An individual-based GPU simulation framework for collective bacterial dynamics in swarms and biofilms

by

Rachel Verla Mok

B.S., Purdue University Calumet (2011)
M.S., Massachusetts Institute of Technology (2013)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Signature redacted

Author ....
Department of Mechanical Engineering
May 10, 2019

Signature redacted

Certified by..............................
Jörn Dunkel
Associate Professor
Thesis Supervisor

Signature redacted

Accepted by ....
Nicolas G. Hadjiconstantinou
Chairman, Department Committee on Graduate Theses

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 13 2019
LIBRARIES
ARCHIVES

An individual-based GPU simulation framework for collective bacterial dynamics in swarms and biofilms

by

Rachel Verla Mok

Submitted to the Department of Mechanical Engineering
on May 10, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

With recent technological advancements, observations and measurements of complex bacterial communities at single-cell resolution are now possible. Guided by these rich experimental data sets, we develop minimal individual-based models to uncover the governing forces driving the dynamics in microbial systems. Our model incorporates the biophysical processes of cell growth and division, viscous drag, bacteria self-propulsion, and mechanical cell-surface and cell-cell interactions through interaction potentials. In particular, our cell-cell interaction potential accounts for hard steric and osmotic repulsion as well as attraction mediated through secreted components which bind cells together. Implementing this model on graphics processing units (GPUs) such that the computational time scales linearly with the system size, we achieve a 10x speedup over a comparable code written on central processing units (CPUs). With this simulation framework, we investigate the collective dynamics of *Bacillus subtilis* swarm expansion and *Vibrio cholerae* biofilm formation. Our experimental and numerical results imply that mechanical cell-cell interactions dominate the swarming motility phases and can account for the emergence of order and structure seen in growing biofilms. Furthermore, this model is used to explore the effectiveness of surface topography on deterring biofilm formation by investigating how locally varying boundary curvature impact the scattering and accumulation dynamics of swimming bacteria. This work shows great promise at increasing our understanding of the physics governing microbial communities, which knowledge is essential to control and inhibit bacterial populations.

Thesis Supervisor: Jörn Dunkel
Title: Associate Professor

Acknowledgments

This thesis would not have been possible without the guidance and support of my advisor, Jörn Dunkel. Jörn, thank you. Thank you for introducing me to this project, which better aligned with my interests, and seeing me through it. I have benefited much from your tutelage. When I was getting ready to have my son and after he was born, Jörn was and continues to be incredibly encouraging, accommodating, and understanding, for which I'm incredibly grateful because it allowed me to enjoy being a new parent in the midst of being a graduate student. I'm also grateful to my other committee members Triantaphyllos R. Akylas and Pierre Lermusiaux for bringing their unique perspectives and insights to my work.

I'm indebted to the many collaborators I've worked with in the various projects that constitute this thesis. I'm especially grateful to Hannah Jeckel, Eric Jelli, Raimo Hartmann, Praveen K. Singh, Knut Drescher, and the other members of the Drescher lab. Their work amounted to an experimental tour de force which allowed for the single-cell observation of bacterial swarm expansion and biofilm formation. These rich data sets were instrumental in steering our modeling efforts. Thanks also to Jan Frederik Totz for performing additional data analysis for the bacterial swarming project. I'm also indebted to Vasily Kantsler for performing the bacteria surface experiments, which allowed us to test theories regarding the effects of boundary curvature on the accumulation dynamics of swimming bacteria. I'm also grateful to the other members of the Dunkel lab for their helpful discussions, and especially for Norbert Stoop and Jonasz Słomka for welcoming me into the group. Special thanks to Boya Song and Philip Pearce whose efforts and insights helped bring the biofilm project across the finish line.

Thanks also to all the people I've worked with over the years as a teaching assistant for 2.005. I'm especially grateful to Pierre Lermusiaux, Jacopo Buongiorno, Kripa Varanasi, Jeffrey Moran, Betar Gallant, Ming Guo, and the other members of the teaching staff for instructing me on what makes an enthusiastic, kind, engaging, and clear teacher.

Thanks also to the many friends I've made here in Boston. Thank you for your kindness, love, support, and for enriching my life.

I am very grateful to my mom and dad for their unwavering love and support which has seen me through all the good and challenging times. They saw in me something I did not see in myself and encouraged me to reach high. My parents have made countless sacrifices to help me finish this PhD, for which I'm eternally grateful. Thanks also to my brothers for their love and humor. I owe much of the person I am today to my family.

Lastly, I must thank my husband Nathan whose constant love, support, patience, insights, and kindness helped me through this journey. I am so grateful for our marriage and for the many sacrifices that he has made to help me complete this thesis. During my time here, our son Samuel was born. Sammy brings so much joy into our lives. I am so very blessed to have the two of you. You both make me happier than I have ever known, and I am a better person because of you. I dedicate this thesis to my family.

My graduate work was generously supported by a James S. McDonnell Foundation Complex Systems Scholar Award, an Edmund F. Kelly Research Award, a MIT-Germany MISTI seed grant, a MIT ODGE Childbirth Accommodation Fund, a MIT OGE Chyn Duog Shiah Memorial Fellowship, and the MIT Mechanical Engineering Department.

Contents

1 Introduction **25**

2 Mechanistic individual-based model **29**

2.1 Growth and division 30
2.2 Bacteria-surface interactions 32
2.3 Bacteria-bacteria interactions 33
 2.3.1 Berne-Pechukas potential 34
 2.3.2 Generalized strength and range parameters 36
 2.3.3 Cell-cell interaction potential 37
2.4 Translation and orientation dynamics 41
2.5 Implementation 43
 2.5.1 Code development 44

3 Bacterial swarming **47**

3.1 Experiments 48
3.2 Simulations 52
 3.2.1 Implementation 55
 3.2.2 Parameters 55
3.3 Results 56

4 Bacterial biofilms **61**

4.1 Experiments 62
4.2 Simulations 65

4.2.1 Implementation 67
4.2.2 Parameters 67
4.3 Results 79
4.3.1 Cell-cell translational forces and rotational dynamics 80
4.3.2 Fitting the WT* 81
4.3.3 Prediction of larger biofilms 83

5 Bacterial surface accumulation 85

5.1 Experiments 87
5.2 Simulations 88
5.2.1 Brownian Dynamics (BD) 88
5.2.2 Run and Tumble (RT) 89
5.2.3 Implementation 90
5.2.4 Parameters 91
5.3 Results 92
5.3.1 Tracking data 92
5.3.2 Optimal sinusoidal boundaries 94
5.3.3 Sinusoidal vs. concave geometries 96

6 Concluding remarks 101

A von Mises-Fisher Distribution 107

B Derivation of the Berne-Pechukas potential 109

C Derivation of generalized strength and range parameters 115

List of Figures

2-1 Parameters related to the growth and division of the cells. When $\ell_p = \ell_{\text{divide}}$, the cell divides. The daughter cells have lengths of ℓ_{d1} and ℓ_{d2} ($\ell_{d1}+\ell_{d2} = \ell_p$, $\ell_{d1} = X\ell_p$ where X is a random number drawn from a normal distribution of mean 0.5 and standard deviation $\sigma_{\text{birth size}}$). $\mathbf{x}_{d1}$ and $\mathbf{x}_{d2}$ are the centers of the daughter cells, calculated as in Eqs. (2.4). $\hat{\mathbf{n}}_{d1}$ and $\hat{\mathbf{n}}_{d2}$ are the orientations of the daughter cells, drawn from a vMF distrbution with mean direction $\hat{\mathbf{n}}_p$ and concentration parameter κ. 31

2-2 Geometric parameters of the overlap coordinate. In the three cases shown, the overlap coordinate $\mathbf{z}_\alpha > 0$, representing contact with the surface. 33

2-3 Plot of $G(\mathbf{x}) = \exp(-\mathbf{x}^\top \mathbf{A}^{-1}\mathbf{x})$ where $\mathbf{A} = \ell^2 \hat{\mathbf{n}}\,\hat{\mathbf{n}}^\top + d^2(\mathbf{I} - \hat{\mathbf{n}}\,\hat{\mathbf{n}}^\top)$ in which $\hat{\mathbf{n}} = \hat{e}_z$. This simplifies to $G(x, y, z) = \exp\left(-\frac{(x^2+y^2)}{d^2} - \frac{z^2}{\ell^2}\right)$. In the plot, $\ell = 2$ and $d = 1$. $G(\mathbf{x})$, which is ellipsoidal in shape and centered on the origin, has the highest values in the center and gradually decays to zero in all directions. 34

2-4 Geometry for the overlap potential between neighboring cells α and β with orientations $\hat{\mathbf{n}}_\alpha$ and $\hat{\mathbf{n}}_\beta$, respectively. α is located at the origin O. The vector $\mathbf{r}_{\alpha\beta}$ points from the center of α to the center of β. 35

2-5 Cell-cell interaction potential parameters. ϵ_0 characterizes the strength of the interaction. λ_r is the range of the repulsion. ν is the relative strength of the RbmA-mediated attraction. λ_a is the effective attractive range. ρ_a is the position of the attractive potential well. $\rho_{\alpha\beta} = r_{\alpha\beta}/\sigma$ is the cell-cell distance normalized by the range parameter σ. Figure from [61]. 39

2-6 Computational time comparison between the CPU (12 processors) and GPU (2816 processors) implementation for the biofilm project, which includes bacterial growth and division and translation and orientation dynamics. The translation and orientation dynamics include cell-surface and cell-cell interactions. The same simulation input parameters are used for the CPU and GPU codes, with both simulations initially starting with one cell. The GPU implementation is significantly faster than the CPU implementation as the number of cells increase. 44

2-7 Computational tile of size $P \times P$ $(P < N)$. The computational tile facilitates the evaluation of the pair-wise interactions. The rows and columns of the computational tile are evaluated in parallel and in sequential order, respectively. Adapted from [99]. 45

2-8 Inputs and outputs of the computational tile of size $P \times P$ $(P < N)$. Inputs to the computational tile include the current sums of $\nabla_{\mathbf{x}_\alpha} V_\alpha$ and $\nabla_{\hat{\mathbf{n}}_\alpha} V_\alpha$ and the state variables ($\mathbf{x}$, $\hat{\mathbf{n}}$, ℓ, d) for the α particle. Loaded into shared memory are the state variables for the β particle. The output of the computational title are the updated sums of $\nabla_{\mathbf{x}_\alpha} V_\alpha$ and $\nabla_{\hat{\mathbf{n}}_\alpha} V_\alpha$. Adapted from [99]. 46

3-1 Adaptive microscopy reveals complete multiscale dynamics of bacterial swarm expansion. (a) Movies at single-cell resolution are acquired at different locations in the swarm, starting from 1 cell to 100 cells, and follow swarm expansion until the agar plate is completely colonized. The number of movies and locations where movies are acquired (indicated by colored squares) are determined adaptively, depending on the detected swarm size. (b)–(d), Top show qualitatively different bacterial behavioral dynamics observed at distinct space-time points, which are marked in (c)–(h) by corresponding magenta symbols. (b)–(d), Bottom demonstrate automated extraction of single-cell positions, orientations, and cell velocities (b) as well as collective behaviors, such as formation of nonmotile clusters (c) and motile rafts (d), corresponding to groups of aligned cells that move in the same direction. Cells assigned to the same nonmotile cluster or motile raft by the classification algorithms share the same color; cells labeled in white have not been identified as belonging to any motile raft or nonmotile cluster. Magenta arrows in (d) indicate the average velocity of a raft. Scale bars 10 μm. (e) Heatmap of the cell density, obtained by averaging single-cell data as in (b)–(d) for each movie at each space-time coordinate. The lag phase, a period following inoculation during which the swarm does not expand, as well as the expansion phase, is indicated. (f)–(h) Additional heatmaps for the cell speed, fraction of cells that are in nonmotile clusters in a given field of view, and fraction of cells that are in motile rafts. A total of 23 statistical observables (Table 3.1) analogous to (e)–(h) are determined at each space-time position. Figure from [66]. 49

3-2 Machine learning the swarming phases from microscopic dynamics. (a) Raw data of one swarm expansion experiment, consisting of ~1,500 space-time points (columns) in a 23-dimensional observation space (rows). Additional replicates are shown in SI Appendix, Fig. S11 of [66]. Color bar indicates relative magnitudes scaled to [0, 1]. In the case of strongly correlated observables with high normalized mutual information (marked by red brackets), only one of them is included in the machine-learning analysis. (b) The values of the 14 remaining observables (rows) are binned into five categories as indicated by the color bar, providing the input data for machine learning. (c) The 2D representation of the data in (b), obtained with t-SNE; k-means clustering robustly identifies five main dynamical phases during swarm expansion across independent experiments ($n = 3$; SI Appendix, Figs. S13–S16 and SI Text of [66]). Phases are labeled with different colors. (d) The emergence of the different phases in time and space during swarm expansion. The dashed lines indicate the approximate outlines of the different phases. Figure adapted from [66]. 52

3-3 Swarming dynamics are dominated by physical cell-cell interactions. (a) Typical images for the phases identified in Fig. 3-2(c): low-density single-cell phase (SC); high-density rafting phase (R) with a high percentage of comoving cells; biofilm phase (B) characterized by long, unseparated cells; and coexistence phases that contain single cells and rafts (SC + R) or rafts and biofilm precursors (R + BP). Scale bar 10 μm. (b) For each phase, simulations were run with the cell shape, motility, and density extracted from the particular phase as input parameters. (c) Detailed quantitative comparisons between experiments (small circles), the particular experimental states shown in (a) (large circles), and simulations (squares; error bars are SDs, $n = 20$) yield good quantitative agreement, except for the B phase, confirming that physical effects determine the four motility-based swarming phases. Figure adapted from [66]. 57

3-4 Experimentally determined data (small symbols) and simulation results with (large square) and without (large circles) hydrodynamic interactions. Simulation results are averaged over 20 runs, with the standard deviation indicated by black lines. In each panel, small circles, squares, and diamonds correspond to the results of the three different experimental replicates. The 5 different colors correspond to the 5 different dynamical phases. Figure from [66]. 59

4-1 Dynamics of *V. cholerae* biofilm formation. (a) Cells constitutively expressing a green fluorescent protein (sfGFP) were imaged with spinning disc confocal microscopy. Images at three different z planes are highlighted. (b) 3D reconstruction of the biofilm shown in (a), where each cell is colored according to the nematic order parameter $S = \langle 3/2(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)^2 - 1/2 \rangle$ in its vicinity. High-time-resolution ($\Delta t = 5-10$ min) imagining allows tracking of cell lineages and discrimination of cells (white) that are not direct descendants of the biofilm founder cell. (c) The extracellular matrix protein RbmA mediates cell-cell adhesion and is distributed throughout the biofilm, as visualized by immunofluorescence. (d) Time-resolved WT* biofilm (a rugose wild-type strain with straight cell shape) growth series. Each cell is colored according to the cellular alignment with the z axis (for the $\Delta rbmA$ mutant see Supplementary Fig. 6 in [61]). Heatmaps showing spatially resolved single-cell measurements of different biofilm structural properties inside (e) WT* and (f) $\Delta rbmA$ biofilms, which are used to characterize biofilm formation ($n > 3$ biofilms, standard deviations are shown in Supplementary Figs. 5 and 7 and the differences among both strains are highlighted in Supplmentary Fig. 8 of [61]) as a function of the distance to the biofilm center (d_{centre}) and the number of cells inside the biofilm (N_{cells}). Figure from [61]. 63

4-2 Experimental (left) and simulation (right) histograms of the dimensionless length $\ell^* = \ell/d$ at $N_{\text{total}} = 300$ cells. The histograms are normalized by dividing the number of cells in each bin by the total number of cells such that the height of the bars sum to 1. The experiments are at flow rate 0.1 μL/min. The averages of the mean and standard deviation of ℓ^* of the 3 experiments are 4.99 and 1.53, respectively. For the 3 simulations, the averages of the mean and standard deviation of ℓ^* are 4.61 and 1.08, respectively. 68

4-3 Characteristic biofilm feature vector for several biofilms of the WT* (red) and the Δ*rbmA*-mutant (blue). The parameters correspond to the ones listed in Table 4.3, where relevant ones are spatially resolved into the core and shell of the biofilm. Figure from [61]. 72

4-4 Parameter scan to test the influence of cell-cell repulsion on the biofilm architechture. Mean MSD values between feature vectors of Δ*rbmA* biofilm experiments ($n > 3$ biofilms) and simulations upon variation of cell-cell interaction strength ϵ_0 and repulsion range λ_r. Inset: effect of parameter variation on the cell-cell interaction potential. Figure from [61]. 73

4-5 Resulting simulated biofilms for different levels of cell-cell repulsion. The figures shows a subset of biofilm renderings corresponding to a range of different values for ϵ_0 and λ_r. The color of each cell corresponds to the nematic order. Figure from [61]. 74

4-6 Parameter scan to test the influence of hard steric cell-cell repulsion on biofilm phenotype. The figure shows MSD values upon variation of $\lambda_{r,\text{steric}}$ and ν_{steric}. Owing to its longer length scale, soft, osmotic pressure-mediated repulsion prevents cells from getting close enough to interact directly through hard steric repulsion. Figure from [61]. . 75

4-7 Increased RbmA production (achieved by increasing the arabinose concentration) decreases the average cell-cell distance in biofilms. Without arabinose, no RbmA is produced and the biofilm architecture is identical to the Δ*rbmA* mutant ($n > 3$ biofilms). Figure adapted from [61]. 76

4-8 Parameter scan to test the influence of RbmA-mediated cell-cell attraction on biofilm phenotype. MSD values for feature vectors of biofilms with arabinose-inducible *rbmA* expression grown at 0.5% arabinose and simulations upon variation of cell-cell interaction strength ν, and well width λ_a and well position ρ_a. Inset: effect of parameter variation on the cell-cell interaction potential. Figure from [61]. 76

4-9 Resulting simulated biofilms for different levels of RbmA-mediated cell-cell attraction. The figure shows biofilm renderings corresponding to selected values of ν and ρ_a while keeping λ_a constant at $\lambda_a = 0.16$. Each cell is colored according to the nematic order. Figure from [61]. 77

4-10 Fitting of the attractive potential parameters. The graphs show the approximate position of the MSD minimum in (ν, ρ_a, and λ_a)-space of the 5% best-matching simulations at increasing arabinose concentrations, corresponding to increasing amounts of secreted RbmA. Each data point refers to the median position of the minimum and the error bars indicate the lower and upper quartile. The position of the WT* is shown in red (solid lines: median values, dashed lines: lower/upper quartiles). Figure adapted from [61]. 78

4-11 Cell-cell interactions of various cell-cell orientations for the $\Delta rbmA$ biofilm, which are modeled by a repulsive interaction potential. The best-fit U for these biofilms has interaction parameters $\epsilon_0 = 10\epsilon_{\text{flow}}(5 \times 10^{-20}\text{J}$, $\lambda_r = 1.65$, and $\nu = 0$. (a) Translational cell-cell interaction forces (red, attractive; blue, repulsive). (b) Rotational cell-cell interaction dynamics of cell β (dashed cell), $|d\hat{\mathbf{n}}/dt| = |(\mathbf{I}-\hat{\mathbf{n}}_\beta\hat{\mathbf{n}}_\beta^\top)(-\boldsymbol{\Omega}\,\partial U/\partial\hat{\mathbf{n}}_\beta)|$, in the case where cell α (solid cell) is held in place (red, clockwise rotation; blue, anticlockwise rotation). The dashed cells are plotted at the average cell-cell distance obtained from the $\Delta rbmA$ experiments. Figure adapted from [61]. 80

4-12 Cell-cell interactions of various cell-cell orientations for a particular level of cell-cell adhesion (0.5% arabinose), which are modeled by an attractive and repulsive interaction potential. The best-fit U for these biofilms has interaction parameters $\epsilon_0 = 10\epsilon_{\text{flow}}(5 \times 10^{-20}\text{J}$, $\lambda_r = 1.65$, $\nu = 0.1$, $\lambda_a = 0.16$, and $\rho_a = 2.75$. (a) Translational cell-cell interaction forces (red, attractive; blue, repulsive). (b) Rotational cell-cell interaction dynamics of cell β (dashed cell), $|d\hat{\mathbf{n}}/dt| = |(\mathbf{I} - \hat{\mathbf{n}}_\beta \hat{\mathbf{n}}_\beta^{\mathsf{T}})(-\boldsymbol{\Omega}\,\partial U/\partial \hat{\mathbf{n}}_\beta)|$, in the case where cell α (solid cell) is held in place (red, clockwise rotation; blue, anticlockwise rotation). RbmA-mediated cell-cell adhesion gives rise to an attractive part (red), acting within the range of experimentally determined average cell-cell distance (dashed cell). Figure adapted from [61]. 81

4-13 (a) Best-fit simulation parameters for varying RbmA and arabinose concentrations (black dots) follow a line in (ν, λ_a, ρ_a) parameter space and cross isosurfaces of average cell-cell distance (see color bar and compare with Fig. 4-7; see Fig. 4-10 for more details about the fitting). The RbmA level of the WT* biofilms is inferred in terms of an effective arabinose concentration by locating the WT* along the line of different arabinose concentrations (blue point), which is very close to the best fit of the WT* (red point). (b) Simulated (best fit) versus experimental WT* biofilm. (c) Comparison of biofilm architectural properties for the WT* experiment (blue) and the WT* simulation prediction (yellow). The architectural properties are spatially resolved for the core (top row) and shell (bottom row) of the biofilm (experiment, $n = 7$; simulation, $n = 10$). Figure adapted from [61]. 82

4-14 Predicting phenotypes of larger biofilms. (a) Rendered WT* experimental (left) and simulated biofilm (right) for approximately 1000 cells. Bottom: normalized MSD of feature vectors for experimental ($n = 7$) and simulated biofilms ($n = 3$). (b) Rendered $\Delta rbmA$ experimental (left) and simulated biofilm (right) for approximately 1000 cells. Bottom: normalized MSD of feature vectors for experimental ($n = 4$) and simulated biofilms ($n = 3$). Error bars are calculated by Gaussian error propagation of the standard deviations of the individual measurements. Figure adapted from [61]. 83

5-1 Typical trajectories of swimming cells for flat, sinusoidal, and semicircle surface geometries as observed in experiment and simulations. The start and end of each trajectory are indicated by the yellow and red circle, respectively. Each trajectory is 10 s long. Bacteria align with the surface in the flat geometry leading to significant surface accumulation for the experiment and simulation models. The sinusoidal ($A = 5.25$ μm, $\lambda = 28$ μm) and concave semicircle ($R = 12$ μm) surface geometries redirect the bacteria away from the surface in the experiments and simulations. Scale bars 10 μm. Figure from [94]. 93

5-2 Segmented raw data for the experiment and simulations, used in the statistical analysis. The segmented trajectories are acquired at 10 fps. The experimental raw data exhibit higher curvature than the simulation raw data likely due to hydrodynamic effects, which are not accounted for in the simulations. Scale bar 10 μm. Figure from [94]. 94

5-3 Mean bacteria surface accumulation for the sinusoidal surface over a range of amplitudes A and wavelengths λ. Accumulation at the surface is measured by comparing the number of bacteria within 5 μm from the surface to the number of bacteria in the same area 50 μm away from the surface (Fig. 5-4). (a) The location of the circles indicate the 20 combinations (A, λ) of the scan, and the size of the circle represents the standard deviation of each point. The white numbers indicate the number of experiments per point. (b, c) 3 simulations were performed for the same pairs (A, λ) as in the experiments and bilinearly interpolated. The BD and RT simulations agree qualitatively with experiment, revealing an optimum max curvature that reduces accumulation. The set of parameters corresponding to the optimum curvature k^* is delineated by the white curve $A = (k^*/4\pi^2)\lambda^2$ where $k^* = 0.31$ μm^{-1}. Typical images for the BD and RT simulations are shown in (d) and (e), respectively, for $A = 7$ μm, $\lambda = 21$ μm (circle), $A = 5.25$ μm, $\lambda = 28$ μm (square), and $A = 1.75$ μm, $\lambda = 49$ μm (triangle). Scale bars 10 μm. Figure from [94]. 95

5-4 Sampled raw data and accumulation histograms for the flat, sinusoidal ($A = 5.25$ μm, $\lambda = 28$ μm), and concave semicircle ($R = 12$ μm) surface geometries for the experiment and simulations. To visualize the spatial cell distributions, the raw data, acquired at 10 fps for 5 min, were projected onto a single wavelength (the flat surface is assumed to have the same wavelength as the semicircle surface) and sampled such that the bulk density is the same in all cases (a) - (c), (e) - (g), and (i) - (k). Both the experiments and simulations qualitatively show a depletion zone above the boundary for the sinusoidal and semicircle geometries. Due to the differences in the surface accumulation, the total cell numbers differ for the three geometries. The accumulation histograms (d), (h), and (l) quantify this effect, with accumulation defined is the ratio of the number of bacteria in each surface bin area (grey region for first bin) to the number of bacteria in a congruent area 50 μm away from the surface (blue region). The results are independent of the shape of the bulk reference area (see Fig. 5-5). Histograms (d), (h) and (l) were computed from 20 independent subsamples of the raw data. The black dashed line indicates the bulk accumulation value of 1. The accumulation histograms show that the concave semicircle geometry is the most efficient at suppressing accumulation in the experiment and simulations. Bin width 5 μm. Scale bars 5 μm. Figure from [94]. 97

5-5 Sampled raw data and surface accumulation bar graphs for the flat, sinusoidal ($A = 5.25$ μm, $\lambda = 28$ μm), and semicircle ($R = 12$ μm) surface geometries for the experiment and simulations. To visualize the spatial cell distributions, the raw data, acquired at 10 fps for 5 min, were projected onto a single wavelength (the flat surface is assumed to have the same wavelength as the semicircle surface) and sampled such that the bulk density is the same in all cases (a) - (c), (e) - (g), and (i) - (k). Due to the differences in the surface accumulation, the total cell numbers differ for the three geometries. Accumulation is defined as the ratio of the number of bacteria in each surface bin area (grey regions) to the number of bacteria in an equally sized area 50 μm away from the surface. Two shapes of equal area are considered for the bulk area: a shape which follows the surface contour (blue) and a rectangle (green). For each geometry in (d), (h), and (l), the blue and green bar show the mean surface accumulation calculated with the surface contour and rectangle as the bulk area, respectively, for 20 independent subsamples of the raw data. The error bars represent the standard deviation. The blue and green bars are nearly equal for each case, demonstrating that the surface accumulation estimation is independent of the shape taken for the bulk reference area. Bin width 5 μm. Scale bars 5 μm. Figure from [94]. 99

A-1 5000 unit vectors drawn from the vMF distribution on the unit sphere in (a) 2D, $\mathbb{S}^1$, and (b) 3D, $\mathbb{S}^2$. $\mu = (1,0)$ and $\mu = (1,0,0)$ for (a) and (b), respectively. $\kappa = 0$ (red points), $\kappa = 10$ (green points), and $\kappa = 100$ (blue points). Note that the vMF distribution is uniform on the sphere $\mathbb{S}^{d-1}$ for $\kappa = 0$ and approaches a point distribution as $\kappa \to \infty$. 108

List of Tables

3.1 List of parameters measured from movie data (observables), including a brief description. Table adaped from [66]. 50
3.2 Summary of bacterial swarming simulation parameters. Table adapted from [66]. 56

4.1 List of calculated single-cell biofilm features. Table adapted from [61]. 69
4.2 List of calculated global biofilm features. Table adapted from [61]. . . 70
4.3 Biofilm parameters and corresponding normalization ranges and weights. Table adapted from [61]. 71
4.4 Key simulation parameters. The parameters $\epsilon_0, \lambda_r, \nu, \lambda_a, \rho_a$, are determined by fitting as described above. Table adapted from [61]. 79

5.1 Summary of bacterial surface accumulation simulation parameters. Table from [94]. 92

Chapter 1

Introduction

When modeling bacterial communities, two main approaches can be taken: continuum and discrete [62, 91, 92, 139]. The continuum approach involves using differential equations to model the dynamics at the population level, taking a top-down approach which omits the details of the individuals [62]. On the other hand, discrete models are a bottom-up approach in which population dynamics emerge from the state and actions of the individuals [62, 91]. Unlike continuum models, discrete models can directly incorporate experimental single-cell observations, which have become possible through technological advances in microscopy, microfluidics, and image analysis [62].

In the 1990s, discrete models for describing bacterial communities started to be developed, with cellular automaton (CA) models being one of the earliest discrete approaches [91, 139]. CA models involve a discretization in space with a lattice, and each rectangular element can represent a biomass of cells, extracellular polymeric substances (EPS), and other molecules. If there are enough nutrients surrounding the rectangular biomass element, it will grow and divide with its resulting mass distributed to a random, neighboring, unoccupied lattice unit. If none of the surrounding grid units are free, a random neighbor is first displaced [91, 106, 107]. Although simple, these CA models can reproduce some of the observed structures seen in growing biofilms [141].

Another class of discrete models is the individual-based model (IBM). Unlike CA models, the agents in IBMs are not limited by the grid and are free to move in any direction [91, 92, 139]. A list of properties for each agent define its state. These agents interact with its environment and each other through a set of defined rules [62, 91, 92, 139]. In IBMs, the population level dynamics arise from the independent actions of the individuals, which allows for the direct investigation of which individual behaviors result in the observed population dynamics. IBMs can also utilize both single-cell and population average data [62]. Taking a more molecular dynamics approach than existing IBMs [9, 10, 43, 52, 65, 109, 109, 113, 114, 127], an individual-based simulation framework is developed that can investigate the underlying forces governing the collective dynamics in such microbial systems as swarms and biofilms.

In Chapter 2, a mechanistic IBM describing collective bacteria dynamics is introduced. The bacteria are modeled as ellipsoids that interact with solid surfaces and each other through interaction potentials. U_{bdy}, an effective steric boundary potential, encodes cell-surface interactions, which can be applied across various geometries. An interaction potential U captures the physical cell-cell interactions which includes cell-cell repulsion from hard steric and osmotic pressure and cell-cell attraction from secreted matrix components. Cell growth and division are also included in the model along with bacterial self-propulsion and viscous drag. A custom highly parallelized code employing graphics processing units (GPUs) is developed to perform the simulations.

Chapter 3 discusses the identification of the space-time phase diagram of bacterial swarm expansion. This work is enabled by a high-throughput adaptive microscopy technique to track the swarming behavior of *Bacillus subtillis* over five orders of magnitude in space and six orders of magnitude in time. The resulting videos are summarized through statistical observables such as cell speed and parameters characterizing the emergence of moving rafts and non-motile clusters. Machine learning is applied to this high-dimensional dataset to identify five main dynamical phases during swarm expansion. To test whether physical forces can account for the different dynamics seen in the phases, two-dimensional (2D) individual-based active matter simulations are performed which reveal that the microscopic swarming motility phases are dominated by physical cell-cell interactions.

Chapter 4 focuses on the emergence of three-dimensional (3D) order and structure in growing biofilms. Biofilms are sessile colonies of bacteria that form on surfaces and are encased by a self-produced extracellular polymer matrix [57]. While some biofilms provide positive benefits such as the treatment of waste water [96], many of these surface-attached microbial communities cause widespread problems to a broad range of industrial equipment and infrastructure, such as food processing facilities [20, 79], ships and pipes [45], and surgical equipment and medical implants [32, 37, 125]. In the medical context, these surface-attached microbial colonies are especially harmful because they can lead to infections including cavities, periodontitis, and cystic fibrosis [25]. Biofilms are quite resilient and hard to remove, being able to withstand many stresses such as changes in temperature, pH, and antibiotic treatments [33, 50]. While previous studies have focused on the genetic and regulatory network of biofilm formation [71, 100], recent developments in confocal microscopy allow for the visualization of biofilm growth at the single-cell resolution [9, 36, 61, 147]. These technical advancements allow for the direct investigation into the physics of biofilms, such as the spatial organization of the individual cells within the biofilm, the mechanisms of bacterial attachment to the surface, and the mechanical properties of the biofilm itself [80, 82, 92, 105]. Based on the single-cell data, a microscopic basis is provided for an effective cell–cell interaction potential, which captures and predicts the growth dynamics, emergent architecture, and local liquid-crystalline order of *Vibrio cholerae* biofilms. The experimental and simulation results imply that local cellular order and global biofilm architecture in these active bacterial communities can arise from mechanical cell–cell interactions, which cells can modulate by regulating the production of particular matrix components.

In Chapter 5, the geometric control of bacterial surface accumulation is examined. Controlling and suppressing bacterial accumulation at solid surfaces is essential for preventing biofilm formation and biofouling. Whereas various chemical surface treatments are known to reduce cell accumulation and attachment, the role of complex surface geometries remains less well understood. Using experiments and simulations, the effects of locally varying boundary curvature on the scattering and accumulation dynamics of swimming *Escherichia coli* bacteria in quasi-two-dimensional microfluidic channels are explored. The experimental and numerical results show that a concave periodic boundary geometry can decrease the average cell concentration at the boundary by more than 50% relative to a flat surface.

Finally, Chapter 6 provides a summary of the IBM for collective bacterial dynamics and reviews the application of this model to investigate the dynamics of bacterial swarming, biofilm formation, and bacterial surface accumulation. Further, extensions of the model are presented along with proposals for future research.

Chapter 2

Mechanistic individual-based model

A bacterium α is modeled as an ellipsoid of half-length ℓ_α and half-width d_α, described by its centroid position $\mathbf{x}_\alpha(t)$ and orientation $\hat{\mathbf{n}}_\alpha(t)$. The cell is assumed to move at a constant self-propulsion speed v_α in the direction of its orientation $\hat{\mathbf{n}}_\alpha$. The cells grow exponentially and follow the adder principle for size homeostasis (Sec. 2.1). An effective steric boundary potential U_{bdy} encodes bacterial surface interactions, which can be applied across various geometries (Sec. 2.2). An interaction potential U captures physical bacteria-bacteria interactions (Sec. 2.3). The overdamped translation and orientation dynamics account for the self-propulsion, viscous drag, cell-surface interactions, and cell-cell interactions acting on a bacterium (Sec. 2.4). A highly parallelized code employing graphics processing units (GPUs) is developed in CUDA to perform the simulations (Sec. 2.5).

In terms of notation, vectors and tensors will be represented in component form, where convenient, with indices that are represented with English alphabet subscripts. To exemplify, the components of the vector $\mathbf{x} = (x_1, x_2, x_3)$ are represented as x_i where $i = 1, 2, 3$. The components of a second-ranked tensor $\mathbf{A}$ are represented as A_{ij} where $i = 1, 2, 3$ and $j = 1, 2, 3$. To label different entities, Greek alphabet subscripts will be used. To exemplify, $\hat{n}_{\alpha i}$ and $\hat{n}_{\beta i}$ refers to a unit vector of label α and another unit vector of label β, respectively.

2.1 Growth and division

How does a cell know when to divide? This is a question that have fascinated biologists for many decades, inspiring many hypotheses and experimental investigations. General theories for the trigger of division include the sizer model, in which the cell divides when it reaches a critical size; the timer model, in which the cell divides after a specific amount of time from birth; and the adder model, in which the cell divides after it has grown a constant size between birth and divison regardless of birth size [68]. Note that the sizer and timer models require feedback while the adder model does not [115]. However, the verification of these models have been hindered by the lack of single-cell data with high spatiotemporal resolution. Overcoming previous limitations, recent experiments have shown that the single-cell data supports the adder model and falsifies the sizer and timer models [15, 128]. Furthermore, a variety of bacterial species have been shown to follow the adder model [69, 115].

In the model, the bacteria are assumed to grow only length-wise, with the bacteria width held constant throughout the simulation. This growth is modeled exponentially as

$$\frac{d\ell_\alpha}{dt} = \frac{\ell_\alpha}{\tau_g}\ln(2) \tag{2.1}$$

for a single bacterium α where ℓ_α is the half-length of the bacteria at time t and τ_g is a growth time constant with units of [s]. Thus, the half-length of the bacteria will grow as

$$\ell_\alpha = \ell_{0\alpha} 2^{t/\tau_g} \tag{2.2}$$

where $\ell_{0\alpha}$ is the initial length of bacterium α. Following the adder model [128], the length added between birth and division (ℓ_{add}) is constant for each cell. Thus, the cells divide when they grow an additional ℓ_{add} from their birth length (division length $\ell_{\text{divide }\alpha} = \ell_{0\alpha} + \ell_{\text{add}}$). At division, a random number X is drawn from a normal distribution of mean 0.5 and standard deviation $\sigma_{\text{birth size}}$, such that the bacterium divides roughly in half. The length of the new daughter bacteria ℓ_{d1} and ℓ_{d2} are

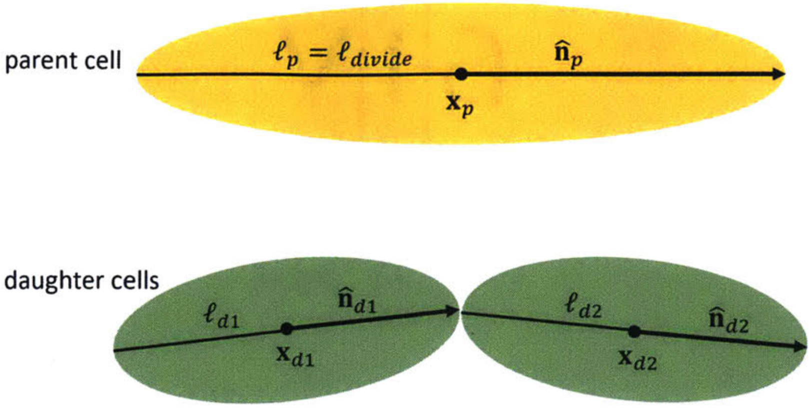


Figure 2-1: Parameters related to the growth and division of the cells. When $\ell_p = \ell_{\text{divide}}$, the cell divides. The daughter cells have lengths of ℓ_{d1} and ℓ_{d2} ($\ell_{d1} + \ell_{d2} = \ell_p$, $\ell_{d1} = X\ell_p$ where X is a random number drawn from a normal distribution of mean 0.5 and standard deviation $\sigma_{\text{birth size}}$). $\mathbf{x}_{d1}$ and $\mathbf{x}_{d2}$ are the centers of the daughter cells, calculated as in Eqs. (2.4). $\hat{\mathbf{n}}_{d1}$ and $\hat{\mathbf{n}}_{d2}$ are the orientations of the daughter cells, drawn from a vMF distrbution with mean direction $\hat{\mathbf{n}}_p$ and concentration parameter κ.

calculated from the length of the parent bacterium ℓ_p

$$\ell_{d1} = X\ell_p \tag{2.3a}$$

$$\ell_{d2} = (1 - X)\ell_p \tag{2.3b}$$

such that the length of the parent cell is conserved. When drawing the random number X, the normal distribution is truncated such that $0.1 < X < 0.9$. The new bacteria centers of the daughter bacteria are

$$\mathbf{x}_{d1} = \mathbf{x}_p - \ell_{d1}\hat{\mathbf{n}}_p \tag{2.4a}$$

$$\mathbf{x}_{d2} = \mathbf{x}_p + \ell_{d2}\hat{\mathbf{n}}_p \tag{2.4b}$$

The new orientations of the daughter bacteria are drawn from a von Mises-Fisher (vMF) distribution, with mean direction corresponding to the direction of the parent cell $\hat{\mathbf{n}}_p$ and concentration parameter κ (see Appendix A for more details about the vMF distrbution). The division lengths of the daughter bacteria are

$$\ell_{\text{divide}_{d1}} = \ell_{d1} + \ell_{\text{add}} \tag{2.5}$$

$$\ell_{\text{divide}_{d2}} = \ell_{d2} + \ell_{\text{add}} \tag{2.6}$$

The parameters related to the growth and division are illustrated in Fig. 2-1.

