# Abstract

# 1 Introduction

Microorganism have been in use in industrial setting for decades to acquire therapeutics, biofuel, chemicals, materials or destroy toxins or waste. Moreover, lately the ability to engineer microorganisms’ genetic networks has led to production of microorganisms that can carry out a synthetically designed task such as pattern formation (<https://www.nature.com/articles/nature03461> ), coupled oscillation (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2838179/> ), population control (<https://www.nature.com/articles/nature02491> ) just to name a few.

The majority of industry at the moment centres on the use of fermentation and thus the microbial strains have been designed for it. Though many improvements have been made to the process, fermentation is still costly and laborious as the conditions in which the microorganisms grow must be closely regulated and maintained (<https://link.springer.com/article/10.1007%2Fs00018-012-0945-1> ). Furthermore, the genetic manipulations required to obtain optimal cellular metabolism and relevant fermentation processes also remain time intensive and complex as interaction between gene regulations, signalling and metabolic processes need to be taken into account. Moreover, the strain development is in large case-specific, meaning high costs and extended time periods to develop each strain for a specific compound production (<https://www.nature.com/articles/nbt.3365.pdf> ). Some other problems have also been recorded mainly in connection with scaling up academic projects, maintaining strain health, performance and optimal conditions (<https://www.nature.com/articles/nbt.3365.pdf> ). To monitor the broth conditions the industry standard is an exogenous sensory system (<https://www.tandfonline.com/doi/abs/10.1080/02648725.1989.10647858> ). Therefore, different approaches of metabolic engineering and implementation have been considered.

## 1.1 Feedback loops as a possible approach

One of such approaches is using synthetically implemented feedback loops in microorganisms to control individual cell and global behaviour with the intent to maintain favourable conditions that would result in optimal production of a specific compound. Feedback loops are naturally occurring in cells and consists of different substances where the concentration of one effects the concentration of another thus resulting in altered metabolism. Such loops can be positive – strengthens the change – or negative – reverses the change – and control a wide range of cell functions such as homeostasis (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2923299/> ). Homeostasis is the process by which the cells retain a stable intracellular environment that allows them to proliferate despite changes in extracellular environment (<https://www.hindawi.com/journals/isrn/2013/645983/> ). It is one of the more important physiological processes that must be taken into account when using microorganisms as part of industrial compound production as too vigorous of production can exceed the cell’s metabolic load capacity (<https://www.sciencedirect.com/science/article/pii/073497509500004A?via%3Dihub> ) or the product itself could be toxic for the cell thus resulting in sub-optimal compound production and microbial death (<https://link.springer.com/article/10.1007%2Fs10482-016-0781-7> ). Exogenous feedback loops where the controller is a computer that control gene expression in microorganisms have been proposed before (<https://pubs.acs.org/doi/full/10.1021/acssynbio.5b00141?src=recsys> , <https://pubs.acs.org/doi/abs/10.1021/acssynbio.5b00135> , <https://www.nature.com/articles/nchembio.1534> , <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4022480/> ). They are more suitable for cases where the cells require minimal genetic modification and for cases where reprogramming or differentiation are the main processes to be controlled (<http://rsif.royalsocietypublishing.org/content/royinterface/13/120/20160380.full.pdf> ). So far, a pressing challenge in synthetic biology is implementation of an endogenous feedback loop which would not require an external sensing mechanisms and computer algorithms to maintain favourable conditions (Fiore et al., 2016).

## 1.2 Implementation of feedback loops in single vs multiple cells

Single cell implementations of feedback loops have been proposed previously. The strategies to implement control over gene expression usually relay on synthetic promoters (<https://www.ncbi.nlm.nih.gov/pubmed/10802621> ) and toggle switches (<https://www.ncbi.nlm.nih.gov/pubmed/15159530> ). The use of one cell systems have faced problems relating to the complexity of such networks and the difficulty in implementing them. Thus, the circuits are very case-specific and can not be adapted easily resulting in costly and time consuming reengineering required. Furthermore, many of the parts are incompatible especially when put in one host cell (<https://www.nature.com/news/2010/100120/pdf/463288a.pdf> ). Moreover, the implementation of a foreign circuit can lead to additional metabolic burden put on a cell which in turn can leads to suboptimal production or even cell death (<http://rstb.royalsocietypublishing.org/content/royptb/371/1701/20160175.full.pdf> ). Single cell feedback control is more suitable when each cell needs to function as an autonomous machine. For example, in medicine for drug delivery (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511399/> ) or disease cell targeting (<http://rsif.royalsocietypublishing.org/content/royinterface/13/120/20160380.full.pdf> ).

Therefore, to overcome the difficulties and for applications where spatial patterns are important multicellular populations that are interconnected by cell-cell communication have been studied (<http://rsif.royalsocietypublishing.org/content/royinterface/13/120/20160380.full.pdf> ). The idea of multiple different cell environment is drawn from nature where coexistence of multiple species and exchange of molecules between them exists (<https://www.nature.com/articles/nrmicro2832> ). Furthermore, bacteria, as a species on which most gene engineering is conducted, exchanges diverse number of molecules with its surroundings to sense the environment and communicate with other cells (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4354409/> ). Furthermore, E. coli in naturally implement a relationship with surrounding cells where some produce compounds that others consume (<https://link.springer.com/article/10.1007%2Fs00248-007-9357-4> ). Consortia-type system scheme is appealing as several of the problems that arise with a single cell genetic circuit can be avoided. For example, resource competition and retroactivity especially in applications dealing with biosynthesis (<https://www.annualreviews.org/doi/pdf/10.1146/annurev-control-060117-105052> ). The use of multiple populations has been shown to increase yield and productivity while decreasing the burden on a single cell (<http://www.cheme.caltech.edu/groups/fha/publications/brenner.pdf> <https://scholarworks.montana.edu/xmlui/bitstream/handle/1/12551/12-001_Synthetic_Escherichia_coli_A1b.pdf?sequence=1> ).

Splitting functional modules between multiple cells can increase system adaptability and modularity as potentially only once cell type must be reengineered (<https://www.annualreviews.org/doi/pdf/10.1146/annurev-control-060117-105052> ). Furthermore, individual design and consequent optimization could lead to simpler construction and wider use (<https://www.nature.com/articles/469171a.pdf> ). Moreover, conceptually the number of components that would need to be implemented in a single cell is lower with multiple different populations than with one cell thus the disruption to the normal physiology would be minimized (<https://www.nature.com/articles/469171a.pdf> ). Thus it is easier to reduce the metabolic burden and maintain cell fitness. Co-culture of multiple strains also results in a more diverse environment which can aid different gene expression pathways (<https://www.sciencedirect.com/science/article/pii/S1096717616300180> ). Consortium has also been shown to be more stable with changing environment (<http://www.cheme.caltech.edu/groups/fha/publications/brenner.pdf> ). As pathways are split, products or by-products of one process that would be detrimental to the process can now be located in two different cells (<https://www.sciencedirect.com/science/article/pii/S1096717616300180> ). This has been shown in fuel desulfurization study (<https://www.sciencedirect.com/science/article/pii/S1096717616000069?via%3Dihub> ).

