Data-driven modeling reveals cell behaviors controlling self-organization during *Myxococcus xanthus* development

Christopher R. Cotter[a], Heinz-Bernd Schüttler[b], Oleg A. Igoshin[c,d,1], and Lawrence J. Shimkets[a,1]

[a]Department of Microbiology, University of Georgia, Athens, GA 30602; [b]Department of Physics and Astronomy, University of Georgia, Athens, GA 30602; [c]Department of Bioengineering, Rice University, Houston, TX 77005; and [d]Center for Theoretical Biological Physics, Rice University, Houston, TX 77005

Edited by E. Peter Greenberg, University of Washington, Seattle, WA, and approved April 24, 2017 (received for review December 20, 2016)

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* uses coordinated movement to generate three-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 unique 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 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.

agent-based simulation | image processing | emergent behavior | fluorescent imaging | cell communication

Collective cell migration is essential for many developmental processes, including fruiting body development of myxobacteria (1) and *Dictyostelium* (2), embryonic gastrulation (3, 4), and neural crest development (5). Conversely, cancer cell metastases represent detrimental migratory events that disseminate dysfunctional cells (6). 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 (7). Much less is known about the system-level coordination of cell migration. Cell movement in these systems is a 3D, dynamic process coordinated by a combination of diverse physical and chemical cues acting on the cells (3, 5, 8). Recent developments in tracking individual cell movement in vivo have provided unprecedented detail and revealed surprising levels of heterogeneity (5, 7). Reverse engineering of how these individual cell movements lead to collective migration patterns has proved difficult. Whereas 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 (9–12). 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 (13) and exchanging outer membrane material (14). Cells are flexible rods that move along their long axis within the biofilm (15). Periodic reversals in direction of movement and a high length-to-diameter aspect ratio allow cells to align with neighbors, move in groups, and follow paths taken by others (16–18). When faced with amino acid limitation, cells self-organize into aggregates much taller than the surrounding biofilm called fruiting bodies (17, 19). Aggregation begins with a burst of cell motility during which cells coalesce into unstable towers a few layers thicker than the surrounding biofilm (20). Within 1 h, towers begin to form spatially stable aggregation centers. Although some aggregates mature into spore-filled fruiting bodies, many initially stable aggregates disseminate back into the biofilm (21). Few data exist on the cues and cell behaviors that lead to these emergent behaviors. Cell-tracking experiments revealed that motility increases outside aggregates (20, 22, 23) and decreases inside (23, 24) whereas statistical image analysis revealed that the area of the aggregate solely determines whether an aggregate will disappear or mature into a fruiting body (19). On their own,

Significance

Coordinated cell movement is critical for a broad range of multicellular phenomena, including microbial self-organization, embryogenesis, wound healing, and cancer metastasis. Elucidating how these complex behaviors emerge within cell populations is frequently obscured by randomness in individual cell behavior and the multitude of internal and external factors coordinating cells. This work describes a technique of combining fluorescent cell tracking with computational simulations driven by the tracking data to identify cell behaviors contributing to an emergent phenomenon. Application of this technique to the model social bacterium *Myxococcus xanthus* suggested key aspects of cell coordination during aggregation without complete knowledge of the underlying signaling mechanisms.

Author contributions: C.R.C., H.-B.S., O.A.I., and L.J.S. designed research; C.R.C. performed research; C.R.C., H.-B.S., O.A.I., and L.J.S. analyzed data; and C.R.C., H.-B.S., O.A.I., and L.J.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: shimkets@uga.edu or igoshin@rice.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1620981114/-/DCSupplemental.

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 toward specific lipids by suppressing reversals when moving up the chemical gradients (25), creating a biased walk. Exopolysaccharides, a major component of the extracellular matrix, also inhibit cell reversals in a concentration-dependent manner (26). However, inhibiting cellular production of known lipid chemoattractants does not diminish aggregation (27, 28), 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 (29), suggests possible contact-based intercellular signaling. In agreement, cells at low cell densities decrease reversal frequency as group size increases (30). However, this reversal suppression does not directly scale to the cell densities typically used in assays of development (22). Thus, whereas 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.

Although computational approaches have been extensively used in hypothesizing models of aggregation (24, 31–36), the lack of quantitative datasets 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 (35), cell alignment (35, 37), active turning (36), density-dependent speed reduction (37), physical jamming (31, 32, 34), and streaming (32, 34) 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 comparisons are rarely performed. For example, Zhang et al.'s (21) analysis of the model in which aggregation is driven by cell alignment and reduced cell speed inside aggregates (24) 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 multilevel 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 (Fig. S1). Under our conditions, aggregation begins 11–12 h after spotting the cells on starvation media. We selected an ∼5-h window that began just before the initiation of aggregation through the period when stable aggregates form (Fig. S1*A*). The beginning of this window was designated time-point zero. About 1 h into this time span aggregation becomes evident. Stable aggregates appear by 1.5 h with a few of the smaller aggregates disappearing by 5 h. Aggregation was not compromised by the use of strains expressing fluorescent proteins or prolonged fluorescent imaging (Fig. S1 *D* and *E*).

Cell-tracking algorithms were developed to track individual fluorescent cells over the 5-h window (Fig. S1*B*). Cell trajectories were subdivided into three movement states: persistent forward, persistent backward, or nonpersistent. 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 backward or forward relative to the direction observed at beginning of the trajectory. The nonpersistent state was assigned when we encountered a velocity too small (less than ∼1 μm/min) or reversal period too high (greater than ∼1 reversal per minute) 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 (Fig. S1*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.

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 (Fig. 1 *A*–*C*). Within aggregates, cells move with only a modest average speed decrease of 1.1-fold relative to outside the aggregates (Fig. 1*A*, blue asterisks). However, the probability for a cell to transition to a nonpersistent state at the end of the run increases 1.8-fold (Fig. 1*D*). Moreover, the average duration of nonpersistent runs doubles inside the aggregates (Fig. 1*B*, red asterisks). Average persistent run duration also decreases inside aggregates by ∼1.5-fold (Fig. 1*B*, blue asterisks). These effects lead to a combined (persistent and nonpersistent) 2-fold decrease in average run distance inside the aggregates vs. outside (Fig. 1*C*, magenta circles). These results are in agreement with other work suggesting that cells reduce movement inside aggregates (24) and provide much more quantitative detail.

