CONSTRAINING MODELS TO DATA: A PRACTICAL APPROACH FOR MASTERING SELF-ORGANIZING SYSTEMS

by

CHRISTOPHER ROBERT COTTER

(Under the Direction of Lawrence Shimkets)

ABSTRACT

Self-organization involves the coordination of cell behavior to generate structures many orders of magnitude greater than the individual cells. Self-organization is a hallmark of embryonic development, is central to the immune response, and is involved in tumor metastases. The multitude of possible behavioral cues, the presence of redundant systems, and high levels of noise obscure coordination mechanisms in self-organizing processes. This dissertation proposes a framework to overcome the challenges of studying self-organizing systems by utilizing *in vivo* time-lapse fluorescent imaging of moving cells and their environment to constrain computer simulations to experimental results. A simple nearest-neighbor sampling technique is described to parameterize simulations with experimental data, eliminating the need for broad assumptions or detailed knowledge about the underlying mechanisms generating cell behavior and behavioral cues. The resulting simulations provide a framework for testing hypotheses that are challenging, costly, or impossible to test via traditional biological experiments. Using this framework, a unified understanding of the cell behavior and behavioral cues required for development in the bacterium *Myxococcus xanthus* was identified, and previously unknown cell behaviors required for development were discovered. These results revealed that decreased cell motility inside the aggregates, a biased walk toward aggregate centroids, and alignment among neighboring cells and

in a radial direction to the nearest aggregate are behaviors that enhance *M. xanthus* development. The simulations also indicated that aggregation is generally robust to perturbations in these behaviors and identified possible compensatory mechanisms. The framework can be applied to answer new questions about *M. xanthus* self-organization using the same fluorescent tracking and simulation framework to compare mutant cell behavior to that of the wild type.

INDEX WORDS: Self-organization, Bacteria, Collective Movement, Image Processing, *Myxococcus xanthus*, Time-Lapse Imaging, Fluorescent Imaging, Cell Communication, Nearest-Neighbor Interpolation, Simulation, Modeling

CONSTRAINING MODELS TO DATA: A PRACTICAL APPROACH FOR MASTERING SELF-ORGANIZING SYSTEMS

by

CHRISTOPHER ROBERT COTTER

B.S., Rochester Institute of Technology, 2010

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

CONSTRAINING MODELS TO DATA: A PRACTICAL APPROACH FOR MASTERING SELF-ORGANIZING SYSTEMS

by

CHRISTOPHER ROBERT COTTER

Major Professor:	Lawrence Shimkets
Committee:	Heinz-Bernd Schüttler
	Scott Dougan
	Vincent Starai

Electronic Version Approved

Suzanne Barbour
Dean of the Grdaute School
The University of Georgia
December 2017

DEDICATION

To the people in my life who not only believed I was capable of so much more than I believed myself, but also provided the resources and support to make me a believer as well:

Jane Cotter, Bob Cotter, Brittany Makufka, Lawrence Shimkets

ACKNOWLEDGEMENTS

So many deserve acknowledgment, this list only scratches the surface.

Brittany Makufka, Jane Cotter, and Bob Cotter, who supported my decision to pursue this degree without hesitation, and with a continued belief I would succeed.

Larry Shimkets, my first mentor and educator to see more in me than mediocracy, and demand that I work to produce great work. I am honored to be his last graduate student.

Friends I gained along the way, Hank Kimbrough and Pattie Rennison, who provided some laughter, solace, and distraction from stress filled years.

Heinz-Bernd Schüttler, who was willing to spend his time tutoring me in the ways of theoretical physics and always ready to meet to discuss results and provide ideas. His contributions to my graduate education far surpass the role of a committee member.

The Microbiology office staff, a rare breed of folks who provide true and honest support for all the graduate students in the department.

The people I haven't met, yet were with me every week through the media they produce, including the McElroy family, Justin, Griffin, Travis, and Clint, and the duo Ross Blocher and Carrie Poppie.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **v**

CHAPTER 1: INTRODUCTION **1**

PREFACE 1

EUKARYOTIC SELF-ORGANIZATION 2

MODELING EMERGENT BEHAVIOR 11

MYXOCOCCUS XANTHUS SELF-ORGANIZATION. 13

PROPOSAL OF A DATA DRIVEN EXPERIMENTAL AND MODELING PARADIGM 20

CHAPTER 2: DATA-DRIVEN MODELING REVEALS CELL BEHAVIORS CONTROLLING SELF-ORGANIZATION DURING MYXOCOCCUS XANTHUS DEVELOPMENT **32**

ABSTRACT 33

INTRODUCTION 33

RESULTS 38

DISCUSSION 65

METHODS 69

SUPPLEMENTAL METHODS 79

CHAPTER 3: CONCLUSIONS **99**

MODELS OF DEVELOPMENT 99

ROBUSTNESS AND COMPENSATORY MECHANISMS 105

TESTING MODELS OF AGGREGATION 106

A MODEL OF CELL-TO-CELL SIGNALING DURING AGGREGATION 111

CONCLUDING REMARKS 114

CHAPTER 1

INTRODUCTION

Preface

Motile cells self-organize into elaborate and remarkably functional structures during embryonic development, tumor metastases, the immune response, and wound healing. The three-dimensional structure emerges from cues passed between cells and simple behaviors encoded within the cellular machinery. The resulting complexity is astounding and found all over the biological spectrum. Examples range from the bacterium *Myxococcus xanthus*, which utilizes motility to generate high cell-density mound structures without any known long-range signals (Yang and Higgs, 2014), to the large-scale rearrangements of cells during gastrulation (Keller, 2005) and neural crest development (Theveneau and Mayor, 2012). Even cancer metastasis is driven in part by cooperative behaviors between individual cancer cells to increase their ability to move out of the original cancer node (Marongiu et al., 2012; Paul et al., 2017). The cues guiding individual cell behaviors within these systems span a broad range of mechanisms consisting of soluble chemical and protein factors (Rogers and Schier, 2011), cell surface modification (Friedl and Mayor, 2017), electrical (Cortese et al., 2014), and physical signals (Charras and Sahai, 2014). Often, specific quantitative biochemical measurements are available for ligand/receptor interactions, binding efficiencies, and the downstream cellular systems these signals interact with, such as transcription machinery. How these individual interactions combine to generate robust coordination at size scales many times larger than the individual cells, however, is often not clear from individual measurements. Identifying the driving factors of biological self-organization requires concurrent understanding of how cells move within their environment, how physical factors affect cell movement, and how cells

transmit and respond to coordinating information. Each of these factors represents a unique challenge to quantify experimentally, confounding the overall challenge of understanding self-organization.

To overcome dataset complexity, this work combines *in vivo* quantification of individual cell behavior and the surrounding environment with computational models of self-organization fully driven by the experimental data. By driving models directly from experimental data, model parameterization is explicitly defined by the experimental data collected, greatly reducing the need for prior knowledge and experimental analysis. With the model parameterized early in the experimental process, hypotheses on how self-organization occurs and what biological features are necessary can then be tested in the simulation space. Coordinated dissection of the environmental cues and cell behaviors can then be performed in a controlled manner and quantitatively compared with experimental results.

Eukaryotic self-organization

Types of Cell Motility

With the exception of a few cell types, such as swimming sperm and ciliated paramecia, eukaryotic motility is generated by asymmetric shape change of the cell body through cytoskeleton rearrangement (Keren et al., 2008; Lauffenburger and Horwitz, 1996). The asymmetry required to produce movement can occur by random fluctuations in the concentrations of proteins within the cell body or formed from asymmetric activation of receptor proteins on the surface of the cell by external factors, such as during chemotaxis (Charras and Sahai, 2014; Rogers and Schier, 2011). While the underlying machinery is conserved, the resulting motility phenotype spans a broad range of behaviors. Movement is typically categorized into one of two major types based on cell morphology, amoeboid or mesenchymal. Amoeboid movement is defined by a rounded or ellipsoid cell shape and a lack of strong surface attachments (Lämmermann and Sixt, 2009). During

amoeboid migration, force is generated either by filopodia protruding along the leading edge of the cell that transiently attach to the underlying substrate, or by cytoskeletal rearrangements which allow cytosolic pressure to form blebs on the trailing edge of the cell membrane that push the cell forward. Amoeboid movement can generate cell speeds as high as 4 μm/min (Condeelis and Segall, 2003). A second type of movement, known as mesenchymal migration, involves the extension of pseudopods from the leading pole of the cell followed by contraction of the cell body to generate forward movement. Mesenchymal movement is primarily used in dense tissues and is tightly coupled with proteolytic rearrangement of the surrounding extracellular matrix (ECM) to generate a path for the cell to move through (Condeelis and Segall, 2003). Mesenchymal movements are much slower than amoeboid movement, with average speeds between 0.1-1 μm/min (Palecek et al., 1997).

Eukaryotic cell movement can be broadly defined as single or collective cell movement. Single cell movement is associated with the lack of strong connections to neighboring cells, allowing them to move in independent directions relative to their neighbors. While lacking concrete connections, single cell movement can still exhibit coordinated behaviors via cell-to-cell contact. Migrating cells extend filopodia up to 100 μm from the cell body that interact with neighboring migrating cells and provide movement cues (Teddy and Kulesa, 2004). During collective movement, cells create tight cell-to-cell junctions that bind the cells together and give a more concrete structure to the ordering of the cells (Friedl and Gilmour, 2009; Friedl and Mayor, 2017; Haeger et al., 2015). Collective movement can take the form of 2-D sheets which move as monolayers across an underlying tissue, such as the movement of a sheet of epidermal cells during wound healing (Chi and Trinkaus-Randall, 2013), or as 3-D strands through a tissue, such as during the creation of vascular, lymphatic, or mammary ducts (Mayor and Etienne-Manneville, 2016). Cells moving collectively often differentiate their movement depending on their location within the structure, particularly in

defining the leading edge of the group. Leader cells, so named for their location at the leading edge of the collective movement, sense most of the motility cues within the environment as well as generate most of the motility force. Follower cells are not passive, however. Tight junctions between cells allow them to participate in generating pulling forces (Trepat et al., 2009) and help stabilize the polarization of leader cells (Stramer et al., 2013; Theveneau and Mayor, 2011). The interactions observed between cells during single and collective cell movement define the building blocks of eukaryotic self-organization.

Physical Effectors of Cell Motility

To achieve a self-organized state, cells must stay attached to each other or to a shared ECM. The ECM varies widely in terms of pore size, rigidity, and adhesiveness depending on the tissue type. Consider, for example, the relative rigidity of bone, cartilage, and brain tissues. Best studied in artificial 3-D collagen matrices, eukaryotic cell velocities can span several orders of magnitude, from a few microns per hour to greater than 10 microns per minute, depending on the collagen matrix makeup. Cell adhesion to the ECM is the major factor in defining motility speeds. Cell speeds are highest with intermediate levels of adhesion. When adhesion is low, a lack of traction against the ECM reduces motility, while at high adhesion, decreased detachment from the ECM at the lagging pole limits speed (DiMilla et al., 1991; Hakkinen et al., 2010). The major effectors of cell-ECM adhesion are integrin availability, ECM stiffness, and ECM pore size. Integrins are responsible for the majority of adhesion. These cell surface receptors bind to all major eukaryotic ECM components including collagens, laminins, and fibronectin (Campbell and Humphries, 2011; Hynes, 2002). Integrin adhesion is modulated by integrin density at the cell surface-ECM junction and integrin receptor affinity. Integrin binding affinity is proportional to the amount of force exerted on the bond, with increased force causing conformation changes in the protein's structure that increase affinity (Jin et al., 2004). Maximum cell adhesion is then, in part, defined by the amount of force

the ECM can withstand before bending. Rigid ECM enhances integrin clustering (Paszek et al., 2009) and allows the cell to exert higher motility forces, increasing integrin receptor affinity. Conversely, lowered rigidity decreases adhesion though reduced integrin recruitment and integrin receptor affinity. ECM rigidity is dependent on collagen content, fiber thickness, and the extent of fiber cross-links within the tissue (Shoulders and Raines, 2009). ECM pore size affects migration by modulating the amount of contact the cell has with its environment and through size exclusion. Migration efficiency is optimal in 3-D matrices with pore sizes equal to the size of the cell (Friedl and Wolf, 2010). As ECM density is decreased, pore size becomes larger. When pore size exceeds the size of the cell, the decrease in cell-ECM adhesion slows motility (Harley et al., 2008). As pore size decreases below the size of the cell, the need to rearrange the cell shape to squeeze through pores become the dominating motility-inhibiting force. Cells can squeeze through pore sizes down to approximately the size of the nucleus, which cannot be compressed (Wolf et al., 2013).

Without consideration of cell state or other biochemical cues that may be acting on the cell, the ECM makeup introduces a complicated, multidimensional landscape of inter-dependent variables that can coordinate behaviors and must be considered when studying self-organizing systems. Quantification of how these factors effect cell-cell coordination is complicated by non-linear motility responses to changes in ECM makeup (Carey et al., 2012; Han et al., 2012; Wolf et al., 2013) and the ability of cells to proteolytically modify the environment. Recent developments in 3-D migration assays using reconstituted ECM enable independent control of matrix pore size, rigidity, and ligand density within an experiment (Hoffmann and West, 2013; Mason et al., 2013; Zaman et al., 2006). Systematically varying each variable and quantifying cell motility changes under each condition could be used to generate a multi-dimensional phase space of cell behavior. However, extracting knowledge from resulting multi-dimensional, non-linear, datasets is currently left to the skill and prior knowledge of the experimentalist.

Coordination Through Cell-to-Cell Signaling

In addition to effects of physical cues, cell motility is affected by active enzymatic modification of the environment to transmit spatially- and temporally-specific information between cells. Perhaps the most widely known form of such cell coordination is chemotaxis. Broadly defined, chemotaxis is a biased movement along the concentration gradient of a chemical substance. Chemotaxis can lead to biased movement up a concentration gradient in the case of attractant molecules, or movement biased down a concentration gradient in the presence of repellants. In eukaryotic cells, the large cell size allows gradient detection via gradated receptor activation along the cell body in proportion to the underlying gradient. Polarized receptor activation causes polarization of the cytoskeletal components, driving motility along the gradient (Swaney et al., 2010). Chemotaxis can be used by cells in an independent manner, such as towards a food source in unicellular eukaryotes. However, by generating the active chemicals themselves, cells can use chemotaxis as a form of communication and coordination, such as the utilization of cyclic-AMP gradients to localize spatially scattered, amoeboid cells, to a central location to form a multicellular fruiting body during *Dictyostelium discoideum* development (Loomis, 2014).

A diffusible gradient is not the only mechanism by which cells coordinate motility. Coordination can also arise by direct interactions between cells or their environment. While such mechanisms are less explored, there are a few concrete examples. Cells in environments containing an adhesion gradient will exhibit higher speeds in the direction of higher adhesion. This speed bias leads to biased movement up the adhesion gradient, a behavior known as haptotaxis (McCarthy et al., 1983). In a similar fashion, the enhancement of motility by increased ECM rigidity is also strong enough to bias cell movement preferentially up stiffness gradients, a behavior termed durotaxis (Discher et al., 2005; Lo et al., 2000). Contact inhibition of locomotion causes motile eukaryotic cells to retract their membrane protrusions when they come into contact with protrusions of

another cell (Abercrombie and Heaysman, 1953; Mayor and Carmona-Fontaine, 2010). The contractions inhibit motility in the direction of the neighboring cell. Contact-inhibition contributes to the coordination of neural crest migration. Neural tube border cells inhibit the formation of membrane protrusions, forcing the neural crest cells to stay within the tube. At the same time, neighboring neural crest cells inhibit motility on contact, forcing cells to move down the tube, away from the main cluster of neural crest cells (Theveneau and Mayor, 2011). Coordinated behaviors can also emerge through the intrinsic nature of the cell shape, motility parameters, and environment. Similar to the creation of sand dunes or the periodic structure of crystals, complex multicellular structures may emerge by completely physical interactions between cells (Darnton et al., 2010; Vedula et al., 2013). For example, cells placed within prefabricated channels of varying width display changes in velocity, direction of motion, and type of motility based on channel size (Vedula et al., 2012). When channel sizes are small, cells exhibit directed collective movement in a single direction. As channel sizes are increased, movement becomes less collective, with the formation of cell vortexes and random movement. These density-dependent changes are reminiscent of physical systems, where increasing particle density leads to a phase change to a more ordered regime. While often overlooked in favor of chemical systems, physical mechanisms of self-organization are sufficient to coordinate cell behavior.

Evidence for cell-to-cell signaling as a coordinating mechanism dates back to early experiments in embryology. Examples include work in invertebrates, in which excision of a portion of the cockroach tibia causes the tibia to elongate back to the correct size (French et al., 1976), and diffusible signal gradients along the body of the hydra that control head formation and cell behavior along the length of the body (Bode, 2011; Meinhardt, 2009; Wolpert, 1969). These and many other examples (Meinhardt, 2009; Oppenheim and Yang, 2005; Restrepo et al., 2014; Swaney et al., 2010) suggest that cell fate is not hard coded. Instead, cells access cues within their local environment to

determine behavior and fate. Even before the discovery of the underlying biochemistry, models of diffusive molecules, termed morphogens, were proposed to provide the cell-cell communication required to relay positional information (Crick, 1970; Stumpf, 1966; Wolpert, 1969). By gradating morphogen concentration in space, local morphogen concentrations can provide sufficient positional information for cells to determine their location within the gradient field. Such a gradient can be formed by secretion of the morphogen from a source tissue or cell and diffusion throughout the surrounding tissue. Cells do not need to sense the direction of the concentration gradient for coordination, a mechanism as simple as a predetermined concentration threshold is sufficient to determine position. In such systems, high concentration thresholds would be activated in cells near the source of the gradient, while progressively lower thresholds would be activated further away from the source. Periodic patterns such as spots and stripes, or non-periodic patterns similar to the red, white, and blue stripes that make up the French flag could be formed depending on concentration threshold levels (Green and Sharpe, 2015; Wolpert, 2011, 2016). In the French flag example, with three separate cell types spatially arranged like the red, white, and blue stripes on the flag, a gradient of morphogen spanning the length of the flag would be detected by cells that have programmed behaviors based on the local morphogen concentration. Predetermined morphogen concentration cutoffs equal to the concentration at distances one-third and two-thirds across the length of the flag would provide sufficient information to generate different cell behaviors in the red, white, and blue portions of the flag. Stops or stripes could be formed in a similar manner, with periodic concentration-behavior dependencies. These positional information (PI) models, however, do not explain how patterns can generate from random initial conditions. The positional signal must be generated outside the pattern, such as by periphery “anchor” cells, which are not part of the process.

Reaction-diffusion (RD) models of patterning (Gierer and Meinhardt, 1972; Meinhardt, 2009; Meinhardt and Gierer, 1974, 2000; Turing, 1952) forgo the need for periphery cells, instead providing a framework in which patterns can spontaneously form without external cues. The general underlying principle is the interaction of two diffusible molecules, a long-range inhibitor (v) and a short-range activator (u). Stable, non-random, patterns of inhibitor and activator concentrations can form with the addition of a few simple rules: (1) The short-range activator stimulates production of itself as well as production of the inhibitor, (2) The activator has a lower diffusion coefficient (D_u) than the inhibitor (D_v), and (3) the inhibitor inhibits production of the activator. These interactions can be written in just two partial derivatives, commonly known as reaction diffusion equations,

$$\begin{pmatrix} \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial t} \end{pmatrix} = \begin{pmatrix} D_u & 0 \\ 0 & D_v \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} + \begin{pmatrix} \mathrm{F}(u,v) \\ \mathrm{G}(u,v) \end{pmatrix},$$

where functions F and G fulfill requirements (1) and (3) above. Remarkably, by varying only the activation and inhibition rates of the two molecules, it is possible to generate a number of patterns that look like those found in biology, including stripes, spots, and waves of varying size from initial random concentrations of the activator and inhibitor.

Positional information and reaction diffusion models provide evidence for how simple mechanisms can coordinate complex patterns similar to those seen in biological systems. Whether most biological mechanisms underlying biological patterns utilize PI or RD mechanisms is still unknown. While clear evidence for the use of morphogen gradients to convey positional information exists in a number of developmental systems (Bode, 2011; Dormann and Weijer, 2006), PI and RD usually fail to capture complexities of gradient generation, diffusion, and sensing. For example, neither mechanism directly explains the ability of many developmental systems to correctly scale to varying embryo size. Scaling is exemplified in classical experiments in which

embryos surgically reduced in size recover to generate small, but well-proportioned and correctly-patterned, animals (Cooke, 1975, 1981; Wolpert, 2011). To account for scaling, currently accepted models require both gradated positional information and oscillations in gene expression (Cooke and Zeeman, 1976; Oates et al., 2012). Furthermore, many known morphogens have complex feedback mechanisms that modify diffusion rates and clearance, affecting the final concentration gradient (Rogers and Schier, 2011). Examples of morphogens forming complexes such as dimers or higher order structures (Eaton, 2008), interacting with antagonists (Yan et al., 2009; Zakin and Robertis, 2010), or being post-translationally modified (Callejo et al., 2006) also exist. The presence of a gradient also does not guarantee that it is utilized to encode positional information, such as the gradient of retinoic acid in limb buds (Rhinn and Dollé, 2012), or the graded distribution of the protein Caudal in Drosophila embryos (Mlodzik and Gehring, 1987; Mlodzik et al., 1985), which have no known patterning role or do not require a gradient to be active, respectively. Coordination mechanisms also change as the environment changes. For example, in early tissue development, where the cell environment is often soft and amorphous, chemotaxis and growth factors coordinate most cell behaviors. As tissue structure forms and the ECM becomes more rigid, examples of mechanical cues become more prevalent (Haeger et al., 2015). Some cell inputs may also dominate over others, with the secondary input only driving behavior in outlier situations or by acting as a redundancy mechanism (Zaman et al., 2006). For example, up-regulation of ECM-degrading enzymes allow cancer cells to transition from non-motile to mesenchymal movement. The transition is typically a key event in the invasion and migration of tumor cells (Wolf et al., 2003). However, blocking the proteolytic enzymes required to generate a path through the ECM required for mesenchymal motility does not completely inhibit metastasis. As a substitute, cells perform amoeboid type movement and search for paths already present within the ECM.

Modeling Emergent Behavior

Characterizing and predicting cell behavior requires analysis of high-dimension variable spaces which include internal, external, physical, and biochemical effectors that are dynamic in both space and time. Computational and mathematical models provide opportunities to summarize these high dimension variable spaces by identifying variables and regions of variable space that lead to the emergence of collective behavior. Models should directly test clearly stated hypotheses about the system under study, be verified by in vivo experiments, and provide predictions that guide future experiments. Models can be broadly categorized into two types: (1) partial differential equations (PDE) and (2) agent based (AB) models. In PDE models, equations describing how a system changes on average over time are used to predict the state at a desired time, location, or both. Perhaps the most well-known model is the RD model discussed earlier. The RD equations model the change in spatial concentration of diffusible morphogens over time. When used to model biological systems at the phenomenological level, as in RD, PDE models can reproduce phenotypic changes with striking accuracy. For example, zebrafish have a set of horizontal stripes that run along their long axis. Mutation in just one gene (*leopard*) causes the formation of spots in place of the stripes. This striped to spotted transition can be explained by the modification of only one variable in RD models (Asai et al., 1999). Furthermore, the RD model correctly phenocopies the shape and size of the spots for *leopard* mutants with graded levels of activity and correctly predict the pattern that regenerates after laser ablation of a section of the stripes (Yamaguchi et al., 2007). The breadth of detailed knowledge already present for PDE systems from the fields of physics and mathematics can also be used to study emergent behaviors in living systems. Well-studied thermodynamic models can explain some complex biological systems using a few simple terms. These works, for example, show that by modeling cell movement as the temperature variable in thermodynamic models, large scale

patterns form within the systems without the need for long range signaling (Takatori and Brady, 2015).

Phenomenological level PDE models provide evidence for mechanisms that create complex patterning through very simple interactions, but provide little insight into the underlying biology. Current experimental data related to the formation of zebrafish stripes, for example, suggest direct cell-to-cell contacts between local cells activate pigment formation, while longer-range contacts via long cell membrane projections inhibit pigmentation (Watanabe and Kondo, 2015). These results suggest stripe formation in zebra fish follows the overarching theme of short-range activators and long-range inhibitors found in RD models, but forgoing any diffusible agents. As such, care must be taken in imparting biological meaning to phenomenological models that recreate biological patterns. When more detailed experimental measurements are present, PDE models are beneficial in situations where processes and flow are the major elements of interest such as modeling diffusion, receptor occupancy, or internal signaling networks (Lai et al., 2004; Saha and Schaffer, 2006).

Agent-based models employ a finite number of interacting entities, known as agents. Agents within the model can represent whole organisms, cells, and/or individual molecules depending on the model scale. Instead of modeling the average change of systems over time and space, as done in PDE models, each of the agents in AB models are individually modeled by thresholds, if-then rules, or non-linear interactions (Bonabeau, 2002). Such interactions are found throughout self-organized biological systems and their non-linear nature are challenging to capture as in differential equations required for PD models. Furthermore, AB models provide an intuitive model of biological systems, which may be easier to communicate across disciplines.

Both AB and PDE modeling techniques require parameterization through experimental data collection. However, the sheer number of possible environmental parameters and cell responses

make it unclear which variables should be studied, or what the important ranges of the variables are (Charras and Sahai, 2014; Doyle et al., 2013). A workaround technique is to utilize models to estimate what range or relative proportional scales parameters must take to reproduce experimental results. Ancillary experimental measurements, such as the shape or size of the pattern, can then be used to fit the unknown variables to match the model (Reeves et al., 2006). These experiments try to bridge the gap between phenomenological models and models of individual cells or systems. Such models are useful for exploring what might be possible within a biological system, but still require extensive testing of model predictions to confirm model assumptions. Furthermore, even in the best-studied model organisms, key experiments, for example, increasing or decreasing the strength of a chemical gradient, may not be possible.

Myxococcus xanthus Self-Organization.

Self-organization is not only found in eukaryotic systems. The bacterium *Myxococcus xanthus* evolved towards a multicellular lifecycle in which cells work together to feed and survive famine (Shimkets, 1999). *M. xanthus* is typically found within biofilms containing millions of cells many layers thick, an environment similar to the high cell density and complex ECM environments of eukaryotic self-organizing systems. During starvation conditions, cells within the biofilm move in a coordinated manner into aggregates of high cell density. A portion of the cells convert into metabolically dormant spores inside the aggregates. The challenges of understanding biological self-organization are well exemplified by the state of understanding of *M. xanthus* development. Similar to eukaryotic systems, *M. xanthus* cells have diverse signaling cascades and behavior cues that could contribute to self-organization. The genome is enriched for signaling systems and secondary metabolites. *M. xanthus* has access to 262 two-component signaling systems, 52 of which are up-regulated during development (Whitworth, 2015), and 21 chemoreceptors, of which 13 create altered developmental phenotypes when deleted (Moine et al., 2014). *M. xanthus* also dedicates 8%

of its genome to secondary metabolite production (Bode and Müller, 2006), suggesting a large capability for the production of unique compounds, most of which are unknown or not well characterized. Like eukaryotic self-organizing systems, *M. xanthus* development defies explanation using typical experimental techniques. For example, (Shi et al., 2008) chose 23 of the 53 up-regulated two-component signaling systems to examine. Despite their up-regulated state, suggesting a role in development, 16 (70%) were not required for cells to generate aggregation patterns visually indistinct from wild type. Furthermore, some secondary metabolites have complicated phenotypes when modified. For example, the major pigment responsible for the yellow color of *M. xanthus*, DKxanthene, affects the speed of aggregation as well as spore maturation (Meiser et al., 2006). The large capability for signaling available to *M. xanthus*, and the lack of clear experimental evidence to challenge hypotheses gives rise to many hypothetical coordinating mechanisms.

Motility

M. xanthus cells lack flagella and are unable to swim in liquid. They are capable of moving along their long axis on solid or semi-solid surfaces with an average speed of about 5 μm/min (Jelsbak and Søgaard-Andersen, 1999, 2002). Motility direction reverses periodically, leading to a back and forth movement pattern. Individual cells on agar pads reverse direction approximately once every 7 minutes (Kearns and Shimkets, 1998; McCleary et al., 1990). Propulsion is accomplished by two independent motility systems. Adventurous (A) motility produces force via the translocation of focal adhesion points along the long axis of the cell (Faure et al., 2016). Direct attachment to the surface allows A-motility to propel the cell forward in the absence of other cells. Social (S) motility is driven by the extension and retraction of type IV pili from the leading pole of the cell. The system is similar to the type IV pili utilized for twitching motility in *Pseudomonas aeruginosa* and *Neisseria* species (Burrows, 2012; Miller et al., 2013). Extended pili attach to extracellular polysaccharides

excreted by neighboring cells (Li et al., 2003). This attachment provides an anchor point by which the cells pull themselves forward. A-motility provides a mechanism for individual cell movement and is most active on rigid surfaces. S-motility requires the presence of other cells, but allows the cells to move efficiently on soft surfaces (Youderian et al., 2003).

Chemotaxis

Biased motility along a chemotactic gradient is performed by modulation of the tumble (for swimming bacteria) or reversal (for gliding bacteria) frequency. Cells moving up an attractant gradient change motility direction less often than when moving down the gradient. Modulating the direction-change frequency causes cells to randomly explore all directions of the gradient on short timescales, but an overall biased movement up the gradient on longer timescales. The opposite behavior is observed in the case of a repellent gradient.

The archetypical chemosensory system involves the Che proteins in *Escherichia coli* and *Salmonella enterica*. Sensing is performed by methyl-accepting chemotaxis proteins (MCP) (Nishiyama et al., 1999). The MCP is a single protein consisting of an extracellular receptor connected via a transmembrane region to a cytoplasmic α-helical coiled-coil domain (Kim et al., 1999; Yeh et al., 1996). CheA, a histidine protein kinase, reversibly binds to the cytoplasmic end of the MCP with the help of the CheW adapter protein (Bourret et al., 1993; Levit et al., 1999; Morrison and Parkinson, 1997). When a ligand binds to the receptor domain of the MCP, conformational changes in the MCP propagate the signal to the bound CheA, activating CheA self-phosphorylation (Levit et al., 1999). The phosphate group is then transferred to the cytoplasmic CheY response regulator (Stewart et al., 2000; Welch et al., 1998). CheY is responsible for transmission of the signal to the motility machinery by diffusion through the cytoplasm. Phosphorylated CheY directly binds to the flagellar motor to increase the probability of a tumble event (Sagi et al., 2003).

Attractant or repellent gradients are sensed by calculating the time derivative of the local chemical concentration. The presence of a time-based chemosensory system is best exemplified by excitation-adaptation experiments (Alon et al., 1999; Macnab and Koshland, 1972). When swimming cells are presented with an abrupt, but uniform, increase in the concentration of an attractant, the cells exhibit an immediate suppression of tumbles, as if they were moving up an attractant gradient. Despite a constant concentration of attractant, tumbling frequency reverts back to the pre-stimulus level within approximately 14 minutes. The adaptation to the presence of the attractant is mediated by methylation of the MCP (Weis and Koshland, 1988). In *E. coli*, ligand binding to the MCP suppresses CheA self-phosphorylation (Levit et al., 1999), leading to a decrease in phosphorylated CheY; thereby reducing the probability of a tumble. Adaptation of the tumble frequency back to basal levels is caused by methylation of the MCP's cytoplasmic coiled-coil helix domain by CheR, a methyltransferase (Springer and Koshland, 1977; Weis and Koshland, 1988). Methylation causes conformational changes in the MCP that reset CheY self-phosphorylation to a basal level in the presence of the higher ligand concentration, returning the cell to a default tumble frequency.