2.2 Bacteria-surface interactions

An effective steric boundary potential U_{bdy} is used to encode bacterial surface interactions across various geometries. U_{bdy} prevents the bacteria from penetrating the boundary and forces them to align parallel with the local surface tangent. This is achieved by penalizing the overlap between the bacteria and the surface exponentially

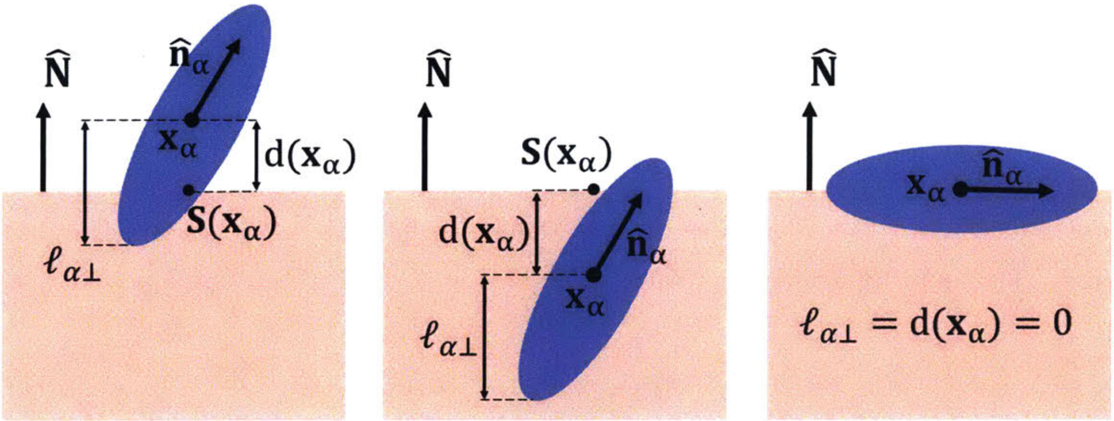
$$U_{\text{bdy}} = \begin{cases} 0 & \text{if } \mathbf{z}_\alpha \leq 0 \quad \text{`no contact'} \\ \epsilon_{\text{bdy}} e^{\mathbf{z}_\alpha/\sigma_{\text{bdy}}} & \text{if } \mathbf{z}_\alpha > 0 \quad \text{`contact'} \end{cases} \tag{2.7}$$

where ϵ_{bdy} is the strength parameter for the bacteria-boundary interaction and σ_{bdy} is a length scale parameter of the order of the bacterial width. The overlap coordinate $\mathbf{z}_\alpha$ of bacterium α is defined as

$$\mathbf{z}_\alpha = \underbrace{\ell_\alpha |\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{N}}(\mathbf{x}_\alpha)|}_{\ell_{\alpha\perp}} + d_\alpha - \underbrace{\hat{\mathbf{N}}(\mathbf{x}_\alpha) \cdot (\mathbf{x}_\alpha - \mathbf{S}(\mathbf{x}_\alpha))}_{\mathsf{d}(\mathbf{x}_\alpha)} \tag{2.8}$$

$\mathbf{S}(\mathbf{x}_\alpha)$ is the point on the surface that is closest to the bacterium's position $\mathbf{x}_\alpha$, and $\hat{\mathbf{N}}(\mathbf{x}_\alpha)$ is the surface normal vector at $\mathbf{S}(\mathbf{x}_\alpha)$. Recall, ℓ_α and d_α are the bacterium's half-length and half-width, respectively. $\ell_{\alpha\perp}$, the first term in $\mathbf{z}_\alpha$, is the projected half-length in the direction of the surface normal. The last term $\mathsf{d}(\mathbf{x}_\alpha)$ is the signed distance of the bacterium's center from the surface. The geometric parameters of the overlap coordinate for three contact cases ($\mathbf{z}_\alpha > 0$) are shown in Fig. 2-2.

The derivatives of the boundary potential with respect to $\mathbf{x}_\alpha$ and $\hat{\mathbf{n}}_\alpha$ are $\partial_{x_{\alpha i}} U_{\text{bdy}} = \partial_{\mathbf{z}_\alpha} U_{\text{bdy}} \partial_{x_{\alpha i}} \mathbf{z}_\alpha$ and $\partial_{\hat{n}_{\alpha i}} U_{\text{bdy}} = \partial_{\mathbf{z}_\alpha} U_{\text{bdy}} \partial_{\hat{n}_{\alpha i}} \mathbf{z}_\alpha$, where $\partial_{\mathbf{z}_\alpha} U_{\text{bdy}} = (\epsilon_{\text{bdy}}/\sigma_{\text{bdy}}) e^{\mathbf{z}_\alpha/\sigma_{\text{bdy}}}$.



$\hat{\mathbf{N}}$
$\hat{\mathbf{n}}_\alpha$
$\mathbf{x}_\alpha$
$d(\mathbf{x}_\alpha)$
$S(\mathbf{x}_\alpha)$
$\ell_{\alpha\perp}$
$\hat{\mathbf{N}}$
$S(\mathbf{x}_\alpha)$
$d(\mathbf{x}_\alpha)$
$\hat{\mathbf{n}}_\alpha$
$\mathbf{x}_\alpha$
$\ell_{\alpha\perp}$
$\hat{\mathbf{N}}$
$\mathbf{x}_\alpha$
$\hat{\mathbf{n}}_\alpha$
$\ell_{\alpha\perp} = d(\mathbf{x}_\alpha) = 0$

Figure 2-2: Geometric parameters of the overlap coordinate. In the three cases shown, the overlap coordinate $\mathbf{z}_\alpha > 0$, representing contact with the surface.

The translational gradient of $\mathbf{z}_\alpha$ is

$$\partial_{x_{\alpha i}} \mathbf{z}_\alpha = \left(\ell_\alpha \frac{\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{N}}(\mathbf{x}_\alpha)}{|\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{N}}(\mathbf{x}_\alpha)|} \hat{n}_{\alpha j} - x_{\alpha j} + S_j \right) \partial_{x_{\alpha i}} \hat{N}_j + \hat{N}_j \partial_{x_{\alpha i}} S_j - \hat{N}_i \tag{2.9}$$

If the surface is flat, both $\hat{\mathbf{N}}$ and $\mathbf{S}$ are constant and independent of the bacterium's position $\mathbf{x}_\alpha$; thus, Eq. (2.9) simplifies to $\partial_{x_{\alpha i}} \mathbf{z}_\alpha = -\hat{N}_i$. Hence, for flat surfaces, the translational force is in the direction of the surface normal. The rotational gradient of $\mathbf{z}_\alpha$ is

$$\partial_{\hat{n}_{\alpha i}} \mathbf{z}_\alpha = \ell_\alpha \frac{\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{N}}(\mathbf{x}_\alpha)}{|\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{N}}(\mathbf{x}_\alpha)|} \hat{N}_i \tag{2.10}$$

2.3 Bacteria-bacteria interactions

An interaction potential U is used to capture physical bacteria-bacteria interactions. This potential accounts for hard steric and osmotic repulsion and cell–cell attraction. Extending the overlap potential for ellipsoidal particles derived by Berne and Pechukas, U applies to nonidentical ellipsoidal particles and includes repulsive and attractive forces.

2.3.1 Berne-Pechukas potential

In their seminal paper [8], Berne and Pechukas derived an overlap potential to model the steric interactions between two molecules, which are approximated as ellipsoids. This overlap potential is obtained by considering the overlap of two ellipsoidal Gaussians. These Gaussians have the form

$$G(x_i) = \exp(-x_i A_{ij}^{-1} x_j) \tag{2.11}$$

where

$$A_{ij} = \ell^2 \hat{n}_i \hat{n}_j + d^2(\delta_{ij} - \hat{n}_i \hat{n}_j) \tag{2.12}$$

Here, $\mathbf{x}$ is a point in space. $\hat{\mathbf{n}}$ is a unit vector in the direction of the principal axis of the ellipsoid. ℓ and d characterize the ellipsoid's half-length and half-width, respectively. δ_{ij} is the Kronecker delta, as defined below

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.13}$$

The inverse of $\mathbf{A}$ is

$$A_{ij}^{-1} = \frac{1}{\ell^2} \hat{n}_i \hat{n}_j + \frac{1}{d^2}(\delta_{ij} - \hat{n}_i \hat{n}_j) \tag{2.14}$$

Note, $\hat{n}_i \hat{n}_j$ is the projector which when applied to any vector will find the component of the vector that is in the same direction as $\hat{n}_i$. On the other hand, $(\delta_{ij} - \hat{n}_i \hat{n}_j)$ is the orthogonal projector which when applied to any vector will find the component of the vector that is on the plane consisting of all vectors that are perpendicular to $\hat{n}_i$ [93]. An example of an ellipsoidal Gaussian is shown in Fig. 2-3.

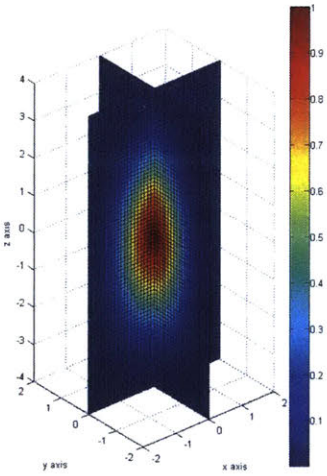


Figure 2-3: Plot of $G(\mathbf{x}) = \exp(-\mathbf{x}^\top \mathbf{A}^{-1} \mathbf{x})$ where $\mathbf{A} = \ell^2 \hat{\mathbf{n}} \hat{\mathbf{n}}^\top + d^2(\mathbf{I} - \hat{\mathbf{n}} \hat{\mathbf{n}}^\top)$ in which $\hat{\mathbf{n}} = \hat{e}_z$. This simplifies to $G(x, y, z) = \exp\left(-\frac{(x^2+y^2)}{d^2} - \frac{z^2}{\ell^2}\right)$. In the plot, $\ell = 2$ and $d = 1$. $G(\mathbf{x})$, which is ellipsoidal in shape and centered on the origin, has the highest values in the center and gradually decays to zero in all directions.

Berne and Pechukas proposed that the steric interaction between two identical molecules is proportional

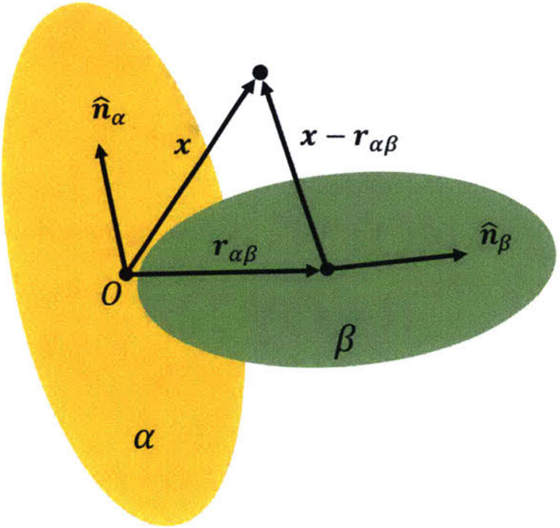


Figure 2-4: Geometry for the overlap potential between neighboring cells α and β with orientations $\hat{\mathbf{n}}_\alpha$ and $\hat{\mathbf{n}}_\beta$, respectively. α is located at the origin O. The vector $\mathbf{r}_{\alpha\beta}$ points from the center of α to the center of β.

to the mathematical overlap of these ellipsoidal Gaussians, which can be calculated through the integration over all space of the product of two ellipsoidal Gaussians. Consider two ellipsoidal Gaussians, α and β, with axial unit vectors, $\hat{\mathbf{n}}_\alpha$ and $\hat{\mathbf{n}}_\beta$ as shown in Fig. 2-4. The vector $\mathbf{r}_{\alpha\beta}$ joins their centers. Note that $\mathbf{r}_{\alpha\beta}$ points from ellipsoid α to ellipsoid β. From the figure, note that the origin of the coordinate system is attached to ellipsoid α. The overlap potential U_{overlap} is determined by

$$U_{\text{overlap}} \propto |\mathbf{A}_\alpha|^{-1/2}|\mathbf{A}_\beta|^{-1/2} \int_{-\infty}^{\infty} \exp[-x_i A_{\alpha ij}^{-1} x_j - (x_i - r_{\alpha\beta i}) A_{\beta ij}^{-1}(x_j - r_{\alpha\beta j})] dx_1 dx_2 dx_3 \tag{2.15}$$

Integrating,

$$U_{\text{overlap}} \propto \pi^{3/2}|\mathbf{A}_\alpha|^{-1/2}|\mathbf{A}_\beta|^{-1/2}|\mathbf{S}|^{-1/2} \exp\left[\frac{1}{4} c_i S_{ij}^{-1} c_j - r_{\alpha\beta i} A_{\beta ij}^{-1} r_{\alpha\beta j}\right] \tag{2.16}$$

where $S_{ij} = [A_{\alpha ij}^{-1} + A_{\beta ij}^{-1}]$ and $c_i = 2A_{\beta ij}^{-1} r_{\alpha\beta j}$. which simplifies to

$$U_{\text{overlap}} \propto \pi^{3/2}|\mathbf{A}_\alpha + \mathbf{A}_\beta|^{-1/2} \exp[-r_{\alpha\beta i}(A_{\alpha ij} + A_{\beta ij})^{-1} r_{\alpha\beta j}] \tag{2.17}$$

which can be represented as

$$U_{\text{overlap}} = \epsilon_0 \epsilon_1 \exp(-r_{\alpha\beta}^2/\sigma^2) \tag{2.18}$$

where $\mathbf{r}_{\alpha\beta} = r_{\alpha\beta}\hat{\mathbf{r}}_{\alpha\beta}$. The proportionality constant ϵ_0 describes the strength of the interaction. The strength parameter $\epsilon_1 = f(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \ell_\alpha, \ell_\beta, d_\alpha, d_\beta)$ is

$$\epsilon_1 \propto |\mathbf{A}_\alpha + \mathbf{A}_\beta|^{-1/2} \tag{2.19}$$

and the range parameter $\sigma = f(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \hat{\mathbf{r}}_{\alpha\beta}, \ell_\alpha, \ell_\beta, d_\alpha, d_\beta)$ is defined as

$$\sigma^2 = \frac{1}{\hat{r}_{\alpha\beta i}(A_{\alpha ij} + A_{\beta ij})^{-1}\hat{r}_{\alpha\beta j}} \tag{2.20}$$

Note that the overlap potential U_{overlap} between two ellipsoids α and β is a function of their axial unit vectors $\hat{\mathbf{n}}_\alpha$ and $\hat{\mathbf{n}}_\beta$ and the vector $\mathbf{r}_{\alpha\beta}$ joining their centers. See Appendix B for the full details of the Berne-Pechukas overlap potential derivation.

2.3.2 Generalized strength and range parameters

A generalized form of the strength parameter ϵ_1, defined in Eq. (2.19), for nonidentical particles is derived by directly expanding the symmetric matrix $A_{\alpha ij} + A_{\beta ij}$ and calculating the determinant such that

$$\begin{aligned}|\mathbf{A}_\alpha + \mathbf{A}_\beta| = &- (d_\alpha^2 + d_\beta^2)(-d_\alpha^2 d_\beta^2 - d_\alpha^2\ell_\alpha^2 - d_\beta^2\ell_\beta^2 - \ell_\alpha^2\ell_\beta^2) \\ &- (d_\alpha^2 + d_\beta^2)(d_\alpha^2 d_\beta^2 - d_\beta^2\ell_\alpha^2 - d_\alpha^2\ell_\beta^2 + \ell_\alpha^2\ell_\beta^2)(\hat{n}_{\alpha i}\hat{n}_{\beta i})^2\end{aligned} \tag{2.21}$$

Dividing each term by $\varpi = -(d_\alpha^2 + d_\beta^2)(-d_\alpha^2 d_\beta^2 - d_\alpha^2\ell_\alpha^2 - d_\beta^2\ell_\beta^2 - \ell_\alpha^2\ell_\beta^2)$,

$$\frac{|\mathbf{A}_\alpha + \mathbf{A}_\beta|}{\varpi} = 1 - \chi^2(\hat{n}_{\alpha i}\hat{n}_{\beta i})^2 \tag{2.22}$$

where

$$\chi = \left[\frac{(\ell_\alpha^2 - d_\alpha^2)(\ell_\beta^2 - d_\beta^2)}{(\ell_\alpha^2 + d_\beta^2)(\ell_\beta^2 + d_\alpha^2)}\right]^{1/2} \tag{2.23}$$

For identically sized particles ($\ell_\alpha = \ell_\beta = \ell$ and $d_\alpha = d_\beta = d$), this reduces to $\frac{(\ell^2-d^2)}{(\ell^2+d^2)}$, which matches Eq. (5b) in [8]. Thus, ϵ_1 is defined as

$$\epsilon_1 = [1 - \chi^2(\hat{n}_{\alpha i}\hat{n}_{\beta i})^2]^{-1/2} = [1 - \chi^2(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)^2]^{-1/2} \tag{2.24}$$

with $\varpi^{-1/2}$ being captured in ϵ_0. The maximum value of ϵ_1 occurs when $\hat{\mathbf{n}}_\alpha$ is parallel to $\hat{\mathbf{n}}_\beta$. ϵ_1 equals the minimum value of 1 when $\hat{\mathbf{n}}_\alpha$ is perpendicular to $\hat{\mathbf{n}}_\beta$.

Cleaver et al. derived a generalized form of the range parameter σ as defined in Eq. (2.20) for nonidentical particles, which is given in Eq. (15) in [21] and is stated below for convenience

$$\sigma = \sigma_0 \left[1 - \chi \left(\frac{\xi(\hat{\mathbf{r}}_{\alpha\beta} \cdot \hat{\mathbf{n}}_\alpha)^2 + \xi^{-1}(\hat{\mathbf{r}}_{\alpha\beta} \cdot \hat{\mathbf{n}}_\beta)^2 - 2\chi(\hat{\mathbf{r}}_{\alpha\beta} \cdot \hat{\mathbf{n}}_\alpha)(\hat{\mathbf{r}}_{\alpha\beta} \cdot \hat{\mathbf{n}}_\beta)(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)}{1 - \chi^2(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)^2}\right)\right]^{-1/2} \tag{2.25}$$

where

$$\sigma_0 = \sqrt{d_\alpha^2 + d_\beta^2}, \tag{2.26}$$

$$\chi = \left[\frac{(\ell_\alpha^2 - d_\alpha^2)(\ell_\beta^2 - d_\beta^2)}{(\ell_\alpha^2 + d_\beta^2)(\ell_\beta^2 + d_\alpha^2)}\right]^{1/2}, \tag{2.27}$$

and

$$\xi = \left[\frac{(\ell_\alpha^2 - d_\alpha^2)(\ell_\beta^2 + d_\alpha^2)}{(\ell_\beta^2 - d_\beta^2)(\ell_\alpha^2 + d_\beta^2)}\right]^{1/2}. \tag{2.28}$$

The maximum value of σ occurs when $\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta$, and $\hat{\mathbf{r}}_{\alpha\beta}$ are all aligned. σ equals the minimum value of σ_0 when $\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta$, and $\hat{\mathbf{r}}_{\alpha\beta}$ are all perpendicular to each other. Appendix C includes the full details of the derivation for the strength and range parameters for nonidentical particles.

2.3.3 Cell-cell interaction potential

As stated in [61], the mechanical interactions between two cells include the steric repulsion, the repulsion mediated by the effective osmotic pressure in the intercellular space, and the interactions of secreted matrix components. In particular, the secreted matrix protein RbmA links cells together, contributing to cell-cell attraction [48, 105, 129]. For two neighboring bacteria α and β centered at $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}_\beta$ with orientations $\hat{\boldsymbol{n}}_\alpha$ and $\hat{\boldsymbol{n}}_\beta$, half-lengths ℓ_α and ℓ_β, and half-widths d_α and d_β, the cell-cell interaction potential U is proposed to be

$$U = \epsilon_0 \epsilon_1 \left(e^{-\frac{\rho_{\alpha\beta}^2}{\lambda_r^2}} + \frac{\nu}{1 + e^{\frac{\rho_a - \rho_{\alpha\beta}}{\lambda_a}}} \right) \tag{2.29}$$

The first term in U describes the combined effects of hard steric and osmotic repulsion. The second term corresponds to cell-cell attraction. The amplitude is set by the interaction strength ϵ_0 and generalized strength parameter ϵ_1 [Eq. (2.24)] which accounts for the instantaneous cell lengths and cell orientations. $\rho_{\alpha\beta} = r_{\alpha\beta}/\sigma$ is the cell-cell distance normalized by the range parameter σ. The vector $\mathbf{r}_{\alpha\beta} = r_{\alpha\beta}\hat{\mathbf{r}}_{\alpha\beta}$ joins the cell centers and is directed from cell α to cell β as shown in Fig. 2-4. The generalized range parameter σ [Eq. (2.25)] is the interaction length scale between nonidentical ellipsoidal cells which depends on the instantaneous cell lengths, the orientation of the cells relative to each other, and the individual cell orientations. λ_r is the range of the repulsion. ν is the relative strength of the RbmA-mediated attraction. λ_a is the effective attractive range. ρ_a is the position of the attractive potential well. Each contribution and parameter in the potential U thus has a well-defined physical meaning. The parameters in the cell-cell potential are assumed to be constant for a given bacterial strain, a simplification that could be relaxed in future models. The schematic diagram of U is shown in Fig. 2-5.

The generalized strength and range parameters are reproduced below for convenience.

$$\epsilon_1 = [1 - \chi^2(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)^2]^{-1/2} \tag{2.30}$$

$$\sigma = \sigma_0 \left[1 - \chi \left(\frac{\xi(\hat{\mathbf{r}}_{\alpha\beta} \cdot \hat{\mathbf{n}}_\alpha)^2 + \xi^{-1}(\hat{\mathbf{r}}_{\alpha\beta} \cdot \hat{\mathbf{n}}_\beta)^2 - 2\chi(\hat{\mathbf{r}}_{\alpha\beta} \cdot \hat{\mathbf{n}}_\alpha)(\hat{\mathbf{r}}_{\alpha\beta} \cdot \hat{\mathbf{n}}_\beta)(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)}{1 - \chi^2(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)^2} \right) \right]^{-1/2} \tag{2.31}$$

where

$$\sigma_0 = \sqrt{d_\alpha^2 + d_\beta^2}, \tag{2.32}$$

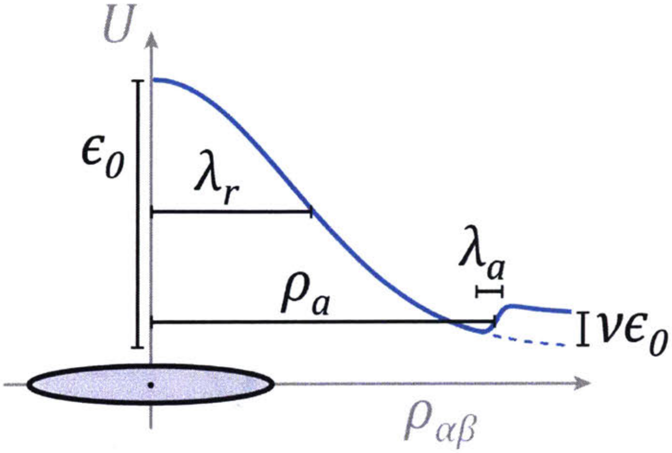


Figure 2-5: Cell-cell interaction potential parameters. ϵ_0 characterizes the strength of the interaction. λ_r is the range of the repulsion. ν is the relative strength of the RbmA-mediated attraction. λ_a is the effective attractive range. ρ_a is the position of the attractive potential well. $\rho_{\alpha\beta} = r_{\alpha\beta}/\sigma$ is the cell-cell distance normalized by the range parameter σ. Figure from [61].

$$\chi = \left[\frac{(\ell_\alpha^2 - d_\alpha^2)(\ell_\beta^2 - d_\beta^2)}{(\ell_\alpha^2 + d_\beta^2)(\ell_\beta^2 + d_\alpha^2)}\right]^{1/2}, \tag{2.33}$$

and

$$\xi = \left[\frac{(\ell_\alpha^2 - d_\alpha^2)(\ell_\beta^2 + d_\alpha^2)}{(\ell_\beta^2 - d_\beta^2)(\ell_\alpha^2 + d_\beta^2)}\right]^{1/2}. \tag{2.34}$$

With respect to $\boldsymbol{x}_\alpha$, key derivatives of the potential are

$$\frac{\partial U}{\partial x_{\alpha i}} = \frac{\partial U}{\partial \rho_{\alpha\beta}}\frac{\partial \rho_{\alpha\beta}}{\partial r_{\alpha\beta}}\frac{\partial r_{\alpha\beta}}{\partial x_{\alpha i}} + \frac{\partial U}{\partial \rho_{\alpha\beta}}\frac{\partial \rho_{\alpha\beta}}{\partial \sigma}\frac{\partial \sigma}{\partial \hat{r}_{\alpha\beta j}}\frac{\partial \hat{r}_{\alpha\beta j}}{\partial x_{\alpha i}} \tag{2.35}$$

$$\frac{\partial U}{\partial \rho_{\alpha\beta}} = \epsilon_0\epsilon_1 \left\{ -\frac{2\rho_{\alpha\beta}}{\lambda_r^2}\exp\left(-\frac{\rho_{\alpha\beta}^2}{\lambda_r^2}\right) + \frac{\nu}{\lambda_a}\frac{\exp\left(\frac{\rho_0 - \rho_{\alpha\beta}}{\lambda_a}\right)}{\left[1 + \exp\left(\frac{\rho_0 - \rho_{\alpha\beta}}{\lambda_a}\right)\right]^2} \right\} \tag{2.36}$$

$$\frac{\partial \rho_{\alpha\beta}}{\partial r_{\alpha\beta}} = \frac{1}{\sigma} \tag{2.37}$$

$$\frac{\partial r_{\alpha\beta}}{\partial x_{\alpha i}} = -\frac{r_{\alpha\beta i}}{r_{\alpha\beta}} \tag{2.38}$$

$$\frac{\partial \rho_{\alpha\beta}}{\partial \sigma} = -\frac{r_{\alpha\beta}}{\sigma^2} \tag{2.39}$$

$$\frac{\partial \sigma}{\partial \hat{r}_{\alpha\beta j}} = \frac{\chi\epsilon_1^2\sigma^3}{\sigma_0^2}\left[\xi(\hat{\boldsymbol{r}}_{\alpha\beta}\cdot\hat{\boldsymbol{n}}_\alpha)\hat{n}_{\alpha j} + \xi^{-1}(\hat{\boldsymbol{r}}_{\alpha\beta}\cdot\hat{\boldsymbol{n}}_\beta)\hat{n}_{\beta j} - \chi(\hat{\boldsymbol{n}}_\alpha\cdot\hat{\boldsymbol{n}}_\beta)\left[\hat{n}_{\alpha j}(\hat{\boldsymbol{r}}_{\alpha\beta}\cdot\hat{\boldsymbol{n}}_\beta) + \hat{n}_{\beta j}(\hat{\boldsymbol{r}}_{\alpha\beta}\cdot\hat{\boldsymbol{n}}_\alpha)\right]\right] \tag{2.40}$$

$$\frac{\partial \hat{r}_{\alpha\beta j}}{\partial x_{\alpha i}} = \frac{(r_{\alpha\beta i})(r_{\alpha\beta j})}{r^3_{\alpha\beta}} - \frac{\delta_{ij}}{r_{\alpha\beta}} \tag{2.41}$$

With respect to $\boldsymbol{\hat{n}}_\alpha$, key derivatives of the potential are

$$\frac{\partial U}{\partial \hat{n}_{\alpha i}} = \frac{\partial U}{\partial \epsilon_1}\frac{\partial \epsilon_1}{\partial \hat{n}_{\alpha i}} + \frac{\partial U}{\partial \rho_{\alpha\beta}}\frac{\partial \rho_{\alpha\beta}}{\partial \sigma}\frac{\partial \sigma}{\partial \hat{n}_{\alpha i}} \tag{2.42}$$

$$\frac{\partial U}{\partial \epsilon_1} = \epsilon_0 \left[\exp\left(-\frac{\rho^2_{\alpha\beta}}{\lambda^2_r}\right) + \frac{\nu}{1+\exp\left(\frac{\rho_0-\rho_{\alpha\beta}}{\lambda_a}\right)}\right] \tag{2.43}$$

$$\frac{\partial \epsilon_1}{\partial \hat{n}_{\alpha i}} = \chi^2 \epsilon_1^3 (\boldsymbol{\hat{n}}_\alpha \cdot \boldsymbol{\hat{n}}_\beta)\hat{n}_{\beta i} \tag{2.44}$$

$$\frac{\partial \sigma}{\partial \hat{n}_{\alpha i}} = \frac{\chi \sigma^3}{2\sigma_0^2}\left(\epsilon_1^2 \frac{\partial \varphi}{\partial \hat{n}_{\alpha i}} + 2\epsilon_1 \varphi \frac{\partial \epsilon_1}{\partial \hat{n}_{\alpha i}}\right) \tag{2.45}$$

where

$$\varphi = \xi(\boldsymbol{\hat{r}}_{\alpha\beta} \cdot \boldsymbol{\hat{n}}_\alpha)^2 + \xi^{-1}(\boldsymbol{\hat{r}}_{\alpha\beta} \cdot \boldsymbol{\hat{n}}_\beta)^2 - 2\chi(\boldsymbol{\hat{r}}_{\alpha\beta} \cdot \boldsymbol{\hat{n}}_\alpha)(\boldsymbol{\hat{r}}_{\alpha\beta} \cdot \boldsymbol{\hat{n}}_\beta)(\boldsymbol{\hat{n}}_\alpha \cdot \boldsymbol{\hat{n}}_\beta) \tag{2.46}$$

$$\frac{\partial \varphi}{\partial \hat{n}_{\alpha i}} = 2\xi(\boldsymbol{\hat{r}}_{\alpha\beta} \cdot \boldsymbol{\hat{n}}_\alpha)\hat{r}_{\alpha\beta i} - 2\chi(\boldsymbol{\hat{r}}_{\alpha\beta} \cdot \boldsymbol{\hat{n}}_\beta)\left[(\boldsymbol{\hat{n}}_\alpha \cdot \boldsymbol{\hat{n}}_\beta)\hat{r}_{\alpha\beta i} + (\boldsymbol{\hat{r}}_{\alpha\beta} \cdot \boldsymbol{\hat{n}}_\alpha)\hat{n}_{\beta i}\right] \tag{2.47}$$

Note,

$$\begin{aligned} \frac{\partial U}{\partial r_{\alpha\beta}} &= \frac{\partial U}{\partial \rho_{\alpha\beta}}\frac{\partial \rho_{\alpha\beta}}{\partial r_{\alpha\beta}} \\ &= \frac{\epsilon_0\epsilon_1}{\sigma}\left\{-\frac{2\rho_{\alpha\beta}}{\lambda_r^2}\exp\left(-\frac{\rho^2_{\alpha\beta}}{\lambda^2_r}\right) + \frac{\nu}{\lambda_a}\frac{\exp\left(\frac{\rho_0-\rho_{\alpha\beta}}{\lambda_a}\right)}{\left[1+\exp\left(\frac{\rho_0-\rho_{\alpha\beta}}{\lambda_a}\right)\right]^2}\right\} \end{aligned} \tag{2.48}$$

$$\begin{aligned} \frac{\partial U}{\partial \sigma} &= \frac{\partial U}{\partial \rho_{\alpha\beta}}\frac{\partial \rho_{\alpha\beta}}{\partial \sigma} \\ &= -\frac{\epsilon_0\epsilon_1 r_{\alpha\beta}}{\sigma^2}\left\{-\frac{2\rho_{\alpha\beta}}{\lambda_r^2}\exp\left(-\frac{\rho^2_{\alpha\beta}}{\lambda^2_r}\right) + \frac{\nu}{\lambda_a}\frac{\exp\left(\frac{\rho_0-\rho_{\alpha\beta}}{\lambda_a}\right)}{\left[1+\exp\left(\frac{\rho_0-\rho_{\alpha\beta}}{\lambda_a}\right)\right]^2}\right\} \end{aligned} \tag{2.49}$$

2.4 Translation and orientation dynamics

In the model, a bacterium α is described by its position $\mathbf{x}_\alpha(t)$, orientation $\hat{\mathbf{n}}_\alpha(t)$, half-length ℓ_α, and half-width d_α. The forces acting on the bacterium are

$$\mathbf{f}_{\text{thrust}_\alpha} = \gamma_1 v \hat{\mathbf{n}}_\alpha \tag{2.50}$$

$$\mathbf{f}_{\text{visc}_\alpha} = -\mathbf{\Gamma}_\alpha \dot{\mathbf{x}}_\alpha \tag{2.51}$$

$$\mathbf{f}_{\text{bdy}_\alpha} = -\frac{\partial U_{\text{bdy}}}{\partial \mathbf{x}_\alpha} \tag{2.52}$$

$$\mathbf{f}_{\text{cell-cell interactions}_\alpha} = -\frac{\partial V_\alpha}{\partial \mathbf{x}_\alpha} \tag{2.53}$$

$\mathbf{f}_{\text{thrust}_\alpha}$ is the self-propulsion force of the bacteria. $\mathbf{f}_{\text{visc}_\alpha}$ is the viscous drag force acting on the bacteria. $\mathbf{f}_{\text{bdy}_\alpha}$ is the force arising from bacteria-surface interactions. $\mathbf{f}_{\text{cell-cell interactions}_\alpha}$ is the force resulting from mechanical interactions between bacteria. $\mathbf{\Gamma}_\alpha$ is a translation friction tensor with units of [kg/s], γ_1 is a friction coefficient with units of [kg/s], U_{bdy}, as described in Sec. 2.2, is the boundary potential with units of [J], and V_α is the total potential of all the pairwise interactions with units of [J] and is expressed by

$$V_\alpha = \sum_{\beta=1,\, \beta\neq\alpha}^{N} U \tag{2.54}$$

where N is the total number of bacteria and U, as described in Sec. 2.3.3, is the interaction potential between two bacteria α and β.

The equation of motion for bacterium α is thus

$$m_\alpha \ddot{\mathbf{x}}_\alpha = \gamma_1 v \hat{\mathbf{n}}_\alpha - \mathbf{\Gamma}_\alpha \dot{\mathbf{x}}_\alpha - \frac{\partial U_{\text{bdy}}}{\partial \mathbf{x}_\alpha} - \frac{\partial V_\alpha}{\partial \mathbf{x}_\alpha} \tag{2.55}$$

where m_α is the mass of bacterium α. Because cells live at low Reynolds number ($Re \approx 10^{-4}$), inertial effects are neglected ($m_\alpha \ddot{\mathbf{x}}_\alpha \to 0$). The resulting overdamped dynamics are

$$\frac{d\mathbf{x}_\alpha}{dt} = \mathbf{\Gamma}_\alpha^{-1} \left[\gamma_1 v \hat{\mathbf{n}}_\alpha - \frac{\partial U_{\text{bdy}}}{\partial \mathbf{x}_\alpha} - \frac{\partial V_\alpha}{\partial \mathbf{x}_\alpha} \right] \tag{2.56}$$

The friction tensor is assumed to have the following form

$$\mathbf{\Gamma}_\alpha = \gamma_0 \left[\gamma_{\|}(\hat{\mathbf{n}}_\alpha \hat{\mathbf{n}}_\alpha^\top) + \gamma_\perp(\mathbf{I} - \hat{\mathbf{n}}_\alpha \hat{\mathbf{n}}_\alpha^\top)\right] \tag{2.57}$$

which accounts for the fact that the bacteria experience more drag when moving perpendicular to their orientation. The inverse is

$$\mathbf{\Gamma}_\alpha^{-1} = \frac{1}{\gamma_0}\left[\frac{1}{\gamma_{\|}}(\hat{\mathbf{n}}_\alpha \hat{\mathbf{n}}_\alpha^\top) + \frac{1}{\gamma_\perp}(\mathbf{I} - \hat{\mathbf{n}}_\alpha \hat{\mathbf{n}}_\alpha^\top)\right] \tag{2.58}$$

For the orientation dynamics, first note that

$$\begin{aligned} \hat{\mathbf{n}} \cdot \hat{\mathbf{n}} &= 1 \\ \frac{d}{dt}(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}) &= \frac{d}{dt}(1) \\ 2\left(\frac{d\hat{\mathbf{n}}}{dt} \cdot \hat{\mathbf{n}}\right) &= 0 \end{aligned} \tag{2.59}$$

which shows that $\frac{d\hat{\mathbf{n}}}{dt}$ and $\hat{\mathbf{n}}$ are perpendicular to each other. Thus, to conserve the unit length of $\hat{\mathbf{n}}$, the orientation dynamics are modeled as

$$\frac{d\hat{\mathbf{n}}_\alpha}{dt} = (\mathbf{I} - \hat{\mathbf{n}}_\alpha \hat{\mathbf{n}}_\alpha^\top)\left[\mathbf{\Omega}_\alpha^{-1}\left(-\frac{\partial U_{\text{bdy}}}{\partial \hat{\mathbf{n}}_\alpha} - \frac{\partial V_\alpha}{\partial \hat{\mathbf{n}}_\alpha}\right)\right] \tag{2.60}$$

where $\mathbf{\Omega}_\alpha$ is a rotational drag tensor with units of [kg m^2/ s] and is approximated as isotropic,

$$\mathbf{\Omega}_\alpha = \omega_0 \gamma_R \mathbf{I} \tag{2.61}$$

$\gamma_0 = 6\pi\mu d$ and $\omega_0 = 8\pi\mu\ell d^2$ are the Stokesian translational and rotational friction coefficients, respectively. Surface adhesion of the cells is captured by increasing the magnitude of the friction tensor $\mathbf{\Gamma}$ by a factor of 30 if the cells are within $1.5d$ of the surface. $\gamma_{\|}$, $\gamma_\perp$, and γ_R are dimensionless geometric parameters characterizing the longitudinal, transverse, and rotational friction parameters of elongated particles that depend only on the aspect ratio $a_\alpha = \ell_\alpha/d_\alpha$. Below are the expressions given in [130] for rod-like macromolecules

$$\frac{2\pi a_\alpha}{\gamma_\parallel} = \ln a_\alpha - 0.207 + \frac{0.980}{a_\alpha} - \frac{0.133}{a_\alpha^2} \tag{2.62a}$$

$$\frac{4\pi a_\alpha}{\gamma_\perp} = \ln a_\alpha + 0.839 + \frac{0.185}{a_\alpha} + \frac{0.233}{a_\alpha^2} \tag{2.62b}$$

$$\frac{\pi a_\alpha^2}{3\gamma_R} = \ln a_\alpha - 0.662 + \frac{0.917}{a_\alpha} - \frac{0.050}{a_\alpha^2} \tag{2.62c}$$

For ellipsoidal particles, the expressions for $\gamma_\parallel$, $\gamma_\perp$, and γ_R are given in [60] as

$$\gamma_\parallel = \frac{8/3}{\frac{2a_\alpha}{1-a_\alpha^2} + \frac{2a_\alpha^2-1}{(a_\alpha)^{3/2}} \ln\left(\frac{a_\alpha+\sqrt{a_\alpha^2-1}}{a_\alpha-\sqrt{a_\alpha^2-1}}\right)} \tag{2.63a}$$

$$\gamma_\perp = \frac{8/3}{\frac{a_\alpha}{a_\alpha^2-1} + \frac{2a_\alpha^2-3}{(a_\alpha)^{3/2}} \ln\left(a_\alpha + \sqrt{a_\alpha^2 - 1}\right)} \tag{2.63b}$$

$$\gamma_R = \frac{2}{3} \frac{a_\alpha^4 - 1}{a_\alpha \left[\frac{2a_\alpha^2-1}{\sqrt{a_\alpha^2-1}} \ln\left(a_\alpha + \sqrt{a_\alpha^2 - 1}\right) - a_\alpha\right]} \tag{2.63c}$$

2.5 Implementation

The growth [Eq. (2.1)], translation [Eq. (2.56)], and orientation [Eq. (2.60)] dynamics are nondimensionalized for the simulations. To perform the simulations, a highly parallelized code employing GPUs is developed in CUDA. The numerical integration is performed using the forward Euler scheme, and $\hat{\mathbf{n}}$ is renormalized at each time step to correct for integration errors. The code is adapted to each project. See Secs. 3.2.1, 4.2.1, and 5.2.3 for specific implementation details for the bacterial swarming, biofilm, and bacterial surface accumulation projects.

