

Pannaree Boonyuen

The Role of Noise and Nutrient Level
in the Regulation of Matrix-Motile Phase Switching
in *Bacillus Subtilis* Biofilm

Supervisor: Diana Fusco

Word count: 5990

I am, Pannaree Boonyuen, of Girton College, undertaking a project titled “The Role of Nutrient Level and Noise in the Regulation of Matrix-Motile Phase Switching in *Bacillus Subtilis* Biofilm”.

I understand the University s definition of plagiarism. I declare that, in accordance with Discipline regulation 6, this dissertation is entirely my own work except where otherwise stated, either in the form of citation of published work, or acknowledgement of the source of any unpublished material.

Sign: PANNAREE BOONYUEN

Date: 01/05/2023

The biofilm data collection was performed by Nikhil Krishan, a PhD in Diana’s lab. The experimental data, codes, and supplementary videos can be found in the Github <https://github.com/pannareeb/SwitchingBsubtilis>

Part III project dissertation: P10

Summary

Biofilm is important for the survival of bacteria in challenging conditions. In a model organism *Bacillus subtilis*, spatiotemporal phenotypic heterogeneity has been observed consistently in its biofilms. Our focus was the re-emergence of *B. subtilis* motile cells at the edge of late grown biofilm, hypothesised to be induced by noises. We created a multi-scaling model framework in which an ODE system of SinI-SinR-SlrR regulatory network was connected to a PDE system representing population growth dynamic. This can be used for evaluating the combined effect of noises and nutrients on cell phases in biofilm in a continuous time and space. It can also be generalised for other behaviours and systems. We found that noises around protein levels was unlikely to explain this late re-emergence of motile cells since noises universally increased the tendency of cells being in a matrix phase, even in a deterministically monostable system. This happened to a greater extent when the complexing rates between the matrix genes repressor, SinR, and its antagonists were high, which can be regulated by nutrient level in such a way that noises increased the width of matrix cell ring. However, the final multi-scaling model could not reproduce the expected pattern of phase switching, despite the optimisation using the experimental data tracking motile and matrix gene expression in a real biofilm. This could be overcome by a larger-scale scanning of several low-confidence parameters, an improved optimisation algorithm, and an adjustment to match timescale of the two sub-models.

Table of Contents

List of Abbreviations and Equations.....	6
Introduction	7
Results	11
Discussion.....	23
Materials and Methods	27
Acknowledgements.....	31
References.....	32
Appendix I: Supplementary Media	34
Appendix II: The derivation of analytical solution of the steady state level of SinR. ...	39
Appendix III: Parameterisation of the effective diffusion coefficient of bacteria and fluorescent coefficients.....	40

List of Abbreviations and Equations

Abbreviations	Definition
(IRL-)ODE	Ordinary Differential Equation (of SinI-SinR-Slrr regulatory network)
(NuBac-)PDE	Partial Differential Equation (of bacterial growth dynamics using agar nutrient)
RFP/ YFP	Red/Yellow Fluorescence Protein
Xss	Steady state level of X
ic	Initial condition

Equations	Uses
(1.1) $\dot{I} = \alpha_0 - I - \delta_I IR$	SinI-SinR-Slrr regulatory network (IRL-ODE)
(1.2) $\dot{R} = \beta_0 - R - \delta_I IR - \delta_L LR$	
(1.3) $\dot{L} = \frac{\gamma}{1+R^2} - L - \delta_L LR$	
(2.1) $\frac{\partial b_{t,x}}{\partial t} = D_b \frac{\partial^2 b_{t,x}}{\partial x^2} + k_g \frac{n_{t,x} b_{t,x}}{K+n_{t,x}} - k_d \frac{n_{t,x} b_{t,x}}{K+n_{t,x}}$	Bacterial growth (b) dynamics using agar nutrient (n) (NuBac-PDE)
(2.2) $\frac{\partial n_{t,x}}{\partial t} = D_n \frac{\partial^2 n_{t,x}}{\partial x^2} - \omega k_g \frac{n_{t,x} b_{t,x}}{K+n_{t,x}} + \theta \omega k_d \frac{n_{t,x} b_{t,x}}{K+n_{t,x}}$	
(3) $\delta_{L,x} = \begin{cases} n_m - v(0.5 - n_x), n_x < 0.5 \\ n_m - v(n_x - 0.5), n_x \geq 0.5 \end{cases}$	Hypothetical non-monotonic link function (n _m = 2.5 and v = 5)
(4.1) $v_b = \sqrt{2 \cdot \frac{n}{K+n} (k_g - k_d) D_b} = \sqrt{2 \cdot 0.271 (k_g - k_d) D_b}$	Speed of experimental biofilm and its relation to the effective diffusion coefficient of bacteria population
(4.2) $D_b = \frac{v_b^2}{2 \cdot 0.271 (k_g - k_d)}$	
(5.1) $\delta_L(n_{t,x}) = cf_1 + cf_2 n_{t,x} + cf_3 n_{t,x}^2 + cf_4 n_{t,x}^3 + cf_5 n_{t,x}^4$	Link function between nutrient and two complexing rates (dILf(n))
(5.2) $\delta_I(n_{t,x}) = cf_6 \cdot \delta_L(n_{t,x})$	
(6) $lossfront = \sum_{t=0}^{t_{141}} (bfront_t - modelfront_t)^2$	Loss function to be minimised in Db parameterisation
(7) $lossall = \sum_{x=0}^{x_{210}} \sum_{t=0}^{t_{141}} (b_{t,x} - (n_y * YFP_{t,x} + n_r * RFP_{t,x}))^2$	Loss function to be minimised in nr-ny parameterisation
(8) $g = \sum_{x=0}^{x_{210}} \sum_{t=0}^{t_{141}} (bMotile_{t,x} - n_y * YFP_{t,x})^2 + (bMatrix_{t,x} - n_r * RFP_{t,x})^2$	Loss function to be minimised in dILf(n) parameterisation

Introduction

Many bacteria species live in biofilm, a group of microbial cells embedded in a self-produced extracellular matrix, consisting of scaffold proteins, polysaccharides, and extracellular DNA (Donlan, 2002). Biofilm allows bacteria to better maintain a stable environment against physical stresses e.g. desiccation, access to nutrients, and change the chemical structure of antibiotics (Sharma et al., 2019). Biofilm colony also exhibits a division of labour, with cells differing in their phenotypes and performing distinct roles (Dragoš et al., 2018). These emergent properties, not observed in planktonic cells found in suspension, confer increased survival chance in harsh or changing conditions (Fazeli-Nasab et al., 2022). As such, studying the regulation of biofilm formation and its properties has implications not only for industrial and medical settings, but also for unveiling mechanisms controlling cellular coordination.

One of the biofilm-forming bacteria is undomesticated *Bacillus subtilis* NCIB 3610, a good model organism due to its simple and diverse genetic modification protocols and the availability of its complete genome (Nye et al., 2017; Su et al., 2020). More importantly, its dynamic phenotypic heterogeneity in a biofilm has been observed both in the natural and laboratory environment (Vlamakis et al., 2008; Chai, Norman, et al., 2010; Chai, Kolter, et al., 2010; Dragoš et al., 2018; Lord et al., 2019). Starting with a planktonic founding population placed onto biofilm-promoting minimal medium (Msgg) agar, some of the cells can transition its phenotypes into matrix-producing phase (Fig. 1.1), in which they turn off genes for motility, e.g. *hag* (encoding flagellin), and instead, produce many protein necessary for matrix formation e.g. accessory scaffold protein TapA. These are two main subpopulations within the growing biofilms: Motile cells and Matrix-producing cells (hereafter, Matrix cells) (Tasaki et al., 2017). Interestingly, as biofilm grows larger, a ring of motile cells always re-emerges at the edge of the biofilm (Fig. 1.1). We hypothesised that, it was the matrix cells at the edge that switch their phenotypes back to motile phase, instead of escaping motile cells from the centre, or motile cells at the edge outcompeting the matrix cells. The underlying mechanisms enabling this second switch from matrix to motile phase were the focus of this research. Although many studies showed that matrix cells at the top of biofilm can irreversibly become sporulated once the nutrient was severely depleted (Vlamakis et al., 2008, Chen et al., 2022), this happened in a reasonably different time and spatial scale from our focused phenomenon, and thus, can be disregarded from our study.

It is known that a network of three proteins, SinI, SinR, and SlrR mainly govern this bidirectional motile-matrix switch. SinR is the main regulator that inhibit matrix genes and

is antagonised by SinI and SlrR. As illustrated in Fig. 1.2, this network has the SinR-SlrR double-inhibition feedback loop, meaning once SinI level is sufficiently high to repress SinR, the system can self-reinforce to further inhibit SinR through SlrR de-repression, leading to the cells being locked in a matrix state (Low SinR). For the first motile-to-matrix switch, it is thought to be SinI that increases due to nutrient depletion, eventually leading to the de-repression of matrix genes (Chen et al., 2022; Fujita et al., 2005; Fujita & Losick, 2005; López et al., 2010). It is, however, less clear what triggers the reverse in this phenotypic switch in the outer matrix cells back to motile phase.

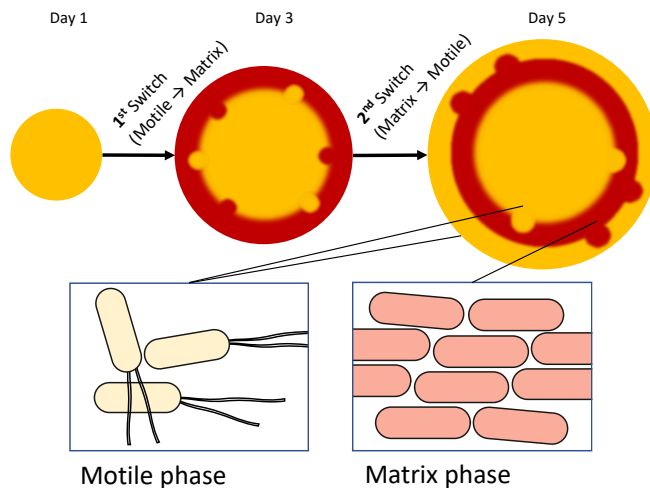


Figure 1.1: A general time-evolution of spatial heterogeneity of *B. subtilis*. In the founding populations, all cells are motile. Then, as biofilm grows, some cells transition into matrix-producing phase, shutting off motile genes (1st switch). Finally, as biofilm grows further, a second ring of motile cells re-emerges (2nd switch)

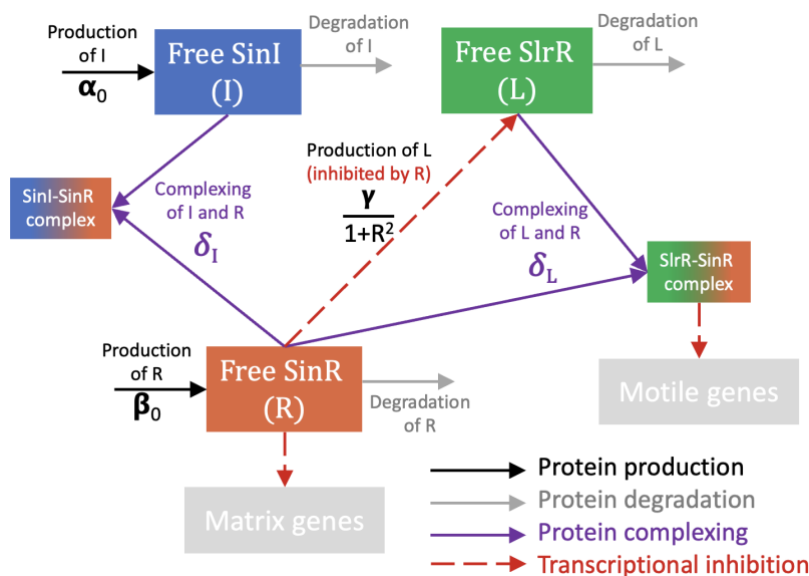


Figure 1.2: The SinI-SinR-SlrR regulatory network, deemed to mainly govern the matrix-motile switch. SinR, working as tetramer, is a transcriptional repressor of matrix genes and SlrR. SinI and SlrR can irreversibly bind to SinR and inhibit its repressive action. Additionally, SlrR can re-direct SinR to motile genes (Chai et al., 2010, Newman et al., 2013).

One emerging proposed mechanism for the observed matrix-to-motile switch is an influence from the stochasticity in the level of SinI, SinR, and SlrR (Lord et al., 2019; Chen et al., 2022). Lord et al. observed two-way phase switching of bacterial cells held in a microfluidic machine, where the nutrient level was maintained over time. More specifically, their stochastic model prediction showed that, with the noise in protein level included, these three proteins were sufficient to recapitulate the distribution of single-cell matrix and

motile phase duration measured from *B. subtilis* and from the *in vivo* reconstructed three-protein system in *E.coli*. However, their experiment and models did not consider the dynamics beyond a single cell system, while it is possible that the noise-responsive nature of the system may interact with environmental signal e.g. nutrients and produce the observed spatiotemporal pattern of the switching behaviours in *B. subtilis* biofilm (Chen et al., 2022). For example, cyclic di-AMP, a signal for nutrient depletion, can induce the activation of matrix genes, by affecting the activity of SinR, but not its level (Fujita & Losick, 2005, Gundlach et al., 2016), and in the context of matrix-to-sporulation transition, a prolonged nutrient depletion conversely decreases SinI and lead to the inhibition of matrix genes (Chai et al., 2011).

To date, it still has not been possible to accurately identify cells have undergone the phenotypic switch in an expanding biofilm, due to a technical difficulty in biofilm imaging and cell segmentation algorithm. Additionally, multiple alternating rings of motile and matrix population in the biofilm was also reported, but more hardly reproducible than the first two rings in Fig 1.1 (Tasaki et al., 2017). Hence, mathematical modelling provides an alternative way to investigate these phenomena in biofilm that involves complex interactions of many biological entities across different scales. We adapted two existing mathematical models: 1) SinI-SinR-SlrR ODE system used by Lord et al., 2019 (IRL-ODE system) and 2) PDE of *B. subtilis* population growth dynamic in agar extracted from Chen et al., 2023 (NuBac-PDE system). They were connected into one multi-scale modelling pipeline (Fig 2A), to investigate the effect of protein interactions, randomness, and nutrient level on the motile-matrix phase switching. We developed a second pipeline (Fig 2B), to parameterise the link function utilising the experimental data tracking *hag* and *TapA* expression, representative of motile and matrix phase, respectively. The final parameterised connected model allowed the investigation of how agar nutrient level may interact with protein level noises and to what extent this interaction can account for the observed spatiotemporal phenotype patterning seen in a real biofilm.

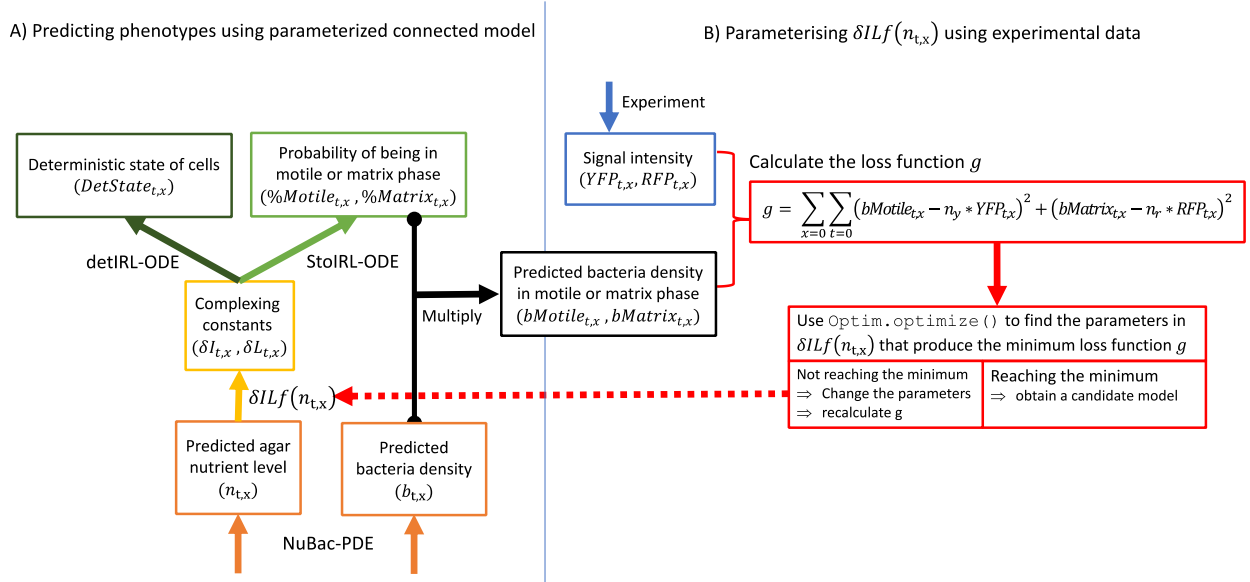


Figure 2: Brief overview of the two modelling pipelines.