Chemotaxis is a balance between ligand binding and MCP methylation. In *E. coli*, when cells are moving up an attractant gradient, ligand binding to the MCP causes suppression of reversals through reduced CheA/CheY phosphorylation. Increased ligand binding causes CheR to methylate the MCP, resetting CheY activity. The reset causes any new decrease in local concentration from the current level to stimulate tumbles (through increased CheY phosphorylation) or suppress tumbles (through suppression of CheY phosphorylation) in the presence of even higher attractant concentrations. Adaptation when moving from higher to lower attractant concentrations occurs in reverse. Lower MCP receptor occupancy causes demethylation via the methylesterase CheB (Anand et al., 1998; Kehry and Dahlquist, 1982). Response to repellents is the reverse process to that of

attractants. Binding of repellents to an MCP causes increased CheA/CheY phosphorylation, leading to an increased probability of tumbling. Adaptation to a repellent is performed by removal of methyl groups from the MCP by CheB (Toews et al., 1979), resetting CheA activity. This adaptation mechanism is sensitive enough for *E. coli* cells to detect a concentration change of only a few molecules in the presence of background concentrations spanning about five orders of magnitude (Sourjik and Berg, 2002).

The role of chemotaxis in *M. xanthus* development is unclear. In early experiments (McVittie and Zahler 1962; Lev 1954), developing myxobacterial cells were covered with a thin layer of agar or dialysis membrane followed by another layer of myxobacteria. The top layer of myxobacteria developed aggregates at approximately the same locations as the bottom layer. Glass beads in place of the underlying layer of developing cells did not affect the locations of the upper layer aggregates, ruling out physical cues and suggesting that positional information diffused through the permeable barrier. Dworkin and Eide (1983) argued that the slow speed of *M. xanthus* relative to the speed of small diffusing molecules would cause the gradient to equalize faster than the cell could move up it. They also provided experimental evidence that *M. xanthus* does not respond to moderate concentration gradients of cyclic-AMP – a diffusible molecule key to D. discoideum development, to amino-acids – a major food source for myxobacteria, or to soluble lysates of vegetative or developing *M. xanthus* cells.

Of the 21 chemoreceptors in the *M. xanthus* genome, FrzCD appears to play a central role in reversal control. The *frzCD* gene is flanked in the genome by a full complement of Che homologues, together known as the Frz system (Blackhart and Zusman, 1985). Cells with a defective Frz system rarely reverse (McBride et al., 1992), do not develop (Zusman, 1982), and are unable to correctly respond to chemoattractants (Kearns and Shimkets, 1998). Like the MCPs in *E. coli*, FrzCD is methylated in the presence of chemoattractants, such as food sources (Shi et al., 1993), and

demethylated in the presence of repellents, such as small chain alcohols (McBride et al., 1992). It is unclear how FrzCD responds to such a broad range of signals. In *E. coli*, each chemoreceptor contains a sensing domain specific to a molecule or molecule family (Falke and Hazelbauer, 2001). Unlike *E. coli* homologues, FrzCD is a soluble cytoplasmic protein (McBride et al., 1992), truncated just before the transmembrane region (McBride et al., 1989). The lack of an external sensing domain suggests signals may be transduced through protein-protein interactions with one or more unknown partners, or directly through an uncharacterized sensing domain.

More recently, *M. xanthus* was shown to chemotactically respond to phosphatidylethanolamine (PE) and diacylglycerol (DAG) lipids (Kearns and Shimkets, 1998). These large, low-solubility molecules generate gradients on agar surfaces (Kearns and Shimkets, 1998). The slow diffusion rate of PE and DAGs generates sharp, slow moving, concentration gradients that may be more suitable for the slow motility of *M. xanthus*. Sensing is specific to the chemical structure, suggesting the response evolved to sense specific cues. Of the known active PE molecules, 16:1ω5c/16:1ω5c (hereafter referred to as 16:1) occurs naturally within the *M. xanthus* membrane, but is rare in other cultured soil bacteria (Curtis et al., 2006). Interestingly, *M. xanthus* is only sensitive to 16:1 as a chemoattractant under developmental (starvation) conditions. However, when all straight chain fatty acids synthesis is inhibited, leading to only 1/10 the normal concentration of 16:1 PE in the cells, growth and development is not affected (Bode et al., 2006). This result suggests there is no required role for 16:1 in development.

C-signaling

An important gene in the study of *M. xanthus* self-organization is *csgA*. Cells lacking the *csgA* gene are unable to develop on their own, with developmental gene expression stopping at or before that of non-motile cells (Kim and Kaiser, 1990). When mixed with wild-type (WT) cells, *csgA* cells sporulate at approximately the same rate as WT (Kim and Kaiser, 1990), suggesting *csgA* cell

development is rescued by the presence of WT cells. Purified extracts of the CsgA protein added back to *csgA* cells also rescues development and is sufficient to induce sporulation in non-motile cells (Kim and Kaiser, 1990). Together, these results led to the hypothesis that CsgA passes between cells, or acts as a ligand on the surface of cell that activates a cognate receptor on neighboring cells during cell-to-cell contact (Jelsbak and Søgaard-Andersen, 2000). In conflict with this hypothesis, CsgA is found associated with the inner membrane (Simunovic et al., 2003), suggesting that CsgA is shielded from contact with neighboring cells by the outer membrane. The contact-dependent CsgA rescue hypothesis also fails to consider enzymatic activity. CsgA is homologous to members of the short-chain alcohol dehydrogenase (SCAD) family of proteins (Baker, 1994). Overexpression of another SCAD, *socA*, in *csgA* cells is able to rescue development (Lee and Shimkets, 1996), and the enzymatic activity of CsgA is required for development (Lee and Shimkets, 1996).

CsgA oxidizes the 2'-OH group of the glycerol backbone of cardiolipin (CL) and phosphatidylglycerol (PG) (Boynton and Shimkets, 2015). The resulting enzymatic product is unstable, quickly breaking down to diacylglycerol (DAG), dihydroxyacetone, and orthophosphate derivatives. As expected, SocA is also able to catalyze the same reactions, but with an approximately 10-fold reduction in efficiency. These results suggest *csgA* cells are blocked in lipid metabolism of CL and PG species found in the inner membrane of the cell (Orndorff and Dworkin, 1980). In agreement, partial glycerides – consisting of monoacylglycerols and DAG – extracted from developing WT cells restore aggregation and sporulation when added to *csgA* cells (Boynton and Shimkets, 2015). Rescue is specific to partial glyceride extracts from developing cells, extracts from vegetative WT or *csgA* cells do not rescue *csgA* development (Boynton and Shimkets, 2015).

Modeling *M. xanthus* self-organization

The challenges of modeling self-organizing systems are well manifested in *M. xanthus* research. A diverse and often conflicting range of computational models exist for *M. xanthus* development.

With many possible mechanisms for coordination of cell movement and a lack of experimental results to constrain hypotheses, the length-to-width ratio of cells (Starruß et al., 2007), cell alignment (Sliusarenko et al., 2007; Starruß et al., 2007), active turning of cell direction (Hendrata et al., 2011), density-dependent speed reduction (Sliusarenko et al., 2007), physical jamming (Holmes et al., 2010; Igoshin et al., 2004; Sozinova et al., 2005), and streaming (Holmes et al., 2010; Sozinova et al., 2005) have all been hypothesized to drive *M. xanthus* development. Each of these examples is backed with a mathematical model that qualitatively replicates aggregation. That is to say, the models generate areas of high cell density that are visually similar in shape and size to *M. xanthus* aggregates. Quantitative analyses comparing simulation to experimental results are rarely performed. When a quantitative analysis was performed for the density-dependent speed reduction model in (Sliusarenko et al., 2007), simulations failed to capture the correct aggregation rate and aggregate count (Zhang et al., 2011). These past experiences reveal the need for an experimental and modeling framework that is constrained by quantitative experimental results. The most desirable approach would implicitly model cell systems that are unknown or cannot be feasibly quantified.

Proposal of a Data Driven Experimental and Modeling Paradigm

Mathematical and computational models are often used to identify and summarize the important aspects of complicated biological experiments. Typically, these models are developed to test hypotheses that are generated by post experimental analysis. Under this paradigm, experimentalists must glean insight directly from the experimental results prior to design, parameterization, and validation of the models. The paradigm fails when traditional experimental techniques are unable to provide clear results due to: (1) noise and non-linear dependencies between cues and behaviors obscuring data trends, (2) results that are ambiguous, routinely resulting in no change or abolishment of the self-organization, and (3) a lack of tools to identify, quantify, and manipulate

the key variables in the system. The challenges of understanding biological self-organization intersect the challenges of the typical modeling paradigm due to the large number of possible hypotheses, conflicting experimental results, or the sheer number of free parameters requiring experimental quantification. These challenges help explain the lack of biological self-organization modeling approaches that are experimentally constrained and quantitative.

This dissertation defines an experimental framework that utilizes *in vivo* imaging of individual cell behavior and the surrounding environment to both parameterize and constrain models of *M. xanthus* development to experimental data. Instead of defining simulations based on experimental conclusions and hypotheses, the framework defines the simulation based on empirical observations from *in vivo* imaging, such as basic cell movement, then drives the simulations with data extracted from quantification of the empirical observations. In this way, simulations are driven by quantitative experimental results while still implicitly modeling cell-systems that are unknown or cannot be feasibly quantified. With the model parameterized early in the experimental process, hypotheses on how self-organization occurs and what biological features are necessary can be tested in the simulation space. The outline of the dissertation is as follows: Chapter 2 defines the experimental framework and applies it to *M. xanthus* development to identify the cell behaviors and cues that drive aggregation. Chapter 3 then concludes by discussing the known signaling and sensing systems within *M. xanthus* in light of the identified cell behaviors and cues used during development.

References

Abercrombie, M., and Heaysman, J.E.M. (1953). Observations on the social behaviour of cells in tissue culture. Exp. Cell Res. *5*, 111–131.

Alon, U., Surette, M.G., Barkai, N., and Leibler, S. (1999). Robustness in bacterial chemotaxis. Nature *397*, 168–171.

Anand, G.S., Goudreau, P.N., and Stock, A.M. (1998). Activation of methylesterase CheB: Evidence of a dual role for the regulatory domain. Biochemistry (Mosc.) *37*, 14038–14047.

Asai, R., Taguchi, E., Kume, Y., Saito, M., and Kondo, S. (1999). Zebrafish *leopard* gene as a component of the putative reaction-diffusion system. Mech. Dev. *89*, 87–92.

Baker, M.E. (1994). *Myxococcus xanthus* C-factor, a morphogenetic paracrine signal, is similar to *Escherichia coli* 3-oxoacyl-[acyl-carrier-protein] reductase and human 17 beta-hydroxysteroid dehydrogenase. Biochem. J. *301*, 311–312.

Blackhart, B.D., and Zusman, D.R. (1985). Cloning and complementation analysis of the "Frizzy" genes of *Myxococcus xanthus*. Mol. Gen. Genet. MGG *198*, 243–254.

Bode, H. (2011). Axis Formation in Hydra. Annu. Rev. Genet. *45*, 105–117.

Bode, H.B., and Müller, R. (2006). Analysis of myxobacterial secondary metabolism goes molecular. J. Ind. Microbiol. Biotechnol. *33*, 577–588.

Bode, H.B., Ring, M.W., Kaiser, D., David, A.C., Kroppenstedt, R.M., and Schwar, G. (2006). Straight-chain fatty acids are dispensable in the myxobacterium *Myxococcus xanthus* for vegetative growth and fruiting body formation. J. Bacteriol. *188*, 5632–5634.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. Proc. Natl. Acad. Sci. *99*, 7280–7287.

Bourret, R.B., Davagnino, J., and Simon, M.I. (1993). The carboxy-terminal portion of the CheA kinase mediates regulation of autophosphorylation by transducer and CheW. J. Bacteriol. *175*, 2097–2101.

Boynton, T.O., and Shimkets, L.J. (2015). *Myxococcus* Csga, *drosophila* Sniffer, and Human Hsd10 are cardiolipin phospholipases. Genes Dev. *29*, 1903–1914.

Burrows, L.L. (2012). *Pseudomonas aeruginosa* twitching motility: Type IV pili in action. Annu. Rev. Microbiol. *66*, 493–520.

Callejo, A., Torroja, C., Quijada, L., and Guerrero, I. (2006). Hedgehog lipid modifications are required for Hedgehog stabilization in the extracellular matrix. Development *133*, 471–483.

Campbell, I.D., and Humphries, M.J. (2011). Integrin structure, activation, and interactions. Cold Spring Harb. Perspect. Biol. *3*.

Carey, S.P., Kraning-Rush, C.M., Williams, R.M., and Reinhart-King, C.A. (2012). Biophysical control of invasive tumor cell behavior by extracellular matrix microarchitecture. Biomaterials *33*, 4157–4165.

Charras, G., and Sahai, E. (2014). Physical influences of the extracellular environment on cell migration. Nat. Rev. Mol. Cell Biol. *15*, 813–824.

Chi, C., and Trinkaus-Randall, V. (2013). New insights in wound response and repair of epithelium. J. Cell. Physiol. *228*, 925–929.

Condeelis, J., and Segall, J.E. (2003). Intravital imaging of cell movement in tumours. Nat. Rev. Cancer *3*, 921–930.

Cooke, J. (1975). Control of somite number during morphogenesis of a vertebrate, Xenopus laevis. Nature *254*, 196–199.

Cooke, J. (1981). Scale of body pattern adjusts to available cell number in amphibian embryos. Nature *290*, 775–778.

Cooke, J., and Zeeman, E.C. (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. J. Theor. Biol. *58*, 455–476.

Cortese, B., Palamà, I.E., D'Amone, S., and Gigli, G. (2014). Influence of electrotaxis on cell behaviour. Integr. Biol. *6*, 817–830.

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

Curtis, P.D., Geyer, R., White, D.C., and Shimkets, L.J. (2006). Novel lipids in *Myxococcus xanthus* and their role in chemotaxis. Environ. Microbiol. *8*, 1935–1949.

Darnton, N.C., Turner, L., Rojevsky, S., and Berg, H.C. (2010). Dynamics of bacterial swarming. Biophys. J. *98*, 2082–2090.

DiMilla, P.A., Barbee, K., and Lauffenburger, D.A. (1991). Mathematical model for the effects of adhesion and mechanics on cell migration speed. Biophys. J. *60*, 15–37.

Discher, D.E., Janmey, P., and Wang, Y. (2005). Tissue cells feel and respond to the stiffness of their substrate. Science *310*, 1139–1143.

Dormann, D., and Weijer, C.J. (2006). Chemotactic cell movement during *Dictyostelium* development and gastrulation. Curr. Opin. Genet. Dev. *16*, 367–373.

Doyle, A.D., Petrie, R.J., Kutys, M.L., and Yamada, K.M. (2013). Dimensions in cell migration. Curr. Opin. Cell Biol. *25*, 642–649.

Dworkin, M., and Eide, D. (1983). *Myxococcus xanthus* does not respond chemotactically to moderate concentration gradients. J. Bacteriol. *154*, 437–442.

Eaton, S. (2008). Multiple roles for lipids in the Hedgehog signalling pathway. Nat. Rev. Mol. Cell Biol. *9*, 437–445.

Falke, J.J., and Hazelbauer, G.L. (2001). Transmembrane signaling in bacterial chemoreceptors. Trends Biochem. Sci. *26*, 257–265.

Faure, L.M., Fiche, J.-B., Espinosa, L., Ducret, A., Anantharaman, V., Luciano, J., Lhospice, S., Islam, S.T., Tréguier, J., Sotes, M., et al. (2016). The mechanism of force transmission at bacterial focal adhesion complexes. Nature *539*, 530–535.

French, V., Bryant, P.J., and Bryant, S.V. (1976). Pattern regulation in epimorphic fields. Science *193*, 969–981.

Friedl, P., and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. Nat. Rev. Mol. Cell Biol. *10*, 445–457.

Friedl, P., and Mayor, R. (2017). Tuning collective cell migration by cell–cell junction regulation. Cold Spring Harb. Perspect. Biol. *9*, a029199.

Friedl, P., and Wolf, K. (2010). Plasticity of cell migration: A multiscale tuning model. J. Cell Biol. *188*, 11–19.

Gierer, A., and Meinhardt, H. (1972). A theory of biological pattern formation. Kybernetik *12*, 30–39.

Green, J.B.A., and Sharpe, J. (2015). Positional information and reaction-diffusion: Two big ideas in developmental biology combine. Development *142*, 1203–1211.

Haeger, A., Wolf, K., Zegers, M.M., and Friedl, P. (2015). Collective cell migration: Guidance principles and hierarchies. Trends Cell Biol. *25*, 556–566.

Hakkinen, K.M., Harunaga, J.S., Doyle, A.D., and Yamada, K.M. (2010). Direct comparisons of the morphology, migration, cell adhesions, and actin cytoskeleton of fibroblasts in four different three-dimensional extracellular matrices. Tissue Eng. Part A *17*, 713–724.

Han, S., Yan, J.-J., Shin, Y., Jeon, J.J., Won, J., Jeong, H.E., Kamm, R.D., Kim, Y.-J., and Chung, S. (2012). A versatile assay for monitoring *in vivo*-like transendothelial migration of neutrophils. Lab. Chip *12*, 3861–3865.

Harley, B.A.C., Kim, H.-D., Zaman, M.H., Yannas, I.V., Lauffenburger, D.A., and Gibson, L.J. (2008). Microarchitecture of three-dimensional scaffolds influences cell migration behavior via junction interactions. Biophys. J. *95*, 4013–4024.

Hendrata, M., Yang, Z., Lux, R., and Shi, W. (2011). Experimentally guided computational model discovers important elements for social behavior in myxobacteria. PLoS ONE *6*, e22169.

Hoffmann, J.C., and West, J.L. (2013). Three-dimensional photolithographic micropatterning: A novel tool to probe the complexities of cell migration. Integr. Biol. Quant. Biosci. Nano Macro *5*, 817–827.

Holmes, A.B., Kalvala, S., and Whitworth, D.E. (2010). Spatial simulations of myxobacterial development. PLoS Comput. Biol. *6*, e1000686.

Hynes, R.O. (2002). Integrins: Bidirectional, allosteric signaling machines. Cell *110*, 673–687.

Igoshin, O.A., Welch, R., Kaiser, D., and Oster, G. (2004). Waves and aggregation patterns in myxobacteria. Proc. Natl. Acad. Sci. *101*, 4256–4261.

Jelsbak, L., and Søgaard-Andersen, L. (1999). The cell surface-associated intercellular C-signal induces behavioral changes in individual *Myxococcus xanthus* cells during fruiting body morphogenesis. Proc. Natl. Acad. Sci. *96*, 5031–5036.

Jelsbak, L., and Søgaard-Andersen, L. (2000). Pattern formation: Fruiting body morphogenesis in *Myxococcus xanthus*. Curr. Opin. Microbiol. *3*, 637–642.

Jelsbak, L., and Søgaard-Andersen, L. (2002). Pattern formation by a cell surface-associated morphogen in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. *99*, 2032–2037.

Jin, M., Andricioaei, I., and Springer, T.A. (2004). Conversion between three conformational states of integrin I domains with a C-terminal pull spring studied with molecular dynamics. Structure *12*, 2137–2147.

Kearns, D.B., and Shimkets, L.J. (1998). Chemotaxis in a gliding bacterium. Proc. Natl. Acad. Sci. *95*, 11957–11962.

Kehry, M.R., and Dahlquist, F.W. (1982). Adaptation in bacterial chemotaxis: CheB-dependent modification permits additional methylations of sensory transducer proteins. Cell *29*, 761–772.

Keller, R. (2005). Cell migration during gastrulation. Curr. Opin. Cell Biol. *17*, 533–541.

Keren, K., Pincus, Z., Allen, G.M., Barnhart, E.L., Marriott, G., Mogilner, A., and Theriot, J.A. (2008). Mechanism of shape determination in motile cells. Nature *453*, 475–480.

Kim, S.K., and Kaiser, D. (1990). Cell motility is required for the transmission of C-factor, an intercellular signal that coordinates fruiting body morphogenesis of *Myxococcus xanthus*. Genes Dev. *4*, 896–905.

Kim, K.K., Yokota, H., and Kim, S.-H. (1999). Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. Nature *400*, 787–792.

Lai, K., Robertson, M.J., and Schaffer, D.V. (2004). The Sonic hedgehog signaling system as a bistable genetic switch. Biophys. J. *86*, 2748–2757.

Lämmermann, T., and Sixt, M. (2009). Mechanical modes of 'amoeboid' cell migration. Curr. Opin. Cell Biol. *21*, 636–644.

Lauffenburger, D.A., and Horwitz, A.F. (1996). Cell migration: A physically integrated molecular process. Cell *84*, 359–369.

Lee, K., and Shimkets, L.J. (1996). Suppression of a signaling defect during *Myxococcus xanthus* development. J. Bacteriol. *178*, 977–984.

Levit, M.N., Liu, Y., and Stock, J.B. (1999). Mechanism of CheA protein kinase activation in receptor signaling complexes. Biochemistry (Mosc.) *38*, 6651–6658.

Li, Y., Sun, H., Ma, X., Lu, A., Lux, R., Zusman, D., and Shi, W. (2003). Extracellular polysaccharides mediate pilus retraction during social motility of *Myxococcus xanthus*. Proc. Natl. Acad. Sci. U. S. A. *100*, 5443–5448.

Lo, C.M., Wang, H.B., Dembo, M., and Wang, Y.L. (2000). Cell movement is guided by the rigidity of the substrate. Biophys. J. *79*, 144–152.

Loomis, W.F. (2014). Cell signaling during development of *Dictyostelium*. Dev. Biol. *391*, 1–16.

Macnab, R.M., and Koshland, D.E. (1972). The gradient-sensing mechanism in bacterial chemotaxis. Proc. Natl. Acad. Sci. U. S. A. *69*, 2509–2512.

Marongiu, F., Doratiotto, S., Sini, M., Serra, M.P., and Laconi, E. (2012). Cancer as a disease of tissue pattern formation. Prog. Histochem. Cytochem. *47*, 175–207.

Mason, B.N., Starchenko, A., Williams, R.M., Bonassar, L.J., and Reinhart-King, C.A. (2013). Tuning three-dimensional collagen matrix stiffness independently of collagen concentration modulates endothelial cell behavior. Acta Biomater. *9*, 4635–4644.

Mayor, R., and Carmona-Fontaine, C. (2010). Keeping in touch with contact inhibition of locomotion. Trends Cell Biol. *20*, 319–328.

Mayor, R., and Etienne-Manneville, S. (2016). The front and rear of collective cell migration. Nat. Rev. Mol. Cell Biol. *17*, 97–109.

McBride, M.J., Weinberg, R.A., and Zusman, D.R. (1989). "Frizzy" aggregation genes of the gliding bacterium *Myxococcus xanthus* show sequence similarities to the chemotaxis genes of enteric bacteria. Proc. Natl. Acad. Sci. *86*, 424–428.

McBride, M.J., Köhler, T., and Zusman, D.R. (1992). Methylation of FrzCD, a methyl-accepting taxis protein of *Myxococcus xanthus*, is correlated with factors affecting cell behavior. J. Bacteriol. *174*, 4246–4257.

McCarthy, J., Palm, S., and Furcht, L. (1983). Migration by haptotaxis of a Schwann cell tumor line to the basement membrane glycoprotein laminin. J. Cell Biol. *97*, 772–777.

McCleary, W.R., McBride, M.J., and Zusman, D.R. (1990). Developmental sensory transduction in *Myxococcus xanthus* involves methylation and demethylation of FrzCD. J. Bacteriol. *172*, 4877–4887.

Meinhardt, H. (2009). Models for the generation and interpretation of gradients. Cold Spring Harb. Perspect. Biol. *1*.

Meinhardt, H., and Gierer, A. (1974). Applications of a theory of biological pattern formation based on lateral inhibition. J. Cell Sci. *15*, 321–346.

Meinhardt, H., and Gierer, A. (2000). Pattern formation by local self-activation and lateral inhibition. Bioessays *22*, 753–760.

Meiser, P., Bode, H.B., and Müller, R. (2006). The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. Proc. Natl. Acad. Sci. *103*, 19128–19133.

Miller, F., Lécuyer, H., Join-Lambert, O., Bourdoulous, S., Marullo, S., Nassif, X., and Coureuil, M. (2013). *Neisseria meningitidis* colonization of the brain endothelium and cerebrospinal fluid invasion. Cell. Microbiol. *15*, 512–519.

Mlodzik, M., and Gehring, W.J. (1987). Expression of the caudal gene in the germ line of *Drosophila*: Formation of an RNA and protein gradient during early embryogenesis. Cell *48*, 465–478.

Mlodzik, M., Fjose, A., and Gehring, W.J. (1985). Isolation of caudal, a *Drosophila* homeo box-containing gene with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage. EMBO J. *4*, 2961–2969.

Moine, A., Agrebi, R., Espinosa, L., Kirby, J.R., Zusman, D.R., Mignot, T., and Mauriello, E.M.F. (2014). Functional organization of a multimodular bacterial chemosensory apparatus. PLOS Genet. *10*, e1004164.

Morrison, T.B., and Parkinson, J.S. (1997). A fragment liberated from the *Escherichia coli* CheA kinase that blocks stimulatory, but not inhibitory, chemoreceptor signaling. J. Bacteriol. *179*, 5543–5550.

Nishiyama, S., Maruyama, I.N., Homma, M., and Kawagishi, I. (1999). Inversion of thermosensing property of the bacterial receptor Tar by mutations in the second transmembrane region. J. Mol. Biol. *286*, 1275–1284.

Oates, A.C., Morelli, L.G., and Ares, S. (2012). Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock. Development *139*, 625–639.

Oppenheim, J.J., and Yang, D. (2005). Alarmins: Chemotactic activators of immune responses. Curr. Opin. Immunol. *17*, 359–365.

Orndorff, P.E., and Dworkin, M. (1980). Separation and properties of the cytoplasmic and outer membranes of vegetative cells of *Myxococcus xanthus*. J. Bacteriol. *141*, 914–927.

Palecek, S., Loftus, J., Ginsberg, M., Lauffenburger, D., and Horwitz, A. (1997). Integrin-ligand binding properties govern cell migration speed through cell-substratum adhesiveness. Nature *385*, 537–540.

Paszek, M.J., Boettiger, D., Weaver, V.M., and Hammer, D.A. (2009). Integrin clustering is driven by mechanical resistance from the glycocalyx and the substrate. PLOS Comput. Biol. *5*, e1000604.

Paul, C.D., Mistriotis, P., and Konstantopoulos, K. (2017). Cancer cell motility: Lessons from migration in confined spaces. Nat. Rev. Cancer *17*, 131–140.

Reeves, G.T., Muratov, C.B., Schüpbach, T., and Shvartsman, S.Y. (2006). Quantitative models of developmental pattern formation. Dev. Cell *11*, 289–300.

Restrepo, S., Zartman, J.J., and Basler, K. (2014). Coordination of patterning and growth by the morphogen DPP. Curr. Biol. *24*, R245–R255.

Rhinn, M., and Dollé, P. (2012). Retinoic acid signalling during development. Development *139*, 843–858.

Rogers, K.W., and Schier, A.F. (2011). Morphogen gradients: From generation to interpretation. Annu. Rev. Cell Dev. Biol. *27*, 377–407.

Sagi, Y., Khan, S., and Eisenbach, M. (2003). Binding of the chemotaxis response regulator CheY to the isolated, intact switch complex of the bacterial flagellar motor: Lack of cooperativity. J. Biol. Chem. *278*, 25867–25871.

Saha, K., and Schaffer, D.V. (2006). Signal dynamics in Sonic hedgehog tissue patterning. Development *133*, 889–900.

Shi, W., Köhler, T., and Zusman, D.R. (1993). Chemotaxis plays a role in the social behaviour of *Myxococcus xanthus*. Mol. Microbiol. *9*, 601–611.

Shi, X., Wegener-Feldbrügge, S., Huntley, S., Hamann, N., Hedderich, R., and Søgaard-Andersen, L. (2008). Bioinformatics and experimental analysis of proteins of two-component systems in *Myxococcus xanthus*. J. Bacteriol. *190*, 613–624.

Shimkets, L.J. (1999). Intercellular signaling during fruiting-body development of *Myxococcus xanthus*. Annu. Rev. Microbiol. 525–549.

Shoulders, M.D., and Raines, R.T. (2009). Collagen structure and stability. Annu. Rev. Biochem. *78*, 929–958.

Simunovic, V., Gherardini, F.C., and Shimkets, L.J. (2003). Membrane localization of motility, signaling, and polyketide synthetase proteins in *Myxococcus xanthus*. J. Bacteriol. *185*, 5066–5075.

Sliusarenko, O., Zusman, D.R., and Oster, G. (2007). Aggregation during fruiting body formation in *Myxococcus xanthus* is driven by reducing cell movement. J. Bacteriol. *189*, 611–619.

Sourjik, V., and Berg, H.C. (2002). Receptor sensitivity in bacterial chemotaxis. Proc. Natl. Acad. Sci. *99*, 123–127.

Sozinova, O., Jiang, Y., Kaiser, D., and Alber, M. (2005). A three-dimensional model of myxobacterial aggregation by contact-mediated interactions. Proc. Natl. Acad. Sci. U. S. A. *102*, 11308–11312.

Springer, W.R., and Koshland, D.E. (1977). Identification of a protein methyltransferase as the *cheR* gene product in the bacterial sensing system. Proc. Natl. Acad. Sci. U. S. A. *74*, 533–537.

Starruß, J., Bley, T., Søgaard-Andersen, L., and Deutsch, A. (2007). A new mechanism for collective migration in *Myxococcus xanthus*. J. Stat. Phys. *128*, 269–286.

Stewart, R.C., Jahreis, K., and Parkinson, J.S. (2000). Rapid phosphotransfer to CheY from a CheA protein lacking the CheY-binding domain. Biochemistry (Mosc.) *39*, 13157–13165.

Stramer, B. m., Dunn, G. a., Davis, J. r., and Mayor, R. (2013). Rediscovering contact inhibition in the embryo. J. Microsc. *251*, 206–211.

Stumpf, H.F. (1966). Mechanism by which cells estimate their location within the body. Nature *212*, 430–431.

Swaney, K.F., Huang, C.-H., and Devreotes, P.N. (2010). Eukaryotic chemotaxis: A network of signaling pathways controls motility, directional sensing, and polarity. Annu. Rev. Biophys. *39*, 265–289.

Takatori, S.C., and Brady, J.F. (2015). Towards a "thermodynamics" of active matter. Phys. Rev. E *91*.

Teddy, J.M., and Kulesa, P.M. (2004). In vivo evidence for short- and long-range cell communication in cranial neural crest cells. Development *131*, 6141–6151.

Theveneau, E., and Mayor, R. (2011). Collective cell migration of the cephalic neural crest: The art of integrating information. Genesis *49*, 164–176.

Theveneau, E., and Mayor, R. (2012). Neural crest delamination and migration: From epithelium-to-mesenchyme transition to collective cell migration. Dev. Biol. *366*, 34–54.

Toews, M.L., Goy, M.F., Springer, M.S., and Adler, J. (1979). Attractants and repellents control demethylation of methylated chemotaxis proteins in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *76*, 5544–5548.

Trepat, X., Wasserman, M.R., Angelini, T.E., Millet, E., Weitz, D.A., Butler, J.P., and Fredberg, J.J. (2009). Physical forces during collective cell migration. Nat. Phys. *5*, 426–430.

Turing, A.M. (1952). The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. *237*, 153–197.

Vedula, S.R.K., Leong, M.C., Lai, T.L., Hersen, P., Kabla, A.J., Lim, C.T., and Ladoux, B. (2012). Emerging modes of collective cell migration induced by geometrical constraints. Proc. Natl. Acad. Sci. *109*, 12974–12979.

Vedula, S.R.K., Ravasio, A., Lim, C.T., and Ladoux, B. (2013). Collective cell migration: A mechanistic perspective. Physiology *28*, 370–379.

Watanabe, M., and Kondo, S. (2015). Is pigment patterning in fish skin determined by the Turing mechanism? Trends Genet. *31*, 88–96.

Weis, R.M., and Koshland, D.E. (1988). Reversible receptor methylation is essential for normal chemotaxis of *Escherichia coli* in gradients of aspartic acid. Proc. Natl. Acad. Sci. U. S. A. *85*, 83–87.

Welch, M., Chinardet, N., Mourey, L., Birck, C., and Samama, J.P. (1998). Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. Nat. Struct. Biol. *5*, 25–29.

Whitworth, D.E. (2015). Genome-wide analysis of myxobacterial two-component systems: genome relatedness and evolutionary changes. BMC Genomics *16*.

