

# Bacterial Quorum Sensing and Pattern Formation Modelling and Computation

Cambridge iGEM 2008

October 28, 2008

## Introduction

We are interested in using peptide-based quorum-sensing systems from gram-positive bacteria to create self-organising biological systems. We would like to demonstrate how to engineer multicellular biological systems that are capable of expressing spatial patterns on a bacterial lawn. This means that we would like to see regular patterns of different levels of gene expression (e.g. of GFP). This would be a first step in the direction of engineering multicellular behaviour that will become immensely useful when thinking about further applications such as self-assembling systems responding to electric input and output (e.g. if one uses a gene that increases cellular electric conductivity).

In the following article, we would like to introduce a model of the agr-QS of *S.aureus* to illustrate how a typical QS system works. Furthermore, the model will predict theoretical values for biological parameters (such as the cell density) that will be verified and we will be also be able to predict how the system will behave if we change a number crucial parameters. This can be extremely useful in informing design decision when building a synthetic device. Finally, we would like to use the model in a hypothetical setup (with a second parallel system) in order to investigate how one might be able to succeed at engineering at biological patterning systems.

## Modelling Bacterial Quorum Sensing Systems - Parameters

Modelling biological systems usually involves finding relations between a great number of variables and parameters. Usually, these systems cannot be solved in an analytical way as one might accomplish with simpler systems. The problem is compounded by the fact that we only have limited knowledge of the specific parameters involved (production rates, decay rates, etc.). For this article, we shall make a number of assumptions:

- 1) we work with enzyme concentrations  $[A]$  etc. and assume enzymes are well-mixed
- 2) we use standard enzyme kinetics to infer relations between enzyme concentrations
- 3) and we scale rate parameters relative to a fixed value.

To expand on the last point, we note that all parameters we will be working with denote rates and thus have the same dimensions. Therefore, we can work with scalar values only. Secondly, we can scale these parameters w.r.t one of them: We shall fix the maximum rate of promoter expression at 1 and determine the other parameters in relation to this rate.

For example, basal expression rates must be significantly lower than the maximum rate of promoter expression. In fact, in the agr-QS system from *Staphylococcus aureus* that we are going to model the p2 promoter seems to exhibit a significant increase [Novick et al., 1995] upon activation in wild-types. We shall take basal rates of 0.05, implying a 20-fold increase promoter activity translating in even a greater increase in AIP production (which is 50-fold according to Gustafsson<sup>1</sup> who has also modelled part of the agr-system). Furthermore, we assume that *all* enzymes decay at the same rate of 0.2.

In addition, phosphorylation and enzyme complex formation rates must be significantly higher than protein expression rates. We shall take phosphorylation rates at 50 (and dephosphorylation rates at 5) and enzyme complex formation rates at 20 (deformation rates at 2). The important thing to remember is that these rates are much higher than the previous ones and will thus allow us to use quasi-steady state assumptions in our model. An example of this method can be found here<sup>2</sup>. To some degree, the quasi-steady state assumption allows us to ignore actual parameter values. This comes in quite handy when modelling biological systems.

To summarise, the parameter values we shall be using are:

maximum promoter expression rate - 1,

basal expression rate -  $\beta = 0.02$ .

enzyme decay rate -  $\delta = 0.2$ ,

enzyme complex formation and deformation rates -  $c_+ = 20$  and  $c_- = 2$ ,

enzyme phosphorylation and dephosphorylation rates -  $p_+ = 50$  and  $p_- = 5$ .

<sup>1</sup>[Gustafsson, 2004] Characterizing the Dynamics of the Quorum-Sensing System in *Staphylococcus aureus*

<sup>2</sup>[Dockery et al. 2001] A Mathematical Model for Quorum Sensing in *Pseudomonas aeruginosa*

## The agr-Quorum Sensing System of *S.aureus*

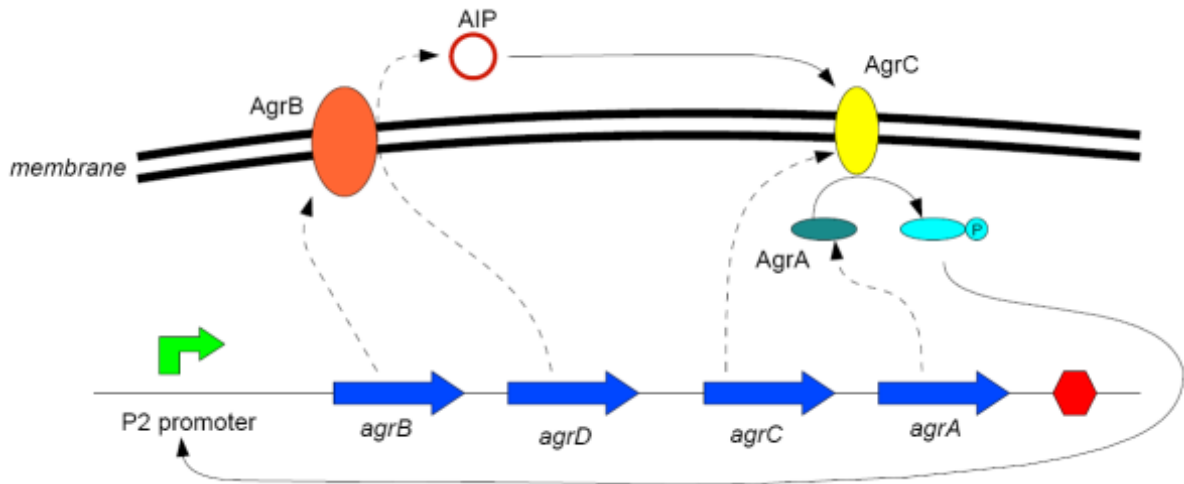


Figure 1: How the agr-QS system works

The agr-QS system is a quorum-sensing system in *S. aureus* and is typical for Gram-positive bacteria. It relies on the signalling peptide AIP (auto-inducing peptide) that is produced by the cell and activates the quorum-sensing system. The QS response is quite tightly regulated as expression occurs only once the local bacterial density passes a fixed threshold value, similar unlike a switch. In a high-density environment the bacteria will exhibit novel behaviour such as virulence or sporulation (both biologically costly but critical and advantageous).

