Engineering Microbes to Sense and Eradicate Pathogens

CoMPLEX MRes Mini-Project 1, January 16, 2014

Author:
Talfan Evans

Supervisors:
Dr. Chris Barnes and Dr. Geraint Thomas

Abstract

This study provides a review, and an investigation into the mechanisms of quorum sensing, used by cells to coordinate collective behaviour. A mathematical model of a synthetically engineered E. Coli is presented, capable of detecting and eradicating the human pathogen Pseudomonas Aeruginosa. Simulations confirm the possibility of cooperative binding of the quorum sensing autoinducer 3-Oxo-HSL to the luxR gene, and show agreement with observed protein expression data.

Contents

1 Introduction

In the past century, the life sciences have enjoyed a period of technological advancement unparalleled in its history, driven largely by improved instrumentation and the constantly evolving demand for better healthcare, allowing us to peer further into the inner machinery of the cell. Biological systems are innately complex, comprising vast networks of interconnected biological components that, when acting in harmony give living organisms the ability to react adaptively to their environments. It is the understanding of this machinery that arguably defines the scope of modern biology.

Cells act upon their environments through protein expression, enacted by the reading of relevant regions of DNA regulated in gene networks by transcription factors, chemical species that bind to specific genes and regulate the activity of RNA polymerase, leading either to expression or repression of proteins. Together, these systems allow cells to sense and react to external stimuli, however additional complexity arises when proteins act as their own transcription factors. This behaviour gives rise to intrinsic feedback, the understanding of which is often impossible by traditional methods, and it is here that modern computing plays a crucial role in handling the enormous quantities of data needed to capture the behaviour of even modest biological systems.

Engineering has not evolved so far on the nano-scale as to be able to manipulate cellular components accurately enough for construction, and for these purposes nature's own tools, restriction and ligation enzymes[7] are exploited, allowing cellular components to be interchanged to produce novel, recombinant organisms. Polymerase Chain Reaction technology[8] can then be utilized to amplify these synthetic DNA/RNA strands into practical quantities. Notable triumphs of genetic engineering include the discovery, and directed use of fluorescent proteins[9], used for *in vivo* monitoring of gene expression.

The aims of synthetic biology are similar to those of any other engineering discipline. The design of modular components able to interact interchangeably with one another is crucial for the construction of complex systems, not only to ease the difficulties of manufacture but also to make them analyzable and deterministic in their function. Specificity and sensitivity are essential requirements to orthogonality, defined by an absence of interaction with native cellular pathways[1], allowing these systems to be approached logically and methodically.

Problems of compatibility may arise where regulatory molecules serve different purposes within two components, thus necessitating the availability of many parts with equivalent function but different specificity[1].

Though the design of biological systems can call on many principles of engineering design, the crucial difference is that of connectivity. Whereas electronic components may interact through wires, cellular components must do so through the medium of diffusable chemical molecules, which are subject to stochasticity and noise. Whereas biological systems rely on complexity and redundancy to ensure robustness, its implementation in artificially engineered circuits is not so trivial. In order for such systems to be viable, components must adhere closely to the principles outlined above, and to be tunable to cope with unpredictable changes.

In this study, a mathematical analysis of one such system, synthesized by Saeidi et al.[2] and capable of sensing and eradicating the human pathogen *Pseudomonas Aeruginosa* is presented. The first section develops and implements a model describing the system used

to detect the pathogen's presence, whereas the second deals with the targeted destruction of those cells. The effect of stochasticity on the system's behaviour is investigated, and an effort is made to determine from available experimental data reasonable estimates to the model parameters. By comparing the performance of the proposed model to the experimental data, the study aims to elucidate the cellular mechanisms of the engineered organism, so that its performance might be further improved for future therapeutic use.

2 Literature Review

The potential of synthetic biology has to date been demonstrated in an extensive range of applications. Early work was concentrated on the computational analogies of the field, with much focus dedicated to producing equivalent Boolean operators[10], soon leading to more ambitious multi-component systems. The repressive effects of some proteins has been exploited to synthesize a toggle switch[11] able to exhibit bistability, the biological equivalent of the electronic flip-flop 1-bit memory device. In combination with bio-logic gates, these components can be assembled to increase storage capacity[12].

Furthermore, synthetic circuits have been developed to perform logical operations, both in single-cells[14, 13] and as multicellular systems[15] using genetically encoded NOR-gate 'chemical wires' to facilitate intercellular communication. Band-pass filters[20] have been developed to combat the noise associated with inter-component communication.

Following further in the footsteps of computational development, much attention has now shifted to producing extra-corporeally programmable and controllable computing units, in a step towards more practical clinical control. Notably, optogenetic glucose controllers have been synthesized [21] allowing the management of diabetes in mice using blue light, and alternating magnetic fields[22] used to control heat-induced therapeutic genes for tumor growth inhibition.

Conversely, much research has been conducted into the dynamic, rather than static behaviour of gene networks, as many cellular processes rely on regular, robust and periodic oscillations to function. Many models have been proposed[17, 18] to establish these circadian rhythms[16], a notable triumph being the Repressilator[19], a three-gene circuit shown to exhibit self-sustained periodic oscillations.

Though synthetic biology as a field is very much in its infancy, the aggregate of these individual developments is already showing profound practical relevance in many industries. The understanding of gene regulatory networks is imperative to the development of robust organisms that will dramatically reduce the cost of the bioprocessing[23] of medicine, biofuels[24] and novel biomaterials.

However, it is the therapeutic potential of synthetic biology that is the main focus of this study, in particular to the development of synthetic systems towards combating infectious diseases and cancers. Weiss et al.[25] demonstrated a system capable of selectively detecting HeLa cells and triggering apoptosis, without affecting non-HeLa types. The system acts as a multi-input AND gate, programmed to detect the presence of six endogenous miRNA species indicative of HeLa presence. The system performs a logical operation such that the system response is only positive in the presence of all six species, releasing human Bcl-2–associated X protein hBax chosen to cause selective apoptosis in the desired cells. Their system demon-

strated on average a 160 fold measured output in the presence of HeLa cells, however the important aspect of the system is it's programmability, where in fact the miRNA profile may be tailored to detect any combination of markers, and the output to any such desired protein expression.

2.1 Quorum Sensing

Quorum sensing is the mechanism of intercellular communication used to coordinate collective behaviour by groups of cells, facilitated by an organic chemical signaling cascade[3]. Signaling chemicals termed autoinducers are released, and are detectable in trace quantities by other cells. The concentration of autoinducer increases with local cell density, triggering internal gene mechanisms at threshold concentrations, causing cells to exhibit a coordinated change in behaviour. To reinforce this mechanism, the autoinducers are often self-activating, causing an exponential increase in production in response to their own detection.

QS is used to coordinate motility, growth, virulence and biofilm formation, an extracellular polymeric substance formed from a mixture of planktonic and adherent communities of bacteria[4]. Biofilms have been shown to act as protective, nutrient-rich barriers that diminish the effect of antimicrobial agents[5] and mask their presence from the host cells.

