Exploring the behaviour of the Repressilator Gene Regulatory Network using *in silico* modelling

Introduction

Genetic circuits, such as circadian rhythms, genetic toggle switches, and synthetic genetic oscillators are fundamental components in understanding and manipulating biological processes. Circadian rhythms regulate physiological and behavioural cycles, while genetic toggle switches exhibit biostability and memory-like behaviour. Synthetic genetic oscillators, inspired by natural systems like the repressilator, generate periodic gene expression patterns. These examples highlight the versatility of genetic circuits and their potential applications in various fields. In this report, I will focus on the *in silico* modelling of a repressilator gene regulatory network (GRN).

Perhaps the most well-known synthetic circuit is the repressilator engineered by Elowitz et al (2000), consisting of three genes interconnected in a negative feedback loop. Specifically, in this system, tetR is repressed by lacI, cI is repressed by tetR, and lacI is repressed by cI. This forms a cyclic pattern where each gene represses a protein that represses a subsequent gene in the cycle. The cyclic nature of the system is shown in Fig 1. This cyclic arrangement contributes to the oscillatory behaviour of the repressilator system. In this model, the behaviour of the network depends on several factors, including cooperativity of the repressor, initial concentration of mRNA for the repressors, rate of translation and rate of mRNA degradation. Depending on the initial conditions, and the values of these parameters, the model either converges to a stable steady state, or an unstable steady state, showing limit-cycle oscillations.

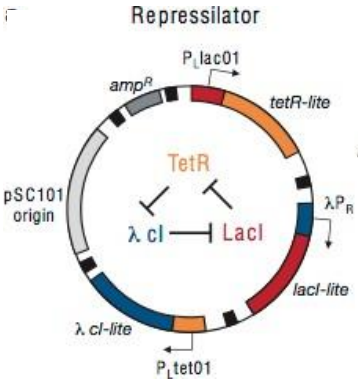

Fig 1 The repressilator system, from Elowitz et al (2000).

Methods

The system is simulated using a deterministic method. The differential equations that make up the model used to simulate the system are:

$$\frac{\mathrm{d}m_i}{\mathrm{d}t} = k_m * \frac{K^n}{K^n + p_j^n} + k_{m0} - k_{dm} * m_i$$

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = k_p * m_i - k_{dp} * p_i$$

$$\begin{pmatrix} i = lacI, tetR, cI \\ j = cI, lacI, tetR \end{pmatrix}$$

where m and p are the concentration of mRNA and protein respectively, k_m is the rate of transcription, K is the activation constant, n is the Hill coefficient, k_{m0} is the "leakiness" of mRNA, k_{dm} is the rate of mRNA degradation, k_p is the rate of translation, and k_{dp} is the rate of protein degradation.

The model does not contain RNA polymerase and nucleotides in transcription, ribosomes and amino acids in translation, and proteases and ribonucleases in degradation, as it assumes that these are all present and available at constant processes. As result, the reactions that involve them can be modelled as mass action processes.

Results and Discussion

Cooperativity

In the mRNA kinetics equations above, n is the Hill coefficient, which represents the cooperativity of binding. A Hill coefficient of 1 represents independent binding, whereas a value greater than 1 represents positive cooperativity. This is where the binding of one molecule facilitates the binding of subsequent molecules at other sites. Varying the value of n helps us understand the non-linear dynamics involved in repression. Fig 2 shows that the system is non-linear. In a linear relationship, the steady state concentration of the tetR protein would decrease proportionally as the concentration of the lacI protein is increased. Additionally, we can see threshold behaviour, especially at high values of n. The curve starts at a high level and decreases gradually, until a threshold is reached. Once the threshold is reached, there is a rapid decrease, eventually levelling off at a lower value. For $n = 1$, the threshold is at a lacI concentration of 25 arbitrary units. The graph also shows saturation. This is where the input has little to no effect on the output once it reaches a certain point. At high cooperativity, saturation occurs at a lacI concentration of 75 units.

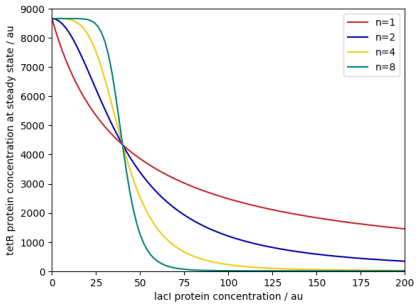

Fig 2 Effect of cooperativity on steady state

The rest of the results in this report are based on a Hill coefficient n value of 2, unless stated otherwise. Repressor proteins, such as lacI, tetR and cI, often display cooperative binding (Whipple et al., 1998). This could be due to several reasons. For example, many repressor proteins exist as multimers, where multiple subunits come together to form the active repressor complex. The interaction between these subunits can lead to cooperative binding. Another reason why repressor proteins may show cooperative binding is due to allosteric effects. Repressor proteins can undergo conformational changes upon binding to the DNA (Flynn et al., 2009). These conformational changes can affect the binding affinity of subsequent repressor molecules to nearby binding sites. In some cases, repressor proteins can bind to multiple DNA sites and form loops that bring distant DNA regions closer together (Hochschild & Ptashne, 1986). This looping allows the repressor proteins to interact with each other and enhance their binding affinity.