Previous observations indicated that cells increase their movement when aggregation initiates (20, 22, 23). To quantify these effects, the mean and 95% confidence intervals for distance, duration, and speed of persistent state runs were calculated in a 20-min sliding window over the length of the experiment (Fig. 1 *E*–*G*). Early in aggregation (ca. 0–1.5 h), the mean persistent run duration outside the aggregates increases ∼1.8-fold (Fig. 1*F*, blue lines), causing an increase in run distance (Fig. 1*G*, blue lines). At ∼1.5 h, run duration transiently returns to levels seen before the onset of aggregation. Soon after, a second transitory increase in run duration occurs. As aggregates mature, run duration gradually decreases back to preaggregation levels. Inside the aggregates, speed and duration remain constant (Fig. 1 *E*–*G*, black lines). Nonpersistent run behaviors are also relatively constant, with run distance varying less than 1.5 μm over the length of the experiment (Fig. S2 *A*–*C*). The probability of transitioning to a nonpersistent state remains about the same, with the exception of a transitory increase outside the aggregates coinciding with the first peak in run duration (Fig. 1*H*). Again, our measurements not only confirm earlier observations but also 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

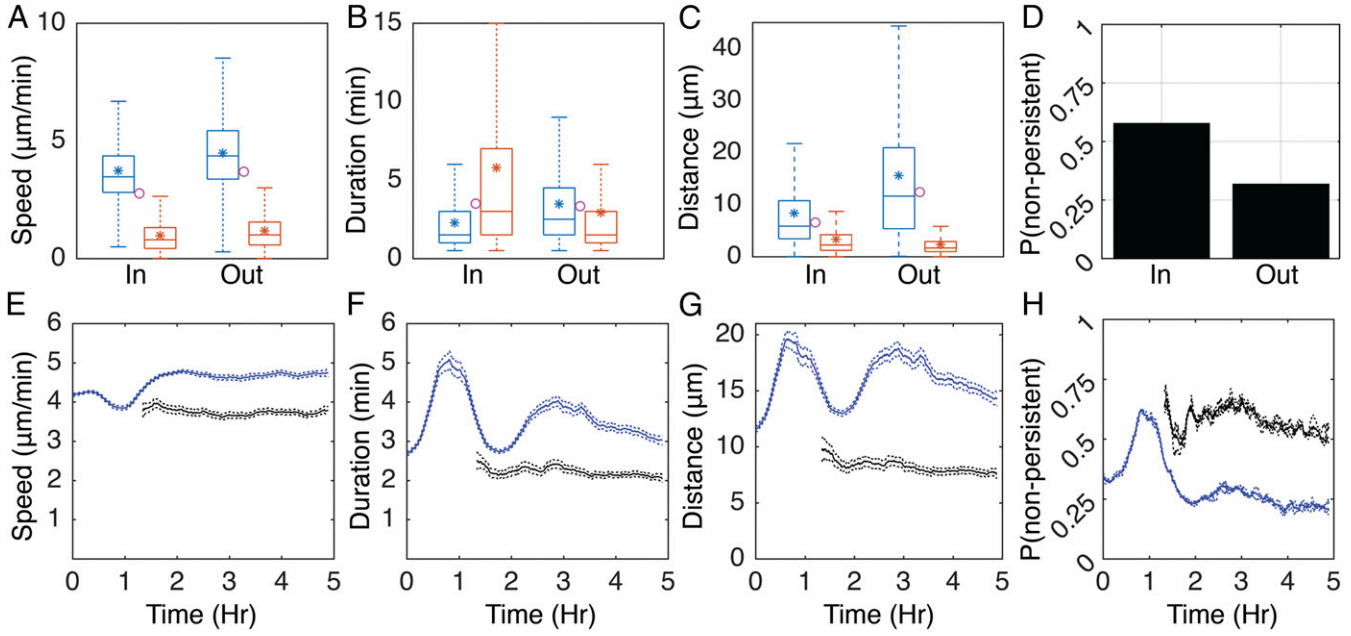

Fig. 1. Run behaviors are dynamic in time and space. (*A*–*C*) Time-integrated distributions of persistent (blue) and nonpersistent (red) run speed (*A*), duration (*B*), and distance (*C*) inside (In) and outside (Out) of the aggregates. Horizontal lines inside the boxes indicate distribution median. 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. Asterisks indicate average. Circles indicate combined (persistent and nonpersistent) average. (*D*) Time-integrated probability of choosing a nonpersistent run after a persistent run inside (In) or outside (Out) of the aggregates. (*E* and *F*) Mean (solid lines) and bootstrapped 95% confidence intervals (dashed lines) for run speed (*E*), duration (*F*), distance (*G*), and probability of choosing a nonpersistent run after a persistent run (*H*) calculated in a 20-min sliding window. Blue lines indicate runs starting outside the aggregates, and black lines, runs inside the aggregates.

dynamics. Agents move in a series of straight lines with properties (persistent vs. nonpersistent, with 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 nonpersistent 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. Because 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 (Fig. S3*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 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 (Fig. 2*A*, blue box).

Previous hypotheses of the mechanistic basis for aggregation predicted that decreased cell movement inside aggregates was the major driver of aggregate growth (21, 24, 31, 32, 38). 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 (Fig. S3*B*). In the resulting simulations, aggregates appear at nearly all expected locations (Fig. 2*B*, *Right*). However, the fraction of cells within the aggregate boundaries by the end of the 5-h window is threefold smaller in simulations compared with experimental results (Fig. 2*B*, *Left*). Addition of time dependence when choosing the state, speed, and duration (Fig. 1 *E*–*H*) does not improve the rate or completeness of aggregation in simulations (Fig. S3 *C* and *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 (21).