QS has also been observed both in interspecies communication. Diggle et al. propose that bacteria are able to detect these 'cues' from other local species either manipulatively, for example to 'coerce' a response, or to facilitate a mutually beneficial symbiosis, by managing resources[29] or avoiding punishment[28]. Inter-kingdom communication, between bacteria and higher organisms has, although very rarely, also been observed, where the marine red algae *Delisea Pulchra* is able to produce a QS-disrupting halogenated furanone that prevents the swarming motility of *Serrantia Liquefaciens*[30].

The detection of these autoinducer molecules has clear significance in the design of systems for selective detection of pathogens, however quorum sensing can also be exploited for offensive purposes through an understanding of the collective behaviour intended, an instructive example being that of *Vibrio Cholerae* infection of the gastrointestinal tract through tainted water. At low cell densities, *V. Cholerae* expresses the chemical toxin corregulated pilus (TCP), facilitating attachment to the intestinal wall[26]. Once attached, the pathogen continues to grow, releasing the virulence factor cholera toxin (CT), causing dehydration and diarrhea. However, *V. Cholerae* employs a quorum sensing system and, when a threshold density is reached coordinates an exchange of virulence factor expression for protease expression, causing the attachment matrix to degrade, thus allowing movement out of the host along with diarrheal excretions.

Duan et al.[27] modified the commensal bacteria Escherichia Coli Nissle 1917 to express the autoinducer CAI-1 secreted by *V. Cholerae*, causing it to release itself from the host system. Their study found that mice pretreated with the engineered Nissle had a 92% survival rate over the untreated mice, of which 0/14 survived.

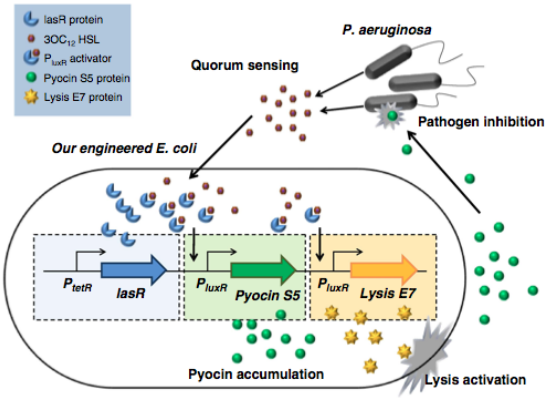


Figure 1: Engineered *E. Coli* system[2].

2.2 Engineering *E. Coli* to Sense and Eradicate *P. Aeruginosa*

Though it has been demonstrated that some native QS systems expropriated, it is not always the case that this is so effective. It is also possible to simply use QS as a detection mechanism, implementing a secondary system to destroy the target.

Saeidi et al. engineered *E. Coli* bacteria to detect autoinducer molecules secreted by the opportunistic pathogen *Pseudomonas Aeruginosa*, an increasingly prevalent human pathogen that is a leading cause of nosocomial infection. Currently, *P. Aeruginosa* accounts for '12% of hospital-acquired urinary tract infections, 10% of blood stream infections, and 8% of surgical wound infections'[31]. It is particularly lethal to immuno-compromised patients, such as those suffering from cystic fibrosis, cancer and ventilator associated pneumonia[32].

Chang et al.[3] attribute the pathogen's intrinsic antibiotic and antimicrobial resistance to its effective efflux systems, while Govan and Deretic[5] show that an alginate biofilm may act as a barrier to potential treatments. Due its fast accession of drug tolerance, conventional therapies more often prove ineffective in treating *P. Aeruginosa*, and so multi-integrated treatments are often preferred, though this approach is undirected and risks the unspecified killing of the delicate balance of symbiotic bacterial cultures present in the human body, often referred to as the microbiome[34].

P. Aeruginosa is unfortunately not the only human pathogen to exhibit such resilience against conventional treatment, other notable examples being the *Staphylococcus* family, and thus the need for a new approach to their eradication is necessary.

The system engineered by Saeidi et al. consists of three principle components, illustrated in Fig. 1.

The first is the QS system, taken from the *P. Aeruginosa* pathogen itself. The $pTetR$ gene constitutively produces $LasR$, which binds to the autoinducer molecule $3-Oxo-C12-HSL$ secreted by the pathogen which is readily diffusible through the membrane of *E. Coli*. Once

bound, the $LasR-HSL$ complex binds to and activates $luxR$ gene, which expresses the target protein.

There are in fact two copies of the $luxR$ gene in each transfected plasmid, where the first is set to express the bacteriocin Pyocin S5, a ribosomally synthesized antimicrobial peptide produced naturally by *P. Aeruginosa*. When released, the Pyocin diffuses towards the target pathogen and degrades its outer membrane, causing leakage of intracellular materials, increased membrane permeability[36], cell surface disruption[35] and eventually lysis.

However, the Pyocin is produced intracellularly and is not diffusible through the cell membrane, so cannot be released. For this purpose, the second $luxR$ produces the Colicin produced E7 Lysis protein, which lysis the host cell and exposes the interior cell contents[37], allowing Pyocin to be released.

Saeidi et al. demonstrated in *in vitro* studies that their system was able to sense and lyse 99% of the viable *P. Aeruginosa* cells, and inhibited biofilm formation by almost 90%. However, *in vivo* studies are yet to be undertaken, and the authors admit that the effectiveness of such systems would be 'highly dependent on the molecular diffusion' of the Pyocin. In aid of this, further work done by Hwang et al.[4] has implemented the previously described system within an *E. Coli* cell capable of chemotactic motility towards the target pathogen through regulation of the $CheZ$ protein in response to autoinducer gradients.

3 Modelling

3.1 Modelling Approaches

3.1.1 Ordinary Differential Equations

Chemical reaction networks can bet handled within the Mass Action Kinetics framework, developed to predict the behaviour of multiple chemical species in dynamic equilibrium. Ordinary differential equations lead naturally from this framework, and are a commonly used method of modelling deterministic dynamics. However, where small numbers of reactants are concerned, ODEs may not be suited to describing the random perturbations present on the microscopic scale.

3.1.2 Stochasticity and the Gillespie Algorithm

These perturbations are stochastic in nature and so cannot be evaluated exactly for any time-course. To deal with discrete systems of this nature, the Gillespie algorithm was developed. In the system, each possible reaction is allocated a *hazard* of occurring, and by summing the total hazard over all reactions, the time to the next reaction can be calculated from an exponential distribution[49].

At this new time-step, a reaction is selected according to weighted probabilities and evaluated, thus updating the species numbers. The algorithm then proceeds in this fashion, updating the reaction hazards until the desired time is reached.

3.1.3 Maximum Likelihood Estimation

Maximum likelihood estimation can be used to deduce unknown parameter values by fitting to empirical data. Unlike other methods, MLE takes into account experimental error, and so is often preferred over other methods.