Simulation of the Repressilator using an ODE model

The simulations were run using the same parameters as in Elowitz and Leibler's paper. The initial number of lacI mRNA was 5, tetR and cI mRNA was 0. The initial number of all three proteins was 0. The other parameters are summarised in table 1. All simulations in this report are run using these values, unless explicitly stated otherwise.

Parameter	Value
k_m	30
K	40
n	2
k_{m0}	0.3 $(0.001 * k_m)$
k_{dm}	0.3466
k_p	6.931
k_{dp}	0.0631 $(0.01 * k_p)$

Table 1 Parameter values used in simulations

Fig 3a shows a reproduction of the results presented by Elowitz and Liebler, using our model. The results show oscillatory behaviour, leading to sustained limit-cycle oscillations after 400 minutes. The simulation results can provide some insights into the gene regulatory network that cannot be inferred by considering the genetic network and its diagram. Firstly, the ODE simulation shows the temporal dynamics of the repressilator network. It shows how the concentrations of each protein changes over time. This is crucial for understanding the system's behaviour and its response to different initial conditions. Additionally, the ODE simulation allows us to analyse the stability of the repressilator network. The simulation allows us to determine whether the system reaches a stable equilibrium or exhibits oscillatory behaviour. Stability analysis provides insights into the robustness and resilience of the network and its sensitivity to perturbations.

The phase plot in Fig 3b shows that the model demonstrates limit-cycle oscillations, after initial transience. Limit-cycle oscillations refer to a sustained, periodic behaviour in dynamic systems, where variables evolve in a repeating cycle without approaching a stable equilibrium.

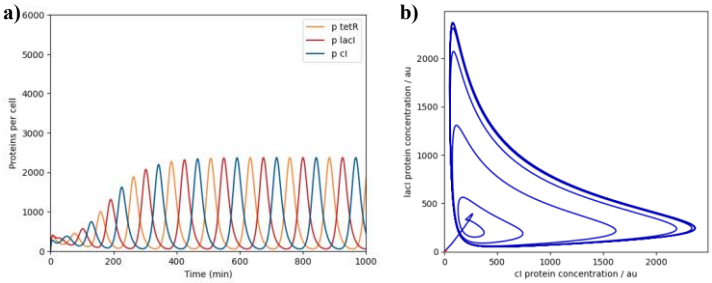

Fig 3 a Deterministic simulation ***b*** Phase plot for deterministic simulation

However, it is important to note that a deterministic ODE-based model has some significant limitations. Firstly, the model ignores stochasticity, and hence does not account for random events and fluctuations that can occur at a molecular level. To address this, a stochastic simulation was also run. Furthermore, the ODE-based models typically treat the system as well-mixed, assuming homogeneity throughout the cells. However, in reality, biological systems often exhibit spatial heterogeneity and compartmentalisation. To address this, a spatial model can be used.

Limit-cycle oscillations and equilibria

As the model assumes that the rate of production, repression, and loss of the three proteins are equal (k_p, k_m, and k_{dp} respectively), the system can enter a steady state if started in a state where the initial number of mRNA and proteins are identical. Fig 4a shows the steady state concentrations of each protein, when the initial number of mRNA for each protein is 5 per cell, and the initial number of each protein is 0. After some initial transient behaviour, the steady state is reached. The steady state concentration of each repressor protein is 240 copies per cell. The steady state

concentration of each mRNA is 2.40 copies per cell. The steady state is unstable, as shown in the phase plot in Fig 4b. In this case, the initial concentration of all the mRNA transcripts and repressor proteins is the same as at steady state, except the initial concentration of tetR protein is reduced by 1%. The phase plot spirals away from the steady state, eventually showing limit-cycle oscillations. This shows that the steady state is unstable.

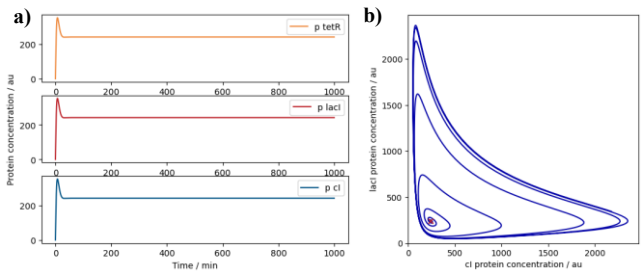

Fig 4 a Protein concentration at steady state ***b*** Phase plot when initial conditions are near steady state. Steady state is marked in red

The repressilator system shows limit-cycle oscillations as it meets the three criteria for a Goodwin Oscillator (Gonze & Ruoff, 2020):

1. **Negative Feedback:** As shown in Fig 1, each gene in the system represses the expression of the next gene, creating a cyclic inhibitory loop.
2. **Non-linearity:** As shown in Fig 2 and discussed above with relation to Hill constant and cooperativity.
3. **Delay:** Delays occur in the repressilator system due to the time it takes for gene expression, protein production, and protein degradation processes to occur. There is also delay in the time it takes for proteins to move within cells, and between cells, as well as to interact with its target.

