Emergence of three-dimensional order and structure in growing biofilms

Raimo Hartmann [1,2], Praveen K. Singh [1,5], Philip Pearce [3,5], Rachel Mok [3,4,5], Boya Song [3],
Francisco Díaz-Pascual [1], Jörn Dunkel [3]* and Knut Drescher [1,2]*

Surface-attached bacterial biofilms are self-replicating active liquid crystals and the dominant form of bacterial life on Earth[1-4]. In conventional liquid crystals and solid-state materials, the interaction potentials between the molecules that comprise the system determine the material properties. However, for growth-active biofilms it is unclear whether potential-based descriptions can account for the experimentally observed morphologies, and which potentials would be relevant. Here, we have overcome previous limitations of single-cell imaging techniques[5,6] to reconstruct and track all individual cells inside growing three-dimensional biofilms with up to 10,000 individuals. Based on these data, we identify, constrain and provide a microscopic basis 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. Furthermore, we show how external fluid flows control the microscopic structure and three-dimensional morphology of biofilms. Our 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. 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.

Vibrio cholerae cells can swim through liquids as isolated individuals, but are more commonly attached to surfaces, where they grow into clonal colonies termed biofilms, with reproducible spatial organization, global morphology and cellular arrangements[7,8]. 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[1-4]. As these active nematic systems operate far from thermodynamic equilibrium[9], there are no relevant fundamental conservation laws known that could be used to characterize the biofilm developmental dynamics. To achieve a detailed qualitative and quantitative understanding of such biologically ubiquitous yet physically exotic bacterial communities, we developed new experimental imaging and image analysis techniques for obtaining high spatiotemporal-resolution data of the biofilm developmental process up to 10^4 cells, representing mid-sized biofilm microcolonies that have already established the architectural state of macroscopic *V. cholerae* biofilms[5]. 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 three-dimensional (3D) image-segmentation technique (see Methods and Supplementary Information) we were able to observe complete 3D biofilm development at cellular resolution with minimal phototoxicity (Fig. 1a,b) and minimal segmentation error (see Supplementary Information). 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 (Fig. 1b–d and Supplementary Movie 1).

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[4,10-17]. 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[10,18]. For *V. cholerae* biofilms, the molecular basis for cell–cell interactions has been intensively investigated: cells are embedded in a self-secreted extracellular matrix composed of the *Vibrio* polysaccharide (VPS), extracellular DNA and proteins[19-21]. 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. 1c)[20,21] and links cells to each other[21-23]; its expression levels are inversely related to cell–cell spacing (Fig. 2a). VPS also weakly binds cells together, yet elevated levels of VPS production do not cause stronger cell–cell attraction or decreased cell–cell spacing (Supplementary Fig. 11). Based on these cell–cell interaction processes, we hypothesized that biofilm architectures are primarily determined by the relative strength of the effective mechanical cell–cell attraction and repulsion forces.

To determine the impact of cell–cell attraction, we quantitatively compared the 3D biofilm architecture dynamics of a rugose wild-type strain with straight cell shape (WT*) with that of a mutant strain ($\Delta rbmA$) with significantly weakened intercellular adhesion (see Methods). Biofilms grown in a low-shear environment approximately display hemispherical symmetry (Fig. 1d), which allows us to characterize the biofilm architectures (Fig. 1e) as a function of the distance to the biofilm centre in the

---

[1]Max Planck Institute for Terrestrial Microbiology, Marburg, Germany. [2]Department of Physics, Philipps-Universität Marburg, Marburg, Germany. [3]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [4]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]These authors contributed equally: Praveen K. Singh, Philip Pearce, Rachel Mok.
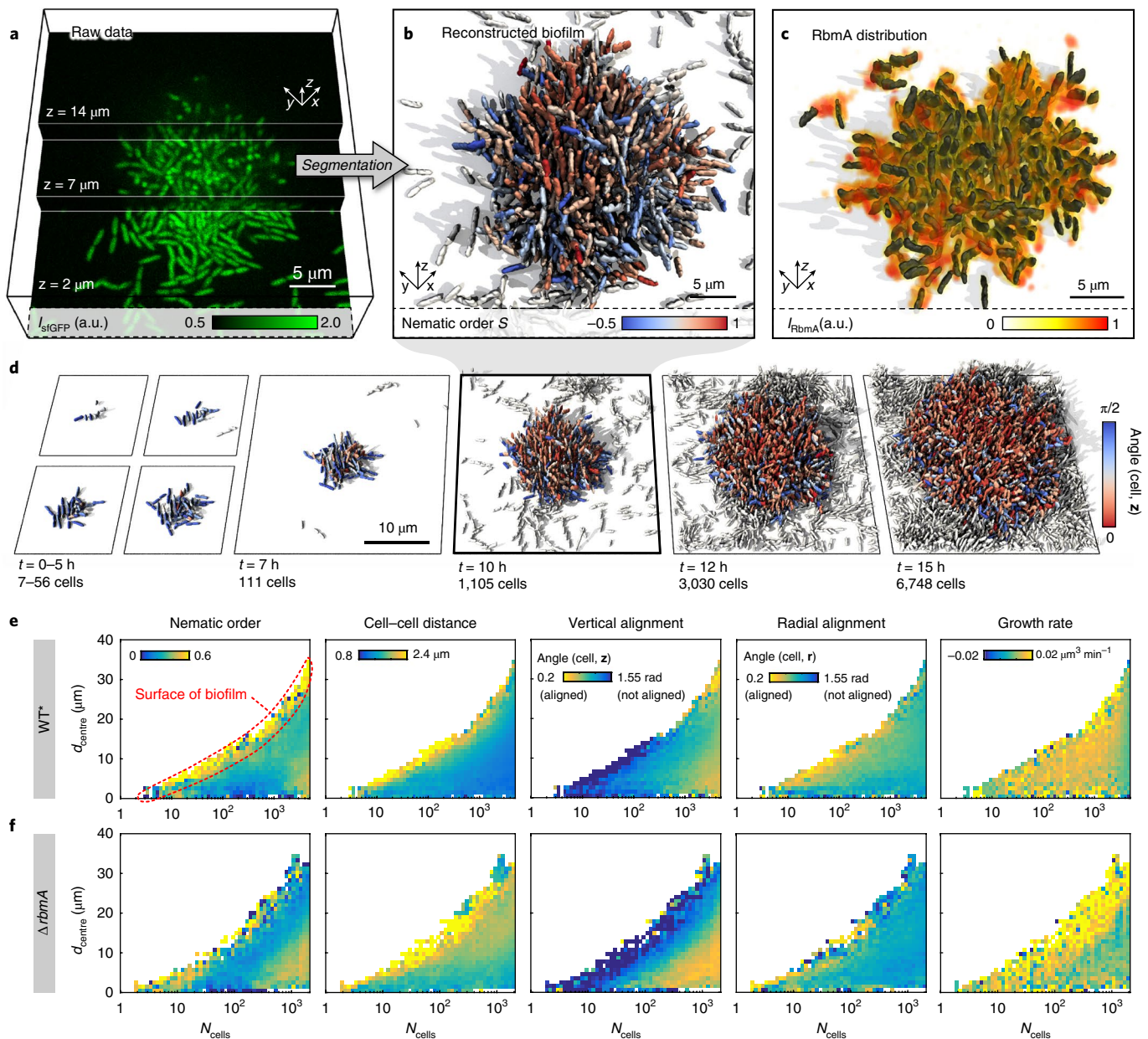*e-mail: dunkel@mit.edu; k.drescher@mpi-marburg.mpg.de

Fig. 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 coloured according to the nematic order parameter $S = \langle 3/2(\hat{n}_i \cdot \hat{n}_j)^2 - 1/2 \rangle$ in its vicinity. High-time-resolution ($\Delta t = 5$–10 min) imaging allowed us to track cell lineages and discriminate 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 growth series. Each cell is coloured according to the cellular alignment with the *z* axis (for the $\Delta rbmA$ mutant see Supplementary Fig. 6). **e,f**, Heatmaps showing spatially resolved single-cell measurements of different biofilm structural properties inside WT* (a rugose wild-type strain with straight cell shape) (**e**) and $\Delta rbmA$ (**f**) 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 Supplementary Fig. 8) as a function of the distance to the biofilm centre (d_{centre}) and the number of cells inside the biofilm (N_{cells}).

basal plane, d_{centre}, using the cell number in the biofilm, N_{cells}, as a quantification of the developmental state. Our 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 (Fig. 1e,f). Interestingly, the cellular growth rate remains homogeneous in space during WT* biofilm development in our conditions and for our biofilm sizes (Fig. 1e and Supplementary Fig. 5), in contrast to theories

assuming steep nutrient gradients inside biofilms[8,10]. The nematic order, cell–cell spacing and cellular orientations with respect to the vertical (**z**) and radial (**r**) directions differ significantly between WT* and $\Delta rbmA$ mutants (Fig. 1e,f and Supplementary Figs. 5–8), revealing the strong effect of cell–cell adhesion on biofilm architecture dynamics.

Based on the high-resolution spatiotemporal data of biofilm development of different bacterial strains, we investigated the
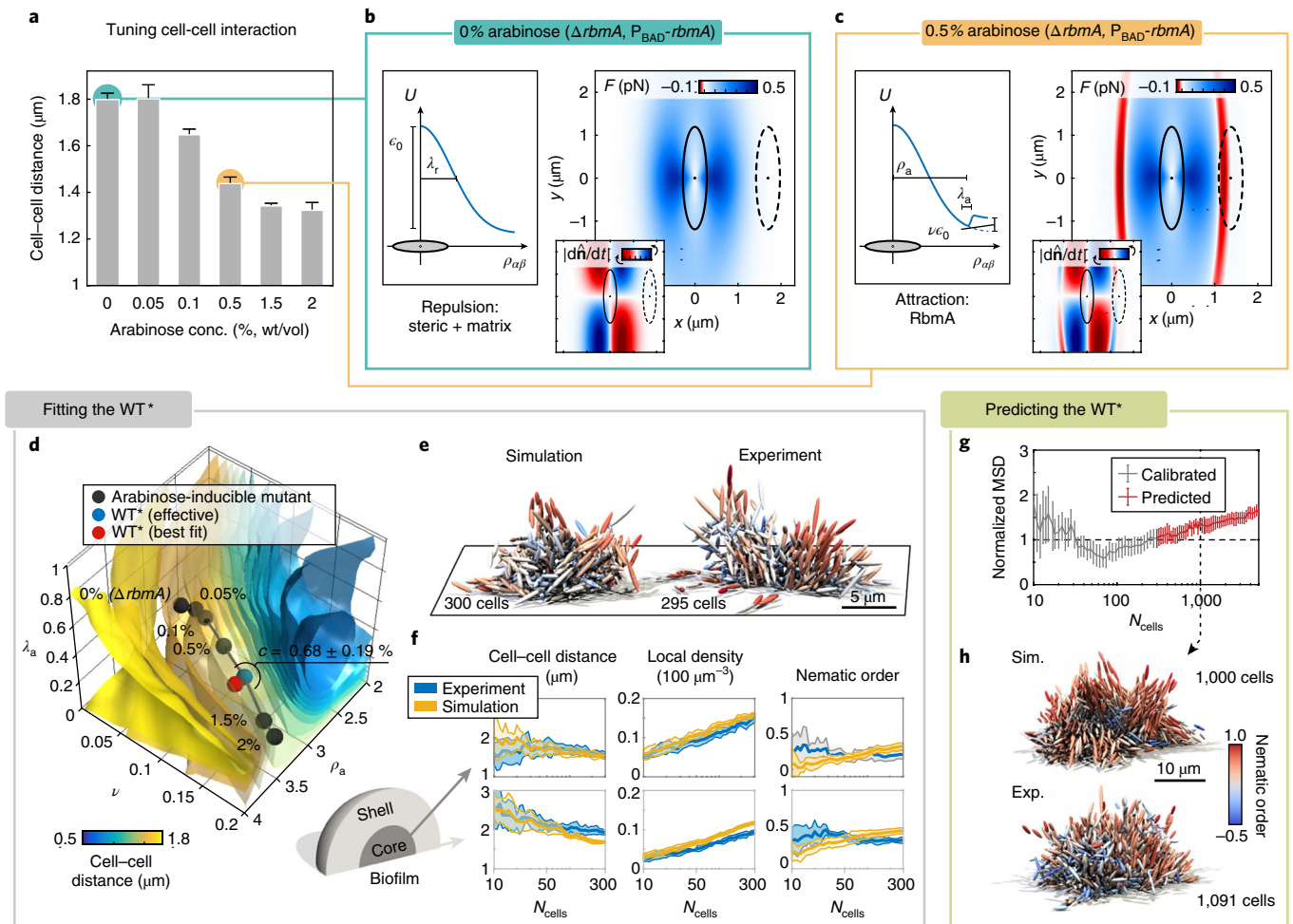
Fig. 2 | Biofilm architecture development is captured by an effective mechanical cell–cell interaction potential. a, Increased RbmA production (achieved by increasing the arabinose concentration, see Methods) decreases the average cell–cell distance in biofilms. Without arabinose, no RbmA is produced and the biofilm architecture is identical to the Δ*rbmA* mutant (*n* > 3 biofilms). **b**, Cell–cell interaction inside Δ*rbmA* mutant biofilms lacking cell–cell adhesion, modelled by the repulsive interaction potential (left) and the resulting cell–cell interaction forces (right) for the best-fit potential and the most prominent cellular orientation (red, attractive; blue, repulsive). Inset, Rotational interaction dynamics (red, clockwise rotation; blue, anticlockwise rotation). For more details and additional orientations see Supplementary Figs. 18 and 19. The dashed cell is plotted at the average cell–cell distance obtained from the corresponding experiment in **a**. **c**, The cell–cell interaction potential (left) and force (right) resulting from the best-fit potential for biofilms with a particular level of cell–cell adhesion (0.5% arabinose). RbmA-mediated cell–cell adhesion gives rise to an attractive part (red), acting within the range of the experimentally determined average cell–cell distance (dashed cell). **d**, Best-fit simulation parameters for varying RbmA and arabinose concentrations (black dots) follow a line in (*v*, λ_a, ρ_a) parameter space and cross isosurfaces of average cell–cell distance (see colour bar, and compare with **a**; for more details about the fitting see Supplementary Fig. 23). 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). **e**, Simulated (best fit) versus experimental WT* biofilm. **f**, 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 (left) and shell (right) of the biofilm (experiment, *n* = 7; simulation, *n* = 10). **g,h**, Simulation predictions of large (*N*cells = 1,000) WT* biofilms (based on the WT* interaction potential calibrated with *N*cells < 300) show quantitative (**g**) and qualitative (**h**) agreement with experiments (experiment, *n* = 4; simulation, *n* = 10). All error bars correspond to standard errors.

hypothesis that the biofilm internal structure and external shape originate from mechanical interactions between cells. Focusing on a minimal model, we describe the effective mechanical interactions in terms of an effective potential that depends on the distance $r_{\alpha\beta}$ between neighbouring cells α and β, and their orientations $\hat{\mathbf{n}}_\alpha$ and $\hat{\mathbf{n}}_\beta$. We made the simplifying assumption that the potential is independent of the biofilm developmental state or nutrient levels. As shown below, 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. Given

the molecular components of the cell–cell interaction and their qualitative effects on attraction and repulsion outlined above, we assume the pair potential

$$U = \epsilon_0 \epsilon_1 \left(e^{-\frac{\rho^2}{\lambda_r^2}} + \frac{\nu}{1 + e^{\left(\frac{\rho_a - \rho}{\lambda_a}\right)}} \right) \quad (1)$$

where $\rho = r_{\alpha\beta}/\sigma$ is the shape-normalized cell–cell distance. The range parameter $\sigma(\hat{\mathbf{r}}_{\alpha\beta}, \hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta)$ depends on the instantaneous cell lengths, the orientation of the cells relative to each other and the

individual cell orientations, and it maps the potential onto non-identical ellipsoidal cells (Supplementary equation (20)). The amplitude is set by the repulsion strength ϵ_0 and instantaneous cell lengths and cell orientations through the strength parameter $\epsilon_1(\hat{\mathbf{n}}_\alpha, \hat{\mathbf{n}}_\beta)$ (Supplementary equation (19)). The first term of the interaction potential describes the combined effects of hard steric and osmotic repulsion with range λ_r (Fig. 2b). The second term corresponds to cell–cell attraction and adds an attractive part of relative depth ν, width λ_a and position ρ_a (Fig. 2c). Each contribution and parameter in the potential U thus has a well-defined physical meaning (schematic diagram in Supplementary Fig. 15 and Supplementary Table 3). We assume here that the interaction parameters are taken to be constant for a given bacterial strain, a simplification that could be relaxed in future models. With these simplifying assumptions, initial estimates of the potential parameters prior to systematic scans can be obtained from basic physical considerations (see Methods).

This potential was then implemented in a particle-based model of biofilm development, in which individual cells were modelled as growing and dividing ellipsoids without self-propulsion (see Supplementary Information), whose interactions are described by U. Bypassing previous limitations of individual-based biofilm models[24,25], all the parameters of our model (cell aspect ratio, division time distribution; Supplementary Table 3) were determined from single-cell properties of experimental biofilms, and the dynamics were solved with a massively parallel computation approach using graphics processing units to evaluate all pairwise interactions (see Supplementary Information). To obtain the key potential parameters $\epsilon_0, \lambda_r, \nu, \lambda_a$ and ρ_a for *V. cholerae* biofilms, we assumed that the attractive term in U can be attributed primarily to RbmA levels, with the VPS acting as a Woods–Saxon background potential (see Methods) akin to the mean-field potential in nucleon models. This assumption is motivated by the experimental findings that increased VPS levels do not increase the cell–cell attraction (Supplementary Fig. 11), yet biofilms that lack RbmA display a small residual mechanical cohesion (Fig. 3e), indicating that VPS does contribute weakly to cell–cell binding. To first obtain the parameters ϵ_0 and λ_r, we fitted the repulsive part of the potential U by comparing experimental $\Delta rbmA$ biofilms, which lack the attractive potential term ($\nu = 0$), with simulated biofilms, using the mean squared difference (MSD) of a feature vector as a metric. The feature vector contains 14 different architectural properties and their temporal variation up to 300 cells (Supplementary Fig. 14), allowing a comprehensive comparison of biofilm architecture and development between simulations and experiment at the same time. Note that even at small sizes, the *V. cholerae* biofilms used in this study produce RbmA and VPS (Supplementary Fig. 10). For $\Delta rbmA$ biofilms we found a broad minimum in the (ϵ_0, λ_r) space as shown in the MSD heatmap (Supplementary Fig. 16), resulting in best-fit simulations that show high similarity to experiments (Supplementary Fig. 17). The effective translational and rotational interaction forces acting on two neighbouring $\Delta rbmA$ cells for the best-fit potential are illustrated in Fig. 2b and Supplementary Fig. 18 for different cellular orientations. The interaction range for two aligned cells is very close to the experimentally observed average cell–cell spacing of the $\Delta rbmA$ mutant (dashed cell).

Because the attraction parameters (ν, λ_a, ρ_a) in potential U depend on the concentration of RbmA, we genetically modified *V. cholerae* so that we could tune the production of RbmA (and therefore tune the strength of the attraction), by adding different concentrations of a compound that induces the *rbmA*-expression construct homogeneously inside the biofilm: arabinose (see Methods and Supplementary Fig. 13). Experimentally, we observed that increasing arabinose concentrations resulted in decreased cell–cell spacing (Fig. 2a), consistent with the assumption that RbmA mediates cell–cell attraction. By fixing the repulsive component (ϵ_0, λ_r) based on the $\Delta rbmA$ biofilms, we then fitted the attractive potential component

(ν, λ_a, ρ_a) for a range of different arabinose concentrations (Fig. 2c,d). The MSD isosurfaces in (ν, λ_a, ρ_a) space and corresponding 3D renderings for simulated and experimental biofilms grown at 0.5% (wt/vol) of arabinose reveal tight fits (Supplementary Figs. 21–23), and the resulting best-fit interaction force displays an attractive region (red) at the average experimental cell–cell distance (Fig. 2c).

