Modelling and analysis
of a genetic oscillator in *E.coli*

Ulrich Janus

A thesis submitted to the University of Warwick for the
degree of Doctor of Philosophy

Mathematics Department, University of Warwick

November 2008

THE UNIVERSITY OF
WARWICK

Contents

List of Figures

3

5

List of Tables

Acknowledgements

Throughout my PhD, I benefited from advice from many people in the mathematics and biology departments, in particular Luca, Martin, Mirela and Philipp in the Mathematics, and Mark, Anja and Ale in Biology, and of course from my supervisors Isabelle Carré and Markus Kirkilionis. I am especially grateful to Isabelle and Luca for the many helpful comments on the thesis draft.

This last line is for Uli for her help and for enduring my prolonged stays on planet research.

Declaration

I, Ulrich Janus, declare that this thesis titled, 'modelling and analysis of a genetic oscillator in *E.coli*' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

In particular, the experimental analysis of a synthetic clock in *E.Coli* was based on the strains provided by the Ninfa laboratory, Department of Biological Chemistry, University of Michigan, School of Medicine, Ann Arbor, Michigan 48109, USA. The experiments were performed with support from, and in the laboratory of Isabelle Carre, Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom.

Signed:
_____

Date:
_____

Abstract

This thesis presents the modelling and analysis of an engineered genetic oscillator in *E.coli*. Genetic oscillators composed of transcriptional feedback loops are the central components of circadian clocks [16]. Thus understanding small genetic oscillators is key for understanding the complex regulatory networks of circadian clocks. In order to monitor clock function, a new colony based imaging assay was set up, based on luminescent transcriptional reporter constructs, that allows for automated data collection over long time spans and for the screening of clock mutants. Clock runs produced damped oscillatory behaviour after starting the clock by removal of the *lac* inducer IPTG or by giving a metabolic stimulus by transferring cells onto fresh agar plates. A detailed mathematical model of the clock was constructed, taking into account discrete and stochastic regulatory binding events at the promoter sites. From this model, using the theory of heterogeneous systems [69, 66], deterministic equations were derived and analysed to yield conditions for the occurrence of stable oscillations based on the system's nullclines. To facilitate the modelling, an algorithm was devised and implemented, that allows for automated construction of Markov chain models of gene activity states based on DNA binding events. In sum, the work constitutes the establishment and analysis of an integrated experimental and modelling system, which opens possibilities for further investigation in order to yield insight into the properties of genetic oscillators.

Introduction

Organisms keep track of time by means of biological clocks, which allows them for example to adjust to the circadian rhythms of light and darkness. Biological clocks are able to incorporate environmental signals to adjust their phase and are robust to noise. At the heart of these clocks lie transcriptional feedback loops, that produce rhythmic patterns of gene expression. Thus understanding genetic oscillators is key for deciphering the complex circadian regulatory networks.

Studying genetic oscillators of circadian clocks is typically complicated by the lack of knowledge of the regulatory interactions, which limits the applicability of detailed mathematical modelling. Recently it has become possible to construct small artificial gene networks from well understood natural components, that are capable of oscillatory behaviour. So far these constructs have mostly been a proof of concept, and their potential for mathematical treatment has not been fully exploited. What is currently lacking is an integrated experimental and modelling system, which can be readily monitored and manipulated in order to study the generation of genetic oscillations in detail.

In order to address this problem, this thesis investigated and studied the properties of a small, artificially engineered genetic oscillator, built from components of the bacterial *lac* and nitrogen systems, that was previously described in [8]. A new imaging assay was set up, based on luminescent transcriptional reporter constructs, to allow for automated monitoring of clock function over long time spans and colony based mutant screening.

The genetic regulation within the engineered circuit was modelled using the theoretical framework of heterogeneous systems [69, 66], which are characterized by the interaction of qualitatively different variables taking finite and infinite values. Building on discrete stochastic Markov chain (MC) models, the theory allows to

derive deterministic rates laws in terms of ordinary differential equations (ODEs). This theory was applied to the clock circuit, constructing a MC model taking into account the molecular details of gene regulation in form of binding of regulatory factors to promoter associated binding sites. The corresponding deterministic rate laws were derived and analysed using standard stability and bifurcation theory for systems of ODEs.

Monitoring the clock function with a reporter construct for one of the two genes in the clock circuit yielded damped oscillatory promoter activity patterns. Oscillations were observed after starting the clock by removal of the *lac* inducer IPTG, and after the application of a metabolic stimulus. Oscillations showed periods of 10-15h, which is in agreement with earlier measurements based on chemostat cultures [8]. In order to facilitate the generation of Markov chain models for genetic regulatory mechanisms, an auxiliary algorithm was devised and implemented, that automatically constructs a Markov chain description of gene activity states, based on the provided information on binding site interactions, including cooperative binding and DNA loop formation. The derived ODE model of the clock circuit was analysed, which yielded conditions for the occurrence of Hopf bifurcations and stable oscillations in terms of the slopes of the system's nullclines. In particular it was shown, that cooperative autoactivation, protein degradation rates and the transcriptional activation threshold for the respective promoters play an important role for the generation of oscillations.

In sum, the work establishes an analytical modelling framework, that is supported by an adjustable experimental system, designed to investigate the molecular basis of genetic oscillations and its properties. The developed algorithm for Markov chain model construction can be regarded as a step towards a more integrated modelling tool to facilitate the detailed modelling of genetic regulatory systems. The derived conditions for oscillations for the analysed clock model correspond well to findings in other studies on genetic oscillators [33]. This suggests, that conclusions from the current and further work on this system should be relevant for understanding the properties of genetic oscillations and thus biological clocks in general.

This thesis is structured into a mathematical (chapters two and three) and an experimental part (chapter four). An introductory chapter (chapter one) presents background and context on modelling and experimental approaches to circadian

clocks, genetic oscillators and genetic regulatory networks in general. It introduces the engineered clock circuit and reviews the mathematical theory of heterogeneous systems, which is used for the modelling. Chapter two reviews the theoretical model framework in more detail, presents the developed algorithm for automated Markov chain construction and also a comparative discussion of models for DNA loop formation mediated by the *lac* repressor. Chapter three contains the derivation and analysis of the model of the clock circuit, where conditions for the occurrence of oscillations are discussed. The fourth and final chapter presents the details of the established imaging assay, and discusses the data collected from the clock runs.

Some additional thoughts on the motivation of the model approach used in this thesis. Finding an appropriate model and fitting the corresponding parameters for a genetic regulatory system is challenging, because of the large number of parameters, the nonlinearity and intrinsic stochasticity of the processes and of the corresponding data being typically sparse, noisy and relying on indirect measurement of gene activity. The problem of parameter estimation has been addressed for example in [35], where Markov chain Monte Carlo (MCMC) based methods using likelihood functions and Bayesian inference have been applied to networks modelled by stochastic differential equations (SDEs). By similar methods applied to simple models of ordinary differential equations (ODEs), the problem of reconstructing transcriptional dynamics from gene expression data based on reporter constructs, like the ones used in this thesis, has been addressed in [24].

In this thesis, no systematic fitting of the parameters was performed. Rather the model parameters are the result of an upscaling process, in which the nonlinearities of the model equations are derived from stochastic models of the molecular interactions, for which parameters were taken from the literature. The model of the genetic circuit was formulated in terms of ODEs, because the corresponding data stems from colony based measurements, i.e. the time-series represents an average over a population of cells. A model formulations in terms of SDEs would be more appropriate for single cell modelling. Further no systematic reconstruction of the transcriptional activity from the reporter data was performed, because the dynamics of the mRNA were neglected in the derived model. This simplification allowed to reduce the dimension of the ODE model to two, making it more accessible for analytical treatment.

In sum, the benefits of the approach used here include the ability to rigorously

derive the nonlinearities in the laws describing gene regulation by upscaling from discrete stochastic models of molecular interactions. Also, the upscaling process is of a modular nature, which allows its application to more complex systems of genetic regulation.

Chapter 1

Genetic networks: modelling approaches and biological context

This chapter provides background on the theoretical context and modelling techniques, with a special focus on the correspondences between discrete stochastic and continuous deterministic model formulations. It also provides background of circadian clocks and experimental approaches, that are employed to investigate them, including the use of artificially engineered circuits to understand transcriptional oscillators.

1.1 Modelling gene networks

1.1.1 General network theory

In many fields of science there has recently been a shift in interest away from studying objects in isolation, and towards investigating interaction and interdependencies among them. This development has led to increased interest in network based model formulations [11]. Biological examples for this include food webs, like predator-prey interactions, networks describing protein interactions, metabolic reaction systems and genetic regulatory pathways. Examples from other areas include information flow in human social networks or the internet, and the complex interdependent dynamics that produce the world climate. This thesis is part of the research initiative

'Unifying networks for science and society' (UniNet, 2005-2008), which has been concerned with reviewing network approaches across scientific disciplines and foster the exchange of network concepts [41].

It has emerged that many complex systems can be seen from a network perspective and that it is often insightful to do so. Improving the robustness of the internet's architecture, the flow of communication in social networks or studying the stability of ecosystems are examples for this. Many sciences have embraced the network perspective recently, while it had long been natural for others - like research on the interconnected functionality of neurons. But in the field of molecular and cellular biology, the rise of the importance of structured interactions has been perceived as a paradigm shift [88].

In molecular biology research the focus has been traditionally on identifying and describing the functionality of a certain enzyme, a piece of genetic code or a part of a cellular organelle. The advances in experimental data acquisition techniques have created a vast and confusing amount of information. So as the functionality of the smallest parts became clearer, the attention shifted on how these parts work together to create the living cell. Networks have thus emerged as a natural way to describe interactions between biological entities. Still, like other abstract or mathematical objects, such as functions or matrices, networks can be applied in many ways. Consequently, a number of different network concepts have arisen in molecular biology.

Networks differ in the nature of the data that they carry and whether their structures are fixed or evolve. Protein interaction networks carry information about which proteins a certain other protein is known to interact with. Weights may indicate the certainty or strength of these interactions. Similar networks exist for genetic regulation, where the network illustrates whether the product of gene A will stimulate or inhibit the expression of a gene B. The information in these networks lies mainly in their structure - the number and connectivity of their nodes, also called the network topology. The network topology in these cases is typically fixed, but is sometimes allowed to change - new species or interactions appear or vanish - to describe selection dynamics on an evolutionary time scale.

In areas such as metabolic reactions, genetic regulation or signaling pathways, there is more data available on the dynamics of these interactions. These dynamics

are typically modelled using kinetic rate laws like mass-action, Michaelis-Menten or Hill-type laws, the latter of which are used to describe cooperative reactions. Network models of these dynamical systems illustrate the interdependencies and interactions of its components. Values defined on nodes and edges may be concentrations and reaction rates, and may change over time. Because of this additional complexity, these networks typically are smaller in size than static interaction networks, to make them tractable for mathematical analysis.

Given the heterogeneity of network models, it is helpful to distinguish four classes, based on the variable types that are attached to the components of the network. These classes are characterized by whether the corresponding networks have static or evolving topologies, and by whether the variables attached to the nodes and links - also called weights - are static or dynamic. Protein interaction networks belong to class I - static networks with static weights. When the evolutionary change of protein interaction is considered, the resulting network model is of class II - evolving networks with static weights. Genetic regulation, signaling and metabolic reaction networks typically belong to class III - static networks with dynamics weights. Models of class IV - evolving networks with dynamic weights - may be used to incorporate dynamics of shorter time-scales on the weights, with slow processes like evolutionary changes of the network structure.

For genetic regulatory networks the situation is complicated by the involvement of qualitatively different players, which are the candidates for network nodes. For one there are the genes, gene modules or operons depending on the specific model emphasis. On the other hand there are the products of gene expression - RNA transcripts and proteins. For genes, it is their state of activity - whether they are transcribed or not, that is of interest. For the expression products, their amount in terms of particle numbers or concentrations is typically what matters. In other words, genes may be described by variables of finite states, gene products by variables of infinite copy numbers. Consequently, there are different types of interactions that may link these nodes. Gene products may interact with each other in a biochemical reaction. Genes are linked to their gene products, which, in turn, may be regulatory factors that influence the activity states of the genes.

The last part of this section reviews work that has been done to describe these hybrid-type networks in more detail and make them accessible to mathematical

treatment. But first, we review general approaches to model genetic networks.

1.1.2 Approaches to model genetic networks

The environment, in which the evolution of modelling techniques for genetic networks progresses, is characterized by the limitations of analytical and computational feasibility on one hand, and the quality of the experimental data on the other. Quality of experimental data may be discussed in terms of the attainable spatial and temporal scales, ease of reproduction by automatization and the grade of quantification and comparability between experiments. Traditionally biological experiments focused on qualitative data to support or rebut descriptive models of biological mechanisms. The process of exploiting the possibilities of the current experimental techniques to increase the quantifiability of data is still ongoing.

The following review on modelling approaches to genetic networks proceeds in a somewhat chronological order. This does not imply that older model techniques are obsolete because they neglect or oversimplify a certain process. Model design is always guided by the problem at hand, the nature of the underlying data, and the urge to keep it simple while avoiding oversimplification. Recent advances in the collection of gene expression data have allowed for more detailed models taking into account even discrete molecular events. Still, all discussed approaches maintain their validity within the appropriate context. For reviews on modelling genetic networks see [34] and [76].

An early approach by Glass, Kaufmann and Thomas [28], [79] uses a Boolean formulation to describe genetic control circuits. There, genes at any time points are either in an 'on' or 'off' state. A gene in an 'on' state may change the activity state of another gene in the next time step. This method appeals by its simplicity and applicability to large networks. Dynamic activity patterns like oscillations and switches and the role of feedback loops in these dynamics were studied with these models [80], [81]. The Boolean network approach has been generalized to allow genes to reside in a number of finite states in order to describe different levels of gene activity. This approach was successfully applied to model gene expression in plants [23], and was used for the reverse engineering of gene regulatory networks from time-series data [43]. The Boolean and related approaches are limited when it is important to capture gradual dynamics of gene regulation. These effects are easier

described by dynamical models, that are based on ordinary differential equations (ODEs).

In these continuous time models the evolution of the variables, the concentrations of gene products, like mRNA transcripts and proteins, are described by systems of coupled differential equations. The variables take values in the real numbers and change gradually over time. Production and degradation of gene products and the influence of regulatory factors on the gene expression are typically modeled by mass-action, Michaelis-Menten or Hill-type equations, which are traditionally used to describe biochemical and enzymatic reactions [71]. This approach is now widely used, for example to study the regulation of circadian clocks [45], [47], or the progression through the cell cycle [82]. The shortcomings of this approach are for one, that the employed reaction kinetics are typically based on heuristics, for example by assuming a 'gene population', of which only a fraction is transcriptionally active at any given time [84]. Secondly, many regulatory factors are present at very low amounts, which limits the feasibility of treating them in terms of deterministically evolving concentrations. Therefore some effort has been dedicated to take the discrete and stochastic elements of gene regulation into account.

Ways to accommodate the stochasticity of molecular interaction when only a few particles are around include the addition of ad-hoc noise terms to differential equations or the decomposition of the deterministic model into discrete stochastic events. The latter approach for example was used to describe different pathway decisions of the cell governed by the noise in gene expression [3]. The complete unfolding of reactions into stochastic equations, and their simulation by the Gillespie algorithm, requires elaborate computations. Also these systems are more difficult to treat analytically than deterministic rate equations are. Motivated by this, recent work has been conducted to develop a theory for connecting the discrete stochastic description of these systems to continuous and deterministic equations. This is reviewed next.

1.1.3   From stochastic to deterministic models of gene expression

In a recent paper by Kepler and Elston [40], the authors study a stochastic model of a gene that is activated by the binding of a transcription factor to a corresponding binding site. When the transcription factor is bound, the gene is expressed at a

higher 'on' rate, if not, then the gene is expressed at a lower 'off' rate. The system is formulated in terms of a Master equation [83, 25], which is a differential equation in the probabilities for the genes to be either in an 'on' or 'off' state. This treatment of genetic regulatory interactions was generalized by Sbano and Kirkilionis in a number of papers ([66, 69, 67, 68]), as discussed below.

In [66], the authors state their perspective of the modelling problem: Consider a number of macro-molecules - called molecular machines, and a number of species of smaller molecules - called communicating molecules. The number of molecular machines is fixed over time, but the molecules may reside in any one of a finite number of discrete states. The communicating molecules on the other hand are defined by their number. These smaller molecules may also influence the transitions of the macro-molecules between states. What we end up with is a connection of two interacting Markov chains (MCs): one that defines the transitions of the macro-molecules between finitely many states, and another, that describes a birth-death process of the communicating molecules. Apart from genetic regulation, other examples for this type of system include the operation of membrane channels or other enzymatic processes, where the enzyme reside in different conformational states.

These models of molecular dynamics are also called systems with finite states and and infinite copy numbers [66], to stress the fact that there are two kinds of variables - those which allow for a continuum approximation, and others which are inherently discrete. More precisely, the infinite copy number may be treated by a continuum limit, whereas the finite states remain discrete, but play a role in the time-scaling argument. In [66] the Master equation (ME) of these systems is studied by deriving a continuum limit and an adiabatic approximation. The continuum limit transforms the infinite copy number - number of particles - into a concentration and transforms the ME into a Fokker-Planck equation (FPE) (for theoretical background on the ME and FPE see [83, 25]). The adiabatic approximation is based on the assumption that the dynamics of the finite states - the states of the macro-molecules - evolve very fast, so that the FPE can be approximated by an expansion in the time-scaling parameter. The leading order term of that expansion corresponds to a differential equation in the concentrations of the communicating molecules and is called the average dynamics.

To summarize, we have one Markov chain that describes the evolution of the

finite states of the macro-molecules, and another one, that describes the birth-death process of the communicating molecules. The continuum limit for the communicating molecules turns the ME into a FPE. An adiabatic approximation of the FPE can be constructed around the invariant measure of the fast evolving MC of the macro-molecules. This yields the average dynamics, which are given by a system of ODEs in the concentrations of the communicating molecules, and whose right hand site is given by a vector field, which stems from the birth-death process, and is averaged against the invariant measure. This model frame is used in [69] for an alternative derivation of kinetic rate laws like Michaelis-Menten or Hill-type equations that are typically used to describe gene regulation and enzymatic reactions.

In two further papers [67, 68], the authors elaborate this theory and study how the associated interaction graphs are affected by the limit processes. Of practical importance for this thesis is in particular an investigation on ways to modularize the MC on the finite states. In typical cases the state space of this MC has the structure of a Cartesian product, and the infinitesimal generator of the MC can be written as a tensor product of smaller matrices. It turns out that the tensor product of the invariant measures of these smaller matrices also solves the steady state condition of the original MC. This can be applied to decompose the MC for the macro-molecules into independent submodules.

We conclude by recalling how this theoretical description is applied to genetic regulation (see figure 1.1). The macro-molecules are the genes or, more precisely, the promoter regions that govern their expression (we will simply talk about 'genes' from here on). The communicating molecules are the gene products, which may be regulatory factors that activate or inhibit the expression of the genes. The feeding back of the regulatory factors onto the activity states of the genes occurs by binding events to enhancer and repressor sites. The state space of a gene may be further complicated by the appearance of regulatory DNA loops.

Summary: Modelling gene networks

In the first section it was described how network models can be classified depending on the type of variables - static or dynamic - that are defined an their nodes and links, and the type of network structure - constant or evolving. Examples were given for popular networks used in molecular biology, like static protein interac-
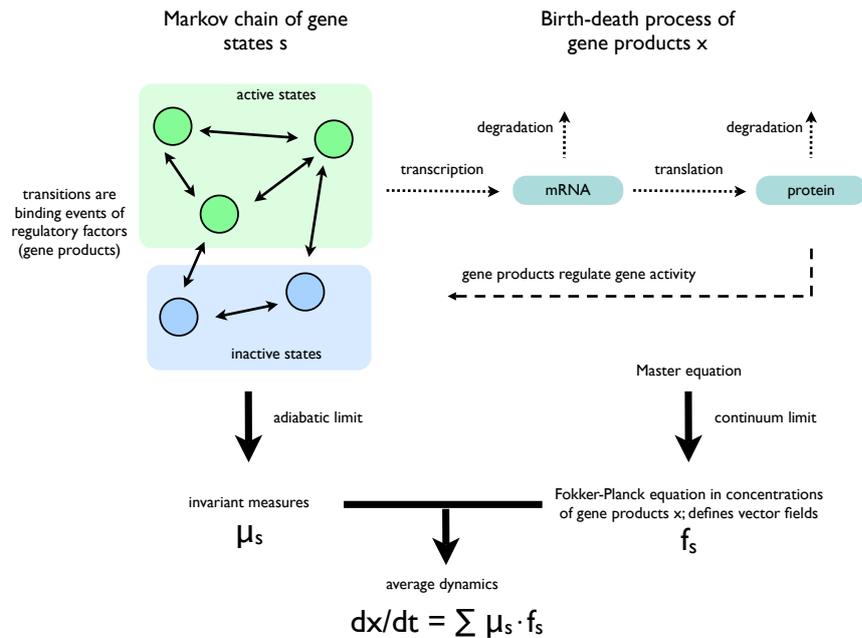
Figure 1.1: From stochastic to deterministic models of gene regulation. The genes reside in transcriptionally active and inactive states s. The transitions between these states are governed by binding events of the gene products (regulatory factors) to the DNA, and are described by a Markov chain (MC). The number of gene products x evolves as a birth-death process defined for each of the gene states. The evolution of the two interacting MCs is described by a Master equation (ME), which is transformed into a Fokker-Planck equation (FPE) by taking a continuum limit, which transforms particle numbers into concentrations of the gene products, also denoted by x. An adiabatic approximation of the FPE can be constructed around the invariant measure μ of the MC of the gene states. This yields a system of ordinary differential equations (ODEs) in the concentration of the gene products, whose right hand side is given by a vector field f_s for each state s averaged over the invariant measure. This ODE system is also called the average dynamics.

tion networks, or networks of dynamical systems like coupled metabolic reaction systems. Confusion may arise when variables of different types are associated to nodes, like discrete variables for gene activity states and continuous variables for the corresponding gene product concentrations.

The second section presented a review on typical model approaches that have been taken to analyse genetic regulatory systems. Historically there has been a development from discrete to continuous deterministic models, and from there to continuous and discrete stochastic models. This development was accompanied by the increased availability of gene expression data and the increase of computational efficiency. Even though it is generally accepted that discrete stochastic events are important in genetic regulation, the complexity of discrete stochastic computation has motivated work to connect this level of modelling with the easier-to-analyse deterministic models - through deriving deterministic equations as appropriate limit processes of discrete and stochastic formulations.

The third section introduced the theory of systems of finite states and infinite copy numbers. This theory is motivated by model situations that are typical for genetic regulation and other molecular processes, where few macro-molecules interact with a larger number of smaller molecules. The number of macro-molecules is fixed, but each macro-molecule is able to switch among a finite number of states described by a Markov chain. The change of the number of smaller molecules is described by a birth-death process for each of the states of the macro-molecules. From these two interconnected Markov chains, average dynamics - a system of coupled ODEs in the concentration of the smaller molecules - can be derived by assuming that the MC of the macro-molecules evolves fast (adiabatic limit), and that the number of smaller molecules can be approximated by concentrations (continuum limit).

In the context of gene regulation, the macro-molecules are the genes. Their states are defined by whether repressors and transcription factors are bound to corresponding regulatory binding sites. Depending on their state, genes may be transcriptionally active or inactive. The smaller molecules are the gene products, which may act on the MC of the macro-molecules as the regulatory factors.

1.2   Oscillating gene networks or "biological clocks"

1.2.1   Circadian clocks

Organisms keep track of time by means of biological clocks. These time keeping mechanisms enable for example mammals to regulate their sleep-wake rhythms, or plants to coordinate leaf movements with the day-night cycle. This section summarizes the general properties of circadian clocks and their molecular basis. The current model of the clock of *Drosophila melanogaster* is given as an example. For recent reviews of the field of circadian clocks, see [20, 16, 90].

Emerging from first studies on *Drosophila melanogaster* circadian clocks have been discovered in a number of other organisms like mammals (mouse, human), plants (*Arabidopsis thaliana*), fungi (*Neurospora crassa*) and cyanobacteria (*Synechococcus elongatus*) - the corresponding model organisms are given inside the parentheses. Cyanobacteria are the only group of bacteria where evidence for a functioning clock has been found. So it seems that clocks are mainly a feature of eukaryotes.

From the studies of the different model organisms a number of paradigms have emerged. It seems that clocks are typically realised within a single cell and do not need communication between cells to function. In higher organisms there can be certain clusters of cells which drive rhythms in other tissues or in the whole organism. But rather than having an exclusive cluster of pacemaking cells, clocks seem to expressed in different tissues if not in all cells of the organism.

Circadian clocks are often described according to a threefold structure. Input pathways sense environmental signals like light and temperature and feed it into the central oscillator. The central oscillator is responsible to produce a stable rhythm and incorporates the environmental signals to adjust its phase. From the central oscillator emanate output pathways, which drive the rhythmic expression of other genes. The observed rhythms are robust in the face of genetic noise, and able to compensate for fluctuations in the environmental conditions.

The molecular basis of clocks is generally composed of a transcriptional negative feedback loop, where a positive element inhibits its own expression indirectly via a negative element. A delay between the expression of the positive element and the consequent self inhibition leads to an oscillatory behaviour. This delay is thought to arise from the biochemical process of gene expression (transcription, translation),

intermediate steps (forming of complexes, phosphorylation) or, in the case of eukaryotes, transport processes out and back into the nucleus.

Although transcriptional feedback loops seem to be an essential element of circadian clocks, it has been shown that circadian oscillations from cyanobacteria clock components can be produced solely by post-ranslational modifications in vitro [51]. Also it has become apparent that circadian oscillations typically rely not just on one negative feedback loop, but multiple interlocking positive and negative feedback loops [16]. The complexity and apparent redundancy of the circadian regulatory network might be necessary to achieve the overall robustness of clocks.

Example: Molecular basis of the clock in *Drosophila*

The core of the *Drosophila* clock consists of a transcriptional negative feedback loop comprising the genes *per* and *tim* (see figure 1.2). The protein products of these two genes form a complex in the cytosol, which, after a number of phosphorylation events, is transported into the nucleus to inhibit *per* and *tim* transcription. This has inspired mathematical models which were able to produce stable oscillations by incorporating transport, per-tim complex formation and phosphorylation reactions [45].

The expression of the *per* and *tim* genes are dependent on the cyc-clk protein complex, where *cyc* is a gene which is expressed at a constant level over time. The per-tim protein complex inhibits the expression of *per* and *tim* by inactivating the cyc-clk complex. The cyc-clk complex also regulates expression of two other genes, *vri* and *Pdp1e*. A second loop is now formed by vri quickly inhibiting and pdp1E - with a time delay - activating the expression of *clk*, which regulates the amount of cyc-clk complexes. How oscillations can arise from these interlocked loops has been explored by model by Kitano [36], where the authors concluded that the interlocked feedback loops increase the robustness of the oscillations.

1.2.2   Experimental approaches to genetic networks

The regulatory interdependencies among the genes in a cell form a highly complex and dynamic network. The product of any gene may trigger directly or indirectly the activation or repression of one or more other genes. These effects may happen fast or with a time delay, they may be switch-like or more gradual. From this variability
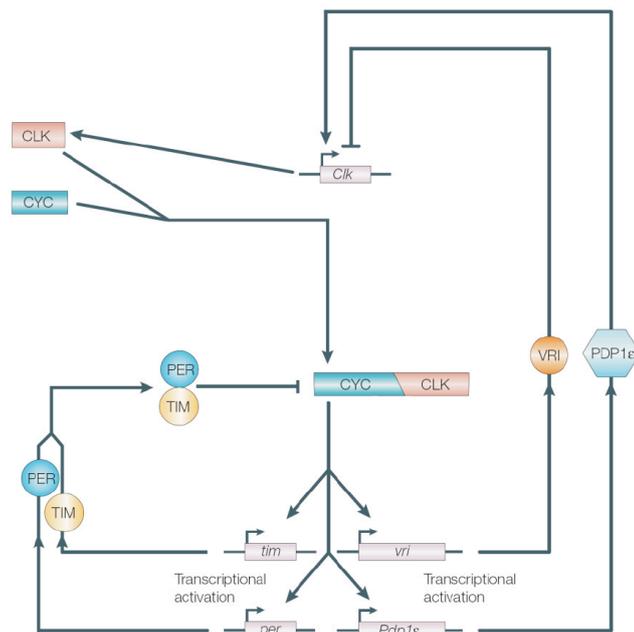
Figure 1.2: Molecular basis of the circadian clock of *Drosophila melanogaster*. Two feedback loops interconnect on the heterodimer cyc-clk. One, cyc-clk activates expression of the *per* and *tim* genes. The per and tim proteins accumulate as heterodimers inhibit the function of the cyc-clk complex. Two, cyc-clk activates the expression of the genes *vri* and *Pdp1ε*. The protein vri accumulates fast and inhibits expression of the *clk* gene and hence the number of cyc-clk complexes. The protein Pdp1ε accumulates more slowly and activates the expression of *clk*. The figure was taken from [16].

of regulatory processes and the sheer size of cellular genomes arises the complexity
of natural gene networks.

Several strategies have been pursued to advance their understanding. One may
take a large-scale genome wide approach, as is characteristic for systems biology. On
a smaller scale one may focus on certain functional submodules of natural systems.
Recently it has also become possible to construct artificial gene circuits from well
understood genes and promoters. We will discuss these approaches in short, with
a special focus on engineered circuits. But first some notes on ways to collect gene
expression data.

Many different ways to measure gene expression have been employed. They
differ in their temporal resolution, the clarity of the signal and whether they can be
automatized or require the manual analysis of samples at certain time intervals.

Recent advances in the technology of microarrays have made available genomic
and proteomic data on a large scale. This data is now being used to elucidate
the dependency relations between genes and to identify certain structural modules
within a gene network [1]. The results of these studies are typically connected
with the static gene interaction networks mentioned earlier, because it is difficult to
obtain precise information on dynamics, although it is possible to obtain time-course
data, as was for example reported in [18]. From these data it is possible to identify
functional groups of genes by Bayesian methods [91]. Unfortunately, microarray
data are typically very noisy.

More accurate dynamical data for the activity of specific genes can be obtained
by mRNA or protein assays, or by reporter based imaging, which has become very
popular because it allows for automatized data acquisition from live organisms.
Imaging techniques are based on reporter proteins like fluorescent proteins (GFP,
YFP, etc.) or bioluminescent proteins of the luciferase family. The promoter region
of the gene of interest can be fused to the gene for the reporter protein, and its
expression can be monitored by fluorescence microscopy or photon-counting cameras.
Cells can be monitored in liquid culture or on a colony basis. There is generally
a payoff between the accuracy of the data and the convenience of its collection.
Therefore mRNA and proteins assays are typically more accurate than reporter
based assays, which in turn are more convenient to conduct.

The data collected for this work stems from bioluminescence assays using lu-

ciferase reporter constructs, where cells were monitored on a colony basis on agar plates [42], see figure 1.3. When the gene of interest is expressed, so is the enzyme luciferase, which reacts with its substrate decanal under light emission. The emitted light is collected by photon-counting camera equipment to produce time-series data of gene expression. This setup allows to automatically collect data over long time periods. Also, it allows to distinguish gene expression patterns of individual colonies, which can be used to identify interesting mutants.
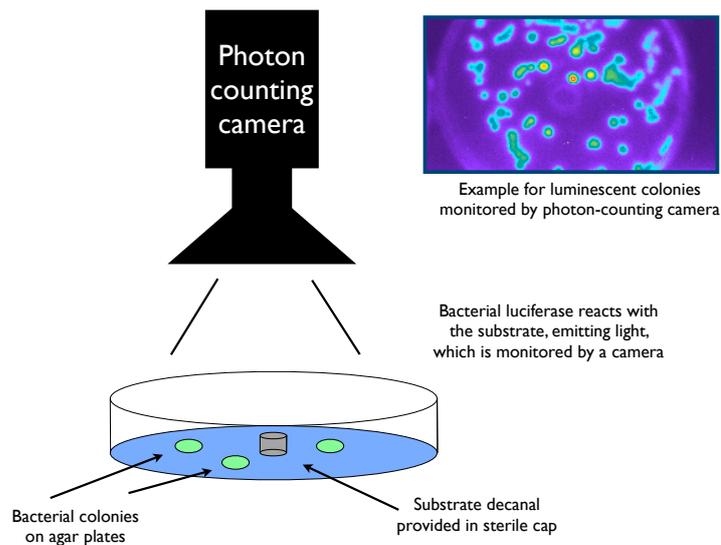
Figure 1.3: Schematic of the bioluminescence assay based on luciferase reporter constructs used for this thesis. Cells carrying luciferase reporter constructs are grown into colonies on agar plates. The substrate decanal for luciferase, which is provided in sterile caps on the agar surface, evaporates and is taken up by the cells, where it reacts with luciferase under the emission of light. The emitted light is monitored by a photon-counting camera. An example of an image of luminescent colonies is given in the upper right corner.

Research on gene regulation traditionally focuses on functional modules of natural systems like regulation of circadian clocks or the cell cycle. The rationale for this lies in the assumption that these systems are governed by a manageable number of genes and that they can be understood on their own - when the impact from all other genes is neglected. This is a popular area for mathematical models, for example for circadian clocks [45, 47]. The fundamental problem in this context is the incomplete knowledge of the system. Information on which genes are important for the regulation of a given system is typically sparse. This limits the possibilities for mathematical modelling of natural gene networks.

An alternative approach to the study of natural systems is the construction and analysis of synthetic gene circuits from well understood natural components [33]. This procedure yields small and self-contained regulatory networks, whose interactions with the rest of the cell's gene regulation network are tractable. Due to isolation from outside influences, the regulation of individual genes can be understood more accurately than would be possible in natural systems. Such a network enters the scope of detailed, dynamical mathematical modelling in close relation to the experimental data.

A number of synthetic gene circuits have been studied to address questions about regulatory mechanisms and dynamics. By coupling three transcriptional repressors, Elowitz and Leibler were able to construct a gene circuit in *E. coli* capable of oscillations [21]. Another system in *E. coli*, exhibiting a toggle switch behavior, was studied by Gardner [26]. Becskei and Serrano used an engineered circuit, also in *E. coli*, to investigate the role of feedback loops on system stability [14]. The role of noise in gene expression was investigated by manipulating expression dynamics of an gene introduced into *B. subtilis* [57]. In *S. cerevisiae* an eukaryotic gene switch was constructed to study the role of positive feedback in cell differentiation [15].

1.2.3 A synthetic genetic oscillator in *E.coli*

This project is based on a synthetic feedback circuit by Atkinson and coworkers, which was shown to exhibit damped oscillations [8]. The circuit consists of two gene modules, which were constructed using elements from the *lac* operon and from the regulatory system of the bacterial nitrogen metabolism. Expression of the genes in the circuit is driven by the two promoters *glnAp2* and *glnKp*, which are both responsive to the transcription factor NRI, but with different activation thresholds. The *glnAp2* promoter was modified to be repressible by the *lac* repressor. We will first review the biological context and the molecular details of the circuits components, and then discuss how they are combined within the engineered clock circuit to produce oscillatory behaviour.

Biological context of the nitrogen metabolism in *E.Coli*

The promoters *glnAp2* and *glnKp* and the structural gene *glnG* of the transcription factor NRI are components of the regulatory system for nitrogen uptake in *E.coli* and

related bacteria [60]. The promoter *glnAp2* is part of the promoter *glnAp*, which in wild type *E. coli* controls the transcription of *glnA*, the structural gene of glutamine synthetase (GS). The promoter *glnKp* controls the expression of *glnK*, the structural gene for GlnK, which is a signal transduction protein. I give a short overview of this system with specific focus on the regulation of the components used in the synthetic clock (see figure 1.4).
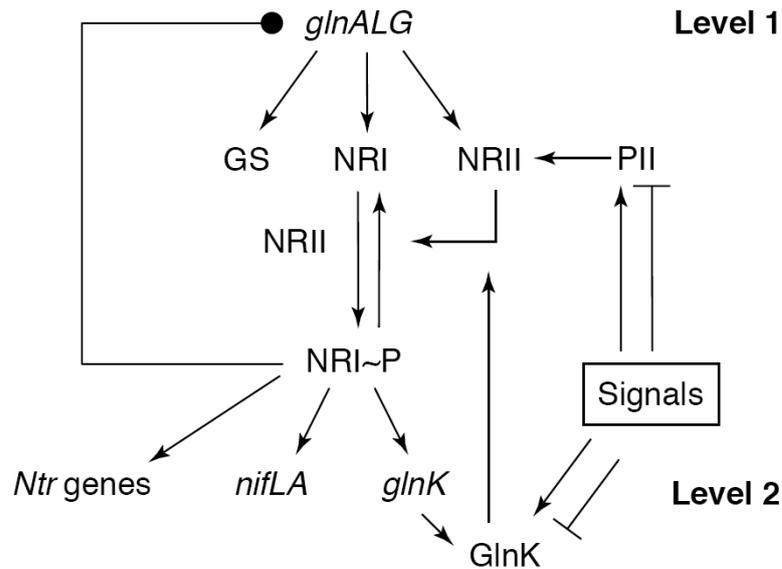
Figure 1.4: Overview of regulatory interactions in the nitrogen metabolism of bacteria. The *glnALG* operon expresses the genes for glutamine synthetase (GS), which is responsible for the nitrogen assimilation, for NRI, which in its phosphorylated form NRI~P is a transcription factor for the operon, and for NRII, which controls the phosphorylation state of NRI. The signal transduction protein PII senses the signals of the carbon and nitrogen levels and regulates the amount of GS by controlling the activity of NRII. It also regulates the activity of GS via the enzyme adenylyltransferase (ATase) (not shown). When the cell is starved for nitrogen, high levels of NRI~P trigger the shift from level 1 to level 2 of nitrogen assimilation. In level 2, the gene *glnK* for a PII like signal transduction protein GlnK is expressed, as well as the *ntr* and *nif genes*, which activates alternative mechanisms for nitrogen assimilation. Pointed arrows denote activation, blunt arrows inhibition. The circle headed arrow denotes activation at low levels and inhibition at very high levels of NRI~P. The boxed 'Signals' denotes the nitrogen and carbon level indicators, glutamine and 2-ketoglutarate (2KG). The figure was adapted from [52].