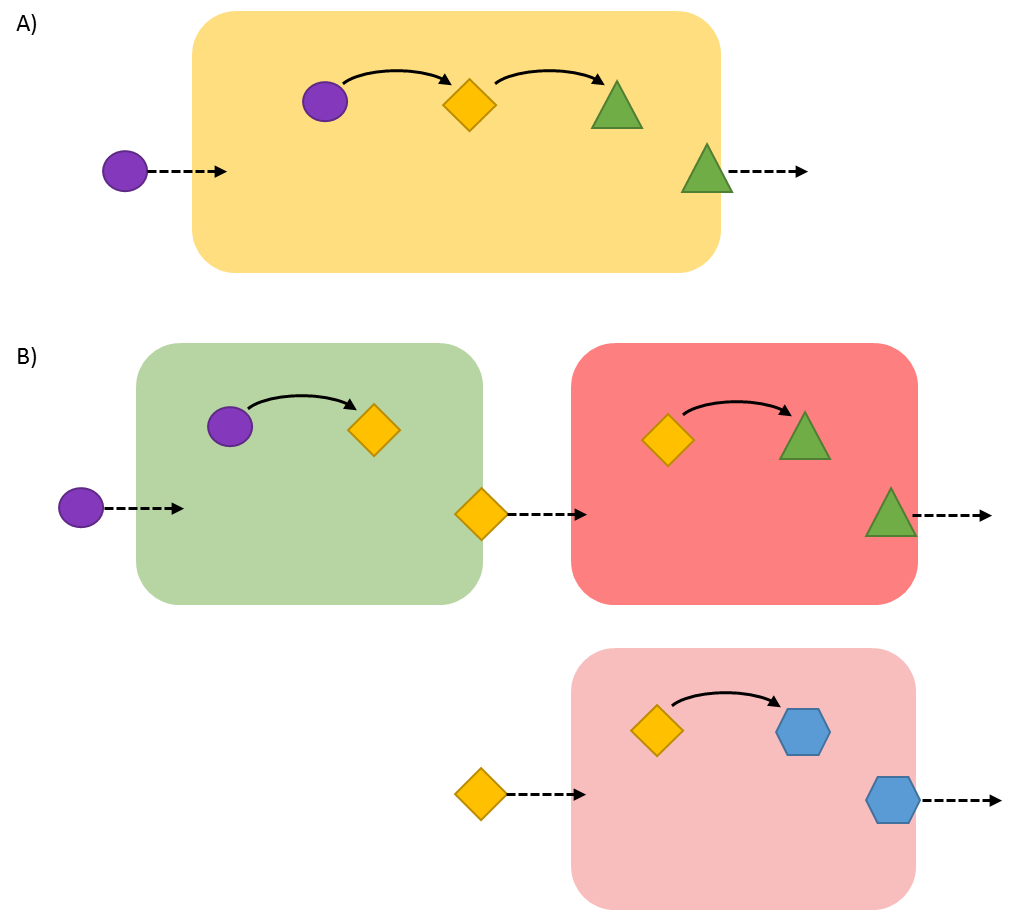
## 1.3 Successful implementations of multi-cellular systems

Effective implementation has been shown in density dependent gene activation (<https://pubs.acs.org/doi/10.1021/sb5002533> ), spatial patterning (<https://www.sciencedirect.com/science/article/pii/S0092867416304081?via%3Dihub> , <https://www.ncbi.nlm.nih.gov/pubmed/15858574> ) and light-dark edge detection (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2775486/> ). Furthermore, studies have been conducted on the implementation of control in microbial consortia to increase overall product’s yield (<http://www.cheme.caltech.edu/groups/fha/publications/brenner.pdf> , <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7798771> ). In multicellular control system each cell has an inner cell feedback loop that controls gene expression but its activation or repression depends on the ensemble state of the whole population which is sensed through diffusible signalling molecules (<http://rsif.royalsocietypublishing.org/content/royinterface/13/120/20160380.full.pdf> ). Consortium is a heterogeneous mixture of two or more microbial subpopulations. As the subpopulations can communicate with each other to control inner-cell expression and the function of producing a substance is spread amongst them the burden on one particular cell is decreased (<https://pubs.acs.org/doi/abs/10.1021/acssynbio.6b00220> ). The classic control scheme in consortium is a sender-receiver type, where the sender is the cell population that sends the signal based on input from the reference signal and the output of the receiver. The receiver is the cell that produces the desired chemical (<https://www.sciencedirect.com/science/article/pii/S0167779912000388?via%3Dihub> ) (Figure consortia).

Recently, it has been shown by constructing a synthetic transcriptional oscillator which was implemented in a microbial consortium consisting of two distinct E. coli populations that through communication and interaction a complex population-level behaviour can be achieved and controlled. The consortium based oscillation showed to be more robust at maintaining the behaviour which points towards efficiency of such systems (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4597888/pdf/nihms-726415.pdf> ). An example of significantly increased ethanol production in co-culture versus mono-culture was presented by Shin et al. (2010 <http://aem.asm.org/content/76/24/8150.full.pdf> ) where the conversion of hemicellulose to ethanol was split between two distinct E. coli strains. The ability of consortium to decrease metabolic burden on a single cell and in turn improve titer and yield was shown through experiments on muconic acid production in E. coli (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4500268/> ). Furthermore, coupling E. coli with other organisms has also shown improved efficiency and increase in yield as two different species are biodiversed and can have different bioconcersion and biosynthesis capabilities. For example, coupling E. coli with yeast – Saccharomyces cerevisiae – has shown improved production of oxygenated isoprenoids (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4867547/pdf/nihms641842.pdf> ). The improvement was mainly achieved as the reactive oxygenated species were unable to inhibit activation of the other pathway due to the segregation. The superiority of co-culture strategy versus mono-culture approach was shown recently for conversion of glycerol to muconic acid (<https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-015-0319-0> ). The study highlighted the split of a biosynthetic pathway and hence lessening the metabolic burden as the main reason for improved biopruduction. However, they also noted that controlling the population is a major challenge and downfall of the approach. Accompanying wet lab experiments, also mathematical and computational models have been designed to better understand the co-culture behaviour and the possible strategies for implementation (<https://www.sciencedirect.com/science/article/pii/S1096717616300180> ). Kerner et al. (2012 <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0034032> ) evaluated culture growth rate and how it can be regulated efficiently to maintain optimal population size and strain ratios. Mathematical models have also been used to analyse construction of gene regulation and signalling mechanisms (Fiore et al., 2016, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4597888/> ).

## 1.4 Challenges of multi-cellular implementation

Despite theoretical and experimental research technical challenges with the implementation of multi-microbial circuits still remain (<https://www.annualreviews.org/doi/pdf/10.1146/annurev-control-060117-105052> ). Problems arise as the behaviour of the system is not well understood especially with changing conditions. The computer models have shown that the system is possible (Fiore et al., 2016) but a deeper understanding of how the *in-silico* design would fare in the real cells is needed. Furthermore, problems due to communication delays – time before the signal reaches the cell and transmission occurs – and can cause cell instability must be explored further (<http://www.cds.caltech.edu/~murray/books/AM05/pdf/am08-complete_22Feb09.pdf> ). For some systems the transporters for intermediate products must be engineered (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4500268/> ) or their number synthetically increased (<http://aem.asm.org/content/76/24/8150.full.pdf> ). The characteristics of intermediates must also be taken into account as some – e.g. phosphorylated metabolites, CoA species – have limited mobility (<https://www.sciencedirect.com/science/article/pii/S1096717616300180> ). Furthermore, problems due to the fact that signal strength depends on special distribution and population size. Thus strict population control must be employed to maintain appropriate signal strength (<https://www.annualreviews.org/doi/pdf/10.1146/annurev-control-060117-105052> ). However, population control has been shown to be a major problem of the approach especially due to growth resource competition between strains (<https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-015-0319-0> ). There are at the moment no well-established techniques for population control and composition manipulation (<https://www.sciencedirect.com/science/article/pii/S1096717616300180> ). To overcome the challenge of engineering a long-term, stable consortium production system more studies are needed (<http://www.cheme.caltech.edu/groups/fha/publications/brenner.pdf> ).



### **Figure one vs multiple: Cartoon representation of a system implemented in a single cell versus two cells**. **A)** A system implemented in a single cell where both reactions occur in one cell to synthesise a product. **B)** The two pathways are split into two cells (green and red cells) and are connected through a diffusing intermediate molecule (yellow diamond). The two red cells (darker red and lighter red) show how such consortium system can be adaptable as the same green cell could potentially in two different processes control production of two different outputs. (<https://www.sciencedirect.com/science/article/pii/S1096717616300180> )



### **Figure consortia: Illustration of consortium and its regulation and output**. There are two distinct cell populations: the green and the red cells. An exogenous signal called Reference in absorbed by the green population of cells which can be called the Controller cells. The Controller cells internalise the reference signal and in accordance with it produce an output (black arrow) which is absorbed by the red cells called Target cells. The Target cells produce a product or a compound. The compound’s concentration is sensed by the Controllers. The output readout from Target cells can be used to evaluate the performance and fine-tune the system for optimal yield. The graph shows the Reference concentration (black line) and product concentration (green line) over time.