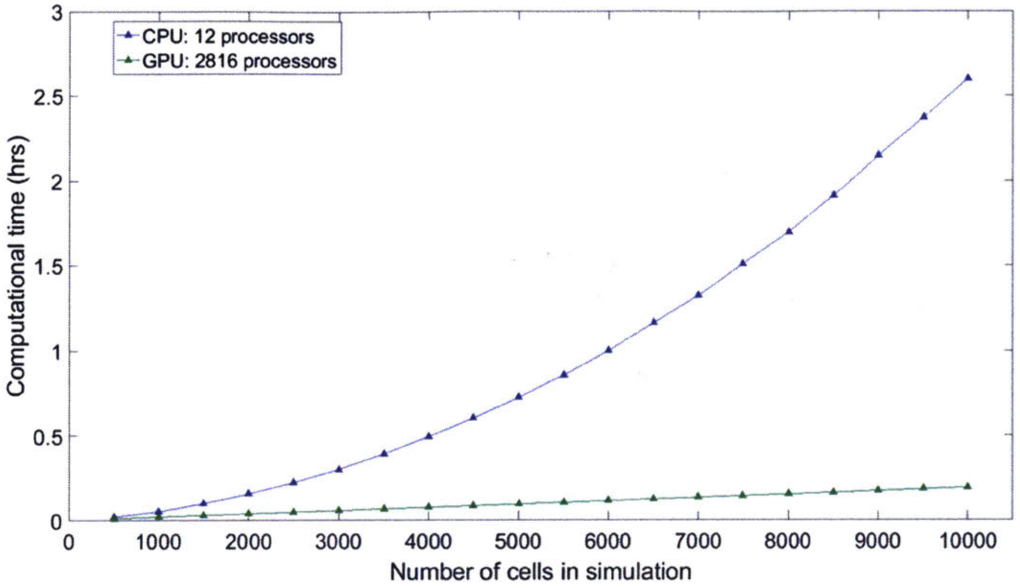


Figure 2-6: Computational time comparison between the CPU (12 processors) and GPU (2816 processors) implementation for the biofilm project, which includes bacterial growth and division and translation and orientation dynamics. The translation and orientation dynamics include cell-surface and cell-cell interactions. The same simulation input parameters are used for the CPU and GPU codes, with both simulations initially starting with one cell. The GPU implementation is significantly faster than the CPU implementation as the number of cells increase.

2.5.1 Code development

Initially, the individual-based code was written in C++ for central processing units (CPUs). It consisted of the following main functions: bacteria growth, neighbor list creation, and time-marching. The bacteria growth and neighbor list creation functions are performed sequentially, whereas the time-marching function is parallelized, utilizing OpenMP. Bacteria grow and divide in the bacteria growth function. The creation of the neighbor lists are facilitated by binning the bacteria according to their physical locations. The neighbor lists are only updated at certain time steps and when a division occurs. In the time-marching function, the cell-surface and cell-cell interactions are calculated and the translation and orientation dynamics are numerically integrated. Bacterial positions, orientations, and sizes are outputted as text files at specific time steps. Basic tests were performed to ensure that the cell-surface and cell-cell interactions were behaving as expected.

In order to speed up the calculation (see Fig. 2-6), the code was then written

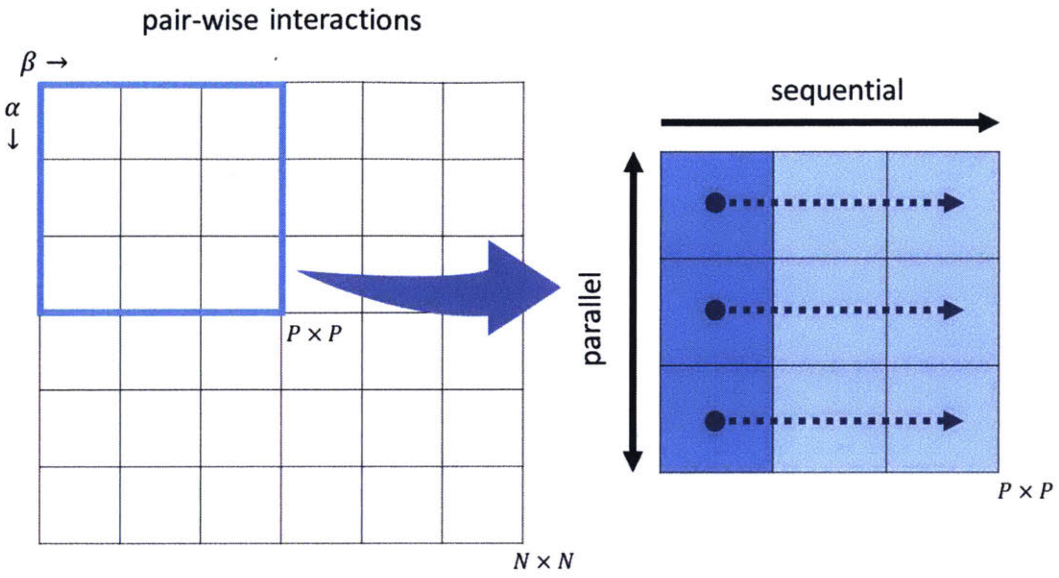


Figure 2-7: Computational tile of size $P \times P$ ($P < N$). The computational tile facilitates the evaluation of the pair-wise interactions. The rows and columns of the computational tile are evaluated in parallel and in sequential order, respectively. Adapted from [99].

for GPUs to take advantage of the inherent parallelism in the problem. The main functions in the GPU code include the bacteria growth, cell-surface interaction calculation, cell-cell interaction calculation, and time-marching. In the bacteria growth function, bacteria grow and divide. $\nabla_{\mathbf{x}_\alpha} U_{\text{bdy}}$ and $\nabla_{\hat{\mathbf{n}}_\alpha} U_{\text{bdy}}$ are calculated in the cell-surface interaction function. All pair-wise cell-cell interactions are evaluated through the CUDA implementation of the all-pairs N-body algorithm [99], described below. All of the functions are performed in parallel except the bacteria growth function to avoid memory collisions when new bacteria are created after division. The output from the code is a text file that includes the bacterial positions, orientations, and sizes at specific time steps. Using the same input parameters and cell configuration, tests were performed to ensure the GPU code could reproduce the final positions and orientations of the cells as the CPU code.

In the CUDA all-pairs N-body algorithm, computational tiles of size $P \times P$ ($P < N$) as illustrated in Fig. 2-7 are utilized to facilitate the evaluation of the $\mathcal{O}(N^2)$ pairwise interactions. The rows and columns of the computational tile are evaluated in parallel and in sequential order, respectively, balancing computational speed and memory use. Inputs for the computational tile are the $2P$ body descriptions and the current sums of $\nabla_{\mathbf{x}_\alpha} V_\alpha$ and $\nabla_{\hat{\mathbf{n}}_\alpha} V_\alpha$, and the outputs are the updated sums of $\nabla_{\mathbf{x}_\alpha} V_\alpha$ and $\nabla_{\hat{\mathbf{n}}_\alpha} V_\alpha$ as illustrated in Fig. 2-8.

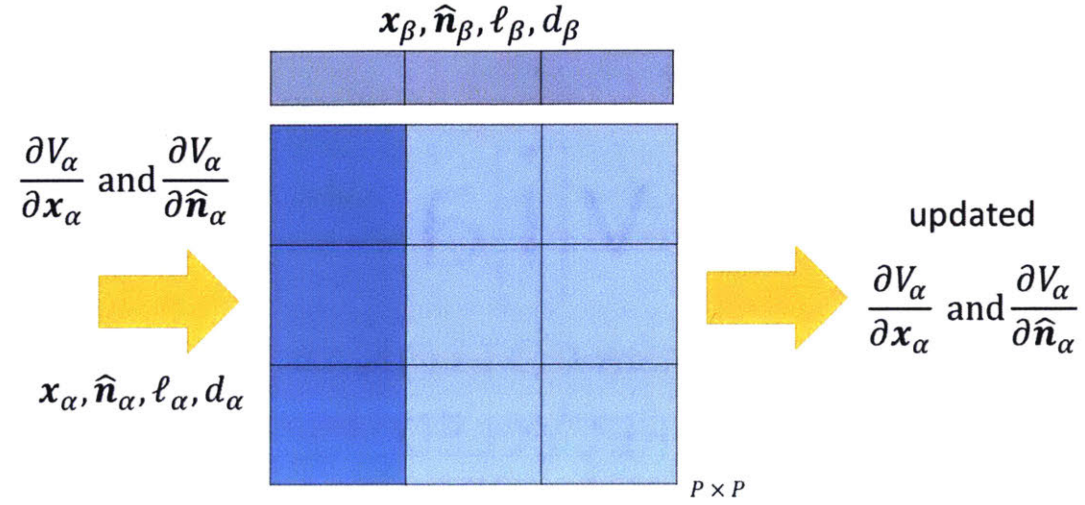

$\boldsymbol{x}_\beta, \hat{\boldsymbol{n}}_\beta, \ell_\beta, d_\beta$
$\frac{\partial V_\alpha}{\partial \boldsymbol{x}_\alpha}$ and $\frac{\partial V_\alpha}{\partial \hat{\boldsymbol{n}}_\alpha}$
$\boldsymbol{x}_\alpha, \hat{\boldsymbol{n}}_\alpha, \ell_\alpha, d_\alpha$
updated
$\frac{\partial V_\alpha}{\partial \boldsymbol{x}_\alpha}$ and $\frac{\partial V_\alpha}{\partial \hat{\boldsymbol{n}}_\alpha}$
$P \times P$

Figure 2-8: Inputs and outputs of the computational tile of size $P \times P$ ($P < N$). Inputs to the computational tile include the current sums of $\nabla_{\mathbf{x}_\alpha} V_\alpha$ and $\nabla_{\hat{\mathbf{n}}_\alpha} V_\alpha$ and the state variables ($\mathbf{x}$, $\hat{\mathbf{n}}$, ℓ, d) for the α particle. Loaded into shared memory are the state variables for the β particle. The output of the computational title are the updated sums of $\nabla_{\mathbf{x}_\alpha} V_\alpha$ and $\nabla_{\hat{\mathbf{n}}_\alpha} V_\alpha$. Adapted from [99].

Chapter 3

Bacterial swarming

Collective migration of flagellated cells across surfaces, termed swarming, is a fundamental bacterial behavior that facilitates range expansions and the exploration of nutrient patches, with profound implications for disease transmission, gene flow, and evolution[1] [24, 43, 73, 90, 118, 123, 135, 137, 150]. Due to its biomedical and ecological importance, bacterial swarming has been widely investigated in microbiology and biophysics as a model system for multicellular self-organization, development, motility, and active matter [1, 4, 11, 12, 18, 19, 26, 31, 38, 41, 53, 64, 85, 117, 133, 140, 145, 149]. Previous studies have revealed important physiological and biophysical factors that control particular aspects of the local swarming behavior, such as the differentiation into distinct cell types [14, 27, 58, 59, 74–76, 101] and the role of osmolarity gradients and surfactant production in maintaining thin liquid films above the surface, through which the cells swim during swarming [28, 67, 72, 77, 108, 112, 122, 146]. However, establishing a comprehensive, causal connection between intracellular, intercellular, and macroscopic behaviors has remained a major challenge due to limitations in data acquisition and analysis techniques suitable for multiscale dynamics. Here, a high-throughput adaptive microscopy approach combined with machine learning allows the identification of key biological and physical mechanisms that determine distinct microscopic and macroscopic collective behavior phases which develop as *Bacillus subtilis* swarms expand over five orders of magnitude in space (Sec. 3.1). This phase identification is combined with particle-based simulations (Sec. 3.2) to infer that cell-cell interactions within each swarming phase are dominated by mechanical interactions (Sec. 3.3).

---

[1] The results of Chapter 3 are published in: Hannah Jeckel, Eric Jelli, Raimo Hartmann, Praveen K. Singh, Rachel Mok, Jan Frederik Totz, Lucia Vidakovic, Bruno Eckhardt, Jörn Dunkel, and Knut Drescher. Learning the space-time phase diagram of bacterial swarm expansion. *Proc. Natl. Acad. Sci. U.S.A.*, 116(5):1489–1494, 2019.

3.1 Experiments

To track the swarming behavior of *B. subtilis* over five orders of magnitude in space at the single-cell level, an adaptive microscope that acquires high-speed movies at times and locations determined by a live feedback between image feature recognition and an automated movement of the scanning area is developed[2] [Fig. 3-1(a)]. This technique allows the imaging of a radially expanding swarm at single-cell resolution in space and time (Fig. 3-1), acquiring movies at a frame rate of 200 Hz over the 10-h duration of a single experiment. Movies are recorded along one line through the swarm [Fig. 3-1(a)], with the length of the line determined adaptively based on the swarm diameter. From each movie, the time-dependent positions, orientations, and velocities of all individual cells are extracted (Fig. 3-1(b) Supplementary Fig. S1 of [66]). To compress, analyze, and visualize this large amount of microscopic time-resolved data, each movie is represented by a list of statistical observables, which include single-cell parameters such as aspect ratio and motility, as well as emergent parameters that characterize the formation of nonmotile clusters and moving rafts [Figs. 3-1(c)–3-1(d)]. The full list of 23 observables extracted at each space-time coordinate is described in Table 3.1. The spatiotemporal evolution of these observables during swarming is visualized in heatmaps (Figs. 3-1(e)–3-1(h) and Supplementary Figs. S2–S6 of [66]), where the color of each pixel is assigned according to an averaged statistical observable of a movie. In the online interactive data explorer (`http://drescherlab.org/data/swarm/`), the space-time heatmap coordinates are linked to the associated microscopic

---

[2] The single-cell experiments and data analysis were performed by Hannah Jeckel, Eric Jelli, Raimo Hartmann, Praveen K. Singh, and Lucia Vidakovic of the Drescher lab. Jan Frederik Totz also assisted in the data analysis.

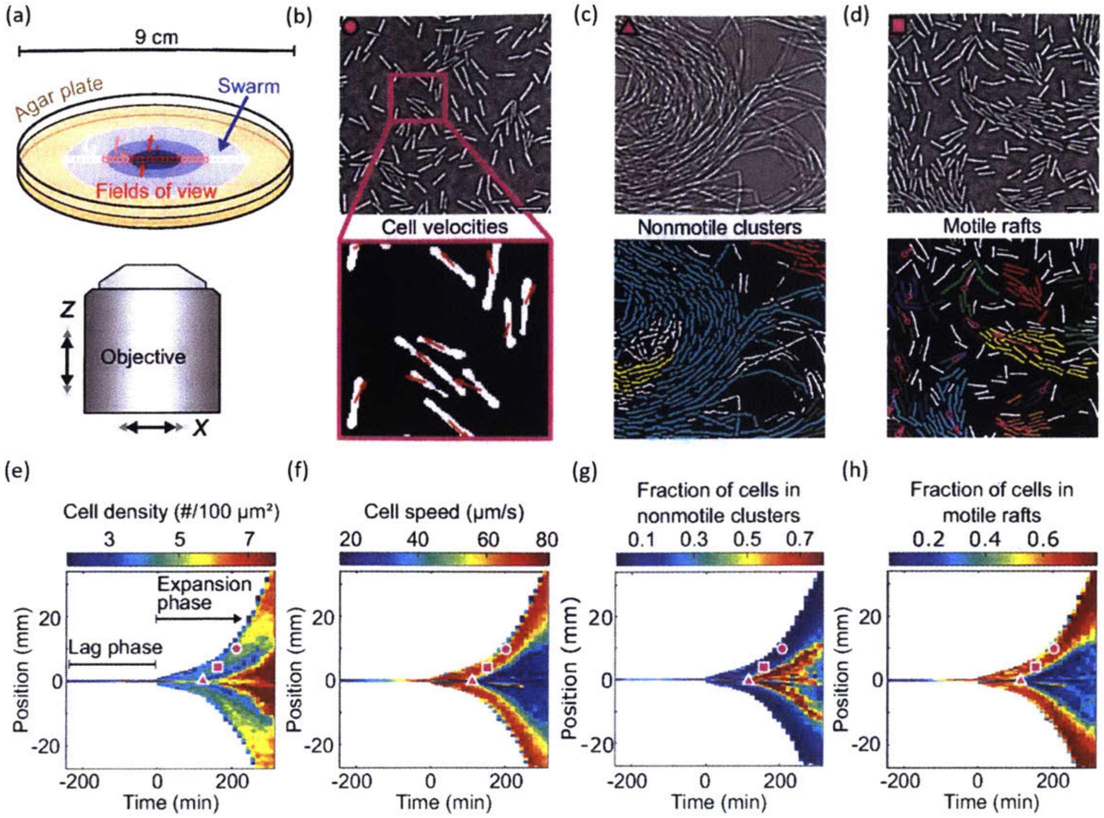


Figure 3-1: Adaptive microscopy reveals complete multiscale dynamics of bacterial swarm expansion. (a) Movies at single-cell resolution are acquired at different locations in the swarm, starting from 1 cell to 100 cells, and follow swarm expansion until the agar plate is completely colonized. The number of movies and locations where movies are acquired (indicated by colored squares) are determined adaptively, depending on the detected swarm size. (b)–(d), Top show qualitatively different bacterial behavioral dynamics observed at distinct space-time points, which are marked in (e)–(h) by corresponding magenta symbols. (b)–(d), Bottom demonstrate automated extraction of single-cell positions, orientations, and cell velocities (b) as well as collective behaviors, such as formation of nonmotile clusters (c) and motile rafts (d), corresponding to groups of aligned cells that move in the same direction. Cells assigned to the same nonmotile cluster or motile raft by the classification algorithms share the same color; cells labeled in white have not been identified as belonging to any motile raft or nonmotile cluster. Magenta arrows in (d) indicate the average velocity of a raft. Scale bars 10 μm. (e) Heatmap of the cell density, obtained by averaging single-cell data as in (b)–(d) for each movie at each space-time coordinate. The lag phase, a period following inoculation during which the swarm does not expand, as well as the expansion phase, is indicated. (f)–(h) Additional heatmaps for the cell speed, fraction of cells that are in nonmotile clusters in a given field of view, and fraction of cells that are in motile rafts. A total of 23 statistical observables (Table 3.1) analogous to (e)–(h) are determined at each space-time position. Figure from [66].

movies within the swarm, to allow for a direct inspection of the connection between microscopic and macroscopic dynamics.

Observable name	Observable description	Units
Aspect ratio	Average aspect ratio of cells	–
Average cell length	Average length of cells	μm
Speed	The most likely cell speed as determined by Rayleigh distribution fit	μm/s
Biomass density	Fraction of area covered by cells	–
Enclosed density	Like Biomass density, but small gaps between cells are removed	–
Cell density	Number of cells per area	# /100 μm^2
Rafting factor	Quantification factor of the rafting behavior of cells, defined as the number of cells within a circle of radius 10 μm swimming in a similar direction (tolerance 10°) divided by all cells within the 10 μm circle.	–
Ratio of rafting cells	Fraction of cells within rafts	–
Average raft size	Average over all raft sizes in the field of view	# cells per raft
Size of largest raft	Size of the largest identified raft in the field of view	# cells
Number of rafts	Number of rafts per area	# /1000 μm^2
Average speed in rafts	Average speed of cells within rafts	μm/s
Ratio of cells speed within/outside of rafts	Average speed of cells within the raft divided by average speed of cells outside of a raft	–
Total raft size	Combined size of all rafts within the field of view	# cells
Average cluster size	Size of non-motile clusters averaged over the field of view. Clusters are defined as a non-motile cell clump which consists of at least 5 neighboring cells with speeds no faster than $\sim 15\mu$m/s.	# cells per cluster
Number of clusters	Number of cluster per area	# /1000 μm^2
Ratio of cells within clusters	Fraction of cells identified to belong to a cluster	–
Total cluster size	Combined size of all clusters within the field of view	# cells
Orthogonal velocity correlation half distance	Distance at which the orthogonal velocity correlation function drops below half its initial value	μm
Parallel velocity correlation half distance	Distance at which the parallel velocity correlation function drops below half its initial value	μm
Local biomass density variation	Variation of local biomass density determined by slicing the image into 64 squares	–
Local enclosed density variation	Variation of local enclosed density determined by slicing the image into 64 squares	–
Density time variation	Variation of local biomass density over time	–

Table 3.1: List of parameters measured from movie data (observables), including a brief description. Table adaped from [66].

During the lag and expansion phase, the swarm shows remarkable behavioral complexity at the microscopic level at different points in space and time [Figs. 3-1(b)–3-1(d)]. The heatmaps [Figs. 3-1(e)–3-1(h)] identify and characterize the microscopic motility behaviors during swarming, indicating a wide range of different behaviors that occur at different space-time points. The strong spatiotemporal variation of each observable indicates the presence of different regimes of bacterial dynamical behaviors. However, some features of the motility behaviors remain hidden when only one or few observables are taken into account, and high-dimensional datasets with many observables that vary in space and time are intrinsically difficult to visualize. Therefore, unsupervised machine learning is applied to identify the dynamical phases from the full set of statistical observables in space and time. To avoid a bias resulting from double counting strongly correlated observables, the pairwise normalized mutual information is first determined. Discounting redündant observables reduced the total number of observables from 23 to 14 (Fig. 3-2(a) and Supplementary Fig. S11 of [66]). To denoise and normalize the data, each of the remaining 14 observables was binned into five categories of equal size (Fig. 3-2(b) and Supplementary Fig. S12 of [66]). After this preprocessing, t-stochastic neighborhood embedding (t-SNE) [87] is used to obtain 2D and 3D representations, followed by the application of k-means clustering to the t-SNE data [Fig. 3-2(c)]. The resulting division into five clusters is robust under variations of target dimensionality and distance metrics used for t-SNE (Supplementary Figs. S13–S16 of [66]). Across independent replicas of the swarming experiment (Supplementary Figs. S13–S16 of [66]), three pure and two coexistence phases are consistently observed. These phases are a single-cell phase (SC) characterized by low cell densities and little collective behavior, a rafting phase (R) exhibiting high fractions of comoving cells, and a biofilm phase (B) where cells are organized in nonmotile structures reminiscent of liquid crystals [61]; the coexistence phases are the mixture of single-cell and rafting behavior (SC + R), as well as the mixture of rafts and biofilm precursors (R + BP), which differ qualitatively [Fig. 3-3(a)] and quantitatively [Fig. 3-3(c)] from the pure phases. The biofilm phase and the rafting phase are consistent with observations of chaining cells in the center of the swarm and rafting cells near its edge [75]. Mapping the distinct phases back onto the space-time heatmap of swarm expansion reveals the complete dynamical phase evolution of bacterial swarming [Fig. 3-2(d)].

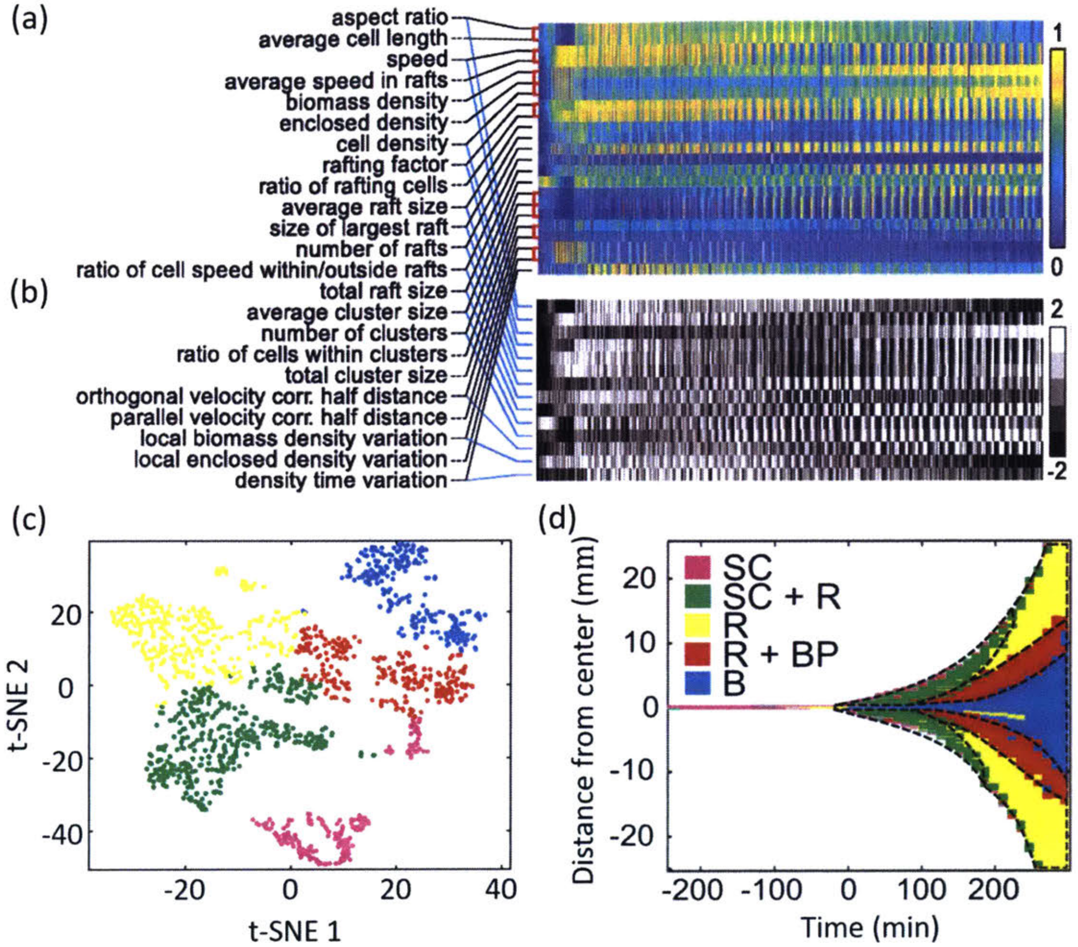


Figure 3-2: Machine learning the swarming phases from microscopic dynamics. (a) Raw data of one swarm expansion experiment, consisting of ~1,500 space-time points (columns) in a 23-dimensional observation space (rows). Additional replicates are shown in SI Appendix, Fig. S11 of [66]. Color bar indicates relative magnitudes scaled to [0, 1]. In the case of strongly correlated observables with high normalized mutual information (marked by red brackets), only one of them is included in the machine-learning analysis. (b) The values of the 14 remaining observables (rows) are binned into five categories as indicated by the color bar, providing the input data for machine learning. (c) The 2D representation of the data in (b), obtained with t-SNE; k-means clustering robustly identifies five main dynamical phases during swarm expansion across independent experiments (n = 3; SI Appendix, Figs. S13–S16 and SI Text of [66]). Phases are labeled with different colors. (d) The emergence of the different phases in time and space during swarm expansion. The dashed lines indicate the approximate outlines of the different phases. Figure adapted from [66].

3.2 Simulations

With identification of the five phases of collective behavior in swarms, it is now possible to identify which cell–cell interaction mechanisms govern the dynamics within each phase and address whether these phases can be explained in terms of common physical principles. To test whether physical forces can account for the dynamics within each behavioral phase and the differences between phases, 2D individual-based active matter simulations are performed. Bacteria are modeled as interacting ellipsoids, all of which share the same minor axis half-length d. Each bacterium α is then described by its major axis half-length ℓ_α, position $\mathbf{x}_\alpha$, and orientation $\hat{\mathbf{n}}_\alpha$. The bacteria are assumed to move at a constant self-propulsion speed v_α in the direction of their orientation. Due to the fact that bacteria swim at low Reynolds number ($Re \approx 10^{-4}$), inertial effects are ignored and the dynamics are approximated as over-damped. Denoting the identity matrix by $\mathbf{I}$, the over-damped translation and orientation dynamics for a single bacterium are

$$\frac{d\mathbf{x}}{dt} = v\hat{\mathbf{n}} - \mathbf{\Gamma}^{-1}\left(\frac{\partial V}{\partial \mathbf{x}}\right) \tag{3.1}$$

$$\frac{d\hat{\mathbf{n}}}{dt} = \left(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}\right)\left(-\mathbf{\Omega}^{-1}\frac{\partial V}{\partial \hat{\mathbf{n}}}\right) \tag{3.2}$$

where use has been made of the fact that $\frac{\gamma_1}{\gamma_0\gamma_\|} \sim 1$ [compare Eq. (2.56)] $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ are

$$\mathbf{\Gamma} = \gamma_0\left[\gamma_\|(\hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) + \gamma_\perp(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\right] \tag{3.3}$$

$$\mathbf{\Omega} = \omega\mathbf{I} \tag{3.4}$$

where γ_0 and ω are the translational and rotational friction coefficients, respectively. $\gamma_\|$ and $\gamma_\perp$ are dimensionless geometric parameters characterizing the longitudinal and transverse friction parameters that depend only on the aspect ratio $a = \ell/d$. The expressions for $\gamma_\|$ and $\gamma_\perp$ given in [130] for rod-like macromolecules Eqs. (2.62a) and (2.62b) are used. V is the total potential of a single cell α for all the N pairwise cell-cell interactions between cell α and cell β ($V = \sum_{\beta=1,\,\beta\neq\alpha}^{N} U$). The steric cell-cell interactions are described in terms of an effective repulsive potential U (Eq. (2.29) with $\lambda_r = 1$ and $\nu = 0$) that depends on the distance $r_{\alpha\beta}$ between neighboring cells α and β, and their orientations $\hat{\boldsymbol{n}}_\alpha$ and $\hat{\boldsymbol{n}}_\beta$ (see Sec. 2.3.3 for more details).

Using d, $\tau_t = \frac{\gamma_0 d^2}{\epsilon_0}$, and ϵ_0, as characteristic length, time, and energy scales, the equations of motion Eqs. (3.1) and (3.2) can be represented in dimensionless form

$$\frac{d\mathbf{x}^*}{dt^*} = v^*\hat{\mathbf{n}} - \left[\frac{1}{\gamma_\parallel}(\hat{\mathbf{n}}\hat{\mathbf{n}}^\top) + \frac{1}{\gamma_\perp}(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^\top)\right]\left(\frac{\partial V^*}{\partial \mathbf{x}^*}\right) \tag{3.5}$$

$$\frac{d\hat{\mathbf{n}}}{dt^*} = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^\top)\left(-\frac{1}{\omega^*}\frac{\partial V^*}{\partial \hat{\mathbf{n}}}\right) \tag{3.6}$$

where superscript * indicates a dimensionless quantity and use has been made of the following definitions and ratios $\mathbf{x}^* = \mathbf{x}/d$, $t^* = t/\tau_t$, $V^* = V/\epsilon_0$, $v^* = v/\frac{\epsilon_0}{\gamma_0 d}$, and $\omega* = \frac{\omega}{\gamma_0 d^2}$. The time scale τ_t is the translation relaxation time, i.e. the time it takes for the bacteria to reach an equilibrium configuration in the absence of self-propulsion.

Anticipated differences exist for the B phase, as the model does not account for the highly elongated and flexible cells. To qualitatively capture these long, incompletely-divided, flexible chains of cells, the model is extended to include growth and division, representing the incompletely-divided cell chains as divided chains of cells. As the cells are non-motile and only interact through steric interactions, the ends of the cells should remain in close proximity during the simulation. Thus, for a cell α

$$\ell_\alpha(t) = \ell_{\alpha 0} e^{ct} \tag{3.7}$$

where c is the growth rate and $\ell_{\alpha 0}$ is the initial length. When a cell reaches its division length $\ell_{\alpha\,\mathrm{divide}}$, the cell will divide. At division, the new positions of the daughter cells (label $d1$ and $d2$) are

$$\mathbf{x}_{d1} = \mathbf{x}_\alpha + \ell_{\alpha\,\mathrm{divide}}/2 \cdot \hat{\mathbf{n}}_\alpha \tag{3.8}$$

$$\mathbf{x}_{d2} = \mathbf{x}_\alpha - \ell_{\alpha\,\mathrm{divide}}/2 \cdot \hat{\mathbf{n}}_\alpha \tag{3.9}$$

respectively. The orientations of the daughter cells are equal to the mother's orientation, and the lengths of the daughter cells are $\ell_{\alpha\,\mathrm{divide}}/2$. The new division length for each daughter cell is drawn from a normal distribution after each division event.

3.2.1 Implementation

In the simulations, new positions and orientations of the bacteria are obtained by numerically integrating the dimensionless translation and orientation dynamics Eqs. (3.5) and (3.6) at each time step. Each 2D simulation runs with a fixed number of cells within a confined square box with periodic boundaries. To ensure the periodic boundaries do not add any undesirable effects, the area of the box is chosen to be 2.25 times as large as the field of view, i.e. the area for which the analysis is executed and videos are generated. Non-motile cells do not self-propel and have a 10-fold higher translation and orientation friction coefficient compared to motile cells, which is motivated by the experimental observation that non-motile cells barely move, even when pushed by several other cells. The cells are initially loaded with random positions and orientations. Analysis of the simulation is performed after it has reached its equilibrium state (after at least 10^5 time steps are executed).

For the R + BP phase, the non-motile cells are arranged into a more realistic configuration than random initial conditions by allowing the non-motile cells to be pushed by the motile cells for 7.5×10^4 time steps. During this time, the translation and orientation friction coefficients of the non-motile cells are equal to the motile cells. After this initialization, the friction coefficients of the non-motile cells are increased by a factor of 100, enforcing that the clusters rarely move as observed in experimental videos.

For the B phase, $N = 50$ non-motile cells are first distributed in a small area ($\sim 12 \times 12$ μm). These cells grow for 2×10^5 time steps. This approximately equals 2 s with a growth factor $c = 2 \times 10^{-5}$, which corresponds to a doubling time of $\sim$ 0.35s. Thus, approximately 5–6 division events per cell can be expected with the final cell number to be about 2500. Once the growth process is completed, the simulation is continued for another 1.5×10^5 time steps for the system to reach equilibrium.

3.2.2 Parameters

A summary of all relevant simulation parameters is given in Table 3.2. The variable parameters for each simulation are the number of cells, their motility, and their shape, which are all directly extracted from the experimental data for each phase. In particular, the number of cells in a simulation is set as the median cell density of a specific phase measured in the experiments. The fraction of non-motile cells in a simulation equals the fraction of cells within non-motile clusters determined in the experiment, and the speed of the motile cells are drawn from a Rayleigh-distribution, whose parameters depend on the experimentally determined bacterial speed. In each phase, all of the aspect ratios of all cells within the field of view of each experimental video are collected. The aspect ratios used for the simulations are drawn from this distribution, after eliminating abnormally small or large values. For the R + BP phase, the long and non-motile cells are simulated by increasing the aspect ratios of the non-motile cells by a factor of 2.

Parameter	Value	Description
d	0.58 μm	Bacteria half-width
ϵ_0	1.575×10^{-16} J	Cell-cell interaction potential strength
γ_0	4.63×10^{-9} kg/s	Translation friction coefficient
ω	1.56×10^{-21} kg $\cdot$m^2/s	Rotation friction coefficient

Table 3.2: Summary of bacterial swarming simulation parameters. Table adapted from [66].

3.3 Results

The simulated dynamics for all of the identified dynamical phases qualitatively agree with their experimental counterparts [Figs. 3-3(a)–3-3(b)]. Additionally, the simulations of the SC, R, SC + R, and R + BP phases are in good quantitative agreement with experiment for many of the measured observables [Fig. 3-3(c) and Fig. 3-4]. As mentioned previously, because the extended simulation model for the B phase represents the long, undivided cells seen in experiments as divided chains of cells, the simulation is not expected to quantitatively match the experiment.

In Fig. 3-4, the simulation results are represented as large symbols, with the error bars indicating the standard deviation based on 20 simulation runs. As shown in Fig. 3-4, there are discrepancies between the simulation and the experiment for

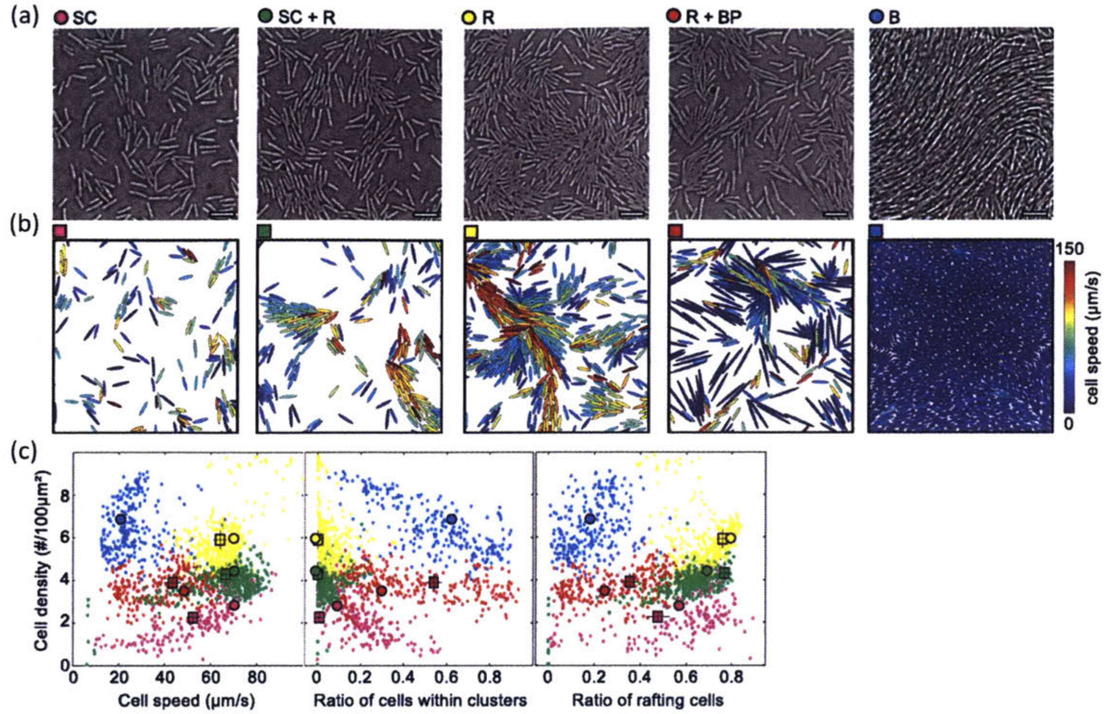


Figure 3-3: Swarming dynamics are dominated by physical cell-cell interactions. (a) Typical images for the phases identified in Fig. 3-2(c): low-density single-cell phase (SC); high-density rafting phase (R) with a high percentage of comoving cells; biofilm phase (B) characterized by long, unseparated cells; and coexistence phases that contain single cells and rafts (SC + R) or rafts and biofilm precursors (R + BP). Scale bar 10 μm. (b) For each phase, simulations were run with the cell shape, motility, and density extracted from the particular phase as input parameters. (c) Detailed quantitative comparisons between experiments (small circles), the particular experimental states shown in (a) (large circles), and simulations (squares; error bars are SDs, $n = 20$) yield good quantitative agreement, except for the B phase, confirming that physical effects determine the four motility-based swarming phases. Figure adapted from [66].

some of the observables (marked in *italic* font). Because a group of cells swimming in the same direction are perfectly aligned in the simulations, the *rafting factor* of the simulations is larger than in experiments for all of the motility phases. This *rafting factor* influences the cutoff value for the determination of rafts, which leads to increases in *number of rafts*, *total raft size*, and *ratio of rafting cells* for the simulations compared to experiments. The *average cluster size* of the simulations are smaller than the experiments for the SC, SC + R, and R phases because the simulations only capture the swarm dynamics not near the swarm front. In the experiments, for the SC and SC + R phases located at the swarm front, the cells at the edge of the swarm hardly move (observed swarm front speed is $0.6 - 5.6$ μm/s), such that they would be classified as non-motile clusters. Further, non-motile clusters form in the experiment due to the growth of non-motile cells or by the gathering of non-motile cells from being pushed around by the motile cells, and only the second is captured in the simulations. The measured *cell length* for the R + BP phase in the simulations are larger than the experiments because of the doubling of the aspect ratios of the non-motile cells. Because the cells in the simulations do not run-and-tumble, the simulation rafts last longer than in experiments, i.e. the rafts split and merge less often. When these rafts move through the field of view, the local density is high at the position of rafts and low in other regions, which leads to a high spatial density fluctuation. On the other hand, the constant splitting and merging of rafts lead to a more homogeneous spatial density distribution in the experiments. Thus, the *biomass density variation* and *enclosed density variation* for the simulation are higher compared to the experiment for the SC + R and R phases.