### Synthesis Equation

Let us at first look at protein synthesis. The RNA Polymerase binds to the promoter and transcribes mRNA,  $[m]$ , at a specific rate (assuming Michaelis-Menten kinetics, which we can deduce from simple chemical kinetics theory) depending on the promoter strength and the inducer concentration,  $[A]$ . It is then translated at a rate depending on the strength of the ribosome binding site and folds into enzyme  $[E]$ . mRNA decay rates are much greater than enzyme decay rates so we can assume that  $[m]$  is in a quasi-steady state. The rate equations (after scaling) can be simplified into a single equation (3) as follows:

$$\frac{\partial[m]}{\partial t} = p \frac{[A]}{1 + [A]} + \beta_m - \delta_m [m] \approx 0 \quad (1)$$

$$\frac{\partial[E]}{\partial t} = r_{bs}[m] - \delta[E] \quad (2)$$

$$\Rightarrow \frac{\partial[E]}{\partial t} = \frac{r_{bs}p}{\delta_m} \frac{[A]}{1 + [A]} + \frac{r_{bs}\beta_m}{\delta_m} - \delta[E] \quad (3)$$

and further scaling of  $[E]$  will leave us with the rate equation,

$$\frac{\partial[E]}{\partial t} = p \frac{[A]}{1 + [A]} + \beta_m - \delta[E],$$

where  $p = 1$  following previous assumptions of maximal promoter expression. By assuming quasi-steady state behaviour, we have gotten rid of the mRNA dynamics. We shall use this synthesis equation in the following sections.

## Recasting the agr-receiver part as a biobrick

We would like to have the receiver part of the agr-QS system in form of a biobrick. It becomes useful to think about inputs and outputs to and from such a biobrick. We shall use PoPS as a universal signal to "connect" biobricks to each other. A constitutive promoter biobrick for instance has no inputs but only PoPS output. Our agr-receiver device can be thought of as follows:

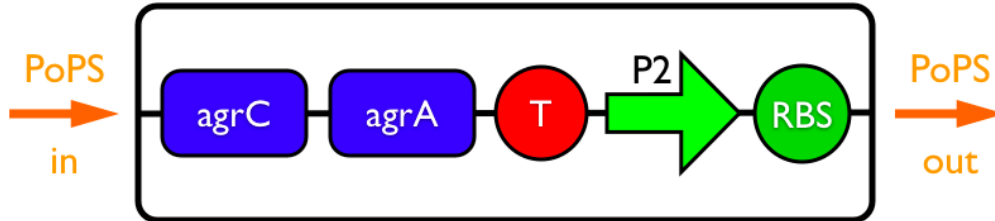


Figure 2: a hypothetical agr-receiver biobrick

In designing the biobrick we use two principles introduced in<sup>3</sup>: Standardisation and characterisation. Here, our biobrick part is modular in the sense that the "interface" between biobricks consists of PoPS in- and output. In essence, this demonstrates a higher-level analogue to the biobrick restriction site standard that is used in assembling the biobrick parts. We will also look at the output, or transfer function, of the biobrick.

## Modelling the agr-quorum sensing response/receiver

To model and test the agr-receiver, we need to put a p2 promoter (and rbs) upstream of the device we just introduced. Let  $[P]$  denote the concentration of AIP,  $[A]$  denote the concentration of AgrA and similarly for the other proteins,  $[CP]$  denote the obvious enzyme complex and  $[A^+]$  be AgrA phosphorylated.

We also need to consider concentrations in extra- and intracellular space carefully. AIP exists (or is only a chemical player) in the extracellular environment whereas cell enzymes (such as AgrA and AgrC) exist only within a cell. Assuming that the local density (i.e. the volume fraction) of cells is  $\rho$  then  $1 - \rho$  will be occupied by AIP with a certain concentration. We assume that concentrations have values relative to extra- or intracellular space i.e. if  $[A] = 0.2$  when  $\rho = 0.1$  then doubling intracellular space to  $\rho = 0.2$  while holding the *total amount* of AgrA constant, will result in  $[A] = 0.1$ .

