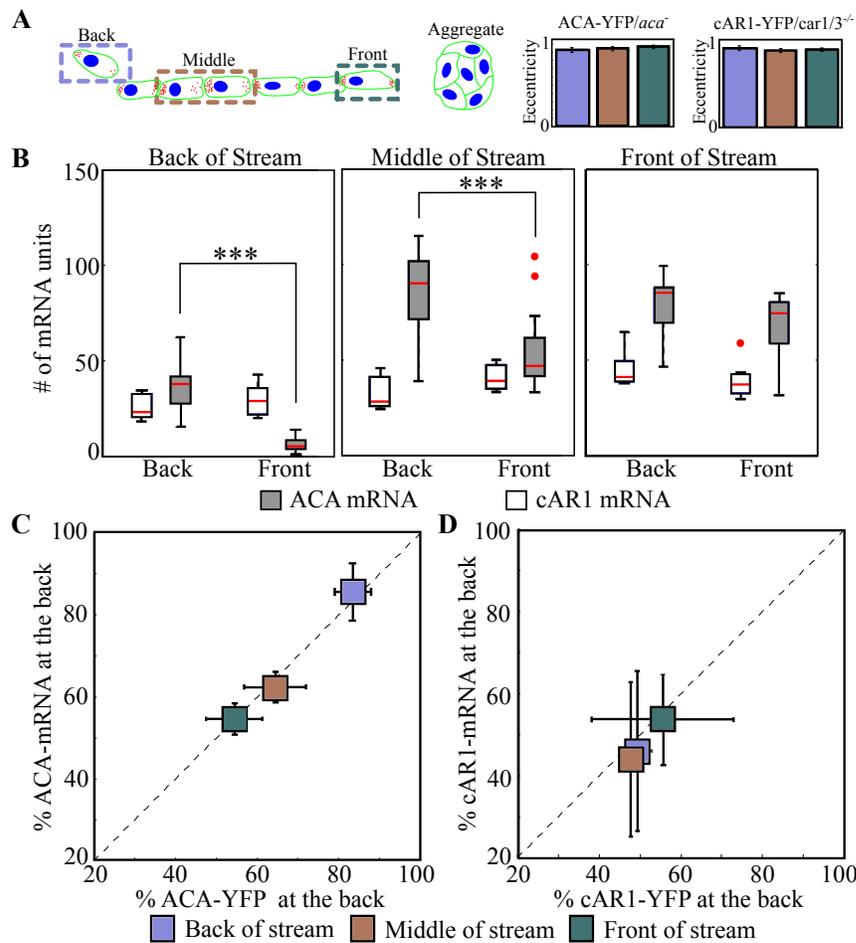


Figure 4.5: The ACA mRNA is asymmetrically distributed at the back of the streaming cells (A) Cartoon depicting the distribution of cells within a stream. Each cell was characterized as either being at the back of a stream, in the middle of stream or near an aggregate based on its position from the aggregate center. (B) The simulated estimate of mRNA units across the cell is plotted for ACA-YFP/*aca*<sup>-</sup> (grey box) and cAR1-YFP/*car1/3*<sup>-/-</sup> (white box) cells. The boxes show the 50% confidence region from the median (red line). The bars cover a region with 99% confidence level from the median. All data points beyond this confidence level are considered as outliers and shown with red dots. The statistical significance is inferred by the t-test, \* represents  $p < 0.05$  and \*\* represents  $p < 0.01$ ,  $n = 15 - 52$ . (C-D) The correlation between mRNA and its corresponding protein at the back of cells is presented for ACA-YFP/*aca*<sup>-</sup> and cAR1-YFP/*car1/3*<sup>-/-</sup> cells (see Experimental Procedures for details),  $n = 12 - 45$ .

In order to understand the functional significance of the ACA mRNA asymmetry, we measured the degree of ACA-YFP protein enrichment at the back of cells in the different stream populations. This value was obtained by measuring the integrated YFP fluorescence intensity at the back of the ACA-YFP/*aca-* cells (Fig. 4.5.C-D). We observed that the degree of ACA mRNA asymmetry strongly correlates with the amount of enriched ACA-YFP at the back of cells (Fig. 4.5.C; Pearsons correlation: 0.99), with cells at the back of streams showing the highest polarized distribution. This correlated polarization differed significantly from what was seen in *cAR1-YFP/car1/3-/-* cells, where we found no correlation between *cAR1* mRNA and *cAR1-YFP* distribution (Fig. 4.5.D; Pearsons correlation: 0.01). As noted, we measured high eccentricity numbers for both ACA-YFP/*aca-* and *cAR1-YFP/car1/3-/-* cells (Fig. 4.5.A, right panel), indicating that all cells within the streams were polarized to the same extent. However, only the ACA mRNA was preferentially enriched at the back of the cells. Together, these findings suggest that the sustained enrichment of ACA-YFP at the back of cells arises from newly translated ACA from the localized ACA mRNA transcripts.

## 4.6 Conclusions

Together, our findings show for the first time localized mRNA in the social amoebae *Dictyostelium discoideum*. We show that the ACA mRNA is specifically localized at the back of polarized cells, and that this is required for the relay of signals during chemotaxis and streaming. Signal relay and streaming are essential

to increase the recruitment range of cells when they migrate to a point source of attractant [33]. We have previously proposed that ACA-containing vesicles are secreted at the back of chemotaxing cells as exosomes [30], where they are required for the formation of streams during chemotaxis. We reason that the localized ACA mRNA allows the local translation and replenishment of ACA protein at the back of cells, where it is necessary to relay signal to neighboring cells. We envision that similar mechanisms are involved in other cellular systems, such as immune cell signaling, where the maintenance of localized protein expression may be necessary to allow for fast spatio-temporal events to occur.

## Chapter 5: A broken-contact order parameter for inferring inter-cellular communication from patterns of externally guided migrating cells

### 5.1 Overview

*This chapter is adapted from the paper in preparation by Parker, Guven, Wang, Ott, and Losert. [51] in which Can Guven and I developed a simulation based migration model for the group migration of Dictyostelium discoideum in the presence of an external chemical gradient. We also defined a spatio-temporal contact network based order parameter to distinguish between individual and cooperative migration, and used this measure to investigate model perturbations.*

We propose an order parameter for investigating the local spatial organization often seen in migrating cell systems. Using a phenomenological model for the group dynamics of *Dictyostelium discoideum*, we observe that monitoring the fraction of broken cell-cell contacts over time resolves the characteristic structures seen in simulations that are not captured by other global order parameters. We then explore the use of this parameter to analyze experiments involving both mechanical and chemical perturbations. We demonstrate that adding adhesive forces

between cells stabilizes streaming structures and diminishes otherwise independent migration. However, the loss of the degradation of the chemoattractant diminishes stream formation, almost eliminating the streaming phase altogether. We discuss the application of this order parameter to experimental and theoretical work on *D. discoideum* migration as well as extensions to other biological and physical systems.

## 5.2 Introduction

Much effort has been made to understand the collective dynamics of the group migration of the slime mold *Dictyostelium discoideum*, which is studied as a model for chemotaxis [36, 95], individual amoeboid motion [90, 96, 97], and group migration [98, 99]. Studying these phenomena provides insights into human inflammatory response to wounds and infections as well as cancer cell migration [100].

*D. discoideum* cells both sense and secrete the same chemoattractant, cyclic adenosine monophosphate (cAMP), providing a bias towards coordinated motion in neighboring cells. For low densities or strong external chemoattractant gradients, uniformly distributed *D. discoideum* cells move independently (see Fig. 5.1.a). However, other conditions allow for chemical signaling between cells, resulting in the formation of chains of migrating cells called “streams” [36, 101, 102] (see Figs. 5.1(b,c)).

Though local in nature, stream formation is crucial for *D. discoideum* to aggregate, which has strong physical and biological implications for the survival of the organism. Changes in cell-cell communication or defected mechano-transductive ability of the cells have been shown to inhibit stream formation [33, 34]. E.g.,

mutant cells with diminished cell-substrate interactions do not form streams and cluster in smaller aggregates, which results in aberrant sporulation (Fig. 5.1(e)). In addition, Garcia et al [34] showed that cells that cannot degrade the chemical signal also display aberrant stream formation, prolonging the self-aggregation of cells.

A current push in experimental and theoretical cell biology is investigating what factors affect the structure and facility of these streams [7]. These transient, local regions of coherence are visible to the eye but means of quantifying streaming behavior is an open area of research. Global order parameters (introduced by Vicsek et al. [52]) have been applied to quantify streams but were not on their own able to account for the local variations in density seen in streaming patterns [102].

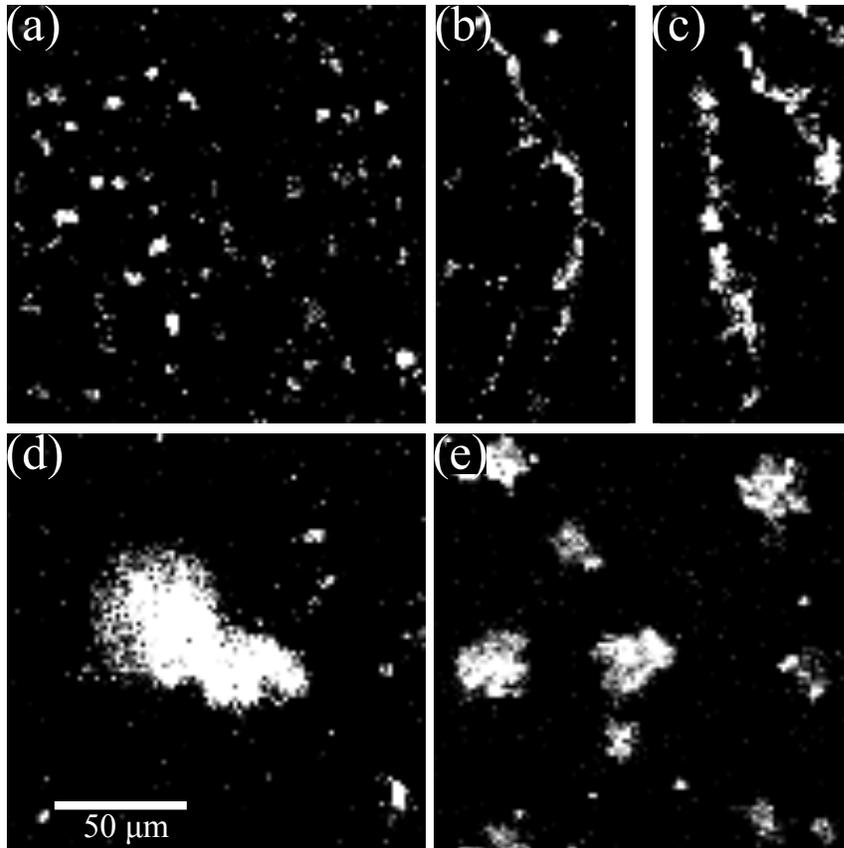


Figure 5.1: (a) Initial state of cells (uniformly distributed). (b-c) Cells moving in streams. (c) Final configuration of wild-type cells is a large aggregate. (d) Mutant cells that cannot stream form smaller aggregates in their final configuration.

In this work, we introduce a local order parameter calculated from the spatio-temporal cell contact network. Using a mathematical model for *D. discoideum* group migration, we demonstrate that the fraction of broken contacts within a window of time can distinguish streaming from individual cell motion, and we use this quantitative characterization to investigate the model’s dependence on experimentally relevant conditions. We then use this model to investigate the effects that degradation and adhesion have on the streaming phase of the model. We find that adhesion stabilizes streams and diminishes individual motion, likely prolonging short-lived stream formation and maintain cell-cell contacts. In contrast, the loss of chemical degradation strongly diminishes streaming to a narrow band of density and external concentration parameters. This suggests that the temporal refinement of the local gradient is required for coordinated chemotaxis.

### 5.3 Stochastic cell migration model

We demonstrate the quantification of streams utilizing a collective *D. discoideum* migration model based on our prior work [102]. In this model the dynamics of a single cell is described with three modules, *gradient sensing* (Eqs. (5.1) and (5.2)), *membrane protrusions* (Eqs. (5.3) and (5.4)), and *center-of-mass motion* (Eqs. (5.5) and (5.6)), supplemented by an equation (Eq. (5.7)) for the diffusion, production, and degradation of the chemoattractant. Each cell  $i$  is represented as a motile disk with a defined front and back, corresponding to the direction of its motion. The cell responds to the external cAMP concentration field,  $C(\mathbf{r}, t)$  by aligning its target

direction,  $\mathbf{g}_i(t)$ , to a perceived gradient direction via the following two equations (where  $|\mathbf{g}_i(t)| = 1$ )

$$\dot{\mathbf{g}}_i(t) = \omega_\phi \mathbf{g}_i(t) \times (\mathbf{H}_i(t) \times \mathbf{g}_i(t)), \quad (5.1)$$

$$\mathbf{H}_i(t) = \frac{l \nabla C(\mathbf{r}_i(t), t)}{\langle C(\mathbf{r}_i(t), t) \rangle_{\text{local}} + K_D} + \sqrt{\eta_\phi} \boldsymbol{\xi}_\phi(t). \quad (5.2)$$

According to (5.1),  $\mathbf{g}_i(t)$  orients itself with the vector  $\mathbf{H}_i(t)$  on a relaxation time scale  $\omega_\phi^{-1}$ . The deterministic contribution to  $\mathbf{H}_i(t)$  (first term on the right hand side of (5.2)) is proportional to the cAMP gradient if  $C$  is low and saturates to  $l(\nabla C/C)$  when  $C$  is large ( $C \gg K_D$ ), consistent with experimental observations [103] ( $l := \sqrt{D/\nu_0}$ , where  $D$  and  $\nu_0$  will be defined subsequently). According to (5.2),  $\mathbf{H}_i(t)$  has a stochastic component modeled by two dimensional white noise of amplitude  $\eta_\phi$  ( $\langle \xi_\phi^j \rangle = 0$ ,  $\langle \xi_\phi^j(t) \xi_\phi^{j'}(t') \rangle = \delta_{jj'} \delta(t - t')$ , where  $j = x, y$  indicates the directional component). This models the stochastic nature of the binding and unbinding of cAMP molecules to receptors [103–106].

