

Spatial modelling of transcription dynamics in bacterial gene regulatory networks



Ruud Stoof

School of Computing Science

University of Newcastle

A thesis submitted for the degree of

Doctor of Philosophy

July 2020

Acknowledgements

I would like to thank: My supervisors. Ángel, for his dedication to and optimism in our research. The many discussions, taking time and walking past an hour later with brilliant suggestion frequently. Anil, for treating me like one of his team members.

The people in ICOS. Fedor and Lewis, for the everyday discussions. Mostly about nonsense, sometimes tangent to our research. Alex, Huseyin, Bill and Nunzia for showing the wet-lab and taking the time to explain basic lab equipment and techniques. Harold, for the discussions about niche mathematics. Paolo, for feedback on progression. Paweł and James with their pleas for open-source software.

The supporting staff. Catherine, Andrew and Neil for the rapid handling of otherwise cumbersome paperwork. Jennifer, for taking the time to help with the grammar and spelling issues.

My Family and friends. Miron and Anna, making sure my lines are secure. Alma, for the many flights to visit and the emotional support. My parents, for providing support and believing in me.

Finally, I wish to express my gratitude for the time that Prof. Guy-Bart Stan and Prof. Natalio Krasnogor have taken to assess my work based on their vast experience.

Abstract

In Synthetic biology, researchers can alter the DNA sequence of organisms such that the behaviour to specific inputs is predictable. Regulatory systems have been ‘hacked’ into doing computation, help with bio-production, aid in personalised medicine and providing highly specific sensors.

A major bottleneck in current synthetic biology is that models fail to predict system behaviour reliably, causing recent progress to be reliant on the trial and error of model-assisted system designs.

One of the reasons for the models to fail is the neglect of Spatial effects. While this neglect simplifies models, recent experimental data shows localised effects.

This work shows that only the combination of 3D cytosol diffusion and the 1D sliding along the chromosome of transcription factors can explain localised effects; the modelling transcription factors initial sliding route after formation reproduces experimental results.

However, one essential assumption for the model described above is the initial location of a functional transcription factor at the encoding gene. While the coupled transcription and translation in prokaryotes are experimentally verified and can lead to the localisation of Transcription Factor proteins, this localisation must be assumed to be transferred to the active dimer form to reproduce the

Abstract

experiment.

To substantiate this assumption, this work expands the limited field of protein dimerisation. A new model is introduced to explain the localisation effect with an extra pathway we call Translation Mediated Dimerisation. Here, the partially formed transcription factors still undergoing translation are thought to meet and form a dimer while still constrained to the mRNA on the other end. Even if this occurs in a minority of events, this can drastically affect non-linear behaviour.

This model allows utilisation of localised effects for the rational design of system dynamics otherwise unavailable, expanding the possibilities and increasing the efficiency of synthetic biology.

Contents

Abstract

Contents	v
1 Introduction	1
1.1 Aims and problem definition	2
1.2 Thesis Outline	3
1.3 Summary of outputs	4
2 Background information	6
2.1 Synthetic biology	8
2.2 Nature designs co-localisation	9
2.3 Modelling	9
2.3.1 ODEs	10
2.3.2 Stochastic simulation	14
2.3.3 PDEs	15
2.4 Boolean based modelling	16
2.5 Alternative modelling approaches to Synthetic Biology	17
2.5.1 Network based modelling	17

CONTENTS

2.5.2	Multi-agent simulations	18
2.5.3	Stochastic discrete models	18
2.6	Related work	19
2.6.1	Empirical spatial transcription factor dynamics	19
2.6.2	Models for spatial dynamics of transcription factors	20
3	Spatial separation as a design parameter	22
3.1	Transcription Factor(TF) Diffusion modelling, 3D search	25
3.1.1	Case study: Spatial separation of a two-gene oscillator	26
3.1.2	Case study: 3D diffusion-only spatial repressilator	31
3.2	Transcription Factor Diffusion modelling; Facilitated diffusion	35
3.2.1	Local search	37
3.2.2	Influence of $k_{b,TF@TG}$, using k_{umr} when at cognate promoter	38
3.2.3	Global search	42
3.3	Transcription factor diffusion modelling; ODEs	43
3.3.1	Assumptions taken by the model.	45
3.3.2	Modelling the search of a transcription factor for its target	46
3.3.3	Quasi steady state mRNA	50
3.3.4	Binding rates	50
3.3.5	System sensitivity to intergenic distance modulation.	51
3.3.6	Nonlinear response regulation to intergenic distance.	57
3.3.7	Inducible system	58
3.4	Space as a design parameter.	58
3.5	Discussion	64
3.6	Future work	66

4	Spatial aspects of TF dimerisation	68
4.1	Related models on transcription and translation	69
4.2	Non-linear effects by Transcription factors	71
4.3	Results	72
4.3.1	New model: Translation mediated dimerisation	72
4.3.2	Engineering α , the translation mediated dimerisation fraction	80
4.3.3	Case study revisited: Effected damping in the repressilator	80
4.3.4	Case study: programmed damping in a genetic toggle switch	84
4.4	Discussion	87
4.5	Future work	88
5	Conclusion	90
	List of symbols	92
	Bibliography	99

CONTENTS

Chapter 1

Introduction

Organisms without a central nervous system seem simple; give a plant some water and sunlight, and it will grow. Bacteria look even more simple, as they appear only to need a carbon source to metabolise. While this predictable behaviour is apparently simple, these organisms thrive in a large range of conditions due to evolution over millions of years, which has led to complex adapting networks adjusting pathways depending on their environment. Over the last two decades, research has developed an understanding of these networks and the relation with the chromosome. Pathways have been drawn out and patterns of genes analysed, incrementally leading to the creation of models about regulation. While checking these models and creating tools for DNA adaptation, a new field arose: synthetic biology. This field alters the DNA sequence to engineer desired dynamics.

However, the networks that are altered are such complex systems that we currently need numerous (over)simplifications to begin to apprehend pathway dynamics and predict behaviour. It is good practice to revisit steps where simplification is undertaken and analyse what effect these assumptions have.

1. INTRODUCTION

1.1 Aims and problem definition

In previous works, it is seen that in bacteria, the location of regulatory genes relative to their target effect system dynamics. Moreover, low distances appear to arise through evolutionary pressure. The central problem in this thesis reads as follows:

While empirical observations in prokaryotes show that the spatial placement of a regulatory gene relative to its target affects its regulatory efficiency, standard modelling techniques do not account for these spatial characteristics. Therefore, standard implementations of synthetic networks are not as reliable as they could be.

In this thesis, I research this problem by introducing mathematical models describing the spatial dynamics of the gene regulatory systems. This work is split into two related mathematical models with a different focus.

- Using the first model, we can answer the following questions to get closer to solving the problem:

What is the reason that gene location is correlated in prokaryotes, where eucaryotes do not show similar such effects? What drives this difference between these groups? How does gene separation affect gene regulatory systems? What is the relation between the amount of gene separation and the advantage for prokaryotes? Finally, if gene separation alters system dynamics, can we use gene separation as a rational design parameter?

- The second model mostly solves a hurdle in the first model:

How can transcription factors reach their active state while remaining at

their encoding gene?

After answering that I dive into:

How Translation Mediated Dimerisation affect non-linear systems? And again, if Translation Mediated Dimerisation alters system dynamics, can we influence it?

1.2 Thesis Outline

- Chapter 2 defines the field of synthetic biology and the importance of [Transcription factors](#). Also, it introduces the modelling frameworks that will be used throughout the work.
- Chapter 3 introduces a spatial expression model for synthetic biology networks. It, first of all, shows how traditional modelling using partial differential equations can not capture in-vivo expression dynamics since there is a lack of modelling specifics and the time scales of different processes are non-comparable. Next, to account for this lack of modelling specifics in traditional modelling, the new model accounts for the spatial dynamics in a special way. The model is an ordinary system of differential equations; however, it differentiates between local [Transcription factors](#) that search along the DNA strand after formation and global [Transcription factors](#) that unbind at some point in the search trajectory to account for empirically observed impact of gene insertion location on regulation efficiency. While the modelling of global [Transcription factors](#) does not contain spatial information, the modelling for local [Transcription factors](#) does rely on the

1. INTRODUCTION

movement relative to the gene encoding it.

- Chapter 4 presents a conceptual model of the dimerisation of **Transcription factors**. In this model, **Transcription factors** can dimerise and fold to their “active” state while being translated. This enables the local **Transcription factors** modelling in Chapter 3, since this would allow the search along DNA directly after formation. This conceptual model takes into account a volume where partially formed proteins can interact with another that is being translated by another ribosome that is nearby. This interaction volume depends on the distance between ribosomes and the length of partially formed proteins.
- Chapter 5 summarises findings and contains final remarks

1.3 Summary of outputs

Portions of the work within this thesis have been documented in the following publications:

- Stoof, R., Wood, A., & Goñi-Moreno, Á. (2019). A Model for the Spatiotemporal Design of Gene Regulatory Circuits. *ACS Synthetic Biology*, 8(9), 2007–2016.[87]
- Stoof, R., & Goñi-Moreno, Á. (2020). Modelling co-translational dimerization for programmable nonlinearity in synthetic biology. *Journal of The Royal Society Interface*, 17(172), 20200561.[86]
- Tas, H., Grozinger, L., Stoof, R., de Lorenzo, V., & Goñi-Moreno, Á. (2021).

Contextual dependencies expand the re-usability of genetic inverters. *Nature Communications*, 12(1), 355.[93]

- Grozinger, L., Amos, M., Goroehowski, T. E., Carbonell, P., Oyarzún, D. A., Stoof, R., . . . Goñi-Moreno, A. (2019). Pathways to cellular supremacy in biocomputing. *Nature Communications*, 10(1), 1–11.[44]
- Stoof, R., Grozinger, L., Tas, H., & Goñi-Moreno, Á. (2020). FlowScatt: enabling volume-independent flow cytometry data by decoupling fluorescence from scattering. *BioRxiv*.[88]

Where Stoof, Grozinger, Tas, and Goñi-Moreno [88] is an application note being adapted for peer review publication about my software tool FlowScatt. FlowScatt is available at: <https://github.com/rstoof/FlowScatt>.

And I made a filtered data-set for Tas et al. [93] available [here](#).

Chapter 2

Background information

*Parts of this background were adapted from our perspective work, [44]
with doi: 10.1038/s41467-019-13232-z.*

The lack of nervous system in organisms forces all “choices” to be hard-coded in the chromosome. While DNA sequences determine these choices, it is not immediately apparent how genes are part of decision trees. The central dogma of biology, see figure 2.1, claims that the flow of genetic information is purely one directional. DNA is used for transcription of RNA which in turn is generally used for translation of proteins.

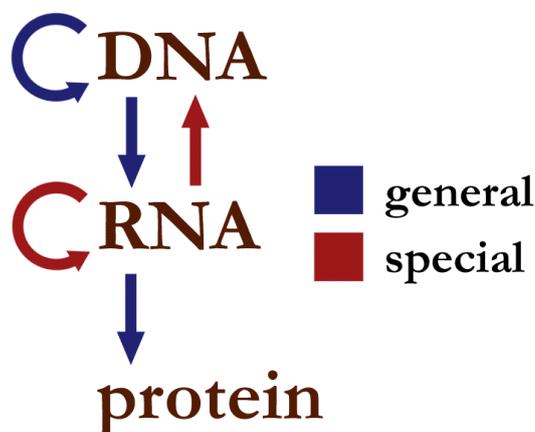


Figure 2.1: The central dogma of molecular biology

An important observation is that the transcription and the translation process are not independent of the environmental conditions. While the information that is available can not be influenced by processes downstream, the rate at which genes are expressed can be influenced.

To maintain homeostasis, cellular pathways often base their expression on surrounding conditions. These pathways are regulated by proteins that have a dynamic structure that is altered by a binding to a specific chemical species. The interaction with a network of gene regulatory components is what allows for the DNA encoded decision trees to express the desired dynamics.

2. BACKGROUND INFORMATION

Transcription factors are an essential part of these gene regulatory components. They have DNA-binding domains specific to certain sequences and when transcription factors encounter these sequences, they have a high binding affinity to them. Upon binding, the transcription factors are split in two categories based on the interaction with RNA-polymerase: they can either up-regulate (activators) or down-regulate (repressor) its binding.

2.1 Synthetic biology

With the ability to alter the chromosome came the ability to make use of these regulatory components; this enabled the engineering of organisms within the field of "Synthetic Biology". The field of man-made biology is nothing new; we have been selecting sheep for their wool, selecting less aggressive wolves to keep as pets and caused some chickens to grow at a staggering rate. What is new in "synthetic" biology is the methodology: whereas, for example, the selection of yeast was made via selection based on phenotypic behaviour, in synthetic biology the underlying process is modelled. Modelling the pathways enables us to know which genes to target.

A strength of using synthetic biology is the proteins that organisms make, specifically the sensory enzymes. The large complexes of amino acids allow an extensive range of possibilities and structures, organisms make use of a wide variety of proteins performing specific task(s).

While all proteins require folding to reach their secondary structure, some highly specific proteins have a binding site. When the binding site is used the structure changes, along with the interaction with gene regulatory components.

These kinds of proteins would also be advantageous to be used as highly sensitive sensors for the regulation of mechanical/electrical machines. However, the direct measurement of a structure reformation would be challenging. To make use of proteins as sensors, researchers need to set up an artificial gene circuit with behaviour that can be recognised by current measurement techniques. In a neat example, researchers [84] showed how to alter bacteria such that they can produce light when they sense the DNT leaking from buried land mines. This shows the extreme versatility of synthetic biology.

2.2 Nature designs co-localisation

Structural analyses on bacterial cells have helped to clarify this internal complexity, where a non-compartmentalised, but highly organised chromosome is compacted [23]. Rather than being randomly dispersed throughout the cell, the chromosome is organised into four large macrodomains and two *non-structured* domains [71, 95]. In addition to this, it is further organised into smaller, more dynamic microdomains [59, 99]. Such chromosomal structure is heavily linked to genetic function; it has been observed that co-regulated genes are often clustered and retained in close proximity, not only in terms of base pairs [49] but also considering the 3D folding of the chromosome [100, 105] .

2.3 Modelling

Mathematical modeling assists the design of synthetic regulatory networks by providing a detailed mechanistic understanding of biological systems. Models

2. BACKGROUND INFORMATION

that can predict the performance of a design are fundamental for synthetic biology since they minimize iterations along the design-build-test lifecycle. Such predictability aspects depend crucially on what assumptions (i.e. biological simplifications) the model considers.

2.3.1 ODEs

Differential equations describe how changes in time will happen when the current state is known (or vice versa, with using temporal changes and a previous state the current state can be determined). The simplest kind are called ordinary differential equations. This case omits spatial dimensions, a so-called 0-D system.

A very basic example for biology is the population size of bacteria. In the simplest scenario bacteria double in a characteristic time (τ_2). If the current population size is Pop

$$Pop(t + \tau_2) = 2Pop(t) \quad (2.1)$$

If we want to know the population size outside of these points, we will need to use a differential equation. If we say that for very small time differences (dt) the population is linear with time, we can say that at each point of time there exists division chance fraction the population (μ) times this time difference. These division events lead to an extra part of the population:

$$dPop(t) = \mu dt Pop(t) \quad (2.2)$$

We can rewrite this to

$$\frac{dPop(t)}{dt} = \mu Pop(t) \quad (2.3)$$

So we know that the derivative of the function Pop (i.e. $Pop'(t)$) is the function Pop itself times μ . The only functions that show this behaviour are exponential functions:

Hence,

$$Pop(t) = Pop(0)e^{\mu t} \quad (2.4)$$

From equation 2.1 we know that $Pop(\tau_2) = 2Pop(0)$, combining this with equation 2.4:

$$2Pop(0) = Pop(0)e^{\mu\tau_2} \quad (2.5)$$

hence,

$$\ln(2) = \mu\tau_2 \quad (2.6)$$

which sets the relationship between the fraction dividing at each time point and the doubling time.

While less intuitive than a doubling time, from here onwards I will use, for notation, a typical τ , where τ is $\frac{\tau_2}{\ln(2)}$. This is easier to combine with other rates as well as allows for more compact and aesthetic differential equations, as seen in 2.7:

$$Pop(t) = Pop(0)e^{\frac{t}{\tau}} = Pop(0)e^{\ln(2) * \frac{t}{\tau_2}} = Pop(0)e^{\mu t} \quad (2.7)$$

While this example shows how non-limited cell populations grow, in biology

2. BACKGROUND INFORMATION

we are often interested in how different (chemical) species interact. To model this in time we can have differential equations that depend on other differential equations, called a system of differential equations. These systems can capture the multitude of interactions in biological environments much more accurately than single differential equations, but are often harder to solve. Luckily, combining simplifying conditions and the usage of numerical simulations, widens the ability to observe the model behaviour within a range of system states within a limited time interval.

A neat example of how a set of differential equations can model real-live interactions are the Lotka-Volterra system of equations.

An example of a Lotka-Volterra system is if we look the population number of Hares (prey), H , which are hunted by foxes (predator), F . We can write the equation for the population of hares as.

$$\frac{dH}{dt} = k_b H - k_{d,H} H - k_h H F \quad (2.8)$$

where

H is the The population of Hares

F is the The population of Foxes

The change in population is dependent on, the first term, a breed rate (k_b) and current population size (H) and, second term, the natural death rate of the population ($k_{d,H}$). The interaction with another equation comes in the third term $-k_h H F$, where the amount of caught hares depends on the number of hares, the number of foxes (F) and a rate describing how efficiently the foxes hunt the Hares (k_h). This equation can be solved, given a constant number of foxes.

However, the Fox population can also change in time:

$$\frac{dF}{dt} = k_h HF - k_{d,F} F \quad (2.9)$$

In this equation the population of foxes can only grow when the foxes catch hares. Also these foxes die with natural death rate $k_{d,F}$.

If we plug in all parameters, we still need the initial population numbers to fully describe the system. In figure 2.2 the initial number are 20 hares and 10 foxes (at $t = 0$). Now we can, using the equations from before, calculate the population numbers for a certain time interval.

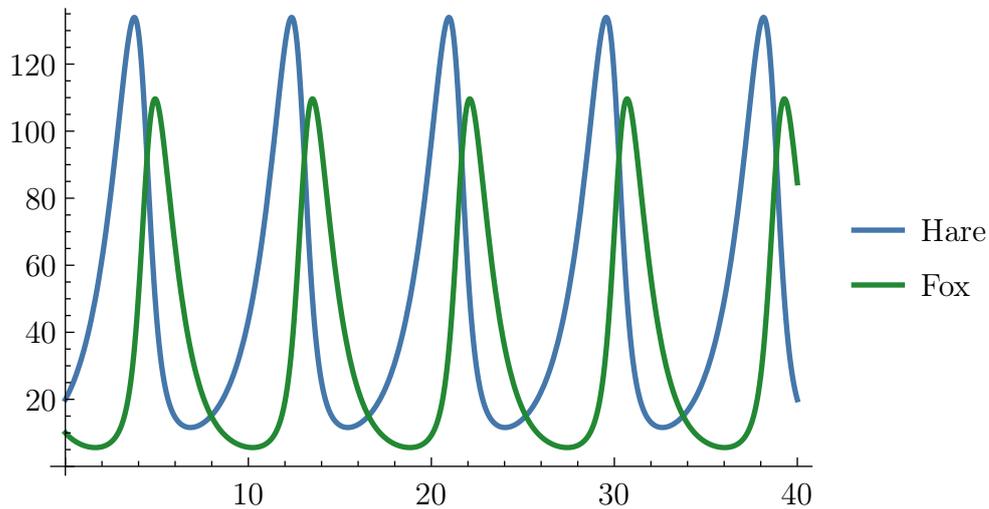


Figure 2.2: Time series simulation of the ODEs describing the predator prey system, where for certain parameters and inertial conditions the population numbers are seen to have periodic solutions.

2. BACKGROUND INFORMATION

2.3.2 Stochastic simulation

In the fox and hare system described in the previous section, the population numbers are assumed to be continuous, which in this case is incorrect. To consider that species are made up by individuals, population numbers change in discrete steps. A way to still adhere to the chemical master equation is to use the Gillespie algorithms.

Gillespie [33] shows a very simple computer algorithm based on chemical reaction rates, which is able to simulate these kinds of discrete processes. An example of a solution to the fox-hare problem as stated above, simulated over time, is given in figure 2.3.

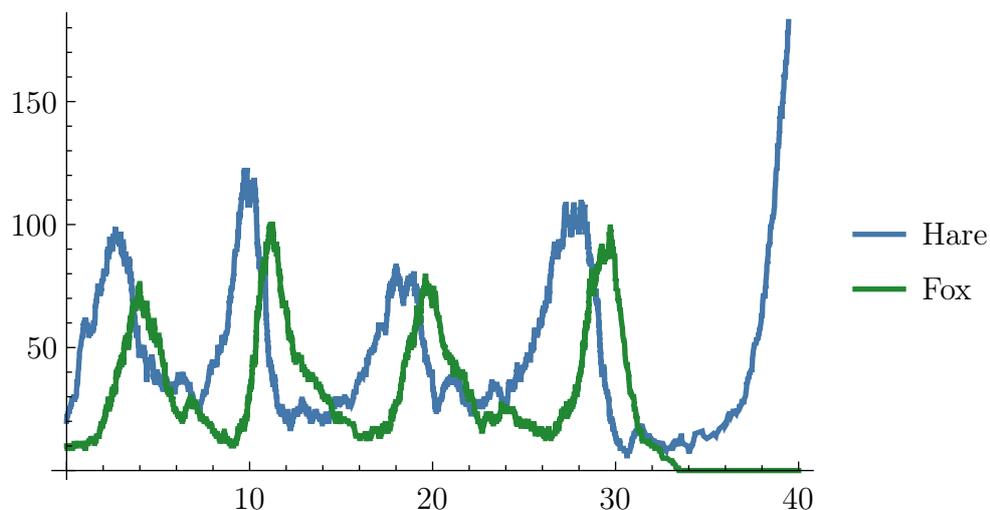


Figure 2.3: Gillespie simulation of the predator prey system. Note that, unlike the ODE model, here when the fox population drops below 1 it does not recover.

The Gillespie algorithm exists of two steps: first it takes the current numbers of populations and the differential equations from the previous sections to calculate the rates at which specific reactions fire. It uses the total of these rate to

determine the time step it will take. On average however a higher total will lead to a smaller time step. However, to take into account the stochastic nature of the reactions, the time step is weighed by a number drawn from a random number generator. The second step is that it takes another random number to pick which reaction to fire weighed by the magnitude of their rates.