Bacteria have to regulate the uptake of ammonia, so that the amount of nitrogen inside the cell is in balance with the amount of intracellular carbon. The cell keeps track of the availability of nitrogen and carbon through indicator molecules,

also referred to as nitrogen and carbon signals. Glutamine, into which ammonia is converted, serves as the nitrogen indicator, while 2-ketoglutarate (2KG) servers as the indicator for carbon. These two signals are overlaid to regulate the activity of the enzyme glutamine synthetase (GS), which catalyzes most of the ammonia assimilation. In this process the signal transduction protein PII [52] plays a central role.

Glutamine synthetase is regulated on two levels. On one level, the expression of its structural gene *glnA* is controlled, on the other level the enzymatic activity of GS is regulated by reversible covalent adenylylation by the enzyme adenylyltransferase (ATase). The expression of *glnA* is dependent on the phosphorylated form of the transcription factor NRI, denoted by NRI~P. The phosphorylation state of NRI in turn is controlled by the kinase/phosphatase NRII [58]. In consequence, the amount of GS is regulated by NRII, while the level of its activity is regulated by ATase. Both these enzymes are affected by the protein PII, which is responsible for the interpretation and transduction of the nitrogen and carbon signals.

The PII protein senses the carbon and nitrogen levels through reversible uridy-lylation and the binding of 2KG - see [52] for details. The net effect is that, when nitrogen levels are low compared to carbon levels, PII increases the activity and expression of GS, by regulating ATase and NRII - and vice versa.

When nitrogen levels are very low, the cell reacts with a stress response response by changing its gene expression pattern - again see [52, 60] for details. The high NRI levels under nitrogen starvation in particular switch on transcription of the gene *glnK* [4]. The gene product GlnK has similar functions as PII, but seems to be tuned to replace PII under nitrogen stress [6, 5]. Consequently, its promoter *glnKp* is weak at low NRI levels, but becomes strong after NRI concentration increases above a certain threshold [4, 17].

Mechanism of NRI dependent transcription at the *glnA* and *glnK* promoters

In wild type *E. coli* the expression of the gene *glnA* is driven by two promoters, *glnAp1* and *glnAp2* [61]. The first promoter *glnAp1* is dependent on the catabolite activator protein (CAP) and is repressed by NRI. The second promoter *glnAp2* is activated by the phosphorylated form of NRI, denoted by NRI~P.

The phosphorylation state of NRI is controlled by the kinase/phosphatase NRII [53]. Only the phosphorylated form of NRI is able for form functional oligomers that expose the regulatory domains, that interact with the σ^{54}-RNA polymerase [44], and stimulate the formation of stable 'open' transcription complexes [77]. This interactions is facilitated by two enhancer sites NRI1 and NRI2 upstream of glnAp2 [54]. The NRI~P bound to the enhancer sites is brought in contact with the σ^{54}-RNA polymerase bound at the promoter by the formation of a DNA loop [77].

At high concentrations of NRI~P the activity of the *glnAp2* promoter is again reduced [73]. Responsible for this effect are three low-affinity NRI~P binding sites (governor sites) NRI3, NRI4, NRI5 [7]. These governor sites are only filled at high NRI~P concentrations and are thought to either interfere with the formation of the activation DNA loop or hinder the successful interaction of NRI~P with the σ^{54}-RNA polymerase transcription complex.

The activation of the promoter *glnKp* works similarly to that of *glnAp2*. Here again, there are two enhancer sites for NRI~P, as needed to facilitate its interaction with the σ^{54}-RNA polymerase, one with a high, the other with a low affinity. The major difference apart from the absence of silencing governor sites, is that the *glnKp* promoter needs a higher NRI~P concentration for activation [4]. This difference in sensitivity to the activator protein is reflected in the multi-level response of *E. coli* to varying degrees of nitrogen starvation [4].

Regulation by the *lac* repressor

The *lac* repressor, which is a component of the repressor module of the engineered clock, regulates the *lac* operon in wild type *E. coli*. Hence, to provide some context, this section gives a short overview of the regulation of the *lac* operon. For a recent review on the lac repressor see [46].

In the presence of glucose and lactose, *E.coli* preferably uses the first. When glucose runs low, and after a short stagnation phase, the bacteria will then feed on lactose. The observation of this growth pattern of bacterial cultures - also called 'diauxic' - led to the investigation of what has become the prototype of operon based gene regulation in bacteria, the *lac* operon. Following the first studies by Jacob and Monod in 1961 [38] the regulation of the operon and the function and structure of its components have been described in a great detail, so that the *lac* operon has

become a popular system for mathematical modelling of gene regulation.

The regulation of the *lac* operon can be described as an AND coupled logic unit with two inputs: One is allolactose, which acts as an indicator of the availability of lactose. The other is cAMP, which acts as an indicator of glucose levels. A shortage of glucose leads to a rise in the concentration of cAMP, which activates CAP by binding to it. By binding to an enhancer site close to the *lac* promoter the CAP-cAMP complex then increases the activity of the promoter, which is very low otherwise. Availability of lactose as a food source leads to the production of the inducer allolactose. When allolactose binds to LacI, the repressor's affinity for binding to the operator sites is strongly reduced, so that it no longer inhibits transcription of the operon's structural genes. So the *lac* operon is only efficiently transcribed when both allolactose and cAMP are present (see figure 1.5). This two-fold regulation leads to the diauxic growth behavior described above.
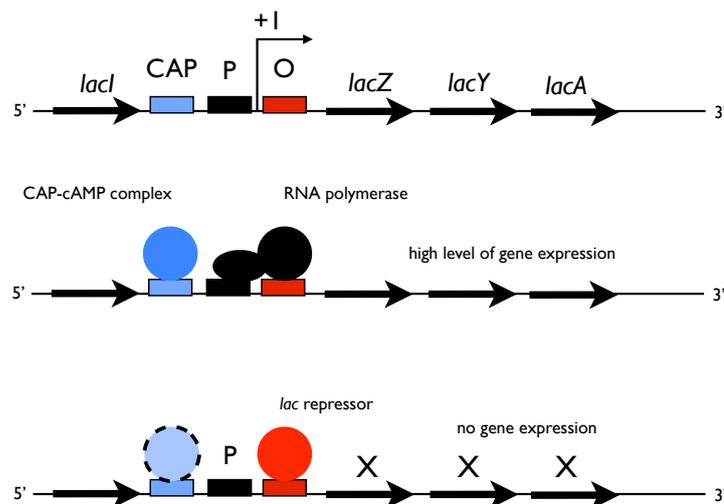
Figure 1.5: Control elements of the *lac* operon. A high level of gene expression only takes place when the *lac* repressor is inactivated due to the presence of the inducer allolactose, and the CAP protein is activated by cAMP. The CAP-cAMP protein complex binds to the CAP site, which enhances the activity of the promoter. The *lac* repressor inhibits transcription by binding to the operator site O, and thereby blocking the binding of RNA polymerase to the promoter site P. So when the *lac* repressor is bound, then no gene expression takes place even when CAP-cAMP is bound as well (indicated with dashed boundary).

The functionality of the *lac* repressor is reflected in its structure. The *lac* repressor is a heterogeneous tetrameric structure, which can be described as a tethered

dimer of dimers (see figure 1.6). The interaction between the two dimers is flexible enough to allow them to adopt different orientations with respect to each other. Each dimer is capable to bind to specific operator sites in the promoter region of the *lac* operon.
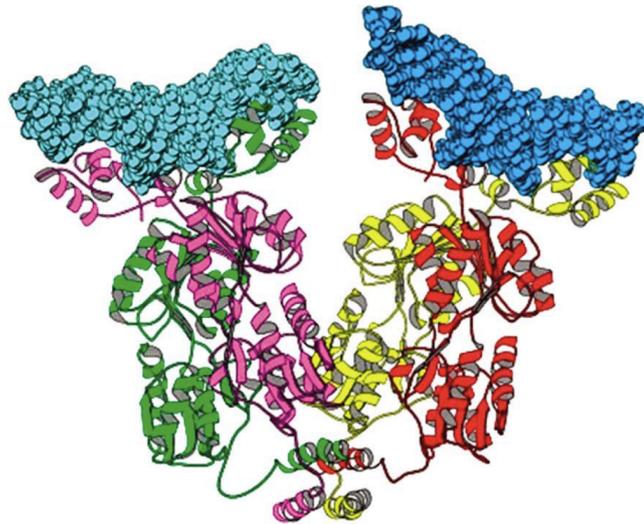
Figure 1.6: A ribbon diagram of the quaternary structure of the *lac* repressor complexed to DNA. Each monomer is drawn in color. The *lac* repressor has the structure of a tethered dimer of dimers. The figure was taken from [46].

When LacI is bound to two operator sites, the DNA forms a loop. In wild type *E. coli* the main operator at position +11 (the position is given in base pairs and relative to the start site of transcription +1) is flanked by two auxiliary sites at positions -82 and +412. When LacI is bound to the main operator it blocks the transcription of the *lac* operon by the polymerase. Removal of one auxiliary operator will lead to a reduced repression, while removal of both reduces the repression about 70-fold, almost completely abolishing repression [56]. There are two possible ways in which the repressor may form the DNA loop. Either a tetrameric repressor binds subsequently to two operator sites with its two DNA binding domains. Alternately dimeric repressor molecules bind independently to two operator sites and form the DNA loop by binding to each other to form a tetrameric repressor.

The effectiveness of the repression is dependent on the binding affinity of LacI to the operator sites and the number, relative positions of these operator sites. The operator sequence within the *lac* operon is not optimized for tight binding [27, 9].

A fully symmetric operator has an binding affinity which is ten times higher than that of the wild type operator [64]. The repression becomes weaker with increasing distance between the operator sites and when distances are small, the exact spacing has a significant effect [50].

The repressor is released from the DNA when an inducer molecule binds to it. The natural inducer is the galactoside allolactose, which is produced in small quantities from lactose by the enzyme β-galactosidase. Another inducer is the chemical compound IPTG, which is usually used for in vitro induction of the *lac* operon - and was also used to regulate the engineered clock. There are also anti-inducers like ONPF, which increase the binding affinity of LacI. All known effector molecules are galactosides and bind to the same effector site of LacI. The effect of some inducers on the binding affinity of the repressor has been quantified and reported in [62, 13].

Design of the gene circuit

The network as described in [8] comprises two modules, an activator and a repressor module, which feed back into each other (see figure 1.7). Each module consists of a gene and a promoter region including certain regulatory binding sites. The modules are constructed from components of the regulatory systems for lactose (*lac* operon) and nitrogen metabolism.

The activator module implements an autoregulatory positive feedback loop. The gene *glnG* was fused to the *glnAp2* promoter sequence. The product of the gene *glnG* is the protein NRI, which is phosphorylated by the kinase NRII. In its phosphorylated form, NRI∼P, it activates the *glnAp2* promoter. Thus NRI activates its own transcription.

A negative feedback loop, is realized via the repressor module, which consists of the *lacI* gene fused to the *glnKp* promoter. Similar to the *glnAp2* promoter, the *glnKp* promoter is activated by NRI∼P binding to two adjacent binding sites. The repressor module regulates the activator module via two *lac* operator sites inserted in the *glnAp2* promoter region. When the *lac* repressor LacI is bound to the promoter proximal operator site, it blocks the transcription of the *glnG* gene.

The *E. coli* strain carrying the modules had mutations in the chromosomal genes *lacI, glnG, glnL* rendering it unable to produce functional LacI, NRI and NRII proteins. So the only source of LacI and NRI was the synthetic clock. In wild
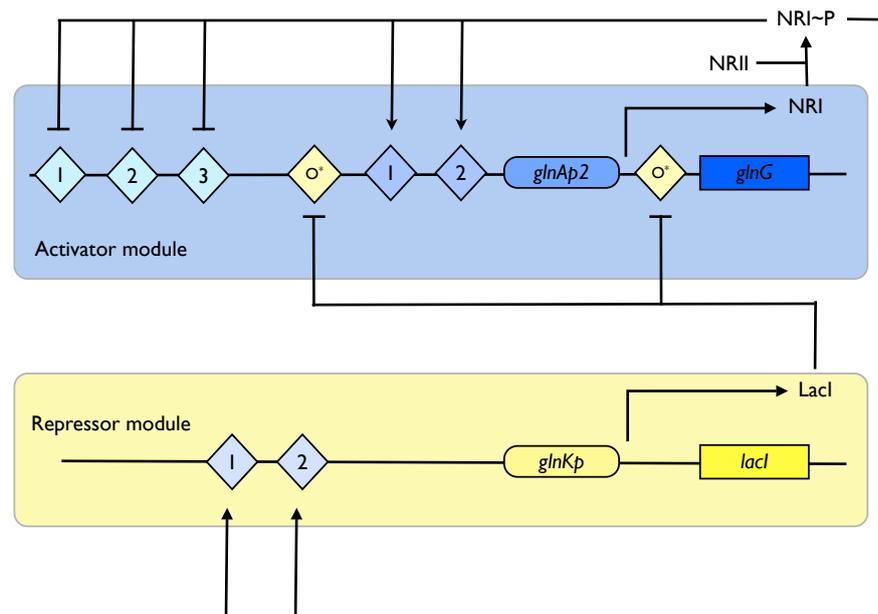
Figure 1.7: Diagram of the regulatory interactions in the synthetic clock, which consists of two gene modules. The activator modules consists of the structural gene *glnG* for the transcription factor NRI, driven by the NRI-responsive *glnAp2* promoter. The repressor module consists of the gene of the *lac* repressor, driven by the *glnKp* promoter. The circuit realizes an auto-activating feedback of the activator module, as the phosphorylated form of NRI, NRI~P binds to enhancer sites in the *glnAp2* promoter, that enhance its activity. A second, negative feedback is realized by NRI~P stimulating the expression of the *lac repressor* via enhancer sites in the *glnKp* promoter. The *lac* repressor then inhibits the expression of NRI by binding to *lac* operator sites inserted within the *glnAp2* promoter of the activation module. Additional low affinity (governor) sites for NRI in the *glnAp2* promoter inhibit its activity when NRI concentrations are high.

type *E. coli*, NRII activity is regulated depending on the amount of nitrogen and carbon available to the cell. In order to assure a constant phosphorylation of NRI, a mutant form NRII2302 was provided by a plasmid [8]. NRII2302 exhibits strong phosphorylation activity of NRI, independent of the cell state.

In their publication [8], the authors demonstrated the circuits ability to produce synchronous damped oscillations in turbidostat cultures (see figure 1.8). The oscillations lasted for about four cycles with a period of 10-20 hours depending on the growth rate of the cells. To synchronize the cells, they were grown in a medium containing the *lac* inducer IPTG, which inactivates the *lac* repressor. The clock was then started by removal of IPTG from the medium. The clock function was monitored by determining the amount of β-galactosidase (LacZ) in the cells in effluent samples from the chemostat at regular intervals.
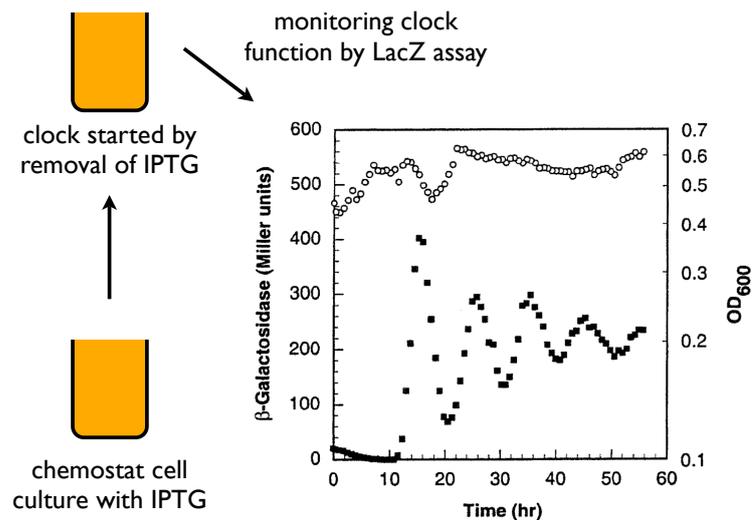
Figure 1.8: Monitoring clock function of the engineered circuit in chemostat cultures using LacZ assays (figure adapted from [8]). To synchronize the clock, cells were grown in the presence of IPTG, which inactivated the *lac* repressor. At time $t = 0$, IPTG was removed from the culture by dilution of the medium. Clock function was measured by assaying LacZ activity (β-Galactosidase, black sqares) from effluent samples at regular time intervals. The cell density was kept approximately constant during the clock run (density measurements are denoted by circles), i.e. the chemostat was used as a turbidostat.

Summary: Biological clocks

The first section gave an overview on research on circadian clocks. Certain paradigms and patterns have emerged form studies of clocks in a number of model organisms. In multi-cellular organisms clocks are functional within isolated single cells, where higher organisms have mechanisms to coordinate clocks running in different tissues. Central oscillators produce rhythms by transcriptional feedback loops, incorporating environmental time-setting signals, and regulate rhythmic expression of other genes via diverse signalling pathways. Observed rhythms are robust in the face of fluctuating environmental conditions, genetic noise, and often even gene knock-outs. As an example, the molecular basis of the clock of *Drosophila* was described in more detail, where multiple interlocking feedback loops work in concert to produce oscillations, a pattern also found in other circadian clocks.

The second section described experimental approaches that have been employed to investigate the complex network of genetic regulation. Large-scale, though mostly static, data can be collected by microarray assays, from which genetic regulatory interactions can be identified. Dynamical expression data with a better time resolution can be obtained, for example, by the use of automated imaging assays based on fluorescent and luminescent reporter molecules like GFP and luciferase, or other protein and mRNA assays, which are more difficult to automatize. Research typically focuses on small to medium-sized functional subsets of genes. Prominent problems in this context include the lack of knowledge of the involved components and of the details of their interactions, which limits the possibilities for mathematical modelling. Alternatively, artificial genetic circuits that are engineered from well known natural components are being studied, which promise to be a better basis for detailed mathematical modelling than natural systems.

The third section introduced a synthetic genetic oscillator, consisting of an activator and a repressor module, which were constructed from the gene of the *lac* repressor and components from the bacterial nitrogen system [8]. The circuit realizes a positive and negative feedback loop. In [8] it was shown to produce damped oscillations in chemostat experiments, where the clock function was measured by LacZ assays of samples taken at periodic intervals. The molecular details of the transcriptional activation and repression of the clock modules were discussed. Some additional background information was provided of the role of the clock components

in wild type bacteria.

Chapter summary

Models for dynamically interacting systems can be described by network systems. Depending on the kind of players and interactions, the nodes, links and structure of the network can be of static or dynamic nature. In genetic networks, variables that describe genes and gene products differ in their degrees of freedom, which means that care has to be taken in the corresponding network model construction. Popular models for gene networks have been based on systems of ODEs, although it has become clear that discrete stochastic events can be important in gene regulation and therefore need to be considered. Recently work has been undertaken to rigorously derive deterministic equations from discrete stochastic formulations, which in gene regulation take the form of systems with variables of finite states and infinite copy numbers.

The molecular basis of circadian clocks is formed by oscillating gene circuits, that feature interlocking transcriptional feedback loops. Natural clocks are built in a way that makes them robust to genetic and environmental noise. Understanding and modelling natural oscillators is difficult, because often important components are not known or interactions poorly understood. An alternative approach is the study of small engineered circuits of well understood natural components, which promise to be a good basis for mathematical modelling.

Project aims

This thesis studies such a synthetic clock, which has been shown to produce damped oscillations by measuring it in chemostat experiments and using LacZ assays. The aims of the project were to construct and analyse a detailed model of the regulatory interactions within this engineered circuit, and to collect data for fitting and optimization of the parameters. The theory of systems with finite states and infinite copy numbers, that was described in section 1.1, was used to study the synthetic oscillator that was introduced in section 1.2. This allowed to derive deterministic equations from the discrete stochastic description of the regulator interactions of the clock components, for which experimental data is available in the literature. A new assay for clock function was set up, based on *in vivo* imaging of a luminescent

reporter construct. This setup allowed automatized data collection and, for future work, offers the possibility to screen for interesting clock functions on a colony basis.

Chapter 2

From stochastic to deterministic models of gene expression

This chapter is divided into three sections. The first section reviews a general modelling framework for molecular systems with finite states and infinite copy numbers, which are discussed in more detail by Sbano [69, 66]. The focus lies here on how deterministic equations can be derived from an underlying discrete stochastic description of the system by a continuum limit and time-scaling argument. The second section deals with the construction of the discrete stochastic model of genetic regulation by DNA binding factors, in particular in the presence of cooperative binding and DNA loop formation, which were not treated in [69, 66]. An algorithm was developed and implemented that constructs the infinitesimal generator for a Markov chain that describes the time evolution of the probabilities of different activity states of one or multiple gene modules. The third section finally applies the techniques laid out in the first two sections to derive and analyse modes of genetic inhibition mediated by the *lac* repressor. This section extends work by Vilar [85] on the DNA looping of the tetrameric repressor by deriving and analyzing models for DNA loop formation mediated by the dimeric form of the repressor.

2.1   From stochastic to deterministic descriptions of molec-ular systems

The discrete state approach to genetic regulation has been reviewed by Kepler and Elston [40] illustrated by examples of simple activating factors. Details on the derivation and expansion of the Master equation are described in the books by van Kampen [83] and Gardiner [25]. The theory for molecular systems with finite states and infinite copy numbers has been studied in detail by Sbano and Kirkilionis in [69, 66, 70, 67, 68].

Molecular systems with finite states and infinite copy numbers

Modelling gene expression involves the description of the evolution of the transcription and translation products, which in turn depend on the activation state of the promoters of the genes in question. On a more abstract level and disregarding the biochemical details one can simplify these complex processes in terms of discrete states and stochastic transitions between them. More precisely, assume there is a fixed number of macromolecucles, for example DNA carrying genetic code or membrane channels, which can reside in a finite number of different states - a gene can be transcriptionally active or inactive, a membrane channel can be open or closed. Further there are smaller molecules that are characterized only by their number, and which are produced or destroyed depending on the state of the macromolecules. As reviewed in [66] the evolution of the species numbers can be described by a Master equation:

$$\frac{\partial P(s, \mathbf{n}, t)}{\partial t} = \sum_{s, s' \in S} L^*_{ss'}(P(s', \mathbf{n}, t)) + \frac{1}{\epsilon} \sum_{s' \in S} K^T_{ss'}(\mathbf{n}) P(s', \mathbf{n}, t), \qquad (2.1)$$

where s denotes the state of the macro-molecules, $s \in S$ is the corresponding finite state space, and N is the number of species of the smaller molecules. Further $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{N}^N$ is the vector that describes the number of particles for each of the N species, and $P(s, \mathbf{n}, t)$ is the probability that at time t the system has n_i particles or the i-th species, $i = 1, \ldots, N$ and is in state s. The matrix K is the infinitesimal generator of the Markov chain that contains the transition rates between the states in S. The small parameter $\epsilon > 0$ describes the time scale of the

evolution of the Markov chain that is defined by K. This parameter is motivated by the notion that the Markov chain K evolves much faster than the production and destruction events of the species. L^* is a matrix of difference operators describing the birth-death process of the species. In our context the matrix L^* will always be diagonal with diagonal elements L_s^* for $s \in S$. Hence equation (2.1) reduces to

$$\frac{\partial P(s, \mathbf{n}, t)}{\partial t} = L_s^*(P(s, \mathbf{n}, t)) + \frac{1}{\epsilon} \sum_{s' \in S} K_{ss'}^T(\mathbf{n}) P(s', \mathbf{n}, t). \qquad (2.2)$$

The structure of the state space S is determined by the geometry of the states that the macro-molecules can reside in, for example conformational states for membrane channels, or binding states for genetic promoter regions. In particular, in the case where the states s are defined by the binding state of a number of binding sites on the macro-molecules, which the communicating molecules can bind to, the structure of S is typically a product space, that is formed by the Cartesian product of the state spaces corresponding to the individual binding sites.

The average dynamics

When the amount of particles is given in terms of concentrations $\mathbf{x} \in \mathbb{R}_+^N$, then the system's dynamics are described by a corresponding Focker-Planck equation (FPE). Also, the FPE can be regarded as an approximation of the master equation (2.1), for when the number of particles is approximated by the concentration $\mathbf{x} = \delta \mathbf{n}$, where δ is the inverse of the system size (volume or average particle numbers). The probability distribution $P(s, \mathbf{n}, t)$ becomes a density $p(s, \mathbf{x}, t)$ and the difference operators L_s^* turn into differential operators $\hat{L}_s^*$. The FPE is given by

$$\frac{\partial p(s, \mathbf{x}, t)}{\partial t} = \hat{L}_s^*(p(s, \mathbf{x}, t)) + \frac{1}{\varepsilon} \sum_{s' \in S} K_{ss'}^T(\mathbf{x}) p(s', \mathbf{x}, t), \qquad (2.3)$$

where the parameter ε is related to ϵ by the time and size scales as discussed in [66]. In chapter 3 the explicit from of the matrices L_s^* and $\hat{L}_s^*$ are given for an example system (see tables 3.1 and 3.2).

From different perspective, the FPE (2.3) can be understood as describing a dynamical system on $\mathbb{X} = \mathbb{R}_+^N \times S$ with elements $(\mathbf{x}, s) \in \mathbb{X}$, whose dynamics are given by a hybrid set of laws. For each $s \in S$ the evolution on $\mathbb{R}_+^N$ is given by a

deterministic vector field,

$$\frac{d\mathbf{x}(t)}{dt} = X^{(s)}(\mathbf{x}(t)). \tag{2.4}$$

The vector field is deterministic, because we neglect diffusion of the particles, which otherwise would yield a corresponding stochastic differential equation. Rather we assume that particles are homogeneously distributed in the system. So the Fokker-Planck equation (2.3) does not contain a diffusion term.

For each $\mathbf{x} \in \mathbb{R}_+^N$ the evolution on S is a finite Markov chain, whose probability distribution evolves according to the FPE

$$\frac{dP(t,s)}{dt} = \sum_{s' \in S} K_{ss'}^T(\mathbf{x}) P(t,s'). \tag{2.5}$$

The vector field in (2.4) is related to the FPE (2.3) by the equation

$$\hat{L}_s^*(p(s,\mathbf{x},t)) = -\nabla \left(X^{(s)}(\mathbf{x}) p(s,\mathbf{x},t) \right). \tag{2.6}$$

If we assume that the Markov chain on S defined by K reaches its equilibrium probability distribution on a time scale that is shorter than the time scale of the evolution of the concentrations $\mathbf{x} \in \mathbb{R}_+^N$, defined by (2.4), we can separate the two processes by an adiabatic approximation of (2.3). In fact, in [66, 69, 70] it is shown that the FPE (2.3) can be solved by looking for an asymptotical solution in ε, with

$$p(s,\mathbf{x},t) = \sum_{n=0}^{\infty} \varepsilon^n p_n(s,\mathbf{x},t).$$

The leading order term of this expansion is the marginal distribution

$$f(\mathbf{x},t) = \sum_{s \in S} p(s,\mathbf{x},t),$$

whose evolution is given by

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = \sum_{s \in S} \hat{L}_s^* \left(\mu_s(\mathbf{x}) f(\mathbf{x},t) \right). \tag{2.7}$$

The vector $\mu(\mathbf{x})$ is the invariant measure of the Markov chain, which is characterized

by

$$K^T(\mathbf{x})\mu(\mathbf{x}) = 0, \qquad \sum_{s \in S} \mu_s(\mathbf{x}) = 1. \tag{2.8}$$

For the systems studied here the Markov chain always evolves to a unique stationary measure, although it is possible also to treat the case of non unique invariant measures [66, 69, 70].

Plugging the expression for $\hat{L}_s^*$ (2.6) into (2.7) yields

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = -\sum_{s \in S} \nabla \left(X^{(s)}(\mathbf{x})\mu_s(\mathbf{x})f(\mathbf{x},t) \right). \tag{2.9}$$

The preceding equation is a Liouville equation for the deterministic dynamical system in $\mathbb{R}^n$, whose trajectories are described by the system of differential equations

$$\frac{d\mathbf{x}(t)}{dt} = \sum_{s \in S} X^{(s)}(\mathbf{x})\mu_s(\mathbf{x}). \tag{2.10}$$

For details of the correspondence of the FPE and ordinary and stochastic differential equations see the books of Gardiner and van Kampen [25, 83].

Decoupling of Markov chains

In the case of a gene, whose expression activity is governed by two types of regulatory factors binding to two distinct operator sites, the state space of the corresponding MC has the structure of a product of the two state spaces corresponding to the two operator sites. More generally let the matrix K generate the Markov chain on the state space $S = S_1 \times \ldots \times S_d$. Then K can be written as a sum of tensor products

$$K = \sum_{l=1}^{d} I^{\otimes(l-1)} \otimes K_l \otimes I^{\otimes(d-l)},$$

where K^l generates a Markov chain on S_l, and $I^{\otimes m}$ is a shorthand for $\underbrace{I \otimes I \otimes \ldots \otimes I}_{m \text{ times}}$.
It can be shown that the unique stationary measure μ of K can be written as

$$\mu = \mu_1 \otimes \ldots \otimes \mu_d,$$

where $\mu_l K_l = 0$ for all l. The decoupling of Markov chains in this fashion is described in more detail in [68]. Definitions and properties for the tensor product can be found

in [48]. This theory will be applied to decompose the system of the engineered clock circuit described in the introductory chapter, which consists of two genes governed by multiple operator sites.

Summary: From stochastic to deterministic descriptions of molecular systems

In this section it was shown how gene regulation and expression can be described as a stochastic system of finite states and infinite copy numbers. From the stochastic description deterministic equations can be derived by taking appropriate limits - an adiabatic limit for fast operator state changes and a continuum limit to move from single particles to concentrations.

The deterministic rate laws derived this way have the advantage to be based on the underlying discrete description and can thus discriminate between different assumption of the dynamics at the molecular scale rather than having to rely on heuristics. From the analytical perspective deterministic equations are preferable to stochastic ones because more can be said about their dynamics.

The main task for setting up a model for a given genetic regulatory network is the generation of the Markov chain. In this section it was shown how Markov chains can be broken up into independent modules with the help of the tensor product. Yet the construction of the Markov chain can be tedious even for smaller systems, especially if it is done by hand. This motivated the fabrication of an algorithm that automatically generates the Markov chain for general models for genetic regulation of operons with any number of binding sites and regulatory DNA looping. This algorithm is presented in the next section.

2.2 Algorithmic derivation of the infinitesimal generator of the Markov chain of the gene states

In our model context the activity state of a gene is defined by the binding state of its operator sites. In order to write down the corresponding ME or FPE (see (2.1), (2.3)) one needs to construct the infinitesimal generator K, which contains the transition rates between the gene states and defines the evolution of the corresponding Markov chain. As an example, consider a gene (or another macro-molecule) that has two

binding sites o_1, o_2, that can either be free, $o_i = 0$ or occupied, $o_i = 1$. Then the corresponding state space S for the macro-molecule is the Cartesian product of the state spaces of the individual binding sites S_1, S_2:

$$S = S_1 \times S_2, \qquad S_i = \{b_i | \ b_i \in \{0,1\}\}, \ i = 1, 2.$$

Writing S out in full, we can number the states as indicated,

$$S = \{(0,0), (0,1), (1,0), (1,1)\} = \{s_1, s_2, s_3, s_4\}.$$

Now, assume that we know the rates at which the gene switches from state s_i to state s_j, and call them ρ_{ij}. We can then write down the infinitesimal generator K, that defines the Markov chain (K, S) by

$$K = \begin{bmatrix} -\rho_{12} - \rho_{13} - \rho_{14} & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{21} & -\rho_{21} - \rho_{23} - \rho_{24} & \rho_{23} & \rho_{24} \\ \rho_{31} & \rho_{32} & -\rho_{31} - \rho_{32} - \rho_{34} & \rho_{34} \\ \rho_{41} & \rho_{42} & \rho_{43} & -\rho_{41} - \rho_{42} - \rho_{43} \end{bmatrix}.$$

This can be simplified by acknowledging that transitions happen by single binding and unbinding events at either one of the two binding sites o_1, o_2. Assume that o_i switches between 0 and 1 with the rates

$$0 \quad \begin{array}{c} \xrightarrow{\lambda_i^+} \\ \xleftarrow[\lambda_i^-]{} \end{array} \quad 1,$$
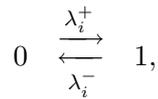
then

$$\rho_{12} = \lambda_2^+, \rho_{21} = \lambda_2^-, \rho_{13} = \lambda_1^+, \rho_{31} = \lambda_1^-,$$

$$\rho_{24} = \lambda_1^+, \rho_{42} = \lambda_1^-, \rho_{43} = \lambda_2^-, \rho_{34} = \lambda_2^+,$$

$$\rho_{23} = \rho_{32} = \rho_{14} = \rho_{41} = 0.$$

Note here that we only consider single binding and unbinding events - transitions that would involve the simultaneous change of binding states of both binding sites are not considered. This is so, because the probabilities of such events are very small. Consequently, the rates that describe such transitions, in this example

$\rho_{23}, \rho_{32}, \rho_{14}, \rho_{41}$, are set to zero. So K can be rewritten as

$$
K = \begin{bmatrix}
-\lambda_2^+ - \lambda_1^+ & \lambda_2^+ & \lambda_1^+ & 0 \\
\lambda_2^- & -\lambda_2^- - \lambda_1^+ & 0 & \lambda_1^+ \\
\lambda_1^- & 0 & -\lambda_1^- - \lambda_2^+ & \lambda_2^+ \\
0 & \lambda_1^- & \lambda_2^- & -\lambda_1^- - \lambda_2^-
\end{bmatrix}. \tag{2.11}
$$

The invariant measure μ for K is a probability distribution the states of the Markov chain that satisfies the equation $K^T \mu = 0$. Sometimes it is more convenient to state some vector v, for which $K^T v = 0$, and make it a probability distribution by an appropriate normalization. For example the invariant measure of K as defined in (2.11) is given by

$$
\mu = \frac{v}{\sum_i v_i}, \qquad v = (1, \lambda_2, \lambda_1, \lambda_2 \lambda_1),
$$

where the substitutions $\frac{\lambda_i^+}{\lambda_i^-} = \lambda_i, i = 1, 2$ were made. If, instead of just two binding sites, there were n of them, the corresponding state space S would consist of 2^n possible states. So, in the presence of multiple binding sites, it is of advantage to be able to compute the state space S and generator K of the markov chain (S, K) automatically.

In order to achieve this an algorithm was created which builds this infinitesimal generator based on the biological model of the regulatory interactions at the promoter regions regulating gene activity. The algorithm takes into account the information on the number of operator sites, what sites must be free or occupied for transcriptional activity, which molecules bind to which sites and whether there are cooperativity effects. It can also be specified under what conditions a DNA loop can be formed, which happens when a molecule binds to two sites on the DNA simultaneously. The algorithm produces the infinitesimal generator and also indicates which states are transcriptionally active.

This section specifies the class of models, for which the algorithm can construct the infinitesimal generator. It describes how the gene states, the possible transitions and the corresponding rates are defined. It explains how cooperative binding and regulatory DNA looping are implemented and discusses the format of the input and output files. The algorithm already provides a certain form of the transition rates on the basis that transitions between states describe binding and unbinding events

of proteins to DNA. A number of examples are provided for illustration.

An overview of the algorithm

The algorithm takes in a description of the binding mechanisms for the operator sites of the operons and the associated activity states of the genes. From this, the state space S of these molecular machines is computed together with the corresponding infinitesimal generator K. The details on how this mathematical construction is implemented is described in this section. The algorithm was implemented as a Perl script and can be downloaded from the Compio website at http://lora.maths.warwick.ac.uk.

The gene states. The state of the Markov chain is defined by the binding states of a number of operator sites $o_1, o_2, \ldots, o_n$ - operator sites in this context are understood to denote all DNA binding sites that affect the transcription of the genes in question. Each of these operators can reside either in a free state $o_i = 0$ or in bound state $o_i = 1$, when a regulating factor is bound to it. The operator sites regulate the activity of a number of genes $p_1, p_2, \ldots, p_m$, where each operator site is associated to exactly one gene. ('Gene' in this context may also stand for transcriptional unit of a group of genes, commonly referred to as operon. Therefore here 'gene' and 'operon' will be used interchangeably.) Each gene may reside in an inactive $p_i = 0$ or active state $p_i = 1$ depending on the binding states of its operators. See figure 2.1 for an illustration.

Example 1: Simplified clock circuit scheme. As an example consider a simplified version of the engineered circuit described earlier (chapter 1) that consists of two genes p_1, p_2. Gene p_1 is driven by two sites o_0, o_1 and p_2 is driven by the operator site o_2. The sites o_0, o_2 are enhancer sites and must be occupied for the promoter to be active. Site o_1 is an inhibitor site and p_1 can only be active if o_2 is free. This scheme is illustrated in figure 2.2. The system has eight possible states $(o_0, o_1, o_2) \in \{0, 1\}^3$, which can be grouped according to their activity profile: Here '*' means that the operon's activity is independent of the binding state of that operator. The first line of the table means, that the enhancer site o_0 needs to be occupied for expression of p_1 to occur, and that the repressor site o_1 must be empty.
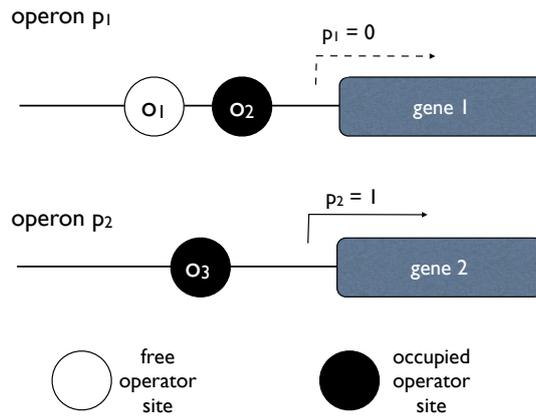
Figure 2.1: The state of genetic operons is defined by the binding states of a number of operator sites. Depending on on the binding profile of the operator sites, the genes can be either transcriptionally active, like gene p_2 in the figure, or transcriptionally inactive, as indicated for gene p_1 in the figure.

	o_0	o_1	o_2
gene p_1 is active when	1	0	*
gene p_2 is active when	*	*	1

Table 2.1: Example 1: Activity of the genes depending on the operator states.