The numerical investigations show that hydrodynamic interactions are not a dominant effect, but that steric interactions and motility suffice to explain the collective behavior among swarming bacterial cells and account for the differences in the distinct dynamical regimes (Fig. 3-4).

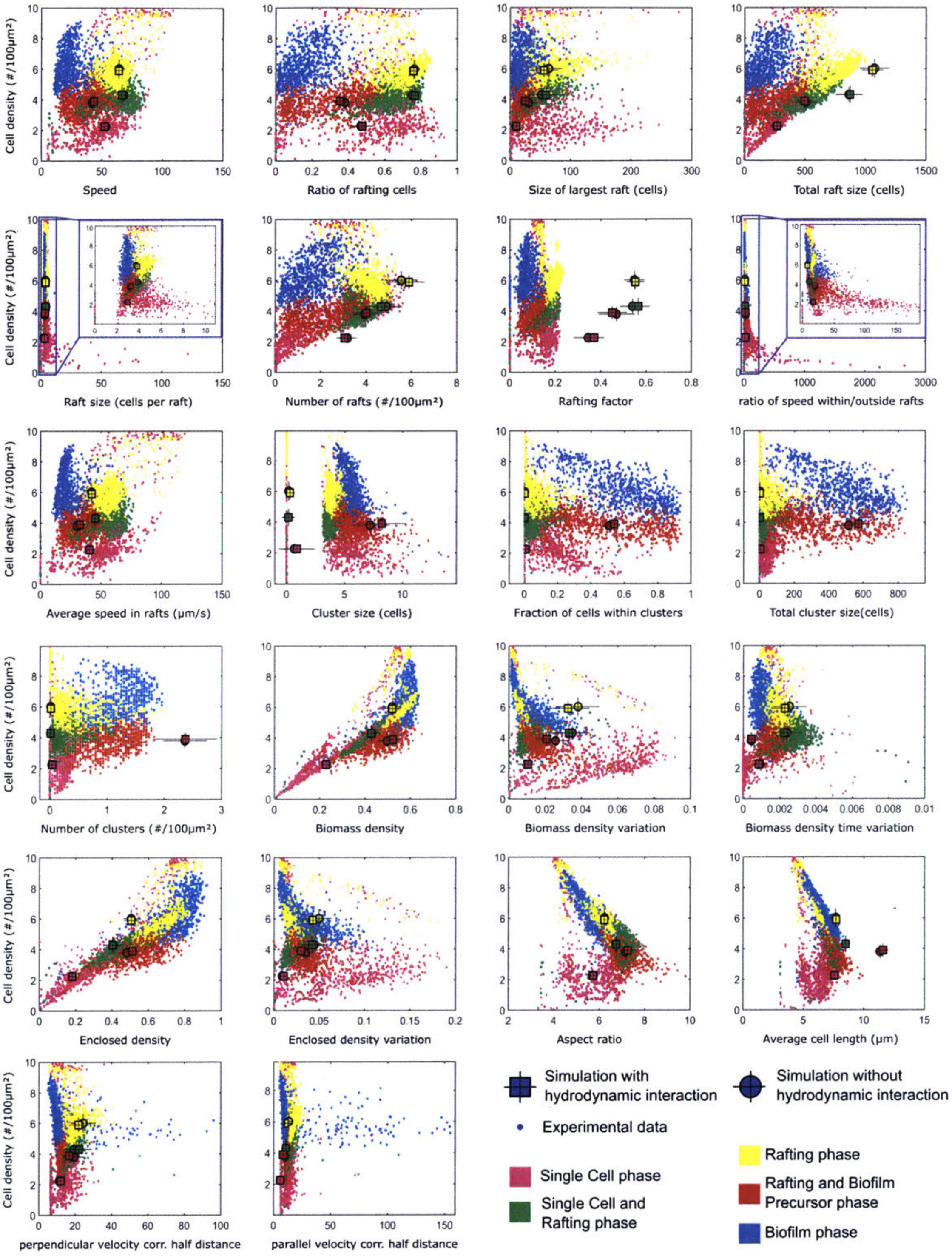


Figure 3-4: Experimentally determined data (small symbols) and simulation results with (large square) and without (large circles) hydrodynamic interactions. Simulation results are averaged over 20 runs, with the standard deviation indicated by black lines. In each panel, small circles, squares, and diamonds correspond to the results of the three different experimental replicates. The 5 different colors correspond to the 5 different dynamical phases. Figure from [66].

Chapter 4

Bacterial biofilms

Vibrio cholerae cells can swim through liquids as isolated individuals, but they are more commonly attached to surfaces, where they grow into clonal colonies termed biofilms, with reproducible spatial organization, global morphology, and cellular arrangements[1] [46, 78]. Biofilm architectures often display striking local nematic order analogous to molecular ordering in abiotic liquid crystals, yet biofilms differ fundamentally in that they are active systems, driven by cell growth and metabolism [34, 56, 138, 151]. As these active nematic systems operate far from thermodynamic equilibrium [90], there are no relevant fundamental conservation laws known that could be used to characterize the biofilm developmental dynamics.

To investigate whether potential-based descriptions can account for the experimentally observed morphologies, new experimental imaging and image analysis techniques are developed to reconstruct and track all individual cells inside growing 3D biofilms (Sec. 4.1) and particle-based simulations implementing an effective cell-cell interaction potential are performed (Sec. 4.2). This combined experimental and simulation analysis implies that local cellular order and global biofilm architecture in these active bacterial communities can arise from mechanical cell-cell interactions, which cells can modulate by regulating the production of particular matrix components (Sec. 4.3). These results establish an experimentally validated foundation for improved continuum theories of active matter and thereby contribute to solving the important problem of controlling biofilm growth.

---

[1] The results of Chapter 4 are published in: Raimo Hartmann, Praveen K. Singh, Philip Pearce, Rachel Mok, Boya Song, Francisco Díaz-Pascual, Jörn Dunkel, and Knut Drescher. Emergence of three-dimensional order and structure in growing biofilms. *Nature Phys.*, 15:251–256, 2019.

4.1 Experiments

To achieve a detailed qualitative and quantitative understanding of such biologically ubiquitous yet physically exotic bacterial communities, new experimental imaging and image analysis techniques are developed for obtaining high spatiotemporal-resolution data of the biofilm development process up to 10^4 cells,[2] representing mid-sized biofilm microcolonies that have already established the architechtural state of macroscopic *V. cholerae* biofilms [36]. By using automated confocal microscopy, with an adaptive live feedback between image acquisition, feature recognition and microscope control, followed by a ground-truth-calibrated, novel 3D image-segmentation technique, the complete 3D biofilm development at cellular resolution with minimal phototoxicity [Figs. 4-1(a)–4-1(b)] and minimal segmentation error can be observed (see Methods and Supplementary Information of [61] for more details). The high temporal resolution ($\Delta t = 5 - 10$ min) allows for cell lineage reconstruction, measurements of local growth rates, and the identification of all cells in a field of view that are not related to the original biofilm founder cell [Figs. 4-1(b)–4-1(d)]

When investigating whether the non-equilibrium dynamics of biofilm development and the emergence of local order can be captured quantitatively through effective cell-cell interaction potentials, it is important to account for the essential biophysical processes—cell growth, cell division, cell-surface interactions and cell-cell interactions [29, 55, 85, 105, 111, 116, 131, 138, 148]. Whereas growth and division are driven by nutrient availability and metabolism, cell-surface and cell-cell attractions are typically mediated by secreted or membrane-associated polysaccharides and proteins [89, 105]. For *V. cholerae* biofilms, the molecular basis for cell-cell interactions has been intensively investigated: cells are embedded in a self-secreted extracellular

---

[2] The single-cell experiments and data analysis were performed by Raimo Hartmann, Praveen K. Singh, and Francisco Díaz-Pascual of the Drescher lab.

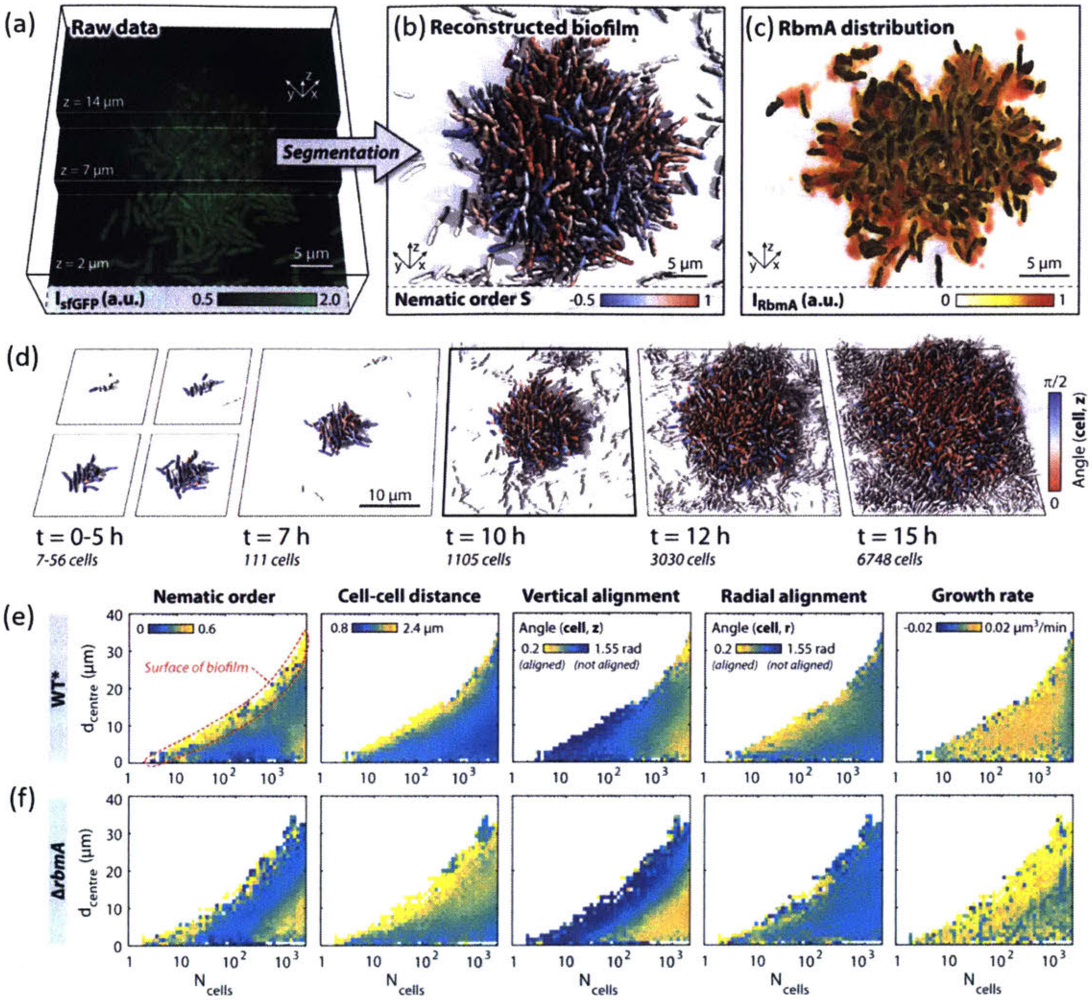


Figure 4-1: Dynamics of *V. cholerae* biofilm formation. (a) Cells constitutively expressing a green fluorescent protein (sfGFP) were imaged with spinning disc confocal microscopy. Images at three different z planes are highlighted. (b) 3D reconstruction of the biofilm shown in (a), where each cell is colored according to the nematic order parameter $S = \langle 3/2(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)^2 - 1/2 \rangle$ in its vicinity. High-time-resolution ($\Delta t = 5 - 10$ min) imagining allows tracking of cell lineages and discrimination of cells (white) that are not direct descendants of the biofilm founder cell. (c) The extracellular matrix protein RbmA mediates cell-cell adhesion and is distributed throughout the biofilm, as visualized by immunofluorescence. (d) Time-resolved WT* biofilm (a rugose wild-type strain with straight cell shape) growth series. Each cell is colored according to the cellular alignment with the z axis (for the $\Delta rbmA$ mutant see Supplementary Fig. 6 in [61]). Heatmaps showing spatially resolved single-cell measurements of different biofilm structural properties inside (e) WT* and (f) $\Delta rbmA$ biofilms, which are used to characterize biofilm formation ($n > 3$ biofilms, standard deviations are shown in Supplementary Figs. 5 and 7 and the differences among both strains are highlighted in Supplmentary Fig. 8 of [61]) as a function of the distance to the biofilm center (d_{centre}) and the number of cells inside the biofilm (N_{cells}). Figure from [61].

matrix composed of the *Vibrio* polysaccharide (VPS), extracellular DNA, and proteins [6, 48, 129]. The osmotic pressure resulting from a high concentration of matrix components in the intercellular space, as well as steric cell-cell interactions, are both expected to contribute to cell-cell repulsion. Cell-cell attraction is primarily mediated by the protein RbmA, which localizes throughout the biofilm [Fig. 4-1(c)] [6, 48] and links cells to each other [47, 48, 88]; its expression levels are inversely related to cell-cell spacing (Fig. 4-7). VPS also weakly binds cells together, yet elevated levels of VPS production do not cause stronger cell-cell attraction or decreased cell-cell spacing (see Supplementary Fig. 11 in [61]). Based on these cell-cell interaction processes, biofilm architectures are hypothesized to be primarily determined by the relative strength of the effective mechanical cell-cell attraction and repulsion forces.

To determine the impact of cell-cell attraction, the 3D biofilm architecture dynamics of a rugose wild-type strain with straight cell shape (WT*) are quantitatively compared with that of a mutant strain ($\Delta rbmA$) with significantly weakened intercellular adhesion (see Methods of [61]). Biofilms grown in a low-shear environment approximately display hemispherical symmetry [Fig. 4-1(d)], which allows the characterization of the biofilm architectures [Fig. 4-1(e)] to be a function of the distance to the biofilm center in the basal plane d_{centre} using the cell number in the biofilm N_{cells} as a quantification of the developmental state. The measurements reveal strong structural differences between the outer biofilm layer and its central part, as well as several distinct architectural phases of the biofilm during growth [Figs. 4-1(e)–4-1(f)]. Interestingly, the cellular growth rate remains homogeneous in space during WT* biofilm development in the conditions and for the biofilm sizes studied here (Fig. 4-1(e) and Supplementary Fig. 5 of [61]), in contrast to theories assuming steep nutrient gradients inside biofilms [105, 147]. The nematic order, cell-cell spacing, and cellular orientations with respect to the vertical ($\mathbf{z}$) and radial ($\mathbf{r}$) directions differ significantly between WT* and $\Delta rbmA$ mutants (Figs. 4-1(e)–4-1(f) and Supplementary Figs. 5–8 of [61]), revealing the strong effect of cell-cell adhesion on biofilm architecture dynamics.

4.2 Simulations

Based on the high-resolution spatiotemporal data of biofilm development of different bacterial strains, the hypothesis that the biofilm internal structure and external shape originate from mechanical interactions between cells is investigated. The cells are modeled as interacting ellipsoids of half-length ℓ and half-width d, described by their position $\mathbf{x}$ and orientation $\hat{\mathbf{n}}$. Because cells live at low Reynolds number ($Re \approx 10^{-4}$), the dynamics are approximated as over-damped, ignoring any inertial effects. Bacterial self-propulsion is neglected as cell movement in *V. cholerae* biofilms is dominated by passive cell displacement from biofilm expansion [61]. Cells can interact with the wall boundary and other cells through interaction potential functions, U_{bdy} and V.

U_{bdy}, Eq. (2.7), encodes cell-surface interactions through a repulsive interaction potential that is proportional to the overlap between a cell and the wall boundary (see Sec. 2.2 for more details). V is the total potential of a single cell α for all the N pairwise cell-cell interactions between cell α and cell β ($V = \sum_{\beta=1,\, \beta\neq\alpha}^{N} U$). The effective mechanical interactions are described in terms of an effective potential U (Eq. (2.29), reproduced below) that depends on the distance $r_{\alpha\beta}$ between neighboring cells α and β, and their orientations $\hat{\boldsymbol{n}}_\alpha$ and $\hat{\boldsymbol{n}}_\beta$ (see Sec. 2.3.3 for more details).

$$U = \epsilon_0 \epsilon_1 \left(e^{-\frac{\rho_{\alpha\beta}^2}{\lambda_r^2}} + \frac{\nu}{1 + e^{\frac{\rho_a - \rho_{\alpha\beta}}{\lambda_a}}} \right) . \tag{4.1}$$

where $\rho_{\alpha\beta} = r_{\alpha\beta}/\sigma$ is the shape-normalized cell-cell distance. σ is the interaction length scale between nonidentical ellipsoidal cells which depends on the instantaneous cell lengths, the orientation of the cells relative to each other, and the individual cell orientations. The amplitude is set by the interaction strength ϵ_0 and strength parameter ϵ_1, which accounts for the instantaneous cell lengths and cell orientations. The first term of the interaction potential describes the combined effects of hard steric and osmotic repulsion with range λ_r. The second term corresponds to cell-cell attraction and adds an attractive part of relative depth ν, width λ_a, and position ρ_a. The potential is assumed to be independent of the biofilm developmental state or nutrient levels. As shown in Sec. 4.3, this simplification suffices to capture the main features of the small to medium-sized biofilms studied here but is expected to become inaccurate at the later stages of biofilm development, when spatiotemporal heterogeneities become relevant.

Denoting the identity matrix by $\mathbf{I}$, the over-damped translational and orientation dynamics for a single cell are

$$\frac{d\mathbf{x}}{dt} = \mathbf{\Gamma}^{-1}\left[-\frac{\partial U_{\text{bdy}}}{\partial \mathbf{x}} - \frac{\partial V}{\partial \mathbf{x}}\right] \tag{4.2}$$

$$\frac{d\hat{\mathbf{n}}}{dt} = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\left[\mathbf{\Omega}^{-1}\left(-\frac{\partial U_{\text{bdy}}}{\partial \hat{\mathbf{n}}} - \frac{\partial V}{\partial \hat{\mathbf{n}}}\right)\right] \tag{4.3}$$

where $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ are

$$\mathbf{\Gamma} = \gamma_m\left[\gamma_{\|}(\hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) + \gamma_{\perp}(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\right] \tag{4.4}$$

$$\mathbf{\Omega} = \omega_m \gamma_R \mathbf{I} \tag{4.5}$$

Here, γ_m and ω_m are the translational and rotational drag coefficients for Stokes' drag in the extracellular matrix for a spheroid ($\gamma_m = 6\pi\mu_m d$, $\omega_m = 8\pi\mu_m \ell d^2$ where μ_m is the viscosity of the extracellular matrix). Surface adhesion of the cells is captured by increasing the magnitude of the friction tensor $\mathbf{\Gamma}$ by a factor of 30 if the cells are within $1.5d$ of the surface. $\gamma_{\|}$, $\gamma_{\perp}$, and γ_R are dimensionless geometric parameters characterizing the longitudinal and transverse friction parameters that depend only on the aspect ratio $a = \ell/d$ of the cell. The expressions for $\gamma_{\|}$, $\gamma_{\perp}$, and γ_R given in [60] for elliptical particles Eqs. (2.63) are used. The instantaneous cell length growth rate for a single cell is

$$\frac{d\ell}{dt} = \frac{\ell}{\tau_g}\ln(2) \tag{4.6}$$

τ_g is the growth time constant (obtained by experimental measurements). The division of the cells follow the adder model [128] as described in Sec. 2.1.

Using d, $\tau_t = \frac{\gamma_m d^2}{\epsilon_0}$, and ϵ_0 as characteristic length, time, and energy scales, Eqs. (4.2), (4.3), and (4.6) are recast into dimensionless form

$$\frac{d\mathbf{x}^*}{dt^*} = \left[\frac{1}{\gamma_{\|}}(\hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) + \frac{1}{\gamma_{\perp}}(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\right]\left[-\frac{\partial U^*_{\text{bdy}}}{\partial \mathbf{x}^*} - \frac{\partial V^*}{\partial \mathbf{x}^*}\right] \tag{4.7}$$

$$\frac{d\hat{\mathbf{n}}}{dt^*} = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\mathsf{T}}) \left[\frac{3}{4a\gamma_R} \left(-\frac{\partial U^*_{\mathrm{bdy}}}{\partial \hat{\mathbf{n}}} - \frac{\partial V^*}{\partial \hat{\mathbf{n}}} \right) \right] \tag{4.8}$$

$$\frac{d\ell^*}{dt^*} = \frac{\tau_t}{\tau_g} \ell^* \ln(2) \tag{4.9}$$

where superscript * indicates a dimensionless quantity and use has been made of the following definitions and ratios $\mathbf{x}^* = \mathbf{x}/d$, $\ell^* = \ell/d$, $t^* = t/\tau_t$, $V^* = V/\epsilon_0$, and $U^*_{\mathrm{bdy}} = U_{\mathrm{bdy}}/\epsilon_0$. Note that τ_t can be interpreted as the translation relaxation time, i.e. a time scale of how long it takes for a bacterium to reach an equilibrium configuration from the cell-cell interaction potential. If the overlap coordinate $\mathbf{z}^* > 0$, the dimensionless boundary potential is $U^*_{\mathrm{bdy}} = U_{\mathrm{bdy}}/\epsilon_0 = \epsilon_r e^{\mathbf{z}^*/\sigma^*_{\mathrm{bdy}}}$ where $\epsilon_r = \epsilon_{\mathrm{bdy}}/\epsilon_0$, $\mathbf{z}^* = \mathbf{z}/d$, and $\sigma^*_{\mathrm{bdy}} = \sigma_{\mathrm{bdy}}/d$.

4.2.1 Implementation

In the simulations, new positions and orientations of the bacteria are obtained by numerically integrating the dimensionless translation and orientation dynamics Eqs. (4.7) and (4.8) and growth law Eq. (4.9) at each time step. The wall boundary is modeled as an xy plane located at the origin with $\hat{\mathbf{N}} = (0, 0, 1)$ and $\mathbf{S} = (0, 0, 0)$. A single cell is initially loaded with $\mathbf{x} = (0, 0, d)$ and $\hat{\mathbf{n}} = (1, 0, 0)$ which grows and divides. The simulation ends when the specified number of cells is reached. The same analysis algorithms used to extract features from the experimental biofilms are used for the simulations.

4.2.2 Parameters

Bypassing previous limitations of individual-based biofilm models [62, 81], the parameters of our model (cell aspect ratio, division time distribution; Table 4.4) are determined from single-cell properties of experimental biofilms. The average half-width $d = 0.2775$ μm and growth time constant $\tau_g = 6130$ s are obtained from experimental measurements. μ_m, the dynamic viscosity of the EPS matrix, is estimated to be 1 Pa·s [102], resulting in a drag coefficient of $\gamma_m = 5.23 \times 10^{-6}$kg/s. To prevent penetration of the cells into the boundary, the strength of the cell-surface

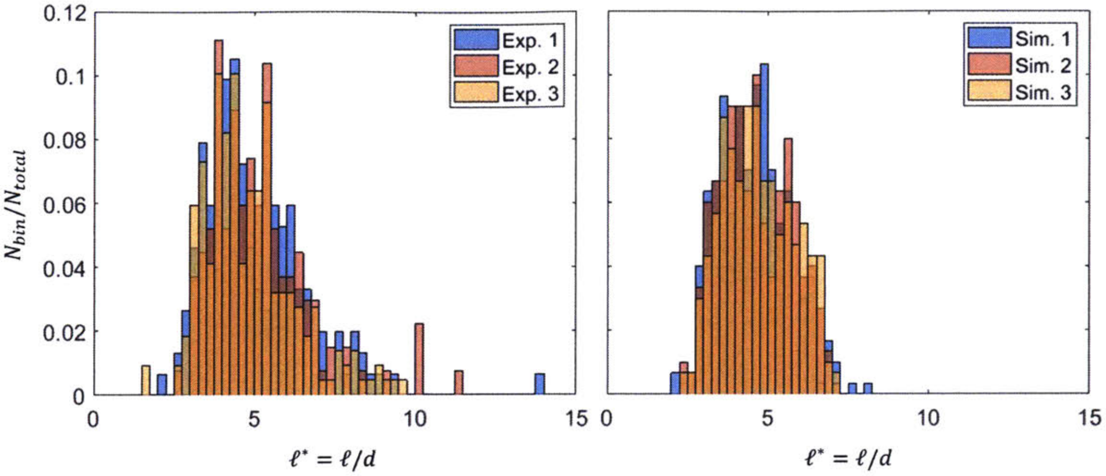


Figure 4-2: Experimental (left) and simulation (right) histograms of the dimensionless length $\ell^* = \ell/d$ at $N_{\text{total}} = 300$ cells. The histograms are normalized by dividing the number of cells in each bin by the total number of cells such that the height of the bars sum to 1. The experiments are at flow rate 0.1 μL/min. The averages of the mean and standard deviation of ℓ^* of the 3 experiments are 4.99 and 1.53, respectively. For the 3 simulations, the averages of the mean and standard deviation of ℓ^* are 4.61 and 1.08, respectively.

interaction is chosen to be $\epsilon_{\text{bdy}} = 10\epsilon_0$, resulting in $\epsilon_r = 10$. The interaction length scale of the boundary potential σ_{bdy} should be of the order to the half-width of a bacterium, which non-dimensionalizes to $\sigma^*_{\text{bdy}} = 1$. The dimensionless length added after division ℓ^*_{add} and dimensionless standard deviation for the birth size $\sigma^*_{\text{birth size}}$ are chosen such that the resulting simulation distributions of dimensionless lengths ℓ^* of 300 cells represent a tighter distribution around the experimental ℓ^* means (Fig. 4-2), as the larger experimental ℓ^* are likely a result of under-segmentation. Because the orientation of daughter cells are highly correlated to the orientation of the parent cell, a high value of the concentration parameter κ for the von Mises-Fisher distribution (used in bacterial division) is chosen, with $\kappa = 100$.

Fitting the cell-cell potential parameters

To obtain the key potential parameters ϵ_0, λ_r, ν, λ_a, and ρ_a for *V. cholerae* biofilms, the attractive term in U is assumed to be attributed primarily to RbmA levels, with the VPS acting as a Woods–Saxon background potential [144] akin to the mean-field potential in nucleon models. This potential is assumed to provide the weak cell-cell binding (with binding strength approximately equal to the Stokes drag felt by

Name	Parameter	Unit	Description
Axial cell position	*height*	μm	Z-coordinate of each cell centroid.
Cell alignment with direction of flow	$angle(\hat{\mathbf{n}}_{\text{cell}}, \mathbf{flow})$	rad	Angle between each cell's major axis (see *cell orientation*) and the direction of the flow.
Radial alignment	$angle(\hat{\mathbf{n}}_{\text{cell}}, \mathbf{r})$	rad	Angle between each cell's major axis (see *cell orientation*) and the radial vector **r** pointing from the biofilm's center of mass projected to the bottom of the biofilm ($z = 0$ μm) to the corresponding cell's center.
Vertical alignment	$angle(\hat{\mathbf{n}}_{\text{cell}}, \mathbf{z})$	rad	Angle between each cell's major axis (see *cell orientation*) and the **z** axis.
Cell volume	v	μm^3	Cell volume as obtained by the segmentation.
Cell size	l, h, w	μm	Cell dimensions (length, height and width), as determined by principal component analysis (PCA) of the segmented cell (equivalent to fitting an ellipsoid into the cell and deriving the three eigenvalues which correspond to the lengths of the main axes).
Cell orientation	$\hat{\mathbf{n}}_{\text{cell}} = (e_1, e_2, e_3)$		Vector of the cell's major axis determined by PCA (eigenvector with largest eigenvalue).
Distance to center of mass	d_{center}	μm	Assuming radially symmetric biofilm growth in spherical coordinates (which is a good estimate for the low flow regime) each cell can be described spatially by its distance to the center of mass projected onto the $z = 0$ μm plane.
Distance to nearest neighbor	$d_{\text{nearest neighbor}}$	μm	Euclidian centroid-to-centroid distance to the nearest cell.
Distance to surface	d_{surface}	μm	Shortest distance of a particular cell to the outer surface of the biofilm.
Local cell density	ρ_{local}	–	Biovolume of cells in a vicinity (defined as a sphere of radius 3 μm around a cell) normalized by the volume of the vicinity.
Local order (nematic order)	S	–	Nematic order parameter $S = \langle 3/2(\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j)^2 - 1/2 \rangle$ in a vicinity defined as a sphere of radius 3 μm around a cell. $\hat{\mathbf{n}}_i$ and $\hat{\mathbf{n}}_j$ are the orientation vectors of cells i and j, respectively. For $S = 1$, the cells are perfectly aligned with their neighbors. For $S = 0$, the cells are randomly oriented [36].
Single cell growth rate	*growth rate*	μm^3/s	Single cell growth rate dv_i/dt of cell i (fixed by cell tracking) with volume v_i as determined by comparing the cell volume in frame N and $N+1$ at times t_N and t_{N+1}: $\Delta v_i/\Delta t_N = (v_{i,N+1} - v_{i,N})/(t_{N+1} - t_N)$. Potential bias caused by over-/under-segmentation and/or cell dispersion/off-shearing was corrected by setting $v_{i,N+1}$ to zero for cells with no children and setting $v_{i,N}$ to zero for cells with no parents. Using this approach, the global biofilm volume $V(t)$, as determined by segmentation, matches $V(t) = \int \sum \Delta v_i dt$.

Table 4.1: List of calculated single-cell biofilm features. Table adapted from [61].

a cell at the edge of the biofilm at low flow rate of 0.1 μL/min) that prevents the disintegration of biofilms from fluid shear acting on $\Delta rbmA$ mutant biofilms. This assumption is motivated by the experimental findings that increased VPS levels do not increase the cell–cell attraction (see Supplementary Fig. 11 of [61]), yet biofilms that lack RbmA display a small residual mechanical cohesion (see Fig. 3e of [61]), indicating that VPS does contribute weakly to cell–cell binding. Note that even at small sizes, the *V. cholerae* biofilms used in this study produce RbmA and VPS. In simulations performed at zero shear, the VPS contribution to cell-cell attraction can be neglected as the Woods-Saxon potential is approximately constant in the bulk of the biofilm.

The values of the cell-cell potential parameters are obtained by comparing simulations with experiments through feature vectors which show the development of the biofilm architecture. The feature vector consists of a set of parameters that represent the phenotype and architecture as fully and accurately as possible. These parameters include the following averaged single cell parameters: local order, vertical alignment, radial alignment, cell-cell distance, local density, and cell length and the following global biofilm parameters: global density, biofilm aspect ratio, and biofilm base circularity (descriptions of these parameters can be found in Tables 4.1 and 4.2). The nematic order, vertical alignment, radial alignment, cell-cell distance, and local density parameters are spatially resolved into the biofilm core ($d_{\text{center}} < \max(d_{\text{center}})/2$) and the biofilm shell ($\max(d_{\text{center}})/2 < d_{\text{center}} < \max(d_{\text{center}})$). To capture the time evolution and logarithic growth of these parameters, biofilms are sampled at 40 log-

Name	Parameter	Unit	Description
Aspect ratio	Z/XY	–	Biofilm height divided by average biofilm base radius.
Biofilm base circularity	B_{circ}	–	Deviation of biofilm cross section from a circle $B_{\text{circ}} = \|1 - Z/XY\|$. The cross section is a slice of the biofilm with a xz plane.
Global cell density	ρ_{biofilm}	μm^{-3}	Number of cells divided by the volume of the biofilm's convex hull.
Biofilm volume	V	μm^3	Sum of the volume of all individual cells.

Table 4.2: List of calculated global biofilm features. Table adapted from [61].

Parameter	Normalization range [min - max]	Weight
Nematic order	0 - 0.8	10
Vertical alignment	0.2 - 1.37 rad	10
Radial alignment	0.2 - 1.37 rad	3
Cell-cell distance	0.8 - 2 μm	20
Local density	0 - 0.2	10
Cell length	1 - 3 μm	1
Global density	0 - 0.015 μm^{-3}	5
Biofilm aspect ratio (Z/XY)	0 - 0.8	1
Biofilm base circularity	0 - 0.8	1

Table 4.3: Biofilm parameters and corresponding normalization ranges and weights. Table adapted from [61].

arithmically spaced cell numbers, ranging from 10 to 300 cells. The different parameters are normalized by the typical data ranges occurring in the data sets and weighted according to their relative importance in representing biofilm phenotypes, as summarized in Table 4.3.

Therefore, these feature vectors (Fig. 4-3) show the temporal variation up to 300 cells of 14 different architechtural properties of the biofilms. For both WT* and Δ*rbmA* biofilms, the nematic order in the core, radial alignment in the core, and cell length is approximately constant throughout the development. The nematic order decreases in the shell, indicting a more random configuration among cells and their neighbors as WT* and Δ*rbmA* biofilms grow. The angle between each cell's major axis and the $\mathbf{z}$ axis (biofilm parameter vertical alignment) decreases in the core and shell for both WT* and Δ*rbmA* biofilms, indicating that the number of vertically aligned cells increases as the biofilm develops, with more vertical cells in the shell than core at 300 cells. As the biofilm grows, the radial alignment in the shell increases in both WT* and especially for Δ*rbmA* biofilms, which indicates that the cells in the shell are aligning less with the $\mathbf{r}$. The cell-cell distance decreases for WT* in the core and shell and increases for Δ*rbmA* in the core, in agreement with the fact that WT* have RbmA proteins linking cells together while Δ*rbmA* do not. In the shell, the cell-cell distance of Δ*rbmA* remains constant. The global density decreases initially and then remains approximately constant for WT* and decreases continuously for Δ*rbmA* biofilms. This indicates that the WT* and especially the

$\Delta rbmA$ biofilms expand rapidly. The decrease in biofilm base circularity for both the WT* and $\Delta rbmA$ biofilms indicates that the deviation between their cross sections and a circle increases as the biofilm develops. The aspect ratio (Z/XY) increases for both biofilms, implying that the biofilms increase in height more than in width. Because of the growth, the number of neighboring cells increase for both biofilms which leads to an increase of the local density in the shell and core. The similarity between a simulation and an experimental biofilm is assessed in terms of the mean square distance (MSD) between the two feature vectors.

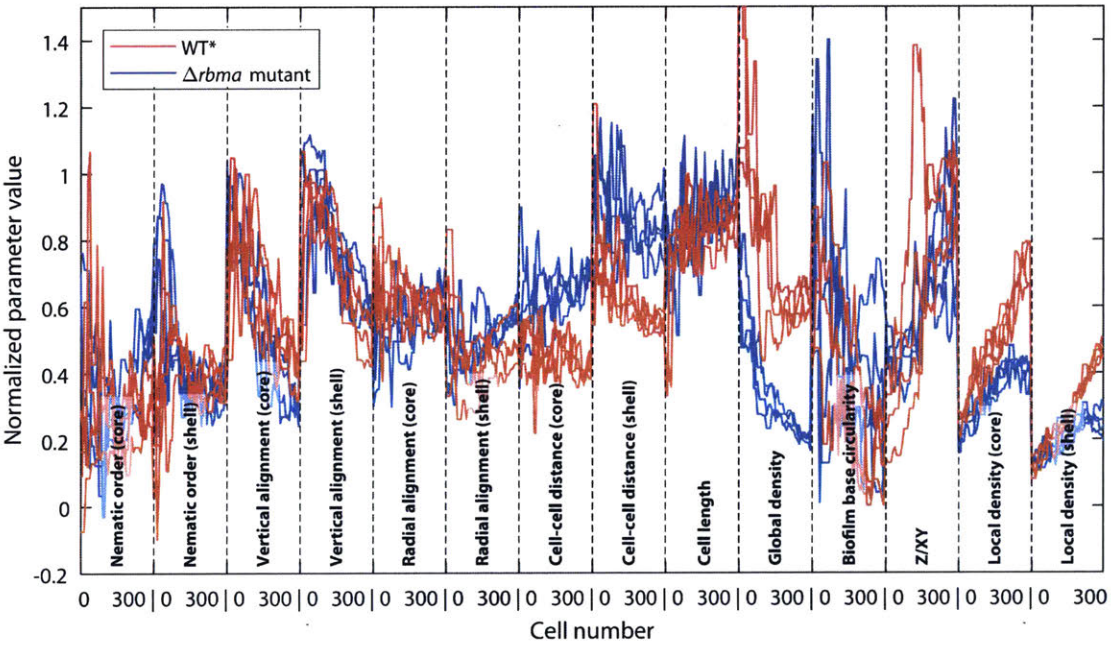


Figure 4-3: Characteristic biofilm feature vector for several biofilms of the WT* (red) and the $\Delta rbmA$-mutant (blue). The parameters correspond to the ones listed in Table 4.3, where relevant ones are spatially resolved into the core and shell of the biofilm. Figure from [61].

Repulsive cell-cell potential parameters

The repulsive parameters ϵ_0 and λ_r of the potential U are obtained by comparing experimental $\Delta rbmA$ biofilms, which lack the attractive potential term ($\nu = 0$), with simulated biofilms, using the MSD of the feature vectors as a metric. The energy scale of cell-cell interactions ϵ_0 is estimated to be within several orders of magnitude of the energy scale of interactions between the cells and the flow ϵ_{flow} such that $\epsilon_{\text{flow}} \leq \epsilon_0 \leq 10^4 \epsilon_{\text{flow}}$ where $\epsilon_{\text{flow}} = 0.005 \times 10^{-18} J$. $\epsilon_{\text{flow}} = \mathbf{F}_{\text{flow}} \cdot 0.1 \mu\text{m}$ where 0.1 μm is the estimated attractive length scale between cells. The Stokes drag force $\mathbf{F}_{\text{flow}}$ is calculated by considering Stokes flow with flow rate 0.1 μL/min in a channel with the dimensions used the experiments past a sphere with the mean cell volume of 0.4 μm^{-3} that is located 2.4 μm above a no-slip boundary, which is the average cell-cell distance at the biofilm surface. The cell-cell distance in the biofilm core, where cell-cell repulsion dominates, is around 1 μm. This corresponds to $\lambda_r = 1.4$ for a representative range parameter of $\sigma = 0.7$ μm, which is the value σ would take for the interaction between two spheres with mean cell volume of 0.4 μm^{-3}. Therefore, λ_r is estimated to be between 0.8 to 2. Simulations resulting from a systematic variation of ϵ_0 and λ_r results in a broad minimum of MSD values (Fig. 4-4), with the corresponding biofilm phenotypes shown in Fig. 4-5, which show that as both ϵ_0 and λ_r increase the cells spread out and become more aligned with their neighbors. Best-fit simulations (with parameters $\epsilon_0 = 10\epsilon_{\text{flow}} = 5 \times 10^{-20}$ J and $\lambda_r = 1.65$) show high similarity to experiments (Fig. 4-5).