With the calibrated simulation, we then inferred an effective arabinose concentration for the WT* of $c = 0.68 \pm 0.19\%$ (wt/vol), by locating the WT* biofilm architecture in the (ν, λ_a, ρ_a) space along the curve of different arabinose concentrations (Fig. 2d). Extracting an effective arabinose concentration and RbmA level for the WT* is based on the simplifying assumption that all cells in the biofilm 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 WT* on the curve of different arabinose concentrations (Fig. 2d). The simulations based on the WT* parameters for biofilms up to 300 cells show good quantitative agreement with experiments (Fig. 2f). Remarkably, these simulations also show architectural 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. 2e, characterized by high local ordering), which are characteristic for biofilms with high concentrations of RbmA. Predictions of the architectural development for larger biofilms ($N_{cells} > 300$) show high quantitative and qualitative agreement with experimental data, for both the WT* (Fig. 2g,h and Supplementary Movie 5) and $\Delta rbmA$ (Supplementary Fig. 24b and Supplementary Movie 5) biofilms up to 10^3 cells. 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. Our combined experimental and theoretical analysis therefore suggests that mechanical interactions between cells suffice to account for the internal cellular order and architecture up to mid-sized *V. cholerae* biofilms.

To determine how external fields can affect the orientational order and morphology of 3D biofilms, we perturbed biofilm growth by applying external flow fields of varying strength, corresponding to shear rates of $\dot{\gamma} = 2$–$2{,}000\,\mathrm{s}^{-1}$, typically encountered by bacteria in natural and man-made environments[26]. At high shear rates (>$600\,\mathrm{s}^{-1}$, corresponding to average flow speeds >$10\,\mathrm{mm\,s}^{-1}$ through the growth chamber), the WT* cells formed smaller, more compact biofilm colonies with droplet-like shapes, compared with low shear environments (Fig. 3a,b, Supplementary Movies 1 and 3). To understand the mechanisms underlying these architectural changes, we investigated both local and global effects of increased shear on biofilms, and changes in matrix production. Exposure to higher shear resulted in a significantly decreased cell–cell spacing and lower growth rate in WT* biofilms (Fig. 3c), but the height-to-width aspect ratio was unaffected when comparing biofilms with similar N_{cells} (Fig. 3e) despite the increased levels of shear stress applied to the top (Fig. 3d). We therefore hypothesized that cells in WT* biofilms at higher shear secrete increased levels of RbmA, allowing increased cell–cell attraction forces to balance shear forces, but leading to a strong reduction in overall growth rate owing to the metabolic cost of increased RbmA production. Using a fluorescent transcriptional reporter for *rbmA* expression, we confirmed that high shear increases RbmA levels (Supplementary Fig. 12), indicating that cells actively modulate the mechanical cell–cell interactions via gene expression.

To explain the observed droplet-like shapes of biofilms grown at high shear rate, we investigated cellular alignment with flow and analysed biovolume flux inside the biofilm using the optical flow method (Fig. 3a–f; see Supplementary Information). We determined that cell alignment with flow increases with increasing shear rate (Fig. 3c–g), and an anisotropic biomass shift downstream occurs at $\dot{\gamma} > 60\,\mathrm{s}^{-1}$ (Fig. 3f), indicating that the observed biofilm
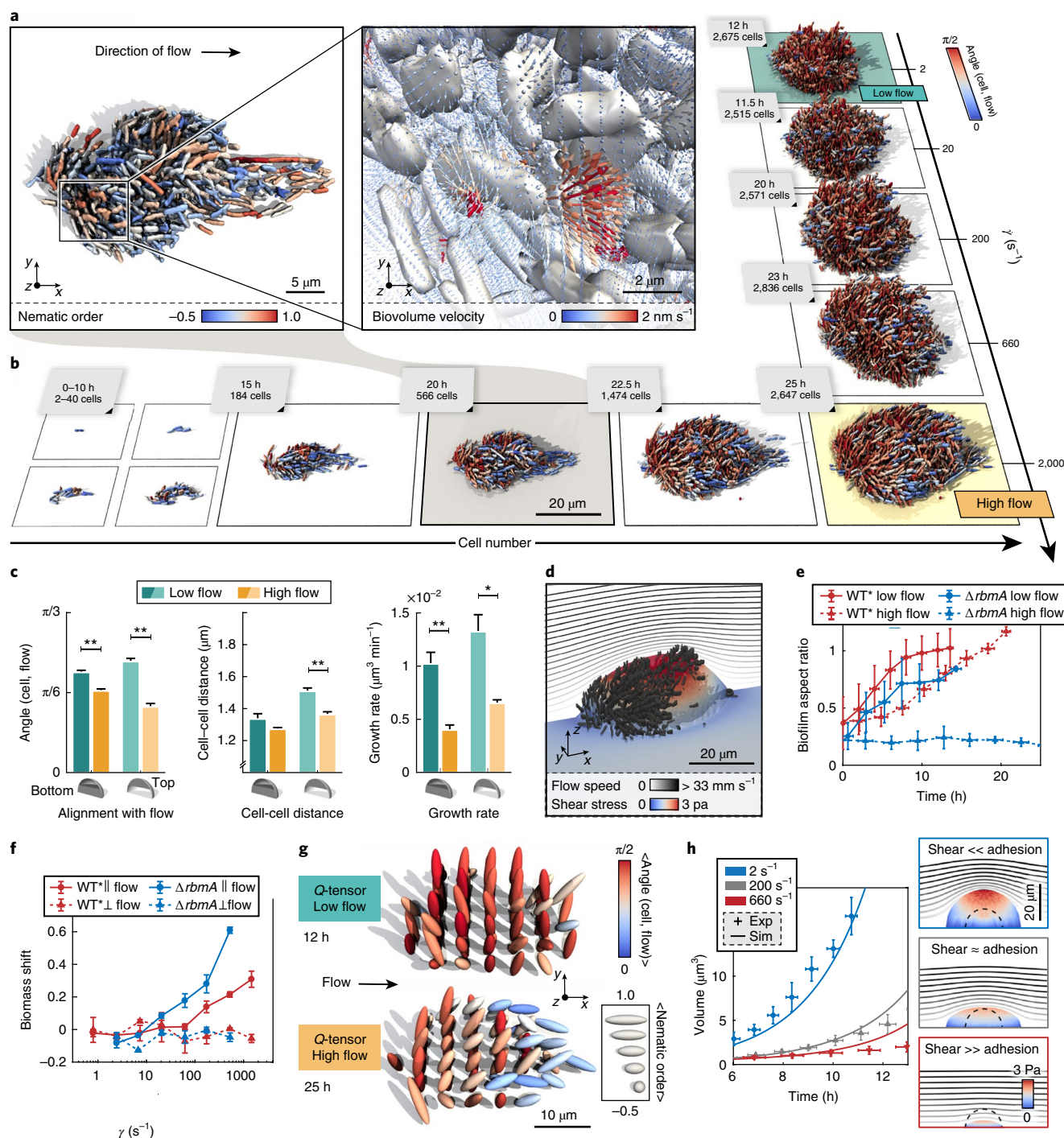
Fig. 3 | Biofilm architecture is shaped by external shear flow. a, WT* biofilms grown under strong shear ($\dot\gamma = 2,000\,\mathrm{s}^{-1}$) display droplet-like shapes. Inset, Biovolume flux field inside the biofilm (see Supplementary Information). **b**, WT* biofilms under high shear ($\dot\gamma = 2,000\,\mathrm{s}^{-1}$) display strong alignment with flow throughout growth, yet biofilms grown in flow with low shear ($\dot\gamma = 2$-$200\,\mathrm{s}^{-1}$) do not show strong architectural modifications. **c**, Quantification of the effect of shear on biofilm architecture: measurements of cellular alignment with flow, cell–cell distance and cell growth rate at the bottom and top of biofilms with $N_{\mathrm{cells}} \sim 800$ cells show that WT* biofilms in high shear are smaller, more compact and display stronger flow alignment. Statistical significance: *$P < 0.05$; **$P < 0.01$ (t-test); error bars indicate standard error ($n = 4$ biofilms). **d**, Simulated shear stress distribution for a WT* biofilm, demonstrating that the region of highest shear is at the top of the biofilm. The streamlines indicate the profile of the external flow. **e**, Biofilm aspect ratio (height/width) increases in time for WT* (red) biofilms, but decreases for $\Delta rbmA$ mutant biofilms (blue) in high flow owing to shear-induced erosion ($n = 4$, error bars indicate standard deviations). **f**, Biomass shift is defined as the fraction of the average total biomass flux through planes parallel (∥) or perpendicular (⊥) to flow (see Supplementary Fig. 2 for details). Positive biomass shift along the flow direction at higher shear rates indicates anisotropic biofilm expansion towards the downstream direction of the external flow. Zero biomass shift perpendicular to the flow indicates no directional bias ($n \geq 3$, error bars indicate standard errors). **g**, The tensorial nematic order parameter (Q tensor, see Supplementary Information) and cellular alignment with the flow direction were measured at equally spaced points inside biofilms at low and high shear rates, indicating the regions in which cells are predominantly aligned with the flow and each other. **h**, Biofilm volumetric growth for $\Delta rbmA$ mutant biofilms is captured by a continuum model (see Supplementary Information) with varying ratios of shear-induced erosion and cell–cell adhesion (experiment: $n = 4$, error bars indicate standard deviations).

shapes are caused by anisotropic expansion of cells aligned with the flow as a result of growth and viscoelastic deformation. Our above measurements regarding increased RbmA levels in WT* biofilms at high shear predict that if RbmA levels are in fact primarily responsible for cell–cell attraction, then most effects of shear on $\Delta rbmA$-mutant biofilms should be explained by shear-induced cell erosion. Indeed, these biofilms showed a reduction in upward growth with higher flow (Fig. 3e), indicating that shear forces are larger than cell–cell attraction forces. This was confirmed by simulations of shear-dependent erosion using a continuum model (see Supplementary Information), which captured the decreased volumetric growth of $\Delta rbmA$-mutant biofilms owing to cell erosion (Fig. 3h and Supplementary Movie 4). Fluid flow therefore strongly affects biofilm architectural development through the effect of shear on growth rate, matrix composition, alignment with flow, biomass shift and shear-induced erosion[27]. These results demonstrate that mechanical interactions at the cellular scale remain important in sculpting biofilm architecture when an external field is applied.

In conclusion, our combined experimental and theoretical analysis shows that the emergence of local nematic order in growing *V. cholerae* biofilms can be captured by an experimentally constrained minimal effective cell–cell interaction potential that translates molecular mechanisms into force parameters. Given the immense complexity of the molecular interactions, metabolism and signalling 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[9,28]. 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[18,29,30] and biophysical research[5,10,31].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41567-018-0356-9.

References

1. Zhou, S., Sokolov, A., Lavrentovich, O. D. & Aranson, I. S. Living liquid crystals. *Proc. Natl Acad. Sci. USA* **111**, 1265–1270 (2014).
2. Hagan, M. F. & Baskaran, A. Emergent self-organization in active materials. *Curr. Opin. Cell Biol.* **38**, 74–80 (2016).
3. Doostmohammadi, A., Adamer, M. F., Thampi, S. P. & Yeomans, J. M. Stabilization of active matter by flow-vortex lattices and defect ordering. *Nat. Commun.* **7**, 10557 (2016).
4. Volfson, D., Cookson, S., Hasty, J. & Tsimring, L. S. Biomechanical ordering of dense cell populations. *Proc. Natl Acad. Sci. USA* **105**, 15346–15351 (2008).
5. Drescher, K. et al. Architectural transitions in *Vibrio cholerae* biofilms at single-cell resolution. *Proc. Natl Acad. Sci. USA* **113**, E2066–E2072 (2016).
6. Yan, J., Sharo, A. G., Stone, H. A., Wingreen, N. S. & Bassler, B. L. *Vibrio cholerae* biofilm growth program and architecture revealed by single-cell live imaging. *Proc. Natl Acad. Sci. USA* **113**, E5337–E5343 (2016).
7. Kragh, K. N. et al. Role of multicellular aggregates in biofilm formation. *mBio* **7**, e00237 (2016).
8. Flemming, H.-C. et al. Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.* **14**, 563–575 (2016).
9. Marchetti, M. C. et al. Hydrodynamics of soft active matter. *Rev. Mod. Phys.* **85**, 1143–1189 (2013).
10. Persat, A. et al. The mechanical world of bacteria. *Cell* **161**, 988–997 (2015).
11. Liu, J. et al. Coupling between distant biofilms and emergence of nutrient time-sharing. *Science* **356**, 638–642 (2017).
12. Rodesney, C. A. et al. Mechanosensing of shear by *Pseudomonas aeruginosa* leads to increased levels of the cyclic-di-GMP signal initiating biofilm development. *Proc. Natl Acad. Sci. USA* **114**, 5906–5911 (2017).
13. Grant, M. A. A., Waclaw, B., Allen, R. J. & Cicuta, P. The role of mechanical forces in the planar-to-bulk transition in growing *Escherichia coli* microcolonies. *J. R. Soc. Interface* **11**, 20140400 (2014).
14. You, Z., Pearce, D. J. G., Sengupta, A. & Giomi, L. Geometry and mechanics of microdomains in growing bacterial colonies. *Phys. Rev. X* **8**, 031065 (2018).
15. Delarue, M. et al. Self-driven jamming in growing microbial populations. *Nat. Phys.* **12**, 762–766 (2016).
16. Seminara, A. et al. Osmotic spreading of *Bacillus subtilis* biofilms driven by an extracellular matrix. *Proc. Natl Acad. Sci. USA* **109**, 1116–1121 (2012).
17. Trejo, M. et al. Elasticity and wrinkled morphology of *Bacillus subtilis* pellicles. *Proc. Natl Acad. Sci. USA* **110**, 2011–2016 (2013).
18. Maier, B. & Wong, G. C. L. How bacteria use type IV pili machinery on surfaces. *Trends Microbiol.* **23**, 775–788 (2015).
19. Teschler, J. K. et al. Living in the matrix: assembly and control of *Vibrio cholerae* biofilms. *Nat. Rev. Microbiol.* **13**, 255–268 (2015).
20. Berk, V. et al. Molecular architecture and assembly principles of *Vibrio cholerae* biofilms. *Science* **337**, 236–239 (2012).
21. Fong, J. C. et al. Structural dynamics of RbmA governs plasticity of *Vibrio cholerae* biofilms. *Elife* **6**, e26163 (2017).
22. Maestre-Reyna, M., Wu, W.-J. & Wang, A. H.-J. Structural insights into RbmA, a biofilm scaffolding protein of *V. Cholerae*. *PLoS ONE* **8**, e82458 (2013).
23. Fong, J. C. N., Karplus, K., Schoolnik, G. K. & Yildiz, F. H. 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**, 1049–1059 (2006).
24. Hellweger, F. L., Clegg, R. J., Clark, J. R., Plugge, C. M. & Kreft, J. U. Advancing microbial sciences by individual-based modelling. *Nat. Rev. Microbiol.* **14**, 461–471 (2016).
25. Lardon, L. A. et al. iDynoMiCS: next-generation individual-based modelling of biofilms. *Environ. Microbiol.* **13**, 2416–2434 (2011).
26. Marcos, Fu,H. C., Powers, T. R. & Stocker, R. Bacterial rheotaxis. *Proc. Natl Acad. Sci. USA* **109**, 4780–4785 (2012).
27. Mitchell, W. H. & Spagnolie, S. E. A generalized traction integral equation for Stokes flow, with applications to near-wall particle mobility and viscous erosion. *J. Comput. Phys.* **333**, 462–482 (2017).
28. Cates, M. E. & Tjhung, E. Theories of binary fluid mixtures: from phase-separation kinetics to active emulsions. *J. Fluid. Mech.* **836**, 1–68 (2018).
29. Singh, P. K. et al. *Vibrio cholerae* combines individual and collective sensing to trigger biofilm dispersal. *Curr. Biol.* **27**, 3359–3366 (2017).
30. Vidakovic, L., Singh, P. K., Hartmann, R., Nadell, C. D. & Drescher, K. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nat. Microbiol.* **3**, 26–31 (2017).
31. Smith, W. P. J. et al. Cell morphology drives spatial patterning in microbial communities. *Proc. Natl Acad. Sci. USA* **114**, E280–E286 (2017).

Author contributions

K.D. and J.D. designed and supervised the study. R.H. and P.K.S. performed experiments. P.K.S. and F.D.-P. created bacterial strains. R.H. developed experimental and analysis software. P.P. developed continuum simulations. R.M. developed cell-based simulation framework. R.M., R.H. and B.S. performed cell-based simulations. R.M., R.H., P.P. and B.S. developed cell–cell potentials. R.H., with the help of P.P., J.D. and K.D., analysed the data. R.H., P.P., J.D. and K.D. wrote the manuscript, with the help of all authors.

Competing interests

The authors declare no competing interests.

Additional information

Methods

Media and cloning approaches. All strains were grown in lysogeny broth (LB) medium supplemented with appropriate antibiotics at 37 °C for normal growth and during cloning. Biofilm experiments with *V. cholerae* were performed in M9 minimal medium, supplemented with 2 mM $MgSO_4$, 100 mM $CaCl_2$, minimum essential medium (MEM) vitamins, 0.5% glucose and 15 mM triethanolamine (pH 7.1). Standard molecular biology techniques were applied to construct plasmids and strains[32]. Restriction enzymes and DNA polymerase enzymes were purchased from New England Biolabs. Oligonucleotides were commercially synthesized by Eurofins. All *V. cholerae* strains used in this study are derivatives of the rugose variant of the wild-type *V. cholerae* O1 biovar El Tor strain N16961 (termed strain KDV148). *V. cholerae* deletion mutations were engineered using the pKAS32 suicide vector harboured in *Escherichia coli* S17-1 *λpir* (ref. [33]). Complementation constructs were inserted at the *lacZ* site with the help of the suicide plasmid pKAS32. The plasmid pNUT542, containing gene coding for the super-folder green fluorescent protein (*sfgfp*) expressed under the control of the P_{tac} promoter, was conjugated into all *V. cholerae* strains except for the complementation strain KDV1082[29]. Plasmid clones were first constructed in the *E. coli* strain Top10 and then mated into *V. cholerae* with the help of an additional *E. coli* strain harbouring the conjugation plasmid pRK600. Arabinose was used as an inducer to control the expression of *rbmA* from the arabinose-regulated promoter P_{BAD}. Details of the strains, plasmids and oligonucleotides are listed in the Supplementary Information.

Strain construction. The rugose variant of the *V. cholerae* N16961 (strain KDV148) displays strong surface attachment and biofilm formation as a consequence of high c-di-GMP production[34]. *V. cholerae* cells are usually characterized by a slightly curved cell shape. To allow *V. cholerae* cells to be modelled by ellipsoids in theory and simulations, we generated a mutant with a straight cell shape (that is, the common bacterial rod shape) by deleting the gene Δ*crvA* according to the method of ref. [35]. In detail, the 1 kb flanking regions of gene *crvA* (VCA1075) were amplified with the oligonucleotides kdo1182/kdo1183 and kdo1183/kdo1184, and the fused PCR product was amplified using kdo1182/kdo1185. The final PCR product was ligated into plasmid pNUT144 (a derivative of pKAS32). The resulting plasmid pNUT961 was conjugated into strain KDV148, to generate the Δ*crvA* deletion mutant, following the selection protocol described in ref. [33]. Finally, cells containing the correct mutation were screened by PCR. Plasmid pNUT542 was conjugated into KDV611 strain to construct strain KDV613 containing the Δ*crvA* deletion (referred to as WT*). The Δ*rbmA* deletion strain (KDV698) was constructed by conjugating plasmid pNUT336 into strain KDV611. The mutant screening was performed by PCR[36].