Cells Perform a Biased Walk Toward the Aggregate Center. Biased walks are found in many types of cell patterning (8, 39, 40). Although chemotaxis has not been implicated in *M. xanthus* aggregation, *M. xanthus* can perform biased walks up specific lipid gradients (25). Bias is created by increasing average run duration when moving up the chemoattractant gradient; conversely, cells decrease average run duration 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 (Fig. S1*A*, green ovals). The results show that persistent runs moving toward the aggregate centroid are longer than runs moving away from it (Fig. 3*A*). This bias is due to an increase in run duration rather than run speed (Fig. 3 *B* and *C*). The probability of transitioning to a nonpersistent state at the end of the run also depends on the run orientation relative to the nearest aggregate (Fig. 3*D*). Inside the aggregates,

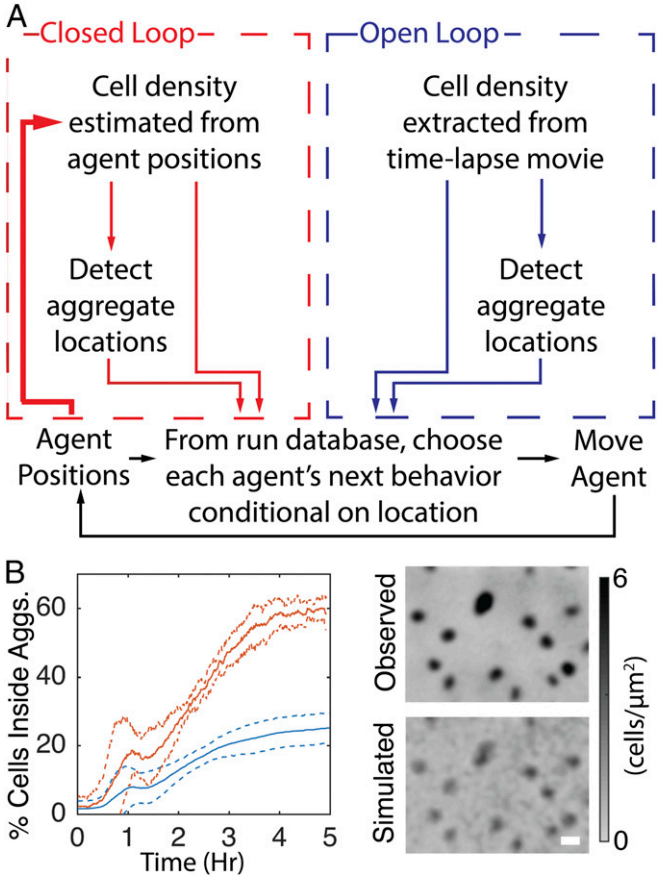

Fig. 2. Reduced movement inside aggregates is not sufficient to fully replicate aggregation in open-loop simulations. (*A*) Overview of open-loop (blue) and closed-loop (red) simulations. The extra path in the closed-loop model is in boldface type to highlight that the agent's positions feed back into 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. (*B*, *Left*) Average (solid lines) and SDs (dashed lines) of the percentage of cells inside experimentally observed aggregate boundaries for experiment (red) and simulation (blue). (*B*, *Right*) Comparison of last frame of representative experientially observed (Observed) cell density with that observed in a simulation.

nonpersistent run durations are 1.5 times longer when moving away from the aggregate centroid (Fig. 3 *E* and *F*). In contrast to a previous report of tangential cell movement inside the aggregates (41), our run durations are longest when pointed toward the aggregate centroid (Fig. S4*A*).

A Biased Walk Toward 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 (Fig. S4 *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 (Figs. S3*E*). The inclusion of the biased walk increases aggregation rate and completeness, leading to a twofold increase in the fraction of agents inside aggregates (Fig. 4*A*). Aggregate density (Fig. 4*B*) and size (Fig. 4*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 (Fig. S3*F*). In these simulations, it is necessary for 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 (Fig. S3*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 (Fig. 2*A*, red box) estimate the density profile from the agent positions by kernel density estimation (KDE) (42). Aggregates were then detected from the agent density profile using the same density cutoff as in experiments. The resulting density profile and aggregate

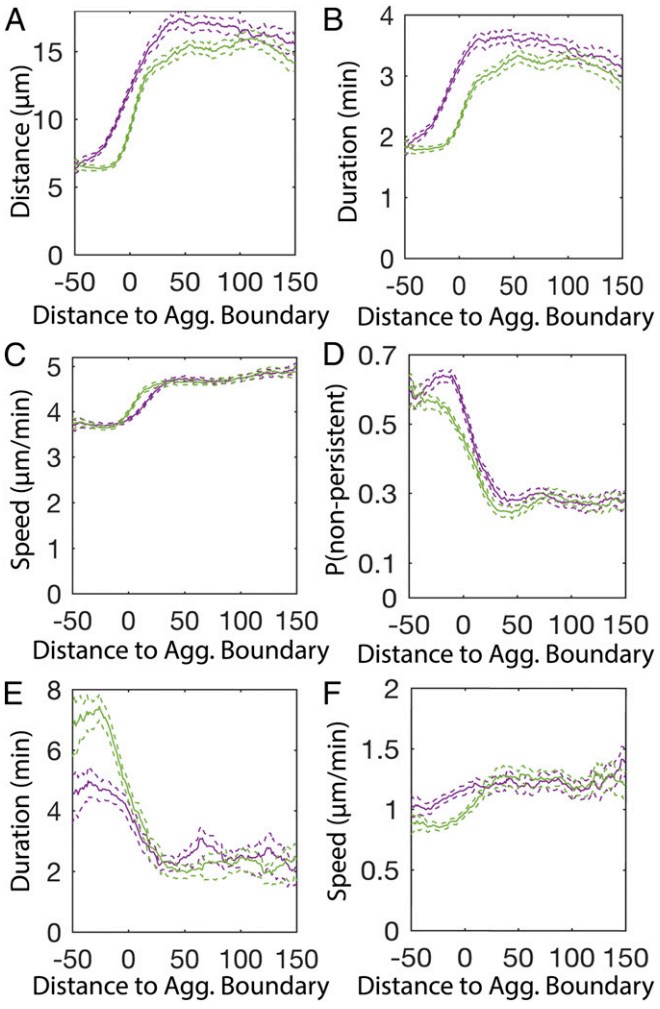

Fig. 3. Cells perform a biased walk toward aggregates. (*A*–*F*) Average (solid lines) and bootstrapped 95% confidence intervals (dashed lines) of persistent run distance (*A*), duration (*B*), speed (*C*), probability of choosing a nonpersistent run (*D*), nonpersistent duration (*E*), and distance (*F*) in a 10-μm sliding window from the beginning of the run. Runs are binned into either pointing toward [$\cos(\beta_n) > 0$ in *A*–*C*, *E*, and *F* or $\cos(\phi_{n-1}) > 0$ in *D*; Fig. S1*C*] the nearest aggregate centroid (purple lines) or pointed away [$\cos(\beta_n) < 0$ in *A*–*C*, *E*, and *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.