2.3.3 PDEs

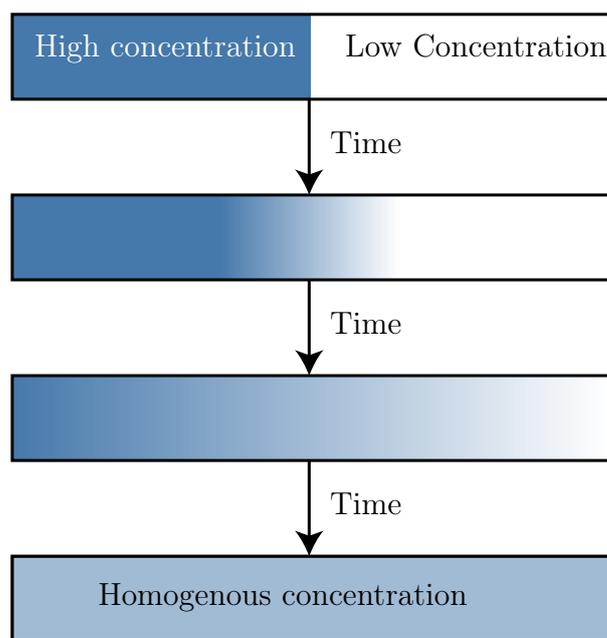


Figure 2.4: Schematic of dynamic of 1D diffusion. The second law of thermodynamic drives the system towards a state with higher entropy. The rate at which this happens is determined by Fick's law.

In general, diffusion can be thought of as the mixing of two different chemical species. On the macroscopic scale it is the tendency of a chemical species to

2. BACKGROUND INFORMATION

move to a lower concentration. This is described in Fick's first law:

$$j(\mathbf{x}, t) = -D\Delta\rho(\mathbf{x}, t) \quad (2.10)$$

where

j is the The flux of a chemical species

D is the Diffusion constant

Δ is the divergence operator

ρ is the Chemical species concentration

This equation describes chemical diffusion very well on the large scale, where one can generally observe the macroscopic properties of chemicals.

On the scale of bacteria, however, this is not always the case. They are so small that the paths individual molecules take can influence the whole organism. In the main body of this work, I will use both the empirical macroscopic laws, such as 2.10, as well as stochastic simulations on the molecular scale.

2.4 Boolean based modelling

A popular approach in synthetic biology is to draw analogies with electronic engineering, a field that over the decades has developed tools for and an intuition in engineers. To build computers, in the semiconductor industry, the Turing complete NAND-gate (only negative when both inputs are negative) is mostly used. A common goal is to accomplish this gate behaviour in living organisms as well. There are several such gates currently known to work, albeit at very specific environmental conditions. These gates consist of a specifically chosen combination of multiple promoters, one or more expressing a transcription factor

regulating the others. For example, an inverter can be made by a regulatable promoter expressing a repressor repressing a repressible promoter. In such a case, if the input promoter is in a "High" state, it represses the other promoter to be in a "Low" state.

Theoretically, one can connect two such gates together and combine the truth tables to get the new dynamics of the system. Moreover, cells with circuits that do computing have been theorised and very simple circuits have been successfully implemented in vivo [19]. However, dynamics arising from such designed circuits often do not comply to the logic of the combinatorial truth tables.

2.5 Alternative modelling approaches to Synthetic Biology

While this work will mostly approach synthetic biology with the methods described above, alternative approaches are commonly used. In the next section, I will touch on some of these alternative approaches that might benefit the reader.

2.5.1 Network based modelling

When working with the vast amounts of data obtained with sequencing techniques, network-based modelling can be fairly efficient. Here, results are not simulated from the ground up, but the model analyses relations between data points. Network analysis algorithms can cope with vast assays of different genetic components[4]. These components are combined from other species, strains and along the phylogenetic tree. Often these genes are clustered in sequence

2. BACKGROUND INFORMATION

similarity, function, co-evolution and expression pattern to find analogous components [3, 83]. These can then iteratively be exchanged to vary the system and approach desired system dynamics.

2.5.2 Multi-agent simulations

In this work, I use single instances of my models. On the population scale, however, multiple individuals have to be modelled. A naive solution would be to run single individual models, like those introduced in this work, multiple times. This naive approach, however, is unable to capture the interactions between individuals. To capture these interactions, multiple instances (or agents) of the same model are set up containing coupling terms between each agent [68]. Frequently, the magnitude of the coupling terms depends on how "close" [46] the individuals are thought to be. Not only could this be applied to solve multi-cell systems, but one might implement this to solve systems with multiple plasmids [57], proteins [43], organelles [18] or other multi-copy components. In short, multi-agent simulation can be a valuable tool to capture the behaviour that arises from coupled systems.

2.5.3 Stochastic discrete models

Section 2.3.2 refers to the traditional Gillespie algorithm to capture the discrete steps in molecular chemical reactions. However, this is a slow and relatively computationally expensive approach. While the Gillespie algorithm guarantees provide one of the stochastic solutions to the chemical master equation, the number of computations it needs is linear with the number of reactions. This linear

computational cost limits the suitability to concentration trajectories of single (low expression) chemical species. However, when attempting to capture common reactions on a large scale, a less expensive method is needed, such as the tau leaping method [34] or the Next Reaction Method [32]). While the tau leaping method exceedingly speeds up computation, it compromises by only approaching an exact solution of the chemical master equation. In this thesis, I only use the simulation on small scales, such that my implementation of the inefficient Gillespie algorithm suffices. Still, if an exact solution is desired and computation efficiency is a concern, a superior alternative would be to use the Next Reaction Method.

2.6 Related work

2.6.1 Empirical spatial transcription factor dynamics

Goñi-Moreno et al. [40] show that there is an influence in expression dynamics, dependent on where a gene is located in the cell. If the regulation and regulated genes are both located on the chromosome, it is seen that the genetic up-regulation is done in a constant and predictable way. If the regulator gene expression comes from a gene in the plasmid, they see significantly different dynamics. The up-regulation in this case is noisy, and varies between the formerly up-regulation given by placement in the chromosome and base expression. The hypothesis here is that the spatial location (which was different from the chromosome and the plasmid) causes this noise effect. It is also observed that genes that work together are seen to cluster together [100], possibly because of this effect.

2. BACKGROUND INFORMATION

For instance, transcription factors are not homogeneously distributed within the volume of a cell [51], but are sometimes localised around their interacting DNA components, e.g. their coding gene and cognate promoter [24].

2.6.2 Models for spatial dynamics of transcription factors

The coupled nature of transcription and translation in prokaryotes [54] is theorised to produce locally high distributions of transcription factors near the site of their expression [55], making co-localisation of transcription factors' coding genes and their target promoters a potential evolutionary solution for the tight control of protein production.

In this scenario, the physical separation of co-regulated genes is revealed as a key parameter for such expression control. It has been shown, for instance, that the further away from the transcription factors' coding gene that the target is located in the chromosome, the less effective the regulation is [58]. Therefore, the strength of a regulatory interaction, either repression or induction, may change with increasing or decreasing this relative distance.

While the field of molecular dynamics is growing quickly in possibilities, it is mostly the descent of the energy landscape from the primary structure to the secondary structure (folding), that gets taken into account. This challenge used to be exceptionally computationally expensive, but is closer to being viable today. These folding dynamics can explain how the transcription factors are highly specific for a genetic sequence, and can explain the binding by specific domains. True molecular dynamics for transcription factors over their whole lifetime, however, remain a distant prospect.

A major aspect for the efficiency of a transcription factor is the search for the target. [Winter et al.](#) show that this search is not a simple process, and it cannot be explained by classical chemistry. Transcription factors are seen to bind to their target at a rate above the limit for 3D-diffusion limited reaction. The DNA binding domain in transcription factors catalyses the search for the target gene. Transcription factors are thought to have two ways of binding to the DNA: one low energy bond that is non-specific to the underlying sequence, and one high energy bond with a very specific DNA sequence. It is thought that the low energy allows lateral movement, guiding the search along the genes.

Some claim that there is no possibility of this behaviour from one gene to another. Moreover, some experiments do not show the need for co-localisation [\[25\]](#). However, experimental results are still inconclusive and contradict each other. In [\[106\]](#), the authors show a model where transcription factors have a lower search time for promoters that are physically closer to where the transcription factor is bound. What is missing however, is how this affects gene circuits. Since this is the only work describing this effect, I will build on this model in [chapter 3](#) to describe how to use space as a design parameter.

Chapter 3

Spatial separation as a design parameter

This chapter is an adaptation of our previously published work, Stoof, Wood, and Goñi-Moreno [87], where I developed the theory, made the models and ran the simulations.

Problem definition: It was previously seen that the spatial separation of two interacting genes affects the dynamics arising from the system. This chapter attempts to capture this influence in a mathematical model, where I show that 3D diffusion alone does not give rise to empirically observed outcomes. I introduce a mathematical model based on facilitated diffusion that *is* able to reproduce these experiments.

When making synthetic constructs, in most cases, sequences are synthesised in plasmids, these plasmids then are to be transformed into the organisms. While plasmids are a quick and reusable way to include the desired genes, plasmids behaviour is different from chromosome inserts in several ways. 1) The copy number of plasmids causes copy number variation. 2) The antibiotics used to keep the plasmid in could affect dynamics, and 3) there is a lack of control of these genes spatial location. This chapter focuses on this third aspect. *This chapter introduces a model that explains the spatial effects of transcription.* I show how the inclusion of genes on the chromosome on a specific location can improve these insertions reliability.

Despite the small volume of a bacterial cell, it has become increasingly apparent that spatial constraints have significant implications for bacterial function [89]. There has been a lot of attention for the spatial dynamics of bacteria on the multi-cellular scale, where bacteria can communicate to form patterns and allow the formation of so-called bio-films [29, 38, 39, 80]. However, on the single-cell level, less research is carried out. Even though back in 2004, when genome sequencing was in its early days, the importance of spatial constraints and its implications for bacterial function as already observed [100]. A supporting argument of the significance of spatial constraints for bacterial function can be found in nature. Where related genes are more likely to be co-located, not only in base pairs [49] but also in considering the 3D folding of the chromosome [100, 105]. Traditional assumptions for spatial effects can not explain this co-localisation.

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

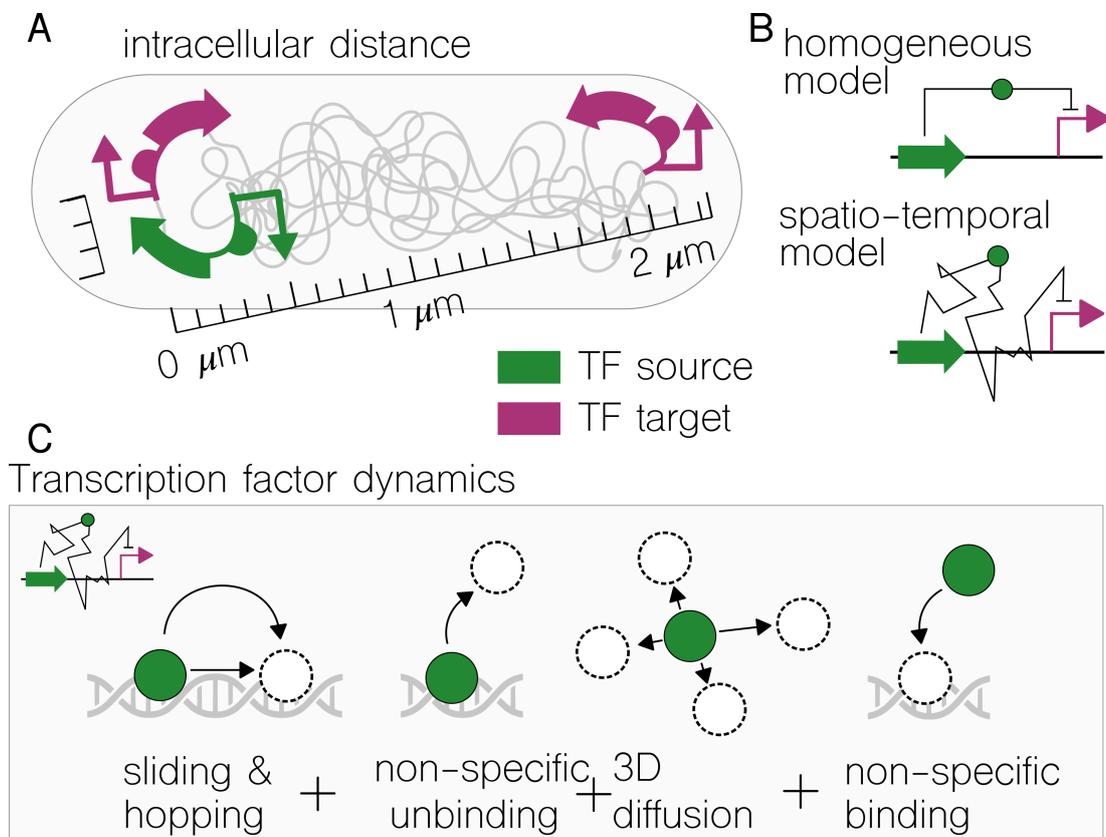


Figure 3.1: Spatiotemporal principles of gene regulation. **(A)** Diagram of three chromosomal insertions: one source of transcription factors (green) and two target regions (purple). One of the target regions is co-located with the source; the other one is spatially separated. This is indicated by rulers along the different combinations. **(B)** A Schematic comparison between *homogeneous* and *spatial* models. While the former assumes that transcription factors are *instantly* available to bind, the latter explicitly simulates the *travel* from source to target. **(C)** Transcription factor dynamics through facilitated diffusion: 1D sliding, hopping, non-specific un-/binding and 3D diffusion. The model adds these dynamics to promoter activity and gene expression events.

Breaking with traditional assumptions of homogeneous distribution complicates the models describing dynamics. To accurately describe the physiological distribution of Transcription Factors involves explicitly modelling 1D diffusion (hopping and sliding) along the chromosome, non-specific interactions (i.e. non-target DNA regions) and 3D diffusion across the cytoplasm [6].

Figure 3.1 shows the spatio-temporal principles that govern gene regulation within the model. Fundamental to this approach is the time it takes for a transcription factor to go from its encoding gene, where it is expressed, to its cognate promoter, where the transcription factor binds. Therefore, these two regions are referred to as the *source* and the *target* respectively. Since transcription factors must actively reach the target, the distance between them (i.e. base-pairs separation in the chromosome) is the key parameter that the model revolves around (Figure 3.1A). This implies that transcription factors will *not* be automatically available to bind to the target after expression, which is the customary view of what I refer to as *homogeneous* models [36, 37] (Figure 3.1B). Importantly, such homogeneous models are inconsistent with having intra-cellular spatial dynamics. Specifically, the basic dynamics of transcription factors searching for the target need to be considered, as summarised in Figure 3.1C: sliding and hopping (combined in our model), 3D diffusion (or *global search* as it is explained next) and non-specific (un)binding to DNA regions that are not the target.

3.1 TF Diffusion modelling, 3D search

As stated before, cells are generally modelled assuming homogeneously mixed contents. This assumption is based on the time constant needed for spatial effects.

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

The reasoning behind this assumption of homogeneously mixing is as follows: Degradation of proteins is typically in the order of hours. In comparison, the typical time, τ , it takes to spread a typical distance of Δx

$$\tau \propto \frac{(\Delta x)^2}{2D}. \quad (3.1)$$

with some small proportionality constant and diffusion constant D . With a rule of thumb diffusion coefficient of $100 \frac{\mu m^2}{s}$ [75] for proteins in water, spatial effects in E.Coli for example decay in $\sim 1s$. Since the cytoplasm consists of around 80% water, this seems like a reasonable value for the diffusion constant. This would indicate that since spatial effects are much shorter than protein degradation, it can therefore be neglected. However, the diffusion of proteins in cells is more complex in cells than in water, affecting the diffusion constant. This increase of the diffusion constant leads to an increasing the time introduced by the spatial effect and therefore spatial effect cannot be neglected.

In this section, I show how standard modelling of diffusion within the cell cannot explain the observed localisation effects as described in 2.2.

3.1.1 Case study: Spatial separation of a two-gene oscillator

As shown in section 2.3.3 it is relatively easy to set up a system of partial differential equations to model spatial effects. A way to study how a method behaves is to look at it in an oscillatory system. In physics, we often use the harmonic oscillator because of its simplicity. In synthetic biology, there is a similar system, the two-gene oscillator[45]. Instead of a phase shift between speed and position,

due to the exchange of potential and kinetic energy leading to, in the two-gene oscillator there is a phase shift in the first and second genes expression due to the repression of the first and the activation of the second gene.

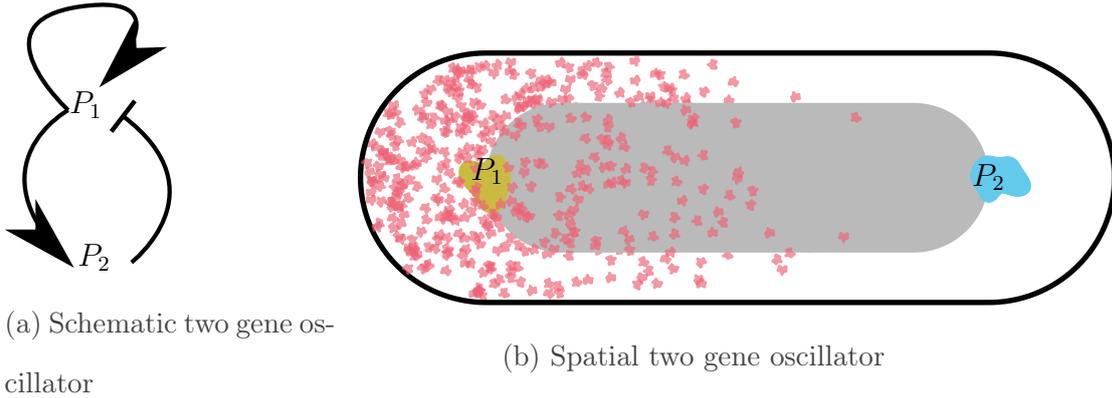


Figure 3.2: **(a)**: Schematic of the two gene oscillator. The “activator” gene with promoter P_1 expresses a transcription factor which increases the expression of P_2 . The “repressor” gene with promoter P_2 expresses a transcription factor that lowers the expression of P_1 . This leads to an oscillatory interaction between the genes. **(b)**: The spatial separation between two genes in the two-gene-oscillator, here in opposite poles of a coli bacterium shaped volume. This is a 2D representation, however full 3D simulations were conducted to get figure 3.3

This leads to a simple set of partial differential equations where:

β_{transl} is the **Rate of transcription**

β_{transc} is the **Rate of translation**

k_{unr} is the **Rate of specific unbinding**

k_{bind} is the **Binding rate**

k_{Dm} is the **mrna degradation rate**

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

k_{Dp} is the Protein degradation rate

D is the Diffusion constant

\mathbf{P} is the Concentration field of the non-inhibited promoters

\mathbf{iP} is the Concentration field of the inhibited promoters

\mathbf{m} is the mRNA concentration field

\mathbf{TF} is the TF field

and

∇^2 is the Laplace operator

The activator promoter(P_1) is repressed to an inactive state(iP_1) by a repressor dimer:

$$\partial t \mathbf{P}_1(t) = k_{\text{unr}} \mathbf{iP}_1(t) - k_{\text{bind}} \mathbf{TF}_2(t)^2 \mathbf{P}_1(t) \quad (3.2)$$

$$\partial t \mathbf{iP}_1(t) = -k_{\text{unr}} \mathbf{iP}_1(t) + k_{\text{bind}} \mathbf{TF}_2(t)^2 \mathbf{P}_1(t) \quad (3.3)$$

The inactive repressor promoter(iP_2) is induced to an active state(P_2) by a activator dimer:

$$\partial t \mathbf{P}_2(t) = k_{\text{bind}} \mathbf{TF}_1(t) \mathbf{iP}_2(t) - k_{\text{unr}} \mathbf{P}_2(t) \quad (3.4)$$

$$\partial t \mathbf{iP}_2(t) = -k_{\text{bind}} \mathbf{TF}_1(t)^2 \mathbf{iP}_2(t) + k_{\text{unr}} \mathbf{P}_2(t) \quad (3.5)$$

mRNA is transcribed from the genes when the promoter is in the active state.

It is modelled to diffuse through the cell and decay over time.

$$\partial_t \mathbf{m}_1(t) = \beta_{\text{transl}} \mathbf{P}_1(t) - D \nabla^2 \mathbf{m}_1(t) - k_{\text{Dm}} \mathbf{m}_1(t) \quad (3.6)$$

$$\partial_t \mathbf{m}_2(t) = \beta_{\text{transl}} \mathbf{P}_2(t) - D \nabla^2 \mathbf{m}_2(t) - k_{\text{Dm}} \mathbf{m}_2(t) \quad (3.7)$$

And finally the transcription factors are created from the mRNA concentrations. This is also modelled to diffuse through the cell and decay over time.

$$\partial_t \mathbf{TF}_1(t) = \beta_{\text{transc}} \mathbf{m}_1(t) - D \nabla^2 \mathbf{TF}_1(t) - k_{\text{Dp}} \mathbf{TF}_1(t) \quad (3.8)$$

$$\partial_t \mathbf{TF}_2(t) = \beta_{\text{transc}} \mathbf{m}_2(t) - D \nabla^2 \mathbf{TF}_2(t) - k_{\text{Dp}} \mathbf{TF}_2(t) \quad (3.9)$$

In short, the one gene indirectly inhibits the other gene which in its place induces the original gene. Since the system is non-linear, it is not easily solvable. The specialised PDE solver FEniCS can solve this system using the Finite Element Method (or FEM). I implement Neumann boundary conditions and trivial initial conditions except for \mathbf{iP}_1 and \mathbf{P}_2 , which have unit probability divided over two spheres of radius 0.09 at $x=1$ at -1 respectively to get figure 3.3.

In figure 3.3, the limit cycle is determined by running a time series simulation of the ODEs given until a periodic solution arrives. Since the diffusion value is not known, it is varied in a big range, from $10^{-3} - 10^1 \frac{\mu\text{m}^2}{\text{s}}$. In the limit where diffusion is fast, the limit cycle arrives as described by a similar well stirred solution, as seen in the regime between $10^0 - 10^1 \frac{\mu\text{m}^2}{\text{s}}$. In the limit where the diffusion is slow, the Transcription Factors degrade before being able to arrive at the other

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

gene, thus not allowing up or down regulation. This is seen in the case around $10^{-3} \frac{\mu\text{m}^2}{\text{s}}$ where there is barely any down regulation of gene one and barely any up regulation of gene 2.