(<https://pubs.acs.org/doi/abs/10.1021/acssynbio.6b00220> )

## 1.5 Purpose and aims of the project

This project looked at a computer model of a specific cell-to-cell controller-receiver feedback system in E. coli that was first described by Fiore et al. (2016). The aim of this project was to, firstly, establish if such a system can be implemented using the computer language gro. Furthermore, by comparing it to outputs obtained by Fiore et al. (2016) the accuracy of the model can be evaluated. Secondly, the implemented model was used to establish robustness of the overall control of the production of molecules and behaviour by varying key parameters in the simplest scenario, the one in which two cells, a controller and a target interact. Furthermore, by testing the model the hypothesis – the system can maintain the output and overall behaviour despite variations in parameters – can be evaluated. The study of how robust the controller system is, is also the novelty of the project. The evaluation of robustness and sensitivity to parameter variations is a valid addition to the overall understanding of such a system and a step forward to understanding the systems behaviour and reliability in real cells where conditions change more frequently and can not be monitored as closely as in computer models. Moreover, the project is also unique as the control system is implemented in a different computer language – gro – which has not been done before.

# 2 Analysis of the Fiore et al. (2016) model

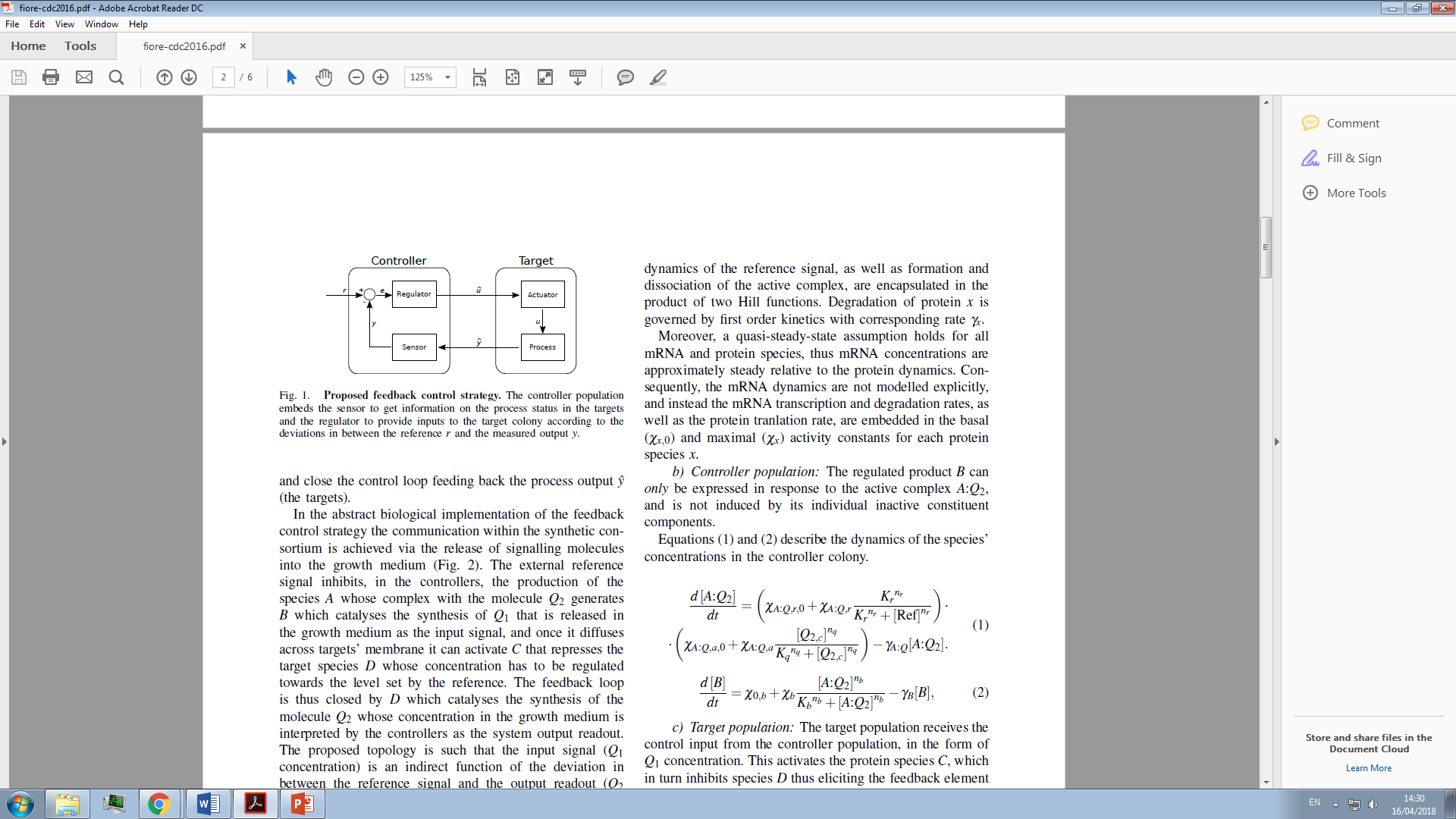
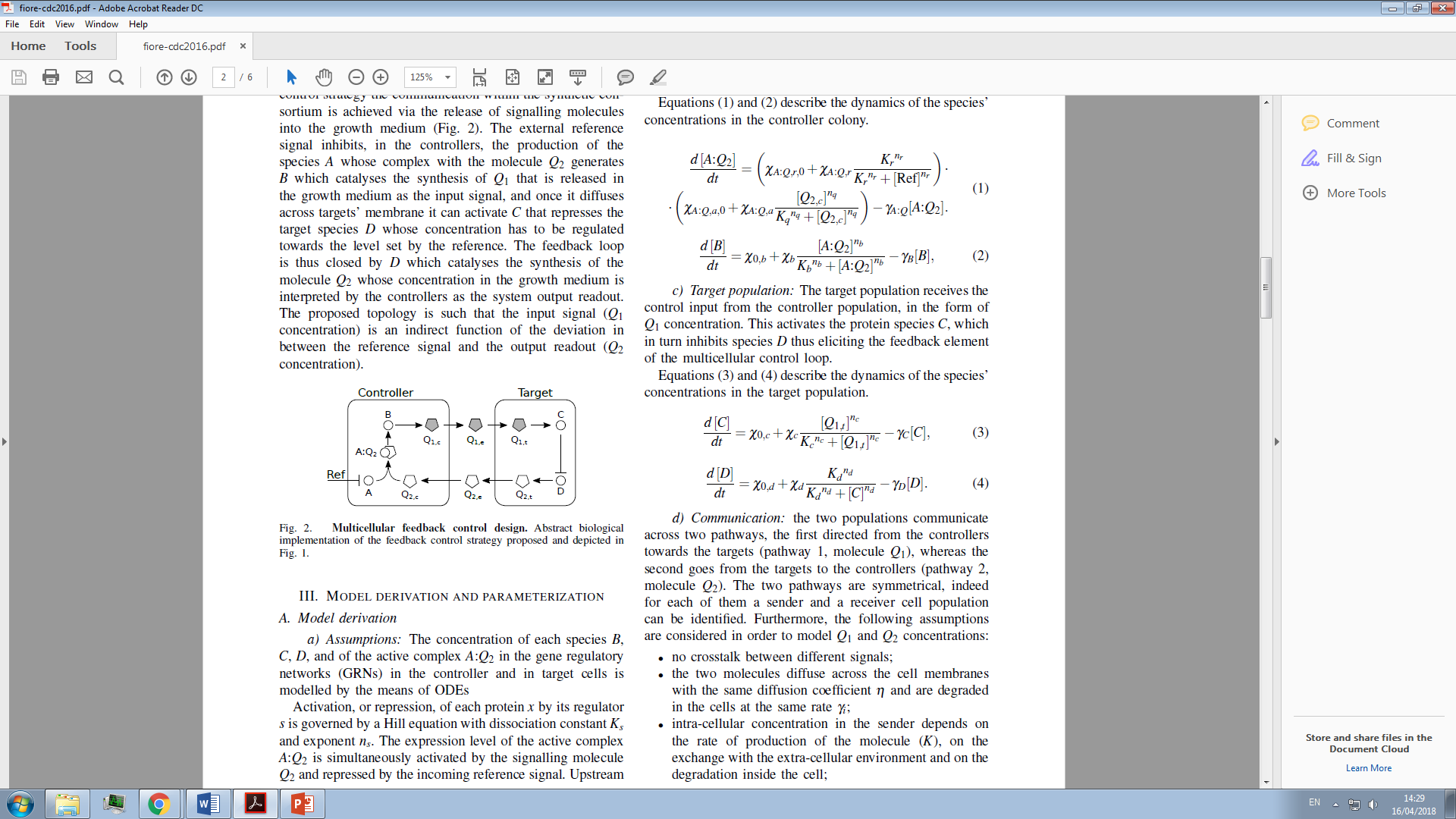
Recently, Fiore et al. (2016) – henceforth Fiore – presented a computer model of a bacterial consortium to assess an endogenous feedback loop control system for compound production and to produce an in-silico proof of concept evaluation. The motivation behind their work was to explore the possibility of simplifying the circuit complexity that is essential in controlling cell behaviour by spreading the necessary functions across two interacting cell populations. They modelled two cells: (1) Controller and (2) Target cell, that are capable of cross membrane communication via signalling molecules.