By assuming the experimental error to be distributed normally, the probability of any data point occurring given the mean and variance of the empirical data can be calculated. The method thus seeks to maximise the likelihood function $\mathcal{L}$, defined as the probability of all predicted data points occurring given prior data, assuming a normal distribution of error. This is stated mathematically as:

$$\mathcal{L} = \prod_{i=1}^{8} \frac{1}{\sigma_i \sqrt{2\pi}} Exp\left[\frac{1}{2\sigma_i^2}\left(x_i - \mu_i\right)^2\right] \tag{1}$$

Where μ_i and σ_i are the mean and variance of the observed data points, and x_i the values predicted by a model for a particular parameter set.

3.2 Reaction Network

The dynamics of the engineered *E. Coli* are represented by the following reactions, adhering to the mass-action kinetics framework:

$$A_o\left(T\right) \overset{\sigma_A(M)}{\rightleftharpoons} A_I \tag{2}$$

$$LasR \underset{V_{0L}}{\overset{k_{dL}}{\rightleftharpoons}} \varnothing \tag{3}$$

$$LasR + A_I \underset{k_{rC}}{\overset{k_{fC}}{\rightleftharpoons}} C \tag{4}$$

$$nC + luxR_P \underset{k_{rB}}{\overset{k_{fB}}{\rightleftharpoons}} nC.LuxR_P\left(B_P\right) \tag{5}$$

$$nC.luxR_P \overset{k_{m\_P}}{\rightarrow} nC.luxR_P + p \tag{6}$$

$$p \overset{k_{R\_P}}{\rightarrow} P_I \tag{7}$$

$$nC + luxR_E \underset{k_{rB}}{\overset{k_{fB}}{\rightleftharpoons}} nC.luxR_E\left(B_E\right) \tag{8}$$

$$nC.luxR_E \overset{k_E}{\rightarrow} nC.luxR_E + e \tag{9}$$

$$e \overset{k_{m\_E}}{\rightarrow} E \tag{10}$$

$$E + M \overset{k_{dM}}{\rightarrow} \varnothing \tag{11}$$

$$P_I \overset{\sigma_{S5}(M)}{\rightarrow} P_o \tag{12}$$

$$P_o \overset{k_{dS5}}{\rightarrow} \varnothing \tag{13}$$

Firstly, the autoinducer produced by the target pathogen (in this case *P. Aeruginosa*) diffuses into the *E. Coli* cell across its membrane. The autoinducer $3OC_{12} - HSL$, which will be referred hereon as A, belongs to the N-Acyl Homoserine Lactone family used commonly by a many quorum sensing bacterial species. Both *V. Fischieri* and *V. Cholerae* employ variants of this molecule, with differing numbers of carbon atoms. Pearson et al.[38] conducted a study on the transport properties of these molecules, and found that in *P. Aeruginosa*, there was indeed evidence of active transport mechanisms, as the inner concentration of A saturated at approximately three times its external concentration. However, *E. Coli* demonstrated no such properties, and A was found to diffuse freely through its membrane. Thus, its diffusion can be modelled as a simple diffusive process according to Darcy's Law, whereby σ_A represents the diffusivity.

Coincidentally, it is possible to represent this diffusive process in the framework of mass-action kinetics, described earlier, as solving Eq. 2 , where A_o and A_I represent extracellular and intracellular concentrations of A respectively, leads to an ODE of equivalent form to Darcy's Law, where the diffusion rate is proportional to the concentration difference across the membrane, and the diffusivity of the A molecule.In the model, the diffusivity is modelled as a function of the membrane integrity $M(t)$ in order to capture the effect of membrane degradation with Pyocin release.

Eq. 3 describes the basal production and degradation of the $LasR$ protein. $LasR$ is expressed constitutively at a rate V_{0L} by the $pTetR$ gene (see Fig. 1), and degrades at a natural rate k_{dL} proportional to its concentration.

Eq. 4 describes the binding of A_I and $LasR$ to form the autoinducer complex C, where the rates of binding and unbinding are denoted k_{fC} and k_{rB} respectively. Claussen et al.[39] propose a kinetic model of this binding mechanism based on *in vitro* data, whereby the $LasR$ monomer is inherently unstable, but dimerizes to form a more stable compound which is resistant to protease degradation. The $LasR$ dimer then binds one molecule of AI in a reversible manner, represented by the two reaction rates shown. However, the rate of formation and degradation of the $LasR$ monomer and dimer is fast relative to its degradation, and thus it is possible to capture the salient dynamics of the process modelling $LasR$ as a single species.

Once bound, the autoinducer complex C binds to the $luxR$ gene promoter region, forming the bound state B, shown in equations 5 and 8, where the two processes are assumed as identical. Similarly to the N-Acyl Homoserine Lactones, the $luxR$ gene is prolific among quorum sensing bacteria. As such, there is fairly substantial literature, however consensus has not been reached over the exact mechanism of autoinducer binding. Hunter et al.[40] proposed a quantitative model, with their parameterisation efforts suggesting a cooperative mechanism whereby two autoinducer molecules bind to a dimerized $LuxR$ protein, which then attached to the $luxR$ gene. However, the full-length $LuxR$ has been not been studied enough to confirm this fact[41] , and other modelling efforts[**?**] assume monomeric binding. The model thus leaves the determination of the binding cooperativity n to parameterisation, explored later in the study.

Equations 6 and 9 illustrate the production of the mRNA coding for the expression of Pyocin and Lysis from the activated $luxR$ genes, where crucially, B is not consumed. Eq. 7 and 10 represent the expression of these proteins from the mRNA strands.

Eq. 11 represents the degradation of the host (*E. Coli*) membrane, where the membrane

integrity M can be considered to react destructively with the E7 Lysis protein, produced intracellularly. Eq. 12 then represents the dissemination of Pyocin from the host cell. As Pyocin is not diffusable in the host membrane, the diffusivity factor σ_{S5} represents the increasing permeability of the membrane as it is degraded. Eq. 13 represents the degradation of the extracellular Pyocin.

4 Characterisation of the Quorum Sensing Circuit

The bound $pLuxR$ promoter region is only an activator for protein expression, and is not consumed in reactions 6 or 9, and thus the reaction network effectively represents two decoupled systems.

These can be regarded as the Quorum Sensing System 1 and Lysing System 2. System 1 detects external autoinducer molecules secreted by *P. Aeruginosa*, and System 2 produces the Pyocin and E7 Lysis necessary for its lysis.

Fig. 2B of Saeidi et al.[2] shows the steady state expression of protein by the $luxR$ gene in response to varying external autoinducer concentration A_0 . In order to characterize the system, the mathematical model developed above is compared to the experimental data presented.

4.1 Details of Protein Production

4.1.1 Motivation

The production of the desired protein P (e.g Pyocin, Lysis, GFP) by the $luxR$ gene is in itself a complex network of many reactions, however it is often preferable to reduce these to a simpler model.