The binding state of o_2 has no influence on the activity of p_1. Similarly, p_2 is active if and only if the enhancer site o_2 is occupied, while the binding state of o_0, o_1 is irrelevant.

Transitions between gene states. The state of the gene is changed when a molecule binds to or is released from one of its operator sites. Only transitions by single binding and unbinding events are considered, because the probability of several operator sites to change state simultaneously is very small. The rates at which an operator switches from the free to the occupied state depends on the amount of molecules in the system that can bind to it and on the rate at which this occurs. It is assumed that operators are molecule specific, which means that only one type of molecule can bind to it. Let ω_i^+ be the binding rate per molecule and x is the number of molecules in the system. When k is the number of molecules that are already bound to other operator sites, then the overall transition rate for
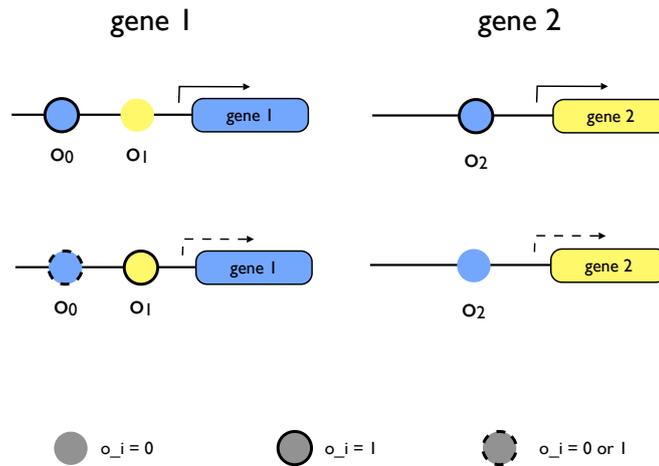
Figure 2.2: Example 1: Operator and promoter states and transition for a simplified version of the engineered clock circuit. Gene 1 corresponds to the gene for the transcription factor NRI, gene 2 corresponds to the gene for the lac repressor LacI.

o_i from the free to the occupied state is:

$$\text{rate of 'free' to 'occupied' transition of operator } o_i \;=\; \omega_i^+(x-k).$$

The rate at which operators switch from the occupied to the free state is determined by the dissociation rate of the bound molecule from the DNA, ω_i^-:

$$\text{rate of 'occupied' to 'free' transition of operator } o_i \;=\; \omega_i^-.$$

Example 1, continued. When we apply the rules laid out above to the example system (figure 2.2), we can write down all possible transitions and their rates in a graphical structure as depicted in figure 2.3. The graph contains all information of the infinitesimal generator K and is also generated by the script in the form of a GraphViz '.dot' file (see below).

Cooperative binding. The binding of two neighbouring operator sites o_i, o_j may happen cooperatively, which means that the binding state of one site will affect the transition rates at the other site. This mechanism is reflected in the model by changing the corresponding transition rates. So when $o_i = 1$ the rate of binding for
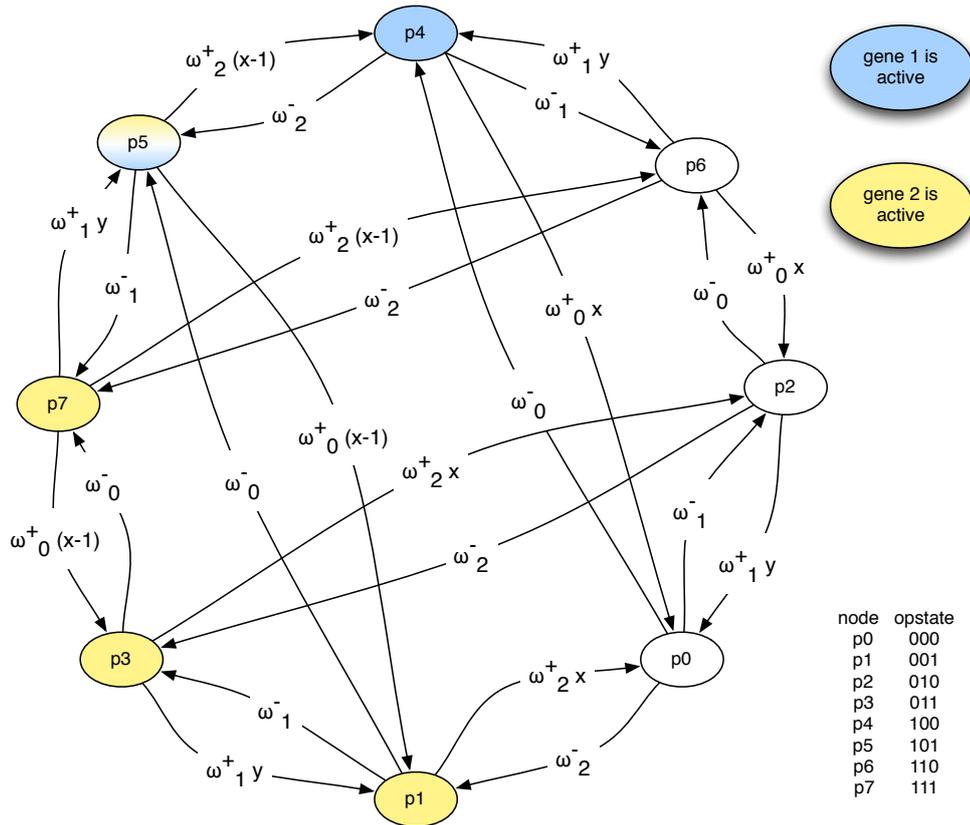
Figure 2.3: Example 1. Transition graph for the operator states of the simplified version of the engineered clock circuit depicted in figure 2.2. Each node indicates a state in the Markov chain that is characterized by the binding states of the three operator sites - o_0, o_1 for gene 1, and o_2 for gene 2. The table next to the graph shows the binding profile for each state in the form o_1, o_2, o_3. 'opstate 101' for example means that $o_0 = 1, o_1 = 0, o_2 = 1$. States for which gene 1 is active are marked in blue, states for which gene 2 is active are marked in yellow. The edges are labelled with the corresponding transition rates.

o_j is scaled by the (non-negatve) cooperativity factor γ_{ij}:

$$\gamma_{ij}\omega_i^+(x-k).$$

This means, that if $\gamma_{ij} > 1$, then the occupation of o_i facilitates the binding to o_j (positive cooperativity). If $\gamma_{ij} < 1$, then the occupation of o_i obstructs the binding to o_j (negative ccoperativity). If $\gamma_{ij} = 1$, then the binding state of o_i does not affect the binding to o_j. Typically the cooperative relation of two binding sites is mutual, which means $\gamma_{ij} = \gamma_{ji}$. An example of cooperative binding between two operator sites is depicted in figure 2.4.
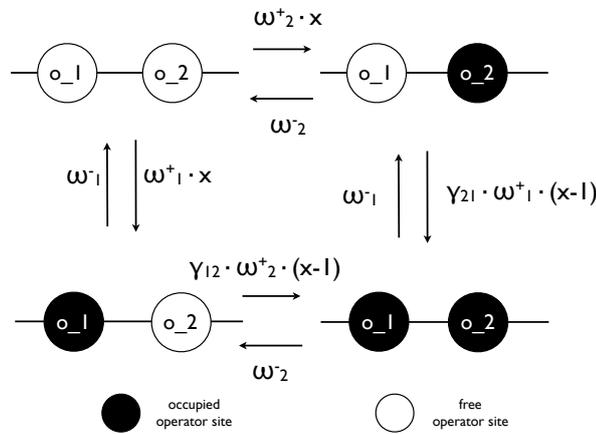
Figure 2.4: Example 2. Illustration of transition rates for cooperative binding. The regulatory factor X, of which x particles are present in the system, binds cooperatively to the operator sites o_1 and o_2. This means that when one of the operator sites, say o_1, is occupied the binding rate to the second operator site o_2 is scaled by the cooperatvity factor γ_{12}.

DNA looping. Regulatory factors that have two active domains for binding onto DNA may mediate the formation of a DNA loop. In the MC model this is reflected by introducing a 'looped state' that involves two operator sites. It needs to be specified which two sites are involved in the DNA loop formation. The rates for the closing and opening of the looped state are denoted by ψ_{ij}^+ and ψ_{ij}^- - the order of the indices ij indicates that the loop is formed by a factor binding to o_j that is already bound to o_i and that the loop is opened by dissociation of the factor from o_j. For an illustration see example 3, which is depicted in figure 2.5.

The rates for the loop formation and opening can be derived from the rate

constants for the involved operators sites, modified by proportionality constants $a(= a_{ij} = a_{ji})$ and $b(= b_{ij} = b_{ji})$:

$$\psi_{ij}^{+} = a\omega_{j}^{+}, \ \psi_{ji}^{+} = a\omega_{i}^{+}, \tag{2.12}$$

$$\psi_{ij}^{-} = b\omega_{j}^{-}, \ \psi_{ji}^{-} = b\omega_{i}^{-}. \tag{2.13}$$

The constant a takes into account the changed entropy gradients for the binding event that forms the loop - the difference in energy that is needed for a molecule in solution to bind to the operator site in question compared to the energy that is required for a molecule that is already bound to a nearby operator site. The constant b describes the change in the dissociation rate caused by the bending of DNA due to the looping (for a detailed discussion on the thermodynamics of DNA looping, see [85]).

Input and output format of the algorithm

This section gives information about the input and output format of the algorithm. Firstly, it shows how the examples presented earlier are translated into the syntax of the input file and how it is done for more general cases. Secondly, it explains the format of the different sections of the output file and how the algorithm presents the information regarding the Markov chain states, the infinitesimal generator and the transition graph.

Input syntax. These are the input files for the examples 1, 2, 3 described earlier. After that the general syntax of the input is discussed.

Input file for the example system 1 as depicted in figure 2.2:

```
# Begin first operon section.
operon activator
# Number and type of operator sites.
# The letter indicates the name of the DNA binding factor
opsites x y
# Define states in which operon is active.
# Separate multiple states by comma. * denotes 0 or 1.
active 1 0

# Begin second operon section.

operon repressor
```
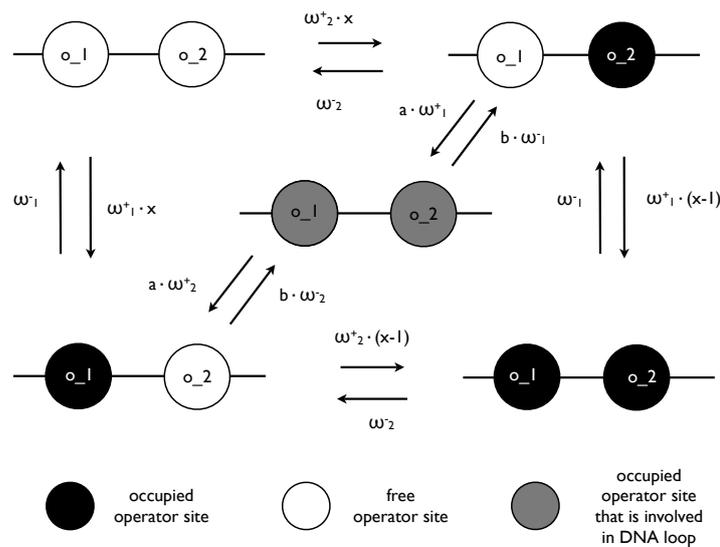
Figure 2.5: Example 3. Illustration of a model for DNA loop formation as it occurs for regulatory DNA binding proteins like the *lac* repressor LacI that have multiple domains for binding to DNA. Here two operator sites o_1, o_2 can be bound the regulatory factor X, whose particle number is denoted by x. The looped state can be attained from any state, where exactly one of the operator sites is occupied. From such a state, for example $(o_1, o_2) = (1, 0)$, there are three possible transitions. First, the operator o_1 can become free and the system goes to the state $(0, 0)$. Second, another free regulatory factor may bind to the second operator and the system moves to state $(1, 1)$ without a loop being formed. Thirdly the factor X that is already bound to the site o_1 can also bind to o_2 with its second DNA binding domain. This yields the state $(1, 1)$ with a DNA loop present. The transition rates $\psi_{ij}^{\pm}$ to and from the loop state are indicated in the diagram as given in equations (2.12, 2.13).

```
opsites x

active 1
```

Input file for the cooperative example scheme 2 as depicted in figure 2.4:

```
# Begin operon section.

operon cooperative

opsites x x

# For example. Both sites must be occupied.

active 1 1

# Specify mutual cooperativity between sites 1 and 2

coop 1 1
```

Input file for the looping example scheme 3 as depicted in figure 2.5:

```
# Begin operon section.

operon Loop

opsites x x

# For example. Site 1 must be free.

active 0 *

# A loop can be formed between sites 1 and 2

loop 1 1
```

In sum the structure of the input file is as follows. The file is structured into one or several operon (gene) definition blocks. Each block begins with the keyword 'operon' followed by the name of the operon. The next line specifies the number and type of operator binding sites. After the keyword 'opsites' follows a list of characters indicating for each operator site the type of molecule that can bind to it. Each operator site can only be targeted by one type of molecule. After the 'operon' and 'opsites' definitions, there may follow further definitions of operon activity (keyword 'active'), of cooperative binding (keyword 'coop'), and of loop formation (keyword 'loop'). It is important though that 'operon' and 'opsites' are defined first.

After the keyword 'active' it can be specified which operator sites must be occupied (marked by '1'), or free (marked by '0'), for the operon to be active. If the activity of that operon is independent of the binding state of an operator site, this

is marked by an asterisk '*'. Cooperative binding between two binding sites can be specified by the keyword 'coop'. The two cooperative sites are set to '1', all other sites must be set to '0'. Similar holds for defining a DNA loop formation with the keyword 'loop'. The two binding sites involved in the DNA loop are set to '1', all other sites are set to '0'. The format for all these definitions - the number and order of operator sites - must follow the one given in the 'opsites' line. If there are more or less binding sites than were specified after 'opsites', the script will reject the input.

Some additional notes on looping and cooperative binding. The loop as defined above is formed by a factor binding first to one operator site and then to the corresponding second site. It is also possible to specify a loop where two molecules first bind to each of the operator sites and then form a loop by dimerization. In that case the two involved operator sites need to be set to '2', not to '1', in the corresponding 'loop' definition. An example for this is given in the next section in the context of a *lac* repressor model. Concerning cooperative binding it is worth noting how it can be used to model operator sites to which different types of molecules can bind. This can be achieved by representing the operator site in question as two sites in the model with cooperative binding. If the cooperativity factor for the sites is set to zero, the binding rate of first operator site will become zero if the second operator is occupied, and vice versa.

Output format.   The script produces an output that is structured into five sections - A, B, C, D and E. Section A simply prints out the information provided by the input file. Because the states of the Markov chain are based on the states of all operator states, the algorithm merges the operator sites of all operons and treats them jointly. The resulting list of operator states and the corresponding activity, cooperative binding and looping definitions are compiled in section B. Section C gives a numbered list of all possible states of the Markov chain. Each state is defined by the occupation state of the operator sites and the presence of DNA loops. For each state it is specified which operons are active. The last entry for each state is a list of accessible states - states to which the Markov chain can move from the current state - and the corresponding transition rates. Section D prints out the actual infinitesimal generator K in a format readable by the mathematical software packages like Maple. Section E finally contains the definition of the transition graph. The nodes

are the states of the Markov chain, the edges denote all the possible transitions and are labelled with the corresponding rates. This graph definition is in a format readable by graph visualization and manipulations software packages such as GraphViz and OmniGraffle.

Below are given two examples on how the output file looks like. First the output generated for example 1 depicted in figure 2.2:

A – Overview - Operon structure –

Operon No. 0, activator, size 2 :
operator sites: x y
active when: 1 0

Operon No. 1, repressor, size 1 :
operator sites: x
active when: 1

B – Overview - operator states considered jointly –

All operator sites: x y x
Operon 0 active for: 1 0 *
Operon 1 active for: * * 1

C – Overview - states of the Markov chain –

MC state [No]; opsite-state; loop-state; active operons; [accessible state: transition rate];
MC state [0]; 0 0 0; ; none; [1: omega_ass[2]*x]; [2: omega_ass[1]*y]; [4: omega_ass[0]*x];
MC state [1]; 0 0 1; ; 1; [0: omega_diss[2]]; [3: omega_ass[1]*y]; [5: omega_ass[0]*(x-1)];
...

D – The infinitesimal generator K –

K:=matrix(8,8,[
-omega_ass[2]*x-omega_ass[1]*y-omega_ass[0]*x,omega_diss[2],omega_diss[1],0,omega_diss[0],0,0,0,
...

E – The .dot file for the transition graph –

digraph TransitionGraph {

p0 -> p1 [label="omega_diss[2]", weight=100];

p0 -> p2 [label="omega_diss[1]", weight=100];

...

The second example shows an extract of the output - sections B and C - for the

loop system, example 3, depicted in figure 2.5. It is given to show how a loop is incorporated into the Markov chain states:

```
...

B – Overview - operator states considered jointly –

All operator sites: x x
Operon 0 active for: 0 *
Looping pair No. 0: 1 1

C – Overview - states of the Markov chain –

MC state [No]; opsite-state; loop-state; active operons; [accessible state: transition rate];
MC state [0]; 0 0; 0; 0; [ 1: omega_ass[1]*x ]; [ 2: omega_ass[0]*x ];
MC state [1]; 0 1; 0; 0; [ 0: omega_diss[1] ]; [ 3: omega_ass[0]*(x-1) ]; [ 4: psi_ass[1,0] ];
MC state [2]; 1 0; 0; none; [ 0: omega_diss[0] ]; [ 3: omega_ass[1]*(x-1) ]; [ 4: psi_ass[0,1] ];
MC state [3]; 1 1; 0; none; [ 1: omega_diss[0] ]; [ 2: omega_diss[1] ];
MC state [4]; 1 1; 1; none; [ 1: psi_diss[1,0] ]; [ 2: psi_diss[0,1] ];

D – The infinitesimal generator K –

...
```

A short description of the algorithm

The algorithm can be structured into five steps that correspond to the sections A, B, C, D and E of the output file discussed earlier. First the model description that was provided in the input file is checked for consistency - for correct length and for allowed characters (section A). Next the model definitions for the individual operons are merged into a joint description in terms of all operator sites (section B). Now the states of the Markov chain are built by considering all possible DNA looping scenarios and compatible operator binding site profiles. Then follows the calculation of all possible transitions.

For each given state all corresponding states are identified to which a transition is possible - named accessible states. Accessible states are characterized in that they differ from the present state in the binding state of only one operator site. The change in the binding state of the operator must be compatible with the DNA looping scenario - either it represents the closing or opening of a loop, or it is not involved in any loop. In the case of a loop formation or opening by a dimerization process as described earlier, only the DNA looping scenario changes while the binding profile of the operator sites is unchanged.

For each of the identified transitions the corresponding rates are calculated. The rates are determined by the operator sites involved in the transition - including cooperative binding factors if applicable. In the case of association rates the number of free molecules is derived from the binding profile of current state. The results of these computations are compiled in section C of the output. From the information the algorithm has generated up to here the infinitesimal generator of the Markov chain and its graphical representation are compiled and presented in sections D and E of the output file. A flowchart representation of the elements of the algorithm is presented in figure 2.6.

Summary: Algorithmic derivation of the infinitesimal generator of the Markov chain of the gene states

This section introduced a piece of modelling software that generates a Markov chain model of regulation of one or several genes or operons. The Markov chain is built on information about the number and type of operator sites and on the required binding states for transcriptional activity. The model is able to account for DNA loop formation and cooperative binding, a feature that was not part of the papers [69, 66], which this work was based on. The generated output includes a definition of the Markov chain's infinitesimal generator and a definition of the corresponding transition graph that can be read by graph manipulation software like GraphViz and Omnigraffle. The features of the program as well as its input and output format were illustrated with a number of examples.

The software can be used to quickly generate stochastic models of small genetic networks where the dynamics of operator sites needs to be taken into account. By automatizing the typically tedious manual derivation of the Markov chain model, the program enlarges the scope of this kind of modelling. In particular it allows to quickly compare the consequences of different model assumptions on the nature of the molecular basis of the genetic regulations. To illustrate this point I will conclude this chapter with an analysis of different models for LacI mediated transcriptional repression, which is presented in the next section.
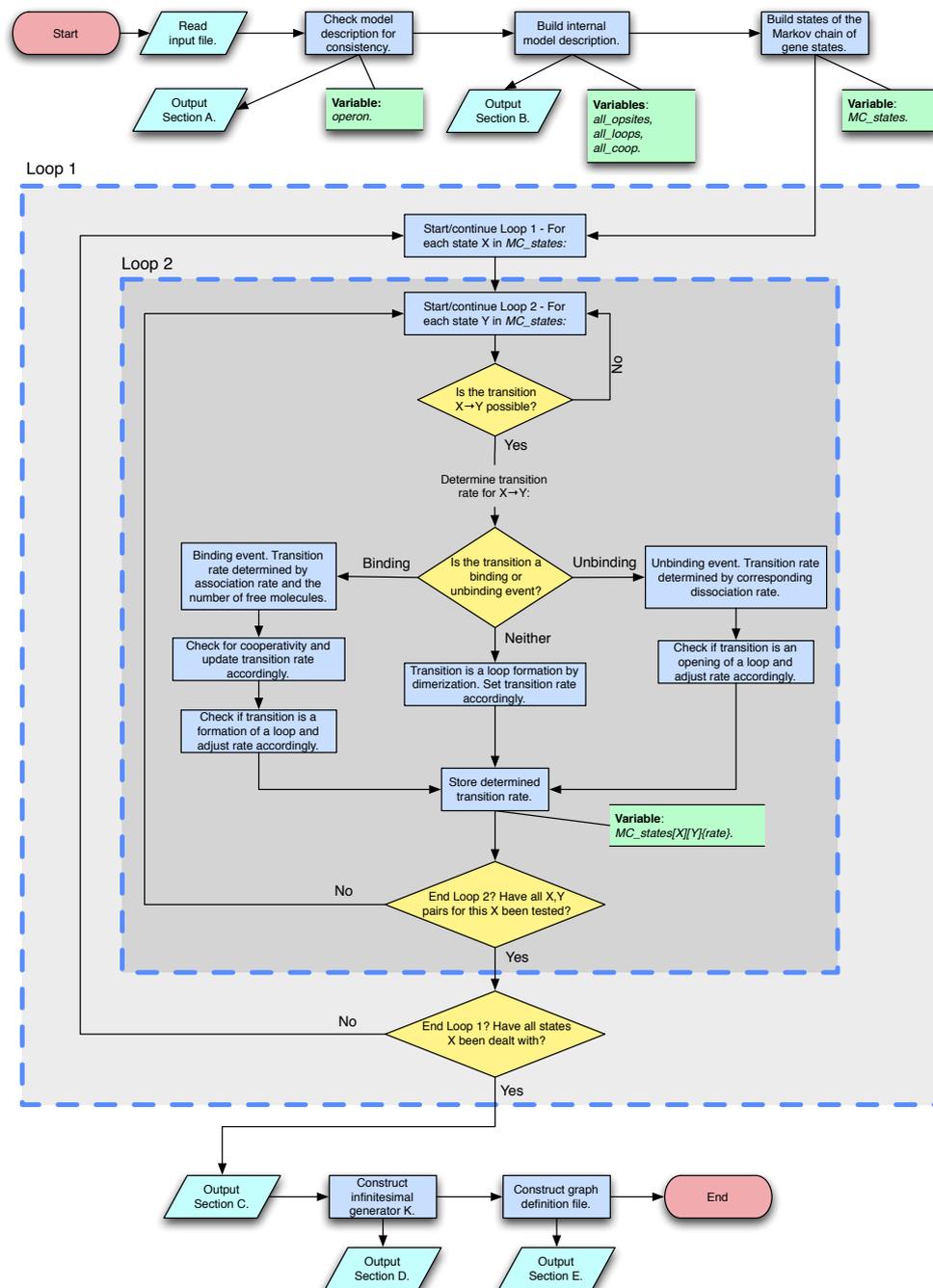
Figure 2.6: Flowchart representation of the elements of Markov chain generating algorithm. Start and endpoint are indicated in red, input and outputs in turquoise, decisions in yellow, processes in blue, and associated variables in green. Arrows indicate the flow of processes. Sometimes, in abuse of notation, arrows also indicate the associated output section of a process. 'Binding' and 'unbinding events' refer to a regulatory factor binding to, or dissociating from, an operator site on the DNA. See the text for further details on the derivation of the transition rates.

2.3 Example: analysis of different models for LacI mediated repression

This section derives and analyses several models for transcriptional repression mediated by the *lac* repressor LacI. In particular it is shown how the introduction of different kinds of DNA looping affects the structure of the transition graph and the properties of the transcriptional repression. The work presented here is based on previous work by Vilar [85] on regulatory DNA looping of the tetrameric *lac* repressor, where DNA looping was discussed from a thermodynamic perspective. Here these models are studied using the theoretical framework or systems of finite states and infinite copy numbers [66, 69] as laid out earlier in this chapter. Also, a loop forming mechanism based on dimerization is considered here, which was not discussed in [85].

Modelling LacI mediated transcriptional repression

The *lac* repressor LacI is a regulatory DNA binding factor that represses the transcriptional activity of the *lac* operon in wild type *E.coli* bacteria. The *lac* operon expresses proteins that are important for the breakdown of the lactose. In conditions when the cell has access to its preferred substrate glucose, the *lac* repressor inhibits the expression of the *lac* operon. When the cells has access to lactose but not to glucose, the *lac* repressor is inactivated by allolactose, a byproduct of the lactose breakdown, and expression of the *lac* operon is switched on. Here we are only interested in how LacI exerts the transcriptional repression. More precisely, in the binding dynamics of LacI to DNA, and in how LacI induces the formation of a DNA loop.

In wild type *E.coli* the promoter region of the *lac* operon has three binding sites for the *lac* repressor - called the main, first and second auxiliary operators. When the *lac* repressor is bound to the main operator site, it blocks the initiation of transcription by the RNA polymerase. The two auxiliary operators lie further upstream ('Upstream' and 'downstream' characterize the location of genetic code relative to the start site of transcription, where transcription is thought to proceed in the downstream direction.) and their binding states do not influence the transcriptional activity directly. Yet the presence of the auxiliary operators increases the

efficiency of repression dramatically [56]. This is due to the ability of LacI to initiate the formation of a DNA loop by binding to two sites on the DNA simultaneously.

The *lac* repressor is a tetrameric molecule that consists of a dimer of dimers. Each of these dimers features a DNA binding domain with a high affinity to the *lac* operator sites. Recently a model of the loop formation by the tetrameric repressor was published that explains how this kind of loop formation increases the overall repression efficiency [85]. But it has been suggested that there are at least two ways in which the DNA loop can be formed, depending on whether the repressor is in its dimeric or tetrameric form, as Lewis notes in [46] on p. 545:

> "There are two plausible mechanisms for looping the DNA that are consistent with the architecture of the *lac* repressor tetramer. The two subunits of the tetramer can bind to the primary operator site and the other dimer subsequently associates with an ancillary operator. Alternately, free repressor dimers could bind to separate operators and a loop would occur when the dimeric repressors associate into a tetramer. [...] Both mechanisms are plausible and depend on the precise physiological conditions."

To illustrate the consequences of these different loop forming mechanisms, this section considers several models for LacI mediated repression. The models assume the existence of either one or two *lac* operator sites and different loop formation mechanisms as indicated in figure 2.7. Model I considers only one operator site that blocks transcription. Model II considers two operator sites, both of which block transcription, but neglects the formation of a DNA loop. The other three models describe different looping mechanisms: Model III describes the dynamics of tetramic DNA loop formation. Model IV describes a loop formation that occurs exclusively by dimerization. Model V also models dimeric loop formation, but allows the loop also to open by the dissociation of the repressor from the DNA. It is assumed that the resulting tetramer breaks up immediately.

Construction of the models. The models for the binding of the *lac* repressor to the DNA (depicted in figure 2.7) are based on the principles laid out earlier (section
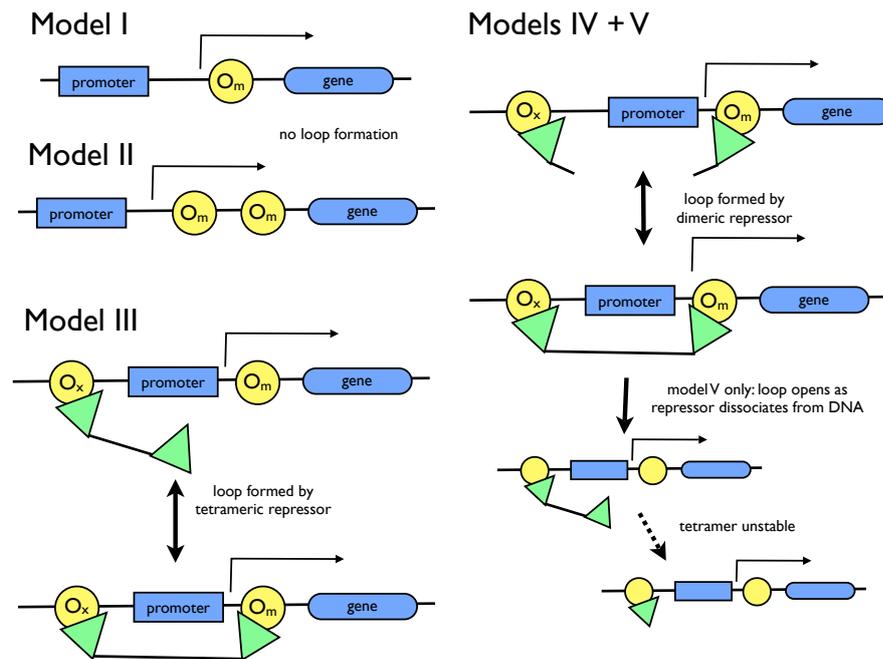
Figure 2.7: Characteristic features of the different models for LacI based transcriptional repression. Model I: one operator, no looping; model II: two operators, both of which block transcription when occupied, no loop formation. Models III-V feature a main and an auxiliary operator between which the *lac* repressor may form a DNA loop. Model III: the DNA loop is formed by tetrameric repressor; model IV: the DNA loop is formed by dimeric repressor; model V: the DNA loop is formed by dimeric repressor as in model IV, but the loop may also open by the dissociation of the repressor from the DNA, after which the resulting tetrameric repressor breaks up into dimers. The main operator sites, which block transcription when occupied, are labelled o_m, the auxiliary operators are labelled o_x.

2.2). The state spaces S_i for the models, $i = 1, \ldots, 5$, are:

$$
\begin{aligned}
S_1 &= \{o_m | \ o_m \in \{0,1\}\} \tag{2.14} \\
&= \{0,1\}, \\
S_2 &= \{(o_{m1}, o_{m2}) | \ (o_{m1}, o_{m2}) \in \{0,1\}^2\} \tag{2.15} \\
&= \{(0,0), (0,1), (1,0), (1,1)\}, \\
S_3 = S_4 = S_5 &= \{(o_x, o_m, l) | \ (o_x, o_m, l) \in \{0,1\}^2 \times \{0\} \cup (1,1,1)\} \tag{2.16} \\
&= \{(0,0,0), (0,1,0), (1,0,0), (1,1,0), (1,1,1)\},
\end{aligned}
$$

where o_m, o_x mark the main and auxiliary operators and l denotes the presence of the DNA loop. The second line of the definition of each state space S_i indicates the ordering of the states $(s_i^1, s_i^2, \ldots)$. For all models the promoter is assumed to be in an active state if and only if the (or both in model II) main operator sites o_m are free ($o_m = 0$ or $o_{m1} = o_{m2} = 0$).

The rate constants for *lac* repressor binding to and unbinding from the operator sites are λ_0^+, λ_0^- for the first operator, and λ_1^+, λ_1^- for the second (in the order as the sites appear in the definitions of S_i (2.14, 2.15, 2.16). The rates for the closing and opening of the DNA loops are $\psi_{ij}^+ = a\lambda_j^+, \psi_{ij}^- = b\lambda_j^-$ (compare equations (2.12), (2.13)), for the tetrameric repressor, and ψ_d^+, ψ_d^-, for the dimeric repressor, as discussed in the previous section 2.2. With the help of the algorithm described in the last section we can construct the infinitesimal generators $K_i, i = 1, 2, 3, 4, 5$ for the models I, II, III, IV and V (see table 2.2). The corresponding transition graphs are depicted in figure 2.8.

The models described here represent extreme cases where the *lac* repressor is exclusively present in either the dimeric or tetrameric form. In a more realistic scenario one would assume some equilibrium relationship between the dimeric and tetrameric repressor molecules. For simplicity's sake this more general scenario is not discussed here.

Model Analysis

The infinitesimal generators $K_i, i = 1, 2, 3, 4, 5$, define the evolution of the probability vector $p_j(t)$, which contains the probabilities for the system to reside in each
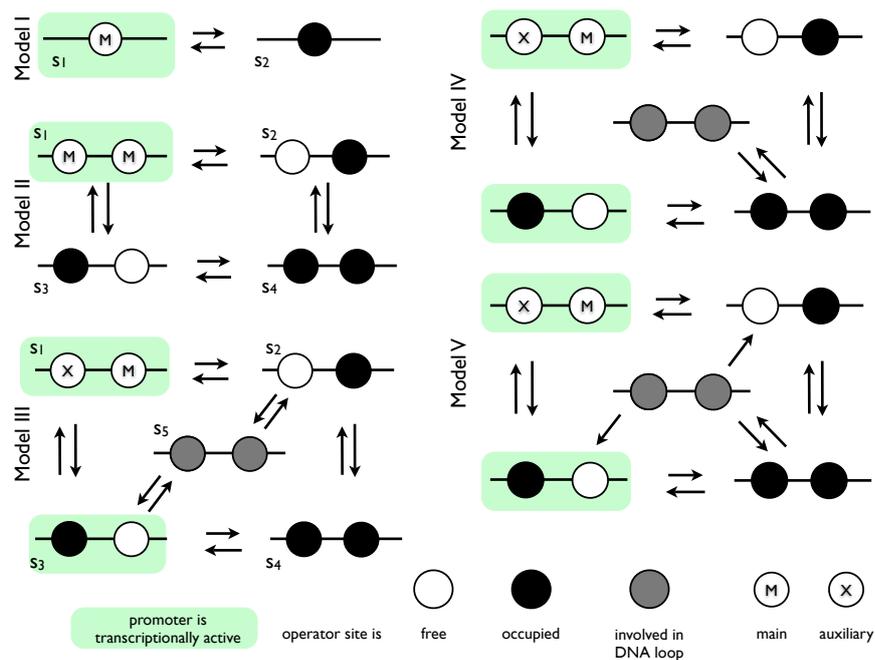
Figure 2.8: Transition graphs for the different *lac* repressor models I, II, III, IV, V (compare figure 2.7). The numbering of states are indicated for models I, II, III. The states in models IV, V are the same as for model III (see definitions (2.14, 2.15, 2.16)). The corresponding transition rates are omitted for clarity and can be inferred from the infinitesimal generators K_i (2.17, 2.18, 2.19, 2.20, 2.21). Also compare the definitions of transition rates in the last section 2.2 and figure 2.5. Operator sites that are free, occupied or involved in a DNA loop are marked in white, black and grey respectively. The main and auxiliary operators are marked by M and X once for each model. States for which transcription occurs are shaded in green.

$$K_1^T = \begin{bmatrix} -x\lambda_0^+ & \lambda_0^- \\ x\lambda_0^+ & -\lambda_0^- \end{bmatrix}, \tag{2.17}$$

$$K_2^T = \begin{bmatrix} -x\lambda_0^+ - x\lambda_1^+ & \lambda_1^- & \lambda_0^- & 0 \\ x\lambda_1^+ & -(x-1)\lambda_0^+ - \lambda_1^- & 0 & \lambda_0^- \\ x\lambda_0^+ & 0 & -\lambda_0^- - (x-1)\lambda_1^+ & \lambda_1^- \\ 0 & (x-1)\lambda_0^+ & (x-1)\lambda_1^+ & -\lambda_0^- - \lambda_1^- \end{bmatrix}, \tag{2.18}$$

$$K_3^T = \begin{bmatrix} -x\lambda_0^+ - x\lambda_1^+ & \lambda_1^- & \lambda_0^- & 0 & 0 \\ x\lambda_1^+ & -(x-1)\lambda_0^+ - \lambda_1^- - a\lambda_0^+ & 0 & \lambda_0^- & b\lambda_0^- \\ x\lambda_0^+ & 0 & -\lambda_0^- - (x-1)\lambda_1^+ & \lambda_1^- & b\lambda_1^- \\ 0 & (x-1)\lambda_0^+ & (x-1)\lambda_1^+ & -\lambda_0^- - \lambda_1^- & 0 \\ 0 & a\lambda_0^+ & a\lambda_1^+ & 0 & -b\lambda_0^- - b\lambda_1^- \end{bmatrix}, \tag{2.19}$$

$$K_4^T = \begin{bmatrix} -x\lambda_0^+ - x\lambda_1^+ & \lambda_1^- & \lambda_0^- & 0 & 0 \\ x\lambda_1^+ & -(x-1)\lambda_0^+ - \lambda_1^- & 0 & \lambda_0^- & 0 \\ x\lambda_0^+ & 0 & -\lambda_0^- - (x-1)\lambda_1^+ & \lambda_1^- & 0 \\ 0 & (x-1)\lambda_0^+ & (x-1)\lambda_1^+ & -\lambda_0^- - \lambda_1^- - \psi_d^+ & \psi_d^- \\ 0 & 0 & 0 & \psi_d^+ & -\psi_d^- \end{bmatrix}, \tag{2.20}$$

$$K_5^T = \begin{bmatrix} -x\lambda_0^+ - x\lambda_1^+ & \lambda_1^- & \lambda_0^- & 0 & 0 \\ x\lambda_1^+ & -(x-1)\lambda_0^+ - \lambda_1^- - a\lambda_0^+ & 0 & \lambda_0^- & b\lambda_0^- \\ x\lambda_0^+ & 0 & -\lambda_0^- - (x-1)\lambda_1^+ - a\lambda_1^+ & \lambda_1^- & b\lambda_1^- \\ 0 & (x-1)\lambda_0^+ & (x-1)\lambda_1^+ & -\lambda_0^- - \lambda_1^- - \psi_d^+ & +\psi_d^- \\ 0 & a\lambda_0^+ & a\lambda_1^+ & \psi_d^+ & -b\lambda_0^- - b\lambda_1^- - \psi_d^- \end{bmatrix}. \tag{2.21}$$

Table 2.2: Infinitesimal generators for the models I, II, III, IV and V of LacI mediated transcriptional repression as depicted in figure 2.8.

respective state s_i at time t, by the Master equation

$$\dot{p}_j = K_i^T p_j.$$

As laid out earlier (compare equation (2.8)) we are interested in the solution of the equation

$$0 = K_i^T p,$$

which is called the invariant measure μ_i of K_i. The invariant measures for the infinitesimal generators K_i are given below. It is convenient to write the invariant measure

$$\mu_i = \frac{v_i}{\sum_j v_i^j},$$

as the normalized form of the vector v_i, where v_i^j is the j-th component of the vector v_i. The vectors v_i are:

$$v_1 = [1,\ x\lambda_0], \tag{2.22}$$

$$v_2 = [1,\ x\lambda_1,\ x\lambda_0,\ x(x-1)\lambda_0\lambda_1], \tag{2.23}$$

$$v_3 = \left[1,\ x\lambda_1,\ x\lambda_0,\ x(x-1)\lambda_0\lambda_1,\ x\frac{a}{b}\lambda_1\lambda_0\right], \tag{2.24}$$

$$v_4 = [1,\ x\lambda_1,\ x\lambda_0,\ x(x-1)\lambda_0\lambda_1,\ x(x-1)\psi_d\lambda_0\lambda_1], \tag{2.25}$$

$$v_5 = \left[1,\ x\lambda_1,\ x\lambda_0,\ x(x-1)\lambda_0\lambda_1 \left[\frac{b\left(\frac{\lambda_0^- + \lambda_1^-}{\psi_d^-}\right)+1}{b\left(\frac{\lambda_0^- + \lambda_1^-}{\psi_d^-}\right)+b\psi_d+1}\right],\right. \tag{2.26}$$

$$\left. x(x-1)\lambda_0\lambda_1 \left[\frac{\psi_d}{b\left(\frac{\lambda_0^- + \lambda_1^-}{\psi_d^-}\right)+b\psi_d+1}\right]\right].$$

In the formulas for v_i the substitutions,

$$\lambda_i = \frac{\lambda_i^+}{\lambda_i^-},\ \psi_d = \frac{\psi_d^+}{\psi_d^-},$$

were made. Note that if we set $b = 0$ in v_5 we retrieve v_4 as expected (Setting $b = 0$ abolishes the alternative loop opening in model V, so it becomes equivalent to model IV, compare figure 2.7).