Wolf, K., Mazo, I., Leung, H., Engelke, K., Andrian, U.H. von, Deryugina, E.I., Strongin, A.Y., Bröcker, E.-B., and Friedl, P. (2003). Compensation mechanism in tumor cell migration: Mesenchymal–amoeboid transition after blocking of pericellular proteolysis. J. Cell Biol. *160*, 267–277.

Wolf, K., Te Lindert, M., Krause, M., Alexander, S., Te Riet, J., Willis, A.L., Hoffman, R.M., Figdor, C.G., Weiss, S.J., and Friedl, P. (2013). Physical limits of cell migration: Control by ECM space and nuclear deformation and tuning by proteolysis and traction force. J. Cell Biol. *201*, 1069–1084.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. *25*, 1–47.

Wolpert, L. (2011). Positional information and patterning revisited. J. Theor. Biol. *269*, 359–365.

Wolpert, L. (2016). Positional Information and Pattern Formation. Curr. Top. Dev. Biol. *117*, 597–608.

Yamaguchi, M., Yoshimoto, E., and Kondo, S. (2007). Pattern regulation in the stripe of zebrafish suggests an underlying dynamic and autonomous mechanism. Proc. Natl. Acad. Sci. *104*, 4790–4793.

Yan, D., Wu, Y., Feng, Y., Lin, S.-C., and Lin, X. (2009). The core protein of glypican dally-like determines its biphasic activity in Wingless morphogen signaling. Dev. Cell *17*, 470–481.

Yang, Z., and Higgs, P. (2014). Myxobacteria (Norfolk, UK: Caister Academic Press).

Yeh, J.I., Biemann, H.P., Privé, G.G., Pandit, J., Koshland, D.E., and Kim, S.H. (1996). High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. J. Mol. Biol. *262*, 186–201.

Youderian, P., Burke, N., White, D.J., and Hartzell, P.L. (2003). Identification of genes required for adventurous gliding motility in *Myxococcus xanthus* with the transposable element mariner. Mol. Microbiol. *49*, 555–570.

Zakin, L., and Robertis, E.M.D. (2010). Extracellular regulation of BMP signaling. Curr. Biol. *20*, R89–R92.

Zaman, M.H., Trapani, L.M., Sieminski, A.L., MacKellar, D., Gong, H., Kamm, R.D., Wells, A., Lauffenburger, D.A., and Matsudaira, P. (2006). Migration of tumor cells in 3D matrices is governed by matrix stiffness along with cell-matrix adhesion and proteolysis. Proc. Natl. Acad. Sci. *103*, 10889–10894.

Zhang, H., Angus, S., Tran, M., Xie, C., Igoshin, O.A., and Welch, R.D. (2011). Quantifying aggregation dynamics during *Myxococcus xanthus* development. J. Bacteriol. *193*, 5164–5170.

Zusman, D.R. (1982). "Frizzy" mutants: A new class of aggregation-defective developmental mutants of *Myxococcus xanthus*. J. Bacteriol. *150*, 1430–1437.

CHAPTER 2

DATA-DRIVEN MODELING REVEALS CELL BEHAVIORS CONTROLLING SELF-ORGANIZATION DURING MYXOCOCCUS XANTHUS DEVELOPMENT[1]

[1] C. R. Cotter. H-B. Schüttler. O. A. Igoshin. L. J. Shimkets. 2017 PNAS. 114(23):E4592-E4601. Reprinted here with permission of the publisher per guidelines at http://www.pnas.org/site/aboutpnas/rightperm.xhtml

Abstract

Collective cell movement is critical to the emergent properties of many multicellular systems including microbial self-organization in biofilms, embryogenesis, wound healing, and cancer metastasis. However, even the best studied systems lack a complete picture of how diverse physical and chemical cues act upon individual cells to ensure coordinated multicellular behavior. Known for its social developmental cycle, the bacterium *Myxococcus xanthus* utilizes coordinated movement to generate 3-dimensional aggregates called fruiting bodies. Despite extensive progress in identifying genes controlling fruiting body development, cell behaviors and cell-cell communication mechanisms that mediate aggregation are largely unknown. We developed an approach to examine emergent behaviors that couples fluorescent cell tracking with data-driven models. A novel feature of this approach is the ability to identify cell behaviors affecting the observed aggregation dynamics without full knowledge of the underlying biological mechanisms. The fluorescent cell tracking revealed large deviations in the behavior of individual cells. Our modeling method indicated that decreased cell motility inside the aggregates, a biased walk toward aggregate centroids, and alignment among neighboring cells and in a radial direction to the nearest aggregate are behaviors that enhance aggregation dynamics. Our modeling method also revealed that aggregation is generally robust to perturbations in these behaviors and identified possible compensatory mechanisms. The resulting approach of directly combining behavior quantification with data driven simulations can be applied to more complex systems of collective cell movement without prior knowledge of the cellular machinery and behavioral cues.

Introduction

Collective cell migration is essential for many developmental processes including fruiting body development of myxobacteria (2014) and *Dictyostelium* (Bretschneider et al., 2016), embryonic gastrulation (Aman and Piotrowski, 2010; Solnica-Krezel and Sepich, 2012), and neural crest

development (Theveneau and Mayor, 2012). Conversely, cancer cell metastases represent detrimental migratory events that disseminate dysfunctional cells (Friedl and Gilmour, 2009). In all these processes, a population of cells leaves its current location and migrates in a coordinated manner to new locations where motility becomes reduced. Remarkable progress has been made in studying the intracellular machinery of these organisms (Park et al., 2016). Much less is known about the system-level coordination of cell migration. Cell movement in these systems is a three-dimensional, dynamic process coordinated by a combination of diverse physical and chemical cues acting on the cells (Aman and Piotrowski, 2010; Delgado and Torres, 2016; Theveneau and Mayor, 2012). Recent developments in tracking individual cell movement *in vivo* have provided unprecedented detail and revealed surprising levels of heterogeneity (Park et al., 2016; Theveneau and Mayor, 2012). Reverse engineering how these individual cell movements lead to collective migration patterns has proven difficult. While computational models are able to test whether a given set of ad-hoc assumptions lead to emergence of observed patterns, these models usually ignore heterogeneity of cell responses, overlook complex behavior dynamics, and rarely perform quantitative comparisons with *in vivo* results (Iber and Zeller, 2012; Masuzzo et al., 2016; Schumacher et al., 2016; Szabó and Mayor, 2016). Therefore, a data-driven modeling framework that integrates multiple levels of experimental observation with quantitative hypothesis testing is needed to uncover the interactions required for emergent behavior. We explored this possibility using a simple bacterial model system.

Emergent behaviors are a central feature of the life cycle of *Myxococcus xanthus*, which occurs within a biofilm many cell layers thick. Cells inside the biofilm are capable of signaling (Kuspa et al., 1992) and exchanging outer membrane material (Pathak et al., 2012). Cells are flexible rods that move along their long axis within the biofilm (Mauriello et al., 2010). Periodic reversals in direction of movement and a high length-to-diameter aspect ratio allows cells to align with

neighbors, move in groups, and follow paths taken by others (Balagam and Igoshin, 2015; Berleman et al., 2016; Lux et al., 2004). When faced with amino-acid limitation, cells self-organize into aggregates much taller than the surrounding biofilm called fruiting bodies (Lux et al., 2004; Xie et al., 2011). Aggregation begins with a burst of cell motility during which cells coalesce into unstable towers a few layers thicker than the surrounding biofilm (Curtis et al., 2007). Within an hour, towers begin to form spatially stable aggregation centers. While some aggregates mature into spore-filled fruiting bodies, many of initially stable aggregates disseminate back into the biofilm (Zhang et al., 2011). Little data exists on the cues and cell behaviors that lead to these emergent behaviors. Cell tracking experiments revealed that motility increases outside aggregates (Curtis et al., 2007; Jelsbak and Søgaard-Andersen, 2002; Thutupalli et al., 2015) and decreases inside (Sliusarenko et al., 2007; Thutupalli et al., 2015) while statistical image analysis revealed that the area of the aggregate solely determines whether an aggregate will disappear or mature into a fruiting body (Xie et al., 2011). On their own, these observations have been unsuccessful in explaining how cells coalesce to form stable aggregates.

Biochemical and genetic experiments have identified systems that could play a role in governing cell behavior during aggregation. Cells chemotax towards specific lipids by suppressing reversals when moving up the chemical gradients (Kearns and Shimkets, 1998), creating a biased walk. Exopolysaccharides, a major component of the extracellular matrix also inhibit cell reversals in a concentration dependent manner (Zhou and Nan, 2017). However, inhibiting cellular production of known lipid chemoattractants does not diminish aggregation (Bode et al., 2006; Kearns et al., 2001), and it is unclear whether exopolysaccharides act as chemoattractants. Induction of developmentally related genes when cells are tightly packed and aligned, but not for randomly positioned cells (Kim, 1990), suggest possible contact-based intercellular signaling. In agreement, cells at low cell densities decrease reversal frequency as group size increases (Shi et al.,

1996). However, this reversal suppression does not directly scale to the cell densities typically used in assays of development (Jelsbak and Søgaard-Andersen, 2002). Thus, while cells undergo behavioral changes indicative of intercellular signaling, conflicting results obscure what these signals are, or how they coordinate cell behaviors to drive aggregation. Computational modeling has frequently been used to bypass the lack of specific mechanistic details but has been largely unsuccessful in spanning the realm between fact and fancy.

While computational approaches have been extensively used in hypothesizing models of aggregation (Hendrata et al., 2011; Holmes et al., 2010; Igoshin et al., 2004; Sliusarenko et al., 2007; Sozinova et al., 2005, 2006; Starruß et al., 2007), the lack of quantitative data sets describing cell movement during aggregation has left the cell behaviors that drive the process conjectural. As a result of these models, cell length to width ratio (Starruß et al., 2007), cell alignment (Sliusarenko et al., 2006; Starruß et al., 2007), active-turning (Hendrata et al., 2011), density-dependent speed reduction (Sliusarenko et al., 2006), physical jamming (Holmes et al., 2010; Igoshin et al., 2004; Sozinova et al., 2005), and streaming (Holmes et al., 2010; Sozinova et al., 2005), have been introduced as cell behaviors required to generate aggregates in simulations. Quantitative comparisons between simulations and experimental results are needed to evaluate whether these simulations fully capture the characteristics of aggregation, but such comparison are rarely performed. For example, Zhang et al's (Zhang et al., 2011) analysis of the model in which aggregation is driven by cell alignment and reduced cell speed inside aggregates (Sliusarenko et al., 2007) revealed that the simulations fail to quantitatively capture the correct aggregation rate, aggregate distribution, and aggregate count. Despite this wealth of work, neither biological experiments nor mathematical models have so far identified the cell behaviors that mediate aggregation.

Here, in the absence of knowledge about the mechanistic basis of the cues directing cells, we identify motility parameters affecting the emergence of aggregates. We developed an approach

that couples multi-level cell tracking (at the level of individual cells within the biofilm and the level of the growing aggregates) with simulations driven by the cell behavior data. Directly including quantified cell behaviors in simulations, rather than averages or artificially generated behavior distributions, allowed full integration of heterogeneity and complex correlations in cell responses. Hypotheses about the cell behaviors driving aggregation were tested in increasingly complex simulations by quantitatively comparing simulations with *in vivo* results. This iterative process allowed us to identify cell behaviors that are sufficient and necessary to match the observed aggregation dynamics and creates opportunities for more powerful comparisons of mutant/parent behavioral differences in future studies.

Results

Cells decrease movement inside aggregates

To quantify cell behavior during development, we used time-lapse microcinematography to measure biofilm cell density, determine aggregate boundaries using a cell-density threshold, and follow individual cells within the biofilm (Figure 2.1). Under our conditions, aggregation begins 11 to 12 hours after spotting the cells on starvation media. We selected an approximately 5-hour window that began just prior to the initiation of aggregation through the period when stable aggregates form (Figure 2.1 A). The beginning of this window was designated time point zero. About 1 hour into this time span aggregation becomes evident. Stable aggregates appear by 1.5 hours with a few of the smaller aggregates disappearing by 5 hours. Aggregation was not compromised by the use of strains expressing fluorescent proteins or prolonged fluorescent imaging (Figure 2.1 D,E).

Cell tracking algorithms were developed to track individual fluorescent cells over the 5-hour window (Figure 2.1 B). Cell trajectories were subdivided into three movement states: persistent forward, persistent backward, or non-persistent. A persistent state was assigned to trajectory segments in which cells were actively moving along their long axis. To account for cell reversals, persistent movements were then further classified as backwards or forwards relative to the direction observed at beginning of the trajectory. The non-persistent state was assigned when we encountered a velocity too small (less than approximately 1 μm/min) or reversal period too high (greater than approximately 1 reversal/min) to accurately detect persistent movement at the spatial and temporal resolution of the time-lapse images. The resulting assignments divide a trajectory into segments. The vector from the beginning of one segment to the next was defined as a run vector (Figure 2.1 C). As such, a new run begins each time a cell changes its movement state. In what follows, we use run vectors to quantify cell motility behavior and to define the behavior of agents in agent-based simulations.

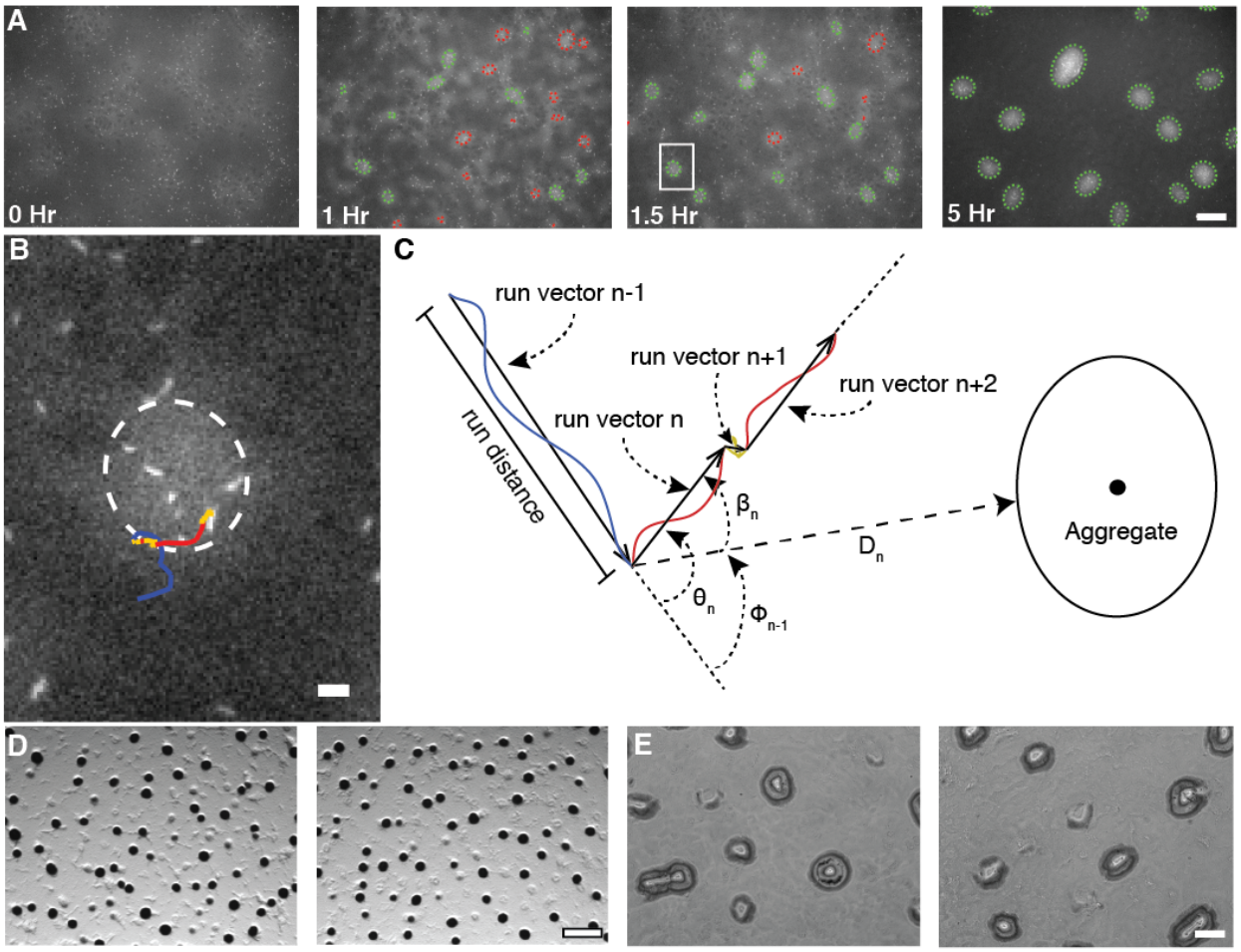

Figure 2.1: Overview of individual cell and aggregate tracking. A) Representative images from time-lapse microcinematography of developing *M. xanthus* cells highly expressing tdTomato mixed 1:2500 with cells weakly expressing eYFP. Cell density is proportional to eYFP fluorescence intensity while, in the same image, individual tdTomato cells are bright enough to detect and track . Detected aggregate boundaries are indicated with dashed green ellipses for stable aggregates and red ellipses for unstable aggregates. Bar represents 100 μm. B) Increased magnification of the image area inside the white box in (A). Line follows a single cell trajectory from the prior 40 minutes to the shown frame. Line color indicates detected cell state. Blue is persistent forward, red is persistent backwards, and yellow is non-persistent movement. Bar represents 10 μm. C) Cell trajectories were segmented into continuous states, the vector pointing from one state to the next is defined as a run vector. Colors are as in B. Run vector distance, speed, duration, distance to nearest aggregate boundary (D_n), angle between two consecutive run vectors (θ_n), and the angle between the nearest aggregate centroid and the ending (ϕ_{n-1}) of the previous and beginning (β_n) of the run vectors are representative of the variables calculated. All angles are in the interval $[-\pi, \pi)$ where $-\pi = \pi$. D) Mixtures of LS3629 and LS3908 on plates containing IPTG and vanillate (left panel) or DK1622 cells without IPTG or vanillate (right panel) produced similar aggregate profiles. Images taken 48 hours post starvation at 25x magnification. Bar represents 500 μm. E) Aggregation profiles were similar after 5 hours of fluorescent imaging (left) and without any fluorescent imaging (right). The phase images were taken at the same time point and magnification as the 5 Hr panel in A. Bar represents 100 μm.

To determine how aggregates affect cell behavior, runs were binned as starting inside or outside the aggregates. In both bins, the speed, duration, and distance of the runs are highly variable (Figure 2.2 A-C). Within aggregates, cells move with only a modest average speed decrease of 1.1-fold relative to outside the aggregates (Figure 2.2 A, blue stars). However, the probability for a cell to transition to a non-persistent state at the end of the run increases 1.8 fold (Figure 2.2 D). Moreover, the average duration of non-persistent runs doubles inside the aggregates (Figure 2.2 B, red stars). Average persistent run duration also decreases inside aggregates by approximately 1.5-fold (Figure 2.2 B, blue stars). These effects lead to a combined (persistent and non-persistent) 2-fold decrease in average run distance inside the aggregates vs. outside (Figure 2.2 C, magenta circles). These results are in agreement with other work suggesting that cells reduce movement inside aggregates (Sliusarenko et al., 2007) and provide much more quantitative detail.

Previous observations indicated that cells increase their movement when aggregation initiates (Curtis et al., 2007; Jelsbak and Søgaard-Andersen, 2002; Thutupalli et al., 2015). To quantify these effects, the mean and 95% confidence intervals for distance, duration, and speed of persistent state runs were calculated in a 20-minute sliding window over the length of the experiment (Figure 2.2 E, F, and G). Early in aggregation (c.a. 0-1.5 hours), the mean persistent run duration outside the aggregates increases approximately 1.8-fold (Figure 2.2 F, blue lines), causing an increase in run distance (Figure 2.2 G, blue lines). At approximately 1.5 hours, run duration transiently returns to levels seen prior to the onset of aggregation. Soon after, a second transitory increase in run duration occurs. As aggregates mature, run duration gradually decreases back to pre-aggregation levels. Inside the aggregates, speed and duration remain constant (Figure 2.2 E, F, and G black lines). Non-persistent run behaviors are also relatively constant, with run distance varying less than 1.5 μm over the length of the experiment (Figure 2.3 A-C). The probability of transitioning to a non-persistent state remains about the same, with the exception of a transitory

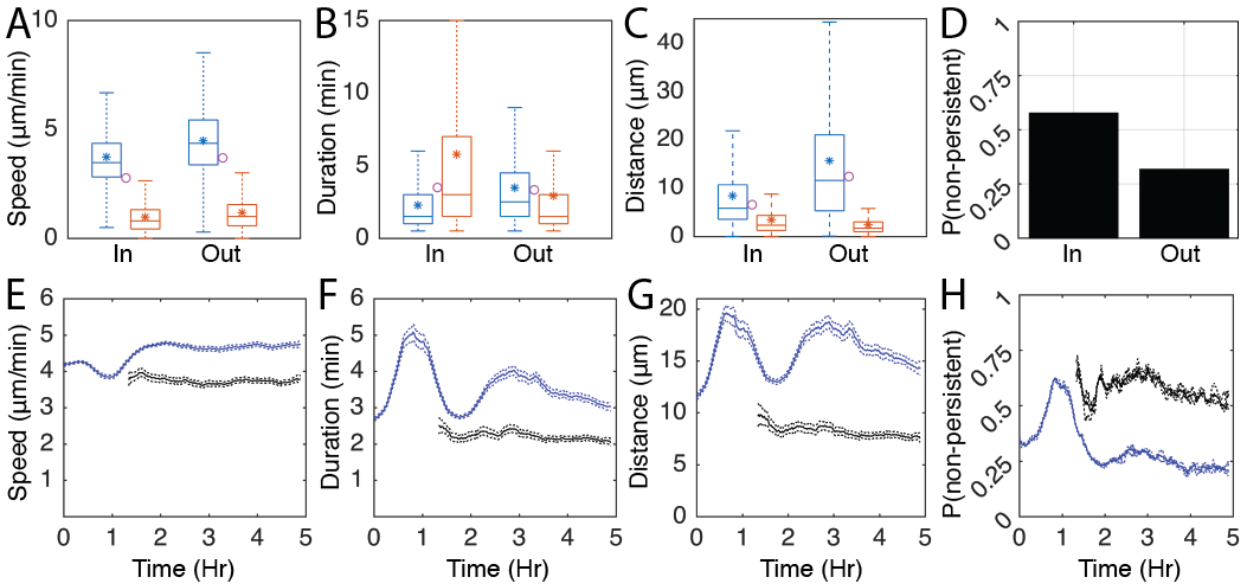

Figure 2.2: Run behaviors are dynamic in time and space. A-C) Time integrated distributions of persistent (blue) and non-persistent (red) run speed (A), duration (B), and distance (C) inside (In) and outside (Out) of the aggregates. Horizontal lines inside the boxes indicate distribution median. Tops and bottoms of each box indicate 75th (q_3) and 25th (q_1) percentiles, respectively. Whiskers extend to the highest and lowest points or $q_3 + 1.5(q_3 - q_1)$ and $q_1 - 1.5(q_3 - q_1)$, whichever is closer to the median. Stars indicate average. Circles indicate combined (persistent and non-persistent) average. D) Time integrated probability of choosing a non-persistent run after a persistent run inside (In) or outside (Out) of the aggregates. E-F) Mean (solid lines) and bootstrapped 95% confidence intervals (dashed lines) for run speed (E), duration (F), distance (G), and probability of choosing a non-persistent run after a persistent run (H) calculated in a 20-min. sliding window. Blue lines indicate runs starting outside the aggregates, black lines, runs inside the aggregates.

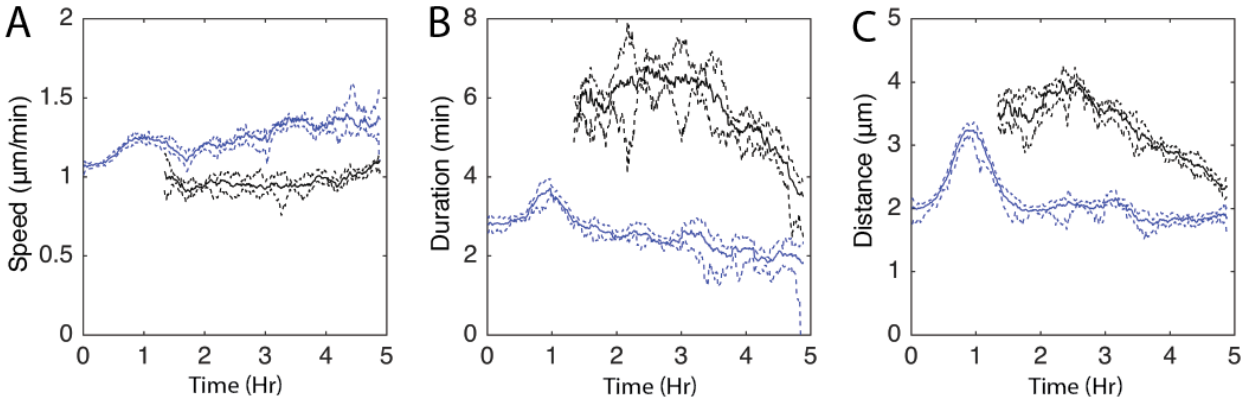

Figure 2.3: Experimental non-persistent run behaviors as a function of time. A-C) Non-persistent average run (solid lines) speed (A), duration (B), distance (C) inside (black lines) and outside (blue lines) the aggregates. Average was performed in a 20-minute sliding window. Dashed lines indicate 95% confidence intervals.

increase outside the aggregates coinciding with the first peak in run duration (Figure 2.2 H). Again, our measurements not only confirm earlier observations but provide greater quantitative detail to facilitate mathematical modeling.

Density-dependent motility decrease is not sufficient for aggregation

To identify the cell behaviors most important to timely and complete aggregation, we developed a data-driven, agent-based simulation technique that couples individual agent behavior with experimentally recorded cell-tracking statistics and biofilm level dynamics. Agents move in a series of straight lines with properties (persistent vs. non-persistent, duration, speed, and turning angle relative to the previous run) sampled from the experimentally measured run distributions. Given that run speed and duration were correlated (Spearman's $\rho = 0.2$ for persistent runs, $\rho = -0.5$ for non-persistent runs), they were sampled as a pair from a joint distribution containing the values from each experimental run. In the simplest model form, agents choose their run states, speeds, durations, and turning angles randomly from a distribution of all experimentally measured run behaviors independent of their location, cell density, or other factors. Since motility of the agents in this model is uncorrelated with their environment, the model does not generate any aggregates. Cells instead approach a steady state of uniform density (Figure 2.4 A). For aggregates to form, cells must coordinate their behavior through external cues.

To model behavior dependent on external cues, agent behavior was chosen conditional on the cell density at their location measured in the fluorescent cell microcinematography experiments. As a consequence, agents behave as if they are within the density profiles from the tracking experiments. This technique facilitates directly comparing different cell-behavior dependencies to the experimental results. Varying the enforced run behavior conditions in simulations can then test different hypotheses on the cues coordinating cell behavior. If the correct cell behavior dependencies are included in the simulations, aggregates should appear at the same

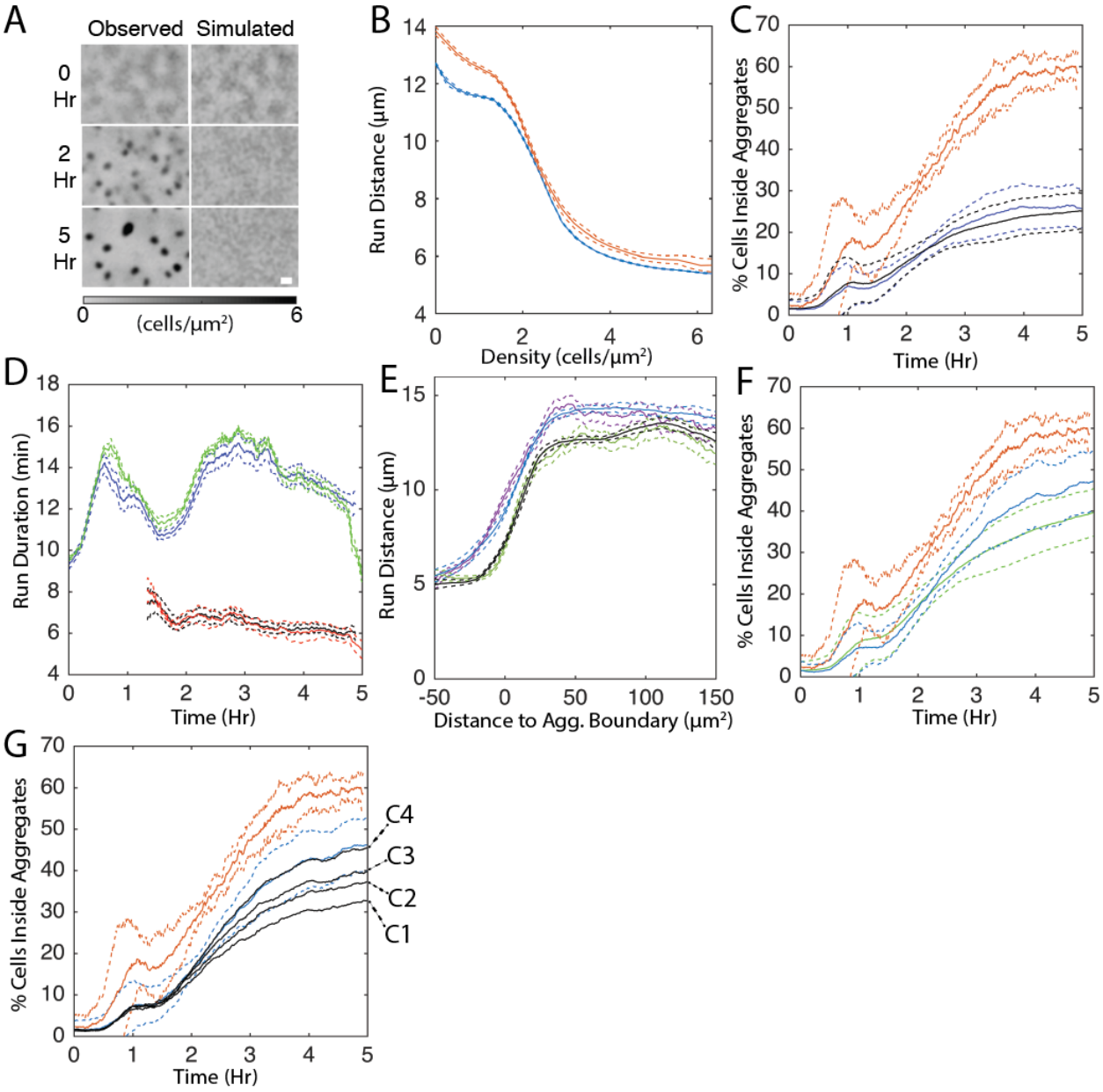
A
Observed
Simulated
0 Hr
2 Hr
5 Hr
0
6
(cells/μm²)
B
Run Distance (μm)
Density (cells/μm²)
C
% Cells Inside Aggregates
Time (Hr)
D
Run Duration (min)
Time (Hr)
E
Run Distance (μm)
Distance to Agg. Boundary (μm²)
F
% Cells Inside Aggregates
Time (Hr)
G
% Cells Inside Aggregates
Time (Hr)
C4
C3
C2
C1

Figure 2.4: Open-loop simulation controls. A) Comparison of experimental results with simulations in which agents are not dependent on any external variables. Representative time courses of experientially observed (Observed) and simulation (Simulated) cell densities over the course of the experimental time window. B) Average run distance in a 2 cells/μm^2 sliding window for simulations in which agent behavior is dependent on local cell density (blue lines) and experimental results (red lines). Dashed lines indicated bootstrapped 95% mean confidence intervals. C) Average (solid) and standard deviations (dashed lines) of the percent of cells inside aggregates for experimental (red lines) and simulations in which agent's behavior is driven by time since the beginning of the experiment and local cell density (blue lines) or only local cell density (black lines). D) Average run duration in a 20-minute sliding window for all runs (persistent and non-persistent) from experimental results inside (black lines) and outside aggregates (blue lines) and open-loop simulations in which agent behaviors were chosen dependent on time since the beginning of the experiment. Green lines indicate agents outside aggregates, red lines, inside. Dashed lines as in B. E) Average run distance in a 10 μm sliding window for simulations in which agent behavior depends on orientation to nearest aggregate. Purple (experimental results) and blue (simulation results) lines indicate runs oriented toward ($\cos(\beta_n) > 0$, see Figure 2.1 C) the nearest aggregate centroid. Green (experimental results) and black (simulation results) lines indicate runs pointed away ($\cos(\beta_n) < 0$) from the nearest aggregate centroid. Negative distances indicate that the run began inside the aggregate. Dashed lines indicated 95% bootstrap confidence intervals. F) Average (solid) and standard deviations (dashed lines) of the percent of cells inside observed aggregate boundaries for experimental (red lines) and simulations with a biased walk and with (blue lines) or without (green lines) time since the begging of the experiment as a dependence for choosing the agent's run state, speed, and duration. G) Average (solid lines) and standard deviations (dashed lines) of percent of cells inside the aggregates in simulations (black lines) in which agents chose their run state, duration, and speed dependent on orientation and distance to the nearest aggregate when the agent was inside an aggregate (C1) or within 25 μm (C2), 50 μm (C3), or 100 μm (C4) of the aggregate boundary. When outside the cutoff distance (C1-C4), no aggregate dependence was used to choose agent behaviors. Blue lines indicate simulations in which aggregate distance and orientation is always included in choosing agent behaviors. Red lines indicate experimental results.

locations, at the same rate, and to the same extent as the respective movie. We call this simulation type "open-loop" to denote that agent behavior is defined solely by the external density profile extracted from a microcinematography experiment (Figure 2.5, blue box).