Firstly, it is in most cases unfeasible due to computational expense, to accurately estimate the large number of rate parameters from data, especially where such data is sparse.

Secondly, there is nothing to guarantee the accuracy or biological relevancy of these estimates. Including more parameters increases the flexibility of the model, and often leads to 'overfitting' of the data.

Thirdly, even in the case of accurate parameter estimation, the dynamics captured by a larger model are usually dominated by a smaller subset of processes, and so the gain in accuracy is not justified over a simpler model that also captures the characteristic behaviour of the system. In addition, the dynamics captured by a simplified model are likely to be more biologically relevant, and their subsequent analysis will certainly be more tractable.

4.2 Steady State Analysis

Using the Briggs-Haldane[42] approximation, the conservation of the binding sites, and considering the relative roll of the mRNA in protein expression, eqs. 5-10 can be reduced (See Appendix) to:

$$\frac{dX}{dt} = V_X . \frac{1}{\left(\frac{K_{\overline{mB}}}{C}\right)^n + 1} - k_X . X + X_0 \tag{14}$$

Where k_X is the degradation rate of target protein X, X_0 its basal production rate, V_X its expression rate and $K_{\overline{mB}}$ the binding dissociation constant. The overall system can be seen also to behave with Hill-type kinetics. In their paper, Saeidi et al. substitute GFP downstream of the $luxR$ gene as a marker to investigate protein production vs. A_0, and find that fitting a Hill-type expression captures the observed response, thus agreeing with the behaviour predicted by the above model.

4.2.1 Autoinducer Complex C

The $LasR$ monomer, produced at a basal rate, is unstable and decays rapidly. However, some of these molecules dimerize to form a more stable complex, which degrades at a much slower rate k_{dL}. In absence of any autoinducer molecules, the $LasR$ dimer will reach a threshold value of $\frac{V_{0L}}{kdL}$.

The $LasR$ dimer (hereafter referred to as $LasR$) associates readily with the autoinducer molecules, and does so rapidly, such that the $LasR - AHL$ complex exists in a quasi-steady state. Two dimer molecules then bind cooperatively to the $luxR$ promoter region, where again the rates of association / dissociation are fast in comparison to the protein expression. The bound state of the $luxR$ gene is therefore also in a quasi-steady state. The rate limiting process determining the bound state of the $luxR$ gene will then be the diffusion of the autoinducer molecules into the cell, with diffusivity calculated (See Appendix) as $\sigma_A = 0.4s^{-1}$.

Since the intracellular autoinducer concentration will only tend to the external concentration, the time to 90% external concentration, $t = 5.76s$ is calculated (See Appendix) as being representative of near equilibrium. Relative to protein expression, this is clearly relatively fast, and so therefore will be the saturation concentration of C.

The steady state concentration of C can then be deduced (See Appendix) as:

$$C_{ss} = \frac{LasR.A_I}{K_{mC}} \tag{15}$$

Where K_{mC} is the autoinducer complex disassociation constant. Similarly, solving for the steady state of $LasR$ and A_I gives:

$$\therefore LasR_{ss} = \frac{V_{0L} + k_{rC}.C}{k_{dL} + k_{fC}.A_I} \tag{16}$$

$$\therefore A_{Iss} = \frac{k_{rC}.C + A_0.\sigma_A}{k_{fC}.LasR + \sigma_A} \tag{17}$$

At steady state, the intracellular concentrations of the metabolites will tend to constant values, and so the consumption of A_I will also be constant. The autoinducer is approximated to be non-degradable, and so in the case of constant consumption, A_0 will continue to diffuse into the cells until an equilibrium is reached. The only steady-state value is therefore:

$$A_{Iss} = A_0 \tag{18}$$

Substituting Eqs. 18 and 16 into Eq. 17 gives:

$$C_{ss} = \frac{A_0.V_{0L}}{K_{mC}.k_{dL}} \tag{19}$$

As expected, the steady state value of C is independent of n. Rather, the exponent n acts only to sharpen the threshold response, as is observed in the next section.

4.2.2 Target Protein X

The steady-state of the expressed protein X can be deduced by the same method. Solution of eq. 14 gives:

$$X_{ss} = V_X.\frac{1}{\left(\frac{K_{mB}}{C}\right)^n + 1} + X_0 = V_X.\frac{1}{\left(\frac{K_{mB}.K_{mC}.k_{d,L}}{A_0.V_{0L}}\right)^n + 1} + X_0 \tag{20}$$

Where the parameters are combined such that $V_X = \frac{V_X}{k_X}$ and $P_0 = \frac{P_0}{k_P}$. This is now equivalent in form to the function fitted by Saeidi et al. to their observed data. The final step in characterizing the QS system is now to estimate the unknown parameters in eq. 20.

4.3 Stochastic Analysis

The intracellular environment is noisy, and reactions between species cannot be assumed to progress in a purely deterministic fashion. To investigate the effects of this stochasticity on the quorum sensing system, the time-courses can be calculated for each molecular species. Fig. compares both the deterministic and stochastic dynamics of the system, calculated using ODEs and the Gillespie algorithm respectively.

Sensible assumptions were made to reflect the parameters' comparative orders of magnitude. It was assumed that in both binding cases, $k_f \gg k_r$, and that these rates were also significantly faster than the production and degradation rates of the $LasR$ protein.

The initial number of $LasR$ molecules is determined by its steady state value $\frac{V_{0L}}{k_{dL}}$, assuming this to be reached during the overnight incubation stage in the absence of autoinducer molecules. B_0, the conserved number of binding sites $N_p = 200$, both bound and unbound, was estimated according to the plasmid used to transfect the genes into the host[47].

The stochastic dynamics follow the deterministic ones closely. All species saturate to their S-S values quickly, with A_I reaching equilibrium in a similar time to that predicted in the absence of the intracellular dynamics. For the number of molecules in consideration, the deterministic model will thus be sufficient.

5 Characterisation of the Lysis Circuit

Saeidi et al. fitted a Hill function to the protein expression vs. autoinducer response, the parameters of which were determined using the MatLab curve fitting toolbox. Consequently,