The equations of this model read:

$$\left\{ \begin{array}{l} \frac{\partial [A]}{\partial t} \\ \frac{\partial [A^+]}{\partial t} \\ \frac{\partial [C]}{\partial t} \\ (1 - \rho) \frac{\partial [P]}{\partial t} \\ \frac{\partial [CP]}{\partial t} \end{array} \right. = \begin{array}{l} \frac{[A^+]}{1 + [A^+]} + \beta - p_+[CP][A] + p_-[A^+] - \delta[A] \\ p_+[CP][A] - p_-[A^+] - \delta[A^+] \\ \frac{[A^+]}{1 + [A^+]} + \beta - c_+[C][P] + c_-[CP][P] - \delta[C] \\ \rho(-c_+[C][P] + c_-[CP] - \delta[P]) \\ c_+[C][P] - c_-[CP] - \delta[CP] \end{array}$$

<sup>3</sup>[Canton+Labno, 2008] Refinement and standardization of synthetic biological parts and devices

Assuming that concentration of  $[CP]$  is in a quasi-steady state we arrive at  $[CP] = \frac{c_+}{c_- + \delta}[C][P]$  and can substitute it back into the system to simplify it into a 4-component system:

$$\left\{ \begin{array}{l} \frac{\partial[A]}{\partial t} \\ \frac{\partial[A^+]}{\partial t} \\ \frac{\partial[C]}{\partial t} \\ (1 - \rho) \frac{\partial[P]}{\partial t} \end{array} \right. = \begin{array}{l} \frac{[A^+]}{1 + [A^+]} + \beta - p_+[CP][A] + p_-[A^+] - \delta[A] \\ p_+[CP][A] - p_-[A^+] - \delta[A^+] \\ \frac{[A^+]}{1 + [A^+]} + \beta - c_+[C][P] + c_-[CP][P] - \delta[C] \\ \rho(-c_+[C][P] + c_-[CP] - \delta[P]) \end{array}$$

Since A and C are surface enzymes, we can also assume that their levels change much slower than the levels AIP or  $A^+$  implying that A and C are in a quasi-steady state. We would then end up with a 2-component system in P and  $A^+$  and could easily determine the dynamics of this system (and show that the systems has two stable steady states (low- and high-density behaviour) and an unstable steady state (the 'switch' density). In the following sections we shall, however, only use minimal assumptions and use the 4-component system in modelling only.

## Characterising the agr-receiver device

Let us now look at how the receiver behaves under a number of conditions. In all cases, we assume that  $[P]$  is held constant (last rate equation would thus fall away).

- We can observe PoPS output at a fixed point in time while varying AIP levels (fig. 3). This is also called a transfer function. Incidentally, PoPS output is identical to PoPS input in this case (both are using the same promoter). The step-like shape of the curve agrees with previously published results<sup>4</sup>.

- We can also look at a surface plot of transfer functions, adding an additional time dimension (fig 4). In this case, we can also clearly see static dynamic performance - PoPS output plotted against time while holding AIP levels constant. This is important in determining the timing of the system output, which can be crucial when using this part in more complex settings.

Throughout the plots in Fig. 3 and Fig. 4 we observe step-like behaviour - this is typical for quorum-sensing systems that react sensitively to local cell densities. Our model seems to work. It is interesting to note that even if we change phosphorylation and complex formation rates to different scalar values (we still require them to be large), the transfer function will only change negligibly. No qualitative change is observed.

## Modelling the agr-sender device

To complete the agr-systems, let us look at the sender that works as follows: AgrB and AgrD lie downstream of a promoter, AgrB is a membrane protein that forms a complex with AgrD and produces AIP. We assume the reaction is  $B + D \rightleftharpoons BD \rightarrow B + P$ . A corresponding biobrick would consist of agrB and agrD with a terminator and would take PoPS input, but would not generate PoPS output since the biobrick would only change AIP levels. Again, we assume quasi-steady state behaviour for BD. The model reads:

<sup>4</sup>[Novick et al. 1995] Cell density control of staphylococcal virulence mediated by an octapeptide pheromone