One consequence of the different loop forming mechanisms of the models can be seen by considering the limit of the invariant measure as the molecule numbers x go

to infinity. These limits $\lim_{x \to \infty} \mu_i = \bar{\mu}_i$ are

$$
\begin{aligned}
\bar{\mu}_1 &= [0,1], \\
\bar{\mu}_2 &= [0,0,0,1], \\
\bar{\mu}_3 &= [0,0,0,1,0], \\
\bar{\mu}_4 &= \frac{1}{1+\psi_d}[0,0,0,1,\psi_d],
\end{aligned}
$$

which means that for high molecule numbers the measure for the tetrameric repressor loop vanishes (model III), while it stays positive for the dimeric repressor loop (models IV and V). The invariatn measure $\bar{\mu}_5$ looks similar to $\bar{\mu}_4$, where $\bar{\mu}_5$ also contains the extra terms in the brackets that appear in the definition of v_5 in equation (2.26)). The states with measure 1 in models 1, 2 and 3 are those where all binding sites are occupied, as should be expected.

Repression levels. The level of transcriptional repression R is typically defined as the fraction of maximal and repressed rates of transcription (see [85]),

$$
R = \frac{T_{max}}{T_{rep}}.
$$

In terms of our model the actual transcription rate equals that of a maximal transcription rate T_{max} times the probability that the gene resides in an active state. So the actual (repressed) transcription rate T_{rep} is described by

$$
T_{rep} = T_{max} \sum_{s_j \in S_{act}^i} \mu_i^j(x),
$$

where $S_{act}^i \subset S_i$ is the subset of states that are transcriptionally active. The transcription is maximal, when there are no repressors around, which is reflected in the equation

$$
T_{max} \sum_{s_j \in S_{act}^i} \mu_i^j(0) = T_{max},
$$

because when $x = 0$, the operator sites are always free and the invariant measure is concentrated in s_1 in all models, $\mu_i^1(0) = 1$, and s_1 is an active state. In sum, the

repression level for model i is of the form

$$R_i = \frac{1}{\sum_{s_j \in S_{act}^i} \mu_i^j}.$$

The repression levels for the models are

$$
\begin{aligned}
R_1 &= 1 + x\lambda_0, & (2.27)\\[4pt]
R_2 &= 1 + (\lambda_0 + \lambda_1)x + x(x-1)\lambda_0\lambda_1, & (2.28)\\[4pt]
R_3 &= \frac{1 + (\lambda_0 + \lambda_1 + \frac{a}{b}\lambda_1\lambda_0)x + \lambda_1\lambda_0 x(x-1)}{1 + \lambda_0 x}, & (2.29)\\[4pt]
R_4 &= \frac{1 + (\lambda_0 + \lambda_1)x + (1+\psi_d)\lambda_0\lambda_1 x(x-1))}{1 + \lambda_0 x}, & (2.30)\\[4pt]
R_5 &= \frac{1 + (\lambda_0 + \lambda_1)x + \dfrac{b\left(\frac{\lambda_0^- + \lambda_1^-}{\psi_d^-}\right) + 1 + \psi_d}{D}\lambda_0\lambda_1 x(x-1))}{1 + \lambda_0 x}, & (2.31)
\end{aligned}
$$

where

$$D = b\left(\frac{\lambda_0^- + \lambda_1^-}{\psi_d^-}\right) + b\psi_d + 1.$$

The number of binding sites and the molecular mechanism of the DNA loop formation is reflected in the structure of the corresponding repression levels and the location of the looping rate constants. The denominators in the fractions correspond to the transcriptionally active states. The order of magnitude of the DNA looping formation parameters a, ψ_d, is typically larger than the rates for *lac* repressor to operator binding [85]. This means that the properties of the repression level will strongly depend where these DNA looping parameters appear. For the repression level of the tetrameric repressor model, R_3, the looping parameters appear in the linear term of the numerator, while for the dimeric repressor models, R_4, R_5 they appear in the quadratic term. The consequence of this is a qualitative difference in the repression level as a function of the number of repressor molecules x.

The repression levels are compared in figure 2.9. The curve for the tetrameric repressor loop model, R_3, has been plotted with parameters from the literature [85], as a benchmark against which to compare the other models. For repression level curves R_1, R_2 the binding constants for the main operators, λ_0, λ_1, were increased to make the repression level for $x = 5$ comparable to R_3. For the dimeric repressor loop model, R_4, the parameter for dimeric loop formation, ψ_d, was varied accordingly.
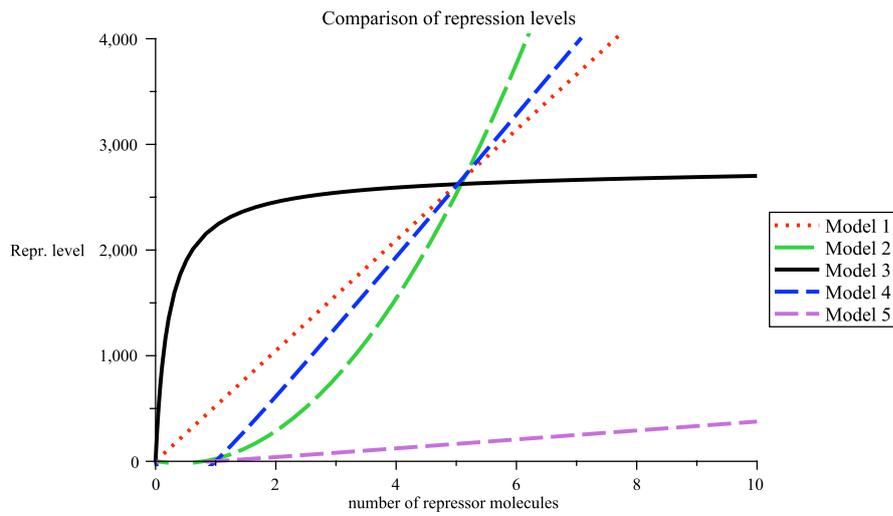
Figure 2.9: Comparison of repression levels R_1, R_2, R_3, R_4, R_5 for the five models for the *lac* repressor (see (2.27, 2.28, 2.29, 2.30, 2.31)). The repression level R_3 was chosen a benchmark and calculated for literature based parameters, $\lambda_0 = \lambda_1 = 4.56, b = 1, a = 596$ [85]. For models without DNA loop formation (R_1, R_2), the binding constants were fitted so that repression levels were comparable for $x = 5$ repressor molecules; for R_1, $\lambda_0 = 523$; for R_2, $\lambda_0 = \lambda_1 = 11$. For the dimeric repressor model, R_4, the dimerization constant ψ_d was the variable parameter. Parameters for R_4, R_5: $\lambda_0 = \lambda_1 = 4.56, \lambda_0^- = \lambda_1^- = \psi_d^- = 0.016\text{s}^{-1}, \psi = 148, b = 0.1$. The unit of the binding constants λ_i and ψ are 'per particle'. The shape of R_3 is determined by its relatively large linear coefficient compared to its smaller quadratic coefficient. The formulas for the other repression levels do not share this behaviour - either their linear and quadratic coefficients are of comparable order of magnitude, or have a larger quadratic coefficient.

Here, a relatively weak loop formation rate sufficed to attain the same repression level as R_3. For R_5, even small values of b, which controls the alternative loop opening mechanism in model V (see figures 2.7, 2.8), decrease the repression level compared to R_4 in such a way, that it can not be compensated for by an increase of the dimeric loop formation constant ψ_d. The reason for this is that positive values for b diminish the impact that the dimeric loop constant has on the repression level.

The repression level curve for tetrameric loop formation, R_3, rises steeply for low values of x, and then flattens out as x increases. In contrast, the curves for all other models retain a relatively steep slope. It has been suggested, that this property makes the repression by tetrameric LacI robust to fluctuations in molecule numbers - in the sense that variations in the number of repressors lead to only small changes in the repression level [85]. The reason that tetrameric repression behaves this way, while dimeric repression does not, lies in the position of the corresponding loop constants in the equation for the repression levels. In R_3, the large linear term lets the repression level rise quickly for low molecule numbers and flatten out as molecule numbers rise, while the large quadratic term in the dimeric repression level, R_4, leads to an approximately linear behaviour.

Another way to illustrate the different behaviours of the models is to look at the typical distribution of the invariant measures on the operator states (figure 2.10). The large loop formation constants increase the probabilities for the Markov chain to reside in the loop states and those that can be directly reached from it. For the tetrameric repressor model (model III) this means that for very low molecule numbers the system is pulled into the loop state and those where only one operator site is occupied. For higher molecule numbers x, the system resides mostly in the state where all operators are occupied - in this range of x the sensitivity of R_3 to changes in x is low. For dimeric repressor (model IV), there are no transcriptionally active states that can be reached from the loop state. This leads to a strong repression, that is sensitive to changes in x. The introduction of the alternative way of loop opening for dimeric repression (model V) makes a transcriptionally active state accessible from the loop state. The consequence of this is a decrease in the repression efficiency.

Generally, it should be recalled that in this theoretical framework an important assumption is a homogeneous spatial distribution of particles, even at low numbers.
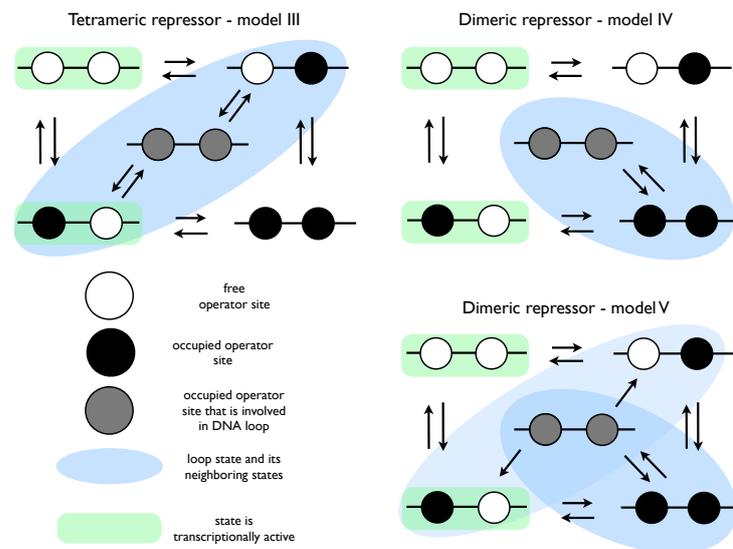
Figure 2.10: Illustration of the effect of the connectivity of the loop state in the models for tetrameric and dimeric repressors. The large constants for loop formation increase the probability that the Markov chain resides in looped states and those that are accessible from it (shaded in blue). In model III the system is pulled towards the tetrameric loop state, while the fully occupied state dominates for higher molecule numbers x. In model IV only the fully occupied state is accessible from the loop state, which yields a strong repression. In model V a transcriptionally active state (shaded in green) can be reached from the looped state - hence the repression efficiency is decreased compared to that of model IV.

That means that diffusion processes are not considered explicitly, but it is assumed that particles are uniformly able to find the binding sites.

The model parameters for the *lac* repressor models were based on the experimental work presented in [55]. Here the authors introduced plasmids carrying *lacZ* genes under the control of the natural *lac* promoter and with different arrangements of *lac* operators, which were introduced into *E.coli* host strains that were *lacZ* and *lacI* deficient. Different numbers of *lac* repressor were provided by additional plasmids. LacZ activity was then measured by standard assays. Such an experimental setup could be used to test further predictions of models like those presented here.

Summary: Analysis of different models for LacI mediated repression

This section presented a number of Markov chain models of the transcriptional inhibition mechanisms of the *lac* repressor. A previously described model of DNA loop formation for the tetrameric repressor [85] was compared to alternative mechanisms based on the dimeric *lac* repressor. For comparison, also models with one or two operator sites without DNA loop formation were discussed. For all models the corresponding infinitesimal generators, invariant measures and repression levels were derived. It was discussed how the different molecular mechanisms are reflected in these expressions.

Because the rate constants for DNA looping are typically much larger than those for normal operator binding, their positions dominate the behaviour of the equation for the repression levels. For dimeric repressors the looped state is accessible only from the state where both two operator sites are occupied, while for tetrameric repressors the looped state is accessible when one of the two operator sites is occupied. Therefore the loop formation constants enter the repression level in front of terms that are either quadratic - for dimeric repressors, or linear - for tetrameric repressors, in the number of molecules x.

Consequently the repression levels for dimeric and tetrameric repressors behave very differently in the context of fluctuating numbers of repressor molecules. The repression curve for tetrameric repressors flattens out when molecule number increase, while the curve for dimeric repressor retains a steep slope. This means that the repression with dimeric loop formation, just as mechanisms without a DNA loop, is more sensitive to variations in the number of molecules than repression that

relies on tetrameric loop formation. This is a general principle of gene regulation mechanisms that involve DNA looping.

Chapter summary

This chapter introduced an algorithm to generate a Markov chain based description of genetic regulatory mechanisms. The algorithm produces a model description which fits into the theoretical framework of molecular systems of finite states and infinite copy numbers as described by Sbano [69], [66]. The algorithm is capable to generate infinitesimal generators for MC models that include multiple genes regulated by multiple operator sites and also takes into account cooperative binding and DNA looping, which were not treated in [69, 66]. The algorithm was applied to compare different DNA looping mechanisms of the *lac* repressor. In particular it was discussed how these mechanisms affect the structure of the corresponding transition graph, its invariant measure and the properties of the repression level in the context of robustness to fluctuations in repressor numbers.

The analysis of the models for the *lac* repressor suggests that dimeric DNA looping does not share the noise robustness properties that tetrameric loop formations confers. This means that if a cell wants to stabilize the level of repression in the presence fluctuating repressor numbers, tetrameric looping should do better than dimeric looping. This result illustrates a principle that distinguishes two general mechanisms for DNA looping. One, a loop between two operator sites is formed by binding of a factor to one site and then binding to the second operator site. Two, two regulating factors bind to two operator sites on the DNA and then form a loop by dimerization. It would be interesting to see if and in which contexts loop formation by dimerization happens *in vivo*. Also, robustness was here used in the sense of the sensitivity of the steady state to parameter changes. In order to extend the analysis, one could also conduct a time-dependent study.

The implementation of the described algorithm represents a step in the effort to facilitate stochastic modelling of genetic regulatory systems. It automatizes the translation of biological models into a mathematical description. This makes it easier and faster to analyse these kind of models. The algorithm as presented still leaves room for improvements and extensions, such as an implementation of the identification of independent submodules as described in section 2.1. Generally it

would be desirable to embed this tool into a larger computational framework, where the constructed models are analyzed further, similar to the integrated modelling tool 'Copasi' for metabolic reaction networks [37].

The next chapter presents an application of the modelling technique described in this chapter to a two gene system that also contains transcriptional regulation by the *lac* repressor. This system is the engineered genetic circuit, that was introduced in chapter 1, and which is inspired by the genetic oscillators that drive circadian clocks.

Chapter 3

Modelling and analysis of the genetic clock

In this chaper, a model for the genetic clock, that was described in chapter 1, is derived and analysed. From Markov chain models for the two promoter regions of the clock, the average dynamics are derived based on the theory described in the previous chapter. The resulting ODE system is analysed in order to investigate its dynamical proterties, in particular the generation of sustained oscillations, which correspond to the existence of a stable limit cycle.

3.1 Derivation of the model for the genetic clock

In order to derive the average dynamics of the detailed model of the genetic regulation for the engineered clock, the Markov chain on the gene states is split into submodules, which are then analysed separately for their invariant measures. From these the invariant measure of the full system can be derived as described in the previous chapter (section 2.1). We illustrate this with an example first, and then continue to derive the model for the full clock circuit.

3.1.1 Average dynamics and decoupling of the Markov chain on the promoter states

In the last chapter, an example of a clock like gene network (example 1, see figure 2.2) was introduced and the graphical representation of its infinitesimal generator K constructed by the described algorithm (figure 2.3). We will conclude the discussion

of that example here by deriving the average dynamics of that system and, in he course of that, illustrate how the Markov chain on the gene states can be decoupled. After that, we will discuss the somewhat more complicated, yet comparable, situation of the engineered clock circuit as introduced in chapter 1 and depicted in figure 3.1.

Example 1, continued

Derivation of the average dynamics.   In the example 1, as depicted in figure 2.2, two genes for proteins X and Y were regulated by three operator sites (compare table 2.1). The activation site o_0 needed to be bound by X for the gene, p_1, for X to be active, site o_1 was a site for the repressor Y, which would inhibit expression of X when bound. Site o_2 was an activating site for the gene p_2. The state of the Markov chain is defined by the triple of the binding states, $s = (o_0, o_1, o_2)$, and the state space given by

$$
\begin{aligned}
S &= \{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)\} \quad (3.1)\\
&= \{s_0, s_1, s_2, \ldots, s_7\}.
\end{aligned}
$$

Further, gene p_1 is active when $s \in S_{p_1} = \{s_4, s_5\}$, and gene p_2 is active when $s \in S_{p_2} = \{s_1, s_3, s_5, s_7\}$.

Recall the infinitesimal generator K for example 1, which is given in table 3.1 (compare its graphical representation in figure 2.3). To be able to define the full master equation,

$$
\frac{\partial P_i(\mathbf{n}, t)}{\partial t} = L_{ii}^*(P_i(\mathbf{n}, t)) + \frac{1}{\epsilon} \sum_{i=0}^{7} K_{ij}^T(\mathbf{n}) P_j(\mathbf{n}, t), \quad (3.2)
$$

we need to specify the birth-death process of the gene products. Transcription and translation are not considered seperately but lumped into one joint expression process. The variable $\mathbf{n} = (n_1, n_2)$ describes the number of particles of type X and Y. The diagonal matrix of difference operators, L^*, can be written as a sum

$$
L^* = L_R^* + L_E^*, \quad (3.3)
$$

where L_R^* is a matrix that describes the reactions for X, Y that do not depend on

the state s of the Markov chain of the genes, and L_E^* describes those reactions, that do [66]. The reactions defining L_R^* correspond here to protein degradation and leaky expression. The proteins are degraded with rates δ_1, δ_2. We consider basal rates of expression β_1, β_2, which describe leaky expression that takes place regardless of the gene state:

$$\emptyset \xrightarrow{\beta_1} X \xrightarrow{\delta_1} \emptyset, \qquad \emptyset \xrightarrow{\beta_2} Y \xrightarrow{\delta_2} \emptyset.$$

In sum this means that $(L_R^*)_{ij} = \delta_{ij} (l_{exp} + l_{deg})$, with

$$l_{deg} = \delta_1 E_{n_1}^{+1} + \delta_2 E_{n_2}^{+1} - (\delta_1 + \delta_2)id,$$

$$l_{exp} = \beta_1 E_{n_1}^{-1} + \beta_2 E_{n_2}^{-1} - (\beta_1 + \beta_2)id,$$

where $E_n^{\pm 1}$ are difference operators acting on a function $f(n)$ as defined by $E_n^{\pm 1} f(n) = f(n \pm 1)$, and δ_{ij} is the Kronecker delta.

The matrix L_E^* describes the regulated gene expression that depends on the activity state of the genes. Molecule X is produced with rate α_1 when the first gene p_1 is active. Molecule Y is produced at rate α_2, when the second gene p_2 is active. This means we have the reactions

$$\emptyset \xrightarrow{\alpha_1} X, \quad s \in S_{p_1} = \{s_4, s_5\}; \qquad \emptyset \xrightarrow{\alpha_2} Y, \quad s \in S_{p_2} = \{s_1, s_3, s_5, s_7\};$$

where S_{p_1}, S_{p_2} are the subsets of the state space S, for which the genes p_1, p_2 are active, respectively. The resulting operator matrix L_E^* is given in table 3.1.

After taking the continuum limit, i.e. transforming the particle numbers n_1, n_2 into the concentrations x, y, the master equation turns into a Fokker-Planck equation, restated here from (2.3),

$$\frac{\partial p_i(\mathbf{x}, t)}{\partial t} = \hat{L}_{ii}^*(p_i(\mathbf{x}, t)) + \frac{1}{\varepsilon} \sum_{j=0}^{7} K_{ij}^T(\mathbf{x}) p_j(\mathbf{x}, t), \tag{3.4}$$

where ε originates from the continuum limit approximation. The matrix operator $\hat{L}^*$ is of the form $\hat{L}^* = \hat{L}_R^* + \hat{L}_E^*$ (compare (3.3)) with

$$\left(\hat{L}_R^* \right)_{ij} = \delta_{ij} \left((\beta_1 - \delta_1)\frac{\partial}{\partial x} + (\beta_2 - \delta_2)\frac{\partial}{\partial y} \right).$$

$$
K^T(\mathbf{n}) =
\begin{pmatrix}
-(n_1\omega_0^+ + n_1\omega_2^+ + n_2\omega_1^+) & \omega_2^- & \omega_1^- & 0 & \omega_0^- & 0 & 0 & 0 \\
n_1\omega_2^+ & -((n_1-1)\omega_0^+ + \omega_2^- + n_2\omega_1^+) & 0 & \omega_1^- & 0 & \omega_0^- & 0 & 0 \\
n_2\omega_1^+ & 0 & -(n_1\omega_0^+ + n_1\omega_2^+ + \omega_1^-) & \omega_2^- & 0 & 0 & \omega_0^- & 0 \\
0 & n_2\omega_1^+ & n_1\omega_2^+ & -(\omega_0^- + (n_1-1)\omega_2^+ + n_2\omega_1^+) & 0 & 0 & 0 & \omega_0^- \\
n_1\omega_0^+ & 0 & 0 & 0 & -(n_1\omega_0^+ + n_1\omega_2^+ + \omega_1^-) & \omega_2^- & \omega_1^- & 0 \\
0 & (n_1-1)\omega_0^+ & 0 & 0 & n_1\omega_2^+ & -(\omega_0^- + n_1\omega_2^+ + \omega_1^-) & 0 & \omega_1^- \\
0 & 0 & (n_1-1)\omega_0^+ & 0 & n_2\omega_1^+ & 0 & -(\omega_0^- + (n_1-1)\omega_2^+ + \omega_1^-) & \omega_2^- \\
0 & 0 & 0 & (n_1-1)\omega_0^+ & 0 & n_2\omega_1^+ & n_1\omega_2^+ & -((n_1-1)\omega_0^+ + \omega_2^- + \omega_1^-)
\end{pmatrix}
$$

$$
L_E^* =
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \alpha_2(E_{n_2}^{-1} - id) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \alpha_1(E_{n_1}^{-1} - id) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \alpha_2(E_{n_2}^{-1} - id) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \alpha_1(E_{n_1}^{-1} - id) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \alpha_2(E_{n_2}^{-1} - id) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \alpha_1 E_{n_1}^{-1} + \alpha_2 E_{n_2}^{-1} - (\alpha_1 + \alpha_2)id & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_2(E_{n_2}^{-1} - id)
\end{pmatrix}
$$

Table 3.1: Matrices K, L_E^* for the Master equation (3.2) for example system 1 (compare its graphical representation in figure 2.3). The infinitesimal generator K (compare figure 2.2). The operator matrix L_E^* with the contributions for protein expression that depend on the activity states of the genes. L_E^* contributes to the definition of L^* (see equation (3.3)) that appears in the Master equation.

$$K^T(\mathbf{x}) = \begin{pmatrix}
-(x\omega_0^+ + x\omega_2^+ + y\omega_1^+) & \omega_2^- & \omega_1^- & 0 & \omega_0^- & 0 & 0 & 0 \\
x\omega_2^+ & -(x\omega_0^+ + \omega_2^- + y\omega_1^+) & 0 & \omega_1^- & 0 & \omega_0^- & 0 & 0 \\
y\omega_1^+ & 0 & -(x\omega_0^+ + x\omega_2^+ + \omega_1^-) & \omega_2^- & 0 & 0 & \omega_0^- & 0 \\
0 & y\omega_1^+ & x\omega_2^+ & -(x\omega_0^+ + \omega_2^- + \omega_1^-) & 0 & 0 & 0 & \omega_0^- \\
x\omega_0^+ & 0 & 0 & 0 & -(\omega_0^- + x\omega_2^+ + y\omega_1^+) & \omega_2^- & \omega_1^- & 0 \\
0 & x\omega_0^+ & 0 & 0 & x\omega_2^+ & -(\omega_0^- + \omega_2^- + y\omega_1^+) & 0 & \omega_1^- \\
0 & 0 & x\omega_0^+ & 0 & y\omega_1^+ & 0 & -(\omega_0^- + x\omega_2^+ + \omega_1^-) & \omega_2^- \\
0 & 0 & 0 & x\omega_0^+ & 0 & y\omega_1^+ & x\omega_2^+ & -(\omega_0^- + \omega_2^- + \omega_1^-)
\end{pmatrix}$$

$$\hat{L}_E^* = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha_2\frac{\partial}{\partial y} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \alpha_2\frac{\partial}{\partial y} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \alpha_1\frac{\partial}{\partial x} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \alpha_1\frac{\partial}{\partial x} + \alpha_2\frac{\partial}{\partial y} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \alpha_2\frac{\partial}{\partial y} & 0
\end{pmatrix}$$

Table 3.2: Matrices K, $\hat{L}_E^*$ for the Focker Plank equation (3.4) for example system 1. These matrices correspond to the matrices of the master equation presented in table 3.1.

The matrices K and $\hat{L}_E^*$ for (3.4) are given in table 3.2. In order to derive the average dynamics, we need to compute the invariant measure of K (see table 3.2). The invariant measure is given by $\mu = \frac{v}{\sum_i v_i}$, where

$$
\begin{aligned}
v \;=\; & (1,\, \omega_2 x,\, \omega_1 y,\, \omega_1 \omega_2 xy,\, \omega_0 x, \\
& \omega_0 \omega_1 x^2,\, \omega_0 \omega_1 xy,\, \omega_0 \omega_1 \omega_2 x^2 y),
\end{aligned}
\tag{3.5}
$$

with the substitutions

$$
\omega_i = \frac{\omega_i^+}{\omega_i^-}, \quad i = 0, 1, 2.
$$

The associated differential equations in terms of the concentrations x, y are

$$
\dot{x} \;=\; \beta_1 + \alpha_1 \sum_{s \in S_{p_1}} \mu_s - \delta_1 x = \beta_1 + \alpha_1 \frac{\omega_0 x}{(1 + \omega_1 y)(1 + \omega_0 x)} - \delta_1 x, \tag{3.6}
$$

$$
\dot{y} \;=\; \beta_2 + \alpha_2 \sum_{s \in S_{p_2}} \mu_s - \delta_2 y = \beta_2 + \alpha_2 \frac{\omega_2 x}{1 + \omega_2 x} - \delta_2 y. \tag{3.7}
$$

Alternative derivation of the invariant measure μ by decoupling of the Markov chain (S, K) into independent submodules. In the equation (3.6) above the sum of invariant measures of the active gene states could be written as a product. This is so, because for this system it is possible to derive the invariant measure by first decoupling the Markov chain into submodules and then solving those separately. As noted in the previous chapter (section 2.1), this procedure makes use of the tensor product (see [48] for the definition and algebraic properties of the tensor product). For one, the binding dynamics for the operator sites o_0, o_2, which molecule X binds to, are independent to those of the operator site o_1, to which molecule Y binds to. Indeed, the generator K can be written as a sum of tensor products:

$$
\begin{aligned}
K \;=\;& I_2 \otimes K_X + K_Y \otimes I_4, \\
=\;& \begin{bmatrix} K_X & 0 \\ 0 & K_X \end{bmatrix} + \begin{bmatrix} -\omega_1^+ y I_4 & \omega_1^+ y I_4 \\ \omega_1^- I_4 & -\omega_1^- I_4 \end{bmatrix}.
\end{aligned}
\tag{3.8}
$$

with I_n being the n-dimensional identity matrix,

$$
K_X = \begin{bmatrix}
-(\omega_0^+ + \omega_2^+)x & \omega_2^+ x & \omega_0^+ x & 0 \\
\omega_2^- & -\omega_0^+ x - \omega_2^- & 0 & \omega_0^+ x \\
\omega_0^- & 0 & -\omega_0^- - \omega_2^+ x & \omega_2^+ x \\
0 & \omega_0^- & \omega_2^- & -(\omega_0^- + \omega_2^-)
\end{bmatrix},
$$

and

$$
K_Y = \begin{bmatrix}
-\omega_1^+ y & \omega_1^+ y \\
\omega_1^- & -\omega_1^-
\end{bmatrix}.
$$

The equality (3.8) holds after an appropriate renumbering of the states, which corresponds to a reordering of rows and columns. The two matrices K_X, K_Y are infinitesimal generators for Markov chains on the two state spaces

$$
S_X = \{(o_0, o_2) \,|\, o_0, o_2 \in \{0,1\}\} = \{(0,0),(0,1),(1,0),(1,1)\} = \{s_0^X, s_1^X, s_2^X, s_3^X\},
$$

$$
S_Y = \{o_1 \,|\, o_1 \in \{0,1\}\} = \{0,1\} = \{s_0^Y, s_1^Y\}.
$$

The corresponding activity of the genes depends on the states as follows. Gene p_1 is active, when the Markov chain (S, K) resides in either one of the states $s \in \{s_4, s_5\} \subset S$ (compare (3.1)). This corresponds to the Markov chain (S_X, K_X) residing in s_2^X or s_3^X, while the Markov chain (S_Y, K_Y) must reside in state s_0^Y. The gene p_2 is active, when the MC (S, K) resides in $s \in \{s_1, s_3, s_5, s_7\} \subset S$. This corresponds to the MC (S_X, K_X) residing in s_1^X or s_3^X, where the MC (S_Y, K_Y) may reside in any state.

Further, since after making the continuum limit we are dealing with concentration and not particle numbers, we even can consider the binding states of the two X specific sites, o_0 and o_2, independently and write K_X as a sum of tensor products:

$$
\begin{aligned}
K_X &= I_2 \otimes K_{X_0} + K_{X_2} \otimes I_2 \\
&= \begin{bmatrix} K_{X_0} & 0 \\ 0 & K_{X_0} \end{bmatrix} + \begin{bmatrix} -\omega_2^+ x I_2 & \omega_2^+ x I_2 \\ \omega_2^- I_2 & -\omega_2^- I_2 \end{bmatrix},
\end{aligned}
\tag{3.9}
$$

where

$$
K_{X_i} = \begin{bmatrix}
-\omega_i^+ x & \omega_i^+ x \\
\omega_i^- & -\omega_i^-
\end{bmatrix}, \quad i = 0, 2.
$$

The equality (3.9), as equality (3.8), holds after an appropriate reordering of rows and columns. We have thus broken down the Markov chain (S, K) into three simple Markov chains, one for each of the binding sites o_0, o_1, o_2. The corresponding invariant measures are

$$\mu_X = \frac{1}{1 + \omega_2 x + \omega_0 x + \omega_0 \omega_2 x^2}(1, \omega_2 x, \omega_0 x, \omega_0 \omega_2 x^2); \tag{3.10}$$

$$\mu_{o_i} = \frac{1}{1 + \omega_i x}(1, \omega_i x), \quad i = 0, 2; \qquad \mu_Y = \mu_{o_1} = \frac{1}{1 + \omega_1 y}(1, \omega_1 y). \tag{3.11}$$

The tensor product of these can be used to find the invariant measure μ of the full system (compare (3.5)) and the one for the MC (S^X, K^X) (compare (3.10)):

$$\mu_X \otimes \mu_Y = ((\mu_X)_1 \mu_Y, \dots, (\mu_X)_4 \mu_Y) = \mu, \tag{3.12}$$

$$\mu_{o_0} \otimes \mu_{o_2} = ((\mu_{o_0})_1 \mu_{o_2}, (\mu_{o_0})_2 \mu_{o_2}) = \mu_X. \tag{3.13}$$

The equations (3.12) and (3.13) hold after an appropriate reindexing. The corresponding terms for the sums of invariant measures in the equations for the average dynamics, (3.6) and (3.7), can be broken down into products of the corresponding terms of the invariant measures of the submodules (equations (3.10),(3.11), (3.12) and (3.13)). For the ODE for x, equation (3.6):

$$(\mu_4 + \mu_5) = \left(\mu_2^X + \mu_3^X\right)\mu_0^Y = \mu_1^{o_0}\left(\sum_i \mu_i^{o_2}\right)\mu_0^Y = \mu_1^{o_0}\mu_0^{o_1} = \frac{\omega_0 x}{(1 + \omega_1 y)(1 + \omega_0 x)}.$$

And for the ODE for y, equation (3.7):

$$(\mu_1 + \mu_3 + \mu_5 + \mu_7) = \left(\mu_1^X + \mu_3^X\right)\left(\sum_i \mu_i^Y\right) = \mu_1^{o_2}\left(\sum_i \mu_i^{o_0}\right) = \mu_1^{o_2} = \frac{\omega_2 x}{1 + \omega_2 x}.$$

We have the following generalization. Assume that instead of three binding sites, we have three groups of binding sites, whose evolution is described by independent Markov chains $(S^1, K^1), (S^2, K^2), (S^3, K^3)$. Let gene p_1 be active if and only if $s' \in S_{act}^1 \subset S^1$ and $s'' \in S_{act}^2 \subset S^2$. Let gene p_2 be active if and only if $s''' \in S_{act}^3 \subset S^3$. Then using the same model and parameters as above, the corresponding average

dynamics for this system can be derived and are of the form

$$\dot{x} = \beta_1 + \alpha_1 \left(\sum_{s' \in S_{act}^1} \mu_{s'}^1 \right) \cdot \left(\sum_{s'' \in S_{act}^2} \mu_{s''}^2 \right) - \delta_1 x \qquad (3.14)$$

$$\dot{y} = \beta_2 + \alpha_2 \left(\sum_{s''' \in S_{act}^3} \mu_{s'''}^3 \right) - \delta_2 y, \qquad (3.15)$$

where $\mu^i, i = 1, 2, 3$ are the invariant measures of the infinitesimal generators $K^i, i = 1, 2, 3$. This decoupling of the Markov chain on the gene states will be used to derive the average dynamics of the engineered circuit, that was described in the introductory chapter. The sum of the invariant measures of the states in $S_{act}^i, i = 1, 2, 3$, will be denoted by the functions $T_{glnA}(x), T_{LacOp}(y)$ and $T_{glnK}(x)$, respectively, indicating their respective correspondence to the *glnAp2* promoter, the *lac* operator sites and the *glnKp* promoter. The dependence on either x or y indicates that the corresponding operator sites are only targeted by either the transcription factor NRI or the repressor LacI.

The Markov chain of the engineered clock

The model of the engineered circuit features seven (including the NRI governor sites) binding sites for NRI, and two binding sites for the *lac* repressor, distributed onto two promoter regions (see figure 3.1). To derive the average dynamics for this system, we will decouple its infinitesimal generator K into submodules as illustrated for the example system 1. In particular, we can again separate the dynamics for the LacI specific binding sites from those, that are specific to NRI. Also, when we apply the continuum approximation, we can separate the binding dynamics for the two promoter regions, *glnAp2* and *glnKp*.

The *lac* repressor binding dynamics have already been discussed in the last chapter, so we can work with the invariant measures derived there. Next we will look into the NRI mediated activation of transcription.