The Controller’s role is to monitor and as its name suggests control gene expression and in turn substance production in Target cell. Figure Fiore et al shows the circuit in detail. The Controller receives an external inducer signal – Reference – which inhibits one of the molecule – A – in Controller and in turn sets the desired level of output from Target cell. The pathway in the Controller cell consists of incorporating the Reference as well as the process output – Q2 – from the Target cell. The latter is a system output readout and is important in sensing the production levels as the product – D – catalyses its production. To communicate the production level desired and to control the synthesis the Controller releases a molecule – Q1 – into the medium. Q1 is an input control signal. The Target is able to absorb this molecule and incorporates it into production of intermediate – C – that inhibits production of desired compound – D. The concentration of D is regulated towards the level established by the Reference. The deviation between the Reference and the output value is used to provide input to Target and to control production of D. The system forms a closed feedback control loop with a symmetrical communication between the cells.

The regulatory networks of gene expression and the concentrations of molecules are represented by ordinary differential equations (ODEs) (Table Fiore ODEs) where the activation or repression of a specific compound is implemented by Hill equations. ODEs are a standard way of describing the dynamics of network regulation over time (<https://link.springer.com/protocol/10.1007%2F978-1-61779-400-1_12> ). The Hill number is set to 2 as it is expected that the response is ultrasensitive and follows a sigmoidal shape. The values of the dissociation constant take varying values depending on the compound and have been selected form the available literature. Hill function is an appropriate modelling technique for describing gene transcription regulations and has been widely used (<https://www.mmnp-journal.org/articles/mmnp/abs/2008/02/mmnp2008204/mmnp2008204.html> , <https://www.sciencedirect.com/science/article/pii/S1007570417301119> ). Degradation of each of the protein is modelled by first order kinetics and degradation rate. The equations for Q1 and Q2 take into account the position of the protein. Thus each protein is represented as three species: intracellular in Controller or Target and extracellular protein. Hence, three separate equations to account for diffusion, absorption and degradation in intracellular and extracellular environment were used. The extracellular behaviour of Q1 and Q2 are modelled by partial differential equations that represent linear diffusion reactions (Table Fiore ODEs: equations 7 and 10). The mRNA translation, transcription and degradation are not modelled explicitly but are instead embedded in the equation by setting a basal and maximal activity constant for each compound. Thus, a quasi-steady-state assumption holds for the model. The values for all parameters were set according to available literature (Table Fiore Parameters).

Fiore carried out two different scenarios: (1) cell-to-cell and (2) multi-cellular scenario. By testing the model with two cells using MATLAB, Fiore evaluated the overall behaviour of the system, carried out parameter tuning experiments and tested the change in output concentration of D with a set value of Reference signal when majority of the parameters are changed simultaneously. They found that the relationship between Reference and output (D) is dynamic and non-linear and can be used to choose the Reference values to achieve desired output level. The predicted performance with changing Reference values was closely followed by the experimental data. Furthermore, despite changes in parameter values the system reaches a steady state. Moreover, by tuning the parameters, an increase in species B dissociation constant and a decrease in Q1 intracellular degradation results in steady-state being achieved faster. They also ran multi-cellular scenarios where two populations of Controller and Target cells were evaluated. Hence the population effect on the output dynamic range was tested with results that support previous claims that tight regulation on population size and ratio is required. To model the consortium BSim framework was used which considers the spatial distribution of cells and diffusing molecules (<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0042790> ).

Fiore study is a preliminary proof-of-concept study that can be used as a base for further studies. They showed with a use of a mathematical model that separating the parts of the circuit is an effective and robust way to implement the feedback control circuit. Especially studies utilising computer programs that take into account spatial distribution.

### **Figure Fiore et al (2016): Illustration of the computer model.** The Controller imbeds the Regulator pathway (molecules: A, A complex with Q2 – A:Q2 – and B) which produces the signalling molecule – Q1 – that diffuses through the medium and is absorbed by Target cell. Reference signal (Ref or r) which is an external inducer, inhibit this pathway in the Controller to set the level of production of output – D – in Target cell. After Target cell absorbs Q1 the Actuator bit of the loop occurs where Q1 activates production of C. C in turn can inhibit production of the output – D. For the Controller cell to be able to sense the production levels, D catalyses the production of Q2 signalling molecule which after diffusion is absorbed by Controller. Hence the Controller can interpret the production levels in Target cell.

D – process to be controlled; C – actuator, steers the process towards the reference value; Q1 – signalling molecule, controls the signal, indirect function of the difference between reference (Ref) and output (Q2); Q2 – signalling molecule, process output readout, its concentration is interpreted by the Controller; subscript e – growth medium, outside the cells; subscript c and t – in Controller and Target cell, respectively. ”

### Table Fiore ODEs: Set of ODEs and PDEs (equation number 7 and 10) that were used in Fiore et al. (2016) model. Adapted from: Fiore et al. (2016)

|  |  |  |
| --- | --- | --- |
| **No** | **Mol.** | **Equation** |
| 1 | A:Q2 |  |
| 2 | B |  |
| 3 | C |  |
| 4 | D |  |
| 5 | Q1,c |  |
| 6 | Q1,t |  |
| 7 | Q1,e |  |
| 8 | Q2,c |  |
| 9 | Q2,t |  |
| 10 | Q2,e |  |

No = number, Mol. = molecule.

### Table Fiore Parameter values: The parameters used to implement the control system. Adapted from: Fiore et al. (2016)

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Value** | **Description** |
|  | 0.1 μM/min | Baseline production of protein x |
|  | 2 μM/min | Maximal production of protein x |
|  | 1.4 min-1 | Degradation of protein x |
|  | 2 | Hill coefficient for a = r, q, b, c, d |
|  | 1 μM | Dissociation constants |
|  | 0.1 μM |
|  | 0.5 μM |
|  | 0.015 μM |
|  | 0.5 μM |
|  | 0.05 min-1 | Synthesis rate of Q1 and Q2 |
|  | 2 min-1 | Cell wall diffusion rate of Q1 and Q2 |
|  | 0.4 min-1 | Internal degradation of Q1 and Q2 |
|  | 0.2 min-1 | External degradation of Q1 and Q2 |
| Θ | 800 μm2 sec-1 | External diffusion rate of Q1 and Q2 |

How control and circuits work inside the cell is not well known and still poses a problem in synthetic biology especially when it comes to characterization of components. Thus engineering feedback control and testing them with computer models has been shown to be useful especially as the preliminary tool (<https://www.nature.com/news/2010/100120/pdf/463288a.pdf> ). Mathematical modelling can help us better understand the circuit topology – how inside the regulatory circuit the components affect and regulate each other – and its role in controlling cell as well as population-based behaviour (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4597888/pdf/nihms-726415.pdf> ). Furthermore, computer models are easier to manipulate and can be used to simulate different scenarios to establish stability, robustness and performance of a system (<http://rsif.royalsocietypublishing.org/content/royinterface/13/120/20160380.full.pdf> ). Furthermore, they can highlight possible problems and give inside into the implementation in living cells. Simulations have been found to be reliable forecasters of behaviours of systems in living cells when parameter values closely resemble biologically realistic values (<https://www.ncbi.nlm.nih.gov/pubmed/22516742> ). Implementing synthetic circuits in living microbes is still a costly, time-consuming and complicated process (<https://www.ncbi.nlm.nih.gov/pubmed/26463592> ).