Tuning cell–cell interaction by inducing *rbmA* expression. To control the timing and rate of RbmA production, an inducible strain (KDV1082) was generated by conjugating plasmid pNUT1519 into the Δ*rbmA* strain KDV698. Plasmid pNUT1519 was created by cloning a P_{tac}-*sfgfp* construct into plasmid pNUT1268. Plasmid pNUT1268 is a derivative of plasmid pNUT542 in which P_{tac}-*sfgfp* was replaced with a construct of P_{BAD}-*rbmA*. P_{BAD}, an arabinose inducible promoter, and the *rbmA* gene was joined by PCR amplification with oligonucleotides kdo1435/kdo1436.

Visualization of secreted RbmA. To visualize RbmA during biofilm growth, the wild-type copy of *rbmA* was exchanged by a FLAG-tagged *rbmA*[16] (with the octapeptide DYKDDDDK) by mating the plasmid pNUT462 into strain KDV148, resulting in *V. cholerae* strain KDV829. Successful FLAG-tagging of RbmA was confirmed by PCR and sequencing. The final strain KDV835 was generated by conjugating the fluorescence protein expression plasmid pNUT542 into strain KDV829. For RbmA visualization in flow chambers, biofilms were grown in M9 medium containing 1 µg ml^{-1} of FLAG tag monoclonal antibody (L5) conjugated to AlexaFluor 555 (Thermo Scientific) and 1 mg ml^{-1} of filter-sterilized bovine serum albumin.

Measuring *rbmA* expression. To measure RbmA production during biofilm growth, the gene *mRuby3* was transcriptionally fused to *rbmA* on the chromosome by introducing plasmid pNUT1401 into strain KDV611. The transcriptional fusion of *rbmA*-*mRuby3* in the resulting strain (KDV1026) was confirmed by PCR and sequencing. The final strain KDV1027 was generated by mating plasmid pNUT542 into strain KDV1026.

Flow chamber biofilm experiments. *V. cholerae* biofilms were grown in microfluidic flow chambers as described in ref. [5] (chamber dimensions: [width; height; length] = [500; 100; 7,000] µm). Flow chambers were constructed from poly(dimethylsiloxane) bonded to glass coverslips using oxygen plasma. The microfluidic design included four independent channels on each coverslip. The manufacturing process for these microfluidic channels guarantees highly reproducible channel dimensions and surface properties. Each channel was inoculated with a culture of a particular *V. cholerae* strain. Cultures were grown overnight at 28 °C in liquid LB medium under shaking conditions, back-diluted 1:200 in LB medium in the morning, and grown to an optical density at 600 nm

of 0.5 before channel inoculation. After inoculation of the channels, the cells were given 1 h to attach to the glass surface of the channel, before a flow of 100 µl min^{-1} M9 medium was initiated for 45 s to wash away non-adherent cells and to remove LB growth medium from the channels. The flow rate was then set to a value between 0.1 and 100 µl min^{-1}, corresponding to an average flow speed $<v>$ between 0.03 and 33 mm s^{-1} and a shear rate $\dot{\gamma}$ between 2 and 2,000 s^{-1} (as indicated) until the end of the experiments. Flow rates were controlled using a high-precision syringe pump (Pico Plus, Harvard Apparatus).

Image acquisition. Single cells were reconstructed from confocal fluorescence image stacks acquired with a Yokogawa CSU confocal spinning disk unit mounted on a Nikon Ti-E inverted microscope, using an Olympus ×100 silicone oil (refractive index = 1.406) objective with NA 1.35, 488 nm and 552 nm lasers (Coherent Sapphire) and an Andor iXon electron-multiplying charge-coupled device (EMCCD) camera. By using this specific objective, heavy distortions at axial positions > 10 µm into the biofilm (owing to refractive index mismatch of biofilms and standard immersion oil) are reduced. The physical resolution was 63.2 nm per pixel in the *x*–*y* plane and was set to 400 nm along the axial direction. Images were acquired every 10 min at very low excitation light intensities with 90 ms exposure time, using the EM gain of the Andor iXon EMCCD camera. A Nikon hardware autofocus was used to correct for focus drift. The hardware was controlled using µManager[37]. During acquisition, live feedback between image acquisition, image analysis and microscope control was used to automatically detect the biofilm height to avoid imaging of empty space below and on top of the biofilm, to eliminate tracking of the *XY* coordinates of non-biofilm-forming cells and to control the temporal resolution (to reduce photobleaching and phototoxicity).

Image processing. Detailed descriptions of image processing, segmentation, segmentation validation, cell tracking, biomass shift, optical flow, 3D visualization, quantitative biofilm features that were measured and the calculation of space–time heatmaps are provided in the Supplementary Information.

Individual cell particle-based model. *Model description and implementation.* The cells were modelled as interacting ellipsoids of half-length ℓ and half-width r, described by their centre position $\mathbf{x}$ and orientation $\hat{\mathbf{n}}$. As cells operate at low Reynolds number (Re ≈ 10^{-4}), we approximated the dynamics as over-damped, ignoring inertial effects. Cells can interact with the wall boundary and other cells through interaction potential functions, U_{bdy} and U (equation (1)). Denoting the identity matrix by I, the over-damped translational and orientational dynamics of a single cell are described by

$$\frac{d\mathbf{x}}{dt} = \Gamma^{-1} \left[-\frac{\partial U_{bdy}}{\partial \mathbf{x}} - \frac{\partial V}{\partial \mathbf{x}} \right] \tag{2}$$

$$\frac{d\hat{\mathbf{n}}}{dt} = (I - \hat{\mathbf{n}}\hat{\mathbf{n}}^T) \left[\Omega^{-1} \left(-\frac{\partial U_{bdy}}{\partial \hat{\mathbf{n}}} - \frac{\partial V}{\partial \hat{\mathbf{n}}} \right) \right] \tag{3}$$

where Γ and Ω are friction tensors and V is the total interaction potential with all other cells as described in the Supplementary Information. The steric interaction between a cell and the wall boundary is modelled with a repulsive interaction potential that is proportional to the overlap between a cell and the wall boundary. The instantaneous cell-length growth rate for a single cell is defined as

$$\frac{d\ell}{dt} = \frac{\ell}{\tau_g} \ln(2) \tag{4}$$

where ℓ is the half-length of the cell at time t and τ_g is the growth time constant. The cell width is constant throughout the simulation. For further details of the particle-based model, see the Supplementary Information.

Simulation implementation. A custom highly parallelized individual cell-based code using graphics processing units was developed to perform the simulations. At each time step, we calculated cell–cell interactions using the all-pairs approach[38] such that all pairwise interactions were evaluated. We used a standard explicit Euler scheme to numerically integrate the translational and orientational dynamics and instantaneous cell-length growth (equations (2), (3) and (4)), as described in the Supplementary Information.

Parameter estimation. Initial order-of-magnitude estimates for systematic scans of the parameters in the pair potential U were obtained from basic physical considerations (see Supplementary Information), before systematic scans of the parameters were computed. The energy scale $\epsilon_0 \sim 0.05–5$ pN µm of the cell–cell interactions was assumed to be within a few orders of magnitude of the energy scale of cell–flow interactions, which were calculated via Stokes drag on a typical cell near the edge of the biofilm at a low flow rate (0.1 µl min^{-1}). The repulsive length scale $\lambda_r \sim 1$ (corresponding to ~1 µm for typically aligned cells) was estimated via the average cell–cell distance in the core of the biofilms, where

cell–cell repulsion dominates. The attraction shift $\rho_a \sim 1$ was estimated via the average cell–cell distance at the edge of the biofilms, where attraction dominates. The attraction width $\lambda_a \sim 0.1$ was determined by considering the standard deviation of experimental cell–cell distances at the edge of the biofilms.

Background potential. Cell–cell adhesion mediated by the VPS matrix component was modelled by a mean-field background Woods–Saxon potential[39], and was assumed to provide the weak cell–cell binding that prevents the disintegration of biofilms owing to fluid shear acting on $\Delta rbmA$ mutant biofilms (which lack the major cell–cell attraction, mediated by RbmA). The mean-field VPS-mediated binding strength was estimated to be approximately equal to the Stokes drag felt by a cell at the edge of the biofilm at low flow rate ($0.1\,\mu l\,min^{-1}$), because significant numbers of cells in the $\Delta rbmA$ background were sheared off at higher flow rate ($100\,\mu l\,min^{-1}$). However, WT* biofilms were found to be robust to this increased fluid shear, suggesting that the increased expression of *rbmA* at higher flow rate (Supplementary Information) increases the RbmA-mediated cell–cell attraction strength by approximately two orders of magnitude above the value predicted at low flow rate. 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.

Comparing simulations with experimental data. The dynamic biofilm architecture was summarized in a feature vector representing key phenotypic and structural properties temporally. The similarity between a vector characterizing a simulation and a real biofilm was assessed in terms of the MSD between them. For details, see Supplementary Information.

Continuum model. The mathematical continuum model of growing biofilms in shear flow is described in the Supplementary Information.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

References

32. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989).
33. Skorupski, K. & Taylor, R. K. Positive selection vectors for allelic exchange. *Gene* **169**, 47–52 (1996).
34. Beyhan, S. & Yildiz, F. H. Smooth to rugose phase variation in *Vibrio cholerae* can be mediated by a single nucleotide change that targets c-di-GMP signalling pathway. *Mol. Microbiol.* **63**, 995–1007 (2007).
35. Bartlett, T. M. et al. A periplasmic polymer curves *Vibrio cholerae* and promotes pathogenesis. *Cell* **168**, 172–185 (2017).
36. Nadell, C. D., Drescher, K., Wingreen, N. S. & Bassler, B. L. Extracellular matrix structure governs invasion resistance in bacterial biofilms. *ISME J.* **9**, 1700–1709 (2015).
37. Edelstein, A. D. et al. Advanced methods of microscope control using μ Manager software. *J. Biol. Methods* **1**, e10 (2014).
38. Nyland, L., Harris, M. & Prins, J. Fast N-body simulation with CUDA. *Simulation* **3**, 677–696 (2007).
39. Woods, R. D. & Saxon, D. S. Diffuse surface optical model for nucleon–nuclei scattering. *Phys. Rev.* **95**, 577–578 (1954).

Emergence of three-dimensional order and structure in growing biofilms

Raimo Hartmann [1,2], Praveen K. Singh [1,5], Philip Pearce [3,5], Rachel Mok[3,4,5], Boya Song[3], Francisco Díaz-Pascual[1], Jörn Dunkel[3]* and Knut Drescher [1,2]*

[1]Max Planck Institute for Terrestrial Microbiology, Marburg, Germany. [2]Department of Physics, Philipps-Universität Marburg, Marburg, Germany. [3]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [4]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]These authors contributed equally: Praveen K. Singh, Philip Pearce, Rachel Mok.
*e-mail: dunkel@mit.edu; k.drescher@mpi-marburg.mpg.de

Emergence of three-dimensional order and structure in growing biofilms

Raimo Hartmann[1,2], Praveen K. Singh[1,†], Philip Pearce[3,†], Rachel Mok[3,4,†], Boya Song[3], Francisco Díaz-Pascual[1], Jörn Dunkel[3,*], Knut Drescher[1,2,*]

[1] Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, DE

[2] Department of Physics, Philipps-Universität Marburg, 35032 Marburg, DE

[3] Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

* Correspondence to: dunkel@mit.edu; k.drescher@mpi-marburg.mpg.de

† Equally contributing authors

Supplementary Information

Table of contents

1. Data Analysis

1.1. Image processing

Image processing was carried out using MatLab (MathWorks). In the following sections, raw data 3D image stacks are referred to as $I_0(t)$ and are matrices of dimension x, y, and z, which were acquired at a time point t (see Supplementary Fig. 1). Any intermediate processing result is labelled with consecutive indices $I_i(t)$. Prior to segmentation, subsequent image stacks were aligned along x, y, and z by image registration (= I_1). Floating cells which were not associated with a biofilm were removed by median filtering along z (= I_2). Image stacks that have been registered and for which floating cells have been removed, were then up-sampled along the z-dimension to obtain equal voxel side lengths (= I_3). Subsequently, noise was removed by filtering as follows. The eigenvalues of each xz- and each yz-plane were obtained by singular value decompositions and the lowest thirds of the calculated values were set to zero to reduce high frequency noise (= I_4). In addition, a 3D-convolution with an averaging kernel was performed (= I_5). Afterwards, the out-of-focus fluorescence was suppressed by slice-wise Top-hat filtering of the image stacks (= I_6, see Supplementary Fig. 1).
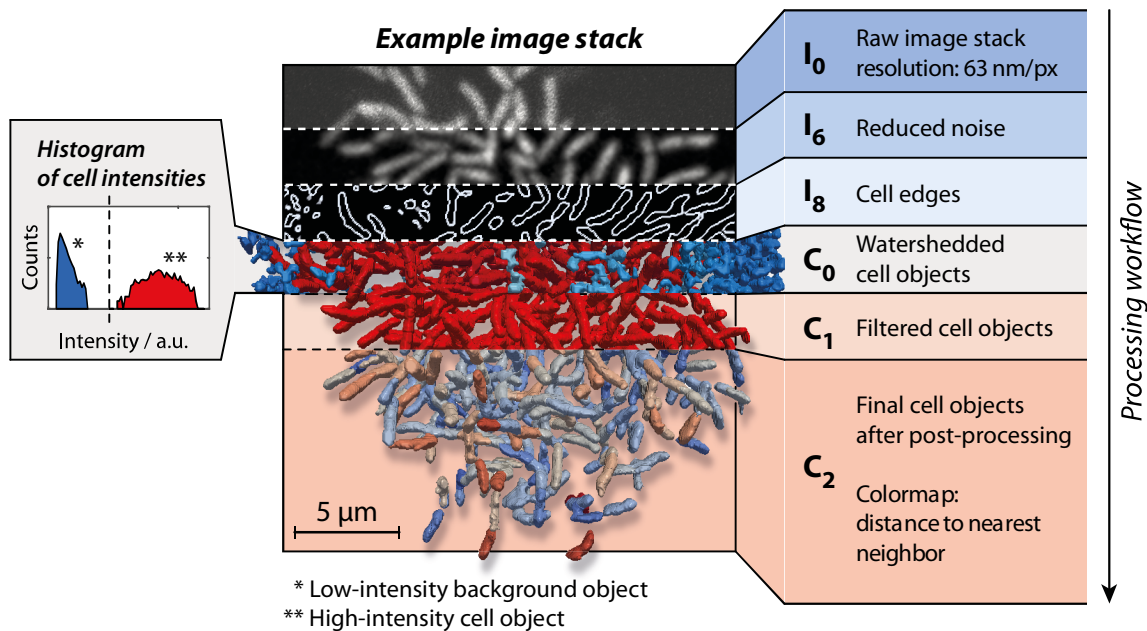
Supplementary Figure 1 – **Image processing pipeline**. Biofilms on spinning disc confocal fluorescence image stacks are processed as illustrated, to obtain 3D models of biofilms based on individual cells.

1.2. Cell detection

Image stacks, $I_6(t)$, were further processed to identify individual cells. First, the gamma value was adjusted to $\gamma = 3$ to enhance the range between signal and background (= I_7). Second, a 3D-edge detection was performed by convolving the data with a 3D Laplacian-of-Gaussian (LoG) kernel of 0.82 µm cube length (corresponding to 13 px) (= I_8). In the obtained matrices $I_8(t)$, zero values represent those locations where the cell's fluorescence intensity values drop approximately by 50% in comparison to its local intensity maxima. These zero crossings were determined to obtain closed surfaces representing the cellular outlines (= I_9). To distinguish between the "inside" (= cell) and "outside" (= background) of such a closed surface, a 3D-Gaussian filter kernel was applied to I_6 with the same spatial dimension as the LoG kernel used for the edge detection. The local maxima of the resulting image stack were identified (= I_{10}) and used as seeds to "flood-fill" the cell interior in I_9, yielding the binary image I_{11} containing all cell clusters. Cell de-clumping was performed by 3D-watershedding. The intensity landscape used

2

for watershedding was obtained by enhancing the local maxima (I_{10}) inside I_6 by a factor of 10, followed by 3D-Gaussian smoothing and masking with I_{11} (= I_{12}). After watershedding, I_{12} was 3D-median filtered and binary cell objects (= C_0) were obtained and processed further, as described in the next sections.

1.3. Cell object processing

For the cell objects C_0 the average underlying intensities in stack I_7 were determined and assigned to each cell. As the edge-detection-based segmentation will also yield objects, which are background, a filtering step is required. Because the average intensities of such background objects (calculated based on I_7) are typically orders of magnitude lower than those of actual cells, the cell objects C_0 were filtered based on these intensity measurements in stack I_7 (= C_1). After neglecting background objects, segmentation results were corrected based on C_1. Over-segmentation was corrected by merging very small cells below a certain volume threshold with neighbouring cells of largest contact area. Clumped cells in multicellular structures (characterized by large volume and low convexity) were dissected by fitting a Gaussian mixture distribution model (GMM) with *N* components (or number of underlying cells) to the corresponding voxel coordinate cloud. *N* was determined by counting the number of skeleton branches above a typical length threshold. The additional cells obtained after GMM clustering were checked for size. Very small objects were merged again with the neighbouring cell of largest contact area; the remaining de-clumped cells were used to replace the original multicellular structure. In addition, cells touching the image edges were removed. The cellular orientations and cell dimensions (length, height and width) were obtained by analysing the principal components (found by principal component analysis, PCA) of each cell volume (in other words: by fitting an ellipsoid into each cell). For additionally calculated cell features see the section on "Biofilm features". Final cell objects are referred to as C_2 (see Supplementary Fig. 1).

1.4. Cell tracking

To determine the single-cell growth rate, the resulting cells were tracked over successive time points. For time point t_i, the parent of each cell was determined by comparison with time point t_{i-1}, and finding the corresponding cell for which volume overlap was maximized and deviations in cell orientation were minimized. All cells that were not related to the biofilm founder cells were excluded from further analysis (at low flow rates a layer of non-related cells is usually formed after longer imaging, *cf.* white cells in Fig. 1b,d and Supplementary Fig. 6), resulting in the analysis of purely clonal biofilms. The accuracy of the segmentation was investigated by segmenting synthetic microscopy image stacks as described in the section on "Validation of biofilm segmentation".

1.5. Biofilm features

Name	Parameter	Unit	Description
Axial cell position	*height*	µm	Z-coordinate of each cell centroid.
Cell alignment with direction of flow	*angle (n_{cell}, flow)*	rad	Angle between each cell's major axis (see *cell size and orientation*) and the direction of the flow.
Radial alignment	*angle (n_{cell}, r)*	rad	Angle between each cell's major axis (see *cell orientation*) and the radial vector *r* pointing from the centre of mass pinned down to the bottom of the biofilm ($z = 0$ µm) to the corresponding cell.
Vertical alignment	*angle (n_{cell}, z)*	rad	Angle between each cell's major axis (see *cell orientation*) and **z**.
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	$n_{cell} = (e_1, e_2, e_3)$		Vector of the cell's major axis determined by PCA (eigenvector with largest eigenvalue).
Distance to centre of mass	d_{centre}	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 centre of mass projected onto the $z = 0$ µm plane.
Distance to nearest neighbour	$d_{nearest\ neighbour}$	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}	µm^{-3}	Number 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 $S = <3/2(n_i \cdot n_j)^2 - 1/2>$ in a vicinity defined as a sphere of radius 3 µm around a cell. n_i and n_j refer to the orientation vectors of cells *i* and *j*, respectively[1].
Single cell growth rate	*growth rate*	µm^{3}s^{-1}	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_{N+1} to zero for cells with no children and setting v_N to zero for cells with no parents. Using this approach, the global biofilm volume $V(t)$, as determined by segmentation, matches $V(t) = \int \sum \Delta v_i dt$.

Supplementary Table 1 – List of calculated single-cell features.

Name	Parameter	Unit	Description
Aspect ratio	Z/XY		Biofilm height divided by average biofilm base radius.
Biofilm base circularity	B_{circ}		Deviation of biofilm base cross section from a circle $B_{circ} = |1\text{-}Z/XY|$.
Global cell density	$\rho_{biofilm}$	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

Supplementary Table 2 – List of calculated global biofilm features.