A) Pipeline of the connected model composed of the bacteria-nutrient growth dynamics (NuBac-PDE), the model of complexing constants as a function of nutrient level (the link function; $\delta ILf(n)$), and the deterministic and stochastic model of SinI-SinR-SirR network (detIRL-ODE or stoIRL-ODE). The final outcomes for each time point and distance point are the deterministic state (DetState) and the bacteria density in motile or matrix states (bMotile, bMatrix).

B) Pipeline of the link function ($\delta ILf(n)$) learning using parameterised stoIRL-ODE and NuBac-PDE outcomes and experimental data. The built-in function `Optim.optimize()` was used as an algorithm to minimise the loss function (g)

Results

1. The deterministic SinI-SinR-SlrR system can be monostable, but become more prone to bistable at high protein interaction rates.

To explore a parameter space that create a bistable SinI-SinR-SlrR system (IRL-ODE; [equation 1.1-1.3](#)), we varied SinR-SinI and SinR-SlrR complexing constants (δ_I and δ_L , respectively), while fixing the production rates of SinI, SinR and SlrR, as in Table 1.1. We represented the two types of stable steady state by SinR level (R_{ss}), a repressor of matrix genes, a motile state (high R_{ss}) and a matrix state (low R_{ss}).

Table 1.1: Parameter values used in the analysis of single cell IRL-ODE model.

Parameter	Definition	Value	Unit	Source
α_0	Constitutive production rate of SinI	85	h^{-1}	Chen et al., 2023
β_0	Constitutive production rate of SinR	100	h^{-1}	Chen et al., 2023
γ	Maximum production rate of SlrR	125	h^{-1}	Chen et al., 2023
δ_I	Complexing rate of SinI and SinR	5	h^{-1}	Scanning result Newman et al., 2013
δ_L	Complexing rate of SlrR and SinR	2	h^{-1}	Scanning result, Newman et al., 2013

Table 1.2: Initial conditions (ic) used deterministic or stochastic IRL-ODE simulations.

Initial numbers of SinI, SinR, and SlrR	$I_0, R_0, \text{ and } L_0$, respectively
Deterministic ic	
Varied in time-evolution and phase plane	$I_0, R_0, L_0 = [10, \text{varied}, 10]$
Stochastic ic	
Low R_0 ic, leading to a deterministic low R_{ss} (matrix phase)	$I_0, R_0, L_0 = [100, 100, 100]$
High R_0 ic, leading to a deterministic high R_{ss} (motile phase)	$I_0, R_0, L_0 = [100, 200, 100]$

The system was bistable at high δ_I and δ_L (Fig. 3.1), based on the number of real positive roots of the equation for R_{ss} (derived in Appendix II). The example δ_L bifurcation diagrams (Fig. 3.2) also showed that the bistable region was smaller when $\delta_I = 1$, compared to when $\delta_I = 5$, and that the systems stayed at a motile state (high R_{ss}) in the monostable cases. These results well-represented the fact that, given the fixed protein production rates, the smaller these complexing rates, the higher level of steady state free SinR to repress matrix genes, and hence, the higher tendency for the system to be locked in a motile system. Consequently, the system has a larger bistable region as SinR production rates (β_0) decreased or SinI or SlrR production rates (α_0, γ) increased, exemplified in Sup. Fig. 1.1. Note that reducing all protein production parameters can force the system to be locked in a globally matrix state (low R_{ss}), exemplified in Sup. Fig. 1.2. However, we focused on only the case where $\alpha_0 = 85, \beta_0 = 100, \gamma = 125$ (as in Table 1.1), as these were values used in Chen et al., 2023, where we extracted the growth dynamics equations, later connected to this IRL-ODE model.

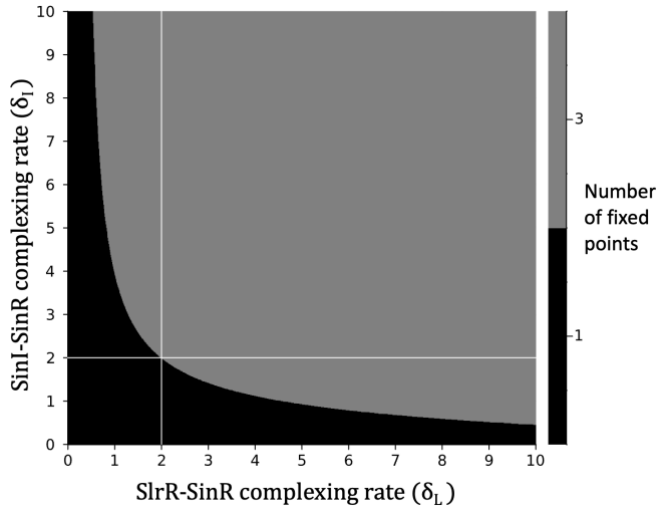


Figure 3.1: The analytical solution of the number of real positive steady-state levels of SinR (R_{ss}), which were the only valid type of R_{ss} . A black region represents a monostable system (1 real positive R_{ss}) and a grey region represents a bistable system (3 real positive R_{ss}). The scanning was performed on the SinR-SinI and SinR-SIrR complexing constants (δ_I and δ_L , respectively). White lines marked δ_I and $\delta_L = 2.0$. Parameters as in Table 1.1, if not stated otherwise.

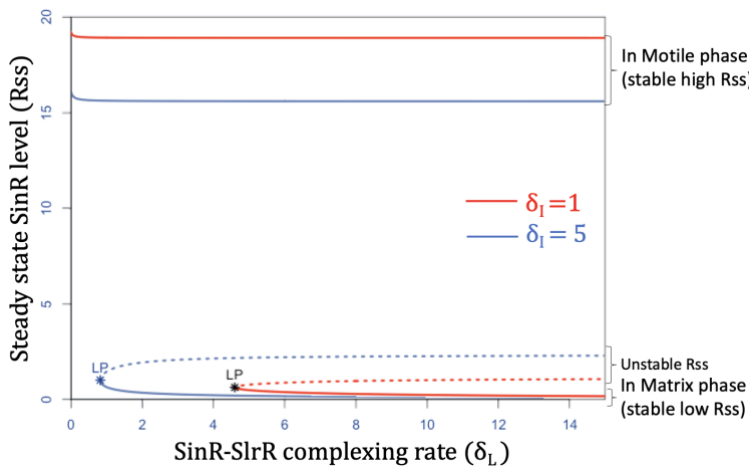


Figure 3.2: The bifurcation diagram showing the steady state level of SinR (R_{ss}) with varying values of δ_L from 0 to 15 when $\delta_I = 1$ or 5. LP: limit points (saddle-node bifurcation points). Dashed line: unstable fixed points. Solid line: stable fixed points. Parameters as in Table 1.1, if not stated otherwise.

2. Matrix-to-motile phase switching in a bistable system can be promoted by an increase in SinR level and SinI-SIrR complexing rate.

As expected, with the chosen parameters from bistability region (Table 1.1), depending on the initial conditions, the system may reach a high R_{ss} (motile phase) or a low R_{ss} (matrix phase), exemplified in the time-evolution plot (Fig. 4.1) for ten different initial numbers of SinR (R_0). The 2D phase planes (Fig. 4.2) of these ten systems highlighted the three fixed points and the fact that both SinI and SlrR level were anti-correlated to SinR i.e. high R_{ss} means low steady state SinI and SlrR level. Therefore, with the fixed parameters, the small increase in the level of SinR at a given time may overcome the inhibitive effect from SinI (e.g. from $R_0 = 5.0$, leading to a matrix low R_{ss} system, to $R_0 = 7.5$, converging to a motile high R_{ss}).

Interestingly, despite δ_I and δ_L similarly affecting the qualitative feature of the system (Fig. 3.1), the value of a high stable fixed point of SinR (high R_{ss}) was varied to a greater extent due to changing δ_I , compared to changing δ_L over the same scale (Fig. 5A). This result represented the greater quantitative influence of the interaction between SinR

and the switch molecule SinI, over the interaction between SinR and SlrR, which involved in the downstream double-inhibition loop (Fig. 1.2). The strong negative correlation between the value of high R_{ss} and δ_I meant that this parameter can be a target for any signal to influence to effectively push a high R_{ss} further away from or closer to a low R_{ss} , which was not majorly changed as δ_I and δ_L changed (Fig. 5B), and hence, affecting the probability of the cells to transition from a motile to matrix phase, in a bistable system.

Figure 4: The dynamics of deterministic SinI-SinR-SlrR (IRL-ODE) systems, beginning with 10 different initial numbers of SinR (R_0): 0.0, 0.1, 0.5, 1.0, 2.5, 5.0, 7.5, 10, 15, and 20. Starting with $R_0 \leq 5.0$, the system converged to a low R_{ss} (matrix phase), and when $R_0 \geq 7.5$, the system went to a high R_{ss} (motile phase). The parameters and initial conditions as in Table 1.1-1.2.

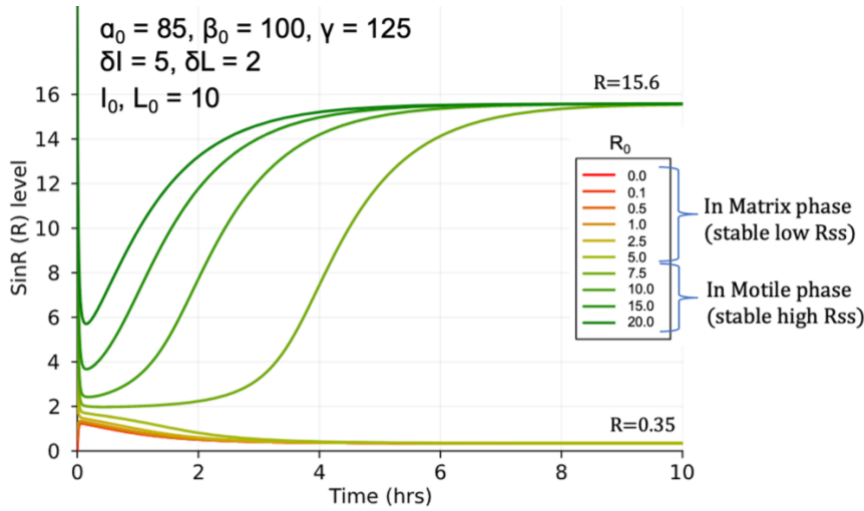
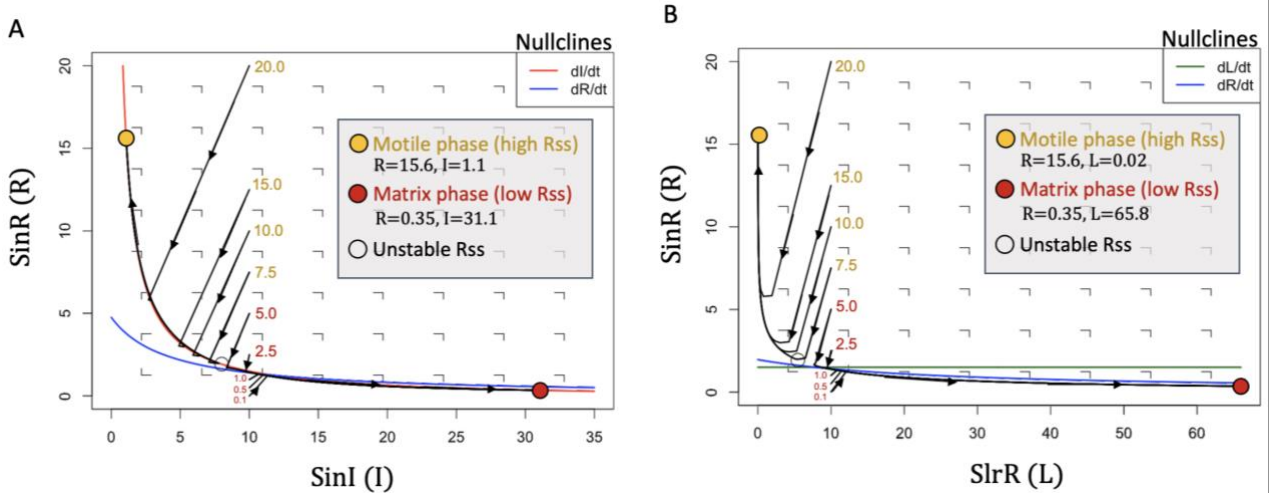


Figure 4.1: The time-evolution plots. Each line starting with different R_0 , stated in the legend. There were two stable steady state: a low R_{ss} ($R_{ss} = 0.35$), representing a motile phase, and a high R_{ss} ($R_{ss} = 15.6$), representing a matrix phase.

Figure 4.2: The 2D phase portrait showing the three fixed points, the nullclines and vector field (A) between SinR-SinI and (B) SinR-SlrR. Each value of R_0 is labelled next to starting points in the graphs, coloured-coded by the type of fixed point to which it converged.



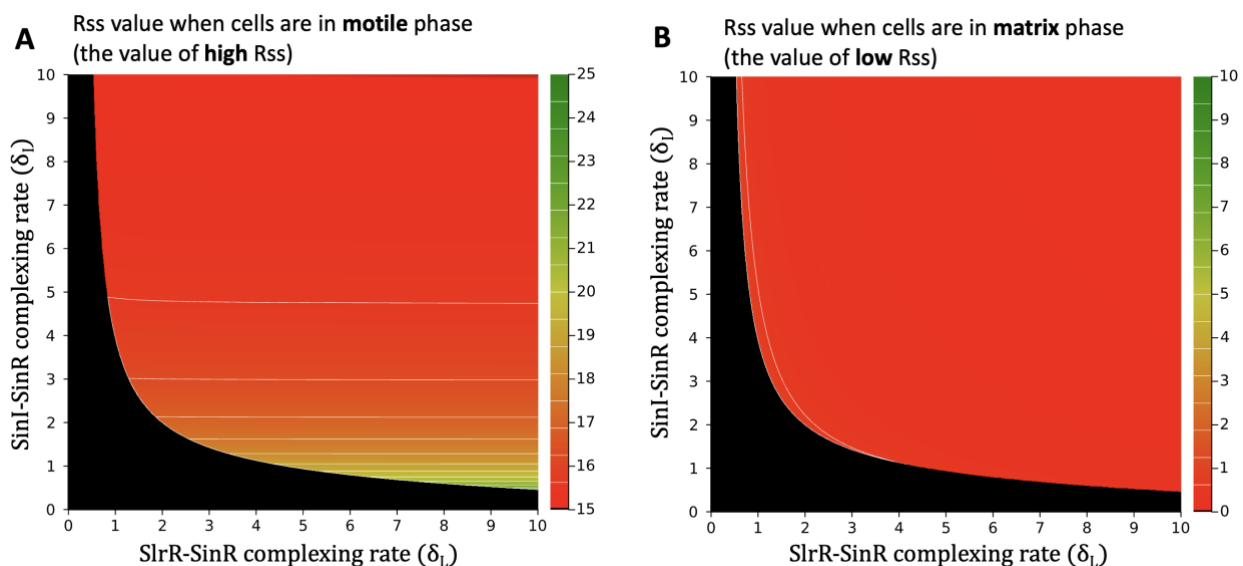


Figure 5: The analytical solution of the values of **A)** a high steady-state level SinR (Rss when cells are in motile phase) and **B)** a low steady-state level SinR (Rss when cells are in matrix phase). Note that each heatmap has a different colour range but the identical step size (both total ranges spanned 10 levels). A black region represents an irrelevant parameter space.

3. Noises can introduce an alternative steady state, even in a deterministically high-SinR monostable system.

To evaluate the effect of noises on cell phase determination, we conducted stochastic simulations of the IRL-ODE system, using Gillespie's method. Accurate estimation of the number of SinI, SinR, and SlrR molecules present in *B. subtilis* was required, and hence, we opted for numerical values on an order of 100 molecules, akin to those employed in Lord et al.'s (2019) study. The two initial conditions able to lead the system with the parameter values in Table 1.1 to either a motile state (high Rss) or (low Rss) were $I_0, R_0, L_0 = [100, 200, 100]$, or $[100, 100, 100]$, respectively (Table 1.2, Sup. Fig 2.1). To generalise the assessment of noise effect, we tested the IRL-ODE system with varying SinI-SinR and SlrR-SinR complexing rate, δ_I and δ_L , each ranging from 0 to 10. If Rss was ≥ 8 molecules (midpoint between two deterministic fixed points), the system was motile; otherwise, matrix phase.