# 3 Methods and materials

## 3.1 Python: preliminary test

The model was first evaluated using Python as a preliminary tool to ensure the equations from Fiore paper can reproduce the behaviour of the system presented in their paper.

To simplify the implementation, only ODEs were used and thus only 2 species of Q1 and Q2; inside Controller (Q1,c, Q2,c) and inside Target (Q2,t, Q1,t) cell, were modelled. The equations that differ from the original Fiore equation set (Table Fiore ODEs) are shown in Table Eq-python. The resulting model hence did not take into account the signal diffusion but instead modelled the system as it would be if signalling was instant. Despite the change, it is still a valid preliminary model as the overall behaviour of the system can still be evaluated and a comparison to the Fiore output can be made thus ensuring their model does work in the way it is described.

To implement the ODEs the SciPy function odenint() was used which is capable of integrating systems of first-order ODEs (<https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.odeint.html> ). The system was used to recreate the cell-to-cell scenario of multi-step function where the concentration of Reference (Ref) signal changes in step like fashion from 0.0 to 1.5 and 3.0 and finally to 2.0. In more details the multi-step experiment is described below.

Code available in supplementary material.

### Table Eq-python: Q1,e = Q1,c & Q2,e = Q2,t ,\*=means differs from original, NONE = the equation was not used at all as the species of Q1 or Q2 does not exist

|  |  |  |
| --- | --- | --- |
| **No** | **Mol.** | **Equation** |
| 5\* | Q1,c |  |
| 6\* | Q1,t |  |
| 7\* | Q1,e |  |
| 8\* | Q2,c |  |
| 9\* | Q2,t |  |
| 10\* | Q2,e |  |

## 3.2 The implementation of the model using gro programming language

To implement the model the programming language gro was used (<https://pubs.acs.org/doi/pdf/10.1021/sb300034m> ). Gro was developed in 2012 by Eric Klavins as a result of increasing construction of synthetic multi-cellular systems but on the other hand, a lack of design frameworks that took into account the gene interactions with signalling molecules and spatial component of multi-cellular behaviour (<https://pubs.acs.org/doi/pdf/10.1021/sb300034m> ). An extension of gro that can be used atop the original version was presented in 2017 by LIA-UPM lab. The extension increases the simulation capabilities as well as presents some additional functions and actions that allow more accurate E.coli simulations (<https://pubs.acs.org/doi/pdf/10.1021/acssynbio.7b00003> ).

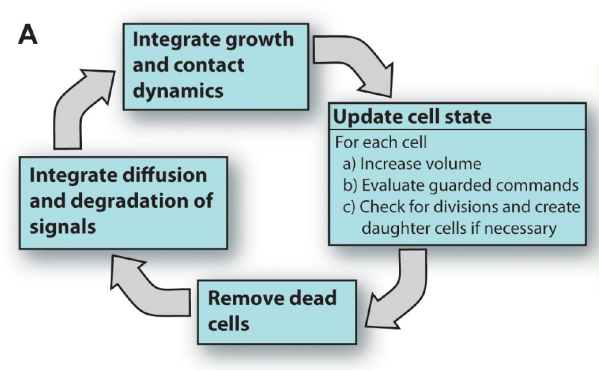
Gro is an open source software package that can be used to simulate and model microorganisms in colonies and their behaviour: growth, cell-to-cell signalling, movement, division, death and more (<https://pubs.acs.org/doi/pdf/10.1021/acssynbio.7b00003> ). Furthermore, it can simulate gene transcription and translation as well as factors that regulate gene expression such as extracellular signals (<https://pubs.acs.org/doi/pdf/10.1021/sb300034m> ). An important feature of gro is the 2D simulation environment which takes into account cell position and density as well as signal diffusion (<https://pubs.acs.org/doi/pdf/10.1021/sb300034m> ). This is one of the reasons the gro was chosen as the modelling tool. Gro simulates microorganisms in a monolayer as they would be observed under a fluorescent microscope. The focus of the program specifications is on *E. coli* and *S. cerevisae* modelling. Hence, this program is an appropriate choice for this project as *E. coli* was used by Fiore and is one of the leading microorganisms used in synthetic biology. Due to gro’s flexibility the behaviour of the system can be addressed on an individual level or on a global level and thus the behaviour of a collection of cells and one cell can be investigated and compared. Another important feature that makes gro a very useful tool is the signalling component in space. As signalling cells can be emitted from individual cells or the environment at a specific diffusion and degradation rate and can be absorbed by other cells at specified rates.

Gro combines parallel computing with a distributed algorithms approaches to enable simulation of microorganisms. The programming language is based on C++. The parameters and actions that describe the system are set by using the gro programming language. To run the simulation the program is executed and proceeds through the following four stages: (1) Parsing step where the syntax is checked for errors and if none are found it can precede to (2) Global Initialization step. In this step the globally set statements are executed, such as the maximum number of all cells, global variables, signal declarations etc. Afterwards, the (3) Program Initialization steps occurs where in each individually set program that specifies a certain behaviour, the non-guarded commands are executed. Lastly, the simulation goes through the (4) Simulation loop where all the guarded commands are evaluated and executed. Guarded commands are those that are only executed if the “guard” which is a Boolean expression is true. Afterwards, each cell in the program is updated. This means that variables such as volume and division status are updated and changed to simulate real-life cell growth and division. Thus, if a cell is big enough a division occurs which approximately halves the volume of the cell and any other numerical value and the programs are passed on to the daughter cell. Furthermore, variables such as concentration of a molecules or concentration of emitting signal are also updated according to the specifications set by the user. Figure gro main loop shows the fourth step in more detail.

The cells are modelled as approximately cylindrical shaped – radius = 0.5 μm and length = 2.0 μm – points in space with an initial volume of 1.57fL. The growth of a cell is modelled by a differential equation of volume change in time and is incorporated in parallel to other sub-processes defined by the user. The number of cells in the colony changes stochastically and models the crowding geometry-like growth observed in real-life. The simulation does not progress in real time but it slows down as the number of cells increase exponentially as more evaluation is required to proceed.

The signals emitted by cells is simulated as a grid point where each point stores the signal concentration that changes over time as the signal diffused or is degraded (<https://pubs.acs.org/doi/pdf/10.1021/acssynbio.7b00003> ).

Gro has been used to simulate E. coli growth rate and has shown to follow the real life data about well (<https://pubs.acs.org/doi/pdf/10.1021/sb300034m> ). Furthermore, many systems have been modelled using gro. For example, gro has been used to model edge detection in microcolony by utilising signal diffusion and consequent differences in concentrations of the signal across the space. This simulation models how cells can sense some traits of the geometry and can recognise if they are near or far from the edge (<https://pubs.acs.org/doi/pdf/10.1021/sb300034m> ). Other simulations have modelled coupled oscillation (<https://depts.washington.edu/soslab/gro/gallery/oscillator.gro> ), morphogenesis (<https://depts.washington.edu/soslab/gro/gallery/morphogenesis.gro> ) and other E. coli colony behaviours. Gro has also been shown to be a useful tool in biology, ecology and epidemiology and can be used to better understand the individual cell as well as colony behaviour (<https://pubs.acs.org/doi/pdf/10.1021/acssynbio.7b00003> ).



### Figure gro main loop. <https://pubs.acs.org/doi/pdf/10.1021/sb300034m>

## 3.3 Gro: constructing the model

After evaluating the validity of Fiore system of equations and their outputs, the system was implemented in gro. The code can be found in the Supplementary material.