3.1.2 Transcriptional activation

This section deals with the modelling of transcriptional activation by the transcription factor NRI. NRI dependent expression rates of the two promoters *glnAp2* and
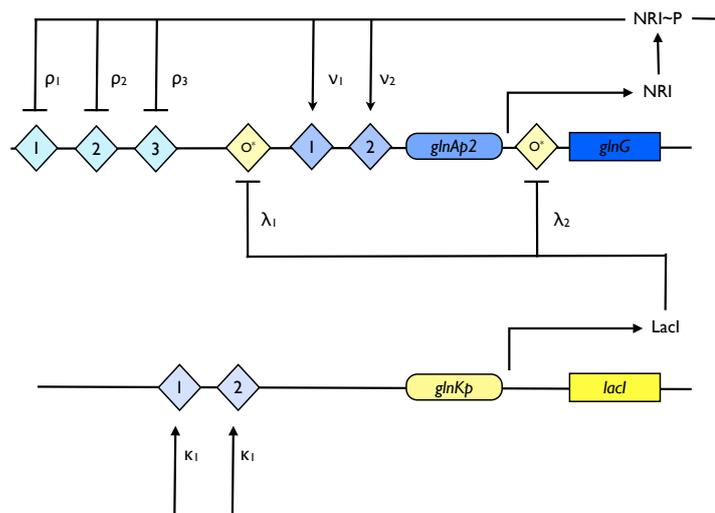
Figure 3.1: Diagram of the regulatory interactions and operator sites in the synthetic clock, compare figure 1.7 in chapter 1. The diagram shows the two gene modules and indicates the binding sites for NRI and LacI. Names for the binding parameters for all binding sites are indicated.

glnKp were derived, based on the invariant measures of corresponding Markov chain models. Parameter ranges were discussed based on published experimental data. For comparison, different models are considered as shown in figure 3.2. Model I, where the gene activity is governed by only one binding site. Model II, where the gene activity is governed by two binding sites and binding happens cooperatively. And a third model for the *glnAp2* promoter, model III, where, in addition to the two enhancer sites, the gene activity is reduced by NRI binding to governor sites at high concentrations.

As described in chapter 1, NRI activates transcription by interacting with the σ^{54}-RNA polymerase, after it has bound to the promoter site, and thereby mediating the transition from the closed into the open transcription complex by the formation of a DNA loop [77]. For the model we neglect the details of this interaction and also the dynamics of the RNA polymerase, but assume that transcription takes place when NRI is bound to the enhancer sites. Experimental data suggest that a certain degree of dimer- or oligomerization is required before NRI can activate transcription [59]. In the model this is reflected insofar, that gene expression is active only when all enhancer sites of the given promoter are occupied. Further, NRI needs to be phosphorylated to activate transcription, and also shows stronger binding activity
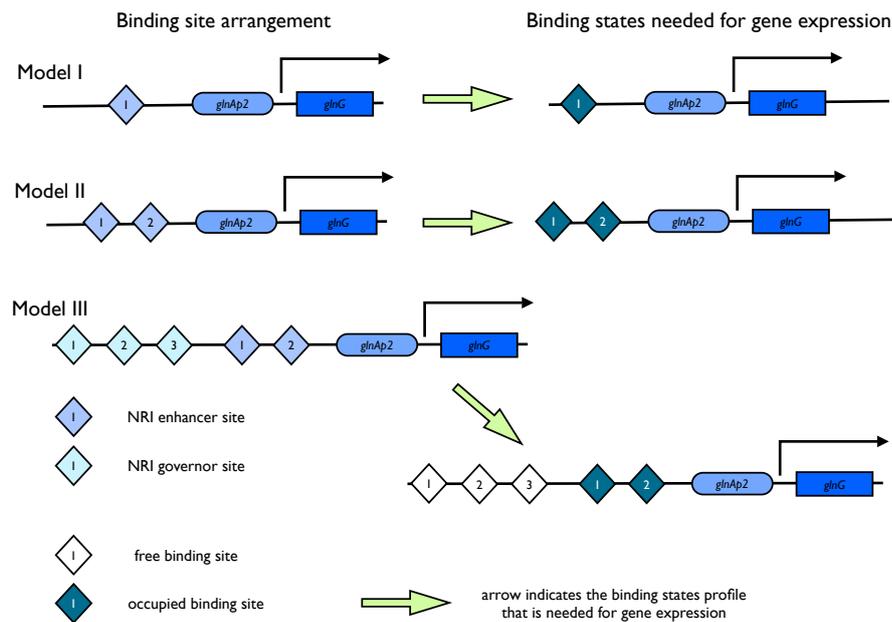
Figure 3.2: Illustration of the three models of NRI mediated activation of gene expression. Model I has only one enhancer site that needs to be occupied for gene expression. Model II features two NRI enhancer sites that both need to be occupied for gene expression and to which NRI binds cooperatively (see text). Model 3 treats the also includes the NRI governor sites, which need to be free for gene expression to occur. NRI molecules also bind cooperatively to neighbouring governor sites. The illustrations are for the situation of the *glnAp2* promoter driven activator module of the engineered clock. Models I and II also apply to the repressor module, which is driven by the *glnKp* promoter, compare figure 3.1.

in its phosphorylated form [87]. As described in chapter 1, the host strain of the engineered clock carries a mutated NRI kinase, which leads to high degrees of NRI phosphorylation [8]. Therefore the phosphorylation dynamics are neglected in the model and NRI is assumed to be mostly phosphorylated.

The construction of the model, and the determination of corresponding parameter ranges, rely on a number of experimental studies of NRI dependent transcriptional activation, which investigated the binding site occupation and transcriptional activation as a function of NRI concentration for different binding site arrangements [4, 7, 54, 72, 87]. In [72, 87], the authors measured the binding affinity of NRI to the enhancer sites of the *glnAp2* promoter and their role in transcriptional activation. In [4], the transcriptional activation of the *glnKp* promoter for different NRI concentrations is compared with activation of the *glnAp2* promoter. The effect of the governor sites of the *glnAp2* promoter was studied in [7, 54]. Concerning the significance of these data it needs to be noted that they were collected under the specific conditions of the given experiment - for example pH value, concentrations of ATP and RNA polymerases or the phosphorylation state of NRI, which differ to varying degrees from the biological conditions *in vivo*. Also, the parameter estimates vary between experiments. Together, this means that the published data can only serve to identify feasible parameter ranges for modelling the engineered circuit, not to yield exact values. Also, it has been reported from scanning force microscopy studies, that NRI forms oligomers up to the size of an octameric structure [89], for which detailed experimental data is lacking, and which is also neglected in the models discussed here.

Figure 3.1 shows the location of the NRI binding sites in the engineered circuit. The *glnAp2* promoter of the activator module features two enhancer sites that activate transcription and three governor sites that inhibit transcription at high NRI concentrations. We consider three models for the activity of the activator module (see figure 3.2). First, a simplified model where transcription is activated by only one binding site (model I). Second, a model for the cooperative binding of NRI to the two enhancer sites (model II). And thirdly, a model that combines the cooperative activation of transcription with the inhibition mediated by the governor sites (model III). For the *glnKp* driven repressor module we consider two models, with one or two enhancer sites (corresponding to models I and II in figure 3.2). The jus-

tifications for the one-enhancer-site model of transcriptional activation are, for one, to illustrate the possible design choices for changing the present engineered clock, or for manufacturing similar gene circuits - indeed, one enhancer site is sufficient to allow for NRI mediated activation of gene expression [87] - and secondly, as an approximation for the cooperative model.

Model I - one enhancer site.   When the expression of the gene is ruled by the binding state of just one enhancer site, the state space is two dimensional, $s \in S_1 = \{0, 1\}$, and transcription is active for $s = 1$. When ν^+, ν^- denote the binding and unbinding rates for that enhancer site, the transpose of the infinitesimal generator of the Markov chain and the corresponding invariant measure are given by

$$
K_1^T = \begin{bmatrix} -x\nu^+ & \nu^- \\ x\nu^+ & -\nu^- \end{bmatrix}, \qquad \mu^I = \frac{1}{1 + \nu x}(1, \nu x), \qquad \frac{\nu^+}{\nu^-} = \nu.
$$

The corresponding rate of transcriptional activity is

$$
T_{glnAp}^1(x) = \mu_2^I(x) = \frac{\nu x}{1 + \nu x}. \tag{3.16}
$$

Model II - cooperative binding to two enhancer sites.   When the expression of the gene is ruled by two binding sites o_1, o_2, to which NRI binds cooperatively, the state space is four-dimensional,

$$
s = (o_1, o_2) \in S = \{(0, 0), (0, 1), (1, 0), (1, 1)\}. \tag{3.17}
$$

If the corresponding binding rates are denoted by $\nu_i^+, \nu_i^-, i = 1, 2$, and the cooperativity constant by ψ, then the transpose of the infinitesimal generator of the Markov chain of the gene and its invariant measure are given by

$$
K_2^T = \begin{bmatrix} -x\nu_1^+ - x\nu_2^+ & \nu_2^- & \nu_1^- & 0 \\ x\nu_2^+ & -(x-1)\psi\nu_1^+ - \nu_2^- & 0 & \nu_1^- \\ x\nu_1^+ & 0 & -\nu_1^- - (x-1)\psi\nu_2^+ & \nu_2^- \\ 0 & (x-1)\psi\nu_1^+ & (x-1)\psi\nu_2^+ & -\nu_1^- - \nu_2^- \end{bmatrix},
$$

$$\mu^{II} = \frac{v}{\sum_i v_i}, \qquad v = (1, \nu_2 x, \nu_1 x, \psi \nu_1 \nu_2 x^2). \tag{3.18}$$

If the gene is expressed only when both enhancer sites are occupied, then the corresponding rate of transcriptional activity is

$$T^2_{glnAp}(x) = \mu^{II}_4(x) = \frac{\psi \nu_1 \nu_2 x^2}{1 + (\nu_1 + \nu_2)x + \psi \nu_1 \nu_2 x^2}. \tag{3.19}$$

A note on the Hill equation. When the cooperativity constant ψ is large compared to the sum of the dissociation constants $\frac{1}{\nu_i}$, such that $\frac{\nu_1 + \nu_2}{\psi \nu_1 \nu_2} = \frac{\frac{1}{\nu_1} + \frac{1}{\nu_2}}{\psi} \approx 0$, then equation (3.19) can be approximated by a Hill equation,

$$T^2_{glnAp}(x) \approx \frac{x^2}{k + x^2}, \qquad k = \frac{1}{\psi \nu_1 \nu_2}, \tag{3.20}$$

that attains its half maximum for $x = \sqrt{k}$. This means that, in this situation, the activation threshold is determined by k, which is the inverse of the product of the binding coefficients and the cooperativity constant.

Model III - enhancer and governor sites. For the governor sites, we assume that binding to neighbouring sites occurs cooperatively with the same cooperativity constant ψ as for the enhancer sites. Since we are treating the amount of NRI as a concentration x, we can decouple the Markov chain for the *glnAp2* promoter into a Markov chain for the enhancer sites, treated in model II, and a Markov chain for the governor sites. The function for transcriptional activity is then derived by taking products of the appropriate invariant measures of these two Markov chains. The state space for the MC of the governor sites is eight-dimensional, with $s = (o_1, o_2, o_3)$, and

$$
\begin{aligned}
s \in S \;=\;& \{0,1\}^3 \\
=\;& \{(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1)\}.
\end{aligned}
\tag{3.21}
$$

The infinitesimal generator K can be quickly constructed by the algorithm described in the previous chapter using the input parameters: opsites x x x, coop 1 1 0, coop 0 1 1; and is therefore omitted here. Its invariant measure is given by

$$\mu^{gov} = \frac{v}{\sum_i v_i}, \; v = (1, \rho_3 x, \rho_2 x, \psi \rho_2 \rho_3 x^2, \rho_1 x, \rho_1 \rho_3 x^2, \psi \rho_1 \rho_2 x^2, \psi^2 \rho_1 \rho_2 \rho_3 x^3). \tag{3.22}$$

The invariant measure for model III is given by the tensor product $\mu^{III} = \mu^{II} \otimes \mu^{gov}$. Consequently, the transcriptional activity for the $glnAp2$ promoter including the enhancer and governor sites is now given by the product of respective invariant measures for the enhancer sites μ^{II} in equation (3.18), and for the governor sites μ^{gov} in equation (3.22):

$$T^3_{glnAp2}(x) = \mu^{II}_4(x)\mu^{gov}_1(x) = \frac{\psi\nu_1\nu_2 x^2}{1 + (\nu_1 + \nu_2)x + \psi\nu_1\nu_2 x^2} \times \qquad (3.23)$$

$$\frac{1}{1 + (\sum_{i=1}^3 \rho_i)x + (\psi(\rho_1\rho_2 + \rho_2\rho_3) + \rho_1\rho_3)x^2 + \psi^2\rho_1\rho_2\rho_3 x^3}.$$

Here it is assumed that binding to any governor site inhibits transcription, so that transcription can take place only when all governor sites are free. This corresponds to the state $s_1 = (0,0,0)$ in S as defined in (3.21) of the Markov chain for the governor sites, which has the corresponding invariant measure $\mu^{gov}_1(x)$ defined in (3.22). The factor μ^{II}_4 is the invariant measure defined in (3.18) of s_4 in S defined in (3.17).

Models for the $glnKp$ promoter. The models for transcriptional activity of the $glnKp$ promoter are of the same form as those for the $glnAp2$ promoter, safe the governor sites (models I and II, compare figure 3.2). Thus, when κ and κ_1, κ_2 denote the binding constants for the one and two enhancer site models, the corresponding transcriptional activity functions are

$$T^1_{glnKp}(x) = \frac{\kappa x}{1 + \kappa x}, \qquad (3.24)$$

$$T^2_{glnKp}(x) = \frac{\psi\kappa_1\kappa_2 x^2}{1 + (\kappa_1 + \kappa_2)x + \psi\kappa_1\kappa_2 x^2}. \qquad (3.25)$$

Figure 3.3 compares the transcriptional activation profiles for the $glnAp2$ and $glnKp$ promoters for literature based parameters. Transcriptional activation of both promoters occurs for NRI concentrations in the nano-molar region. The activation threshold for $glnKp$ is higher than that for $glnAp2$, because one of its binding sites has a weaker affinity for NRI [4]. When NRI concentrations rise above the $glnKp$ activation threshold, the $glnAp2$ activity is reduced again due to the effect of the governor sites, which have a very low affinity for NRI [7]. Generally, the binding constants for the strong NRI sites, like the enhancer sites for $glnAp2$ promoter, are

of the order of $10^{-1} \ldots 10^{1}$ per particle (for a bacterial cell, one particle corresponds to a concentration of about 0.1nM to 1nM), and the cooperativity is estimated to be of the order $10 \ldots 1000$. In wild type *E.coli* typical NRI concentrations range in the order of $10^{-1} \ldots 10^{2}$nM [72, 87].
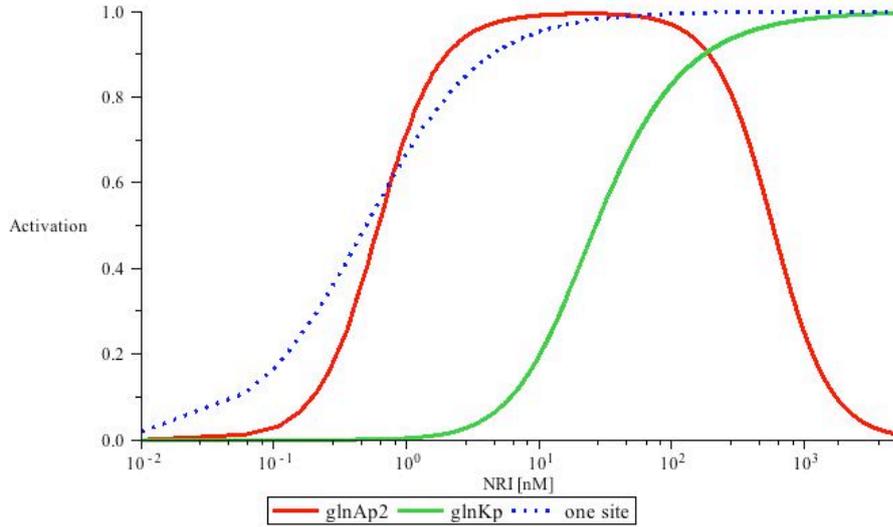

Figure 3.3: Comparison of transcriptional activation mediated by NRI. The *glnAp2* promoter has two relatively strong enhancer sites, while the *glnKp* promoter has a strong and a weak enhancer site, leading to a higher activation threshold for the *glnKp* promoter [4]. At even higher NRI concentrations the *glnAp2* activity is reduced due to the effect of the low affinity governor sites [7]. Parameters are $\nu_1 = \nu_2 = 0.1\text{nM}^{-1}, \psi = 550, \rho_1 = 10^{-4}\text{nM}^{-1}, \rho_2 = 10^{-6}\text{nM}^{-1}, \rho_3 = 10^{-10}\text{nM}^{-1}$ for model T^3_{glnAp2} as defined in (3.23); $\nu = 2\text{nM}^{-1}$ for the single site model T^1 as in equation (3.16) (one site); and $\kappa_1 = 0.1\text{nM}^{-1}, \kappa_2 = 10^{-4}\text{nM}^{-1}$ for the T^2_{glnKp} model for the *glnKp* promoter, see (3.25).

3.1.3   Average dynamics of the clock

Here we assemble the average dynamics of the engineered circuit, using the invariant measures, that were derived for the NRI dependent transcriptional activation and LacI mediated repression of transcription. The decoupling of the Markov chain for the states of the activator and repressor modules of the clock circuit can be performed similarly as for the example system 1 (see text around equations (3.14, 3.15)). Each operator site in example 1 corresponds to a submodule of multiple operator sites in the engineered clock: site o_0 to the NRI mediated regulation of the *glnAp2* promoter, site o_2 to the NRI mediated activation of the *glnKp* promoter, and site o_1 to the LacI mediated repression of the *glnAp2* promoter. Because the *glnAp2*

promoter is only active when both Markov chains for NRI and LacI regulation are in the active states, their respective measures enter the average dynamics of the clock as a product.

Assembling the equations for the NRI and LacI based regulatory interactions, we arrive at the deterministic differential equations for the whole circuit, which describe the time evolution of the concentrations of the NRI, denoted by x, and of LacI proteins, denoted by y (compare the ODE system (3.14, 3.15)):

$$\dot{x} \;=\; \beta_1 + \alpha_1 T_{glnAp}(x) T_{LacOp}(y) - \delta_1 x, \tag{3.26}$$

$$\dot{y} \;=\; \beta_2 + \alpha_2 T_{glnKp}(x) - \delta_2 y, \tag{3.27}$$

where δ_i are the rates of protein degradation, α_i are the maximal expression rates for activated promoters, and β_i are the rates of basal expression, that takes place independently of the promoter state. For the derivation of the average dynamics, the corresponding operator matrix $\hat{L}^*$ was constructed according to the same principles as the one in example 1 (compare table 3.2). The functions $T_{glnAp}, T_{LacOp}, T_{glnKp}$ can be one of the functions derived earlier in this chapter - or derived from those in the previous chapter, in the case for T_{LacOp}. More precisely, T_{glnAp} may be defined as in (3.16), (3.19) or (3.23); T_{glnKp} may be defined as in (3.24) or (3.25). The function T_{LacOp} can be one of the following two choices

$$T^1_{LacOp}(y) \;=\; \frac{1}{1 + y\lambda_0}, \tag{3.28}$$

$$T^2_{LacOp}(y) \;=\; \frac{1 + \lambda_0 y}{1 + (\lambda_0 + \lambda_1 + \frac{a}{b}\lambda_1\lambda_0)y + \lambda_1\lambda_0 y^2}. \tag{3.29}$$

These two functions are derived from the repression levels of the models for the *lac* repressor, that were discussed in the last chapter and stated in equations (2.27) and (2.29). For simplicity we only consider two models for LacI mediated repression (models I and III) out of the five discussed in chapter 2 (see table 2.7). The parameters λ_i are the binding constants for the two *lac* operator sites and a, b are the rates describing the formation and opening of the DNA loop.

A parameter estimation for protein production. The birth-death process for the gene products, as in example 1, is treated as a joint protein expression process,

where transcription and translation are lumped together. The literature typically states parameter estimates for either transcription or translation. In order to motivate how these parameters are combined to yield estimates for the rate of a lumped protein production rate, we quickly discuss the consequence of making a quasi-steady state assumption on the mRNA levels in a simple ODE system describing transcription or a gene and translation of its mRNA transcripts into protein.

Consider the expression of a single gene subject to regulation described by the function T. The number of mRNA transcripts R is described by a concentration r. The number of proteins P is described by a concentration p. We can write down the ad hoc system of differential equations:

$$\dot{r} \ = \ \beta_r + \alpha_r T - \delta_r r, \tag{3.30}$$

$$\dot{p} \ = \ \alpha_p r - \delta_p p, \tag{3.31}$$

where α_r, α_p denote transcription and translation rates, δ_r, δ_p denote degradation rates for mRNA and protein, and β_r denotes leaky transcription that takes place independent of the value of the regulatory function T. In bacteria, there is typically little time delay between transcription and translation - ribosomes may start translating from the mRNA transcripts even while they are being produced by the polymerases. Also, mRNA transcripts typically have a much shorter half-life than proteins, and transcription is a fast process relative to translation. This motivates a quasi–steady-state assumption for the mRNA dynamics, $\dot{r} = 0$. The equilibrium value of r can be computed from equation (3.30):

$$r_{eq} = \frac{\beta_r + \alpha_r T}{\delta_r}.$$

We plug this into the equation for p, (3.31), and get

$$\dot{p} \ = \ \frac{\alpha_p \beta_r}{\delta_r} + \frac{\alpha_p \alpha_r}{\delta_r} T - \delta_p p \tag{3.32}$$

$$= \ \beta + \alpha T + \delta_p p, \tag{3.33}$$

with obvious definitions for α, β.

Feasible parameter ranges were estimated from expression studies of the *lac* operon [39] and other literature on gene expression [78, 57]. Assuming the typical

lifetimes of mRNA transcripts - 2min, and proteins - 1h, the corresponding degradation rates can be derived

$$\delta_p = \frac{\log 2}{3600s} \approx 0.0002s^{-1}, \qquad \delta_r = \frac{\log 2}{120s} \approx 0.006s^{-1}.$$

Estimates for the *lac* operon put mRNA production at $\alpha_r \approx 10^{-3} \ldots 10^{-2}s^{-1}$, with about 5 to 40 proteins produced from every mRNA transcript, which puts the rate of protein production in the order of $\alpha_p \approx 10^{-2} \ldots 10^{-1}s^{-1}$. This yields a lumped expression rate of $\alpha \approx 10^{-1} \ldots 10^{0}s^{-1}$. These parameters of course depend on the exact conditions for the gene in question - like the position on the chromosome or plasmid, accessibility for polymerases, and the overall cell state - availability of ATP, nucleotides, amino acids, or the number of ribosomes.

Summary: Derivation of the model for the genetic clock

To illustrate the construction of a model for expression of multiple genes, the derivation of the average dynamics for the system of example 1, that was introduced in the previous chapter, was discussed. It was shown in detail how the Markov chain of the gene states can be decoupled into submodules, consisting of Markov chains of interdependent groups of operator sites. The invariant measure of the full model can then be derived from those of the submodules with the help of the tensor product.

In order to derive the average dynamics of the engineered clock, the Markov chain was decoupled into NRI mediated regulation of the *glnAp2* and *glnKp* promoters and LacI mediated repression of the *glnAp2* promoter. Functions for NRI dependent transcriptional activity were derived for when the promoter is regulated by one enhancer site, for when it is regulated by two enhancer sites, to which NRI binds cooperatively, and also for when transcriptional activity is reduced by NRI governor sites. Feasible parameter ranges were discussed based on data from the literature. The transcriptional activation functions were compared to illustrate the different activation thresholds for *glnAp2* and *glnKp* promoters.

Finally the average dynamics of the engineered clock were constructed by assembling the transcriptional activation functions from the different submodules into a system of two coupled differential equations, that describe the time evolution of the concentration of the proteins NRI and LacI. Corresponding feasible parameter

ranges for protein production and degradation were discussed based on related literature. The conversion of transcription and translation rates into a lumped protein production rate was motivated by making a quasi-steady-state assumption on the mRNA levels. Having established the overall form of the average dynamics of the clock, the next section will deal with a general analysis of its phase plane by studying the parameter dependent shapes of its nullclines.

3.2   Model analysis: nullclines and stability

In the context of genetic oscillators it is of interest which are the essential features that drive oscillations. This is important for understanding circadian clocks in general and in particular for designing artificially engineered circuits. In this section, the model derived for the engineered clock, that was described in chapter 1, is investigated analytically using the theoretical tools for ordinary differential equations, in particular with respect to the effect of different choices in the arrangement of the regulatory operator sites. In the course of further work, model predictions could be tested by site-directed mutagenesis of the clock's promoter regions.

In a two dimensional system of ordinary differential equations, like the system (3.26, 3.27), information on its dynamical properties, such as steady states or conditions that facilitate the occurrence of limit cycles, can be found by looking at the system's nullclines. Nullclines are the curves in the phase plane, on which the time derivative of one dependent variable is zero. Here it is investigated how the structure of the equations determines the shape of the nullclines, in order to derive conditions that allow for Hopf bifurcations.

Nullcline for LacI.   The nullcline for the concentration of LacI (see figure 3.4) is defined by the solution curve of the equation $\dot{y} = 0$. Setting $\dot{y} = 0$ in the equation (3.27) yields the formula for the nullcline as a function of the NRI concentration x:

$$y = h(x) = \frac{\beta_2 + \alpha_2 T_{glnKp}(x)}{\delta_2}, \tag{3.34}$$

where for any given NRI concentration x, $h(x)$ yields the corresponding LacI concentration y, such that LacI degradation balances the production of LacI governed by the NRI responsive $glnKp$ promoter. Thus the nullcline is determined by the

choice for $T_{glnKp}(x)$, which is out of (3.24) and (3.25). For both choices $T_{glnKp}(x)$ is a monotonically increasing function, and so is $h(x)$. Characteristic values for the nullcline like $h(0), \lim_{x\to\infty} h(x)$ are

$$h(0) = \frac{\beta_2}{\delta_2}, \quad \lim_{x\to\infty} h(x) = \frac{\beta_2 + \alpha_2}{\delta_2},$$

and independent of the choice of $T_{glnKp}(x)$. These values reflect the ratios between the rate of LacI degradation and the minimal and maximal LacI production rates, respectively. The value x, for which the nullcline attains its half maximum, is defined by the equation $T_{glnKp}(x) = \frac{1}{2} - \frac{\beta_2}{2\alpha_2}$, which reduces to $T_{glnKp}(x) \approx \frac{1}{2}$, when $\beta_2 \approx 0$. So the half maximum for $h(x)$ is close to the half maximum for $T_{glnKp}(x)$, which is given by

$$x = \frac{1}{\kappa}, \quad \text{or} \quad x \approx \frac{1}{\sqrt{\psi\kappa_1\kappa_2}},$$

for $T_{glnKp}(x)$ as in (3.24) or in (3.25), respectively; for the approximation compare equation (3.20). In particular the nullcline has no critical points and is defined for all nonnegative x.
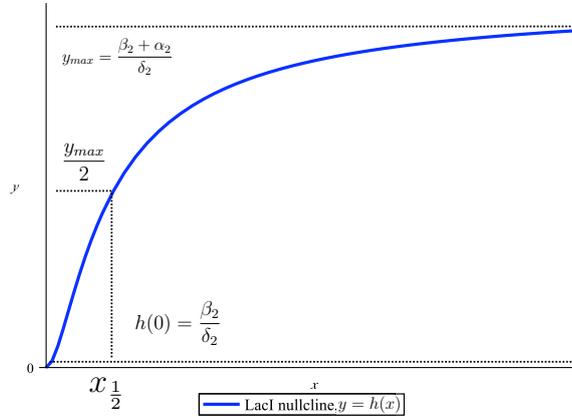

Figure 3.4: Nullcline $h(x)$ for equation (3.27) for the concentration y of the *lac* repressor. The supremum of the nullcline is $y_{max} = \lim_{x\to\infty} h(x)$. The symbol $x_{\frac{1}{2}}$ indicates where the nullcline attains half of its supremum, $h(x_{\frac{1}{2}}) = \frac{y_{max}}{2}$, and is $x_{\frac{1}{2}} \approx \frac{1}{\kappa}$ for T^1_{glnKp}, as in (3.24), or $x_{\frac{1}{2}} \approx \frac{1}{\sqrt{\psi\kappa_1\kappa_2}}$ for T^2_{glnKp}, as in (3.25), for typical parameter values (see text).

Nullcline for NRI.   The nullcline g for NRI is defined by setting $\dot{x} = 0$ in the equation (3.26). This nullcline can also be written as a function of x,

$$y = g(x) = T_{LacOp}^{-1}\left(z(x)\right), \quad z(x) = \frac{-\beta_1 + \delta_1 x}{\alpha_1 T_{glnAp}(x)}, \tag{3.35}$$

where $T_{LacOp}^{-1}(z)$ is defined for $0 < z \leq 1$, which in particular implies $x > \frac{\beta_1}{\delta_1}$. So for each NRI concentration x, the function $g(x)$ gives the corresponding concentration y of LacI, which represses the NRI production governed by the *glnAp2* promoter, such that NRI production and degradation are in balance. The lower boundary for the definition interval for $g(x)$ reflects the fact, that the basal production of NRI cannot be balanced by NRI degradation at low NRI concentrations. The function $z(x)$ is a measure of how much the *glnAp2* promoter needs to be repressed in order to balance NRI production and degradation, where $z = 1$ corresponds to no repression, and $z = 0$ to full repression of the *glnAp2* promoter. The inverse $T_{LacOp}^{-1}(z(x))$ yields the corresponding concentration of repressor molecules. The inverses of $T_{LacOp}(y)$, as given in the equations (3.28) and (3.29), are:

$$T_{LacOp_1}^{-1}(z) \;=\; \frac{1 - z}{\lambda_2 z}, \tag{3.36}$$

$$T_{LacOp_2}^{-1}(z) \;=\; \frac{-[-\lambda_1 + \tilde{\lambda} z] + \sqrt{[-\lambda_1 + \tilde{\lambda} z]^2 + 4\lambda_1 \lambda_2 (1 - z)z}}{2\lambda_1 \lambda_2 z}, \tag{3.37}$$

with

$$\tilde{\lambda} = \lambda_1 + \lambda_2 + \frac{a}{b}\lambda_1 \lambda_2.$$

Note that the inverse $z \mapsto T_{LacOp}^{-1}, (0, 1] \to [0, \infty)$ is a strictly monotone and decreasing function in z for both (3.36) and (3.37). This means, that when $z(x)$ increases up to 1, $g(x)$ will decrease and reach 0. For higher values of z, $g(x)$ is not defined. Other characteristics, like the number of critical points, depend on the choice of the regulatory function T_{glnAp} and the presence of leaky gene expression, described by the rate constant β_1.

On the slope of the NRI nullcline.   A positive slope of the NRI nullcline $g(x)$ means, that an increase in the concentration of NRI needs to be compensated with a corresponding increase of the concentration of the repressor in order to maintain a steady state. A negative slope means, correspondingly, that an increase in the

concentration of NRI needs to be compensated by a decrease in the concentration of repressors. Here we investigate the conditions for the slope of the nullcline to be positive, negative or zero. Generally, by the chain rule we have

$$\frac{d}{dx}g(x) = \frac{d}{dx}T^{-1}_{LacOp}(z(x)) = \underbrace{\left(T^{-1}_{LacOp}\right)'(z(x))}_{<0} z'(x), \tag{3.38}$$

and consequently

$$g'(x) = 0 \iff z'(x) = 0, \tag{3.39}$$

and the sign of the derivative of $g(x)$ depends only on the derivative of $z(x)$. The derivative of $z(x)$ takes the form

$$z'(x) = \frac{\delta_1 T_{glnAp} - (\delta_1 x - \beta_1)T'_{glnAp}(x)}{\alpha_1\left(T_{glnAp}(x)\right)^2}. \tag{3.40}$$

The nullcline $g(x)$ has a nonnegative slope for $z'(x) \leq 0$. This can be written as a fraction of derivatives

$$z'(x) \leq 0 \iff \frac{T'_{glnAp}(x)}{T_{glnAp}(x)} \geq \frac{\delta_1}{\delta_1 x - \beta_1}, \tag{3.41}$$

or, equivalently,

$$\frac{d}{dx}\log\left(T_{glnAp}(x)\right) \geq \frac{d}{dx}\log\left(\delta_1 x - \beta_1\right). \tag{3.42}$$

The NRI nullcline for a single enhancer site at the *glnAp2* promoter. Assume that the NRI mediated activation is described by T^1_{glnAp} as in (3.16), i.e. NRI acts via a single enhancer site. Then the nullcline $g(x)$ has no critical points and is strictly monotone decreasing (see figure 3.5). For T^1_{glnAp} we have from equations (3.35) and (3.40) that

$$z(x) = \frac{\delta_1 \nu x + (\delta_1 - \beta_1 \nu) - \frac{\beta_1}{x}}{\alpha_1 \nu}, \quad z'(x) = \frac{\delta_1 \nu + \frac{\beta_1}{x^2}}{\alpha_1 \nu} > 0,$$

and hence $g'(x) < 0$ for all $x \in D_g = (x_0, x_1]$ for that $g(x)$ is defined. When $\beta_1 = 0$, then $x_0 = 0$ and $\lim_{x \to x_0} g(x) = T^{-1}_{LacOp}\left(\frac{\delta_1}{\alpha_1 \nu}\right)$. Further $x_1 = \frac{\alpha_1}{\delta_1} - \frac{1}{\nu}$, and $g(x_1) = 0$. When $\beta_1 > 0$ then $x_0 = \frac{\beta_1}{\delta_1}$ and $\lim_{x \to x_0} g(x) = \infty$. The value for x_1 increases with β_1.

This means, that when the *glnAp2* promoter is driven by only one enhancer site,

an increase in the concentration of NRI is to be compensated with a decrease in the concentration of repressor in order to maintain a steady state.
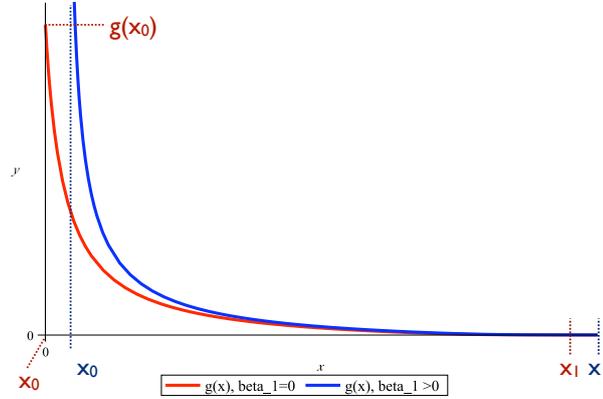

Figure 3.5: Nullcline $g(x)$ for equation (3.26) for the concentration x of NRI, when the regulation at the $glnAp2$ promoter is driven by a single enhancer site, i.e. T_{glnAp} as in (3.16). The nullcline is defined on $x \in (x_0, x_1]$, strictly monotone decreasing in x with $g(x_1) = 0$. For $\beta_1 = 0$, $x_0 = 0$ and $\lim_{x \to x_0} g(x) = T_{LacOp}^{-1}\left(\frac{\delta_1}{\alpha_1 \nu}\right)$; and $x_1 = \frac{\alpha_1}{\delta_1} - \frac{1}{\nu}$. When $\beta_1 > 0$, then $x_0 = \frac{\beta_1}{\delta_1}$ and $\lim_{x \to x_0} g(x) = \infty$, and x_1 increases with β_1.

The NRI nullcline for two enhancer sites at the *glnAp2* promoter ($\beta_1 = 0$).
Assume that the NRI mediated activation of the $glnAp2$ promoter is described by T_{glnAp}^2 as in (3.19), i.e. NRI acts via two enhancer sites. First consider the case when $\beta_1 = 0$, that is, when there is no basal production of NRI. The shape of the nullcline is depicted in figure 3.6. The nullcline $g(x)$ is defined on the interval $D_g = [x_0, x_1]$, with $g(x_0) = g(x_1) = 0$, where x_0, x_1 are the solutions of $z(x) = 1$:

$$x_{0,1} = \frac{1}{2}\left[\left(\frac{\nu_1 + \nu_2}{\psi \nu_1 \nu_2} + \frac{\alpha_1}{\delta_1}\right) \pm \sqrt{\left(\frac{\nu_1 + \nu_2}{\psi \nu_1 \nu_2} + \frac{\alpha_1}{\delta_1}\right)^2 - \frac{4}{\psi \nu_1 \nu_2}}\right], \qquad (3.43)$$

or, when $\frac{\nu_1 + \nu_2}{\psi \nu_1 \nu_2} \approx 0$ (compare equation (3.20)) then

$$x_{0,1} \approx \frac{1}{2}\left[\frac{\alpha_1}{\delta_1} \pm \sqrt{\left(\frac{\alpha_1}{\delta_1}\right)^2 - \frac{4}{\psi \nu_1 \nu_2}}\right].$$

Further, solving $z'(x) = 0$ using equation (3.41), we find by (3.39) that the nullcline $g(x)$ has exactly one local maximum at the position

$$x_{max} = \sqrt{\frac{1}{\psi \nu_1 \nu_2}}, \tag{3.44}$$

which is equal to the approximate x value, for which $T_{glnAp}^2(x)$ attains its half maximum (see equation (3.20)). Therefore $T_{glnAp}^2(x_{max}) \approx \frac{1}{2}$, and

$$g(x_{max}) \approx T_{LacOp}^{-1} \left(\frac{2\delta_1}{\alpha_1 \psi \nu_1 \nu_2} \right). \tag{3.45}$$

This means that for very low NRI concentrations, $x < x_0$, the production of NRI via the *glnAp2* promoter is too weak to be able to compensate for the NRI degradation and hence no steady state can exist. At intermediate concentrations, $x_0 < x < x_{max}$, an increase in x increases the NRI production rate more strongly than the degradation rate. Hence more repressors are needed to balance them, and the nullcline increases. For higher NRI concentrations, $x_{max} < x < x_1$, the opposite is true and less repressors are needed when x increases, and the nullcline decreases correspondingly. When $x > x_1$, then the rate of NRI degradation can no longer be compensated by NRI production, and hence the nullcline is no longer defined.
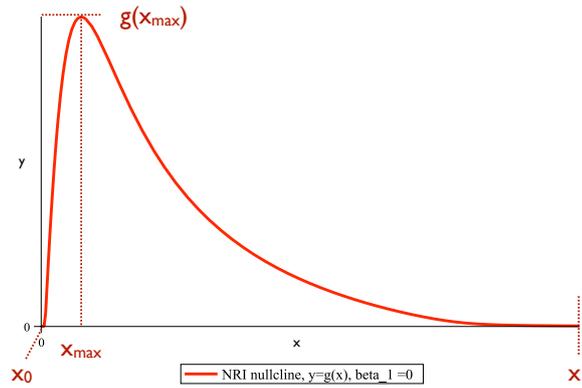
Figure 3.6: Nullcline $g(x)$ for equation (3.26) for the concentration x of NRI, when the regulation at the *glnAp2* promoter is driven by two enhancer sites, i.e. T_{glnAp} as in (3.19). The nullcline $g(x)$ is defined on the interval $D_g = [x_0, x_1]$, with $g(x_0) = g(x_1) = 0$. The nullclines has exactly one local maximum at $x_{max} = \sqrt{\frac{1}{\psi \nu_1 \nu_2}}$ (equation (3.44)). Formulas for x_0, x_1 (equation (3.43)) and $g(x_{max})$ (equation (3.45)) are given in the text.