The “membrane protrusion” module models the cell’s protrusive response  $\mathbf{n}_i(t)$  to its target direction of motion, and with the net direction of new protrusions corresponding to the two equations

$$\dot{\mathbf{n}}_i = \omega_\theta \mathbf{n}_i \times (\mathbf{G}_i \times \mathbf{n}_i), \quad (5.3)$$

$$\mathbf{G}_i = \mathbf{g}_i + \sqrt{\eta_\theta} \boldsymbol{\xi}_\theta(t). \quad (5.4)$$

That is, the unit vector  $\mathbf{n}_i(t)$  responds to a noise-corrupted goal direction  $\mathbf{G}_i(t)$  (with noise amplitude  $\eta_\theta$ ) on a time scale  $\omega_\theta^{-1}$ .

Finally, the center-of-mass motion of cell  $i$ , located at point  $\mathbf{r}_i$  and moving with velocity  $\mathbf{v}_i$ , is modeled by the equations,

$$\dot{\mathbf{v}}_i(t) = -\beta\mathbf{v}_i(t) + \kappa v_0 \mathbf{n}_i(t) + \sum_{j \neq i} \mathbf{f}_{ij}, \quad (5.5)$$

$$\mathbf{f}_{ij}(d_{ij}) = \frac{\mathbf{r}_i - \mathbf{r}_j}{d_{ij}^3} (1 - d_{ij}/2r_{\text{cell}}) k_{rep} \quad (5.6)$$

Here,  $\mathbf{f}_{ij}$  is a “force” with strength  $k_{rep}$  ( $\Theta$  is the unit step function) that prevents the cells (assumed to be of size  $r_0$ ) from occupying the same space. In the absence of mechanical interactions (i.e.,  $k = 0$ ), Eq. (5.5) can be integrated to yield  $\mathbf{v}_i(t) = \kappa v_0 \int_{-\infty}^t e^{-\beta(t-t')} \mathbf{n}_i(t') dt'$ . This identifies the center-of-mass motion as a “sum over protrusions” [90] with a characteristic time scale  $\beta^{-1}$ , where  $\kappa$  is a time-independent parameter selected so that the speed  $|\mathbf{v}_i|$  suitably averaged over cells is  $v_0$  (see supplementary information).

We simulate Eqs. (5.1)-(5.6) in a square chamber with side  $L = 0.33$  mm, cell density  $\rho$ , and periodic boundary conditions for the cell motion (thus, if a cell leaves through one of the four boundaries, it is reintroduced at the opposite boundary). The cells are initially distributed to be not in contact but otherwise random, and the simulation is let run sufficiently long to reach steady state. The dynamics of the

chemoattractant inside the chamber are governed by the following diffusion equation,

$$\partial_t C = D\nabla^2 C - \nu(y)C + s \sum_j \delta(\mathbf{r} - \mathbf{r}_j), \quad (5.7)$$

The first term on the right hand side of Eq. (5.7) represents diffusive spreading of the chemoattractant with diffusivity  $D$ . The second term represents global degradation of the cAMP field by phosphodiesterases secreted by the cells. The third term represents a constant secretion of cAMP from each cell with rate  $s$  [33]. We choose the form of  $\nu(y)$  and the boundary conditions on  $C(x, y, t)$  to mimic behavior experiments performed in no-flow gradient chambers [102]. The boundary conditions are  $C(x, 0, t) = 0$ ,  $C(x, L, t) = C_0$ ,  $C(x + L, y, t) = C(x, y, t)$ , while  $\nu(y) = 4\nu_0 y/L(1 - y/L)$  where  $\nu_0$  is a coefficient reflecting the rate of degradation.

## 5.4 Quantifying coordinated motion in migrating cells

Fig. 5.2.(a-c) show model simulations of the three steady state phases of this model (individual motion, streaming, and aggregation) for a representative set of parameters and three values of concentration strength,  $C_0$ . For high  $C_0$ , the cells do not communicate, moving upward with a diffusive motion in the  $x$ -direction (Fig. 5.2.a). For moderate  $C_0$ , the cell's secretion of cAMP results in a system size stable stream, (Fig. 5.2.b). For low  $C_0$ , these streams become unstable and the cells aggregates (Fig. 5.2.c).

To quantify this behavior, we define a local order parameter,  $F$ , as follows. At time  $t_0$  we consider two cells with  $d_{ij} < 2r_0$  to be in contact, where  $d_{ij} := |\mathbf{r}_i - \mathbf{r}_j|$  is

the distance between cells  $i$  and  $j$  ( $r_0$  is the cell radius). We then look at time  $\Delta t$  in the future and calculate the fraction of those contacts that no longer exist,  $N_{\text{broken}}/N$ , where  $N$  is the total number of contacts at time  $t_0$ . Our order parameter,  $F$ , is then the time average over  $t_0$ ,  $\langle N_{\text{broken}}/N \rangle_{t_0}$ , normalized by the same value achieved for simulations with  $s = 0$  (cells not secreting cAMP). This normalization insures that, though the nominal value of the fraction of broken links increases with  $\Delta t$  due to the motion of the cells, the order parameter itself is sufficiently independent of the time window. Since  $s = 0$  implies no chemical signaling between cells,  $F \approx 1$  implies persistent yet individual motion while  $F < 1$  implies local coordination of cells.

We compare this order parameter to the mean progression index (introduced by Vicsek et al [52]),  $M$ , which is defined as the average of  $\mathbf{v}_i(t)$  averaged over  $t$  and  $i$ , normalized by the average cell speed,  $v_0$ ,  $M := |\langle \mathbf{v} \rangle|/v_0$ .  $M$  is therefore a measure of the cell transport rate along the net displacement of the stream parallel to the chemoattractant gradient.

In Fig.5.2(d), both  $F$  and  $M$  are plotted for varying  $\rho$  and  $C_0$ . For strong external chemical signal or low densities, the fraction of broken links is due to random motion in the  $x$ -direction, with all cells moving upward ( $F \approx 1$ ,  $M \approx 1$ ). As the chemical signal is decreased or density increased, the cells begin to stream, coordinating their lateral position and reducing the number of broken contacts while cells continue to move upward ( $F < 1$ ,  $M \approx 1$ ). As the streams become unstable and collapse into aggregates, the fraction of broken links is small but non-zero and cells stop migrating upwards ( $F \ll 1$ ,  $M \ll 1$ ). As is seen, for regions of aggregation or individual motion, both measures yield the same information. However,  $M$  re-

veals only the transition from motion to aggregation while  $F$  shows a broad region of density and concentration values that yield coordinated streaming cell motion. This clearly demonstrates that  $F$  is capable of more completely characterizing the dynamic phases of *D. discoideum* cell migration.

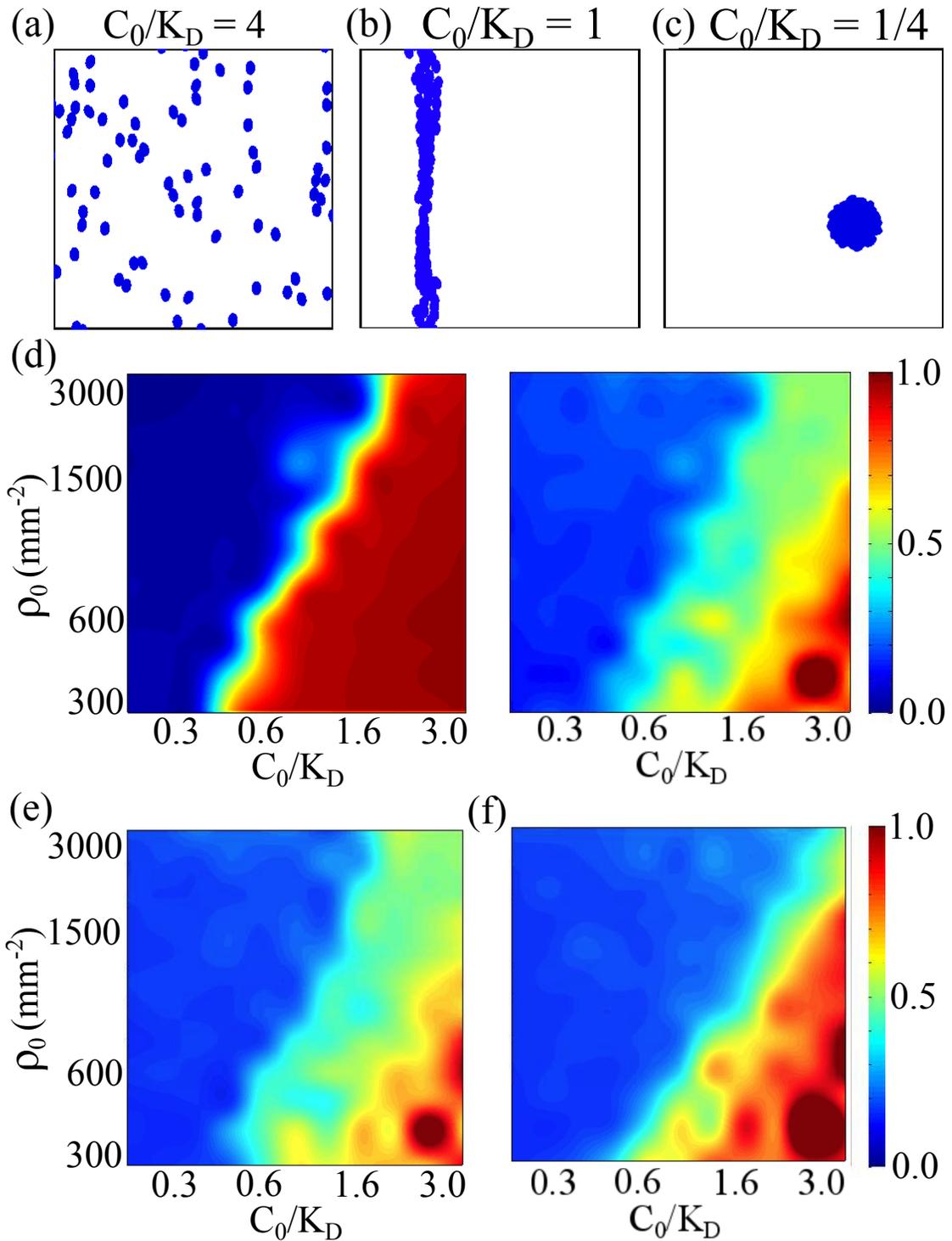


Figure 5.2: (a-c) Snapshots of the simulations demonstrating the three steady state dynamic modes: individual motion, streaming, and aggregates ( $\rho = 750 \text{ mm}^{-2}$ ) (d) The mean progression index (left) and the fraction of broken links (right) for a variety of  $\rho$  and  $C_0$  values. (e) The fraction of broken links for cells with intercellular adhesion (f) The fraction of broken links for cells that do not degrade cAMP

## 5.5 Chemical and mechanical perturbations affect stream stability

We now are positioned to explore perturbations to the experiment and their effects on streaming. Our first perturbation was to modify the chemical interactions by removing degradation of the cAMP signal, which mimics experiments done with PDE1<sup>-</sup> cells. Not only will this affect the externally imposed signal, but will also diminish the cell's ability to locate each other, since degradation removes previously secreted chemoattractant to amplify more recent signals. The second perturbation was to include a small adhesive force between cells, modifying Eq. 5.6 to

$$\mathbf{f}_{ij}(d_{ij}) = \frac{\mathbf{r}_i - \mathbf{r}_j}{d_{ij}^2} (1 - d_{ij}/2r_{\text{cell}}) (k_{\text{rep}}/d_{ij} - k_{\text{adh}}) \Theta(2r_0 - d_{ij}). \quad (5.8)$$

We set the adhesion to be a small effect, with  $k_{\text{adh}}/k_{\text{rep}} = .05$ . This was to keep our model results still informative as it would not draw the cells too close together. Looking at the fraction of broken links for cells with inter-cellular adhesion (Fig. 5.2.e), we see very little change along the aggregate/stream boundary. This suggests that a small amount of adhesion doesn't broadly increase the instability of streams, causing them to collapse into aggregates. However, for high concentrations and low densities, there is a shift towards more cooperative motion as adhesion prevents quick migration away from one's neighbors, thus better relaying of the locally secreted signal. This suggests that adhesion overall broadens the streaming regime.

However, when chemical degradation is turned off (Fig. 5.2.f), the streaming

regime dramatically. This narrowing is in part due to the growth of the individual motion region, suggesting that degradation is needed to produce the sharper gradients that allow cells to find each other in the presence of stronger signals. The narrowing is also due to the shifting of the aggregate/streaming boundary, where cell populations that were once able to distinguish the external signal instead collapse in on themselves. This is likely do to the overwhelming build up of chemoattractant, washing the external signal out. This confirms the vital role for degradation in autocrine signal relay.

## 5.6 Conclusions

In this work, we have shown that monitoring the spatio-temporal contact network is a useful tool for investigating cellular dynamics. Specific to streaming, the contacts broken over time show the ordering in the system due to increases in cell-cell interactions through variation of external parameters. We used this technique to demonstrate the sensitivity of the streams to mechanical and chemical perturbations, demonstrating that adhesive forces can stabilize interactions even in the absence of streaming, while the loss of the ability to degrade the external chemical signal reduces the ability for cells to move cooperatively. Further work will be to model “paracrine” signal relay, which is relevant to modeling neutrophil migration who secrete and respond to LTB4 along with a host of other cytokines [26].

## Chapter 6: Summary and Future Outlook

In this work, I've demonstrated various different ways to characterize pattern sets for exploring biological questions. Biological pattern formation presents special issues due to the inherent noise of cell-to-cell variability but also in the complicated and often non-linear aspects of phenomena. This gives rise to the need to use more outside-of-the-box tools as well as focus on simulation based approaches to better approximate null models to create coherence among pattern sets.

In chapter 2, I demonstrated techniques using topological measures to characterize patterns, allowing for the robust sorting of pattern sets into like groups. We showed that these measures outperform previously used second order functionals, and correctly divided non-interacting protein patterns. From an analysis perspective, this work focused primarily on choosing the right statistical measure, wanting to leverage all moments of the pattern simultaneously.