The reactions were set to mimic the Fiore experiments as well as the natural system in E. coli. The majority of rates were taken from Fiore however, some needed to be changed to fit the modelling environment. The parameters can be found in Table parameters-gro. The values were recorded every 0.05 min (dt = 0.05).

The ODEs were split in two to model the synthesis and the degradation rates separately. The equations were split due to the rules set by the simulation environment. The synthesis is the part of the equations in Table Eq-python that corresponds to the Hill function with exponent ns with dissociation constant Ks. The degradation is the second part of the ODEs where the concentration of a molecule decreases with first order kinetics in accordance with the rate govern by parameter γx.

The initial concentrations of individual molecules were determined by running the simulation for t = 1000min at Reference concentration 0.6 and can be found in Table parameters-gro.

The two cells are set in space where they are set on the same x-axis value but differ in the y-axis positioning. The distance between the centres of the cells on y-axis is 20 μm. This was set so as to ensure that the two cells do not overlap but are in close proximity. The two E. coli simulated in different colours. The Controller is green due to and the Target cell is red (Figure simulation). The colour of the cell is determined by the fluorescent protein in the cell. However, the fluorescent protein does not interfere with the intracellular circuit modelled.

Three signals are declared: (1) reference signal, (2) Q1 signal and (3) Q2 signal. When declaring a signal an extracellular diffusion and degradation rate are required. The two parameters were chosen based on the parameter values that were used by Fiore to describe diffusion and degradation. However, the diffusion rate was amendment to fit the simulation environment. As in gro the signals can be modelled with an in-build function only one species of the signalling molecules is used. Thus there is only a Q1 and Q2 and there is no need to have the other two species.

To model the behaviour three programs were set: (1) Controller cell, (2) Target cell and (3) Main program. In Controller and Target cell programs the intra-cellular circuit of each of the cells was specified. Firstly, the parameter values are declared. To model the synthesis, degradation, absorption and emission of the signals a guarded command that is constantly set to true is used. This results in continuous approximation of the nonlinear equations. To numerically integrate the ODEs and find solutions at given point in time the Euler method with a given initial value is used. Euler methods is generally defined as:

(Butcher 2003, p. 45; Hairer, Nørsett & Wanner 1993, p. 36)

Which in the model results in:

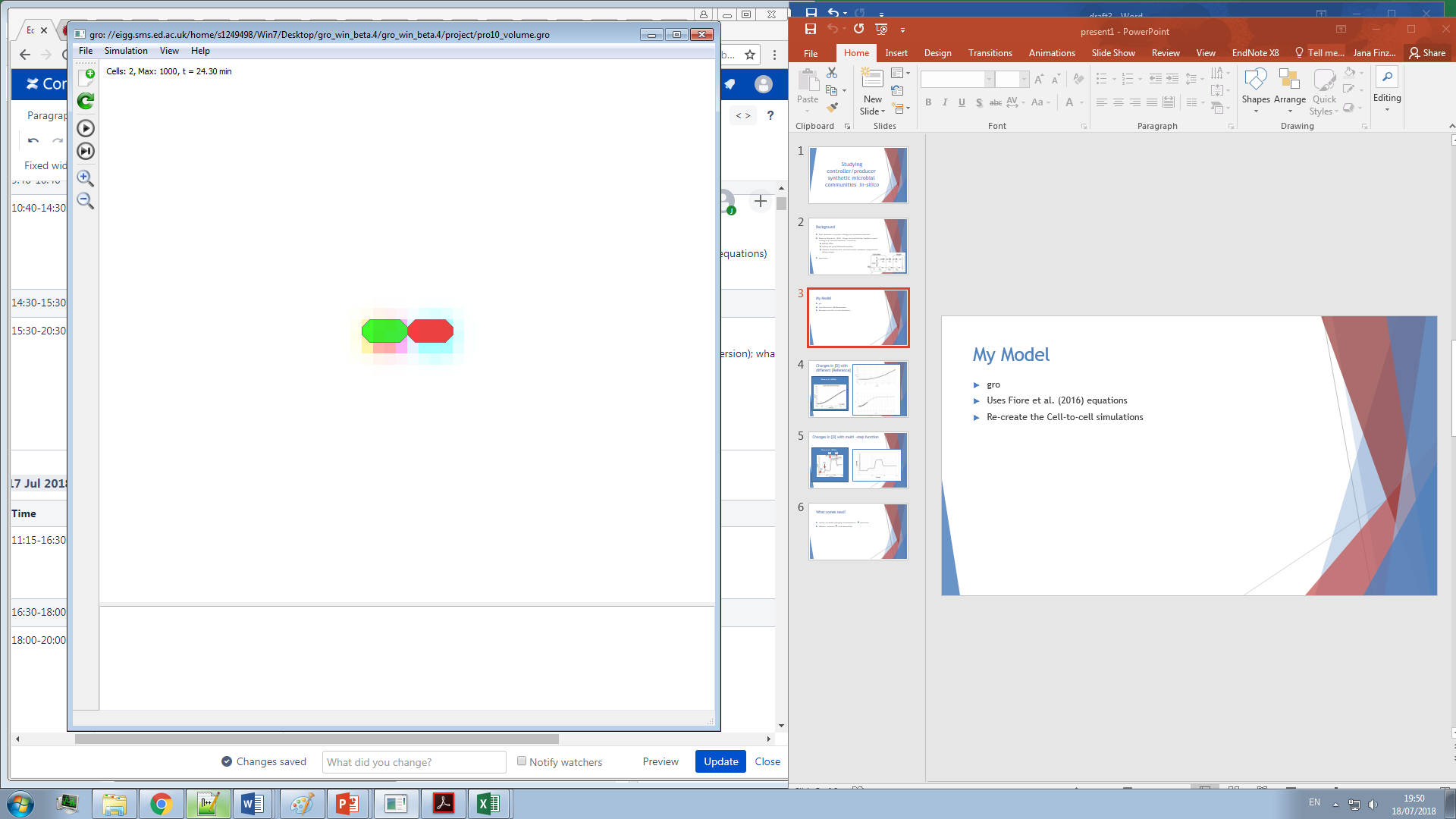
where x is the molecule in question, dt is the time step and f(x) is the ODE.

To model absorption and emission the in-build functions absorb\_signal and emit\_signal are used, respectively. For the functions to work an absorption or an emission rate is required. The rates in this case are represented as amounts of the molecule absorbed or emitted per min. In this model the amount of signal is determined by the concentration of the set signal multiplied by a rate parameter. This part of the model differs greatly from the Fiore model as the programming environment is different. Gro is not able to distinguish the concentration of a signal outside the receiver cell compared to the intracellular concentration in the receiver. Therefore, the absorption rate does not take into account the intracellular concentration of the signal explicitly.

The third program is the Main program which defines the release of Reference signal. The Reference signal is set using the build-in command set\_signal which specifies the position of the extracellular signal and its amount released. The Reference signal is emitted close to the Controller cell. The Main program is by default a special kind of program which is called every cycle and is implemented every time after the values in other programs are updated. Hence, when the extracellular signal is set at a specific concentration the signal is continues and steady.

### Table parameters-gro:

|  |  |  |
| --- | --- | --- |
| **Description** | **Parameter** | **Value** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |



### Figure simulation:

## 3.4 Validating the model

To ensure the model implemented in gro is working without errors, the same experiments as done by Fiore were recreated and compared to Fiore outputs shown in figures.

Firstly, cell-to-cell steady state characterization was modelled. The concentration of Reference was varied in the interval 0.0μM to 3.0μM by 50 steps and the concentration of D after reaching steady state (t = 500min) was obtained. To recreate Fiore Figure 3 the concentration of molecule D at steady state was plotted against concentration of Reference.

To model the multi-step function the change in concentration of Reference signal was changed following the unit step function or Heaviside function u(t):

The function predicts that Reference signal concentrations will change in the following order: 0.0, 1.0, 3.0 and to 2.0μM. The changes occur at t=300min, t=700min and t=1100min and the simulation runs till t=1500min.