Starting with a high R_0 ic $[100, 200, 100]$, regardless of the values of scanned δ_I and δ_L , the deterministic system always reached a motile state (a high Rss), while it was not always the case for the stochastic runs, even in some conditions where the system was deterministically monostable (refer to Fig. 3.1). More specifically, according to Fig. 6.1, only at $\delta_I = 0$ that all stochastic runs converged to a motile state, unaffected from noises. This observation reaffirmed that δ_I has a stronger impact on controlling the system compared to δ_L . However, as either independently δ_I or δ_L increased, there was a fewer chance of a cell resulting in motile state (Fig. 6.1), so the control from δ_L was also evident.

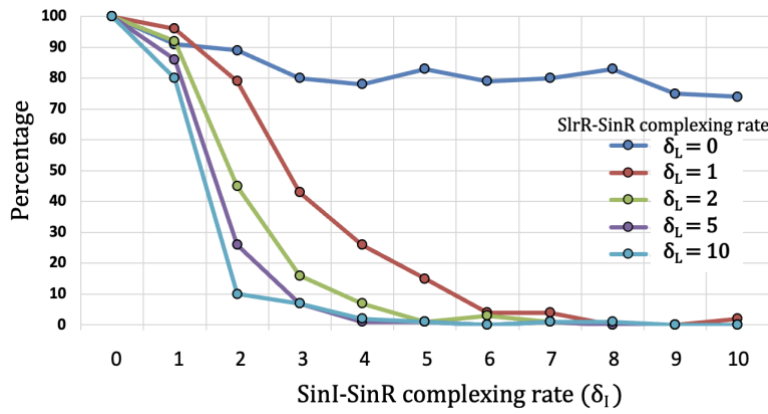


Figure. 6.1: Percentage of a system with a high R0 initial condition, [100,200,100] reaching a motile state (a high Rss). Both complexing rates were varied from 0 to 10, but the results from only five SlrR-SinR complexing rate (δ_L) were shown. The noise shifted the system to the alternative low SinR fixed point, indicative of a matrix cell. The strongest effect was seen at the highest δ_L .

On the contrary, starting with a low R0 ic [100,100,100], most stochastic outcomes agreed with the deterministic outputs over our scanned range of the complexing rates (Table 2). Nevertheless, by comparing Table 2A to Table 2B, we could still saw more noise-inducing downward shift from a deterministic motile system to a matrix steady state (15 out of 26 combinations shifted downward), compared to the upward shift from a deterministic matrix system to a motile state (19 out of 74 combinations shifted upward).

Table 2: Percentages of the IRL system starting with a low R0 [100,100,100] that eventually reached a high steady state SinR or a **motile** state, from (A) deterministic run and (B) 100 stochastic runs. The scanning was carried over varying complexing rates (δ_I and δ_L) from 0 to 10. In both tables, a bordered cells symbolised the deterministic motile system (26 out of 100).

A Deterministic output											
$\delta_I \backslash \delta_L$	0	1	2	3	4	5	6	7	8	9	10
0	100	100	100	100	100	100	100	100	100	100	100
1	100	100	100	100	100	0	0	0	0	0	0
2	100	100	0	0	0	0	0	0	0	0	0
3	100	0	0	0	0	0	0	0	0	0	0
4	100	0	0	0	0	0	0	0	0	0	0
5	100	0	0	0	0	0	0	0	0	0	0
6	100	0	0	0	0	0	0	0	0	0	0
7	100	0	0	0	0	0	0	0	0	0	0
8	100	0	0	0	0	0	0	0	0	0	0
9	100	0	0	0	0	0	0	0	0	0	0
10	100	0	0	0	0	0	0	0	0	0	0

B Stochastic output											
$\delta_I \backslash \delta_L$	0	1	2	3	4	5	6	7	8	9	10
0	100	100	100	100	100	100	100	100	100	100	100
1	92	94	92	92	86	62	65	41	28	32	19
2	84	76	23	2	0	0	0	0	0	0	0
3	85	42	0	1	0	0	0	0	0	0	0
4	86	18	1	0	0	0	0	0	0	0	0
5	82	4	0	0	0	0	0	0	0	0	0
6	89	4	1	0	0	0	0	0	0	0	0
7	73	3	0	0	0	0	0	0	0	0	0
8	83	1	0	0	0	0	0	0	0	0	0
9	88	1	0	0	0	0	0	0	0	0	0
10	77	1	0	0	0	0	0	0	0	0	0

In Table 2B, the numbers in **bold** symbolised the condition that showed the difference in the outcome. The bold number in a bordered cell, thus, means the combination that saw a downward shift from high to low Rss.

Putting together the predicted percentages of cells being in motile phase from both the initial conditions in the context of bifurcation diagram, it is evident that the matrix phase (low Rss) was greatly favoured by noise (Fig. 6.2), and even got introduced in a deterministically motile system. Plus, as predicted in the Result part 2, noise perturbation increased the motile-to-matrix switching to a greater degree as the complexing rates increased.

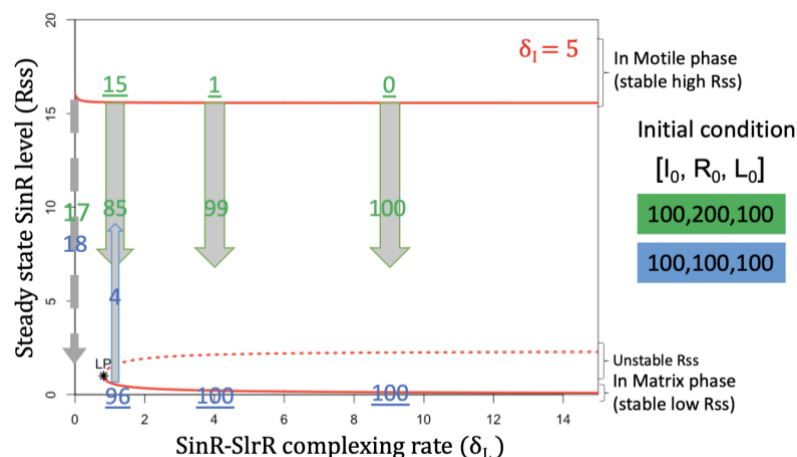


Figure. 6.2: The bifurcation diagram for when $\delta_I = 5$ with the percentages of a system to switch away from the deterministic prediction (in an arrow) or to stay as the deterministic prediction (underlined), annotated for $\delta_L = 1, 4$, and 7 . Initial conditions of the systems were either a **high R_0 ic, $[100, 200, 100]$** or with a **low R_0 ic $[100, 100, 100]$** . The dash arrow means the shift in deterministic monostable system. The noise switched more motile cells to matrix state and kept more cells as matrix state.

It was interesting to see if, for a monostable low-SinR matrix system, the noises also favoured matrix state, or it shifted the equilibrium away from the deterministic fixed points. We additionally carried out the equivalent analysis on this system (Sup. Fig. 2.2), but similar to all the previous results for a monostable high-SinR motile and bistable system, the occurrence of the matrix state.

4. The distribution of the steady state SinR level differed between a deterministic monostable and bistable system and was minimally affected by initial conditions.

As these percentages depended on the selection of the cut-off values as well, examining the distribution of Rss can shed light on how noises interact with the complexing rates to affect the system more quantitatively. Thus, we created the histograms of the steady state SinR (Rss) from stochastic 500 runs starting with either a **high R_0 $[100, 200, 100]$** (Fig. 7.1) or a **low R_0 $[100, 100, 100]$** (Fig. 7.2)

In a monostable motile system (**A** in 7.1 vs 7.2), with $(\delta_I, \delta_L) = (1, 1)$ (), Rss was spread around the mean close to the deterministic high Rss (18.9). Interestingly, despite their opposite deterministic outputs, in other combination of (δ_I, δ_L) , the distributions were also highly similar (**B-D** in 7.1 vs 7.2). More specifically, another peak emerged around to the deterministic low Rss when δ_I and δ_L were high, overall creating a bimodal distribution (**B** and **C**). This low-Rss peak was slightly more favoured compared to a high-Rss peak when the system started with a **low R_0 $[100, 100, 100]$** , and was the only population when (δ_I, δ_L) was as high as (5,2) (Fig 7.2D).

Overall, this result showed that, in the monostable system (i.e. when the complexing rates were too low to allow SinR sequestration), noises spread the values of

steady state SinR level so that, in some instances, it was pushed below a critical point where it was deemed insufficient to effectively inhibit matrix genes (< 8), and therefore, the matrix genes could be stochastically turned on. In the bistable system, the peak at low Rss over dominated the peak at high Rss as the complexing rates increased. Comparing 7.1 to 7.2 also showed that the spread was determined more strongly by the combination of complexing rate parameters δ_I and δ_L , rather than the initial conditions (or deterministic output). Due to these apparent influences from noises and the varying magnitude of the complexing rates, we allowed these rates to be affected by nutrient and included noises in the subsequent investigations of the impact of nutrient on cell phase switching.

Figure 7: The distribution of steady state SinR level (Rss) from 500 runs, using $R_{ss} \geq 8$ as a cut-off for a high Rss, representative of the system reaching a motile state. Only four combinations of complexing rates shown, labelled at the top of each histogram. The deterministic state was also stated for the comparison in each plot. The production rates were as in Table 1.1. The red vertical line marked $R_{ss} = 8$, below which the system was regarded as a matrix cell.

Figure 7.1: starting with a high R_0 [100,200,100]

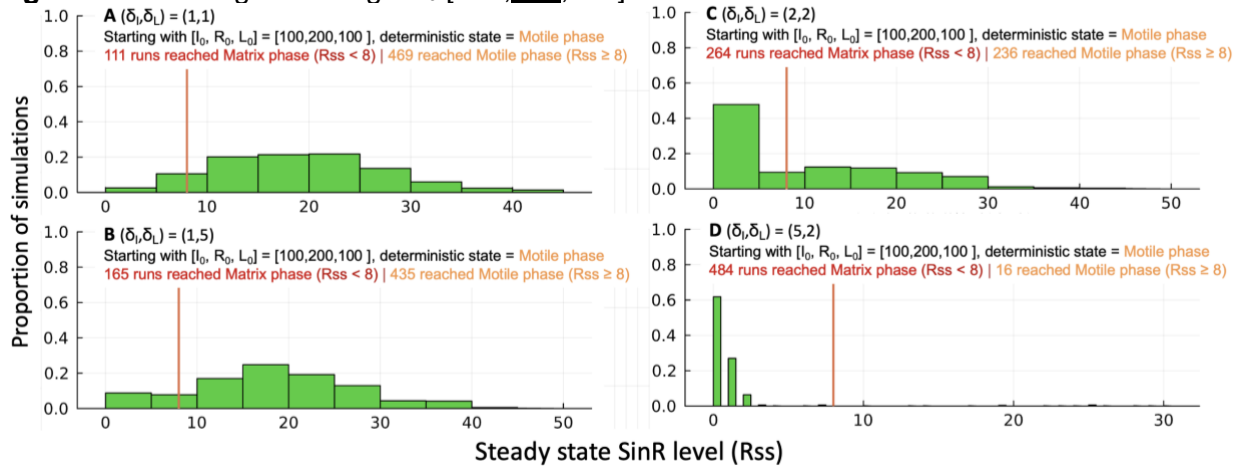
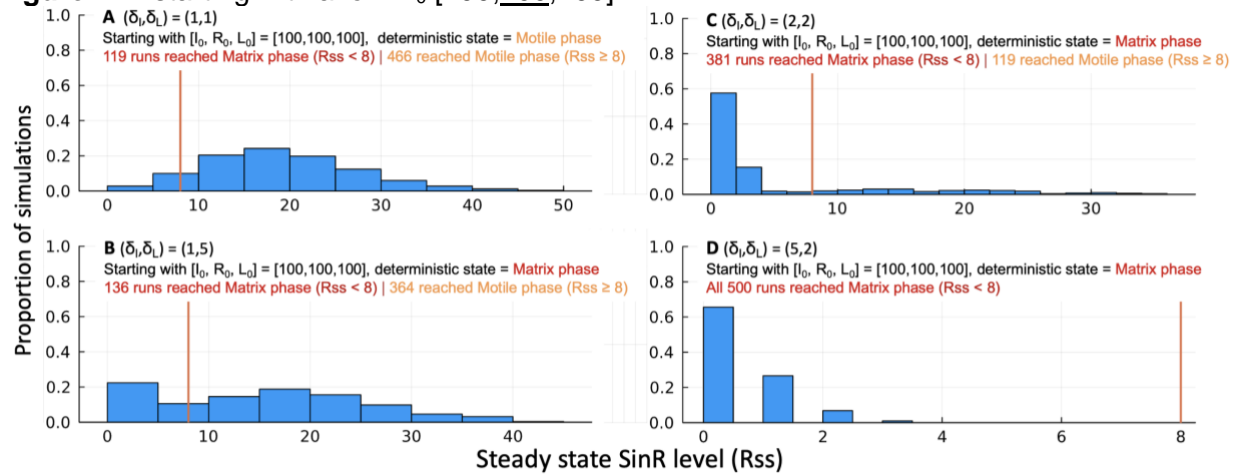


Figure 7.2: starting with a low R_0 [100,100,100]



5. Noises can widen the width of matrix cell ring when agar nutrient, linked to the complexing parameters, deterministically enabled phase shifting.

Aiming to simulate a possible nutrient dynamic in an agar, we extracted the snapshot of the agar nutrient profile experienced by the 16h-grown biofilm, from (Fei et al., 2020). It was then modelled by an exponential curve decaying from the edge to obtain the continuous nutrient level over a distance from the centre (Fig. 8.1A). By imposing a hypothetical link function describing how nutrient profile related to the SinR-SlrR complexing rate parameter (δ_L) (equation 3, Fig. 8.1B), we obtained the pattern of the steady state shifting from a motile state (a high R_{ss}) to a matrix state (a low R_{ss}) at a distance around 3 mm from the centre, and then back to a motile state at a distance around 4.5 mm (Fig 8.2B). This was expected from the way we set the link function. For example, at 3 mm, it was when $\delta_L \geq 0.8$ (Fig 8.2A), which was the threshold of δ_L proved to switch the deterministic steady state from a high R_{ss} to a low R_{ss} with the specific set of parameters in Table 3 (the result not shown).

We then added noises to the simulation to investigate if it changed the emergence position of the first motile-to-matrix and the second matrix-to-motile switching, which would consequently determine the width of matrix cell span. With noises, the probability of being in a motile phase was never more than 0.75 and hit 0.50 at a distance 2 mm (Fig. 8.2C), where δ_L was still as low as 0.25 (Fig. 8.2A). By treating 100 stochastic runs as individual cells and considering 0.50 as a probability threshold of for the emergence of matrix cells, we could make the case that the randomness caused the width of the matrix cell span to increase from 1.5 mm (between 3.0 and 4.5 mm in Fig. 8.2B) to 2.5 mm (between 2.0 and 4.5 mm in Fig. 8.2C).

With these results showing the potential induction of switches by the decrease in nutrient level, which was additionally promoted by its interaction with noises, we aimed to use a spatiotemporally continuous characterisation of nutrient field, along with an experimental fluorescence tracking of motile and matrix subpopulations in a real biofilm to obtain a higher confidence relationship between the nutrient and the complexing parameters.

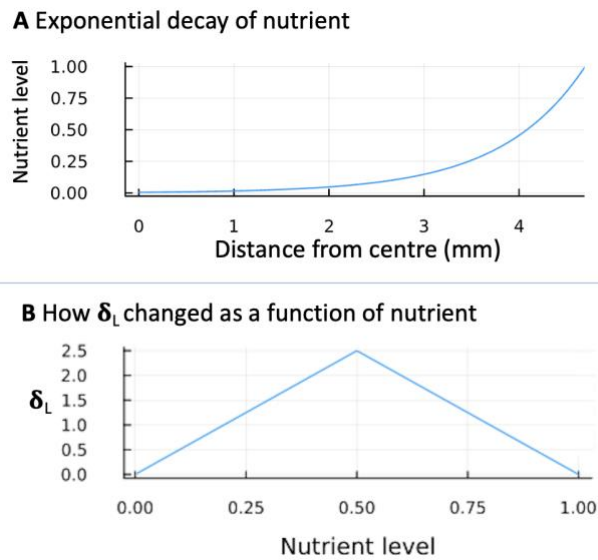


Figure 8.1: The exponentially decaying nutrient profile (**A**) used as the input to equation 6, describing the non-monotonic function relating nutrient level to SlrR-SinR complexing rate (δ_L) (**B**).