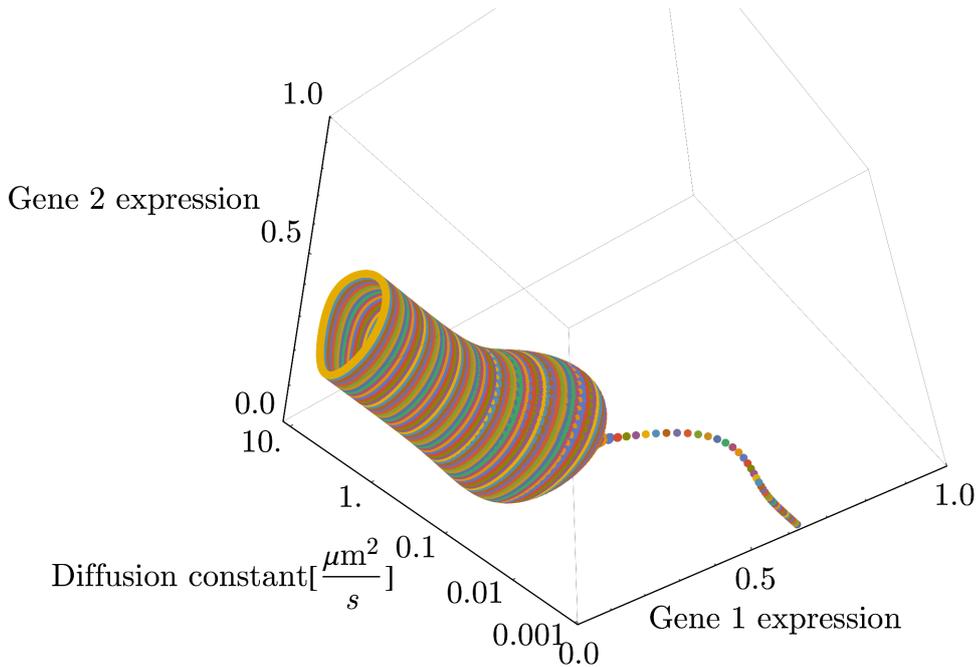


Figure 3.3: The oscillations of the two-gene oscillator using non-physiological parameters for different diffusion rates. Each coloured line is a time-series simulation after reaching a periodic solution. It is seen that the diffusion is only limiting the oscillation up to a value of around $0.1 \mu\text{m}^2/\text{s}$.

While this model shows that the dynamics of the system are dependent on the 3D diffusion of transcription factors, the increase of diffusion constant does only effect dynamics until a point of $0.1 \mu\text{m}^2/\text{s}$. Since physiological diffusion is expected to be higher than this value, I conclude that the 3D spatial model in this example does not add more information than a well stirred model would. To account for spatial effects another model is needed. This is similar to what I find

using the same technique on the repressilator.

3.1.2 Case study: 3D diffusion-only spatial repressilator

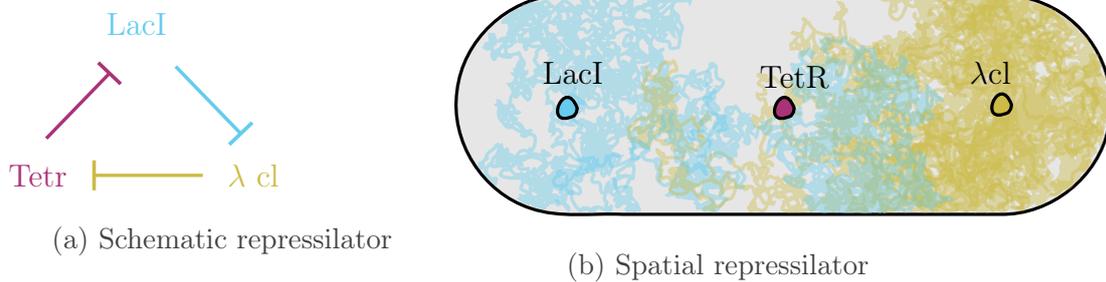


Figure 3.4: **(a)**: Schematic representation of the cyclic repression of the repressilator circuit. This circuit exists of three genes, all of which express transcription factors repressing the promoter of one of the other two genes. The gene encoding the Tet repressor protein starts with the cI promoter and expresses the Tet repressor. The Tet repressor proteins down-regulate the expresses Lac inhibitor protein, since this gene is altered such that it starts with the Tet promoter. The same technique is used such that the Lac inhibitor proteins repress the gene that expresses the Lambda phage repressors. Since the Lambda phage repressors bind to the cI promoter already mentioned this causes a circle of negative feedback. **(b)**: 2D schematic of the 3D spatial simulations of the repressilator. The circuit is placed in a rod shaped volume representing a coli-type bacterium with the two promoters expressing LacI and Lambda phage at opposite poles and the TetR promoter in the middle. The 2D diffusion trajectories represent the 3D simulations resulting in quantised spatial simulation as shown in Figure 3.5

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

The repressilator is a genetic design in which three different parts circularly repress each other, i.e. the first promoter encodes a transcription factor that represses the second, the second represses the third and the third represses the first. This leads to sustainable periodic oscillations between the three species with a $\frac{2\pi}{3}$ phase difference. This system was first shown in [30] and was a starting point for synthetic biology. Although the system expresses a sustainable oscillation and some degree of periodicity, there have been genetic designs improving the amplitude of oscillation [76] and an increase of periodicity. What exactly causes the deviation in the original system is not known.

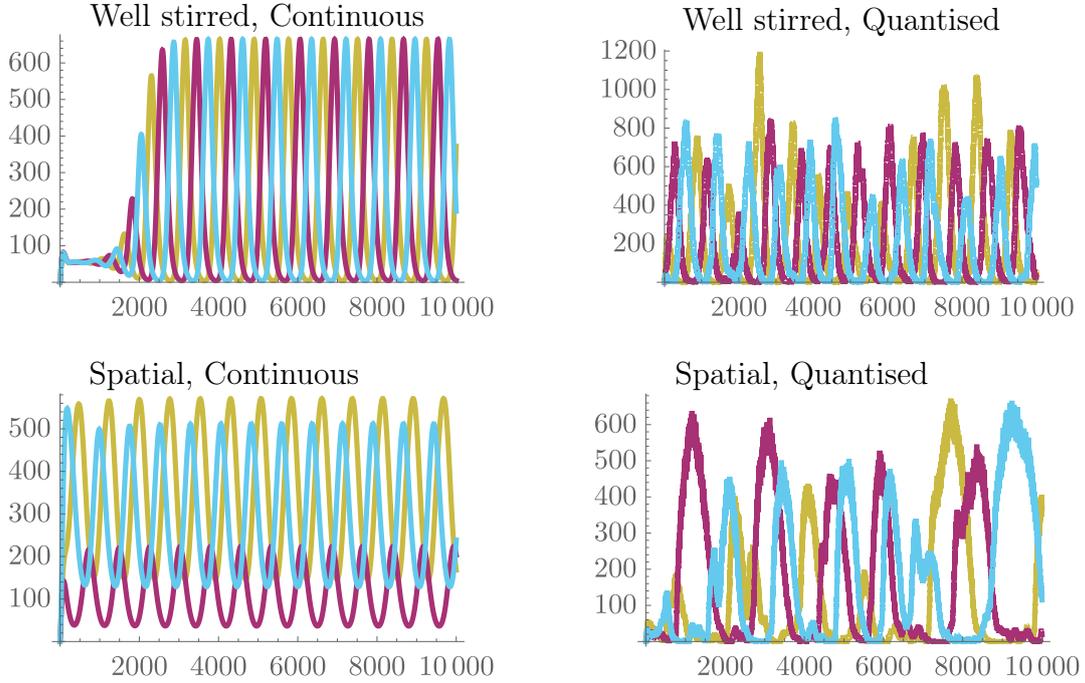


Figure 3.5: Comparison of different simulation methods of the spatial repressilator with an artificially low diffusion constant. The classical well stirred simulation techniques do not allow for (intricate) spatial information as described in Figure 3.4. This results in no differentiation of expression dynamics except for initial conditions for the continuous simulation and fluctuations in expression dynamics in the quantised system to be explained by chance. However for the spatial continuous simulation, since the diffusion distance of the λ cl transcription factors to repress the TerR gene its expression is shorter and to have a lower magnitude. Similar dynamics were found in quantised spatial simulations (Moving simulation available [here](#)). However due to the random nature of this simulation it is harder to discern.

Let us look at an extreme example of a spatial repressilator design. We place two parts of the repressilator in the opposite poles of an E.Coli cell ($\sim 2\mu m$) and

3. SPATIAL SEPARATION AS A DESIGN PARAMETER



(a) Simulation set-up

(b) Time series simulation at different locations compared to the well stirred model.

Figure 3.6: Simulation showing how localised effects due to 3D diffusion are transient. When using physiological values, see Table in Appendix I

one part in the middle. We set up a partial differential equations system in an analogous manner as shown for the two-gene-oscillator shown above. We know that the transcription factors are created at the spatially defined parts of the cell (we assume fast degradation of RNA). The only parts that are now important are diffusion towards the gene and the degradation of the transcription factor. If we model this process using a high parameter for degradation and a low value for the diffusion constant, we should see the maximal effect of 3D-localisation. However, even in this scenario, a homogeneous mixture is reached after only a short delay.

This delay of around 20 seconds for the $2\mu m$ as seen in Figure 3.6b is short relative to the lower bound of the protein degradation time. If we put these delay times in a delay differential equation, we see no significant effect.

To conclude, the 3D diffusion of transcription factors is likely on a much shorter timescale than other system dynamics. Therefore the seen spatial effects are most likely caused by a different mechanism.

3.2 Transcription Factor Diffusion modelling; Facilitated diffusion

To explain spatial effects in bacteria, I investigate another aspect of transcription displacement: the 1D diffusion along the chromosome. While it is widely accepted that the 1D diffusion is needed to explain the incredible efficiency of the transcription factor search, how this affects the spatial distribution remains poorly understood [16, 103, 104]. This transcription factor distribution is known to be higher at the location of both target and source genes [24, 51]. These localised distribution effects have yet to be described with mathematical models. I developed on previous models of *facilitated diffusion* dynamics [65, 106], that add the space-dependent 1D local search [7] to the space-independent 3D diffusion of molecules [7].

Transcription factors are known to have a DNA binding domain. This part of the protein is highly developed to have a structure that is specific in binding to the target gene combination; the binding energy causes the unbinding rate to be very low. Moreover, these binding domains are known to have non-specific binding energies as well. This causes the Transcription factors to (temporarily) remain bound to the chromosome with the DNA binding domain. While the domain is radially bound to the chromosome, it is able to move along the length of the chromosome.

This mechanism of movement along the chromosome is called “sliding” and, along with short sections of 3D-diffusion (hopping), this is thought to affect how a Transcription factor finds its target significantly. The reduced search-space is thought to catalyse the search by orders of magnitude, albeit only in combina-

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

tion with 3D diffusion[104]. The distribution of transcription factors along the chromosome is therefore central to the model.

To analyse this distribution, I examine the location of transcription factors during local search upon their generation from the source. Figure 3.8 records the trajectories of individual regulator molecules along the chromosome while performing 1D diffusion (sliding + hopping) until they unbind into the cytosol. Given the coupling of transcription and translation in bacteria [11, 42, 64], it is safe to assume that newly generated regulators start their search from (nearby) the source coding region (0bp in the graph). The visualisation of 1D movement leads to two conclusions. First of all, the timescale for the local search is of the order of seconds, with many transcription factors spending less than a second bound to the local chromosome. The second conclusion is that the local region is crowded with transcription factors searching for their target, which means that any regulatory interaction will be stronger and/or more efficient within this chromosomal segment. Therefore, the density of transcription factors at a given chromosomal region is correlated to the strength of the regulatory interaction. The cumulative effect of local and global search on such density is shown in Figure 3.7. Global search generates a uniform distribution (i.e. the same value for all chromosomal locations). In contrast to this, the local search leads to a higher concentration of transcription factors around the source region. This suggests a crucial role in the location of the source for determining transcription factor distribution [58].

3.2.1 Local search

In the model by Wunderlich and Mirny [106], transcription factor binding dynamics are modelled with a first-time-passage process[78] to get a typical time, $\tau_{s,1D}$, which a transcription factor needs to find its target due to 1D diffusion:

$$\tau_{s,1D} = \frac{\Delta x}{2\sqrt{D_{1D}/k_{un}}} \quad (3.10)$$

where k_{un} is the [Non-specific DNA unbinding](#)

Δx is the difference binding location and the location of the cognate promoter and D_{1D} is the [1D-diffusion rate](#)

In my model I look at the initial sliding event after formation, where it starts at the gene encoding the Transcription factor “locally” searching along the chromosome until they find the cognate promoter. Reusing equation 3.10 to:

$$\tau_{s,local} = \frac{d}{2\sqrt{D_{1D}/k_{un}}} \quad (3.11)$$

where d is the [Distance along chromosome](#)

It has been previously [96] observed that, after unbinding, transcription factors are likely to rebound after a small 1D-diffusion trajectory. A suggested method to cope with this is to re-scale the unbinding rate to account for these effects. However, if we try to look at this effect in the model suggested by Wunderlich and Mirny [106], there is no chance for transcription factors escaping. Diffusion around the target gene would lead to immediate rebinding, and the transcription factor remains bound over its lifetime. Instead of using the first time passage

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

effects, I suggest that there is a rate of binding when a transcription factor is located at the target gene, $k_{b,TF@TG}$.

3.2.2 Influence of $k_{b,TF@TG}$, using k_{unr} when at cognate promoter

In this section, I show how, while equation 3.11 is accurate for even slight separations, there can be an (effective) specific unbinding rate (k_{unr}) in the case where $d=0$. To allow for this rate, I consider a rate of binding at the target ($k_{b,TF@TG}$).

This rate represents the attraction to the minimal free energy state, depending on the interaction this rate of binding to the target when at the target location is high, but not infinite since the difference in free energy is finite. Since immediate binding is assumed in Wunderlich and Mirny [106] this can only be a lower bound on the local search time in the limit assuming infinite local binding rate, $\lim_{k_{b,TF@TG} \rightarrow \infty}$. In the other limit, $\lim_{k_{b,TF@TG} \rightarrow 0}$ where there is no binding, one can determine the upper bound on the search time. This search time is counter-intuitive measure (as with many limits), that if we do assume that a transcription factor binds, how long we expect it to have take if we assume that the the binding is very low.

In the case where $k_{b,TF@TG} = 0$ the transcription factor freely diffuses along the chromosome until it unbinds. This leads to the concentration distribution depends on the 1D diffusion constant and the rate of non-specific unbinding. This can be captured the following partial differential equation that describes the

density of transcription factors in time and space:

$$\frac{\partial \rho(x, t)}{\partial t} = D_{1D} \frac{\partial^2 \rho(x, t)}{\partial x^2} - k_{\text{un}} \rho(x, t). \quad (3.12)$$

where

ρ is the Chemical species concentration

D_{1D} is the 1D-diffusion rate

k_{un} is the Non-specific DNA unbinding

For the case of a single transcription factor where $t = 0$ is the time of transcription factor expression, and $x = 0$ is the location of the source gene, the initial condition is expressed by the Dirac delta (δ).

$$\rho(x, 0) = \delta(x) \quad (3.13)$$

It is defined to be zero anywhere except at the origin, along with the fact that integration over the origin gives a unitary value. Solving equation 3.12, with initial condition 3.13 gives:

$$\rho(x, t) = \frac{e^{-\frac{x^2}{4D_{1D}t} - tk_{\text{un}}}}{2\sqrt{\pi}\sqrt{D_{1D}t}} \quad (3.14)$$

The integration of t from 0 to ∞ returns the local distribution of Figure 3.7:

$$\int_0^\infty \rho(x, t) dt = \frac{e^{-\sqrt{\frac{k_{\text{un}}x^2}{D_{1D}}}}}{2\sqrt{D_{1D}k_{\text{un}}}} \quad (3.15)$$

Using equation 3.14 and looking at the concentration at location of the cognate promoter x^* multiplied by the binding rate, $k_{\text{b,TF@TG}}$, I determine the fraction of

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

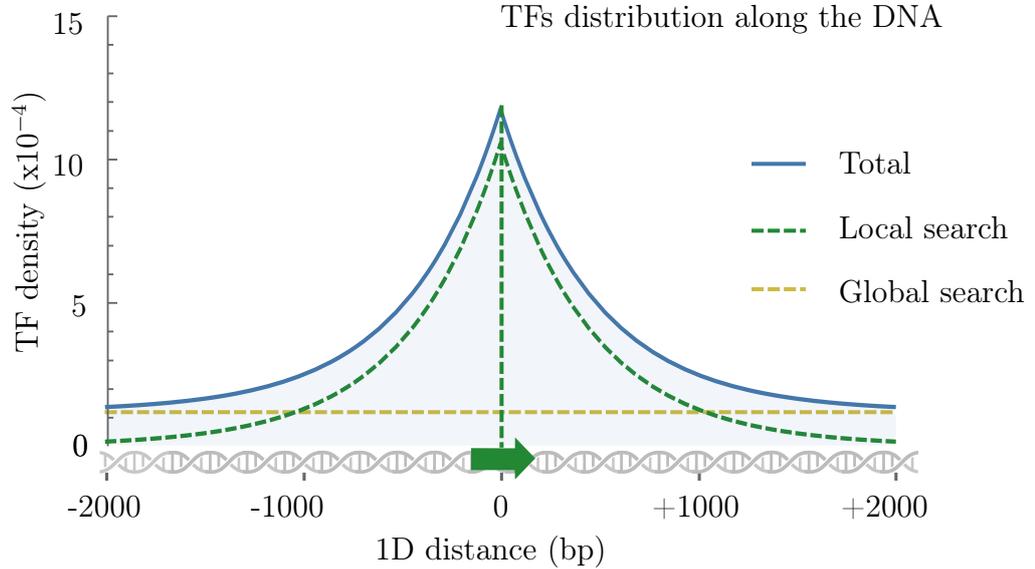


Figure 3.7: Density of transcription factors along the chromosome due to specific modelling dynamics. Global search (yellow line) results in a flat distribution where, due to fast 3D diffusion, the transcription factors are equally spread along the cells volume. Around the source region, local search (green line) favours the accumulation of transcription factors due to one-dimensional sliding after expression. Total distribution is coloured in blue.

eventually binding transcription factor concentration by integrating over time.

$$\int_0^{\infty} k_{b,TF@TG}\rho(x^*, t)dt \quad (3.16)$$

Note that this fraction of bound transcription factor approaches zero in the limit that the binding. However, I can still determine the time it takes for the fraction that *does* bind *to* bind by integrating the binding times and normalising it by the

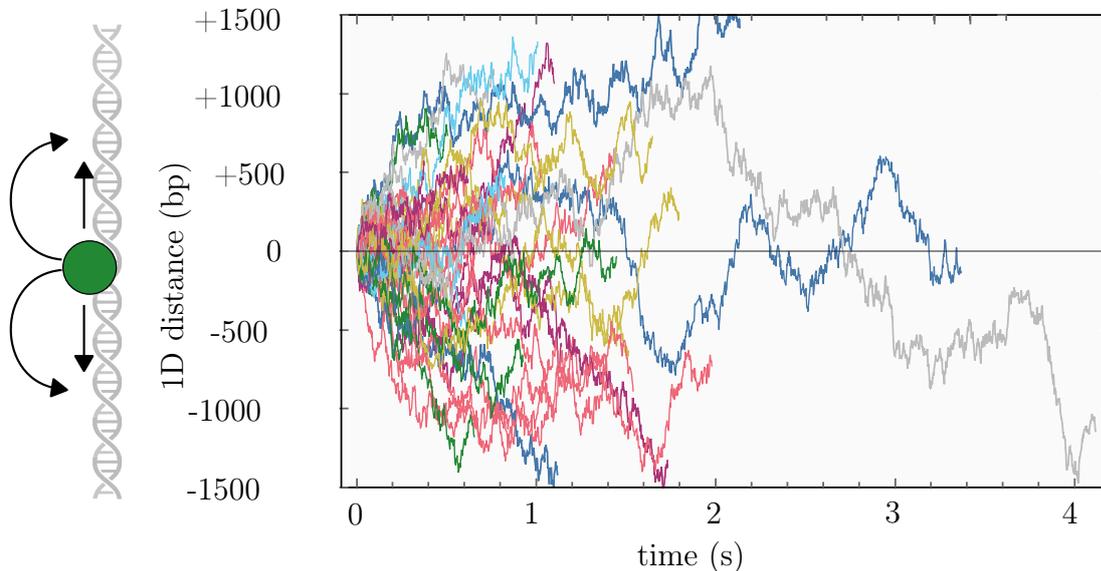


Figure 3.8: Transcription factor diffusion along the chromosome without cognate promoter binding. Simulations of the time (horizontal axis) during which individual transcription factors (colored trajectories) slide and hop along the chromosome and around their encoding gene location (vertical axis). Each simulated line ends when the transcription factor *jumps* to 3D diffusion.

fraction of binding in equation 3.16.

$$\lim_{k_{b,TF@TG} \rightarrow 0} \tau_{s,local} = \frac{\int_0^\infty t k_{b,TF@TG} \rho(x^*, t) dt}{\int_0^\infty k_{b,TF@TG} \rho(x^*, t) dt} = \frac{\int_0^\infty t \rho(x^*, t) dt}{\int_0^\infty \rho(x^*, t) dt} = \frac{\sqrt{k_{un}} |x^*| + \sqrt{D_{1D}}}{2\sqrt{D_{1D} k_{un}}} \quad (3.17)$$

If I compare this against the solution $\frac{|x^*|}{2\sqrt{D_{1D} k_{un}}}$ in Wunderlich and Mirny [106], I see that there is an extra term of $\frac{\sqrt{D_{1D}/k_{un}}}{2\sqrt{D_{1D} k_{un}}}$. This term can be neglected when $|x^*| \gg \sqrt{D_{1D}/k_{un}}$. Here, I assume that the binding rate to the target promoter is very high, so even when $|x^*| \gg \sqrt{D_{1D}/k_{un}}$ is not satisfied, the local search time for distances greater than zero is determined by the lower bound derived

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

in Wunderlich and Mirny [106]. However, I take a different approach when the distance is zero (transcription factor at the target). At this point, the approximation becomes inaccurate. This is why instead of using Equation 3.12 (when the transcription factor is at the target), I use the specific unbinding rate k_{unr} . This is how the model captures the effects of local searches around the target.

I argue that the empirically observed specific unbinding rate (i.e. from the target into the cytoplasm) corresponds to a value [47] that already contain for different specific transcription factor dynamics, like rebinding [96]. This allows me to not know model the specifics (i.e. fast binding after unbinding), for which metrics are not well established experimentally by combining the unbinding rate for cases when $|x^*| \approx 0$ and equation 3.11 otherwise.