A plot of concentration of D over time was constructed to allow comparisons to Fiore Figure 4A. To obtain the desired output the simulation was ran for t=500min at a set Reference concentration – either 0.0, 1.0, 2.0 or 3.0 –to obtain the value of D at steady state. This values were than used to construct the desired output curve.

The comparison was done by observing the shapes of the graphs as well as the concentration values of D reached at steady states in Fiore and gro model.

## 3.5 Evaluating the robustness by varying parameter values

To evaluate robustness, key parameters were varied and a sensitivity calculation was performed.

Changing parameters affects the model’s output and to assess to what degree the change affects the output key parameters are varied (Crick, M.J., Hill, M.D. and Charles, D.: 1987, 'The Role of Sensitivity Analysis in Assessing Uncertainty. In: Proceedings of an NEA Workshop on Uncertainty Analysis for Performance Assessments of Radioactive Waste Disposal Systems, Paris, OECD, pp. 1-258.).. Key parameters in this case are those that are sensitive and have a significant influence on the output results

The parameters that were changed and to what value is shown in Table Parameter change. The same multi-step function as described above was used to obtain the concentrations of D.

To calculate the variation the Sum of squared errors (SSE) was used where the differences between the measured concentration of D and the desired output at time t were calculated, squared and summed up. Next the SSE values of two different parameter values are subtracted from each other and divided by the difference between the actual parameter values. Lastly the value the value obtained was divided by the total number of time points to obtain an average value. The formulas for the calculations are shown below. If the result is equal to 0 than the outputs are identical. Otherwise, there are discrepancies or differences between the parameters.

The parameters are varied by 0.1%, 1%, 5% and 10% from their original value.

Parameters chosen:

**KQ1** 🡪 because Q1 is important in conveying the Controller cell messages and is thus important in regulation of concentration of D, and KQ1 is the most important value in production of Q1

MOGOC LIH NA GRE U METHODS AMPAK DISCUSSION: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.224.7813&rep=rep1&type=pdf> Two things can be looked at: (1) do small changes in the parameter propagate to large changes in the output results and (2) does an uncertainty associated with a parameter have a large contribution to the changes in the results.

### Table Parameter change:

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Original value** | **Alternative values** |
| KQ1 | 0.05 | 0.1% 🡪 0.05005 & 0.04995  1% 🡪 0.0505 & 0.0495  5% 🡪 0.0525 & 0.0475  10% 🡪 0.055 & 0.045 |
| KQ2 |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## 3.6 Data presentation and analysis

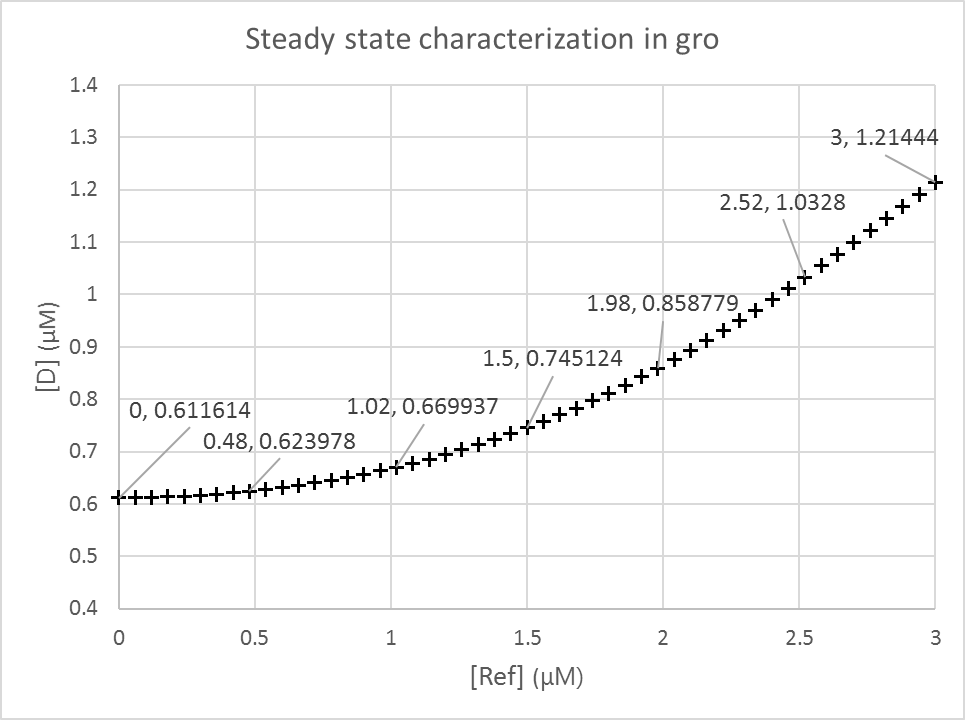
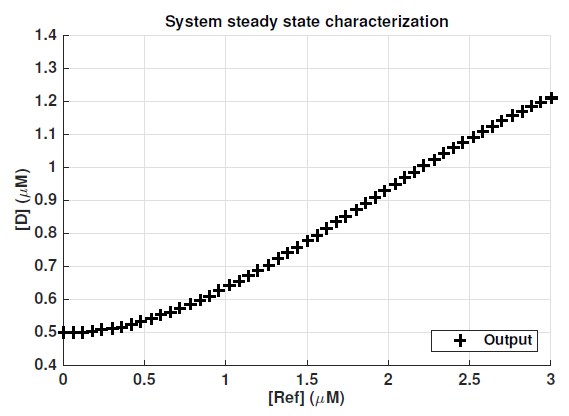
To obtain the data, the gro program allows for direct output of simulation results to Comma-separated values (CSV) files. The CSV files were opened using Microsoft Excel (2016). Furthermore, Microsoft Excel (2016) was used to obtain graphs and carry out any calculations required.

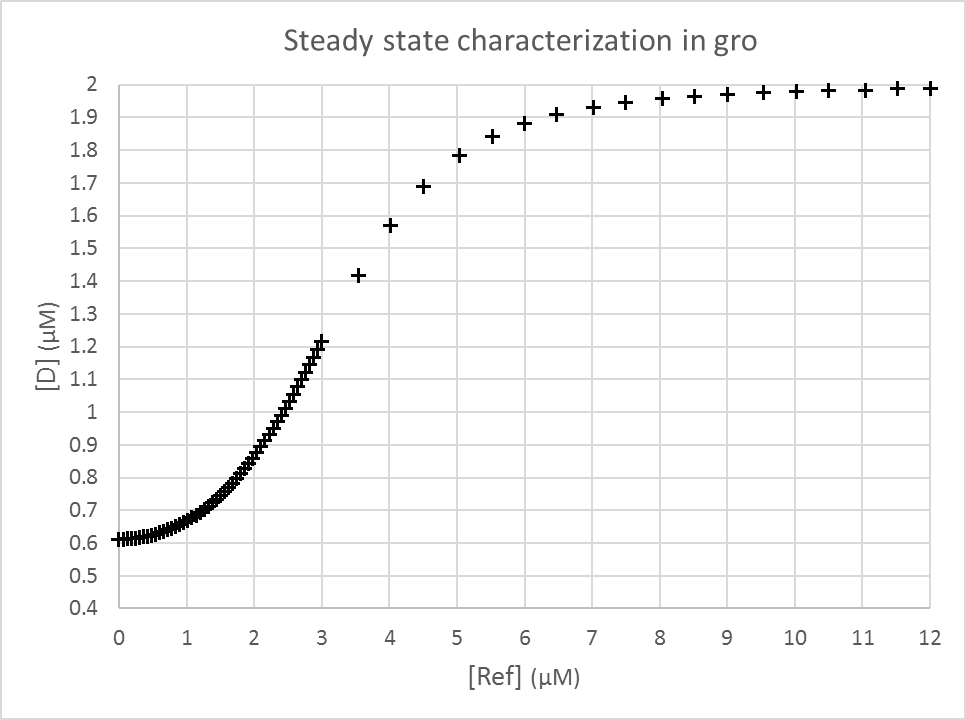
# 4 Results

## 4.1 Validating the computer model

One of the aims of the project was to determine if the model first described by Fiore can be implemented in a different programming environment called gro. To evaluate the new models ability to simulate the behaviour observed by Fiore the outputs of different scenarios were compared amongst the models.