1.6. Optical flow

Pre-processed image stacks (I_5, for details see section on "Image processing") were down-sampled to 1/4 of the initial resolution. The optical flow (= biovolume velocity) vector field (u_x, u_y, u_z) of isolated growing biofilms was determined for each containing voxel using the Horn-Schunck method[2] implemented in MatLab by Mustafa *et al.*[3]. Afterwards, the optical flow was set to zero in the space that did not contain cells. To investigate the effect of the external flow rate on net biomass movement, in terms of moving biovolume, the total biovolume flow through defined planes was calculated (see Supplementary Fig. 2). For external flow of growth medium v_{liquid} pointing in the positive y-direction, the total biovolume flow $I_{BV}^{\parallel}$ through the xz-plane for different y-coordinates, and the total biovolume flow $I_{BV}^{\perp}$ through the yz-plane for different x-coordinates, respectively, was calculated by summation over the biovolume fluxes through the corresponding planes (Eq. (1) and Eq. (2), Supplementary Fig. 2).

The biovolume flow I_{BV} through xz-planes (with plane normal vectors $\hat{\boldsymbol{n}}_y$) parallel to v_{liquid} for different values of y_i is:

$$I_{BV}^{\parallel}(y_i, t) = \sum_{x,z,y=y_i} \hat{\boldsymbol{n}}_y \cdot (u_x(t), u_y(t), u_z(t)) \tag{1}$$

The biovolume flow I_{BV} through yz-planes (with plane normal vectors $\hat{\boldsymbol{n}}_x$) perpendicular to v_{liquid} for different values of x_i is:

$$I_{BV}^{\perp}(x_i, t) = \sum_{x=x_i,z,y} \hat{\boldsymbol{n}}_x \cdot (u_x(t), u_y(t), u_z(t)) \tag{2}$$

To calculate the *biomass shift* (displayed in Fig. 3f of the main paper), the biovolume flow I_{BV} was summed up either along y or x-direction, was normalized by the sum of the absolute values, and the result was averaged over all time points:

$$biomass\_shift^{\parallel} = \left< \frac{\sum_{y_i} I_{BV}^{\parallel}}{\sum_{y_i} |I_{BV}^{\parallel}|} \right> \tag{3}$$

$$biomass\_shift^{\perp} = \left< \frac{\sum_{x_i} I_{BV}^{\perp}}{\sum_{x_i} |I_{BV}^{\perp}|} \right> \tag{4}$$

In Fig. 3f of the main paper, values are given as mean ± standard deviation for $n = 3$ individual biofilms. For better visualization in Supplementary Fig. 2b, the *averaged* biovolume velocity through the xz- and yz-planes is shown, whereas for Fig. 3f the *summed* velocities were used.
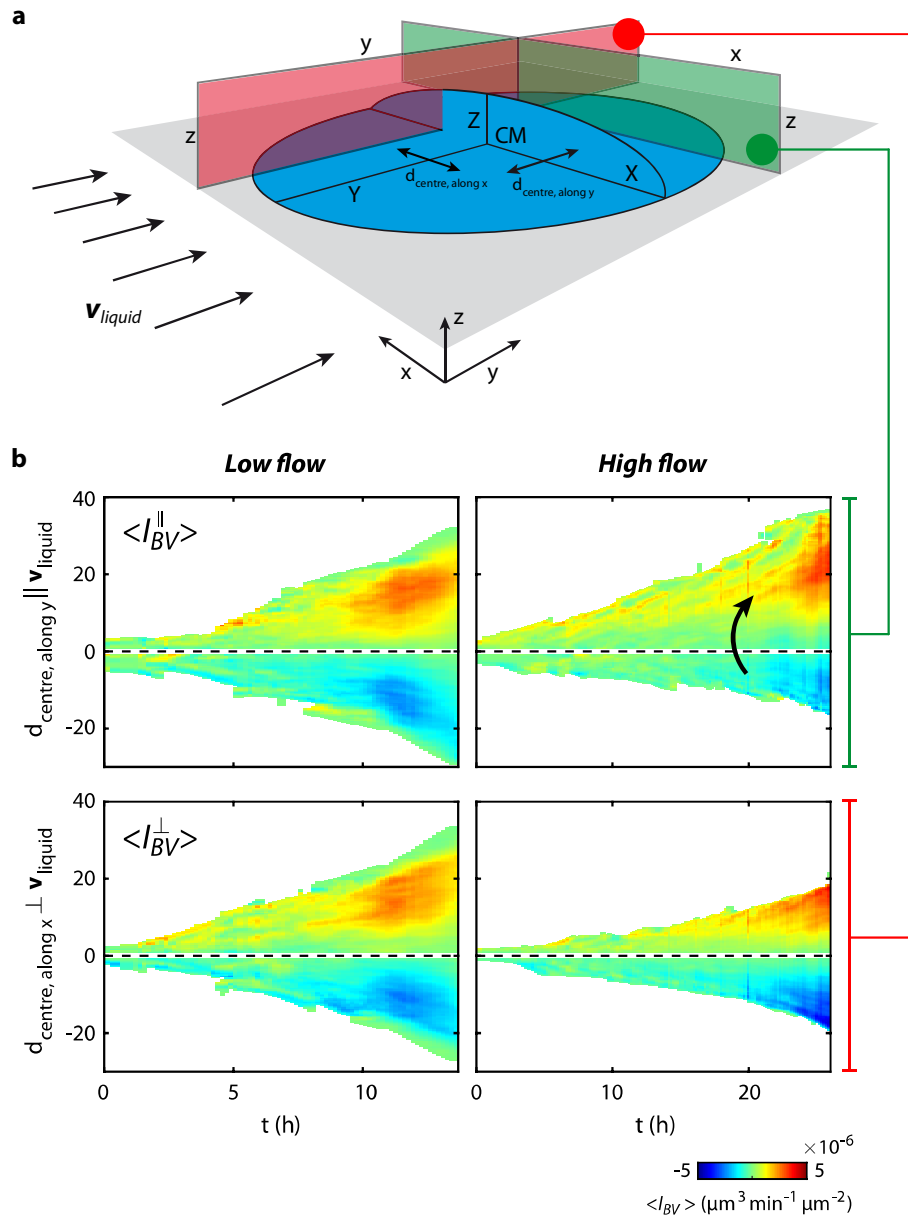
Supplementary Figure 2 – Optical flow measurements of biomass movement inside a biofilm. a, The "optical flow" (= biovolume flow) of a growing biofilm (blue) through the green plane (perpendicular to the direction of flow v_{liquid}, indicated with black arrows) and through the red plane (parallel to the direction of v_{liquid}) was determined. **b**, In panel **a**, the positions of the red and green plane with respect to the centre of mass (CM) projected down to the bottom of the biofilm were varied (red plane: d_{centre}, along x; green plane: d_{centre}, along y). For each time point and value of d_{centre} the average biovolume flow through the corresponding plane is shown as coloured pixel in the heatmap. In the heatmap, red values indicate an average net flux in the positive x or y direction, whereas blue values indicate an average net flux in negative directions. The two heatmaps in the upper row show the biovolume flow with respect to the green plane perpendicular to the direction of the external flow v_{liquid} with distance d_{centre}, along y to the centre. At high flow speed more biomass is moving in the direction of v_{liquid} than at low flow speed, as indicated by the black arrow, and the asymmetry along the $d_{centre} = 0$ axis. The two heatmaps in the bottom row show the flux through the red plane parallel to v_{liquid} (with distance d_{centre}, along x to the centre). The two heatmaps in the bottom row show that irrespective of the flow speed, biofilms grow symmetrically in the direction perpendicular to the flow, as indicated by the symmetric heatmaps around the $d_{centre} = 0$ axis.

1.10. 3D visualization

Segmented biofilms were exported using the mVTK library[4] and rendered in ParaView 5.1.2 (Kitware) using OSPRay rendering with shadows.

Visualization of RbmA

FLAG-labelled RbmA was detected in an additional fluorescence channel. The fluorescence distribution was rendered as a semi-transparent cloud and overlaid with the reconstructed biofilm in ParaView (see Fig. 1c).

1.11. Space-time heatmaps of biofilm development

In Fig. 1, single cell parameters for the WT* and Δ*rbmA* mutant biofilms are visualized spatially and temporally resolved in heatmaps (standard deviations of the presented mean values are shown in Supplementary Fig. 5 and Supplementary Fig. 7, respectively). For these space-time heatmaps, the cells were binned with respect to a spatial descriptor (vertical height z, distance to surface $d_{surface}$, or distance to centre of mass d_{centre}) and the corresponding cell property values were averaged for this particular spatial descriptor across the biofilm at a particular time t, as shown schematically in Supplementary Fig. 3.
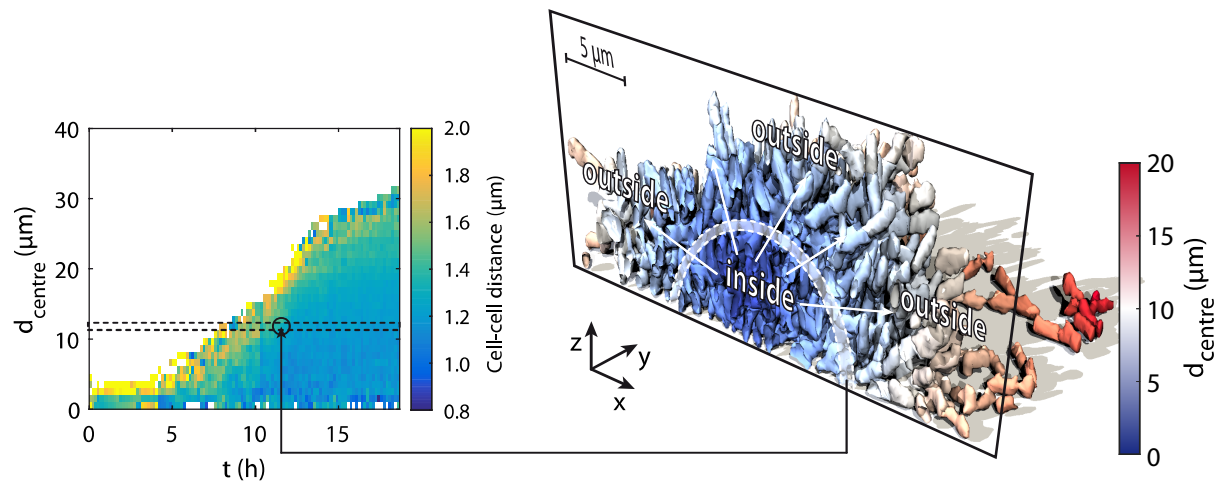
Supplementary Figure 3 – Illustration showing how a single tile in the heatmap (left panel) was generated. In this heatmap, cell-cell distance (color-coded) is measured as a function of the distance to the centre of mass (CM) of the biofilm pinned down to the bottom (d_{centre}), shown on the y-axis of the heatmap. For each time point, all cells inside the biofilm are grouped according to their distance to the CM, as indicated by the white dashed lines in the 3D rendering, approximately corresponding to the black dashed lines in the heatmap. For all cells having a similar value of d_{centre} the cell-cell distances are averaged (right panel) and visualized as a coloured tile in the left panel.

## 1.12. Liquid crystal (Q-tensor) representation of biofilms

To visualize the average cellular local order and alignment, an evenly spaced 3D-grid with 5 µm spacing was overlaid on the biofilm. For each grid node the uniaxial tensorial order parameter $\mathbb{Q}_i = 1/2(3\boldsymbol{n_i} \otimes \boldsymbol{n_j} - \mathbb{I})$ was calculated[5] for each cell i in a 2.5 µm vicinity, where $\boldsymbol{n_i}$ denotes the cell orientation vector. Based on the individual cell tensors the average Q-tensor and its eigenvalues and eigenvectors were determined for each node. In Fig. 3g the resulting values are represented as ellipsoids, where the largest eigenvalue and corresponding eigenvector are represented as the length and major axis orientation of each ellipsoid, respectively. The colour of each ellipsoid corresponds to the angle between the major axis vector $\boldsymbol{n}$ of each ellipsoid and the direction of the flow.

7

1.13. Validation of biofilm segmentation

To benchmark the single-cell segmentation and tracking algorithms, the individual cell-based simulations (See section 3 of this document) were used to synthesize biofilm image stacks, which were convolved with the point spread function of the microscope we used and distorted by the addition of noise, to obtain data as similar as possible to the experimental data. In detail, a simulation up to $N = 2,000$ cells was performed using parameters which correspond to the $\Delta rbmA$ mutant (for biofilm renderings see Supplementary Fig. 4a, for a description of the simulation framework see the section on "Individual Cell-Based Simulations").

The simulated biofilm data were converted into image stacks with a spatial resolution of 62.3 nm/pixel and a temporal resolution of approximately 10 min. The image stacks were blurred using a theoretical point spread function (PSF) corresponding to our microscopy setup (Supplementary Fig. 4b) (Huygens software, Scientific Volume Imaging) and down-sampled along z to match the axial resolution of the biofilm data (0.4 µm/pixel). Finally, the intensity levels were adjusted to the experimental data. To adjust the intensity levels, the average background intensity and the average cell intensity, depending on the axial imaging position z, were extracted from the experimental data. The average background intensity was independent of z ($I_{background} \approx 500$). To record the average cell intensity per plane, the corresponding intensity values were sorted and the highest 2000 values were averaged. The resulting curve $I_{cell}(z)$ was fitted with a bi-exponential function (Supplementary Fig. 4c), which was used to normalize the synthetic image stacks. Poisson-distributed noise was added to mimic noise due to photon detection inside the camera. The noise levels were estimated and generated using the algorithm proposed by Liu *et al.*[6]. The generated images (Supplementary Fig. 4d) were processed as described in the section on "Image processing".

In Supplementary Fig. 4e-h the simulated dataset is compared with the reconstructed one in terms of cell number (Supplementary Fig. 4e), cell displacements (Supplementary Fig. 4f), differences in cellular orientation (Supplementary Fig. 4g), and internal order (Supplementary Fig. 4h). Based on this validation procedure, we determined that our image segmentation algorithms yielded accurate cell segmentation for > 95% of all cells.
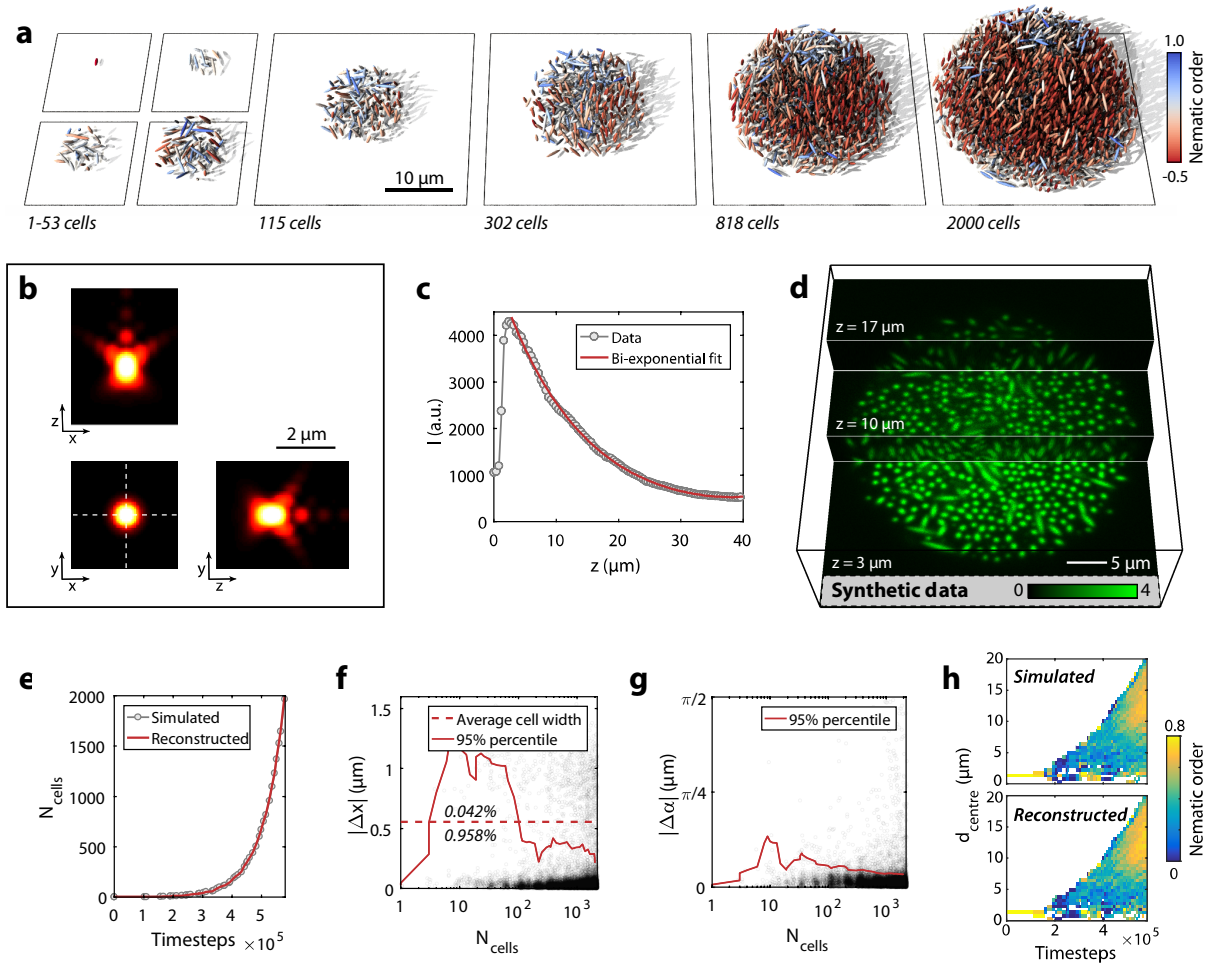
Supplementary Figure 4 – Validation of the single-cell segmentation algorithms using synthetic images. a, Time series of a rendered simulated biofilm, which was used to synthesize microscopic image stacks. **b**, Theoretical point spread function (PSF) of our experimental setup, used for convolution with the simulation data to obtain a more realistic dataset (see **d**). **c**, Experimentally determined typical average cell intensity curve $I_{cell}(z)$ versus the axial position z of the imaged plane. This intensity profile was used to introduce an intensity gradient along z in the synthetic image stacks (see **d**). **d**, Artificial image stacks obtained after convolution with the characteristic PSF (from **b**), intensity fading along z (from **c**), and distortion by typical Poisson-distributed detector noise. **e**, Simulated (grey) vs. reconstructed (red) cell numbers N_{cell}. **f**, Cell displacements Δx after reconstruction. The simulated and reconstructed data was overlaid. For each cell in the simulated dataset, the distance Δx to the closest cell in the corresponding reconstructed data was calculated. 95.8% of all reconstructed cells (for all time points) were closer than one cell width (dashed line) to the theoretical coordinates. Per time point, 95% of all cells deviated less than indicated by the solid red curve. **g**, Differences in cellular orientation (expressed as angle between the major axis of a cell in the simulated dataset and the major axis of the nearest cell in the reconstructed data $\Delta\alpha$). Per time point, 95% of all cells showed a difference in orientation smaller than indicated by the solid red curve. **h**, Spatially resolved distribution of the internal biofilm order of the simulated (top) and reconstructed (bottom) data.