3.2.3 Global search

In the case that a local transcription factor unbinds it becomes a global transcription factor and has to search the whole chromosome, the global search time equation is similar to equation 3.11:

$$\tau_{\text{s,global}} = \frac{M}{2\sqrt{D_{1D}/k_{\text{un}}}} \quad (3.18)$$

where

M is the Chromosome length

D_{1D} is the 1D-diffusion rate

k_{un} is the Non-specific DNA unbinding

3.3 Transcription factor diffusion modelling; ODEs

In this section, I develop an ordinary differential equation model to accurately simulate regulatory interactions in a spatio-temporal setup. For this, I combine the homogeneous (or global) model and the 1D spatial (or local) model as described above. Firstly, the model is described and compared against the traditional homogeneous model. Secondly, the model is used to fit experimental data where inter-genic distance was minimised or enlarged [40]. Finally, distance is highlighted as a potential *design parameter*. That is, intergenic distance is used to fine-tune the predicted performance of synthetic circuits.

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

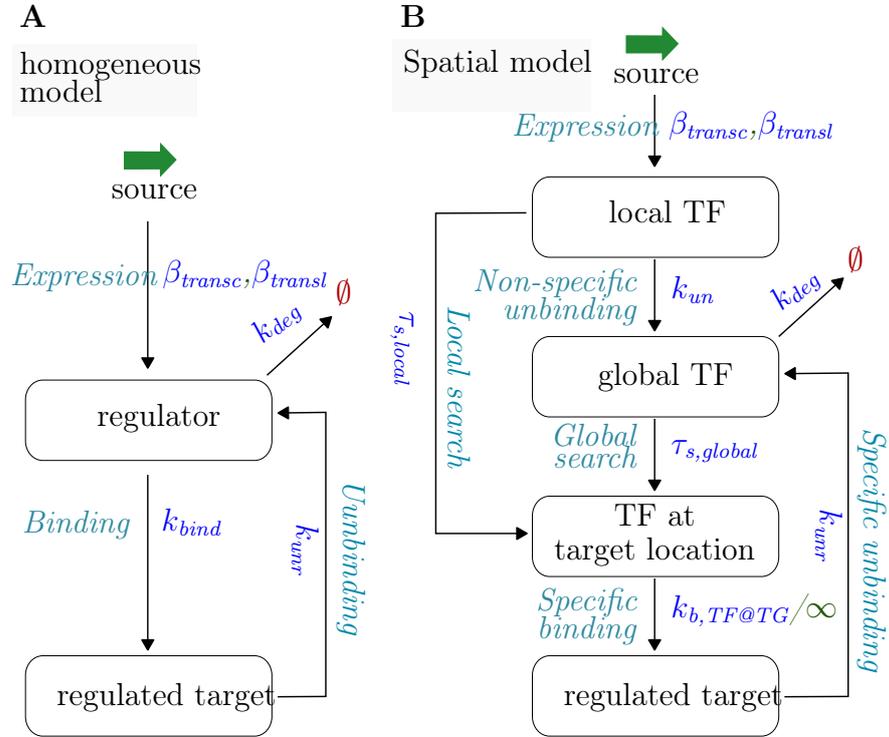


Figure 3.9: Model diagram. **(A)** The commonly used homogeneous model assumes space can be abstracted away. Therefore, transcription factors can directly regulate (i.e. bind) the target promoter after expression. **(B)** The spatio-temporal model presented here explicitly handles spatial dynamics. After expression, transcription factors start a *local search* via one-dimensional sliding along the chromosome. At this stage, a transcription factor may find the target promoter or dissociate into the cytoplasm. In the latter scenario, the transcription factor starts a *global search* which combines both three-dimensional diffusion through the cytoplasm and sliding, until it finds the target promoter or degrades. Transcription factors bind to the target with a *specific binding* rate. The sketch specifies the main rates used in model equations.

3.3.1 Assumptions taken by the model.

There are a number of ground assumptions in the model that are needed to derive system dynamics below.

1. *Transcription and translation are single-particle reactions*

The mRNA-polymerases and ribosomes are not explicitly modeled and the expression of a gene only depends on the promoter availability. This does simplify modelling, but can only represent systems where this promoter availability is limiting.

2. *Global search is a combination of 3D- and 1D-diffusion*

This is generally thought to be needed to catalyse the reaction between the target promoter and transcription factors. Transcriptions slide many times during global search [6, 103, 104], thus reducing the search volume and achieving higher rates than theoretical chemistry without surrounding chromosomes.

3. *Gene sequence length captures the physical distance description*

The chromosome is not explicitly modeled (DNA folding not considered) which reduces the problem from solving a 3D reaction to solving a 1-dimensional one. This should be accurate in the case where a sequence-length below tens of kilo-bases are considered, since chromosome bends are generally larger than that [65].

4. *Transcription and translation are co-localised*

This is generally accepted for prokaryotes [54] (although not without controversy [52]).

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

5. *Transcription factor polymers are formed directly after translation, bind to the chromosome at that moment and perform a local search*

Without this assumption, the model cannot explain spatial effects. Experiments on the polymerisation dynamics of transcription factors are limited.

I developed a model to substantiate this assumption in chapter 4

6. *Transcription factors only bind while sliding*

Since the 1D search space is smaller than the 3D search space this will cause the majority of binding events to be catalysed by 1D-search. Taking this assumption allows to use equation 3.18.

3.3.2 Modelling the search of a transcription factor for its target

The model differs from a homogeneous model in that it considers spatial dynamics (Figure 3.9). Specifically, the model focuses on simulating the diffusion of a transcription factor from its source to its target. To this end, I simulated a regulatory interaction where the expression of transcription factors is controlled. These transcription factors negatively regulate (i.e. repress) a target promoter. In turn, the target promoter controls the expression of a reporter gene. It is assumed by the model that spatial constraints do not affect the mRNAs, but only transcription factors. The former is determined by the following set of Ordinary Differential Equations:

$$\frac{dm_{\text{gfp}}}{dt} = \beta_{\text{transl}} \left(1 - \frac{P_{\text{gfp}}}{1 + \alpha_{\text{leak}}}\right) - (k_{\text{Dm}} + \mu)m_{\text{gfp}} \quad (3.19)$$

$$\frac{dm_R}{dt} = \beta_{\text{transl}}\nu - (k_{\text{Dm}} + \mu)m_R \quad (3.20)$$

where

m is the Population of protein encoding RNA. These are specified for two different encoded protein as shown in the specifier in the subscript, where

GFP is the Reporter protein tagged with a green fluorescent component

R is the Repressor protein

β_{transl} is the **Rate of transcription**

α_{leak} is the Basal transcription of the promoter when repressed

k_{Dm} is the **mrna degradation rate**

P_{gfp} the target promoter (1 when repressed, 0 when free)

and μ is the **Growth rate**.

The model describes a “promoter availability” parameter at the source, ν , with $\nu = 0$ meaning no transcription and $\nu = 1$ meaning full transcription (e.g. constitutive promoter).

Spatial effects are modelled upon mRNA translation. Therefore, mRNA molecules are generated locally to the source gene and do not diffuse away. This assumption draws on the coupling of transcription and translation in prokaryotes [54, 55] along with fast mRNA decay (see Methods). Immediately after the generation of a transcription factor, this is classified as *local* and it is one-dimensional (1D) diffusion along the neighbouring chromosome (sliding + hopping) that determines transcription factors location and dynamics. During local search, transcription factors could unbind and then rapidly *rebind* (hopping) to resume the 1D diffusion [106]. This effect is taken into account by a re-scaled non-specific

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

unbinding rate, which sends a local transcription factor into the global search—local searches would be shorter if rebinding were not considered. After this initial period (timescale of seconds) the regulator will unbind from the chromosome into the cytosol and rapidly lose location autocorrelation due to fast three-dimensional (3D) diffusion. From that moment onward (until it degrades), the transcription factor is classified as *global* (Figure 3.9). The ODEs that calculate the amount of both local and global transcription factors over time are:

$$\frac{dR_{\text{local}}}{dt} = \beta_{\text{transc}} m_R - (k_{\text{Dp}} + \mu) R_{\text{local}} - k_{\text{unr}} R_{\text{local}} - \frac{\gamma(1 - P_{\text{gfp}}) R_{\text{local}}}{\tau_{\text{s,local}}} \quad (3.21)$$

$$\frac{dR_{\text{global}}}{dt} = k_{\text{unr}} R_{\text{local}} - \frac{\gamma(1 - P_{\text{gfp}}) R_{\text{global}}}{\tau_{\text{s,global}}} - (k_{\text{Dp}} + \mu) R_{\text{global}} \quad (3.22)$$

where

R_{local} and R_{global} are the amount of local and global repressors, respectively.

β_{transc} is the

γ is the Chance that a transcription factor binds to the target promoter when both elements are co-located

k_{Dp} is the Protein degradation rate

$\tau_{\text{s,local}}$ is the Local search time (described in Equation 3.11)

$\tau_{\text{s,global}}$ is the Global search time (described in Equation 3.18)

k_{unr} is the Re-scaled specific unbinding rate from the target, which also accounts for re-binding events.

μ is the Growth rate

and

P is the Promoter availability

While 1D diffusion is relatively slow compared to 3D diffusion, local search times can be orders of magnitude shorter than global ones (seconds instead of tens of minutes) if source and target are co-localised. Importantly, 1D diffusion is explicitly simulated, but 3D diffusion is indirectly captured by the model. Since during global search transcription factor location is random with respect to the source, 3D diffusion can be approximated by the definition of an effective binding rate.

When, finally, a repressor binds to the target promoter, the expression of the reporter gene is inhibited. The set of ODEs that describe this is as follows:

$$\frac{dP_{\text{gfp}}}{dt} = -k_{\text{unr}}P_{\text{gfp}} + \gamma(1 - P_{\text{gfp}})\left(\frac{R_{\text{local}}}{\tau_{\text{s,local}}} + \frac{R_{\text{global}}}{\tau_{\text{s,global}}}\right) \quad (3.23)$$

$$\frac{d\text{GFP}}{dt} = \beta_{\text{transc}}m_{\text{gfp}} - (k_{\text{Dp}} + \mu)\text{GFP} \quad (3.24)$$

where

β_{transc} is the Rate of translation

GFP is the Green Fluorescent protein

k_{unr} is the Rate of specific unbinding

k_{Dp} is the Protein degradation rate

m is the mRNA

μ is the Growth rate

P is the Promoter availability

R is the Repressor

$\tau_{\text{s,local}}$ is the local search time

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

and: $\tau_{s,\text{global}}$ is the [Global search time](#)

3.3.3 Quasi steady state mRNA

The system of Ordinary Differential Equations was simplified using the quasi-steady-state assumption for mRNA, m^* . I achieve this by taking equation [3.19](#), replacing the gfp specification with a dummy specification i , setting the derivative to zero and solving for the mRNA species.

$$m_i^* = \frac{\beta_{\text{transl}} \frac{P_i + \alpha_{\text{leak}}}{1 + \alpha_{\text{leak}}}}{k_{\text{Dm}} + \mu} \quad (3.25)$$

replacing equations [3.21](#) and [3.24](#) with:

$$\frac{dR_{\text{local}}}{dt} = \beta_{\text{transc}} m_R^* - (k_{\text{Dp}} + \mu) R_{\text{local}} - k_{\text{un}} R_{\text{local}} - \frac{P_{\text{gfp}} R_{\text{local}}}{\tau_{s,\text{local}}} \quad (3.26)$$

$$\frac{d\text{GFP}}{dt} = \beta_{\text{transc}} m_{\text{gfp}}^* - (k_{\text{Dp}} + \mu) \text{GFP} \quad (3.27)$$

3.3.4 Binding rates

The homogeneous model of [Figure 3.10](#) uses a binding rate of $10^6 \text{ M}^{-1} \text{ s}^{-1}$. Spatial-based binding rates, however, cannot be expressed by such a singular value. In the spatial model, I decompose the binding rate into two different parts: [i] the time it takes for a transcription factor to find the target (i.e. local/global search times), which depends on diffusion and unbinding rates as described in [section 3.2](#) and [ii] the propensity, γ , of a transcription factor to bind after it finds

the target. Figure 3.12 shows the effect of decreasing this parameter tenfold. Unbinding rates are not affected by spatial constraints, and are thus obtained from literature.

3.3.5 System sensitivity to intergenic distance modulation.

Gene expression, and more specifically, promoter activity, reflects the upstream dynamics of the cognate regulatory machinery. In traditional homogeneous models, such promoter activity is determined by the overall number of transcription factors and a constant binding rate. In contrast to this, the addition of spatial resolution to modelling makes transcription factor binding dependent of [i] the location of the transcription factor (i.e. local or global), [ii] diffusion speeds, [iii] the distance between the source of transcription factors and the target promoter, [iv] binding rate and [v] non-specific and specific unbinding rates (Figure 3.10).

For the stochastic simulations in Figures 3.10, 3.12, 3.11 and 3.13, reactions were implemented in the Gillespie algorithm [35] as explained in 2.3.2.

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

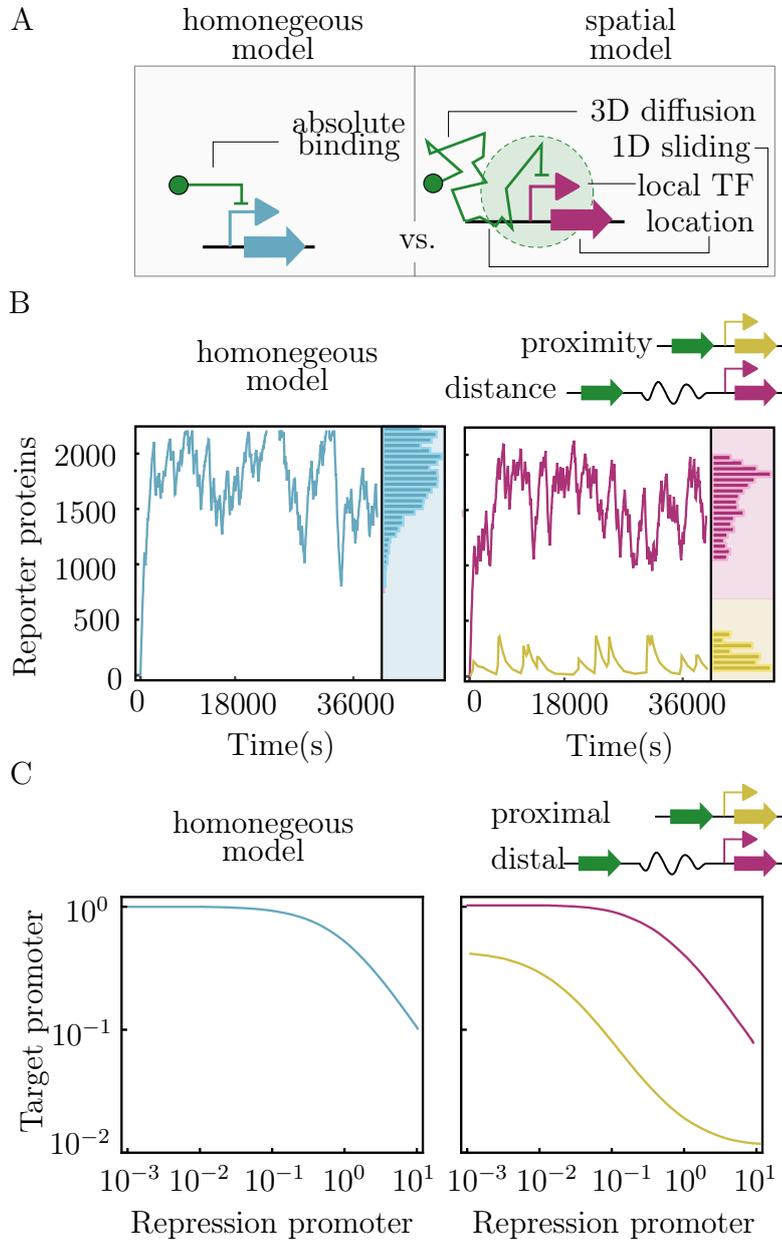


Figure 3.10: (see next page)

Simulations of an inhibition interaction (i.e. repression) using both homogeneous and spatial models were compared (3.10B). For the sake of comparison, the

Figure 3.10 (*previous page*): Analysis and comparison of negative regulation (repression) in bothspatial and homogeneous models. **(A)** The binding of a transcription factor to its cognate promoter is the key difference of both modelling approaches. While a single absolute value is enough in a homogeneous model, binding events depend on 1D/3D sliding/diffusion and the relative distance from transcription factor source in spatial models. **(B)** Time course simulation of gene expression in the two models. The homogeneous simulation is similar to the distant source-target scenario in the spatially-resolved model since both models converge when one-dimensional sliding does not play an important role. **(C)** Characterisation of promoter activity in both source (x axis) and target (y axis) regions. Axes measure the normalised proportion of time (from 0 to 1) that a promoter is in its *active* state. Note that the repression promoter can be “stronger” than the promoter used for normalisation(i.e. > 1). If the repression promoter is high, it reduces the expression of the target promoter. When the target is co-located with the repression promoter, the same repression is reached at a lower expression of the repression promoter.

non-spatial parameters in both models have the same values (e.g., transcription, translation, or molecule degradation). That is, source-target distance (thus transcription factor dynamics) is the only difference between the two models. Simulations returned similar expression levels for both approaches, providing that the relative location of genetic components was distant in the spatial model. This guarantees that when the effect of sliding at the local search is not relevant, the spatially-resolved model will converge to the homogeneous model [73]. Although

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

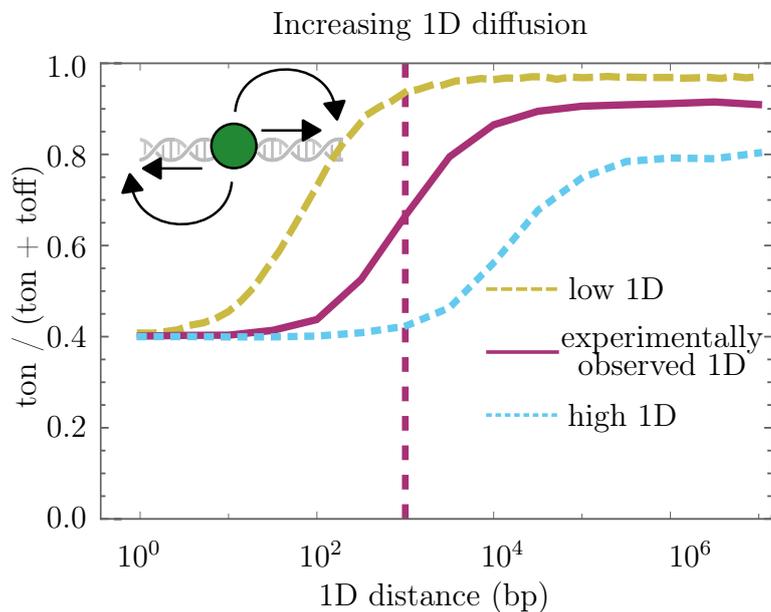


Figure 3.11: Increasing the 1D diffusion coefficient decreases the impact of intergenic distance. The fraction of time the target promoter is active (i.e. without the repressor bound; y axis) increases as the target moves away from the source. However, increasing the 1D diffusion coefficient mitigates (up to a limit) this trend.

the expression levels are qualitatively similar, the simulated DNA components are not far enough to cancel out completely the effects by local search. However, when components were co-localised in the spatial model, the time series simulations were very different. Upon co-localisation, the repression was observed to be stronger (i.e. less reporter expression). A closer look at promoter availability (fraction of time it is in its active form) confirmed this trend - the activity of the source promoter has a faster (and stronger) impact on the target when placed in proximity, since the interaction is not diluted within the volume of the cell (Figure 3.10C). It is not coincidental that the simulations gave similar results be-

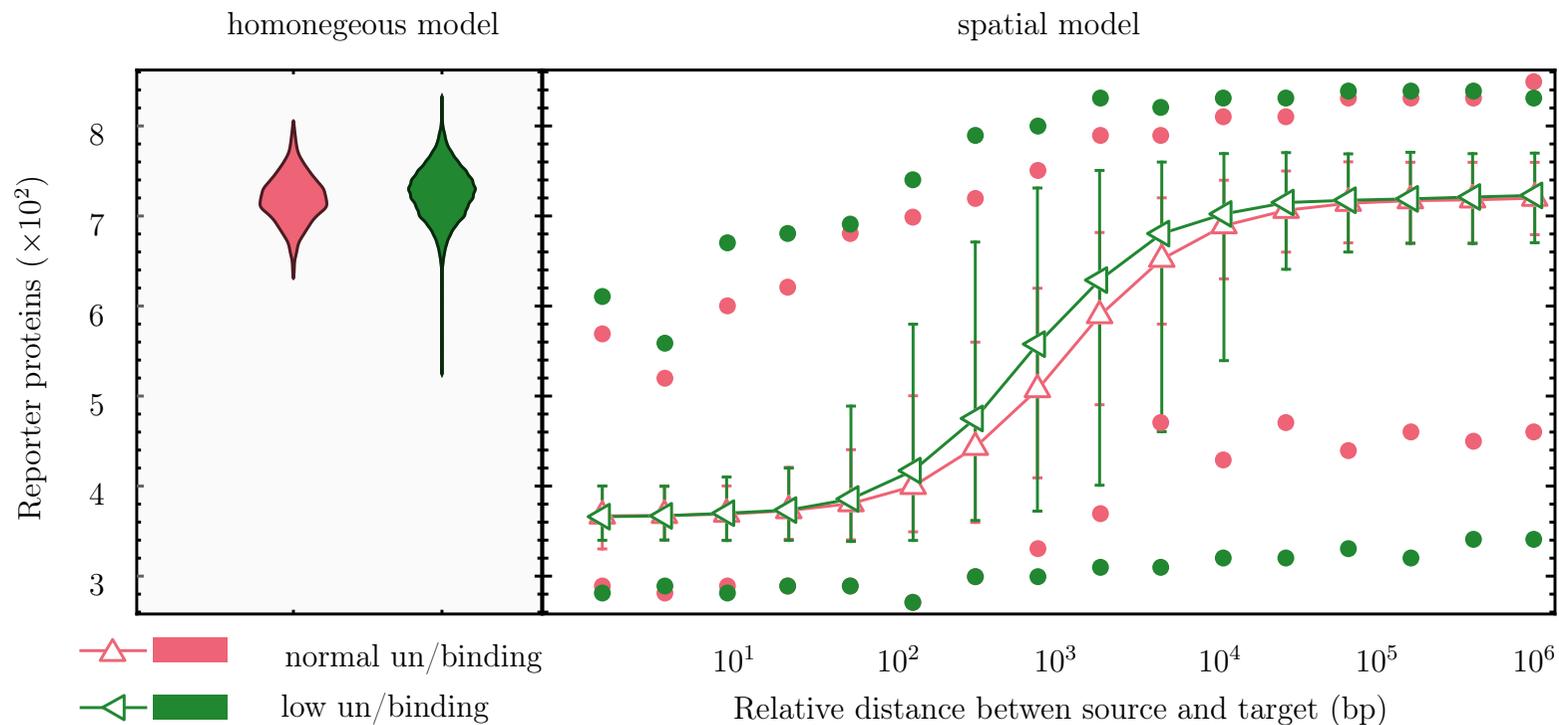


Figure 3.12: Gene expression noise variation in relation to source-target distance. Noise was analysed in both homogeneous and spatial models with two sets of rates for binding and unbinding (*low* rates decrease tenfold in respect to *normal* rates). The relationship between the two simulation runs in the homogeneous model (left) correlates to the spatial model (right) when the relative distance from source to target is high. If such distance is low, noise patterns also overlap. However, at middle distances, the noise corresponding to the *low* set was much bigger. Error bars show noise ranges within standard deviation; dots represent minimum and maximum simulation values.