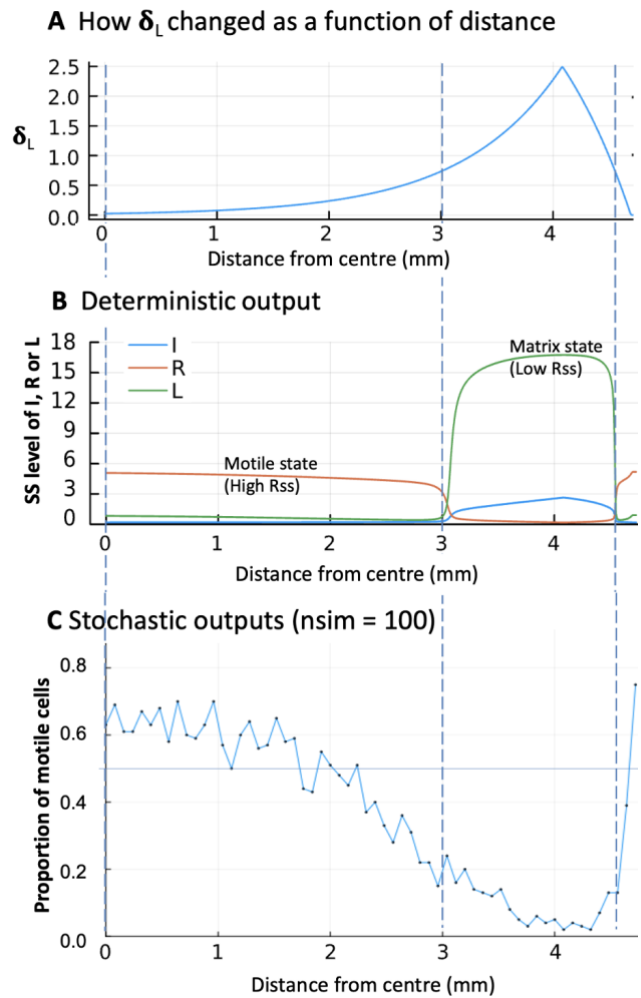


Figure 8.2: The resulting profile of δ_L across the distance of biofilm (**A**), which was used as a parameter in a single deterministic run and in 100 stochastic runs of the IRL system, while other parameters were fixed as in Table 1.3. The steady state level of I, R and L were recorded for a single detIRL run (**B**), while the probability of reaching a motile state was calculated from the proportion of 100 runs that has a high Rss (**C**).

6. The success of the search for the link between nutrient level and the status of SinI-SinR-SlrR network was limited.

With the scaled YFP and RFP signals as proxies for experimental motile and matrix cells, the parameterised biofilm growth NuBac-PDE model (the relevant parameterisation shown in Appendix III), and the single cell IRL-ODE system, it is possible to merge them into one multi-scaling model. It should be noted that certain parameters in the NuBac PDE model were adjusted (as listed in Table 4) from those presented in the original paper to better recapitulate the movement of YFP and RFP signals in our data ([Movie S2](#)).

Table 4: Parameter values used in the bacterial growth dynamics NuBac-PDE model. Asterisks indicated adjusted parameters for the better representation of model to experimental data and the original values were placed in the last column.

Parameter	Definition	Value	Unit	Source
D_n	Diffusion coefficient of nutrient	0.84*	$mm^2 \cdot h^{-1}$	Miyamoto et al., 2018 (3.36)
D_b	Diffusion coefficient of bacteria	0.030	$mm^2 \cdot h^{-1}$	Fitting result (Appendix III)
k_g	Maximum growth rate of bacteria	0.475*	h^{-1}	Chen et al., 2023 (1.90)
k_d	Maximum death rate of bacteria	0.155*	h^{-1}	Chen et al., 2023 (0.62)
ω	Yield coefficient of bacteria	0.10	-	Chen et al., 2023
θ	Fraction of nutrient released by dead bacteria	0.30	-	Chen et al., 2023
K	Half-saturation nutrient for bacterial growth or death	0.0061*	-	Chen et al., 2023 (1.21)

We connected the nutrient predicted from the NuBac-PDE model to the two complexing parameters δ_I and δ_L , using 4th-order polynomial equations (hereafter, dILf(n) function). The coefficients in dILf(n) were optimised by the minimisation of the sum of squared deviation between the experimental density of motile and matrix cells (i.e. the scaled YFP and RFP signals) and the multi-scaling model stochastic prediction of motile and matrix cells (bMotile(t,x) and bMatrix(t,x)), respectively ([equation 8](#)).

However, starting with either coefficients producing a function with a similar shape to equation (6) or ten random initial coefficients, none of the resulting 11 models were identical, representing several local minima in this landscape. 9 of them led to negative δ_I or δ_L , and the remaining two models produced large δ_I or δ_L values (Fig. 11A-B). This can be attributed to the constraint imposed on the range of δ_I or δ_L in the parameterization pipeline, which was restricted to the interval [0, 10]. If the output of dILf(n) fell outside of this range, the δ_I or δ_L was rounded to the nearest bound. However, one consistent aspect was that the magnitude of δ_I was predicted to be higher than δ_L in all models.

7. Stochastically and deterministically simulation of the final multi-scaling model showed a minimal noise effect on phase switching in a biofilm.

Having non-conclusive models, we tried using all 11 parameterised dILf(n) functions as the link between NuBac output ($n(t,x)$) and IRL-ODE parameters (δ_I or δ_L) in the final multi-scaling model, applying the same [0, 10] constraint.

Focusing on the stochastic branch of simulation (refer to Fig. 2A), the predicted density of motile and matrix bacteria (bMotile(t,x) and bMatrix(t,x), respectively) from 5 out of 11 final parameterised connected models replicated some aspects of the experimental counterparts, exemplified in Fig. 12.1, while the others predicted no emergence of matrix cells at all (result not shown).

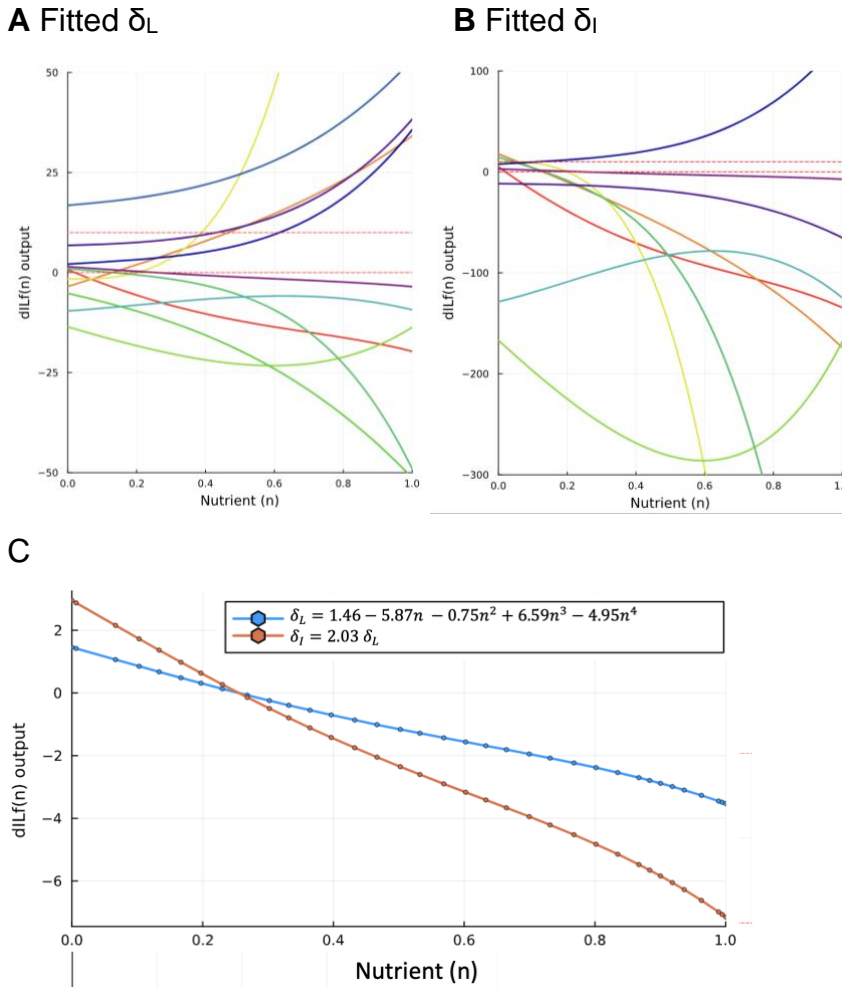
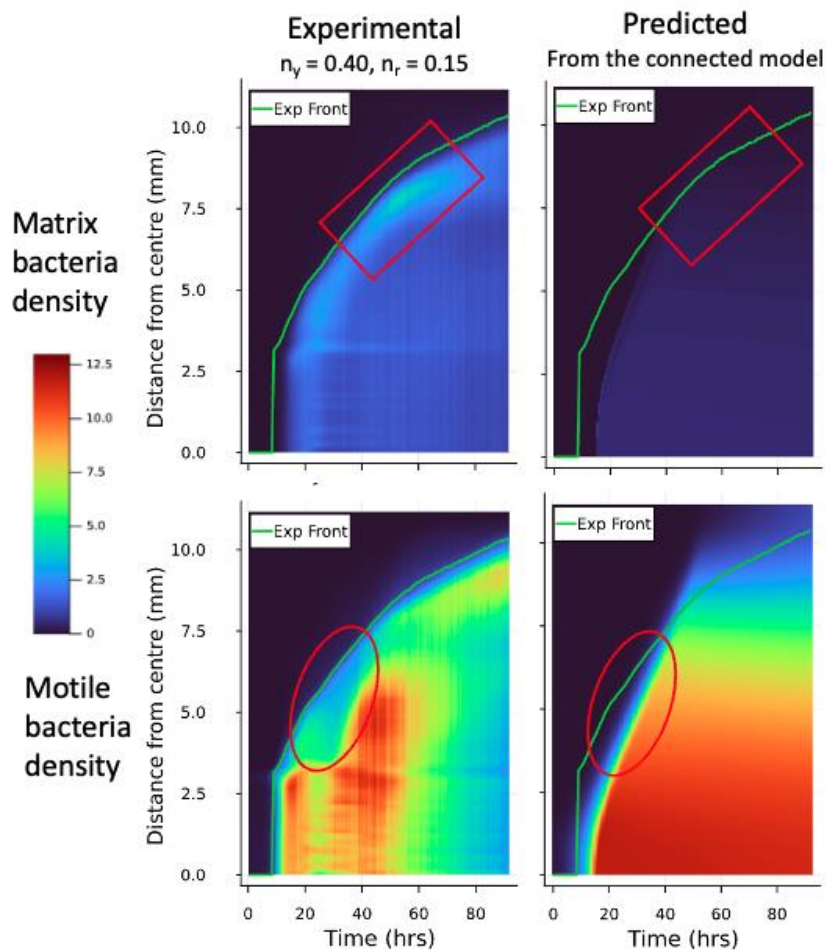


Figure 11: Top two graphs showed highly varied 11 resulting dILf(n) models produced by 11 independent rounds of the optimisation, each starting with different initial coefficients in dILf(n). **A)** fitted- δ_L and **B)** fitted- δ_I . The horizontal red dash line marked the range of allowed δ_L and δ_I used during optimisation (0-10). With the possible range of nutrient produced by PDE, between 0 and 1.0, most models predicted negative values of δ_L and δ_I and some produced δ_L and δ_I . **C)** The parameterised dILf(n) model, used in the multi-scaling connected model (as shown in Fig. 2A) to produce the **example predictions** shown in Fig. 12. This particular model resulted from using a similar shape to equation (6) as the beginning function

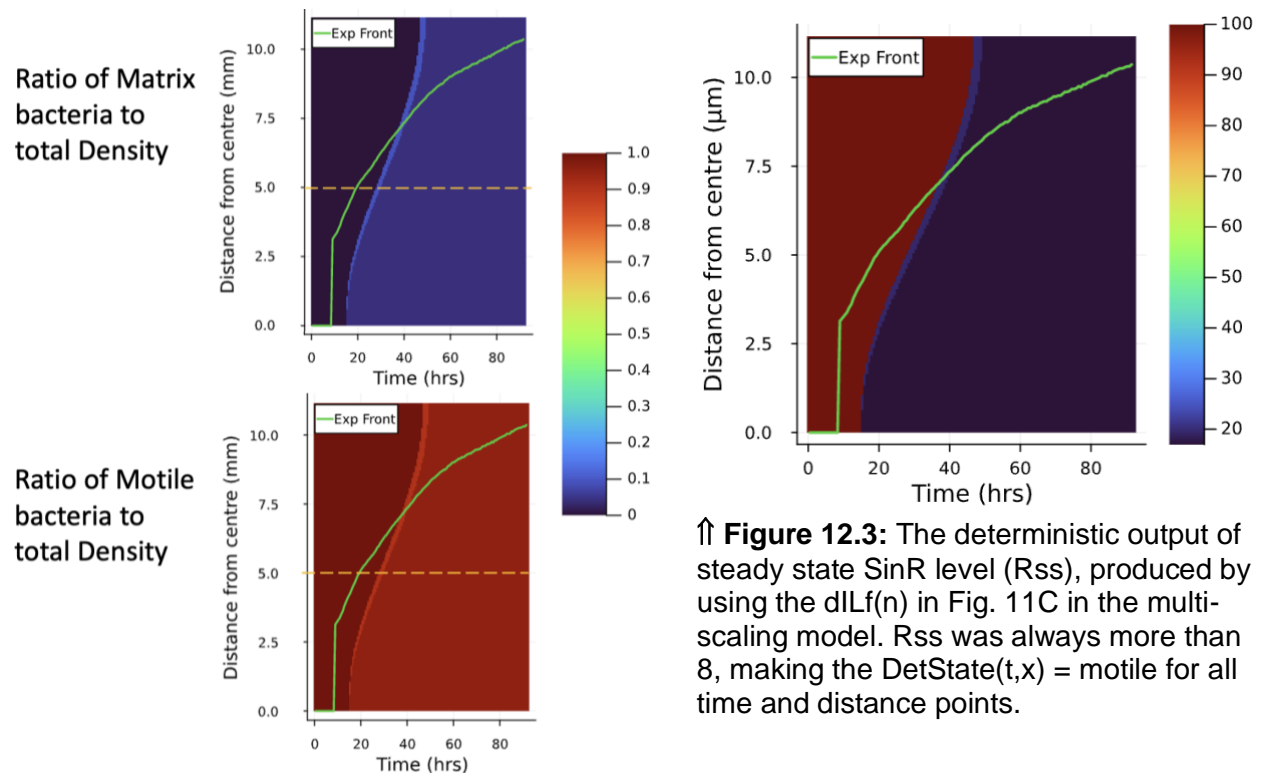
By examining the ratio of matrix cells (or the motile cells) to the total bacteria population, we could assess whether the double-switching pattern (i.e. the first motile-to-matrix and the second matrix-to-motile switch) was reproduced. Unfortunately, all of the models showed only one instance of the increase in the proportion of matrix cells, representing only the first switch (result not shown), except one model (Fig. 11C) that showed a brief fluctuation of the signal, depicted in Fig. 12.2. For example, at the distance 5 mm from the centre (dashed yellow line), the proportion of motile cells decreased from 1.00 to 0.90, and then increased to 0.95. However, this did not represent the two-switching pattern as the drop was brief and the motile proportion during the drop was still far greater.

Focusing on the deterministic branch of simulation (refer to Fig. 2A), although the predicted steady state SinR level (R_{ss}) was decreasing over time, it was never less than 8 in any of 11 models (Fig. 12.3). Consequently, the predicted deterministic state (DetState(t,x)) was also exclusively the high R_{ss} (motile state), whereas we would expect to see the motile-to-matrix switch similar to Result part 5. Due to time constraints, the necessary modifications to the optimization could not be successfully implemented (later in Discussion).



⇐ **Figure 12.1:**

The comparison between the example model prediction (**Right**) and the experimental data (**Left**) of the density of matrix (**Top**) and motile cells (**Bottom**) as a function of growing time and distance from the centre. The model generally correctly predicted 1) the slightly earlier emergence time of motile cells, 2) the smaller matrix subpopulation size, 3) the decrease in motile population (**red ellipsoid**). However, it was unable to show the pattern of matrix cell density that saw the increase at around 7.5mm at 60hrs (**red rectangle**). Green line represents the front of experimental biofilm (bfront), below which the output was valid for interpretation, the same for Fig 12.1-12.3.



↑ **Figure 12.3:** The deterministic output of steady state SinR level (R_{ss}), produced by using the dILf(n) in Fig. 11C in the multi-scaling model. R_{ss} was always more than 8, making the DetState(t,x) = motile for all time and distance points.

↑ **Figure 12.2:** The prediction of the ratio of matrix cells (**Top**) or motile cells (**Bottom**) to the total populations, produced by using the dILf(n) in Fig. 11C in the stochastic multi-scaling model. The brief drop (to 0.90) in motile cells was seen, indicative of the potential two-switching pattern, although the magnitude of the drop was expected to be higher. The orange dashed line was at the distance 5mm from the centre, at which the outcome was mentioned in the main text.