The joint effects of steric cell-cell repulsion and osmotic pressure-mediated cell-cell

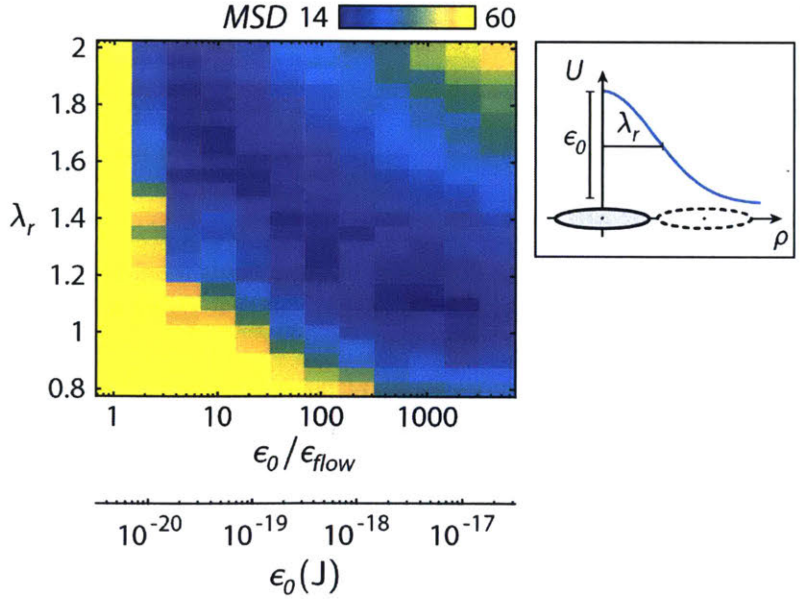


Figure 4-4: Parameter scan to test the influence of cell-cell repulsion on the biofilm architechture. Mean MSD values between feature vectors of $\Delta rbmA$ biofilm experiments ($n > 3$ biofilms) and simulations upon variation of cell-cell interaction strength ϵ_0 and repulsion range λ_r. Inset: effect of parameter variation on the cell-cell interaction potential. Figure from [61].

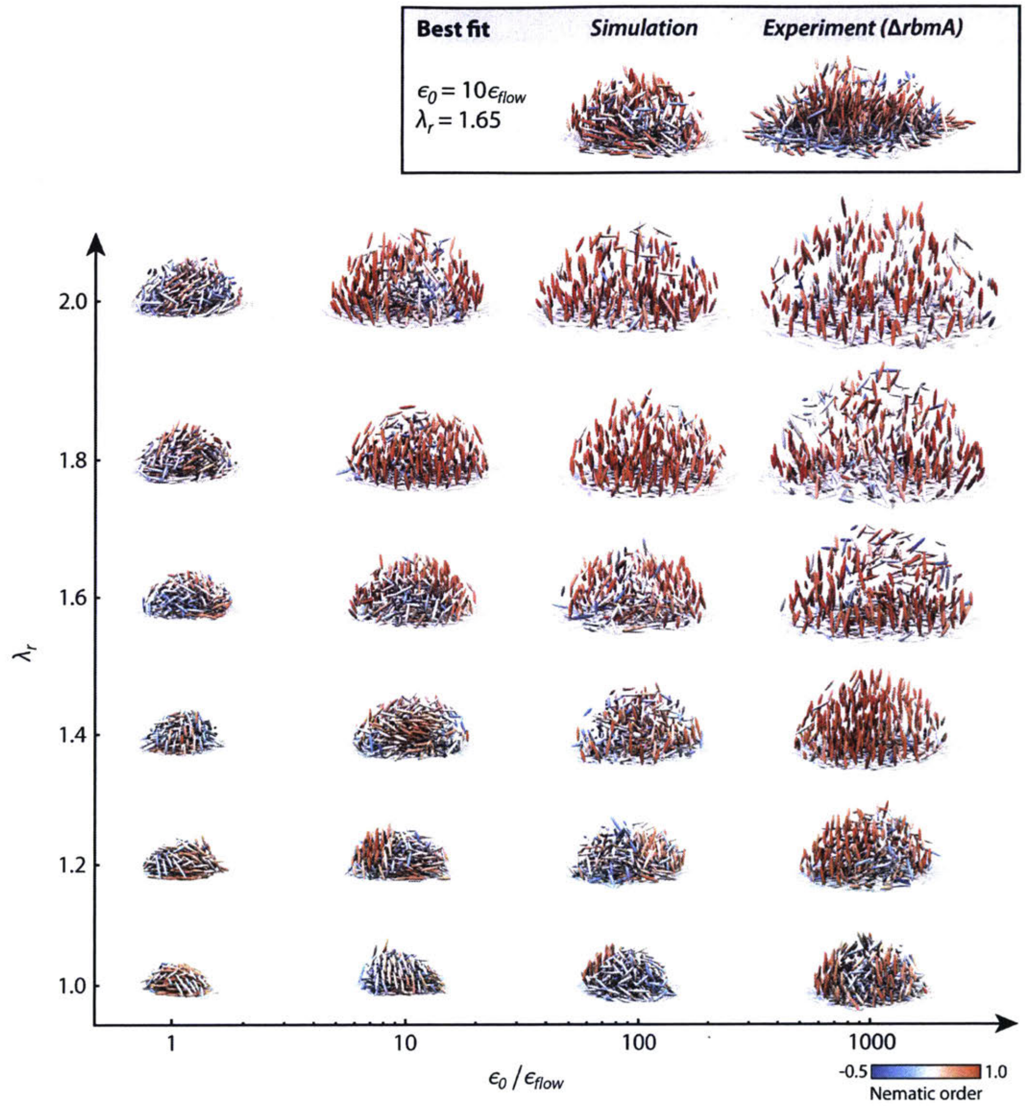

Best fit
Simulation
Experiment (ΔrbmA)
ϵ0 = 10ϵflow
λr = 1.65
2.0
1.8
1.6
1.4
1.2
1.0
λr
1
10
100
1000
ϵ0 / ϵflow
-0.5
1.0
Nematic order

Figure 4-5: Resulting simulated biofilms for different levels of cell-cell repulsion. The figures shows a subset of biofilm renderings corresponding to a range of different values for ϵ_0 and λ_r. The color of each cell corresponds to the nematic order. Figure from [61].

repulsion are modeled with a relatively soft Gaussian function. To test if this model is appropriate, the effect of both contributions to cell-cell repulsion is tested separately where a second, short-ranged, and very strongly repulsive Gaussian function is embedded into the existing potential.

$$U = \epsilon_0 \epsilon_1 \left(\nu_{\text{steric}} e^{-\frac{\rho_{\alpha\beta}^2}{\lambda_{r,\text{steric}}^2}} + e^{-\frac{\rho_{\alpha\beta}^2}{\lambda_r^2}} + \frac{\nu}{1 + e^{\frac{\rho_a - \rho_{\alpha\beta}}{\lambda_a}}} \right) \tag{4.10}$$

The added Gaussian function $\nu_{\text{steric}} e^{-\frac{\rho^2_{\alpha\beta}}{\lambda^2_{r,\text{steric}}}}$ represents steric repulsion and is peaked at $\rho = 0$ and characterized by ν_{steric} and $\lambda_{r,\text{steric}} < \lambda_r$. ϵ_0 and λ_r are fixed to the optimal values obtained by the comparison with the experimental values for the $\Delta rbmA$ biofilm. The exploration of the parameter space is shown in Fig. 4-6. The additional term for hard steric repulsion in Eq. (4.10) does not have any influence on the biofilm phenotype as the MSD values show no variance irrespective of interaction strength ν_{steric} and range $\lambda_{r,\text{steric}}$. This justifies modeling the effects of hard steric and soft, osmotic pressure-mediated repulsion in a combined manner, as the longer length scale of the soft, osmotic pressure-mediated repulsion prevents cells from getting close enough to interact directly through hard steric repulsion.

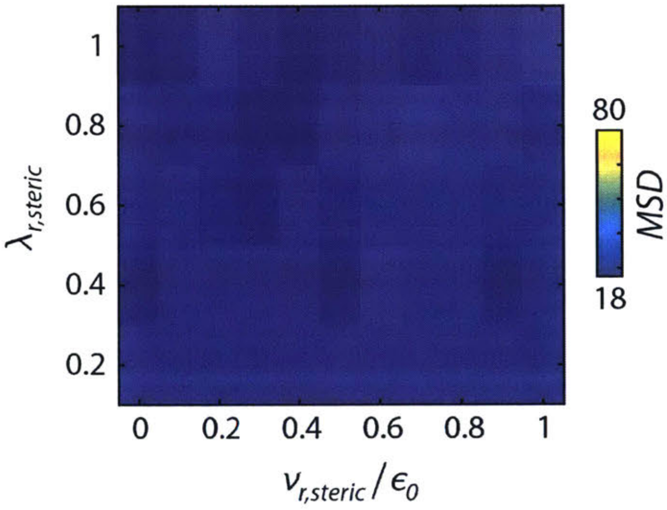


Figure 4-6: Parameter scan to test the influence of hard steric cell-cell repulsion on biofilm phenotype. The figure shows MSD values upon variation of $\lambda_{r,\text{steric}}$ and ν_{steric}. Owing to its longer length scale, soft, osmotic pressure-mediated repulsion prevents cells from getting close enough to interact directly through hard steric repulsion. Figure from [61].

Attractive cell-cell potential parameters

Because the attraction parameters (ν, ρ_a, λ_a) in potential U depend on the concentration of RbmA, *V. cholerae* is genetically modified so that the production of RbmA (and therefore the strength of the attraction) can be tuned, by adding different concentrations of a compound that induces the *rbmA*-expression construct

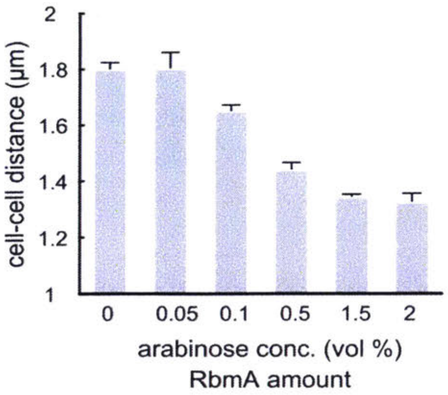


Figure 4-7: Increased RbmA production (achieved by increasing the arabinose concentration) decreases the average cell-cell distance in biofilms. Without arabinose, no RbmA is produced and the biofilm architecture is identical to the $\Delta rbmA$ mutant ($n > 3$ biofilms). Figure adapted from [61].

homogeneously inside the biofilm: arabinose. Experimentally, increasing arabinose concentrations resulted in decreased cell–cell spacing (Fig. 4-7), consistent with the assumption that RbmA mediates cell–cell attraction. The repulsive component (ϵ_0, λ_r) is fixed based on the $\Delta rbmA$ biofilms. The attraction shift ρ_a is estimated for each arabinose concentration by considering the typical cell-cell distance at the edge of biofilms, where attraction dominates, which suggests using values of ρ_a between 2 and 4 (again using $\sigma = 0.7$ μm). The attraction width λ_a is estimated by considering the standard deviation of cell-cell distances near the edge of biofilms, which was found

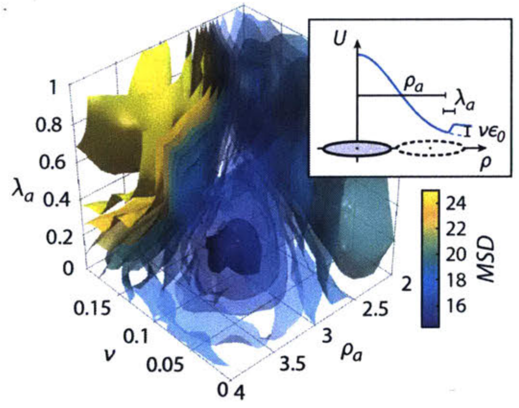


Figure 4-8: Parameter scan to test the influence of RbmA-mediated cell-cell attraction on biofilm phenotype. MSD values for feature vectors of biofilms with arabinose-inducible *rbmA* expression grown at 0.5% arabinose and simulations upon variation of cell-cell interaction strength ν, and well width λ_a and well position ρ_a. Inset: effect of parameter variation on the cell-cell interaction potential. Figure from [61].

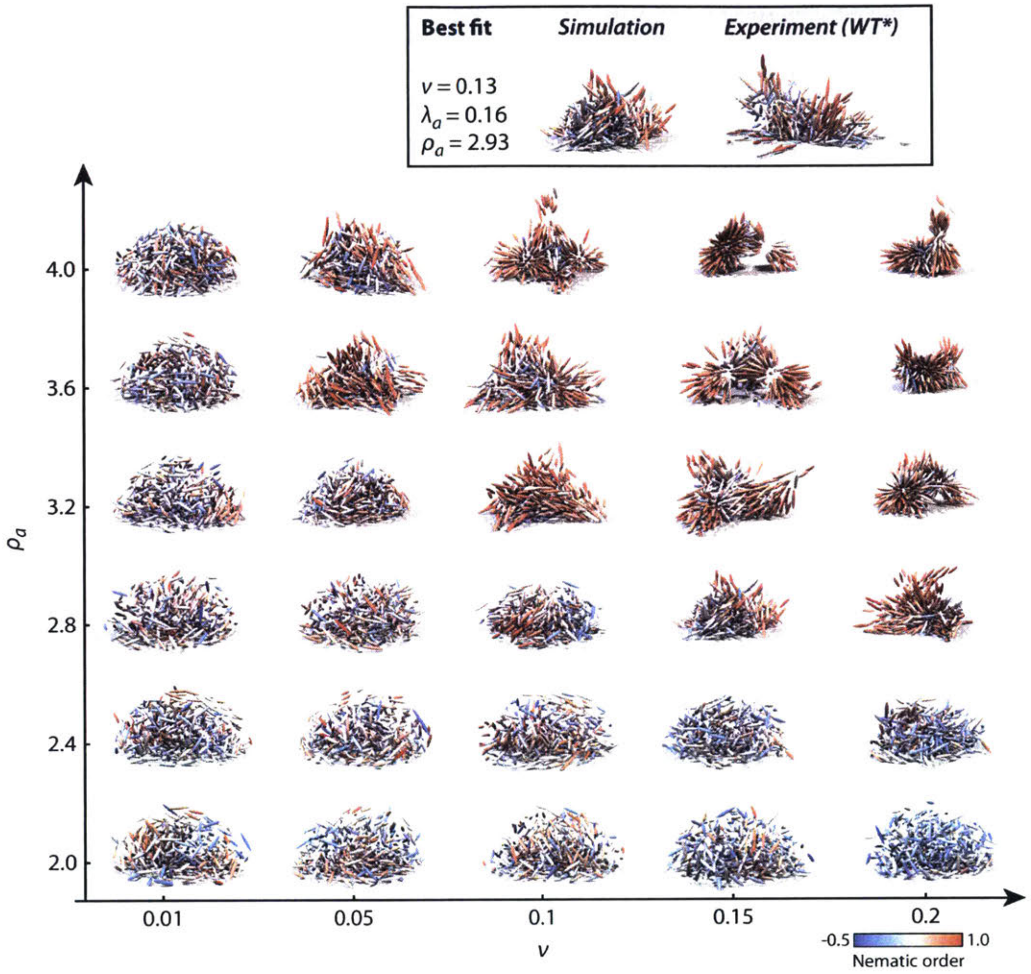


Figure 4-9: Resulting simulated biofilms for different levels of RbmA-mediated cell-cell attraction. The figure shows biofilm renderings corresponding to selected values of ν and ρ_a while keeping λ_a constant at $\lambda_a = 0.16$. Each cell is colored according to the nematic order. Figure from [61].

to be approximately 0.3 μm for all biofilms; therefore, λ_a is expected to be around 0.4 (again using $\sigma = 0.7$ μm).

The attractive potential component (ν, ρ_a, λ_a) are fitted for a range of different arabinose concentrations through comparing the experiments with simulations conducted with a systematic variation of ν, ρ_a, and λ_a using the MSD as a metric. Figure 4-8 shows the MSD values of the scan for the 0.5% (wt/vol), with the corresponding biofilm phenotypes shown in Fig. 4-9. As ν increases at constant ρ_a, the simulated biofilms become more compact, verifying that as the attractive strength increases so does the cell-cell attraction. Similarly, as ρ_a increases at constant ν, the cell-cell distance decreases in the simulated biofilms, indicating an increase in the cell-cell attraction force. As the position of the attractive well increases, the attractive force increases because the repulsive force decreases with increasing distance. For $\nu \rightarrow 0.2$ and $\rho_a \rightarrow 4.0$, the simulated biofilms form star like structures, indicating that the attraction force is too strong compared to the repulsive force in the cell-cell potential. The mean of the corresponding ν, ρ_a, and λ_a of the 5% best matching simulations with the lowest MSD values for each arabinose concentration are defined as the best fit (Fig. 4-10). The best-fit parameters for the experimentally observed WT*-phenotype are $\nu = 0.13$, $\lambda_a = 0.16$, and $\rho_a = 2.93$. A summary of all relevant simulation parameters is given in Table 4.4.

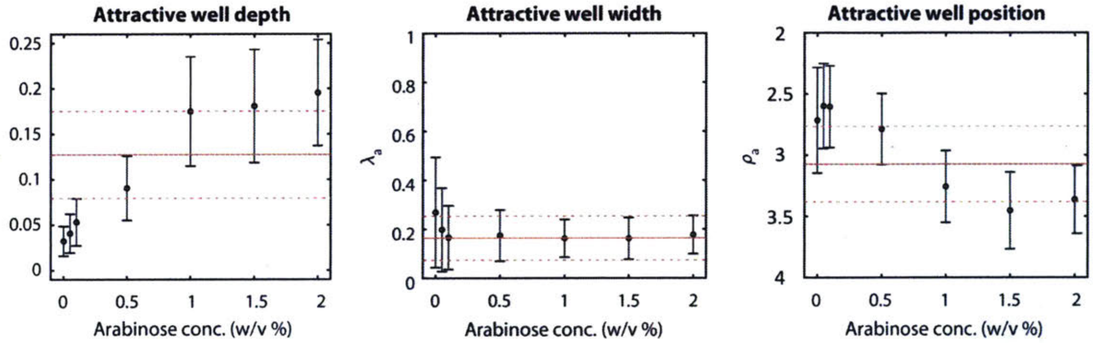


Figure 4-10: Fitting of the attractive potential parameters. The graphs show the approximate position of the MSD minimum in (ν, ρ_a, and λ_a)-space of the 5% best-matching simulations at increasing arabinose concentrations, corresponding to increasing amounts of secreted RbmA. Each data point refers to the median position of the minimum and the error bars indicate the lower and upper quartile. The position of the WT* is shown in red (solid lines: median values, dashed lines: lower/upper quartiles). Figure adapted from [61].

Parameter	Value	Description
d	0.2775 μm	Average half-width of the bacteria from experimental measurements.
τ_g	6130 s	Growth time constant (average cell division time of biofilm-associated cells obtained from experiments).
μ_m	1 Pa·s	Estimate of the dynamic viscosity of the EPS matrix at room temperature [102].
γ_m	5.23×10^{-6} kg/s	Typical drag coefficient for Stokes' drag in EPS matrix ($\gamma_m = 6\pi\mu_m d$).
ϵ_r	10	Ratio comparing the strength of the bacteria-boundary interaction to the strength of the bacteria-bacteria interaction $\epsilon_r = \epsilon_{\text{bdy}}/\epsilon_0$
σ^*_{bdy}	1	Non-dimensional boundary potential length scale parameter.
ℓ^*_{add}	3.65	Non-dimensional length added to bacteria after division.
$\sigma^*_{\text{birth size}}$	0.07	Non-dimensional standard deviation of the normal distribution for the daughter bacteria birth size.
κ	100	Concentration parameter for the von Mises-Fisher distribution for the daughter bacteria division orientation.
τ_t	8.05 s	Translational time scale due to repulsion in matrix (typical time needed for daughter cells in matrix to reach their equilibrium configurations due to repulsion after cell division).
ϵ_0	5×10^{-20} J	Strength of the repulsive part of the cell-cell potential.
λ_r	1.65	Width of the repulsive part of the cell-cell potential (corresponds to 1.16 μm at a typical overlap factor of $\sigma = 0.7$ μm, which is the value it would take for a sphere with the typical mean cell volume of 0.4 μm^3).
ν	0.13 (WT*)	Strength of the attractive part of the cell-cell potential (corresponds to $0.65 \cdot 10^{-20}$ J).
λ_a	0.16 (WT*)	Well-width of the attractive part of the cell-cell potential (corresponds to 0.11 μm at a typical overlap factor of $\sigma = 0.7$ μm).
ρ_a	2.93 (WT*)	Position of the attractive part of the cell-cell potential (corresponds to 2.0 μm at a typical overlap factor of $\sigma = 0.7$ μm).

Table 4.4: Key simulation parameters. The parameters $\epsilon_0, \lambda_r, \nu, \lambda_a, \rho_a$, are determined by fitting as described above. Table adapted from [61].

4.3 Results

For two aligned cells, the cell-cell interaction range for the best-fit U for the $\Delta rbmA$ (Fig. 4-11) and WT* (Fig. 4-12) biofilms acts within the range of the experimentally determined average cell-cell distance (dashed cell). An effective arabinose concentration for the WT* is inferred by locating the WT* biofilm architecture in the (ν, λ_a, ρ_a) space along the curve of different arabinose concentrations [Fig. 4-13(a)]. The simulations based on the WT* parameters for biofilms up to 300 cells show good qualitative [Fig. 4-13(b)] and quantitative [Fig. 4-13(c)] agreement with experiments. Predictions of the architectural development for larger biofilms ($N_{\text{cells}} > 300$) show high quantitative and qualitative agreement with experimental data, for both the WT* and the $\Delta rbmA$ biofilms up to 10^3 cells (Fig. 4-14).

4.3.1 Cell-cell translational forces and rotational dynamics

Cell-cell interaction inside $\Delta rbmA$ mutant biofilms lacking cell-cell adhesion are modeled by a repulsive interaction potential, for which the best-fit U results in the translational forces and rotational dynamics shown in Fig. 4-11. As shown in Fig. 4-11(a), U (with best-fit parameters $\epsilon_0 = 5 \times 10^{-20}$J, $\lambda_r = 1.65$, and $\nu = 0$) results in a repulsive (blue) region of translational forces around the cell, capturing the combined effects of hard steric and osmotic repulsion. With cell α (solid cell) held in place, the best-fit repulsive U results in the rotation of cell β (dashed cell) such that contact between the cells is minimized [Fig. 4-11(b)]. The interaction range for two aligned cells is very close to the experimentally observed average cell-cell spacing of the $\Delta rbmA$ mutant (dashed cells).

For biofilms with cell-cell adhesion, cell-cell interactions are modeled with an attractive and repulsive interaction potential, for which the best-fit U of a particular

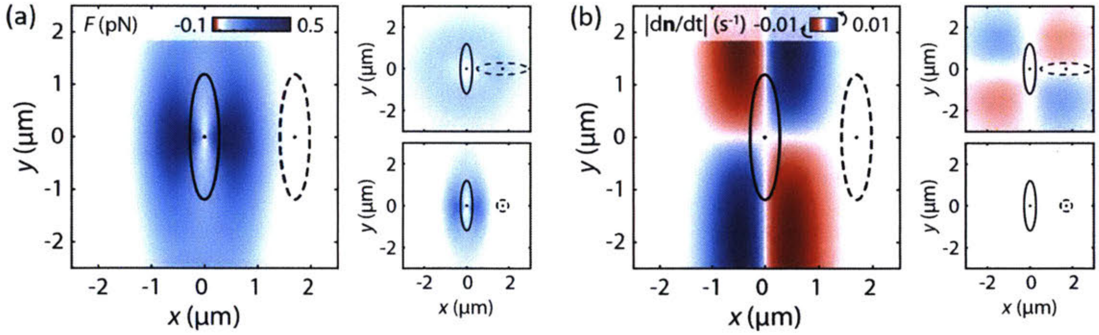


Figure 4-11: Cell-cell interactions of various cell-cell orientations for the $\Delta rbmA$ biofilm, which are modeled by a repulsive interaction potential. The best-fit U for these biofilms has interaction parameters $\epsilon_0 = 10\epsilon_{\text{flow}}(5 \times 10^{-20}$J, $\lambda_r = 1.65$, and $\nu = 0$. (a) Translational cell-cell interaction forces (red, attractive; blue, repulsive). (b) Rotational cell-cell interaction dynamics of cell β (dashed cell), $|d\hat{\mathbf{n}}/dt| = |(\mathbf{I} - \hat{\mathbf{n}}_\beta \hat{\mathbf{n}}_\beta^\top)(-\boldsymbol{\Omega}\, \partial U / \partial \hat{\mathbf{n}}_\beta)|$, in the case where cell α (solid cell) is held in place (red, clockwise rotation; blue, anticlockwise rotation). The dashed cells are plotted at the average cell-cell distance obtained from the $\Delta rbmA$ experiments. Figure adapted from [61].

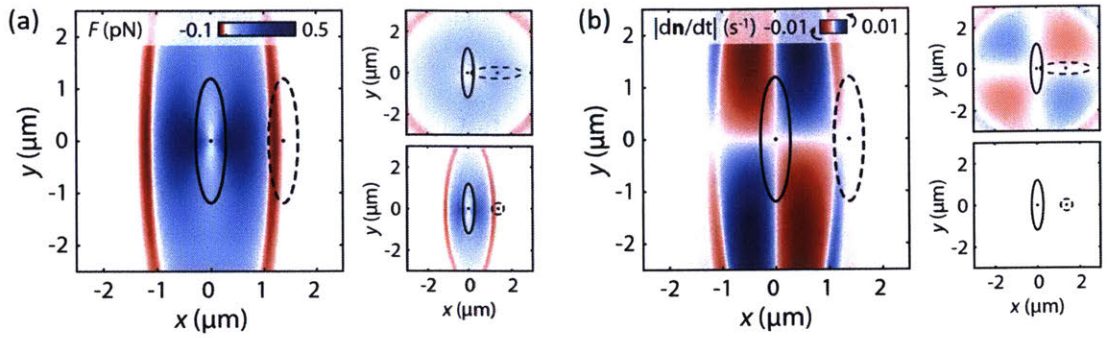


Figure 4-12: Cell-cell interactions of various cell-cell orientations for a particular level of cell-cell adhesion (0.5% arabinose), which are modeled by an attractive and repulsive interaction potential. The best-fit U for these biofilms has interaction parameters $\epsilon_0 = 10\epsilon_{\text{flow}}(5 \times 10^{-20}\text{J}$, $\lambda_r = 1.65$, $\nu = 0.1$, $\lambda_a = 0.16$, and $\rho_a = 2.75$. (a) Translational cell-cell interaction forces (red, attractive; blue, repulsive). (b) Rotational cell-cell interaction dynamics of cell β (dashed cell), $|d\hat{\mathbf{n}}/dt| = |(\mathbf{I} - \hat{\mathbf{n}}_\beta\hat{\mathbf{n}}_\beta^\top)(-\boldsymbol{\Omega}\,\partial U/\partial\hat{\mathbf{n}}_\beta)|$, in the case where cell α (solid cell) is held in place (red, clockwise rotation; blue, anticlockwise rotation). RbmA-mediated cell-cell adhesion gives rise to an attractive part (red), acting within the range of experimentally determined average cell-cell distance (dashed cell). Figure adapted from [61].

level of cell-cell adhesion (0.5% arabinose) results in the translational forces and rotational dyanmics shown in Fig. 4-12. U (with best-fit parameters $\epsilon_0 = 5 \times 10^{-20}$J, $\lambda_r = 1.65$, $\nu = 0.1$, $\lambda_a = 0.16$, and $\rho_a = 2.75$) results in translational forces around the cell with a repulsive (blue) region, capturing the combined effects of hard steric and osmotic repulsion, and an attractive (red) region, capturing the RbmA-mediated cell-cell adhesion [Fig. 4-12(a)]. With cell α (solid cell) held in place, the best-fit U results in a repulsive region where cell β (dashed cell) rotates away from cell α such that the contact between the cells is minimized and an attractive region where cell β rotates toward cell α [Fig. 4-12(b)]. The attractive part of the best-fit U acts within the range of the experimentally determined average cell-cell distance (dashed cell) for aligned cells.

4.3.2 Fitting the WT*

With the calibrated simulation, an effective arabinose concentration for the WT* of $c = 0.68 \pm 0.19\%$ (wt/vol) is inferred by locating the WT* biofilm architecture in the (ν, λ_a, ρ_a) space along the curve of different arabinose concentrations [Fig. 4-13(a)]. Extracting an effective arabinose concentration and RbmA level for the WT* is based on the simplifying assumption that all cells in the biofilms express the same levels of the key matrix components, which represents a minimal model that is in quantitative agreement with the experimental data, as the best-fit (ν, λ_a, ρ_a) values for the WT* are close to the effective (ν, λ_a, ρ_a) values for the WT* on the curve of different arabinose concentrations [Fig. 4-13(a)]. The simulations based on the WT* parameters for biofilms up to 300 cells show good quantitative agreement with experiments [Fig. 4-13(c)]. Remarkably, these simulations also show architechtural properties that were not included in the feature vector used for MSD minimization, such as local density variations and the occurrence of patches of highly aligned cells inside the biofilm (red cells in Fig. 4-13(b), characterized by high local ordering), which are characteristic for biofilms with high concentrations of RbmA.

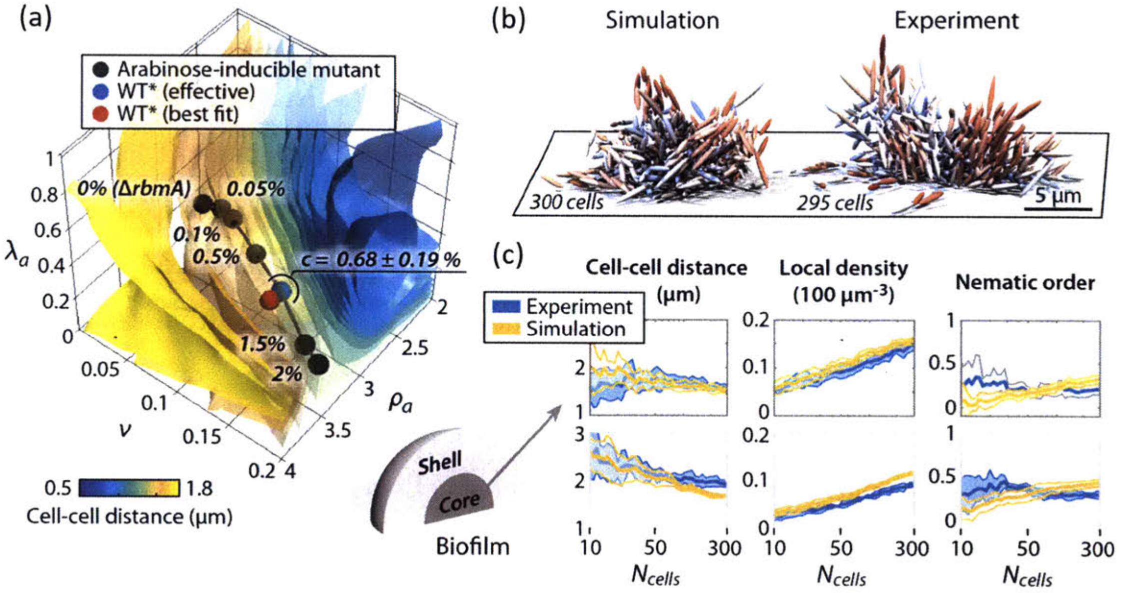


Figure 4-13: (a) Best-fit simulation parameters for varying RbmA and arabinose concentrations (black dots) follow a line in (ν, λ_a, ρ_a) parameter space and cross isosurfaces of average cell-cell distance (see color bar and compare with Fig. 4-7; see Fig. 4-10 for more details about the fitting). The RbmA level of the WT* biofilms is inferred in terms of an effective arabinose concentration by locating the WT* along the line of different arabinose concentrations (blue point), which is very close to the best fit of the WT* (red point). (b) Simulated (best fit) versus experimental WT* biofilm. (c) Comparison of biofilm architectural properties for the WT* experiment (blue) and the WT* simulation prediction (yellow). The architectural properties are spatially resolved for the core (top row) and shell (bottom row) of the biofilm (experiment, $n = 7$; simulation, $n = 10$). Figure adapted from [61].

4.3.3 Prediction of larger biofilms

The cell-cell interaction potential is calibrated based on biofilms growing to 300 cells (Sec. 4.2.2). To test the ability of the calibrated simulations to predict the structure of biofilms with more than 300 cells, a feature vector containing the same parameters as the ones used in Sec. 4.2.2 is used. The MSD between the experimental and simulation feature vectors is calculated for biofilms sampled at 40 intermediate logarithmically spaced cell numbers from 10 to 300 and from 300 to 1000, separately for each chosen number of cells (in contrast to Sec. 4.2.2 where a single MSD between temporal feature vectors is calculated), and is normalized by the average value of the MSD between 10 and 300 cells. Predictions of the architectural development for larger biofilms ($N_{\text{cells}} > 300$) show high quantitative and qualitative agreement with experimental data, for both the WT* and the $\Delta rbmA$ biofilms up to 10^3 cells (Fig. 4-14). These results indicate the remarkable ability of the potential-based simulations calibrated on biofilm development up to 300 cells to predict the development of larger biofilms. To achieve accurate simulation results for very large biofilms ($> 10^3$ cells), spatiotemporal heterogeneity in gene expression, matrix composition, and growth rates probably have to be included in future simulations. The combined experimental and theoretical

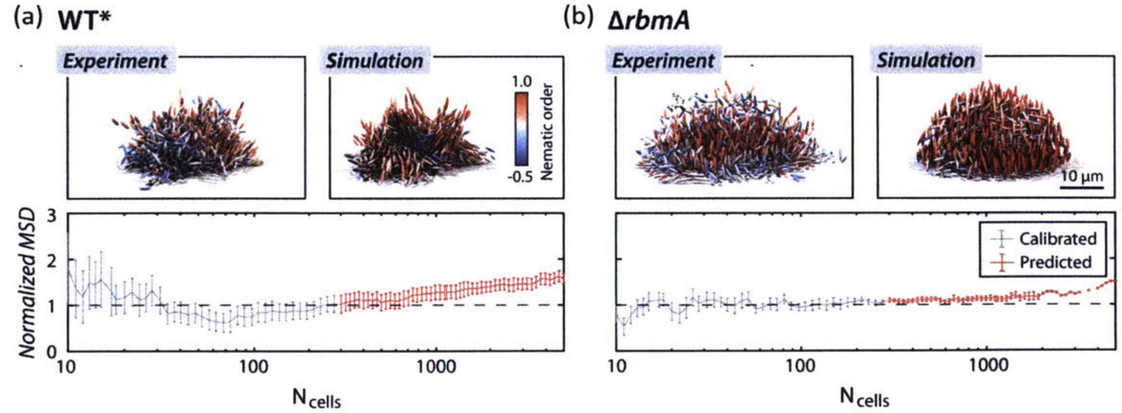


Figure 4-14: Predicting phenotypes of larger biofilms. (a) Rendered WT* experimental (left) and simulated biofilm (right) for approximately 1000 cells. Bottom: normalized MSD of feature vectors for experimental ($n = 7$) and simulated biofilms ($n = 3$). (b) Rendered $\Delta rbmA$ experimental (left) and simulated biofilm (right) for approximately 1000 cells. Bottom: normalized MSD of feature vectors for experimental ($n = 4$) and simulated biofilms ($n = 3$). Error bars are calculated by Gaussian error propagation of the standard deviations of the individual measurements. Figure adapted from [61].

analysis therefore suggests that mechanical interactions between cells suffice to account for the internal cellular order and architecture up to mid-size *V. cholerae* biofilms.

Chapter 5

Bacterial surface accumulation

In the vicinity of surfaces, the behavior of swimming bacteria can change dramatically.[1] In contrast to their approximately straight-line locomotion in bulk fluids, non-tumbling flagellated bacteria typically follow circular trajectories near surfaces, often for an extended period of time [49]. Furthermore, several wild-type peritrichous bacterial strains have been found to exhibit longer run times and smaller mean tumbling angles at the surface compared to their bulk run and tumbles [95]. Exploiting cell-surface interactions, recent studies demonstrated that bacteria can be concentrated by funnel walls [51], drive asymmetric microgears [30, 121], and self-organize into collective vortices [98, 142]. A well-known consequence of bacteria-surface interactions is the accumulation of cells near solid surfaces: Local concentration values for both non-tumbling and tumbling strains near a flat surface can exceed the corresponding bulk concentrations by a factor of 5 or more [7, 83, 84, 95]. Such accumulation increases the possibility cell-surface attachment, facilitating undesirable secondary effects like biofouling and biofilm formation [22, 136].

Over the past two decades, much progress has been made in designing antifouling surfaces based on chemical surface modification [2, 97, 110]. Common surface treatments include released-based coatings in which a biocidal agent (e.g. silver ions, antibiotics, or quaternary ammonium compounds) is released into the environment, hydrophilic polymer coatings, and self-assembled monolayers. However, the antifouling properties of these surface treatments are often temporary because of the depletion of the biocidal substance within the coating or the masking of the coating's chemical functionality by the absorption of biomolecules from the surrounding environment [2, 97, 110]. Further, chemical surface treatments can leach into and have toxic effects on the local ecosystem and have led to the rise of antibiotic- and silver-resistant bacterial strains [2, 110]. Thus, chemical surface modifications alone are unlikely to provide long-term solutions to the antifouling problem. An interesting alternative approach, inspired by the *Nepenthes* pitcher plant, utilizes a lubricant-infused coating that results in a slippery surface [42, 143]. Another nontoxic, persistent solution may be the manipulation of the surface topology to deter bacterial adhesion. Important previous studies of bacterial adhesion on various 2D polydimethylsiloxane (PDMS) patterned surfaces have explored nanoscale tall spatially organized designs [104], a shark skin inspired micrometer high diamond pattern [16], and a nested hierarchically wrinkled surface topography with length scales spanning from tens of nanometers to a fraction of a millimeter [40]. Yet, many aspects of the interplay between complex surface geometries and cell accumulation are not yet well understood.

---

[1] The results of Chapter 5 appear in preprint: Rachel Mok, Jörn Dunkel, and Vasily Kantsler. Geometric control of bacterial surface accumulation. *arXiv preprint arXiv:1903.00926*, 2019.

To contribute to a more detailed understanding, experiments (Sec. 5.1) and simulations (Sec. 5.2) are performed to investigate the effects of locally varying boundary curvature on the scattering and accumulation dynamics of swimming *Escherichia coli* bacteria in quasi-2D microfluidic chambers (Fig. 5-1). To explore the effects of partially convex and concave boundary geometries on the spatial cell distributions, the experiments are complemented with simulations of 2D particle-based models for both Brownian Dynamics (BD) and Run and Tumble (RT) dynamics. The analysis confirms that a minimal steric interaction model [83, 84] suffices to account for the main aspects of the experimental data (Sec. 5.3). Both experimentally observed and simulated cell trajectories illustrate that the non-convex boundary features redirect the bacteria away from the surfaces (Fig. 5-1). Throughout, data from experiments and simulations are analyzed using the same algorithms to compare the observed and predicted surface accumulation (Fig. 5-2). Scanning a range of geometric surface parameters, an optimal curvature is determined that minimizes the bacterial accumulation for a sinusoidal boundary geometry (Fig. 5-3). Furthermore, a strictly concave periodic base geometry is shown to decrease the average cell accumulation near the boundary by more than 50% relative to a flat surface (Figs. 5-4 and 5-5).

5.1 Experiments

Thin microfluidic chambers (4 mm long, 2 mm wide, 3-4 μm thick) were produced by standard soft lithography technique from PDMS (Dow Corning), resulting in a quasi-2D environment.[2] For the sinusoidal geometries [Fig. 5-1(b)], the top and bottom boundaries of each chamber are designed as $\pm A \sin(2\pi x/\lambda)$, and 20 different parameter combinations are investigated with amplitudes $A = [1.75, 3.5, 5.25, 7]$ μm and wavelengths $\lambda = [21, 28, 35, 42, 49]$ μm. The boundaries for the concave semicircle geometry [Fig. 5-1(c)] are designed with radius $R = 12$ μm. To ensure the cell dynamics and statistics are not biased by reflections from the opposing boundary, a large boundary separation distance of 2 mm is chosen. After a 40 s exposure to oxygen plasma (Harrick Plasma, PDC-002) the PDMS chambers were bonded to the glass coverslips initially cleaned in hydrogen peroxide. Dilute suspensions (approximately 10^8 cells/mL) of non-chemotactic *E. coli* cells (strain HCB1733, provided by Howard C. Berg) carrying the pYFP plasmid (Clontech, BD Biosciences), which were washed with 0.1% bovine serum albumin to prevent bacterial adhesion, are loaded into the microfluidic chambers. The device inlets are sealed with unpolymerised PDMS to avoid background fluid flow. The bacteria motion is measured using a Nikon TE2000U inverted microscope with a 40x oil immersion objective (NA 1.3) at 10 frames per second (Evolve Delta, Photometrics) or LSM 510 Zeiss Axiovert 200 M at 3 frames per second (fps). Single-cell trajectory data are reconstructed using a custom Matlab particle tracking script.