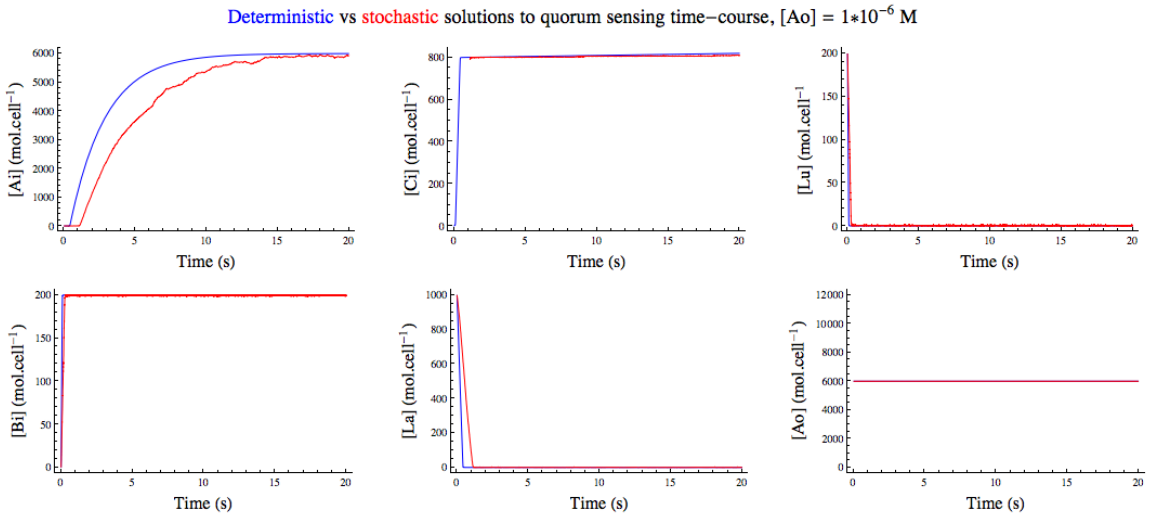


Figure 2: Comparison of stochastic and deterministic dynamics of the quorum sensing system. $N_P = 200$, $\sigma_A = 0.4$, $k_{fC} = 10$, $k_{rC} = 1$, $k_{fB} = 10$, $k_{rB} = 1$, $k_{dL} = 0.001$, $V_0 = 1$.

they use an exponent for $n = 5.78 > 1$, which from a biological perspective would indicate cooperative binding at the $luxR$ promoter site, although to a greater degree than assumed by [40]. The likelihood ratio test can be employed to assess whether the $n > 1$ fit is is significant enough to justify this model over the null hypothesis, which will assume $n = 1$.

5.1 Maximizing the Likelihood

In their supplementary material, Saeidi et al. provide data for the time-courses of protein expression rate for 14 different A_0 ranging from $5 * 10^{-10} - 1 * 10^{-4} M$, however only the first 8 measurements were used, as the data for the last six exhibits uncharacteristic behaviour, discussed in the next section. Saeidi et al. only recorded one data point for each corresponding time, and so the variance σ was optimized with the other parameters in the model.

The variance was optimized along with the model parameters using MLE, according to data from Saeidi et al. Having now established that the concentration of C can be considered as reaching its steady state quickly, eq. 14 can now be solved to give:

$$X(t) = \frac{\left(1 - e^{-K_X.t}\right)}{K_X} * \left(\frac{V_X}{\left(\frac{K_{\overline{m}}}{A_I}\right)^n + 1} + X_0 \right) \tag{21}$$

Where $K_{\overline{m}}$, V_X, X_0 and K_X are pooled constants. To compare the two models ($n = 1$ vs. $n > 1$), both were maximized against the data given by Saeidi et al. using equation 1. In this specific case, it is advantageous to minimize the negative log-likelihood function, as it allows reduction to the simpler and less computationally expensive expression:

$$-log(\mathcal{L}) = Sup \left\{ \frac{log\left(\sigma.\sqrt{2\pi}\right)}{2\sigma^2} + \sum_{i=1}^{N} \left(\mu_{t_i} - X(t_i)\right)^2 \right\} \tag{22}$$

H	-log($\mathcal{L}$)	V_X	K_X	n	K_m	X_0	σ
$H_0\,(n=1)$	3.69	9.75	4.11	1	1.83e-07	0.407	0.0916
$H_1\,(n=2)$	1.87	4.38	2.27	2	1.34e-07	0.390	0.0922
$H_2\,(n>1)$	0.907	2.90	1.69	6.07	1.23e-07	0.403	0.0932

Table 1: Table showing the optimized parameters for each cooperativity hypothesis.

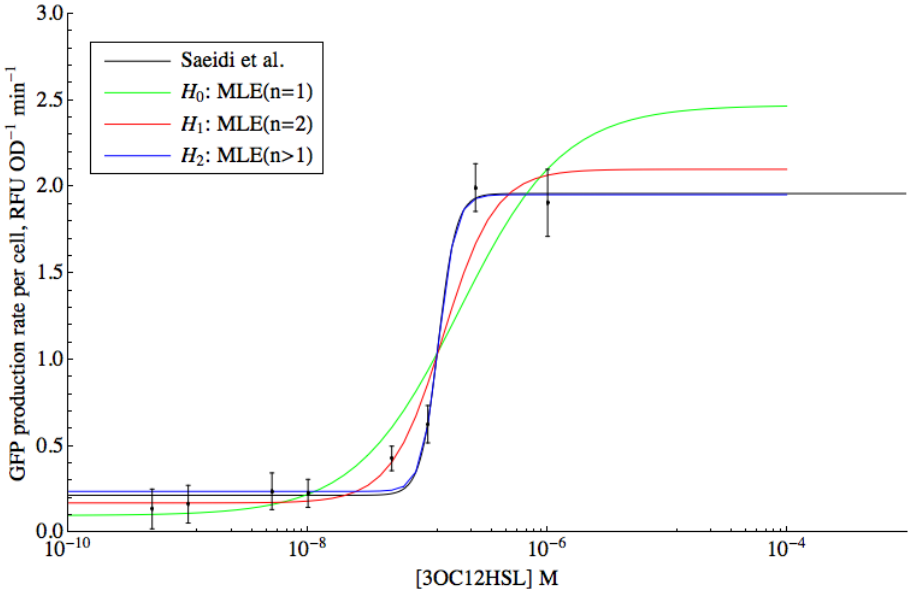


Figure 3: Predicted model for optimized $H_{0,1,2}$ parameters against average protein expression data, shown with error bars from Saeidi et al.

Table 1summarizes the results, where H signifies the hypothesis being tested.

5.2 Likelihood Ratio Testing

Fig. 3 shows the average protein expression rates as compared to the original data presented by Saeidi et al.

The likelihood ratio test relates the best chance of the data occurring under the null hypothesis H_0 to the best chance under no parameter restraints, H. In this case, the null hypothesis $H_0 : n = 1$, given that there is no clear evidence of cooperative binding. The alternative hypotheses H_1 and H_2 (defined previously) allow for the possibility of cooperative binding. The test is defined as:

$$\lambda(x) = -2 * log\left[\frac{sup\{\mathcal{L}(\Theta; x) : \theta \in \Theta_0\}}{sup\{\mathcal{L}(\Theta; x) : \theta \in \Theta\}}\right],\ x \in \mathbb{R}^n_\chi$$