Previous hypotheses of the mechanistic basis for aggregation predicted that decreased cell movement inside aggregates was the major driver of aggregate growth (Gejji et al., 2012; Igoshin et al., 2004; Sliusarenko et al., 2007; Sozinova et al., 2005; Zhang et al., 2011). We tested the hypothesis that the observed decrease in cell movement at the higher cell densities inside aggregates is sufficient to drive aggregation by incorporating density dependence into the simulations. Agents choose their run state, speed, and duration conditional on the experimentally measured local cell density at the beginning of their run. With the addition of this conditionality, agents exhibit a relationship between average run distance and local cell density similar to that of experimental runs (Figure 2.4 B). In the resulting simulations, aggregates appear at nearly all expected locations (Figure 2.5 B, right panel). However, the fraction of cells within the aggregate boundaries by the end of the 5 hr window is three-fold smaller in simulations compared to experimental results (Figure 2.5 B, left panel). Addition of time-dependence when choosing the state, speed, and duration (Figure 2.2 E-H) does not improve the rate or completeness of aggregation in simulations (Figure 2.4 C,D). These results are in agreement with another report indicating that simulations driven solely by local cell density fail to correctly reproduce the number, growth rate, and size of aggregates (Zhang et al., 2011).

Cells perform a biased walk toward the aggregate center

Biased walks are found in many types of cell patterning (Delgado and Torres, 2016; Lander, 2013; Morelli et al., 2012). Though chemotaxis has not been implicated in *M. xanthus* aggregation, *M. xanthus* can perform biased walks up specific lipid gradients (Kearns and Shimkets, 1998). Bias is created by increasing average run duration when moving up the chemoattractant gradient;

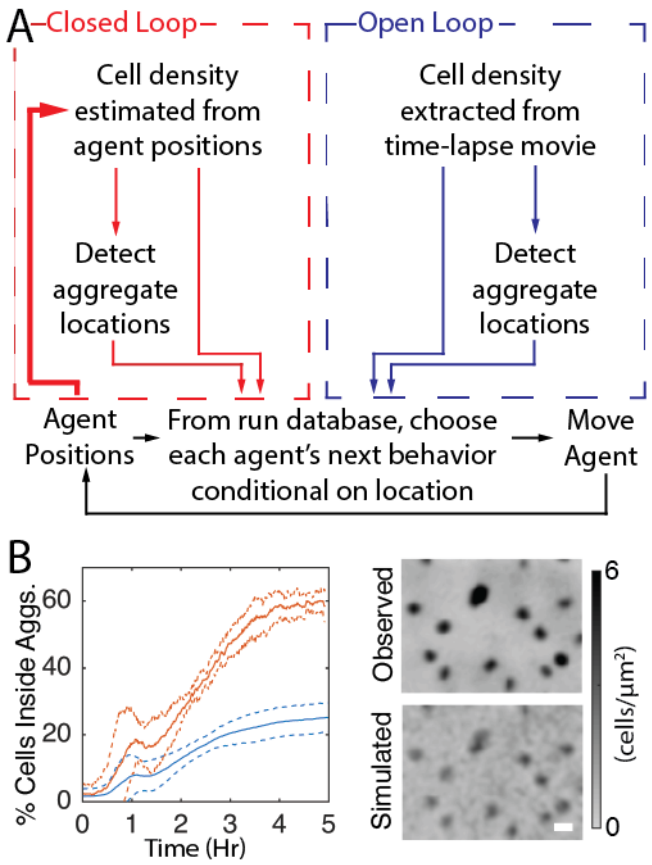

Figure 2.5: Reduced movement inside aggregates is not sufficient to fully replicate aggregation in open-loop simulations. (A) Overview of open-loop (blue) and closed-loop (red) simulations. The extra path in the closed-loop model is bolded to highlight that the agent's positions feed back into the density profile of the biofilm, closing the loop between individual and population level behaviors. (B) Comparison of experimental results with open-loop simulations in which agents reduce average movement proportional to cell density. Left: Average (solid lines) and standard deviations (dashed lines) of the percent of cells inside experimentally observed aggregate boundaries for experimental (red) and simulation (blue). Right: Comparison of last frame of representative experientially observed (Observed) cell density with that observed in a simulation.

conversely, cells decrease average run bias when moving down the gradient. We tested whether cells change their behavior depending on their direction of movement relative to nearby aggregates. Run vectors were quantified with respect to the direction of moment and distance to the nearest stable aggregate (Figure 2.1 A, green ovals). The results show that persistent runs moving toward the aggregate centroid are longer than runs moving away from it (Figure 2.6 A). This bias is due to an increase in run duration rather than run speed (Figure 2.6 B,C). The probability of transitioning to a non-persistent state at the end of the run also depends on the run orientation relative to the nearest aggregate (Figure 2.6 D). Inside the aggregates, non-persistent run durations are 1.5-times longer when moving away from the aggregate centroid (Figure 2.6 E,F). In contrast to a previous report of tangential cell movement inside the aggregates (Sager and Kaiser, 1993), our run durations are longest when pointed toward the aggregate centroid (Figure 2.7).

A biased walk towards aggregates aids in aggregation

To test the importance of the biased walk in aggregation, simulations were performed in which agent's run state, duration, and speed were chosen conditional on the orientation and distance of the agent to the nearest aggregate at the beginning of the run in addition to the local cell density. To account for observed time-dependence in the biased walk (Figure 2.7 B-D), run state, speed and duration were also chosen conditional on time since the beginning of the experiment. As a result, run duration dynamics relative to aggregate location in the simulation matched those in experiments (Figures 2.4 E). The inclusion of the biased walk increases aggregation rate and completeness, leading to a 2-fold increase in the fraction of agents inside aggregates (Figure 2.8 A). Aggregate density (Figure 2.8 B) and size (Figure 2.8 C) in simulations were close to the experimental values. In models with the biased walk, elimination of time-dependence in run properties marginally decreases aggregation (Figure 2.4 F). In these simulations, it is necessary for

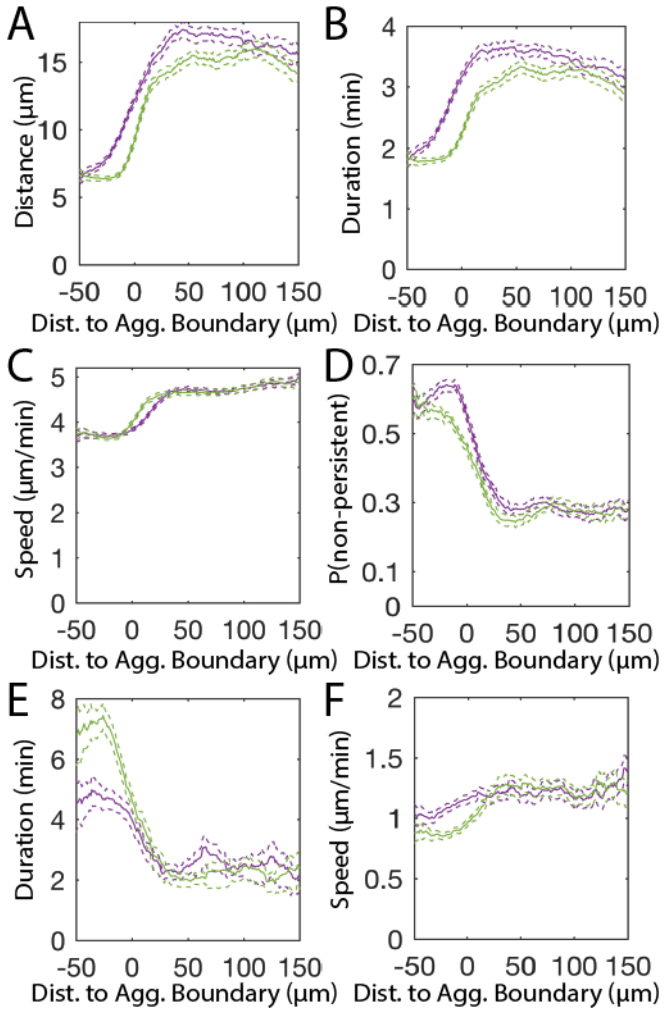

Figure 2.6: Cells perform a biased walk towards aggregates. A-F) Average (solid lines) and bootstrapped 95% confidence intervals (dashed lines) of persistent run distance (A), duration (B), speed (C), probability of choosing a non-persistent run (D), non-persistent duration (E), and speed (F) in a 10 μm sliding window around the distance (dist.) to the nearest aggreagte boundary from the beginning of the runs. Runs are binned into either pointing toward ($cos(\beta_n) > 0$ in A-C,E,F, or $cos(\phi_{n-1}) > 0$ in D, see Figure S1C) the nearest aggregate centroid (purple lines) or pointed away ($cos(\beta_n) < 0$ in A-C,E,F, or $cos(\phi_{n-1}) < 0$ in D) from the nearest aggregate centroid (green lines). Negative distances indicate that the run began inside the aggregate.

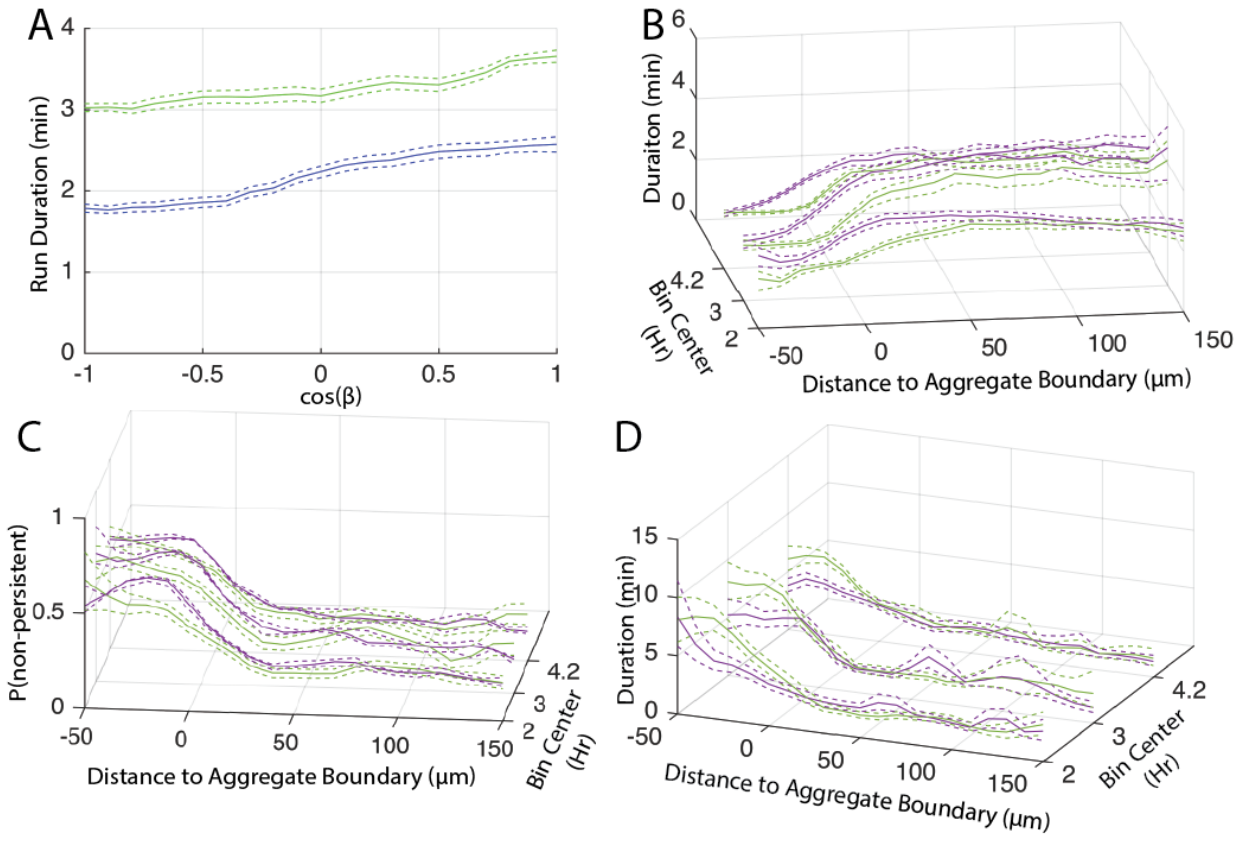

Figure 2.7: Extended biased walk quantification. A) Average and 95% confidence intervals (dashed lines) of persistent run duration as a function of the orientation to the nearest aggregate centroid (β in Figure S1C) for runs starting inside (green lines) and outside (blue lines) aggregates. Cos(β) of 1 indicates running directly towards the aggregate centroid and cos(β) of -1 indicates directly away. B-D) Average and 95% confidence intervals (dashed lines) of persistent run duration (B), probability of choosing a non-persistent run after a persistent run (C), and non-persistent run duration (D). Analysis was binned into 1.5-2.5 hours, 2.5-3.4 hours, and greater than 3.4 hour bins, from front to back, respectively. Purple lines indicate runs oriented toward ($\cos(\beta_n) > 0$ in B,D, $\cos(\phi_{n-1}) > 0$ in C, see Figure S1C) the nearest aggregate centroid and green lines indicate runs pointed away ($\cos(\beta_n) < 0$ in B, D, $\cos(\phi_{n-1}) < 0$ in C) from the nearest aggregate centroid. Negative distances indicate run began inside the aggregate.

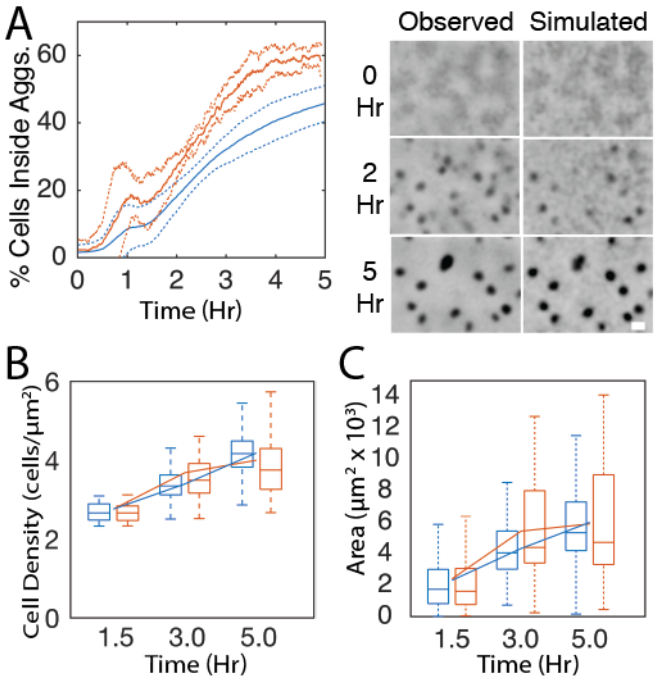

Figure 2.8: A biased walk towards aggregates contributes to aggregation in open-loop simulations. A-C) Comparison of experimental results (red) with simulations (blue) in which agents reduce movement proportional to cell density and perform a biased walk towards aggregates. A) Left: Formatted as in (Figure 2B). Right: Representative time courses of experientially observed and simulation cell densities over the course of the experimental time window. Grayscale is proportional to cell density as in (Figure 2B). B) Distribution of average cell density inside aggregates. C) Distribution of aggregate area. Box plots formatted as in Figure 1A. Line plots indicate mean. All bars are 100 μm.

agents to choose their next behavior conditional on the orientation and distance to the nearest aggregate when up to 100 μm away to achieve full aggregation (Figure 2.4 G).

Closed-loop model of aggregation.

The open-loop simulations identified behaviors that achieve aggregation comparable to that of experimental results. By nature of the technique, aggregate initiation and growth in these simulations were enforced through the continued input of measured cell density profiles. To more stringently test the effect of cell behaviors on aggregation, we closed the loop between agent behavior and the density profile. In contrast to the open-loop simulation's dependence on experimental cell density profile as input, the closed-loop simulations (Figure 2.5 A, red box) estimate the density profile from the agent positions by kernel density estimation (KDE) (Botev et al., 2010). Aggregates were then detected from the agent density profile using the same density cutoff as in experiments. The resulting density profile and aggregate boundaries were used to choose the agent run characteristics, closing the feedback loop between agent behavior and their density profile (Figure 2.5 A). Except for the change in density estimation, the closed-loop model is identical in design to the open-loop model. That is, agents choose their run state, speed, and duration conditional on the local agent density, distance and orientation to the nearest aggregate, and time since the beginning of the experiment. Closed-loop simulations thereby provide a more realistic simulation environment by allowing agents positions to modify the surrounding density profile.

The resulting closed-loop simulations lead to aggregate formation but, as compared to experimental results and open-loop simulations, the fraction of cells in aggregates decreased about two-fold (Figure 2.9 A). Although the resulting average cell density inside the aggregates agrees with experiments (Figure 2.10 A), the aggregate area is smaller than in experimental results (Figure

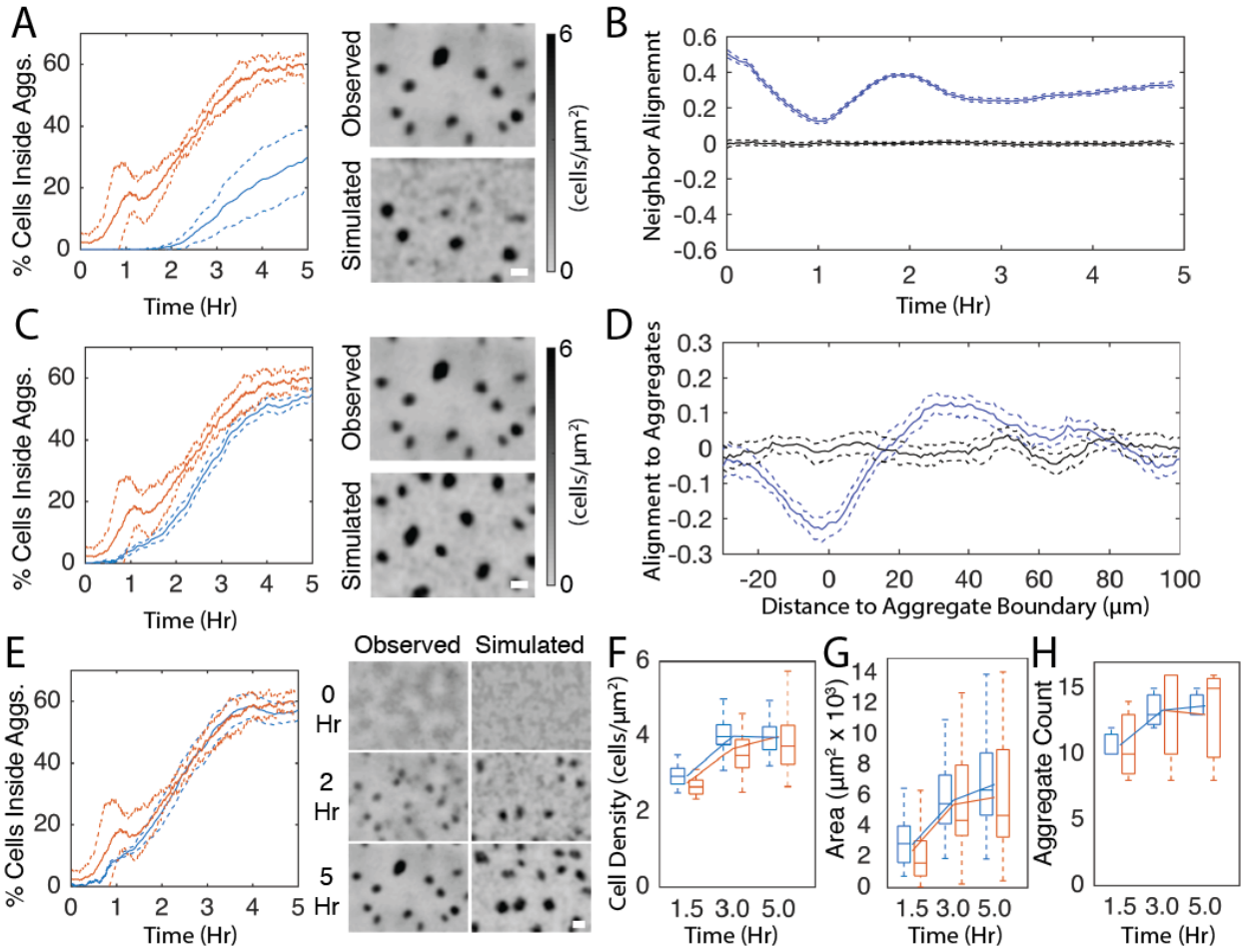
A
% Cells Inside Aggs.
Time (Hr)
Observed
Simulated
(cells/μm²)
B
Neighbor Alignemnt
Time (Hr)
C
% Cells Inside Aggs.
Time (Hr)
Observed
Simulated
(cells/μm²)
D
Alignment to Aggregates
Distance to Aggregate Boundary (μm)
E
% Cells Inside Aggs.
Time (Hr)
Observed
Simulated
0 Hr
2 Hr
5 Hr
F
Cell Density (cells/μm²)
Time (Hr)
G
Area (μm² x 10³)
Time (Hr)
H
Aggregate Count
Time (Hr)

Figure 2.9: Closed-loop simulations reproduce wild-type like aggregation with the addition of cell alignment. A) Simulation results in which agents reduce movement proportional to cell density and perform a biased walk towards aggregates. Left Panel: Average (sold lines) and standard deviation (dashed lines) of the percent of cells inside detected aggregates for experimental (red) and simulation (blue) replicates. Right: Comparison of the last frame of a representative experientially observed (Observed) cell density with a simulation (Simulated). B) Average (solid lines) and 95% confidence intervals (dashed lines) of run vector alignment strength (blue lines) with neighboring run vectors that occurred within ±5 min and 15 μm. Black lines indicate alignment strength with randomly chosen runs. Values may span (-1,1) where 1 indicates all runs are parallel. Likewise, -1 indicates all runs are perpendicular. C) Same as A with the addition that agents in the simulations align their orientation with neighboring agents. D) Alignment strength of run vectors (blue lines) with vector pointing toward nearest aggregate centroid. Black lines indicate alignment strength after randomly shuffling each run's distance to the nearest aggregate. Negative distances indicate that the run began inside an aggregate. Values may span (-1,1) as in B. E-H) In addition to the agent behaviors from simulations in (A) and (C), agents orient toward the nearest aggregate centroid. E) Left panel: Percent of cells inside aggregates as in (A). Right: Comparison of representative experientially observed cell density time progression with that observed in the closed-loop simulation. Grayscale is proportional to cell density as in (A). F) Average cell density inside aggregates. G) Average aggregate area. H) Aggregate count in each replicate. Box plots formatted as in Figure 1A. Lines indicate mean. All bars are 100 μm.

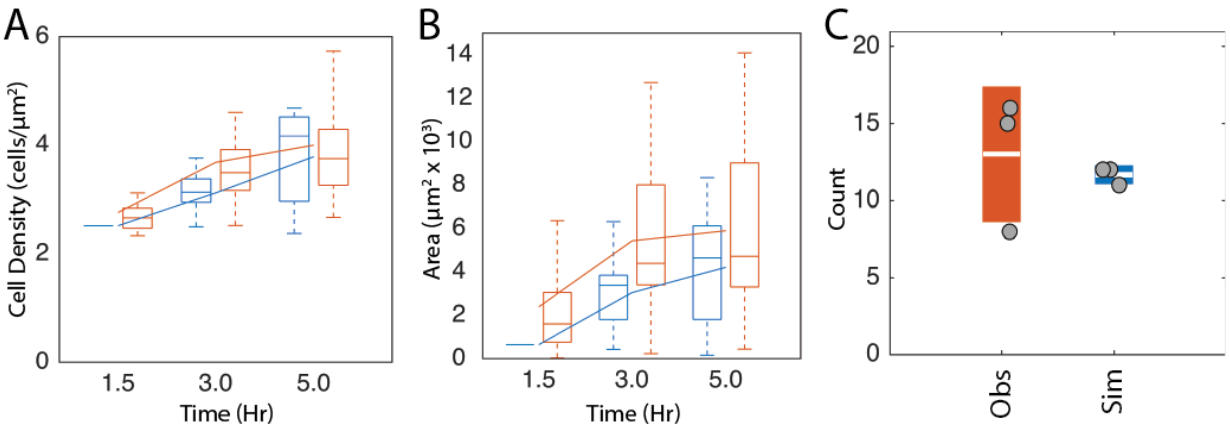

Figure 2.10: Quantification of closed-loop simulations without agent alignment. A and B) Distribution of average cell density inside aggregates (A) and aggregate area (B) in closed-loop simulations without any turning angle dependencies (blue lines). Experimental data in blue. Box plots formatted as described in Figure 2.2. C) Points indicate number of aggregates in each experimental movie (Obs) or simulation (Sim). Boxes indicate sample standard deviation with the white line indicating the sample mean.

2.10 B). Therefore, we hypothesized that additional run properties need to be included to facilitate complete aggregation.

Cell trajectories are aligned within the biofilm

In agreement with other experimental observations (Balagam and Igoshin, 2015; Berleman et al., 2016; Lux et al., 2004; Sliusarenko et al., 2007; Thutupalli et al., 2015), visual inspection of cell trajectories indicates alignment between neighboring paths (Figure 2.11 A, solid boxes). The presence of this alignment has previously been proposed to play a role in aggregation, but has not been experimentally quantified in the high cell densities used in developmental assays. To quantify alignment, we followed (Balagam and Igoshin, 2015) by calculating nematic alignment strength as the correlation of run orientations modulo 180 degrees (with cells moving in the opposite directions still considered aligned) among runs that start within a 15 μm radius and ±5 min of one another. In agreement with visual observations, quantification indicates a correlation in neighboring run orientations (Figure 2.9 B). Furthermore, observations (Figure 2.11, dashed boxes) and quantification of the mean run orientation relative to the nearest aggregate ($\langle \cos(2\beta_n) \rangle$, see Figure 2.1 C) indicate that run vectors outside the aggregate preferentially orient in a direction radial to the nearest aggregate (Figure 2.9 D). Inside the aggregates, runs are biased toward a more tangential orientation. The orientation of cells relative to the aggregates changes with time, with a radial run orientation prevalent at the onset of aggregation and becoming less pronounced as the aggregates mature. In contrast, run orientation inside the aggregates is random early in aggregation and becomes more tangential to the aggregate boundary as the aggregates mature (Figure 2.11 B).

Cell alignment aids in aggregate initiation

The hypothesis that cell alignment improves aggregation was tested in a closed-loop model. Cell alignment was included in simulations by choosing agent turning angles conditional on both the average nematic orientation of neighboring agent runs and the time since the beginning of the

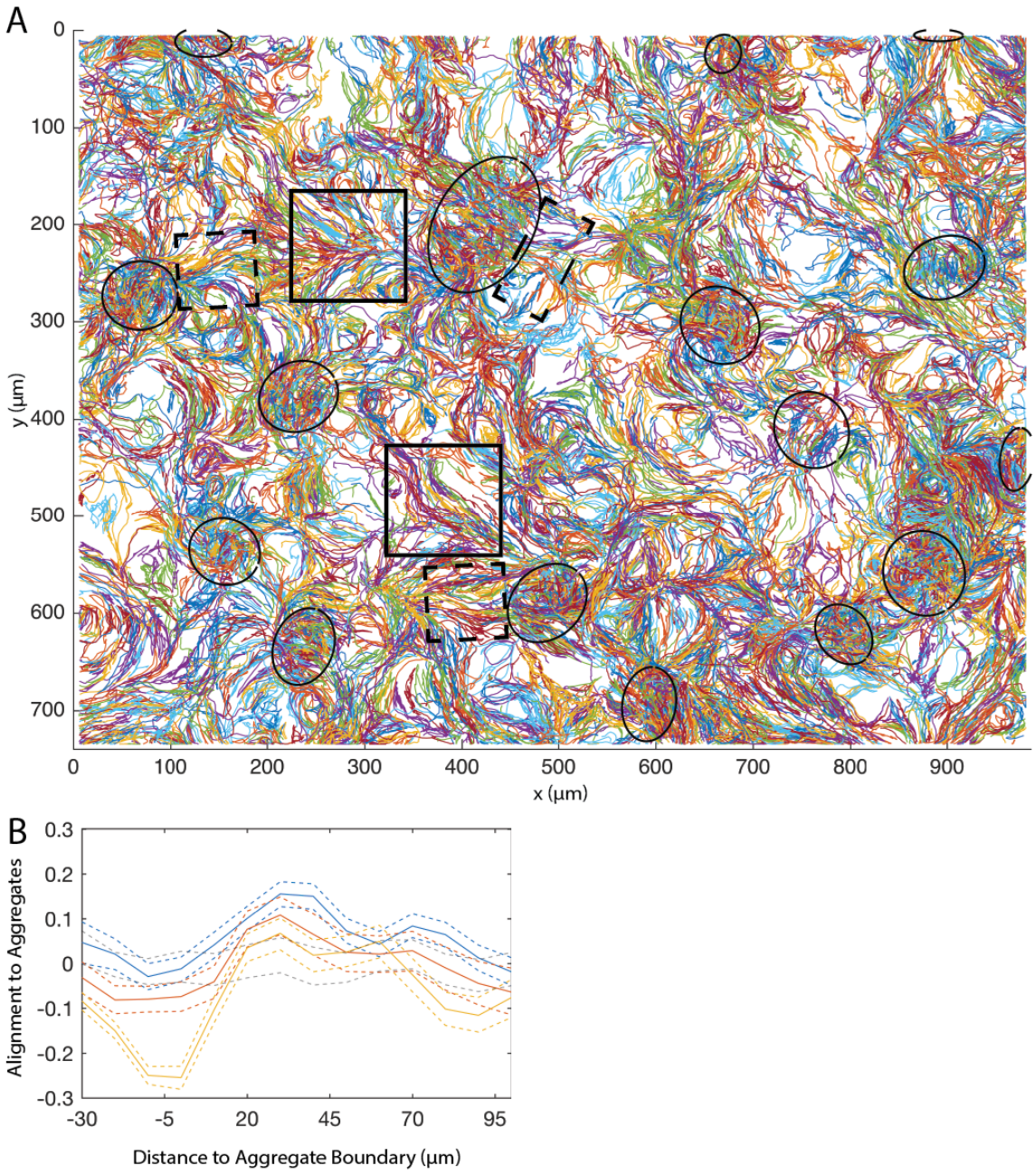

Figure 2.11: Cell trajectories are aligned. A) Plot of all cell trajectories extracted from the movie shown in Figure S1. Trajectories are randomly colored, with colors used multiple times. Ellipses indicate aggregate positions at the end of the experiment. Solid boxes indicate examples of trajectory alignment; dashed boxes indicate examples of trajectories orientated radial to the nearest aggregate boundary. B) Average alignment strength (solid lines) and 95% confidence intervals (dashed lines) of run vectors with vector pointing to the nearest aggregate centroid during hours 1.5-2.5 (blue lines), hours 2.5-3.4 (red lines), and greater than 3.4 hours (yellow lines). Dashed gray lines indicate 95% confidence intervals of alignment strength of a randomly selected set of N runs after randomly shuffling each run's distance to the nearest aggregate. N was equal to the average number of runs in the time bins.

experiment. To allow agents time to align prior to their initiation of aggregation, the simulation was run for 1.5 hours of simulation time using the behavior distribution and turning angles from the first 10 minutes of the experimental results. During this time, agent alignment approaches that seen in the experimental results (Figure 2.12 A). After the 1.5-hour prerun, the simulation was started using agent positions and orientations from the end of the prerun. Addition of neighbor alignment increases aggregation to levels comparable to the open-loop model (Figure 2.9 C). As a control, adding a prerun to simulations without neighbor alignment did not affect aggregation (Figure 2.12 B), confirming that addition of the prerun does not affect aggregation beyond that of aligning the agents.