Steady State vs. Limit-cycle Oscillations

1. Hill Coefficient

The final state of the simulations is dependent on the conditions. By varying the Hill coefficient, rate of mRNA 'leakiness' and rate of protein translation, the system either reaches a steady state or limit-cycle oscillations. Fig 5 shows how the behaviour of the system changes as the value of n varies. Depending on the value of n, the system reaches a constant steady state, or shows limit-cycle oscillations. Fig 5d shows a bifurcation plot of n against the minimum and maximum concentration of tetR protein. The maximum and minimum concentrations are either from steady state or limit-cycle oscillations. In other words, any transient behaviour has been removed. When $n < 1.6$, the minimum and maximum concentration of tetR is the same, i.e. the system reaches a steady state. However, when $n \geq 1.6$, the difference between the minimum and maximum concentration increases as the value of n increases. This shows that the system reaches limit-cycle oscillations, with the peak concentration of tetR increasing with n. In summary, when the cooperativity of repression is less than 1.6, the system reaches steady state. Otherwise, the system is oscillating.

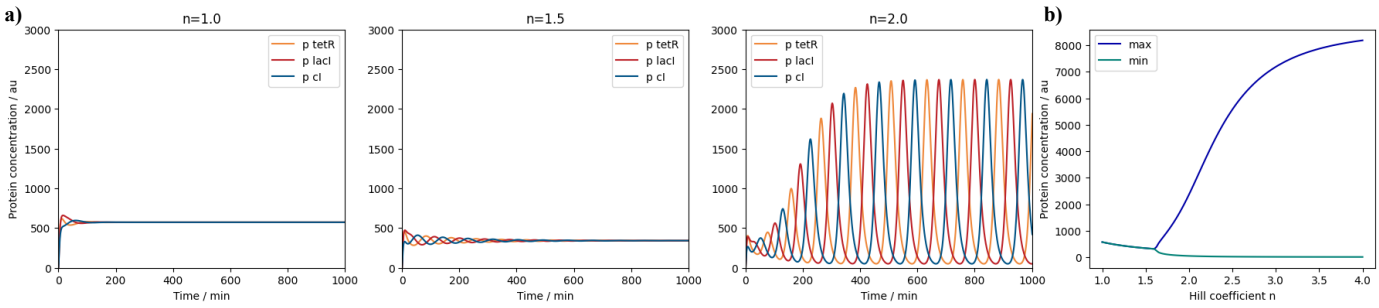

Fig 5 a Protein concentrations when n is varied ***b*** Bifurcation plot of max and min protein levels as n is varied

2. Rate of translation

Fig 6 below shows the effect that the rate of translation has on the system. At high values of k_p, i.e. when rate of translation is high, the system reaches steady state instead of limit-cycle oscillations. Controlling the rate of translation also has an effect on the amplitude and period of the oscillations.