The NRI nullcline for two enhancer sites at the *glnAp2* promoter ($\beta_1 > 0$).

Let again T^2_{glnAp} be defined as in (3.19) and $\beta_1 > 0$. Then $g(x)$ is defined on some interval $D_g = [x_0, x_1]$, where $x_0 = \frac{\beta_1}{\delta_1}$ and $\lim_{x \to x_0} = \infty$, and $g(x_1) = 0$, where x_1 solves $z(x) = 1$. Solving $z'(x) = 0$ for $T^2_{glnAp}(x)$, using (3.41), yields a cubic equation,

$$\delta_1 \psi \nu_1 \nu_2 x^3 + (\beta_1(\nu_1 + \nu_2) - \delta_1)x + 2\beta_1 = 0. \tag{3.46}$$

Note that for $\beta_1 = 0$ the equation becomes quadratic and allows for one positive solution, given in (3.44). For $\beta_1 > 0$, we make use of the fact that equation (3.46) lacks a quadratic term and is therefore of the form:

$$x^3 + 3px + 2q = 0, \quad p = \frac{\beta_1(\nu_1 + \nu_2) - \delta_1}{3\delta_1 \psi \nu_1 \nu_2}, \quad q = \frac{\beta_1}{\delta_1 \psi \nu_1 \nu_2}.$$

(Actually, all cubic equations can be transformed into this form by an appropriate variable substitution.) This equation admits three real solutions if $D = q^2 + p^3 < 0$, and one real and two complex solutions when $D > 0$. Because of the shape of $g(x)$ this means that the nullcline either has two or no critical points for $D > 0$ or $D < 0$ respectively. Explicitly $D < 0$ reads

$$\left(\frac{\beta_1}{\delta_1 \psi \nu_1 \nu_2}\right)^2 + \left(\frac{\beta_1(\nu_1 + \nu_2) - \delta_1}{3\delta_1 \psi \nu_1 \nu_2}\right)^3 < 0, \tag{3.47}$$

which implies $\beta_1(\nu_1 + \nu_2) - \delta_1 < 0$. When we approximate $\frac{(\nu_1 + \nu_2)}{\psi \nu_1 \nu_2} \approx 0$ as in (3.20), then (3.47) reduces to

$$\left(\frac{\beta_1}{\delta_1}\right)^2 - \frac{1}{27\psi \nu_1 \nu_2} < 0.$$

This means that the occurrence of two critical points is facilitated, when the basal transcription rate β_1 is small in relation to the degradation rate δ_1 for NRI, and when the cooperativity constant γ and the binding constants ν_i are not too large. Figure 3.7 illustrates the effect of varying β_1 on the NRI nullcline. The occurrence of two critical points in the NRI nullcline for $\beta_1 > 0$ is necessary for the existence of a stable limit cycle, as discussed further below.

LacI repression and governor sites have little effect on the NRI nullcline.

Assume that T_{glnAp} is as in (3.23), i.e. the reduction of transcription by the NRI governor sites at high concentration of NRI is considered. When the concentration
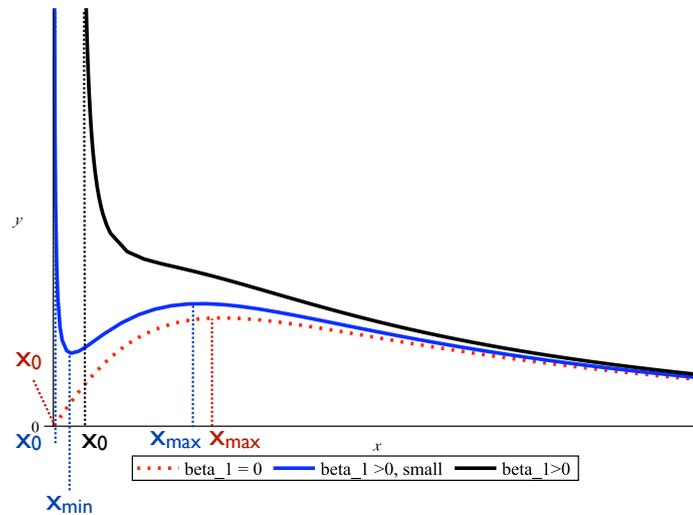
Figure 3.7: Nullcline $g(x)$ for equation (3.26) for the concentration x of NRI, when the regulation at the *glnAp2* promoter is driven by two enhancer sites, i.e. T_{glnAp} as in (3.23), for different values of β_1. For $\beta_1 = 0$, the nullcline is as in figure 3.6. For $\beta_1 > 0$ and small, the nullcline has two critical points, which disappear as β_1 increases. Generally, $g(x)$ has two critical points when condition (3.47) is satisfied. For $\beta_1 > 0$, $g(x)$ is defined on the interval $D_g = (x_0, x_1]$, with $x_0 = \frac{\beta_1}{\delta_1}$ and $\lim_{x \to x_0} g(x) = \infty$. For small β_1, x_{max} in blue is close to the corresponding x_{max} in red; x_{min} in blue is close to the value x_0 in red (compare figure 3.6).
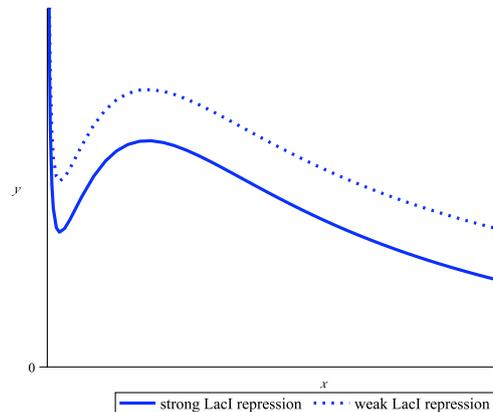
Figure 3.8: Nullcline $g(x)$ for equation (3.26) for the concentration x of NRI, when the regulation at the *glnAp2* promoter is driven by two enhancer sites, i.e. T_{glnAp} as in (3.23) (Compare figure 3.7). Impact of different choices for LacI repression T_{LacOp} is shown. An decrease in the repression efficiency, by changing LacI binding or loop formation parameters, leads to a corresponding increase in the function values of $g(x)$. The qualitative shape of $g(x)$, i.e. the number and position of the critical points, is not affected by the choice of T_{LacOp}.

x is large, so that $T'_{glnAp}(x) \leq 0$ due to the effect of the governor sites (see figure 3.3), then the nullcline cannot have a critical point for that value x. Indeed, because of (3.40) and (3.38) we have that

$$T'_{glnAp}(x) \leq 0 \Rightarrow z'(x) > 0 \Rightarrow g'(x) < 0.$$

Also, the choice of the function for LacI mediated repression, T_{LacOp}, does not affect the qualitative picture for $g(x)$, but, as follows from equation (3.35), only affects its function values as illustrated in figure 3.8. A weaker repression by LacI increases the function values of $g(x)$, while a more effective repression by LacI decreases them.

Nullclines and linear stability. The steady states of the ODE system (3.26, 3.27) are defined by the intersection points of its nullcines $h(x)$ and $g(x)$ (equations (3.34) and (3.35)). In order to identify conditions for the occurrence of limit cycles and Hopf bifurcations, the linear stability properties of the fixed points are investigated. This is done by linear stability analysis, which is based on standard results of ODE theory (see for example the book by Amann [2]). In particular, we are interested in conditions, that can be formulated in terms of the slopes of the nullclines.

The linear stability properties of a steady state are determined by the Jacobian J matrix, which describes the properties of the linearization of the vector field around the steady state (x, y):

$$J = \begin{bmatrix} \alpha_1 T_{LacOp} \frac{dT_{glnAp}}{dx} - \delta_1 & \alpha_1 T_{glnAp} \frac{dT_{LacOp}}{dy} \\ \alpha_2 \frac{dT_{glnKp}}{dx} & -\delta_2 \end{bmatrix}. \qquad (3.48)$$

More precisely, the linear stability properties are determined by the eigenvalues $e_1, e_2 \in \mathbb{C}$ of the matrix J. The eigenvalues are characterized by the existence of a nonzero vector $v_i \in \mathbb{R}^2$ such that $Jv_i = e_i v_i$. Information on the eigenvalues is contained within the trace and determinant of J,

$$tr(J) = e_1 + e_2, \quad det(J) = e_1 e_2.$$

A sufficient condition for a steady state to be unstable is $tr(J) > 0$. Recall that necessary conditions for a Hopf bifurcation are $tr(J) = 0$ and $det(J) > 0$, and that

for sufficiency the derivative of the real part of the eigenvalues with respect to the bifurcation parameter must not be zero.

Explicitly, the equation $tr(J) \geq 0$ reads

$$\alpha_1 T_{LacOp} T'_{glnAp} \geq \delta_1 + \delta_2.$$

Applying the steady state condition $\dot{x} = 0$ to equation (3.26) and plugging this in yields

$$\log' (T_{glnAp}) \geq \left(1 + \frac{\delta_2}{\delta_1}\right) \log' (\delta_1 x - \beta_1). \tag{3.49}$$

Comparing this to equation (3.42) reveals, that the slope for the NRI nullcline $g(x)$ needs to be sufficiently positive for a Hopf bifurcation to occur, or an unstable node or spiral to exist. The condition on the steepness of the slope is weakened when $\frac{\delta_2}{\delta_1}$ is small, i.e. when NRI has a shorter lifetime than the *lac* repressor.

Another necessary condition for a Hopf bifurcation is, that $det(J) > 0$. The sign of the determinant can also be connected with the slope of the nullclines. It holds that

$$det(J) > 0 \iff g'(x) < h'(x). \tag{3.50}$$

Corresponding relations also hold for $det(J) = 0$ and $det(J) < 0$. Indeed, consider the determinant of J, which can be computed from (3.48):

$$det(J) = \delta_2 \left(\delta_1 - \alpha_1 T_{LacOp}(y) T'_{glnAp}(x)\right) - \alpha_1 \alpha_2 T_{glnAp}(x) T'_{glnKp}(x) T'_{LacOp}(y). \tag{3.51}$$

Plugging (3.51) into $det(J) > 0$, using the steady state condition $\dot{x} = 0$ applied to equation (3.26), and rearranging yields

$$\frac{\delta_1 T_{glnAp}(x) - (\delta_1 x - \beta_1) T'_{glnAp}(x)}{\alpha_1 \left(T_{glnAp}(x)\right)^2} > \frac{\alpha_2 T'_{glnKp}(x) T'_{LacOp}(y)}{\delta_2}.$$

Dividing by $T'_{LacOp}(y)$, which is negative, applying the inverse function theorem, and using the equations for the nullclines (3.34), (3.35) and (3.38), we arrive at

$$g'(x) < h'(x),$$

as required.

So if the LacI nullcline has a steeper slope than the NRI nullcline at the intersection point, then $det(J) > 0$, otherwise $det(J) < 0$. If $det(J) > 0$, then condition (3.49) on the steepness of the $g(x)$ decides whether that steady state is stable, unstable, or a candidate of a Hopf bifurcation. If $det(J) < 0$, then the steady state is a saddle point.

Summary: Model analysis - nullclines and stability

This section presented a discussion of the qualitative picture of the two nullclines, $\dot{x} = 0$ and $\dot{y} = 0$, of the clock model defined by equations (3.26) and (3.27). The nullcline $\dot{y} = 0$, parametrized by $h(x)$, is a strictly monotone increasing function for any of the choices for T_{glnKp} out of (3.24), (3.25), i.e. whether the promoter is driven by one or two enhancer sites.

The qualitative properties of the nullcline $\dot{x} = 0$, parametrized by $g(x)$, depend mainly on the number of enhancer sites at the *glnAp2* promoter and the parameter for leaky gene expression for that promoter, β_1. When the activity of *glnAp2* is driven by just one enhancer site (T_{glnAp} as in (3.16)), then $g(x)$ is strictly monotone decreasing. When the activity of *glnAp2* is driven by two enhancer sites (T_{glnAp} as in (3.19)), then $g(x)$ may have up to two critical points. When $\beta_1 = 0$, then $g(x)$ has one local maximum. When $\beta_1 > 0$ then $g(x)$ may have up to two critical points depending on the condition (3.47). Changes in the efficiency of LacI mediated repression of the *glnAp2* promoter (choice of T_{LacOp}) only changes the overall y values of the nullcline, but not the number or position of the critical points. Considering the *glnAp2* governor sites (T_{glnAp} as in (3.19)) also does not change the qualitative picture for g.

The maximum of the LacI nullcline is determined by the maximal expression and degradation rates of LacI. It attains its half maximum for when T_{glnKp} reaches its half maximum. The position of the critical points of the NRI nullcline are mainly determined by the product of cooperativity and binding coefficients for NRI, $\psi\nu_1\nu_2$. The y values for these points depend on the effectiveness of repression and hence of LacI binding and loop formation parameters.

Further, the slope of the NRI nullcline is determined by the relation of the logarithmic derivative of transcriptional activation, T_{glnAp}, and degradation (see equation (3.42)). Linear stability analysis was performed to derive necessary and

sufficient conditions on the slope of the NRI and LacI nullclines at a steady state for it to be stable, unstable or a saddle. The conditions also characterize candidates for Hopf bifurcation points. This proves to be helpful to identify nullcline scenarios that allow for instability and oscillations as discussed next.

3.3   Examples: bifurcations and oscillations

This section will discuss, how oscillations come about in the derived clock model and how this depends on the location of the nullcines. First, an example is given for the existence of a limit cycle that occurs for biologically feasible parameters. Second, it is discussed how parameter dependent changes in the nullcline shapes push the system out of the oscillatory regime. Third, the role of the different activation thresholds of the $glnAp2$ and $glnKp$ promoters with respect to oscillations is discussed.

3.3.1   An example for a stable limit cycle

In the last section it was shown, that if the nullclines intersect with sufficiently steep and positive slopes such that $g'(x) < h'(x)$ and (3.49) hold, then the intersection point is an unstable node or spiral. We can do more than this and show, that if the nullclines intersect only once, i.e. if there is only one unstable steady state and $\beta_1, \beta_2 > 0$, then the system features a stable limit cycle.

The proof relies on the Poincaré-Bendixon theorem, which is a standard result of ODE bifurcation theory (see for example the book of Hale and Kocak [31]). In short, the theorem states that if, for a two-dimensional continuously differentiable vector field, such as (3.26, 3.27), there is a closed and bounded subset $R \subset \mathbb{R}^2$, and a trajectory that is confined in R for all time, then this trajectory is, or spirals into, a closed orbit. The last condition is typically verified by showing, that the vector field on the boundary of R points inwards. That is why R is also called a trapping region.

So let (x^*, y^*) be the only steady state of the system (3.26, 3.27) and unstable. Because $g(x)$ is large for x close to x_0 (compare figure 3.7), the condition (3.50) is fullfilled at the steady state. Therefore it cannot be a saddle. We define R by setting

$$R = \left\{ (x,y) \mid 0 \le x \le \frac{\beta_1 + \alpha_1}{\delta_1}, \ 0 \le y \le \frac{\beta_2 + \alpha_2}{\delta_2} \right\} \setminus U_\epsilon(x^*, y^*), \qquad (3.52)$$

where $\epsilon > 0$, and $U_\epsilon(x^*, y^*)$ is a sufficiently small neighborhood around the steady state, such that the vector field on its boundary points into R. We check the other boundary sides of R in turn (see figure 3.9). Recall the definition of the vector field in x, y, which is given by (3.26) and (3.27), Then on side I, we have

$$x = 0 \Rightarrow \dot{x} = \beta_1 + \alpha_1 T_{glnAp}(0) T_{LacOp}(y) - \delta_1 \cdot 0 = \beta_1 > 0,$$

and for side II,

$$x = \frac{\beta_1 + \alpha_1}{\delta_1} \Rightarrow \dot{x} = \beta_1 + \alpha_1 \underbrace{T_{glnAp}(x) T_{LacOp}(y)}_{<1} - \delta_1 x < \beta_1 + \alpha_1 - \delta_1 \frac{\beta_1 + \alpha_1}{\delta_1} = 0,$$

which means that trajectories on the left and right boundary of R point inwards. Concerning the lower and upper sides (III and IV) we have

$$y = 0 \Rightarrow \dot{y} = \beta_2 + \alpha_2 T_{glnKp}(0) - \delta_2 \cdot 0 = \beta_2 > 0,$$

and

$$y = \frac{\beta_2 + \alpha_2}{\delta_2} \Rightarrow \dot{y} = \beta_2 + \alpha_2 \underbrace{T_{glnKp}(x)}_{<1} - \delta_2 y < \beta_2 + \alpha_2 - \delta_2 \frac{\beta_2 + \alpha_2}{\delta_2} = 0,$$

which means that trajectories on the lower and upper boundaries also point inwards. So all trajectories on the boundary of R point inwards. From the Poincaré-Bendixon it follows then that R contains a stable limit cycle, which attracts all trajectories in R.

An example for a system with a stable limit cycle is given in figure 3.10, which is realized with biologically feasible parameters, as estimated earlier. As is suggested by condition (3.49) the occurrence of a stable limit cycle is facilitated when NRI degrades faster than LacI, as is the case for the example in figure 3.10. The exact values for these degradation rates are not known, but the failure of the actual engineered clock to produce sustained oscillations might be due to the fact the fraction of NRI and LacI lifetimes is not small enough.

Another way to understand the generation of oscillations for this nullcline scenario is based on a separation of time-scale argument and is illustrated in figure 3.11. Assume that x moves on a fast time-scale, while y moves on a relatively slow
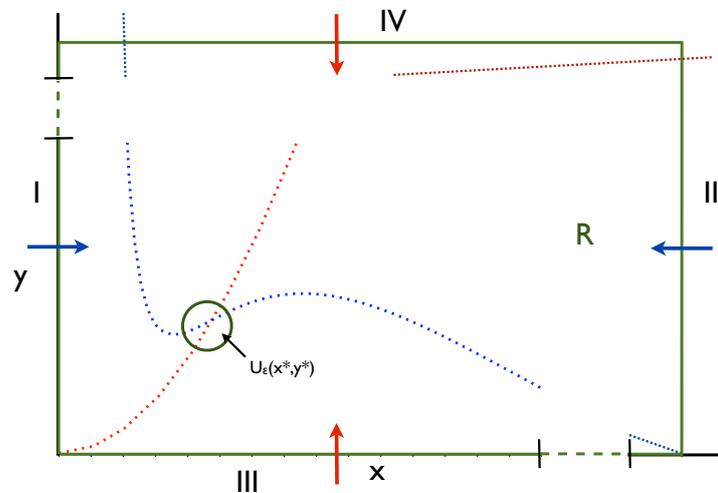
Figure 3.9: Illustration of the trapping region R for when the vector field (3.26), (3.27) has exactly one unstable steady state. The region R, whose boundary is indicated in green, is defined in (3.52), and used to prove the existence of a closed orbit using the Poincaré-Bendixon theorem, see text. The four sides of the box shaped region are labelled I, II, III and IV. Blue and red arrows indicate the direction of the vector field on the boundary side in the x and y directions, respectively. The NRI and LacI nullcilnes are indicated by blue and red dotted curves, whose intersection marks the position of the steady state (x^*, y^*). The dashed intervals indicate that the diagram is not to scale.
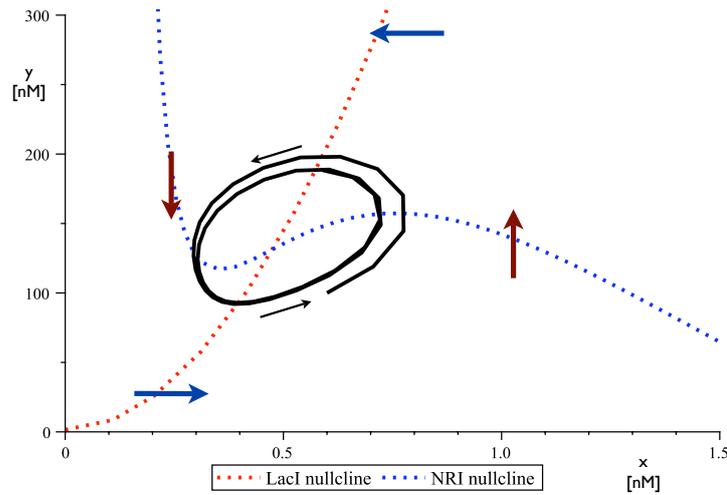
Figure 3.10: Example for a stable limit cycle in the clock model for biologically feasible parameters. The figure shows the phase plane for the ODE system (3.26, 3.27), with the functions $T_{glnAp}, T_{glnKp}, T_{LacOp}$ as in (3.19), (3.25), (3.29). Arrows indicate the flow of the vector field in the phase plane. Red arrows indicate the sign of $\dot{y}$ on the $\dot{x} = 0$ nullcline, and blue arrows the sign of $\dot{x}$ on the $\dot{y} = 0$ nullcline, by pointing in the corresponding direction of the vector field. The NRI and LacI nullclines are shown in blue and red dotted lines. The stable limit cycle is shown in black with arrows indicating the sense of direction that the trajectories follow. Parameters are: $\nu_1 = \nu_2 = \kappa_1 = 0.1, \kappa_2 = 10^{-3}, \psi = 100, \lambda_1 = \lambda_2 = 5, a = 500, b = 1, \beta_1 = 3 \cdot 10^{-4}, \beta_2 = 1.5 \cdot 10^{-4}, \alpha_1 = 12, \alpha_2 = 6, \delta_1 = 0.002, \delta_2 = 0.0001$. Binding parameters are given in nM^{-1}, gene expression parameters in nM·s^{-1}, degradation rates in s^{-1}.

time-scale. Then a stable limit cycle exists that is driven by fast and slow phases of the trajectory. Starting from position 1 in the phase diagram, as indicated in the figure 3.11, the concentration of x increases until it hits the $\dot{x} = 0$ nullcline, which is dotted in blue (fast phase, indicated by a blue arrow). The trajectory then crawls backwards along that nullcline (slow phase, indicated by a red arrow), until it falls off it at position 3, to be caught again by the left branch of the $\dot{x} = 0$ nullcline at position 4. From there the trajectory moves slowly downwards to again fall off the blue nullcline at position 1, where the cycle starts anew.
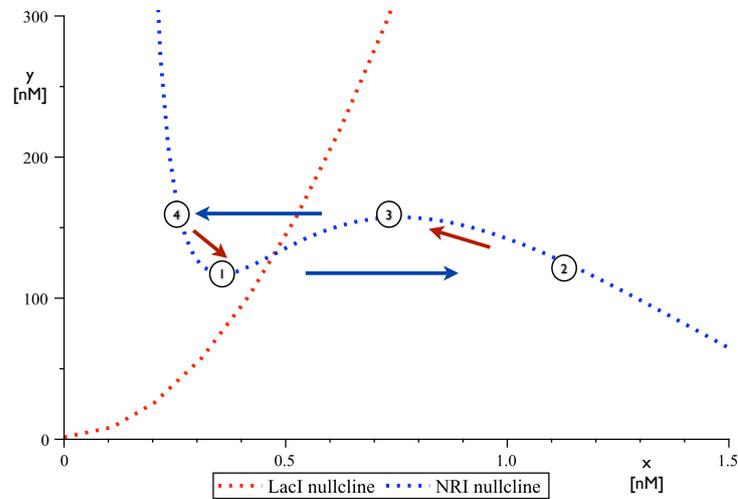
Figure 3.11: Illustration of a stable limit cycle of the system (3.26, 3.27), that is driven by slow and fast phases of the trajectory. The variable x moves on a fast time scale compared to the slower variable y. Nullclines for x and y are shown as blue and red dotted curves, respectively. A trajectory traverses the limit cycle as follows. Starting from position 1 in the phase diagram, x increases until it hits the blue $\dot{x} = 0$ nullcline (fast phase, indicated by a blue arrow). The trajectory then crawls backwards along that nullcline (slow phase, indicated by a red arrow), until it falls off it at position 3, to be caught again by the left branch of the $\dot{x} = 0$ nullcline at position 4. From there the trajectory moves slowly downwards to again fall off the blue nullcline at position 1, where the cycle starts anew.

Sample trajectories for parameters as in figure 3.10 are shown in figure 3.12. The period of oscillations is about 10h, which is comparable to the period found in the experiments. The experimental data also suggest that the period of the oscillations scales with the growth rate of the cells (see chapter 4). This was also observed in previous experiments [8]. This dependency could be explained by assuming that a slower cell growth correlates with slower gene expression and lower degradation

rates due to less dilution. The figure 3.12 shows how the period of the oscillation increases after the reduction of gene expression and degradation rates.
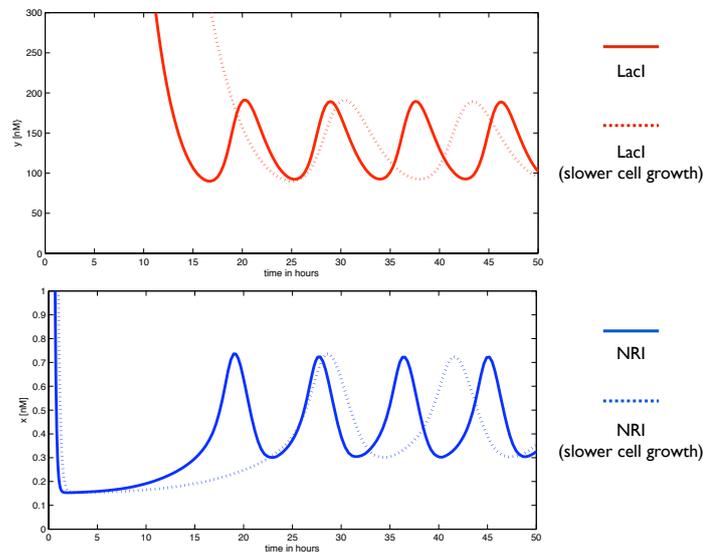

Figure 3.12: Sample trajectories for engineered clock model with parameters as in figure 3.10. Large values for the initial condition $(x_0, y_0) = (100, 10000)$ were chosen to reflect the situation, where the clock is started by the activation of the LacI repressor. After a short lag phase, where proteins degrade, the trajectories approach the limit cycle that is illustrated in figure 3.10. The dotted line indicate trajectories, where the ODE equations have been scaled by the factor 0.75 to reflect slower gene expression and degradation due to a slower cell growth. The dotted curve shows an increased period of about 15h.

Nullcline based sensitivity analysis

This section illustrates how the position of the two nullclines changes with respect to different parameters. The corresponding curves are collected in figures 3.13, 3.14, 3.15 and 3.16. Apart from the oscillatory behaviour, the system may also have one single stable steady state, or exhibit bistability, as illustrated in figures 3.17 and 3.18.

When the curve of the nullcline changes, so does the position of the fixed point. What this means for the stability of the fixed point depends on the circumstances, but a general overview on how small changes in parameters affect the nullcline picture as portrayed in figure 3.10, is given in the figures 3.13, 3.14, 3.15 and 3.16. For example if δ_1 is decreased as shown in figure 3.14 (the black dot indicates the changed position of the steady state), the intersection point moves to the right on
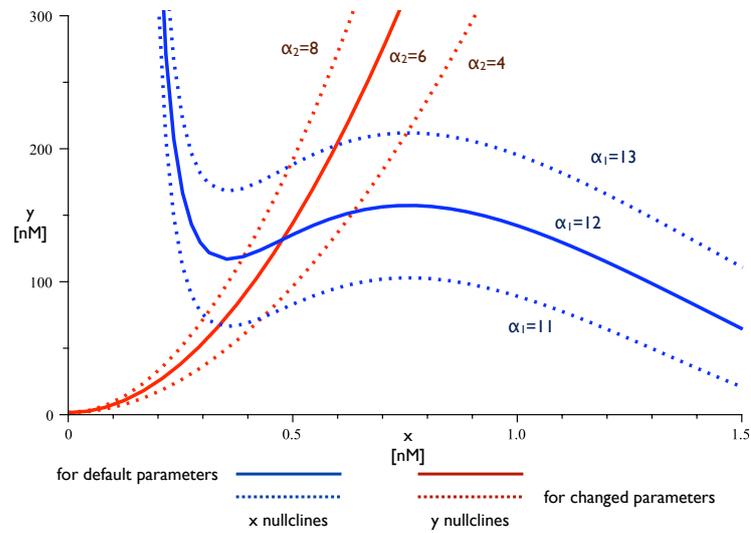
Figure 3.13: Nullcline sensitivity to gene expression rates α_1 and α_2. The parameters α_i (in nMs^{-1}) are indicated next to the corresponding x and y nullclines. All other parameters and the vector field are as given in figure 3.10. The solid curves indicate the nullclines identical to those in figure 3.10. The dotted curves are the nullclines resulting from the different α_i values as indicated.
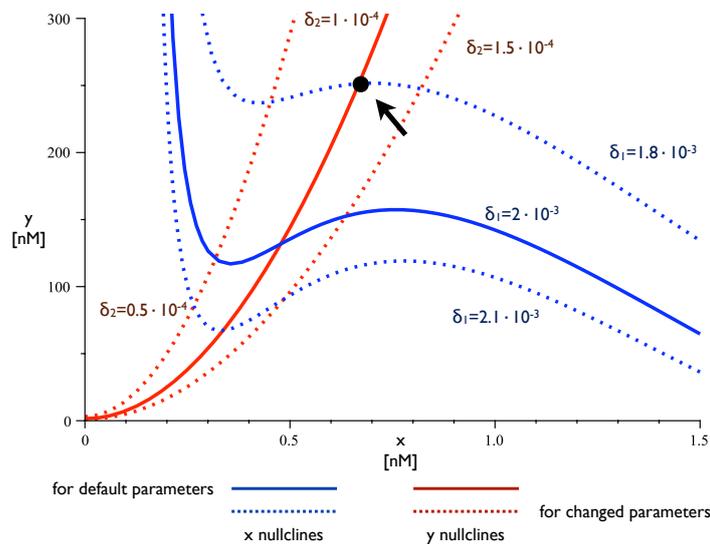
Figure 3.14: Nullcline sensitivity to protein degradation rates δ_1 and δ_2 $[s^{-1}]$. Other details as for figure 3.13. The black dot denotes the nullcline scenario for which sample trajectories are shown in figure 3.17.
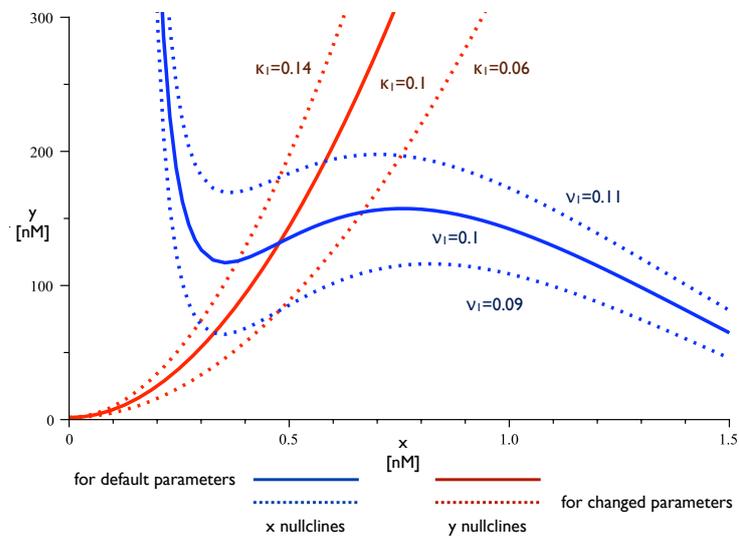
Figure 3.15: Nullcline sensitivity to operator site binding constants ν_1 and κ_1 (in nM^{-1}). Other details as for figure 3.13.
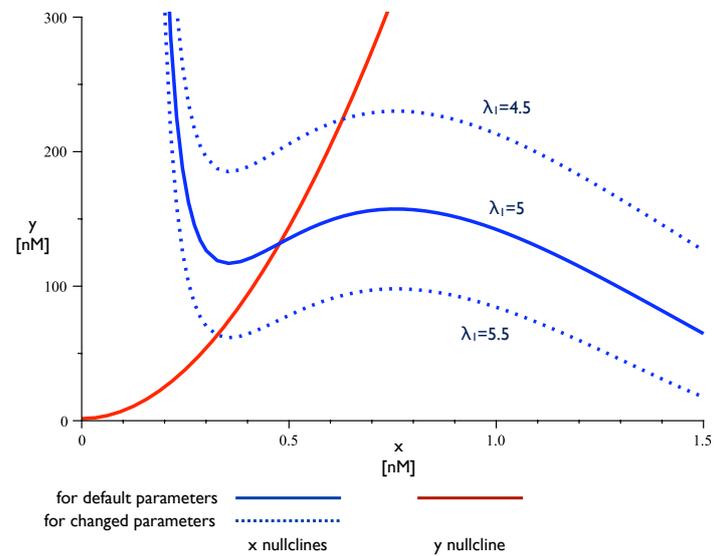
Figure 3.16: Nullcline sensitivity to the *lac* operator binding constant λ_2 (in nM^{-1}). Other details as for figure 3.13.

the NRI (blue) nullcline. As a consequence the slope of the NRI nullcline at the intersection point decreases, so that the limit cycle shrinks and breaks down as the system undergoes a Hopf bifurcation. The steady state becomes a stable spiral. Sample trajectories for this scenario are given in 3.17. As noted earlier, in the case that the nullclines only intersect once, the resulting steady state is either a node or a spiral, but cannot be a saddle point.
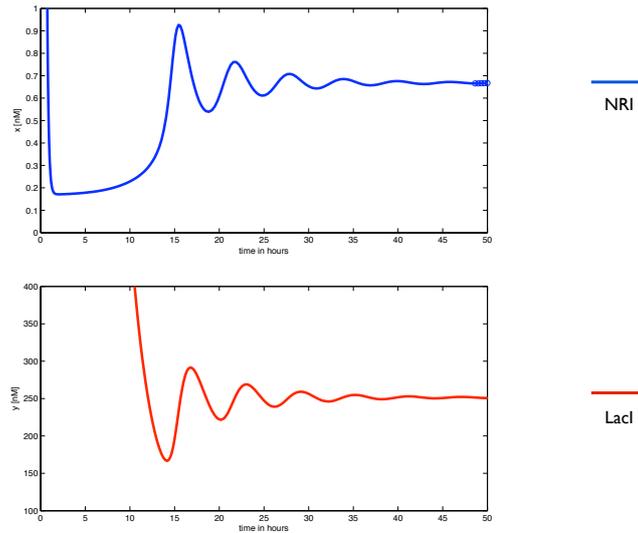

Figure 3.17: Sample trajectory of the nullcline scenario, that is indicated by a black dot in figure 3.14. The degradation rate for NRI is $\delta_1 = 0.0018s^{-1}$. All other parameters are as in figure 3.10, initial conditions as in figure 3.12. The decrease of the NRI degradation rate here has the consequence that the limit cycle collapses and the steady state becomes a stable spiral.

The system can also exhibit bistable behaviour, as is illustrated in figure 3.18. Here the nullclines intersect three times, yielding two stable steady states and an (unstable) saddle point. Trajectories approach one of the two stable steady states depending on the initial conditions. In order to construct this scenario, the parameters for the expression and degradation of LacI, and the binding constants of NRI to the $glnKp$ promoter had to be changed substantially, as is detailed in the legend of figure 3.18. In particular, LacI degradation δ_2 was increased and the maximal expression rate α_2 decreased. Further the binding constants κ_1, κ_2 were increased so that the activation threshold for the $glnKp$ promoter became even lower than that of the $glnAp2$ promoter.
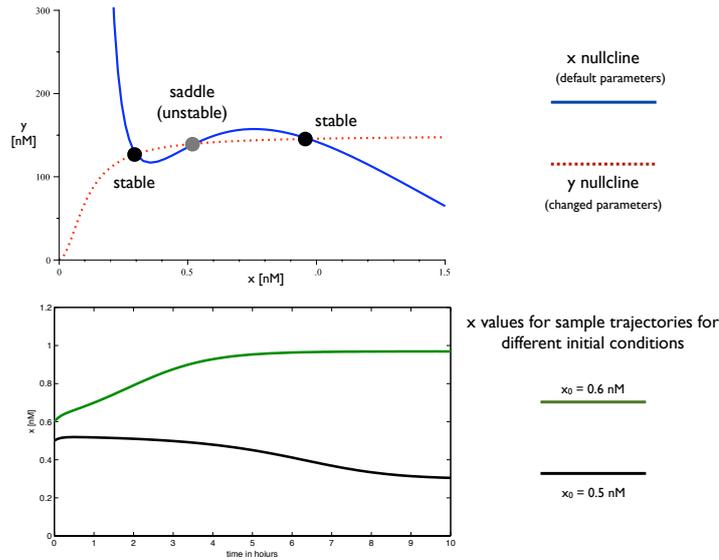
Figure 3.18: Example for a nullcline scenario where the phase plane features two stable fixed points and a saddle. Trajectories approach either stable steady state depending on the initial conditions, which are $y_0 = 100$nM and x_0 as indicated in nM. The NRI nullcline (solid blue curve) has the same shape as in figure 3.10. The LacI nullcline (red dotted line) has changed its position due to the changed parameter choices: $\kappa_1 = \kappa_2 = 1\text{nM}^{-1}, \alpha_2 = 0.3\text{nMs}^{-1}, \delta_2 = 0.002\text{s}^{-1}$. All other parameters are as in figure 3.10.