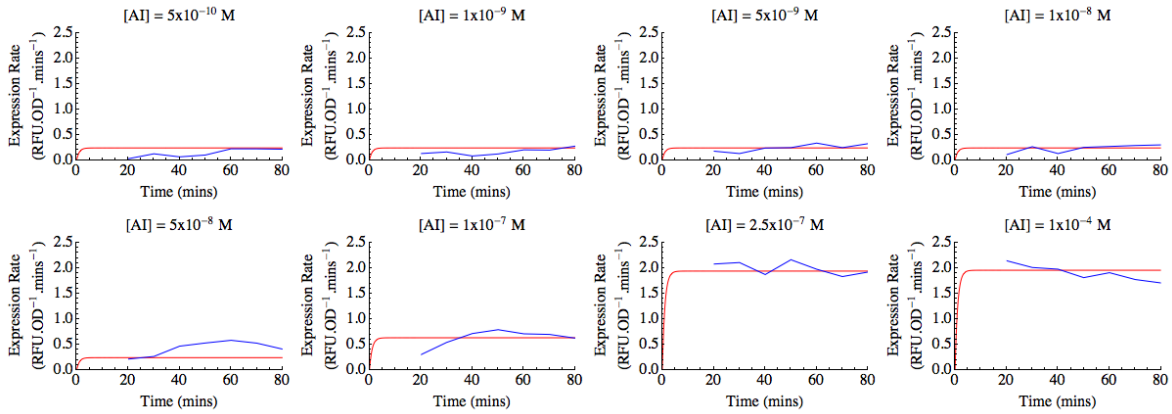


Figure 4: Predicted model for optimized H_2 parameters against observed time-course data.

Giving:

$$\lambda_{01} = 3.66$$

$$\lambda_{02} = 5.586$$

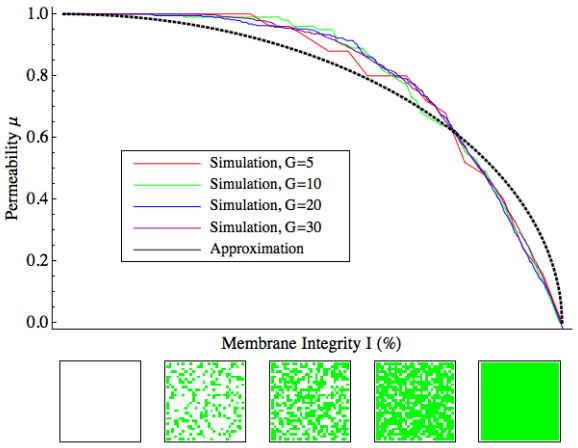
$\lambda_{0.95} = 3.84$, suggesting that only H_2 is statistically significant. However, it is possible that there may be other uncaptured dynamics that can explain this higher than expected level of cooperativity.

Fig. 4 shows the performance of the predicted model against the experimentally observed time-courses.

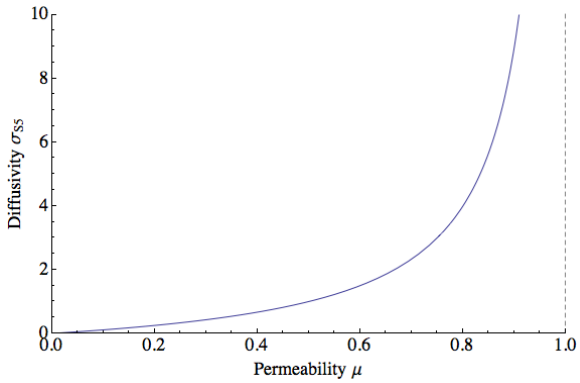
5.3 Membrane Degradation

To compare the model to experimental Pyocin release data, the degradation of the *E. Coli* cells' membranes must me modelled. A simple model for the diffusivity of the membrane $\sigma_{S5}(M)$ might assume a simple linear relationship, however an alternative model is proposed.

Firstly, the membrane is approximated as a discrete grid of interconnected segments M, where one molecule of Lysis reacts destructively with one molecule of M. At full membrane integrity, the Pyocin molecules are too large to diffuse naturally, however when the membrane reacts with the Lysis protein, it becomes permeable, and these regions of permeability are approximated as connected regions of degraded membrane. Fig. 5a shows a stochastic simulation of this process, with a grid size G^2 varying from $G = \{5, 10, 20, 30\}$, assuming Pyocin to be diffusable in a square of $D^2 = 2 \times 2$. Interestingly, the dynamics are seen to be independent of the grid size.



(a) The membrane is degraded by the Lysis protein, forming permeable regions traversable by the Pyocin protein.



(b) Permeability vs. diffusivity.

The permeability μ is defined, and related to the diffusivity as:

$$\mu = \frac{1}{\frac{\sigma_{S5,0}}{\sigma_{S5}} + 1} = \frac{N_D}{(G - D + 2)^2} = \frac{N_D}{N_T} \tag{23}$$

Where N_D represents the number of permeable regions, and the denominator N_T gives the total number of permeable regions possible, given its dimensions and those of the grid. The membrane integrity $I = \frac{M(t)}{M_0}$ represents the proportion of original segments degraded at time t. $\sigma_{S5,0}$ is an empirical coefficient to modulate the base diffusivity.

Whilst the parameter values are not based on empirical data, it is reasonable to assume that the permeable regions would be small in comparison to the membrane. The model also does not represent a time-course, and it should be noted that as the integrity decreases, the probability of any reaction would also decrease exponentially.

Although, the model requires further mathematical analysis, the integrity can be approximated for the purpose of this study by a function of the form:

$$\mu \approx \sqrt{1 - \left(\frac{M(t)}{M_0}\right)^2} = \sqrt{1 - I(t)^2} \tag{24}$$

And the diffusivity is thus given by eq. 23 as:

$$\sigma_A = \frac{\sigma_{S5,0}}{\frac{1}{\mu} - 1} = \frac{\sigma_{S5,0}\sqrt{1 - I(t)^2}}{1 - \sqrt{1 - I(t)^2}} \tag{25}$$

So that when $I = 0$, the membrane is infinitely conductive, leading to instant equilibration between the external and internal chemical concentrations as desired. Fig. 5b shows this relationship.