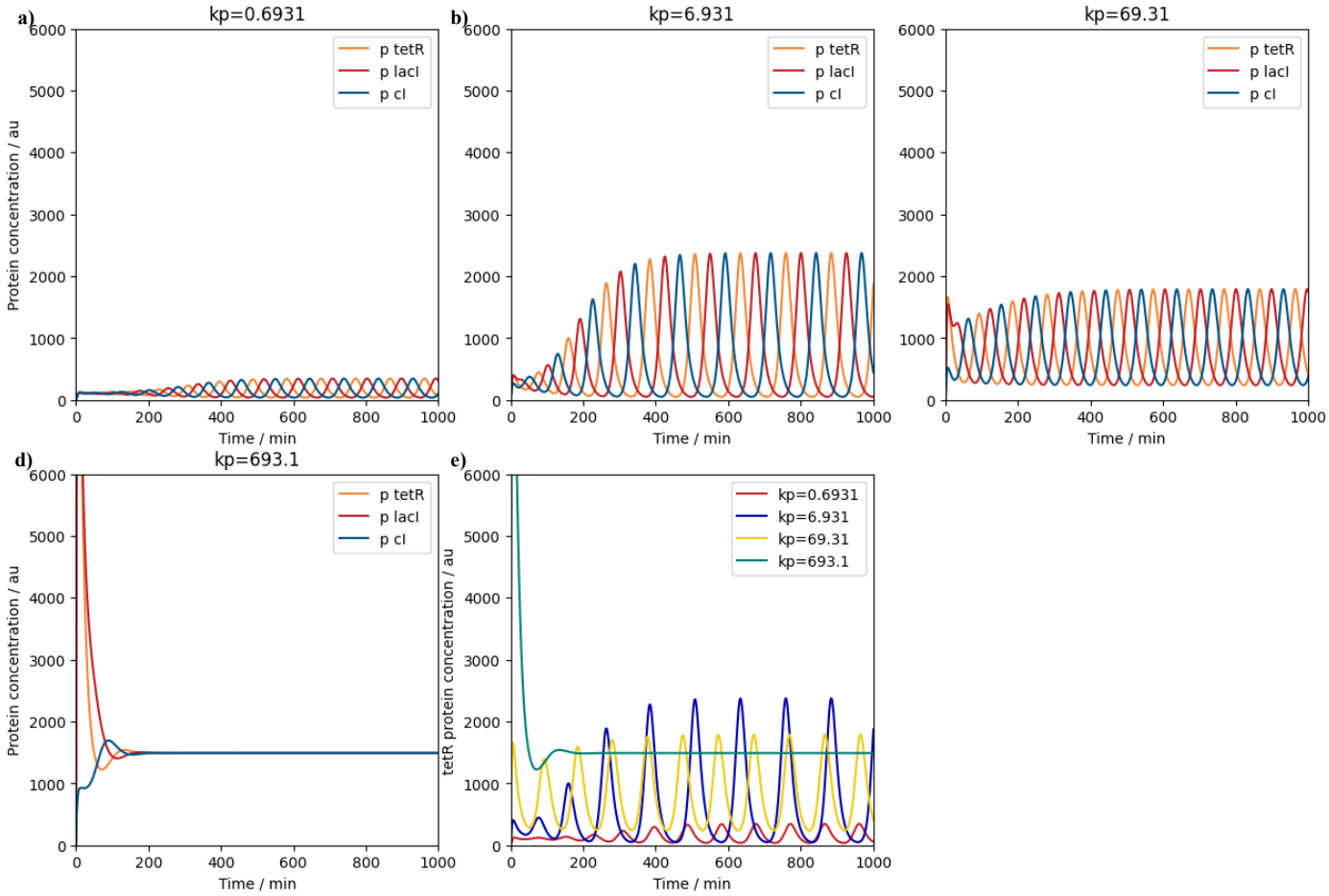

Fig 6 a-d Effect of varying k_p on protein concentration ***e*** Overlay plot of tetR concentration at different k_p values

3. Rate of mRNA leakiness

Fig 7 below shows the effect that the rate of mRNA "leakiness" has on the system. mRNA "leakiness" refers to the unintended transcription of MRNA from a gene. Therefore, k_{m0} represents spurious mRNA production without specific induction signals. Understanding and minimising leakiness is important for precise gene expression control. At high values of k_{m0}, i.e. when rate of mRNA leak is high, the system reaches steady state, instead of limit cycle oscillations. High leakiness in gene expression prevents the establishment of necessary feedback loops for sustained oscillations, leading to steady state behaviour instead of limit cycle oscillations. The continuous production of mRNA overwhelms the regulatory mechanisms, resulting in a stable equilibrium. Leakiness in cells can be controlled through various approaches such as modifying promoter strength, adjusting transcription factor binding sites, optimizing ribosome binding sites, implementing post-transcriptional regulation, or using synthetic biology techniques like CRISPR-based gene editing.