These functionals, however, are hard to interpret visually and are themselves homogeneous functions of each pattern's number intensity. This means we cannot use them to define length scales to investigate clustering in the patterns. So, in Chapter 3, I used a modified pairwise correlation function [78], using simulation techniques to create individualized statistical baselines for each cell to account for

the large scale features associated with protein-membrane interactions and regions of poor image quality. From an analysis perspective, this research focused its attention on the appropriate null models for the set, a key observation for dealing with complicated patterns. By re-normalizing the each cell's pairwise correlation function by its associated null model, I demonstrated that the broad scale heterogeneity in the point pattern sets are dramatically reduced, allowing for the identification of a clustering length scale of about 500 nm. Further using individualized baseline patterns as a reference, I performed cluster analysis on the pattern sets and was able to show that, at that length scale, LAT proteins cluster in small structures of two or three proteins, but that T-cell activation diminishes this hierarchy.

In Chapter 4, I turned my attention to spatial patterns where the point locations aren't known but must be inferred from the heterogenous intensity distribution. The data in question was a set of images from FISH experiments with *Dictyostellium Discoideum* cells that are self aggregating. Two different types of mRNA, ACA and cAR1, had been fluorescently labeled along with their associated proteins and the task at hand was to quantify the relative front/back distribution of these molecules. Adapting techniques to simulate heterogenous Poisson processes, I used a brute force approach to infer the local number densities of each molecule in each image. With this approximated density field, I was then able to measure the front/back distribution of each molecule in both the absolute sense as well as proportionally across the cell. Doing this found that whereas cAR1 mRNA has no preferential location in the cell, ACA mRNA preferentially locates to the rear of the migrating cells. Further, this spatial preference is most prominent in cells that are just joining

streams, and steadily diminishes as the cells approach the aggregate.

Finally, in Chapter 5, applied both the notion of selecting the right measure and the right statistical baseline to study a model of *D. discoideum* migration and the effects that both mechanical and chemical perturbations have on dynamic migration modes, namely individual motion, streaming, and aggregation. In choosing the right measure, I both focused on the need for higher order measures than the chemotactic index as well as the need to consider temporal stability in the pattern, settling on looking at the average fraction of broken contacts within a time window. This was then normalized by the same measure for non-interacting cells, allowing for a pseudo-phase diagram to be made with respect to density and external concentration for each simulation condition. After demonstrating this measure's ability to distinguish the dynamic steady states of the model, I showed that cell-cell adhesion doesn't effect the streaming/aggregate transition but mainly coordinates motion in higher concentration regions. I also showed the lack of chemical degradation all but eliminates the ability of cells to stream, owing to the inability to clarify local gradients.

In regards to the analysis techniques themselves, the obvious extension of these efforts is to multi-color and multi-disperse pattern sets. Current efforts in super-resolution imaging is to fluorescently label more and more molecules, which would allow for investigation into the structure of the molecular complexes that form after T-cell activation. This would allow for the inference and potential identification of what chemical reactions occur downstream of TCR binding events, which could yield dramatic insights in immune system research. Similarly, cell migration assays

with cell mixtures can be done to look at competition and cell sorting, meaning that our spatio-temporal contact network measure should be adapted to networks with different node labels. In both instances, this creates combinatoric complexity in both defining the correct measures and the appropriate null models, which itself requires well defined questions to be answered each time their asked.

In regards to the cell migration model, current work in modeling chemotaxis is focused on the paracrine signal relay seen in neutrophils. The cell response to this type of signaling is much more nuanced and complex than *D. discoideum*, and whether or not the cells move cooperatively towards the sight of the wound is still an open question. Further, neutrophils migrate in the presence of other cell types, rushing to the sight of the wound first which leads to cell sorting. So, having both the control of a simulated model and spatio-temporal analytics that can measure interactions in poly-disperse systems would put us in a unique place to contribute to the understanding of human immune response.

## Chapter A: Appendix B: Supplemental Material for Chapter 2

### A.1 Software

To approximate the 2-D Minkowski functionals of our patterns, we relied exclusively on the software described in [43], which was available online at <http://www.mathe.tu-freiberg.de/inst/stoch/Stoyan/morph2D/>.<sup>1</sup> This program takes as input  $r_{min}$ ,  $dr$ , and  $r_{max}$ . Since our interest is in automation we used the same values for all of our patterns ( $r_{min} = dr = .01, r_{max} = 100$ ).

For both smoothing and applying the Functional Principal Component Analysis, we used the Functional Data Analysis MATLAB packages that are available online at [www.functionaldata.org](http://www.functionaldata.org), and we relied on their description in [64] for implementation. Mimicking the procedure of [45], we first smoothed our functionals using cubic b-splines.

To cluster using Ward’s method, we first utilized MATLAB’s implementation in their “linkage” function. To implement modularity maximization, we used the weighted version of the Fast Modularity algorithm which can be found online at <http://cs.unm.edu/~aaron/research/fastmodularity.htm>. The specifics of

---

<sup>1</sup>At the time of this paper’s submission, this website was down; we are in the process of notifying the appropriate people about this issue

the algorithm are the same as in [73], but maximizes the weighted definition of modularity (for a description of this alteration, see [74]).

Other home made programs were written to compute the second order functionals, simulate point patterns, and implement various portions of the project (either in MATLAB or C). Those interested in discussing these programs should contact the authors.

## A.2 Intensity Scaling

As reported in [70], the Minkowski functionals are homogenous with regard to domain scaling. To be specific, for any parameter  $\lambda > 0$  and domain  $\Omega \subset \mathbb{R}^d$ , the  $n$ -th Minkowski functional  $M_n(|\Omega|)$  satisfies the relation

$$M_n(|\lambda\Omega|) = \lambda^{d-n} M_n(|\Omega|) \tag{A.1}$$

This means two patterns with different overall number intensity will have different Minkowski functionals even if they are the same type of pattern. To address this in our pattern comparison, all patterns were scaled to unit intensity before their Minkowski functionals were approximated.

### A.3 Approximating $g(r)$

To approximate the pairwise correlation function  $g(r)$ , we used the estimator

$$\hat{g}(r) = N^{-1}\lambda^{-1} \sum_{i=1}^N w_i^{-1}(r, dr) \sum_{i \neq j} I(|\vec{r}_i - \vec{r}_j| < r). \quad (\text{A.2})$$

Here,  $I(x)$  is the indicator function and  $\lambda$  is the number intensity. The weight  $w_i$  is the portion of the area of the disc centered on  $\vec{r}_i$  with inner radius  $r$  and outer radius  $r + dr$  that is contained in the pattern window. We found that this method achieved better results than that of [45], where  $g(r)$  is approximated by exploiting its relation to the derivative of Ripley's K-function.

### A.4 Pattern Simulation

Binomial processes were simulated using MATLAB's built in random number generator and scaling the results. MATLAB code to simulate Strauss processes can be found in [107], and Baddelley-Silverman processes were simulated using home-made software based on the procedure described in [41].

## Chapter B: Appendix B: Supplemental materials for Chapter 3

### B.1 Approximating $g(r)$

To approximate the pairwise correlation function  $g(r)$ , we used the estimator

$$\hat{g}(r) = N^{-1} \lambda^{-1} \sum_{i=1}^N w_i^{-1}(r, dr) \sum_{i \neq j} I(|\vec{r}_i - \vec{r}_j| < r). \quad (\text{B.1})$$

Here,  $I(x)$  is the indicator function and  $\lambda$  is the number intensity. The weight  $w_i$  is the portion of the area of the disc centered on  $\vec{r}_i$  with inner radius  $r$  and outer radius  $r + dr$  that is contained in the pattern window, which corrects for edge effects.

### B.2 Theoretical considerations for heterogeneous cluster analysis

#### B.2.1 The heterogeneous pairwise correlation function

In spatial process statistics, each spatial pattern is a stochastic realization of an underlying point process: no two patterns are ever identical, but hallmarks of the point process are instead embedded in the different spatial correlations of the points and their relative distances. Following the formalism of Ref. [78], we consider here a planar point pattern  $\chi = \{\vec{r}_i\}$ , where the first-order intensity of the is pattern  $\lambda(\vec{r})$

with mean value  $\bar{\lambda}$ . We define the second order density  $\rho(\vec{u}, \vec{v})$ , in that  $\rho(\vec{u}, \vec{v})d\vec{u}d\vec{v}$  is the probability that the point pattern  $Y$  has a point in each of the differential rings centered at  $\vec{u}$  and  $\vec{v}$ . Thus, the heterogeneous pairwise correlation function of the pattern  $\chi$  is given by [78]:

$$g_{inh}(\vec{u}, \vec{v}) = \frac{\rho(\vec{u}, \vec{v})}{\lambda(\vec{u})\lambda(\vec{v})} \quad (\text{B.2})$$

We assume sufficient stationarity in the pattern to assert that  $g(\vec{u}, \vec{v}) = g(|\vec{u} - \vec{v}|)$  where  $|\cdot|$  is the euclidean norm. For our purpose, the set of  $N$  patterns  $X = \{\chi_i\}$  has a companion set of  $N$  positive functions  $\Lambda = \{\lambda_i(\vec{r})\}$ . We assume that there exists an underlying process such that  $\langle g_{inh,i}(r) \rangle = g_{inh,\Lambda}(r)$  with probability 1 as  $N \rightarrow \infty$ . If, then, for any  $r_c < \infty$ ,  $g_{inh,\Lambda}(r_c) < 1$ , then we can infer that the underlying, environment filtered process is a more clustered process than random.

To calculate the heterogeneous pairwise correlation function, we blend the approaches of both Ref. [78] and [79]. We take each pattern in turn and simulate 1000 heterogeneous Poisson null models for each pattern (for simulation details, see below). For both the original pattern and all of its random recreations, we calculate the approximate heterogeneous function as

$$\Delta(r) = \frac{\sum_{i=1}^n \sum_{i \neq j} I[r_j \in \partial b_{\vec{r}_i}(r)]}{n(n-1)w(\vec{r}_i)}, \quad (\text{B.3})$$

where  $\partial b_{\vec{\sigma}}(r)$  is differential surface of the Borel set centered at  $\vec{\sigma}$  with radius  $r$ ,  $I[\cdot]$  is the indicator function, and  $w(\vec{r})$  is an edge correction factor (cite, see discussion

below). Then, we approximate the heterogeneous pairwise correlation function for pattern  $i$  as

$$\hat{g}_{inh,i}(r) = \frac{\Delta_{data}(r)}{\langle \Delta_{random}(r) \rangle} \quad (\text{B.4})$$

With the approximated function  $g_{inh,\Delta}(r)$  achieved by an ensemble average.

### B.3 Simulating a heterogeneous Poisson null model for an existing point patten

Starting with a blank window of the same dimensions as the original pattern, we iteratively follow these steps:

- 1) Pick a random location in the simulation window,  $r_{test}^{\vec{}}$
- 2) Calculate the quantity  $p = \lambda(r_{test}^{\vec{}})/max(\lambda(\vec{r}))$ , where  $\lambda(\vec{r})$  is the approximated first-order intensity of the pattern to be randomized.
- 3) Draw a uniformly distributed random number,  $r \in [0, 1]$ . If  $r < p$ , then place a point in the simulated window at location  $\vec{r} = \vec{r}_{test}$ . Otherwise, repeat.

In our work, we continued this process until the number of points,  $N$ , in the two patterns matched. This accept/reject strategy recreates the variation in the local number density of the original pattern, treating it as an un-normalized cumulative probability density function [55]. We approximated the first-order intensity function

through kernel density estimation [37], where

$$\lambda(\vec{r}) = \langle \kappa_h(|x - x_i|) \kappa_h(|y - y_i|) \rangle_i \quad (\text{B.5})$$

where  $\kappa_h(d)$  is the Epanechnikov kernel [108]

$$\kappa_h(d) = \frac{3}{4h} \left(1 - \frac{d^2}{h^2}\right) I[d < h] \quad (\text{B.6})$$

with  $h$  being the “bandwidth” of the estimating kernel.

## Chapter C: Appendix C: Supplemental materials for Chapter 4

### C.1 Experimental procedures

#### C.1.1 Preparation of cells

WT (AX3), ACA-YFP/*aca*, and cAR1-YFP/*car1/3* cells were grown in shaking cultures to  $4 \times 10^6$  cells/ml in HL5 media [2]. They were harvested by centrifugation, washed once in developmental buffer (composition). To allow differentiation, the cells were shaken at 100 rpm for 47 h with pulses of 75 nM cAMP every 6 min [109,110]. The cells were then processed according to the assay performed.

#### C.1.2 Antibodies and Immunoblotting

Whole cell lysates were subjected to a 420% Tris-HCl SDS-PAGE analysis using the Criterion gel system and transferred to Immobilon-P (Millipore). The Immobilon-P was blotted with anti-GFP monoclonal antibody (1:5000; Babco), anti-RPL8 (ARP40215, Aviva Systems Biology, 1:1000) anti-actin (C-11; Santa Cruz Biotechnology, 1:2000) and detection was performed by chemiluminescence using a donkey anti-mouse horseradish peroxidase-coupled antibody (1:5000; GE Healthcare) or an anti-rabbit horseradish peroxidase-coupled antibody (1:10,000;

GE Healthcare) and the ECL Western blotting detection reagents (GE Healthcare).

### C.1.3 Chemotaxis and streaming assays

The chemotaxis assay were performed as previously described [2]. Briefly, 57-h differentiated cells were plated on chambered cover slides as described [33,111] and allowed to adhere and self-stream for 30 min to 1 hr. Alternatively, chemoattractant gradients were generated using a microinjector (Eppendorf) with micropipettes filled with 1 M cAMP. The micropipette was placed in the chambered cover slides and images were captured at specified times. Once the cells started to align in streams and form self-aggregates, they were fixed and processed for in situ hybridization.

### C.1.4 Fluorescent in situ hybridization (FISH)

Vegetative or differentiated cells were fixed in 3% paraformaldehyde (32% (wt/vol)) and permeabilized with Triton X-100 (0.1% vol/vol) in phosphate buffer. A mixture of 48 FISH DNA probes ( 22 nt long) was commercially synthesized (Biosearch technologies) and processed according to the manufacturers protocol. Briefly, fixed cells were hybridized with the FISH probes for 4 hrs at 37C in 10% formamide in 2X saline-sodium citrate (SSC) hybridization buffer. The coverslips were washed three times with 2X and 1X SSC strengths buffers and the nuclei were stained with DAPI. The coverslips were mounted and allow to cure. The slides were observed with a confocal microscope (Axiovert 200; Carl Zeiss, Inc.). Single plane images and Z stacks (1-m confocal slice) were taken using 63 and 100X

plan neofluor objectives (Carl Zeiss, Inc.) and Z stacks were arranged in maximum intensity projections.