Discussion

1. The stochastic shift to matrix phase can be explained by the distribution of steady state SinR.

The present study has shown that the demographic stochasticity, arising from the finite number of protein molecules in the system, can indeed affect the mutually exclusive matrix-producing/motile phenotype of *B. subtilis* cells, that are under a control of three protein regulatory system, SinI-SinR-SlrR (Fig. 2A). Interestingly, the matrix-producing phenotype, represented by a low steady state SinR level, was favoured by these random fluctuations. This was always the case regardless of independent of the initial numbers of SinR protein and the deterministic state, if the interaction between SinI and SinR existed ($\delta_I \neq 0$). Therefore, the re-emergence of motile population at the edge was then unlikely to be caused by noises, but the first motile-to-matrix phenotypic transition may be triggered by noises, in addition to the direct induction of SinI due to early decrease in nutrient.

The distributions of steady state SinR induced by the presence of noise could potentially account for this matrix phase-shifting phenomenon. It seemed to be the combination of two Poisson distributions, typical distribution of stochastic occurrence of rare events in a fixed interval of time (Elf & Ehrenberg, 2003). Since one of the Poisson distributions had a mean close to 0, SinR population was effectively pushed to zero (due to its high right-skewness). While using the midpoint between the two Rss means (0.3 and 15.6) as the threshold for identifying matrix cells as in our analysis would be equivalent to incorrectly assuming a uniform distribution of SinR between the two means. In other words, our threshold ($=8$) could be too high, biasing the occurrence of below-threshold populations. Therefore, scanning over many threshold values to assess the impact of the threshold can improve the reproducibility of the result. An alternative approach would be to consider the steady-state levels of all three proteins, not only SinR, and incorporate the biochemical evidence for the stoichiometry of their binding events to construct a more accurate threshold. For example, the presence of equimolar SinI may be required for the complete disruption of SinR (Milton et al., 2020). Thus, it is possible that SinR must fall further below 8 molecules to be antagonised given that the steady-state SinI level was only 1-2 molecules. This could negate the increase in the likelihood of cells entering in a matrix phase, and therefore, this additional analysis should be performed.

2. SinI-SinR interaction is more important in controlling phase than SlrR-SinR.

The greater influence of SinI-SinR complexing constant (δ_I), compared to SlrR-SinR complexing constant (δ_L), was evident from both the quantitative analysis of its effect on a

value of a high deterministic R_{ss} (Fig. 5) and a more qualitative analysis of its effect on percentage of cells entering a motile phase from stochastic simulation (Fig. 8.1). This agrees with a paper by Lord et al., 2019 concluding that SlrR only prolonged the matrix phase and did not change qualitative properties of the system. Moreover, in all of the parameterised link function that predict δ_I and δ_L , the fitted δ_I has a higher magnitude than δ_L , agreed with the reported higher association constant between SinI-SinR than that of SlrR-SinR (Newman et al., 2013).

It should be noted that our model simplified their binding interactions, so we can include the detailed complexing step of SinI-SinR and SlrR-SinR to be more confident with the conclusion, if the relevant parameters were more well characterised. However, one of the past models in Lord et al., 2019 that included the detailed steps of complexing between SinR-SinI (i.e. 2 SinR can dimerise to form a homodimer SinR, SinI can dismantle SinR dimer, 2 SinR dimers form SinR tetramer, and SinR tetramer can then bind and inhibit a reporter gene) can be matched qualitatively by a simpler model (where 4 SinRs directly bind the reporter gene). Therefore, this comparable analysis can be carried out for SlrR-SinR complexing, to assure that we can safely simplify their interaction.

3. The experimental data produced reasonable statistics but could be more well-described to allow a more accurate abstraction of growth dynamics.

Our result showed that the biofilm front (bfront) expanded with the rate $\sim 112 \mu\text{m.h}^{-1}$, comparable to the expansion velocity reported in Fei et al., 2020 ($180 \mu\text{m.h}^{-1}$). The induction of matrix-producing cells usually accompanies the increase in bacterial density due to its chaining phenotypes and matrix components in which the cells embedded (Tasaki et al., 2020). In our data, the matrix population emerged at around 11-12th hours of biofilm growth (Fig. 9.1 in Appendix III), and this was around the same time point, where the increase in cell density was reported by Tasaki et al., 2020.

However, the NuBac-PDE model fitted by these reliable experimental values overestimated bacterial density for most of biofilm growing time (Sup. Fig. 5). This posed a potential source of error for downstream model fitting, including the parameterisation of fluorescent coefficients and $dILf(n)$ coefficients. This overestimation can be attributed to biofilm wrinkling which occurred after around 18-20 hrs of biofilm growth by inspecting the raw image data. It can also be caused by the decay in fluorescence proteins. Therefore, a better description of both fluorescence reporters in terms of their lifetime *in vivo* and their relationship to the source cell density should be sought after as this will explain if its decay caused this overestimation and allow us to have a more realistic proxies for experimental

motile and matrix subpopulation density. Alternatively, a reporter construct under a constitutive promoter can also be inserted to biofilm colony for a more direct parameterisation of fluorescent coefficients.

4. Several model and pipeline adjustments were required to overcome the limited success of multi-scaling model.

We could show that the noises may increase the span of matrix cells in a biofilm if the system deterministically experienced both the first motile-to-matrix switching and the second matrix-to-motile switching. This double switching can be induced by the decrease in agar nutrient influencing the complexing rates in a non-monotonic way i.e. as the nutrient decreases, the complexing rates increases up until a certain point after which the decreased with decreasing nutrient.

However, the final parameterised multi-scaling model could not produce this pattern either from deterministic or stochastic types of simulations, even when started the optimisation with a non-monotonic function. The matrix subpopulation was predicted to be universally much fewer than motile subpopulation, contrary to the results, showing noise-promoting matrix phase. This mainly due to the unsuccessful parameterisation of the $dILf(n)$ function that linked the nutrient predicted from NuBac-PDE model and the complexing rate parameters in IRL-ODE model. One issue can be the upstream uncertainty in NuBac-PDE parameters, obtained by arbitrarily modifying original values such that the resulting movement ([Movie S2](#)) approximately reproduce the moving pattern of the data ([Movie S1](#)).

Besides, the $dILf(n)$ parameterisation algorithm itself should be adjusted. Because of the time limitation, we only produced two small look-up tables, in which the algorithm searched for the probability of the system being in a motile state (P_{motile}) for a given combination of the complexing rates. The ranges of the complexing rates were only 0 – 10 with the resolution of 1, and consequently, there was a considerable estimation of the complexing rates and P_{motile} . Plus, using a lookup table sacrificed realistic noise inclusion. Therefore, we could either increase the range/resolution of mentioned lookup tables, or run stochastic simulation for every time and distance point, if we had more time or smaller-scale data, if time permits or dealing with a smaller-scale data. Alternatively, a better parameterisation pipeline that penalises the model that produced the complexing rates outside this range can also improve the validity of the $dILf(n)$ model.

Lastly, in our final multi-scaling model, we also assumed that the steady state of the intracellular SinI-SinR-SlrR network (IRL-ODE) has been reached within 1-hr step size of

the growth dynamic simulation (NuBac-PDE). While this was a reasonable assumption for the deterministic simulation branch for one of the initial condition cases (a low R_0 ic), it was less valid for the other case. More importantly, the stochastic simulations required no less than 20-hr time span to correctly differentiate between a motile phase- or matrix phase-converging simulation. Therefore, it is recommended to feed a more correct output from NuBac-PDE, with respect to time, to IRL-ODE system. One way is to adjust the parameters in IRL-ODE so that it evolves much faster.

Materials and Methods

The experimental data, codes, and supplementary videos can be found in the Github, but due to obligated anonymisation, only the code for ODE and PDE included in Appendix IV.

1. Data processing

The solution of motile dual reporter (*phag-YFP* and *pTapA-RFP*) cells was dropped onto the biofilm-promoting Msgg agar, and the fluorescence signal was captured every 40 mins, until the biofilm stopped growing at around 90 hrs. The intensity of fluorescence was calculated by summing the intensity of all the points on the circumference with a given radius and averaging the sum over the area of the ring of 26.6- μm width. We normalized data by subtracting the intensity data at a given time point with the intensity of the edge at that time point. The final processed data is shown in the movie S1.

We tracked the expansion of biofilm by looking at both the movement of the YFP and RFP front (Yfront, Rfront) and the maximum signal (Ymax, Rmax). The position of Yfront or Rfront signal was defined as the distance point furthest from the centre where the signal was higher than 10.0% of Ymax or Rmax, respectively, at a given time point (called front threshold), and the position of Ymax or Rmax as the distance point where the signal was higher than YFP or RFP intensity at any other point at a given time point, respectively.

2. Single cell-level SinI-SinR-SlrR deterministic and stochastic model

We assumed that these three proteins, SinI-SinR-SlrR, were sufficient to elicit the two steady states: the motile and the matrix phase at the single cell level. We chose the following interactions with 5 parameters (Table 1.1), in equation 1.1-1.3.

- Free SinI (I) and free SinR (R) is produced at rate α_0 and β_0 , respectively.
- Free SinI (I) and SlrR (L) can irreversibly bind and inhibit free SinR (R) with the rate δ_I and δ_L , respectively.
- Free SinR (R) can repress the transcription of SlrR (L) with the Hill's repression equation $\gamma/(1+R^2)$
- They degrade exponentially.
- Complex SinR-SinI and SinR-SlrR were not modelled in the system.
- SinR level at steady state (R_{ss}) is a proxy for the phase, similar to Lord et al. 2019.
- A high R_{ss} is the motile phase and a low R_{ss} is the matrix phase.

$$(1.1) \quad \dot{I} = \alpha_0 - I - \delta_I IR$$

$$(1.2) \quad \dot{R} = \beta_0 - R - \delta_I IR - \delta_L LR$$

$$(1.3) \quad \dot{L} = \frac{\gamma}{1+R^2} - L - \delta_L LR$$

We were interested in two cases of initial conditions (Table 1.2) based on the number of protein molecules used in Lord et al. 2019, and because these two sets led to different steady states if the system was bistable. We used OrdinaryDiffEq.jl to JumpProcesses.jl to simulate the model with the deterministic and stochastic interpretation with Gillespie's method, respectively. The number of simulations varied depending on the types of analyses (100-500 simulations), each with 20-hr time span, sufficiently long to allow the correct differentiation between a motile phase- or matrix phase-converging simulation (Sup. Fig. 6). We introduced a hypothetical non-monotonic function of how δ_L depends on nutrient as in equation 3 for one of noise effect investigation.

$$(3) \quad \delta_{L,x} = \begin{cases} n_m - v(0.5 - n_x), & n_x < 5 \\ n_m - v(n_x - 0.5), & n_x \geq 5 \end{cases}$$

3. Biofilm-level Nutrient and Bacteria PDE model

We formed two reaction-diffusion equations for modelling the 1D dynamics of the nutrient (n , unitless) and bacteria density (b , unitless) in an agar (equation 2.1-2.2). At $t = 0$, n was 1.00 at all positions (x), while b started with the initial density of founding bacteria of 0.01 (unitless) if x was within the diameter of the founding population, otherwise $b(t=0,x)$ was zero. The nutrient can be consumed by the bacteria and turned into their growth, while the dead bacteria can release the nutrient back to the agar. The required parameters are described in Table 4, linked in the equation 2.1-2.2 below. We used MethodOfLines.jl to simulate the model with deterministic interpretation.

$$(2.1) \quad \frac{\partial b_{t,x}}{\partial t} = D_b \frac{\partial^2 b_{t,x}}{\partial x^2} + k_g \frac{n_{t,x} b_{t,x}}{K + n_{t,x}} - k_d \frac{n_{t,x} b_{t,x}}{K + n_{t,x}}$$

$$(2.2) \quad \frac{\partial n_{t,x}}{\partial t} = D_n \frac{\partial^2 n_{t,x}}{\partial x^2} - \omega k_g \frac{n_{t,x} b_{t,x}}{K + n_{t,x}} + \theta \omega k_d \frac{n_{t,x} b_{t,x}}{K + n_{t,x}}$$

4. Parameterisation of IRL-ODE and NuBac-PDE models

For the equivalent parameters, their values were extracted from the previous paper by Miyamoto et al., 2018, Newman et al., 2013., and mainly from Chen et al., 2023 (Table 1 and 2). We tested if these extracted values could produce bistability regions and simultaneously explored a wider range of parameter values. We tested parameter values by solving for the number of positive solutions, which was the number of valid steady states, of the polynomials for the fixed points (Appendix II) and by bifurcation diagram plotting for the scanning of one parameter (deBif R package)

To estimate an effective diffusion coefficient of bacteria (D_b) in NuBac model, we derived it from its relationship to the speed of the experimental biofilm front (v_b) and the net

growth rate of bacteria (equation 4.1). We then varied the value of D_b over the same-order of this calculated D_b from equation 4.2. We picked the value of D_b that led to the lowest sum of square of the difference between the experimental front and the front of bacteria produce by the NuBac-PDE, with the front threshold of 10.0% (equation 6). All other parameter values were presented in Table 4.

Derived from NuBac model,

$$(4.1) \quad v_b = \sqrt{2 \cdot \frac{n}{K+n} (k_g - k_d) D_b} = \sqrt{2 \cdot 0.271 (k_g - k_d) D_b}$$

since mean of $(n/(K+n)) = 0.271$ when $n = [0,1]$ and $K=1.21$, and therefore,

$$(4.2) \quad D_b = \frac{v_b^2}{2 \cdot 0.271 (k_g - k_d)}$$

$$(6) \quad lossfront = \sum_{t=0}^{t_{141}} (bfront_t - modelfront_t)^2$$

Following this, we used the obtained D_b to search for the fluorescence coefficients that relates the YFP and RFP signal intensity to the motile and matrix bacteria density (n_r and n_y), using the squared sum of squared deviation between the scaled sum ($= n_y * YFP + n_r * RFP$) of the signals and the model prediction of bacteria density ($b(t,x)$) (equation 7).

$$(7) \quad lossall = \sum_{x=0}^{x_{210}} \sum_{t=0}^{t_{141}} (b_{t,x} - (n_y * YFP_{t,x} + n_r * RFP_{t,x}))^2$$

5. The link dILf(n) function parameterisation and multi-scaling model simulation

After IRL-ODE and NuBac-PDE parameterisation, we aimed to use the IRL-ODE and NuBac-PDE models together, with the pipeline represented in Fig 2A to predict the difference in the outcome from the deterministic and stochastic system. We needed to find the form of the link function dILf(n) that varies the complexing constant based on the nutrient level as evidence showed the SinR activity, not the level, was affected by the nutrient-deprivation signal, c-diAMP. We optimised the parameters in the link function dILf(n) (equation 5.1-5.2), using the least sum of squared deviation between the signals and the model prediction of the bacteria density in motile or matrix subpopulation (equation 8), following the pipeline in Fig 2B.

$$(5.1) \quad \delta_L(n_{t,x}) = cf_1 + cf_2 n_{t,x} + cf_3 n_{t,x}^2 + cf_4 n_{t,x}^3 + cf_5 n_{t,x}^4$$

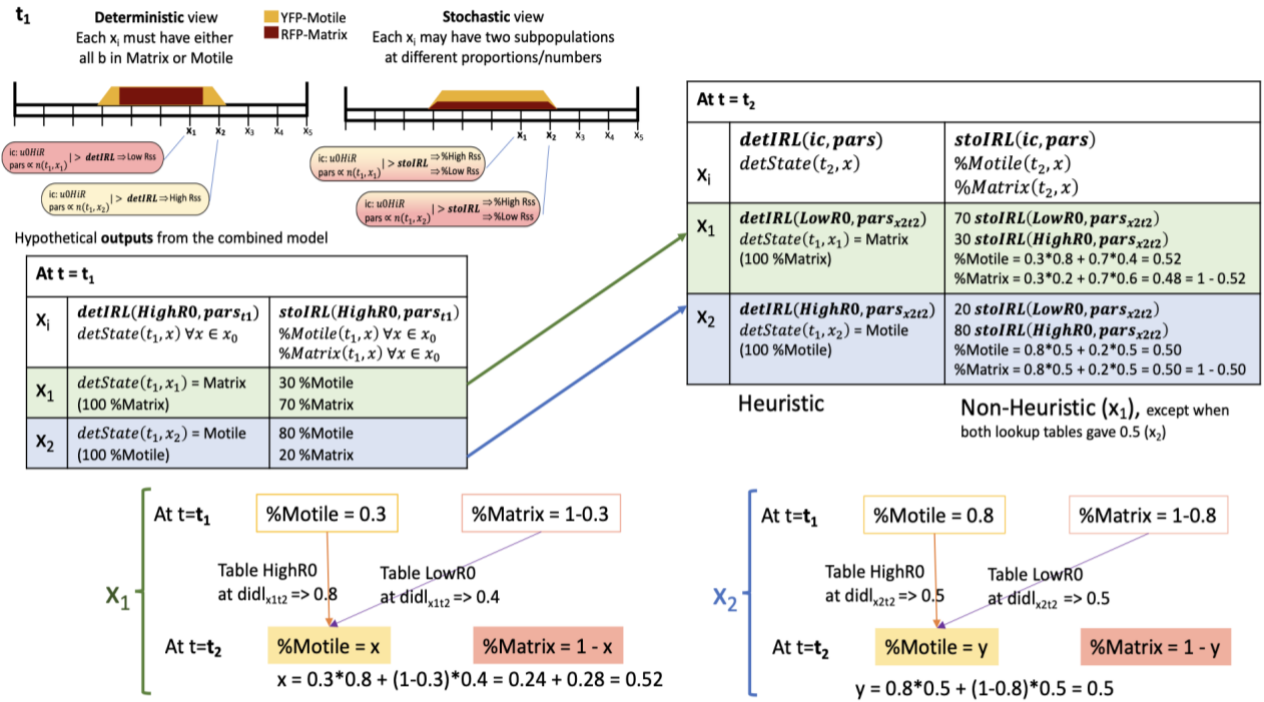
$$(5.2) \quad \delta_I(n_{t,x}) = cf_6 \delta_L(n_{t,x})$$

$$(8) \quad g = \sum_{x=0}^{x_{210}} \sum_{t=0}^{t_{141}} (bMotile_{t,x} - n_y * YFP_{t,x})^2 + (bMatrix_{t,x} - n_r * RFP_{t,x})^2$$

For initial values of the coefficients in equation 5.1-5.2, we started with a function with a similar shape to equation 3 and with ten random initial values between -20 to 20. All candidates of $dILf(n)$ were then used to produce the deterministic cell phase (Detstate), and the density of motile and matrix bacteria from stochastic simulation (bMotile and bMatrix) to investigate the effect of noises, the nutrients, and their potential interaction.