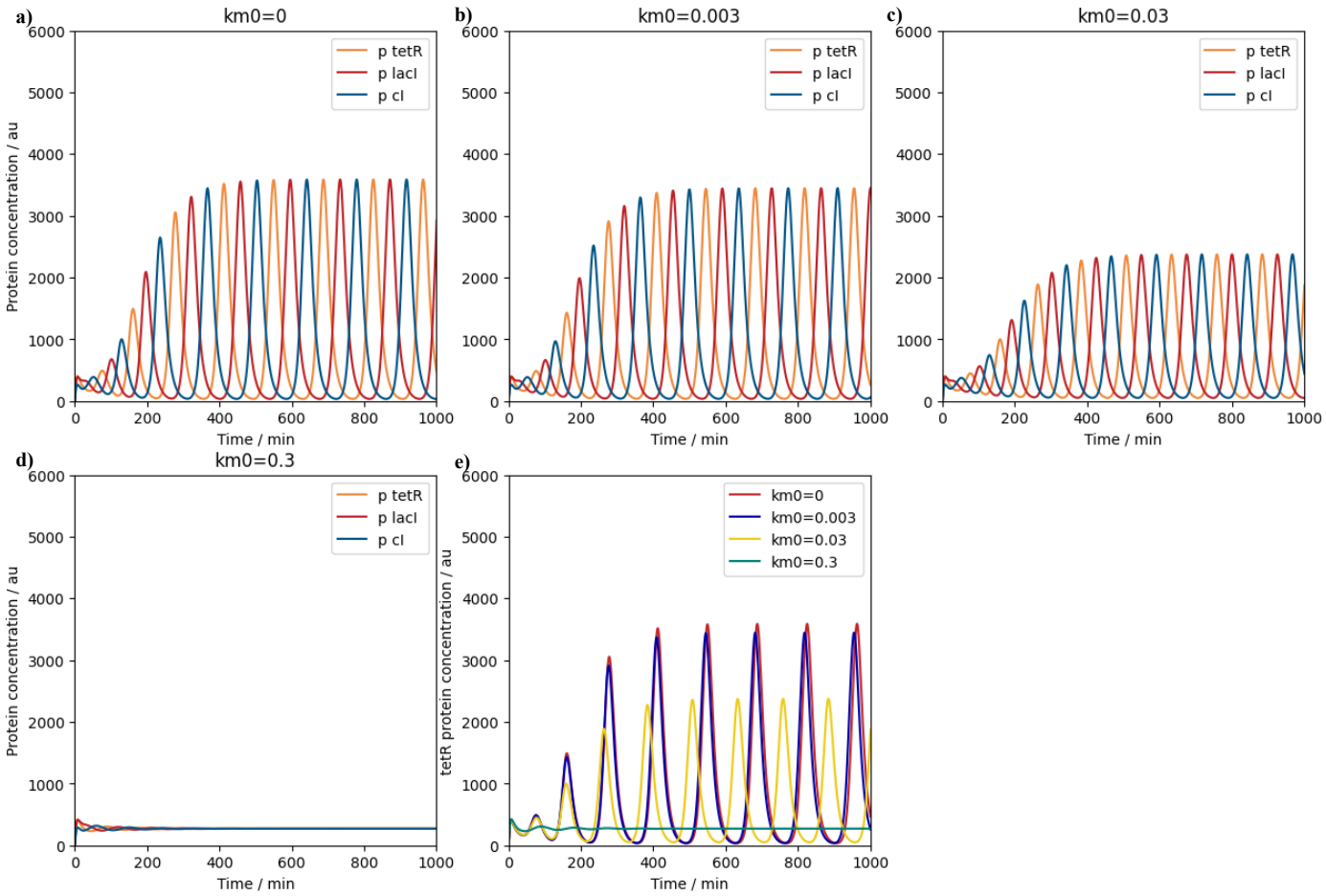

Fig 7 Effect of varying k_{m0} on protein concentration ***a-d*** When k_{m0} is 0, 0.003, 0.03 and 0.3 respectively ***e*** Overlay plot of tetR concentration at different k_{m0} values

Adding a GFP reporter

Adding a green fluorescent protein (GFP) reporter to a gene repressilator system serves as a valuable tool for visualising gene expression dynamics. In this system, the expression of GFP is regulated by the lacI gene, which acts as a repressor. This is shown in Fig 8a. By linking GFP expression to the activity of the repressor, researchers can monitor gene expression levels in real-time. The fluctuating concentration of the lacI repressor leads to corresponding changes in GFP expression, allowing for the visualisation of oscillatory behaviour and analysis of gene expression dynamics. This GFP reporter system provides a direct readout of gene expression and offers insights into the temporal dynamics and regulatory mechanisms of the repressilator model.

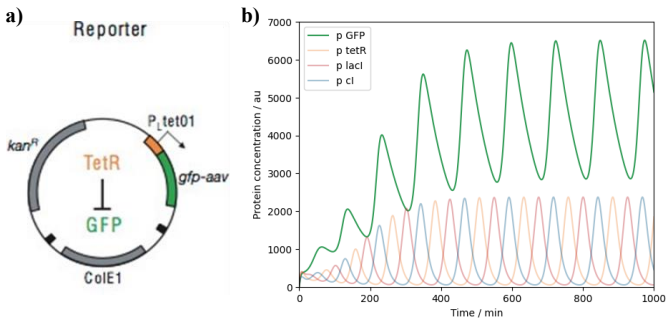

Fig 8 a GFP reporter, from Elowitz and Leibner ***b*** Simulation results after adding GFP reporter.

The model assumes that the rate of transcription and translation of GFP match that for the repressors lacI, tetR and cI. It assumes that the GFP mRNA degrades at the same rate as lacI, tetR and cI mRNA. However, it does not assume that the rate of protein degradation is the same for GFP as it is for the repressors. Instead, the model assumes a GFP half-life of 60 minutes to reflect the longer lifetime compared with the repressor proteins. Hence, the rate of GFP protein degradation, k_{dp-GFP}, is 0.01155 ($ln\ 2/60$). The data from Elowitz and Leibler shows that the lifetime of GFP proteins is roughly 60 minutes. Therefore, the following equations for GFP kinetics apply:

$$\frac{dm_{GFP}}{dt} = k_m * \frac{K^n}{K^n + p_{lacI}^n} + k_{m0} - k_{dm} * m_{GFP}$$

$$\frac{dp_{GFP}}{dt} = kp * m_{lacI} - k_{dp-GFP} * p_{GFP}$$

where k_{dp-GFP} is the rate of degradation of GFP proteins, and all other variables are as previously.

The expression of GFP proteins in the system can be seen in Fig 8b. The simulation results agree with experimental results from Elowitz and Leibler.

Experimental techniques have advanced since the original repressilator experiment. Adding additional reporter proteins, with faster degradation may help to analyse the system in more detail. For example, adding a luciferase reporter protein has some advantages of GFP. Firstly, luciferase-based assays are generally more sensitive than GFP-based assays (Yun & Das Gupta, 2014). Luciferase reactions produce light which can be easily detected and measure with high sensitivity. This makes luciferase particularly useful when studying low-abundance or weakly expressing proteins. High sensitivity also means that small effects of perturbations on the system can be detected.