## C.2 Perturbations

5 M LatA (for 10 or 30 min; Invitrogen) or 400 M CHX (for 10 or 30min; Sigma-Aldrich) was added to cells plated on a cover slip. Cells were fixed and processed for in situ hybridization as described above. Simultaneously, cells were harvested for total RNA isolation (Trizol; Invitrogen) and RT-PCR analysis.

## C.3 RNA isolation and analysis

RNA was isolated using Trizol LS reagent (Invitrogen) and contaminating DNA was removed by treatment with RQ1 DNase (Promega) for 30min at 37C. One microgram of RNA was reverse transcribed using SuperscriptII Reverse Transcriptase (Invitrogen) and random hexamer primers, according to the manufacturers instructions. cDNA was used for PCR reactions in a GeneAmp PCR System 9700 (Applied Biosystems). Densitometry analysis was performed using ImageJ 1.42q (National Institutes of Health).

## C.4 Cycloheximide recovery

Cells were differentiated as described for 4 hrs in shaking flasks containing 2 107 cells/ml. At the end of 4 hrs, 1.6 mM CHX was added to the cells in shaking flask for an additional 2 hrs to inhibit protein synthesis. Cells were then harvested

and washed to remove traces of the drug and resuspended in phosphate buffer. The cells were plated and the recovery of fluorescence was monitored at different time points by imaging using confocal microscopy. In a parallel set of experiment, cells were fixed and processed for in situ hybridization, as described. For western blot analysis,  $2.7 \times 10^6$  cells were harvested at various time points and resuspended in Laemmli buffer [112]. Whole cell lysates were subjected to a 420% Tris-HCl SDS-PAGE as described above.

## C.5 Ribopuromycylation

Ribopuromycylation was performed as previously described [113, 114] and modified for *Dictyostelium*. Differentiated cells were treated with 91  $\mu$ M PMY and 208  $\mu$ M emetine for 10 min at RT. All extraction procedures were performed on ice using reagents pre-chilled to 4°C. Cells were incubated for 2 min with 500  $\mu$ l/well permeabilization buffer (50 mM Tris-HCl, pH 7.5, 5 mM MgCl<sub>2</sub>, 25 mM KCl, 355  $\mu$ M CHX, EDTA-free protease inhibitors, and 10 U/ml RNaseOut containing 0.015% digitonin [Wako Chemicals USA]). After the extraction step, an aliquot was used for western blot analysis. The extracted cells were spun, resuspended in IP buffer (40 mM Tris-HCl, pH 8.0, 2 mM EDTA, 50 mM NaCl, and 1% CHAPS) and Complete protease inhibitor cocktail (Roche), and solubilized for 1 h on ice. After solubilization, 100  $\mu$ l of a 50/50 protein A sepharose CL-4B (GE Healthcare)/IP buffer slurry was added to each sample and incubated using a rotator for 1 h at 48°C to pre-clear the lysate. The protein A sepharose was removed by centrifugation and

15 l of anti-PMY monoclonal antibody was added per milliliter of lysate and incubated on a rotator overnight at 48C. 100 l of a 50/50 protein A sepharose slurry was added and incubated for 2 h at 48C to precipitate puromycylated protein complex. Samples were washed four times with 1 ml IP buffer. The precipitated proteins were released from sepharose beads by adding Laemmli buffer and boiling for 10 min. The resulting samples were run using the Criterion gel system (Bio-Rad Laboratories) using a 420% Tris-HCl gel and immunoblotted with anti-rpl8 polyclonal antibody. RNA associated with the immunoprecipitated complex was dissociated with 10 U/ml RNaseOut containing 0.015% digitonin and used in RT-PCR.

## C.6 Measuring cell polarity

The front/back polarity was determined by manual segmentation, calculating the ratio of the long and short axis of the cell length in the direction of the stream towards an aggregate center. The extent of polarization of the cells was calculated using eccentricity equation  $\epsilon = \sqrt{1 - b^2/a^2}$ . A value of 1 indicates a parabolic or polarized cell shape and a value of 0 indicates a circle or non-polarized cell shape. During this process, we did not take into account cells whose boundaries could not be distinguished from each other in a stream.

### C.6.1 ACA translation originates in the cytoplasm and localizes to the back of cells as they acquire polarity

To visualize the appearance of newly synthesized ACA protein and mRNA in a spatiotemporal fashion, we followed the cellular distribution of the ACA protein and mRNA following recovery after CHX treatment. We treated ACA-YFP/*aca*<sup>-</sup> and cAR1-YFP/*car1/3*<sup>-/-</sup> cells with 1.6 mM CHX for 2 hrs to inhibit protein synthesis. As previously reported [30], this treatment dramatically decreased ACA-YFP levels but did not alter the expression level of cAR1-YFP (Fig. C.1). After CHX treatment, cells were washed and ACA-YFP or cAR1-YFP protein recovery was monitored in live cells using confocal microscopy and by Western analysis. In parallel experiments, at corresponding recovery time points, samples were fixed and hybridized with FISH probes to monitor the appearance and cellular distribution of ACA and cAR1 mRNA. As seen in Figure C.2.A, long-term CHX treatment rendered the cells non-polar and the ACA-YFP signal could not be detected (Fig. C.2.A; 0 min), although a few cells retained ACA mRNA expression (Fig. C.2.B; 0 min). As early as 30 min following CHX removal, when cells remained non-polar, newly synthesized ACA-YFP protein appeared in a vesicular pool within the cytoplasm. The recovery of ACA-YFP expression was also observed 30 min after CHX removal by Western analysis (Fig. C.1). One hour after recovery, ACA-YFP became enriched at the back of the now polarized cells and by 2 hrs, a strong ACA-YFP plasma membrane labeling was evident (Fig. C.2.A). Similarly to the ACA protein recovery, 30 min after CHX removal the ACA mRNA signal strongly appeared

in a random distribution in the cytoplasm of the non-polarized cells. As the cells acquired polarity, within 1-2 hrs after CHX removal, the ACA mRNA signal also became enriched at the back of the cells (Fig. C.2.B). Importantly, under these same conditions, cAR1-YFP and cAR1 mRNA signals remained unchanged throughout the entire recovery time (Fig. C.2.C-D). Using the quantification method described above, we estimated the number of ACA mRNA units in the front and back of cells at different times during CHX recovery. As the cells recovered and acquired polarity, we measured an increase in the number of ACA mRNA units at the back of cells, reaching a maximum 2 hrs after CHX removal (Fig. C.3.A). These findings establish that ACA mRNA first occurs randomly in the cytoplasm and that upon the acquisition of cellular polarity, the ACA translation machinery spatially localizes to the back of migrating, polarized cells.

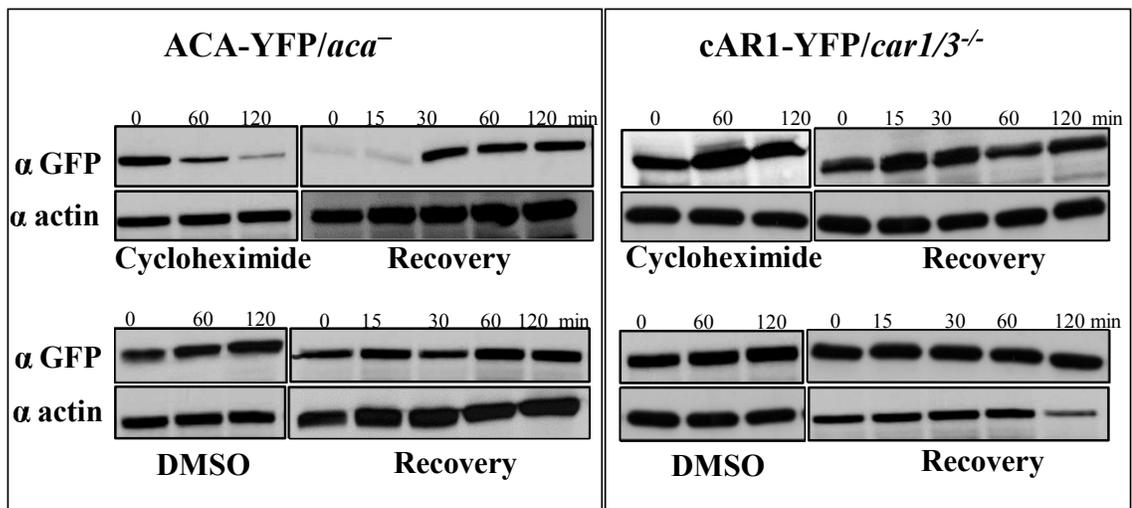


Figure C.1: . Loss of ACA-YFP but not cAR1-YFP after CHX treatment Western analysis showing protein levels of ACA-YFP from *ACA-YFP/aca<sup>-</sup>* cells or cAR1-YFP from *cAR1-YFP/car1/3<sup>-/-</sup>* cells in the presence of 1.6 mM CHX and during the recovery time points. DMSO-treated cells were used as control for this experiment. Representative data of two independent experiments are shown.

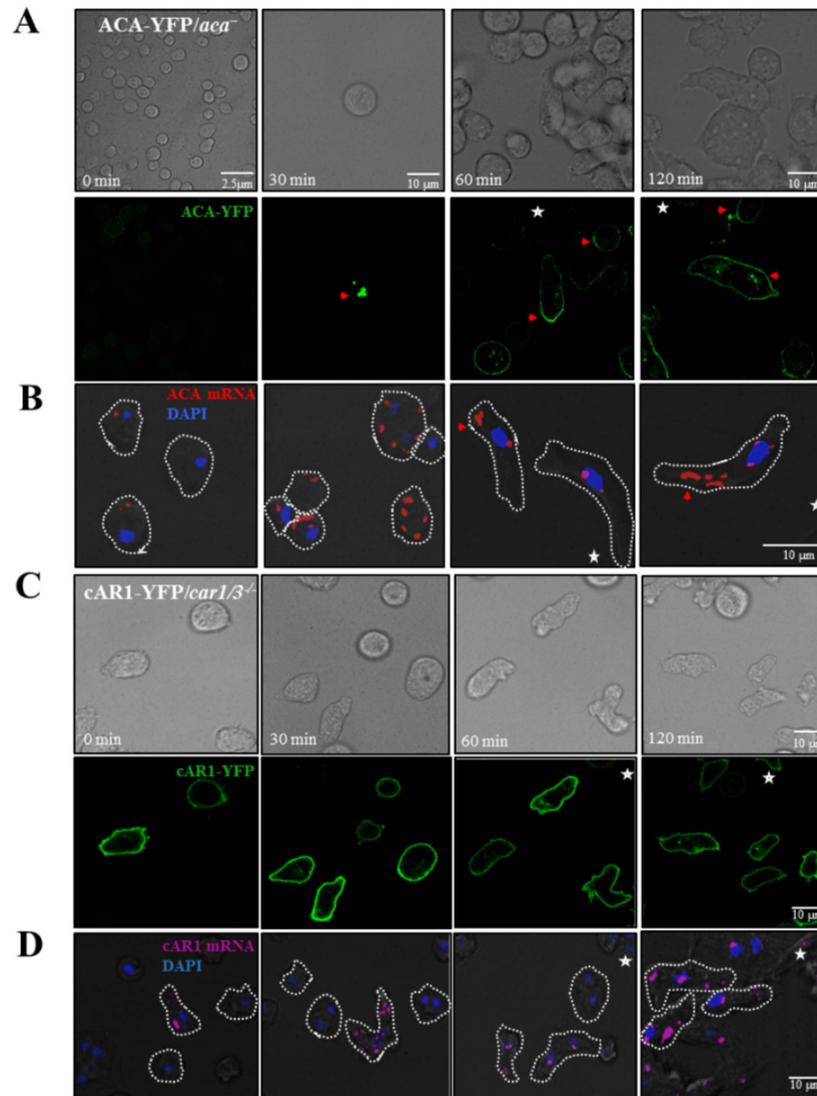


Figure C.2: ACA translation occurs in the cytoplasm followed by localization to the rear of the polarized cells (A) Representative phase contrast (upper panel) and confocal fluorescent (lower panel) images of ACA-YFP/*aca*<sup>-</sup> treated with 1.6 mM CHX for 2 hrs. Fluorescent recovery is monitored after CHX removal. (B) Representative maximum intensity projections of confocal fluorescent images of ACA-YFP/*aca*<sup>-</sup> cells depicting DAPI (nucleus) and ACA mRNA (red). Red arrows indicate the nascent ACA transcription sites after CHX removal. The white star indicates the location of the aggregation center. (C) Representative phase contrast (upper panel) and confocal fluorescent (lower panel) images of cARI-YFP/*car1/3*<sup>-/-</sup> treated with 1.6 mM CHX for 2 hrs. (D) Representative maximum intensity projections of confocal fluorescent images depicting DAPI (nucleus) and cARI mRNA (red). For (A) and (C), red arrows indicate the appearance of the nascent cellular ACA translational sites while white stars indicate the location of the aggregation center. For (B) and (D), white stars indicate the location of the aggregation center.