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

tween the homogeneous and spatial models when the source-target distance was maximised. This is because space-specific parameters in the spatial model were set up to that end. These initial parameter guesses were chosen based on the assumption that the majority of empirically observed (binding/unbinding/ etc.) parameters measured without co-location of genes, however ideally space specific should be used from measurements in future works. Therefore, the comparison informs about what the reduction of source-target distance leads to, which is, as described above, a lower reporter expression (i.e. stronger regulation). However, at high intergenic distances, the majority of transcription factors reach the target promoter via global (instead of local) search, which in our model generates a uniform distribution of transcription factors along the chromosome. As a result, the spatial parameters, that are more relevant during local searches, lose significance and the simulation converges to the homogeneous model. If space-specific parameters had been established so that the proximity scenario had matched the homogeneous model in the first place, the simulation would return the opposite results.

Gene expression noise [48, 63] is decisively influenced by intergenic space separation. A major source of genetic noise is the so-called *bursting* effect [92, 108], a pulse-like expression activity that results from a transcription factor binding and unbinding its cognate promoter. Given that physical source-target separation modifies the binding process through modulating the availability of transcription factors, it can be concluded that it has a direct influence in transcription bursts. Figure 3.12 shows the variation in gene expression noise concerning source-target distance. Results are shown for two sets of binding and unbinding rates: the first one referred to as *normal* (obtained from the experimental literature) and

the second one termed *low* (one-tenth of the previous). The expression profile returned by the homogeneous model shows a trend towards a slightly wider noise pattern in the *low* scenario. This matches the far-away spatial simulation. However, the changes in noise patterns at such high distance are not very relevant, since transcription factors are not efficient (i.e. target promoter always ON). At short separation, the noise patterns under both *normal* and *low* rates also overlap. In contrast, the noise corresponding to the *low* set of rates increases significantly at middle distances. This suggests that protein variability is more than the mere consequence of stochasticity and can be deterministically controlled by modulating intergenic distance alone.

3.3.6 Nonlinear response regulation to intergenic distance.

The effects of source-target separation in reporter expression do not vary in direct proportion to the increase in distance (Figure 3.12). In fact, up to 10^2 base pair (bp) separation, the effects were found to be effectively the same. A similar performance was observed when the separation was above 10^4 bp, when further differences cannot be appreciated. However, the impact of such separation increases in all values in between, almost proportionally, with 10^3 bp being halfway in the overall response curve. Figure 3.12 implies that the origin of such nonlinearity lies in the balance between local and global search: local search predominates in the first region (0- 10^2 bp), within which the location of the target promoter will not make a difference; both local and global searches are combined in the second region (10^2 - 10^4 bp), that shows proportional effects when increasing distance; global search predominates in the third region (10^4 - 10^7) which also re-

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

sulted in a plateau-like response function. Although the thresholds between such regions can be altered by decreasing/increasing 1D diffusion (Figure 3.11), the overall pattern remains the same. It is important to note, however, that the level of repression at long distances (far-right of x-axis) changed substantially when 1D diffusion was modified. This is because global search includes 1D sliding along non-local chromosomal regions. Thus, it is correlated to 1D diffusion in that an increase in this rate would facilitate transcription factors to reach their target during global search. Altogether, this suggests that spatial effects are stable at two source-target relative locations, *very close* and *very far*, while at middle points the regulation would change rapidly.

3.3.7 Inducible system

. For the simulations of Figure 3.13, the target promoter is inducible instead of repressible. This implies the promoter has a leaky expression when a regulator is *not* bound ($P_{\text{gfp}} = 0$). Therefore, Equation 3.19 is re-defined as follows:

$$\frac{dm_{\text{gfp}}}{dt} = \beta_{\text{transc}} \frac{P_{\text{gfp}} + \alpha_{\text{leak}}}{1 + \alpha_{\text{leak}}} - (k_{\text{Dm}} + \mu)m_{\text{gfp}} \quad (3.28)$$

3.4 Space as a design parameter.

A previous study [40] experimentally measured the effects of intergenic distance in a positive regulation (i.e. induction) using components of the TOL pathway [21] of the environmental bacterium *Pseudomonas putida* [13, 70]. Specifically, the gene *xylS* (source), which expresses XylS regulators was inserted in proximity to, or separated from, the promoter *Pm* (target), which is in turn activated by XylS.

Spatial effects were quantified by measuring the expression of *P_m-gfp* fusions in single cells with flow cytometry (Figure 3.13). Results suggested that space could be used as a design parameter for selecting output levels since the performance of the regulatory circuit changed according to spatial configuration. Specifically, when the source-target distance was minimised, reporter expression was fully *on* (i.e. narrow distribution to the right of the plot). In contrast, when such distance was increased, the expression became very noisy (i.e. wide distribution from left to right). Here, I compared the expression distributions in the experimental data with simulations by modulating source-target distance in our spatial model. As it can be observed in the side-by-side comparison of Figure 3.13, the model for spatial regulation presented here gave an accurate reproduction of the experimental information - something that was not possible with homogeneous (i.e. non-spatial) models.

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

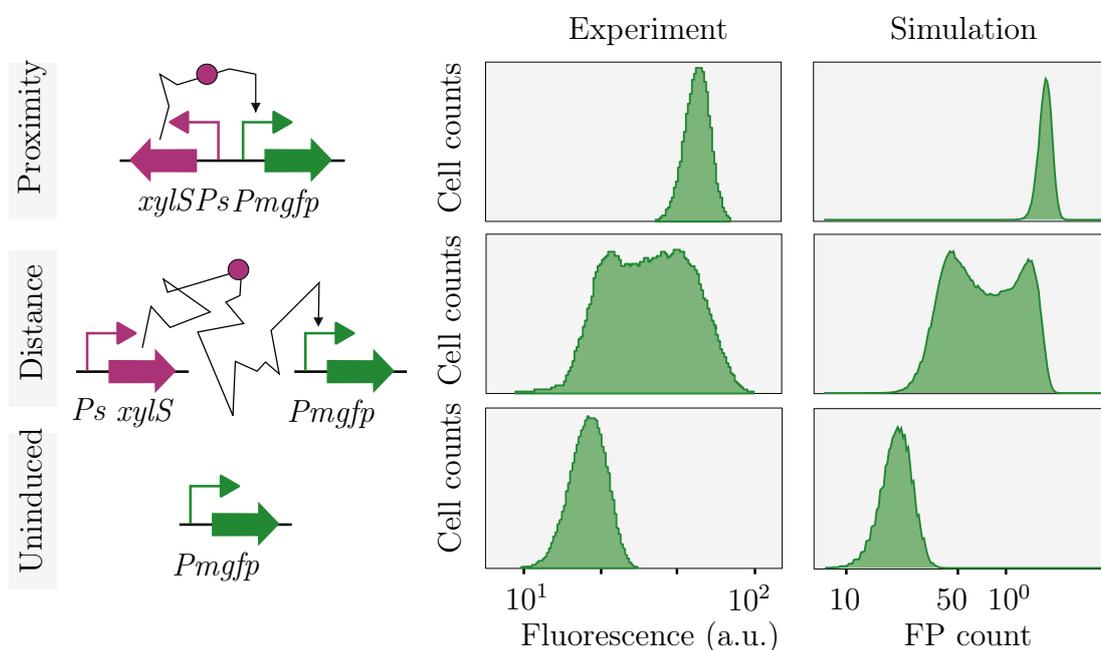


Figure 3.13: Space as a design parameter. Reproducing the experimental results from [Goñi-Moreno et al.\[40\]](#) with the proposed spatial model. These experimental results measured gene expression noise in a construct where the source (gene *xylS*) of transcription factors (XylS) and its cognate target promoter (*Pm*) were engineered to be in either 1) proximity or 2) in distant locations inside the bacteria *Pseudomonas putida*. In vivo flow cytometry results (adapted from [\[40\]](#)) and their in-silico counterpart look alike by modulating the intergenic distance alone.

In the case of co-localisation (Figure 3.13 top row), both source and target were inserted into the same chromosomal location (*attTn7* site), but in diverging orientations. This means that when the transcription factor is finally expressed it is around 1kbp away from the target. In the case of high source-target separation (Figure 3.13 A middle row), I considered the distance to be as large as it could be. This is because the source(*Ps xylS*) was in a plasmid whereas the target

(*Pm gfp*) was located in the plasmid, such that there is no 1D trajectory from source to target and effect of 1D sliding can be ruled out. In this scenario only the global search contributes to binding events. The model uses 1kbp and 10^7 bp as the *close* and *far* distances for the simulations, respectively.

Finally, I illustrate the potential of space as a design parameter in the context of genetic combinatorial circuits. In these, information is transmitted in the form of digital-like values, 0/low expression and 1/high expression, through regulatory interactions. Each of the components of a circuit can be seen as an electronic device that gets a value in the input and returns a value in the output after some processing. To get optimal performance, a key feature for a circuit is that the output of a component must be *compatible* with the input of the next one. Compatible in that the first component's output range (i.e. the gap between the lower and higher values that can be produced as output) must be sufficient to differentiate two distinct values at the input of the second component. Otherwise, that connection will not be able to propagate digital values.

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

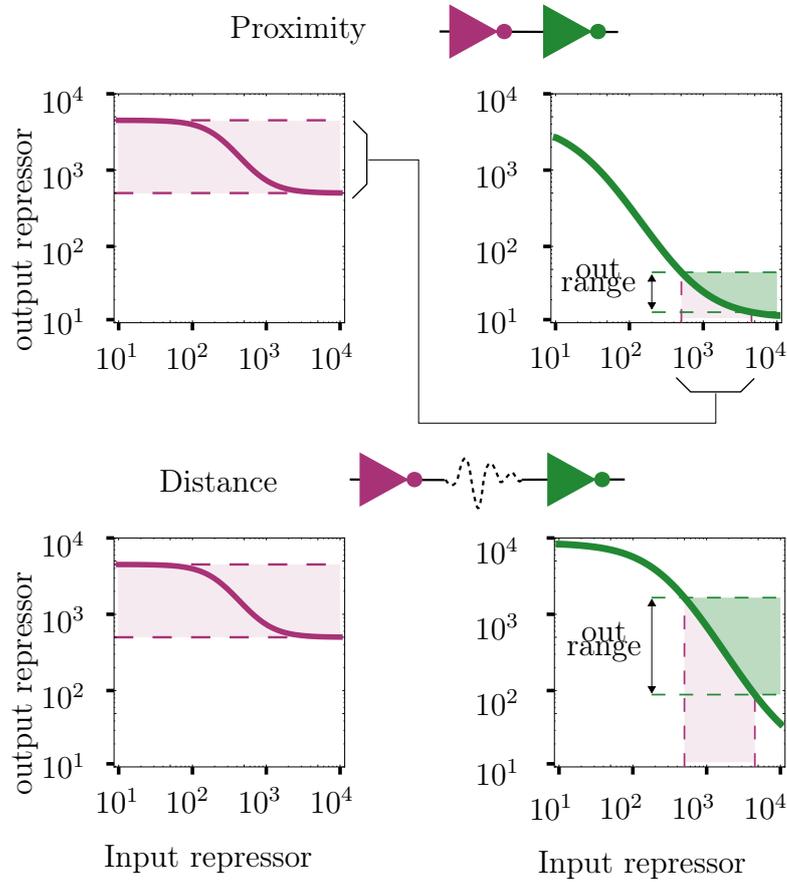


Figure 3.14: Two connected repressions modify their input-output mapping when either co-located or spatially separated. Top graphs show the simulation of the connection when the components of the circuit are placed in proximity; the output of the first (signal b) regulates the input of the second (also, signal b). The output of the second component (signal c) shows a minimal dynamic range (green shaded region). Bottom graphs show the same simulation but changing the location of the components to be far apart. The connection changes and the dynamic range at the output increases. Both examples suggest that space can be effectively used as a design parameter.

Currently, the lack of such compatibility between components is solved by

simply selecting different components. That is, a given genetic logic gate can be replaced by another one if it is not ideal for a specific regulatory interaction. This approach is followed by, for instance, the tool Cello [69], a design automation platform that allows a user to turn high-level specifications into the DNA sequence of the corresponding genetic circuit. However, this procedure is limited by the catalog of available components that could be used to this end - which is, in most of the cases, just a handful. Our take here is that there would be no need to replace the components if optimal performance could be achieved by organising them into a different configuration within the volume of the cell. Figure 3.13 shows the compatibility between two theoretical negative regulations, so-called *genetic inverters*, which are components that *invert* the input signal: if the input is 1, the output would be 0, and vice-versa. The inverters are characterised by their transfer function, which relates the amount of output (y-axis) that correspond to a given input level (x-axis). From the transfer function of the first inverter (violet in Figure 3.13) we can measure its *output range* as the region covered along the y-axis. When this range is turned into the *input range* of the second inverter (green) I can assess their compatibility. As can be observed, by solely changing the inverters from co-located to spatially separated, the output range of the second one improved substantially. Importantly, it was only distance that was modified in the model; the rest of the parameters in the simulation were the same. This suggests that the catalog of *functions* that could be used to replace circuit components may not only be formed by different DNA devices, but by the same ones with a different spatial arrangement. Moreover, this also points out that analog signals can be used to process information beyond mere digital abstractions [19].

3.5 Discussion

The rational design of gene regulatory circuits builds on our mechanistic understanding of gene regulation. Such understanding is commonly captured by mathematical models, that allow for turning mechanistic details into design principles. Although well-studied, gene regulation is still based on unclear dynamics; at least, not clear enough for the rigorous mathematical formalisation that predictive modelling needs. Here, I focus on the implications that the separation between genetic components, within the chromosome, has on their final performance.

Ever since advances in technology allowed for it, the interest in the intracellular spatial organisation of the regulatory machinery has increased [24, 26, 51, 54, 55]. However, current descriptions of how regulatory interactions are affected by intergenic distance are still somewhat controversial. It is the case that similar experimental setups had shown both significant [58] and insignificant [9] effects of such chromosomal separation. Our previous experimental work [40] showed not only a decrease in gene expression, upon inserting the regulator source *far* away, but also an increase of the noise pattern. Therefore, it is both expression and noise that were suggested to be modified by solely altering the spatial configuration of the regulation. There could be many potential causes of these conflicting results, such as growth rates, bacterial species/strains, genetic machinery and measuring methods. Therefore, the formalisation of such dynamics in a modelling framework is essential; model-guided analysis of *spatio-temporal* gene regulation may turn decisive to elucidated these details. Although existing tools such as SmolDyn [2], eGFRD [90], SMeagol [60] and others [82] simulate spatial

dynamics to various extents, they do not focus on gene regulatory circuits, which is the goal of our study.

Our model builds on a previous definition of *facilitated diffusion* [106] that include the classification of transcription factors into two groups: of *local* and *global*. This is a crucial feature of the model. While local TFs perform a one-dimensional search along the chromosome close to their coding gene, global transcription factors undergo rapid three-dimensional diffusion through the cytoplasm. I adapted this definition to include the specific dynamics of gene expression (i.e. transcription and translation) and gene regulation (i.e. transcription factor-promoter interplay). By considering the distribution of transcription factors along the chromosome (due to diffusion) and the binding rates of the system, I calculated the time it takes for a transcription factor to bind its cognate promoter at any given location. While this model was able to reproduce experimental observations accurately (see Figure 3.13) [40] (which seems not to be possible with *non-spatial models*), I identified two areas of improvement that will be the focus of successive studies. Firstly, the model fails to simulate Hill function-based dynamics. This is a direct consequence of diffusion since it makes it very challenging to calculate cooperativity. Secondly, the model does not account for the three-dimensional folding of the chromosome. Rather, it assumes a homogeneous shape along the cell. However, since Hill functions are a phenomenological description, rather than a physical model, and nucleoid structure is not rigorously known, none of these features would make the model physiologically more accurate.

The growing availability of methods for chromosomal insertions [14, 62, 98], allows researchers to transit from plasmid-based to chromosomal-based synthetic systems. While plasmids are metabolically demanding for bacteria [61], chro-

3. SPATIAL SEPARATION AS A DESIGN PARAMETER

mosomal insertions are more efficient [28]. As increasingly precise responses are required from synthetic constructs, each component may need a specific location within the chromosome for optimal performance - an issue that deserves further attention. I advocate for the use of inter-component distance as a design parameter for synthetic circuits, and the use of spatio-temporal modelling to establish the engineering principles of such three-dimensional design.

3.6 Future work

A lot of assumptions were made in the model that should be substantiated. The most far-fetching assumption, that there is co-localisation retention in the dimerisation process, is partially substantiated in chapter 4, but further research is needed.

Another assumption is that the search rate can be characterised by the observed combination of 1-D sliding and short 3-D hops. There is some research [22] indicating that indicates that this is valid, however the molecular dynamics theory of this needs improvement.

To make sure that observed spatial effects are purely due to relative spatial distancing and not due to structural effects, gene copy number or any artifacts due to usage of the plasmid (antibiotics, burden), the experiment could be replicated with some changes: Firstly, both genes should be included into the chromosome to eliminate any plasmid effects. Secondly, instead of a single reporter, it would be beneficial to include a second reporter right next to the location of the regulated gene which instead is constitutively expressed. This would be a good indication of the structural effects on expression (super-coiling, localised crowding etc.).

To confirm the experimental description put forward by the model, I call for and assay of distances that are below the predicted order of magnitude ($< \text{kb}$) and more than kilo-bases.

Chapter 4

Spatial aspects of protein dimerisation

This chapter is an adaptation to our work, Stoof and Goñi-Moreno [86]. The work is currently in peer-review at "Journal of the Royal Society Interface". In this work I came up the initial hypothesis, developed the theory, made the models and did the simulations.

Problem definition:

The assumption in Chapter 3 that a Transcription Factor is located at the encoding gene after it is created is unsubstantiated in Chapter 3. For this assumption to hold, there must be processes that allow the formation of an active form of a transcription factor at the location of its encoding gene. One of these processes is the formation of a dimer, as most transcription factors only function in di-/tri-meric form. This chapter describes an interaction of proteins during the translation process to allow for this di-/tri-merisation. The existence of this process substantiates the assumption in Chapter 3.

For homeostasis, organisms often heavily rely on feedback mechanisms. To reduce the burden of these feedback mechanisms, it is observed that these processes are frequently non-linear. Key processes for synthetic gene circuits, such as the design of bistability or oscillatory patterns, build on non-linear molecular dynamics. However, the molecular details responsible for non-linearity are only partially understood. Non-linear dynamics are mostly described with phenomenological models and the rational design of non-linear interactions remains a distant prospect.

In this chapter, I introduce a mathematical model for the dimerisation of proteins dependent on the transcription and translation processes. I show that the dimerisation process can induce a combination of linear and quadratic processes in the creation process of the proteins. The distribution of which is dependent on molecular and genetic features, such as protein length or gene location within the volume of the cell. Moreover, I revisit Chapter 3's model, implementing the combination of dimerisation and space. I suggest design rules and principles, using these models, section 4.3.2, that may be used to fine-tune non-linear effects in-vivo, resulting in an increase in the computing abilities of biological systems.

4.1 Related models on transcription and translation

Since this is the first mathematical model to describe dimerisation during the transcription and translation processes I attempt to keep the model as simple as possible.

4. SPATIAL ASPECTS OF TF DIMERISATION

Therefore I will not capture the dynamics of the RNA-polymerases nor the capture variations in translation speed by ribosomes. In my model I will assume a random (Poisson distributed) difference in time between ribosome initiation event and not model how the Ribosome Binding site recruits these ribosomes.

However, Tuller et al. [94] show that this is not necessarily the case. The authors claim that genes encoding specific processes encode a “ramp” in the first ~ 40 codons of mRNAs. As the authors state this ramp “may represent an important next stage of translational control that modulates the parameters set by the previous initiation stage” such that the time difference between ribosome initiation events might not be random.

Moreover, the concentration of ribosomes is not constant along the cells. Castellana et al. [12] describe a partial differential equation model of ribosomal sub components and show that that the ribosomes tend to be volume excluded by the nucleoid and driven toward the cell poles. Klumpp et al. [53] confirms that the large ribosome complexes are effected by these crowding effects.

Another effect not captured in the model is transcriptional bursting where it is seen that the arrival of RNA-polymerases is also not constant process. While Pájaro et al. [74] are able to reproduce these bursting dynamics with a partial integral differential equation model, Chong et al. [15] are able to describe bursting from a mechanical model. The authors show that their model of supercoiling induced by a protein called gyrase affects the binding dynamics in a way that reproduces empirical results.

Future modelling could look into how ramping , crowding, supercoiling and other more detailed expression modelling could combine with the dimerisation process.

I do however rely on the model of the expressome. Where the coupled nature of transcription and translation in prokaryotes [54, 81] is theorised to produce locally high distributions of transcription factors near the site of their expression [55].

4.2 Non-linear effects by Transcription factors

It has long been known that the binding of transcription factors is a non-linear mechanism [1]. This non-linearity allows the design of complex synthetic circuits with dynamic behaviour. In the early days of the field, two landmark papers demonstrated that nonlinear dynamics are at the core of relatively simple circuits. The first one implemented a “toggle switch” [31]; a genetic circuit able to switch between two states according to external signals. The second one implemented what was called the “repressilator” [27]; an oscillator based on a circuit of gene transcription repressors. Mathematical models of these two circuits show bistability and reliable oscillations (respectively) only if they account for nonlinear dynamics.