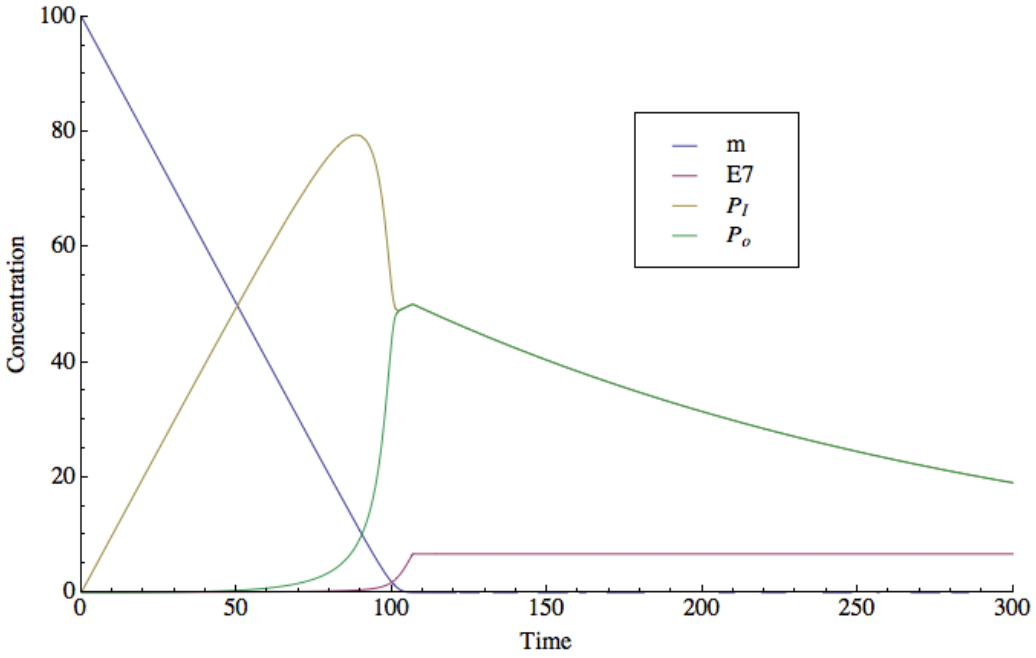


Figure 5: Concentrations (AU) of $P_0(t)$, $P_I(t)$, $E(t)$ and $M(t)$ against time (AU) for a single cell. $k_{S5=0.01}$, $k_{dM} = 0.2$, $\frac{dP}{dt} = \frac{dE}{dt} = 1$, $\sigma_{S5,0} = 0.001$.

5.4 Release of Pyocin

Considering an external autoinducer concentration of $1 * 10^{-6}$, so that the Pyocin / Lysis production is maximal, Fig. 5 demonstrates qualitatively that there exists a delay between internal expression and external pyocin release, as observed by Saeidi et al.

6 Results and Discussion

The accurate modelling of biologically synthesized systems is crucial if these are to be designed efficiently and cost effectively.

In this study, the possibility of cooperative binding of the autoinducer to the $luxR$ promoter was investigated. Whilst a value $n = 2$ of cooperativity did not adhere to a 95% confidence interval given experimental observations, it should be noted that details of the dynamics perhaps not captured by the model might affect this conclusion. The MLE performed in fact showed an optimum cooperativity $n = 6.07$, suggesting that there may indeed be other cooperative mechanisms not accounted for.

The cooperativity formed one element of the proposed model which effectively predicted the expression of protein by the $luxR$ gene. The behaviour of this quorum sensing circuit showed

clear bistability in response to the external autoinducer concentration, showing a switch point at around $A_0 = 1 * 10^{-6} M$.

The second part of the study proposed a novel mechanism of membrane degradation. Whilst a rigorous mathematical analysis remains to be conducted, stochastic simulations showed qualitatively that the diffusivity of the membrane should not increase linearly with decreasing integrity.

The delay between the production and release of intracellular Pyocin observed by Saeidi et al. was also demonstrated qualitatively, and further simulations involving multiple cells should elucidate this behaviour further.

Not addressed in this model was the decrease in protein expression rates observed at high autoinducer concentrations. However, a comparison between Fig. 3A and 4B of Saeidi et al. show that intracellular Pyocin accumulation prevents the growth of the *E. Coli* host, providing a possible mechanism for the explanation of behaviour.

Work remains to quantify the relative rates of protein expression by the $luxR$ gene, converting from the empirical GFP luminescence-based measurements. Additional experimental data is also required for the robust estimation of the model parameters required to evaluate its performance.

7 Appendix

7.1 Simplifying the Protein Expression Model

Beginning with:

$$nC + luxR \underset{k_{rB}}{\overset{k_{fB}}{\rightleftharpoons}} B \overset{k_{fp}}{\rightarrow} B + x \tag{26}$$

Where $B = [nC.luxR]$ and x is the mRNA strand that leads to production of protein X. Equation 26 then leads to:

$$\frac{dC}{dt} = n.\left(B.k_{rB} - C^n.luxR.k_{fB}\right) \tag{27}$$

$$\frac{d\left(luxR\right)}{dt} = B.k_{rB} - C^n.luxR.k_{fB} \tag{28}$$

$$\frac{dB}{dt} = C^n.luxR.k_{fB} - B.k_{rB} \tag{29}$$

$$\frac{dx}{dt} = B.k_{fx} \tag{30}$$

$$\frac{dX}{dt} = k_{fX}.x - k_{d,X}.X \tag{31}$$

Complex B is in an equilibrium state (Briggs-Haldane[42]), $\frac{dB}{dt} = 0$:

$$B = C^n.luxR.\frac{k_{fB}}{k_{rB}} = \frac{C^n.luxR}{K_{mB}} \tag{32}$$

Where $K_{mB} = \frac{k_{rB}}{k_{fB}}$, the binding dissociation constant. The total number of binding sites, bound or unbound is constant, giving the conservation relationship:

$$luxR + B = B_0 \; and \; \therefore B = \frac{C^n.B_0}{K_{mB} + C^n} \tag{33}$$

So the reaction velocity is given by:

$$\therefore \frac{dx}{dt} = \frac{C^n.B_0.k_{fp}}{K_{mB} + C^n} = V_{max} \frac{1}{\left(\frac{K_{\overline{mB}}}{C}\right)^n + 1} \tag{34}$$

Where $V_{max} = B_0.k_{fp}$ and $K_{\overline{mB}} = (K_{mB})\,\frac{1}{n}$ (see Appendix). Therefore, the mRNA coding for X production is seen to be produced via Hill-type kinetics in the presence of the autoinducer complex C. The target protein itself is then produced via first order kinetics in the presence of x-mRNA. A more realistic model incorporates the degradation of the x-mRNA, k_x.

As explained previously, in this case, the role of mRNA can be incorporated into a simplified model. It can be shown (see Appendix) that Eqs. **??** and **??** , when including the effects of a basal rate of protein production X_0, reduce to:

$$\frac{dX}{dt} = V_X.\frac{1}{\left(\frac{K_{\overline{mB}}}{C}\right)^n + 1} - k_X.X + X_0 \tag{35}$$

7.2 Autoinducer Diffusivity

E. Coli cells have a typical length and radius of approximately $2\mu m$ and $0.3\mu m$ respectively[44], giving a surface area of $Surf_c = 4.34 * 10^{-12} m^2$ and a volume of $Vol_C = 5.65 * 10^{-10} L$. The permeability of the wall to the autoinducer molecules is approximately $Perm = 5 * 10^{-6} cm.s^{-1}$[45], and so an estimate of the autoinducer diffusivity is:

$$\sigma_A = \frac{Perm * Surf_C}{Vol_C} \approx 0.4 s^{-1} \tag{36}$$

The time taken for the autoinducer to diffuse into the cell membrane can be calculated from the solution to the diffusion equation across the membrane:

$$A_I = A_0 \left(1 - e^{-\sigma_A t}\right)$$

7.3 Solving for Steady States

The steady state concentration of C can then be deduced by setting the differential of C to zero:

$$\frac{dC}{dt} = 0 = B.k_{rB} + LasR.A_I.k_{fC} - C^n.luxR.k_{fB} - k_{rC}.C \tag{37}$$

Upon substitution of Eq. 32 into 37, the n dependent behaviour disappears, leaving:

$$C_{ss} = \frac{A_0.V_{0L}}{K_{mC}.k_{dL}} \tag{38}$$

References

[1] Slusarczyk A.L., Weiss, R. (2012). Foundations for the design and implementation of synthetic gene circuits.