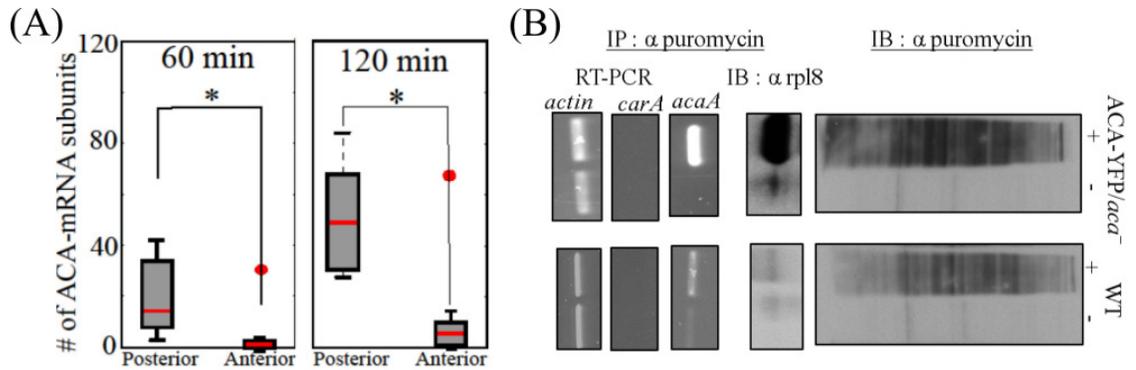


Figure C.3: (A) The simulated estimate of ACA mRNA units 60 and 120 min after CHX removal across the cell is plotted for ACA-YFP/*aca*<sup>-</sup> cells. The boxes show the 50% confidence region from the median (red line). The bars cover a region with 99% confidence level from the median. All data points beyond this confidence level are considered as outliers and shown with red dots. The statistical significance is inferred by the t-test, \* represents  $p < 0.05$ . The data excludes the 0 min time point as these cells are not polarized,  $n = 6 - 15$ . (B) Western analysis of C-terminally puromycylated nascent chains released from ribosomes with and without puromycin (PMY) treatment in ACA-YFP/*aca*<sup>-</sup> and WT cells (upper panel). The complex was immunoprecipitated using the anti-PMY mAb 12D10 and immunoblotted with the anti-rpl8 polyclonal antibody to detect associated ribosomes (middle panel). The associated RNA was dissociated from the complex and the presence of specific mRNAs corresponding to *acaA* and *carA* were detected by RT-PCR using specific primers (bottom panel). This figure is a representative of at least four independent experiments.

We next wanted to assess if ACA is indeed actively translated in cells. For this purpose, we wanted to determine if ACA mRNA is associated with actively translating ribosomes. We adopted the ribopuromylation method [113] to immobilize puromycylated nascent chains on ribosomes in the presence of the chain elongation inhibitors puromycin (PMY) and emetine. We observed high amounts of puromycylated nascent protein chains in both WT and ACA-YFP/*aca*<sup>-</sup> cells (Fig. C.3.B, top panel). The nascent chains bound to the ribosomes were immunoprecipitated with the  $\alpha$ -puromycin antibody and blotted for rpl8 a large ribosomal protein [113, 114]. We found that the immunoprecipitated PMY tethered actively translating ribosome

complex in both WT and ACA-YFP/*aca-* cells, although we routinely observed a much stronger signal in ACA-YFP/*aca-* cells (Fig. C.3.B). Remarkably, RT-PCR analysis of the dissociated complex further revealed that ACA mRNA, but not *cAR1* mRNA, is associated with the actively translating ribosomes (Fig C.3.B). These findings establish that ACA is actively translated in chemotaxing competent cells.

## Chapter D: Appendix: Supplemental materials for Chapter 5

### D.1 Cell motion experiments

#### D.1.1 Cell culture and development

Wild-type *Dictyostelium discoideum* (AX2) and the talin A null (talA-) mutant cells (in an AX2 background) were a generous gift from Carole Parent's lab (NIH/NCI). Cells were cultured in HL-5 at  $1 - 4 \times 10^6$  cells/mL. For differentiation, cells were harvested at  $4 \times 10^6$  cells/mL and transferred to a beaker at  $2 \times 10^7$  cells/mL shaken at 150 rpm. Cells are developed for 5 hours (wild-type) or 6 hours (talA-) in development buffer (5 mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 6.2, 2 mM MgSO<sub>4</sub> and 0.2 mM CaCl<sub>2</sub>). During development, cells are stimulated with pulses of cAMP (50 nM every 6 minutes). In the last 30 minutes of development, CellTracker Green (Invitrogen) (10  $\mu$ g with 0.8% DMSO) was added to cells for fluorescent-labeling. The cells were then washed twice in phosphate buffer (5 mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 6.2) and resuspended in phosphate buffer at  $5 \times 10^5$  cells/mL.

### D.1.2 Chemotaxis assay

In the needle assay experiments, 800  $\mu\text{L}$  phosphate buffer containing  $4 \times 10^6$  cells were dropped in a 2-well Lab-Tek chamber and cells were allowed to settle for 10 minutes. Then, a micropipette containing 0.1 mM cAMP were placed at the center of the imaging field and closed to the bottom of the chamber. An Eppendorf Femtojet system, which was connected to the micropipette, was set at a pressure of 130 hPa to continuously release cAMP from the micropipette as previously described ([33]). The fluorescent images were taken by a Leica SP2 confocal microscope every 6 seconds for 150 minutes. Data from the no-flow linear gradient was taken from Ref. [102].

## D.2 Parameter selection based on experiment

### D.2.1 The Concentration Profile

The concentration profile is given by solving the equations

$$\partial_t C = D \nabla^2 C - \nu(y)C + s_{cell}\rho(\mathbf{r}, t). \quad (\text{D.1})$$

Where the degradation profile  $\nu(y)$  is given by

$$\nu(y) = \nu_0 \frac{y}{L_y} \left( 1 - \frac{y}{L_y} \right). \quad (\text{D.2})$$

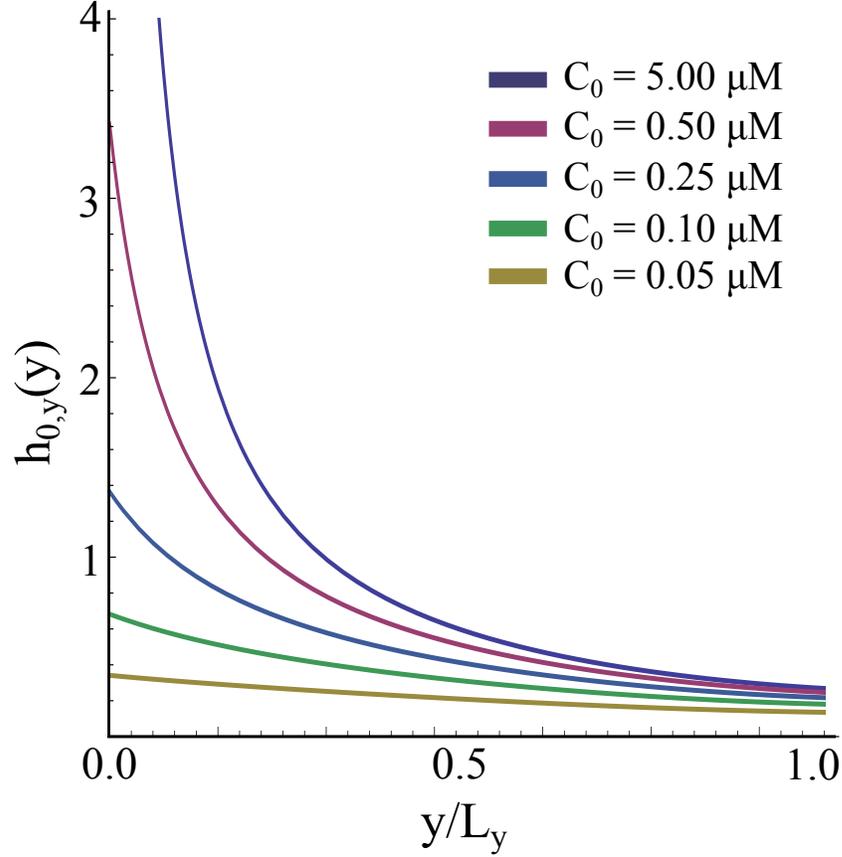


Figure D.1: The y-component of the signal vector seen by the cell in the absence of cell secretion (i.e.  $s_{cell} = 0$ ). As the external concentration is decreased, the signal goes to a constant.

Selecting variables common to experiment and prior modeling [102], we chose  $L_x = 1$  mm,  $L_y = 0.33$  mm,  $\nu_0 = 2.25$  min $^{-1}$ ,  $s_{cell} = 250$  nM, and cell injection rate  $\rho_o v_0 = 7.5$  cells min $^{-1}$  mm $^{-1}$ . Defining the dispersion length  $\Delta_{cAMP} = \sqrt{D/\nu}$ , we modeled the dimensionless chemoattractant signal seen by the cells as

$$\mathbf{h}_0(\mathbf{r}, t) = \Delta_{cAMP} \frac{\nabla C}{\langle C \rangle_{local} + K_{D,cAMP}} \quad (\text{D.3})$$

Where  $K_D = 50$  nM. This for is consistent with the formula for the relative difference of bound cAR1 receptors on the surface of the cell (see Ref. [103]). For  $s_{cell} = 0$ ,

the symmetry in the  $x$  direction allows the steady state equation to reduce to a one dimensional second order differential equation. Fig. D.1 shows the solution to the steady state equation for the  $y$ -component of the signal for varying boundary values  $C_0$ .

$$h_{0,y}(y) = \Delta_{\text{cAMP}} \frac{\partial_y C(y)}{C(y) + K_{D,\text{cAMP}}} \quad (\text{D.4})$$

As can be seen, as  $C_0 \rightarrow K_D$ , the signal decreases towards a constant value.

## D.2.2 Individual cell parameters

Unlike prior work with simpler models [102], our cell model is highly non-linear and methods using linearization or the Fokker-Plank equation are not available to us to fit parameters to experimental data, which is also difficult to come by. Therefore, the cell parameters used in our model are either taken from prior biochemical knowledge or estimated using no-flow chamber experiments with *aca*-mutant cells, i.e. cells that do not perform signal relay and therefore do not interact strongly [33]. This allows us to fit our various parameters based on individual cell tracks. Then, for each cell, we have the equations

$$\dot{\mathbf{g}}_i(t) = \omega_\phi \mathbf{g}_i(t) \times (\mathbf{H}_i(t) \times \mathbf{g}_i(t)) \quad (\text{D.5})$$

$$\dot{\mathbf{n}}_i(t) = \omega_\phi \mathbf{n}_i(t) \times (\mathbf{G}_i(t) \times \mathbf{n}_i(t)) \quad (\text{D.6})$$

$$\dot{\mathbf{v}}_i(t) = -\beta \mathbf{v}_i(t) + \kappa v_0 \mathbf{n}_i(t) \quad (\text{D.7})$$

Where

$$\mathbf{H}_i(\mathbf{r}_i, t) = \Delta_{cAMP} \frac{\nabla C(\mathbf{r}, t)}{\langle C(\mathbf{r}, t) \rangle_{local}} + \sqrt{\eta_\phi} \xi_1(t) \quad (\text{D.8})$$

$$\mathbf{G}_i(t) = \mathbf{g}_i(t) + \sqrt{\eta_\theta} \xi_2(t) \quad (\text{D.9})$$

Where both  $\xi_1(t)$  and  $\xi_2(t)$  are 2D white noise sources whose components have zero mean and correlations  $\langle \xi_i(t) \xi_j(s) \rangle = \delta_{ij} \delta(t - s)$ . We choose  $v_0 = 10 \mu\text{m}/\text{min}$ , which is commonly seen in a chemotaxing cells [36]. The variable  $\kappa$  is fit to ensure that

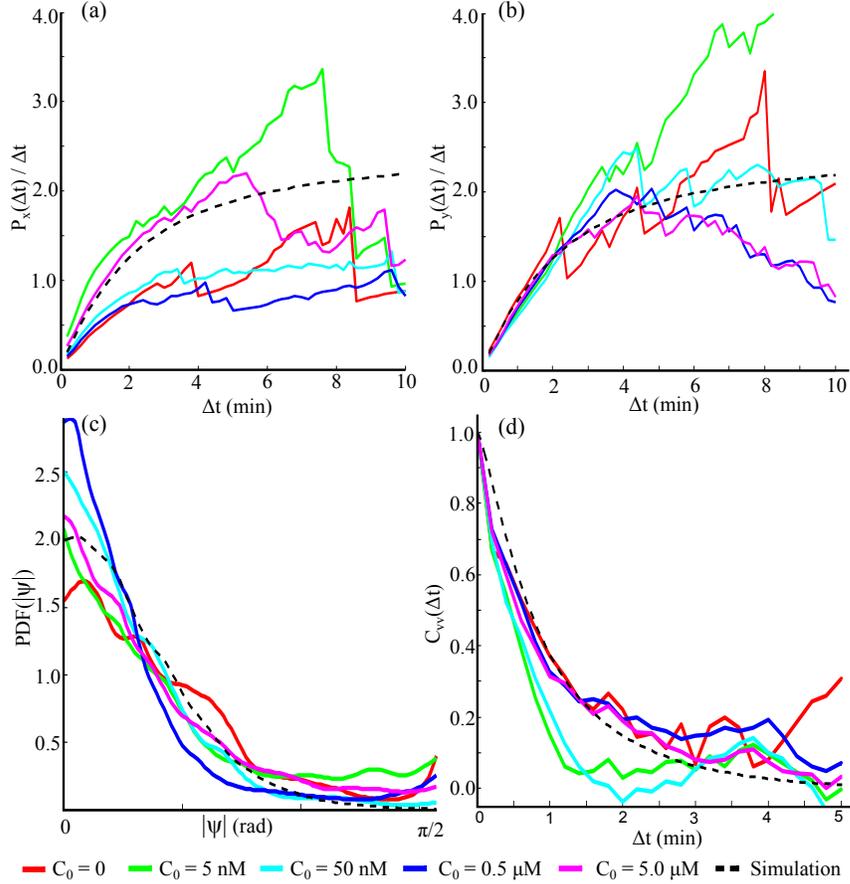


Figure D.2: The signal-to-noise as seen by a cell as a function of its position in the chamber. The noise strength  $\eta_\phi$  is chosen so that the signal-to-noise reaches unity as the cell approaches the end of the chamber (i.e. for  $y = L_y$ )

$v_0$  is the root-mean-square speed of the ensemble. It is time independent and can be calculated analytically (see section), but is calculated numerically before each simulation. We set  $\eta_\phi$  to the ligand binding time  $\approx 1 \text{ s} = 1/60 \text{ min}$  [115].