Controlling the behaviour of the repressilator

1. **IPTG Inducer**

Adding the IPTG inducer into the system reduces lacI activity. When lacI binds to IPTG, it is unable to bind to the promoter region and so repression is inhibited (reference). By adjusting the concentration of IPTG in the system, the proportion of lacI in its bound form can be controlled, thus controlling the level of tetR repression. In the simulation, this is modelled using parameter X, which is the fraction of unbound, active lacI present:

$$p_{lacI-active} = X * p_{lacI}$$

When IPTG is absent, X is 1, and when IPTG is present at saturating levels, X is 0. The equation to model the transcription of tetR is modified accordingly:

$$rate\ of\ m_{tetR}\ production = k_m * \frac{K^n}{K^n + (X * p_{lacI})^n} + k_{m0}$$

Fig 9 shows the effect of adding a lacI inducer to the system. When inducer is at saturating concentration, it means that the concentration of IPTG is high enough to fully occupy and inhibit the lacI repressor protein. This leads to constant derepression of the tetR protein and prevents oscillations from occurring. Instead, the system reaches a steady state where tetR repression is effectively eliminated, resulting in the cessation of oscillations. The period and amplitude of the oscillations, and hence gene expression levels and time intervals, can be controlled by altering the concentration of IPTG.

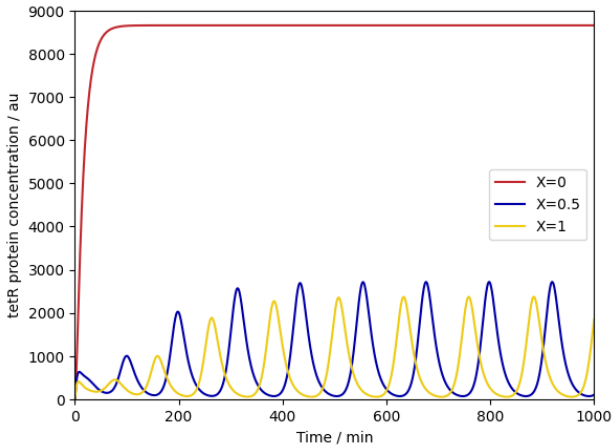

Fig 9 Effect of adding lacI inducer

2. Varying protein degradation

As the repressor involves enzymatic degradation, the rate of protein degradation can be controlled by altering the concentration of protease in the system. In the simulation, this is simplified using parameter Y:

$$rate\ of\ p_i\ loss = Y * k_{dpm} * p_i$$

A high concentration of protease enzyme is represented by a large value of Y. This increases the rate of protein degradation. When the value of Y is set to 0 in the simulation, it means there is no protease enzyme present in the system. Without protease, the degradation of the repressor protein does not occur, resulting in the repressor protein persisting indefinitely. This leads to a constant high level of repressor, preventing the oscillatory behaviour. In other words, without protein degradation, the repressor protein remains abundant and continuously represses its target, preventing the dynamics necessary for oscillations to occur. The period and amplitude of the oscillations, and hence gene expression levels and time intervals, can be controlled by altering the concentration of protease enzyme, and hence the rate of protein degradation.

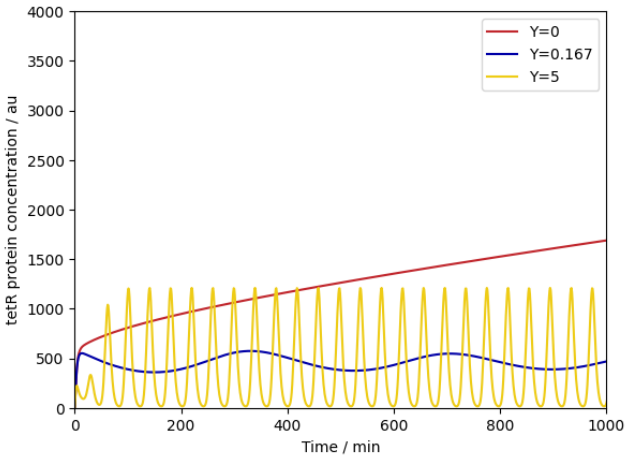

Fig 10 Effect of increasing rate of protein degradation

Stochastic Simulation

A stochastic simulation was also performed. The stochastic model for gene repressilators offers advantages over the deterministic model. It considers inherent noise sources and captures the probabilistic nature of gene regulation, providing a more realistic representation of the system. The stochastic model allows for variability analysis, revealing the range of possible outcomes and distribution of key properties, enhancing our understanding of system behaviour. It also enables the exploration of rare events and noise-induced phenomena that may have functional implications. Overall, the stochastic model enhances our ability to study and interpret gene repressilators by accounting for stochasticity and providing insights into system variability and rare events.

The stochastic makes a simplification to not explicitly model repressor/promoter binding events. This approach is valid under the assumption that the repression binding/unbinding events occur on a significantly faster timescale than the rate of mRNA and protein production events. Hence, the genetic activity can be modelled using the same kinetics as the deterministic ODE model. The results for one run of the stochastic simulation are in Fig 11.