9

2. Detailed Characterization of Biofilm Architecture

2.1. WT* biofilm structure

Following the growth dynamics of the WT* in an unperturbed low-flow environment, we found strong structural differences between the outer biofilm layer and the central part, and several distinct structural phases of the biofilm during growth (Fig. 1e). For each time point we characterized the biofilm spatially with respect to order, density, alignment, and growth to obtain a quantitative description of the WT* phenotype and standard deviations of the properties (Supplementary Fig. 5). Small biofilms of less than 50 cells were generally characterized by a relatively high order, and a low vertical alignment, indicating 2D growth. Then, a transition to three-dimensional growth occurred as indicated by an increasing vertical alignment. Generally, order and cell-cell spacing increased with distance from the biofilm centre for all biofilm sizes. The order at the centre of the biofilm decreased as the cell number increased up to biofilms with more than 1,000 cells, when the order in the centre increased again in agreement with previous structural analyses[1]. This increase in order coincided with a local decrease in cell-cell spacing and a strong local increase in vertical cellular alignment. The radial alignment was higher at the outer parts of small- to medium-sized biofilms and decreased as the biofilms expanded. Surprisingly, the local growth rate, as measured by cell tracking, was constant in space and time, indicating that growth inside the observed biofilms was not limited by diffusion constraints of nutrients in our conditions.
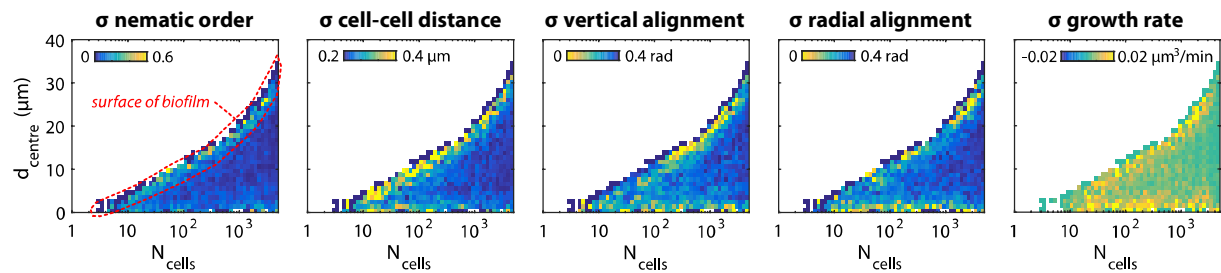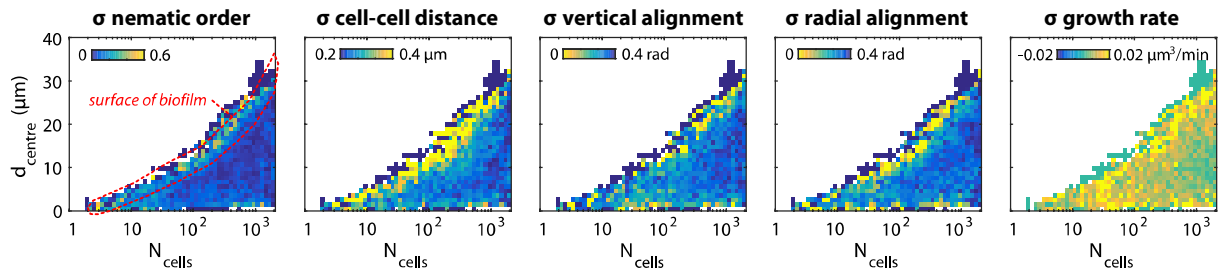

Supplementary Figure 5 – Standard deviations of WT*-data shown in Fig. 1e for $n = 4$ biofilms (characterization of WT* biofilm architecture in an unperturbed low-flow environment)**.**

2.2. ΔrbmA biofilm structure

We internally perturbed our model system by varying the strength of the attractive interaction. This was achieved by modifying the extracellular RbmA levels with an arabinose-inducible promotor (strain KDV1082, with genotype Δ$rbmA$, P_{BAD}-$rbmA$). Biofilms lacking RbmA were more fragile and displayed an increased cell-cell spacing and increased order, in agreement with previous studies (Supplementary Fig. 6, Fig. 1f, Supplementary Movie 2)[1,7]. In comparison to the WT*, the order and vertical alignment inside the biofilm centre was higher, except in the bottom-most layer, where cells were on average oriented parallel to the surface (Fig. 1f). In summary, Δ$rbmA$ mutant biofilms can be partitioned into three parts: the highly ordered core (i) is surrounded by a ~5µm thick, disordered shell (ii), followed by the top outer layer (iii), which again has a slightly higher order. Cells inside the highly ordered core showed strong vertical alignment, whereas in the disordered shell and in the outer layer the vertical alignment was lost. The average cell-cell spacing was more than 30% higher than the WT*. Standard deviations of the properties in Fig. 1f are shown in Supplementary Fig. 7.
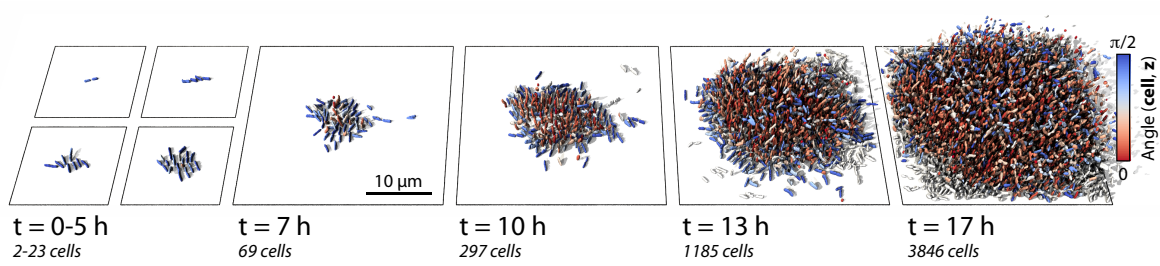
Supplementary Figure 6 – Time-resolved biofilm growth series of the Δ*rbmA*-mutant in a low-flow environment. Each cell is coloured according to the cellular alignment with the *z*-axis pointing away from the substrate plane.
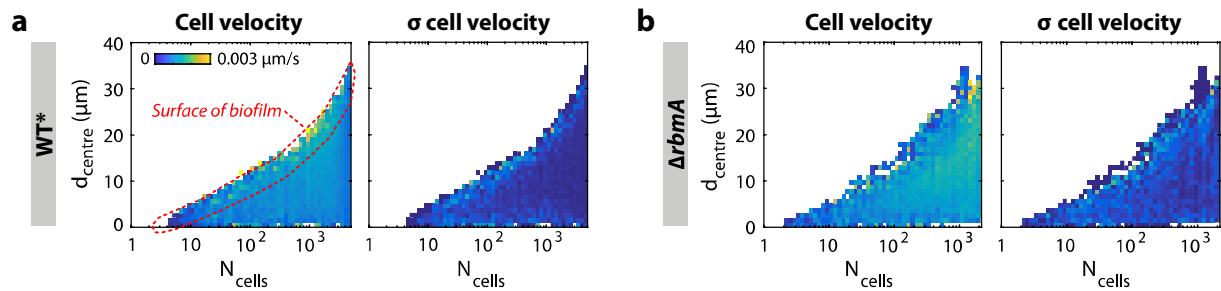

Supplementary Figure 7 – Standard deviations of Δ*rbmA*-data shown in Fig. 1f for $n = 4$ biofilms (characterization of Δ*rbmA*-mutant biofilm architecture in a low-flow environment)**.**
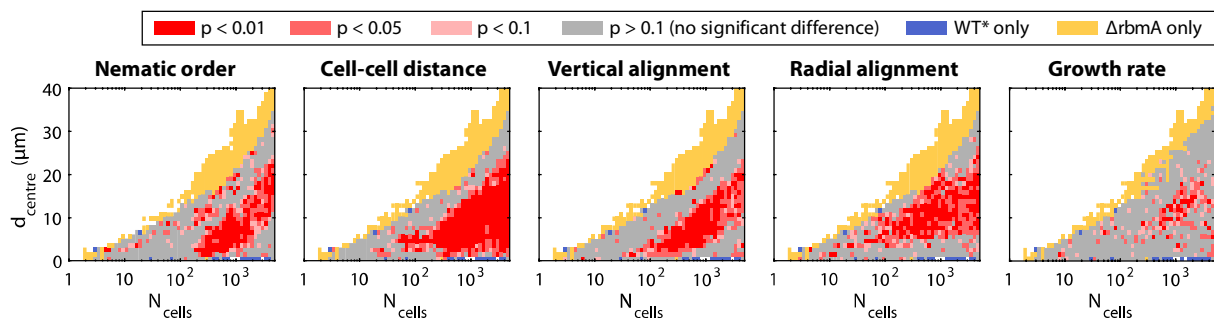
Supplementary Figure 8 – Spatiotemporal differences between WT* and Δ*rbmA* biofilm architecture in a low-flow environment in relation to Fig. 1e,f. The data was compared using a two-tailed t-test, yielding *p*-values.

2.3. Cell motility inside biofilms

V. cholerae cells in rugose biofilms do not display flagellar or type IV pili mediated motility due to high levels of c-di-GMP[8–10], which represses expression of flagella and type IV pili. Twitching motility based on type IV pili has also never been observed for *V. cholerae*[11]. As shown in Supplementary Fig. 9, the individual cell speed inside WT* and Δ*rbmA* biofilms is on the order of nm/s, i.e. much lower than flagella driven motility (~50 µm/s)[11] or pilus-based motility (~1 µm/s)[12], and increases slightly with increasing biofilm sizes. These data indicate that cell movement in *V. cholerae* biofilms is dominated by passive cell displacement due to biofilm expansion rather than active cell motility.

Supplementary Figure 9 – Cell motility inside WT* and Δ*rbmA* biofilms in a low-flow environment. a, Average centroid velocity after linking cells in successive image stacks by cell tracking, and corresponding standard deviation for the WT* (*n* = 4 biofilms). **b,** Corresponding data for the of Δ*rbmA* mutant (*n* = 4 biofilms).

2.4. Spatiotemporal expression of *rbmA* and VPS in WT* biofilms

The *V. cholerae* rugose strain N16961, which was used during this study, is known to continuously produce extracellular matrix, including RbmA, even if the biofilms are small, due to this strain's intrinsically high c-di-GMP levels[13]. The expression of *rbmA* and VPS in an unperturbed low-flow environment were determined in the WT* (strain KDV1027 and KDV1218) using transcriptional fusions of *mRuby3* to the *rbmA* promoter or to the *vpsI* promoter, respectively. To measure *rbmA* and VPS expression in biofilms, the ratio of the background-subtracted *rbmA-mRuby3* fluorescence and *VPS-mRuby3* fluorescence, respectively, were divided by the background-subtracted signal of the constitutively expressed *sfGFP*, per cell (see Supplementary Fig. 10).
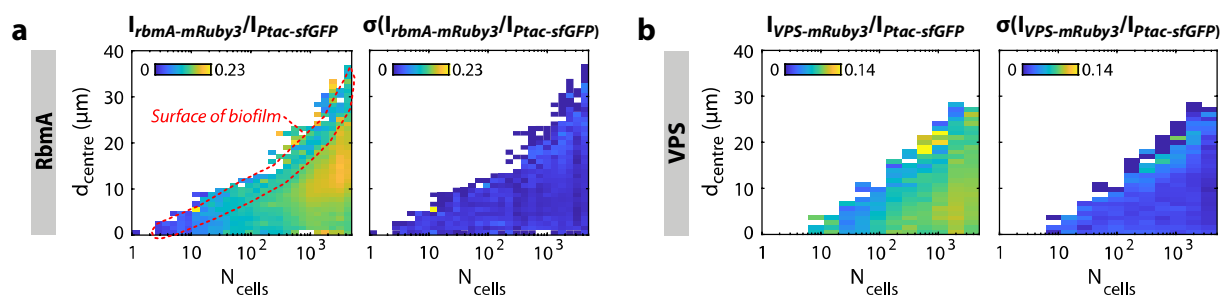

Supplementary Figure 10 – Spatio-temporal expression of *rbmA* and VPS in the WT* strain in a low-flow environment. a, Expression of *rbmA* (left: mean values, right: standard deviation, *n* = 3 biofilms). In each cell the fluorescence of *rbmA-mRuby3* was normalized by the signal of the constitutive P_{tac}-promoter-driven sfGFP-fluorescence signal. **b,** Expression of VPS (left: mean values, right: standard deviation, *n* = 6 biofilms). In each cell the fluorescence of *VPS-mRuby3* was normalized by the signal of the constitutive P_{tac}-promoter-driven sfGFP fluorescence signal.

2.5. Effect of *Vibrio* polysaccharide (VPS) on cell-cell interaction

To test whether VPS can contribute to cell-cell attraction, we compared the average cell-cell spacing in biofilms of cells with normal VPS production (smooth colony phenotype, Δ*rbmA*, KDV383) with a VPS

overproduction strain (rugose colony phenotype, Δ*rbmA*, KDV692). Transcriptome comparisons of smooth and rugose strains indicate strong induction of VPS production in the rugose strain[10]. Comparing the smooth WT and the smooth Δ*rbmA* mutant (or the rugose WT and the rugose Δ*rbmA* mutant), we observe an increase in cell-cell spacing due to lack of RbmA. Comparing the cell-cell spacing for the smooth Δ*rbmA* and rugose Δ*rbmA* (VPS overproduction) strains, we find a similar cell-cell spacing (Supplementary Fig. 11, red bars), indicating that increased VPS production does not mediate an increased cell-cell attraction. In addition, data for the smooth and Rugose WT (KDV103, and, KDV615, respectively) show that enhanced VPS production even slightly increases the cell-cell spacing (Supplementary Fig. 11, blue bars). Increased VPS production therefore does not increase cell-cell attraction.
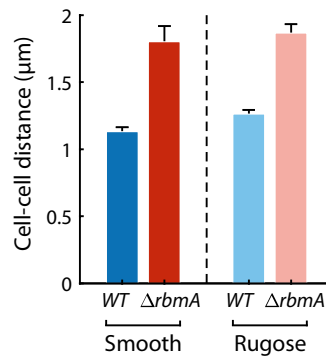
Supplementary Figure 11 – Effect of VPS overproduction on cell spacing in a low-flow environment. The "cell-cell distance" measurement refers to the average cell centroid-centroid distance. Rugose variants produce more VPS compared with smooth strains[10]. Enhanced VPS production does not lead to a decreased cell-cell spacing, indicating that enhanced VPS production is not causing an increased cell-cell attraction. Error bars correspond to the standard error ($n \geq 3$ biofilms).

2.6. Expression of *rbmA* at low vs. high shear rates

The expression of *rbmA* at low shear rate ($\dot{\gamma} = 2$ s^{-1}) and high shear rate ($\dot{\gamma} = 2000$ s^{-1}) was measured in the WT* strain (KDV1027) by calculating the ratio of the background-subtracted fluorescence of the *rbmA-mRuby3* transcriptional reporter, divided by the background-subtracted signal of the constitutively expressed sfGFP, per cell (see Supplementary Fig. 12). Individual cell measurements were averaged per time point and biofilm. Cells in biofilms growing at high shear rates were found to express *rbmA* at higher levels than those in low shear rates, for small biofilms of 10-50 cells, when all cells experience an increased level of shear and there is little deflection of the flow by cells and matrix on the biofilm surface.
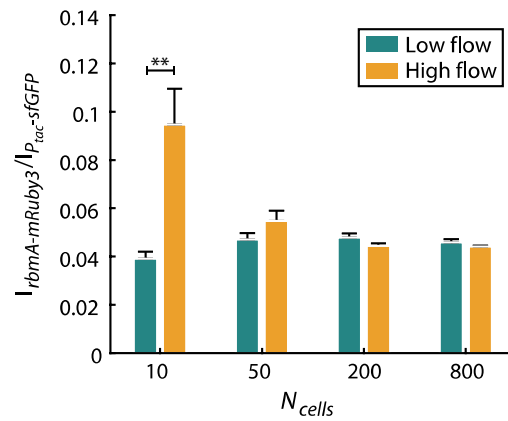
Supplementary Figure 12 – Shear rate-dependent expression of *rbmA*, for different biofilm sizes. In each cell the fluorescence intensity signal of the *rbmA-mRuby3* transcriptional reporter was normalized by the signal of the constitutive P_{tac}-promoter-driven sfGFP fluorescence signal. Error bars correspond to the standard error ($n = 6$ biofilms). Statistical significance: ** is $p < 0.01$ (Mann–Whitney–Wilcoxon test).

2.7. Calibration of arabinose-induced expression of *rbmA*

The expression of *rbmA* (and therefore cell-cell attraction strength) was controlled by the arabinose-inducible promotor P_{BAD} which was reported to be bi-stable for cells in liquid culture[14,15]. However, for using this expression system for our purposes inside biofilms, a homogenous expression pattern is required. Therefore, we analyzed the expression pattern of P_{BAD}-*mRuby3* among individual cells (KDV1228) after 4 h of arabinose-induction (see Methods section in the main manuscript) in (a) shaken liquid cultures of individual cells, and (b) biofilm-associated cells (KDV1231) inside biofilms (300-500 cells) for different levels or arabinose (Supplementary Fig. 13). Bi-stable expression of P_{BAD} appears to be a characteristic behavior of individual cells in liquid culture, but not of biofilm-associated cells, as only few cells were found to be non-responsive to arabinose in biofilms (Supplementary Fig. 13).
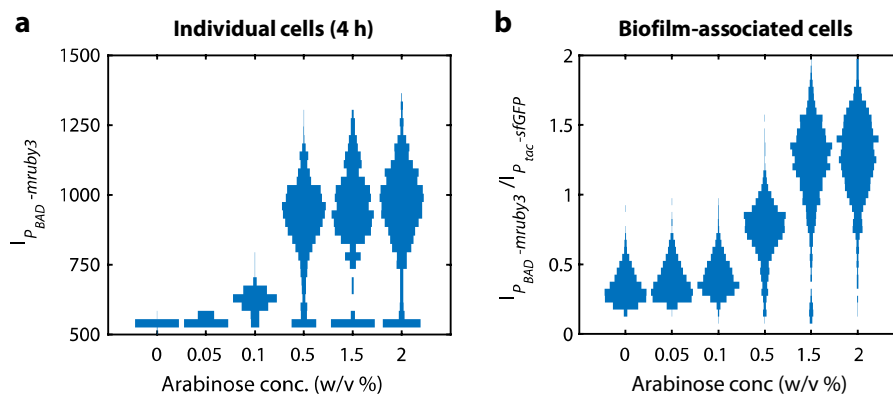
Supplementary Figure 13 – Stability of the expression of the arabinose-inducible promotor P_{BAD}. a, Bi-stable expression of P_{BAD} inside individual cells in shaken liquid culture, 4 h after inoculation. **b,** In contrast, a homogenous expression of P_{BAD} is observed for biofilm-associated cells. Data is shown as "violin plot" distributions.