Firstly, the steady state characterization scenario was implemented in gro. The results are shown in Figure Recreate fig3 together with the original output. From the shape of the graph it can be observed that the gro model outputs also results in a sigmoid curve. The shape of the curve does differ slightly where the concentration of D follows a longer initial concave upwards path compared to the original output. However, by using higher Reference concentrations it can be observed that the concentration of D from the gro model does approach a plateau phase. The concentrations of D at a specific Reference concentration do differ amongst the models. Especially notable is the concentration of D at Reference = 0.0 μM where gro model predicts a 0.1 μM higher concentration of D. After [Ref] = 1.32 μM the concentrations of D are overall higher in original model compared to gro output until [Ref] = 3.00 μM. This is due to the slightly different sigmoid shapes the two outputs follow. The finial D concentration at [Ref] = 3.00 μM is similar for both models – approximately 1.2 μM.





### Figure Recreate fig3:

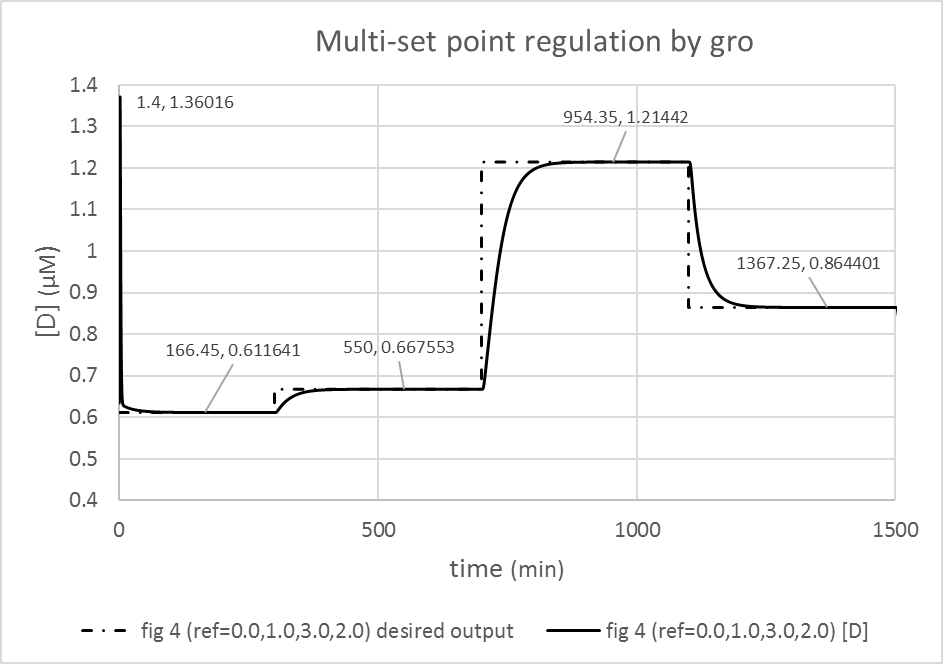
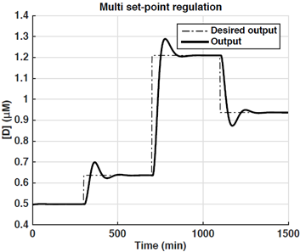
Secondly, the graphs evaluating the changes in concentration of D with multi-step change in concentration of Reference signal over time were compared. The output produced by gro and Fiore are shown in Figure Recreate fig4A. Similar as with Fiore outputs there is zero steady-state error and the settling time varies from 120min to 150min. However, there are some notable differences. Firstly, the gro output has a to reach steady state at the beginning of the simulation when the concentration of D drops from 1.36 μM to the steady state value of 0.61 μM. Furthermore, the gro simulation does not have an overshot as observed by Fiore. Instead, the steady-state is reached in a more gradual way. The concentrations of D reached at every step are comparable to the Fiore outputs, especially the [D] at [Ref] = 3.0 μM.

By comparing the outputs produced by the model implemented in gro to original paper it was observed that there are differences in the outputs as expected due to the different simulation environment. However, the overall behaviour of the system as seen by evaluating concentration of D at different concentrations of Reference signal, is comparable and similar between the models.

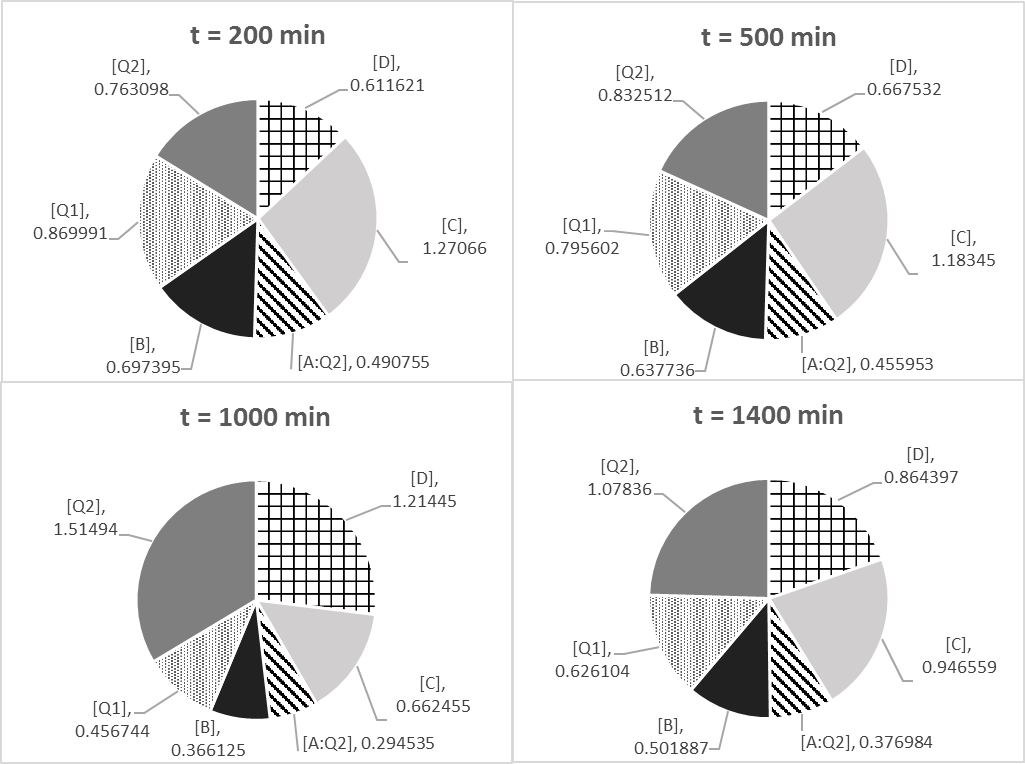
**IF I DECIDE TO PUT IT IN:**

While concentration of D changes also other species in the system change as well and to see how the other molecules change depending on concentration of Reference signal, pie charts of all the species at times t = 200min, 500min, 1000min and 1400min have been created. The times were chosen as all species have reached a steady-state concentration at different concentrations of Reference signal.

It can be observed that the greatest change in concentration values are associated with molecules D and Q2 in Target cell which almost double with [Ref] = 3.0. Concentrations of both molecules increase proportionally and for the same percentage from the initial steady state at [Ref] = 0.0. They increase for 9%, 99% and 41% as [Ref] changes according to the multi-set function. The concentration of the other four molecules decrease in a similar manner.



### Figure Recreate fig4A:



### Figure Species values: not sure yet

## 4.2 Robustness analysis

To evaluate how the behaviour would change and how large is the influence of a parameter value the parameter values were changes. Sensitivity calculation were performed to evaluate how robust the control system is despite the changes.

**KQ1**

> [A:Q2] and [B] --> do not change despite KQ1 changes

> 

>0.1% no change, 1%: 1% increase of decrease, 5%: up to 6% increase and 5% decrease, 10%: up to 12% increase or decrease

> D and Q1 conc. change in the same manner (for the same %)

>

# 5 Discussion

# 6 Conclusion

# References