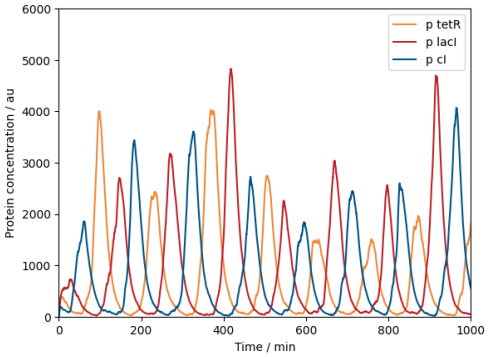

Fig 11 Stochastic simulation results

It is important to note that stochastic simulations have a large source of noise. These sources of noise arise due to the probabilistic nature of molecular interactions and the discrete nature of molecular species. One source of noise in stochastic simulations of the repressilator is the randomness associated with the timing of events. Transcription and translation processes occur in discrete steps, and the timing of these events can vary due to stochasticity. This variability can introduce fluctuations in the concentrations of regulatory proteins, influencing the regulatory interactions within the repressilator circuit. In addition to the inherent noise within the repressilator system, environmental factors such as temperature and pH fluctuations can introduce additional sources of noise. These fluctuations can impact the dynamics of the repressilator circuit. Temperature changes can affect reaction rates, protein folding, and enzymatic activities, while variations in pH can influence the stability and activity of regulatory proteins. Neglecting these environmental effects in our current model limits its ability to capture the full range of dynamics and behaviours observed in real biological systems. Accounting for temperature and pH fluctuations would provide a more realistic understanding of the repressilator system's behaviour in varying cellular environments.

Fig 12 shows the concentration of tetR, lacI and cI proteins respectively over 20 runs. The mean of the 20 runs is also shown. As there are a lot of variations in each simulation, plotting the mean is not very significant or relevant. The peaks observed in the simulations are not synchronised, indicating that the timing of oscillations differs between individual runs. When calculating the mean of multiple simulations, it is important to note that it may not accurately represent the true behaviour of the system due to the asynchrony of peaks. Therefore, analysing the individual simulation results and considering the variations in peak amplitudes, timings, and durations provides a more comprehensive understanding of the stochastic nature of the gene repressilator model and its diverse potential outcomes,

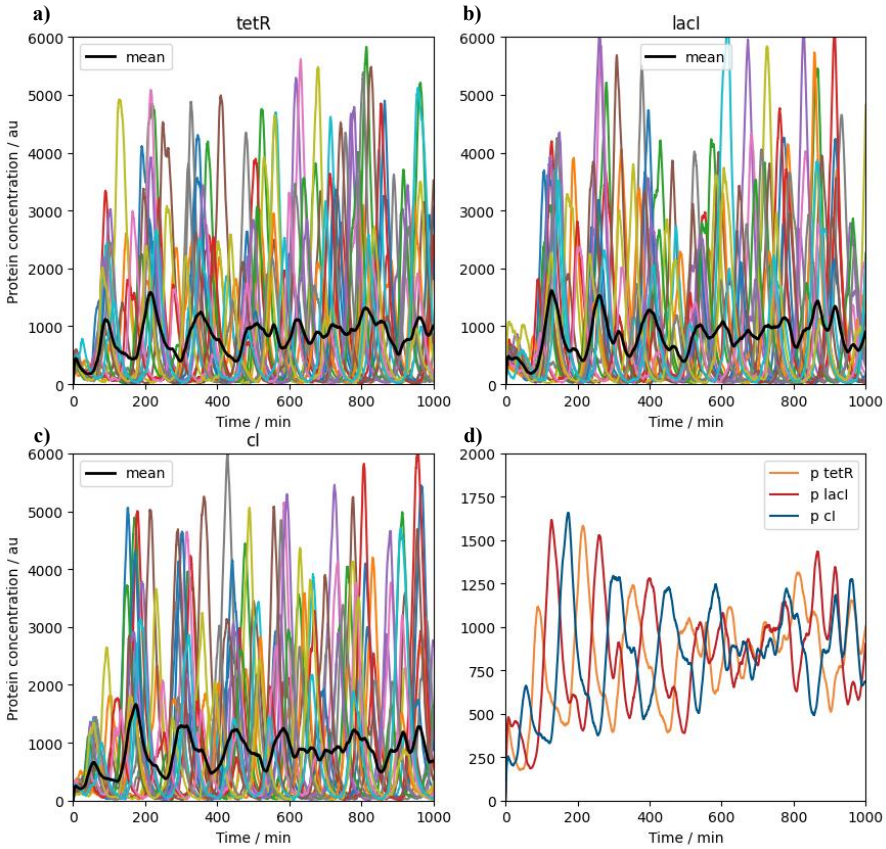

Fig 12 20 runs of a stochastic simulation ***a-c*** tetR, lacI and cI protein concentrations respectively. Each colour is one run. Mean is in black ***d*** Mean concentration of each protein

Analysis of multiple runs

100 separate runs of the stochastic simulation were analysed. Performing multiple runs of a stochastic simulation for a repressilator model provides insights into the variability of its oscillatory behaviour. It helps identify the distribution of peak expression levels and peak-to-peak intervals, capturing the stochastic nature of gene expression dynamics. Additionally, comparing the stochastic simulations with deterministic models allows for understanding the impact of inherent noise on the repressilator's behaviour. Furthermore, exploring the parameter space through multiple simulations aids in optimising the design of the repressilator and understanding the key factors influencing its oscillatory properties. One run of the simulation can be seen in Fig 13a, with the peaks identified. Fig 13b shows the distribution of the amplitude of the peaks. In total, there were 796 peaks in 100 simulation runs. Fig 13c shows the distribution of the oscillation periods for these peaks.