---

[2] The experiments and data analysis were performed by Vasily Kantsler.

5.2 Simulations

To test whether steric surface collisions can account for the experimentally observed cell distributions, 2D particle-based simulations are performed. Focusing on minimal models, hydrodynamic effects [83, 84] are neglected because the small chamber thickness in the experiments strongly suppresses hydrodynamic flows. Similarly, steric cell-cell interactions can be ignored as only dilute bacterial suspensions are considered throughout. The bacteria are modeled as non-interacting ellipsoids of half-length ℓ and half-width d, described by their position $\mathbf{x}(t)$ and orientation $\hat{\mathbf{n}}(t)$. Cells are assumed to move at a constant self-propulsion speed v in the direction of their orientation $\hat{\mathbf{n}}$. An effective steric boundary potential U_{bdy} is used to encode bacterial surface interactions across various geometries. Bacteria in the experiments display occasional stochastic reorientation as they swim [35]. To account for this, simulations for both BD and RT reorientation are performed and compared. In the BD model, bacteria are reoriented through Gaussian rotational noise. In the RT model, a cell moves deterministically for a fixed period of time (run stage) before undergoing a stochastic reorientation event (tumble stage).

5.2.1 Brownian Dynamics (BD)

Denoting the d-dimensional unit matrix by $\mathbf{I}$, the over-damped Langevin equations for a single bacterium with position $\mathbf{x}(t)$ and orientation $\hat{\mathbf{n}}(t)$ in the BD model are

$$\mathrm{d}\mathbf{x} = (v\hat{\mathbf{n}} - \boldsymbol{\Gamma}^{-1}\nabla_{\mathbf{x}}U_{\text{bdy}})\mathrm{d}t \tag{5.1a}$$

$$\mathrm{d}\hat{\mathbf{n}} = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\left((1-\mathsf{d})D_R\hat{\mathbf{n}} - \boldsymbol{\Omega}^{-1}\nabla_{\hat{\mathbf{n}}}U_{\text{bdy}}\right)\mathrm{d}t + \sqrt{2D_R}(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\cdot\mathrm{d}\mathbf{Z} \tag{5.1b}$$

Here, v is the self-swimming speed, D_R the rotational diffusion coefficient, and $\mathbf{Z}$ is a d-dimensional Gaussian random variable of zero mean and variance $\mathrm{d}t$. Note, use has been made of the fact that $\frac{\gamma_{\perp}}{\gamma_0\gamma_{\parallel}} \sim 1$ [compare Eq. 2.56]. The boundary potential U_{bdy}, Eq. (2.7), used for the cell-surface interactions is described in Sec. 2.2. The friction tensor

$$\mathbf{\Gamma} = \gamma_0 \left[\gamma_{\parallel}(\hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) + \gamma_{\perp}(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\right] \tag{5.2}$$

accounts for the fact that the bacteria experience more drag when moving perpendicular to their orientation. Rotational drag is approximated as isotropic,

$$\mathbf{\Omega} = \omega_0 \gamma_R \mathbf{I}. \tag{5.3}$$

γ_0 and $\omega_0 = k_B T / D_R$ are the Stokesian translational and rotational friction coefficients, respectively. k_B is the Boltzmann constant, and T is the temperature. $\gamma_{\parallel}$, $\gamma_{\perp}$, and γ_R are dimensionless geometric parameters characterizing the longitudinal, transverse, and rotational friction parameters of elongated particles that depend only on the aspect ratio $a = \ell / d$. The expressions given in [130] for rod-like macromolecules Eqs. (2.62) are used.

Adopting cell length ℓ and $\tau = \ell / v$ as characteristic length and time scales and defining the following Péclet numbers $P_T \equiv v \ell \gamma_0 / k_B T$ and $P_R \equiv v / D_R \ell$, Eq. (5.1) is recast in nondimensional form. Denoting dimensionless quantities with a superscript * results in

$$\mathrm{d}\mathbf{x}^* = \left(\hat{\mathbf{n}} - \frac{\epsilon_{\mathrm{bdy}}}{k_B T}\frac{1}{P_T}\left[\frac{1}{\gamma_{\parallel}}(\hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) + \frac{1}{\gamma_{\perp}}(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\right] \nabla_{\mathbf{x}^*} U^*_{\mathrm{bdy}}\right) \mathrm{d}t^* \tag{5.4a}$$

$$\mathrm{d}\hat{\mathbf{n}} = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) \left((1 - \mathrm{d})\frac{1}{P_R}\hat{\mathbf{n}} - \frac{\epsilon_{\mathrm{bdy}}}{k_B T}\frac{1}{P_R}\frac{1}{\gamma_R}\nabla_{\hat{\mathbf{n}}} U^*_{\mathrm{bdy}}\right) \mathrm{d}t^* + \sqrt{\frac{2}{P_R}\mathrm{d}t^*}(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) \cdot \mathrm{d}\mathbf{Z} \tag{5.4b}$$

where ϵ_{bdy} characterizes the strength of the bacteria-boundary potential interaction.

5.2.2 Run and Tumble (RT)

During the run stage of the RT model, which lasts a duration τ_{run}, the deterministic motion of a cell is governed by Eq. (5.1) with $D_R = 0$. Rescaling to a dimensionless form using the same characteristic length and time scales as before, the run motion is described by

$$\mathrm{d}\mathbf{x}^* = \left(\hat{\mathbf{n}} - \frac{\epsilon_{\text{bdy}}}{k_B T} \frac{1}{P_T} \left[\frac{1}{\gamma_{\parallel}}(\hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) + \frac{1}{\gamma_{\perp}}(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top})\right] \nabla_{\mathbf{x}^*} U^*_{\text{bdy}}\right) \mathrm{d}t^* \tag{5.5a}$$

$$\mathrm{d}\hat{\mathbf{n}} = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\top}) \left(-\frac{\epsilon_{\text{bdy}}}{k_B T} \frac{1}{P_R} \frac{1}{\gamma_R} \nabla_{\hat{\mathbf{n}}} U^*_{\text{bdy}}\right) \mathrm{d}t^* \tag{5.5b}$$

where the rotational Péclet number due to tumbling is now determined as follows: At the end of τ_{run}, the bacterium undergoes a tumbling event. Let θ be the angle between the previous orientation and new orientation after a tumble. θ is drawn from a von Mises-Fisher (vMF) distribution with the mean angle equal to the original bacterial orientation and concentration parameter κ (see Appendix A for more details about the vMF distrbution). To relate κ to experimental values, note that for weakly tumbling cells $\kappa \gg 1$. In this case, the mean squared angular change per tumble is $\langle \theta^2 \rangle = \tilde{D}_R \tau_{\text{run}} \simeq 1/\kappa$, yielding the effective rotational Péclet number $P_R = v/(\tilde{D}_R \ell)$.

5.2.3 Implementation

Mirror-symmetric confinements parallel to the $y = 0$ line are considered, with the surfaces defined as $s_{y\pm} = f_{\pm}(s_x)$ with $f_-(s_x) = -f_+(s_x)$. $\mathbf{S} = (s_x, s_y)$ denotes a point on the surface. The distance, $\tilde{d}$, of a bacterium at position $\mathbf{x} = (x, y)$ from a surface f is given by the function

$$\tilde{d}(s_x) = \frac{1}{2}\left[(x - s_x)^2 + (y - f(s_x))^2\right] \tag{5.6}$$

where the numerical prefactor 1/2 was chosen for convenience. To find the point on the surface closest to the bacterium, the following equation

$$\frac{\partial \tilde{d}}{\partial s_x} = s_x - x + (s_y - y)\frac{\partial s_y}{\partial s_x} = 0 \tag{5.7}$$

is solved numerically with the bisection method. The second derivative is used to confirm that the surface point found results in a minimum distance. The boundary surface equations for the flat, sinusoidal, and semicircle surfaces used in the simulations are

$$s_{y\pm} = \pm C \tag{5.8}$$

$$s_{y\pm} = \pm A \sin\left(\frac{2\pi}{\lambda} s_x\right) \pm C \tag{5.9}$$

$$s_{y\pm} = \pm\sqrt{R^2 - \left\{\frac{2R}{\pi}\cos^{-1}\left[\cos\left(\frac{\pi}{2R}s_x\right)\right] - R\right\}^2} \pm C \tag{5.10}$$

$C = 1000$ μm is the displacement from the $y = 0$ line for all geometries. Because there is a discontinuity in the derivative of the semicircle geometry at the peaks ($s_x = 2Rn$ for $n = 0, 1, 2, ...$), the boundary potential is neglected for a region of scale $\sim r$ at the peaks for bacteria that are not vertical and the peaks are treated as a flat boundary for bacteria that are vertical to prevent the cells from penetrating the surface.

In the simulations, new positions and orientations of the bacteria are obtained from solving the dimensionless over-damped translation and orientation equations for the BD and RT models, Eqs. (5.4) and (5.5) at each time step. Cells are initially loaded uniformly within the computational domain with random orientations and with random start run times for the RT model. Periodic boundary conditions are applied in the x-direction. Measurements are taken after the simulations had relaxed to a statistical steady-state with constant $\langle y^2 \rangle$.

5.2.4 Parameters

The bacteria are modelled as 1 μm in width and 7 μm in length, accounting for the flagella in addition to cell body length. It is known that *E. coli* move at a speed of approximately 20 μm/s [35], and the run time is typically 1 s [5, 132]. Simulation scans were performed to find ϵ_{bdy} and D_R that resulted in surface accumulations that best matched with the experiments for the sinusoidal surface. For the BD model, we found $\epsilon_{\text{bdy}} = 175$ k_BT and $D_R = 0.08$ rad$^2/s$, and for the RT model $\epsilon_{\text{bdy}} = 1500$ k_BT and $\tilde{D}_R = 0.1$ rad$^2/s$. The fitted near-surface D_R values for both models are of the same order of magnitude as the measured bulk $D_R = 0.057$ rad$^2/s$ for non-tumbling *E. coli* [35]. For both models $\epsilon_{\text{bdy}} \gg k_BT$, indicating that the boundary potential is highly repulsive. The large ϵ_{bdy} is required to prevent the bacteria from penetrating the boundary as the models do not account for the reduction in swimming speed as the cells approach the surface. The fitted ϵ_{bdy} and D_R values obtained for the sinusoidal surface are also used for simulations of the flat and semicircle surfaces. A summary of all relevant simulation parameters is given in Table 5.1.

Parameter	BD	RT	Description
ℓ	3.5 μm	3.5 μm	Bacteria half-length
d	0.5 μm	0.5 μm	Bacteria half-width
v	20 μm/s	20 μm/s	Self-propulsion speed
ϵ_{bdy}	175 k_BT	1500 k_BT	Boundary potential strength parameter
σ_{bdy}	0.5 μm	0.5 μm	Boundary potential scale parameter
$D_R, \tilde{D}_R$	0.08 rad$^2/s$	0.1 rad$^2/s$	Rotational diffusion coefficient
P_T	0.0014	0.0014	Translational Péclet number
τ_{run}	–	1 s	Run time
κ	–	10	Concentration parameter for von Mises-Fisher distribution

Table 5.1: Summary of bacterial surface accumulation simulation parameters. Table from [94].

5.3 Results

For the experiments and simulations, the cell trajectories (Fig. 5-1) of the flat, sinusoidal, and semicircle geometries are segmented (Fig. 5-2) to quantify the bacterial accumulation. To identify optimal sinusoidal surfaces for the reduction of bacterial surface accumulation, a scan is performed over a range of amplitudes and wavelengths (Fig. 5-3). The cell distributions of the optimal sinusoidal and semicircle surfaces are quantified and compared, with the semicircle geometry proving to be the most efficient at reducing bacterial accumulation (Figs. 5-4 and 5-5).

5.3.1 Tracking data

The simulated particle trajectories agree well with the experimental cell trajectories (Fig. 5-1). In the flat geometry, bacteria collide and align with the surface [83, 84], contributing to surface accumulation in the experiments and simulations. Comparing

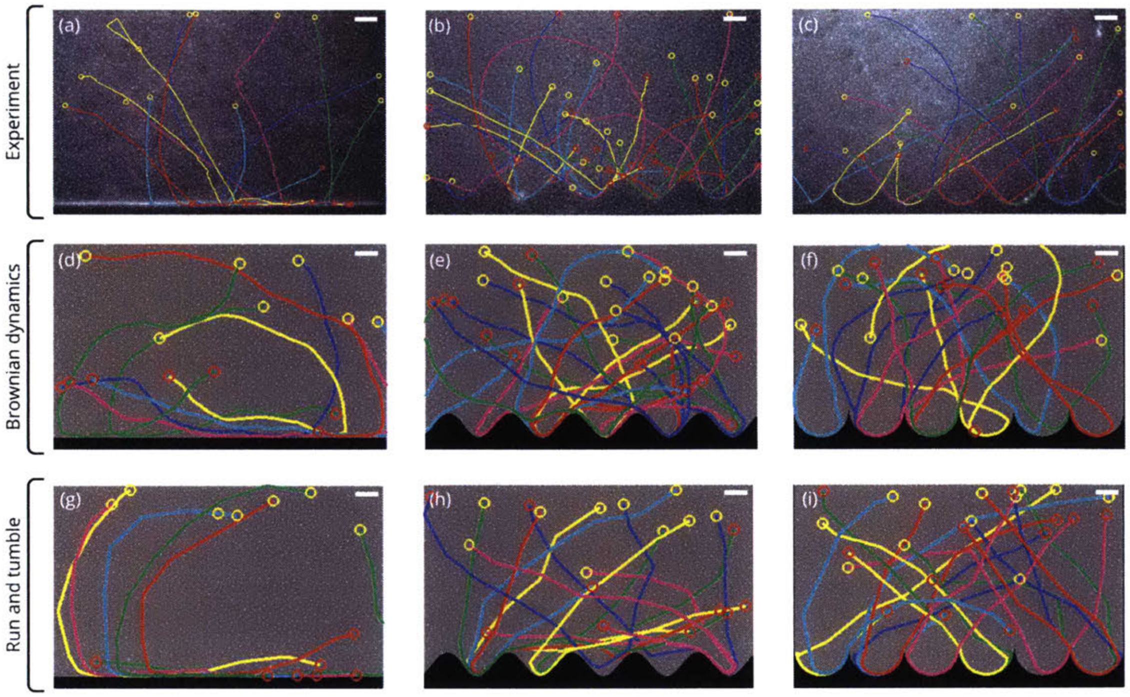


Figure 5-1: Typical trajectories of swimming cells for flat, sinusoidal, and semicircle surface geometries as observed in experiment and simulations. The start and end of each trajectory are indicated by the yellow and red circle, respectively. Each trajectory is 10 s long. Bacteria align with the surface in the flat geometry leading to significant surface accumulation for the experiment and simulation models. The sinusoidal ($A = 5.25$ μm, $\lambda = 28$ μm) and concave semicircle ($R = 12$ μm) surface geometries redirect the bacteria away from the surface in the experiments and simulations. Scale bars 10 μm. Figure from [94].

Figs. 5-1(d) and 5-1(g), we note bacterial residence time at the surface appears shorter in the BD model than the RT model due to the orientation noise. Because of the nonconvex features present in both the sinusoidal and semicircle geometries, the bacteria are redirected away from the surface in both the experiment and simulations, leading to a reduction in surface accumulation. The segmented raw data is normalized to the same frame rate for both the experiment and simulations (Fig. 5-2). Thus, the same analysis algorithms are used to examine the experimental and simulation data. The experimental raw data Figs. 5-2(a)—5-2(c) exhibit higher curvature than the simulation raw data, likely caused by hydrodynamic effects from the chamber walls, which are not accounted for in the simulations.

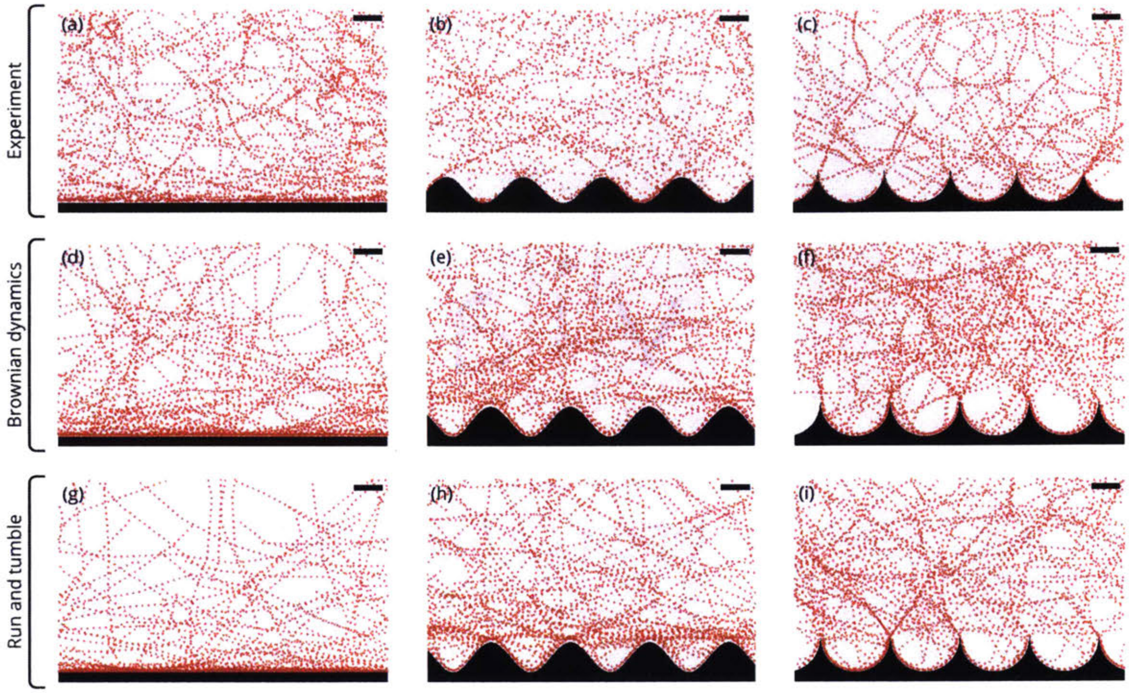


Figure 5-2: Segmented raw data for the experiment and simulations, used in the statistical analysis. The segmented trajectories are acquired at 10 fps. The experimental raw data exhibit higher curvature than the simulation raw data likely due to hydrodynamic effects, which are not accounted for in the simulations. Scale bar 10 μm. Figure from [94].

5.3.2 Optimal sinusoidal boundaries

To determine optimal sinusoidal boundary geometries, a parameter scan over a range of amplitudes A and wavelengths λ is performed. The cell accumulation at the surface is measured in each case [Fig. 5-3(a)]. The bacterial surface concentration is determined from the number of cells between the surface boundary and the boundary contour shifted 5 microns away from the surface (grey regions in Fig. 5-4). The bulk concentration is the number of bacteria in a congruent area 50 μm away from the boundary (blue regions in Fig. 5-4). Accumulation is quantified as the ratio of the surface concentration over the bulk concentration. Figures 5-3(a)—5-3(c) illustrate the resulting mean surface accumulation of the scan. The location and size of the grey circles in Figs. 5-3(a)—5-3(c) designate the 20 combinations (A, λ) and the standard deviation, respectively. In Fig. 5-3(a), the white numbers indicate the number of experiments performed per point. Three simulations were performed for the same

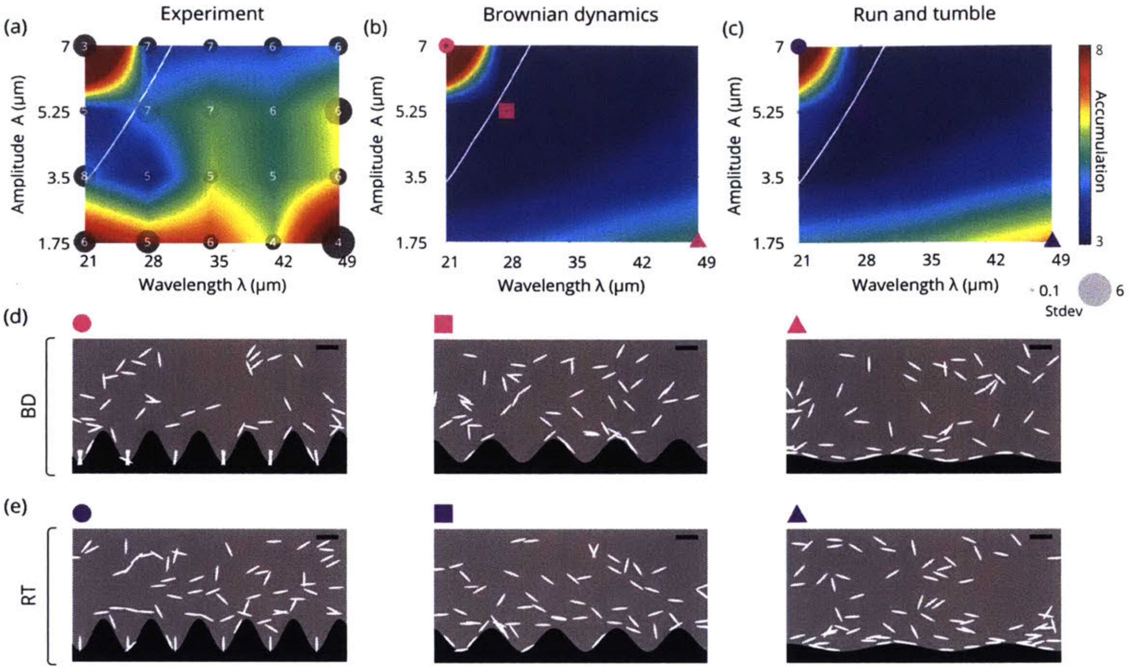


Figure 5-3: Mean bacteria surface accumulation for the sinusoidal surface over a range of amplitudes A and wavelengths λ. Accumulation at the surface is measured by comparing the number of bacteria within 5 μm from the surface to the number of bacteria in the same area 50 μm away from the surface (Fig. 5-4). (a) The location of the circles indicate the 20 combinations (A, λ) of the scan, and the size of the circle represents the standard deviation of each point. The white numbers indicate the number of experiments per point. (b, c) 3 simulations were performed for the same pairs (A, λ) as in the experiments and bilinearly interpolated. The BD and RT simulations agree qualitatively with experiment, revealing an optimum max curvature that reduces accumulation. The set of parameters corresponding to the optimum curvature k^* is delineated by the white curve $A = (k^*/4\pi^2)\lambda^2$ where $k^* = 0.31\ \mu\text{m}^{-1}$. Typical images for the BD and RT simulations are shown in (d) and (e), respectively, for $A = 7\ \mu$m, $\lambda = 21\ \mu$m (circle), $A = 5.25\ \mu$m, $\lambda = 28\ \mu$m (square), and $A = 1.75\ \mu$m, $\lambda = 49\ \mu$m (triangle). Scale bars 10 μm. Figure from [94].

20 combinations (A, λ) as in the experiment for Figs. 5-3(b)—5-3(c).

As evidenced by the mean surface accumulation, both the BD and RT models agree qualitatively with the experiment. Typical still images from the simulations are shown in Figs. 5-3(d)—5-3(e) for $A = 7\ \mu$m, $\lambda = 21\ \mu$m (circle), $A = 5.25\ \mu$m, $\lambda = 28\ \mu$m (square), and $A = 1.75\ \mu$m, $\lambda = 49\ \mu$m (triangle). Due to the steep curvature of the sinusoidal boundary at $A = 7\ \mu$m, $\lambda = 21\ \mu$m, the cells become trapped in the surface pockets, leading to increased accumulation. The BD and especially the RT model can also capture the high accumulation at $A = 1.75\ \mu$m, $\lambda = 49\ \mu$m. Here, the surface is nearly flat and does not deflect the bacteria away from the surface, resulting in high accumulation. Quantitative differences between the experiment and simulations can likely be attributed to hydrodynamic effects. The low-accumulation region in both the experiment and simulations suggests that there exists an optimal curvature for suppressing bacterial accumulation. Characterizing this effect in terms of the maximal local curvature k^* of the sine wave, we find the relation

$$A = (k^*/4\pi^2)\lambda^2 \tag{5.11}$$

After smoothing the experimental values with bilinear interpolation, all the points that are within 15% of the minimum accumulation are fitted to Eq. (5.11), resulting in the optimal maximal curvature $k^* = 0.31\ \mu\text{m}^{-1}$. This is plotted as the white curve in Figs. 5-3(a)—5-3(c).

5.3.3 Sinusoidal vs. concave geometries

Previous work has shown that bacteria can be trapped by convex walls [119]. This suggests that surface accumulation could be suppressed even further by replacing the sinusoidal boundaries with strictly concave structures. To test this hypothesis, the strictly non-convex semicircle geometry ($R = 12\ \mu$m) seen in Fig. 5-1(c) are created. To compare this semicircle surface with the flat and the optimal sinusoidal ($A = 5.25\ \mu$m, $\lambda = 28\ \mu$m) surfaces, the segmented bacteria trajectories (acquired at 10 fps for 5 min) are projected onto one wavelength. The flat surface is set to have the same wavelength as the semicircle geometry. Because the bacterial concentration is different for the three surfaces, the raw data is normalized by using Bernoulli sampling to ensure the bulk density, defined as the density 50 μm away from boundary, is the same in each case. Samples of the resulting normalized data in Fig. 5-4 illustrate the distribution of cells for the three surfaces for the experiment and simulations. The total cell numbers differ between the three geometries reflecting the differences in the surface entrapment. After contact with the sinusoidal and semicircle geometries, the bacteria leave the surface at a particular angle, as evidenced by the inward streaks in Figs. 5-4(b)—5-4(c). This behavior is more clearly reproduced in the RT simulations than the BD simulations. Above the boundaries, there are depletion zones for the

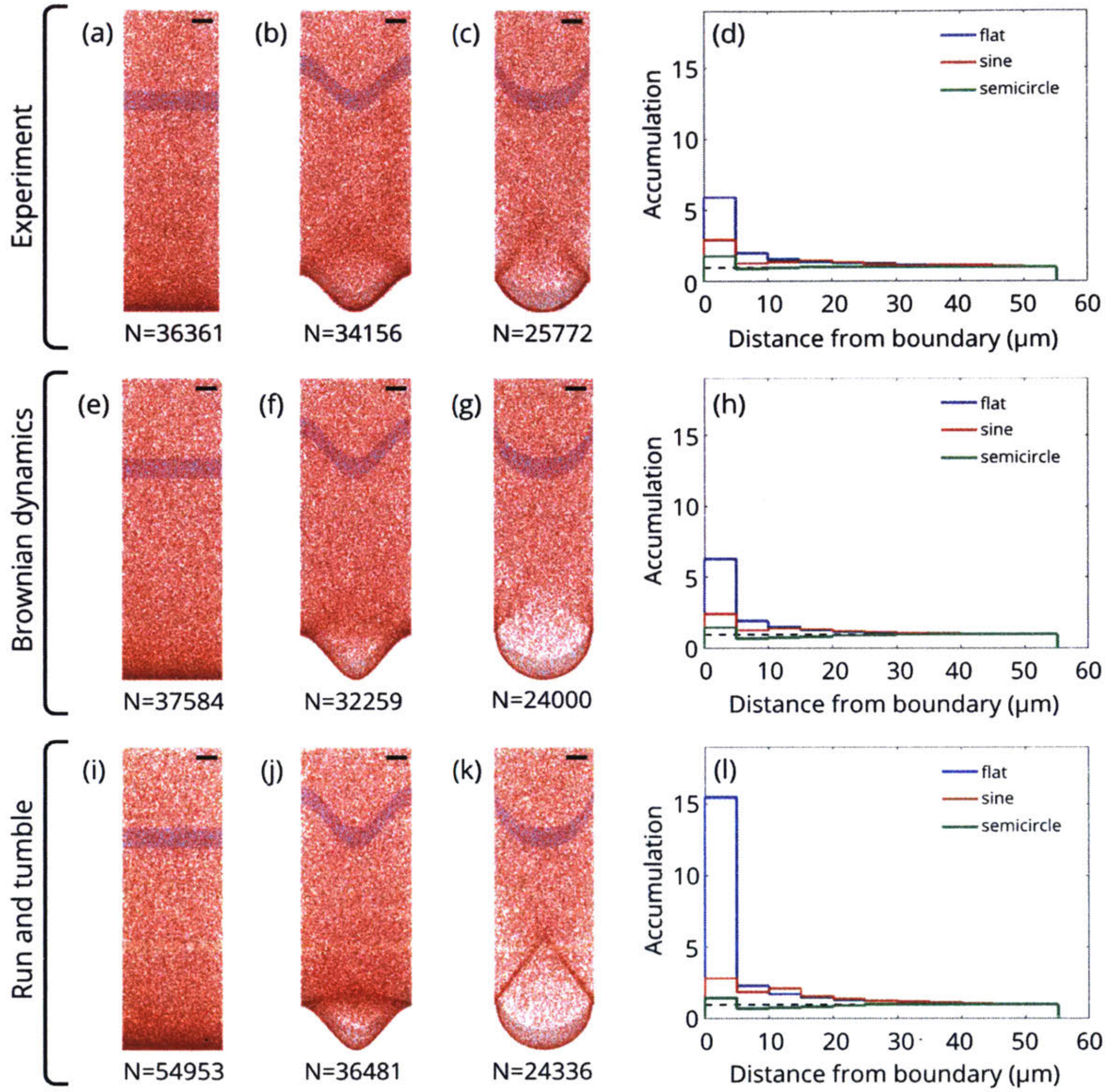


Figure 5-4: Sampled raw data and accumulation histograms for the flat, sinusoidal ($A = 5.25$ μm, $\lambda = 28$ μm), and concave semicircle ($R = 12$ μm) surface geometries for the experiment and simulations. To visualize the spatial cell distributions, the raw data, acquired at 10 fps for 5 min, were projected onto a single wavelength (the flat surface is assumed to have the same wavelength as the semicircle surface) and sampled such that the bulk density is the same in all cases (a) - (c), (e) - (g), and (i) - (k). Both the experiments and simulations qualitatively show a depletion zone above the boundary for the sinusoidal and semicircle geometries. Due to the differences in the surface accumulation, the total cell numbers differ for the three geometries. The accumulation histograms (d), (h), and (l) quantify this effect, with accumulation defined is the ratio of the number of bacteria in each surface bin area (grey region for first bin) to the number of bacteria in a congruent area 50 μm away from the surface (blue region). The results are independent of the shape of the bulk reference area (see Fig. 5-5). Histograms (d), (h) and (l) were computed from 20 independent subsamples of the raw data. The black dashed line indicates the bulk accumulation value of 1. The accumulation histograms show that the concave semicircle geometry is the most efficient at suppressing accumulation in the experiment and simulations. Bin width 5 μm. Scale bars 5 μm. Figure from [94].

sinusoidal and, more prominently, the semicircle geometry. The fact that these depletion zones are also reproduced by the simulations suggests that they arise from the scattering dynamics and not by hydrodynamic effects. The accumulation histograms in Figs. 5-4(d), 5-4(h), and 5-4(l) quantify and compare the bacterial distribution, where the solid line and shaded regions represent the mean and standard deviation of the accumulation ratio for 20 independent subsamples of the raw data.

As before, accumulation is defined as the number of bacteria in the each bin area (grey region for the first bin near the surface) divided by the number of bacteria in the bulk area (blue region), which is 50 μm away from the surface. Each bin is 5 μm wide and follows the boundary contour. The results are independent of the shape taken for the bulk area (Fig. 5-5). The black dashed line at height 1 indicates the bulk reference value. As evident from Fig. 5-4(d), both the flat geometry (blue line) and the partially convex sinusoidal geometry (red line) lead to cell accumulation above the bulk level up to 30 μm from the surface, although this effect is substantially weaker for the sinusoidal geometry. By contrast, except very close to the surface, the distribution of cells for the semicircle geometry (green line) is at the bulk level. Close to the surface, the semicircle geometry decreases the average cell concentration by 70% relative to a flat surface. Thus, the concave semicircle geometry is the most efficient at suppressing accumulation, in agreement with the predictions from the BD and RT models.

Compared to the RT model, the cells in the BD model leave the surface more easily. This can be seen in the histogram curves for the flat geometries (blue lines), which show good agreement between the BD model and experiment, whereas the RT model overestimates the accumulation in the first bin [Fig. 5-5(l)]. Yet, the RT model performs slightly better at replicating the trajectories of the cells after contact with curved surfaces than the BD model [Figs. 5-4(c), 5-4(g), and 5-4(k)]. Thus, bacterial reorientation in the experiments is likely a combination of BD and RT. While the BD and RT underestimate the accumulation for the sinusoidal geometry, they both agree well with experiment for the semicircle geometry, suggesting that near-field hydrodynamics could play a larger role in the bacterial surface entrapment for flat and convex geometries than for concave geometries.

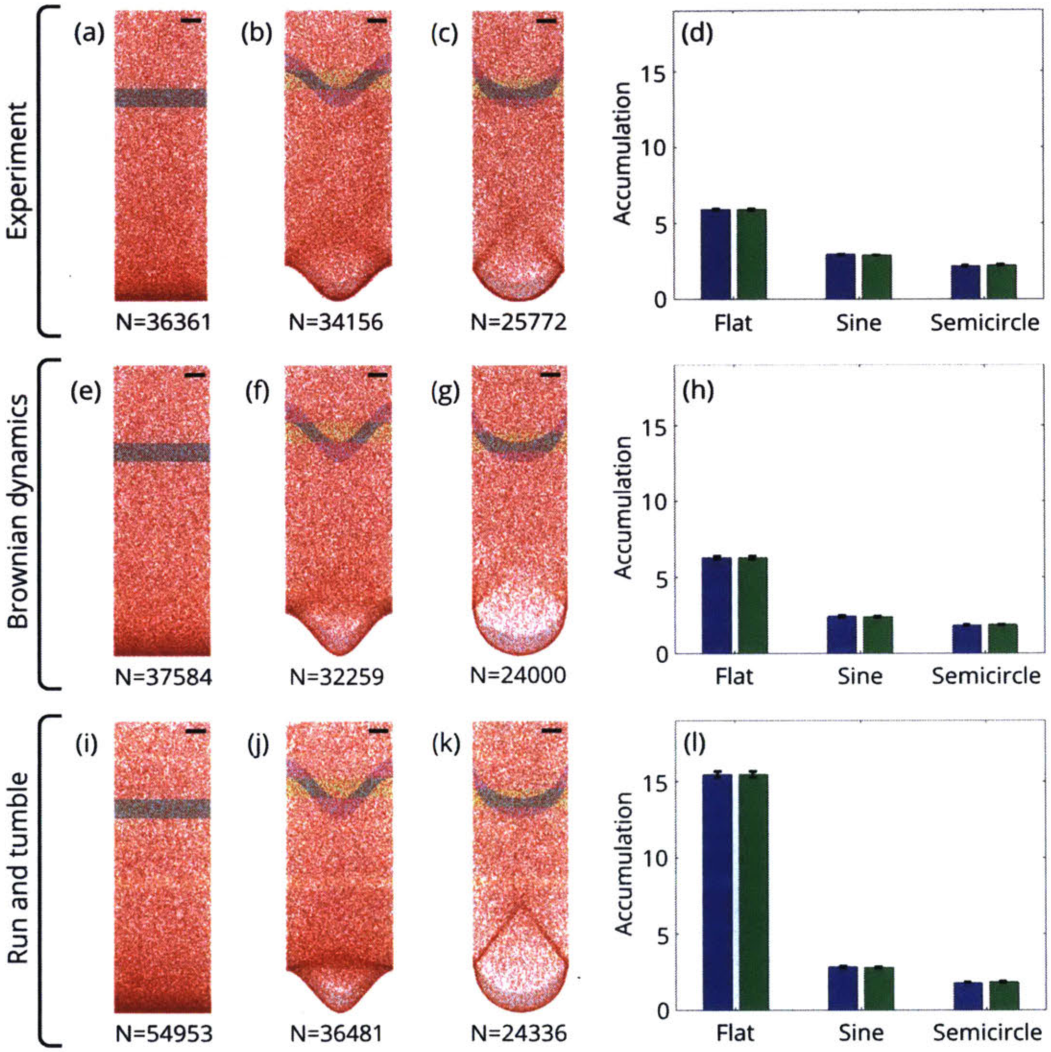


Figure 5-5: Sampled raw data and surface accumulation bar graphs for the flat, sinusoidal ($A = 5.25$ μm, $\lambda = 28$ μm), and semicircle ($R = 12$ μm) surface geometries for the experiment and simulations. To visualize the spatial cell distributions, the raw data, acquired at 10 fps for 5 min, were projected onto a single wavelength (the flat surface is assumed to have the same wavelength as the semicircle surface) and sampled such that the bulk density is the same in all cases (a) - (c), (e) - (g), and (i) - (k). Due to the differences in the surface accumulation, the total cell numbers differ for the three geometries. Accumulation is defined as the ratio of the number of bacteria in each surface bin area (grey regions) to the number of bacteria in an equally sized area 50 μm away from the surface. Two shapes of equal area are considered for the bulk area: a shape which follows the surface contour (blue) and a rectangle (green). For each geometry in (d), (h), and (l), the blue and green bar show the mean surface accumulation calculated with the surface contour and rectangle as the bulk area, respectively, for 20 independent subsamples of the raw data. The error bars represent the standard deviation. The blue and green bars are nearly equal for each case, demonstrating that the surface accumulation estimation is independent of the shape taken for the bulk reference area. Bin width 5 μm. Scale bars 5 μm. Figure from [94].

Chapter 6

Concluding remarks

In this thesis, a mechanistic individual-based model for collective bacteria dynamics is introduced (Chapter 2). This model accounts for biophysical processes such as bacteria growth and division, bacteria self-propulsion, viscous drag, bacteria-surface interactions, and bacteria-bacteria interactions. Specifically, this model utilizes interaction potentials to capture the physical cell-surface and cell-cell interactions. In particular, cell-surface interactions are captured by a repulsive potential, which models the steric interactions between bacteria and a surface. Cell-cell interactions are encoded in a repulsive and attractive potential, which accounts for hard steric and osmotic repulsion and cell-cell attraction through secreted components. A highly parallel version of this model is implemented on GPUs, achieving a 10x speedup compared to the CPU code and allowing for $\mathcal{O}(N)$ scaling. This code is used to investigate a diverse range of bacterial systems. Here, simulations, which are validated by single-cell experimental data, are performed to determine the relevant forces and interactions that determine the dynamics in bacterial swarming, biofilm formation, and bacterial surface accumulation.

Chapter 3 describes an novel adaptive microscopy approach which allowed for the single-cell imaging of swarm dynamics spanning five orders of magnitude in space and six orders of magnitude in time. Because cell proliferation and swarming are both far-from equilibrium biophysical processes, the absence of fundamental conservation laws makes it difficult to identify and characterize qualitatively distinct dynamical phases with conventional equilibrium–thermodynamic approaches. To overcome this conceptual challenge, experiments are combined with particle-based active matter modeling with machine learning to identify and characterize the spatiotemporal evolution of three pure and two coexistence phases during swarm development. This integrated approach suggests that steric interactions and motility are sufficient for explaining the observed dynamics within each phase, which enables a unified conceptual understanding of the emergent multiscale behavioral complexity in swarms in terms of basic biophysical parameters.