Note that, instead of running stochastic IRL-ODE at every time point over every distance point, which slowed down the optimisation, we created a table storing the probability of the cells being in a motile phase (lookup tables, Sup. Table 1) by counting the number of runs, out of 100, with a high Rss a (> 8 molecules). We used two types of initial conditions (Table 1.2), and its use in optimisation algorithm was shown in Fig. 13.

Figure 13: The outputs and the calculation of %Motile and %Matrix from the connected model were exemplified at two time points t_1 (left table) and t_2 (right table) for two distance points, x_1 (green) and x_2 (blue). Table HighR0 and Table LowR0, produced by using a high R0 or a low R0 initial conditions in 100 stochastic IRL-ODE runs, respectively.



Acknowledgements

I would like to thank my supervisor Diana Fusco, Department of Physics, for her advice on constructing a research framework and for resources to the biology of biofilm of *B. subtilis*, mathematical theory, and HPC access. Secondly, I would like to thank Nikhil Krishnan, a PhD in Diana's lab for collecting biofilm image data and converting them into numerical florescent data and for his advice on using HPC. I would also like to thank people in Diana's lab for any questions and discussion related to my work during the general and topic-specific lab meetings.

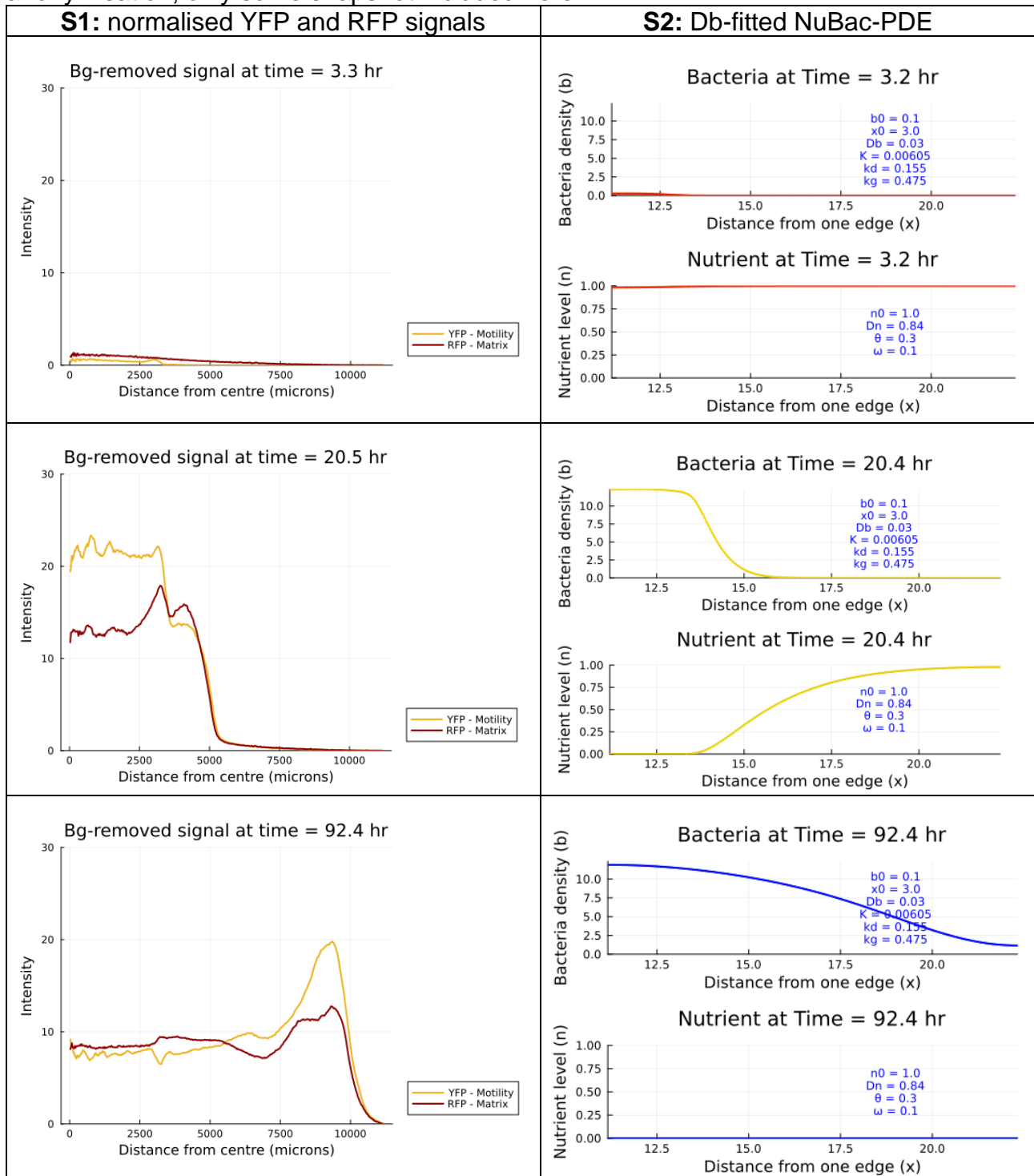
References

- Chai, Y., Kolter, R., & Losick, R. (2010). Reversal of an epigenetic switch governing cell chaining in *Bacillus subtilis* by protein instability: Reversal of an epigenetic switch. *Molecular Microbiology*, no-no. <https://doi.org/10.1111/j.1365-2958.2010.07335.x>
- Chai, Y., Norman, T., Kolter, R., & Losick, R. (2010). An epigenetic switch governing daughter cell separation in *Bacillus subtilis*. *Genes & Development*, 24(8), 754–765. <https://doi.org/10.1101/gad.1915010>
- Chai, Y., Norman, T., Kolter, R., & Losick, R. (2011). Evidence that metabolism and chromosome copy number control mutually exclusive cell fates in *Bacillus subtilis*: Metabolism and chromosome copy number control cell fates in *B. subtilis*. *The EMBO Journal*, 30(7), 1402–1413. <https://doi.org/10.1038/emboj.2011.36>
- Chen, Z., Srivastava, P., Zarazúa-Osorio, B., Marathe, A., Fujita, M., & Igoshin, O. A. (2022). *Bacillus subtilis* Histidine Kinase KinC Activates Biofilm Formation by Controlling Heterogeneity of Single-Cell Responses. *MBio*, 13(1), e01694-21. <https://doi.org/10.1128/mbio.01694-21>
- Donlan, R. M. (2002). Biofilms: Microbial Life on Surfaces. *Emerging Infectious Diseases*, 8(9), 881–890. <https://doi.org/10.3201/eid0809.020063>
- Dragoš, A., Kieseewalter, H., Martin, M., Hsu, C.-Y., Hartmann, R., Wechsler, T., Eriksen, C., Brix, S., Drescher, K., Stanley-Wall, N., Kümmerli, R., & Kovács, Á. T. (2018). Division of Labor during Biofilm Matrix Production. *Current Biology*, 28(12), 1903-1913.e5. <https://doi.org/10.1016/j.cub.2018.04.046>
- Elf, J., & Ehrenberg, M. (2003). Fast Evaluation of Fluctuations in Biochemical Networks With the Linear Noise Approximation. *Genome Research*, 13(11), 2475–2484. <https://doi.org/10.1101/gr.1196503>
- Fazeli-Nasab, B., Sayyed, R. Z., Mojahed, L. S., Rahmani, A. F., Ghafari, M., Antonius, S., & Sukamto. (2022). Biofilm production: A strategic mechanism for survival of microbes under stress conditions. *Biocatalysis and Agricultural Biotechnology*, 42, 102337. <https://doi.org/10.1016/j.bcab.2022.102337>
- Fei, C., Mao, S., Yan, J., Alert, R., Stone, H. A., Bassler, B. L., Wingreen, N. S., & Košmrlj, A. (2020). Nonuniform growth and surface friction determine bacterial biofilm morphology on soft substrates. *Proceedings of the National Academy of Sciences*, 117(14), 7622–7632. <https://doi.org/10.1073/pnas.1919607117>
- Fujita, M., González-Pastor, J. E., & Losick, R. (2005). High- and Low-Threshold Genes in the Spo0A Regulon of *Bacillus subtilis*. *Journal of Bacteriology*, 187(4), 1357–1368. <https://doi.org/10.1128/JB.187.4.1357-1368.2005>
- Fujita, M., & Losick, R. (2005). Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes & Development*, 19(18), 2236–2244. <https://doi.org/10.1101/gad.1335705>
- López, D., Gontang, E. A., & Kolter, R. (2010). Potassium Sensing Histidine Kinase in *Bacillus subtilis*. In *Methods in Enzymology* (Vol. 471, pp. 229–251). Elsevier. [https://doi.org/10.1016/S0076-6879\(10\)71013-2](https://doi.org/10.1016/S0076-6879(10)71013-2)
- Lord, N. D., Norman, T. M., Yuan, R., Bakshi, S., Losick, R., & Paulsson, J. (2019). Stochastic antagonism between two proteins governs a bacterial cell fate switch. *Science*, 366(6461), 116–120. <https://doi.org/10.1126/science.aaw4506>

- Milton, M. E., Draughn, G. L., Bobay, B. G., Stowe, S. D., Olson, A. L., Feldmann, E. A., Thompson, R. J., Myers, K. H., Santoro, M. T., Kearns, D. B., & Cavanagh, J. (2020). The Solution Structures and Interaction of SinR and SinI: Elucidating the Mechanism of Action of the Master Regulator Switch for Biofilm Formation in *Bacillus subtilis*. *Journal of Molecular Biology*, 432(2), 343–357. <https://doi.org/10.1016/j.jmb.2019.08.019>
- Newman, J. A., Rodrigues, C., & Lewis, R. J. (2013). Molecular Basis of the Activity of SinR Protein, the Master Regulator of Biofilm Formation in *Bacillus subtilis*. *Journal of Biological Chemistry*, 288(15), 10766–10778. <https://doi.org/10.1074/jbc.M113.455592>
- Nye, T. M., Schroeder, J. W., Kearns, D. B., & Simmons, L. A. (2017). Complete Genome Sequence of Undomesticated *Bacillus subtilis* Strain NCIB 3610. *Genome Announcements*, 5(20), e00364-17. <https://doi.org/10.1128/genomeA.00364-17>
- Sharma, D., Misba, L., & Khan, A. U. (2019). Antibiotics versus biofilm: An emerging battleground in microbial communities. *Antimicrobial Resistance & Infection Control*, 8(1), 76. <https://doi.org/10.1186/s13756-019-0533-3>
- Su, Y., Liu, C., Fang, H., & Zhang, D. (2020). *Bacillus subtilis*: A universal cell factory for industry, agriculture, biomaterials and medicine. *Microbial Cell Factories*, 19(1), 173. <https://doi.org/10.1186/s12934-020-01436-8>
- Tasaki, S., Nakayama, M., & Shoji, W. (2017). Morphologies of *Bacillus subtilis* communities responding to environmental variation. *Development, Growth & Differentiation*, 59(5), 369–378. <https://doi.org/10.1111/dgd.12383>
- Vlamakis, H., Aguilar, C., Losick, R., & Kolter, R. (2008). Control of cell fate by the formation of an architecturally complex bacterial community. *Genes & Development*, 22(7), 945–953. <https://doi.org/10.1101/gad.1645008>

Appendix I: Supplementary Media

Movie S1: The intensity of normalised YFP and RFP signals over 92 hours of growth from the centre of the biofilm and **Movie S2:** The nutrient and bacteria density level produced by the Db-fitted NuBac-PDE ($Db = 0.03$) can be found on GitHub, but due to anonymisation, only some snapshot included here

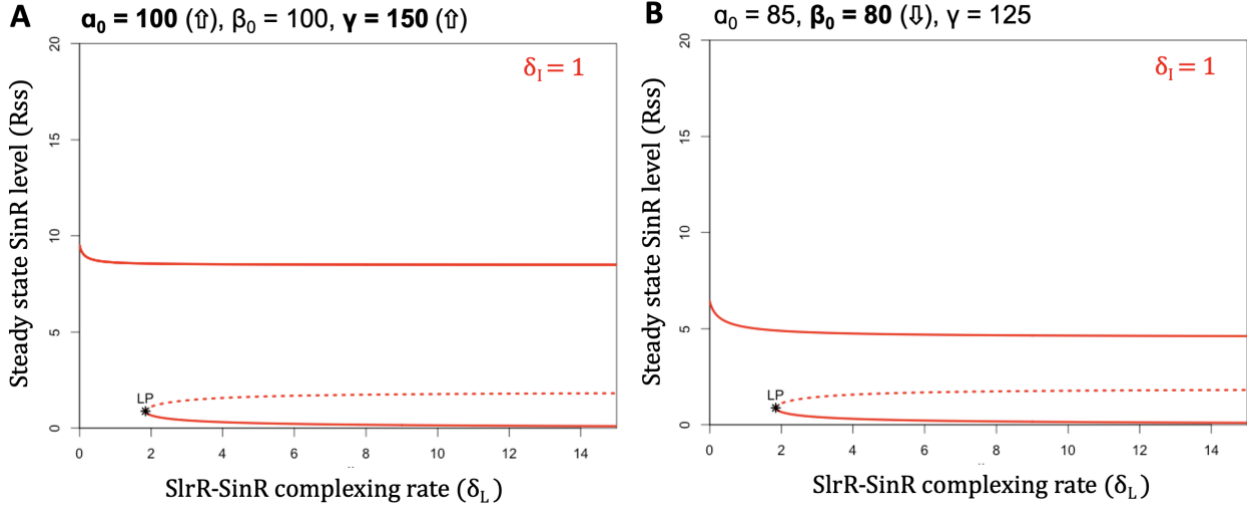


Supplementary Figure 1: Additional δ_L bifurcation diagrams to compare with the main δ_L bifurcation diagram (Fig. 3.2 in the main text). The three protein production parameters were as stated in the figures. LP: limit points (saddle-node bifurcation points). Dashed line: unstable fixed points. Solid line: stable fixed points

Supplementary Figure 1.1: How changes in the protein production rates affect the bistability.

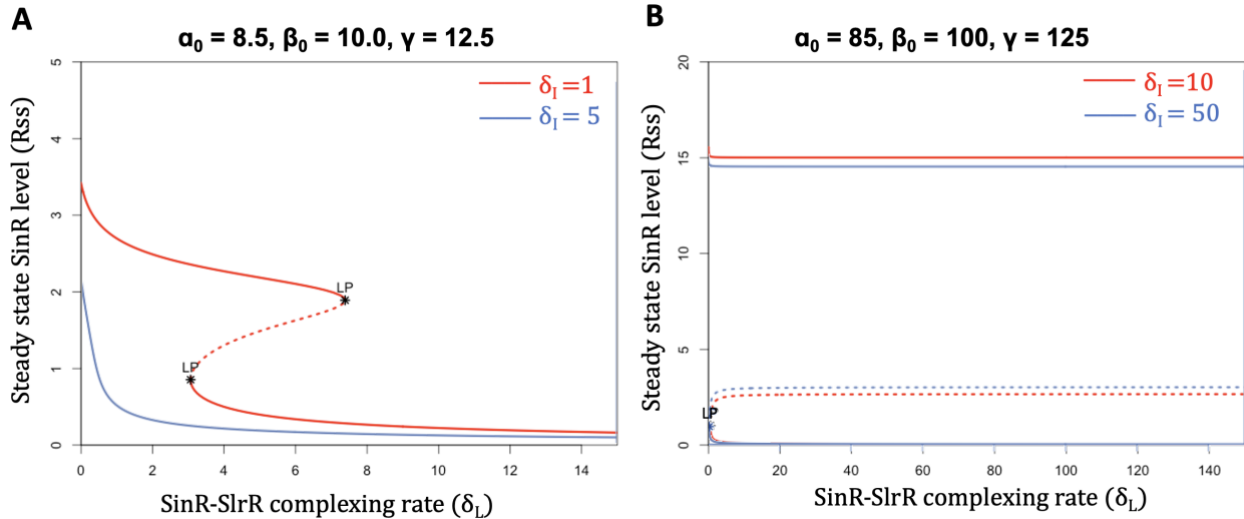
The protein production parameters in **bold** were increased (\uparrow) or decreased (\downarrow) to make a comparison to the default set extracted from Chen et al, 2023 (Table 1.1).