Role of the different activation thresholds for *glnAp2* and *glnKp* promoters

One rationale behind the design of the engineered clock was to exploit the different activation thresholds of the *glnAp2* and *glnKp* promoters (compare figure 3.3): Low NRI concentrations would first trigger the self activation of NRI via the *glnAp2* promoter, and only when NRI concentrations rise higher, NRI production would be inhibited by the *lac* repressor expressed by the *glnKp* promoter. From the linear stability analysis it was learned, on the other hand, that a small fraction of NRI over LacI protein lifetimes, i.e. $\frac{\delta_2}{\delta_1} << 1$, is important for the existence of a stable limit cycle (compare condition (3.49)). Therefore we will discuss here the relation between the activation thresholds and the fraction of protein lifetimes. In particular, how a change in the difference of the activation thresholds affects oscillations, when it is compensated with a corresponding change in the fraction of protein lifetimes, to keep the system in the realm of sustained oscillations.

Assume we are in the situation of figure 3.10, where a stable limit cycle exists.

We would like to know what happens to the oscillations when the difference between the two activation thresholds is changed, while the position of the steady state on the NRI nullcline remains fixed. This is achieved by changing the binding constant of one the binding sites of the $glnKp$ promoter, κ_2, and by compensating this with a corresponding change in the LacI degradation rate δ_2. Rearranging (3.27) for the steady state (x^*, y^*) yields

$$\delta_2 = \delta_2(\kappa_2) = \frac{\beta_2 + T_{glnKp}(x^*)}{y^*}. \tag{3.53}$$

Because $T_{glnKp}(x^*)$ increases with κ_2, so does δ_2, i.e. $\delta_2'(\kappa_2) > 0$.

Example trajectories for different values of κ_2 and corresponding δ_2 are shown in figure 3.19. When κ_2 increases, the gap between the two activation thresholds narrows. As δ_2 increases correspondingly, the fraction of NRI and LacI lifetimes increases as well, which strengthens condition (3.49), and shrinks the piece of the NRI nullcline, where the steady state is unstable. In the example (figure 3.19, red trajectory), this leads to the steady state becoming stable and the oscillations breaking down.

When κ_2 decreases, the gap between the two activation thresholds widens. As δ_2 decreases correspondingly, the fraction of NRI and LacI lifetimes decreases as well, which weakens condition (3.49), and elongates the piece of the NRI nullcline, where the steady state is unstable. This means that oscillations are preserved. Due to the longer lifetime of the LacI protein, the period of the oscillations increases (see the green trajectory in figure 3.19).

Summary: Bifurcations and oscillations

In the first section it was proven that, when the clock model (with T_{glnAp} as in (3.19) and $\beta_1 > 0$) has only one unstable steady state, a stable limit cycle exists. A corresponding example system was shown, based on biologically feasible parameters. The period of oscillations, as shown for sample trajectories, is comparable to that observed in experiments. The emergence of oscillations from the corresponding characteristic nullcline scenario can also be understood in terms of variables evolving on different time scales. In particular, oscillations are favoured, when the concentration of the activating molecule NRI evolves faster then the concentration of the repressor
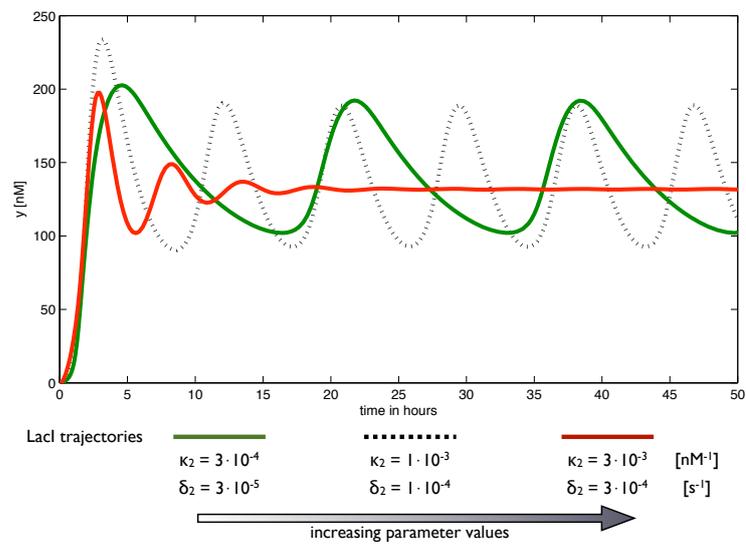
Figure 3.19: The effect of changing the interval between the activation thresholds of the *glnAp2* and *glnKp* promoters. The activation threshold of *glnKp* was changed by changing the binding parameter κ_2 of one of its enhancer sites. In order to keep the position of the steady state fixed, the LacI degradation rate δ_2 was adjusted according to equation (3.53). The black dotted curve shows a LacI trajectory for all parameters as in figure 3.10. Red and green curves show trajectories for decreased and increased values of κ_2 and correspondingly adjusted values for δ_2, respectively. Initial conditions for all trajectories were $x_0 = y_0 = 0$.

molecule LacI.

The second section gave an overview on how the shape of the nullclines is affected by changing the parameters of the system. In particular, with respect to the oscillating system described in the first section, it was shown how a shift in one of the nullclines may take the system out of the region of sustained oscillations. To complete the discussion on possible dynamical behaviors of the system, an example of a nullcline scenario resulting in bistable behaviour was given, where trajectories approach either one of the two stable fixed points depending on the initial conditions.

One intuition behind the design of genetic clock, was to exploit the offset between the two activation thresholds of the *glnAp2* and *glnKp* promoters to produce oscillations. On the other hand, the linear stability analysis revealed that a short lifetime of NRI relative to that of LacI was important for oscillations. A short analysis suggested that indeed, a wider gap between the thresholds facilitates oscillations, while a narrower gap makes oscillations less likely. More precisely, a wider gap between the activation thresholds can be compensated, for example, by a corresponding increase in the LacI lifetime, in order to keep the position of the nullcline intersection fixed. And a longer LacI lifetime increases the interval of instability on the NRI nullcline.

It needs to be stressed that the analysis of the deterministic system neglects the influence of noise in the system, which may alter the qualitative dynamics significantly, see for example [19] and [22]. In particular oscillations in the deterministic system may break down in the stochastic case when the number of particles in the system is too low [30, 12]. Also stochastic effects may lead to oscillating behaviour even when the deterministic case settles down to a steady state. This may happen when small stochastic fluctuations trigger large excursions of the trajectories before they approach the steady state again. In the context of this work this means that for future work it would be very important to compare the deterministic results with corresponding stochastic simulations.

Chapter summary

In the first section 3.1 the average dynamics of the clock model were derived. This was achieved by decoupling the Markov chain of the gene states into submodules describing the NRI dependent regulation of transcription of the *glnAp2* and *glnKp* promoters, and the LacI mediated repression of the *glnAp2* promoter. The invariant

measures of the Markov chains of these submodules were used to construct the average dynamics of the full model, which was illustrated in detail working through an example system introduced in the previous chapter 2.

To complement the LacI models derived in chapter 2, models for NRI activating transcription via one or two enhancer sites, and also inhibiting transcription via NRI governor sites were constructed and their invariant measures derived. Corresponding parameter estimates were taken from the literature to define the different activation thresholds. In particular the activation threshold of the *glnAp2* promoter is lower than that of the *glnKp* promoter, and the NRI induced inhibition of the *glnAp2* promoter appears to be activated at NRI levels, where the *glnKp* promoter is already active. This suggests that for the clock function, the effect of the NRI governor site is not important, as was confirmed by the nullcline analysis of the system. The modelling of the protein production as a single gene expression process, which jointly describes the transcriptional and translational processes, was motivated by making a quasi-steady-state assumption on the mRNA concentration.

Qualitative information of the possible dynamics of a two-dimensional ODE system can be deduced from the corresponding nullclines, whose intersection points are the steady states of the system. Therefore, the second section 3.2 discussed the qualitative shapes of the NRI and LacI nullclines of the clock model for different structural choices like the number of binding sites and the presence of positive basal gene expression rates. In particular it was found that only if the NRI dependent activation was driven by at least two enhancer sites, the NRI nullcline could have an interval of positive slope. Further, a sufficient condition on the parameters was derived for the NRI nullcline to have a positive slope.

Linear stability analysis revealed that a positive slope of the NRI nullcline at the steady state is a necessary condition for the Jacobian matrix to have positive trace, which implies instability of the steady state. A positive trace at the fixed point can be guaranteed for a sufficiently steep slope of that nullcline, where the condition is weakened, when the NRI lifetime is reduced, or the LacI liftetime is increased. Further, when the LacI nullcline intersects the NRI nullcline from below, the resulting steady state will be a node or spiral, while, when the LacI nullcline intersects the NRI nullcline from above, the resulting steady state will be a saddle point.

The third section 3.3 discussed the different nullcline scenarios and corresponding dynamical behaviours of the system, based on the previous nullcline and stability analysis. An example for a stable limit cycle was presented, which was realized with biologically feasible parameters, and for which the period of the oscillations was comparable to that observed in experiments. It was illustrated how changes in parameters lead to different nullcline scenarios. Apart from oscillations, the system can also feature one single steady state, for example a spiral, which corresponds to damped oscillations. Also an example for a bistable system was given, where trajectories approach one of two fixed points depending on the initial conditions.

One rationale behind the construction of the genetic clock was to rely on the different NRI dependent activation thresholds for two promoters, where the activator module (*glnAp2* promoter) would be triggered already for low NRI concentration, while the repressor module (*glnKp* promoter) would be activated only after the NRI levels had risen further. Fixing the position of the steady state on the NRI nullcline and changing the activation threshold for the *glnKp* promoter, while compensating for it by adjusting the degradation rate for LacI, revealed the following: Under the constraint of a fixed steady state, an increase in the activation threshold for *glnKp* can be compensated by an increase in the liftetime of LacI and vice versa. A longer lifetime of LacI, in turn, leads to an increase of the region on the NRI nullcline, where an intersection with the LacI nullcline leads to an unstable steady state. In sum, this suggests that a clock circuit with a wider gap between the two activation thresholds, allows for oscillations in a system with a larger difference in proteins lifeteimes, such that the oscillations are more robust to parameter changes.

The predictions from the mathematical model might be tested experimentally by changing the engineered clock in the following ways. The activation thresholds of the *glnAp2* and *glnKp* promoters could be changed by site-specific mutation of the corresponding enhancer sites, and thereby increasing or decreasing the NRI binding affinity. Degradation rates of proteins in general can be increased by the addition of signal peptides, which mark the proteins for faster degradation by the proteases (*ssrA* tagging, see [33]). Gene expression rates might be changed by relocating the corresponding gene modules on the chromosome (compare [8]), or by mutating polymerase or ribosomal binding sites.

The results so far suggest that for example an increase in the degradation rate

of the activator protein NRI should facilitate oscillations. In order to optimize the accuracy of the model predictions it would be necessary to have a more accurate idea about actual mRNA and protein numbers in in the clock cells. These data could be obtained by performing conventional mRNA and proteins assays during clock experiments.

Chapter 4

Monitoring transcriptional activity of an engineered genetic clock in *E.coli*

This chapter presents the methods and results of experiments for monitoring the activity of the clock modelled in chapter 3 and described in chapter 1. The clock function was monitored by assaying transcriptional regulation of one of the components of the network using a luciferase reporter construct. The data was collected for fitting and optimization of the parameters for the model of the genetic oscillators discussed in the previous chapter, and to enable testing of predictions of the mathematical model. Further work might, for example, investigate the effect of changing the sequence of position of regulatory binding sites in the promoter regions of the clock genes.

In [8] the authors showed that the clock was able to produce oscillations that dampened within three to four cycles. There, clock function was monitored in turbidostat cultures. The cellular clocks were started in a synchronized way by first growing cells in an IPTG containing medium, which was then removed to start the clock. The removal of the IPTG abruptly alters the steady state of the system by activating the *lac* repressor function, and damped oscillations are produced when the clock spirals into the new steady state. To monitor clock function, in [8] the amount of *lac* repressor in the cell was measured indirectly by LacZ assays, because the *lac* repressor regulates the expression of the endogenous *lacZ* gene via the *lac*

operon.

In order to allow for an automated monitoring of the clock over long time spans, and also to open the possibility to screen colonies for interesting clock behaviour, we chose to monitor clock function by an imaging assay based on the luminescence stemming from a luciferase reporter construct, where cells were monitored as colonies on agar plates. The change in the IPTG content to start the clock was achieved by first growing the colonies on medium containing IPTG, and then to transfer the colonies onto fresh plates with IPTG free medium. The colonies were grown on sterile cellophane membranes, which were transferred onto new plates with the colonies on them. This chapter reports the optimization of this assay.

4.1 Experimental methods: establishment of the bioluminescence assay

4.1.1 Molecular cloning methods

Preparation and transformation of competent *E.Coli*

Competent *E.Coli* cells were prepared using the rubidium chloride method described in [29]. Transformation was performed using the 42°C heat shock method described in [65]. Transformed colonies were selected on Luria-Bertani (LB) agar [65] with the appropriate antibiotics (kanamycin sulphate (50μg/ml), ampicillin (100μg/ml), chloramphenicol (12.5μg/ml), gentamicin (20μg/ml)).

Agarose gel electrophoresis and DNA extraction

Separation of DNA fragments by molecular weight was performed by gel electrophoresis as described in [65]. To isolate DNA fragments from agarose gels the appropriate band was cut out of the gel and the DNA extracted by QIAprep Gel Extraction Kit (QIAGEN Ltd, UK).

Molecular methods

Ethanol precipitation.   Ethanol precipitation was used to concentrate DNA or to change buffer as described in [65].

Ligation reactions.   The ligation of DNA fragments into vector plasmids was carried out using T4 DNA ligase (NEB, UK) according to the supplier's protocol. To prepare DNA fragments for blunt ended ligation sticky ends were filled in by the Klenow fragment or T4 Polymerase enzymes according to the supplier's protocol.

Restriction digests.   Restriction endonucleases (Fermentas, Invitrogen, NEB) were used with appropriate restriction buffers according to the protocols provided by the enzyme suppliers.

Site-directed Mutagenesis.   The sequence of promoter constructs was altered by site-directed mutagenesis using the QuikChange method (Stratagene, La Jolla, CA, USA). Efficiency of the method was increased by performing an additional primer elongation step [86]. Correctly mutated plasmids were identified by sequencing plasmids extracted from transformed colonies.

DNA sequencing.   DNA sequencing was carried out by the Warwick University Molecular Biology Service. Sequencing reactions were performed using the BigDye Terminator Version 3.1 system (Applied Biosystems, Warrington, UK) and analyzed with the Applied Biosystems 3130xl Genetic Analyser.

Polymerase Chain Reaction.   DNA was amplified using thermal cycling equipment (Hybaid Co., UK) and the enzymes Taq (Invitrogen, UK) of Pfu turbo polymerase (Stratagene, UK] according to the protocol of the enzyme suppliers.

Extraction of DNA

Extraction of plasmid DNA from *E.Coli*.   DNA extraction from *E.Coli* was performed using the alkaline lysis method with the QIAprep Kit (Qiagen Ltd., UK) according to the manufacturers protocol.

Extraction of *E.Coli* chromosomal DNA.   Chromosomal DNA was extracted from *E.Coli* by phenol:chlorophorm treatment as described in [32].

4.1.2   Construction of luciferase reporter constructs

To study the activity of the engineered clock circuit (see [8]), luciferase based reporters were constructed by fusing the bacterial *luxAB* gene to the modGlnAp promoter region, which drives the activator module of that circuit. The stages of the construction are shown in figure 4.1.
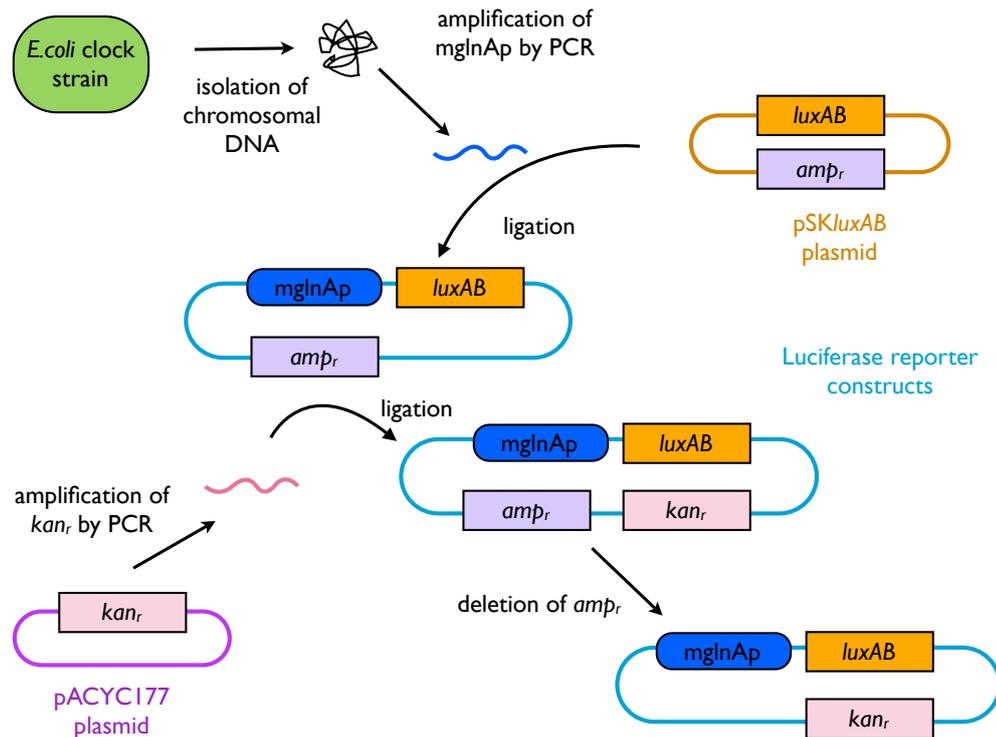

Figure 4.1: Construction of luciferase reporter constructs. To study the activity of the engineered clock circuit, luciferase based reporters were constructed by fusing the bacterial *luxAB* gene to the modGlnAp promoter region (denoted by 'mglnAp' in the figure). The modglnAp promoter region was amplified by PCR from the chromosomal DNA of the *E.coli* clock strain NC12p3415. The promoter region was inserted in front of the *luxAB* gene into the pSK*luxAB* plasmid. A resistance gene for kanamycin was amplified from the pACYC177 plasmid and inserted into the reporter construct. The amp_r gene was deleted to yield a reporter that only confers resistance to kanamycin.

Isolation of the modGlnAp promoter.   The clock circuit was embedded into the chromosome of the *E.coli* strain NC12p3415 (clock strain) [8]. Therefore chromosomal DNA from the clock strain was isolated by phenol:chlorophorm treatment. The modGlnAp promoter region was amplified from the chromosomal DNA by PCR

using the primers PCR-mglnAp (see tables 4.1, 4.2) containing a Shine-Delgano ribosome binding sequence [74].

Fusion of luciferase gene to modGlnAp promoter. Bacterial luciferase (described in [10]), was used as a bioluminescence reporter. The plasmid pSK*luxAB* was cut with restriction enzymes Sal1 and Xho1. The DNA fragment containing the *luxAB* gene was then separated by gel electrophoresis. The modGlnAp region was inserted into the pSK*luxAB* plasmid by blunt ended ligation - the vector and insert were blunt ended by the Klenow fragment after sticky-end ligation had been tried without success. To optimize the distance between the ribosome binding site and the ATG start codon, 10bp of sequence were deleted by the Quikchange method (primer QC-rbs-dist in table 4.1) to yield an optimal distance of 9bp [63].

Changing antibiotic resistance conferred by the reporter. The antibiotic resistance marker of the reporter was changed from ampicillin to kanamycin. The reporter was fitted with a resistance to kanamycin from the plasmid pACYC177. A region containing the kan_r gene on this plasmid was amplified by PCR using the primers PCR-Kan (see table 4.1) and inserted into the psk-mglnAp-*luxAB* reporter plasmid at the sites SacI and XbaI.

The clock strain needed to retain an ampicillin resistance conferring plasmid which carries the gene NRII2302. To avoid having two plasmids conferring the same resistance and the resulting risk of plasmid loss, the ampicillin resistance gene of the reporter construct was disrupted. This was achieved by cutting the reporter plasmid at its two Alw44I restriction sites, one of which lay in the amp_r coding region, and religating it to yield reporter A.

Sequence of reporter constructs. Compared to the expected sequence of the modGlnAp region (expected PCR product in table 4.2, as inferred from [8],[7]) reporter A DNA had an altered sequence (three mutation shown in red in table 4.2). The deletion in the second *lac* operator site and the deletion in the ribosome binding site were fixed by Quikchange using the primer QC-LacOp-rbs (see table 4.1) to yield reporter B.

Primer		Sequence 5'-3'
PCR-mglnAp	FOR	5' GAC CTCGAG AATTGTGAGCGGCTCACAATT GCACCAACATGGTGC
		$\underbrace{}_{xhoI}$ $\underbrace{}_{O^*}$ $\underbrace{}_{NRI_1}$
	REV	5' GAC GTCGAC TTTCTCCTCTTTAAT AATTGTGAGCGGCTCACAATT
		$\underbrace{}_{SalI}$ $\underbrace{}_{rbs}$ $\underbrace{}_{O^*}$
		AAAAAAGATAA AGCCGAAATCTG 3'
		$\underbrace{}_{part\,\sigma^{54}}$
PCR-Kan	FOR	5' ATGC TCTAGA CCTCATCAGTGCCAACATAGTAAG 3'
		$\underbrace{}_{XbaI}$
	REV	5' ATGC GAGCTC TGCCGTGATCTGATCCTTCAA 3'
		$\underbrace{}_{SacI}$
QC-rbs-dist	FOR	5' GGAGAAA GTCGAC GTT ATG AAATTTGG 3'
		$\underbrace{}_{part\,rbs}$ $\underbrace{}_{SalI}$ $\underbrace{}_{start}$
QC-LacOp-rbs	FOR	5' GCGCTCACAATT ATTAAAG AGGAGAAAGTC 3'
		$\underbrace{}_{part\,2nd\,O^*}$ $\underbrace{}_{rbs}$

Table 4.1: Primers used in the construction of luciferase reporter constructs. Names of PCR primers begin with 'PCR', those of Quikchange primers with 'QC'. 'FOR' indicates forward primers, reverse primers are indicated by 'REV'. The modified *glnAp2* promoter region was amplified using the primer 'PCR-mglnAp'. The kanamycin resistance gene was amplified using the primer 'PCR-kan'. The distance between the ribosome binding site (rbs) and the start codon ATG was optimized using the Quikchange primer 'PC-rbs-dist'. The mutations in the *lac* operator and the ribosome binding sites or reporter A were corrected using the Quikchange primer 'QC-LacOp-rbs' - corrected mutations are shown in red. The restriction sites are annotated by the name of the restriction enzyme. The *lac* operator sites are denoted by O^*. NRI_i denote the NRI binding sites. σ^{54} denotes the binding site of the σ^{54} polymerase. Partly represented sites are prefixed with 'part'.

Reporter	Sequence 5'-3'
mglnAp as in [8]	5' AATTGTGAGCGGCTCACAATT GCACCAACATGGTGCTT AATGTTTCCATTGA AGCACTATATTGGTGCAACATTCACATCGTGGTGCAGCCCTTT TGCACGATGGTGCCGCAT GATAA CGCCCTTTTAGGGGGCAATTTAAAAGT TGGCACAGATTTCGCTTT ATCTTTTTT +1 A ATTGTGAGCGGCTCACAATT ATTAAAAGAGGAGAAA 3'
Reporter A	5' AATTGGGAGCGGCTCACAATT GCACCAACATGGTGCTT AATG−TTCCATTGA AGCACTATATTGGTGCAACATTCACATCGTGGTGCAGCCCTTT TGCACGATGGTGCCGCAT GATAA CGCCCTTTTAGGGGGCAATTTAAAAGT TGGCACAGATTTCGCTTT ATCTTTTTT AATTGTGAGCGGCTCACA−TT ATTAAA−AGGAGAAA 3'
Reporter B	5' AATTGGGAGCGGCTCACAATT GCACCAACATGGTGCTT AATG−TTCCATTGA AGCACTATATTGGTGCAACATTCACATCGTGGTGCAGCCCTTT TGCACGATGGTGCCGCAT GATAA CGCCCTTTTAGGGGGCAATTTAAAAGT TGGCACAGATTTCGCTTT ATCTTTTTT AATTGTGAGCGGCTCACAATT ATTAAAAGAGGAGAAA 3'

Labels along the sequences: O^*, NRI_1, NRI_3, NRI_2, NRI_5, NRI_4, σ^{54}, rbs, O^*.

Table 4.2: Sequences of the promoter regions of Reporters A and B are compared with the modified *glnAp2* promoter region of the activator module of the clock circuit. Mutations of the reporters are shown in red. NRI_i denotes the binding sites for NRI. O^* denotes a *lac* operator site, σ^{54} denotes the binding site of the σ^{54} polymerase, *rbs* denotes a ribosomal binding site.

Bacterial strains and growth conditions   The bacterial strains used for monitoring the bioluminescence reporters were $DH5\alpha$, C41(DE3) (described in [49]) and NC12p3415 (clock strain). The clock strain NC12p3415 was provided by the Atkinson laboratory and described in [8]. $DH5\alpha$, C41(DE3) were grown on Luria-Bertani (LB) agar [65]. The clock strain was grown on the medium described in [8], based on the minimum salt base 'W-salts' [75]. More precisely, the growth medium was composed as follows. W-Salts: 10.5g K_2HPO_4, 4.5g KH_2PO_4, 50mg $MgSO_4$ per litre. Supplements: thiamine 0.005% (w/v), tryptophane 0.004% (w/v). Carbon source: glucose 0.4% (w/v). Nitrogen source: glutamine 0.2% (w/v), casein hydrosylate 0.1% (w/v).

4.1.3   In vivo analysis of luciferase transcriptional reporters

Preparation of cells.   Overnight cell cultures were diluted and spread onto standard agar plates containing the growth medium (including antibiotics and IPTG as indicated in figure legends). Plates were incubated at $37°C$ overnight (for about 18h) to yield 20-60 isolated colonies of about 0.5-2mm in diameter, after which the imaging of the bioluminescence was started.

Application of decanal.   The compound n-decanal (Sigma, St. Louis, MO, USA) was used as a substrate for the bacterial luciferase. The substrate was administered as described in [42]: Decanal was dissolved in mineral oil to a concentration of 3% (w/v). 300μl of the decanal-oil solution were applied to a sterile cap of a microtube and placed in the middle of the agar plate. The dish was sealed with Parafilm so the cells were subject to a decanal vapor of constant concentration. Cells were exposed to the decanal right before the start of the imaging.

Transfer of cells during imaging.   To change the growth medium of the cells during image acquisition, cells were grown on a cellophane membrane, which was laid on the agar surface of the Petri dish. Membranes were then transferred onto the new plate after 80-100h of imaging. The exact time of transfer and growth medium are specified in the figure legends.

Image acquisition.   During imaging cells were kept at room temperature (22°C). Images were acquired every thirty minutes with an exposure time of twenty min-

utes. Luminescence of cell colonies was monitored by digital imaging of agar plates using the ORCAII c4742-98 CCD camera system (Hamamatsu (UK) Ltd. Welwyn Garden City, UK) or the liquid nitrogen cooled TEK 512x512DB CCD with an ST138 controller (Princeton Instruments Inc. Trenton, New Jersey). Imaging was controlled by running automated imaging protocols using the Metamorph software package (Molecular Devices Ltd, Wokingham, UK).

Analysis of the imaging data. The light intensities were extracted from 16 bit images which were generated by the cameras using the Metamorph software package. Numerical values were extracted by integrating over selected regions. The regions were drawn by hand to cover a single colony each. To account for different colony sizes the integrated light intensities were normalized by the region area. The background signal was estimated for each image by measuring the light intensity of a region not containing any colonies and subtracted from the colony signal.

For the observed oscillations, the average timing and standard deviations were calculated using the software package Matlab. The data logs extracted from Metamorph were formated by Perl scripts to be readable by Matlab. The timings of the oscillation peaks were obtained by scanning for data points, for which the luminescence values were maximal within an appropriately sized data window centered about these data points.

4.2   Experimental results: monitoring clock function

The transcriptional activity of the activator module of the clock was monitored using the constructed luciferase reporters. In section 4.2.1 the basic properties of the luminescence assay are established. In section 4.2.2 the reporters A and B are compared with respect to their inducibility by IPTG (which relieves repression by LacI). Section 4.2.3 presents the results of monitoring the engineered circuit and investigating its clock function.

4.2.1   Reporter bioluminescence test

Before engaging in monitoring clock activity some basic properties of the bioluminescence reporters were established. In particular for the design of further experiments

it was important to know when the reporter signal would reach a steady state.

To investigate the basic properties of the reporter, reporter A was introduced into the *E.Coli* strain *DH5α*. Diluted cell cultures were spread over LB agar plates to yield about 50 colonies on each plate, which had their bioluminescence monitored (figure 4.2).

After the administration of the luciferase substrate the bioluminescence signal increased to reach a maximum after 5-10 hours. Thereafter the signal slowly decreased until after about 50-70h it reached a steady state. A luminescence signal could still be detected after a week.

The original reporter A carried an ampicillin resistance marker. It was first suspected that the decay in signal was due to the loss of the ampicillin resistance conferring plasmid, as the resistance to ampicillin relies on its degradation which over time creates an ampicillin free microenvironment of the colony. In order to test for this possibility a kanamycin resistance gene was inserted. But cells carrying the kanamycin resistant reporter showed a similar decrease in signal. The peak of the luminescence was shifted a few hours to the right. This might be due to the increased stress for cells growing on media containing kanamycin and ampicillin.

The experiment showed that the reporter was able to produce a reliable bioluminescence signal. The initial peak and following decay in the signal is believed to reflect the uptake of the substrate decanal, possibly followed by the destabilisation of the luciferase enzyme. From the results I concluded that to monitor responses to stimuli, these had to be applied 70h or longer after the administration of the substrate, once the luminescence signal has settled down to a steady state.

4.2.2 Induction test

The next experiment investigated whether the reporter constructs were responsive to changes in the repression efficiency of the LacI repressor. This was checked in particular to investigate the effect of the mutated *lac* operator of reporter A compared to the perfect *lac* operator in reporter B (see reporter sequences in table 4.2).

The promoter regions of the two versions of the reporter differ in the sequence of *lac* operator site. In order to compare the inducibility of the two different constructs by IPTG, both reporter constructs were introduced into *E.coli* cells of the C41(DE3)
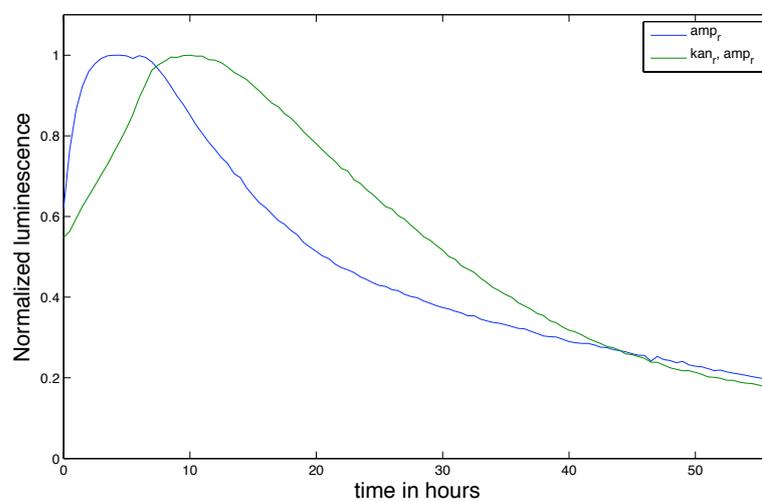
Figure 4.2: Comparison of luminescence patterns for reporter A carrying either ampicillin resistance (amp_r) or both an ampicillin resistance and kanamycin resistance (kan_r). Reporters were introduced into strain $DH5\alpha$ and grown on LB as described in the methods section. The substrate decanal was provided at the start of imaging ($t = 0$). Out of about 50 colonies on each plate the signals of 20 colonies were monitored, all of which showed qualitatively similar behavior. Average luminescence levels are shown for both of these sets of colonies. For comparison the curves were normalized to have a maximum of one.

strain, which carries and expresses the LacI repressor. Cell cultures were grown on LB with antibiotics (ampicillin, kanamycin) and yielded about 50 colonies each. During imaging the colonies were transferred onto fresh LB plates (with antibiotics) that contained IPTG at a concentration of $1mM$ and the reaction monitored. Control colonies were transferred onto fresh plates containing the same medium without any IPTG.

Colonies lying on the same plate all showed similar behaviour. Upon transfer between plates the signal of all colonies, including that of the controls, increased at first and then decayed slowly. The signal increase was more transient and of a lower magnitude for the control colonies than for the induced ones. For control colonies the signal peaked after about 10-15h and ceased to be detectable after 50h. For the induced colonies the signal peaked later at about 25-50h and settled down to a steady state about 100h after the transfer.

On each plate twenty colonies were chosen randomly and their signal averaged. The results are shown in figure 4.3. Induced colonies showed a higher signal than controls for both reporters. The relative increase compared to controls was higher for reporter B (about ten fold on average) than for reporter A (about three fold on average). However, the overall bioluminescence from reporter A was higher than that from reporter B (by a factor of about four).

The data showed that reporter B is more inducible and therefore reacts stronger to changes in LacI repressor activity than reporter A. This behavior was expected as the main *lac* operator in reporter B is the 'perfect' palindromic *lac* operator sequence [64] while the corresponding *lac* operator site in reporter A differs from it by a single base pair mutation. This would suggest that reporter B was best suited to monitor the dynamics of the engineered circuit. But the overall low signal of that reporter leads to a low signal to noise ratio due to the camera limitations. The overall stronger signal of reporter A on the other hand means that the data is less noisy than that of reporter B, while it still shows a sufficiently strong induction. Therefore I decided to use reporter A to measure the engineered clock circuit.

The increase in the signal of the control colonies after the transfer between plates is thought to be the result of the change in the colonies' microenvironment. As the colonies grow over time they will consume nutrients in their immediate vicinity and lower their local concentrations. As a consequence, when a colony is transferred onto

a fresh plate, the concentration of the nutrients in that colony's close environment will suddenly increase. This probably leads to a perturbation of the metabolism of the cells in the colony and overall gene expression activity, and also affects the activity of the *glnAp2* promoter that drives the bioluminescence reporter.

4.2.3   Monitoring the engineered clock circuit

Starting the clock by the removal of the *lac* inducer IPTG

In order to monitor the clock function, reporter A was introduced into the clock strain. Cells were grown on the minimal medium with antibiotics (ampicillin, kanamycin and chloramphenicol). Cells were grown on medium containing varying concentrations of IPTG: 100μM, 1μM and 0μM (as a control). Cells were transferred onto fresh plates without IPTG after 95h.

The response of the colonies to the plate transfer varied with the concentration of IPTG on the pre-transfer plate. A first peak of expression was observed upon transfer, as in the induction experiment shown in figure 4.3, followed by at least one other, suggesting that oscillating functions of the clock had been initiated. With increasing concentration of IPTG the percentage of oscillating colonies decreased to 47 % (1μM) and 15 % (100μM). For colonies showing oscillations the relative timing of the peaks also varied with the IPTG concentration (see table 4.3). Surprisingly, the highest frequency of oscillations was observed when cells were initially grown on medium lacking IPTG. For low IPTG concentrations ($0 - 1\mu$M) the interval between the peaks was about $15h$, for a higher concentration (100μM) the interval varied between $30 - 50$h. The timing of the first signal peak lay shortly after the point of transfer, $0 - 1$h, for most oscillating colonies and did not vary with IPTG concentration. Figure 4.4 shows representative curves for colonies showing damped oscillations.