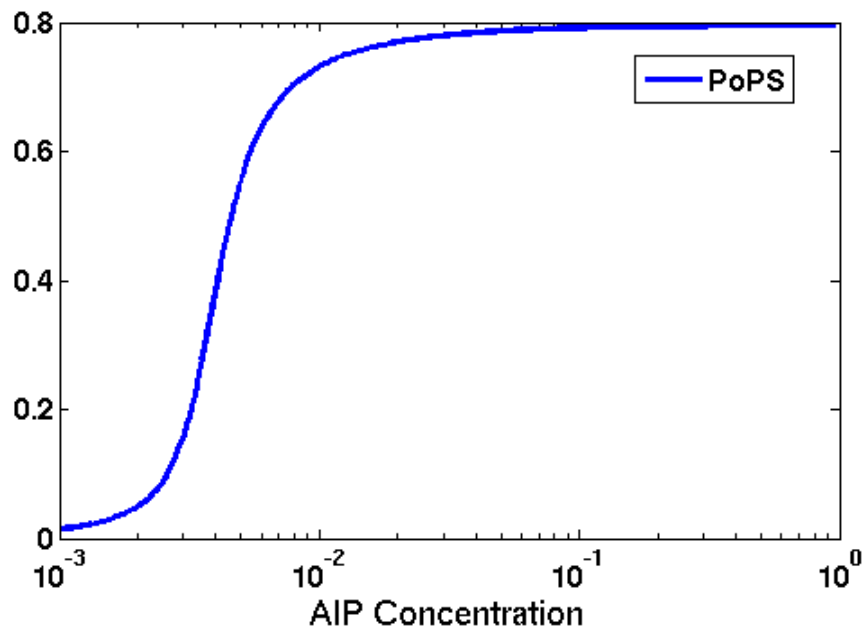


Figure 3: Transfer function plot at  $t=20$

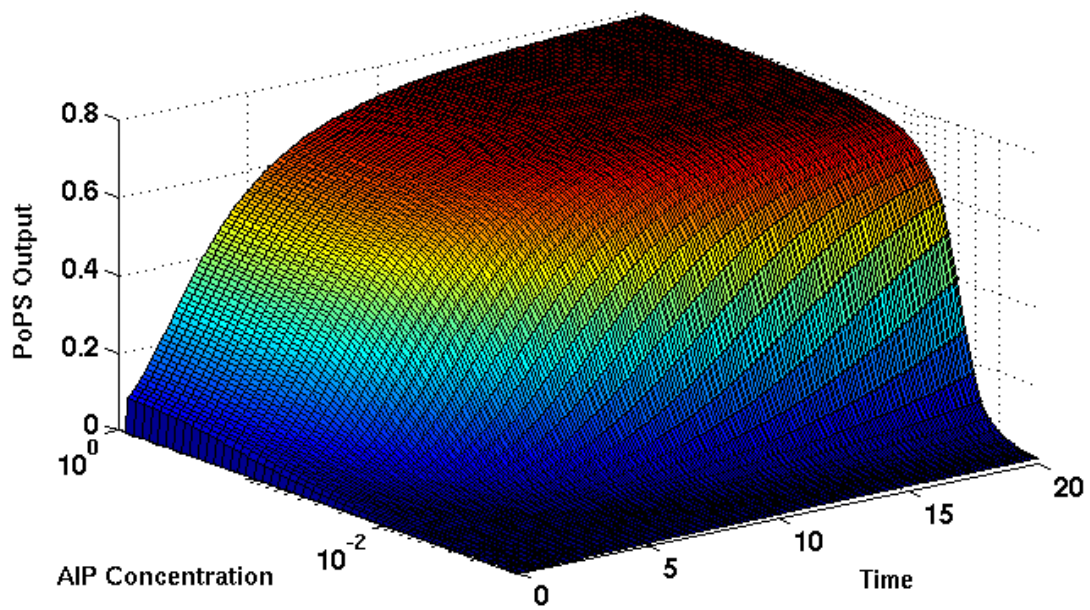


Figure 4: Transfer function surface up to  $t=20$

$$\frac{\partial[BD]}{\partial t} = c_+[B][D] - c_-[BD] - p_{AIP}[BD] - \delta[BD] + \beta \approx 0 \quad (4)$$

$$\Rightarrow [BD] = \frac{c_+[B][D] + \beta}{c_- + p_{AIP} + \delta} \quad (5)$$

$$\left\{ \begin{array}{l} \frac{\partial[B]}{\partial t} = PoPS - c_+[B][D] + c_-[BD] - \delta[B] + \beta \\ \frac{\partial[D]}{\partial t} = PoPS - c_+[B][D] + c_-[BD] - p_{AIP}[BD] - \delta[D] + \beta \\ \frac{\partial[P]}{\partial t} = \frac{\rho}{1-\rho}(p_{AIP}[BD] - \delta[P] + \beta) \end{array} \right.$$

where  $p_{AIP}$  is the production rate of AIP (we take it to be 1). After substituting BD into the rate equations, we end up with a three-component system again. In the wild-type system, AIP synthesis lies downstream of the p2 promoter (fig.1). There is again a term involving  $\rho$  to account for extra- and intracellular medium.

## Modelling the complete agr-quorum sensing system

Let us now consider the complete agr-QS system as illustrated in fig. 1. Clearly, this will be described by solving the equations for the agr-receiver and -sender simultaneously (where  $PoPS$  in the sender part will now be the p2 synthesis term from before). The system of equations to be solved will consist of the rate equations of  $[A] - [D]$  and  $[P]$ . For the rate equation involving  $[P]$  we need to add both contributions from the sender and receiver systems and only take a single decay and basal expression term so that the resulting equation reads:

$$\frac{\partial[P]}{\partial t} = \frac{\rho}{1-\rho}(p_{AIP}[BD] - c_+[C][P] + c_-[CP] + (\beta - \delta)[P])$$

In the lab one can simply<sup>5</sup> put the sender and receiver device downstream of a p2 promoter to recreate the agr-QS system with biobricks. For completeness, the entire model reads:

$$\left\{ \begin{array}{l} \frac{\partial[A]}{\partial t} = \frac{[A^+]}{1 + [A^+]} + \beta - p_+[CP][A] + p_-[A^+] - \delta[A] \\ \frac{\partial[A^+]}{\partial t} = p_+[CP][A] - p_-[A^+] - \delta[A^+] \\ \frac{\partial[C]}{\partial t} = \frac{[A^+]}{1 + [A^+]} + \beta - c_+[C][P] + c_-[CP][P] - \delta[C] \\ \frac{\partial[B]}{\partial t} = \frac{[A^+]}{1 + [A^+]} - c_+[B][D] + c_-[BD] - \delta[B] + \beta \\ \frac{\partial[D]}{\partial t} = \frac{[A^+]}{1 + [A^+]} - c_+[B][D] + c_-[BD] - p_{AIP}[BD] - \delta[D] + \beta \\ \frac{\partial[P]}{\partial t} = \frac{\rho}{1-\rho}(p_{AIP}[BD] - c_+[C][P] + c_-[CP] + (\beta - \delta)[P]) \end{array} \right.$$

and we substitute in the appropriate terms that we arrived at using quasi-steady state assumptions. Remember we took  $\beta = 0.05$ . We solve these equations and plot PoPS output w.r.t. the volume fraction of intracellular space in fig. 5. In the case of *S. aureus*, PoPS output drives a virulence gene.