The analysis of the stochastic gene repressilator model reveals that the amplitude of peaks, representing gene expression levels, exhibits high variability with a mean of 3212 copies per cell and a standard deviation of 1340. This indicates that the expression levels can vary widely from cell to cell, with a relatively broad dispersion of values. On the other hand, the periods of oscillations, measured by the peak-to-peak intervals, show low variability with a mean of 130 minutes and a standard deviation of 21. This suggests that the timing of the oscillations remains relatively consistent, with minimal deviation from the average period. In summary, the gene repressilator model exhibits significant variability in expression levels while maintaining a stable and predictable timing of oscillations.

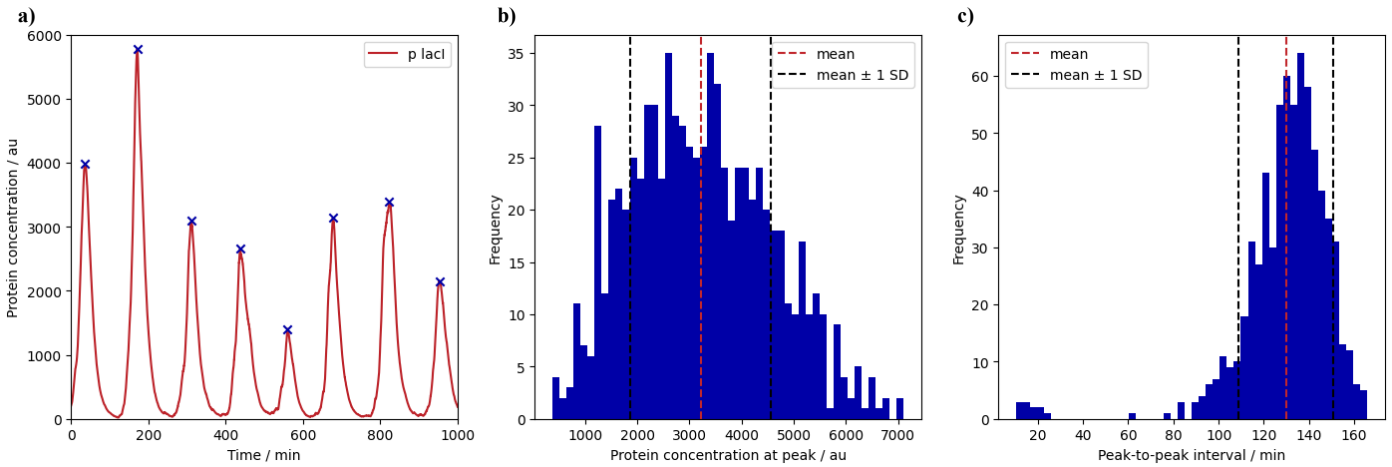

Fig 13 Analysis of 100 stochastic simulations ***a*** Identification of peaks in one run ***b*** Distribution of peak amplitudes in 100 runs ***c*** Distribution of peak periods in 100 runs.

Conclusion

The *in silico* modelling of the repressilator gene regulatory network provides valuable insights into its behaviour. By varying parameters such as cooperativity, protein degradation rates, and the addition of an inducer, we can observe the impact on the system's dynamics. The model accurately reproduces limit-cycle oscillations observed in experimental data and demonstrates the interplay between cooperativity and oscillatory behaviour. Furthermore, the addition of a GFP reporter allows for real-time visualisation of gene expression dynamics. Stochastic simulations capture the inherent noise and variability in the system, highlighting the probabilistic nature of gene regulation. Overall, this study deepens our understanding of the repressilator network and its potential applications in synthetic biology and biotechnology.

References

Flynn, T. C., Swint-Kruse, L., Kong, Y., Booth, C., Matthews, K. S., & Ma, J. (2009). Allosteric transition pathways in the lactose repressor protein core domains: Asymmetric motions in a homodimer. *Protein Science*, *12*(11), 2523–2541. https://doi.org/10.1110/ps.03188303

Gonze, D., & Ruoff, P. (2020). The Goodwin oscillator and its legacy. *Acta Biotheoretica*, *69*(4), 857–874. https://doi.org/10.1007/s10441-020-09379-8

Hochschild, A., & Ptashne, M. (1986). Cooperative binding of λ repressors to sites separated by integral turns of the DNA helix. *Cell*, *44*(5), 681–687. https://doi.org/10.1016/0092-8674(86)90833-0

Whipple, F. W., Hou, E. F., & Hochschild, A. (1998). Amino acid–amino acid contacts at the cooperativity interface of the bacteriophage λ and P22 repressors. *Genes & Development*, *12*(17), 2791–2802. https://doi.org/10.1101/gad.12.17.2791

Yun, C., & DasGupta, R. (2014). Luciferase reporter assay in Drosophila and mammalian tissue culture cells. *Current Protocols in Chemical Biology*, *6*(1), 7–23. https://doi.org/10.1002/9780470559277.ch130149