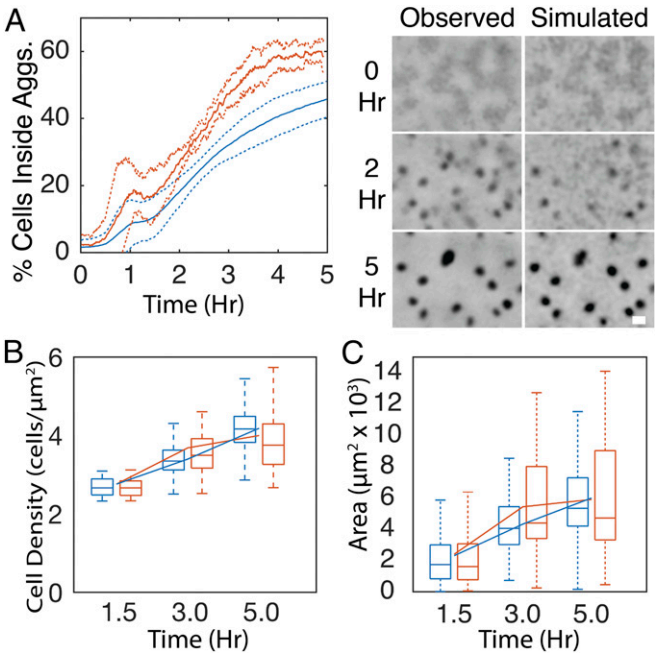

Fig. 4. A biased walk toward aggregates contributes to aggregation in open-loop simulations. (*A–C*) Comparison of experimental results (red) with simulations (blue) in which agents reduce movement proportional to cell density and perform a biased walk toward aggregates. (*A*, *Left*) Formatted as in Fig. 2*B*. (*A*, *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 Fig. 2*B*. (*B*) Distribution of average cell density inside aggregates. (*C*) Distribution of aggregate area. Box plots are formatted as in Fig. 1*A*. Line plots indicate mean. (Scale bars, 100 μm.)

boundaries were used to choose the agent run characteristics, closing the feedback loop between agent behavior and their density profile (Fig. 2*A*, boldface line). 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, compared with experimental results and open-loop simulations, the fraction of cells in aggregates decreased about twofold (Fig. 5*A*). Although the resulting average cell density inside the aggregates agrees with experiments (Fig. S5*A*), the aggregate area is smaller than in experimental results (Fig. S5*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 (16–18, 23, 24), visual inspection of cell trajectories indicates alignment between neighboring paths (Fig. S6*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 ref. 16 by calculating nematic alignment strength as the correlation of run orientations modulo 180° (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 (Fig. 5*B*). Furthermore, observations (Fig. S6, dashed boxes) and quantification of the mean run orientation relative to the nearest aggregate ($\langle\cos(2\beta_n)\rangle$, Fig. S1*C*) indicate that run vectors outside the aggregate preferentially orient in a direction radial to the nearest aggregate (Fig. 5*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 (Fig. S6*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 experiment. To allow agents time to align before the onset of aggregation, the simulation was run for 1.5 h of simulation time, using the behavior distribution and turning angles from the first 10 min of the experimental results. During this time, agent alignment approaches that seen in the experimental results (Fig. S7*A*). After the 1.5-h 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 those of the open-loop model (Fig. 5*C*). As a control, adding a prerun to simulations without neighbor alignment did not affect aggregation (Fig. S7*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 (Fig. S7*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 (Fig. 5*E*). Furthermore, aggregate cell density (Fig. 5*F*), area (Fig. 5*G*), and aggregate count (Fig. 5*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 (Fig. 6). Time dependence of the agent's turning angles was not 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 (Fig. S7*E*). As in open-loop simulations (Fig. 2*B*), removing the biased walk slows the aggregation rate (Fig. 6*B*). However, by removing the time dependencies, closed-loop simulations can be run beyond the length of the of the time-lapse movies. When simulations were continued for another 5 h, agents continue to aggregate, approaching a steady state by 10 h. Even after 10 h, the fraction of cells inside the aggregates and aggregate density are ∼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 h, Fig. 1*F*) and during rapid aggregate growth (ca. 2.5–3.4 h) suggest a possible role for time-dependent run duration. Outside the aggregates, this increase in duration leads to a combined (persistent and nonpersistent) average run distance in the earlier time window that is 1.3 times longer than in the latter (Fig. S7*F*). Inside the aggregates, run distances are