The addition of neighbor alignment in simulations does not cause cells to orient radially with the nearest aggregate (Figure 2.12 C). To include orientation in the simulations, distance to the nearest aggregate boundary and angle to the nearest aggregate centroid were added as dependences on choosing the next turning angle. As a result, the closed-loop model displayed aggregation rates comparable to those of the experimental results (Figure 2.9 E). Furthermore, aggregate cell density (Figure 2.9 F), area (Figure 2.9 G), and aggregate count (Figure 2.9 H) agree with the experimental results. Thus, the closed-loop model revealed one additional feature not discovered in the open-loop model, a requirement for cell alignment. It now becomes possible to perturb the cell behavior dependences included in the closed-loop model to gauge their relative importance.

Behaviors shaping aggregation dynamics

By performing simulations in which the behaviors suggested to be required for aggregation are removed or modified, it is possible to predict phenotypes. To this end, closed-loop simulations were performed in which behaviors identified as necessary to match observed aggregation dynamics were systematically modified (Figure 2.13). Time dependence of the agent's turning angles was not

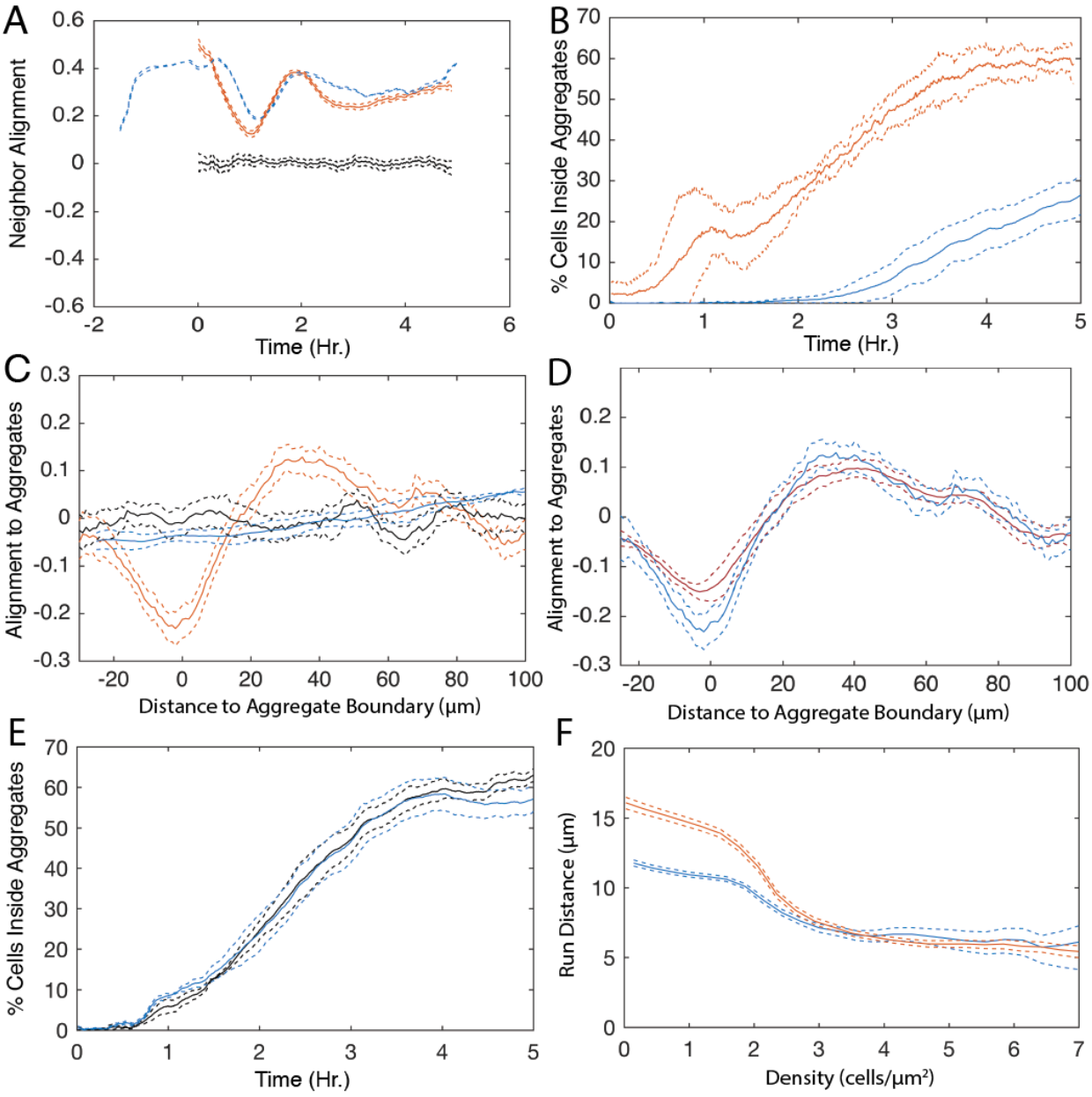
A
Neighbor Alignment
Time (Hr.)
B
% Cells Inside Aggregates
Time (Hr.)
C
Alignment to Aggregates
Distance to Aggregate Boundary (μm)
D
Alignment to Aggregates
Distance to Aggregate Boundary (μm)
E
% Cells Inside Aggregates
Time (Hr.)
F
Run Distance (μm)
Density (cells/μm²)

Figure 2.12: Closed-loop simulation controls. A) Time course of mean (solid lines) and 95% bootstrap confidence intervals (dashed lines) of nematic agent alignment with neighboring runs in a 20-min sliding window for simulations (blue) and experimental (red). Negative values indicate the simulation prerun to provide agents time to align. Black lines indicate analysis performed with randomly chosen runs instead of neighboring runs. B) Average (solid lines) and standard deviation (dashed lines) of the percent of cells inside aggregates for experimental (red lines) and simulation (blue) with a prerun but no turning angle dependencies. C) Average (solid lines) and 95% confidence intervals (dashed lines) of mean run vector alignment with vector pointing to the nearest aggregate centroid in open-loop simulations without any turning angle dependencies (blue lines). Black lines indicate a random distribution as in Figure 5D, red lines indicate observed experimental alignment. D) Average (solid lines) and 95% confidence intervals (dashed lines) of orientation of runs relative to the nearest aggregate centroid. Experimental data in red, closed-loop simulations in which distance and orientation to nearest aggregate centroid was included as dependence for choosing turning angle in blue. -1 indicates all runs are tangent to aggregate centroid, 1 indicates all runs are radial to aggregate centroid. E) Average (solid) and standard deviations (dashed lines) of fraction of cells inside aggregates in simulations with (blue) and without (black) time as a dependence for choosing the agent's next turning angle. F) Average run distance in a 2 cell/μm^2 sliding window for all runs (persistent and non-persistent) that occurred during hours 1.5-2 hours (blue lines) and 2.5-3.4 (red lines). Dashed lines as indicate 95% mean confidence intervals.

included to enable running simulations for times longer than the available experimental data time-window. Simulations indicate that this change does not affect aggregation dynamics (Figure 2.12 E). As in open-loop simulations (Figure 2.5 B), removing the biased walk slows aggregation rate (Figure 2.13 B). However, closed-loop simulations can be run longer than in experimental movies when the time dependences are not included. When simulations were continued for another 5 hours, agents continue to aggregate, approaching a steady state by 10 hours. Even after 10 hours, the fraction of cells inside the aggregates and aggregate density is approximately 30% lower than in experimental results and aggregate boundaries appear less well defined.

The two transient increases in run duration at the onset of aggregation (ca. 0.5-1.25 hours, see Figure 2 F) and during rapid aggregate growth (ca. 2.5-3.4 hours) suggest a possible role for time-dependent run duration. Outside the aggregates, this increase in duration leads to a combined (persistent and non-persistent) average run distance in the earlier time-window that is 1.3 times longer than the latter (Figure 2.12 F). Inside the aggregates, run distances are about the same in both time windows (Figure 2.12 F). To determine the role of these changes in aggregation dynamics, we utilized closed-loop models in which run data only from the 1.5-2 hr or only from the 2.5-3.4 hr window was used to drive agents' behavior for the whole simulation duration. Models based on the short run duration window (1.5-2 hr) produced aggregates at a rate and completeness equivalent to those of the experimental results (Figure 2.13 C). In contrast, agents in simulations utilizing the longer run durations (2.5-3.4 hr) aggregate at a faster rate and to a higher level of completeness than experimental results (Figure 2.13 D). We wondered whether extending the window of longer reversal durations could overcome the need for a biased cell walk. To test this hypothesis, simulations were run using the time windows but without a biased walk towards aggregates. Using the window with longer run durations, agents formed aggregates comparable to experimental results in rate, size, and cell density (compare Figure 2.13 A with Figure 2.13 E). The short run

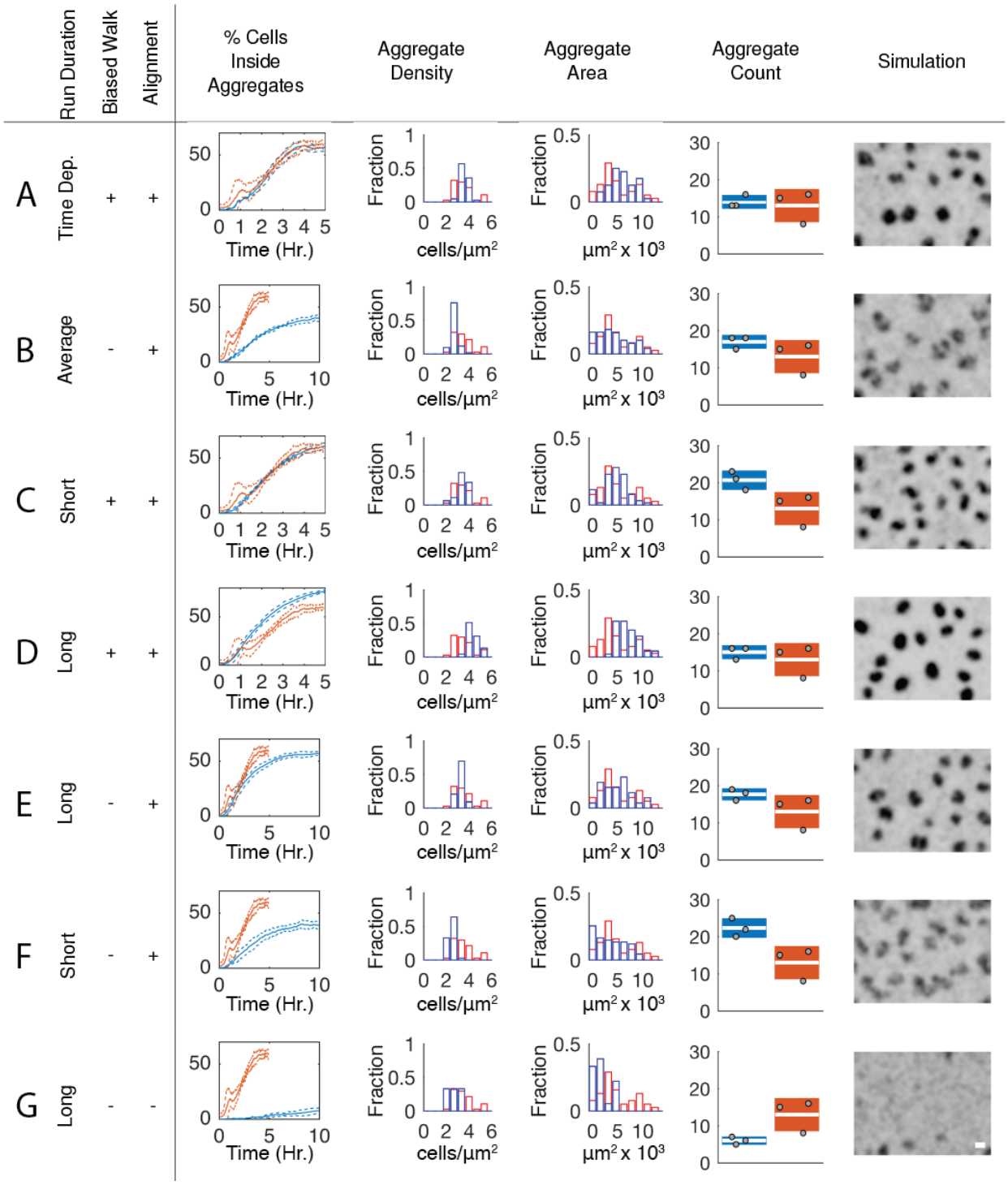
Run Duration
Biased Walk
Alignment
% Cells Inside Aggregates
Aggregate Density
Aggregate Area
Aggregate Count
Simulation
A
Time Dep.
+
+
B
Average
-
+
C
Short
+
+
D
Long
+
+
E
Long
-
+
F
Short
-
+
G
Long
-
-
Time (Hr.)
Fraction
cells/µm²
µm² x 10³

Figure 2.13: Probing interactions shaping aggregation dynamics in closed-loop simulations A-G) Percent cells inside aggregates, aggregate area, cell density inside aggregates, and aggregate count from the last time point in simulations (blue) and experimental (red) results. Aggregate density and area box plots are formatted as in Figure 1. Aggregate count box plots indicate the standard deviation of the replicate counts, white bar indicates the mean count, and each gray dot indicates the count from one replicate. A visual image of the last frame of the simulation was created using a KDE, shading is the same as in Figure 5A. The bar is 100 μm in length. A) Same simulation as in Figure 5E-H. B) Simulations with run behaviors from the entire experimental time span, alignment to neighboring cells and to the nearest aggregate centroid, and without a biased walk towards aggregates. C) Simulations with a biased walk, alignment, and run behaviors chosen from a time window (1.5-2 hours, see Figure 1F) containing short run durations outside of aggregates. D) Simulations as in (C) except run behaviors from a time window (2.5-3.5 hours) containing longer run durations outside aggregates. E) Same as (D) minus the biased walk. F) Same as (C), minus the biased walk. G) Same as (E) minus alignment to neighboring runs and the nearest aggregate centroid.

duration window caused agents to aggregate at a rate and completeness comparable to simulations in which behaviors were chosen from the entire movie but without the biased walk (compare Figure 2.13 B with Figure 2.13 F). Removing both alignment and the biased walk all but abolished aggregation, even when using the longer run duration window (Figure 2.13 G).

Discussion

Identifying cell behaviors that mediate self-organization without a full understanding of the underlying signaling network and motility control mechanisms is a daunting task. Here we developed a framework that integrates datasets of quantified cell behaviors with computer simulations driven by these datasets to reverse-engineer the self-organization process. This approach revealed a set of behaviors that appear to mediate complete aggregation in *M. xanthus*. Our results suggest that cells employ a combination of previously proposed behaviors, such as reduced cell movement inside aggregates, and previously unknown behaviors, including a biased walk toward the aggregate centroid. Remarkably, despite the large heterogeneity observed in individual cell behavior (Figure 2.2 A-C), we found that relatively small changes in average cell behavior, such as a 15% increase in average run duration when moving towards aggregates (Figure 2.6 B), dramatically improved aggregation. At the level of millions of cells, the population can tolerate occasional eccentric behavior provided the average cell behavior engages in the common activity. Live imaging has revealed unexpected heterogeneity and plasticity in stem cell biology (Park et al., 2016) suggesting that heterogeneity may be more widespread than currently appreciated in developmental biology. Large deviation occurs at the expense of resource depletion and would be expected to persist only if it provides an evolutionary benefit. The importance of small changes in average behavior in the face of large deviations from the mean also highlights the utility of large experimental datasets and data-driven simulations to confidently distinguish important cell behaviors from background noise.

To uncover the role of each cell behavior in a dataset with multiple correlated and noisy variables, the framework utilizes two simulation environments (Figure 2.5 A). The open-loop simulation environment assesses the importance of specific cell behaviors by directly overlaying the simulation agents over experimentally measured environments. This overlay provides a

structured way to assess the role of each cell behavior individually. Once the behaviors required to achieve quantitative agreement between open-loop simulations and experimental patterns are identified, closed-loop simulations in which the simulation agents define and modify their environment are employed to study how individual cell behavior shapes the behavior of the population. Through systemically adding and removing dependencies driving cell behavior, simulation results predict essentiality of various cell behaviors.

We believe the framework is generally applicable to many types of cell tracking experiments. The framework can be further generalized to include any additional data on the cell state (e.g. fluorescent gene reporters) or the surrounding environment (e.g. neighboring cells, landmarks, or boundaries) that could be correlated with cell behavior. For example, studies aiming to understand metastatic cancer cell invasion face challenges similar to *M. xanthus* development. Tumor cell state and migration dynamics are correlated with the local microenvironment, cell genetics, and signaling cues (Clark and Vignjevic, 2015). As in *M. xanthus* development, correlations between these cues and heterogeneity in cell response obscure the relationships between the microenvironment, cell state, and migration. Techniques for individual cell imaging and tracking in tumor models are more complex, but the resulting datasets are similar to those used here. For example, multiphoton microscopy enables tracking of individual cells *in vivo* while the second and third harmonic generation signals from the technique allow imaging of the environment, including collagen type I fibers, lipids, and lipid bodies, in the same image. Addition of fluorescent dyes, antibodies, and proteins can further enrich the dataset by concurrently providing information about individual cell state, in some cases down to individual signaling pathways (Ellenbroek and van Rheenen, 2014). Combining the microscopy and cell tracking data with simulations in which the local microenvironments are defined a-priori to identify microenvironment cues of cell behavior would be analogous to the open-loop simulations described here. In cases where datasets

contain a large number of independent variables, or if no clear hypotheses exist, statistical techniques such as correlation analysis, mutual information, or granger causality (Bastos and Schoffelen, 2016; Lock et al., 2014) could be used to generate an initial hypothesis to test in simulations. In systems that have incomplete datasets, hypothesized distributions can be integrated into the agent's behavior. Modification of what defines an agent in the simulation will be specific to each case, but is straightforward.

Application of the framework to development of *M. xanthus* identified decreased cell motility inside the aggregates, a biased walk toward aggregate centroids, alignment with neighboring cells, and cell orientation changes with respect to the aggregate boundaries as behaviors contributing to aggregation. Surprisingly, longer run durations outside of aggregates can compensate for lack of a biased walk towards aggregates (Figure 2.13 E). This observation highlights a possible compensatory mechanism that could make *M. xanthus* development especially robust. Such compensatory behaviors could mask phenotypes in traditional gene knockout experiments, particularly when relying on visual discriminators such as aggregate area or count at the end of the development. Compensation by modulating run durations is a particularly enticing mechanism since *M. xanthus* contains 21 chemoreceptors, of which 13 create altered developmental phenotypes when deleted (Moine et al., 2014), and 2 are thoroughly implicated in both development and reversal control (Curtis et al., 2006; Kearns and Shimkets, 1998; Xu et al., 2008). Furthermore, these cell reversal control pathways can react in timescales of minutes (Kearns and Shimkets, 1998) instead of the longer timescales required for protein level changes. The active role of chemoreceptors in development also suggests the ability to sense chemical gradients, which agrees well with the identification of a biased walk towards aggregates. However, given that no developmental signals have been found yet to guide aggregation, and considering the evidence of contact-mediated reversal control, further studies are needed to unmask the biological mechanisms

of the salient cell behaviors.

This approach could speed up physiological analyses of strains containing genetic deficiencies by applying the same framework to analyze the behavior of fluorescently labeled mutant cells. Open and closed-loop simulations can then be utilized to test whether behavioral differences observed in mutant cells affect aggregation and predict whether these differences compensate for the lack of another behavior. This approach creates a clear path of combining data acquisition with simulations to formulate hypotheses for future rounds of experiments. In this way, the framework can be used to move from a coarse-grain understanding of the behaviors to mechanistic understanding of how cellular machinery, signals, and physical integrations guide emergent cell behaviors.

Methods

Bacterial strains, plasmids and growth conditions

All strains and plasmids used in this study are listed in Supplemental Table S1. *M. xanthus* strains were grown in CYE broth (1% Bacto casitone (Difco), 0.5% yeast extract (Difco), 10 mM 4-morpholinepropanesulfonic acid (MOPS) (pH 7.6), and 0.1% MgSO4) at 32°C with vigorous shaking. Development was induced on 10 ml of TPM agar [10 mM Tris HCl, pH 7.6, 1 mM $KH(H_2)PO_4$, pH 7.6, 10 mM $MgSO_4$, 1.5% agar (Difco)] containing 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 100 μM vanillate in 100 mm diameter petri dishes. pLJS145 was constructed by PCR cloning tdTomato from ptdTomato with primers containing 3' XbaI and 5' KpnI restriction sites and ligated into pMR3487 (Iniesta et al., 2012). pCRC36 was constructed by PCR cloning the eYFP from pEYFP with primers containing 3' NdeI and 5' NheI restriction sites and ligated into pMR3629 (Iniesta et al., 2012). Strains LS3629 and LS3908 were constructed by electroporation (Kashefi and Hartzell, 1995) of plasmids pCRC36 and pLJS145, respectively. Following electroporation, transformants were selected on CYE 1.5% agar plates containing 50 μg/ml kanamycin for pCRC36 or 15 μg/ml oxytetracycline for pLJS145.

Fluorescence time-lapse image capture

Strains LS3908 and LS3629 were grown to exponential phase, mixed 1:2500 (resulting in approximately 500 individually trackable tdTomato cells within the field of view), concentrated to $1.7x10^9$ cells/ml, and 35 μl of the cell mixture was spotted onto TPM agar then dried uncovered in a 32°C incubator. Once dry, the plates were covered, wrapped with parafilm (Bemis Inc., Neenah, WI), and incubated in a heated room. Room temperature varied between 27°C and 29°C, averaging 28°C. Time-lapse images of the spots were acquired using a Leica DM5500B microscope (Leica Microsystems, Wetzlar, Germany) in the same heated room beginning at the indicated times in the TRITC channel at 200x magnification every 30 seconds; a short enough time frame that cells do

not move more than one cell length between images. Data capture was performed using a Flash2.8 (Hamamatsu Photonics, San Jose, CA) camera, a Phoenix-D48CL frame grabber (Active Silicon, Severna Park, MD), and the μManager software (Edelstein et al., 2014). The fluorescence intensity was set to 55%, camera gain set to 255, and exposure time was 600 ms. The mercury lamp was shuttered when not acquiring an image. Imaging was carried out for approximately 6 hours. The time point at which aggregation began varied by up to one hour between replicates. Replicate movies were truncated to synchronize the onset of aggregation and equalize movie length as described in the Supplemental Text, resulting in final movie length of 5 hours. Three replicate movies were created and analyzed in parallel as described below.

Cell Density Estimation

To account for uneven illumination from the microscope's mercury bulb and optics, acquired fluorescent images were normalized to the intensity of the first frame. Images were then Gaussian smoothed to filter the contribution from the individual labeled LS3908 cells and the images were normalized for diminishing fluorescence over the length of the movie by subtracting the mean intensity of each frame. To estimate cell density, the detected cell positions (as described in Cell Tracking) in the last image from each experimental replicate were used to estimate the cell density using a kernel density estimator. Comparing the computed cell density with fluorescence intensity values from the last frame indicated a nonlinear correlation between the two (Figure 2.14 A). To relate these two estimates of cell density (kernel-density and fluorescence-intensity density), a third-degree polynomial was fitted to the data pooled from all three movies using MATLAB's fit function with the robust option set to Bisquare (Figure 2.14 A, red line). The fitted polynomial was used to convert fluorescence-density values to cell densities for all images. Further details for the filters and chosen parameters used are provided in the Supplemental Text.

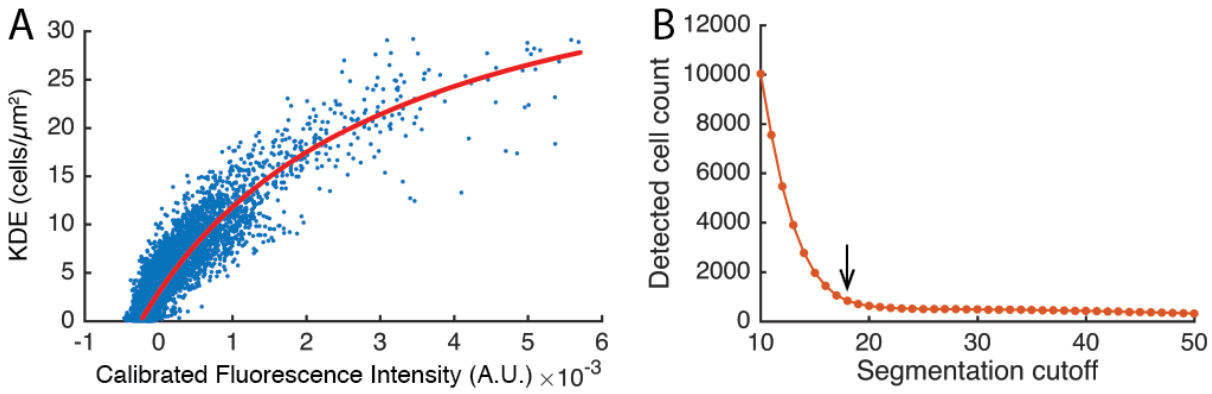

Figure 2.14: Cell density estimation and cell detection parameters. A) Correlation between individual pixel fluorescence intensity in arbitrary units (A.U.) plotted against cell density estimated using a kernel density estimator (KDE) at the same position as the pixel. Red line indicates the fitted line used to relate the two. B) Fluorescence intensity cutoff for cell detection was chosen as the bottom of the "elbow" created when plotted number of cells detected vs. the intensity cutoff. Arrow indicates the cutoff chosen for use in segmentation.

Cell Tracking

To reduce camera sensor noise and fluorescence from the growing aggregates, time-lapse images were band-pass filtered as described in the Supplemental Text. Thereafter, MATLAB function regionprops was used to identify the centroid and orientation of each cell. The segmentation threshold value was chosen by running the segmentation and detection on the first image with threshold values between 10 and 50 in 1 unit increments. Plotting the threshold values versus the number of cells detected (Figure 2.14 B) indicated that the cell count approaches a constant value as the threshold rises above the noise caused by background fluorescence. Visual inspection of the cell detections indicated a threshold value on the edge of the "elbow" (Figure 2.14 B, arrow) provided a good tradeoff between detection of all the cells and little identification of background noise as cells.

To track cell motility between images, we followed procedures established in (Jaqaman et al., 2008). This technique solves the problem of image-to-image linking of detected cells into trajectories by treating the assignments as a linear assignment problem (LAP). In this method, cells are assumed to move, disappear, or appear between two consecutive images. In the move case, a cell will move to a new position in the time between images. Therefore, its positions in the two images should be linked into the same trajectory. If a cell disappears due to leaving the field of view, misdetection, or overlapping with another cell, it should not be linked to a cell in the later image. In a similar fashion, a cell that appears in the later image should constitute a new trajectory. The LAP involves calculating a cost to assigning each of these actions for every cell in the two images. The resulting costs are then used to find an optimal assignment for each cell by minimizing the total cost of assigning all cells to one of the three options. The process is then repeated consecutively for each image from the time-lapse acquisition. We used the Jonker-Volgenant Algorithm (Jonker and Volgenant, 1987) implemented in MATLAB by (Jaqaman et al., 2008) to

solve the LAP (see Figure 1 of (Jaqaman et al., 2008) for an overview of this process). As in (Jaqaman et al., 2008), a second LAP was then performed to relink broken trajectories. A full definition of the cost functions used for linking cells based on the properties of *M. xanthus* motility is described in the Supplemental Materials.

Cell State Detection

Confidently detecting whether a cell is actively moving, stopped, or reversing direction is complicated by noise in the cell trajectories. This noise arises from inaccuracies in detecting the cell position due to the low acquisition magnification and the biological processes that lead to cell movement. We observed that this variability created cell trajectories too noisy for one-dimensional detection techniques (e.g. using tangential speed to detect reversals or a speed cutoff to detect non-moving cells). To detect movement characteristics of the cell reliably, a cell state filter was developed which employs extended Kalman filters (EKF) to estimate the most probable motion model used by the cell between images.

We assume cells use the same movement models as described for cell tracking: persistent forward ($i = 1$), persistent backwards ($i = 2$), and non-persistent ($i = 3$). The EKF estimates state vector $s_t = [x_t, y_t, \mathrm{v_t}, \theta_t]$ consisting of the position (x, y), orientation along the long axis (θ), and speed (v) of the cell in image *t* using the *t-1* state vector and one of the three movement models (f_1 to f_3 in Table S2). The EKF then uses the deviation between the predicted (s_t) state and true cell state in image *t* to calculate the likelihood that each movement was executed by the cell. The model with the maximum likelihood was then assigned as the movement between the two images. A

detailed description of the movement models and EKF algorithm is provided in the Supplemental Materials.

Aggregate Detection and Tracking

A cell density cutoff of 2.32 cells/μm^2 was chosen by visual inspection of aggregate boundaries in the last image of each movie. Aggregates were detected in each frame as areas where cell density exceeded the cutoff. Aggregate boundaries were approximated as ellipsoids with a centroid, major axis, and minor axis calculated using MATLAB's regionprops function. To track aggregate positions from image to image, a LAP was set up similar to that used for cell tracking. Adaptions made to track aggregates are provided in the Supplemental Text.

Run Vectors

Trajectories were divided into runs, which start at the beginning of one contiguous movement state (persistent forward, persistent backwards, non-persistent) and end with the next change of state. Trajectory data prior to the first reversal and after the last reversal were discarded. The average speed (v), period (τ), distance (δ), angle to the nearest aggregate centroid (ϕ), distance to the nearest aggregate boundary (D), ambient cell density (ρ), turning angle (θ), average nematic orientation (explained below) of neighboring runs (γ), and time since beginning of the experiment (T) were calculated for each run vector (Figure 2.1 C).

Average nematic alignment strength was used to quantify trajectory alignment (Figure 2.11 A, solid boxes) at the level of a run. The average nematic alignment strength, denoted as $< \Omega_n >$, is calculated as the average cosine difference between the orientation of run *n* and all runs within a window size of ±5 min and 15 μm radius around the start of run *n*:

$$< \Omega_n > = \frac{1}{N} \sum_{i \in window} \cos(2 * (\chi_n - \chi_i)). \quad [1]$$

In Eq. [1], *N* is the number of runs within the window and χ is the angle of the run relative to the *x*-axis. Due to the lack of motility polarity, the run bearing χ is in the interval $[-\pi, \pi)$ where $-\pi =$

π. Choosing the window size required balancing an *N* large enough to reliably evaluate Eq. [1] while avoiding smoothing out local alignment characteristics. Visual inspection of the trajectories indicated that alignment was stable in time (Figure 2.11 A), allowing the window to be extended in the time dimension to increase N while keeping the spatial search radius around the cell small. The search radius and time window length were chosen by searching the parameter space of possible values and choosing the combination of values that provided the greatest average alignment strength (Figure 2.15 A&B).

Bootstrapping Statistics

Where indicated, 95% confidence intervals were calculated by pooling the data from all 3 replicate movies and bootstrapping parameters using the adjusted percentile method (Davison and Hinkley, 1997) with 1000 bootstrap samples.