Chapter 4 describes the combined experimental and theoretical analysis which showed that the emergence of local nematic order in growing *V. cholerae* biofilms can be captured by an experimentally constrained effective cell–cell interaction potential that translates molecular mechanisms into force parameters. Given the immense complexity of the molecular interactions, metabolism, and signaling that occur between cells, the availability of an experimentally validated potential-based description of biofilm development presents a significant conceptual advance that can provide a microscopic basis for constructing predictive macroscopic continuum theories, by building on coarse-graining techniques recently developed for other classes of active matter [17, 90]. At the same time, a refined model will be needed to account for the spatial heterogeneities and time dependencies that become relevant at the later stages of biofilm development. Such progress is essential for identifying new strategies towards understanding, controlling, and inhibiting biofilm growth under realistic physiological conditions, which remains one of the foremost challenges in biomedical [89, 118, 137] and biophysical research [36, 105, 120].

In Chapter 5, experiments and simulations are used to investigate the scattering and accumulation dynamics of swimming bacteria in the vicinity of curved periodic boundaries. The results demonstrate that a concave boundary can reduce the average cell accumulation by more than 50% relative to a flat surface. Despite the simplifying model assumptions, simulations of a basic steric interaction model can account for experimental observations across the different geometries. In the future, it would be interesting to perform a similar analysis for 2D microtopographic surface designs where hydrodynamic near-field interactions [35, 124] can be expected to become relevant.

Extensions to this work would be to improve the cell-surface adhesion, to include the spatialtemporal heterogeneities that exist in larger biofilms ($>$ 1000 cells), and to capture the effect of flow on biofilms. Specific details of each of these extensions will be discussed in more detail below.

In the current model, cell-surface attachment is represented by an increase in the friction near the surface. To improve upon this, the cell-surface adhesion should be directly modeled. These attachment forces, which can range from a few to hundreds of picoNewtons, are a result of the secreted components, protein structures, and short appendages (such as fimbriae and type IV pilus) that cells use to remain attached to the surface [89, 105, 134]. Cell-surface attachment is also important in the verticalization of biofilms [9]. To capture cell-surface adhesion, an attractive part to the cell-surface potential as in [9] can be added to the model. Or, springs between the cell and the surface can also be used to represent these attachement forces. The location of these adhesive forces on the cell also makes a difference in the microcolony morphogenesis where the asymmetry of the surface attachment induces mechanical tension and buckling of the daughter cells [23, 39].

In order to capture the formation dynamics of larger biofilm systems ($>$ 1000 cells), the model needs to account for spatialtemporal heterogeneities. For example, in larger biofilms, the growth rates of the cells vary depending on where the cell is located, with cells in the center having lower growth rates than cells on the edge due to the availability of nutrients [86]. This could be added to the model by varying the growth rate with position in the biofilm or by modeling the diffusion of the environment's nutrient fields. Also, spatialtemporal heterogeneities in gene expression and matrix composition should be included in the model. Physiological heterogeneity arises in large biofilms as the cells adapt to chemical gradients and local environmental conditions and through cell mutations [126]. This can be added in the model by including gene expression as a state variable of the cells. Matrix composition also varies throughout the biofilm. In *V. cholerae* biofilms in particular, the matrix is composed of four main proteins, where RbmA is distributed homogeneously throughout the biofilm and provides cell-cell adhesion, Bap1 adheres the cells to the surface, and a mixture of VPS, RbmC, and Bap1 encloses the cell clusters [6]. To include these different adhesive forces from the proteins, the cell-cell potential can be adjusted to vary in space. Also, this matrix composition heterogenity will contribute to an increase in the local friction experienced by the cells, with larger amounts of matrix resulting in more friction. This can be captured by varying the viscosity felt by each cell by location, with the highest local viscosity corresponding to highest concentration of cells [103]. By adding age as a state variable, the effects of cell death can also be captured, as the DNA from the lysed cells adds to the structural integrity of the matrix for certain species of bacteria [3].

It would also be interesting to add the effect of external fluid flow into the model, as high fluid shear is known to change the architecture and morphology of developing biofilms [37, 57, 61]. Biofilms under high shear flow show strong alignment with the flow and anisotropic biofilm expansion toward the downstream direction of the flow [61]. In order to capture this in the model, the cell-surface attachment needs to be improved, as discussed above, and the attraction between mother and daughter cells needs to be captured more accurately because it was observed that cells align with the flow due to the fluid drag and because of this asymmetrical attachment between cells after division [103]. This attachment between mother and daughter cells can be modeled with a spring between the polar endpoints of the cells [103].

An exciting application of this simulation framework would be in the design of synthetic multi-cell assemblies and patterns. Using surface-bound nanobodies and antigens that only bind in specific combinations, researchers have developed a cell-cell adhesion library which enables the design of static self-assembling multi-cell structures depending on the cell shape and mixing ratios of the cells expressing the surface-bound nanobodies and antigens [54]. By adding a state variable specifying what kind of nanobody or antigen is expressed on the cell, the simulation framework could be extended to capture this selective cell-cell adhesion by forming a spring between cells of a specific type. The simulations could then provide predictions and insights into the dynamics behind these synthetic self-assembling multicellular patterns and aid in the design of future multicellular morphologies.

Appendix A

von Mises-Fisher Distribution

The von Mises-Fisher (vMF) distribution can be described as a Gaussian distribution of a $(d-1)$-dimensional sphere $\mathbb{S}^{d-1}$ embedded in $\mathbb{R}^d$. As defined in [44], the probability density function of the vMF distribution is

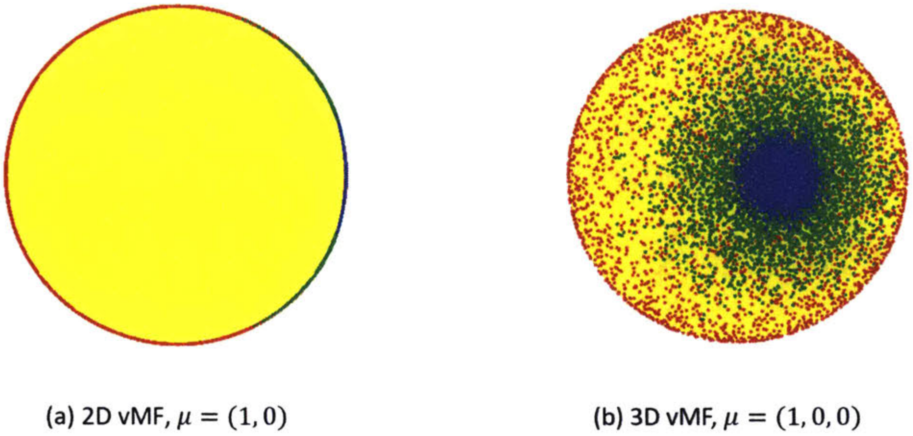
$$f(\mathbf{x}|\mu, \kappa) = c_d(\kappa) e^{\kappa \mu^{\mathsf{T}} \mathbf{x}} \tag{A.1}$$

where $\mathbf{x}$ is a unit random vector ($\mathbf{x} \in \mathbb{R}^d$ and $\|\mathbf{x}\| = 1$), μ is the unit vector ($\|\mu\| = 1$) describing the mean direction, and κ is the concentration parameter ($\kappa \geq 0$). κ describes how concentrated the unit vectors drawn from the vMF distribution are to μ. For $\kappa = 0$, the vMF distribution is uniform on the sphere $\mathbb{S}^{d-1}$. For $\kappa \to \infty$, the vMF distribution approaches a point distribution. This is illustrated in Fig. A-1, which shows points drawn from the 2D and 3D vMF distributions for various values of κ. The normalization constant $c_d(\kappa)$ is

$$c_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)} \tag{A.2}$$

where $I_r(\cdot)$ is the modified Bessel function of the first kind at order r. For $d = 2$, $c_2(\kappa)$ is

$$c_s(\kappa) = \frac{1}{2\pi I_0(\kappa)} \tag{A.3}$$



(a) 2D vMF, $\mu = (1,0)$ **(b) 3D vMF, $\mu = (1,0,0)$**

Figure A-1: 5000 unit vectors drawn from the vMF distribution on the unit sphere in (a) 2D, $\mathbb{S}^1$, and (b) 3D, $\mathbb{S}^2$. $\mu = (1,0)$ and $\mu = (1,0,0)$ for (a) and (b), respectively. $\kappa = 0$ (red points), $\kappa = 10$ (green points), and $\kappa = 100$ (blue points). Note that the vMF distribution is uniform on the sphere $\mathbb{S}^{d-1}$ for $\kappa = 0$ and approaches a point distribution as $\kappa \to \infty$.

For $d = 3$, $c_3(\kappa)$ simplifies to

$$c_3(\kappa) = \frac{\kappa}{4\pi \sinh(\kappa)} = \frac{\kappa}{2\pi(e^{\kappa} - e^{-\kappa})} \tag{A.4}$$

The Matlab program `randvonMisesFisher3.m` [70] is used to draw points from the vMF distribution in the code.

Appendix B

Derivation of the Berne-Pechukas potential

The overlap potential, as described in [8], is obtained by considering the overlap of two ellipsoidal Gaussians, which have the form

$$G(x_i) = \exp(-x_i A_{ij}^{-1} x_j) \tag{B.1}$$

where

$$A_{ij} = \ell^2 \hat{n}_i \hat{n}_j + d^2(\delta_{ij} - \hat{n}_i \hat{n}_j) \tag{B.2}$$

Here, $\mathbf{x}$ is a point in space. $\hat{\mathbf{n}}$ is a unit vector in the direction of the principal axis of the ellipsoid. ℓ and d characterize the ellipsoid's half-length and half-width, respectively. δ_{ij} is the Kronecker delta, as defined below

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{B.3}$$

Note, it is possible to derive an expression for A_{ij}^{-1}. Assuming that A_{ij}^{-1} has a similar form to A_{ij} in that $A_{ij}^{-1} = B_{ij} = \beta_{\parallel}^2 \hat{n}_i \hat{n}_j + \beta_{\perp}^2(\delta_{ij} - \hat{n}_i \hat{n}_j)$. Therefore,

$$\delta_{ik} = A_{ij} B_{jk}$$

$$\begin{aligned}
&= \left(\ell^2\hat{n}_i\hat{n}_j + d^2(\delta_{ij} - \hat{n}_i\hat{n}_j)\right)\left(\beta_{\|}^2\hat{n}_j\hat{n}_k + \beta_{\perp}^2(\delta_{jk} - \hat{n}_j\hat{n}_k)\right) \\
&= \ell^2\beta_{\|}^2\hat{n}_i\hat{n}_j\hat{n}_j\hat{n}_k + \ell^2\beta_{\perp}^2\hat{n}_i\hat{n}_j\delta_{jk} - \ell^2\beta_{\perp}^2\hat{n}_i\hat{n}_j\hat{n}_j\hat{n}_k \\
&\quad + d^2\beta_{\|}^2\hat{n}_j\hat{n}_k\delta_{ij} + d^2\beta_{\perp}^2\delta_{ij}\delta_{jk} - d^2\beta_{\perp}^2\hat{n}_j\hat{n}_k\delta_{ij} \\
&\quad - d^2\beta_{\|}^2\hat{n}_i\hat{n}_j\hat{n}_j\hat{n}_k - d^2\beta_{\perp}^2\hat{n}_i\hat{n}_j\delta_{jk} + d^2\beta_{\perp}^2\hat{n}_i\hat{n}_j\hat{n}_j\hat{n}_k \\
&= (\ell^2\beta_{\|}^2 + \ell^2\beta_{\perp}^2 - \ell^2\beta_{\perp}^2 + d^2\beta_{\|}^2 - d^2\beta_{\perp}^2 + d^2\beta_{\|}^2 - d^2\beta_{\perp}^2 + d^2\beta_{\perp}^2)\hat{n}_i\hat{n}_k + d^2\beta_{\perp}^2\delta_{ik} \\
&= (\ell^2\beta_{\|}^2 - d^2\beta_{\perp}^2)\hat{n}_i\hat{n}_k + d^2\beta_{\perp}^2\delta_{ik} \qquad \text{(B.4)}
\end{aligned}$$

where use has been made of the fact that

$$\begin{aligned}
\hat{n}_i\hat{n}_j\hat{n}_j\hat{n}_k &= \hat{n}_i\hat{n}_1\hat{n}_1\hat{n}_k + \hat{n}_i\hat{n}_2\hat{n}_2\hat{n}_k + \hat{n}_i\hat{n}_3\hat{n}_3\hat{n}_k \\
&= \hat{n}_i\hat{n}_k(\hat{n}_1\hat{n}_1 + \hat{n}_2\hat{n}_2 + \hat{n}_3\hat{n}_3) \\
&= \hat{n}_i\hat{n}_k(1) \\
&= \hat{n}_i\hat{n}_k \qquad \text{(B.5)}
\end{aligned}$$

because $\hat{n}$ is a unit vector. Thus, Eq. (B.4) requires

$$\beta_{\perp}^2 = \frac{1}{d^2} \text{ and } \beta_{\|}^2 = \frac{1}{\ell^2} \tag{B.6}$$

which results in

$$A_{ij}^{-1} = \frac{1}{\ell^2}\hat{n}_i\hat{n}_j + \frac{1}{d^2}(\delta_{ij} - \hat{n}_i\hat{n}_j) \tag{B.7}$$

Berne and Pechukas proposed that the overlap potential is proportional to the mathematical overlap of these ellipsoidal Gaussians, which can be calculated through the integration over all space of the product of two ellipsoidal Gaussians. Consider two ellipsoidal Gaussians, α and β, with axial unit vectors, $\hat{\mathbf{n}}_\alpha$ and $\hat{\mathbf{n}}_\beta$. The vector $\mathbf{r}_{\alpha\beta}$ joins their centers. Note that $\mathbf{r}_{\alpha\beta}$ points from ellipsoid α to ellipsoid β as shown in Fig. 2-4. From the figure, note that the origin of the coordinate system is attached to ellipsoid α. The overlap is determined by

$$U_{\text{overlap}} \propto |\mathbf{A}_\alpha|^{-1/2}|\mathbf{A}_\beta|^{-1/2}\int_{-\infty}^{\infty}\exp[-x_iA^{-1}_{\alpha ij}x_j-(x_i-r_{\alpha\beta i})A^{-1}_{\beta ij}(x_j-r_{\alpha\beta j})]dx_1dx_2dx_3 \tag{B.8}$$

The argument of the exponential function can be simplified to

$$-x_iA^{-1}_{\alpha ij}x_j-(x_i-r_{\alpha\beta i})A^{-1}_{\beta ij}(x_j-r_{\alpha\beta j}) \tag{B.9}$$

$$\begin{aligned}
&= -x_iA^{-1}_{\alpha ij}x_j - x_iA^{-1}_{\beta ij}x_j + x_iA^{-1}_{\beta ij}r_{\alpha\beta j} + r_{\alpha\beta i}A^{-1}_{\beta ij}x_j - r_{\alpha\beta i}A^{-1}_{\beta ij}r_{\alpha\beta j}\\
&= -x_i[A^{-1}_{\alpha ij} + A^{-1}_{\beta ij}]x_j + 2x_iA^{-1}_{\beta ij}r_{\alpha\beta j} - r_{\alpha\beta i}A^{-1}_{\beta ij}r_{\alpha\beta j}\\
&= -x_iS_{ij}x_j + c_ix_i - r_{\alpha\beta i}A^{-1}_{\beta ij}r_{\alpha\beta j}
\end{aligned} \tag{B.10}$$

where $S_{ij} = [A^{-1}_{\alpha ij} + A^{-1}_{\beta ij}]$ and $c_i = 2A^{-1}_{\beta ij}r_{\alpha\beta j}$. Note that use has been made of the fact that $A^{-1}_{\beta ij}$ is symmetric. Eq. (16.11) of [13], which is reproduced below for convenience,

$$\int_{-\infty}^{\infty}\exp(-Y^\top SY + D^\top Y)\prod_{i=1}^{N}dy_i = \sqrt{\frac{\pi^N}{\det(S)}}\exp\left(\frac{1}{4}D^\top S^{-1}D\right) \tag{B.11}$$

in which S is a positive symmetric matrix and D is a complex vector, gives the integration of Eq. (B.8) such that

$$\begin{aligned}
&U_{\text{overlap}}(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \mathbf{r}_{\alpha\beta})\\
&\qquad\propto \pi^{3/2}|\mathbf{A}_\alpha|^{-1/2}|\mathbf{A}_\beta|^{-1/2}|\mathbf{S}|^{-1/2}\exp[-r_{\alpha\beta i}A^{-1}_{\beta ij}r_{\alpha\beta j}]\exp\left[\frac{1}{4}c_iS^{-1}_{ij}c_j\right]\\
&\qquad\propto \pi^{3/2}|\mathbf{A}_\alpha|^{-1/2}|\mathbf{A}_\beta|^{-1/2}|\mathbf{S}|^{-1/2}\exp\left[\frac{1}{4}c_iS^{-1}_{ij}c_j - r_{\alpha\beta i}A^{-1}_{\beta ij}r_{\alpha\beta j}\right]
\end{aligned} \tag{B.12}$$

Note,

$$\begin{aligned}
|\mathbf{A}_\alpha|^{-1/2}|\mathbf{S}|^{-1/2}|\mathbf{A}_\beta|^{-1/2} &= |A_{\alpha ik}S_{km}A_{\beta mj}|^{-1/2}\\
&= |A_{\alpha ik}[A^{-1}_{\alpha km} + A^{-1}_{\beta km}]A_{\beta mj}|^{-1/2}\\
&= |A_{\alpha ik}A^{-1}_{\alpha km}A_{\beta mj} + A_{\alpha ik}A^{-1}_{\beta km}A_{\beta mj}|^{-1/2}
\end{aligned}$$

$$\begin{aligned}
&= |\delta_{im} A_{\beta mj} + A_{\alpha ik} \delta_{kj}|^{-1/2} \\
&= |A_{\beta ij} + A_{\alpha ij}|^{-1/2} \\
&= |\mathbf{A}_\alpha + \mathbf{A}_\beta|^{-1/2} \qquad \text{(B.13)}
\end{aligned}$$

where use has been made of the fact that $|\mathbf{A}||\mathbf{B}||\mathbf{C}| = |\mathbf{A\,B\,C}|$.

Eq. (24) from [63] (with $\mathbf{U} = \mathbf{v} = \mathbf{I}$) gives the inverse of a sum of matrices $\mathbf{A}$ and $\mathbf{B}$, which is reproduced below for convenience

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1} \qquad \text{(B.14)}$$

and is applied to S_{ij}^{-1}

$$\begin{aligned}
S_{ij}^{-1} &= (A_{\beta ij}^{-1} + A_{\alpha ij}^{-1})^{-1} \\
&= A_{\beta ij} - A_{\beta ik} A_{\alpha km}^{-1} (\delta_{mn} + A_{\beta mr} A_{\alpha rn}^{-1})^{-1} A_{\beta nj} \qquad \text{(B.15)}
\end{aligned}$$

This results in

$$\begin{aligned}
\frac{1}{4} c_i S_{ij}^{-1} c_j &= \frac{1}{4}(2A_{\beta it}^{-1} r_{\alpha\beta t})[A_{\beta ij} - A_{\beta ik} A_{\alpha km}^{-1} (\delta_{mn} + A_{\beta mr} A_{\alpha rn}^{-1})^{-1} A_{\beta nj}](2A_{\beta jw}^{-1} r_{\alpha\beta w}) \\
&= r_{\alpha\beta t} r_{\alpha\beta w} A_{\beta ti}^{-1} [A_{\beta ij} - A_{\beta ik} A_{\alpha km}^{-1} (\delta_{mn} + A_{\beta mr} A_{\alpha rn}^{-1})^{-1} A_{\beta nj}] A_{\beta jw}^{-1} \\
&= r_{\alpha\beta t} r_{\alpha\beta w} [\delta_{tj} - \delta_{tk} A_{\alpha km}^{-1} (\delta_{mn} + A_{\beta mr} A_{\alpha rn}^{-1})^{-1} A_{\beta nj}] A_{\beta jw}^{-1} \\
&= r_{\alpha\beta t} r_{\alpha\beta w} [A_{\beta tw}^{-1} - A_{\alpha tm}^{-1} (\delta_{mn} + A_{\beta mr} A_{\alpha rn}^{-1})^{-1} \delta_{nw}] \\
&= r_{\alpha\beta t} r_{\alpha\beta w} [A_{\beta tw}^{-1} - [(\delta_{tm} + A_{\beta tr} A_{\alpha rm}^{-1}) A_{\alpha mn}]^{-1} \delta_{nw}] \\
&= r_{\alpha\beta t} r_{\alpha\beta w} [A_{\beta tw}^{-1} - (A_{\alpha tn} + A_{\beta tr} \delta_{rn})^{-1} \delta_{nw}] \\
&= r_{\alpha\beta t} r_{\alpha\beta w} [A_{\beta tw}^{-1} - (A_{\alpha tn} + A_{\beta tn})^{-1} \delta_{nw}^{-1}] \\
&= r_{\alpha\beta t} r_{\alpha\beta w} [A_{\beta tw}^{-1} - [\delta_{tn} (A_{\alpha nw} + A_{\beta nw})]^{-1}] \\
&= r_{\alpha\beta t} r_{\alpha\beta w} [A_{\beta tw}^{-1} - (A_{\alpha tw} + A_{\beta tw})^{-1}] \\
&= r_{\alpha\beta i} [A_{\beta ij}^{-1} - (A_{\alpha ij} + A_{\beta ij})^{-1}] r_{\alpha\beta j} \qquad \text{(B.16)}
\end{aligned}$$

where use has been made of the fact that $\mathbf{A}_\beta^{-1}$ is symmetric, $\mathbf{I}^{-1} = \mathbf{I}$, and that

$(\mathbf{B})^{-1}(\mathbf{A})^{-1} = (\mathbf{A}\,\mathbf{B})^{-1}$.

Therefore, the overlap potential can be written as

$$U_{\text{overlap}}(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \mathbf{r}_{\alpha\beta})$$

$$\begin{aligned} &\propto \pi^{3/2}|\mathbf{A}_\alpha + \mathbf{A}_\beta|^{-1/2} \exp[r_{\alpha\beta i}[A^{-1}_{\beta ij} - (A_{\alpha ij} + A_{\beta ij})^{-1}]r_{\alpha\beta j} - r_{\alpha\beta i}A^{-1}_{\beta ij}r_{\alpha\beta j}] \\ &\propto \pi^{3/2}|\mathbf{A}_\alpha + \mathbf{A}_\beta|^{-1/2} \exp[-r_{\alpha\beta i}(A_{\alpha ij} + A_{\beta ij})^{-1}r_{\alpha\beta j}] \qquad \text{(B.17)} \end{aligned}$$

Expressing $\mathbf{r}_{\alpha\beta} = r_{\alpha\beta}\hat{\mathbf{r}}_{\alpha\beta}$ where $r_{\alpha\beta}$ is the magnitude of $\mathbf{r}_{\alpha\beta}$ and $\hat{\mathbf{r}}_{\alpha\beta}$ is the unit vector in the direction of $\mathbf{r}_{\alpha\beta}$, Eq. (B.17) can be written as

$$U_{\text{overlap}}(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \mathbf{r}_{\alpha\beta}) = \epsilon_0\epsilon_1(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta) \exp[-r^2_{\alpha\beta}/\sigma^2(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \hat{\mathbf{r}}_{\alpha\beta})] \qquad \text{(B.18)}$$

where ϵ_0 describes the strength of the interaction. ϵ_0 captures the magnitude of the product of $\pi^{3/2}$ and other constants. The strength parameter $\epsilon_1(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta)$ is

$$\epsilon_1(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta) \propto |\mathbf{A}_\alpha + \mathbf{A}_\beta|^{-1/2} \qquad \text{(B.19)}$$

and the range parameter $\sigma(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \hat{\mathbf{r}}_{\alpha\beta})$ is defined as

$$\sigma^2(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \hat{\mathbf{r}}_{\alpha\beta}) = \frac{1}{\hat{r}_{\alpha\beta i}(A_{\alpha ij} + A_{\beta ij})^{-1}\hat{r}_{\alpha\beta j}} \qquad \text{(B.20)}$$

Appendix C

Derivation of generalized strength and range parameters

To derive a generalized form of the strength parameter ϵ_1 defined as

$$\epsilon_1(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta) \propto |\mathbf{A}_\alpha + \mathbf{A}_\beta|^{-1/2} \tag{C.1}$$

for nonidentical particles, the elements of matrix $\mathbf{A}_\alpha + \mathbf{A}_\beta$ are expanded and the determinant is calculated. Note that the determinant of a symmetric matrix with elements B_{ij} is

$$|\mathbf{B}| = B_{11}B_{22}B_{33} - B_{11}(B_{23})^2 - B_{22}(B_{13})^2 - B_{33}(B_{12})^2 + 2B_{12}B_{13}B_{23} \tag{C.2}$$

$A_{\alpha ij} + A_{\beta ij}$ can be rewritten as

$$A_{\alpha ij} + A_{\beta ij} = \eta_\alpha \hat{n}_{\alpha i}\hat{n}_{\alpha j} + \eta_\beta \hat{n}_{\beta i}\hat{n}_{\beta j} + \zeta\delta_{ij} \tag{C.3}$$

where $\eta_\alpha = \ell_\alpha^2 - d_\alpha^2$, $\eta_\beta = \ell_\beta^2 - d_\beta^2$, and $\zeta = d_\alpha^2 + d_\beta^2$. The elements of $\mathbf{A}_\alpha + \mathbf{A}_\beta$ can be calculated as

$$(\mathbf{A}_\alpha + \mathbf{A}_\beta)_{11} = \eta_\alpha(\hat{n}_{\alpha 1})^2 + \eta_\beta(\hat{n}_{\beta 1})^2 + \zeta$$

$$(\mathbf{A}_\alpha + \mathbf{A}_\beta)_{22} = \eta_\alpha(\hat{n}_{\alpha 2})^2 + \eta_\beta(\hat{n}_{\beta 2})^2 + \zeta$$

$$
\begin{aligned}
(\mathbf{A}_\alpha + \mathbf{A}_\beta)_{33} &= \eta_\alpha(\hat{n}_{\alpha 3})^2 + \eta_\beta(\hat{n}_{\beta 3})^2 + \zeta \\
(\mathbf{A}_\alpha + \mathbf{A}_\beta)_{12} = (\mathbf{A}_\alpha + \mathbf{A}_\beta)_{21} &= \eta_\alpha \hat{n}_{\alpha 1}\hat{n}_{\alpha 2} + \eta_\beta \hat{n}_{\beta 1}\hat{n}_{\beta 2} \\
(\mathbf{A}_\alpha + \mathbf{A}_\beta)_{13} = (\mathbf{A}_\alpha + \mathbf{A}_\beta)_{31} &= \eta_\alpha \hat{n}_{\alpha 1}\hat{n}_{\alpha 3} + \eta_\beta \hat{n}_{\beta 1}\hat{n}_{\beta 3} \\
(\mathbf{A}_\alpha + \mathbf{A}_\beta)_{23} = (\mathbf{A}_\alpha + \mathbf{A}_\beta)_{32} &= \eta_\alpha \hat{n}_{\alpha 2}\hat{n}_{\alpha 3} + \eta_\beta \hat{n}_{\beta 2}\hat{n}_{\beta 3}
\end{aligned}
\tag{C.4}
$$

and inserted into Eq. (C.2), which results in

$$
\begin{aligned}
|\mathbf{A}_\alpha + \mathbf{A}_\beta| =&\eta_\alpha\eta_\beta\zeta\{(\hat{n}_{\alpha 2})^2 + (\hat{n}_{\alpha 3})^2 + (\hat{n}_{\beta 2})^2 + (\hat{n}_{\beta 3})^2 - (\hat{n}_{\alpha 3})^2(\hat{n}_{\beta 2})^2 - (\hat{n}_{\alpha 2})^2(\hat{n}_{\beta 3})^2 \\
&- 2(\hat{n}_{\alpha 2})^2(\hat{n}_{\beta 2})^2 - 2(\hat{n}_{\alpha 3})^2(\hat{n}_{\beta 3})^2 - 2\hat{n}_{\alpha 1}\hat{n}_{\alpha 2}\hat{n}_{\beta 1}\hat{n}_{\beta 2} - 2\hat{n}_{\alpha 1}\hat{n}_{\alpha 3}\hat{n}_{\beta 1}\hat{n}_{\beta 3} - 2\hat{n}_{\alpha 2}\hat{n}_{\alpha 3}\hat{n}_{\beta 2}\hat{n}_{\beta 3}\} \\
&+ \eta_\alpha\zeta^2 + \eta_\beta\zeta^2 + \zeta^3 \\
=&\eta_\alpha\eta_\beta\zeta\{(\hat{n}_{\alpha 2})^2 + (\hat{n}_{\alpha 3})^2 + (\hat{n}_{\beta 2})^2 + (\hat{n}_{\beta 3})^2 - (\hat{n}_{\alpha 3})^2(\hat{n}_{\beta 2})^2 - (\hat{n}_{\alpha 2})^2(\hat{n}_{\beta 3})^2 \\
&- 2(\hat{n}_{\alpha 2})^2(\hat{n}_{\beta 2})^2 - 2(\hat{n}_{\alpha 3})^2(\hat{n}_{\beta 3})^2 + (\hat{n}_{\alpha 1})^2(\hat{n}_{\beta 1})^2 + (\hat{n}_{\alpha 2})^2(\hat{n}_{\beta 2})^2 + (\hat{n}_{\alpha 3})^2(\hat{n}_{\beta 3})^2 \\
&- [2\hat{n}_{\alpha 1}\hat{n}_{\alpha 2}\hat{n}_{\beta 1}\hat{n}_{\beta 2} + 2\hat{n}_{\alpha 1}\hat{n}_{\alpha 3}\hat{n}_{\beta 1}\hat{n}_{\beta 3} + 2\hat{n}_{\alpha 2}\hat{n}_{\alpha 3}\hat{n}_{\beta 2}\hat{n}_{\beta 3} \\
&+ (\hat{n}_{\alpha 1})^2(\hat{n}_{\beta 1})^2 + (\hat{n}_{\alpha 2})^2(\hat{n}_{\beta 2})^2 + (\hat{n}_{\alpha 3})^2(\hat{n}_{\beta 3})^2]\} \\
&+ \eta_\alpha\zeta^2 + \eta_\beta\zeta^2 + \zeta^3 \\
=&\eta_\alpha\eta_\beta\zeta[(\hat{n}_{\alpha 2})^2 + (\hat{n}_{\alpha 3})^2 + (\hat{n}_{\beta 2})^2 + (\hat{n}_{\beta 3})^2 - (\hat{n}_{\alpha 3})^2(\hat{n}_{\beta 2})^2 - (\hat{n}_{\alpha 2})^2(\hat{n}_{\beta 3})^2 \\
&+ (\hat{n}_{\alpha 1})^2(\hat{n}_{\beta 1})^2 - (\hat{n}_{\alpha 2})^2(\hat{n}_{\beta 2})^2 - (\hat{n}_{\alpha 3})^2(\hat{n}_{\beta 3})^2 - (\hat{n}_{\alpha i}\hat{n}_{\beta i})^2] \\
&+ \eta_\alpha\zeta^2 + \eta_\beta\zeta^2 + \zeta^3 \\
=&\eta_\alpha\eta_\beta\zeta[1 - (\hat{n}_{\alpha i}\hat{n}_{\beta i})^2] + \eta_\alpha\zeta^2 + \eta_\beta\zeta^2 + \zeta^3 \\
=&- (d_\alpha^2 + d_\beta^2)(-d_\alpha^2 d_\beta^2 - d_\alpha^2\ell_\alpha^2 - d_\beta^2\ell_\beta^2 - \ell_\alpha^2\ell_\beta^2) \\
&- (d_\alpha^2 + d_\beta^2)(d_\alpha^2 d_\beta^2 - d_\beta^2\ell_\alpha^2 - d_\alpha^2\ell_\beta^2 + \ell_\alpha^2\ell_\beta^2)(\hat{n}_{\alpha i}\hat{n}_{\beta i})^2
\end{aligned}
\tag{C.5}
$$

where use has been made of $(u_i n_i)^2 = u_1^2 n_1^2 + u_2^2 n_2^2 + u_3^2 n_3^2 + 2u_1 u_2 n_1 n_2 + 2u_1 u_3 n_1 n_3 + 2u_2 u_3 n_2 n_3$ where u_i and n_i are two vectors and $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1$. Dividing each term by $\varpi = -(d_\alpha^2 + d_\beta^2)(-d_\alpha^2 d_\beta^2 - d_\alpha^2\ell_\alpha^2 - d_\beta^2\ell_\beta^2 - \ell_\alpha^2\ell_\beta^2)$,

$$
\frac{|\mathbf{A}_\alpha + \mathbf{A}_\beta|}{\varpi} = 1 - \chi^2(\hat{n}_{\alpha i}\hat{n}_{\beta i})^2 \tag{C.6}
$$

where

$$\chi = \left[\frac{(\ell_\alpha^2 - d_\alpha^2)(\ell_\beta^2 - d_\beta^2)}{(\ell_\alpha^2 + d_\beta^2)(\ell_\beta^2 + d_\alpha^2)} \right]^{1/2} \tag{C.7}$$

This reduces to $\frac{(\ell^2-d^2)}{(\ell^2+d^2)}$ when $\ell_\alpha = \ell_\beta = \ell$ and $d_\alpha = d_\beta = d$ which matches Eq. (5b) in [8]. Thus, ϵ_1 is defined as

$$\epsilon_1(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta) = [1 - \chi^2(\hat{n}_{\alpha i}\hat{n}_{\beta i})^2]^{-1/2} = [1 - \chi^2(\hat{\mathbf{n}}_\alpha \cdot \hat{\mathbf{n}}_\beta)^2]^{-1/2} \tag{C.8}$$

and $\varpi^{-1/2}$ can be captured in ϵ_0. The minimum value of 1 occurs when $\hat{\mathbf{n}}_\alpha$ is perpendicular to $\hat{\mathbf{n}}_\beta$. The maximum value occurs when $\hat{\mathbf{n}}_\alpha$ is parallel to $\hat{\mathbf{n}}_\beta$.

To derive an expression for a generalized form of the range parameter σ defined as

$$\sigma^2(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \hat{\mathbf{r}}_{\alpha\beta}) = \frac{1}{\hat{r}_{\alpha\beta i}(A_{\alpha ij} + A_{\beta ij})^{-1}\hat{r}_{\alpha\beta j}} \tag{C.9}$$

for nonidentical particles, $\mathbf{A}_\alpha + \mathbf{A}_\beta$ is rewritten as in Eq. (C.3) and $(A_\alpha + A_\beta)^{-1}_{ij}$ is assumed to be of the form

$$(A_\alpha + A_\beta)^{-1}_{ij} = \kappa_1\hat{n}_{\alpha i}\hat{n}_{\alpha j} + \kappa_2\hat{n}_{\beta i}\hat{n}_{\beta j} + \kappa_3\hat{n}_{\alpha i}\hat{n}_{\beta j} + \kappa_4\hat{n}_{\beta i}\hat{n}_{\alpha j} + \lambda\delta_{ij} \tag{C.10}$$

This results in

$$\begin{aligned}
\delta_{ik} =&(\eta_\alpha\hat{n}_{\alpha i}\hat{n}_{\alpha j} + \eta_\beta\hat{n}_{\beta i}\hat{n}_{\beta j} + \zeta\delta_{ij})(\kappa_1\hat{n}_{\alpha i}\hat{n}_{\alpha j} + \kappa_2\hat{n}_{\beta i}\hat{n}_{\beta j} + \kappa_3\hat{n}_{\alpha i}\hat{n}_{\beta j} + \kappa_4\hat{n}_{\beta i}\hat{n}_{\alpha j} + \lambda\delta_{ij}) \\
=&(\eta_\alpha\kappa_1 + \eta_\alpha\lambda + \zeta\kappa_1 + \eta_\alpha\kappa_4\hat{n}_{\alpha j}\hat{n}_{\beta j})\hat{n}_{\alpha i}\hat{n}_{\alpha k} \\
&+ (\eta_\beta\kappa_2 + \eta_\beta\lambda + \zeta\kappa_2 + \eta_\beta\kappa_3\hat{n}_{\alpha j}\hat{n}_{\beta j})\hat{n}_{\beta i}\hat{n}_{\beta k} \\
&+ (\eta_\alpha\kappa_3 + \zeta\kappa_3 + \eta_\alpha\kappa_2\hat{n}_{\alpha j}\hat{n}_{\beta j})\hat{n}_{\alpha i}\hat{n}_{\beta k} \\
&+ (\eta_\beta\kappa_4 + \zeta\kappa_4 + \eta_\beta\kappa_1\hat{n}_{\alpha j}\hat{n}_{\beta j})\hat{n}_{\beta i}\hat{n}_{\alpha k} \\
&+ \zeta\lambda\delta_{ik}
\end{aligned} \tag{C.11}$$

which sets up the following equations

$$\zeta\lambda = 1 \Rightarrow \lambda = \frac{1}{\zeta} \tag{C.12}$$

$$\eta_\alpha \kappa_1 + \frac{\eta_\alpha}{\zeta} + \zeta \kappa_1 + \eta_\alpha \kappa_4 \hat{n}_{\alpha j} \hat{n}_{\beta j} = 0 \tag{C.13}$$

$$\eta_\beta \kappa_2 + \frac{\eta_\beta}{\zeta} + \zeta \kappa_2 + \eta_\beta \kappa_3 \hat{n}_{\alpha j} \hat{n}_{\beta j} = 0 \tag{C.14}$$

$$\eta_\alpha \kappa_3 + \zeta \kappa_3 + \eta_\alpha \kappa_2 \hat{n}_{\alpha j} \hat{n}_{\beta j} = 0 \tag{C.15}$$

$$\eta_\beta \kappa_4 + \zeta \kappa_4 + \eta_\beta \kappa_1 \hat{n}_{\alpha j} \hat{n}_{\beta j} = 0 \tag{C.16}$$

Solving this system of equations for the remaining unknowns $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ results in

$$\kappa_1 = -\frac{\eta_\alpha(\eta_\beta + \zeta)}{\zeta(\eta_\alpha\eta_\beta - \eta_\alpha\eta_\beta(\hat{n}_{\alpha j}\hat{n}_{\beta j})^2 + \eta_\alpha\zeta + \eta_\beta\zeta + \zeta^2)} \tag{C.17}$$

$$\kappa_2 = -\frac{\eta_\beta(\eta_\alpha + \zeta)}{\zeta(\eta_\alpha\eta_\beta - \eta_\alpha\eta_\beta(\hat{n}_{\alpha j}\hat{n}_{\beta j})^2 + \eta_\alpha\zeta + \eta_\beta\zeta + \zeta^2)} \tag{C.18}$$

$$\kappa_3 = \frac{\eta_\alpha\eta_\beta(\hat{n}_{\alpha j}\hat{n}_{\beta j})}{\zeta(\eta_\alpha\eta_\beta - \eta_\alpha\eta_\beta(\hat{n}_{\alpha j}\hat{n}_{\beta j})^2 + \eta_\alpha\zeta + \eta_\beta\zeta + \zeta^2)} \tag{C.19}$$

$$\kappa_4 = \frac{\eta_\alpha\eta_\beta(\hat{n}_{\alpha j}\hat{n}_{\beta j})}{\zeta(\eta_\alpha\eta_\beta - \eta_\alpha\eta_\beta(\hat{n}_{\alpha j}\hat{n}_{\beta j})^2 + \eta_\alpha\zeta + \eta_\beta\zeta + \zeta^2)} \tag{C.20}$$

Thus,

$$\begin{aligned}
&\hat{r}_{\alpha\beta i}(A_\alpha + A_\beta)^{-1}_{ij}\hat{r}_{\alpha\beta j} \\
&\quad = \hat{r}_{\alpha\beta i}(\kappa_1\hat{n}_{\alpha i}\hat{n}_{\alpha j} + \kappa_2\hat{n}_{\beta i}\hat{n}_{\beta j} + \kappa_3\hat{n}_{\alpha i}\hat{n}_{\beta j} + \kappa_4\hat{n}_{\beta i}\hat{n}_{\alpha j} + \lambda\delta_{ij})\hat{r}_{\alpha\beta j} \\
&\quad = \kappa_1(\hat{r}_{\alpha\beta i}\hat{n}_{\alpha i})^2 + \kappa_2(\hat{r}_{\alpha\beta i}\hat{n}_{\beta i})^2 + \kappa_3\hat{r}_{\alpha\beta i}\hat{n}_{\alpha i}\hat{r}_{\alpha\beta j}\hat{n}_{\beta j} + \kappa_4\hat{r}_{\alpha\beta i}\hat{n}_{\beta i}\hat{r}_{\alpha\beta j}\hat{n}_{\alpha j} + \lambda
\end{aligned} \tag{C.21}$$

The range parameter can then be expressed as

$$\sigma(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta, \hat{\mathbf{r}}_{\alpha\beta}) = \left[\kappa_1(\hat{r}_{\alpha\beta i}\hat{n}_{\alpha i})^2 + \kappa_2(\hat{r}_{\alpha\beta i}\hat{n}_{\beta i})^2 + \kappa_3\hat{r}_{\alpha\beta i}\hat{n}_{\alpha i}\hat{r}_{\alpha\beta j}\hat{n}_{\beta j} + \kappa_4\hat{r}_{\alpha\beta i}\hat{n}_{\beta i}\hat{r}_{\alpha\beta j}\hat{n}_{\alpha j} + \lambda\right]^{-1/2} \tag{C.22}$$

which has been confirmed to match Eq. (15) of [21] (stated below for convenience)

$$\sigma = \sigma_0 \left[1 - \chi\left(\frac{\xi(\hat{\mathbf{r}}_{\alpha\beta}\cdot\hat{\mathbf{n}}_\alpha)^2 + \xi^{-1}(\hat{\mathbf{r}}_{\alpha\beta}\cdot\hat{\mathbf{n}}_\beta)^2 - 2\chi(\hat{\mathbf{r}}_{\alpha\beta}\cdot\hat{\mathbf{n}}_\alpha)(\hat{\mathbf{r}}_{\alpha\beta}\cdot\hat{\mathbf{n}}_\beta)(\hat{\mathbf{n}}_\alpha\cdot\hat{\mathbf{n}}_\beta)}{1 - \chi^2(\hat{\mathbf{n}}_\alpha\cdot\hat{\mathbf{n}}_\beta)^2}\right)\right]^{-1/2} \tag{C.23}$$

where

$$\sigma_0 = \sqrt{d_\alpha^2 + d_\beta^2}, \tag{C.24}$$

$$\chi = \left[\frac{(\ell_\alpha^2 - d_\alpha^2)(\ell_\beta^2 - d_\beta^2)}{(\ell_\alpha^2 + d_\beta^2)(\ell_\beta^2 + d_\alpha^2)} \right]^{1/2}, \tag{C.25}$$

and

$$\xi = \left[\frac{(\ell_\alpha^2 - d_\alpha^2)(\ell_\beta^2 + d_\alpha^2)}{(\ell_\beta^2 - d_\beta^2)(\ell_\alpha^2 + d_\beta^2)} \right]^{1/2}. \tag{C.26}$$

The maximum value of σ occurs when $\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta,$ and $\hat{\mathbf{r}}_{\alpha\beta}$ are all aligned. σ equals the minimum value of σ_0 when $\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta,$ and $\hat{\mathbf{r}}_{\alpha\beta}$ are all perpendicular to each other.