Non-linear effects are usually modelled using abstract parameters without bottom-up physical description. While Michaelis-Menten equations, for instance, use integer values for a physical parameter (which has since been described from basic principles), Hill equations allow for unphysical non-integer values (which increase controllability). Although effective, this indicates a fundamental gap in existing mathematical modelling. The lacking description of the molecular details of non-linearity limits the rational design of such behaviour.

4.3 Results

There are several ways the non-linear binding can arise; the conformational change of an operon [56, 85], regulator-promoter interplay [40, 74] (known as “bursting”) and other upstream processes [20, 58]. Here I analyse the monomers-dimer reaction[102]. Upon gene expression, most resulting proteins are “monomers” that need to interact to form “dimers” (or higher-order “oligomers”); it is only the protein in its final form that is active. As a consequence, the relationship between active proteins and gene expression is purely nonlinear: monomers will have to “meet” for dimerisation, such that there is at least a quadratic term. I suggest an alternative pathway for monomers to dimerise: Translation mediated dimerisation.

4.3.1 New model: Translation mediated dimerisation

Transcription (of a gene into RNA) and translation (of RNA into proteins) rarely generate fully functional transcriptional regulators. Rather, resulting proteins—or “monomers”—need to interact with others in order to form “dimers” (4.1A). For example, the repressor protein TetR, which is extensively used in synthetic biology, is a dimer. This suggests that any mathematical model that aims at simulating the dynamics of such a molecule (or that of a genetic circuit regulated by it) must account for the dimerisation of the partially formed regulators generated after translation in order to result in robust predictions.

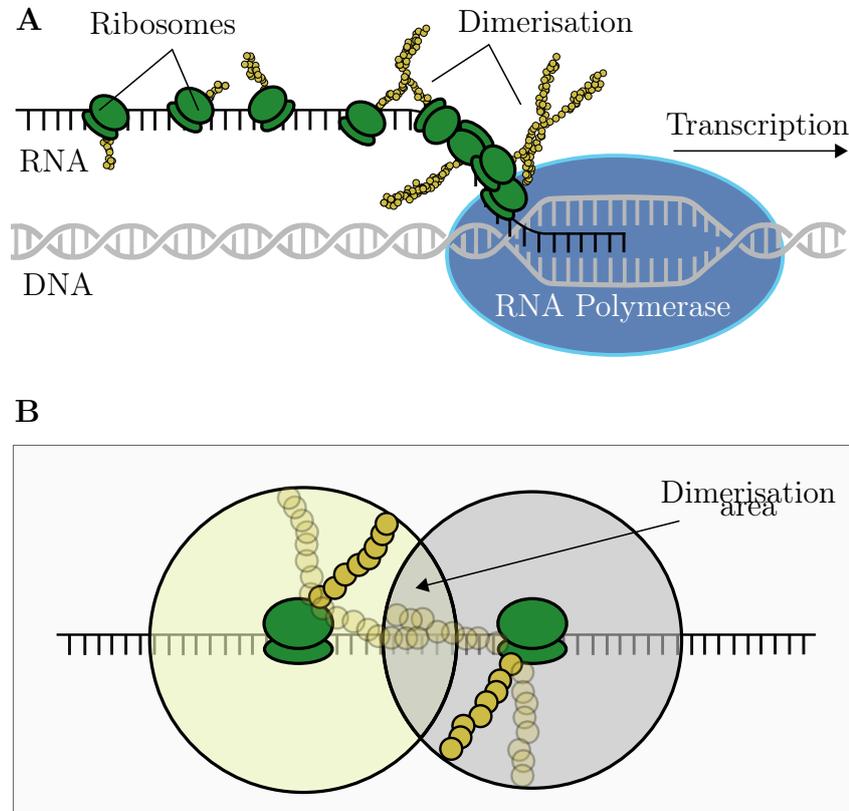


Figure 4.1: **Translation-mediated dimerisation (A)**: Upon DNA transcription by RNA polymerase, ribosomes bind the resulting RNA to translate it into proteins. There is more than one translation processes at any one time, and ribosomes go along the RNA at different speeds, leading to the appearance of "traffic jams". Our model simulates the process by which, when distance among ribosomes is short, partially formed monomers (represented by chains of yellow circles) dimerise with other partially formed monomers as they are being translated. **(B)**: Detail of the translation-mediated dimerisation area. The extension of this region depends on several physical features, as the length of the protein to be translated or the distance between ribosomes; these constraints will affect nonlinearities due to protein dimerisation.

4. SPATIAL ASPECTS OF TF DIMERISATION

Although typical modelling frameworks, for instance, Michaelis-Menten and Hill equations, already account for dimerisation, matching this to its specific molecular mechanisms and features (in contrast to using abstract cooperativity values) is still an overarching challenge. My model adds a detailed mechanism of translation, in which ribosomes bind RNA molecules in an asynchronous fashion; as a result, some ribosomes start translating very close to each other [66]. In this scenario, two partially-formed monomers come into contact and dimerisation starts, which I termed translation-mediated dimerisation (4.1B). In short, when ribosomes are at a long-enough distance, dimerisation will take place in the cytoplasm, as in the cytoplasm both monomers will need to meet. The resulting non-linearity (in final dimer formation) can therefore be predicted and designed by my model by modifying parameters such as the length of the protein to be translated or ribosome availability. A more detailed explanation on this will follow in the case studies, later in this section.

In order to describe translation-mediated dimerisation rates, I firstly assume that the arrival of a ribosome to the RNA is captured by a Poisson process. The distribution of time between binding ($Pr(\Delta t)$) events is then given by:

$$Pr(\Delta t) = \lambda \cdot e^{-\lambda \Delta t} \quad (4.1)$$

where λ is the average binding rate of a ribosome to the RNA (i.e., when translation initiates), and Δt represents the time between such binding events. This time distribution, along with experimentally obtained translation rate of ca. 10 AA/s [97, 107] (AA = amino acids), was used to calculate, Δx , the distance

between two consecutive ribosomes:

$$\Delta x = 3x_{\text{bp}}k_{\text{transl}}\Delta t \quad (4.2)$$

where

k_{transl} is the [Rate of ribosome transcription](#)

and x_{bp} is the [Base pair width](#)

the factor of 3 comes from one amino acid being encoded by 3 bp. The difference in initiation time also gives rise to a difference in length of the chain of amino acids that are already transcribed, Δr :

$$\Delta r = x_{\text{AA}}k_{\text{transl}}\Delta t \quad (4.3)$$

where

x_{AA} is the [Amino Acid width](#), also ca. 3 Å.

This chain of amino acids is the partially formed protein in monomer form. The model assumes Δx and Δr until completing translation, which implies that the separation between ribosomes depends only on the binding times, not on the sliding along the RNA.

To calculate the fraction of dimerisation that takes place during translation, the model defines an area where partially formed monomers can interact (Figure 4.2). While monomers are being formed, they are bound to the ribosome in one end and moving within a sphere around the ribosome, with a radius based on the current length, with the other end. If two consecutive monomers are long enough, the area where the spheres overlap shows where the monomers can interact. In this state, dimerisation may start.

4. SPATIAL ASPECTS OF TF DIMERISATION

To analyse this process I focus on a two consecutive ribosome scenario that can be generalised to the ensemble of all ribosomes. To calculate the chance of dimerisation, first I derive the fraction where these spheres overlap, equation 4.9.

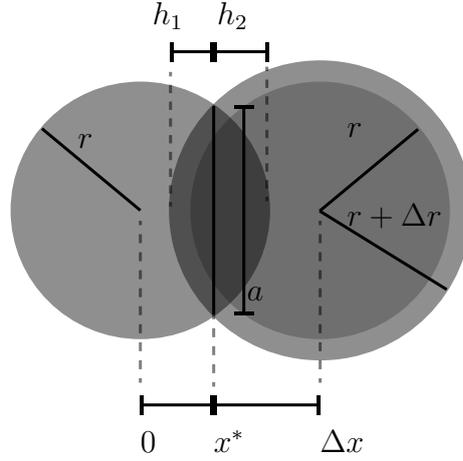


Figure 4.2: Calculation of the spherical overlap, note that the dark grey overlapping section is composed of two spherical domes (seen here is the 2D analogous circular segments)

I choose my system such that the two ribosomes are placed at the origin and at a displacement of Δx with radius r_1 and r_2 respectively. The spheres are then described by:

$$x^2 + y^2 + z^2 \leq r_1^2 \quad (4.4)$$

and

$$(x - \Delta x)^2 + y^2 + z^2 \leq r_2^2 \quad (4.5)$$

Using these equations, the edges of the spheres meet at a circle in the yz plane

at location x determined by:

$$y^2 + z^2 = a^2 = r_1^2 - x^2 = r_2^2 - (x - \Delta x)^2 \quad (4.6)$$

Which can be solved for x^* :

$$x^* = \frac{\Delta x^2 + r_1^2 - r_2^2}{2\Delta x} \quad (4.7)$$

The overlapping volume can be described as a combination of two spherical domes with the disk as described by equation 4.6 as the base. The volume of each such dome is:

$$V = \pi h \frac{h^2 + 3a^2}{6} \quad (4.8)$$

where a is the radius of the disk and h the height of the dome. The value of h differs between the two domes and is $r_1 - x^*$ and $r_2 - (\Delta x) + x^*$ for the ribosome at the origin and the ribosome shifted by Δx respectively. Upon assumption that the translation rate is constant the length of the transcripts of consecutive ribosomes related such that $r_1 = r$ and $r_2 = r + \Delta r$, with Δr as determined by 4.3. This ribosome arrival time also determines the distance between the ribosome, Δx , with equation 4.2.

Plugging this in, along with dividing by the volume of a whole sphere, one can get the following equation for the fraction of overlap of the spheres.

$$\Psi(\Delta t, r) = \begin{cases} \frac{\pi(\Delta r - \Delta x + 2r)^2(-3\Delta r^2 + 2\Delta r\Delta x + \Delta x(\Delta x + 4r))}{12\Delta x \cdot \frac{4}{3}\pi r^3} & \Delta r + 2r > \Delta x \\ 0 & \Delta r + 2r < \Delta x \end{cases} \quad (4.9)$$

4. SPATIAL ASPECTS OF TF DIMERISATION

where the output is a relative value going from 0 (no overlap at all) to 1 (complete overlap of partially formed monomers), Δx , is the distance between two ribosomes, and r is the current length of the least transcribed monomer.

For each ribosome arrival time, Δt , I use the michaelis menten equation to calculate the fraction of dimers after translation. These fractions are multiplied by the the distribution in equation 4.1 to get the weighed dimerisation fraction, α . Finally I integrate over the whole transcription process, from 0AA to the full length of the protein-monomer, to get:

$$\alpha(\Delta t, r) = \lambda \int_0^\infty e^{-\lambda \Delta t} \frac{\int_0^{R_{\text{protein}} - \Delta r} (\Psi(\Delta t, r))^2 dr}{(R_{\text{protein}} - \Delta r)/k_D + \int_0^{R_{\text{protein}} - \Delta r} (\Psi(\Delta t, r))^2 dr} d\Delta t \quad (4.10)$$

where

R_{protein} is the Final length of the protein

k_D is the *Reaction parameter* for the amino acid chains to dimerise when they co-exist in the same volume

and λ is the Typical ribosome arrival time

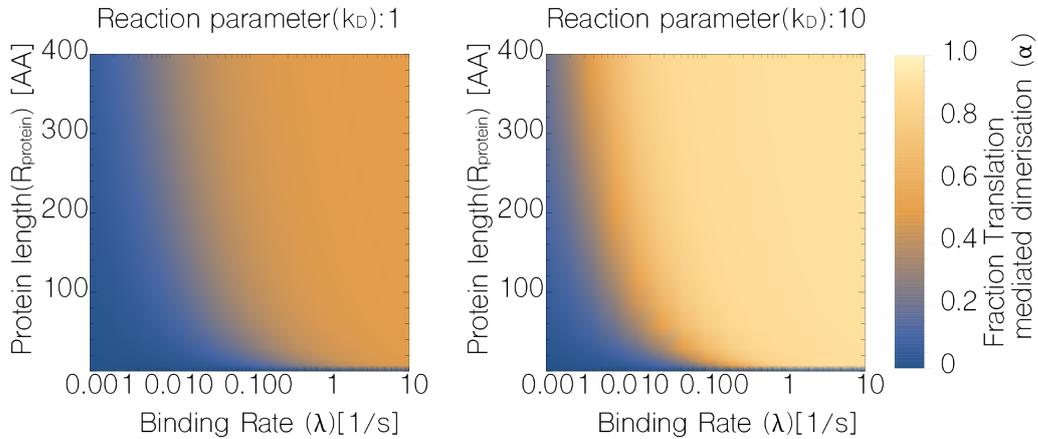


Figure 4.3: **Translation-mediated dimerisation in relation to ribosome binding and protein length.** The fraction of translation-mediated dimerisation (i.e., total dimerisation minus dimerisation in the cellular cytosol), α , as described by equation 4.10, responds primarily to ribosome binding: if binding increases, α also increases. The dimensionless rate *reaction parameter* (k_D) shows an important role: α increases as more likely two partially formed monomers would interact (right plot).

Figure 4.3 shows the resulting fraction of translation-mediated dimerisation (α) given the binding rate of the ribosomes (λ) and the length of the monomers (measured in amino acids). These results are shown for two different values of the reaction parameter (k_D , dimensionless), and used experimentally obtained values for the binding rate [41], $\lambda \approx 0.1s^{-1}$. As observed here, α is minimal (from 0 to ≈ 0.2) at lower values of λ and protein length; although λ is the limiting rate here, since at very low values of this rate, the protein length does not make a difference. This suggests that a very low binding rate (or low ribosome availability) would result in having no translation-mediated dimerisation (or small values which could

4. SPATIAL ASPECTS OF TF DIMERISATION

be neglected). However, as soon as λ increases, α gains importance, also amplified by increasing the reaction parameter k_D . This is because the more λ increases, the more ribosomes will bind to the RNA within a given time interval; as a consequence, ribosomes will be physically closer while translating and partially-formed monomers will tend to overlap more (i.e., higher values at Equation 4.9).

4.3.2 Engineering α , the translation mediated dimerisation fraction

The results shown in Figure 4.3 suggest there is a fragile equilibrium in the fraction of translation-mediated vs. cytosol dimerisation, which, in turn, would impact on the non-linear dynamics of the system. Therefore, different values of α will modify the performance of genetic circuits that build on non-linear reactions to achieve optimal behaviour. In what follows, I analyse how α may be used to alter oscillations and bistability.

This could be achieved by altering the two parameters in Figure 4.3. λ could be influenced by altering how ribosomes bind are recruited. Changing the ribosome binding site [67] or the concentration of available (orthogonal) ribosomes could be regulated[77]. The protein length could also artificially be increases to increase the overlap leading to increased α .

4.3.3 Case study revisited: Effected damping in the repressor

A general requirement in order to obtain oscillations from a biological system is that its kinetic machinery must be non-linear [72]. Here, I modify the well-

known mathematical model of a three-component genetic oscillator [30] with the parameter that represents the fraction of dimerisation that takes place during translation (α). I analyse how this parameter can, by itself, modify the damping of oscillations of the system.

This genetic circuit is formed by a ring of repressors (proteins that negatively regulate a promoter), where each one regulates its successor (Figure 4.4A). As a result, the concentration of all three proteins within the circuit oscillates in time. The system of ODE that represents the oscillator is described as follows. Firstly, the ODE that calculates the concentration of monomeric repressors is:

$$\frac{dR_{i,\text{mono}}}{dt} = k_{\text{on}}(1 - \alpha)P_i - k_D R_{i,\text{mono}} - k_{\text{dim}} R_{i,\text{mono}}^2 \quad (4.11)$$

where $i \in 1, 2, 3$ is a specific repressor protein

R_{mono} is the amount of repressor monomer in the cytosol (i.e., those monomeric regulators that did not dimerise during translation)

k_{on} is the The expression rate of monomers when the promoter P is not repressed (thus fully active)

k_D represents the degradation rate of monomers

k_{dim} is the The rate of dimerisation in the cytosol

and

α is the fraction of monomers that become dimers during translation-mediated dimerisation (which will dissolve as dimers in the cytosol).

Secondly, the ODE that calculate repressor dimers (i.e., the fully active pro-

4. SPATIAL ASPECTS OF TF DIMERISATION

tein) is given by:

$$\frac{dR_{i,\text{dimer}}}{dt} = k_{\text{on}}\alpha P_i - k_{\text{D}}R_{i,\text{dimer}} + k_{\text{dim}}R_{i,\text{mono}}^2 \quad (4.12)$$

where R_{dimer} is the dimerised repressor (note that in this model all three repressors of the system are assumed to be dimers in its final form).

Lastly, the next ODE calculates the fraction of promoter P that is active (i.e., it is not being repressed by any R):

$$\frac{dP_i}{dt} = k_{\text{un}}(1 - P_i) - \frac{1}{\tau_{\text{s,local}}}R_{i-1,\text{dimer}} \quad (4.13)$$

where

k_{un} is the Non-specific DNA unbinding

and $\tau_{\text{s,local}}$ is the local search time

Note that promoter number i is repressed by a repressor number $i - 1$, where the previous element to $i = 1$ is $i = 3$, since the three genetic elements [1,2,3] are arranged in a ring.

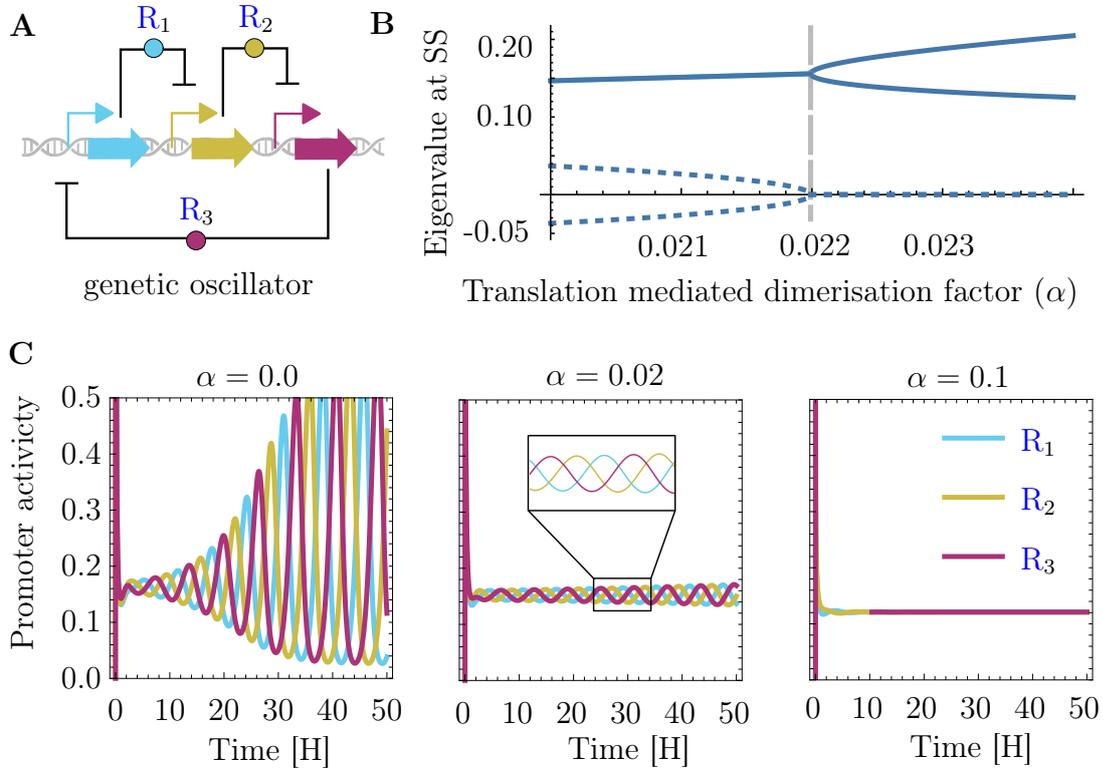


Figure 4.4: **Impact of translation-mediated dimerisation (α) on a genetic oscillator.** (A): Diagram of a three-component genetic oscillator (as in [30])—each repressor protein (R) inhibits the expression of its successor. As a result, the concentration of each repressor oscillates in time. (B): Eigenvalue analysis, which shows bifurcation point around $\alpha = 0.022$. This suggests that the emergence of oscillations is highly dependent on of the balance between translation-mediated and cytosol dimerisation. Solid and dotted lines are the real and imaginary values of the eigenvalues in the jacobian at equilibrium respectively. (C): Time-course simulations of the genetic oscillator at different values of α . Oscillations are damped, and the limit cycle removed, at the bifurcation point as indicated in B.

Figure 4.4 shows how the parameter α alters the damping of the system, to

4. SPATIAL ASPECTS OF TF DIMERISATION

the point that the circuit stops oscillating if the balance between translation-mediated vs. cytosol dimerisation is beyond a bifurcation point (Figure 4.4B). This is, higher values of α lead to a non-oscillating steady state (Figure 4.4C). This bifurcation, beyond which the oscillatory behaviour vanishes is around $\alpha = 0.022$, which indicates that, even if the majority of dimers are formed in the cytosol, the ones that form during translation can drastically affect circuit behaviour.

The new parameter, α , introduces complexity to the mathematical model in that it limits the range of parameter values that generate oscillations—it could be argued that it is more difficult to get reliable oscillations than without it. However, it offers several pathways to modify damping in-vivo, unlike traditional methods for nonlinear dynamics (e.g., Hill coefficients). Therefore, its predictive scope narrows the gap that goes from modelling results to in-vivo experimentation.

4.3.4 Case study: programmed damping in a genetic toggle switch

A genetic toggle switch [31] is a device build from two mutually inhibitory repressors that is able to flip between stable states (Figure 4.5), in which only one of the two proteins is at high expression while the other one is inhibited. The stability of the system, and also the number of stable states, depends intricately on non-linear dynamics. Although there are examples of genetic switches where non-linearity emerges from protein dilution [50, 91], the most common way of achieving non-linear dynamics (and bistability) comes from transcription cooperativity—in which protein dimerisation plays a fundamental role.