Increasing SinI or SlrR production rates (α_0 or γ) as in **A** or decreasing SinR production rates (β_0) as in **B** showed a larger region of bistable region, extending to a lower δ_L compared to the red line in Fig. 3.2, formed using the production parameters in Table 1.1 with $\delta_I = 1$.

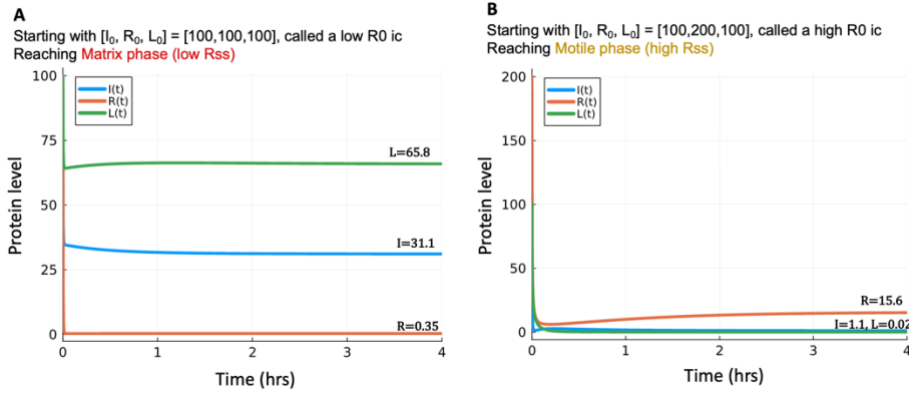


Supplementary Figure 1.2: How the system can possibly be locked in a matrix phase.

Compared to Fig. 3.2, formed using the production parameters in Table 1.1, simultaneously decreasing SinI, SinR, and SlrR production rates (α_0 , β_0 , γ) as in **A** can result in the global matrix state (low Rss) monostable region when $\delta_I = 1$ and $\delta_L > \sim 8$. However, similar to Fig. 3.2, increasing the range of the complexing rates while keeping the production rates constant, as in **B**, did not produce any global matrix state monostable region.

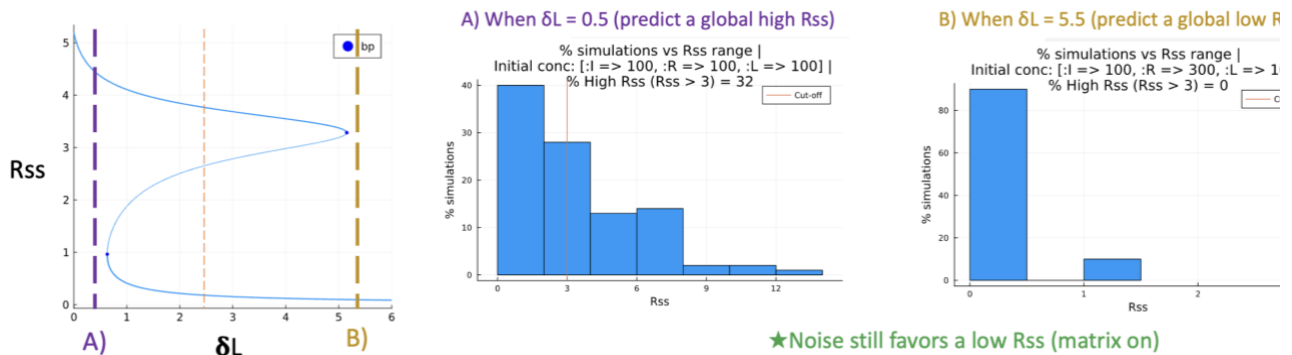


Supplementary Figure 2.1: The time-evolution curves of IRL-ODE system, starting with two

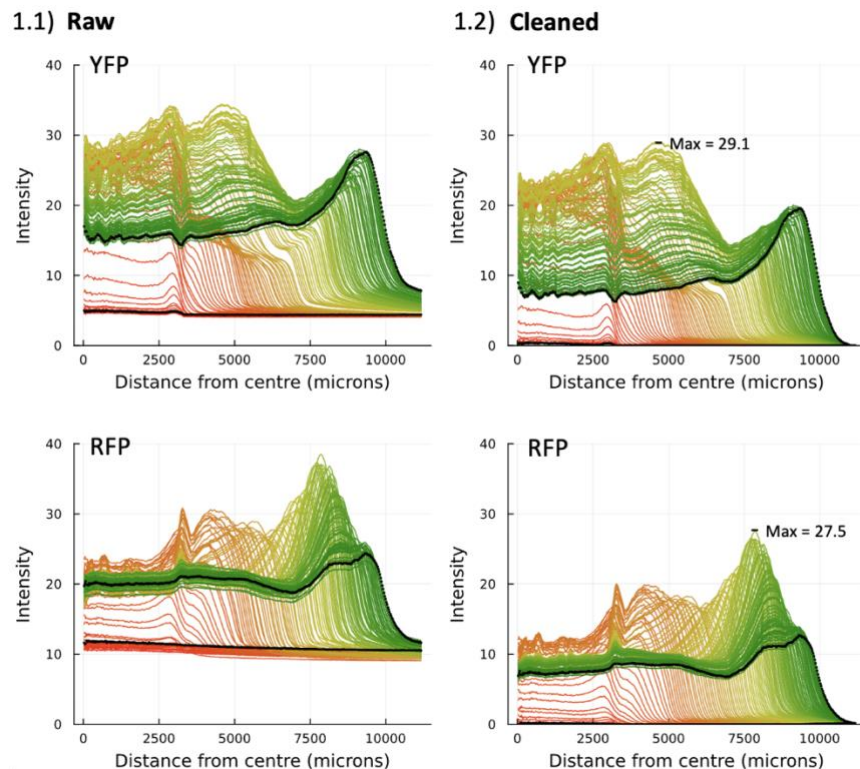


different initial conditions that led to either a matrix phase steady state (A) or a motile phase steady state (B). The numbers were in the order of 100 molecules akin to Lord et al, 2019. The system quickly reached the steady state within 3-4 hrs.

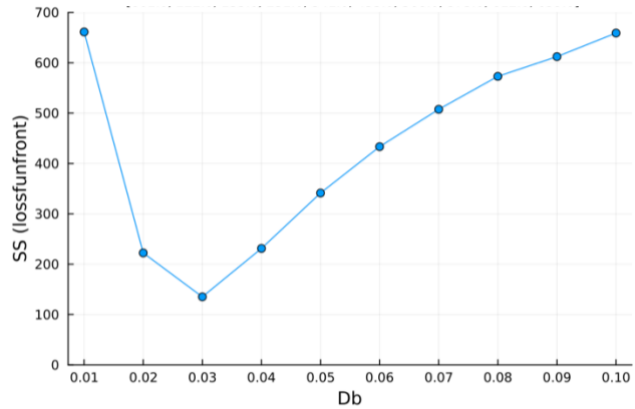
Supplementary Figure 2.2: The single-cell analysis of the noise effect on the steady state SinR distribution when the system was in either monostable motile (A) or monostable matrix (B)



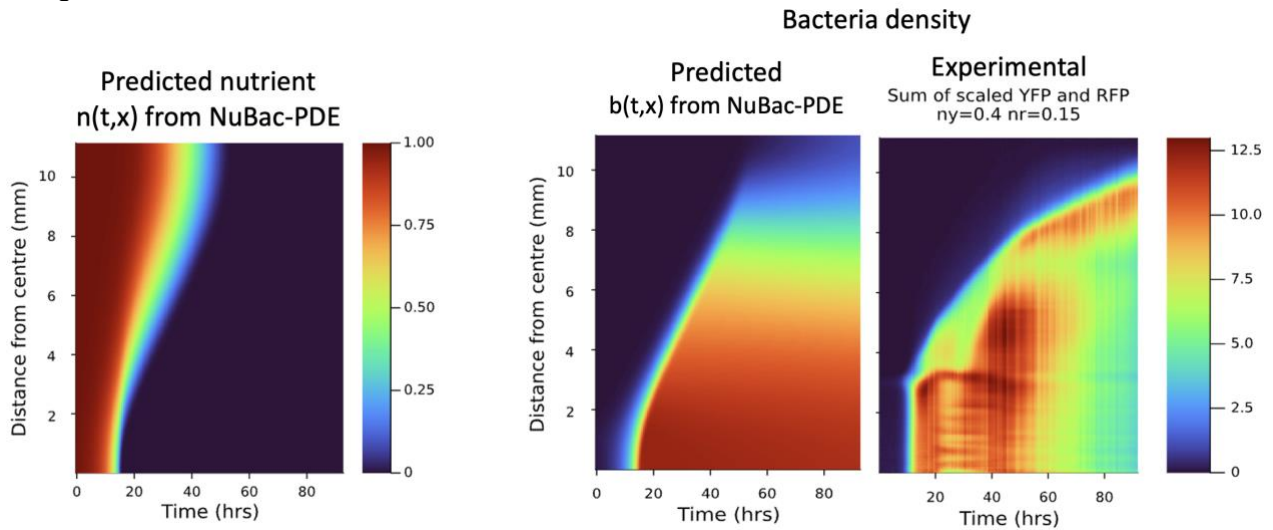
Supplementary Figure 3: The comparison of (1.1) raw and (1.2) cleaned YFP and RFP intensity data. The red-to-green colour transition represents the first to the last time point of the data. Highlighted in black is the trace at the first and the last timepoint. Top graph: YFP-Motile cells. Bottom graph: RFP-Matrix cells.



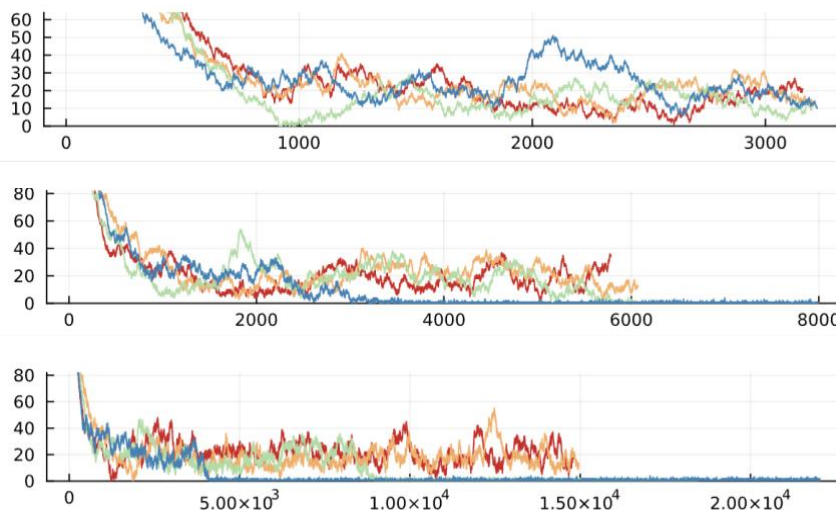
Supplementary Figure 4: The sum of squared deviations between the experimental biofilm front (bfront) and the model front (equation 7). $D_b = 0.03 \text{ mm}^2.\text{h}^{-1}$ gave the smallest deviation .



Supplementary Figure 5: The heatmap of predicted nutrient ($n(t,x)$) and bacterial density ($b(t,x)$) from the final growth dynamic equations (NuBac-PDE). Comparing $b(t,x)$ to the scaled sum of YFP and RFP data, the model overestimated the bacterial density at around 4-7mm from the centre during 20th-30th hours and at almost all distance at 60th hr onwards.



Supplementary Figure 6: Four traces of stochastic simulations for different timespans. The orange and red lines represent the run with SinR at the last time point > 8 and the green and blue lines represent the run with SinR at the last time point ≤ 8 .



A) timespan = 10 hrs
(Not yet reached a steady state)

B) timespan = 20 hrs
(Sufficient to correctly distinguish two cases)

C) timespan = 50 hrs

Supplementary Table 1

The lookup tables for the percentage of the system having a motile phase (%Motile in Fig. 13) as a function of the two complexing constants, created by 100 stoIRL runs, with timespan = 20 hrs.

(A) start with a high R_0 ic [100,200,100]

di\dl	0	1	2	3	4	5	6	7	8	9	10
0	100	100	100	100	100	100	100	100	100	100	100
1	91	96	92	97	90	86	87	84	83	80	80
2	89	79	45	30	32	26	25	22	22	16	10
3	80	43	16	13	9	7	3	5	3	7	7
4	78	26	7	5	2	1	0	3	2	0	2
5	83	15	1	1	1	1	0	0	2	0	1
6	79	4	3	0	2	0	0	0	1	0	0
7	80	4	1	0	0	1	0	0	0	1	1
8	83	0	0	0	0	0	0	0	0	0	1
9	75	0	0	0	0	0	0	0	0	0	0
10	74	2	0	0	0	0	0	0	0	0	0

(B) start with a low R_0 ic [100,100,100]

di\dl	0	1	2	3	4	5	6	7	8	9	10
0	100	100	100	100	100	100	100	100	100	100	100
1	92	94	92	92	86	62	65	41	28	32	19
2	84	76	23	2	0	0	0	0	0	0	0
3	85	42	0	1	0	0	0	0	0	0	0
4	86	18	1	0	0	0	0	0	0	0	0
5	82	4	0	0	0	0	0	0	0	0	0
6	89	4	1	0	0	0	0	0	0	0	0
7	73	3	0	0	0	0	0	0	0	0	0
8	83	1	0	0	0	0	0	0	0	0	0
9	88	1	0	0	0	0	0	0	0	0	0
10	77	1	0	0	0	0	0	0	0	0	0

Appendix II: The derivation of analytical solution of the steady state level of SinR.

The ODE system of SinI-SinR-SIrR (IRL) network

$$\dot{I} = \alpha - I - \delta_I IR \quad (s1)$$

$$\dot{R} = \beta - R - \delta_I IR - \delta_L LR \quad (s2)$$

$$\dot{L} = \frac{\gamma}{(1 + R^2)} - L - \delta_L LR \quad (s3)$$

At the steady state, all the rates are zero;

$$\dot{I} = 0 \Rightarrow I = \frac{\alpha}{1 + \delta_I R} \quad (s4)$$

$$\dot{L} = 0 \Rightarrow L = \frac{\gamma}{(1 + \delta_L R)(1 + R^2)} \quad (s5)$$

$$\dot{R} = 0 = \beta - R - \delta_I IR - \delta_L LR \quad (s6)$$

Substitute (s4) and (s5) in (s6); (Note that I,R,L thereafter are I,R,L at SS level)

$$\beta - R - \delta_I \left(\frac{\alpha}{1 + \delta_I R} \right) R - \delta_L \left(\frac{\gamma}{(1 + \delta_L R)(1 + R^2)} \right) R = 0$$

$$(\beta - R)(1 + R^2)(1 + \delta_I R)(1 + \delta_L R) - \delta_I \alpha R(1 + \delta_L R)(1 + R^2) - \delta_L \gamma R(1 + \delta_I R) = 0$$

$$(\beta + \beta R^2 - R - R^3)(1 + \delta_I R + \delta_L R + \delta_I \delta_L R^2) - \delta_I \alpha R(1 + \delta_L R + R^2 + \delta_L R^3) - \delta_L \gamma R + \delta_I \delta_L \gamma R^2 = 0$$

Summary of all the terms in 5th-order polynomial equation for Rss (s7)

R^0	R^1	R^2	R^3	R^4	R^5
β	$\delta_I \beta$	$\delta_I \delta_L \beta$	$\delta_I \beta$	$\delta_I \delta_L \beta$	$-\delta_I \delta_L$
	$\delta_L \beta$	β	$\delta_L \beta$	$-\delta_I$	
	-1	$-\delta_I$	$-\delta_I \delta_L$	$-\delta_L$	
	$-\delta_I \alpha$	$-\delta_L$	-1	$-\delta_I \delta_L \alpha$	
	$-\delta_L \gamma$	$-\delta_I \delta_L \alpha$	$-\delta_I \alpha$		
		$-\delta_I \delta_L \gamma$			

Appendix III: Parameterisation of the effective diffusion coefficient of bacteria and fluorescent coefficients.