Data-driven Agent-Based Model

An agent-based model consisting of 10,000 agents on a rectangular domain of 986 μm x 740 μm, equal to the microscope field of view, with periodic boundary conditions on each end, was implemented in MATLAB. Each agent represents a single cell sampled from a biofilm of the same average density as experiments (1.1 cells/μm^2), similar to sampling cell behaviors in the biofilm using a small number of fluorescently labeled cells. The random trajectory of a single agent consists of the sequence of reversal locations $(x_i\,, y_i)$, and bearing angles, χ_i, connected by run vectors, $(\Delta x_i\,, \Delta y_i)$, and turning angles, θ_i, beginning at time points T_i. The run vector $(\Delta x_i\,, \Delta y_i)$ is constructed from χ_i, a run speed, v_i, and a run duration τ_i. Since fluorescent images for cell tracking were taken at 30-sec intervals, we have adopted the same time discretization in the simulations with agent's positions along their current run vector updated every $(\Delta t = 30s)$. The agents run variables $(\theta_i,\ v_i\,, \tau_i)$, along with an auxiliary binary variable denoting if the run is persistent or non-persistent, $s_{i,}$ are drawnfrom the reversal PDF, $P(\theta_i, v_i\,, \tau_i\,,\ s_i|\ T_i,\ D_i\,,\ \rho_i,\ \phi_{i-1}, \gamma_i)$. Where T_i is the

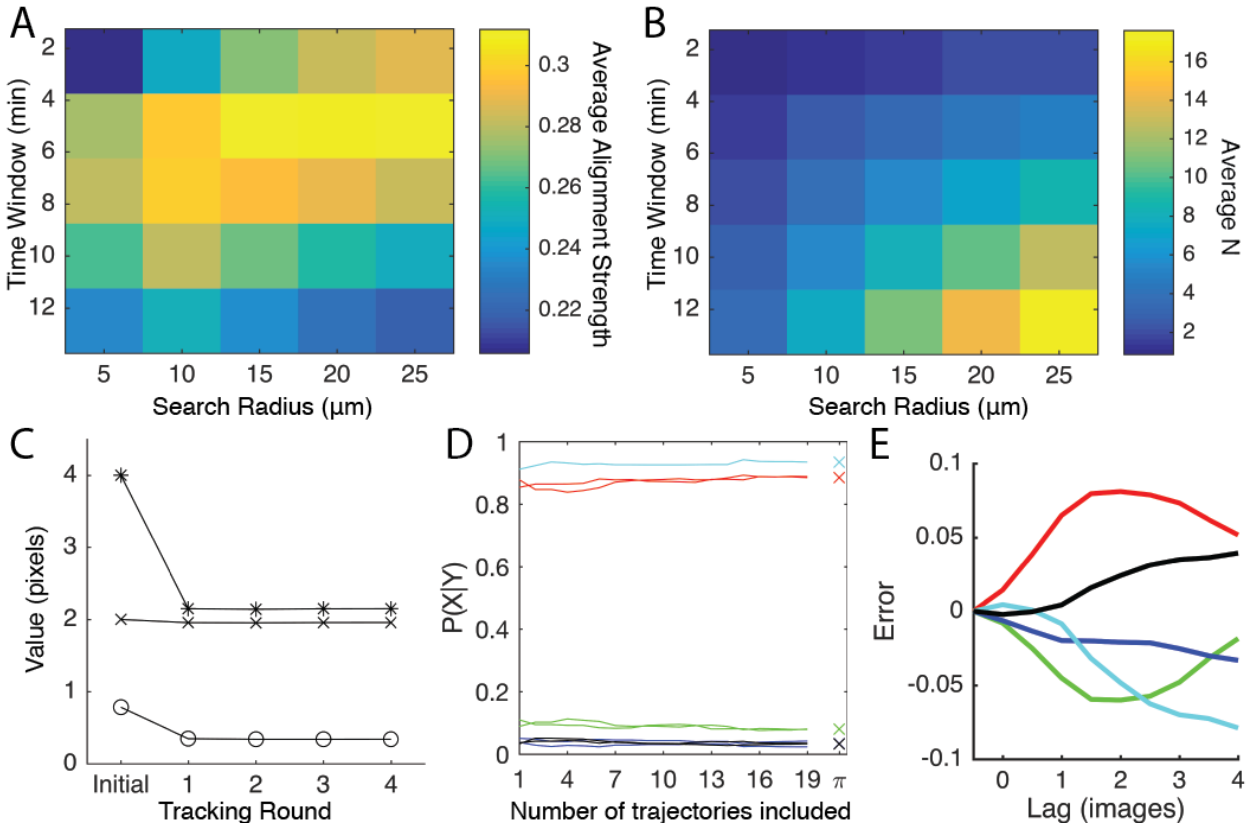

Figure 2.15: Cell tracking and quantification parameter estimation. A&B) Parameter search for alignment window length and size. A) Mean nematic alignment strength for all runs when calculated using the given time window and search radius. B) Mean number of runs used to calculate the nematic alignment strength for each run. C) Progression of variable values for cell tracking calculated using bootstrapping. Stars: standard deviation of the difference between predicted and measured cell position (σ_δ); Open circles: standard deviation of the difference between predicted and measured cell position ($\sigma_{\Delta\theta}$); Crosses: mean difference between predicted and measured cell position in x and y directions. D) Transition probabilities estimated using trajectories from manually assigning trajectory segments as either forward, reverse, or non-persistent movement models. Lines of the same color indicate transition probabilities calculated for forward and backwards movement models for continuing persistent movement (red), reversing direction (green), transitioning from non-persistent to persistent (black), and transitioning from persistent to non-persistent (blue). Cyan line indicates probability of continuing non-persistent movement. Crosses indicate values used in transition matrix π. E) Error between the transition probabilities estimated using trajectories with manually assigned movement models and that estimated using a Markov chain after a lag of t images. Colors indicate continuing persistent movement (red), reversing direction between the two persistent types (green), transitioning from non-persistent to persistent (black), transitioning from persistent to non-persistent (blue), continuing non-persistent movement (cyan).

time since the beginning of the experiment, ρ_i is the local cell density, γ_i is the angle between the cell orientation and the average bearing angle of neighboring runs, and D_i and ϕ_{i-1} are defined in Figure 2.1 C. We used nearest neighbor methods (Hastie et al., 2009), to estimate P by drawing θ_i, a paired (v_i, τ_i), and s_i from experimentally observed runs conditional on $(T_i, D_i, \rho_i, \phi_{i-1}, \gamma_i)$ as described in the Supplemental Materials. This approach incorporates directly from the experimental run database all the information available about P without relying on an explicit re-construction of P on a high-dimensional variable space.

We implemented two alternative modeling approaches, referred to as the *open-loop* and the *closed-loop* model, which differ in how the local cell density (ρ_i) at location (x_t, y_t) and time t was modeled. In the open-loop approach, we used the observed density profile and aggregate locations extracted from each of the three fluorescent and trajectory imaging datasets (movies), as described in the cell density estimation section. In the closed-loop approach, agent positions were initialized from a uniform random distribution. Each time-step, ρ_i was extracted from the current agent positions with a KDE bandwidth of 14 μm. A 14 μm bandwidth provided good agreement between the starting density distributions of the agents and that measured from experimental results (Figure 2.16). Aggregate boundaries and centroids were then calculated from the estimated density profiles, ρ_i in the same manner as for the experimental imaging density data.

The database of experimentally observed runs used to estimate P can comprise of the composite of all runs extracted from all trajectories tracked across all three microcinematography experiments (movies 1-3) reported here, with N_O= 102,972 or else the database may consist only of the runs from all trajectories tracked in each microcinematography movie, with N_O= 36,019, 36,303 or 30,650, respectively. The composite database was used only for the closed-loop simulations. Each open-loop simulation utilized only the single-experiment database for the imaging experiment from which also the input cell density profile was extracted.

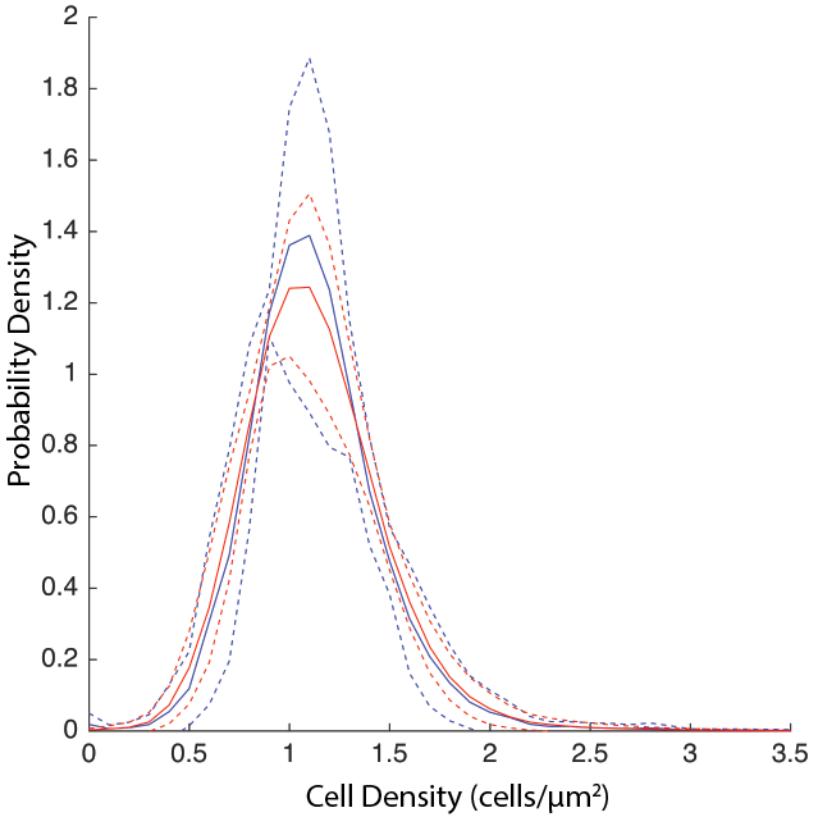

Figure 2.16: Comparison of distribution of biofilm cell densities seen at each (x ,y) position within the field of view in the first image of the experimental replicates (blue) or at the beginning of open-loop simulations (red). For experiments, the density was estimated from the fluorescent images. The blue solid lines indicate mean and dashed lines indicate standard deviation of the density distribution from the experimental replicates. For simulations, the (x,y) location of each of the 10,000 agents was drawn from a 2D probability density generated from the density profile of the first image of an experimental movie. The biofilm cell density was then estimated from the agent positions using a kernel density estimator with a bandwidth of 14 μm. 13 simulation distributions were generated per experimental replicate. The red solid lines indicate mean and dashed lines indicate standard deviation of the density distribution from the replicates.

In open-loop simulations, three independent open-loop simulations were performed for each experimental movie. In the closed-loop, three simulations were performed. Each simulation started from a different random initial configuration of agents. The results from the replicate simulations were then pooled for the subsequent data analyses.

Supplemental Methods

Alignment of Experimental Replicate Movies

The time point at which aggregation began in experiments varied by one hour between replicates. To normalize for timing variation, the fraction of tdTomato cells inside (cell density > 2.32 cells/μm^2) the aggregates (F_t) was calculated for each image (t) in the video. These counts were then normalized using

$$F_t^* = \frac{F_t - \max_{t\in N} F_t}{\max_{t\in N} F_t - \min_{t\in N} F_t} \quad [2]$$

where N is all images in the experimental replicate. The normalized F_t^* counts created curves which spanned from 0 before aggregation began to approximately 1 after aggregation stabilized. The midpoint of aggregation was identified using the F_t^* curves by finding the value of x that minimized the squared error between the F_t^* curve and the function

$$f(t) = \begin{cases} 0 & if\ t < x \\ 0.5 & if\ t = x \\ 1 & if\ t > x \end{cases}, \quad [3]$$

where t is the image time index. To align the replicates, the first frame of the experimental replicate with the minimum x was assigned the time point 0. The beginning of all other replicates were then truncated so that their x was equal to the minimum x. The ends of the aligned replicate were then truncated so that all movies were the same length as the shortest replicate. The final replicate movie length was 590 frames (approximately 5 hours).

Cell Density Estimation

To account for uneven illumination from the microscope's mercury bulb and lens optics, acquired fluorescent images were normalized by dividing each pixel intensity by the intensity of the corresponding pixel in a calibration image. The calibration image was created by taking the average intensity for each pixel from the first 15 frames. The resulting image was then smoothed using MATLAB's (Mathworks, version 2015b) imfilter function with the replicated boundaries option. The mean input filter for imfilter was generated by MATLAB's fspecial function with a radius of 500 pixels. After the illumination-normalization, the images were smoothed using imfilter and an fspecial generated Gaussian filter with a radius of 30 pixels. This smoothing filtered the contribution from the individual labeled (LS3908) cells in the images. The filtered images were then normalized for diminishing fluorescence over the length of the movie by subtracting the mean intensity of each frame.

To estimate cell density, the detected cell positions (as described below) in the last image from each experimental replicate were used to estimate a probability-density function using a kernel density estimator (KDE) described in (Botev et al., 2010). Version 1.3 of the MATLAB function written by the authors was acquired from the MathWorks File Exchange (File ID #17204) and modified to allow for a manually set bandwidth of 23.3 μm. This bandwidth was chosen as the average of all three experiments estimated as described in (Botev et al., 2010). To create a cell density estimate, the probability density function was multiplied by the estimate of the number of cells in the microscope field of view (FOV). Assuming a uniform distribution of cells within the biofilm, a constant number of cells, and that no colony expansion occurs in the experiment, the number of cells was estimated as

$$\#\text{Cells in FOV} = \frac{\text{FOV area}}{\text{Total area}} \times \text{Total \#cells}$$

$$= \frac{7.3\times10^5\ \mu m^2}{5.3\times10^7 \mu m^2} \times 1.7\text{x}10^9 \frac{\text{cells}}{\text{ml}} \times 0.035\ \text{ml} \quad [4]$$

$$= 8.2\times10^5\ \text{cells.}$$

Here the total area was computed as $\pi r^2 = 5.3x10^7 \mu m^2$, where r is the estimated the average radius of the 5 spots measured after drying. The cell density estimate was used to covert the normalized fluorescent intensity values to cell densities as described in the main text.

Bandpass Filter

A bandpass filter was utilized to better identify individual fluorescent cells in each frame by removing high frequency pixel noise from the camera sensor and low frequency changes in fluorescent due to the growing aggregates. The bandpass filter consisted of separately convoluting the image with a Gaussian with a standard deviation of 10 pixels (~5 μm, the approximate cell size) and a boxcar function with a width of 1 pixel. To produce the final filtered image, the boxcar filtered image were subtracted from the Gaussian filtered image (Crocker and Grier, 1996).

Cell Tracking

To track visually indistinguishable cells from image to image, we need to formulate a cost function for linking cells in consecutive frames which relies on the properties of *M. xanthus* motility. To this end, we assume cells use one of three movement models: persistent forward, persistent backwards, or non-persistent. The cost of using each of the models in the LAP is then calculated by measuring the difference between the detected cell positions (x_{t+1}, y_{t+1}) and orientation along the long axis of the cell (θ_{t+1}) and those predicted by the movement models (x^*_{t+1}, y^*_{t+1}, θ^*_{t+1}). The predicted positions of each cell are computed from their position in image *t* as follows:

$$\begin{bmatrix} x^*_{t+1} \\ y^*_{t+1} \end{bmatrix} = \begin{bmatrix} x_t \\ y_t \end{bmatrix} + C_m \begin{bmatrix} \delta x_t \\ \delta y_t \end{bmatrix}. \quad [5]$$

Here C_m is a coefficient specific to the movement model and is 1 for forward, -1 for backwards, and 0 for non-persistent movement. For cells that were tracked in the preceding image, $\delta x_t \; and \; \delta y_t$ are displacements in the previous time interval, i.e.:

$$\begin{aligned} \delta x_t &= x_t - x_{t-1} \\ \delta y_t &= y_t - y_{t-1}. \end{aligned} \qquad [6]$$

For cells that first appear in the image at time *t*, the δx_t and δy_t are estimated based on the orientation of cell major axis and mean-square-displacement of all tracked cells as follows

$$\begin{aligned} \delta x_t &= \left\langle \sqrt{(x_t - x_{t-1})^2} \right\rangle \cos(\theta_t) \\ \delta y_t &= \left\langle \sqrt{(y_t - y_{t-1})^2} \right\rangle \sin(\theta_t). \end{aligned} \qquad [7]$$

Here the angle brackets indicate the average from all cell links assigned in *t-1* to *t* images. If no trajectories contain a *t*-1 position, the averages were substituted with an alternate constant chosen using the bootstrapping technique discussed below. The orientation of cells is assumed to vary little between frames and is thus predicted as $\theta^*_{t+1} = \theta_t$.

The deviations between the measured and predicted cell positions (δ_{xy}) and orientations (δ_θ) each make contributions to the cost of linking cells into the same trajectory. By assuming the deviations are independent and normally distributed, the cost is calculated as

$$-\log\left(P(\delta_{xy}, \sigma_{\mathrm{xy}})P(\delta_\theta, \sigma_\theta)\right). \qquad [8]$$

Here

$$\begin{aligned} \delta_{xy} &= (x_{t+1} - x^*_{t+1})^2 + (y_{t+1} - y^*_{t+1})^2 \\ \delta_\theta &= \frac{1}{2} atan2(\sin(2(\theta_{t+1} - \theta^*_{t+1}), \cos(2(\theta_{t+1} - \theta^*_{t+1}))) \end{aligned} \qquad [9]$$

where atan2 is MATLAB's four-quadrant inverse tangent function and $P(x, \sigma)$ is Gaussian with 0 mean and standard deviation σ (Challa et al., 2011):

$$P(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}. \quad [10]$$

The standard deviations were calculated using bootstrapping methods discussed below. The cost of linking in Eq. [8] is calculated for each movement model for each cell pair between image *t* and *t*+1. The minimum of the costs among the three movement models is assigned as the cost of linking the cell pair into the same trajectory. The costs associated with a trajectory ending or beginning were calculated as described in (Jaqaman et al., 2008).

Cells may be misidentified for a short time (1 to 5 images) due to their overlap with another cell or due to segmentation errors. This can lead to the movements of the same cell being split into multiple trajectories. To address this, we again follow the work of (Jaqaman et al., 2008) and develop a second LAP to connect split trajectories. In this LAP we assume the end of a trajectory could be split due to the errors discussed above or could be a true ending or beginning due to cells entering or leaving the FOV. The cost of assigning the beginning and ending of each trajectory to one of these possibilities was calculated as described below and used in a LAP to find the optimal combination of assignments.

The cost of linking the end of one trajectory with the beginning of another consists of contributions from the distance (δ), change in cell orientation (θ), the angle enclosed between the orientation of the cell and a vector connecting (ϕ), and the time (τ) between the end of one trajectory and beginning of the other trajectory to be linked. We assume the contributions are independent from each other, allowing the cost to be calculated as

$$-\log\big(p_\delta(\delta,\tau)p_\phi(\phi,\tau)p_\theta(\theta,\tau)p_\tau(\tau)\big). \quad [11]$$

In Eq. [11], $p_x(x,\tau)$ was calculated from normalized histogram for each length τ using the previously linked trajectories. The number of bins in the normalized histogram was chosen using the Freedman-Diaconis rule (Freedman and Diaconis, 1981). The average gap length was assumed to

be 1 and was generated from a Poisson distribution with $\lambda = 1$. Only gaps lengths (τ) less than 6 images were considered for closing.

The cost associated with not linking trajectories together was calculated as in (Jaqaman et al., 2008). The resulting LAP was solved as discussed previously. Trajectories that spanned less than 5 minutes (10 consecutive images) were then discarded.

Bootstrapping Unknown Tracking Parameters

Since little data on the behavior of cells inside the biofilm exists, we bootstrapped the unknown standard deviations in Eq. [8] and the alternative displacement for Eq. [7] from the tracking itself. This was done iteratively by performing the tracking, calculating the required variables from the results, and then using them in the next round of tracking. Values for the first round of tracking were chosen based on visual inspection of the time lapse images. These values were $\sigma_{\theta_{t+1},\theta^*_{t+1}} = \frac{\pi}{4}$ radians and $\sigma_{(x_{t+1},y_{t+1}),(x^*_{t+1},y^*_{t+1})} = 4$ pixels for Eq. [10], and 2.5 pixels for the alternative cell displacement used in Eq. [7]. For subsequent tracking rounds the values were generated from the trajectories resulting from the previous round. The iterative tracking continued until the deviations between rounds was less than 1%. This convergence required less than 5 rounds of iteration (Figure 2.15 C). The values from the 5th round were used in the tracking.

Cell State Detection Movement Models and Detection

Given a set of state vectors $S^t = (s^*_{t+1}, s_t, s_{t-1}, \dots, s_1)$ represeting the movement states of a cell up to frame t plus an estimated t+1 state (s^*_{t+1}), the probability of each movement model (M_t) being used between image t-1 and t can be written as

$$P(M_t = i|S^t) = \frac{1}{c} P(S^t|M_t = i)P(M_t = i). \quad [12]$$

Here c is a normalization factor assuring the probabilities of the three movement models sum up to 1. By assuming the transitions between movement models are reasonably Markovian (Figure 2.15

D, discussed below), $P(M_t = i|S^t)$ can be approximated recursively (Challa et al., 2011). Thus, $P(M_t = i)$ is approximated as

$$P(M_t = i) = \sum_{j=1,2,3} P(M_{t-1} = j)\pi_{i,j} \quad [13]$$

with transition probabilities π, who's derivation is discussed below. $P(S^n|M_t = i)$ is approximated using a Markov chain:

$$P(S^n|M_t = i) = \sum_{h=1,2,3} \sum_{j=1,2,3} P(s_{t-1}|M_{t-1} = h)\pi_{h,i} P(s_t|M_t = i)\pi_{i,j} P(s^*_{t+1}|M_{t+1} = j) \quad [14]$$

where s^*_{t+1} is estimated by the EKF using s^*_t as defined in Table S2. s^*_{t+1} augments the estimation of $P(S^n|M_t = i)$ to include available future cell state information. For each of the trajectories generated from the cell tracking, the conditional probabilities and cell state vectors in Eq.[14] were estimated using EKFs (Challa et al., 2011) with the paramaters listed in Table S2 and justified below. If an EKF predicts a movement in the opposite direction to the measurement, it was assigned a probability of 0 for that step. At each step *t,* the movement model with the maximum probability ($P(M_t = i|S^n)$ from Eq. [12]) is chosen and then used to estimate s_t for the next iteration of the EKFs.

The EKFs integrates uncertainty into the model likelihood estimation by adding noise to the movement (f_1 to f_3 in Table S2) and measurement (h in Table S2) functions. The noise is assumed to be Gaussian with a mean of zero and covariance Q for movement and covariance R measurement functions. This noise models influences on cell movement not accounted for in the functions. Typically, these would be developed *a-priori* using an understanding of how the system was measured and how process noise arises. Since little data on the behavior of cells inside the biofilm exists, we instead estimated the covariance matrix Q from the deviations between the

predicted $(x^*_{t+1},\ y^*_{t+1}, \theta^*_{t+1})$ and measured $(x_{t+1},\ y_{t+1}, \theta_{t+1})$ variables from the cell movement tracking (see Eq. [5]). Q was generated from these deviations as

$$Q_m = \begin{bmatrix} <\epsilon_x\epsilon_x> & <\epsilon_y\epsilon_x> & <\epsilon_v\epsilon_x> & <\epsilon_\theta\epsilon_x> \\ <\epsilon_x\epsilon_y> & <\epsilon_y\epsilon_y> & <\epsilon_v\epsilon_y> & <\epsilon_\theta\epsilon_y> \\ <\epsilon_x\epsilon_v> & <\epsilon_y\epsilon_v> & <\epsilon_v\epsilon_v> & <\epsilon_\theta\epsilon_v> \\ <\epsilon_x\epsilon_\theta> & <\epsilon_y\epsilon_\theta> & <\epsilon_v\epsilon_\theta> & <\epsilon_\theta\epsilon_\theta> \end{bmatrix}. \quad [15]$$

Here ϵ is the deviation between the predicted and measured t+1 values that resulted in a linking assignment using model m in the LAP, and angle brackets indicates the mean. The deviation in the cell speed was calculated as

$$\epsilon_v = \frac{1}{\Delta t}\sqrt{\delta_{xy}} - \sqrt{\left(x_t - x_{t+1}\right)^2 + \left(y_t - y_{t+1}\right)^2} \quad [16]$$

where δ_{xy} is from Eq. [9] and Δt is the time between images. Seeing as Q was calculated directly from the trajectories, which include any measurement noise, the calculation was simplified by setting the measurement covariance (R) to 0. Since forward and reverse models only differ in the direction of the movement, their deviations were pooled to create a single matrix used for both their EKFs.

The transition matrix π was generated by manually assigning the movement model for each step from 19 randomly chosen trajectories. π was then calculated from these trajectories. 19 trajectories were sufficient for the probabilities to stabilize to within +/- 0.005 per trajectory added (Figure 2.15 D). We assumed the transitions for persistent forward and backwards models were equal. This was asserted by pooling the forward and reverse transition data and calculating one set of transitions for both models (Figure 2.15 D, crosses). Model transitions were confirmed to be reasonably Markovian by comparing the probability of transitions after a time lag with the Markov chain of the same length with transition probabilities π (Figure 2.15 E).

Aggregate Detection and Tracking

Aggregates were assumed to keep an approximately constant centroid position (x,y) and major (a) and minor (b) axes between images, only deviating by noise. The cost of linking an aggregate in image t with an aggregate in image t+1 in the LAP was calculated as the log-likelihood of the deviation between the centroids (δ_{xy}), major (δ_a), and minor (δ_b) axis of the two aggregates:

$$-\log\Big(P\big(\delta_{xy},\sigma_{\mathrm{xy}}\big)P(\delta_a,\sigma_a)P(\delta_b,\sigma_b)\Big). \quad [17]$$

In Eq. [17] $P(x,\sigma)$ is as in Eq. [10] and the deviations were calculated as:

$$\begin{aligned}\delta_{xy} &= (x_t - x_{t+1})^2 + (y_t - y_{t+1})^2\\ \delta_a &= a_t - a_{t+1}\\ \delta_b &= b_t - b_{t+1}.\end{aligned} \quad [18]$$

Since aggregates are reasonably well spaced and move or grow little between images, precise values of $\sigma_{\mathrm{xy}}, \sigma_a, \sigma_b$ were not vital for accurate tracking. Thus, σ_{xy} was set to 10 μm and σ_a, σ_b to 25 μm and the resulting trajectories were visually inspected to confirm fidelity.

The stability of the aggregates also allows forgoing the second LAP round used in cell tracking. Instead, aggregates in image t that were not linked to an image in t+1 were propagated to image t+1 with the same centroid, major, and minor axis. Propagation was allowed to continue for up to 5 consecutive images. If a propagated aggregate was not linked to a detected aggregate within 5 images the trajectory was ended at the last frame the aggregate was detected.

This study focused only on stable aggregates; defined as aggregate that were present at the end of the experiment (compare green and red ellipses in Figure 2.1 A). Any aggregate that merged into a stable aggregate was also included in the analysis. Merge events were detected as aggregate centroids in image t that ended within the boundaries of another aggregate in image t+1. All other aggregates that did not fit these criteria were discarded from the analysis.

Data-Driven Agent-Based Model

The agent positions, (x_t, y_t) are updated every $\Delta t = 30s$ using:

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} v_i \, \Delta t \cos(\chi_i) \\ v_i \, \Delta t \sin(\chi_i) \end{bmatrix}. \quad [19]$$

Here χ_i is the orientation angle of the agent generated from the orientation of the previous run (χ_{i-1}) and the turning angle of the current run by:

$$\chi_i = \chi_{i-1} + \theta_i \ . \quad [20]$$

Note that subscript *i* denotes the current run, and is only incremented at the end of each run when an agent chooses new run variables (θ_i, v_i , τ_i, and s_i, defined below) whereas *t* denotes simulation time, and is incremented at each simulation time-step.

The run variables (θ_i, v_i , τ_i), along with an auxiliary binary variable denoting if the run is persistent or non-persistent, s_i, are drawn from the conditional reversal PDF of the general functional form

$$P(\theta_i, v_i \, , \tau_i \, , \, s_i | \ T_i, \ D_i \, , \ \rho_i, \ \phi_{i-1}, \gamma_i). \quad [21]$$

In other words, P is assumed to be conditional upon variables $T_i, D_i, \rho_i, \gamma_i$ and ϕ_{i-1}, where T_i is the time since the beginning of the experiment, ρ_i is the local cell density, γ_i is the angle between the cell orientation and the average bearing angle of neighboring runs, and D_i and ϕ_{i-1} are defined in Figure 2.1 C. To calculate γ_i we introduce ω_n, the average bearing angle of neighboring runs at the location of the end of run *n-1*. ω_n is evaluated using the same window as in Eq. [1] using:

$$\omega_n = \frac{1}{2}\text{atan}(\textstyle\sum_{i \in window} \sin(2 * \chi_i) \, , \sum_{i \in window} \cos(2 * \chi_i)), \quad [22]$$

$\gamma_n$is then the smaller of the two angles between $\chi_{n-1}$and ω_n.

Given the above definitions of agent behavior, agent positions for each time step are calculated using the following 4 steps for each of the 10,000 agents:

Step 1: Simulation Initialization: The agent's initial position, $(x_1\ ,\ y_1)$, is chosen. In closed-loop models, $(x_1\ ,\ y_1)$ is drawn uniformly from the rectangular simulation domain. In open-loop models, $(x_1\ ,\ y_1)$ is drawn from a 2D probability density generated from $\hat{\rho}(x, y, 1)$ from the corresponding experimental movie. The run index is initialized to $i = 1$, and a random initial bearing angle χ_1, is drawn uniformly from the interval $[-\pi, +\pi)$.

Step 2: Choose Run Variables: In the closed-loop case, the initial density profile, $\rho(x, y, t)$, is generated from the positions of all 10,000 agents. In the open-loop case, $\rho(x, y, t)$ is the density profile from image t of respective experiment. $\rho_i = \rho(x_t, y_t, t)$ is determined for each agent from the agent's position, (x_t, y_t). $\rho(x, y, t)$ is used to determine aggregate locations, if any, including their centroids and boundaries. Angle χ_i, $(x_t\ ,\ y_t)$, and any detected aggregate centroid locations are used to calculate, D_i and ϕ_i. In the event that $\rho(x, y, t)$ does not yet exhibit any aggregates, D_i and ϕ_i are left undefined. γ_i is calculated as described above (Eq. [22] and surrounding text).

Given (x_t, y_t), $\chi_{i-1}, T_i = t,\ D_i\ ,\ \rho_i,\ \gamma_i, \phi_{i-1}, \gamma_i$ for run index i, a random $(\theta_i, v_i\ , \tau_i, s_i)$ from is drawn from $P(\ \theta_i, v_i\ , \tau_i, s_i |\ T_i,\ D_i\ ,\ \rho_i,\ \phi_{i-1}, \gamma_i\ )$, using the experimental trajectory-based conditional drawing procedure described below. In cases where D_i, and $\phi_{i-1}$are undefined, due to the absence aggregates in $\rho(x, y, T_i)$, or if no runs have occurred within the last 5 minutes of simulation time and 15 μm of the agent's position to calculate γ_i, their respective conditionalities were not enforced in the experimental trajectory-based drawing procedure. θ_i was used to calculate χ_i as described in Eq.[20].

Step 3: Advance the Simulation: v_i and χ_i were used to calculate (x_t, y_t) using Eq. [19] for time steps $to\ t = t + \tau_i$.

Step 4: Checking for Run Termination: If $t \leq\ T_i + \tau_i$, step 3 was repeated on the next simulation time step. If $t >\ T_i + \tau_i$, the run index for that agent was advanced from i to $\ i + 1$, $(x_{i+1}, y_{i+1},\ \chi_i,$

T_{i+1}, D_{i+1} , ρ_{i+1} , ϕ_i) were relabeled as (x_i, y_i, χ_{i-1}, T_i, D_i , ρ_i , ϕ_{i-1}), and Step 2 was repeated on the next step. Steps 2 through 4 were repeated until t reached the simulation termination time.