The first peak in the signal is the expected response following the cell transfer between plates as was seen in the previous section. More precisely, this first peak apparently is not due to the removal of IPTG, but a response to a metabolic stimulus due to the plate transfer. This metabolic stimulus must be masking the response to IPTG removal. The second peak was accredited to the dynamical feedback of the engineered clock circuit. The spread in the relative timing of the peaks is probably
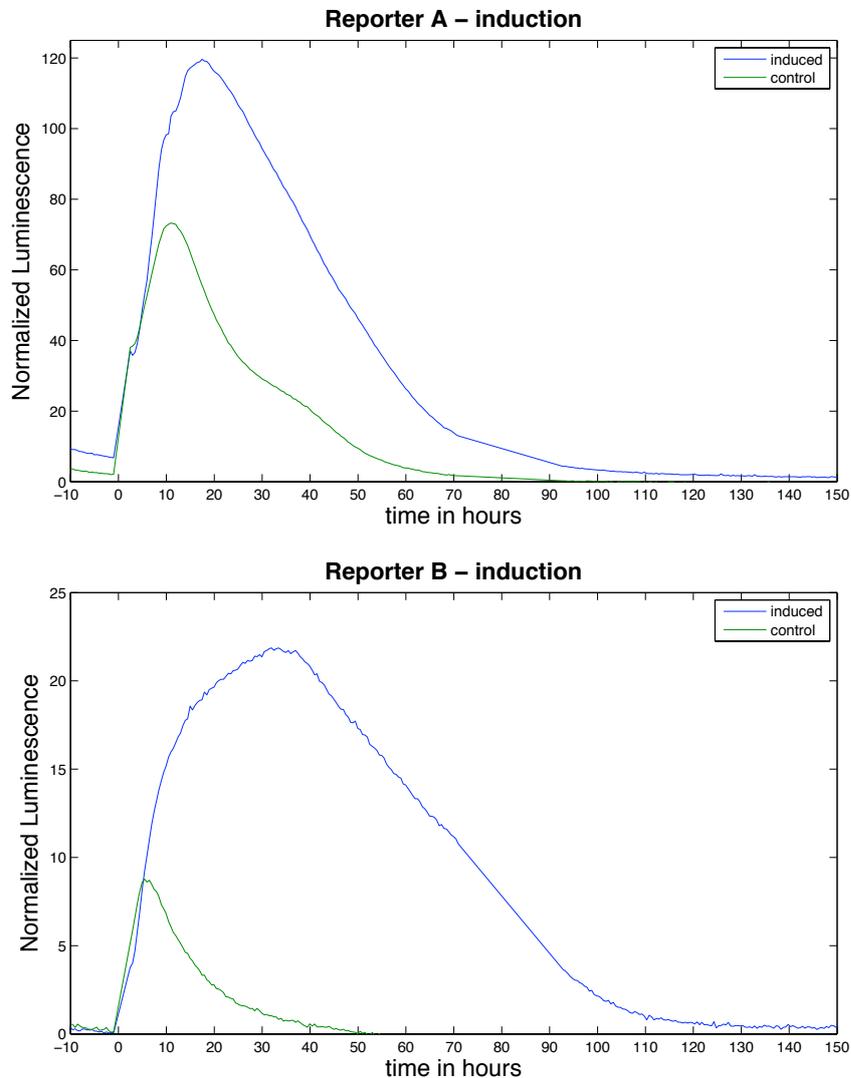
Figure 4.3: Reaction of reporter constructs A and B to *lac* inducer IPTG. The reporter constructs were introduced into the inducible *E.coli* strain C41(DE3) and grown on LB medium containing ampicillin and kanamycin. The medium was changed - as described in the methods section - to contain additionally IPTG at a concentration of 1mM at $t = 0$, 90h hours after the application of the substrate decanal. Control cells were transferred to a medium without any IPTG. The figure shows the averaged signal of 20 randomly chosen colonies from each plate, which contained about 50 colonies each. Colonies from the same plate all showed similar behavior. Luminescence was normalized on a per pixel basis.
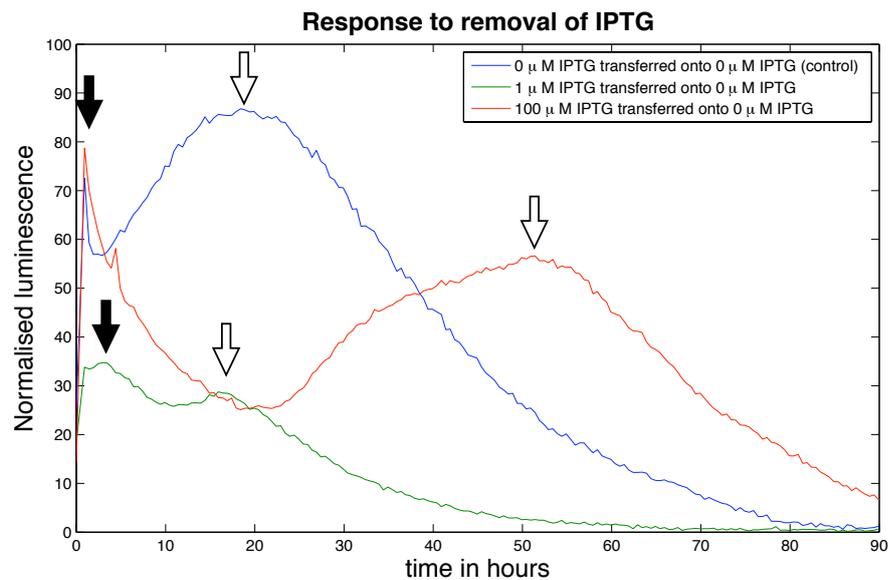
Figure 4.4: Induction of oscillatory behaviour by removal of IPTG. Reporter A was introduced into the clock host strain. Cells were grown on the minimal medium with antibiotics ampicillin, kanamycin, chloramphenicol (as described in the methods) and varying concentrations of IPTG (see legend), then were transferred after 95h onto plates containing medium without IPTG. The number of colonies monitored for each IPTG concentration was 40 (0μM), 34 (1μM) and 40 (100μM). A fraction of the colonies - detailed in table 4.3 - showed damped oscillatory behaviour with two visible peaks. The figure shows representative example curves of the oscillating colonies for the different IPTG concentrations. Solid and open arrows indicate the timing of the first and second peak, respectively.

Table 4.3: Oscillations after removal of IPTG.

IPTG concentration before transfer	0	
No. of monitored colonies	40	
of which showed two peaks	30 (75 %)	
	mean	std. deviation
Time of first peak	0.33h	1.43h
Time of second peak	15.45h	5.37h
Interval between peaks	15.12h	5.24h

IPTG concentration before transfer	1μM	
No. of monitored colonies	34	
of which showed two peaks	16 (47 %)	
	mean	std. deviation
Time of first peak	1.13h	1.06h
Time of first peak	17.65h	2.39h
Interval between peaks	16.53h	2.51h

IPTG concentration before transfer	100μM	
No. of monitored colonies	40	
of which showed two peaks	6 (15 %)	
	mean	std. deviation
Time of first peak	0.5h	0h
Time of second peak	37.67h	7.39h
Interval between peaks	37.17h	7.39h

due to differences in the size and shape of the individual colonies. Cells at different positions within a colony experience different environmental conditions - cells at the rim of a colony have more access to the medium and grow faster, while cells in the centre of the colony experience less favorable conditions and may stop growing or die. Since the signal of a colony is the average of all the cells in it, the colony shape will affect the colony's signal. The decay of the oscillation after the second peak might be the result of the promoter dynamics of the individual cells or due to a desynchronization of the cells within single colonies as they grow during the course of the experiment.

The results suggest that a high IPTG concentration prior to transfer reduces the likelihood of colonies to show more than one clearly defined activation peak. When two peaks can be seen the period of the oscillation seems to increase with the IPTG concentration. In sum transferring colonies between cells whilst removing IPTG from the growth medium triggers damped oscillations. But while a higher pre-transfer concentration of IPTG increases the period of the oscillations it also makes the oscillations less reproducible. This could be because the IPTG is not completely removed from the cells following transfer to new plates.

Surprisingly, colonies on control plates, which were grown on medium without IPTG also before the transfer, also showed damped oscillations. This suggests that the metabolic stimulus, that cells experience upon transfer, perturbs the system in a way that starts the clock.

The exact interpretation of the results is complicated by the overlay of the chemical (removal of IPTG) and metabolical (replenishment of nutrients on fresh plate after transfer) stimulus. That oscillations could only be observed on a fraction of colonies on a plate suggests that the heterogeneity between colonies has a significant effect on the clock's behaviour.

Colony age at the time of transfer affects the period of the oscillations

In [8], the authors reported that the period of the oscillations depend on the growth rate of the cells - doubling times of one and two hours yielded oscillations of the period of about 10h and 20h hours respectively. When colonies grow larger in size, cells in its centre will grow more slowly or die due to worsening environmental conditions. Consequently, the average growth rate of the cells should decrease over

time. This suggests that the age of the colony at the time the clock is triggered would affect the period of the resulting oscillations. More precisely, the higher growth rate of the younger colonies, which contain a higher percentage of rapidly dividing cells, should yield a shorter period of oscillations. Heterogeneity of older colonies may cause damping of oscillations due to different periods.

In order to test this hypothesis, reporter A was introduced into the clock strain and cells grown on the minimal medium, as described in the methods section, with antibiotics ampicillin, kanamycin and chloramphenicol. The colonies were transfered at an earlier stage than in the previous experiment, i.e. after 75h, onto fresh plates that contained the same medium. The bioluminescence response of 40 colonies was monitored.

In about half of the colonies the initial peaks of luminescence was followed by a second peak after about 10-15h, suggesting that the transition had triggered an oscillation of the circuit. The rest of the colonies had only one clearly visible peak in the signal. Signals decayed back to the steady state value about $50h$ after the transfer. The signals of typical colonies are displayed in figure 4.5. The distribution of the relative timing of the peaks is given in table 4.4. The first and second peaks of the signal lay at about 5h and 17.5h after the transfer, yielding an average interval (period) of about 12.5h between the peaks. Signals of oscillating colonies showed a similar timing of the peaks as can be seen of the low standard deviation of 1.91h for the period (interval between peaks.

In sum the experiment showed that the earlier transfer also triggers a damped oscillatory response in the promoter activity of the activator module of the engineered circuit. Oscillations could only be observed in about half of the cells. The timing of the peaks of the oscillating colonies were fairly consistent. The average period of the oscillations was shorter for cells transfered after 75h than for those transferred after 95h, as expected. One issue that makes the interpretation of the results difficult is that the exact effect of the metabolic stimulus on the cells is not clear.
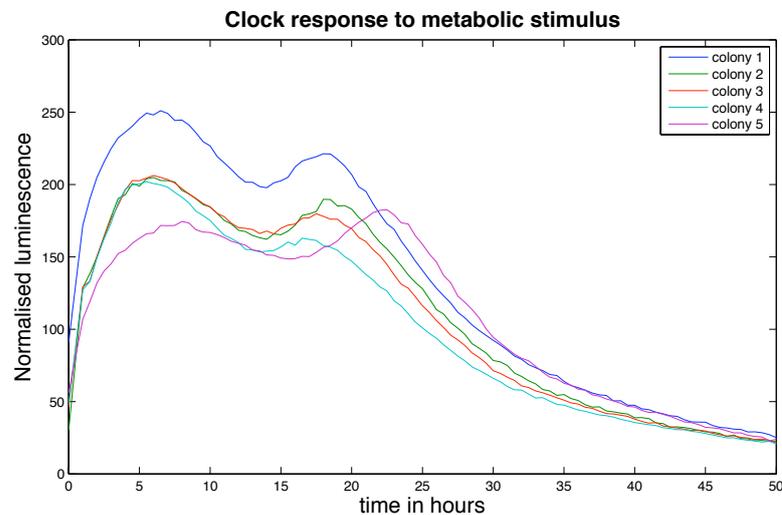
Figure 4.5: Response of the *glnAp:lux* reporter A to the metabolic stimulus after 75h. Reporter A was introduced into the clock strain. Cells were grown on the minimal medium with antibiotics ampicillin, kanamycin, chloramphenicol (as described in the methods). After 75h colonies were transferred to fresh plates containing the same medium - 15h earlier than in the previous experiment (compare figure 4.4). 40 colonies were monitored, 20 of which showed oscillations with two peaks, while the other colonies only showed one peak. The figure shows the bioluminescence response of five typical oscillating colonies. Details on the distribution of the peaks' position are summarized in table 4.4.

Table 4.4: Reaction of activator to metabolic stimulus (no IPTG, transfer after 75h)

No. of monitored colonies	40	
of which showed two peaks	20 (50 %)	
	mean	std. deviation
Time of first peak	5.13h	0.84h
Time of second peak	17.6h	2.65h
Interval between peaks	12.48h	1.91h

Possible causes for the rapid dampening of oscillations

In [8], the authors reported that the clock showed up to three to four cycles of oscillations, while here, only two cycles could be observed. There are several possible causes for this rapid dampening of the oscillations.

- One, over time, the clock of the individual cells desynchronize within the colony and the oscillations are averaged out. In particular, the environmental conditions of the cells in the middle of the colonies lead to slower growth or death of cells - the net result is also a desynchronizing of the individual clocks.

- Two, the stimulus, which the cells experience upon transfer onto a fresh plate is different from the IPTG removal in the chemostat experiments, and therefore leads to different clock response. Also, the IPTG removal by transfer unto a fresh plate might not be complete, because ITPG may still be sticking to the cells or the membrane.

- Three, during the time of the imaging, the clock cells might loose a plasmid that is important for clock function, as is explained below.

The functionality of the engineered circuit relies on the constant phosphorylation of the transcription factor NRI. To assure this the clock strain was fitted with the mutant kinase NRII2302. The gene for this kinase was provided on a plasmid conferring resistance to the antibiotic ampicillin. The way in which the resistance to ampicillin is achieved leads to a degradation of ampicillin in the environment of the cell colony. It was suspected that this change of a colony's microenvironment might cause clock cells to loose the kinase carrying plasmid over time and thus cause the breakdown of the oscillations.

To check whether the plasmid was retained the following experiment was performed. After a typical clock experiment all colonies were washed off and (after sufficient dilution) plated on fresh LB plates containing either kanamycin and ampicillin or only kanamycin as antibiotics. After an overnight incubation the number of growing colonies was counted.

The number of colonies were similar on both plates. The kanamycin plate carried 621 colonies compared to 517 on the kanamycin / ampicillin plate. This means that most of the clock cells that carried the bioluminescence reporter also retained their

ampicillin resistance and the NRII2302 kinase. From that I concluded that the breakdown of oscillations cannot be explained by the loss of the NRII2302 kinase.

Conclusions

It was shown that the constructed luciferase reporters produce a reliable luminescence signal over long time periods. The reporter constructs showed a strong response to the *lac* inducer IPTG. Reporter A, containing a mutant *lac* operator site, was chosen for monitoring the clock circuit because it had stronger luminescence, while still showing good inducibility.

Clock function was measured by synchronizing cells through exposure and subsequent removal of IPTG by transfer between plates. The signal from reporter A showed damped oscilations, whose period increased with the concentration of IPTG. Also, oscillations were observed in control colonies, suggesting that the metabolic stimulus from the transfer onto fresh plates is sufficient to start the clock function. When cells were transfered at an earlier time, a shortening of period was observed, suggesting that the higher average growth rate of younger colonies leads to a shorter period of the clock. In all clock experiments damped oscillations were detected only in a fraction of colonies. An increase in the IPTG exposure resulted in the damped oscillatory behaviour being less reproducible. In order to improve the system it would be important to determine the efficiency of IPTG removal upon transfer, i.e. how much IPTG is still retained by the colonies and the membrane after transfer. This might be tested for example by studying the response of a standard *lac* inducible reporter to the IPTG removal by colony transfer, and compare it to reporter carrying cells, which are grown at different IPTG concentrations. Also, a closer look at the correlation between oscillations and colony size might be insightful.

The motivation for conducting the imaging experiments was to obtain reliable data on which to base the modelling on. The system was also intended as a tool to analyse the consequences of changes to the regulatory DNA sequences, which would have been detected by colony based screening for interesting expression patterns - like longer lasting oscillations. In experiments, the clock produced damped oscillations of fairly consistent period. Only a fraction of colonies on a plate showed oscillatory behaviour suggesting the need for further optimization of the experimental setup. The obtained data serves as a good illustration of the model, as the model is able to

reproduce the damped oscillatory behaviour with a comparable period length (see section 3.3), and thus opens possibilities for further work on that system.

The conducted work highlighted some problems of the imaging assay experiments, in particular the system's sensibility to the growth conditions. Therefore the assay could be further optimized to yield a better control and more homogeneity of the growth conditions of the cells. This could be achieved by monitoring clock function in a chemostat environment as in [8], where the growth medium can be changed in a more controlled manner. Although, one would loose the advantages of the automatization and the possibility of colony screening. Also, further data might be obtained by measuring the level of mRNA transcripts and by monitoring the expression of the second gene in the circuit. With some further work - and a bit of luck - it might still be possible to push the system into the realm of sustained oscillations, which would correspond moving through a Hopf bifurcation in the mathematical model.

Conclusions

This work presented the mathematical modelling and analysis of an engineered genetic oscillator in *E.coli*, in concert with the acquisition and analysis of experimental data on that circuit, based on a newly set up imaging assay. Genetic oscillators are the basis of biological clocks, which allow organisms to maintain stable circadian rhythms. The mathematical modelling took into account the special, heterogeneous nature of genetic regulatory systems, where variables can describe discrete finite gene states, or infinite protein numbers (or concentrations). The newly set up imaging assay for the engineered clock allowed to collect data over long time periods in an automated fashion. Following an introductory chapter (chapter one), the work was divided into three parts: Chapter two described the theoretical modelling framework and introduced a newly developed algorithmic tool for automated model construction. Chapter three presented the model derivation and its analysis. Chapter four discussed the details of the imaging assay and the data collected from the clock cells.

Summary of the presented work. Models with finite, inherently discrete variables in conjunction with variables describing particle numbers or concentrations require special theoretical consideration. In chapter two, a theoretical framework, previously described in [66, 69], was discussed, that allows to derive deterministic and continuous rate equations from discrete stochastic model formulations of molecular interactions. In order to be able to apply this theory to realistic systems, an algorithmic modelling tool was devised and implemented, that automatically constructs Markov chain models for gene activity states based on binding dynamics of regulatory factors to enhancer and operator sites of the respective promoters. In particular, the algorithm is able to account for cooperative binding and DNA loop formation. The source code and example files can be downloaded from the website of the Compbio group at 'lora.maths.warwick.ac.uk'. The developed algorithm was

applied to study different models for repressive DNA loop formation induced by the *lac* repressor. The model analysis suggested, that loop formation based on solely dimeric repressor molecules fails to convey the noise reducing properties of looping by tetrameric rerpressors, which were reported previously in [85].

Chapter three presented the derivation and analysis of models for the engineered clock circuit. To reduce the dimension of the modelling problem, the Markov chain of the gene states of the clock model was decoupled into submodules, to be treated separately, describing the regulation of the promoters *glnAp2* and *glnKp* by the transcription factor NRI, and the repression of the *glnAp2* promoter by the *lac* repressor, LacI. Models for the NRI dependent regulation were constructed, which took into account the cooperative binding of NRI to DNA, and whose parameters were fitted from the literature, where the *glnAp2* and *glnKp* promoters have been studied in detail. The average dynamics for the clock circuit were then derived by reassembling the rate laws for the respective submodules. The transcriptional and translational processes were considered jointly, motivated by a quasi-steady-state assumption on the mRNA concentrations, to arrive at a two dimensional system of coupled ordinary differential equations.

Nullcline and linear stability analysis revealed that the occurrence of a Hopf bifurcation requires the positiveness of the slope of the NRI nullcline. This in turn was only possible for models where the *glnAp2* promoter was driven by at least two enhancer sites. Further, sufficient conditions on the nullcines were derived for sustained ocillations to exist. Other possible behaviour of the system, like the existence of a single stable steady state, corresponding to damped oscillations in the case of a stable spiral, or bistability, were characterized in terms of corresponding nullcline scenarios. The model suggested, that a short NRI lifetime relative to a longer LacI lifetime, as well as the difference in the activation thresholds of the *glnAp2* and *glnKp* promoters, were important for stable oscillations.

In order to fit and optimize the model and to be able to test predictions, a new monitoring assay for the eningeered clock was set up based on luminescent transcriptional reporter constructs. The imaging of colonies growing on a petri dish allowed for automated data collection over long time spans and also opened the possibility to screen for clock mutants. Damped oscillations were observed after starting the clock by the removal of the *lac* inducer IPTG, as was expected from previously re-

ported results based on chemostat culture assays monitoring LacZ activity [8]. Also, damped oscillations were observed after cells were exposed to a metabolic shock due to colony transfer between agar plates and consequent exposure to fresh nutrients. (Colony transfer between plates was used to change the growth medium.) The observed period of oscillations differed with the age of the colonies. On the grounds of previous results [8], it was concluded that different average growth rates of cells within colonies of different ages may be responsible for these period variations. The esablished imaging assay constitutes a decent basis for the further testing of model predictions. For this, the assay may need to be further optimized to improve control of growth conditions. Also the imaging data could be complemented by measurements of mRNA and protein numbers to improve parameter fitting.

Discussion and Outlook. The Markov chain based modelling framework allows to bridge from detailed, discrete stochastic model formulations to deterministic and continuous systems, which are more tractable by mathematical analysis. The presented algorithm, that automatizes the construction of the discrete Markov chain models for the described class of genetic regulatory systems, constitutes a step towards facilitating the application of this kind of modelling to realistic systems and making it more accessible to non-specialists. A possible route for further development would be to extend the algorithm, such that it can treat a larger class of models, or to embed the algorithm into a larger modelling package, such that the derivation of invariant measures and average dynamics can be performed routinely. This would facilitate the comparative analysis of models for gene regulation and related systems, which are able to distinguish between specific molecular mechanisms, and would make such modelling available to a larger part of the multidisciplinary research community. The software package Copasi represents an example for such a modelling tool used mainly for metabolical reaction networks [37]. A comparable tool specifically designed for genetic regulatory networks appears still to be missing.

The analysis of the models of the clock circuit highlighted, that apart from the overall composition of the regulatory dependencies, also the molecular details of the implementations (number of binding sites, DNA looping, cooperativity) are important for determining the qualitative dynamic properties of the system. For the studied clock circuit the analysis suggested that the cooperative nature of the auto-

activation of the activator module is an essential feature for oscillations, while the nonlinearity of other regulatory mechanisms (activation of the repressor module and repression of the activator module) seemed to be less important. The importance of the degradation rates is in accordance with studies of other genetic oscillators [33] and the difference in activation thresholds of the activator and repressor module may mimic the role of delayed negative feedback, like for example in the *Drosophila* clock described in chapter one.

Monitoring the clock using the newly setup imaging assay proved feasible, even though further optimization is needed, like more homogenous growth conditions, to improve the quality of the data. Maintaining homogenous growth conditions for this colony based assay is more difficult than for chemostat cultures, for which, on the other hand, assays based on culture samples are more difficult to automatize. Future work on the established clock assay could include site-specific mutations of the clock's gene modules to study for example the effect of different activation thresholds for the two gene modules of the clock, or increasing the degradation rate of the activator protein NRI, by tagging it with degradation signal peptides. Random mutagenesis of the chromosome of the clock strain might also yield interesting clock behaviour which could be screened for.

On a wider perspective, the current work highlighted some characteristic features of models of genetic regulatory systems, like of the heterogeneity of variables and also the complexity in the details of the corresponding molecular mechanisms. The presented results suggest, that it is important to develop models that are able to precisely distinguish between different molecular mechanisms. On the other hand, biological models found in the literature, are often not precise enough to be translated into mathematical models without some ambiguity and additional assumptions. Making the construction and formulation of mathematically well defined models more routine, should thus help advance the understanding of biological systems.

In particular in the context of models for genetic oscillators, their possibly complex dynamics are sensitive to the molecular details of gene regulation. The advancing ability to construct artificial genetic constructs promises to facilitate the comparison of different mechanisms by modelling and experiment. This may help elucidate the more complex properties of natural biological clocks like temperature

compensation and robustness, by aiming to engineer circuits that are able to mimic some of these advanced properties.

This thesis, being an interdisciplinary work of mathematics and biology, is also characterized by the challenge to bridge between the corresponding research cultures. While biologists typically use descriptive language in order to understand a certain biological function, mathematicians prefer more rigorous and precise formulations in order to capture more general patterns. This two-sidedness presents a challenge in the field of mathematical biology in general. But as the understanding of the cell as a complex molecular machine gains coherency, the description of its workings will naturally become more mathematical. In this respect genetic oscillators are an example for a long standing, and successfully developing field of research for mathematical biology.

Bibliography

[1] F.B. Allison, X. Cui, F.P. Page, and M. Sabripour. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet*, 7(1):55–65, 2006.

[2] H. Amann. *Gewoehnliche Differentialgleichungen*. de Gruyter, Berlin, 1995.

[3] A. Arkin, J. Ross, and H.H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, 149(4):1633–1648, 1998.

[4] M.R. Atkinson, T.A. Blauwkamp, V. Bondarenko, V. Studitsky, and Alexander J. Ninfa. Activation of the *glnA*, *glnK*, and *nac* promoters as *Escherichia coli* undergoes the transition from nitrogen excess growth to nitrogen starvation. *J. Bacteriol.*, 184(19):5358–5363, 2002.

[5] M.R. Atkinson, T.A. Blauwkamp, and A.J. Ninfa. Context-dependent functions of the pii and glnk signal transduction proteins in *Escherichia coli*. *J. Bacteriol.*, 184(19):5364–5375, 2002.

[6] M.R. Atkinson and A.J. Ninfa. Characterization of the glnk protein of *Escherichia coli*. *Molecular Microbiology*, 32(2):301–313, 1999.

[7] M.R. Atkinson, N. Pattaramanon, and A.J. Ninfa. Governor of the *glnAp2* promoter of *Escherichia coli*. *Molecular Microbiology*, 46(5):1247–1257, 2002.

[8] M.R. Atkinson, M.A. Savageau, J.T. Myers, and A.J. Ninfa. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, 113(5):597–607, 2003.

[9] C.P. Bahl, R. Wu, J. Stawinsky, and S.A. Narang. Minimal length of the lactose operator sequence for the specific recognition by the lactose repressor.

Proceedings Of The National Academy Of Sciences Of The United States of America, 74(3):966–970, 1977.

[10] T.O. Baldwin, T. Berends, T.A. Bunch, T.F. Holzman, S.K. Rausch, L. Shamansky, M.L. Treat, and M.M. Ziegler. Cloning of the luciferase structural genes from vibrio harveyi and expression of bioluminescence in *Escherichia coli*. *Biochemistry*, 23(16):3663–3667, 1984.

[11] A.-L. Barabási. *Linked: The New Science of Networks*. Perseus, Cambrdige, Ma, 2002.

[12] N. Barkai and S. Leibler. Biological rhythms: Circadian clocks limited by noise. *Nature*, 403(6767):267–268, 2000.

[13] M.D. Barkley, A.D. Riggs, A. Jobe, and S. Bourgeois. Interaction of effecting ligands with lac repressor and repressor-operator complex. *Biochemistry*, 14(8):1700–1712, 1975.

[14] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, 2000.

[15] A. Becskei, B. Sraphin, and L. Serrano. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *The EMBO Journal*, 20:2528–2535, 2001.

[16] D. Bell-Pedersen, V.M. Cassone, D.J. Earnest, S.S. Golden, E. Hardin, Paul, L. Thomas, Terry, and J. Zoran, Mark. Circadian rhythms from multiple oscillators: Lessons from diverse organisms. *Nature Reviews Genetics Nat Rev Genet*, 6(7):544–556, 2005.

[17] T.A. Blauwkamp and A.J. Ninfa. Physiological role of the glnk signal transduction protein of *Escherichia coli*: survival of nitrogen starvation. *Molecular Microbiology*, 46(1):203–214, 2002.

[18] R.J. Cho, M.J. Campbell, E.A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T.G. Wolfsberg, A.E. Gabrielian, D. Landsman, D.J. Lockhart, and R.W. Davis. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell*, 2(1):65–73, 1998.

[19] R.E. Deville, C.B. Muratov, and E. Vanden-Eijnden. Non-meanfield deterministic limits in chemical reaction kinetics. *J Chem Phys*, 124(23):231102, 2006.

[20] J.C. Dunlap. Molecular bases for circadian clocks. *Cell*, 96(2):271–290, 1999.

[21] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature Nature*, 403(6767):335–338, 2000.

[22] R. Erban, S.J. Chapman, I.G. Kevrekidis, and T. Vejchodsky. Analysis of a stochastic chemical system close to a sniper bifurcation of its mean-field model. *http://arxiv.org/abs/0807.4498v1*, 2008.

[23] C. Espinosa-Soto, P. Padilla-Longoria, and E.R. Alvarez-Buylla. A gene regulatory network model for cell-fate determination during arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *Plant Cell*, 16(11):2923–2939, 2004.

[24] B. Finkenstadt, E.A. Heron, M. Komorowski, K. Edwards, S. Tang, C.V. Harper, J.R.E. Davis, M.R.H. White, A.J. Millar, and D.A. Rand. Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics*, page btn562, 2008.

[25] C.W. Gardiner. *Handbook of Stochastic Methods, 3rd Edition*. Springer, Heidelberg, 2003.

[26] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in *Escherichia coli. Nature Nature*, 403(6767):339–342, 2000.

[27] W. Gilbert and A. Maxam. The nucleotide sequence of the lac operator. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 70(12):3581–3584, 1973.

[28] L. Glass and S.A. Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *Journal of Theoretical Biology*, 39(1):103–129, 1973.

[29] D. Golver, editor. *DNA Cloning, vol. 1.* IRL Press Ltd., London, 1985.

[30] D. Gonze, J. Halloy, and A. Goldbeter. Deterministic versus stochastic models for circadian rhythms. *Journal of Biological Physics*, 28(4):637–653, 2002.

[31] J. K. Hale and H. Kocak. *Dynamics and Bifurcations*. Springer, Heidelberg, 1991.

[32] R.E.W. Hancock. *Hancock Laboratory Methods*. [online]. *URL http://www.cmdr.ubc.ca/bobh/methods.htm*. *Department of Microbiology and Immunology, University of British Columbia, British Columbia, Canada*, 1999. (Last accessed 5 November 2008).

[33] J. Hasty, D. McMillen, and J.J. Collins. Engineered gene circuits. *Nature*, 420(6912):224–230, 2002.

[34] J. Hasty, D. McMillen, F. Isaacs, and J.J. Collins. Computational studies of gene regulatory networks: in numero molecular biology. *Nature Reviews Genetics Nat Rev Genet*, 2(4):268–279, 2001.

[35] E.A. Heron, B. Finkenstadt, and D.A. Rand. Bayesian inference for dynamic transcriptional regulation; the hes1 system as a case study. *Bioinformatics*, 23(19):2596–2603, 2007.

[36] U.R. Hiroki, H. Masatoshi, and K. Hiroaki. Robust oscillations within the interlocked feedback model of drosophila circadian rhythm. *Journal of Theoretical Biology*, 210(4):401–406, 2001.

[37] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. Copasi–a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[38] F. Jacob and J. Monod. On the regulation of gene activity. *Cold Spring Harb. Symp. Quant. Biol.*, 26:193–211, 1961.

[39] D. Kennell and H. Riezman. Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *Journal of Molecular Biology*, 114(1):1–21, 1977.

[40] T.B. Kepler and T.C. Elston. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.*, 81(6):3116–3136, 2001.

[41] M. Kirkilionis. *UniNet website.* [online]. (http://lora.maths.warwick.ac.uk/∼uninet). 2005. (Last accessed 5 November 2008).

[42] T Kondo and M Ishiura. Circadian rhythms of cyanobacteria: monitoring the biological clocks of individual colonies by bioluminescence. *J. Bacteriol.*, 176(7):1881–1885, 1994.

[43] R. Laubenbacher and B. Stigler. A computational algebra approach to the reverse engineering of gene regulatory networks. *J Theor Biol*, 229(4):523–537, 2004.

[44] S.-Y. Lee, A. de la Torre, D. Yan, S. Kustu, B.T. Nixon, and D.E. Wemmer. Regulation of the transcriptional activator ntrc1: structural studies of the regulatory and aaa+ atpase domains. *Genes Dev.*, 17(20):2552–2563, 2003.

[45] J.-C. Leloup and A. Goldbeter. A model for circadian rhythms in drosophila incorporating the formation of a complex between the per and tim proteins. *J Biol Rhythms*, 13(1):70–87, 1998.

[46] B.P. Lewis, C.B. Burge, and D.P. Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20, 2005.

[47] J.C.W. Locke, A.J. Millar, and M.S. Turner. Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*. *Journal of Theoretical Biology*, 234(3):383–393, 2005.

[48] R. Merris. *Multilinear Algebra*. Gordon and Breach Science Publishers, 1997.

[49] B. Miroux and J.E. Walker. Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, 260(3):289–298, 1996.

[50] J. Muller, S. Oehler, and B. Muller-Hill. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *Journal of Molecular Biology*, 257(1):21–29, 1996.

153

[51] M. Nakajima, K. Imai, H. Ito, T. Nishiwaki, Y. Murayama, H. Iwasaki, T. Oyama, and T. Kondo. Reconstitution of circadian oscillation of cyanobacterial kaic phosphorylation in vitro. *Science*, 308(5720):414–415, 2005.

[52] A.J. Ninfa and M.R. Atkinson. Pii signal transduction proteins. *Trends in Microbiology*, 8(4):172–179, 2000.

[53] A.J. Ninfa and B. Magasanik. Covalent modification of the glng product, nri, by the glnl product, nrii, regulates the transcription of the glnalg operon in *Escherichia coli. PNAS*, 83(16):5909–5913, 1986.

[54] A.J. Ninfa, L.J. Reitzer, and B. Magasanik. Initiation of transcription at the bacterial *glnAp2* promoter by purified *E. coli* components is facilitated by enhancers. *Cell*, 50(7):1039–1046, 1987.

[55] S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, and B. Mller-Hill. Quality and position of the three lac operators of *Escherichia coli* define efficiency of repression. *EMBO Journal*, 13(14):3348–3355, 1994.

[56] S. Oehler, E.R. Eismann, H. Kramer, and B. Muller-Hill. The three operators of the lac operon cooperate in repression. *EMBO J.*, 9(4):973–979, 1990.

[57] E.M. Ozbudak, M. Thattai, I. Kurtser, A.D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nat Genet*, 31(1):69–73, 2002.

[58] A.A. Pioszak and A.J. Ninfa. Mutations altering the n-terminal receiver domain of nri (ntrc) that prevent dephosphorylation by the nrii-pii complex in *Escherichia coli. J. Bacteriol.*, 186(17):5730–5740, 2004.

[59] C.S. Porter, K.A. North, B.A. Wedel, and S. Kustu. Oligomerization of ntrc at the glna enhancer is required for transcriptional activation. *Genes Dev.*, 7(11):2258–2273, 1993.

[60] L. Reitzer. Nitrogen assimilation and global regulation in *Escherichia coli. Annual Review of Microbiology*, 57(1):155–176, 2003.

[61] L.J. Reitzer and B. Magasanik. Expression of glna in *Escherichia coli* is regulated at tandem promoters. *PNAS*, 82(7):1979–1983, 1985.

[62] A.D. Riggs, H. Suzuki, and S. Bourgeois. lac repressor-operator interaction : I. equilibrium studies. *Journal of Molecular Biology*, 48(1):67–83, 1970.

[63] S. Ringquist, S. Shinedling, D. Barrick, L. Green, J. Binkley, D.G. Stormo, and L. Gold. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Molecular Microbiology*, 6(9):1219–1229, 1992.

[64] J.R. Sadler, H. Sasmor, and J.L. Betz. A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc Natl Acad Sci U S A.*, 80(22):6785–6789, 1983.

[65] J. Sambrook, E. Frisch, and T. Maniatis. *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbour Laboratory Press, New York, 1989.

[66] L. Sbano and M. Kirkilionis. Molecular systems with infinite and finite degrees of freedom, part I: Continuum approximation. *http://arxiv.org/abs/0802.4259*, 2007.

[67] L. Sbano and M. Kirkilionis. An averaging principle for combined interaction graphs. part I: Connectivity and applications to genetic switches. *http://arxiv.org/abs/0803.0635*, 2008.

[68] L. Sbano and M. Kirkilionis. An averaging principle for combined interaction graphs. part II: modularity and perturbation theory. *Preprint*, 2008.

[69] L. Sbano and M. Kirkilionis. Molecular systems with infinite and finite degrees of freedom, part II: Deterministic dynamics and examples. *http://arxiv.org/abs/0802.4279v1*, 2008.

[70] L. Sbano and M. Kirkilionis. Multiscale analysis of reaction networks. *Theory Biosciences*, 127(2):107–123, 2008.

[71] I. Segel. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. Wiley, John & Sons, 1974.

[72] F.W. Sevenich, J. Langowski, V. Weiss, and K. Rippe. Dna binding and oligomerization of ntrc studied by fluorescence anisotropy and fluorescence correlation spectroscopy. *Nucleic Acids Res*, 26(6):1373–1381, 1998.

[73] S.P. Shiau, B.L. Schneider, W. Gu, and L.J. Reitzer. Role of nitrogen regulator i (ntrc), the transcriptional activator of glna in enteric bacteria, in reducing expression of glna during nitrogen-limited growth. *Journal of Bacteriology*, 174(1):179–185, 1992.

[74] J. Shine and L. Dalgarno. The 3?-terminal sequence of *Escherichia coli* 16s ribosomal rna: Complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A.*, 71(4):1342–1346, 1974.

[75] R.G. Smith, S.Y. Halpern, and B. Magasanik. Genetic and metabolic control of enzymes responsible for histidine degradation in salmonella typhimurium. 4-imidazolone-5-propionate amidohydrolase and n-formimino-l-glutamate formiminohydrolase. *J. Biol. Chem.*, 246(10):3320–3329, 1971.

[76] P. Smolen, D.A. Baxter, and J.H. Byrne. Mathematical modeling of gene networks. *Neuron*, 26(3):567–580, 2000.

[77] W. Su, S. Porter, S. Kustu, and H. Echols. Dna-looping and enhancer activity: Association between dna-bound ntrc activator and rna polymerase at the bacterial glna promoter. *Proceedings of the National Academy of Sciences of the United States of America*, 87(14):5504–5508, 1990.

[78] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *PNAS*, 98(15):8614–8619, 2001.

[79] R. Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3):563–585, 1973.

[80] R. Thomas. Role of feedback circuits: positive feedback circuits are a necessary condition for positive real eigenvalues of the jacobian matrix. *Berichte der Bunsengesellschaft fuer Physikalische Chemie*, 98(9):1148–1151, 1994.

[81] R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks–i. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bulletin of Mathematical Biology*, 57(2):247–276, 1995.

[82] J.J. Tyston, A. Csikasz-Nagy, and B. Novak. The dynamics of cell cycle regulation. *BioEssays*, 24(12):1095–1109, 2002.

[83] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry.* Elsevier B.V., Amsterdam, 1992.

[84] J.M. Vilar, Y.H. Kueh, B. Naama, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992, 2002.

[85] J.M. Vilar and S. Leibler. Dna looping and physical constraints on transcription regulation. *J Mol Biol*, 331(5):981–989, 2003.

[86] W. Wang and B.A. Malcolm. Two-stage pcr protocol allowing introduction of multiple mutations, deletions and insertions using quikchange site-directed mutagenesis. *Biotechniques*, 26(4):680–682, 1999.

[87] V. Weiss, F. Claverie-Martin, and B. Magasanik. Phosphorylation of nitrogen regulator i of *Escherichia coli* induces strong cooperative binding to dna essential for activation of transcription. *Proceedings of the National Academy of Sciences*, 89(11):5088–5092, 1992.

[88] H.V. Westerhoff and B.O. Palsson. The evolution of molecular biology into systems biology. *Nat Biotech*, 22(10):1249–1252, 2004.

[89] C. Wyman, I. Rombel, A.K. North, C. Bustamante, and S. Kustu. Unusual oligomerization required for activity of ntrc, a bacterial enhancer-binding protein. *Science*, 275(5306):1658–1661, 1997.

[90] M.W. Young and S.A. Kay. Time zones: a comparative genetics of circadian clocks. *Nat Rev Genet*, 2(9):702–715, 2001.

[91] M. Zou and S.D. Conzen. A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21(1):71–79, 2005.