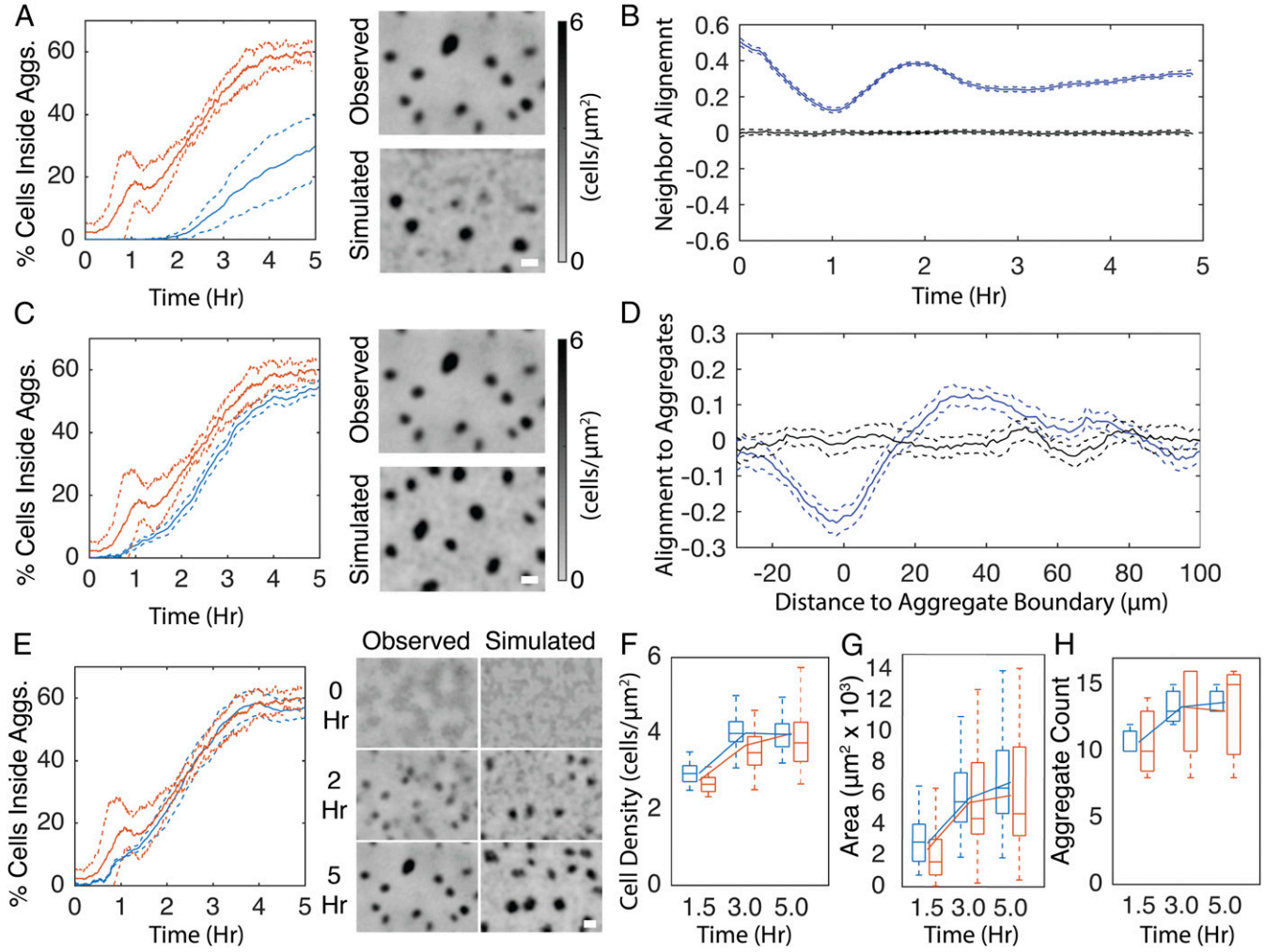

Fig. 5. Closed-loop simulations reproduce wild-type–like aggregation with the addition of cell alignment. (*A*) Simulation results in which agents reduce movement proportional to cell density and perform a biased walk toward aggregates. (*A*, *Left*) Average (solid lines) and SD (dashed lines) of the percentage of cells inside detected aggregates for experimental (red) and simulation (blue) replicates. (*A*, *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*) Percentage of cells inside aggregates as in *A*. (*E*, *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 are formatted as in Fig. 1*A*. Lines indicate mean. (Scale bars, 100 μm.)

about the same in both time windows (Fig. S7*F*). To determine the role of these changes in aggregation dynamics, we used closed-loop models in which run data only from the 1.5- to 2-h or only from the 2.5- to 3.4-h window were used to drive agents' behavior for the whole simulation duration. Models based on the short run duration window (1.5–2 h) produced aggregates at a rate and completeness equivalent to those of the experimental results (Fig. 6*C*). In contrast, agents in simulations using the longer run durations (2.5–3.4 h) aggregate at a faster rate and to a higher level of completeness than in experimental results (Fig. 6*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 toward aggregates. Using the window with longer run durations, agents formed aggregates comparable to those in experimental results in rate, size, and cell density (compare Fig. 6*A* with 6*E*). The short run 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 Fig. 6*B* with 6*F*). Removing both alignment and the biased walk all but abolished aggregation, even when using the longer run duration window (Fig. 6*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 use a