3. Individual Cell-Based Simulations

3.1. Model description and implementation

The cells are modelled as interacting ellipsoids of half-length ℓ and half-width r, described by their position x and orientation $\hat{\boldsymbol{n}}$. Because cells live at low Reynolds number ($Re \approx 10^{-4}$), we approximate the dynamics as over-damped, ignoring any inertial effects. Cells can interact with the wall boundary and other cells through interaction potential functions, U_{bdy} and V. Denoting the identity matrix by $\boldsymbol{I}$, the over-damped translational and orientation dynamics for a single cell are

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{\Gamma}^{-1}\left[-\frac{\partial U_{bdy}}{\partial \boldsymbol{x}} - \frac{\partial V}{\partial \boldsymbol{x}}\right] \tag{5}$$

$$\frac{d\hat{\boldsymbol{n}}}{dt} = (\boldsymbol{I} - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^T)\left[\boldsymbol{\Omega}^{-1}\left(-\frac{\partial U_{bdy}}{\partial \hat{\boldsymbol{n}}} - \frac{\partial V}{\partial \hat{\boldsymbol{n}}}\right)\right] \tag{6}$$

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

$$\boldsymbol{\Gamma} = \gamma_m[\gamma_\parallel(\hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^T) + \gamma_\perp(\boldsymbol{I} - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^T)] \tag{7}$$

$$\boldsymbol{\Omega} = \omega_m\omega_R\boldsymbol{I} \tag{8}$$

Here, γ_m and ω_m are the typical translational and rotational drag coefficients for Stokes' drag in the extracellular matrix for a spheroid ($\gamma_m = 6\pi\mu_m r$, $\omega_m = 8\pi\mu_m lr^2$). Surface adhesion of the cells is captured by increasing the magnitude of the friction tensor $\boldsymbol{\Gamma}$ by a factor of 30 if the cells are within $1.5 \cdot r$ of the surface. $\gamma_\parallel$, $\gamma_\perp$ and ω_R are dimensionless geometric parameters characterizing the longitudinal and transverse friction parameters that depend only on the aspect ratio $a = \ell/r$ of the cell, as given by Han *et al.*[16]:

$$\gamma_\parallel = \frac{8/3}{\dfrac{2a}{1-a^2} + \dfrac{2a^2-1}{(a^2-1)^{3/2}}\ln\left(\dfrac{a+\sqrt{a^2-1}}{a-\sqrt{a^2-1}}\right)} \tag{9}$$

$$\gamma_\perp = \frac{8/3}{\dfrac{a}{a^2-1} + \dfrac{2a^2-3}{(a^2-1)^{3/2}}\ln(a+\sqrt{a^2-1})} \tag{10}$$

$$\omega_R = \frac{2}{3}\frac{a^4-1}{a\left[\dfrac{2a^2-1}{\sqrt{a^2-1}}\ln(a+\sqrt{a^2-1}) - a\right]}, \tag{11}$$

An important property of our proposed orientation dynamics is that $\hat{\boldsymbol{n}} \cdot d\hat{\boldsymbol{n}}/dt = 0$, such that the unit length of $\hat{\boldsymbol{n}}$ is conserved.

The interaction between a cell and the wall boundary is modelled with a repulsive interaction potential, U_{bdy}, that is proportional to the overlap between a cell and the wall boundary. The wall boundary is represented as a plane. To determine this overlap, an overlap coordinate, z_o, is introduced such that $z_o < 0$ implies no contact with the boundary and $z_o > 0$ implies contact with the boundary. The overlap coordinate is defined as

$$z_o = \ell|\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{N}}| + r - \hat{\boldsymbol{N}} \cdot (\boldsymbol{x} - \boldsymbol{S}) \tag{12}$$

where $\hat{\boldsymbol{N}}$ is the unit normal of the plane and $\boldsymbol{S}$ is a point on the plane. In the simulations, we set $\hat{\boldsymbol{N}} = [0, 0, 1]$ and $\boldsymbol{S} = [0, 0, 0]$ such that the wall is an xy plane located at the origin. Thus, repulsion from the wall can be represented by

$$U_{bdy} = \begin{cases} 0 & if\ z_o \leq 0 \\ \epsilon_{bdy} \exp\left(\dfrac{z_o}{\sigma_{bdy}}\right) & if\ z_o > 0 \end{cases} \tag{13}$$

ϵ_{bdy} captures the magnitude of the cell-boundary interaction, and σ_{bdy} is a scale factor of order of the half-width of the cell.

V is the total potential of a single cell α for all the N pairwise cell-cell interactions between cell α and cell β $\left(V = \sum_{\beta=1,\beta\neq\alpha}^{N} U\right)$. The interaction between cell α and cell β is governed by the cell-cell interaction potential introduced in the section "Cell-cell interaction potential".

The instantaneous cell length growth rate for a single cell is

$$\frac{d\ell}{dt} = \frac{\ell}{\tau_g}\ln(2) \tag{14}$$

where ℓ is the half-length of the cell at time t and τ_g is the growth time constant (obtained by experimental measurements). The cell width is constant throughout the simulation. Following the Adder model described by Taheri-Araghi et al.[17], 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. At division, a random number M is drawn from a normal distribution with mean 0.5 and standard deviation $\sigma_{birth\ size}$. Suppose ℓ_p is the length of the cell before division, then the birth lengths of the daughter cells are $M\ell_p$ and $(1 - M)\ell_p$ such that the length of the parent cell is conserved. The new orientations of the daughter cells are drawn from a von Mises-Fisher distribution, with mean direction corresponding to the direction of the parent cell and concentration parameter κ.

If we use r, $\tau_t = \frac{\gamma_m r^2}{\epsilon_0}$, and ϵ_0 as characteristic length, time, and energy scales, we can recast Eq. (5), (6) and (14) in dimensionless form

$$\frac{d\boldsymbol{x}^*}{dt^*} = \left[\frac{1}{\gamma_{\parallel}}(\hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^T) + \frac{1}{\gamma_{\perp}}(\boldsymbol{I} - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^T)\right]\left[-\frac{\partial U_{bdy}^*}{\partial \boldsymbol{x}^*} - \frac{\partial V^*}{\partial \boldsymbol{x}^*}\right] \tag{15}$$

$$\frac{d\hat{\boldsymbol{n}}}{dt^*} = (\boldsymbol{I} - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^T)\left[\frac{3}{4a\omega_R}\left(-\frac{\partial U_{bdy}^*}{\partial \hat{\boldsymbol{n}}} - \frac{\partial V^*}{\partial \hat{\boldsymbol{n}}}\right)\right] \tag{16}$$

$$\frac{d\ell^*}{dt^*} = \frac{\tau_t}{\tau_g}\ell^* \ln(2) \tag{17}$$

where superscript * indicates a dimensionless quantity and use has been made of the following definitions and ratios $\boldsymbol{x}^* = \boldsymbol{x}/r$, $\ell^* = \ell/r$, $t^* = t/\tau_t$, $V^* = V/\epsilon_0$ and $U_{bdy}^* = U_{bdy}/\epsilon_0$. Note that τ_t can be interpreted as the translational 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 $z_o^* > 0$, the dimensionless

boundary potential is $U_{bdy}^* = U_{bdy}/\epsilon_0 = \epsilon_r \exp(z_o{}^*/\sigma_{bdy}^*)$ where $\epsilon_r = \epsilon_{bdy}/\epsilon_0$, $z_o{}^* = z_o/r$, and $\sigma_{bdy}^* = \sigma_{bdy}/r$.

A custom, highly parallelized individual cell-based code employing graphics processing units (GPUs) was developed to perform the simulations. At each time step, we calculate cell-cell interactions using the all-pairs approach[18] such that all pair-wise interactions are evaluated. We use a standard explicit Euler scheme to numerically integrate the dimensionless translational and orientational dynamics, Eq. (15) and (16) and growth law Eq. (17).

The key simulation parameters used for the simulations are shown in Supplementary Table 3.

| Parameter	Value	Unit	Description
r	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 EPS matrix at room temperature[19].
γ_m	5.23	pN·s·µm^{-1}	Typical drag coefficient for Stokes' drag in EPS matrix $(\gamma_m = 6\pi\mu_m r)$.
ϵ_r	10		Ratio comparing the strength of the bacteria-boundary interaction to the strength of the bacteria-bacteria interaction $\epsilon_r = \epsilon_{bdy}/\epsilon_0$.
σ_{bdy}^*	1		Non-dimensional boundary potential length scale parameter.
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).
l_{add}^*	3.65		Non-dimensional length added to bacteria after division.
$\sigma_{birth\,size}^*$	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.
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 0.65·10^{-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).

Supplementary Table 3 – Simulation parameters. All parameters are measured directly from experimental biofilms, except for the potential parameters ν, $\lambda_a, \rho_a, \epsilon_0, \lambda_r$, which were determined by fitting as described in the main text.

3.2. Comparing simulations with experimental data

For the purpose of comparing simulated and experimental biofilms, a set of parameters was chosen to represent the phenotype and architecture as fully and accurately as possible. These parameters include averaged single cell parameters, i.e. local order, vertical alignment, radial alignment, cell-cell distance, local density, cell length, and the following global biofilm parameters: global density, biofilm aspect ratio, and biofilm base circularity (see section on "Biofilm features"). The single cell parameters were determined for the biofilm core ($d_{centre} < \max(d_{centre})/2$) and the biofilm shell ($\max(d_{centre})/2 < d_{centre} < \max(d_{centre})$). The time-evolution of these parameters was compared for biofilms of cell numbers ranging from 10 to 300. To account for logarithmic growth, biofilms were sampled at 40 intermediate logarithmically spaced cell numbers yielding a characteristic biofilm feature vector shown in Supplementary Fig. 14. In addition, each parameter was normalized to a typical range as indicated in Supplementary Table 4. These feature vectors therefore capture biofilm architectural properties (via the different architectural parameters), as well as the temporal biofilm development (via the measurement of these parameters at different cell numbers). The similarity between a vector characterizing a simulation and an experimental biofilm was assessed in terms of the mean square distance (*MSD*) between the two feature vectors. The different parameters were weighted differently to account for their relative importance in representing biofilm phenotypes, as summarized in Supplementary Table 4.

Parameter	Normalization range [min max]	Weight
Local order	0 – 0.8	10
Vertical alignment	0.2 – 1.37	10
Radial alignment	0.2 – 1.37	3
Cell-cell distance	0.8 – 2	20
Local density	0 – 0.2	10
Cell length	1 – 3	1
Global density	0 – 0.15	5
Biofilm aspect ratio (Z/XY)	0 – 0.8	1
Biofilm base circularity	0 – 0.8	1

**Supplementary Table 4 – Biofilm parameters and corresponding normalization ranges and weights.**
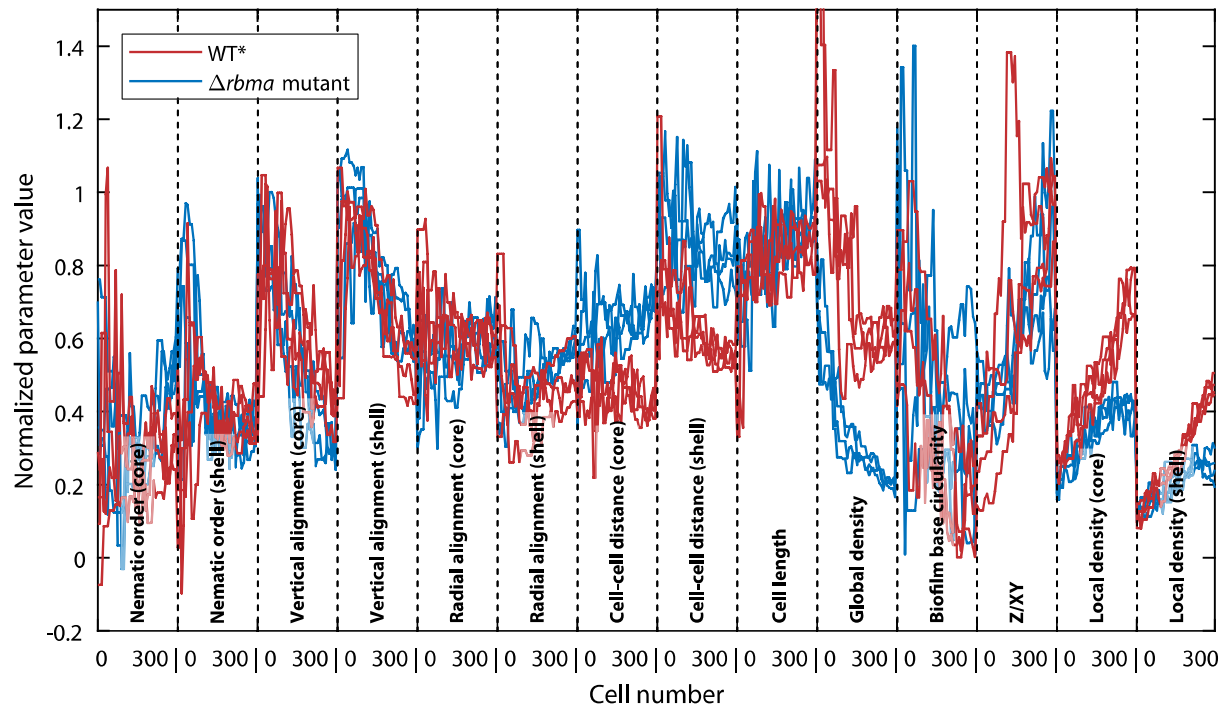
Supplementary Figure 14 – Characteristic biofilm feature vector for several biofilms of the WT* (red) and the Δ*rbmA*-mutant (blue). The parameters correspond to the ones listed in Supplementary Table 4, whereby relevant ones are spatially resolved (core and shell of the biofilm).

3.3. Cell-cell interaction potential

In our model, cells are subject to forces by neighbouring cells due to purely mechanical effects, including steric repulsion and the repulsion mediated by the effective osmotic pressure in the intercellular space, as well as the interaction with secreted matrix components (such as RbmA). While steric interactions and the osmotic pressure in the intercellular space are expected to result in net repulsive forces, RbmA is thought to directly link the cells together, thereby causing net cell-cell attraction[20,21], which we quantitatively confirmed (Fig. 2a). To account for these effects we suggest the cell-cell interaction potential between two particular cells, cell α and cell β, described by equation (18).

$$U = \epsilon_0 \epsilon_1 \left(e^{-\frac{\rho^2}{\lambda_r^2}} + \frac{\nu}{1 + e^{\left(\frac{\rho_a - \rho}{\lambda_a}\right)}} \right) \tag{18}$$

The first term in the interaction potential corresponds to cell-cell repulsion, and the second term corresponds to cell-cell attraction. The vector $\boldsymbol{r}_{\alpha\beta} = r_{\alpha\beta} \hat{\boldsymbol{r}}_{\alpha\beta}$ joins the cell centres and is directed from cell α to cell β, ϵ_0 describes the energy of the cell-cell interaction, ϵ_1 is a strength parameter accounting for the cell orientation configuration, $\rho = r_{\alpha\beta}/\sigma$ is the cell-cell distance normalized by the overlap shape-factor σ, λ_r is the repulsion width, ν is the attraction strength, ρ_a is the attraction shift and λ_a is the attraction width, as summarized in Supplementary Table 3 and Supplementary Fig. 15.
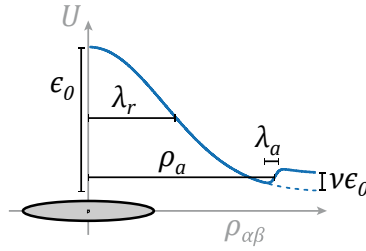
Supplementary Figure 15 – Effects of the cell-cell interaction potential parameters. ϵ_0: Strength of repulsion due to combined effects of steric and matrix repulsion. λ_r: Range of repulsion in cell diameters. ν: Relative strength of RbmA-mediated attraction. λ_a: Effective attractive range of RbmA in cell diameters. ρ_a: Effective distance of the attractive potential well.

The above strength and range parameters lead to different length scales of the cell-cell interaction forces depending on the orientation of the cells (see Fig. 2b,c, Supplementary Fig. 18, and Supplementary Fig. 19, respectively). The generalized forms of the strength and range parameters for non-identical ellipsoids are given by Cleaver *et al.*[22] and are reproduced below for convenience

$$\epsilon_1 = \left[1 - \chi^2 \left(\hat{\boldsymbol{n}}_\alpha \cdot \hat{\boldsymbol{n}}_\beta\right)^2\right]^{-1/2} \tag{19}$$

$$\sigma = \sigma_0 \left[1 - \chi \left(\frac{\xi\left(\hat{\boldsymbol{n}}_\alpha \cdot \hat{\boldsymbol{r}}_{\alpha\beta}\right)^2 + \xi^{-1}\left(\hat{\boldsymbol{n}}_\beta \cdot \hat{\boldsymbol{r}}_{\alpha\beta}\right)^2 - 2\chi\left(\hat{\boldsymbol{n}}_\alpha \cdot \hat{\boldsymbol{r}}_{\alpha\beta}\right)\left(\hat{\boldsymbol{n}}_\beta \cdot \hat{\boldsymbol{r}}_{\alpha\beta}\right)\left(\hat{\boldsymbol{n}}_\alpha \cdot \hat{\boldsymbol{n}}_\beta\right)}{1 - \chi^2\left(\hat{\boldsymbol{n}}_\alpha \cdot \hat{\boldsymbol{n}}_\beta\right)^2}\right)\right]^{-1/2} \tag{20}$$

where

$$\sigma_0 = \sqrt{r_\alpha^2 + r_\beta^2},$$

(21)

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

(22)

and

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

(23)

For two ellipsoids α and β, $\hat{\boldsymbol{n}}_\alpha$, $\hat{\boldsymbol{n}}_\beta$ are their axial unit vectors, ℓ_α, ℓ_β are their half lengths, and r_α, r_β are their half widths.