To fit  $\beta$ , we use that fact that our model in the absence of signal is a Furth model, and fit  $\beta^{-1}$  to the persistence time corresponding with experiments done with aca- mutant cells. The standard way to measure the persistence of a cell is to fit it's mean squared displacement ( $\text{MSD}(\Delta t) := \langle |\mathbf{r}(t + \Delta t) - \mathbf{r}(t)|^2 \rangle$ ) to the Furth model [116–118] of a 2D persistent random walker,

$$\text{MSD}(\Delta t) = 2v^2 t_p (\Delta t - t_p (1 - e^{-\Delta t/t_p})). \quad (\text{D.10})$$

Here,  $v$  is the mean squared value of a component of the cells velocity and  $t_p$  is the persistence time. It can be seen that as  $\lim_{\Delta t \rightarrow \infty} \text{MSD}(\Delta t)/\Delta t = 2v^2 t_p$ , consistent with overall diffusive motion. However, in the no-flow gradient chamber, the presence of an externally imposed gradient causes an overall drift in the direction of the gradient. This means that the velocity components are no longer equal, and the long time behavior of the cell in the lab frame is not diffusive. To measure the persistence of the cells in the presence of a drift, we have to consider the mean squared displacement of individual components (i.e.,  $x_i \in \{x, y\}$ ) in the co-moving frame. We denote the mean squared displacement of a component with  $\overline{\text{MSD}}(\Delta t; x_i)$ . This quantity is defined as

$$\overline{\text{MSD}}(\Delta t, x_i) = \langle (x_i(t + \Delta t) - x_i(t) - \Delta t \langle \mathbf{V} \rangle_{x_i})^2 \rangle. \quad (\text{D.11})$$

where  $\langle \mathbf{V} \rangle_{x_i}$  is the  $x_i$  component of the drift velocity. We define the persistence of the cells as the ratio of the mean squared displacement and the mean squared speed in the co-moving frame,

$$P_{x_i}(\Delta t) = \frac{\overline{\text{MSD}}(\Delta t; x_i)}{\langle v_{x_i}^2 \rangle} \quad (\text{D.12})$$

Here, by dividing out the square of the appropriate velocity component, we remove the effect of the signal's directional bias from the mean squared displacement and can therefore quantify the persistence in the presence of drift. Fig.D.2.a and D.2.b show  $P_x(\Delta t)/\Delta t$  and  $P_y(\Delta t)/\Delta t$  for the aca- mutant cells in a variety of concentrations. The data is very noisy, owing to the low number of long tracks in the experiments. However, there does not seem to be an overall bias with respect to increasing signal. The black dotted line is taken from our simulation with  $\beta = 0.8 \text{ min}^{-1}$  and  $\eta_\theta = 0.25 \text{ min}$ .

The persistence was found to fit well with  $t_p \approx \beta^{-1} = 1.25 \text{ min}$ . It should be noted that this timescale is much shorter than normally seen in *Dictyostellium* discodium migrating in the absence of signal, which normally have a  $t_p \approx 6 \text{ min}$  [116]. Our data here suggests that a gradient changes the persistence dynamics of the cell. We also noted here that overall in the experiments,  $P_x(\Delta t) < P_y(\Delta t)$ , highlighting the overall polarization bias of the ensemble resulting in more persistent motion upwards.

The final variable  $\eta_\theta$  was adjusted to fit our orientation statistics. Fig.D.2(c) and D.2(d) show the angular distribution and autocorrelation of the orientation of

aca- mutant cells in a variety of concentrations, where the orientation is taken as the direction of center-of-mass velocity. Again, the dashed line representing results of our simulation fit right inside the data.

### D.3 The time-scale independence of the fraction of broken links

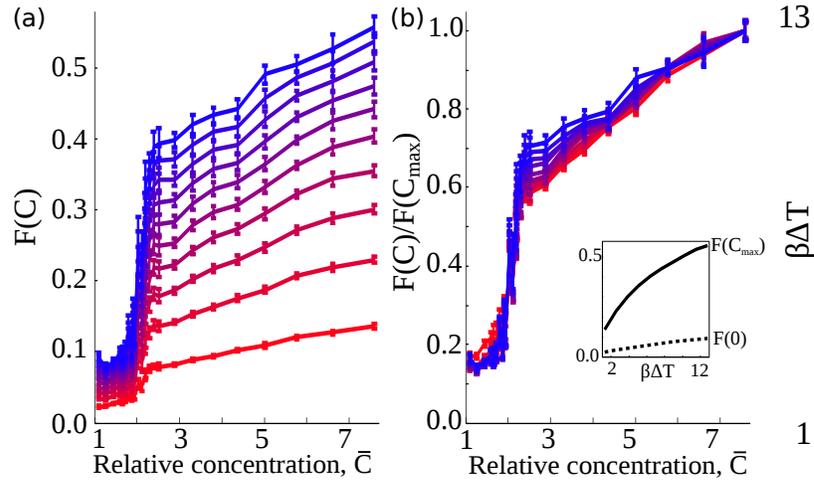


Figure D.3: The fraction of broken links is determined by the time scales of the model and not the actual timescale of the temporal network window

The fraction of broken link measure,  $F(\bar{C})$ , is created using both a length scale and time scale: for us, our length scale was the radius of a cell to make it a contact network, and the time scale decides how far in the future we're looking. Too small a time scale, and the cells will remain in contact mainly due to persistence and not to interactions. Too long a time scale, and an appreciable number of cells will have left the chamber.

In our work, we chose the time scale to be twice the persistence time, roughly corresponding with  $2\beta^{-1}$ , but in general found the functional form of  $F(\bar{C})$  to be unaffected by time scales as long as  $6\beta^{-1}$ , as can be seen in Figure 3.a and 3.b.

What is affected is  $F(0)$ , the number broken for no cell interaction, and to a much smaller extent  $F_{\max}$  (see Figure 3.b inset). Therefore,  $F(\bar{C})/F_{\max}$  is an appropriate and robust measure of the dynamics of the system.

#### D.4 Condition for signal independent ensemble speed

The interaction between the membrane and the forces generated by the polymerizing actin forms protrusions on the cell membrane. The duration of a protrusion and the frequency of protrusions have been measured to be around 20 seconds [119, 120]. Nevertheless, cell motion is has been seen result from a history of protrusions [90]. In our model, we consider that the velocity of a cell  $\mathbf{v}_i$  is driven in the direction of generated protrusions, the form of which in the absence of interactions is given by

$$\dot{\mathbf{v}}_i(t) = -\beta\mathbf{v}_i(t) + \kappa v_0 \mathbf{n}_i(t) \quad (\text{D.13})$$

We set  $\mathbf{v}_i(-\infty) = \mathbf{0}$ , suggesting that all of the cells started from rest and have progressed to their current state over an infinite number of protrusions. This allows us to write the solution to the above equation as

$$\mathbf{v}_i(t) = \kappa v_0 \int_{-\infty}^t e^{-\beta(t-t')} \mathbf{n}_i(t') dt' . \quad (\text{D.14})$$

Here we will calculate the necessary form of  $\kappa$  in order to force the root-mean-square of the speed to be independent of time and consequently the external signal, which

is observed in [36], i.e.  $v_{\text{rms}} := \sqrt{\langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(t) \rangle} = v_0$ , where  $v_0$  is the steady state speed of the ensemble. We calculate the  $v_{\text{rms}}$  from Eq. (D.13),

$$v_{\text{rms}}^2 = \kappa^2 v_0^2 \int_{-\infty}^t e^{-\beta(2t-t'_1-t'_2)} \langle \mathbf{n}_i(t'_1) \cdot \mathbf{n}_i(t'_2) \rangle dt'_1 dt'_2. \quad (\text{D.15})$$

Since the limits of the integrals were the same, we left out the integral sign for the second integral. Next, we replace the correlations of the actin activity with  $\langle \mathbf{n}_i(t'_1) \cdot \mathbf{n}_i(t'_2) \rangle = C_f e^{-\omega_f |t'_1-t'_2|} + C_s e^{-\omega_s |t'_1-t'_2|}$ , where  $\omega_f$  and  $\omega_s$  are the fast and slow processes in the actin activity dynamics and  $C_f$  and  $C_s$  are the associated weights of the corresponding processes in the correlations.

$$v_{\text{rms}}^2 = \kappa^2 v_0^2 (C_f I(\omega_f) + C_s I(\omega_s)), \quad (\text{D.16})$$

where  $I(\omega) := \int_{-\infty}^t e^{-\beta(2t-t'_1-t'_2)-\omega|t'_1-t'_2|} dt'_1 dt'_2$ . In order to satisfy  $v_{\text{rms}} = v_0$ , we get the condition  $\kappa = (C_f I(\omega_f) + C_s I(\omega_s))^{-1/2}$ . Lastly, we calculate the  $I(\omega)$  in order to find the explicit form of  $\kappa$ .

$$I(\omega) = e^{-2\beta t} \int_{-\infty}^t dt'_1 \left( \int_{-\infty}^{t'_1} dt'_2 e^{\beta(t'_1+t'_2)-\omega(t'_1-t'_2)} + \int_{t'_1}^t dt'_2 e^{\beta(t'_1+t'_2)+\omega(t'_1-t'_2)} \right).$$

Carrying out the  $t'_2$  integration for the integrals in the parentheses we get

$$I(\omega) = e^{-2\beta t} \int_{-\infty}^t dt'_1 \left( \frac{e^{(\beta-\omega)t'_1}}{\beta+\omega} e^{(\beta+\omega)t'_1} + \frac{e^{(\beta+\omega)t'_1}}{\beta-\omega} \left( e^{(\beta-\omega)t} - e^{(\beta-\omega)t'_1} \right) \right).$$

The first, second and third integral results  $\frac{e^{2\beta t}}{2\beta(\beta+\omega)}$ ,  $\frac{e^{2\beta t}}{(\beta-\omega)(\beta+\omega)}$  and  $-\frac{e^{2\beta t}}{2\beta(\beta-\omega)}$ , and consequently the prefactor eliminates all the time dependence. When simplified we get  $I(\omega) = (\beta(\beta + \omega))^{-1}$ . Using this expression we can show  $\kappa$  as a function of the time autocorrelations of actin biasing,  $\mathbf{n}(t)$ ,

$$\kappa = \sqrt{\beta} \left( \frac{C_f}{\beta + \omega_f} + \frac{C_s}{\beta + \omega_s} \right)^{-1/2}. \quad (\text{D.17})$$

This variable is independent of time, and only depends on the correlation variables of the second compass and the timescale  $\beta$ , meaning that we can set the ensemble speed while allowing fluctuations in the individual cell's velocity.

## D.5 The parameter regime for two-time scale correlations

The analytical form of  $\kappa$  assumes  $\mathbf{n}$  autocorrelations of the form  $C_f e^{-\omega_f |t'_1 - t'_2|} + C_s e^{-\omega_s |t'_1 - t'_2|}$ . In this section we show the necessary conditions for our model parameters to exhibit have autocorrelations in that form. In the strong external gradients we can linearize equations (D.5) and (D.6),

$$\dot{\phi} = -\omega_\phi(\phi + \xi_\phi) \quad (\text{D.18})$$

$$\dot{\theta} = -\omega_\theta(\theta - \phi + \xi_\theta) \quad (\text{D.19})$$

Then we take the Fourier transform and rewrite the equation for  $\theta$  in the frequency domain

$$\tilde{\theta}(\omega) = \frac{\omega_\theta}{-i\omega + \omega_\theta} \left( \frac{\omega_\phi \tilde{\xi}_\phi(\omega)}{-i\omega + \omega_\phi} + \tilde{\xi}_\theta(\omega) \right), \quad (\text{D.20})$$

where the variable with a tilde are the Fourier transform of the corresponding variable (i.e.,  $\tilde{\theta}(\omega) := \int_{-\infty}^{\infty} e^{i\omega t} \theta(t) dt$ ). We then employ the Wiener-Khintchine theorem,  $\pi\delta(\omega - \omega') S_\theta(\omega) = \langle \tilde{\theta}(\omega) \tilde{\theta}^*(\omega') \rangle$ , where  $S_\theta$  is the power spectrum of  $\theta$

$$S_\theta(\omega) = 2 \int_{-\infty}^{\infty} e^{i\omega t} \langle \theta(t) \theta(0) \rangle dt. \quad (\text{D.21})$$

Using (D.20), we can calculate the correlations as

$$\langle \tilde{\theta}(\omega) \tilde{\theta}^*(\omega') \rangle = \frac{\omega_\theta^2}{\omega^2 + \omega_\theta^2} \left[ \frac{\omega_\phi^2}{\omega^2 + \omega_\phi^2} \langle \tilde{\xi}_\phi(\omega) \tilde{\xi}_\phi^*(\omega') \rangle + \langle \tilde{\xi}_\theta(\omega) \tilde{\xi}_\theta^*(\omega') \rangle \right]. \quad (\text{D.22})$$

The power spectrum of white noise corresponds to the noise intensity,  $\langle \tilde{\xi}_\theta(\omega) \tilde{\xi}_\theta^*(\omega') \rangle = 2\pi\delta(\omega - \omega') \eta_\theta$ . Plugging that in Eq. (D.22) we obtain

$$S_\theta(\omega) = \frac{2\omega_\theta^2 \omega_\phi^2 \eta_\phi}{(\omega^2 + \omega_\theta^2)(\omega^2 + \omega_\phi^2)} + \frac{2\omega_\theta^2 \eta_\theta}{(\omega^2 + \omega_\theta^2)}. \quad (\text{D.23})$$

We associate the estimated correlations of  $\mathbf{n}$  with our fast and slow timescales as  $\langle \theta(\tau) \theta(0) \rangle \approx A_\phi e^{-\omega_\phi |\tau|} + A_\theta e^{-\omega_\theta |\tau|}$  and use in the right-hand side of Eq. (D.21) to

get

$$S_\theta(\omega) = \frac{4A_\phi\omega_\phi}{\omega^2 + \omega_\phi^2} + \frac{4A_\theta\omega_\theta}{\omega^2 + \omega_\theta^2}. \quad (\text{D.24})$$

Finally, solving for  $\eta_\phi$  and  $\eta_\theta$  from the equations (D.23) and (D.24) we get

$$\eta_\phi = \frac{2A_\phi}{\omega_\phi} \left( 1 - \frac{\omega_\phi^2}{\omega_\theta^2} \right) \quad (\text{D.25})$$

$$\eta_\theta = \frac{2}{\omega_\theta} \left( A_\theta - A_\phi \frac{\omega_\phi}{\omega_\theta} \right). \quad (\text{D.26})$$

The noisy intensities are non-negative, which implies  $\omega_\theta \geq \omega_\phi$  and  $A_\theta/A_\phi \geq \omega_\phi/\omega_\theta$ .