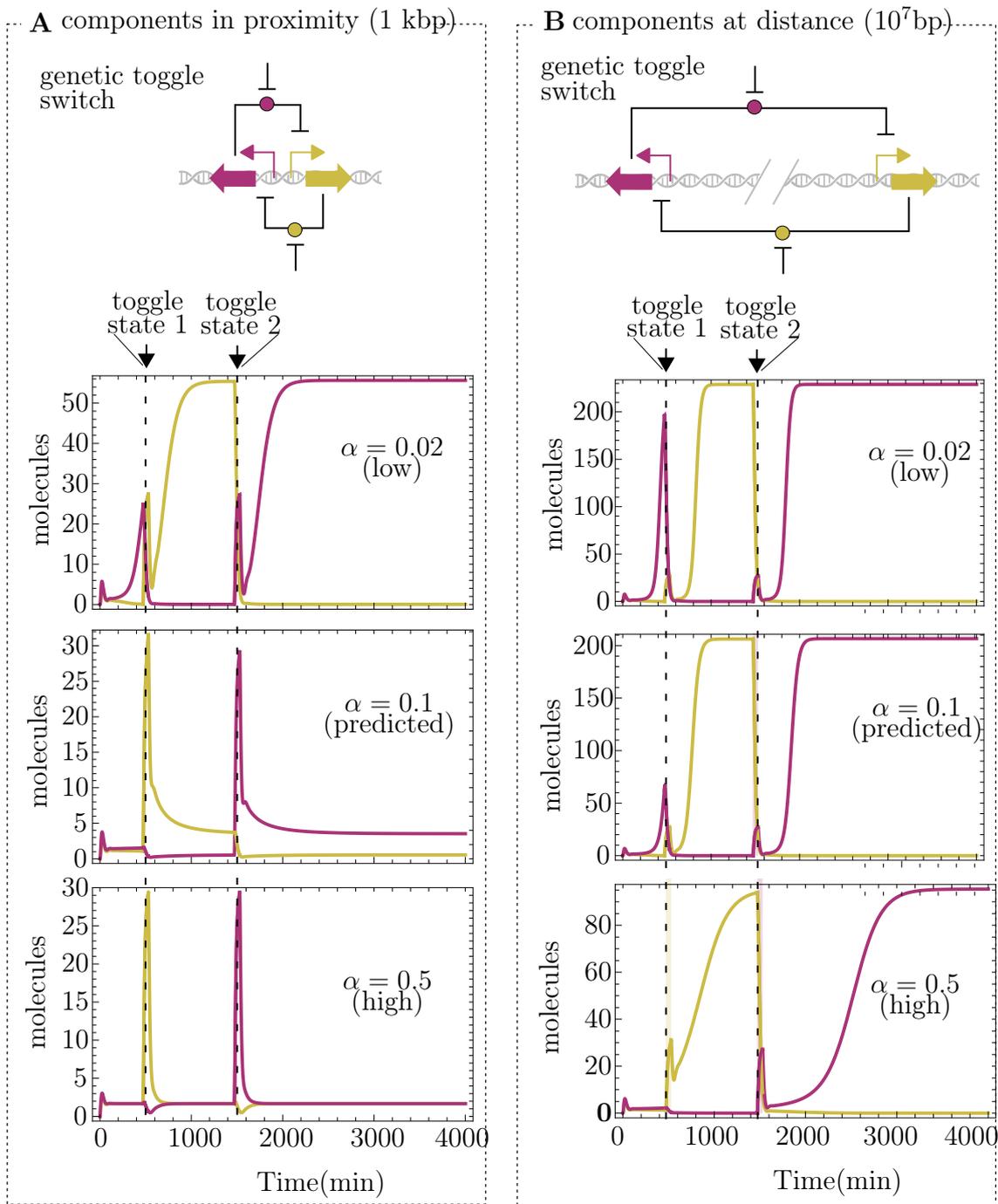


Figure 4.5: **Impact of translation-mediated dimerisation (α) and gene separation on a genetic toggle switch.** (A): Genetic components in proximity. Only in the case of α being low, the damping is low enough to achieve bistability. (B): Genetic components at distance. The physical separation forces proteins to *travel* from its source gene to its target promoter—a nonlinear process which counteracts high α values. A & B The value of $\alpha=0.1$ ("predicted") is our theoretical approximation; up to now, this value has not been experimentally obtained.

4. SPATIAL ASPECTS OF TF DIMERISATION

My model for the toggle switch is specified by equations 4.11, 4.12 and 4.13, but with two (instead of three) repressor proteins (which are also assumed to be dimers). Furthermore, I added complexity to the model by considering the intracellular spatial distribution of genes, based on our recent work on spatio-temporal design [40, 87]. This feature helps differentiating between *local* and *global* repressors; this is, repressors that are in the proximity of, or far from, their encoding gene. The core message was that the distance a protein must “travel” from its source gene to its target promoter, modifies regulation in a predictable fashion. Since the differentiation between local and global proteins intersects with the model I introduce here, where proteins dimerise during translation (i.e., local) or in the cytosol during free diffusion (i.e., global), I analysed the genetic switch considering both dynamics.

Figure 4.5 shows results of simulating the switch with two different spatial setups, proximity (Figure 4.5A) and distance (Figure 4.5B), and three values of translation-mediated dimerisation (α) in each case. Interestingly, the two spatial configurations show different performance for the same level of α , which implies that both phenomena (intergenic separation and protein dimerisation) are both involved in shaping nonlinear dynamics. When the genes of the switch are placed in proximity (Figure 4.5A), only a low value of α reduces damping sufficiently for the system to show bistability. This is because the generation of dimers by translation-mediated dimerisation is fully linear. As soon as this value α increases, more dimerisation takes during translation, leading to system malfunction; although the switching still occurs to some extent (e.g., $\alpha = 0.5$), output values are low, and the intrinsic stochasticity of biological systems would result in unstable states. However, moving genes far apart (Figure 4.5B) restored

the function of the system, even in the case of $\alpha = 0.5$ (a value that seems higher than physically plausible). By increasing the separation between the genes, the dimers formed near the source are not immediately available to bind to the target promoter, since this is now at a long distance. As a result, such linear process is now less relevant: translation mediated dimers must “travel” from source to target resulting in a decreased promoter-binding rate, which, in turn, removes its involvement in total promoter binding events. Unlike the *proximity* scenario, where performance was only achieved at low values for α , the *distance* set up showed bistability at any value. This highlights the role played by intra-cellular distance for circuit design.

Altogether, simulations suggest that the necessary conditions for the emergence of bistability could be rationally designed or fine-tuned, and mapped into biological specifications such as ribosome binding sites (which impacts on α) or inter-genic separation.

4.4 Discussion

The design of increasingly complex biocomputing circuits in cells is a major challenge. The lack of robust predictive modelling—which could accurately foresee the performance of a design before implementation—threatens to undermine the success of the field. Although model-based design [101] is a common practice, accurate predictions are difficult to obtain since gene regulation is still based on unclear dynamics. Here, I focus on modelling protein dimerisation [10] within regulatory interactions, as a way to fine-tune system damping.

The model for dimerisation is based on differentiating between those monomers

4. SPATIAL ASPECTS OF TF DIMERISATION

that interact during translation (what I termed *translation-mediated dimerisation*) and those that dimerise within the cytosol of the cell while diffusing. While the former imposes a linear regime on the reaction, the latter boosts the emergence of non-linear dynamics. By controlling the fraction of each dimerisation type, on-demand fine-tuning damping may be capable. My model simulates this process and suggests a number of routes on how to implement such control in-vivo. For instance, 1) protein length [17] (adding tags or extra amino acids), 2) choice of RBS [79] (stronger RBS will result in closer ribosomes on the mRNA), and 3) ribosome availability [20] (influencing binding rates). The model suggests that considering these features could result in having controllable levels of damping.

Furthermore, I focus on the implication that intergenic distance (i.e., physical distance between two genes within the volume of a cell) has for the system. Based on my model for spatiotemporal design in chapter 3, simulations suggest that the linear dynamics imposed by translation-mediated dimerisation are less important if two interacting genes are far apart. This implies that damping can be also programmed by using spatial constraints.

4.5 Future work

The current work uses the assumption that the free dimer ends are in a flat distribution circular around the bound end. This does not take into account structural qualities of a polypeptide. Ideally, the whole Dimerisation During Translation would be simulated with molecular dynamics. Alternatively, using a model for the dimerisation rates describing molecules with a worm-like chain around the bound end could improve the description of a polypeptide.

To verify the current model, one could try to check whether it is more likely to dimerise with co-located monomers. A hypothesised method to check this is to implement two copies of the same transcription factor with varying fluorescent tags (such as the CFP-YFP pair). If the structure is very specifically designed, the excited fluorescent state might transfer to the other fluorescent tag with Förster resonance (FRET). In the well-mixed model, half of the dimers would have hetero-fluorescent tags with FRET properties, while the other homo-fluorescent tagged dimers would only have a single excitation wavelength. With Translation Mediated Dimerisation, there would be a higher fraction of homo-tagged dimers, thus showing less FRET behaviour. This would require a skilled design of the fluorescent tags, so that their dipoles align correctly, and would require highly precise measurements.

Chapter 5

Conclusion

In this thesis, I showed how the small distances between genes in bacteria can still affect the behaviour of the cell. In a new model I show that a transcription factors' *local search*, in the first seconds after creation, can compete with the *global search*, lasting protein lifetimes. I describe that this local search is dependent on several parameters, one of which is the intergenetic distance between the *source* gene and the *target* gene. I show that in physiological ranges, the search is catalysed when the source and target genes are within 1000 base pairs of each other.

An essential assumption is that after the creation of active transcription factor it is localised at the source gene. While this could be argued for monomers, created by a coupled transcribing-translating *expressome*, models using dimerisation in the cytosol did not allow a localised final form of the transcription factor. I developed a new model to describe spatial effects where partially formed monomers dimerise *during* the translation process. In this model dimers *can* be localised near the encoding gene after creation. I showed that while this process can catalyse the search for a target gene the reaction. Moreover, damping of non-linear

system behaviour is seen to increase in the scenario where transcription factors dimerise during the translation process. I show how the fraction that dimerises during translation increases with the rate that ribosomes bind and the final length of the protein.

These models allow for the rational design of circuits: using co-location to catalyse a circuit or the reduction of either length or RBS to tune system damping of non-linear effects. Moreover, these models can be rapidly tested and be iteratively improved by comparing the designed behaviour using these models and reality.

List of symbols

Term	Symbol	Description	Value	Units	Source	Page List
1D-diffusion rate	D_{1D}	Diffusion of a Transcription factor along the chromosome while remaining bound	varied	$\mu m^2 s^{-1}$	[82]	37 , 39 , 40 , 42 , 43 ,
-Fig 3.13 high			2.62			
-Fig 3.13 low			0.000262			
-General			0.0262			
3D-diffusion rate			2.72	$\mu m^2 s^{-1}$	[82]	
Amino Acid width	x_{AA}		3	\AA		74 ,
Base pair width	x_{bp}		3	\AA		74 ,
Binding rate	k_{bind}	Binding rate transcription factor				27 , 28 , 44 ,
Chance of binding when at TG	γ	Chance that a transcription factor binds to the target promoter when both elements are co-located	varied			48–50 ,
-General		Intermediate binding value used for Fig 3.13 , 3.10B & 3.12	0.1			

Term	Symbol	Description	Value	Units	Source	Page List
-Immediate		Immediate binding used for Fig3.10,3.11,3.14	1			
-Low		Low binding value used for figure 3.12	0.01			
Chemical species concentration	ρ	Chemical species concentration, mostly in context means transcription factor density				16, 39–42,
Chromosome length	M	Length of chromosome, value used here is based on the E. coli chromosome	10^7	bp	[8]	43,
Degradation rate	k_{deg}	Inverse of Transcription factor lifetime				44,
-mrna degradation rate	k_{Dm}	Typical time it takes for a mRNA to be broken down				27, 29, 46,
-Protein degradation rate	k_{Dp}	Typical time it takes for a protein to be broken down				47, 50, 58,
Diffusion constant	D					28, 29, 48–50,
Dimerisation volume	$\Psi(\Delta t, r)$	The normalised volume where co-translating proteins can interact				16, 26, 28, 29,
Distance along chromosome	k_{dim}	The rate of dimerisation in the cytosol				76, 77,
Distance along chromosome	d	Distance between genes where the Transcription factor is encoded and its cognate promoter				80,
						37, 38,

Term	Symbol	Description	Value	Units	Source	Page List
Distance along chromosome	k_{on}	The expression rate of monomers when the promoter P is not repressed (thus fully active)				80,
divergence operator	Δ	divergence chemical species				16,
Doubling time	τ_2	The time it takes for a the population to double in number (Usually related to the time it takes for an individual to be “born” to reproduction)	varied	min	[5]	10, 11,
-Fig 3.8		Value for a slow growing population	120			
-general		Value for a fast growing population	20			
Flux	j	The flux of a chemical species				16,
Global search time	$\tau_{s,global}$	Global search time (described in Equation 3.18)				43, 44, 48–50,
Green Fluorescent protein	GFP	Reporter protein tagged with a green fluorescent component				47, 49, 50,
Growth rate	α	The growth rate of the Population				
Growth rate	μ	$=1/\tau$. Rate at which a Population grows. Inversely proportional to the Typical growth time.	$(\ln(2)\tau_2)^{-1}$			10, 11, 46–50, 58,
Initiation time difference	Δt	Time difference between binding of two consecutive ribosomes				73, 74, 77,

Term	Symbol	Description	Value	Units	Source	Page List
Inter Ribosomal distance	Δx	Distance between two consecutive ribosomes				73–77,
Laplace operator	∇^2					28, 29,
Leakiness production	α_{leak}	Basal transcription of the promoter when repressed	varied			46, 47, 50, 58,
-Fig 3.12			0.005			
-Fig 3.12& 3.13			0.002			
-Zero expression, used in 3.13			0			
local search time	$\tau_{\text{s,local}}$	Local search time (described in Equation 3.11)				37, 42, 44, 48–50, 81,
mRNA	m	Population of protein encoding RNA.				46, 47, 49, 50, 58,
-mRNA field	\mathbf{m}	mRNA concentration field				28, 29,
Natural logarithm	\ln	The logarithm base e				11,
Non-specific DNA unbinding	k_{un}	Rate at which a Transcription factor unbinds from chromosome when it does not interact with its cognate promoter	10	s^{-1}	[106]	37, 39, 40, 42–44, 50, 81,

Term	Symbol	Description	Value	Units	Source	Page List
Partial protein length	r	Length of a protein during its translation, which is time dependent				75–77,
Population	Pop	The count of total organisms of a growing population, generally in the scope of bacteria				10, 11,
-Foxes	F	The population of Foxes				12, 13,
-Hares	H	The population of Hares				12, 13,
Probability distribution	Pr	The Probability density function of the parameter specified				73,
Promoter availability	P	The relative production rate of mRNA of the promoter encoding the protein indicated by the descriptor to its constitutive expression				46–50, 58,
-Inhibited promoter density	iP	Concentration field of the inhibited promoters				80, 81,
-Promoter density	P	Concentration field of the non-inhibited promoters				28, 29,
Protein length	R_{protein}	Final length of the protein	varied			77,
Protein length difference	Δr	Difference in length of partially translated proteins (or monomers)				74, 76, 77,

Term	Symbol	Description	Value	Units	Source	Page List
Rate of binding at the target	$k_{b,TF@TG}$	Rate at which a Transcription factor is attracted to the minimal free energy state(bound) when at the cognate promotor location. Assumed to be very high.				vi, 38, 41, 42, 44,
Rate of specific unbinding	k_{unr}	Re-scaled specific unbinding rate from the target,which also accounts for re-binding events.				vi, 27, 28, 38, 42, 44, 48, 49,
Rate of transcription	β_{transl}					27, 29, 44, 46, 47, 50,
-Rate of ribosome transcription	k_{transl}	The rate of transcription of a single bound ribosome				74,
Rate of translation	β_{transc}					27, 29, 44, 48–50, 58,
Reaction parameter	k_D	<i>Reaction parameter</i> for the amino acid chains to dimerise when they co-exist in the same volume				77–81,
Transcription factor	TF					3, 4
-Global descriptor	global	Descriptor of an instance of Transcription factor indicating that it has unbound from the DNA strand at some point				

Term	Symbol	Description	Value	Units	Source	Page List
-Local descriptor	$local$	Descriptor of an instance of Transcription factor indicating that it has not unbound from DNA since formation				
-Repressor	R	Repressor protein				47–50, 80–82,
-TF field	TF					28, 29,
Transcription factor lifetime	k_{deg}^{-1}	Typical time it takes for a chemical species to be broken down	40	min	[5]	
Translation mediated dimerisation fraction	α	missing				vii, 77–86,
Typical growth time	τ	The time it takes for the population to grow a factor e	$= \tau_2 / \ln(2)$			11,
Typical ribosome arrival time	λ	The typical Initiation time difference				73, 77–79,

Bibliography

- [1] Gary K Ackers, Alexander D Johnson, and Madeline A Shea. Quantitative model for gene regulation by λ phage repressor (gene regulation/repressors and operators/cooperative interaction/thermodynamic model). *Proc. Natl Acad. Sci. USA*, 79(February):1129–1133, 1982. [71](#)

- [2] Steven S Andrews, Nathan J Addy, Roger Brent, and Adam P Arkin. Detailed simulations of cell biology with Smoldyn 2.1. *PLoS Computational Biology*, 6(3):e1000705, 2010. ISSN 1553734X. doi: 10.1371/journal.pcbi.1000705. [64](#)

- [3] Christian Atallah, David James Skelton, Simon J. Charnock, and Anil Wipat. Functional analysis of enzyme families using residue-residue co-evolution similarity networks. *bioRxiv*, 2019. doi: 10.1101/646539. [18](#)

- [4] Holly J Atkinson, John H Morris, Thomas E Ferrin, and Patricia C Babbitt. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PloS one*, 4(2):e4345, 2009. [17](#)

- [5] Subhayu Basu, Yoram Gerchman, Cynthia H. Collins, Frances H. Arnold, and Ron Weiss. A synthetic multicellular system for programmed pat-

BIBLIOGRAPHY

- tern formation. *Nature*, 434:1130, 4 2005. ISSN 0028-0836. doi: 10.1038/nature03461. [94](#), [98](#)
- [6] Otto G Berg, Robert B. Winter, and Peter H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, 20(24):6929–6948, 11 1981. ISSN 0006-2960. doi: 10.1021/bi00527a028. [25](#), [45](#)
- [7] Paul C Blainey, Antoine M van Oijen, Anirban Banerjee, Gregory L Verdine, and X Sunney Xie. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proceedings of the National Academy of Sciences*, 103(15):5752–5757, 2006. [35](#)
- [8] F. R. Blattner. The Complete Genome Sequence of Escherichia coli K-12. *Science*, 277(5331):1453–1462, 9 1997. ISSN 00368075. doi: 10.1126/science.277.5331.1453. [93](#)
- [9] Dena H S Block, Razika Hussein, Lusha W. Liang, and Han N. Lim. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Research*, 40(18):8979–8992, 10 2012. ISSN 03051048. doi: 10.1093/nar/gks694. [64](#)
- [10] Mehdi Bouhaddou and Marc R Birtwistle. Dimerization-based control of cooperativity. *Molecular bioSystems*, 10(7):1824–1832, 2014. [87](#)
- [11] Björn M Burmann, Kristian Schweimer, Xiao Luo, Markus C Wahl, Barbara L Stitt, Max E Gottesman, and Paul Rösch. A NusE: NusG complex links transcription and translation. *Science*, 328(5977):501–504, 2010. [36](#)
- [12] Michele Castellana, Sophia Hsin Jung Li, and Ned S Wingreen. Spatial

- organization of bacterial transcription and translation. *Proc. Natl. Acad. Sci. U. S. A.*, 113(33):9286–9291, 2016. ISSN 10916490. doi: 10.1073/pnas.1604995113. [70](#)
- [13] Max Chavarría, Ángel Goñi-Moreno, Víctor de Lorenzo, and Pablo I Nikel. A Metabolic Widget Adjusts the Phosphoenolpyruvate-Dependent Fructose Influx in *Pseudomonas putida*. *mSystems*, 1(6):e00154–16, 12 2016. ISSN 2379-5077. doi: 10.1128/mSystems.00154-16. [58](#)
- [14] Kyoung-Hee Choi, Jared B Gaynor, Kimberly G White, Carolina Lopez, Catharine M Bosio, RoxAnn R Karkhoff-Schweizer, and Herbert P Schweizer. A Tn7-based broad-range bacterial cloning and expression system. *Nature methods*, 2(6):443, 2005. [65](#)
- [15] Shasha Chong, Chongyi Chen, Hao Ge, and X. Sunney Xie. Mechanism of transcriptional bursting in bacteria. *Cell*, 158(2):314–326, 7 2014. ISSN 10974172. doi: 10.1016/j.cell.2014.05.038. [70](#)
- [16] Frank C Collins and George E Kimball. Diffusion-controlled reaction rates. *Journal of Colloid Science*, 4(4):425–437, 1949. ISSN 00958522. doi: 10.1016/0095-8522(49)90023-9. [35](#)
- [17] Andrew Currin, Neil Swainston, Philip J Day, and Douglas B Kell. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society Reviews*, 44(5):1172–1239, 2015. [88](#)
- [18] Giovanni Dalmaso, Paula Andrea Marin Zapata, Nathan Ryan Brady, and Anne Hamacher-Brady. Agent-based modeling of mitochondria links sub-

BIBLIOGRAPHY

- cellular dynamics to cellular homeostasis and heterogeneity. *PLoS One*, 12(1):e0168198, 2017. [18](#)
- [19] Ramiz Daniel, Jacob R Rubens, Rahul Sarpeshkar, and Timothy K Lu. Synthetic analog computation in living cells. *Nature*, 497(7451):619–623, 2013. ISSN 1476-4687. doi: 10.1038/nature12148. [17](#), [63](#)
- [20] Alexander P S Darlington, Juhyun Kim, José I Jiménez, and Declan G Bates. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nature communications*, 9(1):1–12, 2018. [72](#), [88](#)
- [21] Aitor de las Heras, Sofia Fraile, and Víctor de Lorenzo. Increasing signal specificity of the TOL network of *Pseudomonas putida* mt-2 by rewiring the connectivity of the master regulator XylR. *PLoS genetics*, 8(10):e1002963, 2012. [58](#)
- [22] Michael C DeSantis, Je-Luen Li, and Y M Wang. Protein sliding and hopping kinetics on DNA. *Phys. Rev. E*, 83(2):21907, 2 2011. doi: 10.1103/PhysRevE.83.021907. [66](#)
- [23] Charles J. Dorman. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nature Reviews Microbiology*, 11(5):349–355, 5 2013. ISSN 1740-1526. doi: 10.1038/nrmicro3007. [9](#)
- [24] Peter Dröge and Benno Müller-Hill. High local protein concentrations at promoters: Strategies in prokaryotic and eukaryotic cells. *BioEssays*, 23(2):179–183, 2001. ISSN 02659247. doi: 10.1002/1521-1878(200102)23:2<179::AID-BIES1025>3.0.CO;2-6. [20](#), [35](#), [64](#)