Choosing an agent's next run behaviors: For later reference, we need to determine the overall range spanned by the state variables in $\hat{q}_n$ across the entire database, where $\hat{q}_n \equiv (\hat{\theta}_n, \hat{v}_n, \hat{\tau}_n, \hat{s}_n, \hat{T}_n, \hat{D}_n, \hat{\rho}_n, \hat{\phi}_{n-1}, \hat{\beta}_n, \hat{\gamma}_n)$ for observed runs labeled by $n = 1, \ldots N_O$, with N_O denoting the total number of observed runs in the data base. Here and later, the hat (ˆ) is used to denote variables derived from the microcinematography movies when confusion between microcinematography and simulation derived variables may exist. We also explicitly define β_n, the angle enclosed between ϕ_{n-1} and θ_n (Figure 2.1 C). For non-angular, continuous state variables, these ranges are defined as:

$$\begin{aligned} T_{\mathrm{R}} &= \max_{n,m=1,\ldots N_{\mathrm{O}}} |\hat{T}_n - \hat{T}_m| \\ D_{\mathrm{R}} &= \max_{n,m=1,\ldots N_{\mathrm{O}}} |\hat{D}_n - \hat{D}_m| \\ \rho_{\mathrm{R}} &= \max_{n,m=1,\ldots N_{\mathrm{O}}} |\hat{\rho}_n - \hat{\rho}_m| \end{aligned} \qquad [23]$$

Since angular variables are only defined *modulo* 2π, we restricted each of them to the interval $(-\pi, +\pi]$, before taking their differences. Subject to that modification, the ranges of the angular state variables are then defined by

$$\begin{aligned} \beta_{\mathrm{R}} &= \max_{n,m=1,\ldots N_{\mathrm{O}}} |\hat{\beta}_n - \hat{\beta}_m| \\ \theta_{\mathrm{R}} &= \max_{n,m=1,\ldots N_{\mathrm{O}}} |\hat{\theta}_n - \hat{\theta}_m| \\ \phi_{\mathrm{R}} &= \max_{n,m=1,\ldots N_{\mathrm{O}}} |\hat{\phi}_{n-1} - \hat{\phi}_{m-1}| \\ \gamma_{\mathrm{R}} &= \max_{n,m=1,\ldots N_{\mathrm{O}}} |\hat{\gamma}_n - \hat{\gamma}_m| \end{aligned} \qquad [24]$$

Note that the resulting range values, β_{R}, θ_{R} and ϕ_{R}, are then in fact very close to 2π, since the angles whose differences are being maximized typically cover almost their entire allowed range from $-\pi$ to $+\pi$.

The approach also allows us to incorporate additional, more selective hypotheses about the structure of P and test them against the actually observed collective aggregation behavior. In this manner, we can assess in detail whether the real cells are indeed responding significantly to a specific set of hypothesized condition variables and, if so, how strongly. The approach will be illustrated below for a relatively simple example of assumed dependencies, without any dependence on nematic alignment. Adapting this technique to other combinations of dependencies is straightforward. The conditionality structure assumed in this example is as follows:

Conditionality Hypothesis 1: The state of motion variable, s_i, is conditional upon $(T_i, D_i, \rho_i\,, \phi_{i-1})$ and can be drawn independently of (θ_i, v_i, τ_i) from a conditional PDF, denoted by P_1, of the form

$$P_1(\,s_i|T_i, D_i, \rho_i\,, \phi_{i-1}). \qquad [25]$$

Conditionality Hypothesis 2: The random reversal angle variable, θ_i, is conditional upon $(s_i, T_i, D_i, \phi_{i-1})$ and can be drawn independently of (v_i, τ_i) from a conditional PDF, denoted by P_2, of the form

$$P_2(\theta_i|s_i, T_i, D_i, \phi_{i-1}). \qquad [26]$$

Conditionality Hypothesis 3: The random speed and run period variable pair, (v_i, τ_i), is conditional upon $(s_i, T_i, D_i, \rho_i, \beta_i)$ and can be drawn from a conditional PDF, denoted by P_3, of the form

$$P_3(v_i, \tau_i|\, s_i\,, T_i, D_i, \rho_i, \beta_i). \qquad [27]$$

The overall reversal probability, P, is then formally expressed in terms of P_1, P_2 and P_3 as

$$\begin{aligned} P(\,\theta_i, v_i\,, \tau_i\,,\; s_i|\;T_i,\; D_i\,,\; \rho_i,\; \phi_{i-1}) & \\ = \;\; P_1(s_i \mid T_i, D_i, \rho_i\,, \phi_{i-1}) \;\times\; P_2(\theta_i \mid s_i, T_i, D_i, \phi_{i-1}) & \qquad [28] \\ \times \;\; P_3(\,v_i, \tau_i \mid s_i\,, T_i, D_i, \rho_i, \beta_i)\,. & \end{aligned}$$

The actual random draw of $(\theta_i, v_i\,, \tau_i\,,\ s_i)$ is not executed in a single step following Eq. [28]. Rather, s_i, then θ_i, and then $(v_i\,, \tau_i)$ will be drawn in three successive steps, denoted by Step 2.1, 2.2 and 2.3 below, which implement the corresponding sequence of conditionality hypotheses 1, 2, and 3 stated above, as follows:

Given as input are the values of the conditionality variables $(T_i, D_i, \rho_i\,, \phi_{i-1}, \gamma_i, \beta_i)$, as stated under Step 2 of the random walk algorithm described above, and the observed run database, $\hat{q}_n$ for $n= 1, \dots, N_O$:

Step 2.1: Draw s_i from P_1, Eq. [25]: Find the run index n in the run database for which the tuple of observed variables $(\hat{T}_n, \hat{D}_n, \hat{\rho}_n\,, \hat{\phi}_{n-1})$ most closely matches the tuple of simulation conditionality variables, $(T_i, D_i, \rho_i\,, \phi_{i-1})$, in Eq. [25]. Then set $s_i = \hat{s}_n$ and use it as input to Steps 2.2 and 2.3.

The closest match between the foregoing tuples of observed and simulation conditionality variables is determined here by way of a distance cost function defined by

$$H_1(n) = \frac{\left|\hat{T}_n - T_i\right|}{T_\mathrm{R}} + \frac{\left|\hat{D}_n - D_i\right|}{D_\mathrm{R}} + \frac{\left|\hat{\rho}_n - \rho_i\right|}{\rho_\mathrm{R}} + \frac{\left|\hat{\phi}_{n-1} - \phi_{i-1}\right|}{\phi_\mathrm{R}} \qquad [29]$$

where the variable ranges $T_\mathrm{R}, D_\mathrm{R}, \rho_\mathrm{R}$ and ϕ_R are defined in Eqs. [23] and [24]. The closest match is then defined as the run index n which minimizes $H_1(n)$. In the (very unlikely) event of a tie, with multiple n-values, $n_1, n_2\,, \dots, n_K$, say, giving the same minimal value of H_1, the tie is broken by drawing the n-value randomly with uniform probability from the set $\{n_1, n_2\,, \dots, n_K\}$.

Step 2.2: Draw θ_i from P_2, Eq. [26]: Find the run index n in the run database for which $\hat{s}_n = s_i$ and the tuple of continuous observed variables $(\hat{T}_n, \hat{D}_n, \hat{\phi}_{n-1})$ most closely matches the tuple of continuous conditionality variables, (T_i, D_i, ϕ_{i-1}), in Eq. [26]. Then set $\theta_i = \hat{\theta}_n$.

Analogous to Step 2.1, the closest match is defined here as the run index n which minimizes the metric

$$H_2(n) = \frac{\left|\hat{T}_n - T_i\right|}{T_{\mathrm{R}}} + \frac{\left|\hat{D}_n - D_i\right|}{D_{\mathrm{R}}} + \frac{\left|\hat{\phi}_{n-1} - \phi_{i-1}\right|}{\phi_{\mathrm{R}}} \qquad [30]$$

subject to the constraint that $\hat{s}_n = s_i$, and with any tie to be broken by a uniformly random draw.

Step 2.3: Draw (v_i, τ_i) from P_3, Eq. [27]: Find the run index n in the run database for which $\hat{s}_n = s_i$ and the tuple of continuous observed variables, $\left(\hat{T}_n, \hat{D}_n, \hat{\rho}_n, \hat{\beta}_n\right)$, most closely matches the tuple of continuous conditionality variables, $(T_i, D_i, \rho_i, \beta_i)$, in Eq. [27] Then set $(v_i, \tau_i) = (\hat{v}_n, \hat{\tau}_n)$.

Analogous to Step 2.2, the closest match is defined here as the run index n which minimizes the metric

$$H_3(n) = \frac{\left|\hat{T}_n - T_i\right|}{T_{\mathrm{R}}} + \frac{\left|\hat{D}_n - D_i\right|}{D_{\mathrm{R}}} + \frac{\left|\hat{\rho}_n - \rho_i\right|}{\rho_{\mathrm{R}}} + \frac{\left|\hat{\beta}_n - \beta_i\right|}{\beta_{\mathrm{R}}} \qquad [31]$$

subject to the constraint that $\hat{s}_n = s_i$, and with any tie to be broken by a uniformly random draw.

During the early stages of the simulation the observed or simulated density profiles, $\rho(x, y, t)$, will likely not exhibit any detectable aggregates, thereby leaving the D-, ϕ- and β-variables in Eqs. [25]-[31] undefined. In those cases, we do not enforce the corresponding conditionalities by formally letting the normalization factors D_{R}, ϕ_{R} and β_{R} go to infinity in Eqs. [29]-[31], which is equivalent to simply dropping the D-, ϕ- and β-terms from the respective expressions on the right-hand sides of these equations.

It is imperative here to normalize the absolute difference term of each contributing variable in Eqs. [29]-[31], by dividing by the respective variable range from Eqs. [23] and [24]. For example, in Eq. [31] the four contributing conditionality variables, T_i, D_i, ρ_i and β_i, are each measured in a

different physical unit, and hence they must be non-dimensionalized before they can be added in any meaningful way. Using the variable ranges from Eqs. [23] and [24] as the normalizing divisor has the effect of treating the distance contributions from all four variables on an equal footing. If T_i falls within the biophysically "reasonable" range, defined by the range of the observed $\hat{T}_n$-values, then the dimensionless term $\left|\hat{T}_n - T_i\right| / T_{\mathrm{R}}$ in Eq. [31] falls within the interval [0 1]. The same is true for $\left|\hat{D}_n - D_i\right| / D_{\mathrm{R}}$ if D_i falls within the range of the observed $\hat{D}_n$-values, and likewise for the terms $\left|\hat{\rho}_n - \rho_i\right| / \rho_{\mathrm{R}}$ and $\left|\hat{\beta}_n - \beta_i\right| / \beta_{\mathrm{R}}$. As a consequence, each of the four terms in Eq. [31] carries *a priori* equal weight in contributing to the cost function $H_3(n)$.

References

Aman, A., and Piotrowski, T. (2010). Cell migration during morphogenesis. Dev. Biol. *341*, 20–33.

Balagam, R., and Igoshin, O.A. (2015). Mechanism for collective cell alignment in *Myxococcus xanthus* bacteria. PLOS Comput. Biol. *11*, e1004474.

Bastos, A.M., and Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front. Syst. Neurosci. *9*, 1–23.

Berleman, J.E., Zemla, M., Remis, J.P., Liu, H., Davis, A.E., Worth, A.N., West, Z., Zhang, A., Park, H., Bosneaga, E., et al. (2016). Exopolysaccharide microchannels direct bacterial motility and organize multicellular behavior. ISME J. *10*, 2620–2632.

Bode, H.B., Ring, M.W., Kaiser, D., David, A.C., Kroppenstedt, R.M., and Schwar, G. (2006). Straight-Chain fatty acids are dispensable in the myxobacterium *Myxococcus xanthus* for vegetative growth and fruiting body formation. J. Bacteriol. *188*, 5632–5634.

Botev, Z.I., Grotowski, J.F., and Kroese, D.P. (2010). Kernel density estimation via diffusion. Ann. Stat. *38*, 2916–2957.

Bretschneider, T., Othmer, H.G., and Weijer, C.J. (2016). Progress and perspectives in signal transduction, actin dynamics, and movement at the cell and tissue level: Lessons from *Dictyostelium*. Interface Focus *6*, 20160047.

Challa, S., Morelande, M.R., Musicki, D., and Evans, R.J. (2011). Fundamentals of object tracking (New York: Cambridge Unviersty Press).

Clark, A.G., and Vignjevic, D.M. (2015). Modes of cancer cell invasion and the role of the microenvironment. Curr. Opin. Cell Biol. *36*, 13–22.

Crocker, J., and Grier, D. (1996). Methods of digital video microscopy for colloidal studies. J. Colloid Interface Sci. *310*, 298–310.

Curtis, P.D., Geyer, R., White, D.C., and Shimkets, L.J. (2006). Novel lipids in *Myxococcus xanthus* and their role in chemotaxis. Environ. Microbiol. *8*, 1935–1949.

Curtis, P.D., Taylor, R.G., Welch, R.D., and Shimkets, L.J. (2007). Spatial organization of *Myxococcus xanthus* during fruiting body formation. J. Bacteriol. *189*, 9126–9130.

Davison, A.C., and Hinkley, D. V. (1997). Bootstrap methods and their application (New York, NY: Cambridge University Press).

Delgado, I., and Torres, M. (2016). Gradients, waves and timers, an overview of limb patterning models. Semin. Cell Dev. Biol. *49*, 109–115.

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. Protocol *1*, 1–10.

Ellenbroek, S.I.J., and van Rheenen, J. (2014). Imaging hallmarks of cancer in living mice. Nat. Rev. Cancer *14*, 406–418.

Freedman, D., and Diaconis, P. (1981). On the histogram as a density estimator:L 2 theory. Z. Für Wahrscheinlichkeitstheorie Verwandte Geb. *57*, 453–476.

Friedl, P., and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. Nat. Rev. Mol. Cell Biol. *10*, 445–457.

Gejji, R., Lushnikov, P.M., and Alber, M. (2012). Macroscopic model of self-propelled bacteria swarming with regular reversals. Phys. Rev. E *85*, 021903.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning (New York, NY: Springer New York).

Hendrata, M., Yang, Z., Lux, R., and Shi, W. (2011). Experimentally guided computational model discovers important elements for social behavior in myxobacteria. PloS One *6*, e22169.

Holmes, A.B., Kalvala, S., and Whitworth, D.E. (2010). Spatial simulations of myxobacterial development. PLoS Comput. Biol. *6*, e1000686.

Iber, D., and Zeller, R. (2012). Making sense-data-based simulations of vertebrate limb development. Curr. Opin. Genet. Dev. *22*, 570–577.

Igoshin, O.A., Welch, R., Kaiser, D., and Oster, G. (2004). Waves and aggregation patterns in myxobacteria. Proc. Natl. Acad. Sci. *101*, 4256–4261.

Iniesta, A.A., García-Heras, F., Abellón-Ruiz, J., Gallego-García, A., and Elías-Arnanz, M. (2012). Two systems for conditional gene expression in *Myxococcus xanthus* inducible by isopropyl-β-D-thiogalactopyranoside or vanillate. J. Bacteriol. *194*, 5875–5885.

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L., and Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. Nat. Methods *5*, 695–702.

Jelsbak, L., and Søgaard-Andersen, L. (2002). Pattern formation by a cell surface-associated morphogen in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. U. S. A. *99*, 2032–2037.

Jonker, R., and Volgenant, A. (1987). A shortest augmenting path algorithm for dense and sparse linear assignment problems. Computing *38*, 325–340.

Kashefi, K., and Hartzell, P.L. (1995). Genetic suppression and phenotypic masking of a *Myxococcus xanthus frzF-* defect. Mol. Microbiol. *15*, 483–494.

Kearns, D.B., and Shimkets, L.J. (1998). Chemotaxis in a gliding bacterium. Proc. Natl. Acad. Sci. *95*, 11957–11962.

Kearns, D.B., Venot, A., Bonner, P.J., Stevens, B., Boons, G.J., and Shimkets, L.J. (2001). Identification of a developmental chemoattractant in *Myxococcus xanthus* through metabolic engineering. Proc. Natl. Acad. Sci. U. S. A. *98*, 13990–13994.

Kim, S.K. (1990). Cell alignment required in differeniation of *Myxococcus xanthus*. Science *249*, 926–928.

Kuspa, A., Plamann, L., and Kaiser, D. (1992). Identification of heat-stable A-factor from *Myxococcus xanthus*. J. Bacteriol. *174*, 3319–3326.

Lander, A.D. (2013). How cells know where they are. Science *339*, 923–927.

Lock, J.G., Mamaghani, M.J., Shafqat-Abbasi, H., Gong, X., Tyrcha, J., and Strömblad, S. (2014). Plasticity in the macromolecular-scale causal networks of cell migration. PLoS ONE *9*, e90593.

Lux, R., Li, Y., Lu, A., and Shi, W. (2004). Detailed three-dimensional analysis of structural features of *Myxococcus xanthus* fruiting bodies using confocal laser scanning microscopy. Biofilms *1*, 293–303.

Masuzzo, P., Van Troys, M., Ampe, C., and Martens, L. (2016). Taking aim at moving targets in computational cell migration. Trends Cell Biol. *26*, 88–110.

Mauriello, E.M.F., Mignot, T., Yang, Z., and Zusman, D.R. (2010). Gliding motility revisited: how do the myxobacteria move without flagella? Microbiol. Mol. Biol. Rev. MMBR *74*, 229–249.

Moine, A., Agrebi, R., Espinosa, L., Kirby, J.R., Zusman, D.R., Mignot, T., and Mauriello, E.M.F. (2014). Functional organization of a multimodular bacterial chemosensory apparatus. PLoS Genet. *10*, e1004164.

Morelli, L.G., Uriu, K., Ares, S., and Oates, A.C. (2012). Computational approaches to developmental patterning. Science *336*, 187–191.

Park, S., Greco, V., and Cockburn, K. (2016). Live imaging of stem cells: Answering old questions and raising new ones. Curr. Opin. Cell Biol. *43*, 30–37.

Pathak, D.T., Wei, X., Bucuvalas, A., Haft, D.H., Gerloff, D.L., and Wall, D. (2012). Cell contact-dependent outer membrane exchange in myxobacteria: genetic determinants and mechanism. PLoS Genet. *8*, e1002626.

Sager, B., and Kaiser, D. (1993). Two cell-density domains within the *Myxococcus xanthus* fruiting body. Proc. Natl. Acad. Sci. U. S. A. *90*, 3690–3694.

Schumacher, L.J., Kulesa, P.M., McLennan, R., Baker, R.E., and Maini, P.K. (2016). Multidisciplinary approaches to understanding collective cell migration in developmental biology. Open Biol. *6*, 160056.

Shi, W., Ngok, F.K., and Zusman, D.R. (1996). Cell density regulates cellular reversal frequency in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. U. S. A. *93*, 4142–4146.

Sliusarenko, O., Neu, J., Zusman, D.R., and Oster, G. (2006). Accordion waves in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. *103*, 1534–1539.

Sliusarenko, O., Zusman, D.R., and Oster, G. (2007). Aggregation during fruiting body formation in *Myxococcus xanthus* is driven by reducing cell movement. J. Bacteriol. *189*, 611–619.

Solnica-Krezel, L., and Sepich, D.S. (2012). Gastrulation: Making and shaping germ layers. Annu. Rev. Cell Dev. Biol. *28*, 687–717.

Sozinova, O., Jiang, Y., Kaiser, D., and Alber, M. (2005). A three-dimensional model of myxobacterial aggregation by contact-mediated interactions. Proc. Natl. Acad. Sci. U. S. A. *102*, 11308–11312.

Sozinova, O., Jiang, Y., Kaiser, D., and Alber, M. (2006). A three-dimensional model of myxobacterial fruiting-body formation. Proc. Natl. Acad. Sci. U. S. A. *103*, 17255–17259.

Starruß, J., Bley, T., Søgaard-Andersen, L., and Deutsch, A. (2007). A new mechanism for collective migration in *Myxococcus xanthus*. J. Stat. Phys. *128*, 269–286.

Szabó, A., and Mayor, R. (2016). Modelling collective cell migration of neural crest. Curr. Opin. Cell Biol. *42*, 22–28.

Theveneau, E., and Mayor, R. (2012). Neural crest delamination and migration: From epithelium-to-mesenchyme transition to collective cell migration. Dev. Biol. *366*, 34–54.

Thutupalli, S., Sun, M., Bunyak, F., Palaniappan, K., and Shaevitz, J.W. (2015). Directional reversals enable *Myxococcus xanthus* cells to produce collective one-dimensional streams during fruiting-body formation. J. R. Soc. Interface *12*, 20150049.

Xie, C., Zhang, H., Shimkets, L.J., and Igoshin, O.A. (2011). Statistical image analysis reveals features affecting fates of *Myxococcus xanthus* developmental aggregates. Proc. Natl. Acad. Sci. *108*, 5915–5920.

Xu, Q., Black, W.P., Cadieux, C.L., and Yang, Z. (2008). Independence and interdependence of Dif and Frz chemosensory pathways in *Myxococcus xanthus* chemotaxis. Mol. Microbiol. *69*, 714–723.

Zhang, H., Angus, S., Tran, M., Xie, C., Igoshin, O.A., and Welch, R.D. (2011). Quantifying aggregation dynamics during *Myxococcus xanthus development*. J. Bacteriol. *193*, 5164–5170.

Zhou, T., and Nan, B. (2017). Exopolysaccharides promote *Myxococcus xanthus* social motility by inhibiting cellular reversals. Mol. Microbiol. *103*, 729–743.

(2014). Myxobacteria (Norfolk, UK: Caister Academic Press).

CHAPTER 3

CONCLUSIONS

By constraining computational models of *Myxococcus xanthus* development to experimental data, Chapter 2 identified four behavior-cue correlation that were sufficient to generate aggregation patterns in simulations quantitatively equivalent to that of experiments. These behaviors are: (1) reduced motility inside aggregates, (2) a biased walk toward aggregate centroids, (3) trajectory alignment radial to the aggregate boundary, and (4) trajectory alignment among neighboring cells. Previous *in vivo* imaging reported similar observations, including reduced motility inside the aggregates (Sliusarenko et al., 2007), suppression of reversals outside the aggregates (Jelsbak and Søgaard-Andersen, 2002), and cell alignment within the biofilm (Berleman et al., 2016; Shimkets and Kaiser, 1982). Chapter 2 extended these observations with quantitative measurements extracted from *in vivo* imaging and data driven modeling to test hypotheses on how *M. xanthus* cells aggregate. Without knowledge of the mechanism of the biased walk toward aggregate centroids, most *M. xanthus* aggregation models utilized only short-range signaling mechanisms. Chapter 2 indicated that in addition to local information, simulated cells required knowledge about the distance and direction to the nearest aggregate to match experimental aggregation patterns. Possible mechanisms of aggregation now need to be reevaluated in light of the requirement of long-range information for complete aggregation in *M. xanthus*.

Models of Development

The lack of an identified long-range signal active during *M. xanthus* development led to many models derivative of the traffic-jam hypothesis (Igoshin et al., 2001). In the traffic-jam hypothesis, the presence of some *M. xanthus* cells with reduced motility cause other cells to reduce motility as

well. Over time, cells build up an aggregate in a fashion like that of a buildup of cars in a traffic-jam. Derivatives of this model exist, such as the inclusion of cell-to-cell alignment (Sliusarenko et al., 2007), and short-range direction guidance (Janulevicius et al., 2015). A lack of quantitative comparison to experimental results and differences in free parameters between these models impairs any attempt to correlate them with biological results. As such, the reduction of cell motility inside of aggregates could be due to several mechanisms.

The traffic-jam model need not require biological machinery to process and respond to local cell density information and accomplish aggregation. Many of the properties of *M. xanthus* aggregation emerge in non-living, self-propelled systems. Alignment between non-living, self-propelled rods is well described (Baskaran and Marchetti, 2012; Peshkov et al., 2012; Ramaswamy, 2010; Sumino et al., 2012). For example, Sumino et al. (2012) used a carpet of dynein molecular motors grafted to a glass surface to propel approximately 15 μm long microtubules. With the addition of ATP, the dynein motors propelled the microtubules along their long axis across the dynein motor carpet. Microtubule movement was random, with no preferred direction detected for individual motile microtubules. With a sufficiently high microtubule density (approximately 5 microtubules per 100 μm^2), physical interactions between the microtubules led to aligned streams and vortexes. Alignment was nematic, with microtubules moving in both directions within the streams. Similar stream and vortex like structures are formed in place of aggregates by non-reversing *M. xanthus* mutants in some genetic backgrounds (Blackhart and Zusman, 1985; Zusman, 1982). *M. xanthus* and self-propelled rods share behavior similarities at low cell densities as well. At cell densities low enough that the cells do not completely cover the agar surface, wild type (WT) *M. xanthus* cells generate small clusters of aligned cells. These aligned clusters are also predicted to form in low-density, non-living, self-propelled systems as well (Bain and Bartolo, 2017; Peruani et al., 2012; Starruß et al., 2007). However, clustering is abolished when reversals of motility are

added to models of bendable, self-propelled, rods (Balagam and Igoshin, 2015). Balagam and Igoshin (2015) hypothesized that slime trails laid down by individual *M. xanthus* cells while moving could promote cluster formation in the presence of reversals. When cells encounter slime trails deposited by other cells, they often adopt the path of the slime trail, leading to trail following (Burchard, 1982). When Balagam and Igoshin (2015) added slime trails as a trail-following mechanism to bendable, self-propelled, rod simulations, streams and clusters of cells formed. Whether slime-trail following exists in *M. xanthus* biofilms, many cell layers thick, is not known. However, imaging of the biofilm suggests cells may generate tube-like paths through the extracellular matrix (ECM) that could act as a trail-generating mechanism (Berleman et al., 2016). The requirement of trail following to generate aligned clusters in simulations of bendable rods with reversing motility show that the behavior of *M. xanthus* cells must be accounted for when comparing aggregation to the behaviors of self-propelled, non-living systems.

Generation of correctly spaced aggregates may also be possible without biochemical signals. Thermodynamic models of self-propelled "matter" (Cates and Tailleur, 2015; Takatori and Brady, 2015) describe phase-separation like events in which the systems separate into dense (inside an aggregate) and dilute (outside an aggregate) phases based on the ratio of the particle run length relative to the physical movement constraints from neighboring particles. The presence of aggregate-like patterns in these thermodynamic models suggest that aggregation could occur without any intercellular signaling, instead being driven by changes in the average reversal frequency of the cells and the physical intercellular forces applied between them. It is not known if the bendable body of *M. xanthus* cells and changes in average reversal frequency during aggregation affect the outcome of such thermodynamic models. Thutupalli et al. (2015) has taken the first steps towards unification of *M. xanthus* behavior and thermodynamic models by showing that non-reversing cells produce disordered colonies at low cell density that are similar to a gas-like physical

state. The addition of reversals transforms the colony into a more liquid like state of stable "streams" of cells.

Generation of a biased walk toward the aggregate centroid without long-range signals may be possible as well. McCandlish et al. (2012) described the spontaneous generation of lanes of self-propelled rods, with all rods within a lane moving in the same direction. The spontaneous formation of directional streams in self-propelled rod simulations suggest an aggregation model in which cells could sort into streams with biased movement in one direction based on physical interactions. Aggregates may then form at the intersections of the streams of cells. More work is required to learn if biased lanes can emerge when bending along the long axis of the cell and direction reversals observed in *M. xanthus* cells are taken into account.

Chemotaxis up gradients of self-generated molecules is central to many eukaryotic and bacterial aggregation mechanisms (Budrene and Berg, 1991, 1995; Dormann and Weijer, 2006; Oppenheim and Yang, 2005). In these systems, cells both secrete and chemotactically sense small-diffusible molecules. Random fluctuations in the density of cells cause shallow differences in the concentration of the secreted molecule. Diffusion of the molecule away from areas of high concentration results in weak concentration gradients within the population. All cells move chemotactically up these gradients, increasing cell density at areas of higher chemoattractant concentration. As the cells move up the gradient, they also generate the attractant molecule, creating a feedback loop. Areas of higher chemoattractant concentration attract more cells, which leads to the generation of even higher concentrations of chemoattractant. Evidence against the role of a small, highly diffusible molecule in *M. xanthus* development is perhaps strong enough to discount the small-diffusible chemoattractant hypothesis. The questionable ability of *M. xanthus* to detect even moderate gradients of many highly diffusible molecules would only be exacerbated

by the shallow gradients expected to be present at the onset of aggregation. It is unlikely cells could move up such a gradient in time to reinforce the gradient before it dissipated.

It is possible a sufficiently steep gradient of a highly diffusible molecule could be set up by cells once the initial aggregate locations were defined. Aggregation clusters could initially form using a traffic-jam like model, then further recruit cells to the aggregate foci by production of a diffusible molecule. The modeling in Chapter 2 indicated that a biased walk is not required to generate visible aggregates, only to generate aggregates as quickly and to the same internal cell density as in experiments. Mutants lacking a diffusible signal that works to further recruit cells to the aggregates may display a relatively mild phenotype of slow aggregation. However, the cell movement bias extends approximately 100 μm outside the aggregate. At that distance, diffusible signals from adjacent aggregates could compete for cells. Larger aggregates would generate stronger chemotactic gradients than nearby smaller aggregates, causing cells to move away from the smaller aggregate and toward the larger neighbor. In this model, aggregate disappearance would not only be correlated with aggreagte size, but also with the distance and number of neighboring aggregates, as larger aggregates bias cells away from smaller neighbors. The loss of small aggregates is observed late in the aggregation process, but aggregate loss is not correlated with the properties of neighboring aggregates. The only predictor of aggregate loss is the size of the aggregate its self (Zhang et al., 2011). In all, experimental and theoretical evidence suggests that it is unlikely a small-diffusible molecule plays a role in aggregation.

While gradients established by diffusion of a small molecule followed by chemotactic sensing is the typical model of chemotactic aggregation, other models are possible. Cells secreting a non-diffusing signal would be able to determine their local cell density based on concentration of the molecule. A traffic-jam model could emerge if cells reduced motility as a function of the molecule concentrations. With the use of a chemosensory system, cells could also sense when they leave an

area of high cell density and reverse their motility direction to stay in an aggregation center. How such a mechanism would create biased movement toward an area of higher cell density up to 100 μm away from the aggregate is unclear.

A possible middle ground between fast or no diffusion is a slow-diffusing molecule. Such a molecule would act locally, keeping cells at high cell density locations, on short time scales. Over longer time scales, diffusion of the molecule through the biofilm could lead to biased cell movement toward aggregate locations. Such gradients may evolve too slowly to significantly affect the growth of surrounding aggregates, explaining why the loss of small aggregates is not correlated with the size, distance, or number of neighboring aggregates.

Determining the viability of aggregation models involving physical interactions, local signaling, or long-range signals is best studied using computational models. A general simulation framework addressing the principles of cell behavior could be adjusted to test the validity of each of the hypothetical aggregation mechanisms. The base configuration of the simulation must be parameterized at high cell densities and consider the physical nature of the cells. Specifically, cells must be simulated as bendable rods, self-propelled along their long axis, and capable of physically interacting with one another. The different aggregation models vary only in definition of the environmental factors that affect the probability of stopping and reversing. A model completely driven by physical interactions would be expected to generate aggregates in the proposed modeling framework by correctly tuning the average reversal frequency and physical interactions of the cells. Local and long-range signaling models could be tested by adding appropriate signaling interactions between the simulated cells that alter the probability of stopping and reversing. The validity of each model could be tested by constraining the parameterization of the cell behavior probabilities to the cell run behaviors reported in Chapter 2. Successful models would be expected to correctly reproduce the individual cell behaviors of cell-to-cell alignment, reduced cell motility inside

aggregates, increase in run duration outside of aggregates, and the biased walk toward the aggregate centroids. In addition, models should produce the correct number of aggregates with quantitatively correct size and shape. Even with cell behaviors constrained to experimental results, it is likely multiple models will explain the data. In this case, experimental analyses of mutants using the framework described in Chapter 2, hopefully guided by the simulation results, will be vital to fully distinguish between aggregation models.