Therefore, our model shows two-time velocity autocorrelations, which obey the conditions stated above.

## D.6 Numerical details

The first step in simulating the no-flow gradient chamber is to handle the concentration profile of cAMP. This is done by solving in time the diffusion equation given in (1). This equation is solved with the boundary conditions  $C(x, 0) = 0$ ,  $C(x, L_y) = C_0$ , and  $C(0, y) = C(L_x, y)$ . We picked mesh densities such that each cell is able to resolve at least 5 mesh points, and the contributions of each cell's secreted value  $s_{cell}$  is placed at the cell's rear. The concentration profile is updated using the Forward-Time, Center-Difference strategy, with a time step set to ensure the Courant-Friedrichs-Lewy condition  $D_{\text{cAMP}}\Delta t/(\Delta x\Delta y) < 1/4$ . We then determine the quantity  $\nabla C/|C_{\text{local}}|$  at each mesh point and allow the cell to "measure" the

local signal at it's front, namely

$$\mathbf{H}_i(\mathbf{r}_i, t) = \mathbf{h}_0(\mathbf{r}_i, t) + \xi_\phi(t). \quad (\text{D.27})$$

Where we have used

$$\mathbf{h}(\mathbf{r}, t) = L_y \frac{\nabla C(\mathbf{r}, t)}{\langle C(\mathbf{r}, t) \rangle_{\text{local}} + K_D}, \quad (\text{D.28})$$

$$\xi_\phi(t) = L_y \sqrt{\frac{\eta_\phi}{D_{cAMP}/\nu}} \xi_1(t). \quad (\text{D.29})$$

Where both  $\xi_1(t)$  is a 2D white noise source whose components have zero mean and correlations  $\langle \xi_i(t) \xi_j(s) \rangle = \delta_{ij} \delta(t - s)$ . With these formulas,  $\mathbf{g}(t)$  and  $\mathbf{n}(t)$  evolve under the equations

$$\dot{\mathbf{g}}_i(t) = \omega_\phi \mathbf{g}_i(t) \times (\mathbf{H}_i(t) \times \mathbf{g}_i(t)) \quad (\text{D.30})$$

$$\dot{\mathbf{n}}_i(t) = \omega_\phi \mathbf{n}_i(t) \times (\mathbf{G}_i(t) \times \mathbf{n}_i(t)) \quad (\text{D.31})$$

Where, similar to  $\mathbf{g}(t)$ 's attractor,

$$\mathbf{G}_i(t) = \mathbf{g}_i(t) + \xi_\theta(t), \quad (\text{D.32})$$

$$\xi_\theta(t) = \sqrt{\eta_\theta} \xi_2(t). \quad (\text{D.33})$$

Where  $\xi_2(t)$  is a 2D white noise source with the same mean and correlations as  $\xi_1(t)$ .

Using the definitions of  $G$  and  $H$  and expanding the triple vector product, we can

rewrite Eqns. D.30 and D.31 as

$$\dot{\mathbf{g}} = \omega_\phi(\mathbf{h}_0 - \mathbf{g}(\mathbf{h}_0 \cdot \mathbf{g})) + \omega_\phi(\boldsymbol{\xi}_\phi - \mathbf{g}(\boldsymbol{\xi}_\phi \cdot \mathbf{g})) \quad (\text{D.34})$$

$$\dot{\mathbf{n}} = \omega_\theta(\mathbf{g} - \mathbf{n}(\mathbf{g} \cdot \mathbf{n})) + \omega_\theta(\boldsymbol{\xi}_\theta - \mathbf{n}(\boldsymbol{\xi}_\theta \cdot \mathbf{n})) \quad (\text{D.35})$$

where it is understood that the two compass directions  $\mathbf{g}$  and  $\mathbf{n}$  have unit magnitude.

For simulation purposes, we can continue rewriting our compass equations as

$$\dot{\mathbf{g}} = \omega_\phi \tilde{A}(\mathbf{g})(\mathbf{h}_0 + \boldsymbol{\xi}_\phi) \quad (\text{D.36})$$

$$\dot{\mathbf{n}} = \omega_\theta \tilde{A}(\mathbf{n})(\mathbf{g} + \boldsymbol{\xi}_\theta) \quad (\text{D.37})$$

where  $\tilde{A}(\mathbf{r})$  is a matrix with the form

$$\tilde{A}(\mathbf{r}) = \begin{vmatrix} r_y^2 & -r_x r_y \\ -r_x r_y & r_x^2 \end{vmatrix}.$$

Now, (D.36) and (D.37) can be solved numerically (using Euler-Maruyama [CITATION]). Even though the triple vector product preserves the magnitude of the vectors, both  $\mathbf{g}(t)$  and  $\mathbf{n}(t)$  were normalized after each integration step to remove complications due to rounding error. The remaining step to progress the cells forward involves simple collision detection to include the adhesion and repulsive forces.

## Bibliography

- [1] Principles of genetics biol2250, memorial university of newfoundland. [http://www.mun.ca/biology/desmid/brian/BIOL2250/Week\\_DevoGenetics/1GeneDG.html](http://www.mun.ca/biology/desmid/brian/BIOL2250/Week_DevoGenetics/1GeneDG.html). Accessed: 2015-03-20.
- [2] Keiko U Torii. Two-dimensional spatial patterning in developmental systems. Trends in cell biology, 22(8):438–46, August 2012.
- [3] 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):371–5, June 2013.
- [4] Mikael Hggstrm. Medical gallery of mikael hggstrm 2014. Wikiversity Journal of Medicine, 1(2), 2014.
- [5] Eilon Sherman, Valarie Barr, Suliana Manley, George Patterson, Lakshmi Balagopalan, Itoro Akpan, Carole K. Regan, Robert K. Merrill, Connie L. Sommers, Jennifer Lippincott-Schwartz, and Lawrence E. Samelson. Functional nanoscale organization of signaling molecules downstream of the t cell antigen receptor. Immunity, 35:705 – 720, 2011.
- [6] Gene L Garcia, Erin C Rericha, Christopher D Heger, Paul K Goldsmith, and Carole A Parent. The group migration of Dictyostelium cells is regulated by extracellular chemoattractant degradation. Molecular biology of the cell, 20(14):3295–304, July 2009.
- [7] Colin P McCann, Erin C Rericha, Chenlu Wang, Wolfgang Losert, and Carole Parent. Dictyostelium Cells Migrate Similarly on Surfaces of Varying Chemical Composition. PLoS ONE, 9(2):e87981, 2014.
- [8] King Lam Hui, Chenlu Wang, Brian Grooman, Jessica Wayt, and Arpita Upadhyaya. Membrane Dynamics Correlate with Formation of Signaling Clusters during Cell Spreading. Biophys J, 102(7):1524–1533, April 2012.

- [9] T Gallai and HSM See. Coxeter: Introduction to geometry, 1961.
- [10] Darcy Wentworth Thompson et al. On growth and form. On growth and form., 1942.
- [11] Benoit B Mandelbrot. The fractal geometry of nature. Macmillan, 1983.
- [12] Cris Kuhlemeier. Phyllotaxis. Trends in plant science, 12(4):143–50, April 2007.
- [13] Philip Ball. Shapes: nature’s patterns: a tapestry in three parts. Oxford University Press, 2011.
- [14] Tetsuya Tabata and Yuki Takei. Morphogens, their identification and regulation. Development (Cambridge, England), 131(4):703–12, March 2004.
- [15] A. M. Turing. The Chemical Basis of Morphogenesis. Philosophical Transactions of the Royal Society B: Biological Sciences, 237(641):37–72, August 1952.
- [16] Istvan Lengyel and Irving R Epstein. A chemical approach to designing turing patterns in reaction-diffusion systems. Proceedings of the National Academy of Sciences, 89(9):3977–3979, 1992.
- [17] Enrico Scarpella, Danielle Marcos, Jiří Friml, and Thomas Berleth. Control of leaf vascular patterning by polar auxin transport. Genes & development, 20(8):1015–1027, 2006.
- [18] Didier Reinhardt, Eva-Rachele Pesce, Pia Stieger, Therese Mandel, Kurt Baltesperger, Malcolm Bennett, Jan Traas, Jiří Friml, and Cris Kuhlemeier. Regulation of phyllotaxis by polar auxin transport. Nature, 426(6964):255–260, 2003.
- [19] Michael Cohen, Buzz Baum, and Mark Miodownik. The importance of structured noise in the generation of self-organizing tissue patterns through contact-mediated cell-cell signalling. Journal of the Royal Society, Interface / the Royal Society, 8(59):787–98, June 2011.
- [20] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. Molecular Biology of the Cell, 2002.
- [21] Zeev Pancer and Max D Cooper. The evolution of adaptive immunity. Annual review of immunology, 24:497–518, January 2006.
- [22] Lakshmi Balagopalan, Nathan P Coussens, Eilon Sherman, Lawrence E Samelson, and Connie L Sommers. The LAT story: a tale of cooperativity, coordination, and choreography. Cold Spring Harbor perspectives in biology, 2(8):a005512, August 2010.

- [23] Eric Betzig, George H. Patterson, Rachid Sougrat, Wolf O. Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science, 313(5793):1642–1645, August 2006.
- [24] Amos Gilhar and Richard S Kalish. Alopecia areata: a tissue specific autoimmune disease of the hair follicle. Autoimmunity reviews, 5(1):64–9, January 2006.
- [25] Gyorgy Nagy, Agnes Koncz, and Andras Perl. T- and B-cell abnormalities in systemic lupus erythematosus. Critical reviews in immunology, 25(2):123–40, January 2005.
- [26] Philippe V Afonso, Mirkka Janka-Junttila, Young Jong Lee, Colin P McCann, Charlotte M Oliver, Khaled A Aamer, Wolfgang Losert, Marcus T Cicerone, and Carole A Parent. LTB4 is a signal-relay molecule during neutrophil chemotaxis. Developmental cell, 22(5):1079–91, May 2012.
- [27] Tatyana Chtanova, Marie Schaeffer, Seong-Ji Han, Giel G van Dooren, Marcelo Nollmann, Paul Herzmark, Shiao Wei Chan, Harshita Satija, Kristin Camfield, Holly Aaron, Boris Striepen, and Ellen A Robey. Dynamics of neutrophil migration in lymph nodes during infection. Immunity, 29(3):487–96, September 2008.
- [28] Tamás Németh and Attila Mócsai. The role of neutrophils in autoimmune diseases. Immunology letters, 143(1):9–19, March 2012.
- [29] D. Filmore. It’s a GPCR world. Modern Drug Discovery, 7:24–28, 2004.
- [30] Paul W Kriebel, Valarie A Barr, Erin C Rericha, Guofeng Zhang, and Carole A Parent. Collective cell migration requires vesicular trafficking for chemoattractant delivery at the trailing edge. The Journal of cell biology, 183(5):949–61, December 2008.
- [31] G L Garcia and C A Parent. Signal relay during chemotaxis. J Microsc, 231(3):529–34, September 2008.
- [32] Carole A Parent. Making all the right moves: chemotaxis in neutrophils and Dictyostelium. Current opinion in cell biology, 16(1):4–13, February 2004.
- [33] P.W. Kriebel, V.A. Barr, and C.A. Parent. Adenylyl cyclase localization regulates streaming during chemotaxis. Cell, 112(4):549–560, 2003.
- [34] G.L. Garcia, E.C. Rericha, C.D. Heger, P.K. Goldsmith, and C.A. Parent. The group migration of dictyostelium cells is regulated by extracellular chemoattractant degradation. Mol Biol Cell, 20(14):3295–3304, July 2009.

- [35] S Lee, C A Parent, R Insall, and R A Firtel. A novel Ras-interacting protein required for chemotaxis and cyclic adenosine monophosphate signal relay in Dictyostelium. Molecular biology of the cell, 10(9):2829–45, September 1999.
- [36] Colin P McCann, Paul W Kriebel, Carole A Parent, and Wolfgang Losert. Cell speed, persistence and information transmission during signal relay and collective migration. J Cell Sci, 123(Pt 10):1724–31, May 2010.
- [37] Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. Statistical Analysis and Modeling of Spatial Point Patterns. Statistics in Practice. John Wiley and Sons, New York, 2008.
- [38] Chiho Kaito, Ulf Dieckmann, Akira Sasaki, and Fugo Takasu. Beyond pairs: Definition and interpretation of third-order structure in spatial point patterns. Journal of theoretical biology, 372:22–38, February 2015.
- [39] K. Schladitz and A. J. Baddeley. A Third Order Point Process Characteristic. Scandinavian Journal of Statistics, 27(4):657–671, December 2000.
- [40] Stefano Borgani, Silvio A Bonometto, Massimo Persic, and Paolo Salucci. Correlation functions of matter from galaxy rotation curves. The Astrophysical Journal, 374:20–28, 1991.
- [41] A. J. Baddeley and B. W. Silverman. A cautionary example on the use of second-order methods for analyzing point patterns. Biometrics, 40(4):pp. 1089–1093, 1984.
- [42] J. P. Diggle. Statistical Analysis of Spatial Point Patterns. Academic Press, New York, 1983.
- [43] Klaus R. Mecke and Dietrich Stoyan. Morphological characterisation of point patterns. Biometrical Journal, 47(4):473–488, 2005.
- [44] Joshua Parker, Eilon Sherman, Matthias van de Raa, Devaraj van der Meer, Lawrence E. Samelson, and Wolfgang Losert. Automatic sorting of point pattern sets using Minkowski functionals. Phys. Rev. E, 88:022720, Aug 2013.
- [45] Janine Illian, Erica Benson, John Crawford, and Harry Staines. Principal component analysis for spatial point processes assessing the appropriateness of the approach in an ecological context. In Adrian Baddeley, Pablo Gregori, Jorge Mateu, Radu Stoica, Dietrich Stoyan, P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, and S. Zeger, editors, Case Studies in Spatial Point Process Modeling, volume 185 of Lecture Notes in Statistics, pages 135–150. Springer New York, 2006.
- [46] Beifang Chen. A Simplified Elementary Proof of Hadwiger’s Volume Theorem. Geometriae Dedicata, 105(1):107–120, April 2004.
- [47] David J. Strauss. A model for clustering. Biometrika, 62(2):467–475, 1975.