- [25] Johan Elf. Fast Evaluation of Fluctuations in Biochemical Networks With the Linear Noise Approximation. *Genome Research*, 13(11):2475–2484, 2003. ISSN 1088-9051. doi: 10.1101/gr.1196503. [21](#)
- [26] Johan Elf, Gene-Wei G.-W. Li, and X Sunney Xie. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science*, 316(5828):1191–1194, 5 2007. ISSN 0036-8075. doi: 10.1126/science.1141967. [64](#)
- [27] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 1 2000. ISSN 0028-0836. doi: 10.1038/35002125. [71](#)
- [28] Jacob A. Englaender, J. Andrew Jones, Brady F. Cress, Thomas E. Kuhlman, Robert J. Linhardt, and Mattheos A.G. G. Koffas. Effect of Genomic Integration Location on Heterologous Protein Expression and Metabolic Engineering in *E. coli*. *ACS Synthetic Biology*, 6(4):710–720, 4 2017. ISSN 21615063. doi: 10.1021/acssynbio.6b00350. [66](#)
- [29] David R Espeso, Esteban Martínez-García, Víctor de Lorenzo, and Ángel Goñi-Moreno. Physical Forces Shape Group Identity of Swimming *Pseudomonas putida* Cells. *Frontiers in microbiology*, 7:1437, 2016. [23](#)
- [30] Jordi Garcia-Ojalvo, Michael B Elowitz, and Steven H Strogatz. Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):10955–60, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0307095101. [32](#), [81](#), [83](#)

BIBLIOGRAPHY

- [31] Timothy S Gardner, Charles R Cantor, James J Collins, Timothy S Gardner, and Charles R Cantor. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767):339–342, 2000. ISSN 00280836. doi: 10.1038/35002131. [71](#), [84](#)
- [32] Michael A Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, 104(9):1876–1889, 2000. [19](#)
- [33] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976. ISSN 00219991. doi: 10.1016/0021-9991(76)90041-3. [14](#)
- [34] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of chemical physics*, 115(4):1716–1733, 2001. [19](#)
- [35] Daniel T Gillespie and Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 93(5):2340–2361, 1977. ISSN 0022-3654. doi: 10.1021/j100540a008. [51](#)
- [36] Angel Goñi-Moreno and Martyn Amos. Continuous computation in engineered gene circuits. *Biosystems*, 109(1):52–56, 2012. [25](#)
- [37] Angel Goñi-Moreno and Martyn Amos. A reconfigurable NAND/NOR genetic logic gate. *BMC systems biology*, 6(1):126, 2012. [25](#)
- [38] Angel Goni-Moreno and Martyn Amos. DiSCUS: a simulation platform

- for conjugation computing. In *International Conference on Unconventional Computation and Natural Computation*, pages 181–191. Springer, 2015. [23](#)
- [39] Angel Goñi-Moreno, Miguel Redondo-Nieto, Fernando Arroyo, and Juan Castellanos. Biocircuit design through engineering bacterial logic gates. *Natural Computing*, 10(1):119–127, 2011. [23](#)
- [40] Ángel Goñi-Moreno, Ilaria Benedetti, Juhyun Kim, and Víctor Víctor de Lorenzo. Deconvolution of Gene Expression Noise into Spatial Dynamics of Transcription Factor–Promoter Interplay. *ACS Synthetic Biology*, 6(7):1359–1369, 7 2017. ISSN 2161-5063. doi: 10.1021/acssynbio.6b00397. [19](#), [43](#), [58](#), [60](#), [64](#), [65](#), [72](#), [86](#)
- [41] Thomas E Gorochofski, Irina Chelysheva, Mette Eriksen, Priyanka Nair, Steen Pedersen, and Zoya Ignatova. Absolute quantification of translational regulation and burden using combined sequencing approaches. *Molecular Systems Biology*, 15(5):1–15, 5 2019. ISSN 1744-4292. doi: 10.15252/msb.20188719. [79](#)
- [42] J Gowrishankar and R Harinarayanan. Why is transcription coupled to translation in bacteria? *Molecular microbiology*, 54(3):598–603, 2004. [36](#)
- [43] Juliette Griffié, Ruby Peters, and Dylan M Owen. An agent-based model of molecular aggregation at the cell membrane. *Plos one*, 15(2):e0226825, 2020. [18](#)
- [44] Lewis Grozinger, Martyn Amos, Thomas E. Gorochofski, Pablo Carbonell, Diego A. Oyarzún, Ruud Stoof, Harold Fellermann, Paolo Zuliani, Huseyin

BIBLIOGRAPHY

- Tas, and Angel Goñi-Moreno. Pathways to cellular supremacy in biocomputing. *Nature communications*, 10(1):1–11, 2019. ISSN 20411723. doi: 10.1038/s41467-019-13232-z. [5](#), [6](#)
- [45] Raúl Guantes and Juan F. Poyatos. Dynamical Principles of Two-Component Genetic Oscillators. *PLoS Computational Biology*, 2(3):e30, 3 2006. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0020030. [26](#)
- [46] Anna-Karin Gustavsson, David D van Niekerk, Caroline B Adiels, Franco B du Preez, Mattias Goksör, and Jacky L Snoep. Sustained glycolytic oscillations in individual isolated yeast cells. *The FEBS journal*, 279(16): 2837–2847, 2012. [18](#)
- [47] Petter Hammar, Mats Walldén, David Fange, Fredrik Persson, Özden Baltekin, Gustaf Ullman, Prune Leroy, and Johan Elf. Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nature Genetics*, 46(4):405–408, 4 2014. ISSN 15461718. doi: 10.1038/ng.2905. [42](#)
- [48] Anders S Hansen and Erin K O’shea. Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Molecular systems biology*, 9(1):704, 2013. [56](#)
- [49] R Hershberg, E Yegerlotem, and H Margalit. Chromosomal organization is shaped by the transcription regulatory network. *Trends in Genetics*, 21(3): 138–142, 3 2005. ISSN 01689525. doi: 10.1016/j.tig.2005.01.003. [9](#), [23](#)
- [50] Daniel Huang, William J Holtz, and Michel M Maharbiz. A genetic bistable

- switch utilizing nonlinear protein degradation. *Journal of biological engineering*, 6(1):9, 2012. [84](#)
- [51] Akira Ishihama, Ayako Kori, Etsuko Koshio, Kayoko Yamada, Hiroto Maeda, Tomohiro Shimada, Hideki Makinoshima, Akira Iwata, and Nobuyuki Fujita. Intracellular concentrations of transcription factors in *Escherichia coli*: 65 species with known regulatory functions. *Journal of Bacteriology*, pages JB–01579, 2014. [20](#), [35](#), [64](#)
- [52] Juhyun Kim, Angel Goñi-Moreno, Belén Calles, and Víctor de Lorenzo. Spatial organization of the gene expression hardware in *Pseudomonas putida*. *Environmental microbiology*, 2019. [45](#)
- [53] Stefan Klumpp, Matthew Scott, Steen Pedersen, and Terence Hwa. Molecular crowding limits translation and cell growth. *Proceedings of the National Academy of Sciences of the United States of America*, 110(42):16754–9, 2013. ISSN 1091-6490. doi: 10.1073/pnas.1310377110. [70](#)
- [54] R Kohler, R A Mooney, D J Mills, R Landick, and P Cramer. Architecture of a transcribing-translating expressome. *Science*, 356(6334):194–197, 4 2017. ISSN 0036-8075. doi: 10.1126/science.aal3059. [20](#), [45](#), [47](#), [64](#), [71](#)
- [55] Grigory Kolesov, Zeba Wunderlich, Olga N Laikova, Mikhail S Gelfand, and Leonid A Mirny. How gene order is influenced by the biophysics of transcription regulation. *Proceedings of the National Academy of Sciences*, 104 (35):13948–13953, 8 2007. ISSN 0027-8424. doi: 10.1073/pnas.0700672104. [20](#), [47](#), [64](#), [71](#)

BIBLIOGRAPHY

- [56] Tino Krell, Wilson Terán, Obdulio López Mayorga, Germán Rivas, Mercedes Jiménez, Craig Daniels, Antonio Jesús Molina-Henares, Manuel Martínez-Bueno, María Trinidad Gallegos, and Juan Luis Ramos. Optimization of the Palindromic Order of the TtgR Operator Enhances Binding Cooperativity. *Journal of Molecular Biology*, 369(5):1188–1199, 2007. ISSN 00222836. doi: 10.1016/j.jmb.2007.04.025. [72](#)
- [57] Stephen M Krone, Ruinan Lu, Randal Fox, Haruo Suzuki, and Eva M Top. Modelling the spatial dynamics of plasmid transfer and persistence. *Microbiology (Reading, England)*, 153(Pt 8):2803, 2007. [18](#)
- [58] Thomas E Kuhlman and Edward C Cox. Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Molecular Systems Biology*, 8(1):610, 9 2012. ISSN 1744-4292. doi: 10.1038/msb.2012.42. [20](#), [36](#), [64](#), [72](#)
- [59] Marco Cosentino Lagomarsino, Olivier Espéli, and Ivan Junier. From structure to function of bacterial chromosomes: Evolutionary perspectives and ideas for new experiments. *FEBS Letters*, 589(20PartA):2996–3004, 10 2015. ISSN 00145793. doi: 10.1016/j.febslet.2015.07.002. [9](#)
- [60] Martin Lind En, Vladimi Curi, Alexis Boucharin, David Fange, Johan Elf, Martin Lindén, Vladimir Ćurić, Alexis Boucharin, David Fange, and Johan Elf. Simulated single molecule microscopy with SMeagol. *Bioinformatics*, 32(15):2394–2395, 2016. doi: 10.1093/bioinformatics/btw109. [64](#)
- [61] Juergen Mairhofer, Theresa Scharl, Karoline Marisch, Monika Cserjan-Puschmann, and Gerald Striedner. Comparative Transcription Profiling

- and In-Depth Characterization of Plasmid-Based and Plasmid-Free Escherichia coli Expression Systems under Production Conditions. *Applied and Environmental Microbiology*, 79(12):3802–3812, 2013. ISSN 0099-2240. doi: 10.1128/AEM.00365-13. [65](#)
- [62] Laura J Marinelli, Graham F Hatfull, and Mariana Piuri. Recombineering: A powerful tool for modification of bacteriophage genomes. *Bacteriophage*, 2(1):5–14, 2012. [65](#)
- [63] Harley H McAdams and Adam Arkin. It’s a noisy business! Genetic regulation at the nanomolar scale, 2 1999. ISSN 01689525. [56](#)
- [64] O. L. Miller, Barbara A. Hamkalo, and C. A. Thomas. Visualization of bacterial genes in action. *Science*, 169(3943):392–395, 1970. ISSN 00368075. doi: 10.1126/science.169.3943.392. [36](#)
- [65] Leonid Mirny, Michael Slutsky, Zeba Wunderlich, Anahita Tafvizi, Jason Leith, and Andrej Kosmrlj. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*, 42(43):434013, 10 2009. ISSN 1751-8113. doi: 10.1088/1751-8113/42/43/434013. [35](#), [45](#)
- [66] Namiko Mitarai and Steen Pedersen. Control of ribosome traffic by position-dependent choice of synonymous codons. *Physical biology*, 10(5):56011, 2013. [74](#)
- [67] Dokyun Na and Doheon Lee. Rbsdesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics*, 26(20):2633–2634, 2010. [80](#)

BIBLIOGRAPHY

- [68] Jonathan Naylor, Harold Fellermann, Yuchun Ding, Waleed K Mohammed, Nicholas S Jakubovics, Joy Mukherjee, Catherine A Biggs, Phillip C Wright, and Natalio Krasnogor. Simbiotics: a multiscale integrative platform for 3d modeling of bacterial populations. *ACS Synthetic Biology*, 6(7):1194–1210, 2017. [18](#)
- [69] Alec A.K. K. Nielsen, Bryan S. Der, Jonghyeon Shin, Prashant Vaidyanathan, Vanya Paralanov, Elizabeth A. Strychalski, David Ross, Douglas Densmore, and Christopher A. Voigt. Genetic circuit design automation. *Science*, 352(6281):aac7341–aac7341, 4 2016. ISSN 0036-8075. doi: 10.1126/science.aac7341. [63](#)
- [70] Pablo I Nickel, Esteban Martínez-García, and Víctor de Lorenzo. Biotechnological domestication of pseudomonads using synthetic biology. *Nature Reviews Microbiology*, 12(5):368, 2014. [58](#)
- [71] Hironori Niki, Yoshiharu Yamaichi, and Sota Hiraga. Dynamic organization of chromosomal DNA in Escherichia coli. *Genes and Development*, 14(2):212–223, 1 2000. ISSN 08909369. doi: 10.1101/gad.14.2.212. [9](#)
- [72] Béla Novák and John J Tyson. Design principles of biochemical oscillators. *Nature reviews Molecular cell biology*, 9(12):981–991, 2008. [80](#)
- [73] Joris Pajmans and Pieter Rein ten Wolde. Lower bound on the precision of transcriptional regulation and why facilitated diffusion can reduce noise in gene expression. *Physical Review E*, 90(3):1–14, 9 2014. ISSN 15502376. doi: 10.1103/PhysRevE.90.032708. [53](#)

- [74] Manuel Pájaro, Antonio A. Alonso, Irene Otero-Muras, and Carlos Vázquez. Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting. *Journal of Theoretical Biology*, 421: 51–70, 2017. ISSN 10958541. doi: 10.1016/j.jtbi.2017.03.017. [70](#), [72](#)
- [75] Rob Phillips, Jane Kondev, and Julie Theriot. *Physical Biology of the Cell*. 2008. [26](#)
- [76] Laurent Potvin-Trottier, Nathan D Lord, Glenn Vinnicombe, and Johan Paulsson. Synchronous long-term oscillations in a synthetic gene circuit. *Nature*, 538(7626):514–517, 2016. ISSN 0028-0836. doi: 10.1038/nature19841. [32](#)
- [77] Oliver Rackham and Jason W Chin. A network of orthogonal ribosome-mrna pairs. *Nature chemical biology*, 1(3):159–166, 2005. [80](#)
- [78] Sidney Redner. *A guide to first-passage processes*. Cambridge University Press, 2001. [37](#)
- [79] Benjamin Reeve, Thomas Hargest, Charlie Gilbert, and Tom Ellis. Predicting translation initiation rates for designing synthetic biology. *Frontiers in bioengineering and biotechnology*, 2:1, 2014. [88](#)
- [80] Sergi Regot, Javier Macia, Núria Conde, Kentaro Furukawa, Jimmy Kjellén, Tom Peeters, Stefan Hohmann, Eulàlia de Nadal, Francesc Posas, and Ricard Solé. Distributed biological computation with multicellular engineered networks. *Nature*, 469(7329):207, 2011. [23](#)

BIBLIOGRAPHY

- [81] Jeffrey W. Roberts. Syntheses that stay together. *Science*, 328(5977):436–437, 2010. ISSN 00368075. doi: 10.1126/science.1189971. [71](#)
- [82] Hugo G Schmidt, Sven Sewitz, Steven S Andrews, and Karen Lipkow. An Integrated Model of Transcription Factor Diffusion Shows the Importance of Intersegmental Transfer and Quaternary Protein Structure for Target Site Finding. *PLoS ONE*, 9(10):e108575, 10 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0108575. [64](#), [92](#)
- [83] Benno Schwikowski, Peter Uetz, and Stanley Fields. A network of protein–protein interactions in yeast. *Nature biotechnology*, 18(12):1257–1261, 2000. [18](#)
- [84] Benjamin Shemer, Noa Palevsky, Sharon Yagur-Kroll, and Shimshon Belkin. Genetically engineered microorganisms for the detection of explosives’ residues. *Frontiers in microbiology*, 6:1175, 2015. [9](#)
- [85] Steven Stayrook, Peera Jaru-Ampornpan, Jenny Ni, Ann Hochschild, and Mitchell Lewis. Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. *Nature*, 452(7190):1022–1025, 2008. ISSN 14764687. doi: 10.1038/nature06831. [72](#)
- [86] Ruud Stoof and Ángel Goñi-Moreno. Modelling co-translational dimerization for programmable nonlinearity in synthetic biology. *Journal of the Royal Society Interface*, 17(172):20200561, 2020. [4](#), [68](#)
- [87] Ruud Stoof, Alexander Wood, and Ángel Goñi-Moreno. A Model for the Spatiotemporal Design of Gene Regulatory Circuits. *ACS Synthetic Biology*,

- 8(9):2007–2016, 9 2019. ISSN 2161-5063. doi: 10.1021/acssynbio.9b00022.
[4](#), [22](#), [86](#)
- [88] Ruud Stoof, Lewis Grozinger, Huseyin Tas, and Ángel Goñi-Moreno. FlowScatt: enabling volume-independent flow cytometry data by decoupling fluorescence from scattering. *bioRxiv*, 2020. doi: 10.1101/2020.07.23.217869. [5](#)
- [89] Ivan V Surovtsev and Christine Jacobs-Wagner. Subcellular organization: a critical feature of bacterial cell replication. *Cell*, 172(6):1271–1293, 2018. doi: 10.1016/j.cell.2018.01.014. [23](#)
- [90] Koichi Takahashi, S. Tanase-Nicola, and P. R. ten Wolde. Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proceedings of the National Academy of Sciences*, 107(6):2473–2478, 2 2010. ISSN 0027-8424. doi: 10.1073/pnas.0906885107. [64](#)
- [91] Cheemeng Tan, Philippe Marguet, and Lingchong You. Emergent bistability by a growth-modulating positive feedback circuit. *Nature chemical biology*, 5(11):842, 2009. [84](#)
- [92] Yuichi Taniguchi, Paul J Choi, Gene-Wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X Sunney Xie. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, 2010. [56](#)
- [93] Huseyin Tas, Lewis Grozinger, Ruud Stoof, Victor de Lorenzo, and Ángel Goñi-Moreno. Contextual dependencies expand the re-usability of genetic inverters. *Nature communications*, 12(1):1–9, 2021. [5](#)

BIBLIOGRAPHY

- [94] Tamir Tuller, Asaf Carmi, Kalin Vestsigian, Sivan Navon, Yuval Dorfan, John Zaborske, Tao Pan, Orna Dahan, Itay Furman, and Yitzhak Pilpel. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, 141(2):344–354, 2010. ISSN 00928674. doi: 10.1016/j.cell.2010.03.031. [70](#)
- [95] Michèle Valens, Stéphanie Penaud, Michèle Rossignol, François Cornet, and Frédéric Boccard. Macrodome organization of the Escherichia coli chromosome. *EMBO Journal*, 23(21):4330–4341, 10 2004. ISSN 02614189. doi: 10.1038/sj.emboj.7600434. [9](#)
- [96] Jeroen S. van Zon, Marco J Morelli, Sorin Tănase-Nicola, and Pieter Rein ten Wolde. Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophysical journal*, 91(12):4350–4367, 2006. ISSN 00063495. doi: 10.1529/biophysj.106.086157. [37](#), [42](#)
- [97] Ulla Vogel and K F Jensen. The RNA chain elongation rate in Escherichia coli depends on the growth rate. *Journal of Bacteriology*, 176(10):2807–2813, 1994. ISSN 0021-9193. doi: 10.1128/JB.176.10.2807-2813.1994. [74](#)
- [98] Haifeng Wang, Marie La Russa, and Lei S Qi. CRISPR/Cas9 in genome editing and beyond. *Annual review of biochemistry*, 85:227–264, 2016. [65](#)
- [99] Xindan Wang, Paula Montero Llopis, and David Z Rudner. Organization and segregation of bacterial chromosomes. *Nature Reviews Genetics*, 14(3):191–203, 3 2013. ISSN 1471-0056. doi: 10.1038/nrg3375. [9](#)
- [100] P B Warren and Pieter Rein Ten Wolde. Statistical analysis of the spatial distribution of operons in the transcriptional regulation network of Es-

- cherichia coli. *Journal of Molecular Biology*, 342(5):1379–1390, 2004. ISSN 00222836. doi: 10.1016/j.jmb.2004.07.074. [9](#), [19](#), [23](#)
- [101] Leandro Watanabe, Tramy Nguyen, Michael Zhang, Zach Zundel, Zhen Zhang, Curtis Madsen, Nicholas Roehner, and Chris Myers. iBioSim 3: a tool for model-based genetic circuit design. *ACS synthetic biology*, 8(7): 1560–1563, 2018. [87](#)
- [102] Martin Welch, Jens Christian Brasen, Christopher T. Workman, and Thomas Sams. Origin of cooperativity in the activation of dimeric transcription factors. *Physical Review Research*, 2(1):1–10, 2020. doi: 10.1103/physrevresearch.2.013151. [72](#)
- [103] Robert B. Winter and Peter H. von Hippel. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 2. The Escherichia coli Repressor-Operator Interaction: Equilibrium Measurements. *Biochemistry*, 20(24): 6948–6960, 11 1981. ISSN 15204995. doi: 10.1021/bi00527a029. [35](#), [45](#)
- [104] Robert B. Winter, Otto G Berg, and Peter H. von Hippel. Diffusion-Driven Mechanics of Protein Translocation on Nucleic-Acids .3. the Escherichia-coli-Lac Repressor-Operator Interaction - Kinetic Measurements and Conclusions. *Biochemistry*, 20(24):6961–6977, 1981. ISSN 0006-2960. doi: 10.1021/bi00527a030. [21](#), [35](#), [36](#), [45](#)
- [105] Matthew A Wright, Peter Kharchenko, George M Church, and D. Segre. Chromosomal periodicity of evolutionarily conserved gene pairs. *Proceedings of the National Academy of Sciences*, 104(25):10559–10564, 6 2007. ISSN 0027-8424. doi: 10.1073/pnas.0610776104. [9](#), [23](#)

BIBLIOGRAPHY

- [106] Zeba Wunderlich and Leonid A Mirny. Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Research*, 36(11):3570–3578, 6 2008. ISSN 03051048. doi: 10.1093/nar/gkn173. [21](#), [35](#), [37](#), [38](#), [41](#), [42](#), [47](#), [65](#), [95](#)
- [107] R. Young and H. Bremer. Polypeptide chain elongation rate in *Escherichia coli* B/r as a function of growth rate. *Biochemical Journal*, 160(2):185–194, 1976. ISSN 02646021. doi: 10.1042/bj1600185. [74](#)
- [108] Chenghang Zong, Lok-hang So, Leonardo A Sepúlveda, Samuel O Skinner, and Ido Golding. Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. *Molecular systems biology*, 6(1):440, 2010. [56](#)