<sup>5</sup>Biologists will clearly agree with this statement.



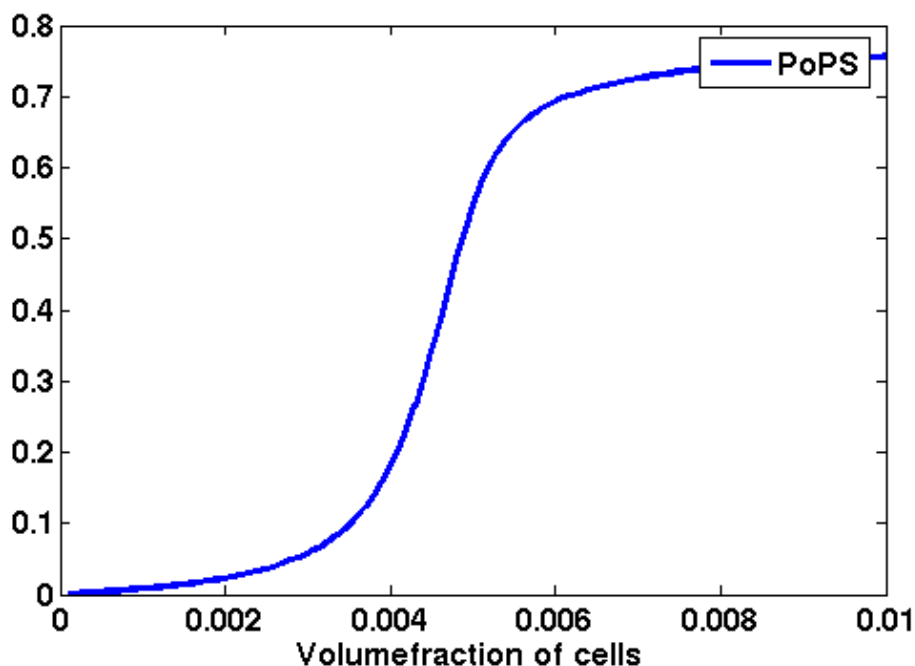


Figure 5: PoPS output of agr-QS system at  $t=20$

While the curve is surprisingly robust when changing initial conditions and most parameters, it is extremely sensitive when varying basal rates of expression (assuming maximum expression is fixed at 1). The shape of the curve and particularly the threshold density at which the system switches behaviour, is therefore largely dependent on the basal rate of expression. The QS-system switches from low- into high-density behaviour at a local density of roughly 0.005 per unit volume.

While we were not able to find data that would verify this estimate, we can consider the gram-negative *E.coli*. In ideal conditions, *E.coli* grows to roughly  $(2-5) \times 10^9$  cells/ml in LB until it reaches stationary phase. Assuming that a single *E.coli* cell has a volume on the order of  $1 \mu m^3$ , the cell density at stationary phase will be on the order of  $10^{-3}$  per unit volume. Assuming that growth in ideal conditions implies higher resulting cell densities, this number will be an upper limit to cell densities, and the *E.coli*-QS system must be activated at roughly those densities (or slightly below). This validates that our guess of  $5 \times 10^{-3}$  is in the correct order of magnitude.

Finally, it is interesting to note that we have arrived at this result by purely theoretical considerations. We have seen that the basal rate of expression plays a critical role in the system response. This in turn will be useful in informing possible design choices for engineered applications: To change the threshold density one can simply increase basal rates of expression, e.g. by constitutively expressing the agr-system in parallel to the agr-genes that are regulated by  $p_{lux}$ . As an example, if one was to increase basal expression rates (i.e.  $\beta$ ) the threshold value would decrease. This makes it extremely interesting to engineered applications in Synthetic Biology.

## Pattern formation

Is it possible to engineer bacteria that exhibit multicellular behaviour and can express spatial patterns (e.g. on a bacterial lawn)?

We would like to use AIP quorum-sensing systems to demonstrate how this might be accomplished. AIP-based quorum-sensing systems have the advantage that there exist numerous derivatives in gram-positive bacteria that are mutually exclusive. This stands in contrast with OHL-based systems (in gram-negative bacteria) that exhibit a high occurrence of interference between OHL derivatives.

Before we try to model such behaviour, let us look at a simple mathematical model for pattern formation.

## Reaction-diffusion systems

The theory of patterns is vast and it would (and does) extend beyond the scope of this project. However, we will look at a very simple mathematical system that is capable of generating patterns. This might give us some intuition in how patterns are formed even though in our case there might be no direct application.

**Turing Patterns** Turing's idea was to look at systems in the absence of diffusion. He required such a system to have a stable steady-state and upon addition of diffusion, required this to become unstable. A simple example is the Schnakenberg system that is analysed in *Mathematical Biology* by J. Murray.

$$\begin{cases} \frac{\partial A}{\partial t} = \gamma(\alpha - A + A^2B) + \nabla^2 A \\ \frac{\partial B}{\partial t} = \gamma(\beta - A^2B) + d\nabla^2 B \end{cases}$$

This is a simple system that is capable of creating diffusion-driven patterns. The mathematical analysis goes as follows: One considers the system in the absence of diffusion and looks for a stable steady-state. Next, one considers the same system with the diffusion term added and substitutes in a trial solution that is composed of a Fourier series while requiring the steady-state to be unstable. This will result in a number of further conditions (inequalities) and it is easy to find the (Turing) space of parameters that will satisfy all conditions for pattern generation. A gallery of patterns produced by Schnakenberg system can be found on the next page.