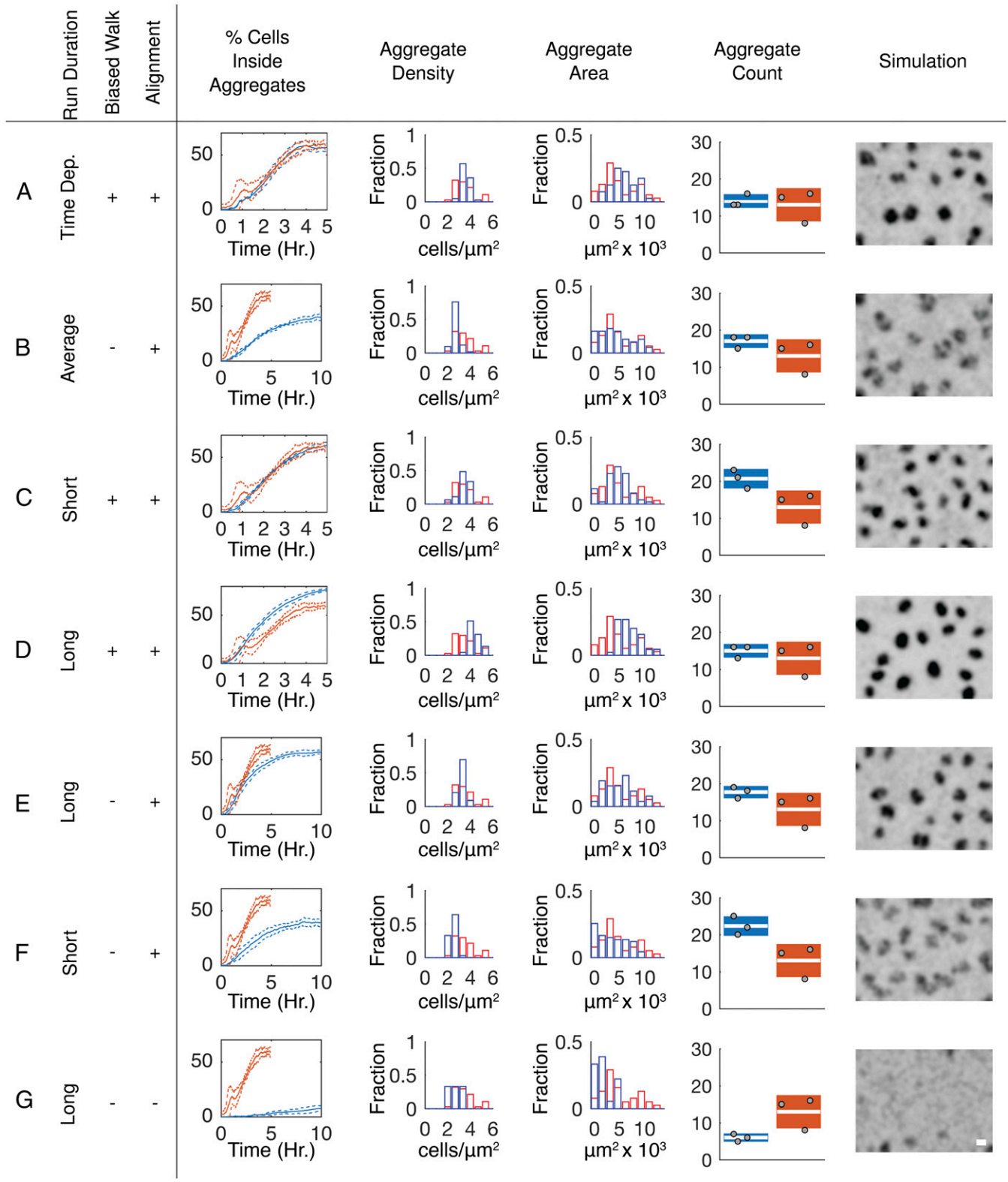

Fig. 6. Probing interactions shaping aggregation dynamics in closed-loop simulations. (*A*–*G*) Percentage of cells inside aggregates, aggregate area, cell density inside aggregates, and aggregate count from the last time point in simulations (blue) and experimental (red) results. Aggregate density and area box plots are formatted as in Fig. 1. Aggregate count box plots indicate the SD 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 Fig. 5*A*. (Scale bar, 100 μm.) (*A*) Same simulation as in Fig. 5 *E*–*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 toward aggregates. (*C*) Simulations with a biased walk, alignment, and run behaviors chosen from a time window (1.5–2 h; Fig. 1*F*) containing short run durations outside of aggregates. (*D*) Simulations as in *C* except run behaviors from a time window (2.5–3.5 h) 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.

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 (Fig. 1 *A*–*C*), we found that relatively small changes in average cell behavior, such as a 15% increase in average run duration when moving toward aggregates (Fig. 3*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 (7), 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 uses two simulation environments (Fig. 2*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 used 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 (43). 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 that used here. For example, multiphoton microscopy enables tracking of individual cells in vivo and 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 (44). Combining the microscopy and cell-tracking data with simulations in which the local microenvironments are defined a priori could be used to identify microenvironment cues of cell behavior that 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 (45, 46) 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 toward aggregates (Fig. 6*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 because *M. xanthus* contains 21 chemoreceptors, of which 13 create altered developmental phenotypes when deleted (47), and 2 are thoroughly implicated in both development and reversal control (25, 48, 49). Furthermore, these cell-reversal control pathways can react in timescales of minutes (25) 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 toward 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 used 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-grained understanding of the behaviors to a 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 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% $MgSO_4$] at 32 °C with vigorous shaking. Development was induced on 10 mL 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 (50). pCRC36 was constructed by PCR cloning the eYFP from pEYFP with primers containing 3′-NdeI and 5′-NheI restriction sites and ligated into pMR3629 (50). Strains LS3629 and LS3908 were constructed by electroporation (51) 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 1:2,500 (resulting in ∼500 individually trackable tdTomato cells within the field of view), concentrated to 1.7×10^9 cells/mL, and 35 μL of the cell mixture was spotted onto TPM agar and then dried uncovered in a 32 °C incubator. Once dry, the plates were covered, wrapped with parafilm (Bemis Inc.), 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) in the same heated room beginning at the indicated times in the TRITC channel at 200× magnification every 30 s, 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) camera, a Phoenix-D48CL frame grabber (Active Silicon), and the μManager software (52). The fluorescence intensity was set to 55%, camera gain was

set to 255, and exposure time was 600 ms. The mercury lamp was shuttered when not acquiring an image. Imaging was carried out for ∼6 h. The time point at which aggregation began varied by up to 1 h between replicates. Replicate movies were truncated to synchronize the onset of aggregation and equalize movie length as described in *SI Methods*, resulting in final movie length of 5 h. 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 (Fig. S8*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 (Fig. S8*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 *SI Methods*.

Cell Tracking. To reduce camera sensor noise and fluorescence from the growing aggregates, time-lapse images were band-pass filtered as described in *SI Methods*. Thereafter, the 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 vs. the number of cells detected (Fig. S8*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" (Fig. S8*B*, arrow) provided a good tradeoff between detection of all of the cells and little identification of background noise as cells.

To track cell motility between images, we followed procedures established in ref. 53. 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 (54) implemented in MATLAB by the authors of ref. 53 to solve the LAP (see figure 1 of ref. 53 for an overview of this process). As in ref. 53, 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 *SI Methods*.

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 nonmoving cells). To detect movement characteristics of the cell reliably, a cell-state filter was developed that 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 backward ($i=2$), and nonpersistent ($i=3$). 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–f_3 in Table S2). The EKF then uses the deviation between the predicted (s_t) state and the 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 *SI Methods*.