Bibliography

[1] Gil Ariel, Amit Rabani, Sivan Benisty, Jonathan D Partridge, Rasika M Harshey, and Avraham Be'er. Swarming bacteria migrate by Lévy Walk. *Nature Commun.*, 6:8396, 2015.

[2] Indrani Banerjee, Ravindra C. Pangule, and Ravi S. Kane. Antifouling coatings: Recent developments in the design of surfaces that prevent fouling by proteins, bacteria, and marine organisms. *Adv. Materials*, 23(6):690–718, 2011.

[3] Kenneth W. Bayles. The biological role of death and lysis in biofilm development. *Nature Rev. Microbiol.*, 5(9):721, 2007.

[4] Avraham Be'er, Shinji K. Strain, Roberto A. Hernândez, Eshel Ben-Jacob, and E.-L. Florin. Periodic reversals in *Paenibacillus dendritiformis* swarming. *J. Bacteriol.*, 195(12):2709–2717, 2013.

[5] Howard C. Berg and Douglas A. Brown. Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature*, 239:500 – 504, 1972.

[6] Veysel Berk, Jiunn C. N. Fong, Graham T. Dempsey, Omer N. Develioglu, Xiaowei Zhuang, Jan Liphardt, Fitnat H. Yildiz, and Steven Chu. Molecular architecture and assembly principles of *Vibrio cholerae* biofilms. *Science*, 337(6091):236–239, 2012.

[7] Allison P. Berke, Linda Turner, Howard C. Berg, and Eric Lauga. Hydrodynamic attraction of swimming microorganisms by surfaces. *Phys. Rev. Lett.*, 101:038102, 2008.

[8] Bruce J. Berne and Philip Pechukas. Gaussian model potentials for molecular interactions. *J. Chem. Phys.*, 56(8):4213–4216, 1972.

[9] Farzan Beroz, Jing Yan, Yigal Meir, Benedikt Sabass, Howard A. Stone, Bonnie L. Bassler, and Ned S. Wingreen. Verticalization of bacterial biofilms. *Nature Phys.*, 14:954–960, 2018.

[10] Andrew E. Blanchard and Ting Lu. Bacterial social interactions drive the emergence of differential spatial colony structures. *BMC Sys. Bio.*, 9(1):59, 2015.

[11] Thomas Böttcher, Hunter L Elliott, and Jon Clardy. Dynamics of snake-like swarming behavior of *Vibrio alginolyticus*. *Biophys. J.*, 110(4):981–992, 2016.

[12] Vasil Bratanov, Frank Jenko, and Erwin Frey. New class of turbulence in active fluids. *Proc. Natl. Acad. Sci. U.S.A.*, 112(49):15048–15053, 2015.

[13] Kevin Cahill. *Physical Mathematics.* Cambridge University Press, 2013.

[14] Cinzia Calvio, Francesco Celandroni, Emilia Ghelardi, Giuseppe Amati, Sara Salvetti, Fabrizio Ceciliani, Alessandro Galizzi, and Sonia Senesi. Swarming differentiation and swimming motility in *Bacillus subtilis* are controlled by *swrA*, a newly identified dicistronic operon. *J. Bacteriol.*, 187(15):5356–5366, 2005.

[15] Manuel Campos, Ivan V. Surovtsev, Setsu Kato, Ahmad Paintdakhi, Bruno Beltran, Sarah E. Ebmeier, and Christine Jacobs-Wagner. A constant size extension drives bacterial cell size homeostasis. *Cell*, 159(6):1433–1446, 2014.

[16] Michelle L. Carman, Thomas G. Estes, Adam W. Feinberg, James F. Schumacher, Wade Wilkerson, Leslie H. Wilson, Maureen E. Callow, James A. Callow, and Anthony B. Brennan. Engineered antifouling microtopographies – correlating wettability with cell attachment. *Biofouling*, 22(1):11–21, 2006.

[17] Michael E. Cates and Elsen Tjhung. Theories of binary fluid mixtures: from phase-separation kinetics to active emulsions. *J. Fluid Mech.*, 836:P1, 2018.

[18] Chong Chen, Song Liu, Xia-qing Shi, Hugues Chaté, and Yilin Wu. Weak synchronization and large-scale collective oscillation in dense bacterial suspensions. *Nature*, 542(7640):210, 2017.

[19] Xiao Chen, Xu Dong, Avraham Be'er, Harry L. Swinney, and H. P. Zhang. Scale-invariant correlations in dynamic bacterial clusters. *Phys. Rev. Lett.*, 108:148101, 2012.

[20] R. A. N. Chmielewski and J. F. Frank. Biofilm formation and control in food processing facilities. *Compr. Rev. Food Sci.*, 2(1):22–32, 2003.

[21] Douglas J. Cleaver, Christopher M. Care, Michael P. Allen, and Maureen P. Neal. Extension and generalization of the Gay-Berne potential. *Phys. Rev. E*, 54:559–567, 1996.

[22] Jacinta C. Conrad. Physics of bacterial near-surface motility using flagella and type IV pili: implications for biofilm formation. *Res. Microbiol.*, 163(9):619 – 629, 2012.

[23] Benjamin J. Cooley, Sheri Dellos-Nolan, Numa Dhamani, Ross Todd, William Waller, Daniel Wozniak, and Vernita D. Gordon. Asymmetry and inequity in the inheritance of a bacterial adhesive. *New J. Phys.*, 18(4):045019, 2016.

[24] Matthew F. Copeland and Douglas B. Weibel. Bacterial swarming: a model system for studying dynamic self-assembly. *Soft Matter*, 5(6):1174–1187, 2009.

[25] J. W. Costerton, Philip S. Stewart, and E. P. Greenberg. Bacterial biofilms: A common cause of persistent infections. *Science*, 284(5418):1318–1322, 1999.

[26] Christopher R. Cotter, Heinz-Bernd Schüttler, Oleg A. Igoshin, and Lawrence J. Shimkets. Data-driven modeling reveals cell behaviors controlling self-organization during *Myxococcus xanthus* development. *Proc. Natl. Acad. Sci. U.S.A.*, 114(23):E4592–E4601, 2017.

[27] Ruth Daniels, Jos Vanderleyden, and Jan Michiels. Quorum sensing and swarming migration in bacteria. *FEMS Microbiol. Rev.*, 28(3):261–289, 2004.

[28] Nicholas C. Darnton, Linda Turner, Svetlana Rojevsky, and Howard C. Berg. Dynamics of bacterial swarming. *Biophys. J.*, 98(10):2082 – 2090, 2010.

[29] Morgan Delarue, Jörn Hartung, Carl Schreck, Pawel Gniewek, Lucy Hu, Stephan Herminghaus, and Oskar Hallatschek. Self-driven jamming in growing microbial populations. *Nature Phys.*, 12(8):762, 2016.

[30] R. Di Leonardo, L. Angelani, D. Dell'Arciprete, G. Ruocco, V. Iebba, S. Schippa, M. P. Conte, F. Mecarini, F. De Angelis, and E. Di Fabrizio. Bacterial ratchet motors. *Proc. Natl. Acad. Sci. U.S.A.*, 107(21):9541–9545, 2010.

[31] Christopher Dombrowski, Luis Cisneros, Sunita Chatkaew, Raymond E. Goldstein, and John O. Kessler. Self-concentration and large-scale coherence in bacterial dynamics. *Phys. Rev. Lett.*, 93:098103, 2004.

[32] Rodney M. Donlan. Biofilm formation: A clinically relevant microbiological process. *Clin. Infect. Dis.*, 33(8):1387–1392, 2001.

[33] Rodney M. Donlan and J. William Costerton. Biofilms: Survival mechanisms of clinically relevant microorganisms. *Clin. Microbiol. Rev.*, 15(2):167–193, 2002.

[34] Amin Doostmohammadi, Michael F. Adamer, Sumesh P. Thampi, and Julia M. Yeomans. Stabilization of active matter by flow-vortex lattices and defect ordering. *Nature Commun.*, 7:10557, 2016.

[35] Knut Drescher, Jörn Dunkel, Luis H. Cisneros, Sujoy Ganguly, and Raymond E. Goldstein. Fluid dynamics and noise in bacterial cell–cell and cell–surface scattering. *Proc. Natl. Acad. Sci. U.S.A.*, 108(27):10940–10945, 2011.

[36] Knut Drescher, Jörn Dunkel, Carey D. Nadell, Sven van Teeffelen, Ivan Grnja, Ned S. Wingreen, Howard A. Stone, and Bonnie L. Bassler. Architectural transitions in *Vibrio cholerae* biofilms at single-cell resolution. *Proc. Natl. Acad. Sci. U.S.A.*, 113(14):E2066–E2072, 2016.

[37] Knut Drescher, Yi Shen, Bonnie L. Bassler, and Howard A. Stone. Biofilm streamers cause catastrophic disruption of flow with consequences for environmental and medical systems. *Proc. Natl. Acad. Sci. U.S.A.*, 110(11):4345–4350, 2013.

[38] Jörn Dunkel, Sebastian Heidenreich, Knut Drescher, Henricus H. Wensink, Markus Bär, and Raymond E. Goldstein. Fluid dynamics of bacterial turbulence. *Phys. Rev. Lett.*, 110:228102, 2013.

[39] Marie-Cécilia Duvernoy, Thierry Mora, Maxime Ardré, Vincent Croquette, David Bensimon, Catherine Quilliet, Jean-Marc Ghigo, Martial Balland, Christophe Beloin, Sigolène Lecuyer, et al. Asymmetric adhesion of rod-shaped bacteria controls microcolony morphogenesis. *Nature Commun.*, 9(1):1120, 2018.

[40] Kirill Efimenko, John Finlay, Maureen E. Callow, James A. Callow, and Jan Genzer. Development and testing of hierarchically wrinkled coatings for marine antifouling. *ACS Appl. Materials Inter.*, 1(5):1031–1040, 2009.

[41] Perry W. Ellis, Daniel J. G. Pearce, Ya-Wen Chang, Guillermo Goldsztein, Luca Giomi, and Alberto Fernandez-Nieves. Curvature-induced defect unbinding and dynamics in active nematic toroids. *Nature Phys.*, 14:85–90, 2017.

[42] Alexander K. Epstein, Tak-Sing Wong, Rebecca A. Belisle, Emily Marie Boggs, and Joanna Aizenberg. Liquid-infused structured surfaces with exceptional anti-biofouling performance. *Proc. Natl. Acad. Sci. U.S.A.*, 109(33):13182–13187, 2012.

[43] F. D. C. Farrell, O. Hallatschek, D. Marenduzzo, and B. Waclaw. Mechanically driven growth of quasi-two-dimensional microbial colonies. *Phys. Rev. Lett.*, 111:168101, 2013.

[44] Nicholas I. Fisher. *Statistical analysis of circular data.* Cambridge University Press, 1995.

[45] H.-C. Flemming. Biofouling in water systems—cases, causes and countermeasures. *Appl. Microbiol. Biotechnol.*, 59(6):629–640, 2002.

[46] Hans-Curt Flemming, Jost Wingender, Ulrich Szewzyk, Peter Steinberg, Scott A Rice, and Staffan Kjelleberg. Biofilms: an emergent form of bacterial life. *Nature Rev. Microbiol.*, 14(9):563, 2016.

[47] Jiunn C. N. Fong, Kevin Karplus, Gary K. Schoolnik, and Fitnat H. Yildiz. Identification and characterization of RbmA, a novel protein required for the development of rugose colony morphology and biofilm structure in *Vibrio cholerae. J. Bacteriol.*, 188(3):1049–1059, 2006.

[48] Jiunn C. N. Fong, Andrew Rogers, Alicia K. Michael, Nicole C. Parsley, William-Cole Cornell, Yu-Cheng Lin, Praveen K. Singh, Raimo Hartmann, Knut Drescher, Evgeny Vinogradov, Lars E. P. Dietrich, Carrie L. Partch, and Fitnat H. Yildiz. Structural dynamics of RbmA governs plasticity of *Vibrio cholerae* biofilms. *eLife*, 6:e26163, 2017.

[49] P. D. Frymier, R. M. Ford, H. C. Berg, and P. T. Cummings. Three-dimensional tracking of motile bacteria near a solid planar surface. *Proc. Natl. Acad. Sci. U.S.A.*, 92(13):6195–6199, 1995.

[50] C. A. Fux, J. W. Costerton, P. S. Stewart, and P. Stoodley. Survival strategies of infectious biofilms. *Trends Microbiol.*, 13(1):34 – 40, 2005.

[51] Peter Galajda, Juan Keymer, Paul Chaikin, and Robert Austin. A wall of funnels concentrates swimming bacteria. *J. Bacteriol.*, 189(23):8704–8707, 2007.

[52] Pushpita Ghosh, Jagannath Mondal, Eshel Ben-Jacob, and Herbert Levine. Mechanically-driven phase separation in a growing bacterial colony. *Proc. Natl. Acad. Sci. U.S.A.*, 112(17):E2166–E2173, 2015.

[53] Francesco Ginelli, Fernando Peruani, Markus Bär, and Hugues Chaté. Large-scale collective properties of self-propelled rods. *Phys. Rev. Lett.*, 104:184502, 2010.

[54] David S. Glass and Ingmar H. Riedel-Kruse. A synthetic bacterial cell-cell adhesion toolbox for programming multicellular morphologies and patterns. *Cell*, 174(3):649–658, 2018.

[55] Matthew A. A. Grant, Bartłomiej Wacław, Rosalind J. Allen, and Pietro Cicuta. The role of mechanical forces in the planar-to-bulk transition in growing *Escherichia coli* microcolonies. *J. R. Soc. Inter.*, 11(97):20140400, 2014.

[56] Michael F. Hagan and Aparna Baskaran. Emergent self-organization in active materials. *Curr. Opin. Cell Bio.*, 38:74 – 80, 2016.

[57] Luanne Hall-Stoodley, J. William Costerton, and Paul Stoodley. Bacterial biofilms: from the natural environment to infectious diseases. *Nature Rev. Microbiol.*, 2:95–108, 2004.

[58] Lina Hamouche, Soumaya Laalami, Adrian Daerr, Solène Song, I. Barry Holland, Simone J. Séror, Kassem Hamze, and Harald Putzer. *Bacillus subtilis* swarmer cells lead the swarm, multiply, and generate a trail of quiescent descendants. *mBio*, 8(1), 2017.

[59] Kassem Hamze, Sabine Autret, Krzysztof Hinc, Soumaya Laalami, Daria Julkowska, Romain Briandet, Margareth Renault, Cédric Absalon, I. Barry Holland, Harald Putzer, and Simone J. Séror. Single-cell analysis in situ in a *Bacillus subtilis* swarming community identifies distinct spatially separated subpop-

ulations differentially expressing hag (flagellin), including specialized swarmers. *Microbiol.*, 157(9):2456–2469, 2011.

[60] Yilong Han, Ahmed Alsayed, Maurizio Nobili, and Arjun G. Yodh. Quasi-two-dimensional diffusion of single ellipsoids: Aspect ratio and confinement effects. *Phys. Rev. E*, 80:011403, 2009.

[61] Raimo Hartmann, Praveen K. Singh, Philip Pearce, Rachel Mok, Boya Song, Francisco Díaz-Pascual, Jörn Dunkel, and Knut Drescher. Emergence of three-dimensional order and structure in growing biofilms. *Nature Phys.*, 15:251–256, 2019.

[62] Ferdi L. Hellweger, Robert J. Clegg, James R. Clark, Caroline M. Plugge, and Jan-Ulrich Kreft. Advancing microbial sciences by individual-based modelling. *Nature Rev. Microbiol.*, 14(7):461, 2016.

[63] H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1):53–60, 1981.

[64] Bella Ilkanaiv, Daniel B. Kearns, Gil Ariel, and Avraham Be'er. Effect of cell aspect ratio on swarming bacteria. *Phys. Rev. Lett.*, 118:158002, 2017.

[65] Pahala Gedara Jayathilake, Prashant Gupta, Bowen Li, Curtis Madsen, Oluwole Oyebamiji, Rebeca González-Cabaleiro, Steve Rushton, Ben Bridgens, David Swailes, Ben Allen, A. Stephen McGough, Paolo Zuliani, Irina Dana Ofiteru, Darren Wilkinson, Jinju Chen, and Tom Curtis. A mechanistic Individual-based Model of microbial communities. *PLOS ONE*, 12(8):1–26, 2017.

[66] Hannah Jeckel, Eric Jelli, Raimo Hartmann, Praveen K. Singh, Rachel Mok, Jan Frederik Totz, Lucia Vidakovic, Bruno Eckhardt, Jörn Dunkel, and Knut Drescher. Learning the space-time phase diagram of bacterial swarm expansion. *Proc. Natl. Acad. Sci. U.S.A.*, 116(5):1489–1494, 2019.

[67] Daria Julkowska, Michal Obuchowski, I. Barry Holland, and Simone. J. Séror. Comparative analysis of the development of swarming communities of *Bacillus subtilis* 168 and a natural wild type: Critical effects of surfactin and the composition of the medium. *J. Bacteriol.*, 187(1):65–76, 2005.

[68] Suckjoon Jun, Fangwei Si, Rami Pugatch, and Matthew Scott. Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review. *Rep. Prog. Phys.*, 81(5):056601, 2018.

[69] Suckjoon Jun and Sattar Taheri-Araghi. Cell-size maintenance: universal strategy revealed. *Trends Microbiol.*, 23(1):4–6, 2015.

[70] Sungkyu Jung. Generating von Mises Fisher distribution on the unit sphere(S2). *https://www.stat.pitt.edu/sungkyu/software/randvonMisesFisher3.pdf*, 2009.

[71] Ece Karatan and Paula Watnick. Signals, regulatory networks, and materials that build and break bacterial biofilms. *Microbiol. Mol. Bio. Rev.*, 73(2):310–347, 2009.

[72] Wan-Ju Ke, Yi-Huang Hsueh, Yu-Chieh Cheng, Chih-Ching Wu, and Shih-Tung Liu. Water surface tension modulates the swarming mechanics of *Bacillus subtilis*. *Front. Microbiol.*, 6:1017, 2015.

[73] Daniel B Kearns. A field guide to bacterial swarming motility. *Nature Rev. Microbiol.*, 8(9):634, 2010.

[74] Daniel B. Kearns, Frances Chu, Rivka Rudner, and Richard Losick. Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Mol. Microbiol.*, 52(2):357–369, 2004.

[75] Daniel B. Kearns and Richard Losick. Swarming motility in undomesticated *Bacillus subtilis*. *Mol. Microbiol.*, 49(3):581–590, 2003.

[76] Daniel B. Kearns and Richard Losick. Cell population heterogeneity during growth of *Bacillus subtilis*. *Genes Dev.*, 19(24):3083–3094, 2005.

[77] Kyunghoon Kim and Jung Kyung Kim. Visualization of biosurfactant film flow in a *Bacillus subtilis* swarm colony on an agar plate. *Int. J. Mol. Sci.*, 16(9):20225–20238, 2015.

[78] Kasper N. Kragh, Jaime B. Hutchison, Gavin Melaugh, Chris Rodesney, Aled E. L. Roberts, Yasuhiko Irie, Peter Ø. Jensen, Stephen P. Diggle, Rosalind J. Allen, Vernita Gordon, and Thomas Bjarnsholt. Role of multicellular aggregates in biofilm formation. *mBio*, 7(2), 2016.

[79] C. Ganesh Kumar and S. K. Anand. Significance of microbial biofilms in food industry: a review. *Int. J. Food Microbiol.*, 42(1):9 – 27, 1998.

[80] Guillaume Lambert, Andrew Bergman, Qiucen Zhang, David Bortz, and Robert Austin. Physics of biofilms: the initial stages of biofilm formation and dynamics. *New J. Phys.*, 16(4):045005, 2014.

[81] Laurent A. Lardon, Brian V. Merkey, Sónia Martins, Andreas Dötsch, Cristian Picioreanu, Jan-Ulrich Kreft, and Barth F. Smets. iDynoMiCS: next-generation individual-based modelling of biofilms. *Environ. Microbiol.*, 13(9):2416–2434, 2011.

[82] Sigolene Lecuyer, Roman Stocker, and Roberto Rusconi. Focus on the physics of biofilms. *New J. Phys.*, 17(3):030401, 2015.

[83] Guanglai Li, James Bensson, Liana Nisimova, Daniel Munger, Panrapee Mahautmr, Jay X. Tang, Martin R. Maxey, and Yves V. Brun. Accumulation of swimming bacteria near a solid surface. *Phys. Rev. E*, 84:041932, 2011.

[84] Guanglai Li and Jay X. Tang. Accumulation of microswimmers near a surface mediated by collision and rotational brownian motion. *Phys. Rev. Lett.*, 103:078101, 2009.

[85] Jintao Liu, Rosa Martinez-Corral, Arthur Prindle, Dong-yeon D. Lee, Joseph Larkin, Marçal Gabalda-Sagarra, Jordi Garcia-Ojalvo, and Gürol M. Süel. Coupling between distant biofilms and emergence of nutrient time-sharing. *Science*, 356(6338):638–642, 2017.

[86] Jintao Liu, Arthur Prindle, Jacqueline Humphries, Marçal Gabalda-Sagarra, Munehiro Asally, D. Lee Dong-yeon, San Ly, Jordi Garcia-Ojalvo, and Gürol M. Süel. Metabolic co-dependence gives rise to collective oscillations within biofilms. *Nature*, 523(7562):550, 2015.

[87] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *J. Mach. Learn. Res.*, pages 2579–2605, 2008.

[88] Manuel Maestre-Reyna, Wen-Jin Wu, and Andrew H.-J. Wang. Structural insights into RbmA, a biofilm scaffolding protein of *V. Cholerae*. *PLOS ONE*, 8(12), 2013.

[89] Berenike Maier and Gerard C. L. Wong. How bacteria use type IV pili machinery on surfaces. *Trends Microbiol.*, 23(12):775 – 788, 2015.

[90] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, and R. Aditi Simha. Hydrodynamics of soft active matter. *Rev. Mod. Phys.*, 85:1143–1189, 2013.

[91] M. R. Mattei, L. Frunzo, B. D'acunto, Y. Pechaud, F. Pirozzi, and G. Esposito. Continuum and discrete approach in modeling biofilm development and structure: a review. *J. Math. Bio.*, 76(4):945–1003, 2018.

[92] Marco G. Mazza. The physics of biofilms—an introduction. *J. Phys. D.: Appl. Phys.*, 49(20):203001, 2016.

[93] Carl D. Meyer, editor. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[94] Rachel Mok, Jörn Dunkel, and Vasily Kantsler. Geometric control of bacterial surface accumulation. *arXiv preprint arXiv:1903.00926*, 2019.

[95] Mehdi Molaei, Michael Barry, Roman Stocker, and Jian Sheng. Failed escape: Solid surfaces prevent tumbling of *Escherichia coli*. *Phys. Rev. Lett.*, 113:068103, 2014.

[96] Robert Nerenberg. The membrane-biofilm reactor (MBfR) as a counter-diffusional biofilm process. *Curr. Opin. Biotech.*, 38:131 – 136, 2016.

[97] T R Neu. Significance of bacterial surface-active compounds in interaction of bacteria with interfaces. *Microbiol. Rev.*, 60(1):151 – 166, 1996.

[98] Daiki Nishiguchi, Igor S. Aranson, Alexey Snezhko, and Andrey Sokolov. Engineering bacterial vortex lattice via direct laser lithography. *Nature Commun.*, 9(1):4486, 2018.

[99] Lars Nyland, Mark Harris, and Jan Prins. Fast N-body simulation with CUDA. In Hubert Nguyen, editor, *GPU Gems 3*, chapter 31, pages 677 – 696. Addison-Wesley Professional, Upper Saddle River, 2007.

[100] George O'Toole, Heidi B. Kaplan, and Roberto Kolter. Biofilm formation as microbial development. *Annual Review of Microbiology*, 54(1):49–79, 2000.

[101] Joyce E. Patrick and Daniel B. Kearns. Swarming motility and the control of master regulators of flagellar biosynthesis. *Mol. Microbiol.*, 83(1):14–23, 2012.

[102] Leonid Pavlovsky, John G. Younger, and Michael J. Solomon. In situ rheology of *Staphylococcus epidermidis* bacterial biofilms. *Soft Matter*, 9:122–131, 2013.

[103] Philip Pearce, Boya Song, Dominic J. Skinner, Rachel Mok, Raimo Hartmann, Praveen K. Singh, Jeffrey S. Oishi, Knut Drescher, and Jörn Dunkel. Flow-induced symmetry breaking in growing bacterial biofilms. *bioRxiv*, 2019.

[104] David Perera-Costa, José Morales Bruque, María Luisa González-Martín, Antonio Cándido Gómez-García, and Virginia Vadillo-Rodríguez. Studying the influence of surface topography on bacterial adhesion using spatially organized microtopographic surface patterns. *Langmuir*, 30(16):4633–4641, 2014.

[105] Alexandre Persat, Carey D. Nadell, Minyoung Kevin Kim, Francois Ingremeau, Albert Siryaporn, Knut Drescher, Ned S. Wingreen, Bonnie L. Bassler, Zemer Gitai, and Howard A. Stone. The mechanical world of bacteria. *Cell*, 161:988–997, 2015.

[106] Cristian Picioreanu, Mark C. M. van Loosdrecht, and Joseph J. Heijnen. Mathematical modeling of biofilm structure with a hybrid differential-discrete cellular automaton approach. *Biotech. Bioeng.*, 58(1):101–116, 1998.

[107] Cristian Picioreanu, Mark C. M. van Loosdrecht, and Joseph J. Heijnen. A new combined differential-discrete cellular automaton approach for biofilm modeling: Application for growth in gel beads. *Biotech. Bioeng.*, 57(6):718–731, 1998.

[108] Liyan Ping, Yilin Wu, Basarab G. Hosu, Jay X. Tang, and Howard C. Berg. Osmotic pressure in a bacterial swarm. *Biophys. J.*, 107(4):871 – 878, 2014.

[109] Nikodem J. Popławski, Abbas Shirinifard, Maciej Swat, and James A. Glazier. Simulation of single-species bacterial-biofilm growth using the Glazier-Graner-Hogeweg model and the CompuCell3D modeling environment. *Math. Biosci. Eng.*, 5(2):355, 2008.

[110] Lars D. Renner and Douglas B. Weibel. Physicochemical regulation of biofilm formation. *MRS Bulletin*, 36(5):347–355, 2011.

[111] Christopher A. Rodesney, Brian Roman, Numa Dhamani, Benjamin J. Cooley, Parag Katira, Ahmed Touhami, and Vernita D. Gordon. Mechanosensing of shear by *Pseudomonas aeruginosa* leads to increased levels of the cyclic-di-GMP signal initiating biofilm development. *Proc. Natl. Acad. Sci. U.S.A.*, 114(23):5906–5911, 2017.

[112] Shmuel M. Rubinstein, Ilana Kolodkin-Gal, Anna Mcloon, Liraz Chai, Roberto Kolter, Richard Losick, and David A. Weitz. Osmotic pressure can regulate matrix gene expression in *Bacillus subtilis*. *Mol. Microbiol.*, 86(2):426–436, 2012.

[113] Timothy J. Rudge, Fernán Federici, Paul J. Steiner, Anton Kan, and Jim Haseloff. Cell polarity-driven instability generates self-organized, fractal patterning of cell layers. *ACS Synth. Biol.*, 2(12):705–714, 2013.

[114] Timothy J. Rudge, Paul J. Steiner, Andrew Phillips, and Jim Haseloff. Computational modeling of synthetic microbial biofilms. *ACS Synth. Biol.*, 1(8):345–352, 2012.

[115] John T. Sauls, Dongyang Li, and Suckjoon Jun. Adder and a coarse-grained approach to cell size homeostasis in bacteria. *Curr. Opin. Cell Bio.*, 38:38 – 44, 2016.

[116] Agnese Seminara, Thomas E. Angelini, James N. Wilking, Hera Vlamakis, Senan Ebrahim, Roberto Kolter, David A. Weitz, and Michael P. Brenner. Osmotic spreading of *Bacillus subtilis* biofilms driven by an extracellular matrix. *Proc. Natl. Acad. Sci. U.S.A.*, 109(4):1116–1121, 2012.

[117] Marina Sidortsov, Yakov Morgenstern, and Avraham Be'er. Role of tumbling in bacterial swarming. *Phys. Rev. E*, 96:022407, 2017.

[118] Praveen K. Singh, Sabina Bartalomej, Raimo Hartmann, Hannah Jeckel, Lucia Vidakovic, Carey D. Nadell, and Knut Drescher. *Vibrio cholerae* combines individual and collective sensing to trigger biofilm dispersal. *Curr. Bio.*, 27(21):3359–3366, 2017.

[119] O. Sipos, K. Nagy, R. Di Leonardo, and P. Galajda. Hydrodynamic trapping of swimming bacteria by convex walls. *Phys. Rev. Lett.*, 114:258104, 2015.

[120] William P. J. Smith, Yohan Davit, James M. Osborne, Wook Kim, Kevin R. Foster, and Joe M. Pitt-Francis. Cell morphology drives spatial patterning in microbial communities. *Proc. Natl. Acad. Sci. U.S.A.*, 114(3):E280–E286, 2017.

[121] Andrey Sokolov, Mario M. Apodaca, Bartosz A. Grzybowski, and Igor S. Aranson. Swimming bacteria power microscopic gears. *Proc. Natl. Acad. Sci. U.S.A.*, 107(3):969–974, 2010.

[122] Andrey Sokolov, Igor S. Aranson, John O. Kessler, and Raymond E. Goldstein. Concentration dependence of the collective dynamics of swimming bacteria. *Phys. Rev. Lett.*, 98:158102, 2007.

[123] Kwangmin Son, Douglas R. Brumley, and Roman Stocker. Live from under the lens: exploring microbial motility with dynamic imaging and microfluidics. *Nature Rev. Microbiol.*, 13(12):761, 2015.

[124] Saverio E. Spagnolie and Eric Lauga. Hydrodynamics of self-propulsion near a boundary: predictions and accuracy of far-field approximations. *J. Fluid Mech.*, 700:105–147, 2012.

[125] Shilpi Srivastava and Atul Bhargava. Biofilms and human health. *Biotechnol. Lett.*, 38(1):1–22, 2016.

[126] Philip S. Stewart and Michael J. Franklin. Physiological heterogeneity in biofilms. *Nature Rev. Microbiol.*, 6(3):199, 2008.

[127] Maciej H. Swat, Gilberto L. Thomas, Julio M. Belmonte, Abbas Shirinifard, Dimitrij Hmeljak, and James A. Glazier. Multi-scale modeling of tissues using CompuCell3D. In Anand R. Asthagiri and Adam P. Arkin, editors, *Computational Methods in Cell Biology*, volume 110 of *Methods in Cell Biology*, pages 325 366. Academic Press, 2012.

[128] Sattar Taheri-Araghi, Serena Bradde, John T. Sauls, Norbert S. Hill, Petra Anne Levin, Johan Paulsson, Massimo Vergassola, and Suckjoon Jun. Cell-size control and homeostasis in bacteria. *Curr. Bio.*, 25:385–391, 2015.

[129] Jennifer K. Teschler, David Zamorano-Sánchez, Andrew S. Utada, Christopher J. A. Warner, Gerard C. L. Wong, Roger G. Linington, and Fitnat H. Yildiz. Living in the matrix: assembly and control of *Vibrio cholerae* biofilms. *Nature Rev. Microbiol.*, 13:255–268, 2015.

[130] M. Mercedes Tirado, Carmen López Martínez, and José García de la Torre. Comparison of theories for the translational and rotational diffusion coefficients of rod-like macromolecules. Application to short DNA fragments. *J. Chem. Phys.*, 81(4):2047–2052, 1984.

[131] Miguel Trejo, Carine Douarche, Virginie Bailleux, Christophe Poulard, Sandrine Mariot, Christophe Regeard, and Eric Raspaud. Elasticity and wrinkled morphology of *Bacillus subtilis* pellicles. *Proc. Natl. Acad. Sci. U.S.A.*, 110(6):2011–2016, 2013.

[132] Linda Turner, William S. Ryu, and Howard C. Berg. Real-time imaging of fluorescent flagellar filaments. *J. Bacteriol.*, 182(10):2793–2801, 2000.

[133] Linda Turner, Rongjing Zhang, Nicholas C. Darnton, and Howard C. Berg. Visualization of flagella during bacterial swarming. *J. Bacteriol.*, 192(13):3259–3267, 2010.

[134] Hannah H. Tuson and Douglas B. Weibel. Bacteria–surface interactions. *Soft Matter*, 9:4368–4380, 2013.

[135] Jordi van Gestel, Franz J. Weissing, Oscar P. Kuipers, and Akos T. Kovács. Density of founder cells affects spatial pattern formation and cooperation in *Bacillus subtilis* biofilms. *ISME J.*, 8(10):2069, 2014.

[136] M. C. van Loosdrecht, J. Lyklema, W. Norde, and A. J. Zehnder. Influence of interfaces on microbial activity. *Microbiol. Rev.*, 54(1):75–87, 1990.

[137] Lucia Vidakovic, Praveen K. Singh, Raimo Hartmann, Carey D. Nadell, and Knut Drescher. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nature Microbiol.*, 3(1):26, 2018.

[138] Dmitri Volfson, Scott Cookson, Jeff Hasty, and Lev S. Tsimring. Biomechanical ordering of dense cell populations. *Proc. Natl. Acad. Sci. U.S.A.*, 105(40):15346–15351, 2008.

[139] Qi Wang and Tianyu Zhang. Review of mathematical models for biofilms. *Solid State Commun.*, 150(21):1009 – 1022, 2010.

[140] Henricus H. Wensink, Jörn Dunkel, Sebastian Heidenreich, Knut Drescher, Raymond E. Goldstein, Hartmut Löwen, and Julia M. Yeomans. Meso-scale turbulence in living fluids. *Proc. Natl. Acad. Sci. U.S.A.*, 109(36):14308–14313, 2012.

[141] Julian W. T. Wimpenny and Ric Colasanti. A unifying hypothesis for the structure of microbial biofilms based on cellular automaton models. *FEMS Microbiol. Eco.*, 22(1):1–16, 1997.

[142] Hugo Wioland, Francis G. Woodhouse, Jörn Dunkel, and Raymond E. Goldstein. Ferromagnetic and antiferromagnetic order in bacterial vortex lattices. *Nature Phys.*, 12:341 – 345, 2016.

[143] Tak-Sing Wong, Sung Hoon Kang, Sindy K. Y. Tang, Elizabeth J. Smythe, Benjamin D. Hatton, Alison Grinthal, and Joanna Aizenberg. Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity. *Nature*, 477:443 – 447, 2011.

[144] Roger D. Woods and David S. Saxon. Diffuse surface optical model for nucleon-nuclei scattering. *Phys. Rev.*, 95:577–578, 1954.

[145] Xiao-Lun Wu and Albert Libchaber. Particle diffusion in a quasi-two-dimensional bacterial bath. *Phys. Rev. Lett.*, 84:3017–3020, 2000.

[146] Yilin Wu and Howard C. Berg. Water reservoir maintained by cell growth fuels the spreading of a bacterial swarm. *Proc. Natl. Acad. Sci. U.S.A.*, 109(11):4128–4133, 2012.

[147] Jing Yan, Andrew G. Sharo, Howard A. Stone, Ned S. Wingreen, and Bonnie L. Bassler. *Vibrio cholerae* biofilm growth program and architecture revealed by single-cell live imaging. *Proc. Natl. Acad. Sci. U.S.A.*, 113(36):E5337–E5343, 2016.

[148] Zhihong You, Daniel J. G. Pearce, Anupam Sengupta, and Luca Giomi. Geometry and mechanics of microdomains in growing bacterial colonies. *Phys. Rev. X*, 8:031065, 2018.

[149] H. P. Zhang, Avraham Be'er, E.-L. Florin, and Harry L. Swinney. Collective motion and density fluctuations in bacterial colonies. *Proc. Natl. Acad. Sci. U.S.A.*, 107(31):13626–13630, 2010.

[150] Kun Zhao, Boo Shan Tseng, Bernard Beckerman, Fan Jin, Maxsim L. Gibiansky, Joe J. Harrison, Erik Luijten, Matthew R Parsek, and Gerard C. L. Wong. Psl trails guide exploration and microcolony formation in *Pseudomonas aeruginosa* biofilms. *Nature*, 497(7449):388, 2013.

[151] Shuang Zhou, Andrey Sokolov, Oleg D. Lavrentovich, and Igor S. Aranson. Living liquid crystals. *Proc. Natl. Acad. Sci. U.S.A.*, 111(4):1265–1270, 2014.