### *Influence of cell-cell repulsion on biofilm architecture*

Biofilms grown from $\Delta rbmA$ cells, which lack RbmA proteins, were simulated by setting the strength of the attractive part of the cell-cell potential to zero ($\nu = 0$), as cell-cell attraction is primarily due to the levels of RbmA. Before computing systematic parameter scans for obtaining the repulsion-parameters ϵ_0 and λ_r of the interaction potential U, we estimated the parameters using the physical considerations: The energy scale of the cell-cell interactions was estimated to be within several orders of magnitude of the energy scale of interactions between the cells and the flow. Therefore, we simulated biofilms for values of ϵ_0 ranging from $1 \cdot \epsilon_{flow}$ to $10^4 \cdot \epsilon_{flow}$, where the cell-flow interaction energy $\epsilon_{flow} = 0.005$ pN·µm was calculated by considering Stokes flow (with flow rate 0.1 µL/min in a channel with the dimensions used in the experiment) past a sphere with the typical cell volume of 0.4 µm^3, being locating 2.4 µm above a no-slip boundary (as is typical for cells in the outer shells of biofilms). A typical cell-cell distance in the core of biofilms, where cell-cell repulsion dominates, is around 1 µm, which corresponds to $\lambda_r = 1.4$ (for a typical overlap shape-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). Therefore, we used values of λ_r from 0.8 to 2. The resulting *MSD* values upon a systematic variation of the cell-cell interaction energy ϵ_0 and repulsion width λ_r are visualized in Supplementary Fig. 16. Corresponding biofilm architecture phenotypes are shown in Supplementary Fig. 17. The identified parameters for the best fit between simulations and the experimentally observed $\Delta rbmA$-phenotype are $\epsilon_0 = 10 \cdot \epsilon_{flow}$ ($5 \cdot 10^{-20}$ J) and $\lambda_r = 1.65$. The resulting translational and rotational forces are visualized in Fig. 2b, Supplementary Fig. 18a, and Supplementary Fig. 19a, respectively.
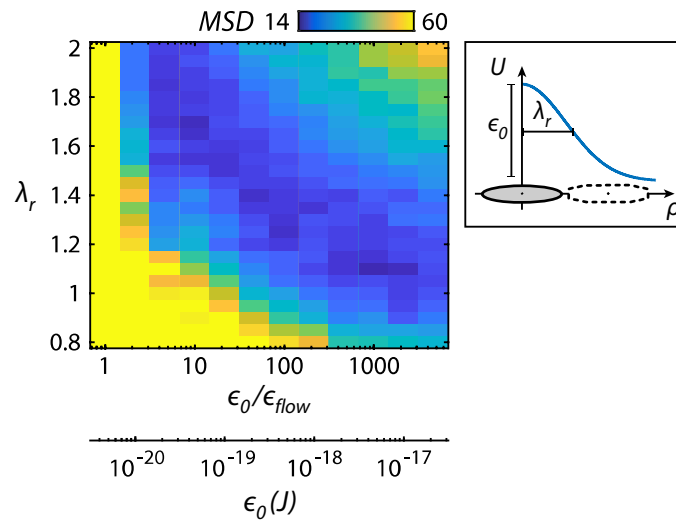
Supplementary Figure 16 – Parameter screen to test the influence of cell-cell repulsion on biofilm architecture phenotype. a, Mean *MSD* values between feature vectors of Δ*rbmA* mutant biofilm simulations and experiments (see section on "Comparing simulations with experimental data") upon variation of cell-cell interaction strength ϵ_0 and repulsion range λ_r. Inset: effect of parameter variation on the cell-cell interaction potential.
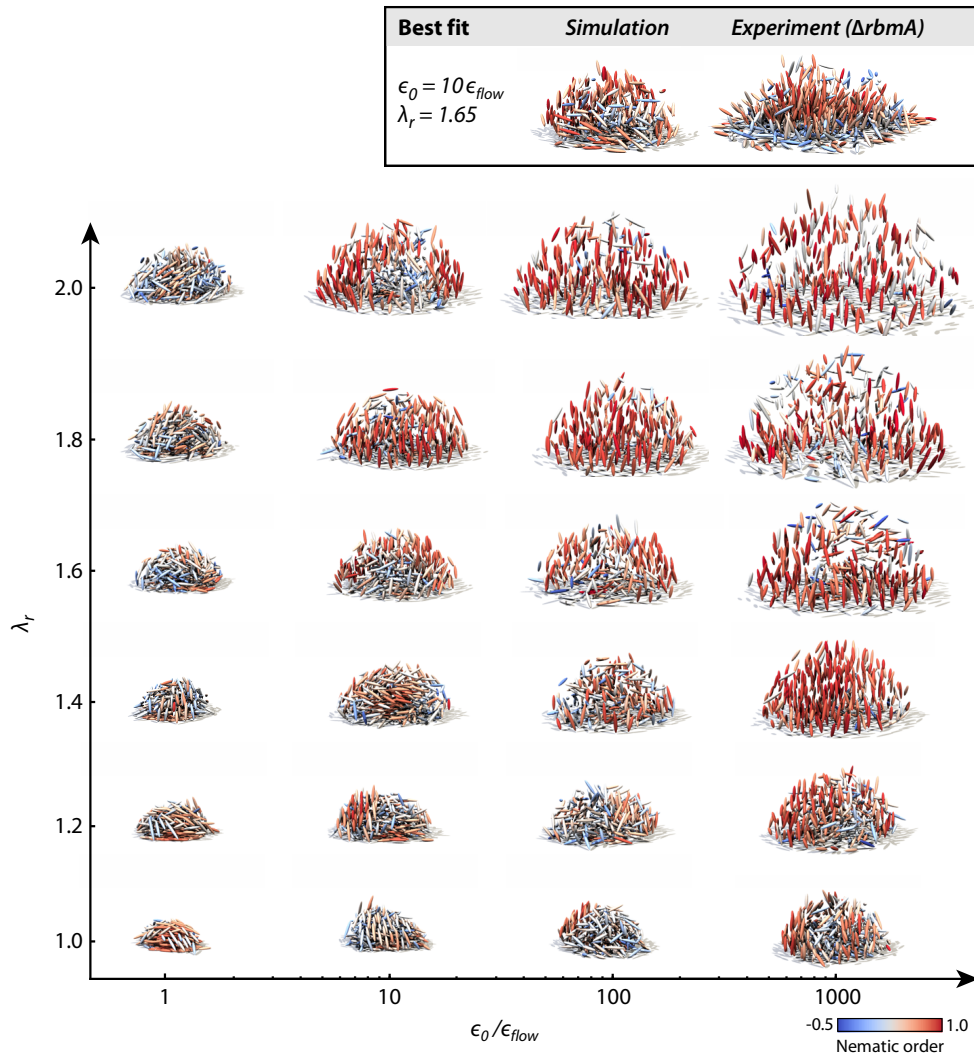
Supplementary Figure 17 – Resulting simulated biofilms for different levels of cell-cell repulsion. The figure shows a subset of biofilm renderings corresponding to a range of different values for ϵ_0 and λ_r (*cf.* Fig. 2b and Supplementary Fig. 16). The colour of each cell corresponds to the nematic order.
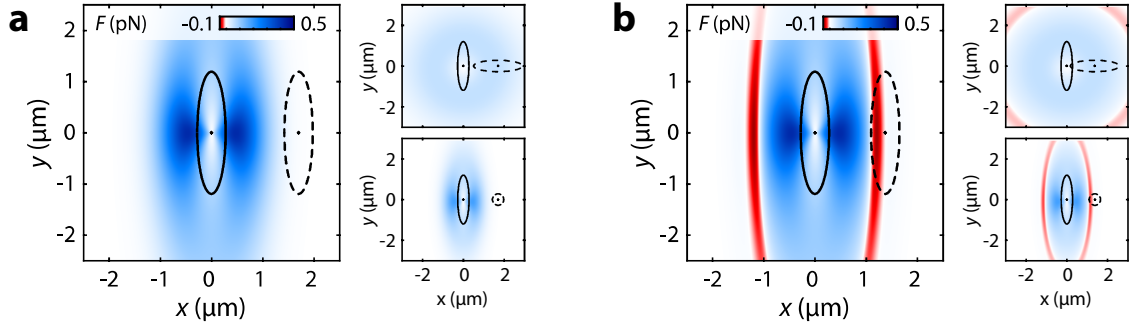
Supplementary Figure 18 – Translational cell-cell interaction forces. a, Translational forces for $\epsilon_0 = 10 \cdot \epsilon_{flow}$ ($5 \cdot 10^{-20}$ J), $\lambda_r = 1.65$, and $\nu = 0$ (corresponding to the $\Delta rbmA$-mutant, *cf.* Fig. 2b) for different orientations. **b,** Translational forces for $\epsilon_0 = 10 \cdot \epsilon_{flow}$ ($5 \cdot 10^{-20}$ J), $\lambda_r = 1.65$, $\nu = 0.13$, $\lambda_a = 0.164$, and $\rho_a = 2.93$ (corresponding to an arabinose concentration of 0.5%, *cf.* Fig. 2c) for different orientations.
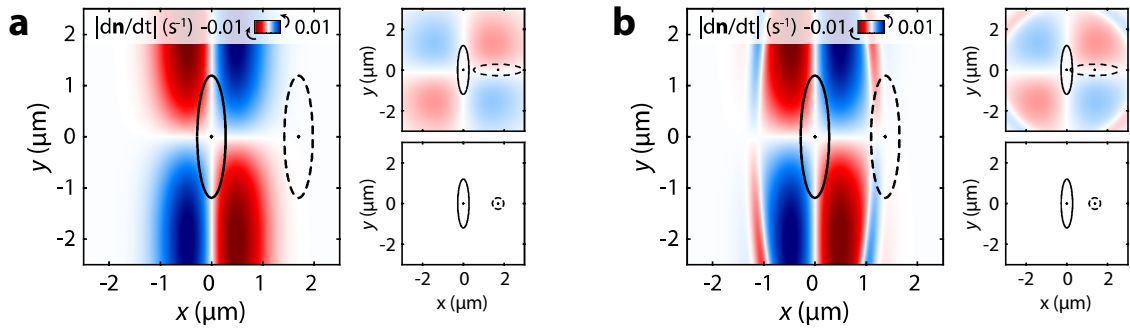
Supplementary Figure 19 – Rotational cell-cell interaction dynamics. $|\mathrm{d}\mathbf{n}/\mathrm{d}t| = |(\mathbf{I} - \hat{\mathbf{n}}_\beta \hat{\mathbf{n}}_\beta)(-\mathbf{\Omega}^{-1} \partial U / \partial \hat{\mathbf{n}}_\beta)|$ acting on two neighbouring cells. **a,** Rotational dynamics for $\epsilon_0 = 10 \cdot \epsilon_{flow}$ ($5 \cdot 10^{-20}$ J), $\lambda_r = 1.65$, and $\nu = 0$ (corresponding to the $\Delta rbmA$-mutant, *cf.* Fig. 2b) for different orientations. **b,** Rotational dynamics for $\epsilon_0 = 10 \cdot \epsilon_{flow}$ ($5 \cdot 10^{-20}$ J), $\lambda_r = 1.65$, $\nu = 0.13$, $\lambda_a = 0.164$, and $\rho_a = 2.93$ (corresponding to an arabinose concentration of 0.5%, *cf.* Fig. 2c) for different orientations. A positive speed (indicated in red) results in a clockwise rotation.
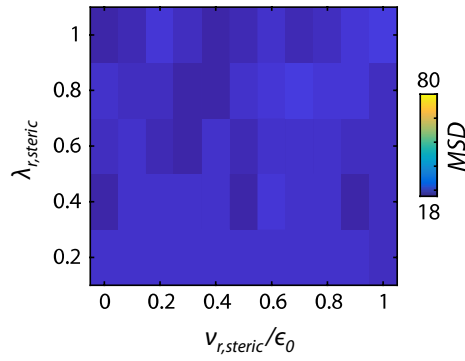
Influence of steric and osmotic pressure-mediated cell-cell repulsion on biofilm phenotype

The joint effects of steric cell-cell repulsion and osmotic pressure-mediated cell-cell repulsion were modelled with a relatively soft Gaussian function, *cf.* Eq. (18). To test if this model is appropriate, we performed the following analysis: To check the effect of both contributions to cell-cell repulsion separately, we embedded a second, short-ranged and very strongly repulsive part, representing steric repulsion, into the existing potential by adding another Gaussian function peaked at $\rho = 0$ characterized by ν_{steric} and $\lambda_{r,steric} < \lambda_r$, see Eq. (24). Here, ϵ_0 and λ_r are fixed to the optimal values obtained by comparison with the experimental values for the $\Delta rbmA$-mutant.

$$ U = \epsilon_0 \epsilon_1 \left(e^{-\frac{\rho^2}{\lambda_r^2}} + \nu_{steric} \cdot e^{-\frac{\rho^2}{\lambda_{r,steric}^2}} + \frac{\nu}{1 + e^{\left(\frac{\rho_a - \rho}{\lambda_a}\right)}} \right) \tag{24} $$

The exploration of the parameter space is shown in Supplementary Fig. 20. The additional term for hard-steric repulsion in Eq. (24) does not have any influence on the biofilm phenotype as the *MSD* values show no variance irrespective of interaction strength ν_{steric} and range $\lambda_{r,steric}$. This justifies modelling the effects of hard steric and soft osmotic pressure-mediated repulsion in a combined manner, as cells are unlikely to interact exclusively via hard and very short-ranged steric repulsion (the matrix prevents them from getting close enough together).
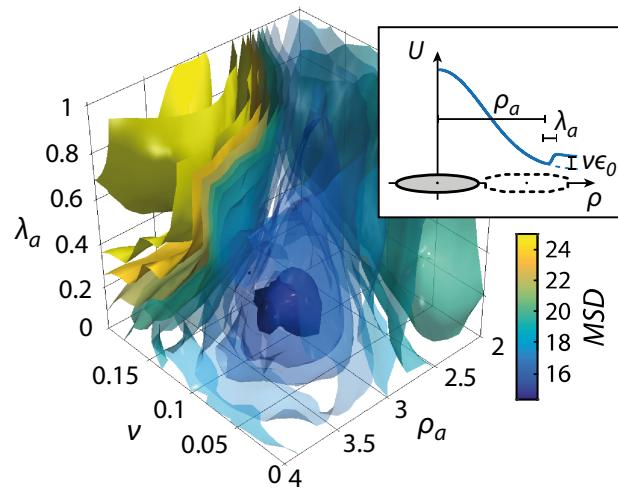
Supplementary Figure 20 – Parameter screen to test the influence of hard steric cell-cell repulsion on biofilm phenotype. The figure shows *MSD* values upon variation of v_{steric} and $\lambda_{r,steric}$. Owing to its longer length scale, soft, osmotic pressure-mediated repulsion prevents cells from getting close enough to interact directly through hard steric repulsion.
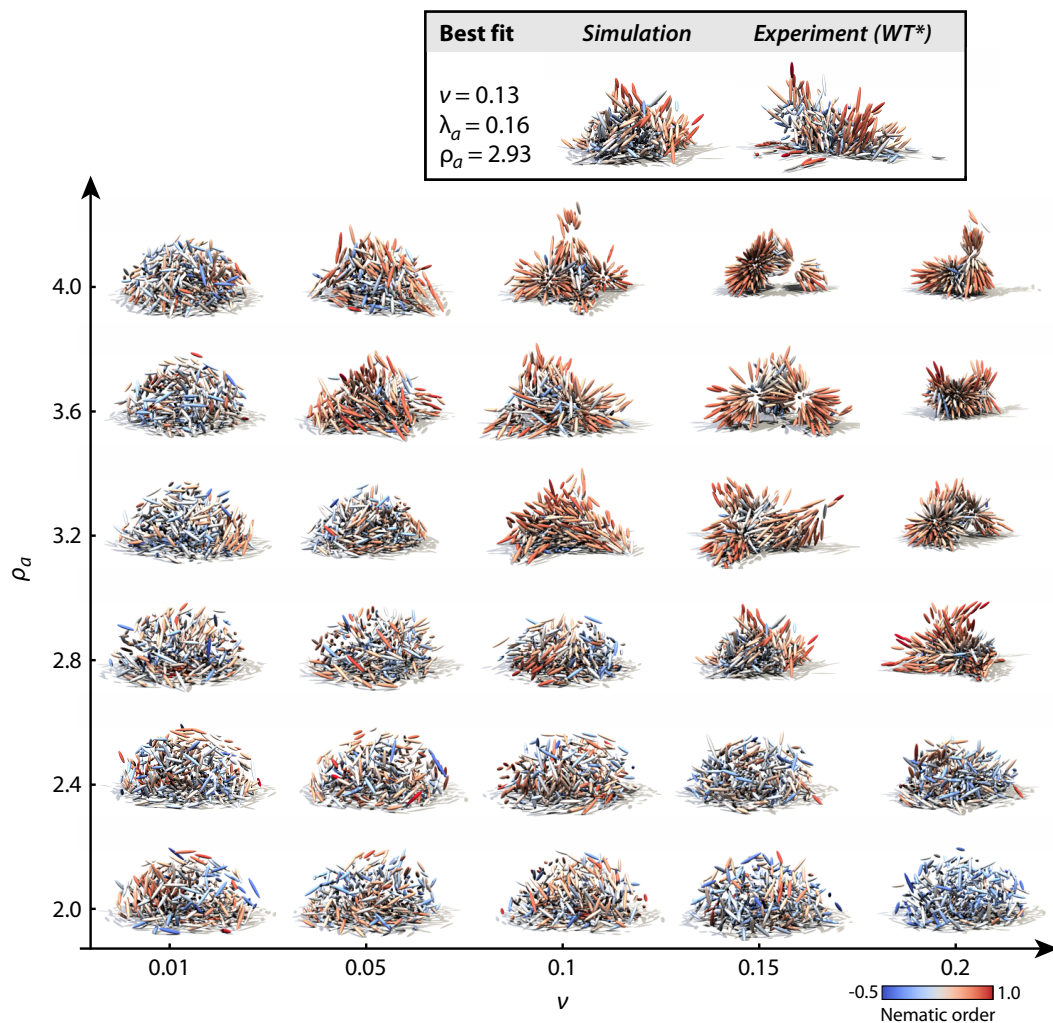
Influence of RbmA-mediated cell-cell attraction on biofilm phenotype

The attractive part in the cell-cell potential is thought to be governed by RbmA. Having fixed the cell-cell repulsion parameters ϵ_0 and λ_r for the $\Delta rbmA$ phenotype, the remaining parameters influencing cell-cell attraction (namely v, ρ_a, and λ_a) were determined by finding the best fits when comparing the simulations with biofilms of different RbmA concentrations (by varying the expression of the *rbmA* gene using the arabinose-inducible P_{BAD} promoter). Prior to systematic parameter scans with simulations, we estimated the parameters using the following considerations: The attraction shift ρ_a was estimated for each arabinose concentration by considering the typical cell-cell distance at the edge of biofilms, where attraction dominates. These estimates suggested using values of ρ_a between 2 and 4 (again using $\sigma = 0.7$ µm). The attraction width λ_a was estimated by considering the standard deviation of cell-cell distances near the edge of biofilms, which was found to be approximately 0.3 µm for all biofilms; therefore, we expected λ_a to be around 0.4 (again using $\sigma = 0.7$ µm).

The resulting *MSD* values resulting from simulations conducted with a systematic variation of v, λ_a and ρ_a are shown in Supplementary Fig. 21. The resulting translational and rotational forces for selected conditions are visualized in Fig. 2c, Supplementary Fig. 18b, and Supplementary Fig. 19b, respectively. Corresponding biofilm phenotypes are shown in Supplementary Fig. 22. The 5% best matching simulations (5% lowest *MSD* values) for a given condition were chosen and the mean of the corresponding values for v, λ_a and ρ_a was defined as the best fit. In Supplementary Fig. 23 the obtained mean values ± standard deviation are displayed. The best-fit parameters for the experimentally observed WT*-phenotype were $v = 0.13$, $\lambda_a = 0.16$ and $\rho_a = 2.93$.

Supplementary Figure 21 – Parameter screen 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 (w/v) and simulations (see section "Comparing simulations with experimental data) 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.
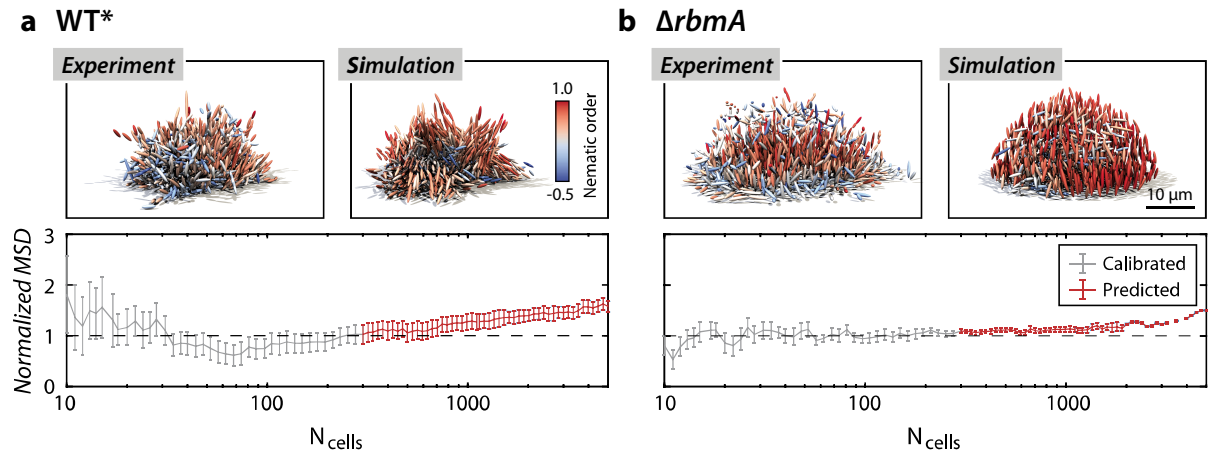

Supplementary Figure 22 – Resulting simulated biofilms for different levels of RbmA-mediated cell-cell attraction. The figure shows biofilm renderings corresponding to selected values of ν and ρ_a while keeping λ_a constant at $\lambda_a = 0.16$ (*cf.* Fig. 2e). Each cell is coloured according to the nematic order.
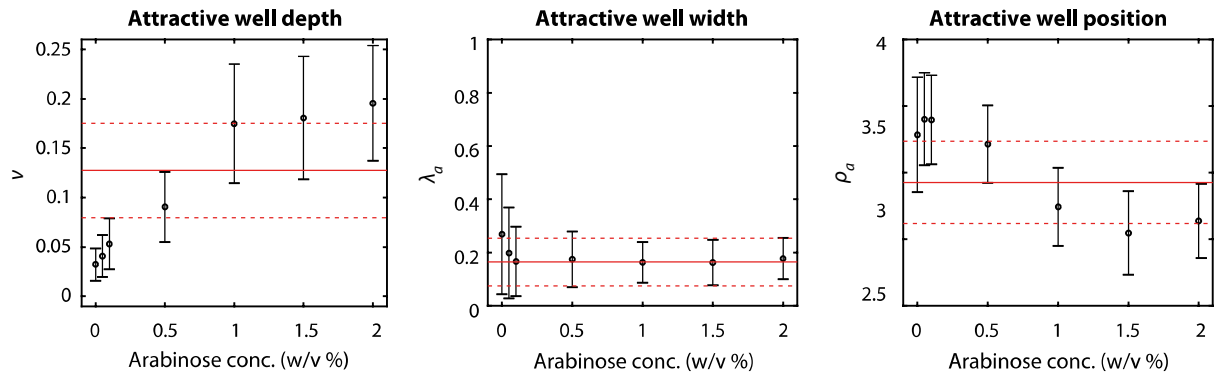
Supplementary Figure 23 – Fitting of the attractive potential parameters. The graphs show the approximate position of the *MSD* minimum in (ν, λ_a and ρ_a) -space of the 5% best-matching simulations at increasing arabinose concentrations, corresponding to increasing amounts of secreted RbmA. Each data 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).