The colony of *B. subtilis* engineered with motile phase and matrix phase fluorescence reporters (*phag*-YFP and *pTapA*-RFP, respectively), grown over 90 hours to the diameter of around 22 mm, was imaged every 40 minutes to track intensity of both signals. Using only the linear front movement between 13.2 and 52.8 hrs, the speed of bfront was around $112 \mu\text{m} \cdot \text{h}^{-1}$, which corresponded to D_b of $0.01 - 0.10 \text{ mm}^2 \cdot \text{h}^{-1}$ (equation 4.1-4.2 and 2.1-2.2). We used the minimisation of the sum of squared deviation between aforementioned experimental data front (bfront) and the front of predicted bacteria density ($b(t,x)$) from the NuBac-PDE model (equation 6), which turned out to be $0.03 \text{ mm}^2 \cdot \text{h}^{-1}$ (Fig. 9.2). It should be noted that certain parameters in the NuBac PDE model were adjusted (as listed in Table 4) from those presented in the original paper to better recapitulate the movement of YFP and RFP signals in our data (Movie S2).

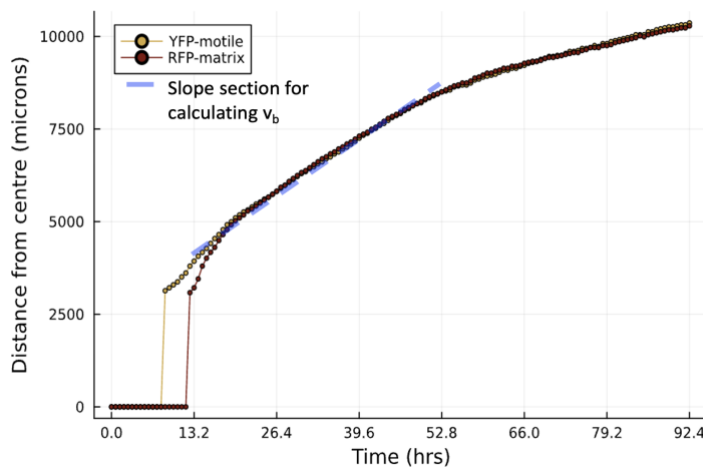


Figure 9.1: The movement of the fronts of motile (YFP) or matrix subpopulations (RFP) with the front position defined as the furthest position from the centre where the signal exceeded 10% of the corresponding maximum signal. The speed of the front movement was calculated using the position of the front from 13.2 to 52.8 hrs.

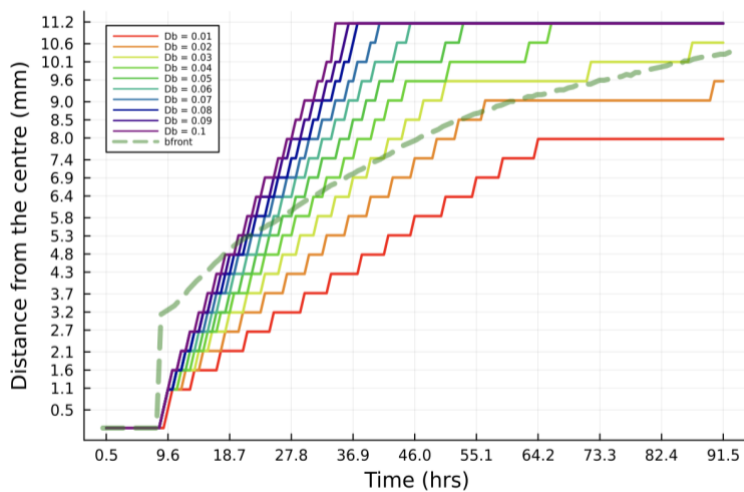


Figure 9.2: The comparison plot between the experimental front position (bfront, the green dashed line) and ten model fronts, generated by the model using bacterial diffusion coefficients (D_b) ranging from 0.01 to 0.10, as in the legend. The fronts were defined as 10% of the maximum signals or model prediction. $D_b = 0.03$ produced the least sum of squared deviation (Sup. Fig. 4)

For the parameterisation of the final multi-scaling model, it was a prerequisite to estimate each experimental bacterial subpopulation from the YFP and RFP signal

intensity. However, the density of the motile and matrix subpopulations cannot be quantitatively inferred from the corresponding signal as we lacked the information of their relationship to the density of source cells (fluorescence coefficients, called n_y and n_r for YFP and RFP intensity, respectively). Therefore, we assumed that the predicted bacterial density from the D_b -parameterised NuBac-PDE model ($b(t,x)$) must be represented by the sum between n_y -scaled YFP and n_r -scaled RFP, called scaled sum ($= n_y \cdot \text{YFP} + n_r \cdot \text{RFP}$). Using the same principle for optimising parameters, we calculated the squared sum of the squared deviation between the two for different combination of n_y and n_r . From the two-way n_y - n_r scanning result (Fig. 10), although the combinations that produced the least sum of squared deviation was $n_r = 0.05$ and $n_y = 0.45$, this pair weighed YFP intensity 9-fold higher than RFP intensity, and hence, likely to result in the model that excessively biased for the existence of motile cells. Thus, we looked at the combinations that produced $<1\%$ more than the least deviation and picked the pair that had the ratio of n_r to n_y closest to 1. We eventually obtained the optimised $n_r = 0.15$ and $n_y = 0.40$.

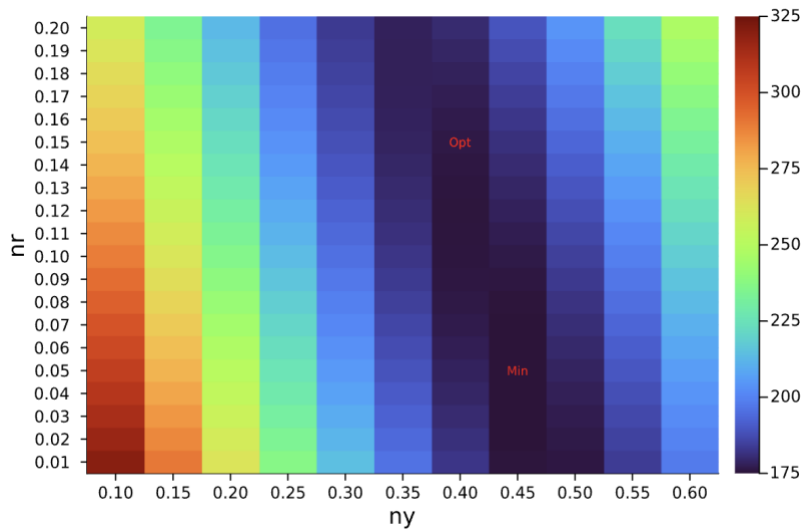


Figure 10: The sum of squared deviation (loss) between the model prediction ($b(t,x)$) and the sum of n_y -scaled YFP and n_r -scaled RFP.

“Min” marked the combination with the smallest deviation.
“Opt” marked the actual chosen pair of n_r and n_y , that produced the deviation within 1% of the smallest loss, but n_r/n_y was closest to 1, in order to minimise bias to one of the signals.

Appendix IV: Julia codes

Example Julia code for SinI-SinR-SIrR model

```
using Catalyst, Latexify, DifferentialEquations, JumpProcesses
using Plots, DataFrames, StatsPlots
using ColorSchemes
using Random, Distributions
using CSV, DelimitedFiles
using LinearAlgebra
using PolynomialRoots, Polynomials
using Setfield, BifurcationKit

#Part 1: Create the IRL model
simplestIRL = @reaction_network begin
     $\alpha_0$ , 0 --> I #constitutive SinI production
     $\beta_0$ , 0 --> R #constitutive SinR production
     $\gamma/(1+(R^2))$ , 0 --> L #SlrR production is a function of SinR level, hill coeff = 2
    1, I --> 0 #exponential dilution/degradation of SinI
    1, R --> 0 #exponential dilution/degradation of SinR
    1, L --> 0 #exponential dilution/degradation of SlrR
     $\delta_I$ , I + R --> 0 #complexing of SinR and SinI (C)
     $\delta_L$ , L + R --> 0 # complexing of SinR and SlrR (K)
end  $\alpha_0$   $\beta_0$   $\gamma$   $\delta_I$   $\delta_L$ 

latexify(convert(ODESystem, simplestIRL))> render
Graph(simplestIRL)

#Part 2: ODE run
#2.1 Deterministic Run and visualisation
u0 = [:I => 100, :R => 200, :L => 100] #initial conditions of proteins
tspan = (0.0, 5) #timespan set to ensure the steady state is reached
p = Dict{: $\alpha_0$  => 85, : $\beta_0$  => 100, : $\gamma$  => 125, : $\delta_I$  => 5, : $\delta_L$  => 2}
myIRLprob = ODEProblem(simplestIRL, u0, tspan, p) #create ODE problem
sol = solve(myIRLprob) #solve sol.u is the I,R,L and sol.t is the time
sslevelODE = round.(sol.u[end], digits=3) #collect the ss level of each protein

#time-evolution plot
TimeEvoDet = plot(sol, xaxis = "Time", yaxis = "Concentration", w = 2, title = "Plot of simplest IRL
system \n  $\alpha_0$ ,  $\beta_0$ ,  $\gamma$ ,  $\delta_I$ ,  $\delta_L$  = $(p[:\alpha_0]), $(p[:\beta_0]), $(p[:\gamma]), $(p[:\delta_I]), $(p[:\delta_L]) \n initial conc. =
```

```

$(u0)\nSS level of I, R, L = $(sslevelODE)",legend=:topleft,titlefontsize =10, size =
(600,500),guidefontsize = 15,xtickfontsize=12, ytickfontsize = 12, lw = 4, legendfontsize = 10)

#2.2 Stochastic Run and visualisation

nsim = 100
u0 = [:I => 100, :R => 100, :L => 100]
tspan = (0.0,10.0)
p = Dict{:α_0=> 85,:β_0=>100, :γ=>125, :δ_I=>5,:δ_L=>2)
Solu = Vector{Vector{Vector{Float64}}}(undef, nsim)
Solt = Vector{Vector{Float64}}(undef,nsim)
for i in 1:nsim
    discretelRLprob = DiscreteProblem(simplestIRL,u0,tspan,p)
    jumpIRLprob = JumpProcesses.JumpProblem(simplestIRL,discretelRLprob,
JumpProcesses.Direct())
    jsol = solve(jumpIRLprob, JumpProcesses.SSAStepper())
    Solu[i] = jsol.u #sol.u is Vector{Vector{Float64}}
    Solt[i] = jsol.t #sol.t is Vector{Float64}
    println("Fin- $(i)")
end

#find statistics
arrayrun = Array{Matrix{Float64}}(undef, nsim, 1)
lastpoint = Vector{Vector{Float64}}(undef,nsim)
avgtenlast = Matrix{Float64}(undef,nsim,3)
avg100last = similar(avgtenlast)
for i in 1:nsim
    arrayrun[i] = mapreduce(permutedims, vcat, Solu[i]) #take Solu_o[i] -> transpose each -> vcat them
    lastpoint[i] = arrayrun[i][end,:]
    avgtenlast[i,:] = mean(arrayrun[i][(end-9):end,:], dims = 1)
    avg100last[i,:] = mean(arrayrun[i][(end-99):end,:], dims = 1)
end
dflastpoint = mapreduce(permutedims, vcat, lastpoint)
ssLsd = round(std((dflastpoint[:,1])), digits = 3)
ssRsd = round(std((dflastpoint[:,2])), digits = 3)
ssLsd = round(std((dflastpoint[:,3])), digits = 3)
#plot R at steady state (RSS) of each simulation and count the number of runs that reach high RSS
co = 8

```

```

count_ZeroR = sum(x -> x == 0, dflastpoint[:,2])
count_highR = sum(x -> x >= co, dflastpoint[:,2])
count_lowR = sum(x -> x < co, dflastpoint[:,2])

histogram(dflastpoint[:,2], title = "% simulations vs Rss range |\n Initial conc: $(u0) |\n % High Rss (Rss
> $(co)) = $(count_highR) ", xlabel = "Rss", ylabel = "% simulations", label = false, titlefontsize = 10)

vline!([co], label = "Cut-off")

```

Example Julia code for growth dynamic NuBac-PDE model

```

using ModelingToolkit
using MethodOfLines, DomainSets
using Symbolics
using CSV

#Part 1: the NuBacPDE model
#1.1 Construct state variables and parameters

@parameters t x #independent variables
@variables b(..) n(..) #dependent variables
#define derivatives
Dt = Differential(t)
Dx = Differential(x)
Dxx = Differential(x)^2
@register founder(x, x0,xmax,bstart)
#Space and time domains
xmax = 11.160121*2 #diameter of biofilm, based on data we collect
x0 = 3. # diameter of founding populations, based on data we collect
tmax = 92.6 #growing tome of biofilm, based on data we collect
domains = [t ∈ Interval(0.0,tmax),x ∈ Interval(0.0,xmax)]
#pcf is for the modification of the parameters
pcf_1 = Dict(:bstart => 1, :nstart => 1, :D_b=> 1, :D_n => 1, :kg => 1, :kd =>1, :K => 1,:D_b_fin => 0.2)
pcf_2 = Dict(:bstart => 1, :nstart => 1, :D_b=> 1, :D_n => 1, :kg => 4, :kd =>10, :K => 0.8,:D_b_fin =>
0.2)
pcf_3 = Dict(:bstart => 400/4, :nstart => 4/4, :D_b=> 1/4, :D_n => 1/4, :kg => 1/4, :kd =>1/4, :K =>
0.02/4,:D_b_fin => 0.03)
#final pcf
pcf = pcf_3

```

```

bstart = 0.001 * pcf[:bstart]  #(taken from paper), unitless
nstart = 1. * pcf[:nstart]  #normalised to 1 (taken from paper), unitless
D_b = pcf[:D_b_fin]  #0.075 * pcf[:D_b] #we will vary this to fit with the data
 #exactly from the data; D_b = 0.001787256032 #Diffusion coeff of bacteria ; unit = mm2 h-1
D_n = 3.36 * pcf[:D_n]  #Diffusion coeff of nutrient (glycerol conc 0.1-0.5%); unit = mm2 h-1
kg = 1.90 * pcf[:kg]  #maximum growth rate of bact; unit = h-1
kd = 0.62 * pcf[:kd]  #maximum death rate of bact; unit = h-1
K = 1.21 * pcf[:K]  # Half-saturation nutrient for bacterial growth or death; no unit
θ = 0.3  #Fraction of nutrient released from dead bacteria; no unit
ω = 0.1  #Yield coefficient of bacteria (how much consumed nu become growth); no unit
 #1D PDE and #we discretize space, leave the time undiscretized using MethodOfLines
eqb = Dt(b(t,x)) ~ D_b * Dxx(b(t,x)) + kg * b(t,x)*(n(t,x)/(n(t,x)+K)) - kd * b(t,x)*(n(t,x)/(n(t,x)+K))
eqn = Dt(n(t,x)) ~ D_n * Dxx(n(t,x)) - ω*kg * b(t,x)*(n(t,x)/(n(t,x)+K)) + θ*ω*kd * b(t,x)*(n(t,x)/(n(t,x)+K))
eqs = [eqb,eqn]

ics = [b(0,x) ~ founder(x,x0,xmax,bstart), n(0,x)~ nstart]  #Initial conditions
bcs = [Dx(n(t,xmax))~ 0., Dx(b(t,xmax))~ 0., Dx(n(t,0))~ 0., Dx(b(t,0))~ 0.]  #and boundary conditions
ic_bcs = vcat(ics, bcs)

@named pde_system = PDESystem(eqs,ic_bcs,domains,[t,x],[b(t,x),n(t,x)])  #build PDE system
latexify(pde_system)|> render

dx = 0.05314343333  #define stepsize of x - what if it is 11.160121*2 /420 = 0.05314343333
discretization = MOLFiniteDifference([x => dx], t, approx_order = 2)  #t is continuous
prob = discretize(pde_system,discretization)  #form discretised PDE problem

 #1.2 call solver
 #here we can use a normal ODE solver to solve our discretized PDE
sol = solve(prob, Tsit5(), saveat = 0.05)
#writedlm("directsol.txt",sol)
#sol_coarse = solve(prob, Tsit5(), saveat = 0.5)

#visualise
discrete_x = sol[x]
discrete_t = sol[t]
soln = sol[n(t,x)]
solb = sol[b(t,x)]
after_mid = ceil(Int,size(solb,2)/2)+1
halfb = solb[11:13:1843,after_mid:(end-1)]'
halfn = soln[11:13:1843,after_mid:(end-1)]'

```