We can see that the type of patterns created is highly dependent on our parameter values, particularly  $\alpha$  and  $\beta$ . Can we arrive at similar patterns with our AIP-QS system?

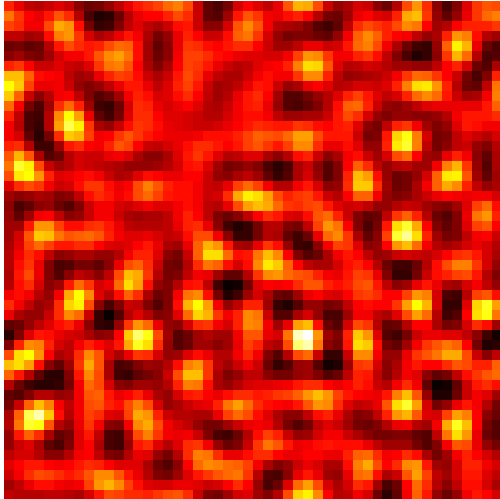


Figure 6:  $d = 27.5, \alpha = 0.2, \beta = 1$

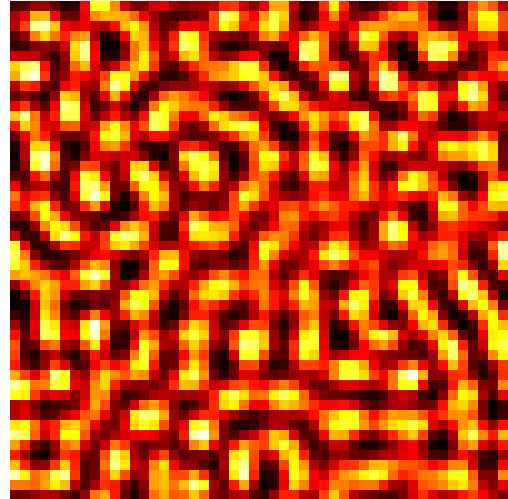


Figure 9:  $d = 20, \alpha = 0, \beta = 1.4$

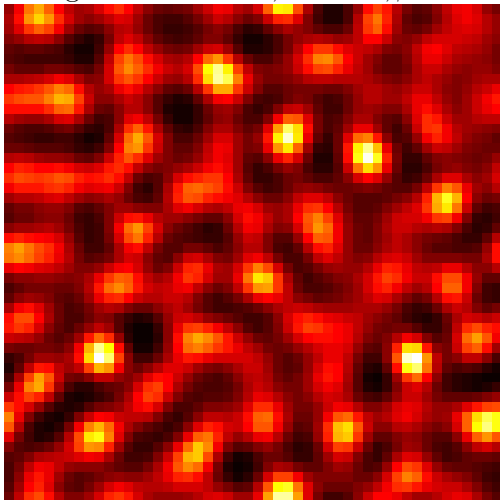


Figure 7:  $d = 29.5, \alpha = 0.2, \beta = 0.6$

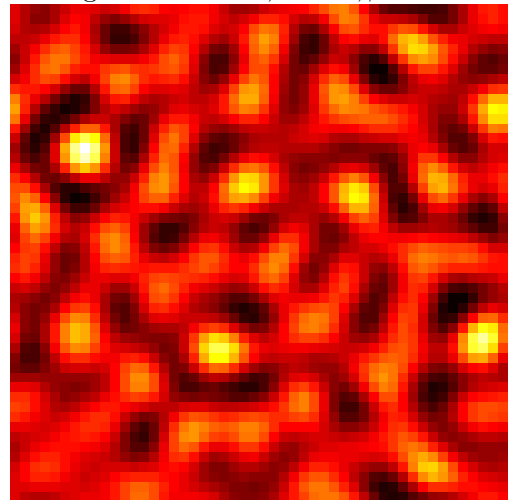


Figure 10:  $d = 21.5, \alpha = 0.2, \beta = 0.4$

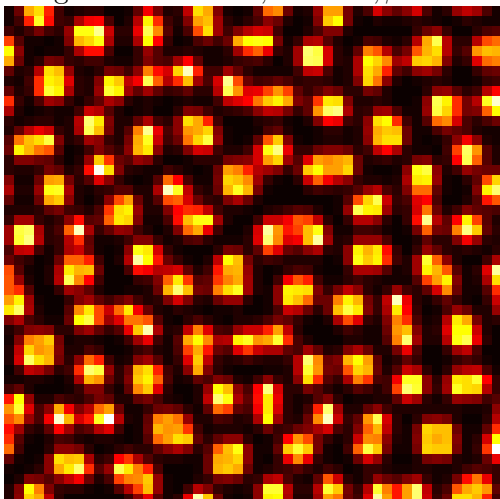


Figure 8:  $d = 30, \alpha = 0, \beta = 12$

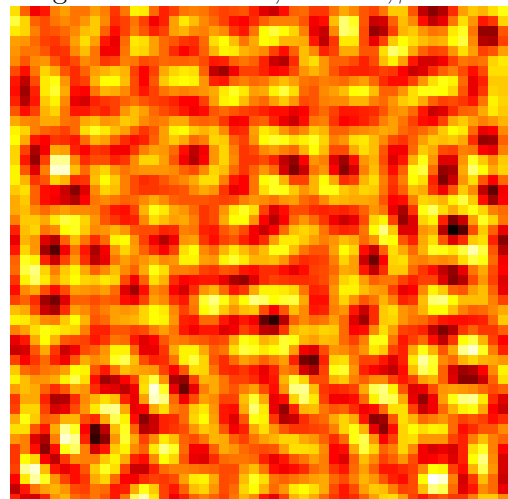


Figure 11:  $d = 30, \alpha = 0, \beta = 2$

## Pattern formation in Synthetic Biology

How can we engineer bacteria to express spatial patterns? (It is interesting to note that many bacteria do form patterns, particularly when exposed to different nutrient levels) We would like to express patterns on a bacterial lawn with constant bacterial density everywhere and we can think of patterns as spatial distributions of PoPS output (downstream of which we can put GFP) or cell differentiation. Since we have modelled the agr-system and know that non-interfering AIP-based systems exist, we would like to construct a pattern generating system as follows: Both the AIP-A and AIP-B systems inhibit production of the "competitor" molecule. Furthermore, both systems are self-activating. This would be a more complex analogue to the simple reaction-diffusion system introduced previously.