Robustness and Compensatory Mechanisms

The flat agar surfaces on which developmental assays are performed are significantly different than the soil, dung, and ocean environments various myxobacterial species can be found. This plethora of substrates on which myxobacteria live may be a driving factor behind the need to keep two independent surface motility systems. It stands to reason that if evolutionary pressures exist to drive the upkeep of such complicated mechanisms of motility – one optimized for soft surfaces, the other optimized for rigid surfaces (Shi and Zusman, 1993) – similar pressures could drive the upkeep of more than one method of aggregation. Simulations in Chapter 2 demonstrated that an increase in cell reversal period outside the aggregates is sufficient to overcome the need for a biased walk, and that the biased walk is sufficient to drive aggregation without the need for increased run durations outside the aggregates. Increased run duration outside the aggregates is a central prediction of aggregation models derived from only physical interactions. The increase in run durations is equivalent to an increase in temperature in thermodynamic-backed aggregation models, and is a key change required to drive the system into stable and concurrent phases of dilute and concentrated cells. Conversely, biased movement toward an aggregate is more conducive to models of aggregation relying on chemotaxis. Multiple mechanisms of aggregation may help explain why so little about *M. xanthus* aggregation has been elucidated by mutational studies. Only genes that either abolish motility control (such as the Frz system) or are required for all aggregation

systems could be easily identified using mutant screens. The possibility of multiple aggregation mechanisms further highlights the need for experimental frameworks that allow the quantification of cell behavior and cue correlations without the need for global phenotypic changes or abolishment of the self-organization.

Testing models of aggregation

Scientists often try to fit experimental results to "toy-models", named for their deliberately simplistic nature, ignoring details in an effort to concisely explain an overarching hypothesis about the system under study. Analysis of *M. xanthus* aggregation highlights the caveats of applying toy models to self-organizing systems. Even with quantitative knowledge of the behaviors and general cues driving *M. xanthus* development, it is still possible to fit the results to toy-models involving, cell-to-cell, slowly-diffusible, or non-diffusible signaling. Previous knowledge of the biochemical signaling and motility control mechanisms should help differentiate between hypotheses. However, available experimental results generate conflicting evidence, providing multiple points for and against each model type. The conflicting experimental results suggest that some of the described systems are not related to *M. xanthus* development, instead playing roles in other unknown processes, or represent the presence of redundant aggregation mechanisms. Further experimentation under the experimental framework proposed by this thesis must be undertaken to identify the correct aggregation hypothesis. The multitude of possible behavior and cue combinations that could play a role in aggregation, and the lack of clear conclusions from currently available experimental results, make identifying mutants that merit further study a daunting task. A few promising leads are discussed below.

A number of experimental observations tie exopolysaccharide (EPS) regulation and lipid sensing to motility control and development. EPS is a key component of the extracellular matrix of the *M. xanthus* biofilm and consists of monosaccharides including galactose, glucosamine, glucose,

rhamnose, and xylose (Behmlander and Dworkin, 1994). EPS is required for cohesion between cells (Shimkets, 1986a), development (Shimkets, 1986b), and social motility (Shimkets, 1986a; Yang et al., 2010). The large-polymer nature and its importance in adhesion between cells suggest it would not be naturally diffusible.

Phosphatidylethanolamine (PE) and diacylglycerol (DAG) are chemoattractants for *M. xanthus* cells (Kearns and Shimkets, 1998), and are slowly diffusible in agar (Kearns and Shimkets, 2001). The slow diffusion rate may allow these lipids to act in a manner similar to the slow-diffusion model of aggregation discussed above. Bioactivity of lipids during development is already documented. Cells blocked in the synthesis of isovaleryl-coenzyme A, a precursor for the creation of *iso*15:0 fatty acid, and the triacylglycerol TG1 generate aggregate-like mounds of cells but very few spores. Aggregation and sporulation can be rescued by co-development with WT cells (Downard et al., 1993; Toal et al., 1995) or by the addition of *iso*15:0 or TG1 at physiological concentrations (Bhat et al., 2014), suggesting *iso*15:0 and TG1 act as signals during development, but not necessarily as the transmitters of aggreagte location information. While these lipids appear to affect sporulation more than aggregation, other lipids may also play a role in development. Kearns et al. (2001) identified the PE molecule 16:1ω5c/16:1ω5c (hereafter referred to as 16:1) as a potent chemoattractant at physiological levels. 16:1 is only active as a chemoattractant under starvation conditions, a key trigger of aggregation. However, inhibition of production of all but approximately 2% of the normal levels of 16:1 does not eliminate aggregation (Bode et al., 2006), suggesting it is not an essential component of development. While 16:1 does not appear to be a key component of aggregation, there are many untested lipids found in *M. xanthus* cells during development that could be active (Curtis et al., 2006).

The recent identification of cardiolipin (CL) and phosphatidylglycerol (PG) as substrates for CsgA, and that partial glycerides extracted from developing WT cells restore development to *csgA*

cells (Boynton and Shimkets, 2015) suggest lipid metabolism may play a role in any *csgA* cell behavior defects and provide an avenue for further study of lipid chemotaxis during development. That the partial glycerides rescue development when added uniformly suggests they do not directly provide aggregate location information, such as by acting as a chemoattractant. At the very least, enzymatic activity downstream of CsgA would be required to locally convert the partial glycerides into a specific chemoattractant.

A key sensory system for both EPS regulation and lipid chemotaxis is the chemosensory system Dif (Yang et al., 2000). The *dif* operon contains six genes, *difABCDEG*. DifA, C, E, and D make up a typical prokaryotic chemosensory system comprising of a methyl-accepting chemotaxis protein, adapter protein, histidine protein kinase, and response regulator, respectively (Bellenger et al., 2002; Yang et al., 1998). DifG is homologous to a chemotaxis protein not found in *E. coli*, but present in other prokaryotes, such as *Bacillus*, and typically found in archaeal chemotaxis pathways (Rosario and Ordal, 1996; Szurmant et al., 2004). DifG may be involved in dephosphorylation of CheY homologues (such as DifD) and in adaptation regulation (Wadhams and Armitage, 2004). DifB is homologous to a family of proteins of unknown function (Black and Yang, 2004). Key proteins missing in the operon are the methylesterase and methyltransferase, which are required for adaptation to chemoattractants in the typical chemosensory system (Wadhams and Armitage, 2004). Indeed, DifA is not methylated, suggesting the Dif system is unable to directly sense concentration gradients (Xu et al., 2008, 2011). Mutations in the main sensory complex, DifACE, abolish EPS production. Surprisingly, the response regulator, DifD, is not required for EPS production. Instead, *difD* mutants overproduce EPS, suggesting a novel and complex regulatory pathway. Since EPS is required for development, mutants lacking a functional DifACE signaling system do not form aggregates under starvation conditions. The inability to aggregate can be bypassed by the addition of EPS purified from WT cells (Chang and Dworkin, 1994). EPS is also

required for PE sensing (Bonner et al., 2005; Kearns et al., 2000). However, the role of Dif in PE sensing is more complicated than a lack of EPS, as addition of purified WT EPS to a *difA* mutant is not sufficient to restore PE sensing (Kearns et al., 2000). DifA, C, and E are also all required for methylation of FrzCD in the presence of lipid chemoattractants (Xu et al., 2008), suggesting adaptation to PE concentrations is accomplished through FrzCD methylation.

EPS and the Dif pathway also have a direct effect on motility control. Deletion of *difA, difC,* or *difE* leads to a 5-fold increase in the single cell basal reversal period from approximately 7 min to 37 min between reversals (Bonner et al., 2005). Interestingly, the change in reversal period can be fixed by incubating the cells with purified WT EPS prior to performing the reversal assay (Kearns et al., 2000).

FibA is an extracellular EPS-associated protein homologous to the M4 family of zinc metalloproteases. While the function of FibA is unknown, it is required for PE sensing (Kearns et al., 2002). Mutant *fibA* cells develop normally, with the exception of creating long ridge-like aggregates instead of the typical ovals when developed at higher than normal cell densities (Kearns et al., 2002). Individual *fibA* cells have a 2-fold increase in the single cell reversal period relative to WT cells. Bonner et al. (2006) reported that while mutants lacking either *fibA* or *pilA* individually create aggregates and spores, a *fibA pilA* double mutant completely abolishes development. The PilA protein is the sole component of pili extended by the type IV pili S-motility system. While mutations in many of the type IV pili machinery genes are unable to develop, mutants that allow for the extension, but not retraction of pili (e.g. *pilT, pilS)*, or mutants lacking *pilA* are able to generate aggregates and spores (Bonner et al., 2006; Wu et al., 1998; Yang et al., 2010). A notable correlation between type IV pili mutants and development is EPS production. Mutants that lack *pilT and pilS* produce at least as much EPS as WT cells. A *pilA* mutant makes less than half WT levels of EPS, but still more than the non-developing mutants (Black et al., 2006; Wu et al., 1998;

Yang et al., 2010). Mutants that lack *fibA* produce WT levels of EPS (Bonner et al., 2006), suggesting a *pilA fibA* double mutant should be equivalent to a *pilA* mutant in EPS production. The lack of development in the double mutant suggests *fibA* and *pilA* cells could be using complementary mechanisms to drive development.

Several lines of evidence inexplicitly tie the *csgA* null mutation and motility during aggregation: (1) Developmental gene expression is blocked at approximately 6 hours after starvation for both non-motile and *csgA* cells. This time point is also approximately when cells begin to move into aggregation centers. (2) Non-motile developmental gene expression can be rescued by the addition of exogenous CsgA protein, suggesting that signaling via CsgA requires motility. (3) The methylation of FrzCD during development is contingent on the presence of *csgA* (Søgaard-Andersen and Kaiser, 1996), correlating *csgA* expression with reversal adaptation. Correlations with motility, a possible role in lipid metabolism, and the ability to rescue *csgA* cells by co-development with WT cells (Lee and Shimkets, 1996), suggest CsgA may be part of a signaling mechanism required for cell behavior coordination. While signaling mechanisms in which *csgA* is deficient in a cell-cell contact dependent mechanism was previously proposed (Jelsbak and Søgaard-Andersen, 2000), almost all the reported results on *csgA* cell behavior are consistent with a general abrogation in motility control. Jelsbak and Søgaard-Andersen (2002) provided some evidence that *csgA* development is not fully complemented when mixed with WT cells. The authors tracked GFP labeled *csgA* or WT cells mixed into a WT biofilm at a ratio of 1:500. The GFP cells were then tracked for 90 minutes starting at 9 hours after starvation. By 9 hours, stable aggregates have already formed. In these experiments, the small number of GFP labeled *csgA* cells are completely surrounded by WT cells within the biofilm, providing ample ability for the WT cells to complement any defects caused by the *csgA* mutation. Comparison of the behavior of GFP labeled WT and *csgA* cells indicated *csgA* cells reversed 1.5 times less often than WT cells at 9 hours into development.

While reported experimental error suggests the difference may be significant, Jelsbak and Søgaard-Andersen (2002) interpreted a difference of 1.5 times to be within the range of fully complemented to WT behavior patterns. Simulations in Chapter 2 indicate that a 1.5-fold difference in reversal frequency is sufficient to overcome the need for a biased walk, suggesting the difference may not be WT-like, and could be masking other defects in *csgA* cell behavior.

A model of cell signaling during aggregation

EPS and lipids have strong experimental evidence tying them to motility control, aggregation, and chemotaxis. These ideas are combined in the model proposed in Figure 3.1. At the onset of development, CsgA participates in the conversion of membrane cardiolipin (CL) and phosphatidylglycerol (PG) into DAG. The DAGs are either released by the cell, where they can be utilized as a slow-diffusible signal sensed via a FibA, EPS, DifA dependent mechanism, or they interact inside the cell with FrzCD to modify the reversal period of the cell (Figure 3.1 green arrows). As the green arrows suggest, it is also possible DAGs affect FrzCD from both inside and outside the cell. The Dif system is also required for EPS production (Figure 3.1, red arrows). The lack of development in a *fibA pilA* double mutant represents a second aggregation mechanism. The presence of extended pili is central to this second mechanism. Cells that are able to produce pili are unaffected by addition of the *fibA* mutation. These include WT, or mutants that can extend, but not retract pili, for example, *pilT* (Bonner et al., 2006; Wu et al., 1997). Aggregation is abolished by the addition of the *fibA* mutation to backgrounds that make sufficient EPS to develop, but no pili, such as *pilA* and *pilH* (Bonner et al., 2006; Wu et al., 1998). Extended pili could provide an avenue for medium range sensing by extending out from the cell to interact with the ECM or nearby cells. Extended pili could also change cell motility by acting as a tether to other cells or the ECM. The following experiments are suggested to test the validity of the Figure 3.1 model.

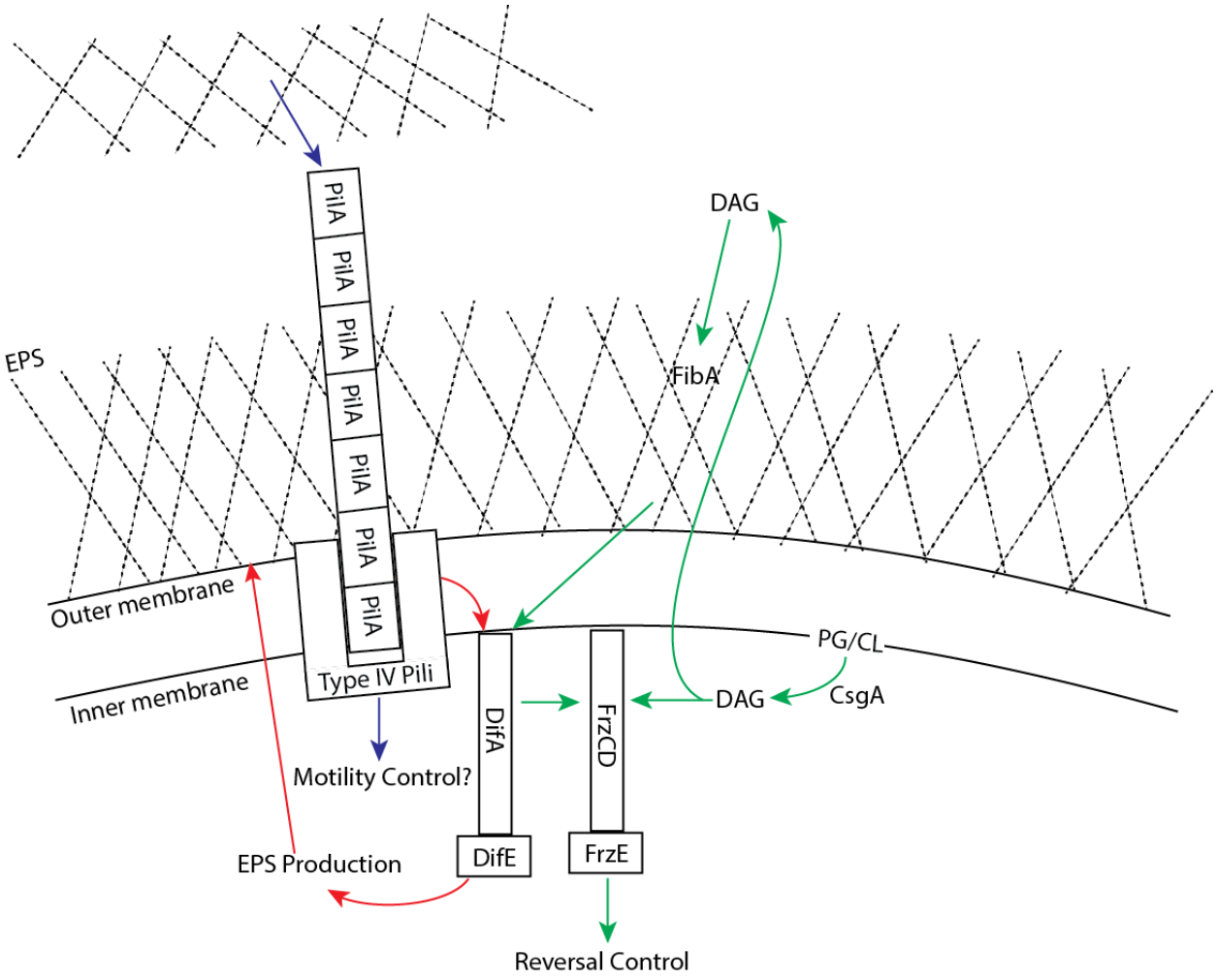

Figure 3.1: Toy model of possible signaling mechanisms utilized during aggregation. Exopolysaccharides (EPS) are required for development. EPS production is regulated by the Dif chemosensory system via interactions with the type IV pili machinery (red arrows). Green arrows represent a chemotactic aggregation model in which CsgA converts membrane phosphatidylglycerol (PG) and cardiolipin (CL) into DAG diacylglycerol (DAG), which is then transported outside the cell to act as a chemical signal, or interact with FrzCD intracellularly to affect the cell reversal period. Extracellular DAGs are sensed via a FibA, ECM, DifA mechanism, which affects the cell reversal period though FrzCD. A second, type IV pili dependent aggregation mechanism is also proposed (blue arrows).

To fully test whether *csgA* cells are complemented by WT cells, a detailed analysis of cell behaviors during the onset of aggregation should be performed. Cell-tracking and modeling techniques described in Chapter 2 should be repeated for fluorescent *csgA* cells in a *csgA* biofilm and in a WT biofilm. When aggregation is induced in a biofilm consisting of a mixture of 50:50 WT to *csgA* cells, *csgA* cells sporulate at an approximately 1:1 ratio with WT cells (Lee and Shimkets, 1996). This result suggests two possible mechanisms of complementation. In the first, *csgA* cells randomly enter the aggregates, or are not well-trapped by the aggregates, but are more likely to become spores than WT cells. The second possibility is that csgA cells move into, and stay in aggregates, at the same rate as WT cells.

The lack of development in the *fibA pilA* double mutant suggests *fibA* or *pilA* single mutants could be using complementary mechanisms to drive development. The importance of FibA in lipid chemotaxis suggests a possible chemotactic mechanism. How a PilA mediated mechanism may function is less clear. Mutants capable of generating pili do not require *fibA* to generate aggregates, suggesting pili play a role in aggregation beyond inducing EPS production. A complementation experiment of adding purified WT EPS to each of the *pil* mutants that are deficient in EPS production would confirm that the lack of EPS is their sole developmental defect. If EPS restores development, type IV pili machinery mutants that are unable to generate pili (pili$^-$) may be utilizing the *fibA* aggregation mechanism. In this case, adding purified EPS to pili$^-$ *fibA* double mutants would not generate aggregates. The independent aggregation mechanisms hypothesis also suggests the possibility of significantly different cell behaviors between *pilA, fibA,* and WT cells during development. Fluorescent-cell tracking of *pilA, fibA* individual mutants or the double *pilA fibA* mutant during development in homogeneous or mixed into WT biofilms and subsequent data-driven simulations provide a second avenue of analysis of the aggregation mechanisms.

Concluding Remarks

The framework described in this dissertation proposes a technique to simplify the identification of the cell behaviors and behavioral cues coordinating self-organized biological processes. Chapter 2 elucidated the framework and applied it to *M. xanthus* development. Analysis of *M. xanthus* development confirmed some behavior cue combinations already discussed in the literature, such as reduced motility inside the aggregates and cell alignment within the biofilm. The analysis also uncovered an unexpected cell behavior, a biased cell walk toward aggregates by cells as far as 100 μm away from the nearest aggregate boundary. This discovery identifies the need for analysis of self-organized systems which include quantitative analysis of cell behavior within the system and simulations of the self-organization tightly constrained to experimental results. The discussion of short-range, small-diffusible, and slow-diffusible molecule signaling and related aggregation models in this chapter highlight that any of these methods could, in theory, produce the cell behaviors and behavioral cues identified in Chapter 2. Further analysis using both simulations constrained by quantitative experimental data, as well as quantitative analysis of mutant *M. xanthus* strains, will be required to fully uncover the underlying mechanisms of aggregation.

References

Bain, N., and Bartolo, D. (2017). Critical mingling and universal correlations in model binary active liquids. Nat. Commun. *8*.

Balagam, R., and Igoshin, O.A. (2015). Mechanism for collective cell alignment in *Myxococcus xanthus* bacteria. PLOS Comput. Biol. *11*, e1004474.

Baskaran, A., and Marchetti, M.C. (2012). Self-regulation in self-propelled nematic fluids. Eur. Phys. J. E *35*.

Behmlander, R.M., and Dworkin, M. (1994). Biochemical and structural analyses of the extracellular matrix fibrils of *Myxococcus xanthus*. J. Bacteriol. *176*, 6295–6303.

Bellenger, K., Ma, X., Shi, W., and Yang, Z. (2002). A CheW homologue is required for *Myxococcus xanthus* fruiting body development, social gliding motility, and fibril biogenesis. J. Bacteriol. *184*, 5654–5660.

Berleman, J.E., Zemla, M., Remis, J.P., Liu, H., Davis, A.E., Worth, A.N., West, Z., Zhang, A., Park, H., Bosneaga, E., et al. (2016). Exopolysaccharide microchannels direct bacterial motility and organize multicellular behavior. ISME J. *10*, 2620–2632.

Bhat, S., Ahrendt, T., Dauth, C., Bode, H.B., and Shimkets, L.J. (2014). Two lipid signals guide fruiting body development of *Myxococcus xanthus*. MBio *5*, e00939-13.

Black, W.P., and Yang, Z. (2004). *Myxococcus xanthus* chemotaxis homologs difd and difg negatively regulate fibril polysaccharide production. J. Bacteriol. *186*, 1001–1008.

Black, W.P., Xu, Q., and Yang, Z. (2006). Type IV pili function upstream of the Dif chemotaxis pathway in *Myxococcus xanthus* EPS regulation. Mol. Microbiol. *61*, 447–456.

Blackhart, B.D., and Zusman, D.R. (1985). "Frizzy" genes of *Myxococcus xanthus* are involved in control of frequency of reversal of gliding motility. Proc. Natl. Acad. Sci. *82*, 8767–8770.

Bode, H.B., Ring, M.W., Kaiser, D., David, A.C., Kroppenstedt, R.M., and Schwar, G. (2006). Straight-chain fatty acids are dispensable in the myxobacterium *Myxococcus xanthus* for vegetative growth and fruiting body formation. J. Bacteriol. *188*, 5632–5634.

Bonner, P.J., Xu, Q., Black, W.P., Li, Z., Yang, Z., and Shimkets, L.J. (2005). The Dif chemosensory pathway is directly involved in phosphatidylethanolamine sensory transduction in *Myxococcus xanthus*: Dif pathway mediates PE chemotaxis. Mol. Microbiol. *57*, 1499–1508.

Bonner, P.J., Black, W.P., Yang, Z., and Shimkets, L.J. (2006). FibA and PilA act cooperatively during fruiting body formation of *Myxococcus xanthus*. Mol. Microbiol. *61*, 1283–1293.

Boynton, T.O., and Shimkets, L.J. (2015). *Myxococcus* Csga, *drosophila* Sniffer, and Human HSD10 are cardiolipin phospholipases. Genes Dev. *29*, 1903–1914.

Budrene, E.O., and Berg, H.C. (1991). Complex patterns formed by motile cells of *Escherichia coli*. Nature *349*, 630–633.

Budrene, E.O., and Berg, H.C. (1995). Dynamics of formation of symmetrical patterns by chemotactic bacteria. Nature *376*, 49–53.

Burchard, R.P. (1982). Trail following by gliding bacteria. J. Bacteriol. *152*, 495–501.

Cates, M.E., and Tailleur, J. (2015). Motility-induced phase separation. Annu. Rev. Condens. Matter Phys. *6*, 219–244.

Chang, B.Y., and Dworkin, M. (1994). Isolated fibrils rescue cohesion and development in the Dsp mutant of *Myxococcus xanthus*. J. Bacteriol. *176*, 7190–7196.

Curtis, P.D., Geyer, R., White, D.C., and Shimkets, L.J. (2006). Novel lipids in *Myxococcus xanthus* and their role in chemotaxis. Environ. Microbiol. *8*, 1935–1949.

Dormann, D., and Weijer, C.J. (2006). Chemotactic cell movement during *Dictyostelium* development and gastrulation. Curr. Opin. Genet. Dev. *16*, 367–373.

Downard, J., Ramaswamy, S.V., and Kil, K.-S. (1993). Identification of esg, a genetic locus involved in cell-cell signaling during *Myxococcus xanthus* development. J. Bacteriol. *175*, 7762–7770.

Igoshin, O.A., Mogilner, A., Welch, R.D., Kaiser, D., and Oster, G. (2001). Pattern formation and traveling waves in myxobacteria: Theory and modeling. Proc. Natl. Acad. Sci. *98*, 14913–14918.

Janulevicius, A., Loosdrecht, M. van, and Picioreanu, C. (2015). Short-range guiding can result in the formation of circular aggregates in myxobacteria populations. PLOS Comput. Biol. *11*, e1004213.

Jelsbak, L., and Søgaard-Andersen, L. (2000). Pattern formation: Fruiting body morphogenesis in *Myxococcus xanthus*. Curr. Opin. Microbiol. *3*, 637–642.

Jelsbak, L., and Søgaard-Andersen, L. (2002). Pattern formation by a cell surface-associated morphogen in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. *99*, 2032–2037.

Kearns, D., and Shimkets, L. (2001). Directed movement and surface-borne motility of *Myxococcus* and *Pseudomonas*. Methods Enzymol. *336*, 94–102.

Kearns, D.B., and Shimkets, L.J. (1998). Chemotaxis in a gliding bacterium. Proc. Natl. Acad. Sci. *95*, 11957–11962.

Kearns, D.B., Campbell, B.D., and Shimkets, L.J. (2000). *Myxococcus xanthus* fibril appendages are essential for excitation by a phospholipid attractant. Proc. Natl. Acad. Sci. *97*, 11505–11510.

Kearns, D.B., Venot, A., Bonner, P.J., Stevens, B., Boons, G.-J., and Shimkets, L.J. (2001). Identification of a developmental chemoattractant in *Myxococcus xanthus* through metabolic engineering. Proc. Natl. Acad. Sci. *98*, 13990–13994.

Kearns, D.B., Bonner, P.J., Smith, D.R., and Shimkets, L.J. (2002). An extracellular matrix-associated zinc metalloprotease is required for dilauroyl phosphatidylethanolamine chemotactic excitation in *Myxococcus xanthus*. J. Bacteriol. *184*, 1678–1684.

Lee, K., and Shimkets, L.J. (1996). Suppression of a signaling defect during *Myxococcus xanthus* development. J. Bacteriol. *178*, 977–984.

McCandlish, S.R., Baskaran, A., and Hagan, M.F. (2012). Spontaneous segregation of self-propelled particles with different motilities. Soft Matter *8*, 2527.

Oppenheim, J.J., and Yang, D. (2005). Alarmins: Chemotactic activators of immune responses. Curr. Opin. Immunol. *17*, 359–365.

Peruani, F., Starruß, J., Jakovljevic, V., Søgaard-Andersen, L., Deutsch, A., and Bär, M. (2012). Collective motion and nonequilibrium cluster formation in colonies of gliding bacteria. Phys. Rev. Lett. *108*.

Peshkov, A., Aranson, I.S., Bertin, E., Chaté, H., and Ginelli, F. (2012). Nonlinear field equations for aligning self-propelled rods. Phys. Rev. Lett. *109*, 268701.

Ramaswamy, S. (2010). The mechanics and statistics of active matter. Annu. Rev. Condens. Matter Phys. *1*, 323–345.

Rosario, M.M., and Ordal, G.W. (1996). CheC and CheD interact to regulate methylation of *Bacillus subtilis* methyl-accepting chemotaxis proteins. Mol. Microbiol. *21*, 511–518.

Shi, W., and Zusman, D.R. (1993). The two motility systems of *Myxococcus xanthus* show different selective advantages on various surfaces. Proc. Natl. Acad. Sci. *90*, 3378–3382.

Shimkets, L.J. (1986a). Correlation of energy-dependent cell cohesion with social motility in *Myxococcus xanthus*. J. Bacteriol. *166*, 837–841.

Shimkets, L.J. (1986b). Role of cell cohesion in *Myxococcus xanthus* fruiting body formation. J. Bacteriol. *166*, 842–848.

Shimkets, L.J., and Kaiser, D. (1982). Induction of coordinated movement of *Myxococcus xanthus* cells. J. Bacteriol. *152*, 451–461.

Sliusarenko, O., Zusman, D.R., and Oster, G. (2007). Aggregation during fruiting body formation in *Myxococcus xanthus* is driven by reducing cell movement. J. Bacteriol. *189*, 611–619.

Søgaard-Andersen, L., and Kaiser, D. (1996). C factor, a cell-surface-associated intercellular signaling protein, stimulates the cytoplasmic Frz signal transduction system in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. *93*, 2675–2679.

Starruß, J., Bley, T., Søgaard-Andersen, L., and Deutsch, A. (2007). A new mechanism for collective migration in *Myxococcus xanthus*. J. Stat. Phys. *128*, 269–286.

Sumino, Y., Nagai, K.H., Shitaka, Y., Tanaka, D., Yoshikawa, K., Chaté, H., and Oiwa, K. (2012). Large-scale vortex lattice emerging from collectively moving microtubules. Nature *483*, 448–452.

Szurmant, H., Muff, T.J., and Ordal, G.W. (2004). *Bacillus subtilis* CheC and FliY are members of a novel class of CheY-P-hydrolyzing proteins in the chemotactic signal transduction cascade. J. Biol. Chem. *279*, 21787–21792.

Takatori, S.C., and Brady, J.F. (2015). Towards a "thermodynamics" of active matter. Phys. Rev. E *91*.

Thutupalli, S., Sun, M., Bunyak, F., Palaniappan, K., and Shaevitz, J.W. (2015). Directional reversals enable *Myxococcus xanthus* cells to produce collective one-dimensional streams during fruiting-body formation. J. R. Soc. Interface *12*.

Toal, D.R., Clifton, S.W., Roe, B.A., and Downard, J. (1995). The esg locus of *Myxococcus xanthus* encodes the E1α and E1β subunits of a branched-chain keto acid dehydrogenase. Mol. Microbiol. *16*, 177–189.

Wadhams, G.H., and Armitage, J.P. (2004). Making sense of it all: bacterial chemotaxis. Nat. Rev. Mol. Cell Biol. *5*, 1024–1037.

Wu, S.S., Wu, J., and Kaiser, D. (1997). The *Myxococcus xanthus pilT* locus is required for social gliding motility although pili are still produced. Mol. Microbiol. *23*, 109–121.

Wu, S.S., Wu, J., Cheng, Y.L., and Kaiser, D. (1998). The *pilH* gene encodes an ABC transporter homologue required for type IV pilus biogenesis and social gliding motility in *Myxococcus xanthus*. Mol. Microbiol. *29*, 1249–1261.

Xu, Q., Black, W.P., Cadieux, C.L., and Yang, Z. (2008). Independence and interdependence of Dif and Frz chemosensory pathways in *Myxococcus xanthus* chemotaxis. Mol. Microbiol. *69*, 714–723.

Xu, Q., Black, W.P., Nascimi, H.M., and Yang, Z. (2011). DifA, a methyl-accepting chemoreceptor protein-like sensory protein, uses a novel signaling mechanism to regulate exopolysaccharide production in *Myxococcus xanthus*. J. Bacteriol. *193*, 759–767.

Yang, Z., Geng, Y., Xu, D., Kaplan, H.B., and Shi, W. (1998). A new set of chemotaxis homologues is essential for *Myxococcus xanthus* social motility. Mol. Microbiol. *30*, 1123–1130.

Yang, Z., Ma, X., Tong, L., Kaplan, H.B., Shimkets, L.J., and Shi, W. (2000). *Myxococcus xanthus dif* genes are required for biogenesis of cell surface fibrils essential for social gliding motility. J. Bacteriol. *182*, 5793–5798.

Yang, Z., Lux, R., Hu, W., Hu, C., and Shi, W. (2010). PilA localization affects extracellular polysaccharide production and fruiting body formation in *Myxococcus xanthus*: PilA localization modulates EPS production. Mol. Microbiol. *76*, 1500–1513.

Zhang, H., Angus, S., Tran, M., Xie, C., Igoshin, O.A., and Welch, R.D. (2011). Quantifying aggregation dynamics during *Myxococcus xanthus* development. J. Bacteriol. *193*, 5164–5170.

Zusman, D.R. (1982). "Frizzy" mutants: A new class of aggregation-defective developmental mutants of *Myxococcus xanthus*. J. Bacteriol. *150*, 1430–1437.