- [48] Eilon Sherman, Valarie Barr, and Lawrence E Samelson. Super-resolution characterization of TCR-dependent signaling clusters. Immunological reviews, 251(1):21–35, January 2013.
- [49] Joshua Parker, Eilon Sherman, Valarie Barr, Lawrence E. Samelson, and Wolfgang Losert. Heterogeneous cluster analysis of lat protein point pattern sets reveal changes in clustering upon t-cell activation. to be submitted to Physical Review E, 2015.
- [50] Satarupa Das, Joshua Parker, Peter Kriebel, Can Guven, Dan Larson, Wolfgang Losert, and Carole Parent. Aca-mrna is spatially localized to the back of chemotaxing dictyostelium cells maintaining aca polarity and its translation during signal relay. to be submitted, 2015.
- [51] Joshua Parker, Can Guven, Chenlu Wang, Edward Ott, and Wolfgang Losert. A broken-contact order parameter for inferring inter-cellular communication from patterns of externally guided migrating cells. to be submitted to Physical Review E, 2015.
- [52] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel Type of Phase Transition in a System of Self-Driven Particles. Phys. Rev. Lett., 75:1226–1229, Aug 1995.
- [53] C Joanne Wang, Adriel Bergmann, Benjamin Lin, Kyuri Kim, and Andre Levchenko. Diverse sensitivity thresholds in dynamic signaling responses by social amoebae. Sci Signal, 5(213):ra17, February 2012.
- [54] Richard Law, Janine Illian, David F. R. P. Burslem, Georg Gratzer, C. V. S. Gunatilleke, and I. A. U. N. Gunatilleke. Ecological information from spatial patterns of plants: insights from point process theory. Journal of Ecology, 97(4):616–628, 2009.
- [55] Thorsten Wiegand and Kirk Moloney. Rings, circles, and null-models for point pattern analysis in ecology. Oikos, 104(2):209–229, 2004.
- [56] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-Exciting Point Process Modeling of Crime. Journal of the American Statistical Association, 106(493):100–108, March 2011.
- [57] Anthony C. Gatrell, Trevor C. Bailey, Peter J. Diggle, and Barry S. Rowlingson. Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. Trans. Inst. Br. Geogr., 21(1):256–274, 1996.
- [58] Roger Bivand and Stefan Szymanski. Modelling the spatial impact of the introduction of compulsory competitive tendering. Regional Science and Urban Economics, 30(2):203 – 219, 2000.
- [59] Y. Ogata. Seismicity analysis through point-process modeling: A review. Pure and Applied Geophysics, 155:471–507, 1999.

- [60] S. J. L. Billinge. The atomic pair distribution function: past and present. Zeitschrift fur Kristallographie, 219:117–121, 2004.
- [61] W.S. Kendall. Statistics of the galaxy distribution. Journal of the American Statistical Association, 98:248–248, January 2003.
- [62] E. Lachmanovich, D. E. Shvartsman, Y. Malka, C. Botvin, Y. I. Henis, and A. M. Weiss. Co-localization analysis of complex formation among membrane proteins by computerized fluorescence microscopy: application to immunofluorescence co-patching studies. Journal of microscopy, 212(Pt 2):122–131, November 2003.
- [63] J. O. Ramsay and B. W. Silverman. Functional Data Analysis. Springer Series in Statistics. Springer, New York, 2nd edition, June 2005.
- [64] J. O. Ramsay, Giles Hooker, and Spencer Graves. Functional Data Analysis with R and MATLAB. Springer, New York, 1st edition, 2009.
- [65] A. Tscheschel and D. Stoyan. Statistical reconstruction of random point patterns. Comput. Stat. Data Anal., 51(2):859–871, 2003.
- [66] J. Schmalzing, M. Kerscher, and T. Buchert. Minkowski Functionals in Cosmology. In S. Bonometto, J.R. Primack, and A. Provenzale, editors, Dark Matter in the Universe, page 281, 1996.
- [67] M. Kerscher and A. Tikhonov. Morphology of the local volume. Astron. & Astrophysics, 509:A57, 2010.
- [68] Stany Gallier. A stochastic pocket model for aluminum agglomeration in solid propellants. Propellants, Explos., Pyrotech., 34(2):97–105, 2009.
- [69] Enrico Calzavarini, Martin Kerscher, Detlef Lohse, and Federico Toschi. Dimensionality and morphology of particle and bubble clusters in turbulent flow. Journal of Fluid Mechanics, 607:13–24, 2008.
- [70] K. R. Mecke. Additivity, convexity, and beyond: applications of Minkowski functionals in statistical physics. Lecture Notes in Physics, 554:111–184, 2000.
- [71] P. C. Mahalanobis. On the generalised distance in statistics. In Proceedings National Institute of Science, India, volume 2, pages 49–55, April 1936.
- [72] Jr. Ward. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58:236–244, 1963.
- [73] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. Phys. Rev. E, 70:066111, Dec 2004.
- [74] M. E. J. Newman. Analysis of weighted networks. Phys. Rev. E, 70:056131, Nov 2004.

- [75] R.J. Robb, W.C. Greene, and C.M. Rusk. Low and high affinity cellular receptors for interleukin 2. implications for the level of tac antigen. J Exp Med, 160(4):1126–46, 1984.
- [76] A. Marroquin, J. A. Garcia, J. Garrido, and V. L. Mateos. Neyman-Scott cluster model for daily rainfall processes in lower extremadura (Spain): Rainfall Generating Mechanisms. Theoretical and Applied Climatology, 52(3-4):183–193, 1995.
- [77] Jesper Dall and Michael Christensen. Random geometric graphs. Physical Review E, 66(1):016121, 2002.
- [78] A. J. Baddeley, J. Mller, and R. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica, 54(3):329–350, 2000.
- [79] Eric Marcon, Florence Puech, and Stéphane Traissac. Characterizing the Relative Spatial Structure of Point Patterns. International Journal of Ecology, 2012:1–11, 2012.
- [80] Robert T Abraham and Arthur Weiss. Jurkat T cells and development of the T-cell receptor signalling paradigm. Nature reviews. Immunology, 4(4):301–8, April 2004.
- [81] Edith Gabriel. Estimating Second-Order Characteristics of Inhomogeneous Spatio-Temporal Point Processes. Methodology and Computing in Applied Probability, 16(2):411–431, July 2013.
- [82] C. Comas, F. J. Rodriguez-Cortes, and J. Mateu. Second-order analysis of anisotropic spatiotemporal point process data. Statistica Neerlandica, 69(1):49–66, February 2015.
- [83] Yicong Wu, Peter Wawrzusin, Justin Senseney, Robert S Fischer, Ryan Christensen, Anthony Santella, Andrew G York, Peter W Winter, Clare M Waterman, Zhirong Bao, Daniel A Colón-Ramos, Matthew McAuliffe, and Hari Shroff. Spatially isotropic four-dimensional imaging with dual-view plane illumination microscopy. Nature biotechnology, 31(11):1032–8, November 2013.
- [84] Marisol Corral-Debrinski. mrna specific subcellular localization represents a crucial step for fine-tuning of gene expression in mammalian cells. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1773(4):473 – 475, 2007.
- [85] Christine E. Holt and Simon L. Bullock. Subcellular mrna localization in animal cells and why it matters. Science, 326(5957):1212–1216, 2009.
- [86] Stavroula Mili and Ian G Macara. RNA localization and polarity: from A(PC) to Z(BP). Trends in cell biology, 19(4):156–64, April 2009.

- [87] Andrea M. Femino, Fredric S. Fay, Kevin Fogarty, and Robert H. Singer. Visualization of single rna transcripts in situ. Science, 280(5363):585–590, 1998.
- [88] Jeffrey M Levisky and Robert H Singer. Fluorescence in situ hybridization: past, present and future. Journal of cell science, 116(Pt 14):2833–8, July 2003.
- [89] Z Xiao, N Zhang, D B Murphy, and P N Devreotes. Dynamic distribution of chemoattractant receptors in living cells during chemotaxis and persistent stimulation. The Journal of cell biology, 139(2):365–74, October 1997.
- [90] Meghan K Driscoll, Colin McCann, Rael Kopace, Tess Homan, John T Fourkas, Carole Parent, and Wolfgang Losert. Cell shape dynamics: from waves to migration. PLoS Comput Biol, 8(3):e1002392, March 2012.
- [91] Tatjana Trcek, Jeffrey A Chao, Daniel R Larson, Hye Yoon Park, Daniel Zenklusen, Shailesh M Shenoy, and Robert H Singer. Single-mRNA counting using fluorescent in situ hybridization in budding yeast. Nature protocols, 7(2):408–19, February 2012.
- [92] Hendrik Deschout, Francesca Cella Zanacchi, Michael Mlodzianoski, Alberto Diaspro, Joerg Bewersdorf, Samuel T Hess, and Kevin Braeckmans. Precisely and accurately localizing single emitters in fluorescence microscopy. Nature methods, 11(3):253–66, March 2014.
- [93] P J Van Haastert. Sensory adaptation of Dictyostelium discoideum cells to chemotactic signals. The Journal of cell biology, 96(6):1559–65, June 1983.
- [94] B Varnum-Finney, N A Schroeder, and D R Soll. Adaptation in the motility response to cAMP in Dictyostelium discoideum. Cell motility and the cytoskeleton, 9(1):9–16, January 1988.
- [95] C A Parent and P N Devreotes. A cell’s sense of direction. Science, 284(5415):765–70, April 1999.
- [96] Liang Li, Simon F Norrelykke, 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.
- [97] Nicholas P Barry and Mark S Bretscher. Dictyostelium amoebae and neutrophils can swim. Proc. Natl. Acad. Sci. U.S.A., 107(25):11376–80, June 2010.
- [98] M H Cohen and A Robertson. Wave propagation in the early stages of aggregation of cellular slime molds. J. Theor. Biol., 31(1):101–18, April 1971.
- [99] M J Potel and S A Mackay. Preaggregative cell motion in Dictyostelium. J. Cell. Sci., 36:281–309, April 1979.

- [100] Stephanie Alexander, Gudrun E Koehl, Markus Hirschberg, Edward K Geissler, and Peter Friedl. Dynamic imaging of cancer growth and invasion: a modified skin-fold chamber model. Histochem. Cell Biol., 130(6):1147–54, December 2008.
- [101] Cornelis J Weijer. Collective cell migration in development. J. Cell. Sci., 122(Pt 18):3215–23, September 2009.
- [102] Can Guven, Erin Rericha, Edward Ott, and Wolfgang Losert. Modeling and measuring signal relay in noisy directed migration of cell groups. PLoS Comput. Biol., 9(5):e1003041, May 2013.
- [103] Peter J M van Haastert and Marten Postma. Biased random walk by stochastic fluctuations of chemoattractant-receptor interactions at the lower limit of detection. Biophys. J., 93(5):1787–96, September 2007.
- [104] Azadeh Samadani, Jerome Mettetal, and Alexander van Oudenaarden. Cellular asymmetry and individuality in directional sensing. Proc Natl Acad Sci U S A, 103(31):11549–54, August 2006.
- [105] Danny Fuller, Wen Chen, Micha Adler, Alex Groisman, Herbert Levine, Wouter-Jan Rappel, and William F Loomis. External and internal constraints on eukaryotic chemotaxis. Proc Natl Acad Sci U S A, 107(21):9656–9, May 2010.
- [106] Yuan Xiong, Chuan-Hsiang Huang, Pablo A Iglesias, and Peter N Devreotes. Cells navigate with a local-excitation, global-inhibition-biased excitable network. Proc. Natl. Acad. Sci. U.S.A., 107(40):17079–86, October 2010.
- [107] W. Martinez. Computational statistics handbook with MATLAB. Chapman & Hall/CRC, 2001.
- [108] V. A. Epanechnikov. Non-Parametric Estimation of a Multivariate Probability Density. Theory of Probability & Its Applications, 14(1):153–158, January 1969.
- [109] P Devreotes, D Fontana, P Klein, J Sherring, and A Theibert. Transmembrane signaling in Dictyostelium. Methods in cell biology, 28:299–331, January 1987.
- [110] C A Parent and P N Devreotes. Constitutively active adenylyl cyclase mutant requires neither G proteins nor cytosolic regulators. The Journal of biological chemistry, 271(31):18333–6, August 1996.
- [111] Paul W Kriebel and Carole A Parent. Group migration and signal relay in Dictyostelium. Methods in molecular biology (Clifton, N.J.), 571:111–24, January 2009.
- [112] U K Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227(5259):680–5, August 1970.

- [113] Alexandre David, Brian P Dolan, Heather D Hickman, Jonathan J Knowlton, Giovanna Clavarino, Philippe Pierre, Jack R Bennink, and Jonathan W Yewdell. Nuclear translation visualized by ribosome-bound nascent chain puromycylation. The Journal of cell biology, 197(1):45–57, April 2012.
- [114] Alexandre David, Jack R Bennink, and Jonathan W Yewdell. Emetine optimally facilitates nascent chain puromycylation and potentiates the ribopuromycylation method (RPM) applied to inert cells. Histochemistry and cell biology, 139(3):501–4, March 2013.
- [115] Gabriel Amselem, Matthias Theves, Albert Bae, Carsten Beta, and Eberhard Bodenschatz. Control parameter description of eukaryotic chemotaxis. Phys. Rev. Lett., 109(10):108103, September 2012.
- [116] Liang Li, Edward C Cox, and Henrik Flyvbjerg. ‘Dicty dynamics’: Dictyostelium motility as persistent random motion. Phys Biol, 8(4):046006, August 2011.
- [117] M H Gail and C W Boone. The locomotion of mouse fibroblasts in tissue culture. Biophys. J., 10(10):980–93, October 1970.
- [118] Reinhold Fürth. Die Brownsche Bewegung bei Berücksichtigung einer Persistenz der Bewegungsrichtung. Mit Anwendungen auf die Bewegung lebender Infusorien. Zeitschrift für Physik, 2:244–256, 1920.
- [119] Natalie Andrew and Robert H Insall. Chemotaxis in shallow gradients is mediated independently of PtdIns 3-kinase by biased choices between random protrusions. Nat Cell Biol, 9(2):193–200, February 2007.
- [120] Leonard Bosgraaf and Peter J M Van Haastert. Navigation of chemotactic cells by parallel signaling to pseudopod persistence and orientation. PLoS One, 4(8):e6842, 2009.