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 *SI Methods*.

Run Vectors. Trajectories were divided into runs, which start at the beginning of one contiguous movement state (persistent forward, persistent backward, nonpersistent) and end with the next change of state. Trajectory data before 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 the beginning of the experiment (T) were calculated for each run vector (Fig. S1*C*).

Average nematic alignment strength was used to quantify trajectory alignment (Fig. S6*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 $-\pi=\pi$. 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 (Fig. S6*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 (Fig. S9 *A* and *B*).

Bootstrapping Statistics. Where indicated, 95% confidence intervals were calculated by pooling the data from all three replicate movies and bootstrapping parameters using the adjusted percentile method (55) with 1,000 bootstrap samples.

Data-Driven Agent-Based Model. An agent-based model consisting of 10,000 agents on a rectangular domain of 986 μm × 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 in 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. Because fluorescent images for cell tracking were taken at 30-s intervals, we have adopted the same time discretization in the simulations with agents' positions along their current run vector updated every $(\Delta t=30s)$. The agents' run variables (θ_i, v_i, τ_i), along with an auxiliary binary variable denoting whether the run is persistent or nonpersistent, s_i, are drawn from the reversal probability density function (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 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 Fig. S1*C*. We used nearest-neighbor methods (56), 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 *SI Methods*. This approach incorporates directly from the experimental run database all of the information available about P without relying on an explicit reconstruction of P on a high-dimensional variable space.

We implemented two alternative modeling approaches, referred to as the open-loop model 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 *Cell Density Estimation*. 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 (Fig. S10). 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 be composed of the composite of all runs extracted from all trajectories tracked across all three microcinematography experiments (three movies) 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 used only the single-experiment database for the imaging experiment from which also the input cell density profile was extracted.

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.

ACKNOWLEDGMENTS. The research reported here was supported by the National Science Foundation under Awards MCB-1411891 (to L.J.S.) and MCB-1411780 (to O.A.I.).