[2] Saeidi, N., Wong,, C.K. (2011). Engineering microbes to sense and eradicate Pseudomonas Aeruginosa, a human pathogen.

[3] Chang W., Small D.A., Toghrol F., Bentley W.E. (2005). Microarray analysis of toxicogenomic effects of peracetic acid on Pseudomonas aeruginosa.

[4] Hwang, I.Y., Tan, M.H., Koh, E., Ho, C.L., Poh, C.L., & Chang, M.W. (2013). Reprogramming Microbes to Be Pathogen-Seeking Killers.

[5] Govan, J.R., Deretic, V. (1996). Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia.

[6] www.biobricks.org

[7] Arber, W., Linn, S. (1969). DNA Modification and Restriction.

[8] Mullis, K., Faloona, F. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.

[9] Chalfie, M., Tu, Y. (1994). Green fluorescent protein as a marker for gene expression.

[10] Arkin, A., Ross, J. (1994). Computational functions in biochemical reaction networks.

[11] Gardner, T.S., Cantor, C.R., Collins, J.J. (2000). Construction of a genetic toggle switch in Escherichia coli.

[12] Ajo-Franklin, C.M., Drubin, D.A. et al. (2007). Rational design of memory in eukaryotic cells.

[13] Rinaudo, K., Bleris, L., Maddamsetti, R., Subramanian, S., Weiss, R., Benenson, Y. (2007). A universal RNAi-based logic evaluator that operates in mammalian cells.

[14] Auslander, S., Auslander, D., Muller, M., Wieland, M., Fussenegger, M. (2012). Programmable single-cell mammalian biocomputers.

[15] Tamsir, A., Tabor, J.J., Voigt, C.A. (2011). Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'.

[16] Bell-Pedersen, D., Cassone, V.M. , Earnest, D.J. et al. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms.

[17] Leloup, J.C., Goldbeter, A. (1998). A model for circadian rhythms in Drosophila incorporating the formation of a complex between the PER and TIM proteins.

[18] Tyson, J.J., Hong, C.I., Thron, C.D., Novak, B. (1999). A simple model of circadian rhythms based on dimerization and proteolysis.

[19] Elowitz, M.B., Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators.

[20] Sohka, T. et al (2009). An externally tunable bacterial band-pass filter.

[21] Ye, H., Daoud-El Baba, M., Peng, R.W. & Fussenegger, M. (2011). A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice.

[22] Yamaguchi, M., Ito, A., Ono, A., Kawabe, Y, Kamhihira, M. (2013). Heat-Inducible Gene Expression System by Applying Alternating Magnetic Field to Magnetic Nanoparticles.

[23] Anesiadis, N., Cluett, W.R., Mahadevan, R. (2008). Dynamic metabolic engineering for increasing bioprocess productivity.

[24] Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S.B., Keasling, J.D. (2010). Microbial production of fatty- acid-derived fuels and chemicals from plant biomass.

[25] Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., Benenson, Y. (2011). Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells.

[26] Zhu, J., Mekalanos, J.J. (2003). Quorum sensing-dependent biofilms enhance colonization in Vibrio cholerae.

[27] Duan, F., March, J.C. (2010). Engineered bacterial communication prevents Vivrio Cholerae virulence in an infant mouse model.

[28] West, S.A., Pen, I. & Griffin, A.S. (2002). Conflict and cooperation—cooperation and competition between relatives.

[29] Sachs, J.L., Mueller, U.G., Wilcox, T.P. & Bull, J. J. (2004). The evolution of cooperation.

[30] Givskov, M., DeNys, R., Manefield, M., Gram, L., Maximilien, R., Eberl, L., Molin, S., Steinberg, P.D. & Kjelleberg, S. (1996). Eukaryotic interference with homo- serine lactone-mediated prokaryotic signaling.

[31] Va Delden, C., Iglewski, B.H. (1998). Cell-to-Cell Signaling and Pseudomonas aeruginosa Infections.

[32] Dockery, J.D., Keener, J.P. (2000). A Mathematical Model for Quorum Sensing in Pseudomonas Aeruginosa.

[33] Chang, W., Small, D. A., Toghrol, F., Bentley, W. E. (2005). Microarray analysis of Pseudomonas aeruginosa reveals induction of pyocin genes in response to hydrogen peroxide.

[34] Pflughoeft, K.J., Versalovich, J. (2012). Human Microbiome in Health and Disease.

[35] Ling, H., Saeidi, N., Rasouliha, B.H., Chang, M.W. (2010). A predicted S-type pyocin shows a bactericidal activity against clinical Pseudomonas aeruginosa isolates through membrane damage.

[36] Uratani, Y., Hoshino, T. (1984). Pyocin Ri Inhibits Active Transport in Pseudomonas aeruginosa and Depolarizes Membrane Potential.

[37] http://parts.igem.org/Part:BBa_K117000

[38] Pearson, J.P.,Van Delden, C.,Iglewski, B.H. (1999). Active Efflux and Diffusion Are Involved in Transport of Pseudomonas aeruginosa Cell-to-Cell Signals.

[39] Claussen, A., Jakobsen, T.H., Bjarnsholt, T., Givskov, M., Welch, M., Ferkinghoff-Borg, J., Sams, T. (2013). Kinetic Model for Signal Binding to the Quorum Sensing Regulator LasR.

[40] Hunter, G.A.M., Vasquez, F.G., Keener, J.P. (2013). A mathematical model and quantitative comparison of the small RNA circuit in the Vibrio harveyi and Vibrio cholerae quorum sensing systems.

[41] Reversible Acyl-Homoserine Lactone Binding to Purified Vibrio fischeri LuxR Protein Urbanowski, M.L., Lostroh, C.P., Greenberg, E.P. (2004). J. Bacteriol., 186(3):631.

[42] Briggs, G. E., and Haldane, J. B. (1925). A Note on the Kinetics of Enzyme Action.

[43] Michaelis, L., Menten, M. (1913). Die kinetik der invertinwirkung.

[44] Kaplan, H.B., Greenberg, E.P. (1985). Diffusion of Autoinducer Is Involved in Regulation of the Vibrio fischeri Luminescence System.

[45] Rubinow, S.I. (2002). Introduction to Mathematical Biology.

[46] http://www.genomics.agilent.com/biocalculators/calcODBacterial.jsp?_requestid=425426

[47] http://parts.igem.org/Part:pSB1C3

[48] Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evo- lution of coupled chemical reactions.

[49] Wilkinon, D.J. Stochastic Modelling for Systems Biology.