3.4. Predicting the structure of larger biofilms

The cell-cell interaction potential was calibrated based on biofilms with up to 300 cells (section "Comparing simulations with experimental data"). 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 section 3.2 was used. The *MSD* between the experimental and simulation feature vectors was 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 section 3.2 where a single *MSD* between temporal feature vectors was calculated), and was normalized by the average value of the *MSD* between 10 and 300 cells. For cell numbers between 300 and 1000, the *MSD* at each cell number was found to be close to the average value of the *MSD* between 10 and 300 cells for WT* and $\Delta rbmA$ biofilms, and strong qualitative agreement was observed between simulated and experimental biofilms with up to 1000 cells (Supplementary Fig. 24, Supplementary Movie 5). 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.

Supplementary Figure 24 – 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 ($n = 3$). These data are also shown in Fig. 2g,h. **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 were calculated by Gaussian error propagation of the standard deviations of the individual measurements.

4. Continuum Model

## 4.1. Mathematical model of a growing biofilm in shear flow

For this model, we assume that the biofilm colony is located in the centre of a rectangular channel. Fluid flows through the channel in the x-direction with a prescribed flow rate. Flow is modelled by the Stokes equations

$$\nabla \cdot \boldsymbol{u} = 0, \qquad \mu \nabla^2 \boldsymbol{u} = \nabla p \tag{25}$$

with a no-slip condition ($\boldsymbol{u} = \boldsymbol{0}$) applied on the channel walls and on the surface of the biofilm. The biofilm surface $z = S(x, y, t)$ grows or shrinks in the direction of its outer unit normal $\boldsymbol{n}$, with velocity

$$\begin{aligned} \mathbf{U} &= \alpha(\boldsymbol{x}, t)\boldsymbol{n} - \beta |(\mathbb{I} - \boldsymbol{nn})\sigma \cdot \boldsymbol{n}|\boldsymbol{n} \\ &= \alpha(\boldsymbol{x}, t)\boldsymbol{n} - \beta\tau\boldsymbol{n} \end{aligned} \tag{26}$$

where the first term on the right hand side represents movement owing to cell growth and division, and the second term represents erosion proportional to the shear stress τ imposed on the surface by the fluid; here σ is the stress tensor of the fluid. For a hemispherical biofilm growing uniformly in the radial direction at a rate r_0, with a base centred at the origin, $\alpha(\boldsymbol{x}, t) = r_0 r$, where $r = |\boldsymbol{x}|$. This leads to exponential volumetric growth in time, $V(t) = V_0 \exp(3r_0 t)$, where V_0 is the volume of the biofilm at $t = 0$. The growth rate r_0 is found by fitting an exponential curve to experimental data of volume versus time for a $\Delta rbmA$-mutant in a channel with an average flow velocity of $<v>$ = 0.03 mm/min (Supplementary Fig. 25).
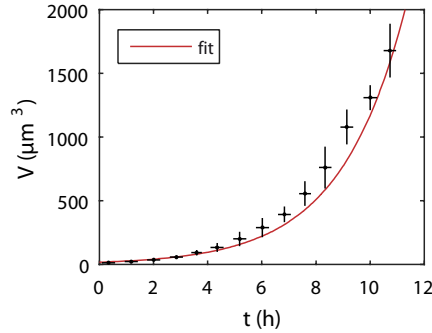
Supplementary Figure 25 – Volume of $\Delta rbmA$-mutant (black) growing at a very low flow speed ($<v>$ = 0.03 mm/min). Exponential fit (red), error bars correspond to the standard deviation (n = 6 biofilms).

4.2. Numerical method

The surface of the biofilm is represented by the zero level set of a level set function ϕ, which satisfies the Hamilton-Jacobi equation[23]

$$\phi_t + a(\boldsymbol{x}, t)|\nabla\phi| = 0. \tag{27}$$

Here $a = |\boldsymbol{U}|$ is the normal speed of the surface.

Equation (27) is solved using the Level Set Toolbox[24] in Matlab, which approximates the time derivative with an explicit total variation diminishing Runge-Kutta integration scheme. The spatial grid is uniform in each direction, with a smaller step size in the z-direction. A symmetry condition ($\nabla\phi \cdot \boldsymbol{n} = 0$) is applied along the xz-plane through the centre of the biofilm. At all other boundaries, ghost nodes are added and filled with data linearly extrapolated from the computational boundary[24].

At each timestep, the steady Stokes equations (25) are solved using the finite-element package deal.II[25], with the surface height $z = S(x, y, t)$ used as input to denote the grid points at which the no-slip condition should be applied. The finite-element mesh is locally refined in all directions around the biofilm surface. A uniform velocity value is imposed at the inflow boundary, and the flow is allowed to develop in an entrance region. A symmetry condition ($\boldsymbol{u} \cdot \boldsymbol{n} = 0$) is applied on the xz-plane through the centre of the biofilm. The stress tensor σ at the biofilm surface is output for use in equation (26) at the next timestep. The value of σ is interpolated to the level set spatial grid. The total drag force in the direction of the flow on a hemisphere is checked to agree with the value calculated by Pozrikidis *et al.*[26].

The shear stress applied to real biofilms is calculated by incorporating experimental data from wild-type biofilms into the continuum model. The convex hull of the set of cell centroids at a specific time point is extracted, after some outlying cells and cells not descended from the biofilm population have been removed. The surface is smoothed by binarizing the volumes inside and outside the convex hull, then applying a Gaussian filter. The flow and the biofilm surface are assumed to be symmetric about the midplane of the biofilm, as before (i.e. flow is only simulated on one half of the experimental domain, chosen arbitrarily).

5. Supplementary Data

Strain	Genotype/Relevant features	Reference
	E. coli	
S17-1 λpir	ΔlacU169 (ΦlacZΔM15), recA1, endA1, hsdR17, thi-1, gyrA96, relA1, λpir	De Lorenzo et al.[27]
TOP10	mcrAΔ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15ΔlacX74deoRrecA1 araD139Δ(ara-leu)7697 galU galK rpsL endA1 nupG	Invitrogen
	V. cholerae	
KDV101	wild type strain N16961 (O1 El Tor, Sm^R)	Meibom et al.[28]
KDV103	N16961 lacZ:P$_{tac}$-mKOKappa	Drescher[1]
KDV115	N16961 vpvC^{W240R} lacZ:P$_{tac}$-mKOKappa	Drescher[1]
KDV148	N16961 vpvC^{W240R} (matrix hyperproducer phenotype)	Drescher[1]
KDV383	N16961 ΔrbmA, lacZ:P$_{tac}$-mKOKappa	Drescher lab
KDV611	N16961 vpvC^{W240R}, ΔcrvA	This work
KDV613	N16961 vpvC^{W240R}, ΔcrvA contains plasmid pNUT542	This work
KDV692	N16961 vpvC^{W240R}, ΔrbmA contains plasmid pNUT542	This work
KDV698	N16961 vpvC^{W240R}, ΔcrvA, ΔrbmA	This work
KDV815	N16961 vpvC^{W240R}, ΔrbmA, ΔcrvA contains plasmid pNUT542	This work
KDV829	N16961 vpvC^{W240R}, rbmA::3xFLAG	This work
KDV835	N16961 vpvC^{W240R}, rbmA::3xFLAG contains plasmid pNUT542	This work
KDV1026	N16961 vpvC^{W240R}, ΔcrvA, rbmA:mRuby3	This work
KDV1027	N16961 vpvC^{W240R}, ΔcrvA, rbmA:mRuby3 and plasmid pNUT542	This work
KDV1082	N16961 vpvC^{W240R}, ΔrbmA, ΔcrvA contains plasmid pNUT1519	This work
KDV1218	N16961 vpvC^{W240R}, ΔcrvA contains plasmid pNUT844	This work
KDV1228	N16961 vpvC^{W240R}, ΔcrvA contains plasmid pNUT1734	This Work
KDV1231	N16961 vpvC^{W240R} lacZ:P$_{tac}$-mKOKappa and plasmid pNUT1734	This Work

Supplementary Table 5 – Bacterial strains used in this study.

Plasmids	Origin, Marker	Comments	References
pNUT015	pR6K, Amp^R	pKAS32	Skorupski et al.[29]
pNUT144	pR6K, Amp^R, Kan^R	pKAS32 with Kan^R	Drescher et al.[30]
pNUT336	pR6K, Amp^R	pKAS32 ΔrbmA (vc0928)	Nadell et al.[31]
pNUT462	pR6K, Amp^R	pNUT144 rbmA::3xFLAG (translational fusion)	Drescher lab stock
pNUT542	pSC101*, Gent^R	P$_{tac}$-sfgfp	Singh et al.[32]
pNUT844	p15a, Gent^R	P$_{tac}$-sfgfp, P$_{vpsL}$-mRuby2	Drescher lab stock
pNUT961	pR6K, Amp^R, Kan^R	pNUT144 ΔcrvA (vca1075)	This work
pNUT1268	pSC101*, Gent^R	pNUT542 with P$_{BAD}$-rbmA to replace P$_{tac}$-sfgfp	This work
pNUT1401	pR6K, Amp^R, Kan^R	pNUT144 rbmA:mRuby3 (transcriptional fusion)	This work
pNUT1519	pSC101*, Gent^R	pNUT1268 with P$_{tac}$-sfgfp	This work
pNUT1734	pSC101*, Gent^R	pNUT1268 with sfgfp to replace rbmA	This work

Supplementary Table 6 – Plasmids used in this study.

Name	Sequence	Function
KDO1182	GGGTCTAGAGCTCGATATAGCGATCCACCACTTC	Plasmid pNUT961 construction
KDO1183	GAGTTTTACCCACACCGTCCCTG-CAACATATTTATGTTTAGCC	Plasmid pNUT961 construction
KDO1184	GGGACGGTGTGGGTAAAACTCAGACCAAACAAAG	Plasmid pNUT961 construction
KDO1185	TTTGCGGCCGCCCGCATGCCGTCATAGAAACCAG	Plasmid pNUT961 construction
KDO1433	TTTAAAGTTAGACAACGCAATATATCCTAG-TTATAAAAAATTTAACGCCA	Plasmid pNUT1268 construction
KDO1434	CTAGGATATATTGCGTT-GTCTAACTTTAAAGGATCTATCATGAACAAA	Plasmid pNUT1268 construction
KDO1435	TTATTTTTTTACCACTGTCATTGACTGTTC	Plasmid pNUT1268 construction
KDO1436	ATGCATAAATAACAAAAACTGCTAAACGTT	Plasmid pNUT1268 construction
KDO1667	TTGCGGCCGCGTTACAAGAACCCGGAAGAATGTG	Plasmid pNUT1401 construction
KDO1716	CCTAGTGGCTTATTTTTTTACCACTGTCATTGACTGTTCC	Plasmid pNUT1401 construction
KDO1717	CAGTGGTAAAAAAATAAGCCACTAGGAGGTGGTTGATGGTG	Plasmid pNUT1401 construction
KDO1718	GAACTGTATAAATAAATTTACCTAGTCACTTAGTCG-TATGTATAAAAAACCGC	Plasmid pNUT1401 construction
KDO1719	CTAAGTGACTAGGTAAATTTATTTATACAG-TTCATCCATACCACCACCCAG	Plasmid pNUT1401 construction
KDO1720	GGTTGCTAGCGTCTACCTTGGCAGCACTAAAGGTACC	Plasmid pNUT1401 construction
KDO389	GCTTGCTCAATCAATCACCGGATCC	Plasmid pNUT1519 construction
KDO578	TTCACACCTCCTGTACGCGCGGC	Plasmid pNUT1519 construction
KDO2162	ATAACTAGGATATATTGCGATGAGCAAAGGTGAA-GAACTGTTTAC	Plasmid pNUT1734 construction
KDO2163	GCTCATCGCAATATATCCTAGTTATAAAAAATTTAAC-GCCAG	Plasmid pNUT1734 construction
KDO2164	GAGCTCTACAAGTAAATGCATAAA-TAACAAAAACTGCTAAACGTTATTC	Plasmid pNUT1734 construction
KDO2165	GTTATTTATGCATTTACTT-GTAGAGCTCGTCCATGCCGTGAG	Plasmid pNUT1734 construction

Supplementary Table 7 – DNA oligonucleotides used in this study. Sequences are given in the 5'→3' direction.

6. References

1. Drescher, K. *et al.* Architectural transitions in *Vibrio cholerae* biofilms at single-cell resolution. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E2066–E2072 (2016).
2. Horn, B. K. P. & Schunck, B. G. Determining optical flow. *Artif. Intell.* **17,** 185–203 (1981).
3. M. A. Mustafa. Horn-Schunck optical flow method for 3-D images. (2014). Available at: https://de.mathworks.com/matlabcentral/fileexchange/37053-horn-schunck-optical-flow-method-for-3-d-images. (Accessed: 2nd November 2017)
4. Flandin, G. mVTK: a MATLAB VTK library. (2017). Available at: http://www.artefact.tk/software/matlab/mvtk/. (Accessed: 2nd November 2017)
5. Sonnet, A. M. & Virga, E. G. Dynamics of dissipative ordered fluids. *Phys. Rev. E* **64,** 031705 (2001).
6. Liu, X., Tanaka, M. & Okutomi, M. Single-image noise level estimation for blind denoising. *IEEE Trans. Image Process.* **22,** 5226–5237 (2013).
7. Yan, J., Sharo, A. G., Stone, H. A., Wingreen, N. S. & Bassler, B. L. *Vibrio cholerae* biofilm growth program and architecture revealed by single-cell live imaging. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E5337–E5343 (2016).
8. Beyhan, S., Tischler, A. D., Camilli, A. & Yildiz, F. H. Transcriptome and phenotypic responses of *Vibrio cholerae* to increased cyclic di-GMP level. *J. Bacteriol.* **188,** 3600–3613 (2006).
9. Watnick, P. I., Lauriano, C. M., Klose, K. E., Croal, L. & Kolter, R. The absence of a flagellum leads to altered colony morphology, biofilm development and virulence in *Vibrio cholerae* O139. *Mol. Microbiol.* **39,** 223–235 (2001).
10. Yildiz, F. H., Liu, X. S., Heydorn, A. & Schoolnik, G. K. Molecular analysis of rugosity in a *Vibrio cholerae* O1 El Tor phase variant. *Mol. Microbiol.* **53,** 497–515 (2004).
11. Utada, A. S. *et al. Vibrio cholerae* use pili and flagella synergistically to effect motility switching and conditional surface attachment. *Nat. Commun.* **5,** 4913 (2014).
12. Conrad, J. C. Physics of bacterial near-surface motility using flagella and type IV pili: Implications for biofilm formation. *Res. Microbiol.* **163,** 619–629 (2012).
13. Beyhan, S. & Yildiz, F. H. Smooth to rugose phase variation in *Vibrio cholerae* can be mediated by a single nucleotide change that targets c-di-GMP signalling pathway. *Mol. Microbiol.* **63,** 995–1007 (2007).
14. Fritz, G. *et al.* Single cell kinetics of phenotypic switching in the arabinose utilization system of *E. Coli. PLoS One* **9,** e89532 (2014).
15. Megerle, J. A., Fritz, G., Gerland, U., Jung, K. & Rädler, J. O. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.* **95,** 2103–2115 (2008).
16. Han, Y., Alsayed, A., Nobili, M. & Yodh, A. G. Quasi-two-dimensional diffusion of single ellipsoids: Aspect ratio and confinement effects. *Phys. Rev. E* **80,** 011403 (2009).
17. Taheri-Araghi, S. *et al.* Cell-size control and homeostasis in bacteria. *Curr. Biol.* **25,** 385–391 (2015).
18. Nyland, L., Harris, M. & Prins, J. Fast n-body simulation with CUDA. *Simulation* **3,** 677–696 (2007).
19. Pavlovsky, L., Younger, J. G. & Solomon, M. J. In situ rheology of *Staphylococcus epidermidis* bacterial biofilms. *Soft Matter* **9,** 122–131 (2013).
20. Fong, J. C. *et al.* Structural dynamics of RbmA governs plasticity of *Vibrio cholerae* biofilms. *Elife* **6,** e26163 (2017).
21. Persat, A. *et al.* The mechanical world of bacteria. *Cell* **161,** 988–997 (2015).
22. Cleaver, D. J., Care, C. M., Allen, M. P. & Neal, M. P. Extension and generalization of the Gay-Berne potential. *Phys. Rev. E* **54,** 559–567 (1996).

23. Osher, S., Fedkiw, R. & Piechor, K. Level set methods and dynamic implicit surfaces. *Appl. Mech. Rev.* **57,** B15 (2004).

24. Mitchell, I. M. The flexible, extensible and efficient toolbox of level set methods. *J. Sci. Comput.* **35,** 300–329 (2008).

25. Arndt, D. *et al.* The deal.II library, version 8.5. *J. Numer. Math.* **25,** 137–145 (2017).

26. Pozrikidis, C. Shear flow over a protuberance on a plane wall. *J. Eng. Math.* **31,** 29–42 (1997).

27. De Lorenzo, V. & Timmis, K. N. Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. *Methods Enzymol.* **235,** 386–405 (1994).

28. Meibom, K. L. *et al.* The *Vibrio cholerae* chitin utilization program. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 2524–9 (2004).

29. Skorupski, K. & Taylor, R. K. Positive selection vectors for allelic exchange. *Gene* **169,** 47–52 (1996).

30. Drescher, K., Nadell, C. D., Stone, H. A., Wingreen, N. S. & Bassler, B. L. Solutions to the public goods dilemma in bacterial biofilms. *Curr. Biol.* **24,** 50–55 (2014).

31. Nadell, C. D., Drescher, K., Wingreen, N. S. & Bassler, B. L. Extracellular matrix structure governs invasion resistance in bacterial biofilms. *ISME J.* **9,** 1700–1709 (2015).

32. Singh, P. K. *et al. Vibrio cholerae* combines individual and collective sensing to trigger biofilm dispersal. *Curr. Biol.* **27,** 3359–3366.e7 (2017).

nature research

Corresponding author(s):   Knut Drescher

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

▶ Experimental design

1. Sample size

Describe how sample size was determined.

> Each of the n replicates of each experiment is the average of hundreds to thousands of bacterial cells as a sample size. Each experiment was replicated independently at least 3 times, but usually n>3. The n for each experiment is indicated in each figure.

2. Data exclusions

Describe any data exclusions.

> No data were excluded.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

> Each experiment was replicated n times (and n is given in each figure for each experiment). Although the exact quantitative results differ between replicates, the qualitative results were the same, so that it is reasonable to state that the "replication was successful".

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

> There were many bacterial cells within each of the n replicates. Because of the large sample size for each replicate, a representative number of samples were collected for each replicate. There was no allocation of samples into experimental groups, beyond conducting independent biological replicates.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

> Blinding of group allocation is irrelevant to our data analysis, because there was no allocation to experimental groups, beyond collecting n replicates, all of which were analyzed by software equally.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

> Matlab (version R2016b) was used to analyze data. A detailed description of the analysis is provided in the Materials and Methods section.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

> All materials that were used in this study are available from standard commercial suppliers or from the corresponding author.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

> All antibodies that were used are listed in the Materials and Methods section, including part numbers and supplier information. Validation: Biofilms grown from strains that do not contain the FLAG modification to RbmA did not show any antibody signal.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

> No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

> No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

> No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

> No commonly misidentified cell lines were used.

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

> No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

> This study did not involve human research participants.