1. Yang Z, Higgs PI, eds (2014) *Myxobacteria* (Caister Academic, Norfolk, UK).
2. Bretschneider T, Othmer HG, Weijer CJ (2016) Progress and perspectives in signal transduction, actin dynamics, and movement at the cell and tissue level: Lessons from *Dictyostelium*. *Interface Focus* 6:20160047.
3. Aman A, Piotrowski T (2010) Cell migration during morphogenesis. *Dev Biol* 341:20–33.
4. Solnica-Krezel L, Sepich DS (2012) Gastrulation: Making and shaping germ layers. *Annu Rev Cell Dev Biol* 28:687–717.
5. Theveneau E, Mayor R (2012) Neural crest delamination and migration: From epithelium-to-mesenchyme transition to collective cell migration. *Dev Biol* 366:34–54.
6. Friedl P, Gilmour D (2009) Collective cell migration in morphogenesis, regeneration and cancer. *Nat Rev Mol Cell Biol* 10:445–457.
7. Park S, Greco V, Cockburn K (2016) Live imaging of stem cells: Answering old questions and raising new ones. *Curr Opin Cell Biol* 43:30–37.
8. Delgado I, Torres M (2016) Gradients, waves and timers, an overview of limb patterning models. *Semin Cell Dev Biol* 49:109–115.
9. Szabó A, Mayor R (2016) Modelling collective cell migration of neural crest. *Curr Opin Cell Biol* 42:22–28.
10. Iber D, Zeller R (2012) Making sense-data-based simulations of vertebrate limb development. *Curr Opin Genet Dev* 22:570–577.
11. Schumacher LJ, Kulesa PM, McLennan R, Baker RE, Maini PK (2016) Multidisciplinary approaches to understanding collective cell migration in developmental biology. *Open Biol* 6:160056.
12. Masuzzo P, Van Troys M, Ampe C, Martens L (2016) Taking aim at moving targets in computational cell migration. *Trends Cell Biol* 26:88–110.
13. Kuspa A, Plamann L, Kaiser D (1992) Identification of heat-stable A-factor from *Myxococcus xanthus*. *J Bacteriol* 174:3319–3326.
14. Pathak DT, et al. (2012) Cell contact-dependent outer membrane exchange in myxobacteria: Genetic determinants and mechanism. *PLoS Genet* 8:e1002626.
15. Mauriello EMF, Mignot T, Yang Z, Zusman DR (2010) Gliding motility revisited: How do the myxobacteria move without flagella? *Microbiol Mol Biol Rev* 74:229–249.
16. Balagam R, Igoshin OA (2015) Mechanism for collective cell alignment in *Myxococcus xanthus* bacteria. *PLoS Comput Biol* 11:e1004474.
17. Lux R, Li Y, Lu A, Shi W (2004) Detailed three-dimensional analysis of structural features of *Myxococcus xanthus* fruiting bodies using confocal laser scanning microscopy. *Biofilms* 1:293–303.
18. Berleman JE, et al. (2016) Exopolysaccharide microchannels direct bacterial motility and organize multicellular behavior. *ISME J* 10:2620–2632.
19. Xie C, Zhang H, Shimkets LJ, Igoshin OA (2011) Statistical image analysis reveals features affecting fates of *Myxococcus xanthus* developmental aggregates. *Proc Natl Acad Sci USA* 108:5915–5920.
20. Curtis PD, Taylor RG, Welch RD, Shimkets LJ (2007) Spatial organization of *Myxococcus xanthus* during fruiting body formation. *J Bacteriol* 189:9126–9130.
21. Zhang H, et al. (2011) Quantifying aggregation dynamics during *Myxococcus xanthus development*. *J Bacteriol* 193:5164–5170.
22. Jelsbak L, Søgaard-Andersen L (2002) Pattern formation by a cell surface-associated morphogen in *Myxococcus xanthus*. *Proc Natl Acad Sci USA* 99:2032–2037.
23. Thutupalli S, Sun M, Bunyak F, Palaniappan K, Shaevitz JW (2015) Directional reversals enable *Myxococcus xanthus* cells to produce collective one-dimensional streams during fruiting-body formation. *J R Soc Interface* 12:20150049.
24. Sliusarenko O, Zusman DR, Oster G (2007) Aggregation during fruiting body formation in *Myxococcus xanthus* is driven by reducing cell movement. *J Bacteriol* 189:611–619.
25. Kearns DB, Shimkets LJ (1998) Chemotaxis in a gliding bacterium. *Proc Natl Acad Sci USA* 95:11957–11962.
26. Zhou T, Nan B (2017) Exopolysaccharides promote *Myxococcus xanthus* social motility by inhibiting cellular reversals. *Mol Microbiol* 103:729–743.
27. Kearns DB, et al. (2001) Identification of a developmental chemoattractant in *Myxococcus xanthus* through metabolic engineering. *Proc Natl Acad Sci USA* 98:13990–13994.
28. Bode HB, et al. (2006) Straight-chain fatty acids are dispensable in the myxobacterium *Myxococcus xanthus* for vegetative growth and fruiting body formation. *J Bacteriol* 188:5632–5634.
29. Kim SK, Kaiser D (1990) Cell alignment required in differentiation of *Myxococcus xanthus*. *Science* 249:926–928.
30. Shi W, Ngok FK, Zusman DR (1996) Cell density regulates cellular reversal frequency in *Myxococcus xanthus*. *Proc Natl Acad Sci USA* 93:4142–4146.
31. Igoshin OA, Welch R, Kaiser D, Oster G (2004) Waves and aggregation patterns in myxobacteria. *Proc Natl Acad Sci USA* 101:4256–4261.
32. Sozinova O, Jiang Y, Kaiser D, Alber M (2005) A three-dimensional model of myxobacterial aggregation by contact-mediated interactions. *Proc Natl Acad Sci USA* 102:11308–11312.
33. Sozinova O, Jiang Y, Kaiser D, Alber M (2006) A three-dimensional model of myxobacterial fruiting-body formation. *Proc Natl Acad Sci USA* 103:17255–17259.
34. Holmes AB, Kalvala S, Whitworth DE (2010) Spatial simulations of myxobacterial development. *PLoS Comput Biol* 6:e1000686.
35. Starruß J, Bley T, Søgaard-Andersen L, Deutsch A (2007) A new mechanism for collective migration in *Myxococcus xanthus*. *J Stat Phys* 128:269–286.
36. Hendrata M, Yang Z, Lux R, Shi W (2011) Experimentally guided computational model discovers important elements for social behavior in myxobacteria. *PLoS One* 6:e22169.
37. Sliusarenko O, Neu J, Zusman DR, Oster G (2006) Accordion waves in *Myxococcus xanthus*. *Proc Natl Acad Sci USA* 103:1534–1539.
38. Gejji R, Lushnikov PM, Alber M (2012) Macroscopic model of self-propelled bacteria swarming with regular reversals. *Phys Rev E Stat Nonlin Soft Matter Phys* 85:021903.
39. Lander AD (2013) How cells know where they are. *Science* 339:923–927.
40. Morelli LG, Uriu K, Ares S, Oates AC (2012) Computational approaches to developmental patterning. *Science* 336:187–191.
41. Sager B, Kaiser D (1993) Two cell-density domains within the *Myxococcus xanthus* fruiting body. *Proc Natl Acad Sci USA* 90:3690–3694.
42. Botev ZI, Grotowski JF, Kroese DP (2010) Kernel density estimation via diffusion. *Ann Stat* 38:2916–2957.
43. Clark AG, Vignjevic DM (2015) Modes of cancer cell invasion and the role of the microenvironment. *Curr Opin Cell Biol* 36:13–22.
44. Ellenbroek SIJ, van Rheenen J (2014) Imaging hallmarks of cancer in living mice. *Nat Rev Cancer* 14:406–418.
45. Lock JG, et al. (2014) Plasticity in the macromolecular-scale causal networks of cell migration. *PLoS One* 9:e90593.
46. Bastos AM, Schoffelen J-M (2016) A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front Syst Neurosci* 9:175.
47. Moine A, et al. (2014) Functional organization of a multimodular bacterial chemosensory apparatus. *PLoS Genet* 10:e1004164.
48. Xu Q, Black WP, Cadieux CL, Yang Z (2008) Independence and interdependence of Dif and Frz chemosensory pathways in *Myxococcus xanthus* chemotaxis. *Mol Microbiol* 69:714–723.
49. Curtis PD, Geyer R, White DC, Shimkets LJ (2006) Novel lipids in *Myxococcus xanthus* and their role in chemotaxis. *Environ Microbiol* 8:1935–1949.
50. Iniesta AA, García-Heras F, Abellón-Ruiz J, Gallego-García A, 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.
51. Kashefi K, Hartzell PL (1995) Genetic suppression and phenotypic masking of a *Myxococcus xanthus frzF-* defect. *Mol Microbiol* 15:483–494.
52. Edelstein AD, et al. (2014) Advanced methods of microscope control using μManager software. *Protocol* 1:1–10.
53. Jaqaman K, et al. (2008) Robust single-particle tracking in live-cell time-lapse sequences. *Nat Methods* 5:695–702.
54. Jonker R, Volgenant A (1987) A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing* 38:325–340.
55. Davison AC, Hinkley DV (1997) *Bootstrap Methods and Their Application* (Cambridge Univ Press, New York).
56. Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning* (Springer, New York), 2nd Ed.
57. Chen H, Keseler IM, Shimkets LJ (1990) Genome size of *Myxococcus xanthus* determined by pulsed-field gel electrophoresis. *J Bacteriol* 172:4206–4213.
58. Challa S, Morelande MR, Musicki D, Evans RJ (2011) *Fundamentals of Object Tracking* (Cambridge Univ Press, New York).
59. Crocker J, Grier D (1996) Methods of digital video microscopy for colloidal studies. *J Colloid Interface Sci* 310:298–310.
60. Freedman D, Diaconis P (1981) On the histogram as a density estimator: L 2 theory. *Probab Theory Relat Fields* 57:453–476.