In the following part, we assume that nutrients are abundant (we do not model nutrient consumption) and that the bacterial density throughout the lawn is constant and high by setting  $\rho = 0.01$ . Note that we are disregarding the fact that these bacteria now live on a 2d surface. We would arrive at a more detailed estimate for bacterial densities in a bacterial lawn if we introduced a diffusion term to account for signalling molecules diffusing across the nutrient layer.

Furthermore, the synthesis terms in this model are bounded in our system. Meinhardt mentioned<sup>6</sup> stripe-like patterns can be generated by such systems where activator production is bounded. While he was analysing a simple two-component system, we would like to generate patterns by using two agr-systems in parallel, acting as activator and inhibitor respectively.

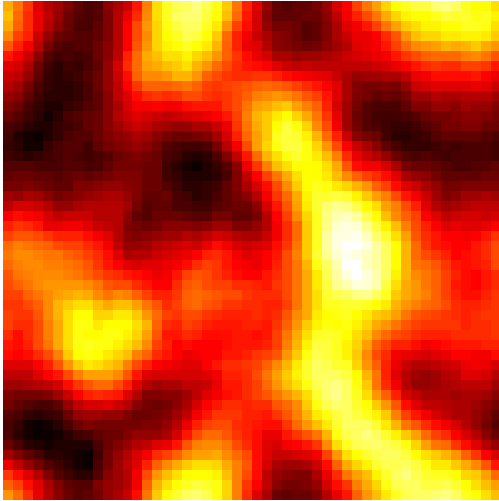
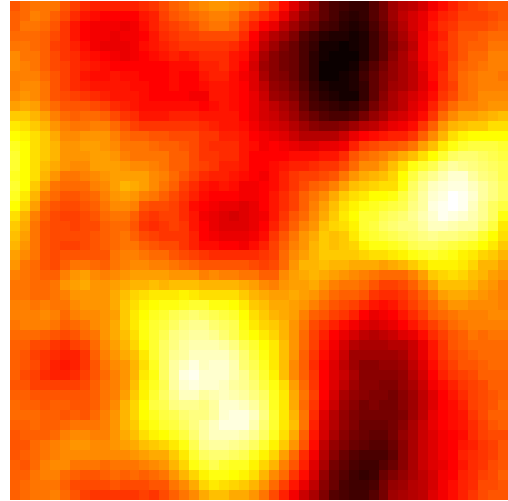
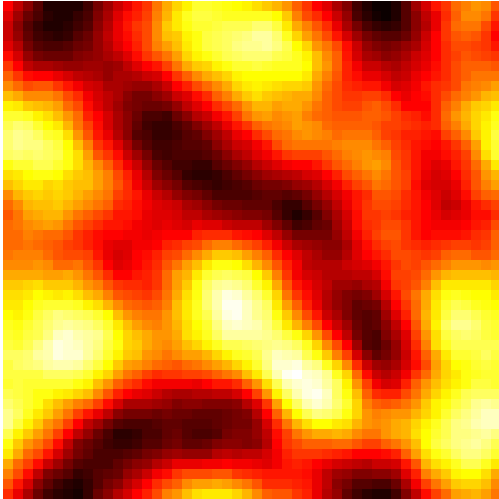
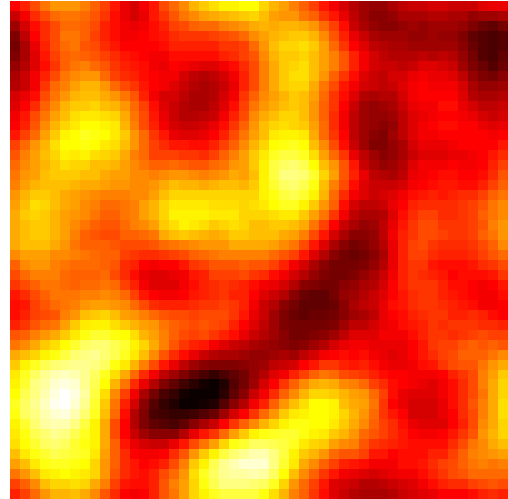
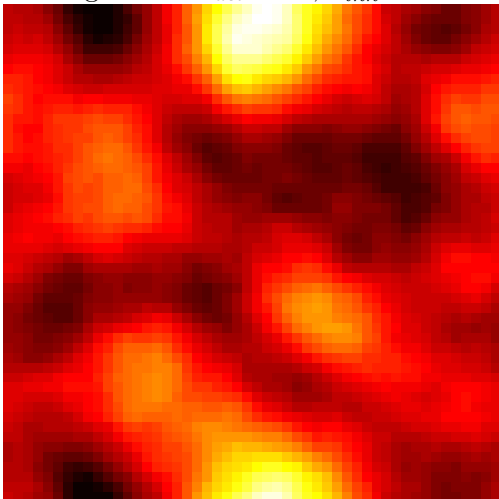
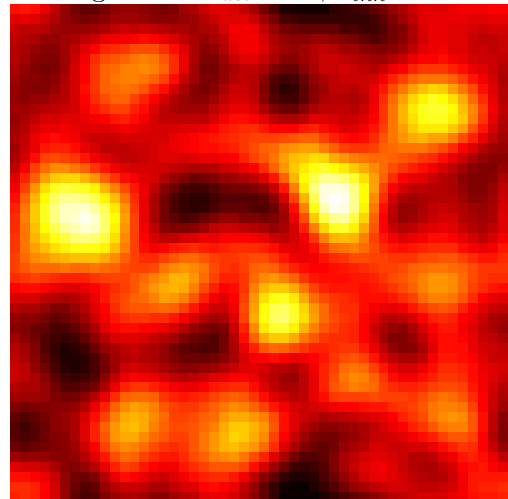
The rate equations for one of the parallel agr-systems (with peptide  $P_1$ ) read as follows:

$$\left\{ \begin{array}{l} \frac{\partial[A]}{\partial t} = \frac{[A^+]}{1 + [A^+]} f(P_2) + \beta - p_+[CP][A] + p_-[A^+] - \delta[A] \\ \frac{\partial[A^+]}{\partial t} = p_+[CP][A] - p_-[A^+] - \delta[A^+] \\ \frac{\partial[C]}{\partial t} = \frac{[A^+]}{1 + [A^+]} f(P_2) + \beta - c_+[C][P] + c_-[CP][P] - \delta[C] \\ \frac{\partial[B]}{\partial t} = \frac{[A^+]}{1 + [A^+]} f(P_2) - c_+[B][D] + c_-[BD] - \delta[B] + \beta \\ \frac{\partial[D]}{\partial t} = \frac{[A^+]}{1 + [A^+]} f(P_2) - c_+[B][D] + c_-[BD] - p_{AIP}[BD] - \delta[D] + \beta \\ \frac{\partial[P_1]}{\partial t} = \frac{\rho}{1 - \rho} (p_{AIP}[BD] - c_+[C][P] + c_-[CP] + (\beta - \delta)[P]) + d_1 \nabla^2 P_1 \end{array} \right.$$

where  $f(x) = \frac{40}{0.1 + x}$ . The equations for the the second agr-derivative system with peptide  $P_2$  read analogously with  $P_1$  and  $P_2$  interchanged (and independent variables for intracellular enzymes). Note that we have added a diffusion term to account for spatial diffusion across the plane.

Solving those equations with random initial conditions in levels of  $P_i$  gives us:

<sup>6</sup>[Meinhard, 1992] Rep. Prog. Phys., Pattern formation in biology: a comparison of models and experiments

Figure 12:  $d_{act}=401, d_{inh} =350$ Figure 15:  $d_{act}=441, d_{inh} =350$ Figure 13:  $d_{act}=421, d_{inh}=350$ Figure 16:  $d_{act}=371, d_{inh}=350$ Figure 14:  $d_{act}=431, d_{inh}=350$ Figure 17:  $d_{act}=500, d_{inh}=350$ 

Some regularity in the patches of differentiation can be observed although these "patterns" depend to a greater degree on initial conditions compared to the patterns seen previously in the Schnakenberg system. However, this is a promising indicator that demonstrates how coordinated multicellular behaviour might originate from an engineered system.

## Further Questions

**A concise model** We have just tried to go from intra- to intercellular behaviour by naively applying the primarily intracellular agr-QS model to the intercellular case. We cannot say how accurate this approach is since we did not run experiments. However, we can consider a more concise and simpler version of our pattern-generating system. The most important feature of our envisioned systems is that both "populations" inhibit or compete against each other and that the auto-synthesis term is bounded. Assuming that competition can be modelled by simple mass-action, a simplified version of our model would read:

$$\begin{cases} \frac{\partial A}{\partial t} = p_1 \frac{A}{1+A} - k_1 AB - \delta A + d_1 \nabla^2 A \\ \frac{\partial B}{\partial t} = p_2 \frac{B}{1+B} - k_2 AB - \delta B + d_2 \nabla^2 B \end{cases}$$

Of course, one can also try other systems that incorporate the main features of our system. Further work can be done on the dynamics that these systems exhibit.

**Other pattern-generating systems** There is a vast number of pattern-generating systems. While we try to model two competitively inhibiting populations, patterns might be created by a completely different system. In particular, we are interested in the question how pattern generation might occur using a single signalling molecule only.

**Bounded patterns** Patterning is everywhere in biology - one can think of the whole development process as "pattern formation". Here, pattern generation is switched on by higher-level mechanisms and patterns are expressed in bounded domains (e.g. if we regard organs as patterns, these are clearly separated from other organs). How does this work? How can we generate patterns that are constrained to a domain? Can we engineer a system that might exhibit this kind of behaviour?

## Notes

All equations have been solved numerically with a simple difference method - this method is easy to implement and gives accurate results, even though one must be careful that stability does not become an issue.